Universität Rostock

Traditio et Innovatio

Dissertation
zur Erlangung des akademischen Grades
Doktor-Ingenieur (Dr.-Ing.)

Customized Workflow Development and Omics Data Integration Concepts in Systems Medicine

Promotionsgebiet Systembiologie und Bioinformatik

Fakultät für Informatik und Elektrotechnik
Universität Rostock

Markus Wolfien

Geboren am 15. Februar 1989 in Haldensleben
Matrikelnummer: 212206712
Datum der Abgabe: 17.07.2020
Datum der Verteidigung: 30.11.2020

Gutachter

Prof. Olaf Wolkenhauer	Universität Rostock
Prof. David Ellinghaus	Christian-Albrechts-Universität zu Kiel
Jun.-Prof. Stefan Simm	Universitätsmedizin Greifswald

https://doi.org/10.18453/rosdok_id00002868

"Do you have a favourite saying?" asked the boy.

"Yes" said the mole

"What is it?"

"If at first you don't succeed, have some cake."

"I see, does that work?"

"Every time."

The boy, the mole, the fox and the horse
Charlie Mackesy

Acknowledgements

Thanks to ...

... the Department of Systems Biology and Bioinformatics (SBI), which was for me the best place I can imagine to start a scientific career, find the best colleagues, and do fascinating research projects. From early on, my advisor Olaf Wolkenhauer gave me many opportunities to get engaged with publications, proposals, and the freedom to do multiple things at the same time and develop own ideas. Please keep it this way, it enabled me to grow as a person and researcher.

... Dagmar Waltemath and Martin Scharm, who are the two persons why I initially joined the SBI as a student researcher. I had a great time in the SEMS project around SBML and SEDML. Maybe I would have never come across the joy of computational projects without this position. I can still remember the job interview with Martin and hope you do not regret to hire me back than.

... Ulf Schmitz, for guiding and introducing me into the brave world of RNA bioinformatics, and agreeing on supervising my Masters thesis, which was a great start into my PhD work.

... the whole SBI crew: Ali Salehzadeh-Yazdi, for being yourself. You made me a more critically thinking researcher about the science we are doing. I enjoyed the philosophical discussions about networks, phenotypes, and Nietzsche. Holger Henning, for showing me the shiny world of AI, ML, and DL. Andrea Bagnacani, what would be the numerous trainings and trips to Freiburg without you. Saptarshi Bej, Faiz Khan, Shailendra Gupta, Sherry Freiesleben, and Suchi Smita for many fascinating conversations around the coffee machine. The inhabitants of the "*Hochleistungslabor*", including Martin Scharm (yet again), Tom Theile, Martin Peters, Tom Gebhardt, and Julia Scheel. You have been wonderful office mates and I am glad to be your beer representative.

... all the student researchers at the SBI, it was always a pleasure to actually learn from you. Thanks to Mariam Nassar, Tobias Plewka, Christian Krienke, Matti Hoch, Muhammad Naveez, David Brauer, and Maximilian Hillemanns. Maybe on coincidence,

many of you became my long-term colleagues afterwards. Thank you, Amy Sheeran for the excellent language proofreading, I never knew there were so many rules for using commas.

... the RTC, my second, hidden working group, especially Paula Müller, Frauke Hausburg, Anna Skorska, Sophie Kussauer, Julia Jung, Ulrike Müller-Ruch, Madeleine Bartsch, Ralf Gäbel, Christian Rimmbach, Praveen Vasudevan, and Heiko Lemcke for continuously providing me with new fascinating data around the heart. Robert David and Cajetan Lang, thanks to both of you for numerous scientific meetings around climbing spots. I want to keep that spirit. Gustav Steinhoff for setting up and involving me in the PERFECT Phase III clinical trial and keeping the persistance of taking the project further for already a long duration.

... Virginia Bolowski, Peggy Sterling, and Mandy Klingbeil for organizing the things that often cannot be organized.

... my project partners in iRhythmics, de.STAIR, GB-XMAP, and the RBC: Wolfgang Hess, Steffen Lott, Steve Hoffmann, Konstantin Riege, David Ellinghaus, Florian Uellendahl-Werth, Sören Mucha, Björn Grüning, Berenice Batut, Rolf Backofen, and of course the whole Galaxy community.

... the FBN in Dummerstorf, namely Anne-Marie Galow, Andreas Höflich, Ronald Brunner, and Tom Goldammer, for providing high-throughput data in a high-throughput manner.

... my international collaborators, Arash Yavari from Oxford University, and Takajuki Asahara and Amankeldi Salybekov from Tokai University (Japan) for your outstanding experimental methods.

... my friends, and family. Life was continuously changing throughout the thesis. I became husband and a proud father. Thank you for showing me every day what matters most. Whatever will be next, as long as my path involves who I have in my life right now, I am happy to take it.

The dissertation was written as part of the research project "*iRhythmics*". The funding of the project is obtained from the European Social Fund (ESF) as part of the qualification program "*Promoting young researchers in an excellent research network*", an Excellence Research Program of the State of Mecklenburg-Western Pomerania (ESF/14-BM-A55-0027/18).

Abstract

Background: The ever-increasing amount and diversity of biological and medical data is a major challenge in computational analyses. Today, to investigate such heterogeneous data, computational methods have to be combined into comprehensive analysis workflows for seamless, swift, and transparent computation or downstream processing. As an approach to using and integrating mathematical as well as computational models, systems medicine has shown potential to investigate complex data from the molecular level to the level of the organism. Therefore, my interdisciplinary research project utilizes state-of-the-art computational approaches to improve the analysis of heterogeneous data and support medical translational research. It focuses specifically on cardiovascular diseases and cardiac regeneration.

Methods: Workflow development, defined as the successive use of software applications, has emerged as one of the most essential concepts for transparent, reusable data analysis. In the quickly growing landscape of high-throughput sequencing analysis tools, it is increasingly important to retrieve results that are as comprehensible as possible to ensure high-quality, reproducible research. In this work, numerous workflows have been developed for the general processing of bulk RNA sequencing (RNA-Seq), single-cell sequencing experiments, and non-coding RNA identification. All workflows are open access and publicly available via Galaxy and FairdomHub. For example, TRAPLINE includes in addition to basic quality control, genomic alignment, quantification, normalization, and differential expression procedures, also advanced capabilities for micro-RNA target prediction, protein-protein interaction identification, and gene set enrichment analyses. Further applications of RNA velocity calculations and downstream network approaches (e.g., weighted gene co-expression, gene ontology term, and pathway enrichment analyses) have been shown to be vital for the computational validation and monitoring of previously identified differentially expressed transcripts. In addition to the development and maintenance of workflows a contribution to publicly available online training materials for sequencing data analysis was made to further facilitate scientific education. Mathematical concepts of machine

learning and univariate meta-analyses have been successfully implemented to independently investigate the role of cell therapies in cardiac regeneration.

Results: To enable deeper insights into the transcriptomic changes in cardiac regeneration, an analysis workflow for RNA-Seq data was initially developed, tested, critically evaluated, and validated through the comparison of stem cell-derived cardiomyocyte subtypes. The induced sinoatrial bodies, which belong to the category of *in vitro* generated cardiac pacemaker cells after external *Tbx3* administration and *Myh6*-based antibiotic selection, yield the most promising transcriptomic similarity in comparison to a human sinoatrial node. Furthermore, the RNA-Seq counts were used to independently validate the hub-gene 2 AMPK via weighted gene co-expression analysis revealing it to be essential in heart rate generation. In contrast, network and gene set enrichment analyses of alternative mRNA-based strategies for the cardiac reprogramming of human mesenchymal stromal cells into cardiomyocytes showed only an incomplete reprogramming without beating cells. Single-nuclei RNA-Seq was applied to obtain the first characterization of an entire adult mammalian heart and provided realistic cell-type distributions combined with RNA velocity kinetics. To address the efficacy of cell-based therapies for the treatment of ischemic myocardial damage in mice, univariate meta-analysis and the subsequent meta regression analysis showed a positive effect in cardiac regeneration. The significant moderators of this observed effect that were identified are the cell origin (allogen, syngen, xenogen), amount of the injected cells, and gender. The preclinical results are in line with the findings made in the clinical trial data in which the applied machine learning model could identify a novel biomarker signature. These biomarkers are used for the preoperative selection of responsive patients, who receive a therapeutic application of purified allogenic cells after myocardial infarction.

Impact: Throughout this PhD-project eight workshops on RNA-Seq data analysis were taught to more than one hundred PhD students, postdocs, and professors as part of the German Network for Bioinformatics Infrastructure (de.NBI) training courses. In total, two bachelor's and three master's theses were supervised. Moreover, this thesis served as the basis for two international patents, contributed to twenty peer-reviewed manuscripts, generated computational resources with more than two thousand downloads, and obtained findings that resulted in three successfully funded project proposals (de.STAIR, GB-XMAP, iRhythmics) with a total funding volume of more than three million Euros.

Zusammenfassung

Hintergrund: Die ständig wachsende Menge und Vielfalt biologischer und medizinischer Daten erhöht die Komplexität der nachfolgenden Computeranalysen erheblich. Daher müssen diese Berechnungsmethoden zu umfassenden Analyse "*Workflows*" kombiniert werden, um eine nahtlose, schnelle und transparente Berechnung oder Weiterverarbeitung zu gewährleisten. Die Systemmedizin als Ansatz zur Verwendung und Integration von mathematischen und computergestützten Modellen hat bereits ein großes Potenzial zur Untersuchung komplexer Daten von molekularen bis zu organismischen Ebenen bewiesen. Daher nutzt diese interdisziplinäre Arbeit modernste integrative Ansätze, um die heterogene Datenanalyse ebenfalls zu verbessern und die translationale Forschung von Herz-Kreislauf-Erkrankungen und Herzregeneration zu unterstützen.

Methoden: Die Workflowentwicklung, hier definiert als die sukzessive Verwendung von Softwareanwendungen, erwies sich als eines der wichtigsten Konzepte für eine transparente und wiederverwendbare Datenanalyse. Insbesondere in der stark wachsenden Landschaft von Analysewerkzeugen für die RNA Sequenzierung (RNA-Seq) wird es immer wichtiger, die jeweiligen Ergebnisse so verständlich wie möglich abzurufen, um eine qualitativ hochwertige und reproduzierbare Forschung sicherzustellen. In dieser vorliegenden Arbeit wurden zahlreiche Workflows für die allgemeine Verarbeitung der RNA-Seq bzw. Einzelzellsequenzierungsexperimente (scRNA-Seq) und die nichtkodierende RNA Identifizierung entwickelt. Alle Workflows werden öffentlich über Galaxy und FairdomHub bereitgestellt. Zum Beispiel umfasst TRAPLINE neben grundlegenden Qualitätskontroll-, Normalisierungs- und differentiellen Expressionstestverfahren auch erweiterte Funktionen für die Vorhersage von mikro RNA Zielgenen, die Identifizierung von Protein-Protein Interaktionen und Anreicherungsanalysen von Genen. Weitere Anwendungen von RNA Geschwindigkeitsberechnungen, nachgeschalteten Netzwerkansätzen wie gewichteter Genkoexpression und Signalwegsanreicherungsanalysen haben sich als entscheidend für die Validierung und Evaluation der zuvor identifizierten differentiell exprimierten Transkripte erwiesen. Neben der Entwicklung und Pflege von Workflows wurde ein Beitrag zu öffentlich verfügbarem online Schulungs-

material für die Sequenzdatenanalyse geleistet, um die wissenschaftliche Ausbildung zu verbessern. Konzepte des maschinellen Lernens und univariate Metaanalysen wurden zusätzlich implementiert, um die Rolle von Zelltherapien bei der Herzregeneration zu untersuchen.

Resultate: Um tiefere Einblicke in die transkriptomischen Veränderungen der Herzregeneration zu ermöglichen, wurde ein Analyseablauf für die RNA-Seq entwickelt, kritisch bewertet und durch einen biologischen Vergleich von stammzellabgeleiteten Kardiomyozytensubtypen validiert. Die induzierten Sinusknoten, die zur Kategorie der *in vitro* generierten Herzschrittmacherzellen nach externer *Tbx3* Verabreichung und *Myh6* basierter Antibiotikaselektion gehören, ergaben die erfolgversprechendste transkriptomische Ähnlichkeit im Vergleich zu einem menschlichen Sinusknoten. Darüber hinaus wurde die RNA Expression verwendet, um das Hubgen 2 AMPK unabhängig über eine gewichtete Koexpressionsanalyse zu validieren, was letztendlich bestätigt durch weitere Experimente darauf hinweist, dass es für die Herzfrequenzerzeugung ein wesentlicher Faktor ist. Im Gegensatz dazu zeigten Genanreicherungsanalysen alternativer mRNA basierter Strategien zur kardialen Reprogrammierung menschlicher Stromazellen in Kardiomyozyten nur eine unvollständige Reprogrammierung ohne schlagende Schrittmacherzellen. Weiterhin wurde die scRNA-Seq angewendet, um die erste Charakterisierung eines erwachsenen Säugetierherzens zu erhalten und realistische Zelltypverteilungen in Kombination mit RNA Geschwindigkeitskinetiken bereitzustellen. Um eine generelle Antwort auf die Wirksamkeit zellbasierter Therapien zur Behandlung von Myokardschäden bei Mäusen zu geben, zeigte die durchgeführte univariate Metaanalyse und die anschließende Metaregressionsanalyse einen positiven Effekt bei der Herzregeneration. Diese präklinischen Ergebnisse konnten mit den Ergebnissen der klinischer Studie PERFECT validiert werden, in dem das angewandte Modell des maschinellen Lernens eine Biomarkersignatur identifizieren konnte. Diese Biomarker werden zur präoperativen Auswahl von reaktiven Patienten verwendet, die nach einem Myokardinfarkt eine Behandlung mit allogenen Zellen erhalten.

Auswirkung: Im Rahmen dieser Arbeit wurden acht Trainings innerhalb des Deutschen Netzwerks für Bioinformatik Infrastruktur (de.NBI) für mehr als einhundert Doktoranden und Postdocs zur RNA-Seq Datenanalyse abgehalten. Insgesamt wurden zwei Bachelor- und drei Masterarbeiten betreut. Darüber hinaus diente diese Arbeit als Grundlage für zwei internationale Patente, zwanzig von Experten begutachteten Manuskripten und erzeugte Datenressourcen mit bereits mehr als zweitausend Downloads. Die Ergebnisse führten zu drei erfolgreich finanzierten Projektvorschlägen (de.STAIR, GB-XMAP, iRhythmics) mit einem Gesamtfinanzierungsvolumen von mehr als drei Millionen Euro.

Theses

- Workflow development facilitates the reusability of computational data analysis procedures (Section 2.1)

- Galaxy is a sustainable data analysis framework for genomic and transcriptomic investigations in the cardiac research field and beyond (Section 2.1)

- Single-nuclei RNA sequencing analyses can uncover in-depth information about cell type compositions and RNA kinetics in adult mammalian hearts (Section 2.1)

- Signaling network analysis can support the evaluation of reprogrammed cardiac subtypes (Section 2.2)

- Co-expression analyses of RNA sequencing data can validate hub-genes responsible for heart rate influence (Section 2.2)

- The potential of cell therapies for cardiac regeneration can be investigated using preclinical studies (Section 2.3)

- Patient stratification for stem-cell therapy after myocardial infarction is possible through an integrative dataset from human peripheral blood samples (Section 2.3)

Contents

1 Computational workflow development in systems medicine

Systems medicine is an interdisciplinary approach in which clinicians closely cooperate with experts from biology, biostatistics, computer science, engineering, and mathematics to develop novel methods for patient data analysis that may ultimately lead to an improved diagnosis, treatment, and therapy. This chapter introduces systems medicine in general and computational workflows in particular to present state-of-the-art approaches that can be combined for clinical data processing.

1.1 Motivation of this work

Current computational and experimental technologies for the generation, investigation, and analysis of medical and research data already contribute to clinical and translational research, as well as healthcare in general. The intention of this work is to facilitate the ease of use for such novel, often highly technical approaches within the clinical setting of cardiac regeneration.

This work arose from collaborations with clinical partners from the Rostock University Medical Center and the Reference and Translation Center for Cardiac Stem Cell Therapy (RTC) in Rostock, as well as pre-clinical partners from the University of Göttingen, University of Oxford (United Kingdom), Tokai University (Japan), and computational partners at the University of Freiburg and the Friedrich-Loeffler-Institute in Jena.

Despite significant progress, cardiovascular diseases (CVD) such as heart insufficiency, cardiac arrhythmia, coronary heart disease, and cardiac infarction continue to be by far the leading cause of death in Germany, accounting for 37.2%.[1] Significantly more patients had to be treated because of heart diseases in hospitals than several years ago; in 2017 there were more than 1.71 million hospitalizations. Mecklenburg-Western-Pommerania is among the regions with the highest mortality rate triggered by CVD. The nearly-unchanged mortality rate after CVD strongly indicates the need for new therapeutic approaches in which novel treatments such as autologous stem cells in combination with computational assistance, are considered to have a high potential in cardiac regeneration (Steinhoff et al., 2017a).

Today, clinical care is increasingly centered around concepts such as personalized medicine, precision medicine, P4 medicine (predictive, preventive, personalized, and participatory), and systems medicine, which are different names to illustrate the common global effort to establish a more patient-related, accurate, and systematic approach in medicine (Apweiler et al., 2018; Hood et al., 2012; Trachana et al., 2018; Wolkenhauer, 2014). These concepts have the common aim of improving diagnoses, obtaining targeted therapies, finding individualized prognostic markers, supporting early detection, and ultimately preventing diseases. The routes by which these goals should be achieved are similar, and one of their shared core elements is the integration of data from different sources, including

[1] https://www.herzstiftung.de/pressemappe-herzbericht-2018.html

conventional patient data, clinicopathological parameters, molecular and genetic data, as well as data generated by additional omics technologies (Apweiler et al., 2018).

To what extents can a *systems medicine* approach improve the clinical landscape and, possibly, analyze such complex data in an improved manner? First, we need to clarify the term *system* itself; it essentially means that something refers to a set of inter- and intra-related objects. Such objects or so-called features can be defined as closely connected, patient-derived parameters (e.g., blood sample values, medications, genetic background, transcriptomic profiles) in a *medicine*-related context. Therefore, a systems medicine approach identifies parameters, objects, or features within a system or larger set of objects and features that are most important, significant, or relevant for unveiling a given phenotype. In addition to the growing amount of patient data, there is already ample preclinical data of CVD available that have been obtained from *in vitro* cell studies or *in vivo* studies mainly conducted in mice (Hausburg et al., 2017; Leopold et al., 2020). This vast amount of data, commonly obtained from patients and corresponding pre-clinical studies, is too complex and heterogeneous for human beings to comprehensively interpret without technological support (Topol, 2019). In addition, it is impossible for a single clinician to keep track of the broad spectrum of newly published data and discoveries to reliably recall and utilize that information at all time (Dilsizian and Siegel, 2014). Therefore, the current translational systems medicine approaches, including diverse algorithms, tools, and data resources, are the focus of the advances in implementing automatized biomedical and healthcare analytics in clinical settings and will play a significant role in clinical decision-making and the implementation of data-driven medicine (Leopold et al., 2020; Shameer et al., 2017). Here, *data integration* refers to analyzing and combining the different multilayered information (e.g., genomics, transcriptomics, and proteomics) by means of functional networks, artificial intelligence, and further systems medicine approaches. Taken together, these diverse data types, which are derived from preclinical and clinical studies that have been generated for CVD, must be integrated into and interpreted at the cellular and organismic levels (Currie and Delles, 2018; Steinhoff et al., 2017a).

For this purpose, suitable approaches must be developed, integrated, and validated in order to be able to accurately recapitulate the behavior of an individualized, patient-derived system (Sutton et al., 2020). Each part within the approach also must be supported by computational, mathematical, and statistical tools. Without low entrance barriers and easy-to-use experimental protocols, the challenge of proper, transparent, and reproducible data analyses will remain a bottleneck (Lott et al., 2017; Spjuth et al., 2016; Sutton et al., 2020). With respect to the number of steps in the analysis, the complexity of decisions

regarding tool selection has also increased; hence, there has been a call for a systematic method to chain computational tools, which is commonly referred to as *workflow* or pipeline development (Lampa et al., 2013). By facilitating workflows, this thesis also contributes to their development, education, and application (Batut et al., 2018; Wolfien et al., 2016, 2019; Section 2.1). Such reusable computational workflows can be seen as an analogy of the strict laboratory protocols in biochemistry that ensure the transparency and reproducibility of the experimental methods (Peng, 2011).

Bioinformatics and systems biology workflows are widely applied in research investigations and have shown promising results in various research fields to facilitate the investigation of biological phenomena (Akat et al., 2014; Yang et al., 2014). In this interdisciplinary PhD project, this knowledge is transferred to systems medicine approaches, in which workflows are utilized to address biological and medical questions, as well as to overcome computational obstacles.

1.1.1 Medical perspective and the necessity of workflows in cardiac research

Why do we have to combine research areas within the field of CVD, i.e., cardiac regeneration? A primer on heart physiology is provided to critically evaluate the complex molecular and clinical phenotypes of these biological phenomena. Systems medicine approaches are introduced to present the enhanced biological insights of routine experimental and clinical work.

In agreement with the German CVD statistics and the "*Heart Disease and Stroke Statistics*" published annually by the American Heart Association, CVDs cause more deaths each year than all forms of cancer and chronic lower respiratory disease combined (Benjamin et al., 2018). Among CVDs, coronary heart disease is the leading cause (43.8%), followed by cardiac infarction (16.8%), heart failure (9.0%), high blood pressure (9.4%), and other CVDs (17.9%). Heart failure represents the final common phenotype and results from a diverse range of inherited and acquired cardiac insults; it affects around 26 million individuals worldwide. Individuals with severe heart failure have a dismal prognosis with a worse five-year adjusted mortality than many cancers. To date, allogeneic heart transplantation remains the only available treatment option for patients with end-stage heart failure who are symptomatic, despite optimal medical and device-related therapy for cardiac resynchronization (Kobashigawa, 2017). In spite of advances in surgical

techniques, perioperative management, and immunomodulation, a major limitation to its wider application is donor organ scarcity: in Europe in 2017, only 548 donor organs were successfully engrafted, while 1,141 patients are on the active Eurotransplant waiting list, and only 861 are newly registered.[2] Twenty percent of the patients died after three years, before they could undergo heart transplantation. Even for those who received transplants, positive, long-term outcomes are limited by complications (in association with immunosuppression, including malignancy, infection, renal advisor dysfunction, and allograft vasculopathy; Peyster et al., 2018).

In view of these limitations, highly innovative approaches are currently under exploration with the ultimate goal of establishing a safe, durable cellular replacement and repair to injured or diseased myocardium, in addition to *in vitro* disease modelling and drug development applications (Jain and Bansal, 2015; Simkin and Seifert, 2018). A key requirement of these regenerative approaches of the heart is to ensure a highly reliable, robust generation of fully functional cardiomyocytes with physiological properties as close as possible to their natural counterparts, as we have shown in recent publications (Hausburg et al., 2017; Müller et al., 2020; Section 2.2). Current progress in understanding the biology of stem cell pluripotency and endogenous repair mechanisms has fostered a deeper understanding of its remarkable therapeutic potential for tissue repair or replacement (Broughton et al., 2018). Such novel stem cell-based approaches are urgently required to effectively treat the growing burden of disorders characterized by irreversibly damaged or diseased tissue, resulting in the loss of organ or tissue function that is associated with a rapidly ageing population (Hausburg et al., 2017). Furthermore, through the production of induced pluripotent stem cells (iPSCs) from autologous cells (cells from an individual that can be transferred back after a modification), regenerative strategies hold promise in providing truly patient-specific therapies for structural and functional repair in cardiac diseases (Broughton et al., 2018; Steinhoff et al., 2017a).Why, though, must numerous experimental procedures, clinical data, and computational methods be combined to solve this complex puzzle of putative comorbid disease and regeneration mechanisms arising around the heart (Fig. 1.1; Section 2.3)?

To answer this question, we must first know what is particular about the heart that has led it to persist as one of the most fascinating and fundamental human organs. In literature, it justifies the source of poetry; in philosophy, it manifests the origin of the soul; and in medicine, starting in early embryonic development, it is the engine of our lives. In actuality, the heart can be considered as a vital muscular organ in higher vertebrates

[2] https://www.eurotransplant.org/patients/

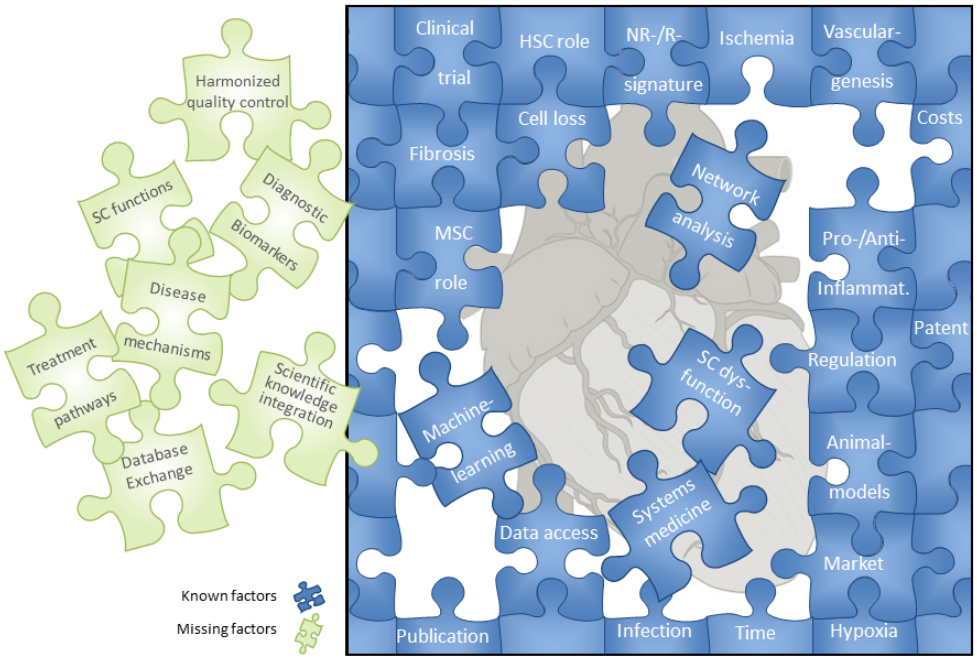
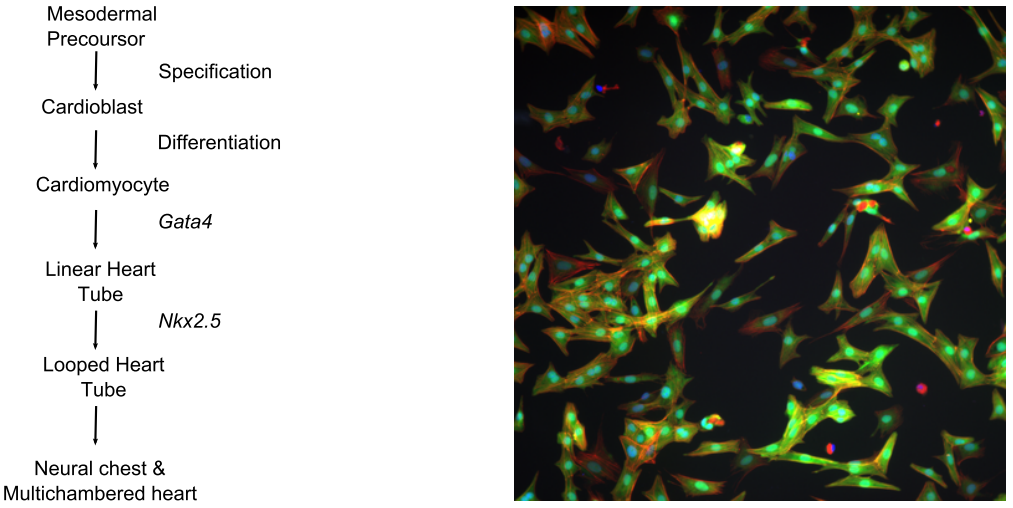


Figure 1.1: Increasing data and analysis heterogenity in clinical studies. To be able to solve the complete picture of a disease, there has to be a combination of numerous clinical datasets, computational analyses, and cross-species data integration (Steinhoff et al., 2017a).

that pumps blood within vessels of the circulatory system through the body. The blood provides the body with oxygen and nutrients and removes metabolic waste. In early developmental stages, the heart develops from embryonic mesodermal precursor cells and after specification, they differentiate into mesothelium, endothelium, and myocardium (Moorman and Christoffels, 2003; Fig. 1.2a). The scheme illustrates that heart tube formation depends on the cardiac transcription factors *Gata4* and *Nkx2.5* and, finally, leads to a functional heart. The average heart rate for a human is about 72 beats per minute. In contrast, a mouse heart beats up to 670 times per minute.

The myocardium itself consists of interlacing bundles of muscle cells, *i.e.*, cardiomyocytes (CM), which comprise the cardiac muscle. CM are presented in Fig. 1.2b[3] by showing a confocal microscopy image with actinin (red), myosin (green), and nuclei (blue) staining. The cells are striated with narrow dark and light bands due to the parallel arrangement of actin and myosin filaments that extend from end to end of each myocyte. In comparison, CM are narrower and much shorter than skeletal muscle cells; they are approximately 0.02 mm wide and 0.1 mm long (Andre et al., 2018).

[3] http://biofrontiers.colorado.edu



(a) Development of the heart. **(b)** Confocal microscopy of cardiomyocytes.

Figure 1.2: Characteristics of cardiomyocytes. (a) Scheme of the evolutionary heart development including necessary transcription factors *Gata4, Nkx2.5* (Olson and Srivastava, 1996). **(b)** Confocal microscopy of cardiomyocytes. Staining: red - actin (Texas Red), green - myosin (GFP), and blue - nucleus (DAPI).

Cardiovascular diseases affect the heart rate

The highly complex phenotype of the mammalian heart, given its four chambers, requires the generation of specific muscle and non-muscle cell types, including CM of the left and right atria as well as the left and right ventricles, a conduction system, pacemaker, vascular smooth muscles, and endo- and epicardial cells (Meilhac et al., 2014; Sahara et al., 2015). Cardiovascular diseases include a broad range of disorders in addition to heart and vessel dysfunction. Therefore, the accompanying problems have serious effects on patients' quality of life. These patients have to deal with limited physical capacity and have a lifelong dependency on medication or technical aids. When untreated, CVDs often lead to highly serious complications and end-stage heart failure.

In the following, two crucial transcription factors (TF) are considered to be those that determine cardiovascular fate: the TF MesP1 (mesoderm posterior 1; Chiapparo et al., 2016) and the surface molecule VEGFR2 (vascular growth factor receptor 2; He et al., 2016). Further development is achieved from multipotent cardiac progenitor cells (Musunuru et al., 2010) and can be divided into two main origins: i) the first (primary) heart field, demarcating a $Nkx2.5^{+}$ / $Hcn4^{+}$ cell population, which forms the cardiac crescent (Später et al., 2013), and ii) the second heart field, demarcating a $Nkx2.5^{+}$ / $Isl1^{+}$ cell population derived from the pharyngeal mesoderm and lying medially and posteriorly to the primary heart field (George et al., 2015). Moreover, the decisive role of the tertiary heart field during

avian pacemaker development of the sino-atrial (SA) node has been reported (Bressan et al., 2013).

The automaticity of the heart-beat is crucial for life: the regular contraction of the heart and the resulting continuous blood flow throughout the body are induced by electrical impulses, which highly specialized cells within the heart initiate and conduct. These cells represent the so-called cardiac conduction system (Kennedy et al., 2016). The SA node is the dominant pacemaker in the human heart and was originally described in 1907 as a sub-epicardial structure located at the right atrium and superior *vena cava* (Silverman and Hollman, 2007). In particular, the SA node integrates the activity of pacemaker cells in a compact region of the right atrium with only a few thousand cells in humans (Fig. 1.3; Boyett et al., 2000). Functionally, these cells depolarize and produce action potentials almost synchronously, beginning within the SA node and spreading towards the atrioventricular (AV) node and the Hisbundle. It has been estimated that only about 1% of the cells in the SA node act as the leading pacemaker (Boyett et al., 2003). The SA cells differ from the working myocardial cells in their content of ion channels and gap junction proteins (Bartos et al., 2015). In particular, they are rich in hyperpolarization-activated cyclic nucleotide-gated cation channel 4 (Hcn4) and t-type calcium channel (Cav3.1). The main gap junction protein in the SA node is connexin45 (Cx45), which plays an essential role in regulating cardiac conduction velocity (Davis et al., 1995). The voltage clock (cyclic activation and deactivation of membrane ion channels, caused by the "*funny current*" Hcn4) and the Ca^{2+} clock (rhythmic spontaneous sarcoplasmic reticulum Ca^{2+} release) function synergistically to generate SA-node automaticity (Joung et al., 2009). The upstroke of the action potential (AP) results from Ca^{2+}-channels as opposed to voltage-gated Na^{+}-channels, which are typical for working myocardial cells (Bartos et al., 2015).

In this respect, it is assumed that the complexity of the human cardiovascular system implies the occurrence of a diverse number of cell types and their intact physiological interplay. Single-cell analyses of entire mammalian hearts from adult mice might serve as the basis of a first insight into such a complex tissue (Wolfien et al., 2020a; Section 2.1.6). This in turn may lead to better explanations of previously identified causes and the influencing parameters for the diseases of individual patients (Steinhoff et al., 2017b). Obviously, the limited regenerative potential of the human heart reveals the decisive issue of recovery, taking into account the extremely low turnover rate of 1% at the age of 25, which decreases to 0.45% at the age of 75 (Traister et al., 2018). Accordingly, less than 50% of human CM are replaced during an average lifespan of 75 years (Hausburg et al., 2017).

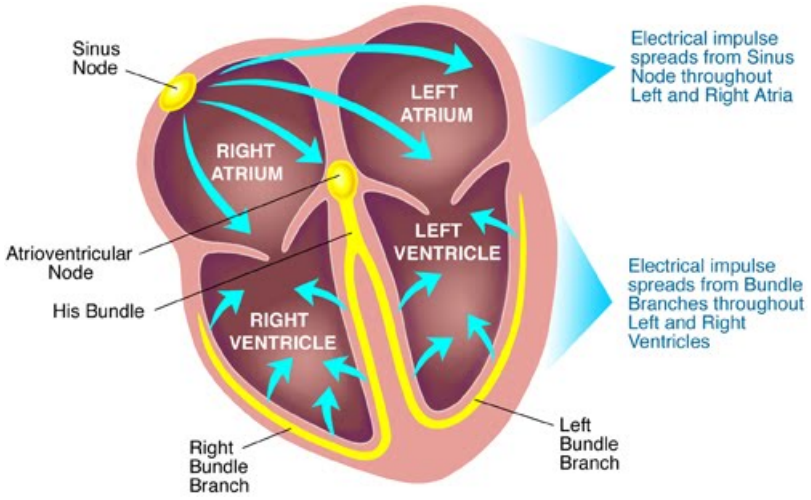


Figure 1.3: Schematic illustration of heart-beat generation. The electrical impulse is generated in the SA node, spreads towards the AV node, reaches the ventricles via branches and, finally, leads to heart contraction (Martini et al., 2020).

Interestingly, the percentage of CM situated in mitosis and cytokinesis is at its maximum in infants, which suggests the significant regenerative potential of the heart in children and adolescents (Mollova et al., 2013). Additional studies have shown that upon transition from mononucleate to the mature binucleate state, CM lose their proliferative potential during a short postnatal period (Foglia and Poss, 2016; Paradis et al., 2014). Thus, the negligible regeneration potential of the myocardium causes an unfeasible functional repair and deleterious remodeling of the affected tissue post-myocardial infarction. However, resident cell populations, such as cardiac progenitor cells or preexisting CM, may offer conceivable sources for myocardial repair after an injury (Bersell et al., 2009; Galow et al., 2020).

Currently, many research groups worldwide aim for preventive methods, therapies, or aftercare operations of cardiac arrhythmia with three primary different approaches:

Medication: Medications are helpful in selected cases but are not as effective as permanent pacing (Tracy et al., 2012).

Artificial (electrical) pacemaker: Controlling symptomatic sick sinus syndrome (the inability of the SA node to generate a heart rate that meets an individual's physiological needs) usually involves the implantation of an artificial pacemaker, a small, battery-operated device that supports the electrical conduction of the heart, resulting in a regular rhythm driven by stimulating electrical impulses (Dobrzynski et al., 2007). The pacemaker is

implanted under the skin through a small incision and is connected directly to the heart through wires that are implanted at the same time. The electric impulses are conducted via leads to the heart and are timed to generate regular intervals, just like natural heart impulses. While the electrical pacemaker can relieve symptoms, offer a better quality of life, and improve survival in certain cases (John and Kumar, 2016), it also has some clear disadvantages, such as the hazard of infections, limited battery lifespan, tearing leads, and interference with electromagnetic devices.

Biological pacemakers: Another approach this thesis also investigates relies on the transplantation of *in vitro*-generated SA-like cells derived from pluripotent stem cells (PSCs) (Kehat et al., 2002). In general, there are two principal strategies: the first is virus-based gene transfer, which aims to convert the resident cells of the heart into cells with pacemaker properties (Raghunathan et al., 2020). The second are cell-based strategies, in which *in vitro* pre-processed cells are transplanted into the heart as pacemakers (Jung et al., 2014; Rimmbach et al., 2015; Wolfien et al., 2016). Notably, when generated from patient-derived iPSCs, such cells may become essential for personalized *in vitro* drug testing.

Systems medicine approaches in cardiac stem cell research

As the 2013 Nobel Prize laureate for Chemistry showed, modelling and simulation became standard techniques in the life sciences to support research about biological, chemical, and, more recently, medical investigations. The research field of systems biology and, subsequently, systems medicine, consists of an iterative cycle of data-driven modelling, resulting in model-driven experimentation (Wolkenhauer, 2014). A possible motivation for researchers and clinicians could be enhancing their specific, deep molecular field of view and transferring their data to the macroscopic area, *i.e.*, to the aimed tissue level, related phenotypes, and, ultimately, diseases. Nevertheless, the clinical impact of the cardiac models is moderate, unless they achieve a high level of detail, as with the virtual heart or the Cancer, Heart, and Soft Tissue Environment initiative (CHaSTE). However, instructive examples of the early, successful integration of computational approaches to answer biological questions include publications such as those of Dowell *et al.* (2014), Gutschner *et al.* (2013), and, more recently, our own contributions (Steinhoff et al., 2017a,b; Wolfien et al., 2019, 2020c; Yavari et al., 2017). These publications strongly support the benefit and effectiveness of computational and experimental collaborative work by using systems medicine approaches for network prediction and lncRNA analyses at the preclinical and clinical scales.

Although the field of systems medicine is a new, evolving extension of systems biology with regards to interdisciplinary, broad research projects and collaborations, it has already provided results that have an impact on clinical practice (Apweiler et al., 2018; Topol, 2019; Wolkenhauer et al., 2013). Therefore, cardiac research should also draw on the benefits of knowledge about the computational integration, analysis, and statistical interpretation of multiscale data. Moreover, medical use cases or more complex, real-world clinical designs can help computational scientists test and improve their existing approaches, which could ultimately result in well-suited, interdisciplinary implementations for new diagnosis, prognosis, and therapeutic methods (Section 2.3).

Due to advances in genomics, transcriptomics, and proteomics technologies, hereafter referred to as "*omics*", life science research now approaches a more detailed molecular level at the single nucleotide resolution. On its own, each of these technologies has contributed numerous advances in the field. However, each individual technology cannot capture the full biological complexity across all the layers of most human diseases; therefore, further integration of multiple levels with a combined approach is necessary to provide a more comprehensive view of the underlying biological phenomenon (Karczewski and Snyder, 2018). The analysis of complex datasets from numerous different sources (such as omics) is inevitable in the life sciences and will also become more important in the daily clinical routine (Lott et al., 2017; Steinhoff et al., 2017a; Triantafyllidis and Tsanas, 2019).

Taken together, computational approaches are already established in many branches at different omics scales and hold great promise for more extensive use if they are jointly integrated. Building upon these ideas, this thesis mainly utilizes transcriptomics data derived from RNA-sequencing (RNA-Seq) as well as microarray experiments, both of which are not currently clinical routine measurements, but they may become inevitable in the future.

Transcriptomics and the influences of miRNAs and lncRNAs

Transcriptomics refers to the study of expressed genes, the active part of the DNA, including their structures, functionality, and all considerable subclasses of currently known RNA classes. Such so-called RNA transcripts can be messenger RNA (mRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), and several non-coding RNAs that are either produced in a single cell or a population of cells (Fig. 1.4). The following two non-coding RNA types are especially emphasized according to their regulatory relevance.

microRNAs (miRNAs): First discovered in 1993, these are small, non-coding RNA molecules (containing about 22 nucleotides), and they can be found in plants, animals, bacteria, and some viruses (Tjaden et al., 2006). Their main functions are RNA silencing and the post-transcriptional regulation of gene expression (Sontheimer and Carthew, 2005), meaning that they can regulate mRNA transcripts. miRNAs are normally transcribed by the RNA polymerase II, which usually binds to a promoter site near the DNA sequence and encodes what will become an 80 nucleotide RNA stem-loop. The resulting transcript is capped with a specific nucleotide at the 5' end and is polyadenylated with multiple adenosines, forming a poly(A) tail. Therefore, these immature miRNAs are part of a several hundred nucleotide-long miRNA precursor called primary miRNA (pri-miRNA) (Cai et al., 2004). This poly(A) tail is one of the requirements for most of the standard RNA-Seq procedures. In contrast to pri-miRNA, mature miRNA are based on their length and missing poly(A) tail, which is less detectable with RNA-Seq approaches (Liu et al., 2019). The measured influence of miRNAs is therefore only based on the approximation that every pri-miRNA also becomes a mature miRNA due to post-processing within the cell.

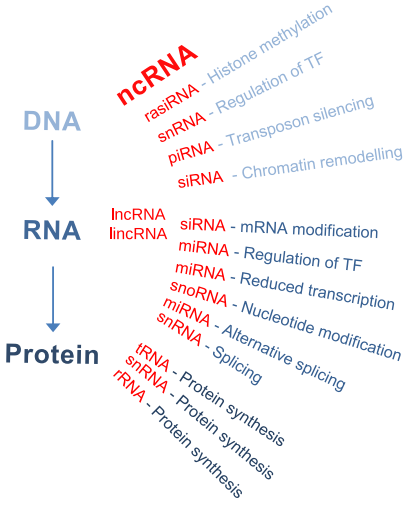


Figure 1.4: Illustration to show the complexity and versatile role of ncRNA subtypes. Visualization of currently known ncRNA subtypes (Wolfien et al., 2019).

long non-coding RNAs (lncRNAs): Large-scale sequencing efforts of cDNA in mammalian cells have identified widespread mRNAs of long transcripts that appear to fall outside the classes of housekeeping or short RNAs, such as miRNAs or ncRNAs (Carninci et al., 2005). From these initial transcript maps, lncRNA species whose loci lie within and between protein coding genes have been identified. While lncRNAs remain the most enigmatic ncRNA species in terms of function, much effort is now focused on their functional characterization and their molecular mechanisms in different cell types (Ilott and Ponting,

2013; Wolfien et al., 2019). Beginning with a variety of screens and expression analyses, it becomes increasingly evident that changes in the expression levels of many lncRNAs are correlated with developmental processes and disease states, but the majority of lncRNAs await further verification (Kung et al., 2013). The number of validated mammalian representatives is very limited; only around 150 lncRNAs in mice and 200 in human are currently annotated in the lncRNA-database (Amaral et al., 2011). Generally, lncRNAs can be detected through NGS approaches and separated into lncRNAs with or without poly(A) tails. It is assumed that more than 80 % of lncRNAs have poly(A) tails (Yang et al., 2011).

To utilize the full potential of omics technologies, integrative methods that combine data from multiple technologies have emerged as critical statistical and computational approaches (Wolfien et al., 2019). The key challenge in developing such approaches is the identification of effective yet representative mathematical models or combinations to provide a comprehensive systems view that most accurately reflects the underlying biology. An ideal approach can answer a biological or medical question, identify important features, and predict outcomes by harnessing heterogeneous data across several dimensions of biological variation (Zitnik et al., 2019). For this reason, this PhD research project uses and integrates multiple computational approaches for the analyses of complex, heterogeneous data types and these may serve as a blue print for later studies (Section 2.3.1).

1.1.2 Computational obstacles in workflow development and data integration

Analyzing sparse, incomplete, and heterogeneous data in a reproducible, transparent manner is a highly demanding and challenging task. It is important to investigate whether either a single workflow within a current data analysis framework (e.g., Galaxy) or the combination of different data analysis workflows across computing platforms can provide a proper, meaningful analysis of multidimensional preclinical and clinical data.

With the rapid growth of health-related data, including genomic, proteomic, imaging, and further clinical meta information, the arduous task of data integration can be overwhelming because of the complexity regarding the so-called 4 Vs of data: variety, veracity, velocity, and volume (Frey, 2018). In comparison to single-omics interrogations, multi-omics investigations provide an even higher level of complexity, but researchers can likewise gain a greater understanding of the flow of information between the different regulatory layers;

there are investigations from the original cause of the disease (genetic, environmental, or developmental) to the functional consequences or relevant interactions possible (Civelek and Lusis, 2014). Multi-omics studies, by their nature, rely on large numbers of comparisons, tailored statistical analyses, and a considerable investment of time, skilled manpower, and money (Hasin et al., 2017).

Why has data analysis complexity risen exponentially? For example, if we consider a single data analysis step, it is clear that there are multitudes of combinations necessary to choose the most suitable type of analysis algorithm, the actual tool to obtain proper results, and optimized parameters to use (Fig. 1.5). Therefore, the time and manpower needed in a multi-omics investigation for tool comparisons, benchmarks, and implementations quickly increases by magnitudes. In addition to the thorough, transparent process of tool selection and parameter optimization, the computational reproducibility of the results must be guaranteed, which is the ability to exactly reproduce results given the same data. In a previous survey, 90% of researchers acknowledged a "*reproducibility crisis*" (Baker, 2016), meaning that numerous computational protocols used for research are not readily reproducible because not all of the steps and parameters involved in the analysis are scripted in a machine-readable format or are mentioned at all. Often, results can be reproduced only with help from a study's authors, which requires a substantial time investment or is impossible (Beaulieu-Jones and Greene, 2017).

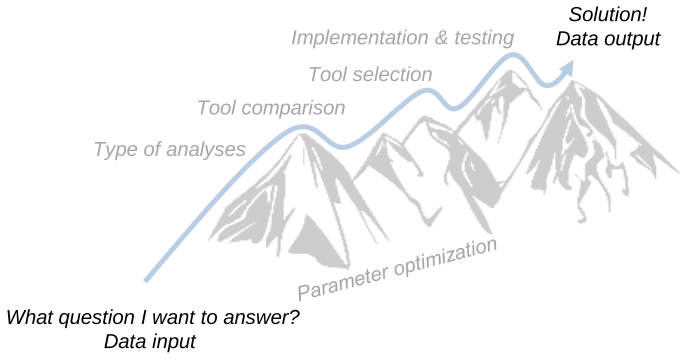


Figure 1.5: Path of a usual data analysis strategy for a single task. Starting from a specific research question and dataset with numerous possibilties of tool selections to obtain the final data output (Lott et al., 2017).

One approach to taming data size and diversity is data integration that collectively and concurrently utilizes available datasets to create a joint model, which commonly uses a mathematical concept (often depicted as a network) as a starting point to represent different investigational layers (Gligorijević and Pržulj, 2015). Gligorijevic *et al.* (2015) have also

described the increasing abundance of biological and medical data as one important aspect of the increasing popularity of data integration approaches in the past decade, and they have noted that understanding cellular process and molecular interactions by integrating molecular networks is just one of the challenges ahead. Some examples of these heterogeneous data types include next generation sequencing (NGS) data (O'Leary et al., 2016); mutation data from The Cancer Genome Atlas (TCGA; Weinstein et al., 2013); functional annotations through gene set enrichment analysis (GSEA); ontologies such as the gene ontology (GO; Ashburner et al., 2000) and disease ontology (DO; Bello et al., 2018); pathway data such as WikiPathways (Slenter et al., 2018) or BioCarta (Nishimura, 2001); as well as further databases for miRNA target prediction (e.g., miRanda)[4] or protein-protein interaction databases (e.g., BioGrid).[5] These databases add additional value to molecular networks, and they must be incorporated into any data integration framework to increase the reliability of newly discovered scientific knowledge that cannot be gained from any single dataset alone. However, it is important to be aware that many databases are not manually curated and might introduce certain errors in the networks.

To fulfill this overarching goal of data integration, suitable approaches must overcome many computational challenges, such as sparse or missing data with heterogeneous formats and multiple dimensionalities that increase the complexity, noisiness, and mutual concordance between the samples under investigation. In addition, a suitable solution strategy that is supported by means of machine learning (ML) algorithms should be designed to address complex tasks and data types (Triantafyllidis and Tsanas, 2019). Nevertheless, these challenges are even greater when working with heterogeneous clinical patient data instead of highly standardized preclinical studies or cell culture experiments because patients have different ages, medications, comorbidities, and genetic backgrounds. Although more and more community efforts and data-driven challenges (e.g., Dream and Kaggle challenges) for the integration of multiple data types have been initiated, such vastly collaborative efforts cannot be considered for every research study (Scharm et al., 2014; Waltemath et al., 2016; A Modelers tale).[6] This thesis seeks a balance of suitable and easily applicable approaches that can be used for the small- and large-scale analysis of preclinical and clinical data.

[4] https://omictools.com/miranda-tool
[5] https://thebiogrid.org/
[6] https://figshare.com/articles/A_Modeler_s_Tale/3423371

1.2 Combining computational methods to integrate heterogeneous data types

The underlying technical and algorithm-specific concepts are introduced and critically evaluated in this subchapter to provide the reader a deeper understanding of the importance and challenges of using these methods.

Accessing and retrieving high-quality datasets is the first great challenge to address in a clinical and experimentally driven research field like the investigation of CVD (Steinhoff et al., 2017a). An analysis of high-throughput data, together with patient phenotype information, can only lead to the identification of robust sets of candidate genes, proteins, and pathways if one uses controlled and manually curated high-quality data (Müller-Ruch et al., 2020). Data quality at such an advanced level of investigation is key because, without high-quality biological data, almost every subsequent computational analysis and its hypothesis generation are obsolete (Uellendahl-Werth et al., 2020). In light of these aspects, the data under investigation should be obtained under *good practice* (GxP; Del Mazo-Barbara et al., 2016), but of course GxP can neither assure the usage of appropriate or scientifically relevant methods nor guarantee the scientific significance of analyses or examinations. However, in order to build a high-level association of the underlying processes involved in the disease pathology, it is necessary to integrate various classes of heterogeneous information and to explore the complex relationships between entities such as the diseases themselves, candidate genes, proteins, interactions, and pathways (Hausburg et al., 2017; Lysenko et al., 2016). This is referred to as a "*connective workflow*" development in Section 2.1.4, in which the use and combination of multiple workflows for data analysis is presented.

1.2.1 Sequencing data analysis for complex, molecular investigations

Today, the molecular characterization and identification of transcriptomic profiles can be easily realized via RNA-Seq.

The rapid development of sequencing technologies made it possible for biomedical disciplines to outrival the physical sciences in terms of data-generation capability. The combined

output of today's genomics studies has already surpassed the data acquisition rate of entire scientific domains, such as astronomy, or internet platforms, such as YouTube or Twitter (Stephens et al., 2015). This is reflected by the numerous applications (e.g., single-nucleotide polymorphisms, single-cell analysis, epigenetic modifications, copy number variants, differential expression, and alternative splicing) in research and diagnostics, paired with a high-quality control at the single-base level that is also the principal motivation for an intensified use of NGS in recent years (Bahassi and Stambrook, 2014).

First, why should one consider NGS technology to acquire deeper molecular insights? In comparison to other high-throughput methods such as microarray, NGS technologies enable genome-wide investigations of various phenomena for new, highly sensitive investigations of either genomic or transcriptomic regions to explore the current status of a single cell or tissue. The technical advances of NGS methods reveal a revolutionary tool for deeper insights into transcriptomic analyses (Wang et al., 2009).

In particular, these developments include computational improvements in transcription start site mapping, strand-specific measurements, gene fusion detection, small RNA characterization, and the detection of alternative splicing events (Bloom et al., 2009; Sultan et al., 2008). One major application with an enhanced effectiveness over previous technologies is the RNA-Seq technology (Cloonan et al., 2009). RNA-Seq uses NGS devices to sequence, map, and quantify a population of transcripts (Mortazavi et al., 2008). The range of RNA-related sequencing experiments and subsequent data analyses has steadily increased, and researchers thus must weigh and decide on specific technologies and combinations of experiments, which can easily become more and more complex (Adiconis et al., 2013; Kukurba and Montgomery, 2015; Podnar et al., 2014; Wolfien et al., 2019). Several laboratories have provided evidence that sequencing library preparation and RNA-Seq datasets themselves are technically easily reproducible, while offering a broad, dynamic detection range, which makes this platform more reliable in the detection of transcripts with low abundance (Ozsolak et al., 2009; Uellendahl-Werth et al., 2020). Ongoing developments promise even further advances in the application of RNA-Seq towards approaches that allow RNA quantification from very small amounts of cellular materials on a single-cell level (Chaudhry et al., 2019; Galow et al., 2020; Schaum et al., 2018; Wolfien et al., 2020a,b). Nevertheless, there are many different NGS approaches, and each has advantages and disadvantages (Liu et al., 2012a). In this thesis, the different Illumina Genome Analyzers were used and, for further explanation, briefly compared with other sequencing platforms (Tab. 1.1). The Illumina sequencer achieves parallelization by the so-called bridge amplification of DNA fragments method. These fragments are immobilized onto

the flow cell of the instrument at a concentration that promotes a dense array of non-overlapping fragment colonies. Each fragment colony is then sequenced according to a single base at a time by the cyclical addition of fluorescent-labeled nucleotides that are conjugated with a reversible terminator (Bentley et al., 2008). Our example experimental and computational workflow, starting from sample preparation to data analysis, is shown in Fig. 1.6 and may serve as a starting point for RNA-related analyses to characterize novel RNAs, especially for newly identified ones in clinics (see also Section 2.1.4).

Another important aspect to consider is that a bulk RNA-Seq experiment (sequencing from heterogeneous cell type mixtures, e.g., tissues) can only reflect the current state of the mRNA and may not necessarily give specific hints of a joint mRNA-miRNA regulation in future states. A set of experiments would be needed to acquire a more robust time-resolute analysis. However, on the single-cell RNA-Seq level, novel approaches, such as the RNA velocity analysis, could at least partially overcome this need for multiple time points (La Manno et al., 2018). Here, the ratio between spliced and unspliced RNA is used to calculate the direction and speed for each cell in an approximated future time step (4h to 6h, depending on the cell type investigated). In the underlying thesis, this analysis was also applied and evaluated on single-cell data for an entire murine heart (Section 2.1.6).

Platform	**Method**	**Read length (bp)**	**Throughput**	**Reads**	**Runtime**
SOLiD 5500xl	Seq. by ligation	2 x 60	95 GB	800 M	6 d
Illumina HiSeq2500	Seq. by synthesis	2 x 125	1 TB	4 B	6 d
Illumina X-ten	Seq. by synthesis	2 x 150	1.8 TB	6 B	$<$ 3 d
PacBio RSII	Single molecule	15 K	1 GB	55 K	4 h
Oxford Nanopore	Single molecule	200 K	1.5 GB	$>$ 100 K	Up to 48 h

Table 1.1: Comparison of different NGS technologies. Abbreviations are: base pairs (bp), sequencing (Seq.), gigabyte (GB), terabyte (TB), thousand (K), million (M), billion (B), hours (h), days (d; Lu and Zhan, 2018; Quail et al., 2012).

There is an extensive list of examples of NGS success stories (e.g., The Cancer Genome Atlas, 100.000 Genomes Project, Genome10k, Tabula muris consortium), but there are still bottlenecks with this emerging technology: transparent and properly reproducible data analysis strategies (Lott et al., 2017). With respect to the number of data analysis steps, the complexity of decisions regarding tool selection has also increased, thus the call for a systematic workflow development and management frameworks (Lampa et al., 2013).

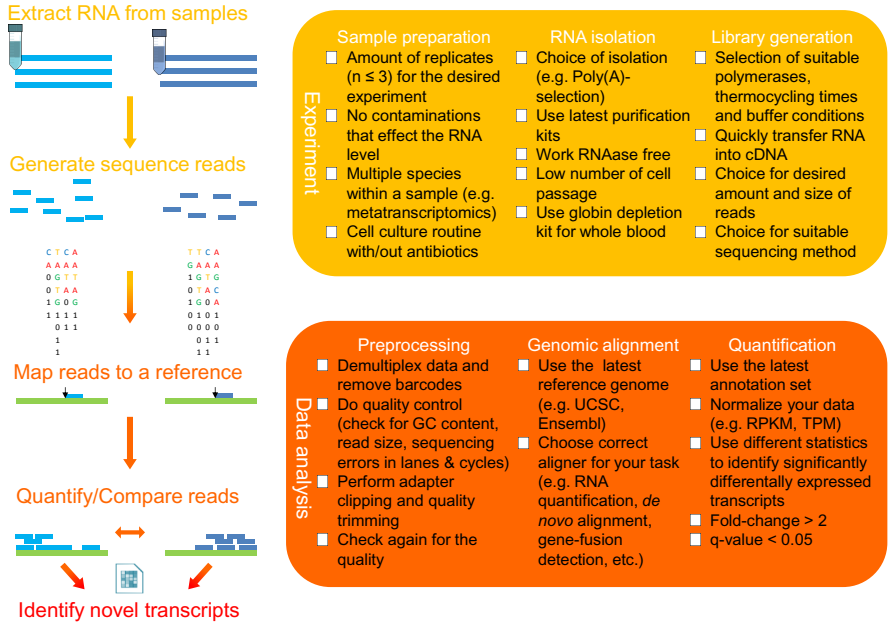


Figure 1.6: Integrated experimental, computational workflow with specific checkboxes for RNA-Seq analysis steps. On the left hand side a common process from RNA extraction to genome alignment, quantification, and DE detection can be seen. The right hand side shows experimental as well as computational hints depicted as checkboxes (Wolfien et al., 2019).

However, the gap of fully standardized and automated methodologies has steadily decreased for the analysis and interpretation of RNA-Seq data (Afgan et al., 2018). In general, before analyzing an RNA-Seq dataset, it has to be taken into consideration that the variety of experimental RNA-Seq protocols, study designs, and the characteristic properties of the organisms under investigation greatly affect downstream and comparative analyses (Conesa et al., 2016). Today, there are broadly accepted conventions about best-practices in RNA-Seq data analysis (e.g., frequently cited tools for data analyses), which is why this thesis also critically compares and evaluates state-of-the-art bioinformatics approaches based on the fact that such computationally intense data processing should be easily accessible as well as modular for an easier exchange of specific tools. Another interesting aspect investigated here is whether such analysis can be conducted on a personal computer with few computational resources (e.g., a limited amount of CPU and RAM) or only at an advanced computing node or servers. For this reason, a workflow is presented that integrates the best-performing data analysis, data evaluation, and annotation methods in a **T**ransparent, **R**eproducible, and **A**utomated **P**ipe**LINE** (**TRAPLINE**) for RNA-Seq

data processing (suitable for Illumina, SOLiD, and Solexa sequencing platforms; Wolfien et al., 2016). This workflow uses a structured pre-selection, classification, and integration of well-suited tools within modularized data analysis approaches that are embedded in ready-to-use computing infrastructures, while using different experimental data as use cases and for validation (Section 2.1.1).

Introduction of the most common workflow management frameworks
de.NBI and ELIXIR are initiatives that support the expansion and further development of accessible workflow frameworks.

Galaxy (Afgan et al., 2016, 2018) and the *Galaxy-RNA-Workbench* (Fallmann et al., 2019; Grüning et al., 2017): The Galaxy project is a framework that makes advanced computational tools accessible without the need for extensive prior training. Galaxy seeks to make data-intensive research more transparent and reproducible by providing a web-based environment in which users can perform computational analyses and have all of the details automatically tracked for later inspection, publication, or reuse. It is applicable for non-computational users on a public server and contains explanatory features for interactive Galaxy tours for beginners and the Galaxy "*Tool Shed*", which contains more than 3,500 tools, for advanced users. Galaxy is free to use (open source) and includes a broad community with over 125 public servers available for various tasks, as well as pre-built Docker/rkt images for local use and a constantly maintained international training network (Batut et al., 2018). New tools need to be *.xml* wrapped to be integrated.

KNIME (de la Garza et al., 2016): The Konstanz Information Miner (KNIME) is a modular environment that enables the visual assembly and interactive execution of data pipelines. It is designed as a teaching, research, and collaboration platform and enables the simple integration of new algorithms and tools, as well as data manipulation or visualization methods as new modules or nodes. KINME contains a grid and user support environment; however, the execution of workflows on high-performance clusters is only available with the commercial version. The workflows are interoperable and can be represented as Petri nets, which enables a hierarchy of workflows, e.g., meta nodes can wrap a sub-workflow into an encapsulated new workflow. The framework also enables "*HiLiting*" (selecting and highlighting several rows in a data table, and the same rows are also highlighted in all other views that show the same data table).

Chipster (Gentleman et al., 2004): Chipster is a user-friendly analysis software for high-throughput data, and it currently contains more than 450 tools. Its intuitive graphic user

interface enables researchers to access a powerful collection of data analysis and integration tools to visualize data interactively. Users can collaborate by sharing analysis sessions and workflows. A desktop application user interface based on *Java* is also available. Chipster has strong support and easily integrates *R*-based tools (e.g., from BioConductor). Another advantage is the freely available and open source client-server system and its numerous visualizations (interactive and static).

Snakemake (Koster and Rahmann, 2012): Snakemake is a workflow engine that provides a readable, *Python*-based workflow definition language and a powerful execution environment that scales from single-core workstations to multi-node compute clusters without modifying the workflow. Snakemake is based on a readable Python workflow definition language and allows efficient resource usage, but it is only available on Linux. A computationally advanced command-line based framework interoperates with any installed tool or available web service, and jobs can be visualized as directed acyclic graphs.

In addition to the computational frameworks presented, Poplawski *et al.* (2015) and Lachmann *et al.* (2020) have completed a systematic search and evaluation of further workflow management frameworks with a focus on RNA-Seq data analysis. As this thesis is the work of an associated partner of the RNA Bioinformatics Center (RBC) in Freiburg, it uses the Galaxy framework for workflow development, because its main European server is hosted and developed at the University of Freiburg.

Technical illustration of cloud computing frameworks

After choosing, implementing, and setting up the data analysis workflow within an appropriate software framework, a reasonable computing environment must be selected. In general, computing environments can be web-based (free of charge community cloud computing), offline, and hybrid solutions (e.g., private and commercial cloud computing). According to the National Institute of Standards and Technology, cloud computing is defined "as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction". Due to the rapidly increasing amount of computational data being created, large consortia, namely public-private partnerships, have been established to share objectives, resources, costs, risks, and responsibilities between academia and industrial partners (e.g., International Cancer Genome Consortium, 100,000 Genomes Project: Granados Moreno et al., 2017). The most frequently used commercial cloud services are Google's and Amazon's Web Services private cloud-computing infrastructures.

Due to concerns about data safety, security, and privacy, cloud computing has a low adoption rate within the healthcare system (Griebel et al., 2015). An emerging solution to deploy computational workflows, including all the necessary tools and dependencies, is software channels and containers such as Bioconda, Docker, rkt, or NextFlow (Ranganathan et al., 2019). These "*software containers*" have emerged as a possible solution for many of the concerns mentioned above, as they allow the packaging of software and workflows in an isolated, self-contained system, which simplifies the distribution and execution of tools in a portable manner across a wide range of computational platforms, such as Galaxy and KNIME (Di Tommaso et al., 2015). The technology combines several areas from systems research, especially operating system virtualization, cross-platform portability, modular reusable elements, versioning, and a "*DevOps*" philosophy (Boettiger, 2015). Wolfien *et al.* (2016), Schulz *et al.* (2016), Gruening *et al.* (2017), and Fallmann *et al.* (2019) have demonstrated the successful implementations of a Galaxy/Docker based workflow with discrete software applications for the analysis of NGS data.

Setting up a Galaxy instance and TRAPLINE

The RNA workbench is an example of a Galaxy instance. It contains a collection of more than 50 RNA-centric data analysis tools.[7] The second example of a utilized Docker container refers to the aforementioned TRAPLINE workflow. The tools used are conserved in this running environment and can be used to apply the workflow. It should be noted that the tool versions used in the TRAPLINE container are not the current ones; rather, they refer to the versions of the published manuscript. The tools of the RNA workbench are continously updated. Both data analysis suites can be installed under OSX and Windows using the graphical tool Kitematic[8], with the following command on Unix systems or directly at the Docker console (grey box).

These Galaxy Docker containers are "*read-only*", which means that all changes, processes, and computed results from a session are lost after restart. This mode is useful to present Galaxy to colleagues or to run workshops with it. However, it can also be configured as a running compute server using external folders for storage as well as compute clusters or cloud environments for processing. Due to the modular system, it is also possible to install all or only a few tools of the RNA workbench or more advanced tools from the "*Tool Shed*" on available Galaxy servers. When using the RNA workbench Docker container, the user has full administration rights, which enables an independent customization of tools as well as potential additional users, including their restrictions. In this thesis, Docker containers

[7] https://galaxyproject.org/use/rna-workbench/
[8] https://kitematic.com

are utilized to share workflows, training material, and complete processing frameworks to enable users to analyze their RNA-Seq datasets on their own desired system.

The following entries show examples of a docker command that can be used on an UNIX or Docker console to run (or first download) a Docker container utilizing a fully-fledged Galaxy instance of the RNA workbench (3.4k downloads, last accessed December 14, 2020) or the publication version of TRAPLINE (1.1k downloads, last accessed December 14, 2020) for RNA-Seq data analysis:

```
docker run -d -p 8080:80 bgruening/galaxy-rna-workbench
docker run -d -p 8080:80 mwolfien/trapline
```

The expression *docker run* starts the container. The argument *-d* will start the docker container in *Daemon mode*, which is a type of long-running program. The argument *-p 8080:80* opens the port 80 (inside of the container) towards the port 8080 on your local host. To bind both of the ports together an Apache web server is running inside the container to cast the output to this local port on your host computer. With this options you can access your Galaxy instance via http://localhost:8080 after executing one of the commands above. The Galaxy Admin User has the username *admin@galaxy.org* and the password *admin*.

Galaxy as well as the RNA workbench are designed as community projects and, thus, new users can easily contribute to both platforms with workflows, new tools, and training material keeping the frameworks up-to-date and valuable for research. Based on the basic Galaxy infrastrcture all components and addons such as tools, workflows, visualizations, interactive tours, atoms, and training material can be easily integrated into any other of the more than 150 available Galaxy instances for teaching, learning, or exploratory purposes.

1.2.2 Network approaches to explaining the alteration of gene patterns

After quantifying the gene expression of samples, network approaches are another central concept in systems medicine because they combine existing knowledge about classical linear pathways, ontologies, and disease phenotypes with in vitro and in vivo experimental data of various organisms to enhance the overall utility of the available data. In particular, a

list of differentially expressed transcripts needs to be further evaluated, and the underlying biological meaning of the transcripts has to be characterized and enriched.

Biological networks can span many levels, such as genes, transcripts, proteins, metabolites, organelles, cells, organs, organisms or even social and ecosystems (Steinhoff et al., 2017a). A biological network (or graph) essentially consists of nodes (or vertices) and edges (or links); nodes usually represent discrete biological entities at molecular (e.g., genes, transcripts, proteins, metabolites, drugs) or phenotypic (e.g., diseases, ontologies, pathways) levels, whereas edges represent physical, functional, or chemical relationships between pairs of entities (Vidal et al., 2011). In general, they appear to exhibit architecture described mathematically as "scale free," in which most nodes have some links but a small fraction of nodes, the so-called "*hubs,*" which are highly interconnected (Lusis and Weiss, 2010). In recent decades, networks have been extensively used as a mathematical framework/tool for computational modeling and analyzing omics data (Aittokallio and Schwikowski, 2006); for this reason, this thesis also utilizes different network analysis approaches for downstream processing (Section 2.2).

In general, functional network analyses can include studies about gene interaction (Kikkawa, 2018; Liu et al., 2012b), gene co-expression (Saha et al., 2017), protein-protein interaction (PPI; Hakes et al., 2008), and metabolic interaction (Scharm et al., 2020) that have already revealed valuable biological insights into different aspects within the cell machinery or organism. However, the comprehensive understanding of a biological system can only be achieved by a common, multi-purpose, and integrative analysis from all these network types; developing such an integrated network representation that can capture all associations by including all molecular details remains one of the major challenges in functional network integration (Luo et al., 2017). Throughout this dissertation, network approaches are utilized to enhance RNA-Seq expression data with further databases for an investigation of the topological structure of over-represented transcripts as well as GSEA performed via Enrichr (Chen et al., 2013) or ClueGO (Bindea et al., 2009; Section 2.2.1 and Section 2.2.2).

Gene co-expression analyses

The alteration of gene co-expression patterns in biological samples is described for numerous conditions (e.g., diseases, treatments) and has been proposed as one mechanism for rewiring and extending transcription regulatory networks (Farahbod and Pavlidis, 2019). In particular, if they do not share common primary or significantly differentially expressed transcripts, it can be assumed that such co-expression analysis approaches are poised

to become unavoidable in interpreting gene networks (Menche et al., 2015). Gene co-expression network-based approaches, such as Weighted Gene Co-expression Network Analysis (WGCNA) are one of the most powerful and widely used co-expression analysis approaches. These are commonly used in analyzing microarray and RNA-Seq data, especially for identifying functional modules and hub-like genes (Langfelder and Horvath, 2008). WGCNA or newer implementations, such as Cemitools (Russo et al., 2018), are a so-called guilt-by-association (GBA) approach for constructing co-expression networks that are subsequently used for cluster (module) identification of highly correlated genes (Langfelder and Horvath, 2008). However, there may be major topological differences between RNA-Seq and microarray co-expression in the form of low overlaps between hub-like genes from each network due to changes in the correlation of expression noise within different technologies (Ballouz et al., 2015). This is why it has thus far not been possible to integrate both gene expression technologies with this method, unless some investigations show comparability and transferability between RNA-Seq and microarray data (Wolff et al., 2018). In contrast to protein-protein interactions, transcriptional co-expression is neither a discrete property of a pair of genes nor an actual physical interaction, and it does not give links to clean regulatory relationships (Farahbod and Pavlidis, 2019). The co-expression between transcripts is only derived from the correlation matrix of an entire data set, which means that closely co-expressed pairs of transcripts have a higher likelihood of being part of a larger pattern (here, part of a certain cluster or module) within sets of expressed transcripts. These issues take into consideration that the ability to extract specific (*i.e.*, biochemical or physical) interactions from differential co-expression is limited; nevertheless, this limitation does not necessarily detract from the potential utility of the tissue- or other context-specific data to improve the relevance of functional predictions from co-expression studies (Farahbod and Pavlidis, 2019).

In summary, WGCNA is a popular method for the identification of co-expressed transcripts in which experimentally obtained gene expression levels are used to predict common clusters that most likely share a common biological function. In Section 2.2.3, the applicability and power of interpretation for such transcriptional co-expression studies are investigated via preclinical datasets. WGCNA was used to construct a co-expression network around the induced sinoatrial bodies (iSABs) to investigate potential hub-genes for cardiac pacemaking.

1.2.3 Machine learning and meta-analysis models to assess cardiac outcomes

In terms of identifying meaningful information from large amounts of data, machine learning (ML) has evolved into one of the most widespread and dominant approaches in the life sciences. Machine learning approaches gain experience through an applied learning procedure from several types of input data using mathematical assumptions. Dimensional reduction, the classification of data into distinct groups, and feature selection are several common scenarios of ML approaches that are now used on a daily basis.

The third class of computational approaches utilized in this thesis is statistical models for meta-analyses and ML concepts. Here, the aim is to use and apply these approaches to larger, heterogeneous datasets instead of data from *in vitro* experiments or in-bred mice only. It is also an additional showcase of *connective workflows* for patient outcome prediction.

Why are ML approaches necessary in the clinic? Optimal patient care in clinical routine settings requires a rapid, fact-based decision-making process to retrieve a suitable diagnosis in consideration of the available treatment options. This has been true for centuries; however, the amount of patient data available that can be quickly generated has changed in recent years due to ongoing hospital digitalization and advanced technological breakthroughs, such as improved imaging devices, in-depth blood analysis techniques, and more efficient data analysis approaches. These *in silico* methods under the overarching umbrella of Artificial Intelligence (AI) in particular hold great potential for an improved, accurate medical field (Topol, 2019). In this thesis, AI algorithms were applied on a clinical Phase III trial study because in this trial, common linear statistical methods showed only a limited prediction outcome for the therapy response stratification (Steinhoff et al., 2017b; Section 2.3.3).

In agreement with the observations in this thesis, a recent analysis has shown that real-life digital health interventions (e.g., clinical decision support systems, web-based strategies, mobile apps, or disease monitoring) that incorporate ML are highly useful and effective (Triantafyllidis and Tsanas, 2019). These independent ML algorithms can be seen as an extension of previously established statistical approaches to disease risk assessment, such as the recommendations by the American Heart Association/American College of Cardiology (ACC/AHA) that can predict the prognostic risk of CVD based solely on common risk factors such as cholesterol, age, smoking, and diabetes (Benjamin et al.,

2018). Nonetheless, many patients are not identified by these classical linear prediction models, and some patients are unnecessarily treated (Kwon et al., 2018; Weng et al., 2017). These models may thus oversimplify complex, high-dimensional datasets by using too few parameters, or they may not consider non-linear interactions among the parameters measured. With the rise of highly efficient ML algorithms, alternative approaches beyond classical linear prediction models have been developed, and they have the potential to use more complex, so-called "*Big Data*" for a better prognosis and diagnosis (Obermeyer and Emanuel, 2016).

A primer on AI approaches

In the last decade, we have obsered rapid progress in the development of AI, which is defined as a computational entity's independent learning based only on available given information of any kind. Today, the vast amount of information obtained from patients or pre-clinical studies is too complex and heterogeneous for humans to comprehensively interpret without any technological support (Dilsizian and Siegel, 2014). With the dawn of AI, including the concepts of ML and deep learning (DL), the supportive analysis of high-dimensional patient-specific information enables clinicians to improve their diagnostic, prognostic, and therapeutic decisions.

In particular, an AI algorithm relies on a computer system to learn the provided input data by minimizing the error between predicted and observed outcomes to ultimately unravel the most important and often non-linear interactions between measurements (Dreiseitl and Ohno-Machado, 2002). Such approaches can, for example, significantly improve the accuracy of CVD risk prediction and increase the number of patients identified that could benefit from a preventive treatment or avoid unnecessary treatments (Steinhoff et al., 2017b; Weng et al., 2017).

Numerous AI architectures have been developed and are used to classify, impute, predict, and cluster datasets based on so-called *features*. Such features include relevant patient-specific information, such as medical traits or clinical measurements including blood test parameters, MRI images, or smart watch sensory data; these measuremetns allow a multi-parametric assessment of diseases (Cal-Gonzalez et al., 2018). The main difference between ML and DL approaches is the selection of important features for decisions. In ML, this process is manually performed by a domain expert or specialized feature selection algorithms. A major benefit of classical ML is an improved understanding of the importance of each individual feature in itself because a domain expert manually pre-selected it. However, feature extraction is not trivial and can be highly biased based on

the curator; therefore, the advantage of DL is the absence of manual feature selection. The result often leads to robust models with less bias based on prior knowledge (Wang et al., 2019). In addition, DL explainability algorithms have become more prominent in research to close this gap (Holzinger et al., 2019). In contrast to ML, DL algorithms are based on non-linear, multi-layered networks for automatic feature extraction and classification (Shah et al., 2019). Briefly, a DL algorithm consists of internal nodes that represent learned features that are automatically utilized and weighted due to their importance for a certain decision. In this thesis, only classical ML approaches such as *Random Forest*, *Naïve Bayes*, or *Boosting* have been utilized because the "*Good clinical practice*" (GCP) controlled clinical trial data was already highly curated, and the explainability of the results was essential at all times during the analysis.

The technical implementation of an AI algorithm that utilizes features for a prediction is called a *model*. The underlying mathematical concepts of such a model seek combinations of linear and non-linear decision boundaries or patterns in a given set of information to try to separate individual data points, such as patients or diseases. For example, if a data point represents a patient, the corresponding features are the clinically measured data, and the classification label can refer to the stage of a specific disease, diagnosis, or treatment option. In addition to classification scenarios of patients into disease and healthy states, ML can be used to uncover features that are important for the choice between these states that can subsequently be used to further explain the underlying biological phenomenon, which is called "*feature selection*." In contrast to supervised ML, where the ground-truth of the labels is known, unsupervised ML models need no specific ground-truth to train the actual model. Therefore, these statistical learning analyses assume that there are naturally occurring subclasses within data (e.g., patient cohorts, disease subtypes) that behave differently across a number of populations and across varying scenarios (e.g., varying treatments, ethnologies, environments). Thus, an ML study usually emphasizes finding an intrinsic structure within patient phenotypic data, which can then be evaluated retrospectively and prospectively to predict treatment outcomes and guide clinical trial design (Harrer et al., 2019). By applying non-linear approaches, such as t-distributed stochastic neighbor embedding (t-SNE; van der Maaten and Hinton, 2008) or Uniform Manifold Approximation and Projection (UMAP; McInnes et al., 2018) for dimensional reduction, distinct groups can be discovered and independently used to justify a specific hypothesis (Steinhoff et al., 2017b). Unsupervised ML has become an invaluable asset to test and evaluate novel classification hypotheses of a disease or clinical syndrome, and it should be a mandatory analysis for a robust, independent validation strategy. Nevertheless, currently, ML drives little in health care (Rajkomar et al., 2019), which is why this

dissertation fosters state-of-the-art applications for a transfer from proof-of-concept models towards bedside applications in cardiology (Section 2.3).

The pitfalls of AI in medicine

Despite the promise of AI, there are several challenges that complicate it successful application in studies in clinical routine scenarios. The common denominator in most AI applications is high-quality data, which is commonly known as the "*garbage-in and garbage-out*" principle (Steinhoff et al., 2017a). It is therefore important to generate data according to standardized guidelines and understand its origin, especially when the data is generated from multiple study sites. The points raised below address the most common pitfalls.

Not yet standardized data management procedures: GCP and "*Good laboratory practice*" (GLP) guidelines define how standardized clinical processes and high-level medical research should be conducted to achieve high-quality, trustworthy, and reusable data (Müller-Ruch et al., 2020; Steinhoff et al., 2017a; Verma, 2013). Nevertheless, it is difficult to determine how and what extent research groups follow such guidelines. This phenomenon also applies to many AI-related medical studies because it is either challenging or impossible for other research groups to repeat those studies with their own datasets. This aspect clearly generates the need for an environment that allows the management and sharing of de-identified generated heterogeneous datasets and computational models in the context of the actual experiments. A framework for publishing findable, accessible, interoperable, and reusable (FAIR) data, operating procedures, and models for the medical community has to be widely established and fully enable researchers and doctors to organize, share, and publish data, models, and protocols for the enhanced reproducibility and reusability of research results (Wolstencroft et al., 2017). Current data management guidelines applied in this thesis involve sample acquisition from informed study patients who gave their written consent according to the Declaration of Helsinki (approval by the Ethical committee, Rostock University Medical Center 2009; No. HV-2009-0012). Analyses and examinations were performed before the unblinding of the PERFECT trial[9] and under careful adherence to the protection of data privacy (pseudonyms). Preclinical data of murine experiments were stored in the publicly available Sequence Read Archive (SRA; Accession Number SRS1064711)[10] and Arrayexpress (Accession Number E-MTAB-8751, E-MTAB-8848).[11]

[9] https://clinicaltrials.gov/ct2/show/NCT00950274
[10] https://www.ncbi.nlm.nih.gov/sra
[11] https://www.ebi.ac.uk/arrayexpress/

Processed and published data is also accessible through the FairdomHub project website of the iRhythmics project.[12]

Data access and generation: Novel AI models require large, independent amounts of retrospective data for validation and evaluation, possibly generated under the same comparable conditions. Nevertheless, old in-house data might have been acquired with suboptimal or different instruments, protocols, or employees, or it may have been only partially archived without raw formats (Papp et al., 2018). Even though recent hospital digitalization endeavors are highly welcome given their promise of increased patient care, on their own, they cannot be considered a remedy for AI applications. Without appropriate, applicable standardization processes, the veracity of digitized "*Big Data*" may still degrade predictive AI performance (Papp et al., 2018).

Limited amount of available multi-centric data: Multi-centric data are generally difficult to access due to restrictive hospital policies, and data generation itself is often not under comparable conditions even if the same protocols and devices are used. Why is this the case? First, there is a certain element of reluctancy in data generation, *i.e.*, the time and duration of sample taking or the speed of sample processing until freezing. Second, local hospital rules, bureaucracy, and sharing processes may appear overcomplicated and time-consuming, which may delay successful research that builds on multi-centric collaborations (Papp et al., 2018). Finally, even if the willingness to share is present and the data undergoes local anonymization processes, some data subsets (e.g., sequencing or imaging data) may still reveal certain characteristics of individuals and therefore might not be able to be used in the final analysis. All these factors together appear to challenge the establishment of a comprehensive, multi-centric dataset, which could increase AI-related research (Cal-Gonzalez et al., 2018). The lack of multi-centric data is generally considered one of the major reasons that few AI solutions have thus far been integrated into clinical routine practice (Topol, 2019).

Evaluation and validation of AI models: There is a certain element of bias in the selection of AI methods, which is typically driven by prior expertise and familiarity with AI tools or the popularity of certain AI methods that might not be optimal for a given study. The "*no free lunch theorem*" states that there is no superior AI approach in general; rather, the ideal AI approach is data- and application-specific (Adam et al., 2019). This suggests that one should test multiple AI models (e.g., Random Forest, Naïve Bayes, Boosting) over the available data to understand the underlying characteristics and the method's

[12] https://fairdomhub.org/projects/28

applicability. AI frameworks that have been used in such a comparative strategy include computational packages such as caret[13] or mlbench[14] for the *R* programming language, or the *Python*-based scikit-learn package.[15] Furthermore, different performance metrics such as the "*area under the curve*" (AUC), "*receiver operating characteristics*" (ROC), and the F2-score have been applied. However, these metrics can also make established model performances difficult to compare among numerous research groups because different AI tools tend to utilize different metrics for the training process (Mohseni et al., 2018). The lack of proper cross-validation in single-center studies is one of the major concerns of AI-driven predictive models (Papp et al., 2018). In the underlying work for this study, all ML models were 10-fold cross-validated. Finally, predictive models' lack of interpretability is a general concern for clinicians because an understanding of how actual predictive models operate and predict decisions must always be possible (Rudin, 2019). Validation with an independent yet similarly generated dataset should be considered the gold standard (Chen et al., 2020).

Diagnostic support for clinical decision making

Taken together, the key to accurate AI models is high-quality clinical data generated under GxP, and, if possible stored in a standardized format (e.g., Fast Healthcare Interoperability Resources [FHIR] format) and digitally available as so-called electronic health record (EHR) data. It has already been shown that AI models that combine different sources of data, such as images, patient records, and clinical parameters, perform better than models working on only a single dataset (Haas et al., 2017; Mirza et al., 2019). Such an approach for medically relevant AI models based on multiple data types is widely known as a clinical decision support system (CDSS). A CDSS leverages EHR data to assist clinicians in basic actions, including alerting, reminding, rejecting orders, interpreting, predicting, diagnosing, assisting, and suggesting (Beeler et al., 2014). According to Beeler *et al.*, CDSS software incorporates the generic steps of input, processing, and output: (i) health professionals involved in healthcare enter the patient-specific data, (ii) it is processed and linked to knowledge stored in a database, and (iii) notifications are communicated back to the clinicians. Although several CDSS approaches have already been successfully applied, they may also introduce errors in some cases (Beeler et al., 2014). This occurs, for example, due to a high number of alerts and relatively small sample sizes because CDSSs are no yet fully established in all medical fields or even incorporated into the current clinical routine, which may change in the future (Semenov et al., 2019; Vinks et al., 2020).

[13] http://topepo.github.io/caret/index.html
[14] https://cran.r-project.org/web/packages/mlbench/mlbench.pdf
[15] https://scikit-learn.org/stable/

Meta-analyses as the gold standard of measuring primary clinical values

In contrast to high-dimensional multivariate ML models, there have also been univariate mathematical models in the form of meta regression analyses performed in this thesis. Since these analyses compare measurements across multiple studies by means of objective and quantitative methods, they provide less biased estimates on a specific topic, if the included studies were selected appropriately (Higgins et al., 2019). This data integration approach is called "*horizontal*" data integration because its data type is common in contrast to the formerly integrated "*vertical*" omics data (Xia et al., 2013). Meta-analyses contain a quantitative, formal, and epidemiological study design used to systematically assess previous research studies to derive conclusions about that body of research (Haidich, 2010). Typical outcomes of a meta-analysis include a more precise estimate of the treatment effect (in this study, the change of the left ventricular ejection fraction to assess the cardiac regeneration capacity) or other clinically relevant surrogate measurements (e.g., increase of viability or lifespan). In contrast to a small, individual study the statistical power of a large, systemic study is therefore increased, while likewise decreasing individual- and study-specific biases (Xia et al., 2013). The examination of the variability or heterogeneity within the study outcomes, as well as potential study biases, are an additional value that contributes to the benefits of meta-analyses to approach a consolidated and quantitative review of a larger body of literature (Haidich, 2010). According to Haidich *et al.* (2010), rigorously conducted meta-analyses are useful tools in evidence-based medicine because the need to integrate findings from many studies ensures that meta-analytic research is desirable and the large body of research now generated makes conducting this research feasible. The downside of meta-analyses is similar to that of ML models: the data generation and extraction, which may occur because of high manual curation effort and biased inclusion and exclusion criteria (Desai et al., 2020). One possibility to avoid such a bias in meta-analyses is provided by standardized guidelines from the PRISMA and Cochrane consortia (da Costa, BR and Juni, P, 2014; Higgins et al., 2019).

Another major advantage of a meta-analysis is that it generates a precise estimate of the observed effect size based on different mathematical approaches (in this thesis, random and fixed effects models) with considerably higher statistical power, which might be important, if the power of the primary study was limited (e.g., through a small sample size; Lee, 2018). Furthermore, meta-analyses can reveal the actual source of variation, a bias of potentially missing studies, and different effects among subgroups by investigating specific moderator values across multiple studies. In Section 2.3.2, a meta-analysis is conducted to investigate the regenerative potential of CM subtypes after cardiac infarction, which might be used to clarify the applicability of injected cells as a source for heart regeneration.

1.3 Objectives

The overall goal of this thesis is to support medical translational research by utilizing state-of-the-art computational methods to facilitate the adoption of systems medicine approaches in pre-clinical and clinical settings for cardiac regeneration. Exemplarily, data integration workflows for pre-clinical research data in mice and clinical routine data are developed to highlight the beneficial impact of such an integrative analysis that requires several computational steps, i.e., NGS data processing, single-cell analysis, co-expression, network analysis as well as ML classification and feature selection. In particular, patiens' regenerative response is predicted, and its detailed molecular mechanisms, which influence the heart rate and cardiac repair, are investigated.

The molecular investigation of a disease and its relevance in patients also must be determined, evaluated, and validated in preclinical and clinical studies because, in most cases, only cellular or animal experiments can explain a disease phenotype in humans that might have never been uncovered in the heterogeneous patient data alone (Steinhoff et al., 2017a). Here, a modularized workflow scheme is developed to utilize preclinical and clinical measurements, as well as low- and high-throughput datasets. This connective workflow scheme can serve as a data analysis and integration blue print for upcoming studies that investigate common or individual aspects in pre-clinical and clinical data regarding the same underlying biological phenomenon.

The working hypotheses for the different sections in the dissertation are the following:

- Tailor-made and transparent computational workflows for each data type are necessary to develop the most appropriate and reproducible data analysis strategy.

- Biological mechanisms of the heart rate and cardiac response after myocardial infarction can be uncovered using sequencing and network analyses on transcriptomics data.

- Cell therapies have a positive effect on cardiac repair after a myocardial infarction.

To address these working hypotheses, in Section 2.1, the thesis starts with the development of an easily reproducible computational RNA-Seq data analysis workflow, including a benchmark of different analysis tools to show the necessity of an easily accessible imple-

mentation within a transparent data analysis framework. It contains the development of an RNA-Seq data analysis workflow itself, which consists of the actual comparison of tools for the relevant analysis steps of preprocessing, genomic alignment, quantification, and differential expression (DE) detection. The implementation of the workflow into a specific workflow management framework such as Galaxy, *i.e.*, the RNA-Workbench and interactive Galaxy tours, as well as portable, containerized dissemination concepts such as Docker, are presented.

In Section 2.2, various applications of the previously developed RNA-Seq workflow (e.g., for the quantification of gene expression or the identification/annotation of ncRNAs) are performed and extended to additional data annotation databases to validate their overall use and to investigate the underlying biological concepts. For this reason, the results of TRAPLINE are incorporated into and further evaluated with several network analyses (e.g., topological analysis, GSEA) and co-expression analysis approaches (e.g., weighted gene co-expression analysis [WGCNA]). A characterization of and distinction between different murine cardiac cell types is achieved, and subpopulations can be attributed to different cardiac functionalities.

Section 2.3 provides a detailed investigation of data analysis approaches that have been combined to connect the results of preclinical and clinical data for cardiac regeneration studies. Here, low- and high-throughput datasets, as well as patient-specific meta-data, are used to identify i) molecular biomarkers and moderator values for a given group of patients, ii) commonly and individually regulated signaling pathways in therapy-responding patients from protein and gene expression data, iii) network features that are relevant in the regenerative patient subgroup, and iv) relevant patient-specific features based on ML that are associated with cardiac regeneration. The integration and visualization of clinical trial data complements this section to ultimately show the applicability and usability of the suggested connective workflows to investigate such a complex biological phenomenon.

Each section of this cumulative thesis contains peer-reviewed publications in indexed international journals or books. Each publication is introduced with a brief background description, a statement of my personal contribution to this work, and a summary of the results and the impact. Additional current publications can be retrieved in section titled "*Contributions in Peer-reviewed Publications*" (Section 3.4) and in my ORCID, ResearchGate, and Scopus profiles.

2 Published and peer-reviewed scientific work

2.1 Workflow development for RNA-Seq data analysis

*This section contains the **detailed development** of a transparent and easily accessible RNA-Seq data analysis workflow using the Galaxy platform. The intended usage of the workflow also addresses the wide data processing range between a common personal computer and a more advanced server infrastructure to allow every user a self-driven, autonomous data analysis. The workflow is separated into (i) data preprocessing, (ii) the identification of significantly differentially expressed transcripts that may include newly uncharacterized RNAs, (iii) gene annotation clustering, and (iv) a ready-to-use gene expression interaction network file of important transcripts, including database information of detected miRNAs and lncRNAs. In the following, the experience gained from developing such a workflow was used to give further advice to related data analysis fields (e.g., single-cell experiments) and workflow development in general. In particular, workflows for bulk and single-nuclei RNA-Seq were developed and applied. The concept of connective workflows was introduced to characterize novel ncRNAs. At the end of this section, a community effort is presented that shows the development of a sustainable Galaxy training material for life science data analyses (e.g., RNA-Seq).*

2.1.1 TRAPLINE, an RNA-Seq data analysis workflow in Galaxy

Wolfien, M., Rimmbach, C., Schmitz, U., Jung, J.J., Krebs, S., Steinhoff, G., David, R., and Wolkenhauer, O. (2016).
TRAPLINE: A standardized and automated pipeline for RNA sequencing data analysis, evaluation and annotation.
BMC Bioinformatics. IF: 2.970, Citations (December 14, 2020): 16

Molecular data generated from NGS devices provide means to acquire deeper insights into cellular functions. However, the lack of standardized, automated methodologies poses a challenge for the analysis and interpretation of RNA-Seq data.

In this manuscript, I critically compared, benchmarked, and evaluated state-of-the-art bioinformatics approaches (e.g., tools for preprocessing, genome alignment, transcript quantification, and differentially transcript detection). Subsequently, I developed and implemented a comprehensive data analysis workflow in Galaxy that integrates the best-performing tools for data analysis, data evaluation, and annotation methods in a **T**ransparent, **R**eproducible and **A**utomated **P**ipe**LINE** (**TRAPLINE**). This workflow is suitable for RNA-Seq data generated with Illumina, SOLiD, and Solexa platforms. Comparative transcriptomics analyses with TRAPLINE result in a set of DE genes, their corresponding protein-protein interactions, splice variants, promoter activity, predicted miRNA-target interactions, and files for SNP calling. I selected the BioGRID database for possible protein-protein interactions and the microRNA.org database for miRNA-mRNA interactions because these allowed an easily integration in Galaxy. Here, the value of the proposed workflow is demonstrated by characterizing the transcriptome of in-house stem cell-derived antibiotic-selected cardiac bodies (‘*aCaBs*’). Furthermore, findings made during this work have been patented.

In summary TRAPLINE supports NGS-based research by providing a workflow that requires no bioinformatics skills on the commmand-line, decreases the processing time of the analysis, and works in the cloud, as well as at local systems with minor computational performance. The workflow is implemented in the biomedical-research platform Galaxy and is freely accessible via TRAPLINE[1] or the specific Galaxy manual page.[2] A Docker container with more than 1,000 downloads is available at Docker-Hub.[3]

[1] https://www.sbi.uni-rostock.de/RNAseqTRAPLINE
[2] https://usegalaxy.org/u/mwolfien/p/trapline-manual
[3] https://hub.docker.com/r/mwolfien/trapline

Wolfien et al. BMC Bioinformatics (2016) 17:21
DOI 10.1186/s12859-015-0873-9

BMC Bioinformatics

SOFTWARE Open Access

TRAPLINE: a standardized and automated pipeline for RNA sequencing data analysis, evaluation and annotation

Markus Wolfien[1*], Christian Rimmbach[2], Ulf Schmitz[3,6], Julia Jeannine Jung[2], Stefan Krebs[4], Gustav Steinhoff[2], Robert David[2*] and Olaf Wolkenhauer[1,5]

Abstract

Background: Technical advances in Next Generation Sequencing (NGS) provide a means to acquire deeper insights into cellular functions. The lack of standardized and automated methodologies poses a challenge for the analysis and interpretation of RNA sequencing data. We critically compare and evaluate state-of-the-art bioinformatics approaches and present a workflow that integrates the best performing data analysis, data evaluation and annotation methods in a **T**ransparent, **R**eproducible and **A**utomated **P**ipe**LINE** (TRAPLINE) for RNA sequencing data processing (suitable for Illumina, SOLiD and Solexa).

Results: Comparative transcriptomics analyses with TRAPLINE result in a set of differentially expressed genes, their corresponding protein-protein interactions, splice variants, promoter activity, predicted miRNA-target interactions and files for single nucleotide polymorphism (SNP) calling. The obtained results are combined into a single file for downstream analysis such as network construction. We demonstrate the value of the proposed pipeline by characterizing the transcriptome of our recently described stem cell derived antibiotic selected cardiac bodies ('aCaBs').

Conclusion: TRAPLINE supports NGS-based research by providing a workflow that requires no bioinformatics skills, decreases the processing time of the analysis and works in the cloud. The pipeline is implemented in the biomedical research platform Galaxy and is freely accessible via www.sbi.uni-rostock.de/RNAseqTRAPLINE or the specific Galaxy manual page (https://usegalaxy.org/u/mwolfien/p/trapline—manual).

Keywords: RNA sequencing, NGS data processing, Data evaluation, Bioinformatics workflow, Galaxy, TRAPLINE, Stem cells

Background

In comparison to other high-throughput methods, Next Generation Sequencing (NGS) technologies enable genome-wide investigations of various phenomena, including single-nucleotide polymorphisms, epigenetic events, copy number variants, differential expression, and alternative splicing [1]. RNA sequencing (RNAseq) uses the NGS technology for discovering novel RNA sequences, and quantifying all transcripts in a cell [2, 3]. Like genome tiling arrays, an RNAseq experiment can capture evidence for yet unannotated genes and isoforms. The utility of RNAseq to uncover new transcripts is well documented [3–8]. Several laboratories have provided evidence that cDNA library preparation and RNA sequencing sets are technically well reproducible and in contrast to microarrays RNAseq offers a broader dynamic range, which makes this platform more sensitive in the detection of transcripts with low abundance [9].

The steady increase of publications involving RNAseq experiments generated a need for statistical and computational tools to analyze the data. Basically, all RNAseq

* Correspondence: markus.wolfien@uni-rostock.de; robert.david@med.uni-rostock.de
Markus Wolfien, Christian Rimmbach and Ulf Schmitz are first author
Robert David and Olaf Wolkenhauer are senior author
[1]Department of Systems Biology and Bioinformatics, University of Rostock, 18057 Rostock, Germany
[2]Reference und Translation Center for Cardiac Stem Cell Therapy (RTC), University of Rostock, Rostock 18057, Germany
Full list of author information is available at the end of the article

analyses involve the following tasks: pre-processing, quality control, read mapping and further analyses like differential expression (DE) analysis, single nucleotide polymorphism (SNP) analysis or gene isoform and splicing variant detection. However, the availability of tools following a standardized analysis protocol are limited [10].

A number of software packages and pipelines have already been introduced to deal with these tasks. These software packages are mainly based on programming languages like *C*, *Python* or *R* and require advanced expertise in programming or computer science for proper implementation and use or they do not provide advanced analytical tools like gene network inference methods, miRNA-target predictions and/or the integration of protein-protein interactions [11–16]. Additionally, the possibility of discovering alternatively spliced genes or promoter activity would be desirable. Furthermore, there is no common RNAseq data analysis strategy, despite the obvious need for such a standardized pipeline [17]. The increased dependence on computational approaches in life sciences has revealed grave concerns about the accessibility and reproducibility of the computed results [5]. Galaxy is a free web-based platform for omics research that addresses the following needs [18, 19]:

- *Accessibility*: Galaxy enables users to perform integrative omics analyses by providing a unified, web-based interface for obtaining omics data and applying computational tools to analyze these data. Learning a programming language or the implementation details is not necessary.
- *Reproducibility*: Galaxy produces metadata about every possible analysis step and automatically tracks descriptive information about datasets, tools, and parameter values to ensure reproducibility. User annotations and tagging is possible at each step of the pipeline.
- *Transparency*: Galaxy includes a web based framework for sharing models including datasets, histories, workflows and repositories. It also allows users to communicate and discuss their experimental results in an online forum.

Implementation

Using Galaxy, we developed a comprehensive, Transparent, Reproducible and Automated analysis PipeLINE, named TRAPLINE, for RNAseq data processing (optimized for Illumina FASTQ reads, but also suitable for other sequencing platforms like SOLiD or Solexa), evaluation and prediction. The predictions are based on modules which are able to identify protein-protein interactions, miRNA targets and alternatively splicing variants or promoter enriched sites. A schematic representation of the analysis pipeline is illustrated in Fig. 1. TRAPLINE can be accessed via the published Galaxy page of TRAPLINE (https://usegalaxy.org/u/mwolfien/p/trapline—manual) or via www.sbi.uni-rostock.de/RNAseqTRAPLINE.

TRAPLINE implements the following tools and resources: (i) FASTQ quality trimmer, FASTXclipper and FastQC for pre-processing and quality control, (ii) TopHat2 for read mapping, (iii) Picard Toolkit for read correction and SNP identification, (iv) Cufflinks2/Cuffdiff2 for DE analysis, splicing and promoter testing (v) the Database for Annotation, Visualization and Integrated Discovery (DAVID) for gene annotation and functional classification, (vi) miRanda for miRNA target prediction, (vii) BioGRID for protein-protein interactions and, finally, a compiling module for ready to use network construction files. For detailed instructions regarding the usage of TRAPLINE please see the manual in the Additional file 1.

Results

To show the effectiveness of our automated pipeline, we exemplarily applied TRAPLINE to RNAseq data generated from our recently described antibiotic selected cardiac bodies ("aCaBs"), which are highly pure clusters of mouse embryonic stem cell (mESC) derived cardiomyocytes generated via *Myh6* promoter based antibiotic selection plus a standardized differentiation protocol (Additional file 2: Figure S1) [20, 21]. Their RNA expression profiles were compared to control embryoid bodies (EBs) derived from the same cell line without administration of the antibiotic.

TRAPLINE includes state-of-the-art quality control processes

TRAPLINE analyses RNAseq reads obtained from Illumina, SOLiD and Solexa platforms with the help of "FASTQ Groomer" [22] that converts the specific formats as a first step. In the following pre-processing step, adapter sequences, which have been added to the 5' and 3' ends of the cDNA fragments during the sample preparation phase, are being clipped (no influence towards other platforms). In the Illumina sequencing procedure, sequences are extended on both ends by - 62 nucleotide long adapters that may influence the results of the subsequent analysis [23]. These adapters are only used during the Illumina bridge amplification procedure to immobilize the cDNA transcripts. In TRAPLINE we implemented the tool "FASTXClipper" (http://hannonlab.cshl.edu/fastx_toolkit/index.html) for this purpose.

It is necessary to discriminate sequencing errors from biological variation by using quality scores (Q) [24]. Therefore, in the last pre-processing step uncalled and wrongly called bases are removed (Quality Trimming). Standard approaches rely on the associated quality

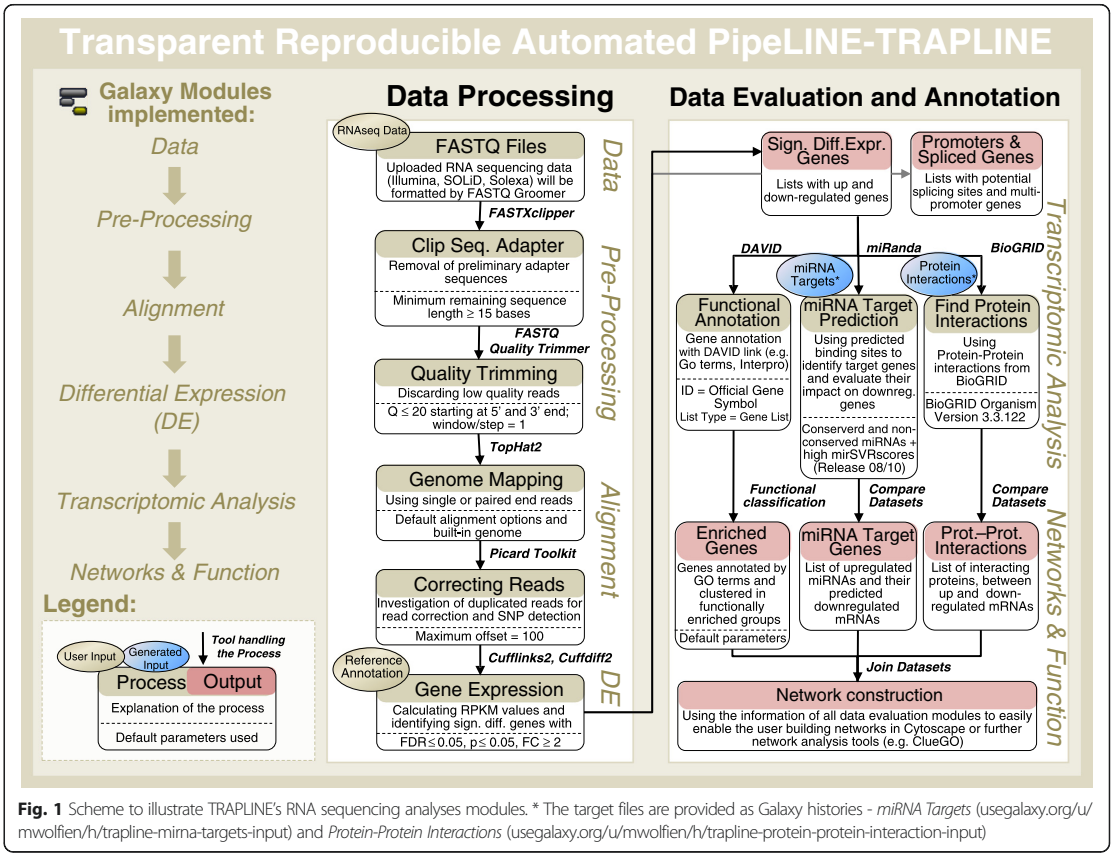


Fig. 1 Scheme to illustrate TRAPLINE's RNA sequencing analyses modules. * The target files are provided as Galaxy histories - *miRNA Targets* (usegalaxy.org/u/mwolfien/h/trapline-mirna-targets-input) and *Protein-Protein Interactions* (usegalaxy.org/u/mwolfien/h/trapline-protein-protein-interaction-input)

scores to retain the read, or a portion of it, if the score is above a predefined threshold [22]. As suggested by Mbandi et al. [24] we excluded reads with a score $Q < 20$ to ensure reliable genome mapping results. For the purpose of discarding low quality reads, we implemented the widely used tool "FASTQ Quality Trimmer" which returns a quality control report for each dataset that was analyzed. The effects achieved by quality trimming our data are shown in Additional file 3: Figure S2.

We compared the fraction of mapped reads with and without applying data pre-processing (Fig. 2a). Our observations confirm the findings of Chen et al. [25], who demonstrated the necessity of applying the described pre-processing steps in a read-mapping benchmark.

TopHat2 - the most accurate alignment tool in Galaxy

To select the most suitable read alignment tool, we analyzed the overall mapped transcript coverage on the genome (accuracy) of the most commonly used alignment tools, which are based on the exon first approach. Figure 2b shows the results of our comparison between BWA [26], Bowtie [27], Bowtie2 [28], and TopHat2 [29], and their average accuracy in the mapping of six different datasets. The overall alignment accuracy of the mapped reads to the reference genome is between 70 % and 85 %. Bowtie2 and TopHat2, that share a similar algorithm, produce a significantly higher accuracy in comparison to the BWA and Bowtie alignment tools (based on a significance level $\alpha = 0.05$). In our case, the Bowtie2 alignment algorithm was able to map in average 2.5 million more reads to the genome than the BWA/Bowtie algorithm (total amount reads: 24–26 million). Our observations are consistent with the results of Kim et al. [30], who found that TopHat2 generates more accurate alignments than competing tools, using fewer computational resources. Because of the significantly superior mapping accuracy of TopHat2, in contrast to Bowtie/BWA, and the additional functionality to find splice junctions and promoter regions, we decided to include TopHat2 into TRAPLINE. The outputs of TopHat2 are BAM files which contain the aligned reads to the reference genome and text files summarizing the accuracies of the mapped reads for each FASTQ file.

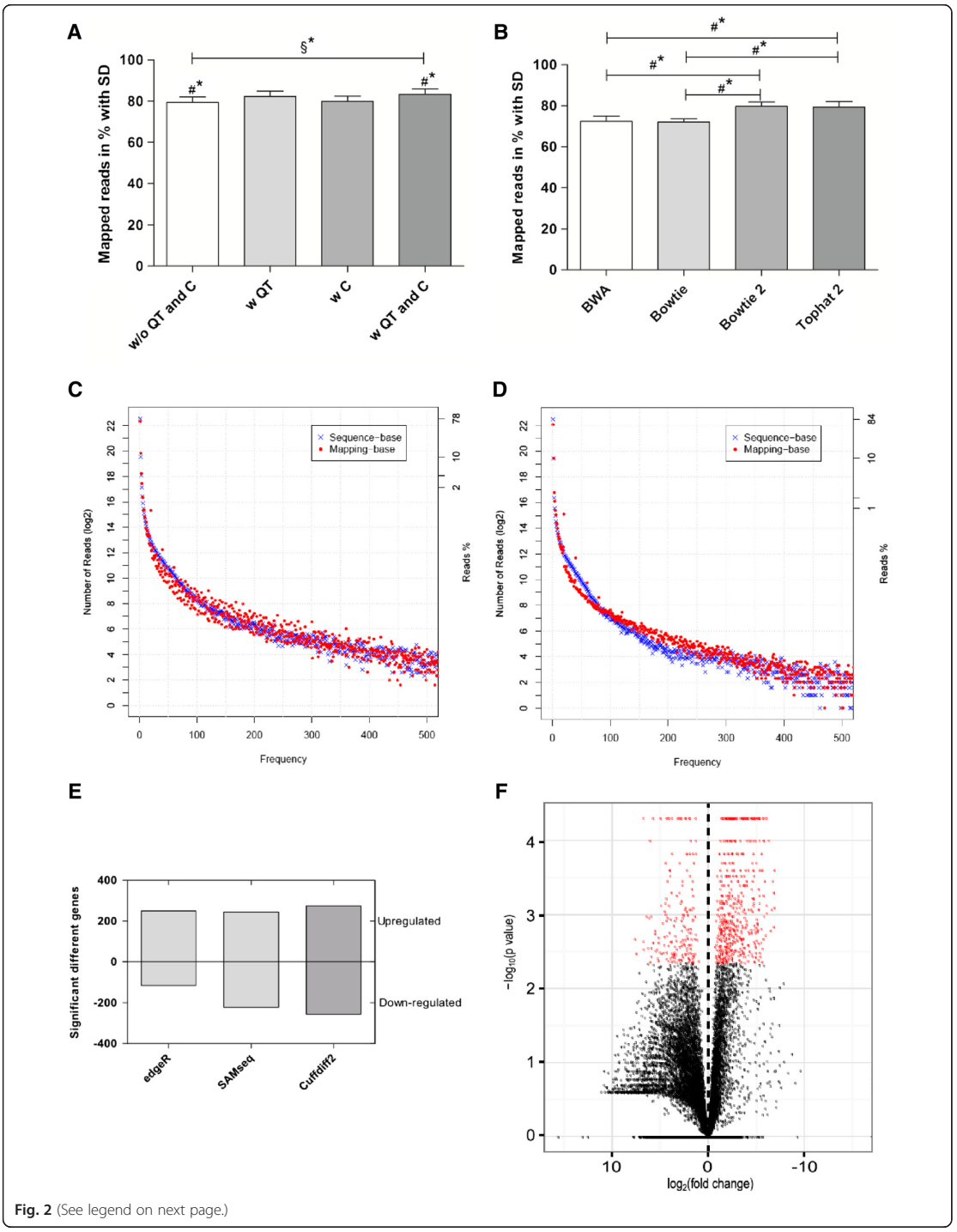


Fig. 2 (See legend on next page.)

(See figure on previous page.)
Fig. 2 Evaluating the different analyses modules of TRAPLINE. **a** A fraction of mapped reads with and without applying pre-processing modules (QT: quality trimming; C: clipping). TopHat2 was used for genome mapping. Error bars indicate the standard deviation. Asterisks indicate a significant difference: # Welch's *t*-test with α = 0.05; § ANOVA with α = 0.05; (*n* = 6). **b** Comparison of different genome mapping tools. The bars indicate the transcript accuracy of the reads aligned to the genome in %, including the standard deviation. Marks indicate significant difference: # Welch's *t*-test with α = 0.05, Bonferroni test with α = 0.05; (n = 6). **c** and **d** Comparison of read correction procedure by Picard Toolkit, before (**c**) and after (**d**), to visualize and correct for multiple RNA sequences in the experimental datasets. RSeQC shows the two specific read duplication correction possibilities: "Sequence-base" reads have the same nucleotide sequence (blue), "Mapping-base" reads have the same mapped sequence, but are aligned to different locations on the genome (red). **e** Comparison of three different DE analysis tools (edgeR, SAMseq and Cuffdiff2), after read mapping with Bowtie (edgeR, SAMseq) TopHat2 (Cuffdiff2). The total number of significantly differentially expressed genes is based on FDR < 0.05 and divided into upregulated and downregulated genes. **f** Vulcano plot illustrating significantly differentially expressed genes (red dots: FC≥2; p≤0.05)

Correction of reads is necessary for SNP detection

The presence of duplicates is a major issue in single/paired short reads from NGS platforms. PCR amplification is one of the major sources of duplicates, which are usually introduced during sequencing library amplification [31]. These duplicates might have a serious impact on research applications, especially towards SNP detection, because they can confound the expression data of a particular gene and, therefore, are usually removed [32]. A popular tool for this task is "MarkDuplicates" from the Picard toolkit (https://github.com/broadinstitute/picard), which finds the 5' coordinates and mapping orientations of each read pair and removes them. During this procedure, the tool considers all clipping that has taken place as well as any gaps or jumps in the alignment. To investigate the influence of the duplicate removing step with Picard tools, we determined the read duplication rates and the number of reads mapped to the same location using the "RSeQC" *Python* module [33]. The results are visualized in Fig. 2c and d. RSeQC uses two strategies to determine read duplication rates: (i) sequence based (blue dots), which means reads with identical sequences are regarded as duplicated reads; (ii) mapping based (red dots) which expresses reads that are mapped to exactly the same genomic location. A comparison of both figures clearly shows an elevation of the mapping-base. The red dots are refined in Fig. 2d in contrast to Fig. 2c, meaning that there were many aligned reads with a similar mapping sequence but with a different location on the genome, which was corrected by "MarkDuplicates". The corrected bam-files can be further investigated by a SNP calling analysis software such as GATK [34] or CRISP [35].

Cuffdiff2 adds value to the standard DE analysis

The different DE analysis methods are based on (i) negative binomial models, such as edgeR, DEseq, baySeq, (ii) non-parametric approaches, such as SAMseq, NOIseq and (iii) transcript-based detection methods, such as Cuffdiff2 and EBSeq [36]. We compared the performances of the most widely used DE tools from each group, which are Cuffdiff2 [37], edgeR [38] and SAMseq [39], to show how they compare and to underline the DE analysis efficiency of TRAPLINE. Prior to the analysis reads were mapped to the reference genome with different methods (Bowtie for edgeR/SAMseq and TopHat2 for Cuffdiff2), because each tool has different data input requirements. Figure 2e shows only slight differences between the applied DE methods. All tools nearly identified the same amount of genes as significantly upregulated among in aCaBs compared to EBs (~250), however different amounts of genes were classified as downregulated. In general, the statistical approaches used by edgeR and SAMseq are more liberal in defining significant differences than the Cuffdiff2 algorithm [17]. In agreement with our results, these widely used methods have recently been compared by several research groups [17, 36, 40, 41]. Cuffdiff2 estimates expression at transcript-level resolution and controls the variability and read mapping ambiguity by using a beta negative binomial model for fragment counts [37]. Furthermore, the tool enhances the comparability between experiments, because it uses the derived "reads per kilobase per million" (RPKM) mapped reads metric [3] which normalizes for both gene size (more reads or fragments can be mapped to larger genes) and the total number of reads or fragments (per million mapped). Seyednasrollah et al. [17] stated Cuffdiff2 as the most conservative DE method with the lowest false positive rate. Therefore, we included Cuffdiff2 for RNAseq DE analysis in TRAPLINE to retrieve precise results with highly significant genes.

As default setting Cuffdiff2 considers genes as significant for $p \leq 0.05$, and a fold change (FC) higher than two. Another reason to integrate Cuffdiff2 into TRAPLINE is the possibility to determine differential splicing events and to perform differential promoter testing [42]. This possibility qualifies the pipeline to investigate for genes with two or more splice variants and genes producing two or more distinct primary transcripts (multi-promoter genes). Multiple splice and promoter isoforms are often co-expressed in a given tissue [3].

We have performed a performance test between TRAPLINE and other tools. A summary of the ratio of mapped reads, discarded reads and significantly differentially

expressed genes obtained with the indicated tools is shown in Additional file 4: Table S1.

Gene annotation, miRNA target prediction and protein-protein interactions with TRAPLINE

Additionally, we included three data annotation and prediction steps into TRAPLINE. First, filtering modules were implemented to scan the list of differentially expressed genes and extract sets of upregulated and downregulated genes. Additionally, users receive a link to DAVID [43] to evaluate the functional influences of the significantly upregulated/downregulated transcripts within their data. In general, DAVID finds Gene Ontology terms (GO terms), signaling pathways (based on databases like Panther, KEGG, Biocarta, etc.) or protein domains (e.g. based on InterPro) that are predominantly associated with lists of genes (e.g. from a DE analysis). Moreover, DAVID performs a functional annotation clustering analysis that groups these terms into functionally related clusters which gives the user a first and quick insight into the biological impact of the discovered differences [44]. Second, TRAPLINE includes modules for miRNA target prediction that use significantly upregulated and downregulated miRNAs and automatically spot possible targets among the downregulated or upregulated mRNAs in the analyzed datasets. For this purpose we provide formatted text files of conserved and non-conserved miRNAs and their predicted targets for different species (human, mouse, rat, fruitfly and nematode), based on the latest version of the microRNA.org database (*release 2010*; [45]). The files can be obtained via a Galaxy history and have to be uploaded as TRAPLINE "miRNA targets" input. Third, we implemented a module which is able to identify verified interactions between proteins of significantly upregulated and downregulated mRNAs. The protein-protein interactions are based on data from peer-reviewed publications deposited in the BioGRID database (*release 3.3.122;* [46]). Similar to the miRNA targets, we provide protein-protein interactions from five different species (human, mouse, rat, fruitfly and nematode) in the form of Galaxy history files and will continuously extend the species.

Identifying transcriptomic differences of EBs and aCaBs

A step by step description on how to use TRAPLINE is provided in the ***Supplementary Material*** section. In summary, users upload FASTQ files from a RNAseq experiment, select the reference genome for the species under investigation and run the pipeline to obtain the significantly differentially expressed transcripts. Optionally, one can upload the provided miRNA target and protein interaction files to use the full potential of TRAPLINE. Exemplarily, we applied the developed pipeline on RNAseq data from our murine ESC derived aCaBs [21] in comparison to control EBs.

We uploaded in total six datasets (as fastqsanger files), the murine reference annotation (as gtf or gff3 that can be obtained from http://geneontology.org/page/reference-genome-annotation-project), the mm9 miRNA targets file (from the provided Galaxy history), the mm9 protein interactions file (also from the history) and ran TRAPLINE with the default parameter settings. After a processing time of ~10 h we retrieved the results. We found ~550 significantly differentially expressed transcripts, 260 of which were upregulated. The volcano plot shown in Fig. 2f illustrates the results of the DE analysis. It shows the ratio of the significantly differentially expressed genes (red) against the non-significant genes (black). At this point one might want to lower the cutoff of the p-value to obtain less reads marked as significant, which is easily possible by tuning the corresponding Cuffdiff2 parameter. However, we took the 260 upregulated genes as input for the subsequent functional classification analysis with DAVID, which revealed several annotation clusters. The first three annotation clusters contain 160 genes in total and suggest a biological impact on the cytoskeleton, actin and the contractile fibers (Additional file 5: Table S2). Based on the annotated biological processes described by GO terms, we created a network to show the links and significance of each GO term using the Cytoscape application ClueGo [47]. The network is shown in Fig. 3 and illustrates the 260 upregulated genes that are associated with enriched biological processes. The distribution of significant biological processes is illustrated in Additional file 6: Figure S3. These genes could be a starting point for subsequent analyses.

We also predicted miRNA interactions of downregulated mRNAs. Their associated GO terms suggest an impact on cardiac cell differentiation. Exemplarily, we show the significantly upregulated miRNA "mmu-mir369" with 5.522 predicted targets which include 57 genes that are downregulated in aCaBs (Additional file 7: Table S3). These 57 genes were functionally classified by DAVID and reveal a high probability to affect the cell cycle and to support cell differentiation. Among these genes is "Atp1a2" which is known to negatively regulate heart function [48]. Furthermore, we analyzed the ~550 significantly differentially expressed mRNAs and identified ~230 verified protein interactions, 10 splice variants and 12 multi promoter regions (Additional file 8: Table S4).

Discussion

We developed TRAPLINE for RNAseq data analysis to link differentially expressed transcripts to the corresponding phenotypic changes and biological phenomena. There exist other tools such as the Bioconductor packages edgeR and DEseq [49], that are with no doubt

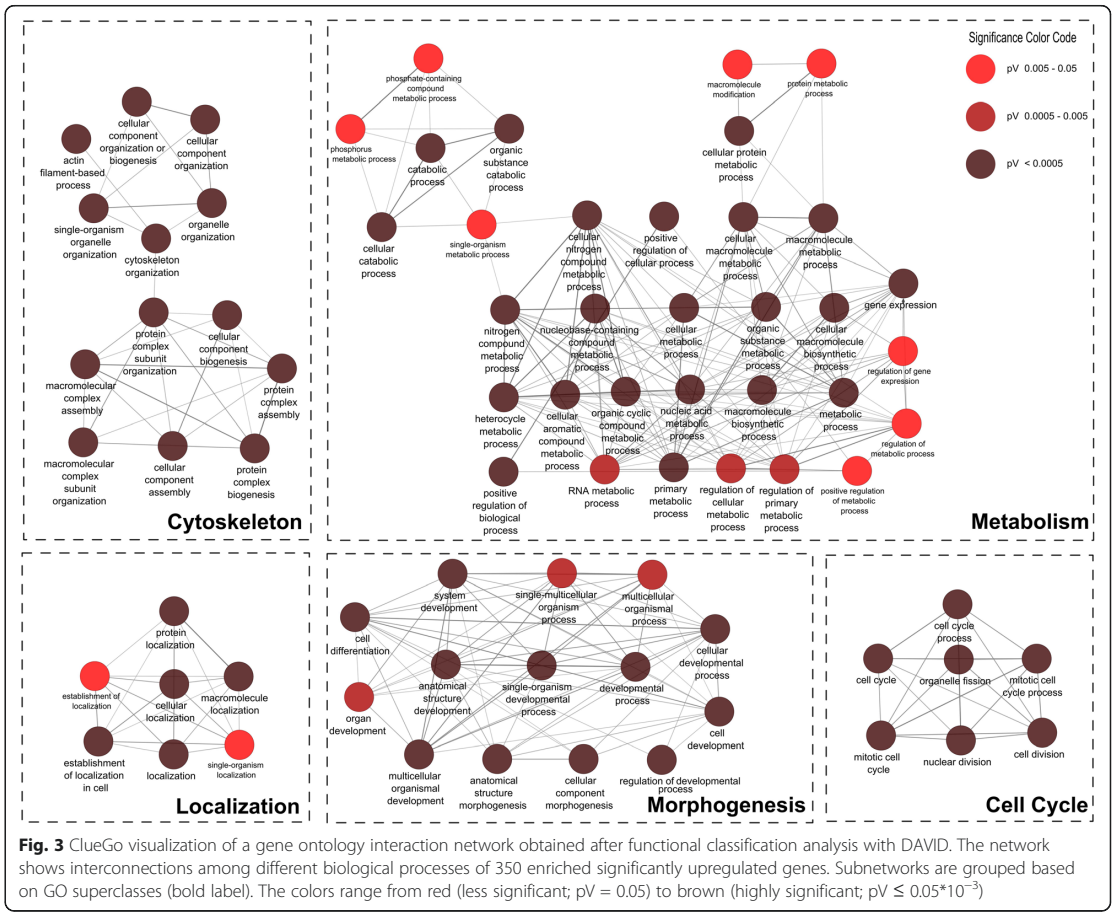


Fig. 3 ClueGo visualization of a gene ontology interaction network obtained after functional classification analysis with DAVID. The network shows interconnections among different biological processes of 350 enriched significantly upregulated genes. Subnetworks are grouped based on GO superclasses (bold label). The colors range from red (less significant; pV = 0.05) to brown (highly significant; pV ≤ $0.05*10^{-3}$)

valuable resources to support the analysis of NGS data. Our pipeline, however, includes pre-processing and genome mapping modules and, is furthermore, easily applicable. TRAPLINE mainly addresses researchers with limited or no programming skills e.g. in *R* or *Python*. We are confident that the graphical user interface of TRAPLINE, which is implemented in the Galaxy platform, greatly supports the accessibility of our RNAseq data analysis pipeline to users with no computational background. Furthermore, we have carefully selected a set of best performing interconnected modules that evade compatibility or file formatting issues. The entire RNAseq data analysis workflow can thus be performed in one go without losing the flexibility that experienced users appreciate when being enabled to adjust module parameters to their own needs. Different to other automated workflows like MeV, Chipster, RobiNA or Grape our pipeline is additionally predicting spliced variants, enriched promoter sites, miRNA targets and protein-protein interactions to enable users getting a comprehensive insight to the analyzed samples [11–14]. There are several other Galaxy pipelines available online, for example the widely used Oqtans workbench [15]. Oqtans is a collection of tools without a pre-defined pipeline. In contrast, our work for the first time introduces an automated Galaxy workflow that includes detailed data analysis and data annotation on a public Galaxy server. TRAPLINE is using all benefits of Galaxy and is independent of computational resources (*i.e.* no need for high performance computers). Researchers can access and share their data and the results worldwide via the internet, however Galaxy also offers private accounts and the possibility to install a local Galaxy instance on a private machine, which is beneficial in case of limited internet connectivity. Moreover, Galaxy enables a synchronous work, e.g. four read mapping tasks at a time are possible. In our case study the time for the analysis was reduced to 10 h in comparison to a desktop PC requiring 24 h (Additional file 9: Table S5). Additionally, to accomplish a transparent computing speed analysis, we performed a comparison between a standard TRAPLINE run

at the public Galaxy instance and a local desktop PC based on a randomly selected publicly available SRA dataset (BioProject:PRJNA292442; SRA study: SRP062238) [50].

The implementation of Cuffdiff2 for detecting differentially expressed genes enhances the comparability between various RNAseq experiments, because the method is accompanied by RPKM normalization [37]. Nevertheless, it has to be considered that the RPKM value for a gene from a deep library may have more statistical meaning than an equivalent value from a more shallow library [51].

It is known that spatial biases along the genome exist, resulting in a non-uniform coverage of expressed transcripts [3]. Especially when using Cufflinks, it has been shown that DE analysis attempting to correct for differences in gene length have the tendency of introducing a bias in the per gene variances, in particular for lowly expressed genes [52]. These spatial biases hinder comparisons between genomic regions and will therefore adversely affect any analysis where such a comparison is integrated. To overcome this problem the current version of Cufflinks2 has an integrated bias correction algorithm [53]. In our investigated datasets there was no need for a bias correction, therefore, we turned this feature off (Additional file 10: Figure S4). It can be reimported manually by setting the respective Cufflinks parameter.

With respect to the biological reliability of the results, the number of our above described 550 significantly differentially expressed genes could be further reduced based on p-value and fold change adjustments. Please be aware that the performance of our pipeline was evaluated based on the Illumina sequencing platform that was used to generate the experimental data. Additionally, it is possible to apply different multiple testing correction method like Bonferroni or Benjamini-Hochberg [54]. Using the same parameters, all three applied methods deliver similar results for differentially expressed genes. With the default parameter values, the pipeline also considers genes which are only slightly up or downregulated ($|FC| \geq 2$). The gene annotation clustering approach enables enrichment in information and a pointer to the biological relevance of the apparently large number of differentially expressed genes. Gene Ontology terms and especially the gene set enrichment analysis performed by DAVID are established methods for gaining first insights into phenotype variations between the tested experimental conditions [43]. Interestingly, the first three enriched GO term clusters in our case study relate to biological processes concerning the cytoskeleton and actin regulation which are two core factors of cardiomyocytes and thus provide a proof of principle for our pipeline (Additional file 5: Table S2).

After successful DE analysis, there are several possibilities for further data evaluation and characterization of the transcripts. As we already showed, the GO terms and differentially expressed mRNAs can be visualized as interaction networks using Cytoscape. miRanda predictions have the largest relative overlap with other miRNA prediction algorithms/tools [55], which is why we chose to include miRanda predictions into TRAPLINE in the first place. A SNP analysis with respective tools can also be done by simply using the SNP output of TRAPLINE. Additionally, a co-expression network analysis could be performed to identify co-expressed mRNAs that are simultaneously dis-regulated [56].

Conclusion

Taken together, our proposed pipeline includes all relevant RNA sequencing data processing modules, is easily applicable, and needs no time consuming installation processes. TRAPLINE guides researchers through the NGS data analysis process in a transparent and automated state-of-the-art pipeline. Experimentalists will be able to analyze their data on their own without learning programming skills or advanced computational knowledge. The data can be accessed worldwide and can optionally be shared among researchers. Gaining quickly in-depth insights into the biology underlying the investigated data, our work for the first time introduces an automated Galaxy workflow including detailed data processing, data evaluation and annotation modules (www.sbi.uni-rostock.de/RNAseqTRAPLINE).

Availability and requirements

Project name: TRAPLINE

Project home page: https://usegalaxy.org/u/mwolfien/p/trapline—manual

Operating system(s): Platform independent

License: Galaxy Web Portal Service Agreement (https://usegalaxy.org/static/terms.html)

Materials and methods

Cell culture and aCaB-Generation

Murine ES cell lines described previously [57] were grown in high glucose DMEM with stable glutamine (GIBCO) containing 10 % FBS Superior (Biochrom), 100 μM non-essential amino acids (GIBCO), 1 % Penicillin/Streptamycin (GIBCO) and 100 μM β-Mercaptoethanol (Sigma) in presence of 1000 U/mL of Leukemia inhibitory factor (LIF, Milllipore). Differentiation of aCaBs was performed in hanging drop culture for two days using 1000 cells as starting material for one EB in Iscove's basal medium (Biochrom) containing 10 % FBS (Biochrom), 100 μM non-essential amino acids (GIBCO), 1 % Penicillin/Streptamycin (GIBCO) and 450 μM 1-Thioglycerol. For additional 4 days, the cells were differentiated in suspension culture, and at day 6 of differentiation consistently 15 EBs were seeded on one well of a 24-well-plate. Antibiotic

selection with 400 μg/mL G418 (Biochrom) was initiated at day 8 post seeding. 4 days thereafter, aCaBs were isolated via treatment with 6000 U/mL Collagenase IV (GIBCO) for 30 min. To obtain single cells for subsequent experiments, the bodies were further dissociated with 100 % Accutase (Affimetrix) for 15 min. To ensure successful generation of aCaBs, potential mycoplasma contamination was routinely controlled twice a week using the PCR based MycoSPY kit system (Biontex).

RNA-Sequencing

For library generation and sequencing, cultured adherent cells were drained from the culture medium, washed and directly lysed by addition of lysis buffer. 1 ul of this lysate was used for cDNA Synthesis and amplification with the SMARTer kit (Clontech, Mountain View CA, USA) according to the manufacturer's instructions. In brief, cDNA synthesis was initiated by annealing a polyA-specific primer and adding a reverse transcriptase with terminal transferase activity. The newly synthesized first strand cDNA is then tailed first with a homopolymer stretch by terminal transferase and then with a specific amplification tag by template switching. The resulting double-tagged cDNA was amplified by PCR, fragmented by sonication (Bioruptor, Diagenode, Liege Belgium; 25 cycles 30 s on/30 s off) and converted to barcoded Illumina sequencing libraries using the NEBnext Ultra DNA library preparation kit (New England Biolabs, Ipswich MA, USA). After PCR enrichment the libraries were purified with AmpureXP magnetic beads (Beckman-Coulter, Brea CA, USA) and quantified on a Bioanalyzer 2100 (Agilent, Santa Clara CA, USA). Libraries were pooled at equimolar amounts and sequenced on an Illumina GenomeAnalyzer IIx in single-read mode with a read-length of 78 nucleotides and a depth of 21 million to 32 million raw reads per replicate.

Additional files

Additional file 1: TRAPLINE manual: Step by Step instructions for the usage. (DOC 35 kb)

Additional file 2: Figure S1. Flowchart for aCaB Generation. Cartoon is displaying sequential steps for the generation of aCaBs, combining *Myh6*-promoter selection and an additional cell-dissociation step [21]. (TIF 90 kb)

Additional file 3: Figure S2. Visualization for RNA transcript quality control and comparison of per base quality score Q. The images are taken before (A) and after (B) quality trimming procedure (removes reads with $Q \leq 20$) to estimate the effect of trimming. The quality score Q is plotted to the read position by using the FastQC package in Galaxy (*http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*). The color indicates the quality of the read: "red" low quality, "orange" median quality, "green" good quality. Red line expresses the mean of the measured values (yellow boxes are inter-quartile range) and the blue line represents the mean quality. (ZIP 81 kb)

Additional file 4: Table S1. Performance comparison of TRAPLINE *vs* other tools. (DOC 28 kb)

Additional file 5: Table S2. Example for DAVID functional gene annotation clustering of significantly differentially expressed genes from aCaBs and EBs. (DOC 48 kb)

Additional file 6: Table S3. Exemplarily we show a result of a miRNA target prediction analysis of TRAPLINE. (TIF 84 kb)

Additional file 7: Figure S3. Pie chart illustrating enriched biological processes of upregulated genes in the aCaB derived cardiomyocytes. The chart presents the enriched GO superclasses. (DOC 26 kb)

Additional file 8: Table S4. Exemplarily results of protein-protein interaction prediction, splice variants and multi promoter regions. (DOC 37 kb)

Additional file 9: Table S5. Benchmarking results of TRAPLINE performed on a public Galaxy server and on a local desktop PC (based on computing speed). (DOC 31 kb)

Additional file 10: Figure S4. A comparison of experiments without (A) and with (B) bias correction performed with the help of Cufflinks2. The dots represent the dependency of the log ratio of two FPKM values (M) and their mean average (A). The MA plot is a common method to investigate the biases of datasets [53]. (PPTX 236 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MW, CR, US, JJJ, GS and RD developed the idea. MW developed the workflow and did the Galaxy implementation. CR and JJJ designed and conducted the experiments. SK performed the RNA sequencing. MW analyzed the experimental data. CR, US, RD helped with the analysis and biological interpretation of the data. OW provided the general framework for bioinformatics study. MW, US, RD, OW contributed in drafting the manuscript, revising it critically and approved the final version. MW, CR and US share first authorships. RD and OW are joint senior authors on this work. All authors read and approved the final manuscript.

Acknowledgements

This work has been funded by the Federal Ministry of Education and Research Germany (FKZ 0312138A, FKZ 316159 and FKZ 02NUK043C) and the State Mecklenburg-Western Pomerania with EU Structural Funds (ESF/IV-WM-B34-0030/10 and ESF/IV-BM-B35-0010/12), by the DFG (DA 1296–1), the German Heart Foundation (F/01/12), by the FORUN Program of Rostock University Medical Centre (889001) and the EU funded CaSyM project (grant agreement #305033).

Author details

[1]Department of Systems Biology and Bioinformatics, University of Rostock, 18057 Rostock, Germany. [2]Reference und Translation Center for Cardiac Stem Cell Therapy (RTC), University of Rostock, Rostock 18057, Germany. [3]Gene & Stem Cell Therapy Program, Centenary Institute, 2050 Camperdown, Australia. [4]Gene Center Munich, LMU Munich, 81377 Munich, Germany. [5]Stellenbosch Institute of Advanced Study (STIAS), Wallenberg Research Centre at Stellenbosch University, 7602 Stellenbosch, South Africa. [6]Sydney Medical School, University of Sydney, Sydney, NSW 2006, Australia.

Received: 5 September 2015 Accepted: 22 December 2015
Published online: 06 January 2016

References

1. Hayden EC. Genome sequencing: the third generation. Nature. 2009; 457(7231):768–9.
2. Morozova O, Hirst M, Marra MA. Applications of New Sequencing Technologies for Transcriptome Analysis. Annu Rev Genom Hum G. 2009;10: 135–51.
3. Mortazavi A, Williams BA, Mccue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5(7): 621–8.
4. Hu Y, Wang K, He XP, Chiang DY, Prins JF, Liu JZ. A probabilistic framework for aligning paired-end RNA-seq data. Bioinformatics. 2010;26(16):1950–7.
5. Pepke S, Wold B, Mortazavi A. Computation for ChIP-seq and RNA-seq studies. Nat Methods. 2009;6(11):S22–32.

6. Ramskold D, Wang ET, Burge CB, Sandberg R. An Abundance of Ubiquitously Expressed Genes Revealed by Tissue Transcriptome Sequence Data. Plos Computational Biology. 2009;5(12):e1000598.
7. Wilhelm BT, Landry JR. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods. 2009; 48(3):249–57.
8. Wilhelm BT, Marguerat S, Goodhead I, Bahler J. Defining transcribed regions using RNA-seq. Nat Protoc. 2010;5(2):255–66.
9. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008;18(9):1509–17.
10. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57–63.
11. Howe EA, Sinha R, Schlauch D, Quackenbush J. RNA-Seq analysis in MeV. Bioinformatics. 2011;27(22):3209–10.
12. Kallio MA, Tuimala JT, Hupponen T, Klemela P, Gentile M, Scheinin I, et al. Chipster: user-friendly analysis software for microarray and other high-throughput data. BMC Genomics. 2011;12:507.
13. Knowles DG, Roder M, Merkel A, Guigo R. Grape RNA-Seq analysis pipeline environment. Bioinformatics. 2013;29(5):614–21.
14. Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, et al. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res. 2012;40(Web Server issue):W622–7.
15. Sreedharan VT, Schultheiss SJ, Jean G, Kahles A, Bohnert R, Drewe P, et al. Oqtans: the RNA-seq workbench in the cloud for complete and reproducible quantitative transcriptome analysis. Bioinformatics. 2014;30(9):1300–1.
16. Kalari KR, Nair AA, Bhavsar JD, O'Brien DR, Davila JI, Bockol MA, et al. MAP-RSeq: Mayo Analysis Pipeline for RNA sequencing. BMC Bioinformatics. 2014;15.
17. Seyednasrollah F, Laiho A, Elo LL. Comparison of software packages for detecting differential expression in RNA-seq studies. Brief Bioinform. 2013; 16(1):59–70.
18. Blankenberg D, Hillman-Jackson J. Analysis of next-generation sequencing data using Galaxy. Methods Mol Biol. 2014;1150:21–43.
19. Goecks J, Nekrutenko A, Taylor J, Team G. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biology. 2010, 11(8).
20. Rimmbach C, Jung JJ, David R. Generation of Murine Cardiac Pacemaker Cell Aggregates Based on ES-Cell-Programming in Combination with Myh6-Promoter-Selection. J Vis Exp. 2015.
21. Jung JJ, Husse B, Rimmbach C, Krebs S, Stieber J, Steinhoff G, et al. Programming and isolation of highly pure physiologically and pharmacologically functional sinus-nodal bodies from pluripotent stem cells. Stem Cell Reports. 2014;2(5):592–605.
22. Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A. Manipulation of FASTQ data with Galaxy. Bioinformatics. 2010;26(14): 1783–5.
23. Del Fabbro C, Scalabrin S, Morgante M, Giorgi FM. An extensive evaluation of read trimming effects on Illumina NGS data analysis. PLoS ONE. 2013; 8(12):e85024.
24. Mbandi SK, Hesse U, Rees DJ, Christoffels A. A glance at quality score: implication for de novo transcriptome reconstruction of Illumina reads. Front Genet. 2014;5:17.
25. Chen C, Khaleel SS, Huang H, Wu CH. Software for pre-processing Illumina next-generation sequencing short read sequences. Source Code Biol Med. 2014;9:8.
26. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
27. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.
28. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
29. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7(3):562–78.
30. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14(4):R36.
31. Kozarewa I, Ning ZM, Quail MA, Sanders MJ, Berriman M, Turner DJ. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G plus C)-biased genomes. Nat Methods. 2009;6(4):291–5.
32. Xu HB, Luo X, Qian J, Pang XH, Song JY, Qian GR, et al. FastUniq: A Fast De Novo Duplicates Removal Tool for Paired Short Reads. PLoS ONE. 2012;7(12): e52249.
33. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics. 2012;28(16):2184–5.
34. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20(9):1297–303.
35. Bansal V. A statistical method for the detection of variants from next-generation resequencing of DNA pools. Bioinformatics. 2010;26(12): i318–24.
36. Soneson C, Delorenzi M. A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics. 2013;14.
37. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2013;31(1):46. −+.
38. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11(3):R25.
39. Li J, Tibshirani R. Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. Stat Methods Med Res. 2013;22(5):519–36.
40. Kvam VM, Liu P, Si Y. A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. Am J Bot. 2012;99(2): 248–56.
41. Nookaew I, Papini M, Pornputtapong N, Scalcinati G, Fagerberg L, Uhlen M, et al. A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in Saccharomyces cerevisiae. Nucleic Acids Res. 2012;40(20):10084–97.
42. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511–5.
43. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1): 44–57.
44. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. 2007;8(9):R183.
45. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. Nucleic Acids Res. 2008;36(Database issue): D149–53.
46. Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, et al. The BioGRID interaction database: 2015 update. Nucleic Acids Res. 2015;43(Database issue):D470–8.
47. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009;25(8): 1091–3.
48. Swift F, Tovsrud N, Sjaastad I, Sejersted OM, Niggli E, Egger M. Functional coupling of alpha(2)-isoform Na(+)/K(+)-ATPase and Ca(2+) extrusion through the Na(+)/Ca(2+)-exchanger in cardiomyocytes. Cell Calcium. 2010; 48(1):54–60.
49. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106.
50. Luxan G, Casanova JC, Martinez-Poveda B, Prados B, D'Amato G, MacGrogan D, et al. Mutations in the NOTCH pathway regulator MIB1 cause left ventricular noncompaction cardiomyopathy. Nat Med. 2013; 19(2):193–201.
51. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics. 2010;11.
52. Oshlack A, Wakefield MJ. Transcript length bias in RNA-seq data confounds systems biology. Biol Direct. 2009;4.
53. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. 2011; 12(3):R22.

54. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met. 1995; 57(1):289–300.
55. Ritchie W, Flamant S, Rasko JE. Predicting microRNA targets and functions: traps for the unwary. Nat Methods. 2009;6(6):397–8.
56. Iancu OD, Kawane S, Bottomly D, Searles R, Hitzemann R, McWeeney S. Utilizing RNA-Seq data for de novo coexpression network inference. Bioinformatics. 2012;28(12):1592–7.
57. David R, Groebner M, Franz WM. Magnetic cell sorting purification of differentiated embryonic stem cells stably expressing truncated human CD4 as surface marker. Stem Cells. 2005;23(4):477–82.

2.1.2 Customized workflow development and data modularization concepts

Lott, S.C., **Wolfien, M.**, Riege, K., Bagnacani, A., Wolkenhauer, O., Hoffmann, S., and Hess, W.R. (2017).
Customized workflow development and data modularization concepts for RNA-Sequencing and metatranscriptome experiments.
Journal of Biotechnology. IF: 3.142, Citations (December 14, 2020): 6

RNA-Seq has become a widely used approach to study ample quantitative and qualitative aspects of transcriptome data. In this article, an overview is given about the variety of RNA-Seq protocols, experimental study designs, and to what extend the characteristic properties of the organisms under investigation affect downstream and comparative analyses.

Here, I focussed on the increasing importance for well-suited developed scientific workflows, as well as slecting the most appropriate data analysis frameworks. The reader learns about the benefits, while using workflows in data analysis frameworks. In addition, it includes a technical illustration what has to be done to develop better workflows as a bioinformatician for the end user. Current computational technologies, such as Galaxy[4], Docker[5], and BioConda[6], are introduced for a scientific use.

In summary, we explained the impact of structured pre-selection, classification, and integration of best-performing tools within modularized data analysis workflows and ready-to-use computing infrastructures. Examples for workflows and specific use cases were highlighted; these contain analyses for prokaryotic, eukaryotic, and mixed dual RNA-Seq (meta-transcriptomics) experiments. The authors are summarizing their expertise in the German Network for Bioinformatics Infrastructure (de.NBI)[7] about "*Structured Analysis and Integration of RNA-Seq experiments*" (de.STAIR)[8] and their integration into the RNA Bioinformatics Center.[9]

[4] https://usegalaxy.org/
[5] https://www.docker.com/
[6] https://bioconda.github.io/
[7] https://www.denbi.de/
[8] https://www.sbi.uni-rostock.de/research/projects/detail/47
[9] https://www.denbi.de/network/rna-bioinformatics-center-rbc

Journal of Biotechnology 261 (2017) 85–96

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

Review

Customized workflow development and data modularization concepts for RNA-Sequencing and metatranscriptome experiments

Steffen C. Lott[a,1], Markus Wolfien[b,1], Konstantin Riege[c,1], Andrea Bagnacani[b,1], Olaf Wolkenhauer[b,d], Steve Hoffmann[c], Wolfgang R. Hess[a,*]

[a] Genetics and Experimental Bioinformatics, Faculty of Biology, University of Freiburg, Schänzlestr. 1, 79104 Freiburg, Germany
[b] Department of Systems Biology & Bioinformatics, University of Rostock, Ulmenstr. 69, 18057 Rostock, Germany
[c] Transcriptome Bioinformatics Group, LIFE Research Complex, University Leipzig, Härtelstrasse 16-18, 04107 Leipzig, Germany
[d] Stellenbosch Institute of Advanced Study (STIAS), Wallenberg Research Centre at Stellenbosch University, 7602 Stellenbosch, South Africa

ARTICLE INFO

Keywords:
Workflow development
Transcriptomics
Regulatory RNA
Metatranscriptomics

ABSTRACT

RNA-Sequencing (RNA-Seq) has become a widely used approach to study quantitative and qualitative aspects of transcriptome data. The variety of RNA-Seq protocols, experimental study designs and the characteristic properties of the organisms under investigation greatly affect downstream and comparative analyses. In this review, we aim to explain the impact of structured pre-selection, classification and integration of best-performing tools within modularized data analysis workflows and ready-to-use computing infrastructures towards experimental data analyses. We highlight examples for workflows and use cases that are presented for pro-, eukaryotic and mixed dual RNA-Seq (*meta*-transcriptomics) experiments. In addition, we are summarizing the expertise of the laboratories participating in the project consortium "Structured Analysis and Integration of RNA-Seq experiments" (de.STAIR) and its integration with the Galaxy-workbench of the RNA Bioinformatics Center (RBC).

1. Introduction: RNA related sequencing experiments and associated data analyses

1.1. The variety of transcript types and regulatory RNAs

RNA is in the form of mRNAs, tRNAs and rRNAs the type of molecule that interconnects the mechanisms involved in the readout of genetic information from the genome to protein. However, different types of RNA participate in a wide variety of additional processes. These include RNAs involved in the regulation of multiple physiological processes, following similar principles, often through forming sequence-specific base pairings with cellular RNA or DNA targets (Gorski et al., 2017). Among the types of RNAs fundamental to these RNA-based systems are miRNAs in eukaryotic cells, small RNAs (sRNAs) in bacteria and archaea, but also CRISPR RNAs (crRNAs), which are at the heart of the prokaryotic immune mechanism. All these RNAs act by using seed sequences that are presented through a particular ribonucleoprotein complex.

Different, yet highly relevant classes of regulatory RNA are antisense transcripts that often play gene expression modulating functions (Georg and Hess, 2011; Yelin et al., 2003) in all three domains of life and long non-coding RNAs (lncRNAs) in eukaryotes that often impact epigenetic status and chromosome organization (Jégu et al., 2017).

1.2. RNA-Seq experiments may come in different flavours: RNA-Seq, dRNA-Seq, dual RNA-Seq, ribosomal profiling, metaRNA-Seq

Most RNA-Seq experiments are performed to measure differential gene expression. For this, RNA-Seq targets the composition of the entire transcriptome in a sample using next-generation sequencing techniques. By quantifying and comparing the transcriptome composition between samples of a time series or from different tissues or cell types, differences in gene expression are detected (Wang et al., 2009). Therefore, RNA-Seq can be applied to any kind of cell from any kind of organism, even without prior knowledge about its genome sequence. However, RNA-Seq is a powerful tool not only to analyze quantitative changes in gene expression. With RNA-Seq exon/intron boundaries as well as alternatively spliced transcript variants can be detected and quantified or post-transcriptional modifications identified. In more specialized RNA-Seq protocols, information is obtained about the suite of active transcriptional start sites (TSSs) or particular RNA classes such as sRNAs or miRNAs.

* Corresponding author.
E-mail address: wolfgang.hess@biologie.uni-freiburg.de (W.R. Hess).
[1] Shared first authors.

http://dx.doi.org/10.1016/j.jbiotec.2017.06.1203
Received 21 February 2017; Received in revised form 22 June 2017; Accepted 26 June 2017
Available online 01 July 2017

One of the RNA-Seq variants targeting a particular fraction of the transcriptome is Ribo-Seq, or Ribosomal profiling (Ingolia et al., 2012). This approach targets specifically mRNA sequences protected by the ribosome during the process of translation. Therefore, it provides information on the complement of actively translated mRNAs at a certain moment, on the presence of signals for translation and on the regulation of protein synthesis. Other RNA-Seq variants target a particular fraction of the transcriptome, e.g., after size fractionation, specifically miRNAs and other sRNAs (Aldridge and Hadfield, 2011).

One widely applied specialized RNA-Seq protocol is called differential RNA-Seq (dRNA-Seq) (Sharma et al., 2010), a prolific experimental approach for the identification of all active TSSs at single-nucleotide resolution (Borries et al., 2012). As this method not only identifies TSSs linked to an mRNA but all TSSs, it is a superior approach for the detection of bacterial sRNAs. The dRNA-Seq protocol was first applied to the human pathogen *Helicobacter pylori* (Sharma et al., 2010) and rapidly applied to other bacteria, such as *E. coli* (Thomason et al., 2015), *Salmonella enterica* serovar Typhimurium (Kröger et al., 2012), *Streptococcus pyogenes* (Deltcheva et al., 2011), *Xanthomonas* (Schmidtke et al., 2012) and various cyanobacteria (Mitschke et al., 2011a; Mitschke et al., 2011b; Kopf et al., 2014; Kopf et al., 2015a; Kopf et al., 2015b; Voss et al., 2013; Pfreundt et al., 2014; Voigt et al., 2014). Whereas dRNA-Seq initially was primarily developed for bacteria, it has been applied to archaea (Jäger et al., 2009; Babski et al., 2016) and to eukaryotic cells. In a variant of dRNA-Seq, called "dual RNA-Seq", the primary transcriptomes of a bacterial pathogen together with that of its eukaryotic host cells are analyzed in parallel (Westermann et al., 2016).

More recently, this methodology has been expanded for the analysis of complex environmental assemblages of organisms belonging to diverse species from all three domains of life (Hou et al., 2016). Metatranscriptomic differential RNA-Seq (mdRNA-Seq) as well as metatranscriptomic RNA-Seq (metaRNA-Seq) are protocols to analyze the highly complex transcript pools of entire biological communities or microbiomes.

1.3. Too many good options? data analysis tools and algorithms for RNA-Seq experiments

The range of RNA related sequencing experiments and subsequent data analyses are steadily increasing. Researchers have to weigh and decide on specific technologies and combinations of experiments which are getting more and more complex (Adiconis et al., 2013; Kukurba and Montgomery, 2015; Podnar et al., 2014). Likewise, newly developed tools, which are rapidly published for many sequencing data analysis tasks, are also based on more sophisticated algorithms. A current collection of commonly used tools of any kind can be obtained at omic.-tools and bio.tools (please see also Table 1 for the list of online resources and websites that are mentioned or discussed in this review).

Tools, which are no longer maintained or were never designed to cope with evolving RNA-Seq protocols and the rapidly increasing amount of available sequence data from first (Sanger), second (454, Solexa, Illumina) and third generation sequencing approaches (IonTorrent, SOLiD, Nanopore, PacBio, SMRT) become outdated over time. Another phenomenon is that data analysis tools being continuously maintained may change their behavior and parameters over time. Therefore, the reuse of formerly generated workflows is frequently not simple or not possible at all to adapt towards certain analyses at the present time. Another major challenge is the comparison, benchmarking, selection and integration of the most appropriate tools, which is time-consuming and needs computational domain expertise. Depending on the number of samples, the scale of time series and sequencing depth, computations may require heavy computational resources such as cluster, grid and cloud computing solutions. An adaptive management of available computing resources by load balancers and queuing systems is often inevitable in creating analysis workflows.

The German Bioinformatics Network Infrastructure (de.NBI) as well as the European Network ELIXIR are aiming at supporting and training scientists with respect to diverse bioinformatics questions. In particular, the de.STAIR project focusses on the needs of the experimental researchers for robust data analysis tools and, therefore, develops tailor-made workflows for RNA-Seq experiments and further downstream data integration approaches to facilitate the accessibility of the latest bioinformatic tools, the most suitable analysis approaches and flexible computing environments.

The de.STAIR service is highlighted in Fig. 1 and represents the data analysis elements of the workflows (which are an extension of the recommendations by Conesa et al. (2016) that are developed within the infrastructure of the RBC. We offer workflows for pro-, eukaryotic and mixed dual RNA-Seq (*meta*-transcriptomics) experiments for multiple input layers, like raw fastq, quality controlled fastq, sam/bam files, etc. The data analyses procedures cover preprocessing, alignment and further advanced downstream analyses such as alternative and non-linear splicing, differential expression, epigenetic analyses and many more. The output from the workflows includes quality reports, calculations and predictions for novel transcripts, probabilities of differentially expressed transcripts and transcript characterizations like annotations, such as GO, KEGG, Panther, wiki pathways, HMDB, DisGeNet, Reactome and methods to visualize the results (e.g. volcano plots, heatmaps, PCAs, networks, sashimi plots).

The following sections provide detailed insights towards the used technology for the distribution and modularization of the workflows, the necessity for integration of transcriptomic data and specific examples of the potential of our service.

2. Scientific workflows and modules for RNA-Sequencing

Reduced costs and increased accuracy of biological sequencing enabled the investigation of biological phenomena at a high resolution. Unless the low entrance barriers and easy-to-use experimental protocols, the challenge of proper, transparent and reproducible data analyses are still a bottleneck (Spjuth et al., 2016). With respect to the number of data analysis steps, the complexity of decisions on tool selection is increasing likewise, hence calling for systematic workflow development and management frameworks (Lampa et al., 2013).

2.1. Choosing the most appropriate workflow management framework

The de.NBI and ELIXIR initiatives are supporting the expansion and further development of accessible workflow frameworks:

Galaxy (Afgan et al., 2016) and the Galaxy-RNA-Workbench (Grüning et al., 2017): The Galaxy project is a framework that makes advanced computational tools accessible without the need of prior extensive training. Galaxy seeks to make data-intensive research more accessible, transparent and reproducible by providing a web-based environment in which users can perform computational analyses and have all of the details automatically tracked for later inspection, publication, or reuse.

- Applicable for non-computational user's on a public server, explanatory interactive Galaxy tours, Galaxy "Tool Shed" for advanced user's (contains more than 3.500 tools), free to use (open-source), broad community with over 80 public servers available for various tasks, pre-build Docker/rkt images, international training network, new tools need to be xml wrapped to be integrated

KNIME (Berthold et al., 2009; de la Garza et al., 2016): The Konstanz Information Miner is a modular environment, which enables easy visual assembly and interactive execution of a data pipeline. It is designed as a teaching, research and collaboration platform, which enables simple integration of new algorithms and tools as well as data manipulation or visualization methods in the form of new modules or

Table 1
Web sites and resources that have been mentioned or discussed in this review (accessed at 17th Feb. 2017).

Organizations	
Name	Web address
Bioconda	https://bioconda.github.io/
de.NBI	https://www.denbi.de/
de.STAIR	http://destair.bioinf.uni-leipzig.de/
Docker	https://www.docker.com/
Elixir	https://www.elixir-europe.org/
rkt	https://coreos.com/rkt
RNA Bioinformatics Center (RBC)	https://www.denbi.de/rbc

Tools	
Name	Web address
Bcheck	http://rna.tbi.univie.ac.at/bcheck/
BioConductor	https://www.bioconductor.org/
bio.tools	https://bio.tools/
Blast	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Blat	https://genome.ucsc.edu/cgi-bin/hgBlat
Bowtie2	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
BWA	http://bio-bwa.sourceforge.net/
Chipster	http://chipster.csc.fi/
ConDeTri	https://github.com/linneas/condetri
CopraRNA	https://dx.doi.org/10.1093/nar/gku359
CRT	https://dx.doi.org/10.1186/1471-2105-8-209
Cufflinks/Cuffdiff	http://cole-trapnell-lab.github.io/cufflinks/
Cutadapt	https://github.com/marcelm/cutadapt
DESeq2	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
DNApi	https://github.com/jnktsj/DNApi
edgeR	https://bioconductor.org/packages/release/bioc/html/edgeR.html
Emboss	http://emboss.sourceforge.net/
FastQC	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
FASTX-Toolkit	http://hannonlab.cshl.edu/fastx_toolkit/
featureCounts	http://subread.sourceforge.net/
Galaxy	https://usegalaxy.org
Galaxy-RNA-Workbench	https://github.com/bgruening/galaxy-rna-workbench
GLASSgo	http://rna.informatik.uni-freiburg.de/GLASSgo/Input.jsp
Gorap	https://github.com/rna-hta-jena/gorap
HiSat2	https://ccb.jhu.edu/software/hisat2/index.shtml
HTSeq-count	http://www-huber.embl.de/HTSeq/doc/count.html
Infernal	http://eddylab.org/infernal/
IntaRNA	https://dx.doi.org/10.1093/bioinformatics/btn544
Kallisto	https://pachterlab.github.io/kallisto/
KNIME	https://www.knime.org/
Kraken package	https://ccb.jhu.edu/software/kraken/
NcDNAlign	https://doi.org/10.1016/j.ygeno.2008.04.003
NGS QC Toolkit	http://www.nipgr.res.in/ngsqctoolkit.html
omic.tools	https://omictools.com/
PAREsnip	https://dx.doi.org/10.1093/nar/gks277
PrinSeq	https://sourceforge.net/projects/prinseq/files/
Prokka	https://dx.doi.org/10.1093/bioinformatics/btu153
RAST	https://dx.doi.org/10.1186/1471-2164-9-75
Reapr	http://www.sanger.ac.uk/science/tools/reapr
RNAlien	http://rna.tbi.univie.ac.at/rnalien/
RNAcounter	https://pypi.python.org/pypi/rnacounter
RNAhybrid	https://dx.doi.org/10.1093/nar/gkl243
RNAmmer	https://dx.doi.org/10.1093/nar/gkm160
RNAplex	https://dx.doi.org/10.1093/bioinformatics/btn193
RNApredator	https://dx.doi.org/10.1093/nar/gkr467
RNASEG	https://dx.doi.org/10.1186/1471-2105-15-122
RNAz	https://www.tbi.univie.ac.at/software/RNAz/
Sailfish	http://www.cs.cmu.edu/~ckingsf/software/sailfish/
Salmon	https://combine-lab.github.io/salmon/
Segemehl	http://www.bioinf.uni-leipzig.de/Software/segemehl/
Sickle	https://github.com/najoshi/sickle
SILVA	https://www.arb-silva.de/
SIPHT	https://dx.doi.org/10.1007/978-1-61779-949-5_1
Skewer	https://github.com/relipmoc/skewer
Snakemake	https://snakemake.readthedocs.io/en/stable/
SortMeRNA	https://dx.doi.org/10.1093/bioinformatics/bts611
Star	https://github.com/alexdobin/STAR
TargetRNA2	https://dx.doi.org/10.1093/nar/gku317
TopHat2	https://ccb.jhu.edu/software/tophat/manual.shtml

(continued on next page)

Table 1 *(continued)*

Tools	
Name	Web address
Trapline	https://www.sbi.uni-rostock.de/TRAPLINE
Trimmomatic	http://www.usadellab.org/cms/?page=trimmomatic
TriplexRNA	https://triplexrna.org/
tRNAscan-SE	http://lowelab.ucsc.edu/tRNAscan-SE/
TSSAR	https://dx.doi.org/10.1186/1471-2105-15-89
UPARSE	https://dx.doi.org/10.1038/nmeth.2604
VSEARCH	https://dx.doi.org/10.7717/peerj.2584

nodes.

- Modular, grid and user support environment, workflows are interoperable and represented as Petri nets, hierarchy of workflows possible, e.g., meta nodes can wrap a sub-workflow into an encapsulated new workflow, framework enables "*hiliting*" (selecting and highlighting several rows in a data table and the same rows are also highlighted in all other views that show the same data table), execution of workflows on high performance clusters only within the commercial version

Chipster (Gentleman et al., 2004; Kallio et al., 2011): Chipster is a user-friendly analysis software for high-throughput data (contains currently more than 360 tools). Its intuitive graphical user interface enables biologists to access a powerful collection of data analysis and integration tools, and to visualize data interactively. Users can collaborate by sharing analysis sessions and workflows.

- Desktop application user interface available (Java), strong support and easy integration of R based tools (e.g. from BioConductor), freely available and open source client-server system, about 25 different visualizations (interactive and static)

Snakemake (Köster and Rahmann, 2012): Snakemake is a workflow engine that provides a readable Python-based workflow definition language and a powerful execution environment that scales from single-core workstations to compute clusters without modifying the workflow.

- Readable Python-based workflow definition language, efficient resource usage, available on Linux, computationally advanced command line based framework, interoperates with any installed tool or available web service, jobs can be visualized as directed acyclic graph

In addition, a systematic search and evaluation of further workflow management frameworks with a focus on RNA-Seq data analysis was done by Poplawski et al. (Poplawski et al., 2016).

2.2. Technical illustration of cloud computing frameworks

After choosing and setting up the analysis workflow within an appropriate framework one has to decide on a reasonable computing environment. In general, computing environments can be distinguished between web-based (free of charge community cloud computing), offline, and hybrid solutions (e.g., private and commercial cloud computing). According to the National Institute of Standards and Technology cloud computing is defined "as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released

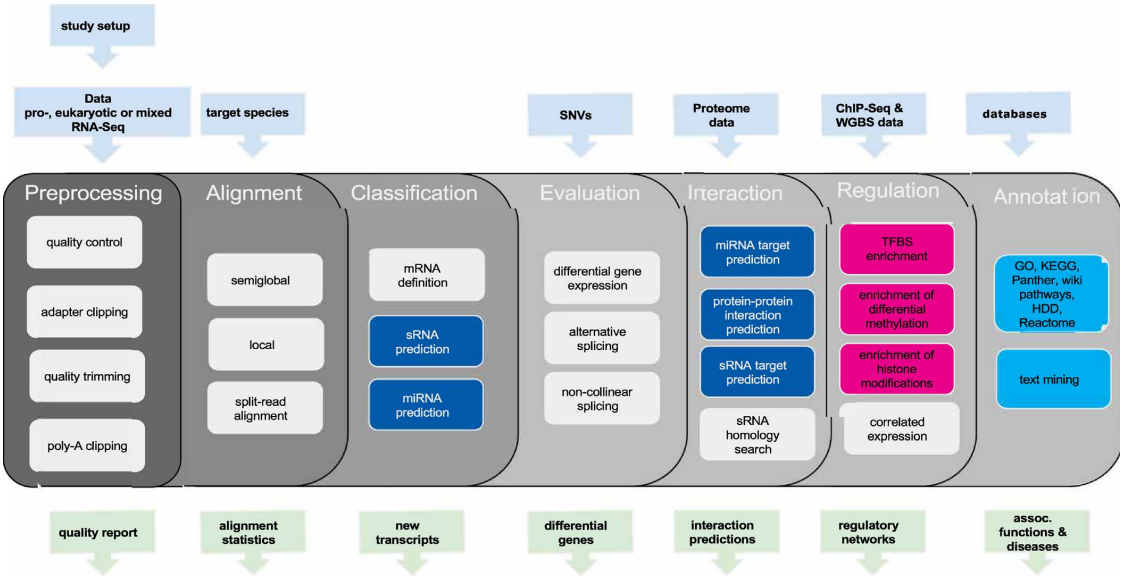


Fig. 1. Overview of possible data input formats (top), modularized workflow elements separated by certain tasks (middle; dark blue: *in silico* predictions; pink: linkage of epigenetic data; light blue: integration of further databases) and examples of subsequently generated output (bottom).

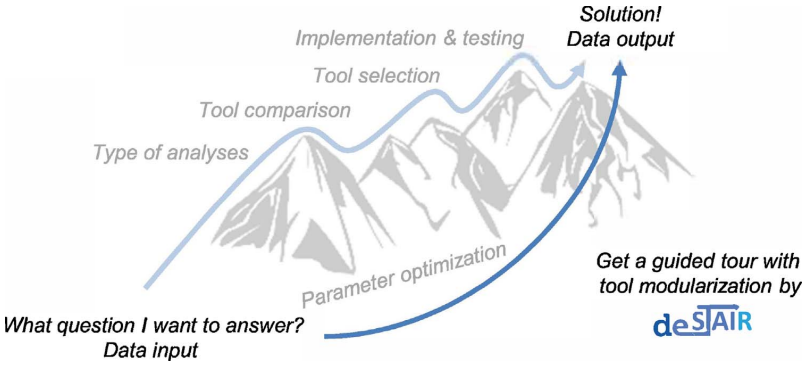


Fig. 2. Conceptualization of the de.STAIR project's recommendation system. The figure shows how the use of our guided and modularized workflows allow end users to achieve their goals by actively choosing a trajectory path that optimizes their analysis. A recommendation system is responsible for the guidance of the trajectory, leaving the user the choice of which tool to operate with. Such a system lets researchers carry out their analyses while providing the benefits of using a modularized workflow instead of developing data analysis pipelines from scratch.

with minimal management effort or service provider interaction". Due to the vastly increasing amount of computational data being created, large consortia, namely public-private-partnerships, are established to share objectives, resources, costs, risks, and responsibilities between academia and industrial partners (e.g., International Cancer Genome Consortium, 100.000 Genomes Project) (Granados Moreno et al., 2017). The most frequent used commercial cloud services are Google's and Amazon's Web Services private cloud-computing infrastructures. Due to concerns on data safety, security and privacy, cloud computing is rather weakly adopted within the healthcare system (Griebel et al., 2015). An emerging solution to deploy the workflows, including all necessary tools and dependencies, are software channels and containers like Bioconda, Docker or rkt. These containers are emerging as a possible solution for many of the formerly addressed concerns, as they allow the packaging of workflows in an isolated and self-contained system, which simplifies the distribution and execution of tools in a portable manner across a wide range of computational platforms such as Galaxy and KNIME (Di Tommaso et al., 2015). The technology combines several areas from systems research, especially operating system virtualization, cross-platform portability, modular reusable elements, versioning, and a "DevOps" philosophy (Boettiger, 2015). Exemplarily, Wolfien et al. (2016) and Schulz et al. (2016) demonstrated successful implementations of a Galaxy/Docker based workflow with discrete software applications for the analysis of NGS data (Schulz et al., 2016; Wolfien et al., 2016).

2.3. How to develop better workflows for end users

Workflow management frameworks and cloud computing services are bridging the gap between tool developers and end users, aiming towards an easy applicable and up-scalable computational data analysis. This in turn allows for an improved data reproducibility, process documentation, and monitoring of submitted jobs. Finally, workflows facilitate the use of state-of-the-art computational tools which would be hard to access for non-experts without graphical user interface frameworks. However, the use of workflows could be even more simplified for experimental researchers by strengthening the specific focus on the addressed research hypothesis and lessening the effort for the selection of the most appropriate tool. The selection and benchmarking of new tools by the bioinformatician is a crucial step for establishing and updating applicable data analyses workflows for non-computational experts with the help of *modularized* workflow development. Starting with a hypothesis or research question, the user will be guided to the necessary input data type and the most suitable software solution will be provided as a modularized workflow (Fig. 2). Therefore, the comparison and the selection of existing tools as well as their implementation into the computing infrastructure will be omitted for the end-user, which will save time and guarantees an expert driven data analysis. With respect to our documentation, the parameters have to be adjusted and optimized to obtain the final results.

In order to adapt workflows over time, we recommend keeping up with the changes in the tools by a registration in bio.tools (Ison et al., 2016), where they are described by means of the EDAM Ontology (Ison et al., 2013). This ontology enables the characterisation of a tool's input formats, output formats, and parameter types. Our proposed software layer would therefore leverage on the EDAM Ontology to infer what tool and parametrization can be used to carry out the desired task. This approach is tool implementation agnostic, which means that if the tool changes, its EDAM terms change, and therefore our recommendation for the most appropriate tool can change accordingly. This software layer implements a *recommendation system*, which empowers the user to decide on specific modules of the workflow to run against the provided input data. The result will be an expert driven, tailor-made workflow to perform the most appropriate computational data analysis.

In order to enhance the usability, the workflows can be showcased by means of a Galaxy Tour: an interactive guide that illustrates how the main components of the workflow connect in relation to real-life user tasks (Grüning et al., 2017).

3. Pitfalls of RNA-Seq data analyses

In this section we will showcase the importance of workflow development for differential gene expression analysis of standard RNA-Seq protocols (independent of rRNA depletion or poly-A selection methods during library preparation). This workflow will be modularly implemented by de.STAIR, taking care of parameter settings for different tools.

3.1. Draft of an RNA-Seq data processing workflow

Sequencing technologies that involve DNA amplification steps, including RNA-Seq analyes, can cause asymmetric sequence amplification due to the inherent GC bias (Aird et al., 2011; Chen et al., 2013). A highly efficient approach to recalculate the initial real number of transcripts after the amplification and sequencing step, is called "digital RNA sequencing" (Shiroguchi et al., 2012). This procedure combines the specific tagging of sequences with unique barcodes with a distinct strategy for post-processing analysis. The tagging takes place before the amplification is performed, whereas the unique barcodes are counted after the sequencing to retrieve the original number of transcripts (Shiroguchi et al., 2012).

Assuming demultiplexed RNA-Seq data, the processing usually starts with raw reads provided in fastq format. In addition to the sequence of nucleotides, the fastq format also provides a quality value, i.e. a Phred score, for each of the sequenced bases. In the first step, an evaluation of these quality values as well as the calculation of the GC content, read duplication levels and contaminations are crucial for any further analyses. The quality visualization tools FastQC ("Babraham

Bioinformatics – FastQC A Quality Control tool for High Throughput Sequence Data," n.d.) or NGS QC Toolkit (Patel and Jain, 2012) calculate multiple quality statistics for read data which can be used to adjust parameters for downstream analysis. For example, possible remaining adapter sequences can be detected from a basic *k-mer* analysis or overrepresented sequences.

While the adapter sequence should always be documented and thus known to anyone who works with a specific sequencing dataset, in practice this is rarely the case. In such scenarios, adapters may also be automatically predicted using DNApi (Tsuji and Weng, 2016). To clip them off, various tools such as fastx_clipper from the FASTX-Toolkit, Cutadapt (Martin, 2011), Skewer (Jiang et al., 2014) and Reaper from the Kraken package (Wood and Salzberg, 2014) can be used. Often adapter clippers are already integrated into trimming software like PrinSeq (Schmieder and Edwards, 2011), Trimmomatic (Bolger et al., 2014) and ConDeTri (Smeds and Künstner, 2011). After the removal of adapters, a quality trimming step is recommendable. Removing low quality parts of a read, such as homopolymers, improves the reliability of downstream analysis. Despite the fact that most of the tools for quality trimming use an almost similar approach, usually a sliding window, they have been shown to perform quite differently. Thus, an adjustment of parameters and thresholds is often necessary to obtain optimal results. In a study performed with reads from Illumina sequencing experiments, the tools Trimmomatic, ConDeTri or Sickle (najoshi, n.d.) perform best with loose cutoffs (average Phred score ~20), whereas for others, e.g. PrinSeq, a more strict cutoff (average Phred score ~25) needs to be used. The chosen quality threshold is crucial for maximizing the relative number of trimmed reads alignable to the reference and to increase specificity for, e.g., SNP calling (Del Fabbro et al., 2013). In order to perform trimming and mapping, the encoding of the Phred scores is necessary to be known. While classic Sanger sequences, as well as Illumina sequences (CASAVA ≥ 1.8) are usually encoded in the so called Phred + 33 scheme, Solexa (Phred + 59) and older Illumina sequences (Phred + 64) often need to be transcoded (commonly known as conversion from Illumina to Sanger) by fastq_quality_converter ("FASTX-Toolkit," n.d.) or EMBOSS (Rice et al., 2000) to work with today's software.

Over the last decade several read alignment algorithms were developed to replace traditional sequence aligners like BLAST (Altschul et al., 1990) and BLAT (Kent, 2002), which are limited in dealing with huge amounts of sequencing data. Furthermore, most of the state-of-the-art mapping algorithms take care of intronic regions and allow split-read alignments. Some of the most popular tools are BWA (Li and Durbin, 2009), Bowtie2 (Langmead and Salzberg, 2012), TopHat2 (Kim et al., 2013) followed up by HiSat2 (Kim et al., 2015). All of them are based on Burrows-Wheeler transform methods and seed-extend based mapping techniques. Other aligners like STAR implement suffix arrays as index of the reference for efficient mapping (Dobin et al., 2013). Using a similar approach, Segemehl is a multi-split-read aligner based on enhanced suffix arrays, which is capable of processing InDels during the seed search and thus is suitable for mapping also short or contaminated reads and can be subsequently used to detect circular RNAs (Hoffmann et al., 2014, 2009). Therefore, reads with increased error rates towards their 3′ ends or biases in the nucleotide composition can still be mapped using Segemehl (Hansen et al., 2010). In an exhaustive study 11 different alignment programs were reviewed regarding accuracy/mismatch-frequency, splice site detection and performance. The underlying algorithms were described to either truncate reads or allow for mismatches and that mapping performance or accuracy in splice site detection is lost for higher mapping rates and increased sensitivity (Engström et al., 2013). More recent methods, including Kallisto (Bray et al., 2016), Sailfish (Patro et al., 2014) and Salmon (Patro et al., 2017) for the quantification of RNA-Seq reads follow the tendency to use 'alignment free' quantification methods for faster and resource-sparing RNA-Seq analysis. Such quasi-mapping techniques are based on light-weight-alignment- (Salmon) or pseudo-alignment-algorithms (Kallisto) and efficiently use the structure of a reference sequence without performing full base-to-base alignments. This allows reporting of all potential alignments (multi-mappings) without increased running time as compared to other modern mappers. To infer likely seed positions of reads, fragment mapping information can be obtained from a reference indexed only once, using suffix arrays or approximately matching paths in a De Bruijn graph and other efficient data structures like k-mer hashes (Srivastava et al., 2016). Quasi-mapping lends itself for coarsed tasks like transcript quantifications, clustering and isoform prediction, thereby performing with similar accuracy to traditional approaches (Robert and Watson, 2015).

This variety of options underscores the need to choose appropriate aligners for specific research questions. Aiming for non-coding transcript identification, especially miRNAs, reads should preferably not accumulate mismatches in seed regions, but can be truncated. Furthermore, most mapping tools allow for a multiple or unique mapping strategy. Regarding the first strategy, reads may be aligned to multiple regions such as domain sharing paralogous genes and pseudogenes as well as regions of low complexity in genomic references or numerous isoforms in transcriptomic references (Conesa et al., 2016). While transcript quantification based on multi-mapped reads for non-coding RNA (ncRNA) classes sharing similar sequences is advisable, it is generally not for other genes because the quantification becomes more challenging and the base wise accuracy decreases.

Other important questions with consequences for the data analysis arise from the species under investigation. de.STAIR aims to suggest suitable tools and proper default parameter settings.

3.2. Differential gene expression analysis

Downstream analysis aims at solving a wide range of questions such as the detection of differentially expressed genes, splice isoforms, and identification of up- and downregulated pathways or single nucleotide variant (SNV) enrichments. Differential gene expression analysis tools often start with (normalized) per gene read counts from different RNA-Seq samples. Existing tools for read quantification differ in terms of counting strategy, parametrization and runtime. For example, the widely used software HTSeq-count (Anders et al., 2015) counts reads and split read fragments in a rather static manner and offers comparably few parameters. Other tools such as RNAcounter and featureCounts (Liao et al., 2014) have a higher level of flexibility and are usually faster.

To compare two or more samples, it is essential to take into account varying library sizes and differing transcript lengths caused by multiple isoforms or SECIS element activity (Conesa et al., 2016). To remove such biases, a number of different measures for the normalized quantifications of reads such as the "reads/fragments per kilobase million (R/FPKM)" or "transcripts per million (TPM)" have been introduced. Meanwhile, a number of tools for the detection of differentially expressed genes have been published. The softwares edgeR (Robinson et al., 2010) and DESeq2 (Love et al., 2014) are using a similar statistical approach for variance stabilization transformation. This transformation becomes necessary as the variance of read counts essentially grows and thus depends on the number of counted reads. DESeq2 may be adjusted to take higher variances of non-differentially expressed genes into account as deduced from replicate samples collected from different species or patients. An alternative tool for the detection of differentially expressed genes is Cuffdiff (Trapnell et al., 2013). Rapaport et al. observed a reduced sensitivity compared to various other tools and concluded that the Cufflinks specific normalization process, including alternative isoform expression and transcript lengths, may be a possible reason (Rapaport et al., 2013). In a recent study to investigate host-pathogen interactions with a special focus on expressed long ncRNAs (lncRNAs), an additional filter for differentially expressed genes was proposed (Klassert et al., 2017; Riege et al., 2017). The authors concluded that minimum TPM values ought to be used to obtain a

good set of significantly expressed genes, which are able to eliminate potential biases due to transcript lengths in normalized read counts. To further facilitate the integration of RNA-Seq experiments with other data such as those derived from the analysis of epigenetic modifications or transcription factor binding sites, de.STAIR is implementing tools and workflows to quickly identify differentially methylated regions even in larger datasets (Jühling et al., 2016) and to integrate these regions with the results from RNA-Seq experiments, e.g., to obtain correlated differentially methylated regions or significant transcription factor (binding site) alterations.

4. From the identification of regulatory RNAs in bacteria and their targets to the characterization of natural microbial communities

Bacterial regulatory small RNAs (sRNAs) are crucial for the post-transcriptional regulation of gene expression. Indeed, these are involved in almost all responses to environmental changes (Marchfelder and Hess, 2012). Bacterial sRNAs mediate cross-regulation between bacterial mRNAs because one mRNA can be targeted by multiple sRNAs, as known for most sRNAs and most sRNAs have multiple targets (Bossi and Figueroa-Bossi, 2016). However, sRNA genes are not commonly annotated during genome analysis and the identification of their targets requires substantial additional effort. In this chapter, we describe state-of-the art approaches to deal with these issues.

4.1. The identification of sRNAs in bacteria by computational prediction within datasets generated by differential RNA-Seq (dRNA-Seq)

Major problems arise from the fact that regulatory RNAs in bacteria are extremely heterogeneous. Their length varies between 40 nt for the *Escherichia coli* sRNA tpke70 (Hershberg et al., 2003) and more than 800 nt long for the *Salmonella* sRNA STnc510 (Sittka et al., 2008). Even the very conception of a regulatory RNA as being non-coding has been challenged with the discovery of dual function sRNAs, i.e., sRNAs that have a regulatory function and also encode a functional peptide or small protein (Gimpel and Brantl, 2017). Examples include the RNAIII of *Staphylococcus aureus*, which is a regulatory RNA of 514 nt and encodes the 26 amino acid δ hemolysin (Benito et al., 2000; Boisset et al., 2007; Vandenesch et al., 2012) or the 227 nt SgrS sRNA of enteric bacteria that encodes the 43 amino acid functional polypeptide SgrT (Wadler and Vanderpool, 2007). Moreover, regulatory RNAs may derive by processing from larger mRNAs, UTRs as well as ncRNAs (Chao et al., 2012; De Lay and Garsin, 2016; Lalaouna et al., 2015; Miyakoshi et al., 2015) and include regulatory elements such as marooned riboswitches (De Lay and Garsin, 2016).

Computational approaches for the prediction of bacterial sRNAs and their genes can be classified into *de novo* approaches and approaches that utilize comparative information. *De novo* approaches may combine the search for promoters, specific transcription factor binding sites and Rho-independent terminators in intergenic regions (Argaman et al., 2001; Chen et al., 2002; Lenz et al., 2004). For *trans*-encoded sRNAs, comparative strategies start with a conserved sequence from an intergenic region. Then, homologs from closely related species are clustered and compared in pairwise or multiple alignments, which are subsequently scored according to predicted RNA structural features. These include thermodynamic stability values derived from the consensus folding of aligned sequences, e.g. by using the tool RNAz (Washietl et al., 2005; Washietl and Hofacker, 2004). Based on such strategies, sRNAs were predicted for different sets of closely related model cyanobacteria (Ionescu et al., 2010; Voss et al., 2009a), in Rhizobiales (Madhugiri et al., 2012; Voss et al., 2009b), in the haloarchaeon *Haloferax volcanii* (Babski et al., 2011) and for picocyanobacteria of the *Prochlorococcus-Synechococcus* lineage (Axmann et al., 2005). Using the NcDNAlign algorithm (Rose et al., 2008) sRNAs were successfully predicted in *Pseudomonas* (Sonnleitner et al., 2008), relying on the usage of comparative information from 10 genomes. Comparative approaches are widely used to increase the reliability of a prediction due to the underlying statistical possibilities. By considering this principle, e.g. SIPHT (Livny, 2012), Infernal (Nawrocki et al., 2009; Nawrocki and Eddy, 2013a) and RNAlien (Eggenhofer et al., 2016) have been developed. In this context, “comparative” does not only mean the comparison of instantly generated sequences identified by an alignment tool of choice, but can also relate to well-annotated knowledge as existing, e.g., in the Rfam database (Nawrocki et al., 2014).

The annotation of ncRNAs enables researchers to carry out extended functional and differential expression studies. Therefore, new members of RNA families as provided by the Rfam database (Nawrocki et al., 2014) can be identified by secondary structure constrained homology searches. For this task, the GORAP pipeline is being developed (rna-hta-jena, n.d.), which is mainly based on the Infernal package (Nawrocki et al., 2009; Nawrocki and Eddy, 2013b) and uses multiple in-house filters for taxonomic information, RNA family specific thresholds, structure and sequence properties, which can be governed by the user. GORAP comprises modular, specialized software for the detection of specific RNA families: tRNAscan-SE (Lowe and Eddy, 1997) for tRNAs, RNAmmer (Lagesen et al., 2007) for rRNAs, Bcheck (Yusuf et al., 2010) for RNase P RNAs and CRT (Bland et al., 2007) for CRISPR RNAs. It was successfully applied on different whole genome assemblies for bacterial species (Lechner et al., 2014; Möbius et al., 2015; Sachse et al., 2014) and fungi (Linde et al., 2015; Schwartze et al., 2014).

The most powerful approach for the detection of bacterial sRNAs relies on RNA-Seq and especially dRNA-Seq (Sharma et al., 2010). Here, it is advantageous to generate dRNA-Seq data together with a parallel classical RNA-Seq approach. Then, workflows such as the “TSS annotation regime” (TSSAR) (Amman et al., 2014) can be applied, which utilizes both types of datasets and is considering the local expression rate from RNA-Seq and associates peaks from dRNA-Seq to define TSSs. With the aid of the corresponding genome annotation file, the TSSs then can be classified as (1) gTSS (gene TSS: TSS of an annotated gene) and/or (2) aTSS (antisense TSS: located on the reverse strand of an annotated gene) and/or (3) iTSS (internal TSS: located within an annotated gene) or (4) oTSS or nTSS (for orphan or non-coding TSS: e.g., sRNAs transcribed from an intergenic region) (Čuklina et al., 2016).

The majority of published bacterial genome sequences were automatically annotated by computational services like RAST (the Rapid Annotation Server) (Aziz et al., 2008) or Prokka (Seemann, 2014). These annotation regimes provide rapid insight into the composition and arrangement of genes or regulatory elements for a given genome. Additional information from dRNA-Seq can help to improve the existing annotation by correcting the 5′ ends of modelled genes and adding precise information on the TSSs (Čuklina et al., 2016). However, by combining the information from dRNA-Seq and RNA-Seq it is possible to define transcriptional units (TU) based on real expression data, which is very advantageous for the identification of operons, the full lengths of sRNAs and asRNAs, as well as the identification of divergent but overlapping transcripts due to alternative TSSs or maturation events. Such TUs can be efficiently identified using the software package “RNASEG” (Bischler et al., 2014).

During computer-aided data analysis and the experimental setup, several pitfalls should be kept in mind. So, a correct parameter adjustment to optimize estimation numbers of TSS, respectively TU prediction, is important for the discussed tools.

4.2. Analysis of sRNAs in natural microbial communities

Metatranscriptomic differential RNA-Seq (mdRNA-Seq) as well as metatranscriptomic RNA-Seq (metaRNA-Seq) are protocols to analyze highly complex transcription compositions of biological communities. Hou et al. (Hou et al., 2016) proposed a bioinformatic workflow for mdRNA-Seq/metaRNA-Seq data to perform taxonomic assignments, prediction of TSSs (including classification into gTSS, iTSS, aTSS and oTSS) and the analysis of associated promoter sequences and regulatory

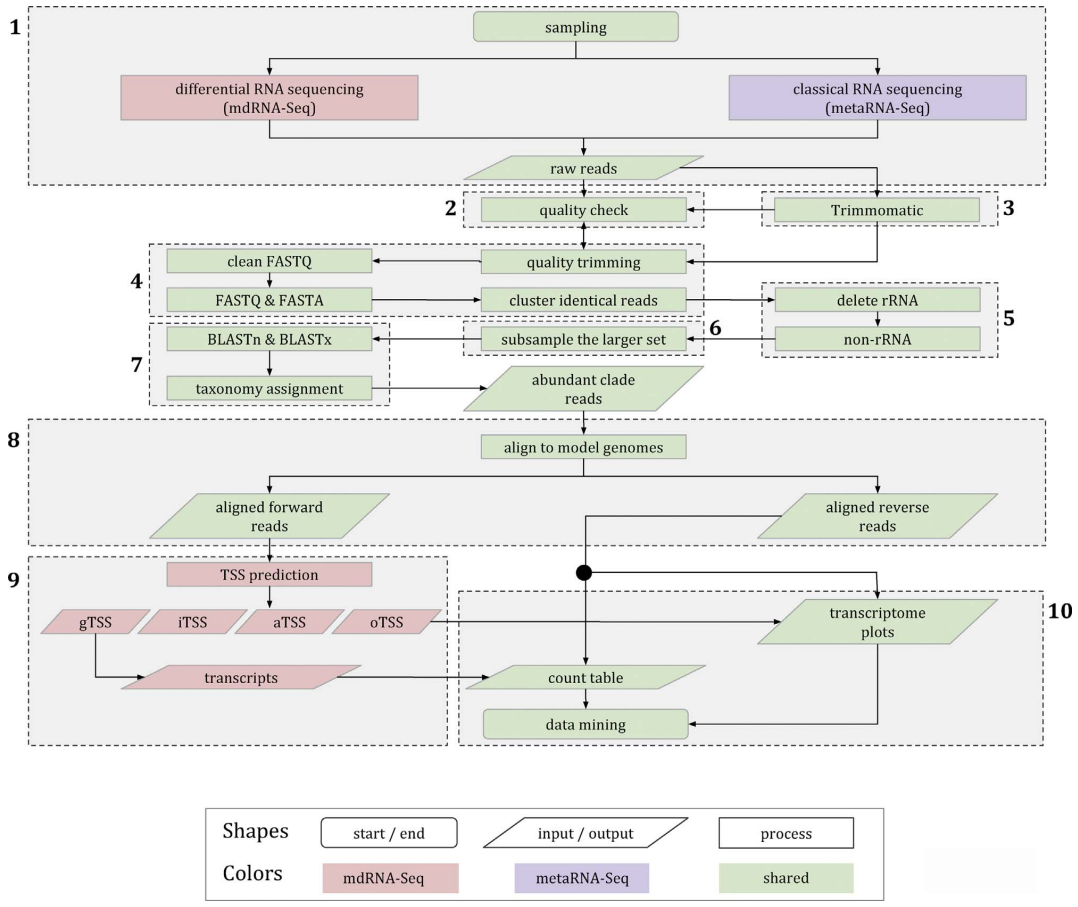


Fig. 3. Bioinformatic workflow for mdRNA-Seq analysis (modified from Hou et al., 2016). **Step 1,** the sampled data as mdRNA-Seq and metaRNA-Seq are used as input and both are quality controlled in **Step 2** using FastQC and in **Step 3** become trimmed and barcode/adapter removed using Trimmomatic (Bolger et al., 2014). **Step 4,** the modified data are transferred from FASTQ to FASTA and identical reads are clustered. **Step 5** divides the sequences into two groups "non-rRNA" and "rRNA" by using SortMeRNA (Kopylova et al., 2012) and only the "non-rRNA" reads are used for further analysis. **Step 6,** through the sequencing and quality filtering, the two datasets (mdRNA-Seq/metaRNA-Seq) can differ in the number of reads. The subsampling procedure is tackling this problem to adjust the libraries to identical sizes. **Step 7,** taxonomy assignment – for further analysis the reads are assigned to their common lowest ancestor (BLASTn/BLASTx). **Step 8**, the pre-grouped reads of interest, e.g., reads assigned to *Prochlorococcus* are used for a reassignment step against a well annotated version of a *Prochlorococcus* model genome. **Step 9,** with the aid of the mdRNA-Seq information, a TSS prediction is performed as well as classified into gTSS, iTSS, aTSS and oTSS. **Step 10,** by considering the TSS prediction and the complete sequencing information, a transcriptome plot can be drawn.

motifs that are finally complemented by further functional analyses utilizing KEGG. This workflow was successfully applied to analyze data coming from the northern Gulf of Aqaba in the Red Sea (Hou et al., 2016). By using this workflow the authors located genome-wide TSS, regulatory elements in the promoter regions and intergenic regions and improved the genome annotation for several non-model organisms belonging to all three domains of life. Fig. 3 shows a slightly modified and updated workflow for this analysis. For example, "Cutadapt" (Martin, 2011) was replaced by "Trimmomatic" (Bolger et al., 2014), based on its enhanced usability and computational power (see also Skewer for alternative) (Jiang et al., 2014). As input both types of datasets were used to perform a quality curation, followed by a global read assignment approach (BLASTn/BLASTx) and a TSS prediction in combination of all sequences is applied to draw a transcriptome plot.

In the metatranscriptomic analysis of natural microbial populations (Pfreundt et al., 2016a) it is important to consider the community structure, i.e., the numerical relationships among taxa. In this way it is possible to differentiate "active" taxa with high transcriptional activity from "non-active" community members, which are present but show little-to-none gene expression. A workflow has been developed by Pfreundt et al. (2016b) for the community composition analysis based on 16S amplicon quantification using the UPARSE pipeline (Edgar, 2013) and taxonomic classification using the SILVA SSU taxonomy database (Quast et al., 2013).

4.3. Approaches for the prediction of the regulatory targets of bacterial sRNAs

In view of the high number of different sRNAs in any given bacterial genome, the identification of their regulatory targets is critical for their further characterization. Therefore, the reliable computational prediction of sRNA targets has become an important field of research.

The main challenges for reliable predictions are the small number of interacting sequence elements between a sRNA and its frequently distant target mRNA as well as imperfect complementarity, which can reside in various sections of the sRNA. In addition, a few sRNAs have

protein-binding rather than mRNA-binding functions, some targets are recognized by the joint action of two different modules in the interacting RNA molecules (kissing hairpins) and some sRNAs have single or only very few targets whereas others control multiple different mRNAs. There are different approaches to compute interactions between a given sRNA and its target mRNAs. In the following, a selection of sRNA prediction tools will be presented.

RNAplex (Tafer and Hofacker, 2008)

- based on: Energy folding algorithm; fast detection of possible hybridization sites
- main features: Extension of minimum energy folding algorithm to two sequences (Li et al., 2012); up to 10–27 times faster than RNAhybrid (Krüger and Rehmsmeier, 2006)

IntaRNA (Busch et al., 2008)

- based on: Minimization of extended hybridization energy of two interacting RNAs
- main features: Accessibility of binding sites; user-specified seed (Li et al., 2012), freely available web server

RNApredator (Eggenhofer et al., 2011; Tjaden et al., 2006)

- based on: RNAplex
- main features: Target site accessibility (Li et al., 2012); at least three orders of magnitude faster than RNAup or IntaRNA; freely available web server

PAREsnip (Folkes et al., 2012)

- based on: transcriptome, degradome, sRNAome and genome data
- main features: applicable on large as well as on small-scale experiments

TargetRNA2 (Kery et al., 2014; Tjaden et al., 2006)

- based on: Conservation of the sRNA, secondary structure of the sRNA and mRNA target, hybridization energy between the interacting sRNA/mRNA
- main features: freely available web server; identification of *trans*-acting sRNA targets

CopraRNA (Wright et al., 2014, 2013)

- based on: Phylogenetic information, IntaRNA predictions
- main features: freely available web server; p-value; mRNA region plot; sRNA region plot; functional enrichment (DAVID annotation)

During the comparison of the most popular target prediction tools, CopraRNA ranked on top from several perspectives (Pain et al., 2015). CopraRNA is especially good with regard to the number of false-positives (Wright et al., 2013). This can be ascribed to the consideration of multiple sRNA homologs instead of only a single sRNA. At the same time, the need for multiple sRNA orthologs is also a major disadvantage of CopraRNA because, in some cases, the sRNA of interest may be restricted to a single species or, more frequently, the potential homologs are difficult to find. To overcome this disadvantage and to find sRNA homologs in a reliable way and avoiding descriptor based approaches (Macke et al., 2001), the GLASSgo algorithm is being developed. GLASSgo is currently integrated into the web server providing the Freiburg RNA Tools and can be freely used without limitations.

GLASSgo provides an approach to detect, extract and evaluate potential sRNAs from scratch. This workflow works for sequences coming from dRNA-Seq/RNA-Seq as well as mdRNA-Seq/metaRNA-Seq experiments (Fig. 4 – **(a) "Input"**). Followed by a preprocessing step, which contains "Quality Control" with FastQC, "Adapter + Barcode removal" with Trimmomatic (Bolger et al., 2014), "Sequence Trimming" with Trimmomatic (Bolger et al., 2014) and finally "Sequence Mapping" with Segemehl (Hoffmann et al., 2009; Otto et al., 2014) or VSEARCH (Rognes et al., 2016). The last step depends on the used sequencing protocol. For dRNA-Seq/RNA-Seq all transcripts are mapped against a reference genome, whereas mdRNA-Seq/metaRNA-Seq needs a preselection to assign the sequences with respect to their associated genome. At the end of the "Preprocessing" step, a SAM file (Li et al., 2009) is needed for the "TU-Prediction" tool RNASEG (Bischler et al.,

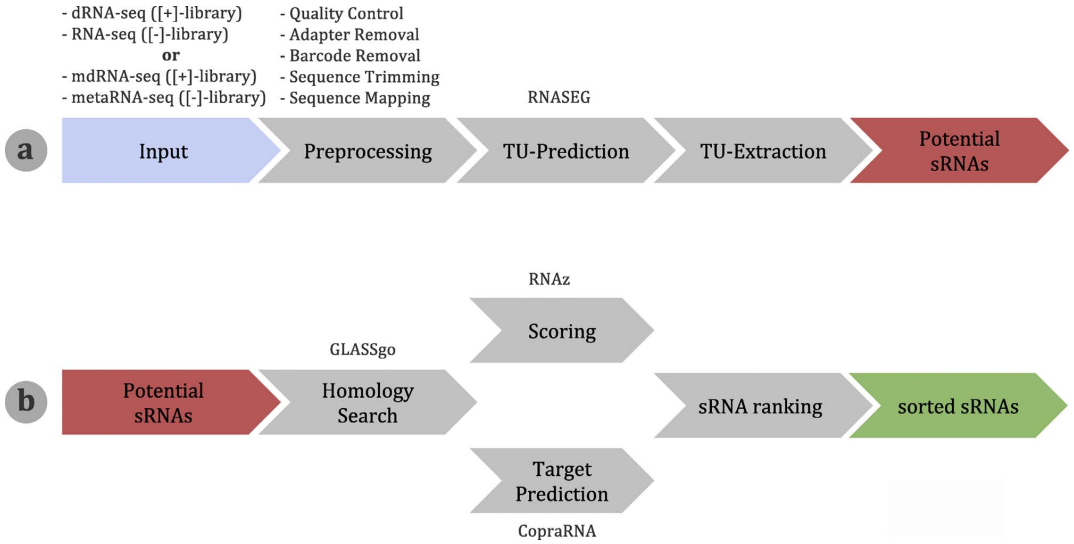


Fig. 4. Workflow to predict intergenic located sRNAs from scratch. **(a)** The first part of the workflow describes the data handling, Transcriptional Units (TU) prediction, extraction of intergenic located TUs and the setup of the potential sRNAs list. **(b)** The second part focuses on the evaluation of the previously extracted sRNAs. It is based on a homology search to enhance the evidence by using comparative approaches like RNAz (Gruber et al., 2009) and CopraRNA (Wright et al., 2014, 2013).

2014). Under the condition that the reads are Poisson distributed, Bischler et al. tried to define sharp borders between the start and end site of a transcript. This is called a "Transcriptional Unit" (TU) and the interval (genome coordinates + strand orientation) of a TU can be extracted from the final result table. The suggested workflow was designed to predict intergenic located sRNAs and therefore the "TU-Extraction" procedure takes only these types of TUs into account. The first part (a) of the workflow (Fig. 4) is available and the potential set of sRNA TUs serve as input for the second part (b).

Each predicted potential sRNA TU is used as query to perform a "Homology Search" with GLASSgo. It returns a trustworthy set of homologs and the query sequence itself as FASTA format. These sets are analyzed independently with RNAz (Gruber et al., 2009; Washietl et al., 2005) as well as CopraRNA (Wright et al., 2014, 2013) and finally the outcomes of both algorithms are correlated to set up a descend ranked table (best to worst). The sRNA candidates with the highest ranked potential among the "sorted sRNAs" can be used to carry out experimental tests.

5. The consortium

The members of the de.STAIR consortium "Structured Analysis and Integration of RNA-Seq Experiments" aim at supporting the research community with tools and workflows to enhance the overall integration of transcriptomic data towards additional regulative, predictive and annotation potential. To enable maximum suitability, interconnectivity, and accessibility for the developed approaches and services, de.STAIR provides dedicated training programs and materials for bioinformaticians and other life scientists and, ultimately, is lowering the bars to RNA-Seq data analysis as a whole. These aims are supported by the development of tools for the analysis of gene regulatory networks as well as the prediction and identification of miRNA-RNA interactions (TriplexRNA) based on high throughput data (Amirkhah et al., 2015; Khan et al., 2014; Lai et al., 2013; Schmitz et al., 2016; Schmitz and Wolkenhauer, 2016), the development of transparent and automated pipelines for RNA-Seq (Wolfien et al., 2016) and mdRNA-Seq analysis (Hou et al., 2016) as well as the ongoing integration of specific tools with the the existing RNA workbench of the RBC.

Funding

Financial support for this work by the German Federal Ministry for Education and Research (BMBF) program de.NBI-Partner (Grant 031L0106, 02NUK043C) is greatly acknowledged.

Acknowledgments

We acknowledge the partners and management of the German Network for Bioinformatics Infrastructure (de.NBI) for continuous support and guidance. In addition, we want to thank our past, present and future users who utilize and trust in the provided services.

References

Adiconis, X., Borges-Rivera, D., Satija, R., DeLuca, D.S., et al., 2013. Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nat. Methods 10, 623–629.
Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., et al., 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 44, W3–W10.
Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., et al., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. 12, R18.
Aldridge, S., Hadfield, J., 2011. Introduction to miRNA profiling technologies and cross-platform comparison. Methods in Molecular Biology. pp. 19–31.
Altschul, S.F., Gish, W., Miller, W., Myers, E.W., et al., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Amirkhah, R., Schmitz, U., Linnebacher, M., Wolkenhauer, O., et al., 2015. MicroRNA-mRNA interactions in colorectal cancer and their role in tumor progression. Genes. Chromosomes Cancer 54, 129–141.
Amman, F., Wolfinger, M.T., Lorenz, R., Hofacker, I.L., et al., 2014. TSSAR: TSS annotation regime for dRNA-seq data. BMC Bioinf. 15, 89.
Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq – a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169.
Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., et al., 2001. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. Curr. Biol. 11, 941–950.
Axmann, I.M., Kensche, P., Vogel, J., Kohl, S., et al., 2005. Identification of cyanobacterial non-coding RNAs by comparative genome analysis. Genome Biol. 6, R73.
Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., et al., 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genom. 9, 75.
Babraham Bioinformatics – FastQC A Quality Control tool for High Throughput Sequence Data [WWW Document]. URL http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. (Accessed 14 February 2017).
Babski, J., Tjaden, B., Voss, B., Jellen-Ritter, A., et al., 2011. Bioinformatic prediction and experimental verification of sRNAs in the haloarchaeon *Haloferax volcanii*. RNA Biol. 8, 806–816.
Babski, J., Haas, K.A., Näther-Schindler, D., Pfeiffer, F., et al., 2016. Genome-wide identification of transcriptional start sites in the haloarchaeon *Haloferax volcanii* based on differential RNA-Seq (dRNA-Seq). BMC Genom. 17, 629.
Benito, Y., Kolb, F.A., Romby, P., et al., 2000. Probing the structure of RNAIII, the *Staphylococcus aureus agr* regulatory RNA, and identification of the RNA domain involved in repression of protein A expression. RNA 6, 668–679.
Berthold, M.R., Cebron, N., Dill, F., Gabriel, et al., 2009. KNIME – the Konstanz information miner: version 2.0 and beyond. ACM SIGKDD Explor. Newsl. 11, 26–31.
Bischler, T., Kopf, M., Voß, B., 2014. Transcript mapping based on dRNA-seq data. BMC Bioinf. 15, 122.
Bland, C., Ramsey, T.L., Sabree, F., Lowe, M., et al., 2007. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. BMC Bioinf. 8, 209.
Boettiger, C., 2015. An introduction to Docker for reproducible research. Oper. Syst. Rev. 49, 71–79.
Boisset, S., Geissmann, T., Huntzinger, E., Fechter, P., et al., 2007. *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. Genes Dev. 21, 1353–1366.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.
Borries, A., Vogel, J., Sharma, C.M., 2012. Differential RNA sequencing (dRNA-Seq): deep-sequencing-based analysis of primary transcriptomes. Tag-Based Next Generation Sequencing. pp. 109–121.
Bossi, L., Figueroa-Bossi, N., 2016. Competing endogenous RNAs: a target-centric view of small RNA regulation in bacteria. Nat. Rev. Microbiol. 14, 775–784.
Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34, 525–527.
Busch, A., Richter, A.S., Backofen, R., 2008. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. Bioinformatics 24, 2849–2856.
Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C.M., et al., 2012. An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. EMBO J. 31, 4005–4019.
Chen, S., Lesnik, E.A., Hall, T.A., Sampath, R., et al., 2002. A bioinformatics based approach to discover small RNA genes in the Escherichia coli genome. Biosystems 65, 157–177.
Chen, Y.-C., Liu, T., Yu, C.-H., Chiang, T.-Y., et al., 2013. Effects of GC bias in next-generation-sequencing data on de novo genome assembly. PLoS One 8, e62856.
Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., et al., 2016. A survey of best practices for RNA-seq data analysis. Genome Biol. 17, 13.
Čuklina, J., Hahn, J., Imakaev, M., Omasits, U., et al., 2016. Genome-wide transcription start site mapping of *Bradyrhizobium japonicum* grown free-living or in symbiosis – a rich resource to identify new transcripts, proteins and to study gene regulation. BMC Genom. 17, 302.
De Lay, N.R., Garsin, D.A., 2016. The unmasking of junk RNA reveals novel sRNAs: from processed RNA fragments to marooned riboswitches. Curr. Opin. Microbiol. 30, 16–21.
Del Fabbro, C., Scalabrin, S., Morgante, M., Giorgi, F.M., 2013. An extensive evaluation of read trimming effects on Illumina NGS data analysis. PLoS One 8, e85024.
de la Garza, L., Veit, J., Szolek, A., Röttig, M., et al., 2016. From the desktop to the grid: scalable bioinformatics via workflow conversion. BMC Bioinf. 17, 127.
Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., et al., 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602–607.
Di Tommaso, P., Palumbo, E., Chatzou, M., Prieto, P., et al., 2015. The impact of Docker containers on the performance of genomic pipelines. PeerJ 3, e1273.
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., et al., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.
Edgar, R.C., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods 10, 996–998.
Eggenhofer, F., Tafer, H., Stadler, P.F., Hofacker, I.L., 2011. RNApredator: fast accessibility-based prediction of sRNA targets. Nucleic Acids Res. 39, W149–54.
Eggenhofer, F., Hofacker, I.L., Höner Zu Siederdissen, C., 2016. RNAlien – unsupervised RNA family model construction. Nucleic Acids Res. 44, 8433–8441.
Engström, P.G., Steijger, T., Sipos, B., Grant, G.R., et al., 2013. Systematic evaluation of spliced alignment programs for RNA-seq data. Nat. Methods 10, 1185–1191 RGASP Consortium.
FASTX-Toolkit [WWW Document]. URL http://hannonlab.cshl.edu/fastx_toolkit/index.html. (Accessed 14 February 2017).
Folkes, L., Moxon, S., Woolfenden, H.C., Stocks, M.B., et al., 2012. PAREsnip: a tool for rapid genome-wide discovery of small RNA/target interactions evidenced through

degradome sequencing. Nucleic Acids Res. 40, e103.
Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., et al., 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80.
Georg, J., Hess, W.R., 2011. cis-Antisense RNA, another level of gene regulation in bacteria. Microbiol. Mol. Biol. Rev. 75, 286–300.
Gimpel, M., Brantl, S., 2017. Dual-function small regulatory RNAs in bacteria. Mol. Microbiol. 103, 387–397.
Gorski, S.A., Vogel, J., Doudna, J.A., 2017. RNA-based recognition and targeting: sowing the seeds of specificity. Nat. Rev. Mol. Cell Biol. 18, 215–228.
Grüning, B.A., Fallmann, J., Yusuf, D., Will, S., et al., 2017. The RNA workbench: best practices for RNA and high-throughput sequencing bioinformatics in Galaxy. Nucleic Acids Res Epub ahead of print.
Granados Moreno, P., Joly, Y., Knoppers, B.M., 2017. Public-private partnerships in cloud-computing services in the context of genomic research. Front. Med. 4, 3.
Griebel, L., Prokosch, H.-U., Köpcke, F., Toddenroth, D., et al., 2015. A scoping review of cloud computing in healthcare. BMC Med. Inform. Decis. Mak. 15, 17.
Gruber, A.R., Findeiß, S., Washietl, S., Hofacker, I.L., et al., 2009. RNAz 2.0. Biocomputing 2010. pp. 69–79.
Hansen, K.D., Brenner, S.E., Dudoit, S., 2010. Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res. 38, e131.
Hershberg, R., Altuvia, S., Margalit, H., 2003. A survey of small RNA-encoding genes in Escherichia coli. Nucleic Acids Res. 31, 1813–1820.
Hoffmann, S., Otto, C., Kurtz, S., Sharma, C.M., et al., 2009. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. PLoS Comput. Biol. 5, e1000502.
Hoffmann, S., Otto, C., Doose, G., Tanzer, A., et al., 2014. A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. Genome Biol. 15, R34.
Hou, S., Pfreundt, U., Miller, D., Berman-Frank, I., et al., 2016. mdRNA-Seq analysis of marine microbial communities from the northern Red Sea. Sci. Rep. 6, 35470.
Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., et al., 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat. Protoc. 7, 1534–1550.
Ionescu, D., Voss, B., Oren, A., Hess, W.R., et al., 2010. Heterocyst-specific transcription of NsiR1, a non-coding RNA encoded in a tandem array of direct repeats in cyanobacteria. J. Mol. Biol. 398, 177–188.
Ison, J., Kalas, M., Jonassen, I., Bolser, D., et al., 2013. EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats. Bioinformatics 29, 1325–1332.
Ison, J., Rapacki, K., Ménager, H., Kalaš, M., et al., 2016. Tools and data services registry: a community effort to document bioinformatics resources. Nucleic Acids Res. 44, D38–D47.
Jäger, D., Sharma, C.M., Thomsen, J., Ehlers, C., et al., 2009. Deep sequencing analysis of the *Methanosarcina mazei* Gö1 transcriptome in response to nitrogen availability. Proc. Natl. Acad. Sci. U. S. A. 106, 21878–21882.
Jégu, T., Aeby, E., Lee, J.T., 2017. The X chromosome in space. Nat. Rev. Genet. 18, 377–389.
Jühling, F., Kretzmer, H., Bernhart, S.H., Otto, C., et al., 2016. metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. Genome Res. 26, 256–262.
Jiang, H., Lei, R., Ding, S.-W., Zhu, S., 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinf. 15, 182.
Köster, J., Rahmann, S., 2012. Snakemake – a scalable bioinformatics workflow engine. Bioinformatics 28, 2520–2522.
Kallio, M.A., Tuimala, J.T., Hupponen, T., Klemelä, P., et al., 2011. Chipster: user-friendly analysis software for microarray and other high-throughput data. BMC Genomics 12, 507.
Kent, W.J., 2002. BLAT – the BLAST-like alignment tool. Genome Res. 12, 656–664.
Kery, M.B., Feldman, M., Livny, J., Tjaden, B., 2014. TargetRNA2: identifying targets of small regulatory RNAs in bacteria. Nucleic Acids Res. 42, W124–9.
Khan, F.M., Schmitz, U., Nikolov, S., Engelmann, D., et al., 2014. Hybrid modeling of the crosstalk between signaling and transcriptional networks using ordinary differential equations and multi-valued logic. Biochim. Biophys. Acta 1844, 289–298.
Kim, D., Pertea, G., Trapnell, C., Pimentel, H., et al., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36.
Kim, D., Langmead, B., Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360.
Klassert, T.E., Bräuer, J., Hölzer, M., Stock, M., et al., 2017. Differential effects of vitamins A and D on the transcriptional landscape of human monocytes during infection. Sci. Rep. 7, 40599.
Kopf, M., Klähn, S., Scholz, I., Matthiessen, J.K.F., et al., 2014. Comparative analysis of the primary transcriptome of Synechocystis sp. PCC 6803. DNA Res. 21, 527–539.
Kopf, M., Klähn, S., Scholz, I., Hess, W.R., et al., 2015a. Variations in the non-coding transcriptome as a driver of inter-strain divergence and physiological adaptation in bacteria. Sci. Rep. 5, 9560.
Kopf, M., Möke, F., Bauwe, H., Hess, W.R., et al., 2015b. Expression profiling of the bloom-forming cyanobacterium *Nodularia* CCY9414 under light and oxidative stress conditions. ISME J. 9, 2139–2152.
Kopylova, E., Noé, L., Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28, 3211–3217.
Kröger, C., Dillon, S.C., Cameron, A.D.S., Papenfort, K., et al., 2012. The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. Proc. Natl. Acad. Sci. U. S. A. 109, E1277–86.
Krüger, J., Rehmsmeier, M., 2006. RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res. 34, W451–4.
Kukurba, K.R., Montgomery, S.B., 2015. RNA sequencing and analysis. Cold Spring Harb. Protoc. 2015, 951–969.
Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.-H., et al., 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35, 3100–3108.
Lai, X., Schulz, C., Seifert, F., Dolniak, B., et al., 2013. The role of microRNA regulation in the early inflammatory response: miR-146a and NF-κB signaling in lung inflammation. Pneumologie 67.
Lalaouna, D., Carrier, M.-C., Semsey, S., Brouard, J.-S., et al., 2015. A 3′ external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise. Mol. Cell 58, 393–405.
Lampa, S., Dahlö, M., Olason, P.I., Hagberg, J., et al., 2013. Lessons learned from implementing a national infrastructure in Sweden for storage and analysis of next-generation sequencing data. Gigascience 2, 9.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.
Lechner, M., Nickel, A.I., Wehner, S., Riege, K., et al., 2014. Genomewide comparison and novel ncRNAs of Aquificales. BMC Genom. 15, 522.
Lenz, D.H., Mok, K.C., Lilley, B.N., Kulkarni, R.V., et al., 2004. The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. Cell 118, 69–82.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., et al., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.
Li, W., Ying, X., Lu, Q., Chen, L., 2012. Predicting sRNAs and their targets in bacteria. Genom. Proteom. Bioinform. 10, 276–284.
Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930.
Linde, J., Duggan, S., Weber, M., Horn, F., et al., 2015. Defining the transcriptomic landscape of *Candida glabrata* by RNA-Seq. Nucleic Acids Res. 43, 1392–1406.
Livny, J., 2012. Bioinformatic discovery of bacterial regulatory RNAs using SIPHT. Methods Mol. Biol. 905, 3–14.
Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.
Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.
Möbius, P., Hölzer, M., Felder, M., Nordsiek, G., et al., 2015. Comprehensive insights in the *Mycobacterium avium* subsp. *paratuberculosis* genome using new WGS data of sheep strain JIII-386 from Germany. Genome Biol. Evol. 7 (9), 2585–2601.
Macke, T.J., Ecker, D.J., Gutell, R.R., Gautheret, D., et al., 2001. RNAMotif, an RNA secondary structure definition and search algorithm. Nucleic Acids Res. 29, 4724–4735.
Madhugiri, R., Pessi, G., Voss, B., Hahn, J., et al., 2012. Small RNAs of the *Bradyrhizobium/Rhodopseudomonas* lineage and their analysis. RNA Biol. 9, 47–58.
Marchfelder, A., Hess, W., 2012. Regulatory RNAs in Prokaryotes. Springer Science & Business Media.
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10.
Mitschke, J., Georg, J., Haas, I., Sharma, C.M., et al., 2011a. An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. Proc. Natl. Acad. Sci. U. S. A. 108, 2124–2129.
Mitschke, J., Vioque, A., Haas, F., Hess, W.R., et al., 2011b. Dynamics of transcriptional start site selection during nitrogen stress-induced cell differentiation in *Anabaena* sp. PCC7120. Proc. Natl. Acad. Sci. U. S. A. 108, 20130–20135.
Miyakoshi, M., Chao, Y., Vogel, J., 2015. Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. EMBO J. 34, 1478–1492.
najoshi, GitHub – najoshi/sickle: Windowed Adaptive Trimming for fastq files using quality [WWW Document]. URL https://github.com/najoshi/sickle. (Accessed 15 February 2017).
Nawrocki, E.P., Eddy, S.R., 2013a. Computational identification of functional RNA homologs in metagenomic data. RNA Biol. 10, 1170–1179.
Nawrocki, E.P., Eddy, S.R., 2013b. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29, 2933–2935.
Nawrocki, E.P., Kolbe, D.L., Eddy, S.R., 2009. Infernal 1.0: inference of RNA alignments. Bioinformatics 25, 1335–1337.
Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., et al., 2014. Rfam 12.0: updates to the RNA families database. Nucleic Acids Res. 43, D130–D137.
Otto, C., Stadler, P.F., Hoffmann, S., 2014. Lacking alignments? The next-generation sequencing mapper segemehl revisited. Bioinformatics 30, 1837–1843.
Pain, A., Ott, A., Amine, H., Rochat, T., et al., 2015. An assessment of bacterial small RNA target prediction programs. RNA Biol. 12, 509–513.
Patel, R.K., Jain, M., 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One 7, e30619.
Patro, R., Mount, S.M., Kingsford, C., 2014. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. Nat. Biotechnol. 32, 462–464.
Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., et al., 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14, 417–419.
Pfreundt, U., Kopf, M., Belkin, N., Berman-Frank, I., et al., 2014. The primary transcriptome of the marine diazotroph *Trichodesmium erythraeum* IMS101. Sci. Rep. 4, 6187.
Pfreundt, U., Spungin, D., Bonnet, S., Berman-Frank, I., et al., 2016a. Global analysis of gene expression dynamics within the marine microbial community during the VAHINE mesocosm experiment in the southwest Pacific. Biogeosciences 13, 4135–4149.

Pfreundt, U., Van Wambeke, F., Caffin, M., Bonnet, S., et al., 2016b. Succession within the prokaryotic communities during the VAHINE mesocosms experiment in the New Caledonia lagoon. Biogeosciences 13, 2319–2337.
Podnar, J., Deiderick, H., Huerta, G., Hunicke-Smith, S., 2014. Next-Generation Sequencing RNA-Seq library construction. Curr. Protoc. Mol. Biol. 106 (4.21), 1–19.
Poplawski, A., Marini, F., Hess, M., Zeller, T., et al., 2016. Systematically evaluating interfaces for RNA-seq analysis from a life scientist perspective. Brief. Bioinform. 17, 213–223.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., et al., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–6.
Rapaport, F., Khanin, R., Liang, Y., Pirun, M., et al., 2013. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome Biol. 14, R95.
Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the european molecular biology open software suite. Trends Genet. 16, 276–277.
Riege, K., Hölzer, M., Klassert, T.E., Barth, E., et al., 2017. Massive effect on LncRNAs in human monocytes during fungal and bacterial infections and in response to vitamins A and D. Sci. Rep. 7, 40598.
Robert, C., Watson, M., 2015. Errors in RNA-Seq quantification affect genes of relevance to human disease. Genome Biol. 16, 621.
Rognes, T., Flouri, T., Nichols, B., Quince, C., et al., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584.
Rose, D., Hertel, J., Reiche, K., Stadler, P.F., et al., 2008. NcDNAlign: plausible multiple alignments of non-protein-coding genomic sequences. Genomics 92, 65–74.
rna-hta-jena, GitHub – rna-hta-jena/gorap: Genomewide ncRNA Annotation Pipeline [WWW Document]. URL https://github.com/rna-hta-jena/gorap/. (Accessed 15 February 2017).
Sachse, K., Laroucau, K., Riege, K., Wehner, S., et al., 2014. Evidence for the existence of two new members of the family Chlamydiaceae and proposal of *Chlamydia avium* sp. nov. and *Chlamydia gallinacea* sp. nov. Syst. Appl. Microbiol. 37, 79–88.
Schmidtke, C., Findeiss, S., Sharma, C.M., Kuhfuss, J., et al., 2012. Genome-wide transcriptome analysis of the plant pathogen *Xanthomonas* identifies sRNAs with putative virulence functions. Nucleic Acids Res. 40, 2020–2031.
Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27, 863–864.
Systems Medicine: Methods in Molecular Biology. In: Schmitz, U., Wolkenhauer, O. (Eds.), Springer, New York.
Schmitz, U., Naderi-Meshkin, H., Gupta, S.K., Wolkenhauer, O., et al., 2016. The RNA world in the 21 st century – a systems approach to finding non-coding keys to clinical questions. Brief. Bioinform. 17, 380–392.
Schulz, W.L., Durant, T.J.S., Siddon, A.J., Torres, R., 2016. Use of application containers and workflows for genomic data analysis. J. Pathol. Inform. 7, 53.
Schwartze, V.U., Winter, S., Shelest, E., Marcet-Houben, M., et al., 2014. Gene expansion shapes genome architecture in the human pathogen *Lichtheimia corymbifera*: an evolutionary genomics analysis in the ancient terrestrial mucorales (Mucoromycotina). PLoS Genet. 10, e1004496.
Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069.
Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., et al., 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature 464, 250–255.
Shiroguchi, K., Jia, T.Z., Sims, P.A., Xie, X.S., 2012. Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes. Proc. Natl. Acad. Sci. U. S. A. 109, 1347–1352.
Sittka, A., Lucchini, S., Papenfort, K., Sharma, C.M., et al., 2008. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. PLoS Genet. 4, e1000163.
Smeds, L., Künstner, A., 2011. ConDeTri – a content dependent read trimmer for Illumina data. PLoS One 6, e26314.
Sonnleitner, E., Sorger-Domenigg, T., Madej, M.J., Findeiss, S., et al., 2008. Detection of small RNAs in *Pseudomonas aeruginosa* by RNomics and structure-based bioinformatic tools. Microbiology 154, 3175–3187.
Spjuth, O., Bongcam-Rudloff, E., Dahlberg, J., Dahlö, M., et al., 2016. Recommendations on e-infrastructures for next-generation sequencing. Gigascience 5, 26.
Srivastava, A., Sarkar, H., Gupta, N., Patro, R., 2016. RapMap: a rapid, sensitive and accurate tool for mapping RNA-seq reads to transcriptomes. Bioinformatics 32, i192–i200.
Tafer, H., Hofacker, I.L., 2008. RNAplex: a fast tool for RNA-RNA interaction search. Bioinformatics 24, 2657–2663.
Thomason, M.K., Bischler, T., Eisenbart, S.K., Förstner, K.U., et al., 2015. Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. J. Bacteriol. 197, 18–28.
Tjaden, B., Goodwin, S.S., Opdyke, J.A., Guillier, M., et al., 2006. Target prediction for small, noncoding RNAs in bacteria. Nucleic Acids Res. 34, 2791–2802.
Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., et al., 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 31, 46–53.
Tsuji, J., Weng, Z., 2016. DNApi: a de novo adapter prediction algorithm for small RNA sequencing data. PLoS One 11, e0164228.
Vandenesch, F., Lina, G., Henry, T., 2012. Staphylococcus aureus hemolysins, bi-component leukocidins, and cytolytic peptides: a redundant arsenal of membrane-damaging virulence factors? Front. Cell. Infect. Microbiol. 2, 12.
Voigt, K., Sharma, C.M., Mitschke, J., Lambrecht, S.J., et al., 2014. Comparative transcriptomics of two environmentally relevant cyanobacteria reveals unexpected transcriptome diversity. ISME J. 8, 2056–2068.
Voss, B., Georg, J., Schön, V., Ude, S., et al., 2009a. Biocomputational prediction of non-coding RNAs in model cyanobacteria. BMC Genom. 10, 123.
Voss, B., Hölscher, M., Baumgarth, B., Kalbfleisch, A., et al., 2009b. Expression of small RNAs in Rhizobiales and protection of a small RNA and its degradation products by Hfq in *Sinorhizobium meliloti*. Biochem. Biophys. Res. Commun. 390, 331–336.
Voss, B., Bolhuis, H., Fewer, D.P., Kopf, M., et al., 2013. Insights into the physiology and ecology of the brackish-water-adapted Cyanobacterium *Nodularia spumigena* CCY9414 based on a genome-transcriptome analysis. PLoS One 8, e60224.
Wadler, C.S., Vanderpool, C.K., 2007. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. Proc. Natl. Acad. Sci. U. S. A. 104, 20454–20459.
Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63.
Washietl, S., Hofacker, I.L., 2004. Consensus folding of aligned sequences as a new measure for the detection of functional RNAs by comparative genomics. J. Mol. Biol. 342, 19–30.
Washietl, S., Hofacker, I.L., Stadler, P.F., 2005. Fast and reliable prediction of noncoding RNAs. Proc. Natl. Acad. Sci. U. S. A. 102, 2454–2459.
Westermann, A.J., Förstner, K.U., Amman, F., Barquist, L., et al., 2016. Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. Nature 529, 496–501.
Wolfien, M., Rimmbach, C., Schmitz, U., Jung, J.J., et al., 2016. TRAPLINE: a standardized and automated pipeline for RNA sequencing data analysis, evaluation and annotation. BMC Bioinf. 17, 21.
Wood, D.E., Salzberg, S.L., 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15, R46.
Wright, P.R., Richter, A.S., Papenfort, K., Mann, M., et al., 2013. Comparative genomics boosts target prediction for bacterial small RNAs. Proc. Natl. Acad. Sci. U. S. A. 110, E3487–96.
Wright, P.R., Georg, J., Mann, M., Sorescu, D.A., et al., 2014. CopraRNA and IntaRNA: predicting small RNA targets: networks and interaction domains. Nucleic Acids Res. 42, W119–23.
Yelin, R., Dahary, D., Sorek, R., Levanon, E.Y., et al., 2003. Widespread occurrence of antisense transcription in the human genome. Nat. Biotechnol. 21, 379–386.
Yusuf, D., Marz, M., Stadler, P.F., et al., 2010. Bcheck: a wrapper tool for detecting RNase P RNA genes. BMC Genom. 11, 432.

2.1.3 A guide of best practices for RNA-Seq analysis in Galaxy

Gruening, B.A., ..., **Wolfien, M.**, ..., Wolkenhauer, O., ...,
and Backofen, R. (2017).
The RNA workbench: best practices for RNA and high-throughput sequencing bioinformatics in Galaxy.
Nucleic Acids Research. IF: 10.727, Citations (December 14, 2020): 21

RNA-based regulation has become a major research topic in molecular biology. The analysis of epigenetic and expression data is therefore incomplete if RNA-based regulation is not considered. Thus, it is increasingly important but not yet standard to combine RNA-centric data and analysis tools with other types of experimental data, such as RNA-Seq or ChIP-Seq. Here, the authors present the RNA workbench, a comprehensive set of analysis tools and consolidated workflows that enable the researcher to combine these two worlds.

I contributed to this article by incorporating my experience gained during the development of TRAPLINE (Section 2.1.1). In particular, I defined, revised, and tested workflows for RNA-Seq data analyses and incorporated these into the underlying training material. Based on the dockerized Galaxy framework, the RNA workbench guarantees simple access, easy extension, flexible adaption to personal, and security needs, as well as sophisticated analyses that are independent of command-line knowledge. Currently, it includes more than 50 bioinformatics tools that are dedicated to different research areas of RNA biology including RNA structure analysis, RNA alignment, RNA annotation, RNA-protein interaction, ribosome profiling, RNA-Seq analysis, and RNA target prediction.

The workbench is developed and maintained by experts in RNA bioinformatics and the Galaxy framework. Together with the growing community evolving around this platform, they are committed to keep the workbench up to date for future standards and needs, providing researchers with a reliable and robust framework for RNA data analysis. The code is continiously updated and publicly available for contribution at github (Fallmann et al., 2019).[10]

[10] https://github.com/bgruening/galaxy-rna-workbench

W560–W566 *Nucleic Acids Research, 2017, Vol. 45, Web Server issue* *Published online 5 June 2017*
doi: 10.1093/nar/gkx409

The RNA workbench: best practices for RNA and high-throughput sequencing bioinformatics in Galaxy

Björn A. Grüning[1,2,*], Jörg Fallmann[3], Dilmurat Yusuf[4], Sebastian Will[5], Anika Erxleben[1], Florian Eggenhofer[1], Torsten Houwaart[1], Bérénice Batut[1], Pavankumar Videm[1], Andrea Bagnacani[6], Markus Wolfien[6], Steffen C. Lott[7], Youri Hoogstrate[8], Wolfgang R. Hess[7], Olaf Wolkenhauer[6], Steve Hoffmann[3], Altuna Akalin[4], Uwe Ohler[4,9], Peter F. Stadler[3,5,10,11] and Rolf Backofen[1,2,12,*]

[1]Bioinformatics Group, Department of Computer Science, University of Freiburg, Georges-Koehler-Allee 106, D-79110 Freiburg, Germany, [2]Center for Biological Systems Analysis (ZBSA), University of Freiburg, Habsburgerstr. 49, D-79104 Freiburg, Germany, [3]Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstr. 16-18, D-04107 Leipzig, Germany, [4]Berlin Institute for Medical Systems Biology, Max-Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, D-13125, Berlin, Germany, [5]Institute for Theoretical Chemistry, University of Vienna, Währingerstrasse 17, A-1090 Vienna, Austria, [6]Department of Systems Biology and Bioinformatics, University of Rostock, Ulmenstr. 69, D-18051 Rostock, Germany, [7]Genetics and Experimental Bioinformatics, Faculty of Biology, University of Freiburg, Schänzlestr. 1, D-79104 Freiburg, Germany, [8]Department of Urology, Erasmus University Medical Center, Wytemaweg 80, 3015 CN Rotterdam, Netherlands, [9]Departments of Biology and Computer Science, Humboldt University, Unter den Linden 6, D-10099 Berlin, [10]Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, D-04103 Leipzig, Germany, [11]Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA and [12]BIOSS Centre for Biological Signaling Studies, University of Freiburg, Schänzlestr. 18, D-79104 Freiburg, Germany

Received March 02, 2017; Revised April 13, 2017; Editorial Decision April 28, 2017; Accepted May 31, 2017

ABSTRACT

RNA-based regulation has become a major research topic in molecular biology. The analysis of epigenetic and expression data is therefore incomplete if RNA-based regulation is not taken into account. Thus, it is increasingly important but not yet standard to combine RNA-centric data and analysis tools with other types of experimental data such as RNA-seq or ChIP-seq. Here, we present the RNA workbench, a comprehensive set of analysis tools and consolidated workflows that enable the researcher to combine these two worlds. Based on the Galaxy framework the workbench guarantees simple access, easy extension, flexible adaption to personal and security needs, and sophisticated analyses that are independent of command-line knowledge. Currently, it includes more than 50 bioinformatics tools that are dedicated to different research areas of RNA biology including RNA structure analysis, RNA alignment, RNA annotation, RNA-protein interaction, ribosome profiling, RNA-seq analysis and RNA target prediction. The workbench is developed and maintained by experts in RNA bioinformatics and the Galaxy framework. Together with the growing community evolving around this workbench, we are committed to keep the workbench up-to-date for future standards and needs, providing researchers with a reliable and robust framework for RNA data analysis. Availability: The RNA workbench is available at https://github.com/bgruening/galaxy-rna-workbench.

INTRODUCTION

Since recent advances in high-throughput sequencing (HTS) emphasized the importance and versatile role of (non-coding) RNAs, there is high demand for integrated computational analyses investigating RNA-mediated regulation. Previously existing workbenches (such as *miARma-Seq* (1) *RAP* (2) and the UEA Small RNA Workbench (3)) were focused on providing tools for the analysis of RNA deep sequencing data and do not contain RNA centric tools.

We addressed these needs by developing the RNA workbench. Based on the Galaxy framework (4) it combines a

*To whom correspondence should be addressed. Email: backofen@informatik.uni-freiburg.de
Correspondence may also be addressed to Björn A. Grüning. Tel: +49 761 2037460; Fax: +49 761 2037462; Email: gruening@informatik.uni-freiburg.de

Nucleic Acids Research, 2017, Vol. 45, Web Server issue **W561**

comprehensive set of tools for the analysis of RNA structures, RNA alignments, RNA–RNA and RNA–protein interactions, RNA sequencing, ribosome profiling, genome annotation and many more. So far, we integrated more than 50 RNA-related tools, including suites like the ViennaRNA package, covering this broad variety of use-cases (a complete list of tools can be found on GitHub). Every available tool works as a single building-block that can be connected with other tools to create computational pipelines. Datasets can be incorporated in a similar manner, facilitating an intersection of diverse data sources such as DNA methylation with RNA-seq experiments. Input and output datasets can be defined by the user, and can be as diverse as the adapted set of tools. Established data types for sequence and/or structure information are accepted as input. Output data types follow the same principle, can be converted to different formats, or ultimately used to draw plots and create figures. The workbench provides tools for visualizations of RNA structure datasets, such as dot-bracket strings, and RNA 2D or 3D structures. The workbench also covers a broad range of RNA secondary structure prediction and analysis tools such as RNAfold (5) or LocARNA (6,7).

GOALS OF THE RNA WORKBENCH

The main driving force behind the development of the RNA workbench is the goal to establish a central, redistributable workbench for scientists and programmers working with RNA-related data, and build a sustainable community around it. This platform is unique in combining available tools, workflows and training material, as well as providing easy access for experimentalists. Simultaneously, it serves as a central hub for programmers, which can easily integrate and deploy their existing or novel tools and workflows. The RNA workbench is based on three pillars: (i) a comprehensive set of RNA-bioinformatics tools, (ii) easy and stable dissemination via Galaxy and *Docker* and (iii) a set of predefined workflows and associated descriptions/training material. The latter is needed for two reasons: first, it facilitates the use of the RNA workbench for researchers with limited bioinformatics experience, and second, it allows to integrate the workbench in the daily lab work by combining RNA-related analysis tasks with workflows for RNA-seq analysis.

Building on the shoulders of giants

In order to achieve long-term sustainability, we provide the essentials of our work on *BioConda* (https://bioconda.github.io) and *BioContainers* (8) (http://biocontainers.pro) for reproducible deployments of tools into Galaxy. Using easy-to-distribute packages for all tool dependencies also enables automatic continuous integration tests for all developed tools and the workbench. After a tool passes the tests and gets accepted it will be made available via an automatic deployment into the Galaxy ToolShed (https://toolshed.g2.bx.psu.edu) (9). From the ToolShed, Galaxy administrators can easily install desired tools and workflows.

Easily accessible and reproducible analysis platform

For the fast dissemination of the RNA workbench, as well as for an easy integration with other HTS analysis tasks, we implemented the RNA workbench within the Galaxy framework. A major advantage of relying on Galaxy as the core framework is that it is possible to leverage its scalability, which enables the RNA workbench to run on single CPU installations as well as on large multi-node high performance computing environments. Furthermore, Galaxy provides researchers with means to reproduce their own workflow analyses, enabling them to rerun entire pipelines, or publish and share them with others. The RNA workbench is containerized, *i.e.*, administrators can deploy it via *Docker*. That makes it possible to have all tool installation dependencies already resolved, while still keeping maintenance tasks to a minimum. The provided layer of virtualization also allows the handling of user-defined input data in a secure and compartmentalized way, a key requirement for researchers working on sensitive data (*e.g.* patient data in clinics). Running the containerized RNA workbench simply requires installing *Docker* and starting the Galaxy RNA workbench image. Furthermore, containerizing Galaxy enables a customized Galaxy instance with a selected subset of tools dedicated to specific data analysis tasks, while keeping deployment and installation simple.

RNA-BIOINFORMATICS TOOLS

In its current state, the RNA workbench includes more than 50 tools covering all aspects of RNA research. In a community effort, these tools will be kept up-to-date and adapted to future needs. New tools and new ways to visualize data provided to the user will also be integrated. A current overview of tools available in the RNA workbench can be found at http://bgruening.github.io/galaxy-rna-workbench/.

In the following, we will highlight a few of the integrated tools.

The ***ViennaRNA*** package (5) consists of a suite of tools centered around the prediction of secondary structures of RNAs based on the thermodynamic Turner energy model. Thus, it covers prediction of optimal and suboptimal structures from single sequences as well as alignments, prediction of ensemble base pair probabilities, accessibility of sequences, and RNA–RNA interaction prediction. Importantly, predictions can be flexibly controlled by hard and soft structure constraints; the latter enables the inclusion of structure probing data.

AREsite2 (10) is a resource for the investigation of AU, GU and U-rich elements (ARE, GRE, URE) in human and model organisms. It provides information on genomic location, genomic context, RNA secondary structure context and conservation of annotated motifs in the whole gene body including introns. It is integrated into the RNA workbench via its REST interface, which provides search results directly in Galaxy for further analysis.

LocARNA (6,7) provides a comparative analysis of multiple (unaligned) RNAs by simultaneous folding and alignment, implementing a fast variant of the Sankoff algorithm. Beyond pairwise and multiple alignments, it computes reliabilities of alignment columns and provides very fast analysis

W562 Nucleic Acids Research, 2017, Vol. 45, Web Server issue

by simultaneous folding and matching. Finally, *LocARNA* supports anchor and structure constraints, which improve its applicability in practice.

doRiNA (11) is a database of RNA interactions in post-transcriptional regulation. The combined action of RNA-binding proteins (RBPs) and microRNAs (miRNAs) is believed to form the backbone of post-transcriptional regulation. *doRiNA* is implemented as data source tool inside the RNA workbench. This means that the Galaxy user is redirected to the post-transcriptional interaction database and can make selections using the optimized doRiNA interface. Once the selection is done, the data is streamed directly to Galaxy and can be freely analyzed with other tools.

The ***Infernal*** (12) tool suite can construct probabilistic models, also called covariance models (CM), that represent the sequence and structure of an RNA family from a multiple sequence alignment with consensus secondary structure. The covariance model can be used to find more members of this RNA family via homology search.

PARalyzer (13) generates a high resolution map of interaction sites between RNA-binding proteins and their targets. The algorithm utilizes the deep sequencing reads generated by the PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation) protocol. The use of photoactivatable nucleotides in the PAR-CLIP protocol results in more efficient crosslinking between the RNA-binding protein and its target relative to other CLIP methods; in addition a nucleotide substitution occurs at the site of crosslinking, providing for single-nucleotide resolution binding information. *PARalyzer* utilizes this nucleotide substitution in a kernel density estimate classifier to generate the high resolution set of protein-RNA interaction sites.

FuMa (14) can generate an integration report on predicted fusion genes from most RNA-seq fusion gene detection software. It automatically orders the result based on the frequencies of the fusion genes such that frequently predicted fusion genes can be extracted.

WORKFLOWS

One of the core concepts of the RNA workbench is the definition of standard workflows as a minimal set of building blocks around which a researcher can compose and tailor specific pipelines. For example, a researcher wants to analyze the effects of an RNA-binding protein (RBP) in regard to expression levels in wild-type compared to knockout or knockdown of the RBP of interest. In this case, one needs to combine the detection of differentially expressed genes in the two conditions with the information of publicly available CLIP-data, as provided for example by the *doRiNA* (11) database, to differentiate between direct and indirect targets. Workflows for the analysis of differentially expressed genes are part of the RNA workbench, as well as an interface to *doRiNA*, such that it becomes an easy task to design a new workflow combining these analysis steps.

In Galaxy, workflows are typically created in two different ways: (i) from an existing history, which stores all tools applied in a previous analysis together with all pertinent parameters, or (ii) from scratch, using a graphical editor via drag-and-drop of tools from the tool panel into the workflow editor. Within workflows, tools can be freely combined to ensure a maximum of flexibility in their usage and connectivity between different analysis steps, e.g. RNA structure analysis tools and RNA-seq data analysis. Various format converters embedded in Galaxy allow combining diverse analysis outputs. Easy sharing of workflows with other Galaxy users guarantees highly reproducible and transparent research. In other words, the workflows ensure that all analysis steps, tools and parameters of an experiment are documented and visible to researchers, readers and reviewers. Workflows can also be submitted to the Galaxy ToolShed or *myexperiment.org* (15) for further distribution. The RNA workbench currently includes publicly available standard workflows for RNA data analysis, e.g. for RNA-seq. These workflows contain all required steps such as quality control, mapping, differential expression analysis, and visualization of results. Provided workflows can easily be extended or modified, e.g. to use other read mappers available in Galaxy.

In the following, we will describe two sample workflows, one closely related to the detection of ncRNAs, which is a common task in RNA-related research. The other workflow is related to the analysis of RNA-seq data and is often needed as a subworkflow for more complex analysis tasks. These workflows are well annotated and described in the RNA workbench and extended by interactive Galaxy *tours*.

Analysis of (unaligned) non-coding RNAs

An important task is to test for the existence of a functional structure in a non-coding RNA. However, the secondary structure of structured non-coding RNAs is not significantly more stable compared to random sequences (16). Thus, putative functional structures can only be detected using information about conservation. Our workflow for non-coding RNAs performs the typical analysis steps required to detect conserved secondary structures, given a set of unaligned RNA sequences. It computes a sequence and a structure-based alignment by *MAFFT* (17) and *LocARNA*, respectively, and analyzes them with *RNAcode* (18) and *RNAz* (19) with appropriate parameter settings. *RNAz* and *RNAcode* both work on a given alignment. *RNAz* tests whether a consensus secondary structure is significantly conserved, whereas *RNAcode* differentiates coding from non-coding RNAs. Together these tools provide information, whether the RNAs are related and conserve a common secondary structure. In addition, a covariance model is built from the *LocARNA* alignment and subsequently used to search the given sequence database for RNAs with similar sequence- and structure-conservation. This workflow resembles the core of *RNAlien* (20), which is based on the same tools and is integrated into the RNA workbench. Going beyond the presented workflow, *RNAlien* automatically gathers sequences via homology search starting from a single sequence and constructs RNA family models in an iterative process.

To give an other example, in the context of μORFs detection, RNA-seq analysis, the identification of non-coding RNAs with *RNAcode* and *RNAz* and the detection of transcription start sites can be used to determine new, short transcripts that are expressed and do not exhibit secondary

Nucleic Acids Research, 2017, Vol. 45, Web Server issue **W563**

structure conservation (i.e. are likely not functional ncRNAs). Subsequent analysis of Ribo-seq data can then provide additional evidence for a new transcript that may code for a small protein. For all these tasks, partial workflows and required tools are already integrated in our RNA workbench, which implies that it is easy to set up a new workflow for a more complex task.

RNA-seq analysis: trimming, mapping and read count

As mentioned before, the analysis of RNA-centric data like CLIP-seq requires the combination with other type of data, and very often RNA-seq. For that reason, we provide a standard RNA-seq workflow that can easily be combined with other workflows. The RNA-seq workflow (as shown in Figure 1) takes a list of RNA-seq datasets as input and successively executes a series of analysis steps - adapter & quality trimming, mapping to a reference genome and read count per annotated gene. The input allows two conditions, e.g. treatment versus control and it also accepts single-end and paired-end reads for each condition. At the trimming step, the workflow employs *Trim Galore!* (21,22) to perform adapter trimming. Then, *TopHat2* (23) is used to map the trimmed reads against the reference sequences, which should be provided by the user. As last step, the workflow executes *HTSeq-count* (24) to generate read counts per annotated gene for each condition and for each sequencing type. A reference annotation in Gene Transfer Format (GTF), *e.g.* provided by Ensembl (25) is required at this step. The final read counts can be used for the downstream assessment of differential expression using tools like *DESeq2* (26). The current workflow can serve as a template that can be modified by the user according to different needs, for instance, replacement of tools or modification of the wrapping strategy.

IMPLEMENTATION

The workbench is implemented as portable virtualized container based on Galaxy. The Galaxy framework allows for reproducible and transparent scientific research which makes it easy to access, deploy and scale—conceptualized as a web service. The foundation of the workbench container is a generic Galaxy *Docker* instance (http://bgruening.github.io/docker-galaxy-stable/). On-top of this, pre-configured Galaxy tools can be automatically installed from the Galaxy ToolShed using the Galaxy API *BioBlend* (27). In Galaxy, tool dependencies are automatically resolved via *BioConda*, which is the bioinformatics channel for the *Conda* package manager. *BioConda* facilitates software packaging and enables installation at a user level, keeping track of different versions of the same software in virtual environments. These features are in line with the scope of Galaxy; maintaining large numbers of dependencies in a reproducible way. Therefore, all available tools within the RNA workbench are also distributed as *BioConda* packages and *BioContainers*, which are persistent, frozen, containerized versions of *Conda* packages. The RNA workbench ships with a variety of tools, tours, documentation, workflows and data that have been added as additional layers on top of the generic *Docker* instance. During development, the software has been tested extensively in a continuous integration setup (CI) at different levels: Galaxy itself, tool integration in Galaxy (IUC, galaxytools channels), dependencies (*BioConda*) and at the workbench level. Together with a strict version management on all levels, this contributes to a high degree of error-control and reproducibility. The RNA workbench started in January 2015 - with constant development over 2 years, and extensive testing in local and public Galaxy instances, such as the Freiburg Galaxy instance, the MDC instance in Berlin and Erasmus MC's Galaxian. More than 500 users accessed the RNA tools during the last two years and the virtualized *Docker* instance was already downloaded >500 times. Moreover, due to an open and transparent development process, there is a growing community that contributes to our workbench, which guarantees the sustainability of the RNA workbench project and maintenance of the underlying *Docker*/*rkt* images.

USING THE RNA WORKBENCH

Installation: The RNA workbench can be installed under OSX and Windows using the graphical tool Kitematic (https://kitematic.com), or with the following Linux command:

```
docker run -d -p 8080:80 bgruening/galaxy-rna-workbench
```

This installation is production-ready and can be configured to use external computer clusters or cloud environments. Due to the very modular system, it is also possible to install all or only a few tools of the RNA workbench on available Galaxy servers. Just get in contact with your local Galaxy administrator. When using the RNA workbench *Docker* image, the user has full administration rights, which enables customization independent of potential user restrictions.

Training

For self-empowering the user, documentation and training of the RNA workbench are important. We included an extensive set of documentation in traditional formats, *e.g.* tool descriptions and 'README' files.

We also provide training sessions around HTS data analyses and RNA-seq data analysis. The training materials ranging from the introduction to Galaxy, to usage and maintenance of Galaxy and the RNA workbench are freely accessible for self-paced studies at the Galaxyproject Github repository (http://galaxyproject.github.io/training-material). This training material is constantly improved and extended in an international community effort, including ELIXIR and EMBL. For HTS data analyses we provide training as a specific introduction to the topic with self-explanatory presentation slides, a hands-on training documentation describing the analysis workflow, all necessary input files ready-to-use via *Zenodo*, a Galaxy *Interactive Tour*, and a tailor-made Galaxy *Docker* image for the corresponding data analysis.

To provide an even more intense training experience within the RNA workbench, we also included interactive training such as the *Galaxy Interactive Tours*. Such tours guide users through an entire analysis in an interactive

W564 Nucleic Acids Research, 2017, Vol. 45, Web Server issue

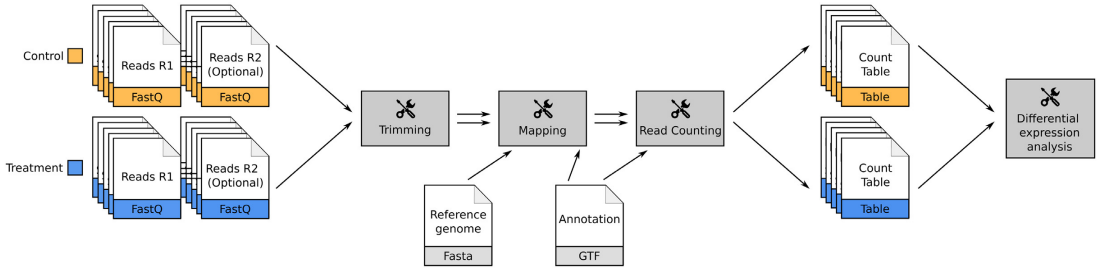


Figure 1. The workflow for analyzing RNA-seq data. The workflow tolerates single-end and paired-end reads derived from different conditions. It employs *TopHat2* for mapping and *HTSeq-count* to create the read counts. The final outputs contain read count per annotated gene for each condition and for each sequencing type.

and explorative way. It combines advantages from training videos and detailed protocols. Production of training videos is very time-consuming and tend to become outdated very soon, due to tool version changes or renewed workflows. In contrast to conventional screencasts, a Galaxy *Interactive Tour* can be easily updated and improved to guide the Galaxy user step-by-step, e.g. through a whole HTS analysis starting from uploading the data to using complex analysis tools. Exemplary, the RNA workbench currently integrates two Galaxy *Interactive Tours*. The first one introduces a new user to the Galaxy interface and its usage with an RNA-seq example dataset. The second one illustrates secondary structure prediction of RNA molecules using parts of the *ViennaRNA* package. To show how Galaxy *Interactive Tours* can interactively guide users through the necessary steps of HTS analyses, the tours are also provided as online screencasts.

Visualization

Following data reduction as a key element of explorative research, there is a need for meaningful figures and visualizations that summarize results. The RNA workbench includes standard interactive plotting tools to draw bar charts and scatter plots from all kinds of tabular data and allows for connections to *Integrated Genome Browser* (29) and *UCSC* (30) like any other Galaxy instance. On top of this, we included three visualizations specific to RNA research. An interactive DotPlot visualization for secondary structures in EPS format (Figure 2b), a 2D visualization for the common dot-bracket format (Figure 2a) and a 3D visualization capable of visualizing PDB, SDF and MOL files containing three-dimensional coordinates (Figure 2c).

COMMUNITY

The RNA workbench project is an open source project that strives to create a community interested in accessible and reproducible RNA-related research. Knowing that real sustainability can only come true with a strong community we are aiming at more open participation, reward, and inclusion. We are working together with Galaxy, *BioConda*, *BioContainers* and *BioJS* and coordinating efforts to not reinvent the wheel but joining forces to create the new generation of bioinformatics infrastructure together. In the RNA workbench community, we practice the organizations on GitHub, IRC, and Gitter and welcome everyone to contribute on every level to improve the entire stack from documentation to tools and scientific workflows. Support will be provided through the same channels.

DISCUSSION

In this work, we present the RNA workbench, maintained and developed by a constantly growing community. The presented workbench is unique as it allows to easily combine RNA-centric analysis with other types of experiments. It provides a set of tools, each one being available as *BioConda* package as well as a Docker/rkt container (*BioContainers*). Based on the Galaxy *Docker* project, the proposed web server is more than the sum of its parts. It offers a comprehensive virtualized RNA workbench that can be deployed on every standard Linux, Windows and OSX computer, but can at the same time employ high-performance- or cloud-computing infrastructure.

Major advantages of our approach to deliver a dockerized workbench for RNA centric analysis are the ease of installation, the high number of pre-included tools, the flexibility in regard to extension with other tools and workflows and the high reproducibility and transparency of workflows. All tools that are available on the Galaxy *Toolshed* can be installed along with their automatically resolved dependencies with a single click in the Galaxy interface. Best practice pipelines for the analysis of RNA-seq data are provided with the *Docker* image and can easily be modified, extended or combined with other analysis pipelines via Galaxy's workflow editor GUI.

The RNA workbench was designed as a community project, and as such it is easy for users to contribute to the workbench with workflows, new tools and training material, keeping the workbench up-to-date and valuable for research. Moreover, all components such as tools, workflows, visualizations, interactive tours and training material can be easily integrated into any available Galaxy instance for teaching, learning or exploratory purposes.

The main difference to existing solutions such as *miARma-Seq* (1), *RAP* (2) and the UEA Small RNA Workbench (3) is that our RNA workbench combines the realm of RNA-centric analysis on sequence and structure level with modern high-throughput sequence analysis. In this re-

Nucleic Acids Research, 2017, Vol. 45, Web Server issue **W565**

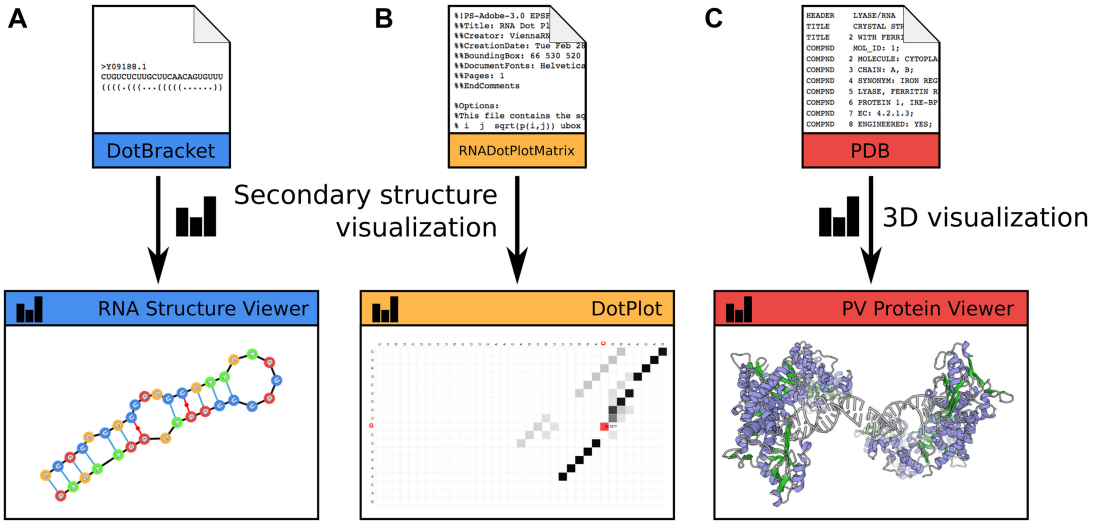


Figure 2. RNA structure visualization: The figure shows visualization for an *IRE1* RNA sequence, retrieved from the Rfam database (28), via different backends integrated into the toolbox. (**A**) Secondary structure encoded in dot-bracket notation, can be displayed by the RNA structure viewer. (**B**) Base pairing probabilities are visualized as DotPlot. (**C**) Tertiary/Quaternary structure information encoded in protein-database format is rendered via Protein Viewer.

gard we provide well established tools for RNA structure prediction, analysis and visualization together with read mappers and expression analysis tools for HTS analysis.

ACKNOWLEDGEMENTS

We thank the de.NBI and ELIXIR projects for supporting bioinformatics infrastructure. Thanks also to the Galaxy community, especially to the Freiburg Galaxy Team, for developing, maintaining and supporting this great framework. We also like to acknowledge the *BioConda* and *BioContainers* community for setting new standards in reproducible software deployments. Thanks also to the BioJS community for great discussions about scientific visualizations and how we can make them more accessible. Moreover, the authors acknowledge the support of many upstream developers that helped us to integrate their tools into the RNA workbench and accepted patches.

FUNDING

Collaborative Research Center 992 Medical Epigenetics [DFG grant SFB 992/1 2012]; German Federal Ministry of Education and Research [BMBF grants 031 A538A/A538C RBC, 031L0101B/031L0101C de.NBI-epi, 031L0106 de.STAIR (de.NBI)]; Center for Translational Molecular Medicine (CTMM), TraIT project [05T-401 to Y.H.]. Funding for open access charge: German Government.
Conflict of interest statement. None declared.

REFERENCES

1. Andrés-León,E., Núñez-Torres,R. and Rojas,A.M. (2016) miARma-Seq: a comprehensive tool for miRNA, mRNA and circRNA analysis. *Scientific Rep.*, **6**, 25749.
2. D'Antonio,M., De Meo,P.D., Pallocca,M., Picardi,E., D'Erchia,A.M., Calogero,R.A., Castrignanò,T. and Pesole,G. (2015) RAP: RNA-Seq analysis pipeline, a new cloud-based NGS web application. *BMC Genomics*, **16**, S3.
3. Stocks,M.B., Moxon,S., Mapleson,D., Woolfenden,H.C., Mohorianu,I., Folkes,L., Schwach,F., Dalmay,T. and Moulton,V. (2012) The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics*, **28**, 2059–2061.
4. Afgan,E., Baker,D., van den Beek,M., Blankenberg,D., Bouvier,D., Cech,M., Chilton,J., Clements,D., Coraor,N., Eberhard,C. *et al.* (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.*, **44**, W3–W10.
5. Lorenz,R., Bernhart,S.H., Honer Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
6. Will,S., Reiche,K., Hofacker,I.L., Stadler,P.F. and Backofen,R. (2007) Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comput. Biol.*, **3**, e65.
7. Will,S., Joshi,T., Hofacker,I.L., Stadler,P.F. and Backofen,R. (2012) LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. *RNA*, **18**, 900–914.
8. da~Veiga~Leprevost,F., Grüning,B.A., Aflitos,S.A., Röst,H.L., Uszkoreit,J., Barsnes,H., Vaudel,M., Moreno,P., Gatto,L., Weber,J. *et al.* (2017) BioContainers: an open-source and community-driven framework for software standardization. *Bioinformatics*, doi:10.1093/bioinformatics/btx192.
9. Blankenberg,D., Von Kuster,G., Bouvier,E., Baker,D., Afgan,E., Stoler,N., Taylor,J. and Nekrutenko,A. (2014) Dissemination of scientific software with Galaxy ToolShed. *Genome Biol.*, **15**, 403.
10. Fallmann,J., Sedlyarov,V., Tanzer,A., Kovarik,P. and Hofacker,I.L. (2016) AREsite2: an enhanced database for the comprehensive investigation of AU/GU/U-rich elements. *Nucleic Acids Res.*, **44**, D90–D95.

11. Blin,K., Dieterich,C., Wurmus,R., Rajewsky,N., Landthaler,M. and Akalin,A. (2015) DoRiNA 2.0–upgrading the doRiNA database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res.*, **43**, D160–D167.
12. Nawrocki,E.P. and Eddy,S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, **29**, 2933–2935.
13. Corcoran,D.L., Georgiev,S., Mukherjee,N., Gottwein,E., Skalsky,R.L., Keene,J.D. and Ohler,U. (2011) PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol.*, **12**, R79.
14. Hoogstrate,Y., Bottcher,R., Hiltemann,S., van der Spek,P.J., Jenster,G. and Stubbs,A.P. (2016) FuMa: reporting overlap in RNA-seq detected fusion genes. *Bioinformatics*, **32**, 1226–1228.
15. Goble,C.A., Bhagat,J., Aleksejevs,S., Cruickshank,D., Michaelides,D., Newman,D., Borkum,M., Bechhofer,S., Roos,M., Li,P. *et al.* (2010) myExperiment: a repository and social network for the sharing of bioinformatics workflows. *Nucleic Acids Res.*, **38**, W677–W682.
16. Rivas,E. and Eddy,S.R. (2001) Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics*, **2**, 8.
17. Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.
18. Washietl,S., Findeiss,S., Muller,S.A., Kalkhof,S., von Bergen,M., Hofacker,I.L., Stadler,P.F. and Goldman,N. (2011) RNAcode: robust discrimination of coding and noncoding regions in comparative sequence data. *RNA*, **17**, 578–594.
19. Gruber,A.R., Neubock,R., Hofacker,I.L. and Washietl,S. (2007) The RNAz web server: prediction of thermodynamically stable and evolutionarily conserved RNA structures. *Nucleic Acids Res.*, **35**, W335–W338.
20. Eggenhofer,F., Hofacker,I.L. and Honer Zu Siederdissen,C. (2016) RNAlien—unsupervised RNA family model construction. *Nucleic Acids Res.*, **44**, 8433–8441.
21. Krueger,F. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisufite-Seq) libraries.
22. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, doi:10.14806/ej.17.1.200.
23. Kim,D., Pertea,G., Trapnell,C., Pimentel,H., Kelley,R. and Salzberg,S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.*, **14**, R36.
24. Anders,S., Pyl,P.T. and Huber,W. (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
25. Aken,B.L., Achuthan,P., Akanni,W., Amode,M.R., Bernsdorff,F., Bhai,J., Billis,K., Carvalho-Silva,D., Cummins,C., Clapham,P. *et al.* (2017) Ensembl 2017. *Nucleic Acids Res.*, **45**, D635–D642.
26. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
27. Sloggett,C., Goonasekera,N. and Afgan,E. (2013) BioBlend: automating pipeline analyses within Galaxy and CloudMan. *Bioinformatics*, **29**, 1685–1686.
28. Nawrocki,E.P., Burge,S.W., Bateman,A., Daub,J., Eberhardt,R.Y., Eddy,S.R., Floden,E.W., Gardner,P.P., Jones,T.A., Tate,J. *et al.* (2015) Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.*, **43**, D130–D137.
29. Thorvaldsdottir,H., Robinson,J.T. and Mesirov,J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics*, **14**, 178–192.
30. Tyner,C., Barber,G.P., Casper,J., Clawson,H., Diekhans,M., Eisenhart,C., Fischer,C.M., Gibson,D., Gonzalez,J.N., Guruvadoo,L. *et al.* (2017) The UCSC Genome Browser database: 2017 update. *Nucleic Acids Res.*, **45**, D626–D634.

2.1.4 Linking workflow development and the annotation of ncRNAs

Wolfien, M., Brauer, D. L., Bagnacani, A., and Wolkenhauer, O. (2019).
Workflow development for the functional characterization of ncRNAs.
Methods in Molecular Biology. Downloads: 1,000; Citations (December 14, 2020): 2

The functional role of ncRNAs is not yet entirely understood but intensively investigated and, thus, already revealed interesting regulatory potential in various biological processes and clinical use cases. Research efforts have identified numerous ncRNAs and multiple RNA subtypes, which are attributed to diverse functionalities known to interact with different functional layers ranging from DNA and RNA to proteins. These diverse functionalities strongly hamper the functional prediction for newly identified ncRNAs. However, current bioinformatics and systems biology approaches show promising results to facilitate an identification of these diverse ncRNA functionalities.

In this book chapter, I developed an experimental and computational strategy to identify and functionally characterize ncRNAs by using RNA-Seq data, as well as further databases (e.g., STRING, Reactome, GO, LncRBase). This strategy includes analyses from transcriptome-wide association studies, GBA, molecular network analyses, and artificial intelligence guided predictions. These integration of diverse tools is summarized as "*connective workflows*" because a combination of single data analysis workflows is needed for a proper characterization of such diverse ncRNAs.

In summary, we show current experimental NGS protocols for an identification of ncRNAs, give an overview of sequencing data analysis workflows, as well as available computational environments, and provide state-of-the-art approaches to functionally characterize ncRNAs. A strategy is presented to cover the identification and functional characterization of unknown ncRNA transcripts by using *connective workflows*.

Chapter 5

Workflow Development for the Functional Characterization of ncRNAs

Markus Wolfien, David Leon Brauer, Andrea Bagnacani, and Olaf Wolkenhauer

Abstract

During the last decade, ncRNAs have been investigated intensively and revealed their regulatory role in various biological processes. Worldwide research efforts have identified numerous ncRNAs and multiple RNA subtypes, which are attributed to diverse functionalities known to interact with different functional layers, from DNA and RNA to proteins. This makes the prediction of functions for newly identified ncRNAs challenging. Current bioinformatics and systems biology approaches show promising results to facilitate an identification of these diverse ncRNA functionalities. Here, we review (a) current experimental protocols, i.e., for Next Generation Sequencing, for a successful identification of ncRNAs; (b) sequencing data analysis workflows as well as available computational environments; and (c) state-of-the-art approaches to functionally characterize ncRNAs, e.g., by means of transcriptome-wide association studies, molecular network analyses, or artificial intelligence guided prediction. In addition, we present a strategy to cover the identification and functional characterization of unknown transcripts by using connective workflows.

Key words Workflow, ncRNA, Transcript identification, Experimental RNA discovery, Data analysis, Next Generation Sequencing, Network analysis, Co-expression analysis, Machine learning

1 The Missing Link to Functionally Characterize ncRNAs

Before the 1980s, RNAs were seen as macromolecules that primarily support the protein synthesis and have been considered to be dormant—their regulative potential was unheard of. Our current knowledge about the human organism assumes 2% of the genome to encode for functional protein-coding RNAs, messenger RNAs (mRNAs), and more than 60% of the transcriptional output can be attributed to ncRNAs, which shows the actual importance of this formerly unrecognized molecule class [1]. Nowadays, we also know that the regulation of gene expression by noncoding RNAs (ncRNAs) via mRNA/ncRNA interactions is an essential and widespread phenomenon that occurs in almost all biological domains and, therefore, has become a basic principle in biology [2]. As of

Xin Lai et al. (eds.), *Computational Biology of Non-Coding RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1912, https://doi.org/10.1007/978-1-4939-8982-9_5,

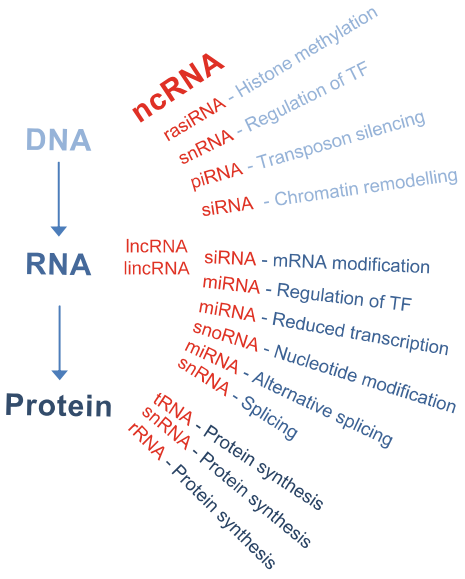


Fig. 1 Illustration to show the complexity and versatile role of ncRNA subtypes

early 2018, further ncRNA subtypes, e.g., circular RNAs (circRNAs), microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), long noncoding RNAs (lncRNAs), and many more, have been recently characterized [3]. The currently known ncRNA subtypes and their respective functionalities are summarized in Fig. 1.

Next Generation Sequencing (NGS) technologies provide an attractive platform for the identification and quantification of genomes as compared to other high-throughput technologies and, furthermore, have been widely implemented for various applications such as DNA sequencing, de novo genome sequencing, epigenomics and transcriptomics profiling, and chromatin immunoprecipitation sequencing [4]. The application of NGS in clinical areas includes the identification of genetic variants, somatic or inherited mutations, as well as epigenetic changes to analyze an individual's disease-specific genome or tissue-specific transcriptome, where a comprehensive match of variants, like single-nucleotide polymorphisms (SNPs), can be easily detected [5]. Only very few of the ncRNAs under investigation can be interpreted and are known to be actionable, which means that a notable amount will be of either unknown or novel clinical importance and, therefore, holds the biological need and multifarious capability in a functional characterization [6].

Although state-of-the-art experimental technologies have yielded promising results in finding and characterizing novel ncRNAs, they are still subject to certain limitations, because the

expression of most ncRNAs is lower than mRNA expression and, moreover, ncRNAs show tissue/stage-specific expression patterns [7, 8]. High-throughput sequencing generates an enormous amount of data and, thus, requires substantial computational power [9]. Algorithms for the identification are complementing experimental methods, allowing for a more focused approach for specific organisms and cell types. The following threefold difficulties of computational prediction that arise through the biological circumstances will be discussed throughout this chapter.

1. Variety of ncRNA subtypes: Advances in sequencing technologies have led to the discovery of a multitude of ncRNA subtypes, in which some are highly conserved (e.g., miRNAs, circRNAs) and others are generally lacking conservation across species, such as lncRNAs [10, 11].
2. Amount of uncharacterized ncRNAs: Only considering two of the most common ncRNA subtypes, many thousands of miRNAs have been discovered in many organisms. According to miRBase (http://www.mirbase.org/), there are currently 2,694 mature miRNAs in the human genome that are individually predicted to target hundreds of genes across multiple pathways [12]. In addition, the recently discovered lncRNAs are comprehensively summarized within the LNCipedia database (https://lncipedia.org/), which currently consists of 120,353 human lncRNA transcripts that are obtained from different sources, e.g., RefSeq, Ensembl, and Noncode [13]. As of March 2018, the LncRNA Database (http://www.lncrnadb.org/), a repository of lncRNAs curated from evidence and supported by the literature, lists 184 biologically validated lncRNAs in humans [14].
3. Versatility of functionality: The diverse biological impact of (l)ncRNAs toward multiple layers, such as chromatin remodeling (signal and/or scaffold), chromatin interactions, competing endogenous mRNAs, and natural antisense transcripts, shows that interactions at genomic, transcriptomic, and protein levels are possible and, thus, are very difficult to encounter by a single computational algorithm [15].

2 Experimental and Computational Identification of Novel ncRNAs

The first step in the characterization of ncRNAs—the discovery itself—is the most crucial one in the process, but is often based on inadequate sequencing experiments. The determination of the scientific problem rather than the affordability of the technology should drive the investigation [16]. For this reason, we are going to highlight common experimental and computational practices for

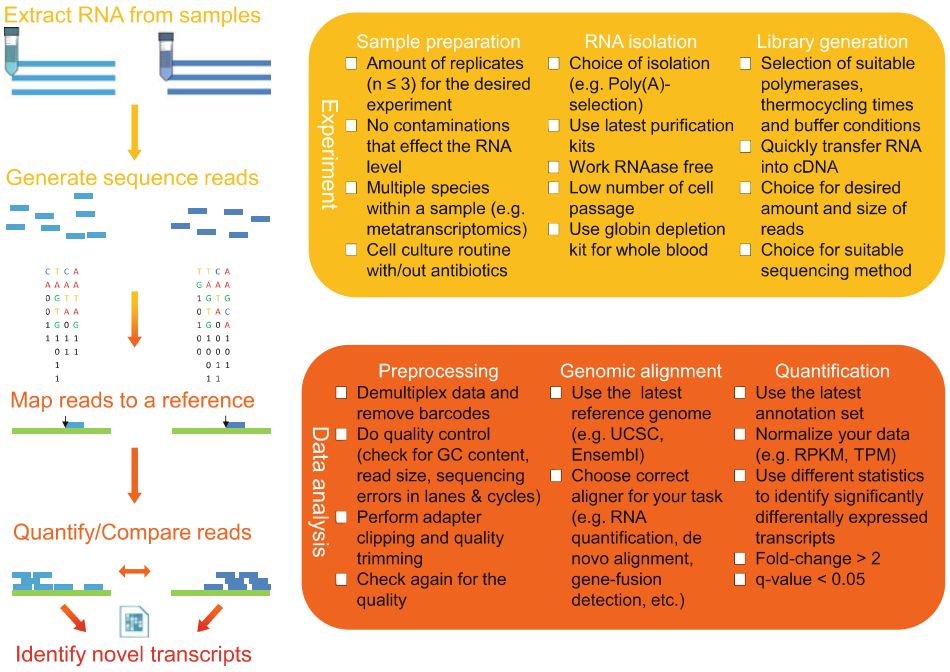


Fig. 2 Integrated experimental and computational workflow with specific checkboxes for the identification of ncRNAs from RNA-Seq datasets

ncRNA discovery. As an example, a workflow starting from sample preparation to data analysis and processing of already known and novel transcripts can be seen in Fig. 2.

2.1 Experimental Procedures for Proper ncRNA Identification

The initial step toward a successful experiment starts with the process of sample preparation, which involves already numerous decisions that are specifically dependent on the RNA subtype and species of interest. An overview of the extensive sequencing methodologies for the different platforms has been recently published by Tripathi et al. [17]. As an example to ensure clinical grade quality of the NGS technology, the Korean Society of Pathologists developed laboratory guidelines for NGS cancer panel testing procedures and requirements for clinical implementation of NGS [18]. The suggested laboratory part addresses important issues across multistep NGS cancer panel tests including the choice of the gene panel and platform, sample handling, nucleic acid management, sample identity tracking, library preparation, sequencing, data analysis, and reporting to the patient.

In research-oriented sequencing experiments, at least three biological replicates per condition are necessary to enable meaningful downstream statistical comparisons for the genome-wide detection of significant differences. In a recent review, Ouzain et al. [19]

found that for exploratory analyses of homogeneous samples (e.g., in cell lines or tissues from genetic mouse mutants and controls) in a highly controlled experimental setup, three biological replicates at high read depth can provide on the one hand sufficient reads to detect novel lncRNAs and on the other sufficient power to detect statistical differences between the conditions. However, the number of samples required will depend strongly on the biological and technical variability of an experiment and, of course, in a real-world setting involving the use of clinical patient samples many more samples would be required, but are rarely obtained, to achieve a similar statistical power [19]. Some other experimental aspects that have to be considered with caution are the isolation of the sample via commercially available purification kits itself, influence of the cell culture passage and tissue origin, poly(A)-tail selection of the RNA transcripts, globin depletion while working with native blood samples, or effect of antibiotics in cell culture in general [20].

After the RNA isolation step, the cDNA library preparation has to be done, which evokes the risk of biases resulting from genomes with high or low GC content. This should be in fact avoided by optimizing the preparation step through careful selection of polymerases for PCR amplification, thermocycling, condition, and buffer optimization [17]. The total amount of reads is yet another important parameter in determining the genomic coverage in RNA sequencing (RNA-Seq), because during the sequencing experiment different reads are generated from different RNA libraries and, thus, the overall coverage is defined by the number of times a genomic region, at the single base pair level, is covered by a read [16]. The combination of the desired read length and the amount of reads defines the throughput of an instrument in number of bases per run. NGS technologies are still, unless very low, prone to sequencing errors, but these most randomly occurring incorrect base calls (probability approximately 0.0001) can be compensated by sequencing the same region multiple times, which would result in an increased coverage. Increasing the read coverage likewise increases the confidence of existing variations (e.g., SNPs) in the genome/transcriptome under investigation. However, it was shown that a too high coverage might be problematic as well, because the absolute number of sequencing errors will jointly increase with the coverage and will impact the quality of the genome assembly [21]. The fractionation and sizing of the reads (based on the application of interest, e.g., de novo genome alignments will need longer reads), especially the impact of the sequencing depth and read length on single-cell RNA sequencing data, have to be considered [22].

2.2 A User-Friendly Environment for Computational Data Analysis

The application of computational analyses in the life science plays an increasingly important role. A comprehensive database of such analysis tools for different omics datasets can be obtained from the OMICtools community (https://omictools.com/), which aims to accelerate the selection of the most appropriate tools for specific use

cases. An additional service of guided data analysis is provided by the German Network for Bioinformatics Infrastructure—de.NBI (https://www.denbi.de/)—which is a national infrastructure providing comprehensive, high-quality bioinformatics guidance to users in life science research and biomedicine. The European-wide bioinformatics support is coordinated and integrated by ELIXIR (https://www.elixir-europe.org/), which sustains bioinformatics resources across its member states and enables users in academia and industry to access computational services.

Reduced costs and increased accuracy of sequencing experiments enable the investigation of biological phenomena at a high resolution [23]. Despite the low technological entrance barriers of the already presented specialized experimental protocols, the challenge of proper, transparent, and reproducible data analyses is still a bottleneck [24]. With respect to the number of data analysis steps, including preprocessing, genomic alignment, and quantification, the complexity in tool selection, implementation, and benchmarking is increasing likewise, hence calling for more systematic approaches such as tool management frameworks [23, 25]. This means that many NGS tools being installed on desktop computers are inadequate for interdisciplinary collaborations and many researchers are eagerly looking for easily accessible cloud-computing solutions to provide scalable processing environments for sequencing data [26].

One of these cloud-computing solutions for RNA-centric research is the RNA workbench [27]. This platform is unique in combining available tools, workflows, and training material as well as providing easy access for experimentalists. The RNA workbench is built upon the Galaxy project, which is a framework that makes advanced computational tools accessible without the need of prior extensive training [28]. Galaxy seeks to make data-intensive research more accessible, transparent, and reproducible by providing a Web-based environment in which users can perform computational analyses and have all of the details automatically tracked for later inspection, publication, or reuse. In order to achieve long-term sustainability, it provides the essential resources on sustainable platforms such as BioConda (https://bioconda.github.io) and BioContainers [29], which are emerging solutions to deploy complete data analysis workflows, including all necessary tools and dependencies. Running the containerized RNA workbench simply requires installing Docker (https://www.docker.com/) and starting the Galaxy RNA workbench image [27]. For example, Wolfien et al. [30] and Schulz et al. [31] demonstrated successful implementations of a Galaxy/Docker-based workflow with discrete software applications for the analysis of NGS data. The provided layer of virtualization also allows the handling of user-defined input data in a secure and compartmentalized way, which is a key requirement for researchers working on sensitive data (e.g., patient data in clinics) [27].

2.3 Best Practices for Sequencing Data Analysis

Most NGS technologies currently available are based on sequencing a large number of fragments (thousands to millions) in parallel within a single-flow cell that pools multiple samples per run. Assuming demultiplexed RNA-Seq data, which means that the individual samples have been separated by specific barcodes, the processing usually starts with the quality control of the raw reads provided in the fastq format. In addition to the sequence of nucleotides, the fastq format also provides a quality value, i.e., Phred score, for each of the sequenced bases. In the first step, an evaluation of these quality values as well as the calculation of the GC content, read duplication levels and contaminations are crucial for any further analysis. The quality control tools FastQC [32], NGS QC Toolkit [33], or Qualimap2 [34] calculate multiple quality statistics and create visualizations for sequencing data, which can be used to fine-tune parameters and further downstream processing steps.

Adapter sequences that are added during the experimental steps do not provide any additional information and can therefore be removed by using various tools such as Cutadapt [35], Skewer [36], or Reaper from the Kraken package [37]. Often adapter clippers are already integrated into quality score trimming software like Trimmomatic [38] or TrimGalore! [39]. After removing the adapters, a quality trimming step that is removing low-quality parts of a read (usually quality score <20 is removed) is essential and improves the reliability of the subsequent analyses [23].

Numerous algorithms have been developed to align the individual reads onto a reference genome. The most popular tools are aligners like TopHat2 [40], HiSat2 [41], STAR [42], or Segemehl [43]. A comparison of different alignment tools regarding accuracy/mismatch frequency, splice site detection, and performance was already done [44]. Alignment-free quantification methods for the quantification of RNA-Seq such as Kallisto [45] and Salmon [46] are faster and additional resource-sparing analyses, because they efficiently use the structure of a reference sequence without performing full base-to-base alignments, which is most time consuming. Nevertheless, these alignment-free quantification algorithms perform only with similar accuracy compared to traditional approaches when applied to ordinary tasks like transcript quantifications, clustering, and isoform prediction [47]. With respect to the identification of noncoding transcripts, the reads should preferably not accumulate mismatches in seed regions, but can be truncated. Furthermore, most mapping tools allow for a multiple or unique mapping strategy, which means that reads may be aligned to multiple regions, pseudogenes, and regions of low complexity in genomic references or numerous isoforms in transcriptomic references [23, 48].

After aligning the reads onto the reference genome (e.g., from UCSC or Ensembl), they have to be quantified and compared

across different samples to finally obtain the differentially expressed genes. To be able to compare two or more samples from the same or different sequencing runs, it is inevitable to normalize for the varying library/read size and length. For this reason, different statistical methodologies like *transcripts per million* (TPM) or *reads/fragments per kilobase per million mapped reads* (R/FPKM) have been developed [48]. Commonly used tools for the detection of differentially expressed genes are Cuffdiff [49], DESeq2 [50], and Sleuth [51] that are likewise used subsequently after applying TopHat2, STAR, or Kallisto. A recent study about the investigation of host-pathogen interactions, with a special focus on lncRNAs, showed the incorporation of an additional filter for enhancing the accuracy of differentially expressed genes [52].

Being able to get further self-paced training within the complex field of sequencing data analysis and the usage of evolving complex workflows, the Galaxy training network, which is a community-driven framework, enables interested users modern, interactive training for data analytics in life sciences and, therefore, facilitates the general use of NGS [53]. The Galaxy training network community combines online tutorials with a Web-based analysis framework to empower biomedical researchers to perform computational analyses themselves through a Web browser without the need to install software or search for tutorial datasets (http://galaxyproject.github.io/training-material/).

3 Linking Novel ncRNAs to Already Known Features

Ultimately, the functionality of (l)ncRNAs should be determined and/or tested by using experimental approaches, however, such experimental approaches like gene knockdown, overexpression, or CRISPR-Cas editing are typically too time and cost intense for an application toward an extensive pool of identified candidates [54]. Fortunately, numerous promising, cost-efficient in silico methods have been developed to overcome this experimental bottleneck by mathematical algorithms, structure and topological based features, network methodologies, or machine learning approaches. These individual approaches often do not require any novel generated experimental data, but are reusing publicly available data obtained from databases or repositories, e.g., Gene Ontology (http://www.geneontology.org/), STRING (https://string-db.org/), or Reactome (https://reactome.org/). A specific overview about computational life science databases was summarized from Hall et al. [55]. A comprehensive database for lncRNAs in human and mouse is LncRBase (http://bicresources.jcbose.ac.in/zhumur/lncrbase/), which hosts information on basic lncRNA transcript features, with additional details on genomic location, overlapping small noncoding RNAs, associated repeat elements,

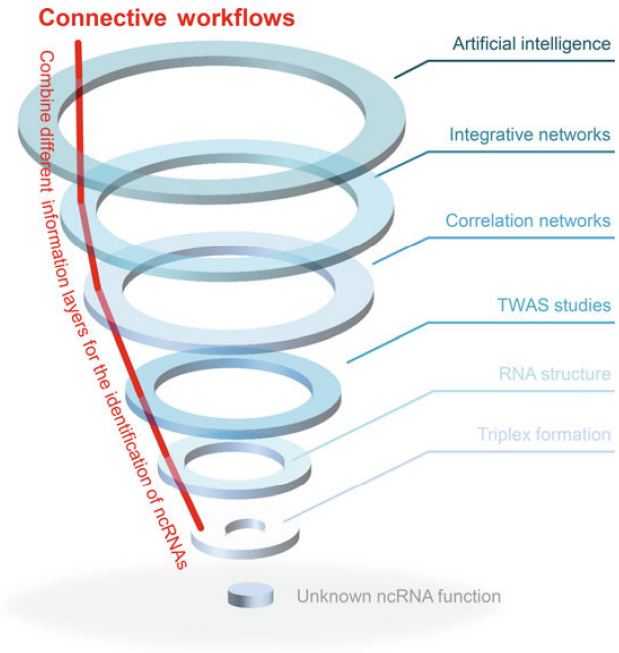


Fig. 3 Using connective workflows for the functional characterization of ncRNAs

imprinted genes, and lncRNA promoter information [56]. Users can also search for microarray probes mapped to specific lncRNAs and associated disease information as well as search for lncRNA expression in a wide range of tissues. In the following, we will show current approaches to integrate commonly used mathematical algorithms, topological features, and network modeling to characterize newly identified or uncharacterized ncRNAs (Fig. 3).

3.1 Combining Genomics Knowledge with Transcript Information: SNPs and TWAS

Genome-wide association studies (GWAS) are already applied to large numbers of individuals in a population or across multiple populations and result in associating their individual genome-wide genotypic variations to the personal respective phenotype. In a complementary manner, it is possible to perform a transcriptome-wide association study (TWAS) to identify significant expression-trait associations [57]. The obtained results from such a study showcase the power of integrating genotype and gene expression (mRNA and ncRNA) information together with phenotype data to gain insights into the genetic basis of complex traits.

Recently, Lopez-Meastre et al. [58] proposed a method that identifies, quantifies, and annotates SNPs without any reference genome from RNA-Seq data. The basis of their study is to identify the variants related to a phenotype, whereas SNPs are called de

novo from the reads, without separating the steps of assembly and SNP calling. The clear advantage is that it can be applied to non-model species, without a reference genome being available. It was likewise shown that the SNP calling methods could be tailored to have a good precision, meaning that most of the reported SNPs are true SNPs. Clearly, only SNPs from transcribed regions can be targeted, but they arguably correspond to those with a more direct functional impact. RNA-Seq experiments may also provide very high depth at specific loci and, therefore, allow discovering infrequent alleles in highly expressed genes. Finally, pooling samples is already extensively used in DNA-Seq (sometimes termed Pool-Seq) [59]. They outlined that, even though the case studies presented included only two replicates, the method can be applied to any number of replicates and the key contribution is that they are able to produce a list of SNPs stratified by their impact on the protein sequence [58].

3.2 Using RNA-Triplexes to Refine Cooperative miRNA Target Prediction Possibilities

The biological phenomenon of cooperating miRNAs, which is a pair of miRNAs that is able to synergistically regulate mutual targets (mRNAs) to compel a more effective target repression, has been recently reported and shown to be meaningful for the functional characterization of formerly uncharacterized miRNAs [60, 61]. Their workflow can be used to identify so-called RNA triplexes and determine the respective functionality of cooperative target regulation by two miRNAs for each triplex [61]. The algorithm and the underlying data have been implemented in the triplexRNA database (https://triplexrna.org/), which contains predicted RNA triplexes composed of two cooperatively acting microRNAs (miRNAs) and their mutual target mRNAs for humans and mice. They derived experimentally confirmed miRNA-gene interactions from miRTarBase, a manually curated, literature-based database of validated miRNA-target interactions based on different experimental methods (release 6.0) [62] and complemented the set of validated miRNAs by predictions from the highly sensitive miRNA-target prediction algorithm miRanda [63]. An exemplary use case scenario for this application can be obtained from Lai et al. [64] who show the integration of bioinformatics, structural and kinetic modeling, as well as experimental validation to study the cooperative regulation of E2F1 by miRNA pairs in the context of anticancer chemotherapy resistance.

3.3 From RNA Structure to Biological Functionality

Structural versatility is another explanation for ncRNAs to have various different functions and, therefore, specific algorithms attributed to the primary, secondary, or tertiary structure provide new insights into the respective functionalities. Veneziano et al. [65] reviewed the computational methods for RNA structure prediction that have been adopted to analyze the structure of circRNAs, small ncRNA, as well as lncRNAs and have been already

shown to provide indispensable information for further in-depth investigation. Another current review by Yan et al. [66] focuses on mainstream RNA structure prediction methods at the secondary and tertiary levels. They conclude that ongoing improvements in the accuracy of ncRNA structural prediction contribute to reliable predictions for the tertiary structure of small RNA molecules, but are lacking in the accurate prediction for the structure of large RNA molecules or those with complex topological structures [66]. A third review by Guo et al. [67] likewise concludes that the structures predicted by the computational methods still retain a high false-positive rate and the distinct structure–function relationships for many lncRNAs are still unknown, but with respect to the structure level of lncRNAs components discovered in the lncRNA secondary structures are of great value for further analysis, especially based on high-throughput sequencing technologies [67].

Two collective approaches to accumulate structure-based knowledge about RNAs are the Rfam database (http://rfam.sanger.ac.uk), which categorizes ncRNAs and their conserved primary sequence and RNA secondary structure through the use of multiple sequence alignments, consensus secondary structure annotation, as well as covariance models. In addition, with respect to the structure of lncRNAs, there is LNCipedia (http://www.lncipedia.org), a novel database for human lncRNA transcripts and genes [68]. LNCipedia offers 21,488 annotated human lncRNA transcripts obtained from different sources and includes basic transcript information, the gene structure, secondary structure information, protein coding potential, and miRNA-binding sites.

3.4 Integrative Network Approaches to Investigate the Regulative Potential of ncRNAs

Recent advances in systems biology shed light into the regulation of different pathways that investigate the interaction of genes, resulting in biological networks depicted as graphs. Especially, ncRNAs are highly connected within gene interaction networks and can therefore influence numerous targets to drive a specific biological response and the fate of cells or tissues. It has been shown that ncRNAs are particularly relevant in various research fields (e.g., cancer), in which ncRNAs act as main drivers or suppressors and can be seen as key regulators of physiological programs in developmental and disease-specific contexts [1]. It has become increasingly difficult to investigate ncRNA functionality in an isolated manner, because, e.g., miRNAs usually target mRNAs of various genes and, likewise, the mRNA of each gene can be targeted by multiple miRNAs that means these ncRNAs naturally link associated genes into regulatory networks [69, 70].

By the means of the previously discussed transcriptomics technologies, such as RNA-Seq, the activity of genes can be measured and integrated into a molecular network of choice to gain deeper insights into the gene expression in general. The recently published

KeyPathwayMiner enables the extraction and visualization of interesting subnetworks from a larger network based only on a series of gene expression datasets [71]. After applying the tool KeyPathwayMiner, one is able to identify the important subnetworks within a constructed large-scale molecular protein-protein interaction (PPI) network (e.g., based on all interactome information) to demonstrate known molecular interactions between significantly upregulated genes [72]. Once such a network is developed, even more sophisticated mathematical models can be applied to the network. Based on the available information, one can employ ordinary differential equations (ODE), discrete modeling, or hybrid modeling (composed of ODE and logic sub-modules) to dynamically analyze the networks for stimulus-response behavior and in silico perturbations [73]. It has been shown that these kinds of network approaches are able to identify disease-specific regulatory cores within large gene networks and, moreover, to predict receptor signatures associated with certain diseases [74].

In general, network enrichment methods combine experimental transcriptomic and proteomic data to be able to extract subnetworks from data-derived setups. Nevertheless, the integration of time series expression data with such network approaches is still challenging, thus limiting the identification of time-dependent responses. To overcome this limitation, Wiwie et al. combined human-augmented clustering with a novel approach for network enrichment to find temporal expression prototypes that are mapped to a network [75]. Their developed Time Course Network Enrichment (TiCoNE) methodology investigates enriched prototype pairs that interact more often than expected by chance.

So far, network-based computational studies investigate the hypothetic functions of lncRNAs through identifying molecules interacting with them, but since there are only a few molecular interactions known for multitudes of lncRNAs the application of these methods is rather difficult.

3.5 Correlation Networks to Link Uncharacterized ncRNAs to Known Annotations

It is well accepted that co-expressed genes are more likely to be co-regulated and, therefore, functionally related [76]. Thus identifying co-expressed protein-coding genes can help to assign the functions of uncharacterized ncRNAs [54], which has been successfully applied to study protein-coding genes, like the mammalian γ2 AMPK, that regulate the intrinsic heart rate [77].

For this reason, correlation networks are considered to be increasingly important bioinformatics applications, especially the weighted gene co-expression network analysis, which describes the correlation patterns among genes across multiple samples. This weighted correlation network analysis (WGCNA) is a guilt-by-association (GBA) approach for constructing co-expression networks based on gene expression data that is subsequently used for finding clusters (modules) of highly correlated genes [78]. This

analysis supports not only the identification of co-regulated ncRNAs within specific modules, but also the investigation of intramodular hub genes (e.g., transcription factors) that may connect different pathways for relating different clustered modules together or even associate them to external sample traits (e.g., environmental factors) [78].

Based on public RNA-Seq datasets of four solid cancer types, Li et al. utilized WGCNA and proposed a strategy for exploring the functions of lncRNAs altered in more than two cancer types [79]. WGCNA in combination with DAVID (the database for annotation, visualization, and integrated discovery https://david.ncifcrf.gov/home.jsp) [80] can be used to explore the underlying associated functions of the identified modules. Their results indicate that cancer-expressed lncRNAs show high tissue specificity and likely play key roles in the multistep development of human cancers, covering a wide range of functions in genome stability maintenance, signaling, cell adhesion and motility, morphogenesis, cell cycle, immune, and inflammatory response, whereas the lncRNAs are lower expressed than protein-coding genes.

In addition to gene set enrichment analyses performed by the well-known DAVID tool, one can also use Enrichr (http://amp.pharm.mssm.edu/Enrichr), which currently contains a large collection of diverse gene set libraries available for analysis and download [81]. In total, Enrichr currently contains 234,849 annotated gene sets from 128 gene set libraries. New features have been added to Enrichr including the ability to submit fuzzy sets and upload BED files, improved application programming interface, and visualization of the results as clustergrams [82]. Overall, Enrichr is a resource for curated gene sets and a search engine that accumulates biological knowledge for further in-depth discoveries.

3.6 Artificial Intelligence-Guided Identification of ncRNA Functionality

Identifying meaningful information from huge data in bioinformatics, machine learning (ML) or artificial intelligence (AI) has evolved to one of the most dominant approaches in this area. Supervised and unsupervised ML algorithms are likewise using training data to look for characteristic patterns, build a mathematical model around the training data, and, finally, predict new data or data that has been left out for training (e.g., normalized by ten-fold cross-validation). The data is usually preprocessed by removing features with low variance and high correlation for initial dimension reduction and, therefore, following best practice recommendations [83]. Frequently used algorithms are support vector machines (SVMs), Bayesian networks (BNs), random forest (RF), boosting, or hidden Markov models (HMM) that have been applied across various omics fields [84, 85]. Small clinical datasets are often prone to overfitting, which is the reason why it is important to choose classifiers that are suitable for training on small datasets for a comparison of features given little training and choose the most

appropriate algorithm according to accuracy and robustness toward overfitting [86]. Classical ML algorithms have limitations in processing the extensive amount of raw data that made researchers to predefine sets of suitable high-abstraction-level features, which can be used by the algorithms [87]. This time-consuming step could only be managed with considerable good domain expertise. In contrast to supervised ML algorithms, unsupervised approaches need no specific ground truth to train the actual model, but based on their nonlinear dimensional reduction they are less effective to identify a specific set of important features. These unsupervised statistical learning approaches, such as t-distributed stochastic neighbor embedding (t-SNE), assume that there are naturally occurring subclasses within sample sets that behave differently yet reproducibly across a number of populations and varying scenarios (e.g., treatment/control case, environments) [88]. In the following, we highlight specific ML-based tools and algorithms that have been recently applied to characterize ncRNAs.

iSeeRNA [89] is an SVM-based classifier, which can accurately and quickly identify lncRNAs from expression datasets (e.g., RNA-Seq) by using conservation open reading frame- and nucleotide sequence-based features in order to appropriately classify lncRNAs from protein-coding genes. Due to the lack of the aforementioned annotation for novel transcripts, they did not include homology search-based features, because this would enlarge the false-positive rate for predictions. In addition, customized SVMs for other species of interest can be trained and built upon on own datasets.

Xiao et al. predicted functions of lncRNAs through the construction of a regulatory PPI network between lncRNAs and protein-coding genes [90]. By integrating RNA-Seq data, they have been able to construct transcript profiles for the lncRNAs and protein-coding genes. After applying their Bayesian network approach, which implies dependency relations between lncRNAs and protein-coding genes, toward the initial regulatory network, a refined network was built. The integration of the highly connected coding genes and a single given lncRNA was subsequently used to predict functions of the lncRNA through functional enrichment within the PPI network [90].

The identification of ncRNAs within genomic regions can be done by a classification tool that was developed based on a hybrid RF and logistic regression model to classify short ncRNA sequences as well as long complex ncRNA sequences [91]. This classifier was trained and tested on a dataset with an achieved accuracy of 92.11%, sensitivity of 90.7%, and specificity of 93.5%. The authors also introduced a so-called SCORE feature, which is generated based on a logistic regression function that combines five significant features (structure, sequence, modularity, structural robustness, and coding potential) to enable an improved characterization of

lncRNAs. They showed that the use of SCORE improved the performance of the formerly used RF-based classifier in the identification of Rfam lncRNA families [91].

Deep learning (DL), a subtype of machine learning algorithms, has emerged recently on the basis of big data, the power of parallel and distributed computing, and even more sophisticated mathematical algorithms. DL algorithms have overcome the former limitations of manually, handcrafted feature selection and are making major advances in diverse fields such as image recognition, speech recognition, and natural language processing [92]. Recently, there have been studies published that focused in particular on deep learning algorithms for the prediction of ncRNAs [93]. The main advantage of using DL approaches is that they do not require pre- and post-processing classification steps to handle the raw big data formats.

The developers of *deepTarget*, an end-to-end machine learning framework for miRNA target prediction, showed that even without any known features there are substantial performance boosts over existing miRNA target detectors [93]. DeepTarget uses deep recurrent neural networks and does not depend on any sequence alignment processing, which is being considered as indispensable in many bioinformatics workflows to identify meaningful differences between samples. As highlighted in the previous section, the numerous alignment algorithms involve parameter optimization strategies and the obtained results are often not reproducible, because the different alignment methods are not intercomparable in terms of allowance for mismatches and base-to-base comparison.

MiRTDL is another novel miRNA target prediction algorithm and based on convolutional neural networks (CNN) [94]. The authors showed that their CNN-based approach automatically extracts the most important information from the formerly created balanced training datasets and is then applied to 1606 experimentally validated miRNA target pairs. Finally, their results indicate that miRTDL performs better in comparison to existing target prediction algorithms and achieves significantly higher sensitivities.

4 Connective Workflow Development

Due to the advances in omics technologies, life science research is coming toward a detailed molecular level at single-nucleotide resolution. Each on its own, these technologies have contributed numerous advances in the field of genomics, transcriptomics, proteomics, and metabolomics. However, each technology individually cannot capture the entire biological complexity across all the given layers of most human diseases and, therefore, needs further integration of multiple levels as a combined approach to provide a

more comprehensive view of the underlying biological phenomenon [95].

We already presented workflow management frameworks and cloud-computing services that are responsible for bridging the gap between tool developers and end users and showed that workflows facilitate the use of state-of-the-art computational tools, which would be difficult to access for nonexperts without graphical user interface frameworks [23]. However, the use of single workflows for specific tasks (e.g., the presented RNA-Seq analysis workflow) can be even more facilitated by the assembly of multiple workflows into a single connective workflow. Such universal connective workflows could be used to apply multilayered approaches and can incorporate several independent algorithms to test/benchmark different workflows against each other (Fig. 3). This would facilitate the certainty of the obtained knowledge, because independent algorithms, such as WGCNA and ML/DL, can be combined for the functional characterization of novel transcripts. The anticipated strategy for interoperable standards of workflows, namely the common workflow language, joins command-line tools across multiple platforms to workflows and, likewise, offers a modular concept for functional workflows that are built around containerized software solutions (documentation available at https://www.commonwl.org/). In order to adapt tools and workflows over time and ensure reusability and sustainability, we recommend keeping up track with the changes in the tools by a registration in platforms such as OMICtools (https://omictools.com/) or bio.tools (https://bio.tools/) [96], where tools are described by means of the meta-descriptive EDAM Ontology [97].

5 Conclusion

Each and every in silico predicted functionality of an ncRNA should be experimentally validated. Computational methods are used to accelerate the generation of new hypotheses or to narrow down specific molecular candidates. The interplay of model-driven experimentation and data-driven modeling could be seen as a guiding principle for the integration of the different interdisciplinary needs and expertise to achieve a task like characterization of a novel ncRNA [98].

The ideal validation of a new transcript would likewise involve the same layer of identification (e.g., ncRNAs validated via real-time PCR or Northern blot) as well as other layers of identification (e.g., ncRNA was co-immunoprecipitated with a protein) and, ultimately, a functional clarification by means of a knockout or overexpression experiment. If the elimination of an ncRNA affects a biological process, which is required for the proper development or homeostasis of the organism, can be checked very feasible with

the advent of CRISPR/Cas technology [99]. In contrast to the golden path of experimental characterization, Palazzo and Lee [100] identified a scenario where experimental validation cannot give meaningful results, because a given ncRNA may only have a small impact on a biological process and, thus, results only in a small reduction of a relevant feature (e.g., reducing the number of offspring by 0.1%). Such small effects would be difficult to detect in a laboratory setting, but would be strongly selected against in the natural environment and would indicate that the specific ncRNA has a function [100].

The life sciences are mainly driven by technological and algorithmic developments that will both have steady impact toward the field. The best and most recent examples belong to the development of the NGS methodology and the still accelerating pace of AI algorithms within all research fields. The joint potential of both fields was demonstrated by the study of Scarano et al. [101], who showed the applicability of strand-specific RNA-Seq data in gene prediction. They also implied that libraries covering different organs, tissues, developmental stages, and a range of stress conditions are necessary to get meaningful annotation-specific genes. However, there is still a growing need in individual developments of both fields such as nanopore direct RNA-Seq, which allows for single-molecule sequencing that circumvents reverse transcription or amplification steps and, therefore, enables a real-time RNA-Seq technology [102]. Equally important are new DL algorithms that can be trained on genomic or transcriptomic data. Those algorithms can currently build predictive models of RNA-processing events such as splicing, transcription, and polyadenylation. When applied to clinical data, the algorithms were able to identify mutations and flag them as pathogenic, even though they have never seen clinical data for training [103].

To address the continuous growing number of identified ncRNAs by means of sequencing experiments, it is inevitable to look for computational guidance to rank, predict, or score the most likely functionally important transcripts. From experimental design to computational data analysis, it still needs specific domain expertise that can be satisfied only by interdisciplinary research teams or large community efforts.

Acknowledgments

We acknowledge the partners and management of the German Network for Bioinformatics Infrastructure (de.NBI) for continuous support and guidance. Financial support for this work by the German Federal Ministry for Education and Research (BMBF) and European Social Fund (ESF) is greatly acknowledged (Grant 031L0106C, 02NUK043C, ESF/14-BM-A55-0027/18).

References

1. Anastasiadou E, Jacob LS, Slack FJ (2017) Non-coding RNA networks in cancer. Nat Rev Cancer 18:5–18. https://doi.org/10.1038/nrc.2017.99
2. Delihas N (2015) Discovery and characterization of the first non-coding RNA that regulates gene expression, micF RNA: a historical perspective. World J Biol Chem 6:272. https://doi.org/10.4331/WJBC.V6.I4.272
3. Schmitz U, Naderi-Meshkin H, Gupta SK et al (2016) The RNA world in the 21st century—a systems approach to finding non-coding keys to clinical questions. Brief Bioinform 17:380–392. https://doi.org/10.1093/bib/bbv061
4. Tripathi R, Chakraborty P, Varadwaj PK (2017) Unraveling long non-coding RNAs through analysis of high-throughput RNA--sequencing data. Non-coding RNA Res 2:111–118. https://doi.org/10.1016/J.NCRNA.2017.06.003
5. Xuan J, Yu Y, Qing T et al (2013) Next-generation sequencing in the clinic: promises and challenges. Cancer Lett 340:284–295. https://doi.org/10.1016/j.canlet.2012.11.025
6. Metzker ML (2010) Sequencing technologies—the next generation. Nat Rev Genet 11:31–46. https://doi.org/10.1038/nrg2626
7. Bernhart SH, Hofacker IL (2009) From consensus structure prediction to RNA gene finding. Briefings Funct Genomics Proteomics 8:461–471. https://doi.org/10.1093/bfgp/elp043
8. Derrien T, Johnson R, Bussotti G et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22:1775–1789. https://doi.org/10.1101/gr.132159.111
9. Moran VA, Perera RJ, Khalil AM (2012) Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. Nucleic Acids Res 40:6391–6400. https://doi.org/10.1093/nar/gks296
10. Bejerano G, Pheasant M, Makunin I et al (2004) Ultraconserved elements in the human genome. Science 304:1321–1325. https://doi.org/10.1126/science.1098119
11. Johnsson P, Lipovich L, Grandér D, Morris KV (2014) Evolutionary conservation of long non-coding RNAs; sequence, structure, function. Biochim Biophys Acta 1840:1063–1071. https://doi.org/10.1016/J.BBAGEN.2013.10.035
12. Hammond SM (2015) An overview of microRNAs. Adv Drug Deliv Rev 87:3–14. https://doi.org/10.1016/j.addr.2015.05.001
13. Volders P-J, Verheggen K, Menschaert G et al (2015) An update on LNCipedia: a database for annotated human lncRNA sequences. Nucleic Acids Res 43:D174–D180. https://doi.org/10.1093/nar/gku1060
14. Quek XC, Thomson DW, Maag JLV et al (2015) lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. Nucleic Acids Res 43:D168–D173. https://doi.org/10.1093/nar/gku988
15. Fang Y, Fullwood MJ (2016) Roles, functions, and mechanisms of long non-coding RNAs in cancer. Genomics Proteomics Bioinformatics 14:42–54. https://doi.org/10.1016/j.gpb.2015.09.006
16. Vincent AT, Derome N, Boyle B et al (2017) Next-generation sequencing (NGS) in the microbiological world: how to make the most of your money. J Microbiol Methods 138:60–71. https://doi.org/10.1016/J.MIMET.2016.02.016
17. Tripathi R, Sharma P, Chakraborty P, Varadwaj PK (2016) Next-generation sequencing revolution through big data analytics. Front Life Sci 9:119–149. https://doi.org/10.1080/21553769.2016.1178180
18. Kim J, Park W-Y, Kim NKD et al (2017) Good laboratory standards for clinical next-generation sequencing cancer panel tests. J Pathol Transl Med 51:191–204. https://doi.org/10.4132/jptm.2017.03.14
19. Ounzain S, Micheletti R, Beckmann T et al (2015) Genome-wide profiling of the cardiac transcriptome after myocardial infarction identifies novel heart-specific long non-coding RNAs. Eur Heart J 36:353–68a. https://doi.org/10.1093/eurheartj/ehu180
20. Ryu AH, Eckalbar WL, Kreimer A et al (2017) Use antibiotics in cell culture with caution: genome-wide identification of antibiotic-induced changes in gene expression and regulation. Sci Rep 7:7533. https://doi.org/10.1038/s41598-017-07757-w
21. Ekblom R, Wolf JBW (2014) A field guide to whole-genome sequencing, assembly and annotation. Evol Appl 7:1026–1042. https://doi.org/10.1111/eva.12178
22. Rizzetto S, Eltahla AA, Lin P et al (2017) Impact of sequencing depth and read length on single cell RNA sequencing data of T cells.

Sci Rep 7:12781. https://doi.org/10.1038/s41598-017-12989-x

23. Lott SC, Wolfien M, Riege K et al (2017) Customized workflow development and data modularization concepts for RNA-sequencing and metatranscriptome experiments. J Biotechnol 261:85–96. https://doi.org/10.1016/j.jbiotec.2017.06.1203
24. Spjuth O, Bongcam-Rudloff E, Dahlberg J et al (2016) Recommendations on e-infrastructures for next-generation sequencing. Gigascience 5:26. https://doi.org/10.1186/s13742-016-0132-7
25. Lampa S, Dahlö M, Olason PI et al (2013) Lessons learned from implementing a national infrastructure in Sweden for storage and analysis of next-generation sequencing data. Gigascience 2:9. https://doi.org/10.1186/2047-217X-2-9
26. Celesti A, Celesti F, Fazio M et al (2017) Are next-generation sequencing tools ready for the cloud? Trends Biotechnol 35:486–489. https://doi.org/10.1016/J.TIBTECH.2017.03.005
27. Grüning BA, Fallmann J, Yusuf D et al (2017) The RNA workbench: best practices for RNA and high-throughput sequencing bioinformatics in Galaxy. Nucleic Acids Res 45:D626–D634. https://doi.org/10.1093/nar/gkx409
28. Afgan E, Baker D, van den Beek M et al (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res 44:W3–W10. https://doi.org/10.1093/nar/gkw343
29. da Veiga Leprevost F, Grüning BA, Alves Aflitos S et al (2017) BioContainers: an open-source and community-driven framework for software standardization. Bioinformatics 33:2580–2582. https://doi.org/10.1093/bioinformatics/btx192
30. Wolfien M, Rimmbach C, Schmitz U et al (2016) TRAPLINE: a standardized and automated pipeline for RNA sequencing data analysis, evaluation and annotation. BMC Bioinformatics 17:21. https://doi.org/10.1186/s12859-015-0873-9
31. Schulz W, Durant T, Siddon A, Torres R (2016) Use of application containers and workflows for genomic data analysis. J Pathol Inform 7:53. https://doi.org/10.4103/2153-3539.197197
32. FASTQC (2010) Babraham Institute. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 20 Jun 2018
33. Patel RK, Jain M (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One 7:e30619. https://doi.org/10.1371/journal.pone.0030619
34. Okonechnikov K, Conesa A, García-Alcalde F (2016) Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32:292–294. https://doi.org/10.1093/bioinformatics/btv566
35. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10. https://doi.org/10.14806/ej.17.1.200
36. Jiang H, Lei R, Ding S-W, Zhu S (2014) Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15:182. https://doi.org/10.1186/1471-2105-15-182
37. Wood DE, Salzberg SL (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:R46. https://doi.org/10.1186/gb-2014-15-3-r46
38. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
39. TrimGalore! (2012) Babraham Institute. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/. Accessed 20 Jun 2018
40. Kim D, Pertea G, Trapnell C et al (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:R36. https://doi.org/10.1186/gb-2013-14-4-r36
41. Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357–360. https://doi.org/10.1038/nmeth.3317
42. Dobin A, Davis CA, Schlesinger F et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635
43. Hoffmann S, Otto C, Doose G et al (2014) A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. Genome Biol 15:R34. https://doi.org/10.1186/gb-2014-15-2-r34
44. Engström PG, Steijger T, Sipos B et al (2013) Systematic evaluation of spliced alignment programs for RNA-seq data. Nat Methods

10:1185–1191. https://doi.org/10.1038/nmeth.2722

45. Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525–527. https://doi.org/10.1038/nbt.3519
46. Patro R, Duggal G, Love MI et al (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14:417–419. https://doi.org/10.1038/nmeth.4197
47. Robert C, Watson M (2015) Errors in RNA-Seq quantification affect genes of relevance to human disease. Genome Biol 16:177. https://doi.org/10.1186/s13059-015-0734-x
48. Conesa A, Madrigal P, Tarazona S et al (2016) A survey of best practices for RNA-seq data analysis. Genome Biol 17:13. https://doi.org/10.1186/s13059-016-0881-8
49. Trapnell C, Hendrickson DG, Sauvageau M et al (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol 31:46–53. https://doi.org/10.1038/nbt.2450
50. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8
51. Pimentel H, Bray NL, Puente S et al (2017) Differential analysis of RNA-seq incorporating quantification uncertainty. Nat Methods 14:687–690. https://doi.org/10.1038/nmeth.4324
52. Riege K, Hölzer M, Klassert TE et al (2017) Massive effect on LncRNAs in human monocytes during fungal and bacterial infections and in response to vitamins A and D. Sci Rep 7:40598. https://doi.org/10.1038/srep40598
53. Batut B, Hiltemann S, Bagnacani A et al (2017) Community-driven data analysis training for biology. bioRxiv: 225680. doi: https://doi.org/10.1101/225680
54. Signal B, Gloss BS, Dinger ME (2016) Computational approaches for functional prediction and characterisation of long noncoding RNAs. Trends Genet 32:620–637. https://doi.org/10.1016/J.TIG.2016.08.004
55. Smalter Hall A, Shan Y, Lushington G, Visvanathan M (2013) An overview of computational life science databases & exchange formats of relevance to chemical biology research. Comb Chem High Throughput Screen 16:189–198
56. Chakraborty S, Deb A, Maji RK et al (2014) LncRBase: an enriched resource for lncRNA information. PLoS One 9:e108010. https://doi.org/10.1371/journal.pone.0108010
57. Gusev A, Ko A, Shi H et al (2016) Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet 48:245–252. https://doi.org/10.1038/ng.3506
58. Lopez-Maestre H, Brinza L, Marchet C et al (2016) SNP calling from RNA-seq data without a reference genome: identification, quantification, differential analysis and impact on the protein sequence. Nucleic Acids Res 44:e148. https://doi.org/10.1093/nar/gkw655
59. Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. Nat Rev Genet 15:749–763. https://doi.org/10.1038/nrg3803
60. Lai X, Bhattacharya A, Schmitz U et al (2013) A systems' biology approach to study microRNA-mediated gene regulatory networks. Biomed Res Int 2013:703849. https://doi.org/10.1155/2013/703849
61. Schmitz U, Lai X, Winter F et al (2014) Cooperative gene regulation by microRNA pairs and their identification using a computational workflow. Nucleic Acids Res 42:7539–7552. https://doi.org/10.1093/nar/gku465
62. Chou C-H, Chang N-W, Shrestha S et al (2016) miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Res 44:D239–D247. https://doi.org/10.1093/nar/gkv1258
63. Betel D, Koppal A, Agius P et al (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol 11:R90. https://doi.org/10.1186/gb-2010-11-8-r90
64. Lai X, Gupta SK, Schmitz U et al (2018) MiR-205-5p and miR-342-3p cooperate in the repression of the E2F1 transcription factor in the context of anticancer chemotherapy resistance. Theranostics 8:1106–1120. https://doi.org/10.7150/thno.19904
65. Veneziano D, Nigita G, Ferro A (2015) Computational approaches for the analysis of ncRNA through deep sequencing techniques. Front Bioeng Biotechnol 3:77. https://doi.org/10.3389/fbioe.2015.00077
66. Yan K, Arfat Y, Li D et al (2016) Structure prediction: new insights into decrypting long

noncoding RNAs. Int J Mol Sci 17:132. https://doi.org/10.3390/IJMS17010132

67. Guo X, Gao L, Wang Y et al (2016) Advances in long noncoding RNAs: identification, structure prediction and function annotation. Brief Funct Genomics 15:38–46. https://doi.org/10.1093/bfgp/elv022
68. Volders P-J, Helsens K, Wang X et al (2013) LNCipedia: a database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res 41:D246–D251. https://doi.org/10.1093/nar/gks915
69. Ebert MS, Sharp PA (2012) Roles for MicroRNAs in conferring robustness to biological processes. Cell 149:515–524. https://doi.org/10.1016/J.CELL.2012.04.005
70. Yamamura S, Imai-Sumida M, Tanaka Y, Dahiya R (2018) Interaction and cross-talk between non-coding RNAs. Cell Mol Life Sci 75:467–484. https://doi.org/10.1007/s00018-017-2626-6
71. Alcaraz N, Kücük H, Weile J et al (2011) KeyPathwayMiner: detecting case-specific biological pathways using expression data. Internet Math 7:299–313. https://doi.org/10.1080/15427951.2011.604548
72. Hausburg F, Jung JJ, Hoch M et al (2017) (Re-)programming of subtype specific cardiomyocytes. Adv Drug Deliv Rev 120:142–167. https://doi.org/10.1016/j.addr.2017.09.005
73. Khan FM, Schmitz U, Nikolov S et al (2014) Hybrid modeling of the crosstalk between signaling and transcriptional networks using ordinary differential equations and multi-valued logic. Biochim Biophys Acta 1844:289–298. https://doi.org/10.1016/J.BBAPAP.2013.05.007
74. Khan FM, Marquardt S, Gupta SK et al (2017) Unraveling a tumor type-specific regulatory core underlying E2F1-mediated epithelial-mesenchymal transition to predict receptor protein signatures. Nat Commun 8:198. https://doi.org/10.1038/s41467-017-00268-2
75. Wiwie C, Rauch A, Haakonsson A, et al (2017) Elucidation of time-dependent systems biology cell response patterns with time course network enrichment. arXiv.org arXiv:1710.10262
76. Stuart JM, Segal E, Koller D, Kim SK (2003) A gene-coexpression network for global discovery of conserved genetic modules. Science 302:249–255. https://doi.org/10.1126/science.1087447
77. Yavari A, Bellahcene M, Bucchi A et al (2017) Mammalian γ2 AMPK regulates intrinsic heart rate. Nat Commun 8:1258. https://doi.org/10.1038/s41467-017-01342-5
78. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559. https://doi.org/10.1186/1471-2105-9-559
79. Li S, Li B, Zheng Y et al (2017) Exploring functions of long noncoding RNAs across multiple cancers through co-expression network. Sci Rep 7:754. https://doi.org/10.1038/s41598-017-00856-8
80. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57. https://doi.org/10.1038/nprot.2008.211
81. Chen EY, Tan CM, Kou Y et al (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14:128. https://doi.org/10.1186/1471-2105-14-128
82. Kuleshov MV, Jones MR, Rouillard AD et al (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 44:W90–W97. https://doi.org/10.1093/nar/gkw377
83. Caicedo JC, Cooper S, Heigwer F et al (2017) Data-analysis strategies for image-based cell profiling. Nat Methods 14:849–863. https://doi.org/10.1038/nmeth.4397
84. Kuhn M (2008) Building predictive models in *R* using the caret package. J Stat Softw 28:1–26. https://doi.org/10.18637/jss.v028.i05
85. Ray SS, Maiti S (2015) Noncoding RNAs and their annotation using metagenomics algorithms. Wiley Interdiscip Rev Data Min Knowl Discov 5:1–20. https://doi.org/10.1002/widm.1142
86. Saeb S, Lonini L, Jayaraman A, et al (2016) Voodoo machine learning for clinical predictions. bioRxiv: 059774. https://doi.org/10.1101/059774
87. Yu N, Cho KH, Cheng Q, Tesorero RA (2009) A hybrid computational approach for the prediction of small non-coding RNAs from genome sequences. In: 2009 International Conference on Computational Science and Engineering. IEEE, pp 1071–1076
88. van der ML, Hinton G (2008) Visualizing Data using t-SNE. J Mach Learn Res 9:2579–2605
89. Sun K, Chen X, Jiang P et al (2013) iSeeRNA: identification of long intergenic non-coding RNA transcripts from transcriptome sequencing data. BMC Genomics 14(Suppl 2):S7.

https://doi.org/10.1186/1471-2164-14-S2-S7
90. Xiao Y, Lv Y, Zhao H et al (2015) Predicting the functions of long noncoding RNAs using RNA-Seq based on Bayesian network. Biomed Res Int 2015:1–14. https://doi.org/10.1155/2015/839590
91. Lertampaiporn S, Thammarongtham C, Nukoolkit C et al (2014) Identification of non-coding RNAs with a new composite feature in the Hybrid Random Forest Ensemble algorithm. Nucleic Acids Res 42:e93. https://doi.org/10.1093/nar/gku325
92. Abbas Q, Raza SM, Biyabani AA, Jaffar MA (2016) A review of computational methods for finding non-coding RNA genes. Genes (Basel) 7:113. https://doi.org/10.3390/genes7120113
93. Lee B, Baek J, Park S, Yoon S (2016) deepTarget: end-to-end learning framework for microRNA target prediction using deep recurrent neural networks. arXiv.org arXiv:1603.09123
94. Cheng S, Guo M, Wang C et al (2016) MiRTDL: a deep learning approach for miRNA target prediction. IEEE/ACM Trans Comput Biol Bioinform 13:1161–1169. https://doi.org/10.1109/TCBB.2015.2510002
95. Karczewski KJ, Snyder MP (2018) Integrative omics for health and disease. Nat Rev Genet 19:299–310. https://doi.org/10.1038/nrg.2018.4
96. Ison J, Rapacki K, Ménager H et al (2016) Tools and data services registry: a community effort to document bioinformatics resources. Nucleic Acids Res 44:D38–D47. https://doi.org/10.1093/nar/gkv1116
97. Ison J, Kalas M, Jonassen I et al (2013) EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats. Bioinformatics 29:1325–1332. https://doi.org/10.1093/bioinformatics/btt113
98. Wolkenhauer O (2014) Why model? Front Physiol 5:21. https://doi.org/10.3389/fphys.2014.00021
99. Doudna JA, Charpentier E (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 346:1258096. https://doi.org/10.1126/science.1258096
100. Palazzo AF, Lee ES (2015) Non-coding RNA: what is functional and what is junk? Front Genet 6:2. https://doi.org/10.3389/fgene.2015.00002
101. Scarano D, Rao R, Corrado G (2017) In silico identification and annotation of non-coding RNAs by RNA-seq and de novo assembly of the transcriptome of Tomato Fruits. PLoS One 12:e0171504. https://doi.org/10.1371/journal.pone.0171504
102. Garalde DR, Snell EA, Jachimowicz D et al (2018) Highly parallel direct RNA sequencing on an array of nanopores. Nat Methods 15:201–206. https://doi.org/10.1038/nmeth.4577
103. Webb S (2018) Deep learning for biology. Nature 554:555–557. https://doi.org/10.1038/d41586-018-02174-z

2.1.5 Reproducible analyses to understand RNA interactions

Bagnacani, A., **Wolfien, M.**, and Wolkenhauer, O. (2019).
Tools for Understanding miRNA-mRNA Interactions for reproducible RNA Analysis.
Methods in Molecular Biology. Downloads: 1.1k, Citations (December 14, 2020): 0

MicroRNAs (miRNAs) are an integral part of gene regulation at the post-transcriptional level. In particular, miRNA-mRNA interactions in gene expression analyses became increasingly important to gain insights into the underlying regulatory mechanisms. As a result, we are confronted with a growing landscape of tools, while standards for reproducibility and benchmarking lag behind. This work identifies the challenges for reproducible RNA analysis and highlights best practices on the processing and dissemination of scientific results and connects it to new information. Here, we exemplarily use the TriplexRNA database[11] to show the importance of tools embedded into a larger processing framework, such as Galaxy.

In this work, I defined use cases for miRNA-mRNA interactions, as well as selections of relevant tools or combinations of RNA-Seq analyses into workflows. We think that the success of a tool does not solely depend on its performances: equally important is how a tool is received and supported in a community. For this reason, I graded different workflows based the achieved results, their accessibility, ease-of-use, and applicabilty for an RNA-Seq test case scenario. Our basis of such a transparent computational platform for sharing workflows and processing tools around RNA-centric data analysis was the Galaxy framework. Since there is currently no tool for miRNA cooperativity on mRNA integrated into Galaxy, we chose the TriplexRNA as a well suited candidate to test our hypotheses.

In summary, we used the community guidelines to extend the Galaxy portfolio of RNA tools with the integration of the TriplexRNA database to identify miRNA-mRNA relationships. Our findings are also providing a staring point for the development of a recommendation system, to guide users in the choice of tools and workflows.

[11] https://triplexrna.org

Chapter 8

Tools for Understanding miRNA–mRNA Interactions for Reproducible RNA Analysis

Andrea Bagnacani, Markus Wolfien, and Olaf Wolkenhauer

Abstract

MicroRNAs (miRNAs) are an integral part of gene regulation at the post-transcriptional level. The use of RNA data in gene expression analysis has become increasingly important to gain insights into the regulatory mechanisms behind miRNA–mRNA interactions. As a result, we are confronted with a growing landscape of tools, while standards for reproducibility and benchmarking lag behind. This work identifies the challenges for reproducible RNA analysis, and highlights best practices on the processing and dissemination of scientific results. We found that the success of a tool does not solely depend on its performances: equally important is how a tool is received, and then supported within a community. This leads us to a detailed presentation of the RNA workbench, a community effort for sharing workflows and processing tools, built on top of the Galaxy framework. Here, we follow the community guidelines to extend its portfolio of RNA tools with the integration of the TriplexRNA (https://triplexrna.org). Our findings provide the basis for the development of a recommendation system, to guide users in the choice of tools and workflows.

Key words miRNA–mRNA interactions, Gene regulation, RNA workbench, Galaxy, Database

1 Introduction to Computational Data Analysis Challenges in the Life Sciences

MicroRNAs (miRNAs) play a key role in gene regulation at the post-transcriptional level. Their presence can be correlated with the progression of diseases [1, 2], and are therefore used for the design of diagnostic and prognostic markers [3–5]. For these reasons, transcript quantification and discovery by RNA sequencing (RNA-Seq) is at the basis of diverse experiments in life science research [6]. RNA-Seq is a high-throughput technique that does not require predetermined DNA probes to known genes, and is therefore a key technology for the discovery of new exons, splice variants, and small RNAs [7].

The functional characterization of miRNA–mRNA interactions implies the use of specialized computational tools for RNA-Seq data analysis. This, from both users and developers' sides, is a process entailing challenges whose solutions do not yet leverage

Xin Lai et al. (eds.), *Computational Biology of Non-Coding RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1912, https://doi.org/10.1007/978-1-4939-8982-9_8,

on a unique corpus of standards, but rather a set of community-based best practices [8, 9].

From a developer's perspective, the need for designing and implementing a new computational tool for RNA analysis is dictated by the lack of the desired functions within all available software tools. This might be the case when the investigation is novel, or when available tools miss the necessary parametrization that allows room for testing the hypothesis under scrutiny. Another scenario can be that of achieving the desired computational approach by chaining existing tools in a *workflow*. Nonetheless, developers first look for tools already implementing part of the devised strategy, not to reinvent the wheel. In turn, choosing to implement new tools by leveraging on third-party modules, workflows, and frameworks can narrow down the types of input data format that are processed throughout the analysis. However, the use of a restricted set of data formats should not be seen as a limit, but rather an effort to provide de facto standards for reproducible analyses [8].

From a user's perspective, the need for adopting a computational tool for RNA analysis is dictated by the high availability of RNA-centric data, and by the applicability and usability of software tools. While it has become increasingly cheap to produce next-generation sequencing (NGS) data, the required efforts to gain insights by mining it have likewise increased in time and resources [10]. Furthermore, such tasks entail the adoption of statistical and computational methods, which are rarely the domain of a life science curriculum. As a result, users are often confronted with a new technical jargon, spanning from software interface tutorials to operative system-dependent installation instructions. Under such light, software solutions look like seas of alternatives to choose from.

Not surprisingly, community forums such as BioStars (https://www.biostars.org) and Stack Overflow (https://stackoverflow.com) have become hubs for sharing expertise across users of diverse expertise. Multiple RNA-centric tools in fact endorsed lively communities as support platforms, or started their very own development right from decentralized repositories such as GitHub (https://github.com) or BitBucket (https://bitbucket.org).

Although these platforms represent a valuable starting point to approach data analysis problems, there is still a high demand for providing guidance, and ultimately hands-on training on how to face such tasks. As a response, in recent years multiple training initiatives have started delivering topic- and/or tool-specific hands-on sessions and workshops to life scientists [11]. The European Life Sciences Infrastructure for Biological Information (ELIXIR, https://www.elixir-europe.org) training program TeSS (https://tess.elixir-europe.org), the Global Organisation for Bioinformatics Learning, Education & Training (Goblet, https://www.mygoblet.org), the German Network for Bioinformatics Infrastructure (de.NBI, https://www.denbi.de), the Software Carpentry

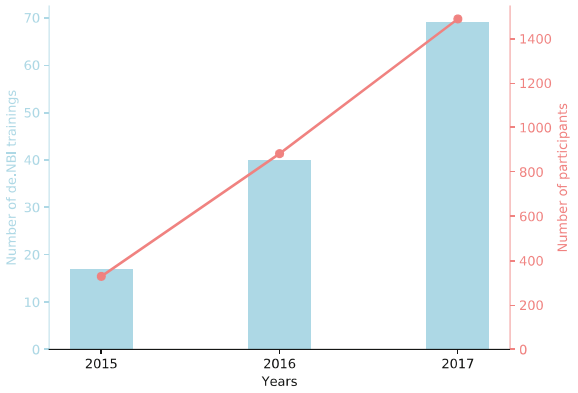


Fig. 1 Attendance to training courses organized by the German Network for Bioinformatics Infrastructure (de.NBI) between 2015 and 2017. The graph shows how the number of participants grows with the number of provided training courses. Trainings provided by de.NBI include topic- and/or tool-specific courses, focusing on computational approaches for life science data analyses. Further information is available at https:/www.denbi.de/training

(https://software-carpentry.org) initiative, and the Galaxy Training Network (https://galaxyproject.org/teach/gtn) are examples of active training providers across the globe. Figure 1 highlights how the attendance to topic- and tool-specific training events has increased in the last years within the de.NBI network.

However, the role of trainers becomes twofold: on the one hand they propose a biological problem and a computational solution that allows for tailored parametric options, while on the other they make sense of the whole plethora of software solutions, and provide a structured overview of tools, workflows, libraries, and frameworks to users. This interaction brings diverse technical jargons together, establishing a community where disparate expertise converge.

From a community's perspective, the adoption of a selected tool, workflow, or framework consolidates a computational approach into an established practice, paving the path for robust and reproducible data analysis [8]. Indeed, the success of a tool does not solely depend on how its algorithm performs: usability, as well as interoperability to different computing environments, plays a key role on the dissemination of the tool itself, and the community around it finally shapes its acceptance.

In this section, we overviewed the challenges that developers and users face when approaching life science problems computationally. Within this framework, the investigation of miRNA–mRNA interactions, for the acquisition of a better understanding of the phenomena and implications, represents a

subproblem holding the same traits: the need for usable software solutions developing around a set of standards, and a lively community able to share its diverse expertise, and provide guidance and best practices for reproducible science.

2 Online Catalogs of Bioinformatic Tools

In response to the growing amount of software solutions for downstream analysis in the life sciences, tools have been organized into registries: online indexes of tools, collected and categorized on the basis of their functions and target analyses.

Tool categorization has been implemented with diverse approaches and for different communities: for bioinformatic tools, established examples are the *Molecular Biology Database List* published by *Nucleic Acids Research*'s *Database Issues* [12, 13], the *Bioinformatics Links Directory* [14, 15], the *EMBRACE Registry* [16, 17], and *BioCatalogue Web services* [18].

In recent years, two new players have entered the scene: *OMICtools* [19] (https://omictools.com) and *bio.tools* [20] (https://bio.tools).

OMICtools is a Web-based search engine for biomedical resources. Its inventory is manually curated, and organized under a tailor-made taxonomy. Tools can be browsed by submitting keywords in a user-friendly input form. The whole inventory's descriptions are meant for human readers, and are compiled by both users and developers. Coupled with a catalog of tools, the service additionally integrates an editor where users can provide their review and feedback. Such a feature makes OMICtools not just a tool registry, but also a community hub.

Bio.tools is ELIXIR's effort of providing a manually curated registry of bioinformatic tools. By offering as well a user-friendly Web-based search engine to browse for computational resources, it furthermore provides the consultation of its catalog to both human and software agents. This is made possible by its Application Program Interface (API), which serves tool descriptions organized by means of the EMBRACE Data And Methods ontology (EDAM) [21]: a structured vocabulary of terms and their relations within the domain of bioinformatic (a) operations, (b) identifiers and data types, (c) data formats, and (d) topics.

From a developer's perspective, such catalogs represent a channel for the dissemination of software solutions, while users can directly benefit from the structured overview they offer for all available tools pertaining a specific biological problem or analysis.

Indeed, if a community of users promotes the adoption of an already implemented tool, this becomes established, and future users will save time and resources by channeling their research in

the direction of testing hypotheses, rather than implementing new prototypes apt at addressing the same set of questions.

Tools however might be modified to accommodate new features. For this reason, new *similar* tools might be implemented to make room for specialized parameters and functionalities.

Tool similarity and redundant functionalities are not synonyms of tool *duplication*. Similar tools can be used by different communities to investigate diverse aspects of a biological problem, as well as to benchmark their combined outcomes. For example, several algorithms have been devised and implemented for aligning individual sequence reads against a reference genome. This is the case of genomic alignment tools such as TopHat2 [22], HiSat2 [23], STAR [24], or Segemehl [25]. Indeed, the reason behind the development of all these tools is not to merely rebrand the problem of sequence alignment with different names, but rather to address specific applications, approaches, and parameter features such as accuracy, sensitivity, and splice-site detection.

A benchmark of the aforementioned tools has been in fact provided in terms of their compared performances [26]. Under this light, tool similarity represents a feature, rather than a problem of redundancy.

3 A Platform for Understanding miRNA–mRNA Interactions: The RNA Workbench

In a computational analysis, each tool represents a logical step toward the final result. Here, tools are *modules*, whose outputs become the inputs for the next modules at each successive iteration. Complex biological questions are addressed *computationally* by organizing tools into *workflows* which, in turn, comprise the minimal set of combinable modules whose functions are tailored for a specific analysis. As a consequence, each tool should be designed to be simple, extensible, easy to be maintained, and repurposed by other developers [27, 28]. Such design principles are necessary, albeit not sufficient, to make research reproducible. Indeed, even when sharing resources such as code, data, and parameter settings, the variability of computing environments makes it difficult to reproduce results. During workflow design, it is in fact common to leverage on specific module versions, and one or more software libraries.

These requirements constitute the set of *dependencies* that have to be satisfied at every run of the analysis. Moreover, reproducibility is tempered by the lack of standards, and the use of large dataset or multiple data sources (e.g., integrative omics experiments) [29].

For RNA analyses, all the aforementioned challenges have been addressed by the RNA workbench [8], a Galaxy [30] instance tailored for RNA-centric data analyses. The workbench combines a comprehensive set of tools for the analysis of RNA structures,

Table 1
An overview of the most prominent tools provided by the RNA Workbench, for the understanding of miRNA–mRNA interactions

Tool name	Tool application
ViennaRNA [31]	A suite of tools aiming at the prediction of RNA secondary structures. The ViennaRNA package covers predictions for optimal and suboptimal structures from single sequences, as well as sequence alignments, predictions of ensemble base-pair probabilities, and RNA–RNA interactions. Furthermore, it enables hard and soft constraint parametrizations for the prediction of RNA structures
LocARNA [32, 33]	Realizes a variant of the Sankoff algorithm [34] for comparative analysis of unaligned RNAs by simultaneously folding and aligning
PARalyzer [35]	Provides high-resolution maps of the interaction sites occurring between RNA-binding proteins and their targets
RNAz [36]	Predicts structurally conserved and thermodynamically stable RNA secondary structure
doRiNA [37]	A database for the investigation of the regulatory pattern occurring between RNA-binding proteins (RBPs) and miRNAs. Information about this phenomenon is accessible directly through Galaxy for the incorporation of doRiNA-based queries in custom computational pipelines

genomic aligners, RNA–RNA and RNA–protein interactions, RNA-Seq, ribosome profiling, genome annotation, and more. Table 1 provides an overview of the most prominent computational approaches for understanding miRNA–mRNA interactions offered by the RNA workbench. An up-to-date overview of its growing set of tools can be found online, at https://bgruening.github.io/galaxy-rna-workbench.

The key reasons for the development of a comprehensive RNA analysis framework relying on Galaxy are its scalability, which enables the RNA workbench to run on single-CPU installations as well as on large multi-node high-performance computing environments, and workflow reproducibility, which provides researchers with a means to share their own analyses with colleagues, as well as import third-party workflows.

Tool and tool-version dependencies are resolved via BioConda (https://bioconda.github.io): the bioinformatics channel for the Conda (https://conda.io) package manager. BioConda facilitates version-aware software packaging, enabling installation at user level. Finally, the workbench is containerized with Docker (https://www.docker.com) to deal with different installation environments and provide a platform agnostic suite for RNA analysis. This layer of virtualization also allows the handling of user-defined input data in a secured manner, which represents a crucial requirement for the analysis of sensitive data (e.g., patient data in clinics).

In this section we illustrated how the dissemination of tools and best practices can benefit from a comprehensive and system-agnostic environment, able to cope with software-dependency resolution. For these reasons, the Galaxy community provides technical guidance on how to bring *stand-alone* services to their ample portfolio of topic-specific tools. This is achieved through Planemo (https://planemo.readthedocs.io): a set of command-line utilities to help developers building and publishing tools in Galaxy.

4 From Stand-Alone RNA Tool to a Comprehensive Environment: The TriplexRNA

Among the RNA tools that investigate miRNA-mediated gene regulation is the TriplexRNA database. The aim of this database is consistent with the study and characterization of our understanding of miRNA–mRNA interactions. However, as with plenty of RNA-centric software tools available, the TriplexRNA is *stand-alone*, and not integrated with the growing community around the RNA workbench.

The TriplexRNA focuses on investigating a regulatory pattern involving a pair of miRNAs, cooperatively inhibiting the expression of a mutual target gene [38, 39]. The *triplex* model is not a computational artifact: Sætrom et al. experimentally validated this complex, further characterizing its structural constraints in 2007 [40]; additional experimental evidence confirmed this regulatory mechanism, whose knowledge has been harvested from the literature and computationally simulated in 2012 [38]. The results of this study were finally gathered in a dedicated database in 2014 [39], the TriplexRNA, accessible at https://triplexrna.org.

The database contains all cooperating miRNA pairs in human and mouse, as well as their target genes, graphical illustrations of triplex secondary structures, their Gibbs free energies, and predicted equilibrium concentrations. A link with known human diseases is provided by a dedicated interactive interface, to search for cooperative miRNA pairs linked to KEGG pathways [41]. These features allow the identification and testing of disease-specific miRNAs that could be used as diagnostic and prognostic markers.

In this section, we implement a TriplexRNA *wrapper* and show its usage with Planemo, to provide all functionalities of the original database within the Galaxy Web interface.

4.1 Queries for Understanding Cooperative miRNA Regulation

In its online stand-alone version, the TriplexRNA organizes all its functionalities behind a unique *entry point*: a form, consisting of a sentence which is adapted to reflect the desired interrogation (Fig. 2). Upon user submission, the results are returned within an interactive table, with per-column filters and selectors for easy consultation of the retrieved results (Fig. 3). Moreover, for promoting data reuse, and enabling researchers to integrate database

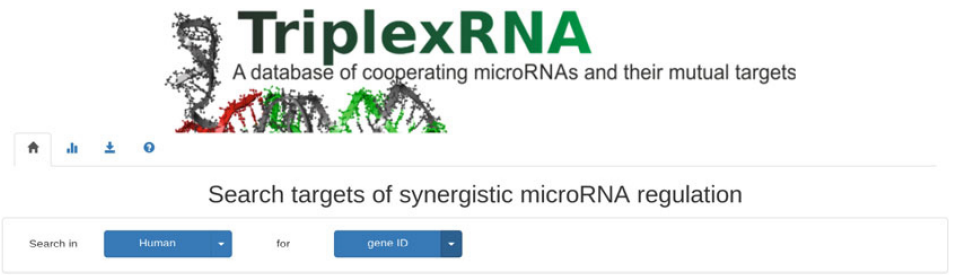


Fig. 2 The TriplexRNA form. The form constitutes a unique entry point, whereby queries are launched depending on the composed sentence. The default operation is to look for genes subjects of concerted miRNA regulation in humans. This query can be modified by selecting the desired organism, and gene or miRNA of interest

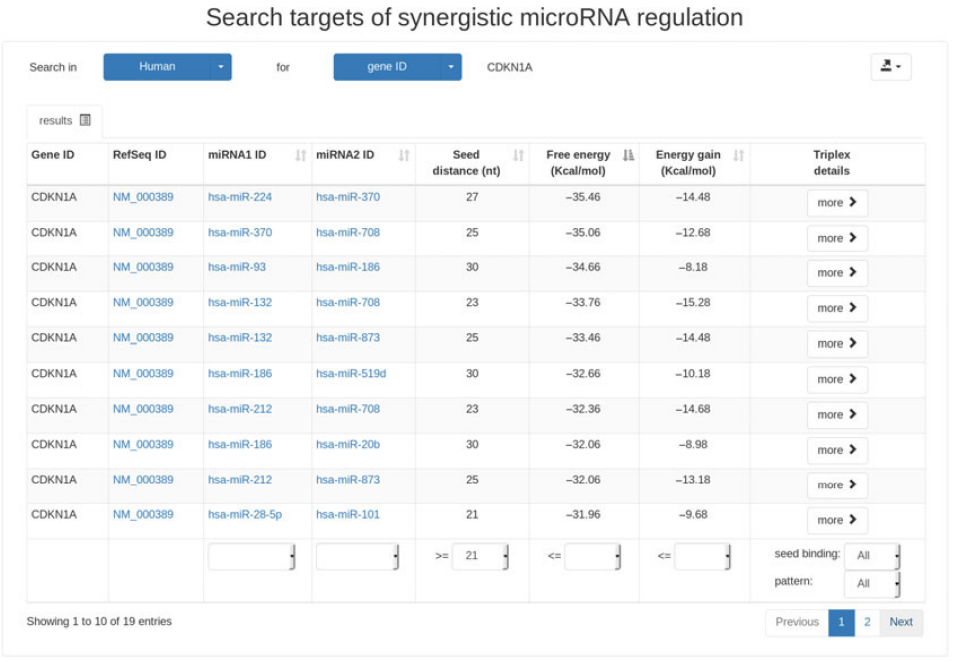


Fig. 3 The TriplexRNA result table. In this example, the table presents all putative triplexes involving the target gene CDKN1A. The table provides per-column filters and selectors to narrow down the entries to the results of interest. More information on the meaning of each column can be found in the help section at https:/triplexrna.org

queries within their computational pipelines, the TriplexRNA implements an API for retrieving results as HTML (https://www.w3.org/html), CSV (https://tools.ietf.org/html/rfc4180), and JSON (https://www.json.org). Each of these *standard*

Table 2
The TriplexRNA queries, implemented to investigate cooperative miRNA regulation. All functionalities are available through the standard Web interface, as well as API requests. API requests are launched from user pipelines, using an interrogation path which encodes the query parameters in a URI, for programmatic database retrieval. The table describes all TriplexRNA queries, and their API counterparts for retrieving data in JSON format

Query function	Interrogation path
Single gene query: retrieve all RNA triplexes of organism *O*, involving target gene *X*	triplexrna.org/JSON/*O*/gene/*X*
Multiple gene query: retrieve all RNA triplexes of organism *O*, involving the genes *X*, *Y*, *Z*	triplexrna.org/JSON/*O*/genes/*X*/*Y*/*Z*
Triplex query: retrieve all details of organism *O*'s RNA triplex *T*	triplexrna.org/JSON/*O*/triplex/*T*
Pathway query: retrieve all RNA triplexes of organism *O*, involved in KEGG pathway *K*	triplexrna.org/JSON/*O*/pathway/*K*
Single miRNA query: retrieve all RNA triplexes of organism *O*, involving miRNA *M*	triplexrna.org/JSON/*O*/mirna/*M*
miRNA pair query: retrieve all RNA triplexes of organism *O*, involving miRNAs *M* and *N*	triplexrna.org/JSON/*O*/mirna/*M*/*N*
Targets of cooperative miRNA pair query: retrieve all RNA triplexes of organism *O*, involving miRNA pair *M* and *N*, and targeting genes *X*, *Y*, *Z*	triplexrna.org/JSON/*O*/mirna/*M*/*N*/targeting/*X*/*Y*/

representations is returned depending on the selected *interrogation path*, i.e., a URI (https://www.w3.org/Addressing) which replaces the original user-operated Web form (entry point), for programmatic database access. Table 2 offers an overview of the TriplexRNA functionalities, their meaning, and the corresponding interrogation path that users shall adopt to automate data retrieval.

The integration of the TriplexRNA database functionalities within Galaxy leverages on the interrogation paths from the aforementioned table. These are implemented in a Python 2.7 (https://www.python.org) wrapper, available at https://github.com/bagnacan/triplexrna-planemo. The wrapper maps each database query to a specific command-line invocation, which is triggered once the corresponding arguments have been called.

Galaxy users do not interact directly with tool command lines; conversely, they invoke their functions by means of their dedicated Web interfaces. The tool wrapper helps developers in moving toward this solution: the next step toward the integration of the TriplexRNA within Galaxy leverages in fact on the wrapper's options and parameter semantics, to define the tool's very own Galaxy Web interface.

4.2 Integrating RNA Cooperativity Investigations in Galaxy

The integration of novel tools within Galaxy is carried out through Planemo.

In the first part of this section we explained the aim of its set of command-line utilities. Here, we overview their practical application, in relation to the TriplexRNA wrapper functions outlined in Table 2.

Planemo semi-automates the creation of a tool's Web page through the *planemo tool_init* function. This creates a draft XML (https://www.w3.org/XML) file, which includes the mandatory sections describing the tool in terms of its package dependencies, input/output formats, and mode of operations. For the creation of the TriplexRNA XML file, we used Planemo version 0.48.0, which generates the code shown in Fig. 4. All sections are then parsed by Galaxy, and rendered together as a dedicated Web interface that users can adopt to interact with the underlying tool.

The rendering is performed after a validation phase. This operation is carried out by *planemo lint*, which parses the developer-provided descriptions, to test for inconsistencies within the XML content. The linting phase has to be repeated at every successive editing of the XML file.

The TriplexRNA XML descriptor file is available at https://github.com/bagnacan/triplexrna-planemo.

`<tool id="t" name="T" version="0.1.0">`	
`<requirements>` `</requirements>`	this section contains the tool's package dependancies

`<command detect_errors="exit_code"><![CDATA[`
`TODO: Fill in command template.`
`]]></command>`
	this section contains the full command line options that are executed upon user query invocation

`<inputs>` `</inputs>`	this section defines how parameters are passed from the web interface to the underlying tool's command line
`<outputs>` `</outputs>`	this section defines the type and format of the results retrieved from the command's exectution

`<help><![CDATA[`
`TODO: Fill in help.`
`]]></help>`
	this section provides an explanation of all parameter's meaning

`</tool>`	

Fig. 4 The Planemo draft XML file. This includes the mandatory descriptions for a tool's integration in Galaxy. Each description is enclosed in a dedicated XML section, which provides a mean for the characterization of the underlying command-line tool, in terms of its environment requirements, input/output formats, and modes of operation. An overview of additional XML section is provided on the Planemo documentation, available at https:/planemo.readthedocs.io

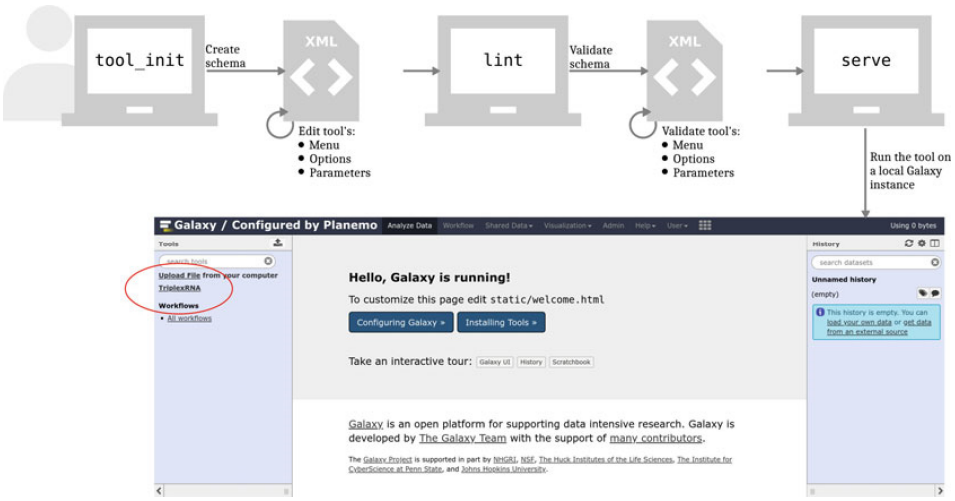


Fig. 5 The first steps toward integrating a novel tool within Galaxy using Planemo. (1) Planemo *tool_init* creates a scaffold XML file, filled by the user to accommodate each parameter option. (2) The provided XML is parsed and validated by Planemo *lint*. (3) Ultimately, Planemo *serve* includes the tool's wrapper and XML file for execution within a local Galaxy instance

Once all tool dependencies, input/output formats, and modes of operation are defined, the tool can be tested for running within Galaxy, by using the *planemo serve* command. This phase runs a local Galaxy instance, whose list of tools includes the one that underwent the previous edit and linting phases.

Figure 5 shows the aforementioned tool_init, lint, and serve steps we took to (1) create and edit the TriplexRNA XML descriptor for the underlying tool wrapper, (2) iteratively edit and validate the tools input/output options, and (3) launch the tool within a local Galaxy instance.

Planemo also provides utilities to test the new tool against predefined input datasets, assess the correct execution of each mode of operation, as well as publish the tool in the Galaxy Tool Shed (https://toolshed.g2.bx.psu.edu), for a later inclusion within dedicated Galaxy instances. These phases are not within the scope of the present work; however, they are mandatory for the development of new Galaxy tools. Further documentation on these topics can be found at https://planemo.readthedocs.io.

5 Discussion

The investigation and functional characterization of miRNA–mRNA interactions can be assessed computationally. A plethora of RNA software tools have been implemented for analyzing

their mechanisms; however, their solutions leverage on mathematical and statistical foundations: subjects that are rarely within the focus of life science curricula. As a result, before answering a biological question, users are required to look for available solutions by delving into online catalogs of tools, and make sense of their data formats, parametrization semantics, and software dependencies, ultimately facing a novel technical jargon.

This scenario has highlighted the need for creating a corpus of shared bioinformatics expertise, for the dissemination of computational approaches and best practices, aimed at addressing biological problems. The challenge has been dealt with by numerous Web platforms and training initiatives around the globe, which have been proved successful in establishing communities for discussing topic-specific problems, discussing best practices, and creating frameworks for the harmonization and reuse of shared processing tools and workflows [11]. In particular, for RNA analyses, the coherence of computational environments, data formats, interoperability, and reproducibility has been addressed by the Galaxy [30] community with the creation of the RNA workbench [8]. Such platform acts as a hub for developers as well as users: on the one hand to easily integrate and deploy novel or existing tools and workflows, and on the other to readily address biological questions, regardless of the background experience in administering computing environments. Following their community guidelines, we showcased Planemo and Galaxy, by realizing a wrapper for the TriplexRNA database [38, 39]. Finally, we provided a starting point for the inclusion of this stand-alone RNA tool as a new building block of the broad portfolio of tools within the RNA workbench.

The RNA workbench is growing, incorporating further tools for RNA analysis, and enriching the platform of diverse interoperable methods for understanding miRNA–mRNA interactions. However, this is coming at a cost, because as the list of tools increases, it is necessary to accommodate them in a structured catalog that must be both *clear* and *usable*: it should group the tools under pertinent and agreed-upon terms, to avoid forcing its users open each category before finding the desired tool, and abstract enough to accommodate novel entries, therefore not forcing its users to endlessly browse through an overwhelming list of categories. Failing to meet these trade-offs will systematically penalize the overall user experience, hindering tools below a long list of browsable fuzzy terms. Albeit on a smaller scale, the problem of organizing tools into catalogs [19, 20], as we highlighted in Section 2, reappears in the Galaxy framework.

In practice, both *Galaxy Main* (https://usegalaxy.org) and the *Galaxy EU* (https://usegalaxy.eu) public instances solve the problem of ambiguous terms by organizing their tools under pertinent and agreed-upon terms; however, due to the large number of tools,

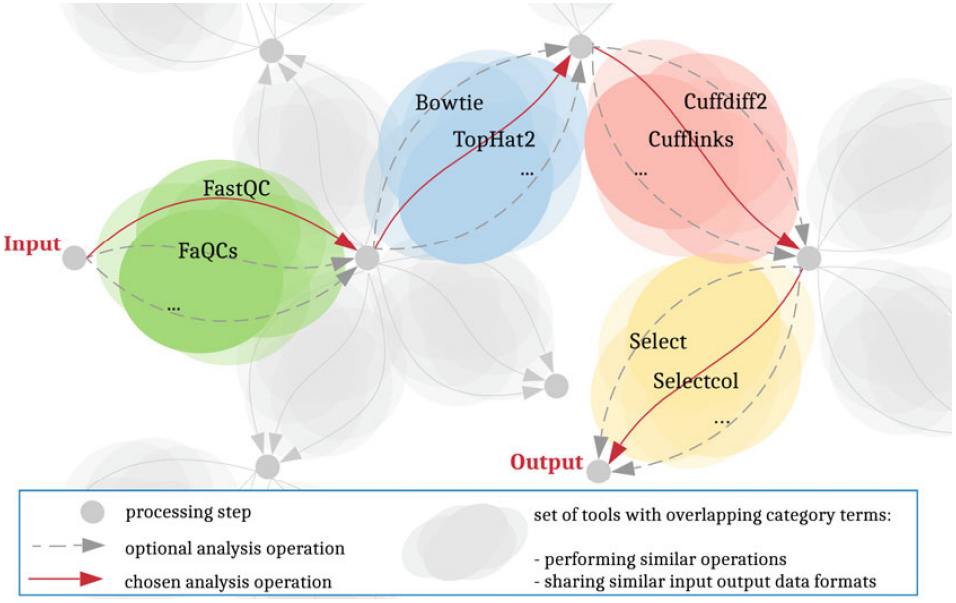


Fig. 6 Design of the Galaxy recommendation system. Workflow analyses are represented as paths over a graph comprising all known analysis pipelines. Here, each node represents a state of the processed data, and each edge a tool's function, whose allowed inputs and generated outputs correspond to the states of data it connects. In this framework, the next recommended tool must be pertinent with the scope of the desired analysis

novice users are left with no choice other than to browse most of the categories before finding the desired tool. This usability problem is partly overcome with the addition of a *search bar*; however, this feature requires the users to already know the name of the desired tool, which, especially for novice users, is not always the case. Moreover, such a solution does not advertise newly incorporated tools, therefore missing the promotion of possible alternatives and benchmarking.

To address the usability problem, we propose the integration of a *recommendation system*, aiming at providing guidance toward the completion of the desired analysis, while at the same time hiding all tools and categories behind a unique interface, where suggestions for next tools are loaded at each iteration. Figure 6 shows a candidate approach for the design of such a system. Here, we define a node as a state of the processed data, representing the starting or landing point of an edge: a tool's function, whose input and output (I/O) formats correspond to the specific nodes it connects. Within this framework, we define the next candidate tool as *pertinent* in terms of (a) allowed I/O data formats, accepted and generated by the tool, and (b) the original tool's categorization, attributed by

the instance's administrator to organize tools into a list of agreed-upon terms.

Such system leverages on a knowledge base built from anonymized Galaxy user-generated analyses, on top of which a predictive model *chains* tools into putative computational workflows on the basis of tool usage and occurrence. The model is further refined by the aforementioned definition of a tool's pertinence: here, allowed I/O data formats are retrieved using BioBlend (https://bioblend.readthedocs.io), while categories are obtained using a *web crawler* against Galaxy EU's public Web interface (https://usegalaxy.eu).

A proof of concept of both the predictive model and the tool's pertinence refinement are available at https://github.com/anuprulez/similar_galaxy_workflow.

We argue that frameworks such as the Galaxy Main, Galaxy EU, and topic-specific instances such as the Galaxy RNA workbench would benefit from a recommendation system, because it would improve the overall user experience, and therefore increase its acceptance and adoption across diverse communities, as a reference framework for shared best practices and reproducible analyses.

Acknowledgments

The authors would like to thank the de.NBI and ELIXIR initiatives, for their support in the bioinformatics infrastructure. Thanks also to the Galaxy community, for developing, maintaining, and providing guidance on the use of this comprehensive framework. A warm thank you goes to the RBC Freiburg group, in particular to Anup Kumar, Björn Grüning, and Rolf Backofen for their efforts and commitment in improving the Galaxy framework.

References

1. Lu J, Getz G, Miska EA et al (2005) MicroRNA expression profiles classify human cancers. Nature 435:834–838
2. Croce CM, Calin GA (2005) miRNAs, cancer, and stem cell division. Cell 122:6–7
3. Mitchell PS, Parkin RK, Kroh EM et al (2008) Circulating microRNAs as stable blood-based markers for cancer detection. PNAS 105:10513–10518
4. Chen X, Ba Y, Ma L et al (2008) Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res 18:997–1006
5. Cho WCS (2010) MicroRNAs: potential biomarkers for cancer diagnosis, prognosis and targets for therapy. Int J Biochem Cell Biol 42:1273–1281
6. Linsen SEV, de Wit E, Janssens G et al (2009) Limitations and possibilities of small RNA digital gene expression profiling. Nat Methods 6:474–476
7. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63
8. Grüning BA, Fallmann J, Yusuf D et al (2017) The RNA workbench: best practices for RNA and high-throughput sequencing bioinformatics in Galaxy. Nucleic Acids Res 45:W560–W566. https://doi.org/10.1093/nar/gkx409
9. Conesa A, Madrigal P, Tarazona S et al (2016) A survey of best practices for RNA-seq data analysis. Genome Biol 17:13
10. Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB (2011) The real cost of

sequencing: higher than you think! Genome Biol 12:125

11. Batut B, Hiltemann S, Bagnacani A, et al (2017) Community-driven data analysis training for biology. bioRxiv: 225680
12. Burks C (1999) Molecular biology database list. Nucleic Acids Res 27:1–9
13. Galperin MY, Rigden DJ, Fernández-Suárez XM (2015) The 2015 nucleic acids research database issue and molecular biology database collection. Nucleic Acids Res 43:D1–D5
14. Fox JA, Butland SL, McMillan S, Campbell G, Ouellette BFF (2005) The bioinformatics links directory: a compilation of molecular biology web servers. Nucleic Acids Res 33:W3–W24
15. Brazas MD, Yim D, Yeung W, Ouellette BFF (2012) A decade of web server updates at the bioinformatics links directory: 2003–2012. Nucleic Acids Res 40:W3–W12
16. Pettifer S, Thorne D, McDermott P, Attwood T, Baran J, Bryne JC, Hupponen T, Mowbray D, Vriend G (2009) An active registry for bioinformatics web services. Bioinformatics 25:2090–2091
17. Pettifer S, Ison J, Kalaš M et al (2010) The EMBRACE web service collection. Nucleic Acids Res 38:W683–W688
18. Bhagat J, Tanoh F, Nzuobontane E et al (2010) BioCatalogue: a universal catalogue of web services for the life sciences. Nucleic Acids Res 38:W689–W694
19. Henry VJ, Bandrowski AE, Pepin A-S, Gonzalez BJ, Desfeux A (2014) OMICtools: an informative directory for multi-omic data analysis. Database (Oxford). https://doi.org/10.1093/database/bau069
20. Ison J, Rapacki K, Ménager H et al (2016) Tools and data services registry: a community effort to document bioinformatics resources. Nucleic Acids Res 44:D38–D47
21. Ison J, Kalaš M, Jonassen I, Bolser D, Uludag M, McWilliam H, Malone J, Lopez R, Pettifer S, Rice P (2013) EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats. Bioinformatics 29:1325–1332
22. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:R36
23. Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357–360
24. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21
25. Hoffmann S, Otto C, Doose G et al (2014) A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. Genome Biol 15:R34
26. Engström PG, Steijger T, Sipos B et al (2013) Systematic evaluation of spliced alignment programs for RNA-seq data. Nat Methods 10:1185–1191
27. Möller S, Prescott SW, Wirzenius L et al (2017) Robust cross-platform workflows: how technical and scientific communities collaborate to develop, test and share best practices for data analysis. Data Sci Eng 2:232–244
28. Sandve GK, Nekrutenko A, Taylor J, Hovig E (2013) Ten simple rules for reproducible computational research. PLoS Comput Biol 9: e1003285
29. Goecks J, Nekrutenko A, Taylor J (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 11:R86
30. Afgan E, Baker D, van den Beek M et al (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res 44:W3–W10
31. Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL (2011) ViennaRNA Package 2.0. Algorithms Mol Biol 6:26
32. Will S, Joshi T, Hofacker IL, Stadler PF, Backofen R (2012) LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. RNA 18:900–914
33. Will S, Reiche K, Hofacker IL, Stadler PF, Backofen R (2007) Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. PLoS Comput Biol 3:e65
34. Zuker M, Sankoff D (1984) RNA secondary structures and their prediction. Bltn Mathcal Biol 46:591–621
35. Corcoran DL, Georgiev S, Mukherjee N, Gottwein E, Skalsky RL, Keene JD, Ohler U (2011) PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. Genome Biol 12:R79
36. Gruber AR, Findeiß S, Washietl S, Hofacker IL, Stadler PF (2010) Rnaz 2.0: improved noncoding RNA detection. Pac Symp Biocomput 15:69–79
37. Blin K, Dieterich C, Wurmus R, Rajewsky N, Landthaler M, Akalin A (2015) DoRiNA 2.0—upgrading the doRiNA database of RNA

interactions in post-transcriptional regulation. Nucleic Acids Res 43:D160–D167

38. Lai X, Schmitz U, Gupta SK, Bhattacharya A, Kunz M, Wolkenhauer O, Vera J (2012) Computational analysis of target hub gene repression regulated by multiple and cooperative miRNAs. Nucleic Acids Res 40:8818–8834
39. Schmitz U, Lai X, Winter F, Wolkenhauer O, Vera J, Gupta SK (2014) Cooperative gene regulation by microRNA pairs and their identification using a computational workflow. Nucleic Acids Res 42:7539–7552
40. Sætrom P, Heale BSE, Snøve O, Aagaard L, Alluin J, Rossi JJ (2007) Distance constraints between microRNA target sites dictate efficacy and cooperativity. Nucleic Acids Res 35:2333–2342
41. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30

2.1.6 Identification of rare cardiac cell types from single-nuclei RNA-Seq

Wolfien M., Galow A.M., Müller P., Bartsch M., Brunner R.M., Goldammer T., Wolkenhauer O., Hoeflich A., David R. (2020).
Single Nuclei Sequencing of an entire Mammalian Heart: Cell Type Composition and Velocity.
Cells. IF: 5.656, Citations (December 14, 2020): 1

In-depth analyses on the cellular level are indispensable to expand our understanding of complex tissues, such as the mammalian heart, because often a small amount of cells can determine the cell fate of a whole tissue. Single-nucleus sequencing (snRNA-Seq) allows for the exploration of cellular compositions and, with respect towards the cardiac tissue, it overcomes major hurdles of single-cell RNA-Seq in terms of size restriction and cell viability. In this study, we used snRNA-Seq to investigate for the first time an entire, adult mammalian heart to characterize its overall cell composition and RNA kinetics.

I developed the computational workflow for the analysis of the snRNA-Seq data with respect to the basic data processing (via Kallisto and Bustools), rare cell type identification (Harmony), and RNA-velocity analysis (velocyto.R). In particular, a customized, genomic index for the murine genome mm10 (Ensembl build 98) was build and uploaded on Zenodo (Downloads: 816, accessed at December 14, 2020)[12] to be able to utilize the RNA-velocity analysis. After further genomic alignment and quantification of the *.fastq* files, I integrated the data into *R* and performed further downstream analyses with *Seurat*. The nuclei were clustered, annotated, and ultimately the RNA-velocity approach was applied to investigate the RNA kinetics between the spliced and unspliced RNA transcripts. I supported the generation of Fig.1 and Fig.2, as well as the development of the supplementary online material at the publicly available iRhythmics FairdomHub instance.[13]

In summary, we present the first report of sequencing an entire, adult mammalian heart, providing realistic cell-type distributions combined with RNA-velocity kinetics to characterize cellular interrelations. Interestingly, mature CM appeared to originate not only from a single lineage but also from an additional endothelial direction. We found a cell population (CM-like endothelial cells) that comprises endothelial markers, and markers clearly related to CM function.

[12] https://zenodo.org/record/3623148
[13] https://doi.org/10.15490/fairdomhub.1.study.713.1

cells

MDPI

Communication

Single-Nucleus Sequencing of an Entire Mammalian Heart: Cell Type Composition and Velocity

Markus Wolfien [1,†], Anne-Marie Galow [2,†], Paula Müller [3,4,†], Madeleine Bartsch [3,4], Ronald M. Brunner [2], Tom Goldammer [2,5], Olaf Wolkenhauer [1,6], Andreas Hoeflich [2,*] and Robert David [3,4,*]

1 Department of Systems Biology and Bioinformatics, University of Rostock, 18051 Rostock, Germany; markus.wolfien@uni-rostock.de (M.W.); olaf.wolkenhauer@uni-rostock.de (O.W.)
2 Institute of Genome Biology, Leibniz Institute for Farm Animal Biology (FBN), 18196 Dummerstorf, Germany; galow@fbn-dummerstorf.de (A.-M.G.); brunner@fbn-dummerstorf.de (R.M.B.); tom.goldammer@uni-rostock.de (T.G.)
3 Reference and Translation Center for Cardiac Stem Cell therapy (RTC), Department of Cardiac Surgery, Rostock University Medical Center, 18057 Rostock, Germany; paula.mueller@uni-rostock.de (P.M.); madeleine.bartsch@med.uni-rostock.de (M.B.)
4 Department of Life, Light, and Matter of the Interdisciplinary Faculty at Rostock University, 18059 Rostock, Germany
5 Molecular Biology and Fish Genetics, Faculty of Agriculture and Environmental Sciences, University of Rostock, 18059 Rostock, Germany
6 Stellenbosch Institute of Advanced Study, Wallenberg Research Centre, Stellenbosch University, 7602 Stellenbosch, South Africa
* Correspondence: hoeflich@fbn-dummerstorf.de (A.H.); robert.david@med.uni-rostock.de (R.D.)
† These authors contributed equally.

Received: 23 December 2019; Accepted: 25 January 2020; Published: 28 January 2020

Abstract: Analyses on the cellular level are indispensable to expand our understanding of complex tissues like the mammalian heart. Single-nucleus sequencing (snRNA-seq) allows for the exploration of cellular composition and cell features without major hurdles of single-cell sequencing. We used snRNA-seq to investigate for the first time an entire adult mammalian heart. Single-nucleus quantification and clustering led to an accurate representation of cell types, revealing 24 distinct clusters with endothelial cells (28.8%), fibroblasts (25.3%), and cardiomyocytes (22.8%) constituting the major cell populations. An additional RNA velocity analysis allowed us to study transcription kinetics and was utilized to visualize the transitions between mature and nascent cellular states of the cell types. We identified subgroups of cardiomyocytes with distinct marker profiles. For example, the expression of Hand2os1 distinguished immature cardiomyocytes from differentiated cardiomyocyte populations. Moreover, we found a cell population that comprises endothelial markers as well as markers clearly related to cardiomyocyte function. Our velocity data support the idea that this population is in a trans-differentiation process from an endothelial cell-like phenotype towards a cardiomyocyte-like phenotype. In summary, we present the first report of sequencing an entire adult mammalian heart, providing realistic cell-type distributions combined with RNA velocity kinetics hinting at interrelations.

Keywords: snRNA-seq; RNA velocity; cluster analysis; cardiomyocytes; seurat

1. Introduction

Single-cell sequencing allows for an in-depth characterization of complex tissues and their cell types [1]. However, there are two major issues when it comes to the cardiovascular system, namely,

(i) the difficulty of dissociating the adult mammalian heart tissue without damaging constituent cells and (ii) technical limitations regarding cell capture techniques leading to an underrepresentation of individual cell types (i.e., cardiomyocytes) due to their large cell size and irregular shape [2]. Whereas research efforts aim to avoid these issues by relying on embryonic and neonatal murine hearts or focusing on non-myocyte populations in adult mouse hearts, we desisted from single-cell Ribonucleic acid sequencing (RNA-seq) and instead conducted single-nucleus RNA-seq (snRNA-seq), which has been shown to present similar transcriptomic results [3]. Currently, existing studies on adult mammalian hearts concentrate only on selected substructures such as the ventricle [4] or the conduction system [5]. To our knowledge, we present the first snRNA-seq analysis of an entire adult mammalian heart.

Recently, a method was established to predict even future states of individual cells using single-cell or single-nucleus data. The relative abundance of nascent (unspliced) and mature (spliced) mRNA in these datasets is exploited to predict the rates of gene splicing and degradation. The time derivative of the gene expression state is calculated on the basis of these gene splicing events and is referred to as RNA velocity [6]. The RNA velocity analysis of our snRNA-seq data allowed us to study transcription kinetics and revealed details about the dynamics and interconnectedness of our identified cell clusters.

2. Materials and Methods

2.1. Isolation of Nuclei

To avoid potential aberrations due to inbreeding, we relied on an outbred mice strain (Fzt:DU) [7]. Mice were handled in accordance with Directive 2010/63/EU on the protection of animals and with the Scientific Committee supervising animal experiments in the Leibniz-Institute for Farm Animal Biology (FBN), Dummerstorf, Germany. Whole hearts were harvested from 4 male mice (12 weeks) after cervical dislocation. The hearts were pooled and nuclei isolated using the Nuclei PURE Prep isolation kit (Sigma-Aldrich, Darmstadt, Germany) according to the manufacturer's protocol. All work was carried out on ice. In brief, hearts were rinsed with ice cold PBS, minced thoroughly, and preincubated in 10 mL freshly prepared lysis buffer for 10–15 min before the tissue was further homogenized using a gentleMACS dissociator (Miltenyi Biotec, Bergisch Gladbach, Germany). Cell debris and clumps were removed by using 40 µm strainers. To purify the nuclei, lysate samples were mixed with 18 mL chilled sucrose cushion solution, layered on 10 mL pure 1.8 M sucrose cushion solution in a 50 mL Beckman ultracentrifuge tube, and centrifuged for 45 min at 30,000× *g* and 4 °C. Nuclei pellets were resuspended in 5 mL chilled PBS containing 1% BSA and 0.2 U/µL RNase inhibitor and cell debris was removed by a final filtration step. After centrifugation for 8 min at 600× *g* and 4 °C, the supernatant was carefully removed and nuclei were resuspended in 3 mL Nuclei PURE storage buffer. The samples were transferred to cryotubes, snap-frozen in liquid nitrogen, and stored at −80 °C until processing.

Sequencing was conducted by Genewiz (Leipzig, Germany) on the 10xGenomics system (Carlsbad, CA, USA). Single nuclei were captured in droplet emulsions and snRNA-seq libraries were constructed as per the 10x Genomics protocol using GemCode Single-Cell 3' Gel Bead and Library V3 Kit (Carlsbad, CA, USA). RNA was controlled for sufficient quality on an Agilent 2100 Bioanalyzer system (Santa Clara, CA, USA) and quantified using a Qubit Fluorometer (Waltham, MA, USA). Libraries were subsequently sequenced on the NovaSeq 6000 Sequencing System (Illumina, San Diego, CA. USA).

2.2. Computational Data Analysis

The snRNA-seq fastq data files were aligned with kallisto (v.0.46) to the generated mm10 genome (Ensembl release 98) index. The UNIX source code containing the detailed steps of the generation is provided at our FairdomHub/iRhythmics instance (https://doi.org/10.15490/fairdomhub.1.study.713.1). Additionally, the latest version of the complete index build was shared at Zenodo for further reuse (https://doi.org/10.5281/zenodo.3623148). This index contains the spliced and unspliced transcript annotations of the mm10 murine needed for RNA velocity analysis. The kallisto alignment files were

Cells 2020, 9, x FOR PEER REVIEW 3 of 6

index contains the spliced and unspliced transcript annotations of the mm10 murine needed for RNA velocity analysis. The kallisto alignment files were subsequently quantified with bustools (v.0.39.3) as previously described [8]. Subsequently, transcripts were integrated into R by using the BUSpaRse R-package (v.0.99.25) to be able to use the downstream processing tool Seurat (v.3.1.1). For clustering, dimensionality was initially reduced by principal component analysis and numbers of the most variable principal components were selected using heuristic methods implemented in Seurat. For an improved UMAP clustering and identification of small subgroups, we included the upstream processing algorithm harmony (v.1.0) [9]. The RNA velocity was processed with the velocyto.R (v.0.6) [8] package. Sets of well-known marker genes were used to assign cell types of the generated clusters, as summarized in our computational R script. In addition, novel cell cluster markers recently identified by other groups working with single-nucleus data[4] were applied and found to be transferable to our dataset.

The detailed experimental protocol, computational scripts, top 100 transcripts per cluster as well as the expression of the top markers for our identified clusters can be accessed from FairdomHub/iRhythmics. Raw data is provided in the Single Cell Expression Atlas via Arrayexpress (Accession ID: E-MTAB-8751).

3. Results and Discussion

Single-nucleus analysis included a total of 8635 nuclei and 22,568 genes in which each cell exhibits an average total expression of 2662.6 reads. The analysis revealed 24 distinct clusters as a UMAP representation showing a global connectivity among the groups (Figure 1). The largest clusters can be attributed to populations of endothelial cells (28.8%), fibroblasts (25.3%), and cardiomyocytes (22.8%) containing ~2500, ~2200, and ~2000 nuclei, respectively.

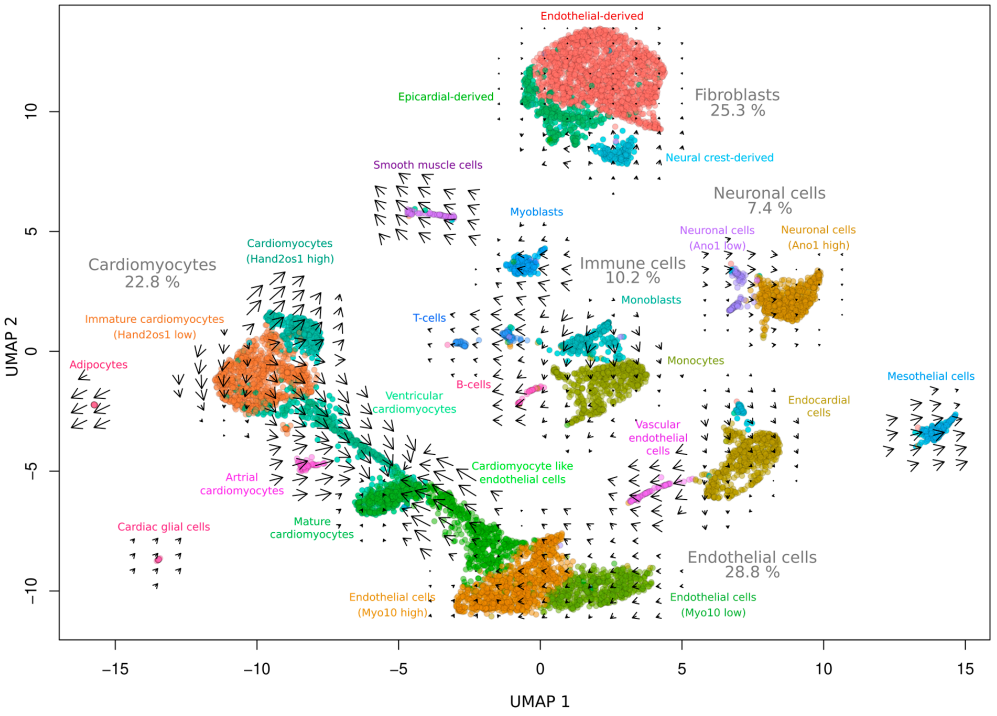


Figure 1. Single-nucleus transcriptome characteristics of pooled whole Fzt:DU mice hearts ($n = 4$). UMAP clustering of snRNA-seq data (8635 nuclei) reveals 24 distinct clusters for the indicated cell types. The arrows represent RNA velocity kinetics visualizing the direction and acceleration between mature and nascent mRNA. The percentages represent the nuclei ratio.

Figure 1. Single-nucleus transcriptome characteristics of pooled whole Fzt:DU mice hearts ($n = 4$). UMAP clustering of snRNA-seq data (8635 nuclei) reveals 24 distinct clusters for the indicated cell types. The arrows represent RNA velocity kinetics visualizing the direction and acceleration between mature and nascent mRNA. The percentages represent the nuclei ratio.

Interestingly, our data contradict earlier studies based on flow cytometry that suggest a much higher proportion of endothelial cells of up to 55% [10]. This disparity might be due to different isolation protocols on the one hand, and the fact that we used single-nucleus instead of single-cell isolation protocols, on the other hand. However, more recent findings, that based on single-nucleus sequencing [5], are in accordance with our data, so that we further assume that this kind of holistic approach may yield more robust results than approaches relying on single marker genes. Moreover, we not only observed various immune cells but also identified cells of neuronal origin (7.4%) and cardiac glial cells (0.2%) representing the innervated system of the heart and confirming the comprehensiveness of our data. The wealth of data enabled the identification of further cell-type markers that, in addition to the standard markers, facilitated the annotation of clusters, thereby providing novel reference points for us and the research community (Figure 2).

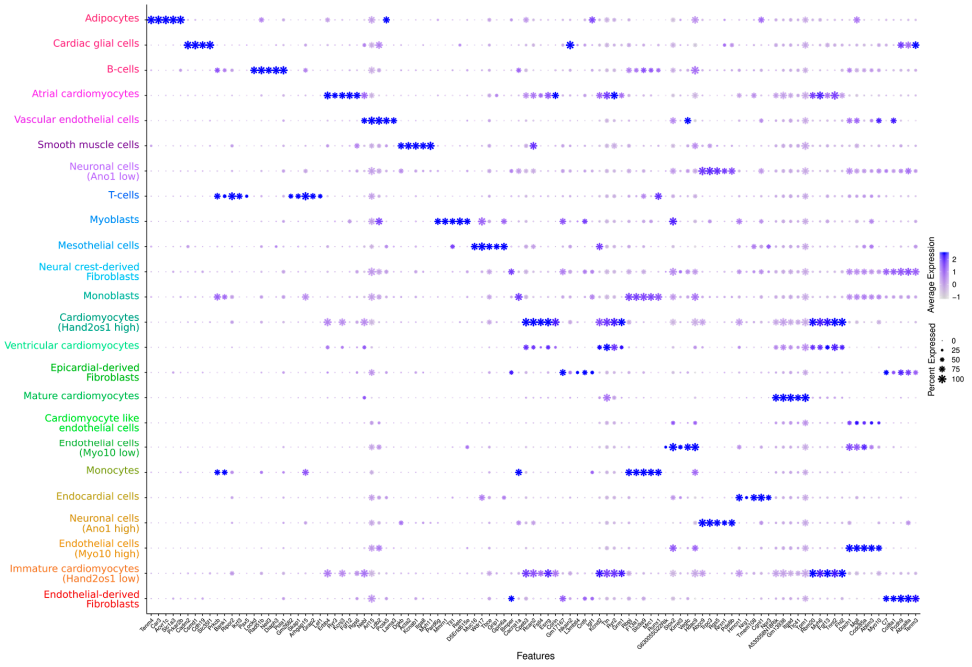


Figure 2. Dot-plot representation of the gene expression marker genes for the identified cell types. The size of dots represents the relative gene expression in percent for each cluster, e.g., a value of 100 means that each cell within this cell type expressed this gene. The color indicates the average expression level for the indicated gene per cell type. The color of the clusters is taken from Figure 1. A dot plot for the most significant gene per cluster as well as an extended visualization of the top 10 markers per cluster can be obtained at our FairdomHub/iRhythmics instance.

Our additional RNA velocity analysis of the snRNA-seq allowed us to study transcription kinetics (Figure 1). The indicated arrows show the direction and the velocity for future cell states. For example, immune cells undergo intense transformation processes upon maturation and activation and, therefore, show a high velocity (lengthy arrows) in our Fzt:DU mice. A quick turnover of RNA was also shown for smooth muscle cells, which have to adapt frequently to changing demands on the vascular pressure, confirming the physiological relevance of our data.

We furthermore identified subgroups of cardiomyocytes with distinct marker profiles and could visualize their developmental course by RNA velocity analysis (Figure 1). In particular, we found the expression of Hand2os1 [11–13]—a long non-coding RNA—that orchestrates heart development by dampening HAND2 expression, to distinguish immature cardiomyocytes from fully differentiated cardiomyocyte populations (Figure 2). Besides the mature atrial and

cardiomyocyte populations (Figure 2). Besides the mature, atrial, and ventricular cardiomyocytes, there is another Hand2os1 high cardiomyocyte population with a 1.5-fold expression enrichment apparently originating from the Hand2os1 low population. Based on very recent findings of de Soysa et al. [14], who identified *Hand2* as a specifier of outflow tract cells but not right ventricular cells during embryonal development, we assume that this population represents cells of the outflow tract.

Interestingly, mature cardiomyocytes appeared to originate not only from a single lineage but also from an additional endothelial direction (Figure 1). We found a cell population (cardiomyocyte-like endothelial cells) that comprises endothelial markers (e.g., Flt1, Dach1) as well as markers clearly related to cardiomyocyte function (e.g., Ryr2, Tpm1, Ttn, Gja1, and Myh6). The dual role of this population can also be recognized in the dot plot (Figure 2). Although the population lacked other typical cardiomyocyte markers (e.g., Tnnt2), together with the velocity data our results suggest a trans-differentiation process from an endothelial cell-like phenotype towards a cardiomyocyte-like phenotype, supporting previous findings [15].

As our data apparently include the findings of other studies, we are confident that our whole heart single-nucleus analysis of the outbred Fzt:DU mouse strain at present provides the most accurate representation of cell types in an adult mammalian heart and can be used as a reference for further comparative studies.

Author Contributions: Conceptualization, M.W., A.-M.G., R.M.B., and R.D.; methodology, M.W., A.-M.G., P.M., M.B., R.M.B., T.G., A.H., and R.D.; formal analysis, M.W. and A.-M.G.; investigation, M.W., A.-M.G., P.M., M.B., R.M.B., and R.D.; resources, T.G., O.W., A.H. and R.D.; data curation, M.W., A.-M.G., A.H., and O.W.; writing—original draft preparation, M.W., A.-M.G., T.G., A.H., and R.D.; writing—review and editing, M.W., A.-M.G., and R.D.; visualization, M.W., A.-M.G., O.W., T.G., A.H., and R.D.; supervision, R.D.; project administration, A.H. and R.D.; funding acquisition, O.W., A.H., and R.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the EU Structural Fund (ESF/14-BM-A55-0024/18), the "Deutsche Forschungsgemeinschaft" - DFG (DA1296/6-1), the German Heart Foundation (F/01/12), the FORUN Program of Rostock Medical University (889001/889003), the Josef and Käthe Klinz Foundation (T319/29737/2017), the Damp Foundation (2016-11), and the Federal Ministry of Education and Research - BMBF (VIP+00240, 031L0106C).

Conflicts of Interest: The authors declare no conflict of interest. The funders were not involved in study design, data collection and interpretation, and manuscript preparation.

References

1. Schaum, N.; Karkanias, J.; Neff, N.F.; May, A.P.; Quake, S.R.; Wyss-Coray, T.; Darmanis, S.; Batson, J.; Botvinnik, O.; Chen, M.B.; et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nat. Nat. Publ. Group* **2018**, *562*, 367–372.
2. Ackers-Johnson, M.; Tan, W.L.W.; Foo, R.S.-Y. Following hearts, one cell at a time: Recent applications of single-cell RNA sequencing to the understanding of heart disease. *Nat. Commun.* **2018**, *9*, 4434. [CrossRef]
3. Bakken, T.E.; Hodge, R.D.; Miller, J.A.; Yao, Z.; Nguyen, T.N.; Aevermann, B.; Barkan, E.; Bertagnolli, D.; Casper, T.; Dee, N.; et al. Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS ONE* **2018**, *13*, e0209648. [CrossRef] [PubMed]
4. Linscheid, N.; Logantha, S.J.R.J.; Poulsen, P.C.; Zhang, S.; Schrölkamp, M.; Egerod, K.L.; Thompson, J.J.; Kitmitto, A.; Galli, G.; Humphries, M.J.; et al. Quantitative proteomics and single-nucleus transcriptomics of the sinus node elucidates the foundation of cardiac pacemaking. *Nat. Commun.* **2019**, *10*, 2889. [CrossRef]
5. Hu, P.; Liu, J.; Zhao, J.; Wilkins, B.J.; Lupino, K.; Wu, H.; Pei, L. Single-nucleus transcriptomic survey of cell diversity and functional maturation in postnatal mammalian hearts. *Genes Dev.* **2018**, *32*, 1344–1357. [CrossRef] [PubMed]
6. La Manno, G.; Soldatov, R.; Zeisel, A.; Braun, E.; Hochgerner, H.; Petukhov, V.; Lidschreiber, K.; Kastriti, M.E.; Lönnerberg, P.; Furlan, A.; et al. RNA velocity of single cells. *Nature* **2018**, *560*, 494–498. [CrossRef] [PubMed]
7. Dietl, G.; Langhammer, M.; Renne, U. Model simulations for genetic random drift in the outbred strain Fzt:DU. *Arch. Anim. Breed.* **2004**, *47*, 595–604. [CrossRef]

8. Melsted, P.; Booeshaghi, A.S.; Gao, F.; Beltrame, E.D.V.; Lu, L.; Hjorleifsson, K.E.; Gehring, J.; Pachter, L. *Modular and Efficient Pre-Processing of Single-Cell RNA-Seq*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, USA, 2019; p. 673285.
9. Korsunsky, I.; Millard, N.; Fan, J.; Slowikowski, K.; Zhang, F.; Wei, K.; Baglaenko, Y.; Brenner, M.; Loh, P.-R.; Raychaudhuri, S. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **2019**, *16*, 1289–1296. [CrossRef] [PubMed]
10. Pinto, A.R.; Ilinykh, A.; Ivey, M.J.; Kuwabara, J.T.; D'antoni, M.L.; Debuque, R.; Chandran, A.; Wang, L.; Arora, K.; Rosenthal, N.A.; et al. Revisiting cardiac cellular composition. *Circ. Res. Lippincott Williams Wilkins* **2016**, *118*, 400–409. [CrossRef] [PubMed]
11. Han, X.; Zhang, J.; Liu, Y.; Fan, X.; Ai, S.; Luo, Y.; Li, X.; Jin, H.; Luo, S.; Zheng, H.; et al. The lncRNA Hand2os1/Uph locus orchestrates heart development through regulation of precise expression of Hand2. *Development* **2019**, *146*, dev176198. [CrossRef] [PubMed]
12. Ritter, N.; Ali, T.; Kopitchinski, N.; Schuster, P.; Beisaw, A.; Hendrix, D.A.; Schulz, M.H.; Müller-McNicoll, M.; Dimmeler, S.; Grote, P. The lncRNA Locus Handsdown Regulates Cardiac Gene Programs and Is Essential for Early Mouse Development. *Dev. Cell* **2019**, *50*, 644–657.e8. [CrossRef] [PubMed]
13. Anderson, K.M.; Anderson, U.M.; McAnally, J.R.; Shelton, J.M.; Bassel-Duby, R.; Olson, E.N. Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. *Nature* **2016**, *539*, 433–436. [CrossRef] [PubMed]
14. De Soysa, T.Y.; Ranade, S.S.; Okawa, S.; Ravichandran, S.; Huang, Y.; Salunga, H.T.; Schricker, A.; Del Sol, A.; Gifford, C.A.; Srivastava, D. Single-cell analysis of cardiogenesis reveals basis for organ-level developmental defects. *Nature* **2019**, *572*, 120–124. [CrossRef] [PubMed]
15. Condorelli, G.; Borello, U.; De Angelis, L.; Latronico, M.; Sirabella, D.; Coletta, M.; Galli, R.; Balconi, G.; Follenzi, A.; Frati, G.; et al. Cardiomyocytes induce endothelial cells to trans-differentiate into cardiac muscle: Implications for myocardium regeneration. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10733–10738. [CrossRef] [PubMed]

2.1.7 Community-driven data analysis training with Galaxy

Batut, B., Hiltemann, S., Bagnacani, A., ..., **Wolfien, M.**, ..., and Gruening, B.A. (2018).
Community-driven Data Analysis Training for Biology.
Cell Systems. IF: 8.640, Citations (December 14, 2020): 7

One of the major challenges resulting from the numerous biomedical datasets is neither the amount of data itself, nor the computational resources, or the required storage space, but the general lack of trained and skilled researchers to analyze these large amounts of data. Eliminating this problem requires the development of easy accessible and comprehensive educational resources. Here, a community-driven framework is presented that enables modern, interactive teaching of data analytics in life sciences and facilitates the development of training materials.

To achieve this goal of a comprehensive collection of well-suited and informative trainings, example workflows and datasets, including a wide range of expertise, are needed. I contributed to the trainings about RNA-Seq data analyses, as well as "*Quality control*" and "*Mapping*". The training material is accessible at Galaxy training[14] and open for contributions of any kind. The key feature of this material represents a continuously improved collection of self-driven tutorials. In addition, these tutorials are tought person to person at international conferences and workshops. A list of the eight RNA-Seq data analysis trainings that I jointly gave with the de.STAIR project members in Freiburg, Jena, and Rostock can be assed at the de.STAIR training page[15] or in the Appendix.

In summary, we integrated numerous data analysis tutorials into a unified web-based analysis framework, in which biomedical researchers learn to utilize complex computations themselves through an interactive interface without the need to install software or search for example datasets. The ultimate goal is to expand the breadth of training materials to include fundamental statistical and data-science topics to be able to precipitate a complete re-engineering of undergraduate and graduate curricula in life sciences.

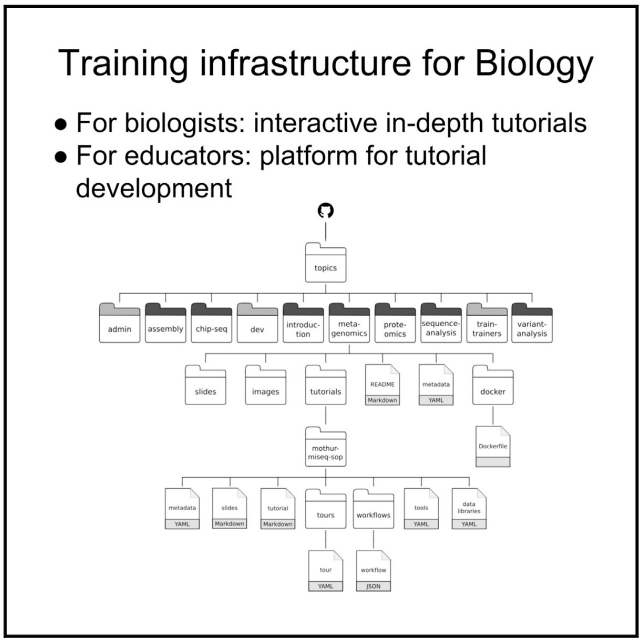
[14] https://training.galaxyproject.org
[15] https://github.com/destairdenbi/trainings

Report

Cell Systems

Community-Driven Data Analysis Training for Biology

Graphical Abstract



Highlights

- Web-based, community-maintained infrastructure for data analysis training
- Incorporates latest tools and types of biomedical data
- Supports trainers as much as it supports trainees

Authors

Bérénice Batut, Saskia Hiltemann, Andrea Bagnacani, ..., Rolf Backofen, Anton Nekrutenko, Björn Grüning

Correspondence

backofen@informatik.uni-freiburg.de (R.B.), anton@nekrut.org (A.N.), gruening@informatik.uni-freiburg.de (B.G.)

In Brief

We developed an infrastructure that facilitates data analysis training in life sciences. It is an interactive learning platform tuned for current types of data and research problems. Importantly, it provides a means for community-wide content creation and maintenance and, finally, enables trainers and trainees to use the tutorials in a variety of situations, such as those where reliable Internet access is unavailable.

Batut et al., 2018, Cell Systems *6*, 752–758
June 27, 2018
https://doi.org/10.1016/j.cels.2018.05.012

CellPress

CellPress

Cell Systems
Report

Community-Driven Data Analysis Training for Biology

Bérénice Batut,[1,19] Saskia Hiltemann,[2,19] Andrea Bagnacani,[3] Dannon Baker,[4] Vivek Bhardwaj,[5] Clemens Blank,[1] Anthony Bretaudeau,[6] Loraine Brillet-Guéguen,[7] Martin Čech,[8] John Chilton,[8] Dave Clements,[4] Olivia Doppelt-Azeroual,[9] Anika Erxleben,[1] Mallory Ann Freeberg,[10] Simon Gladman,[11] Youri Hoogstrate,[2] Hans-Rudolf Hotz,[12] Torsten Houwaart,[1] Pratik Jagtap,[13] Delphine Larivière,[8] Gildas Le Corguillé,[14] Thomas Manke,[15] Fabien Mareuil,[9] Fidel Ramírez,[15] Devon Ryan,[15] Florian Christoph Sigloch,[1] Nicola Soranzo,[16] Joachim Wolff,[1] Pavankumar Videm,[1] Markus Wolfien,[3] Aisanjiang Wubuli,[17] Dilmurat Yusuf,[1] Galaxy Training Network,[18] James Taylor,[4] Rolf Backofen,[1,*] Anton Nekrutenko,[8,20,*] and Björn Grüning[1,*]

[1]Bioinformatics Group, Department of Computer Science, Albert-Ludwigs-University Freiburg, Georges-Köhler-Allee 106, Freiburg 79110, Germany
[2]Erasmus Medical Centre, Wytemaweg 80, Rotterdam 3015 CN, the Netherlands
[3]Department of Systems Biology and Bioinformatics, University of Rostock, Ulmenstraße 69, Rostock 18051, Germany
[4]Johns Hopkins University, 3400 N Charles Street, Mudd Hall 144, Baltimore 21218, MD, USA
[5]Department of Biology, Albert-Ludwigs-University, Schänzlestraße 1, Freiburg 79104, Germany
[6]INRA, UMR IGEPP, BIPAA/GenOuest, INRIA/Irisa - Campus de Beaulieu, 35042 RENNES Cedex, France
[7]CNRS, UMPC, FR2424, ABiMS, Station Biologique, Roscoff, France
[8]The Pennsylvania State University, 505 Wartik Lab, University Park, PA 16802, USA
[9]Bioinformatics and Biostatistics HUB, Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI, USR 3756 Institut Pasteur et CNRS), Institut Pasteur, 25-28 Rue du Docteur Roux, 75015 Paris, France
[10]European Bioinformatics Institute, Hinxton, Cambridge, UK
[11]Melbourne Bioinformatics, The University of Melbourne, Melbourne, VIC 3010, Australia
[12]Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, Basel 4058, Switzerland
[13]Biochemistry, Molecular Biology and Biophysics, University of Minnesota Medical School, 420 Delaware Street SE, Minneapolis, MN 55455, USA
[14]PMC, CNRS, FR2424, ABiMS, Station Biologique, Place Georges Teissier, Roscoff 29680, France
[15]Max Planck Institute of Immunobiology and Epigenetics, Stübeweg 51, Freiburg 79108, Germany
[16]Earlham Institute, Norwich Research Park, Norwich NR4 7UZ, UK
[17]Leibniz Institute for Farm Animal Biology (FBN), Wilhelm-Stahl-Allee 2, Dummerstorf 18196, Germany
[18]https://galaxyproject.org/teach/gtn/
[19]These authors contributed equally
[20]Lead Contact
*Correspondence: backofen@informatik.uni-freiburg.de (R.B.), anton@nekrut.org (A.N.), gruening@informatik.uni-freiburg.de (B.G.)
https://doi.org/10.1016/j.cels.2018.05.012

SUMMARY

The primary problem with the explosion of biomedical datasets is not the data, not computational resources, and not the required storage space, but the general lack of trained and skilled researchers to manipulate and analyze these data. Eliminating this problem requires development of comprehensive educational resources. Here we present a community-driven framework that enables modern, interactive teaching of data analytics in life sciences and facilitates the development of training materials. The key feature of our system is that it is not a static but a continuously improved collection of tutorials. By coupling tutorials with a web-based analysis framework, biomedical researchers can learn by performing computation themselves through a web browser without the need to install software or search for example datasets. Our ultimate goal is to expand the breadth of training materials to include fundamental statistical and data science topics and to precipitate a complete re-engineering of undergraduate and graduate curricula in life sciences. This project is accessible at https://training.galaxyproject.org.

INTRODUCTION

Rapid development of DNA-sequencing technologies has made it possible for biomedical disciplines to rival the physical sciences in data production capability. The combined output of today's genomics studies has already surpassed the data acquisition rate of entire scientific domains such as astronomy or Internet platforms such as YouTube or Twitter (Stephens et al., 2015). Yet biology is different from astronomy (and other quantitative disciplines) in one fundamental aspect: the lack of computational and data analysis training in standard biomedical curricula. Many biomedical scientists do not possess the skills to use or even access existing analysis resources. Such paucity of training also negatively affects the ability of biological investigators to collaborate with their statistics and

mathematics counterparts because of the inability to speak each other's language. In addition, an estimated one-third of biomedical researchers do not have access to proper data analysis support (Larcombe et al., 2017). The only way to address these deficiencies is with training. The need for such training cannot be overstated: while the majority (>95%) of researchers work or plan to work with large datasets, most (>65%) possess only minimal bioinformatics skills and are not comfortable with statistical analyses (Larcombe et al., 2017) (Williams and Teal, 2017) (Barone et al., 2017). This overwhelming need drives the demand, which, at present, greatly exceeds supply (Attwood et al., 2017). In a recent survey (Community Survey Report, 2013), over 60% of biologists expressed a need for more training, while only 5% called for more computing power. Thus one can assume that the true bottleneck of the current data deluge is not storage or processing power but the knowledge and skills to utilize already existing resources and tools. It is necessary to point out that there are great existing sources of training, such as online teaching materials provided by Johns Hopkins, University of Utah, Rosalind, and others. These are valuable entry points to the field of biomedical data analysis. The type of learning we are describing in this report is complementary to these resources and provides an interactive environment allowing researchers to learn and "play" using pre-configured, data, tools, and computational resources. Importantly, our approach is community driven and thus does not rely on a particular principal investigator, research group, or institution making it potentially more robust and sustainable.

Since 2006, our team has been pondering the question of how to enable computationally naive users to perform complex data analysis tasks. We attempted to solve this problem by creating a platform, Galaxy (http://galaxyproject.org; Afgan et al., 2016), that provides access to hundreds of tools used in a wide variety of analysis scenarios. It features a web-based user interface while automatically and transparently managing underlying computation details. It can be deployed on a personal computer, heterogeneous computer clusters, as well as computation systems provided by Amazon, Microsoft, Google, and other clouds, such as those running OpenStack. Over the years, a community has formed around this project, providing it with an ever-growing, up-to-date set of analysis tools and expanding it beyond life sciences.

These features of Galaxy attracted many biomedical researchers, making it well suited for use as a teaching platform. Here we describe a community-driven effort to build, maintain, and promote a training infrastructure designed to provide computational data analysis training to biomedical researchers worldwide.

RESULTS AND DISCUSSION

Our goal is to develop an infrastructure that facilitates data analysis training in life sciences. At a minimum, it needs to provide an interactive learning platform tuned for current datasets and research problems. It should also provide means for community-wide content creation and maintenance, and, finally, enable trainers and trainees to use the tutorials in a variety of situations, such as those where a reliable Internet access is not an option.

Interactive Learning Tailored to Research Problems

We produced a collection of hands-on tutorials that are designed to be interactive and are built around Galaxy. The hands-on nature of our training material requires that a trainee has two web-browser windows open side by side: one pointed at the current tutorial and the other at a Galaxy instance. We build most tutorials around a "research story": a scenario inspired by a previously published manuscript or an interesting dataset (with the caveat that some more technical materials do not lend themselves to this goal). To make training comprehensive, we aim to cover major branches of biomedical big-data applications, such as those listed in Table 1. Please note that, while we are using Galaxy as an analysis platform, it is not the only way to analyze biomedical data. Thus we design tutorials to teach underlying concepts that will be useful outside Galaxy.

As an example, suppose that a researcher is interested in learning about metagenomic data analyses. The category "Metagenomics" at https://training.galaxyproject.org presently contains a set of introductory slides, two hands-on tutorials, and HTML-based slides designed as a brief (10–20 min) introduction to the subject. In addition, every hands-on tutorial contains background information and explains how it influences data analysis (e.g., Figure 1). This background story is included to account for situations when tutorials are used for self-teaching in the absence of an instructor who would provide a formal introduction. After the introduction, the hands-on part of the tutorial begins and is laid out in a step-by-step fashion with explanations (boxes in Figure 1) of what is being done inside Galaxy, which parameters are critical, and how modifying parameters affects downstream results. The first step in this progression is usually a description of the datasets and how to obtain them. We invested a large effort in creating appropriate datasets by downsampling original published data, which is necessary since real-world datasets are usually too big for tutorials. Our goal was to make datasets as small as possible while still producing an interpretable result. We use Zenodo (http://www.zenodo.org), an open data archiving and distribution platform, to store the tutorial datasets and to provide them with stable digital object identifiers (DOIs) that can be used to credit their authors and for citation purposes.

Tutorials start with a list of prerequisites (typically other tutorials within the site) to account for the variation in trainees' backgrounds, a rough time estimate, questions addressed during the tutorial, learning objectives, and key points. These components help trainees and instructors to keep track of the training goals. For example, the learning objectives are single sentences describing what a trainee will be able to do as a result of the training (Via et al., 2013). Throughout the tutorials, question boxes (Figure 1) are added as an effective way to motivate the trainees (Dollar et al., 2007; Scheines et al., 2005) and guide self-training. The training material is distributed under a Creative Commons BY 4.0 (https://creativecommons.org/licenses/by/4.0/) license: its contents can be shared and adapted freely as long as appropriate credit is given. Efforts have been made also in the direction of ensuring website accessibility to disabled persons by regular evaluation with WAVE (http://wave.webaim.org), a web accessibility evaluation tool, and by automatic checking for alternative text for the images.

Table 1. Topics Available in the Galaxy Training Material Website (https://training.galaxyproject.org) with Their Target Users and Available Tutorials

Topic	Target	Tutorials
Galaxy Server administration	Admin	Galaxy database schema; Docker and Galaxy; advanced customization of a Galaxy instance
Assembly	Biol	Introduction to genome assembly, De Bruijn graph assembly, Unicycler assembly
ChIP-seq data analysis	Biol	Identification of the binding sites of the T cell acute lymphocytic leukemia protein 1, identification of the binding sites of the estrogen receptor
Development in Galaxy	Dev	Contributing with GitHub, tool development and integration into Galaxy, Tool Shed: sharing Galaxy tools, Galaxy interactive tours, Galaxy interactive environments, visualizations: charts plugins, Galaxy Webhooks, visualizations: generic plugins, BioBlend module, a Python library to use Galaxy API, tool dependencies and Conda, tool dependencies and containers, Galaxy code architecture
Epigenetics	Biol	DNA methylation
Introduction to Galaxy	Biol	Galaxy 101, from peaks to genes, multisample analysis, options for using Galaxy, IGV introduction, getting data into Galaxy
Metagenomics	Biol	16S microbial analysis with mothur, analyses of metagenomics data - the global picture
Proteomics	Biol	Protein FASTA database handling, metaproteomics tutorial, label-free versus labelled - how to choose your quantitation method, detection and quantitation of N termini via N-TAILS, peptide and protein ID, secretome prediction, peptide and protein quantification via stable isotope labeling
Sequence Analysis	Biol	Quality control, mapping, genome annotation, RAD-seq reference-based data analysis, RAD-seq de novo data analysis, RAD-seq to construct genetic maps
Train the trainers	Inst	Creating a new tutorial - writing content in Markdown; creating a new tutorial - defining metadata; creating a new tutorial - setting up the infrastructure; creating a new tutorial - creating Interactive Galaxy Tours; creating a new tutorial - building a Docker flavor for a tutorial; good practices to run a workshop
Transcriptomics	Biol	De novo transcriptome reconstruction with RNA-seq, reference-based RNA-seq data analysis, differential abundance testing of small RNAs

Admin, Galaxy administrators; Biol, biomedical researchers; Dev, tool and software developers; Inst, instructors and tutorial developers; RAD-seq, restriction site-associated DNA sequencing; RNA-seq, RNA sequencing. The scripts for the extraction of such information are available in GitHub (https://github.com/bebatut/galaxy-training-material-stats). This table displays content current as of 28th Sep, 2017.

Keeping trainees engaged is critical, particularly for self-training. To this end, we aim to provide interactive tours for each tutorial: using instruction bubbles, each tutorial step can be performed by the user directly inside Galaxy, guiding learners to the needed tools while also allowing exploration of the framework's functionalities. Tours can be created directly in the browser using our tour creator plugin (https://zenodo.org/record/830481).

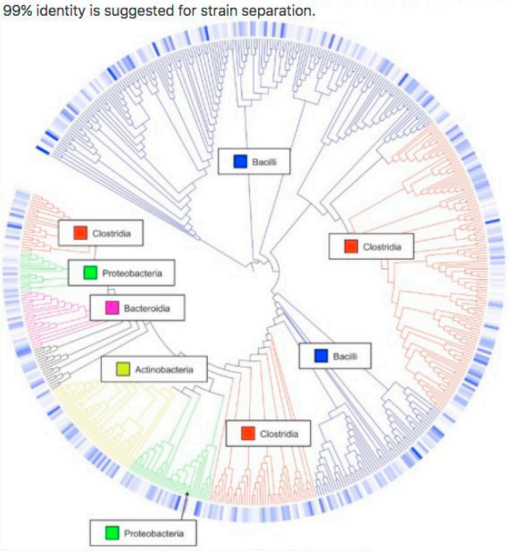
Infrastructure to Facilitate Community-Led Content Development

To build a comprehensive collection of training materials covering the spectrum of topics in the life sciences, we must leverage community expertise, as no single group can possibly know it all. To achieve this goal, we built an infrastructure that makes tutorial creation a convenient, hassle-free process and enables transparent peer-review and curation to guarantee high-quality and current content. In implementing these requirements, we took inspiration from the Software and Data Carpentry (SDC) projects (Wilson, 2014). In SDC, materials are openly reviewed and iteratively developed on GitHub (https://github.com/) to capture the breadth of community expertise. SDC delivers training via online tutorials with hands-on sections, which offer better training support than videos because trainees who are actively participating learn more (Dollar et al., 2007). This format is also adapted to face-to-face courses and self-training, as the content is openly accessible online. The content of these web pages is easy to edit, thus reducing the contribution barrier. The tutorials are developed in Markdown, a plain text markup language, which is automatically transformed into web-browser-accessible pages. Using these strategies, we created a GitHub repository (https://github.com/galaxyproject/training-material) to collect, manage, and distribute training materials. The architecture of this infrastructure is shown in Figure 2 (center), with the process for developing a tutorial illustrated at the bottom of the figure. To create a new tutorial, the main repository is "forked" (duplicated into a user-controlled space) within GitHub by an individual developing the tutorial. The developer then proceeds to write the content using Markdown, as explained in our guide at https://training.galaxyproject.org/topics/contributing (itself consisting of several tutorials). The guide contains detailed information on technical and stylistic aspects of tutorial development. After settling on a final version of the tutorial (circles 1–10, bottom of Figure 2), a "pull request" is created against the original repository. When a new pull request is issued, this is an indication that a new tutorial is ready to be

A

Background: Operational Taxonomic Units (OTUs)

In 16S metagenomics approaches, OTUs are clusters of similar sequence variants of the 16S rDNA marker gene sequence. Each of these clusters is intended to represent a taxonomic unit of a bacteria species or genus depending on the sequence similarity threshold. Typically, OTU cluster are defined by a 97% identity threshold of the 16S gene sequence variants at species level. 98% or 99% identity is suggested for strain separation.



(Image credit: Danzeisen et al. 2013, 10.7717/peerj.237)

B

Hands-on: Cluster mock sequences into OTUs

First we calculate the pairwise distances between our sequences

- **Dist.seqs** with the following parameters
  - "fasta" to the fasta from Get.groups
  - "cutoff" to `0.20`

Next we group sequences into OTUs

- **Cluster** with the following parameters
  - "column" to the dist output from Dist.seqs
  - "count" to the count table from Get.groups

Now we make a *shared* file that summarizes all our data into one handy table

- **Make.shared** with the following parameters
  - "list" to the OTU list from Cluster
  - "count" to the count table from Get.groups
  - "label" to `0.03` (this indicates we are interested in the clustering at a 97% identity threshold)

And now we generate intra-sample rarefaction curves

- **Rarefaction.single** with the following parameters
  - "shared" to the shared file from Make.shared

Question

How many OTUs were identified in our mock community?

▶ Click to view answer

Figure 1. Key Elements of an Interactive Tutorial

(A) A fragment of introductory material within a tutorial.

(B) A "hands-on" element in the upper box contains instructions for running a tool inside Galaxy. The question box at the bottom contains an answer field that can be toggled.

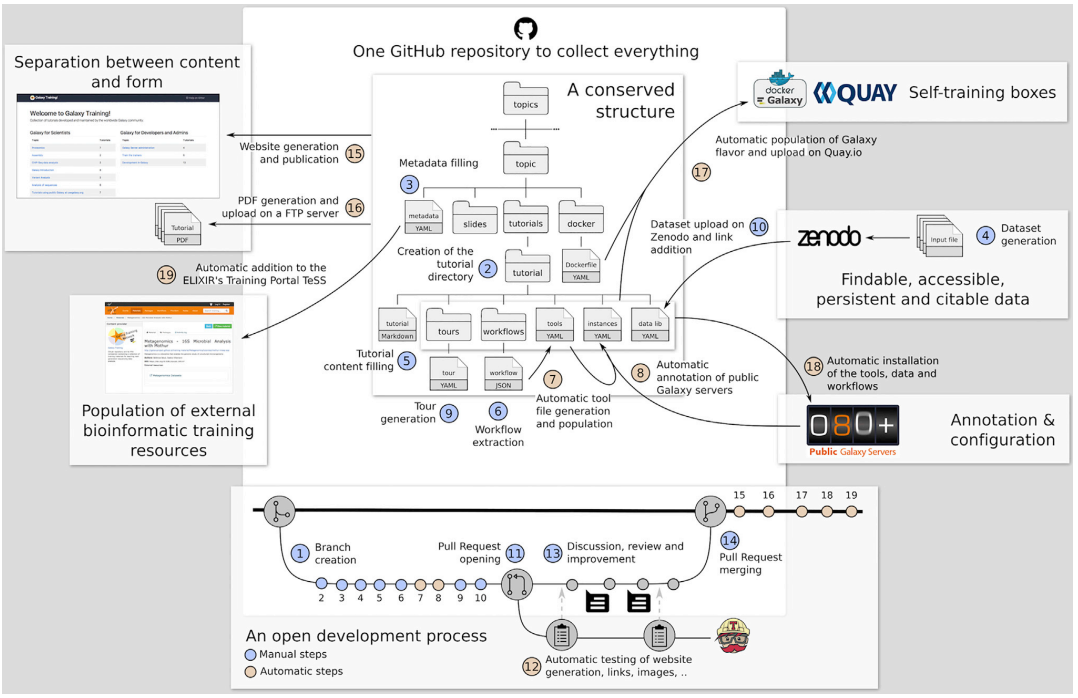


Figure 2. Structure and Development of Content in GitHub (http://github.com/galaxyproject/training-material)
The material is organized in different topics, each topic in a dedicated directory. Inside each topic's directory, the structure is the same: a metadata file, a directory with the topic introduction slide decks, a directory with the tutorials, and a directory with the Dockerfile describing the details to build a container for the topic that would contain a dedicated Galaxy instance with all tools relevant for the tutorials. Inside the topic directory, each tutorial related to the topic has its own subdirectory with several files: a tutorial file written in Markdown with hands-on, an optional slides file to support the tutorial, a directory with Galaxy interactive tours to reproduce the tutorial, a directory with workflows extracted from the tutorial, a file with the links to the input data needed for the tutorial, and a file with the description of needed tools to run the tutorial. The process of development of new content is shown at the bottom.

reviewed by the editorial team. The team then makes suggestions on the new contents, these suggestions are discussed, and the content is edited accordingly. A decision is then made whether to accept the pull request. At the same time the pull request is first created, the newly added content is automatically tested for HTML generation and all links and images are verified. When the pull request is accepted, the new tutorial becomes a part of the official training material portfolio, and the entire site is regenerated. This open strategy for content creation started paying off early as we already have over 60 individuals contributing and editing content within the GitHub repository.

This infrastructure has been developed in accordance with the FAIR (findable, accessible, interoperable, reusable) principles (Wilkinson et al., 2016). Each tutorial, slide deck, and topic is complemented by numerous metadata described in a standard, accessible, interoperable format (YAML; http://yaml.org/). The metadata are used to automatically populate the TeSS training portal at the European Life Sciences Infrastructure for Biological Information (ELIXIR; https://tess.elixir-europe.org), ensuring global reach (Beard et al., 2016). Each topic, tutorial, and slide deck has as metadata a reference to a topic in the EDAM ontology (Ison et al., 2013), a comprehensive catalog of well-established, familiar concepts that are prevalent within bioinformatics and computational biology. These references can be used to represent relationships among the materials and make them more findable and searchable.

Using the framework described above, we relaunched the Galaxy Training Network (GTN; https://galaxyproject.org/teach/gtn). This growing network currently consists of 33 scientific groups (https://galaxyproject.org/teach/trainers) invested in Galaxy-based training. The GTN regularly organizes training events worldwide (Figure S1) and offers best practices for developing Galaxy-based training material, advice on computer platform choice to use for training, and a catalog of existing training resources for Galaxy (Table 1).

As of writing (March 2018) 64 individuals contributed to development of infrastructure and tutorials (http://bit.ly/gxy_tr_people). Of these, only 18 individuals are associated with the two largest Galaxy Project installations in the United States (http://usegalaxy.org) and Germany (http://usegalaxy.eu). This ratio (46/18 ≈ 2.5) is an indicator of community engagement and our goal is to increase it.

Ensuring Accessibility of Tutorials

Most training materials hosted within the GTN resource are intended to be used side by side with the Galaxy framework. However, the main public Galaxy instances (e.g., https://usegalaxy.org or https://usegalaxy.eu) are occasionally subject to unpredictable load, may be inaccessible due to network problems in remote parts of the world, or may not have all the tools necessary for completing the tutorials. To account for these situations, we have developed a Docker-based framework for creating portable, on-demand Galaxy instances specifically targeted for a given tutorial. Docker (https://www.docker.com) is a container platform that provides lightweight virtualization by executing "images" (files that include everything needed to run a piece of software) isolated from the host computer environment. An individual creating a new tutorial lists all tools that are required to complete it in a dedicated configuration file (tools file, Figure 2). For example, a metagenomics tutorial uses the mothur (Schloss et al., 2009) set of tools as well as visualization applications such as Krona (Ondov et al., 2015). The corresponding Galaxy tools are listed in a configuration file that is a part of the metagenomics tutorial. This file is used to install these Galaxy tools and their dependencies into a base Galaxy Docker image (containing essential Galaxy functionality and a core set of tools) to create a dedicated ''on-demand'' Galaxy instance that can then be used on any trainer's or trainee's computer. The Docker image also contains input data, tours, and workflows.

A Vision for the Future

Life sciences are on a trajectory toward becoming an entirely data-driven scientific domain. A growing understanding that biomedical curricula must be modernized to reflect these changes is gaining attention (Hitchcock et al., 2017). Our project represents one of the first fully open, ''grass-roots'' attempts at unifying and standardizing heterogeneous training resources around the Galaxy platform. While it may not be appropriate to all, our multi-year experience with teaching workshops at various skill levels can be summarized as the following set of recommendations, which we use as guiding principles. These recommendations may also be useful for the development of alternative frameworks as well as for curriculum planning:

1. Require quantitative training. No one expects biomedical researchers to rival their colleagues in departments of mathematics or statistics. However, background level statistical reasoning must be included in all training materials and general statistical courses must become a part of undergraduate and graduate education. This would have an enormous positive impact on the quality of biomedical research because researchers with basic understanding of quantitative concepts will not, for example, perform an RNA sequencing experiment without a sufficient number of replicates. While our current set of tutorials lacks in-depth statistical analyses of the data, we are planning to change this. Our integration with Jupyter is the first step in this direction (Grüning et al., 2017).
2. Demystify computational methodologies. Fundamental principles, limitations, and assumptions of molecular experimental techniques are typically well understood by biomedical researchers even when proprietary reagent kits are used. This is not the case with software tools, which are often treated as black boxes. We argue that fundamental principles of bioinformatic techniques (e.g., read mapping, read assembly) must be understood by experimentalists as this will also lead to an increase in overall quality of research output.
3. Advocate the fundamental virtues of open and transparent research. Open and transparent data analysis (e.g., through the use of open-source software) promotes replication and validation of results by independent investigators. It also speeds up research progress by facilitating reuse and repurposing of published analyses to different datasets or even to other disciplines. We advocate openness as a basic principle for computational analysis of biomedical data.

The infrastructure presented here has been developed to support training using Galaxy, a powerful tool for teaching bioinformatics concepts and analysis, but such a model is not only limited to Galaxy. It could be applied to bioinformatics training more generally (and to other disciplines as well) to support learners and instructors in this ever-changing landscape that is the life sciences.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes one figure and can be found with this article online at https://doi.org/10.1016/j.cels.2018.05.012.

ACKNOWLEDGMENTS

The authors are grateful to the Freiburg Galaxy and Core Galaxy teams as, without these resources, this work would not be possible. Adoption of Galaxy Tours has been accelerated with the introduction of Galaxy Tour Builder (https://zenodo.org/record/830481) by William Durand (https://tailordev.fr). This project was supported by Collaborative Research Centre 992 Medical Epigenetics (DFG grant SFB 992/1 2012), German Federal Ministry of Education and Research (BMBF grant 031 A538A RBC [de.NBI]), NIH grants U41 HG006620 and R01 AI134384-01, as well as NSF grant 1661497.

AUTHOR CONTRIBUTIONS

B.B., S.H., and B.G. developed the conceptual foundation for the training infrastructure, developed the proof of principle, and outlined the software process. B.B., S.H., R.B., and A.N. wrote the manuscript. A. Bagnacani, D.B., V.B., C.B., A. Bretaudeau, L.B.-G., M.Č., J.C., D.C., G.T.N., O.D.-A., A.E., M.A.F., S.G., Y.H., H.-R.H., T.H., P.J., D.L., G.L.C., T.M., F.M., F.R., D.R., F.C.S., N.S., J.T., J.W., P.V., M.W., A.W., and D.Y. contributed software components and tutorials and also edited and commented on the manuscript.

CellPress

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 28, 2017
Revised: March 10, 2018
Accepted: May 18, 2018
Published: June 27, 2018

REFERENCES

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., et al. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. *44*, W3–W10.

Attwood, T.K., Blackford, S., Brazas, M.D., Davies, A., and Schneider, M.V. (2017). A global perspective on evolving bioinformatics and data science training needs. Brief. Bioinform. https://doi.org/10.1093/bib/bbx100.

Barone, L., Williams, J., and Micklos, D. (2017). Unmet needs for analyzing biological big data: a survey of 704 NSF principal investigators. PLoS Comput. Biol. *13*, e1005755.

Beard, N., Attwood, T., and Nenadic, A. (2016). TeSS – training portal. F1000Res. *5*, https://doi.org/10.7490/f1000research.1112652.1.

Community Survey Report – 2013. EMBL Australia Bioinformatics Resource. https://www.embl-abr.org.au/news/braembl-community-survey-report-2013/.

Dollar, A., Steif, P.S., and Strader, R. (2007). Enhancing traditional classroom instruction with web-based Statics course. In: 2007 37th Annual Frontiers in Education Conference - Global Engineering: Knowledge without Borders, Opportunities without Passports.

Grüning, B.A., Rasche, E., Rebolledo-Jaramillo, B., Eberhard, C., Houwaart, T., Chilton, J., Coraor, N., Backofen, R., Taylor, J., and Nekrutenko, A. (2017). Jupyter and Galaxy: easing entry barriers into complex data analyses for biomedical researchers. PLoS Comput. Biol. *13*, e1005425.

Hitchcock, P., Mathur, A., Bennett, J., Cameron, P., Chow, C., Clifford, P., Duvoisin, R., Feig, A., Finneran, K., Klotz, D.M., et al. (2017). The future of graduate and postdoctoral training in the biosciences. Elife *6*, https://doi.org/10.7554/eLife.32715.

Ison, J., Kalas, M., Jonassen, I., Bolser, D., Uludag, M., McWilliam, H., Malone, J., Lopez, R., Pettifer, S., and Rice, P. (2013). EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats. Bioinformatics *29*, 1325–1332.

Larcombe, L., Hendricusdottir, R., Attwood, T.K., Bacall, F., Beard, N., Bellis, L.J., Dunn, W.B., Hancock, J.M., Nenadic, A., Orengo, C., et al. (2017). ELIXIR-UK role in bioinformatics training at the national level and across ELIXIR. F1000Res. *6*, https://doi.org/10.12688/f1000research.11837.1.

Ondov, B.D., Bergman, N.H., and Phillippy, A.M. (2015). Krona: interactive metagenomic visualization in a web browser. In Encyclopedia of Metagenomics, K.E. Nelson, ed. (Springer), pp. 339–346.

Scheines, R., Leinhardt, G., Smith, J., and Cho, K. (2005). Replacing lecture with web-based course materials. J. Educ. Comput. Res. *32*, 1–25.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. *75*, 7537–7541.

Stephens, Z.D., Lee, S.Y., Faghri, F., Campbell, R.H., Zhai, C., Efron, M.J., Iyer, R., Schatz, M.C., Sinha, S., and Robinson, G.E. (2015). Big data: astronomical or genomical? PLoS Biol. *13*, e1002195.

Via, A., Blicher, T., Bongcam-Rudloff, E., Brazas, M.D., Brooksbank, C., Budd, A., De Las Rivas, J., Dreyer, J., Fernandes, P.L., van Gelder, C., et al. (2013). Best practices in bioinformatics training for life scientists. Brief. Bioinform. *14*, 528–537.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. Sci. Data *3*, 160018.

Williams, J.J., and Teal, T.K. (2017). A vision for collaborative training infrastructure for bioinformatics. Ann. N. Y. Acad. Sci. *1387*, 54–60.

Wilson, G. (2014). Software carpentry: lessons learned. F1000Res. *3*, 62.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
GitHub software development and distribution platform		github.com
Galaxy Project		galaxyproject.org
Galaxy main public site US		usegalaxy.org
Galaxy main public site EU		usegalaxy.eu
Galaxy Training Materials	This paper	http://github.com/galaxyproject/training-material

CONTACT FOR REAGENT AND RESOURCE SHARING

The Lead Contact is Anton Nekrutenko (anton@nekrut.org). There are no restrictions on the use of reported resources. This resource is licensed under the Creative Commons Attribution 4.0 International License.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

No experiments have been performed in the framework of this study.

METHOD DETAILS

Training materials are developed using Markdown markup language and served from the GitHub platform. Extensive description of the development process and content structure can be found at https://github.com/galaxyproject/training-material/blob/master/CONTRIBUTING.md.

QUANTIFICATION AND STATISTICAL ANALYSIS

No quantification or statistical analyses have been performed in this study.

DATA AND SOFTWARE AVAILABILITY

All tutorials are available from http://galaxyproject.github.io/training-material/.

2.2 Application and validation of workflows via network analysis and modeling

*This section contains the **detailed results** and **further applications** of the RNA-Seq analysis workflows that were developed. The focus is the integration of RNA-Seq data and network analysis approaches to investigate and evaluate the interaction between RNA transcripts. The results and **validation** experiments include the identification of significantly differentially expressed genes, gene annotation clustering, and the gene expression interaction network generation of important genes for cardiomyocyte differentiation. In addition, the results obtained from TRAPLINE are the basis for a WGCN analysis that was conducted to investigate the co-expression of a transcript of interest. The analysis thus gives insights into the SA node and the influence in relation to the heart rate. Finally, another community effort to generatw a large scale whole-cell model is presented.*

2.2.1 Evaluation of cardiomyocyte subtypes for cardiac regeneration

Hausburg, F., Jung, J.J., Hoch, M., **Wolfien, M.**, Yavari, A., Rimmbach, C., and David, R. (2017).
(Re-)programming of subtype specific cardiomyocytes.
Advanced Drug Delivery Reviews. IF: 16.361, Citations (December 14, 2020): 6

Adult CM possess a highly restricted intrinsic, regenerative potential. They yield a major barrier to the effective treatment of a range of chronic, degenerative cardiac disorders, which are characterized by cellular loss and/or irreversible dysfunctions underliying the majority of deaths in developed countries. In this article, we highlight both stem cell programming and direct cell reprogramming that hold promise as novel, potentially curative approaches to address this therapeutic challenge. We also reflect on the advent of induced pluripotent stem cells that started other stem cell source investigations beside embryonic stem cells to enable an autologous CM production. Different clinical scenarios will require the generation of highly-pure, specific cardiac subtypes.

Here, I applied and transfered the RNA-Seq data analysis workflow TRAPLINE to a cardiac use case, in which different stem cell-derived CM subtypes are compared. Likewise data analysis and systems-based network approaches are used to enhance nodal cell programming. I developed this systems-based data analysis workflow in Cytoscape[16] to identify enriched subnetworks. The molecular information (e.g., gene expression data) of these subnetworks are subsequently linked to characterize phenotypic processes (e.g., contraction, electrophysiology, metabolism) of CM cell types via integrating external databases (e.g., STRING, BioGrid), and text mining tools.

In summary, we discuss the cardiovascular stem cell and programming field, including a focus on the emergent topic of pacemaker cell generation for the development of biological pacemakers and *in vitro* drug testing.

[16] https://cytoscape.org/

Advanced Drug Delivery Reviews 120 (2017) 142–167

Contents lists available at ScienceDirect

Advanced Drug Delivery Reviews

journal homepage: www.elsevier.com/locate/addr

(Re-)programming of subtype specific cardiomyocytes☆

Frauke Hausburg [a,b,1], Julia Jeannine Jung [a,b,1], Matti Hoch [c], Markus Wolfien [c], Arash Yavari [d,e,f], Christian Rimmbach [a,b], Robert David [a,b,*]

[a] Reference and Translation Center for Cardiac Stem Cell Therapy (RTC), Department of Cardiac Surgery, Rostock University Medical Center, Schillingallee 69, 18057 Rostock, Germany
[b] Department Life, Light and Matter of the Interdisciplinary Faculty at Rostock University, Albert-Einstein-Straße 25, 18059 Rostock, Germany
[c] Department of Systems Biology and Bioinformatics, University of Rostock, Ulmenstraße 69, 18057 Rostock, Germany
[d] Experimental Therapeutics, Radcliffe Department of Medicine, University of Oxford, UK
[e] Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, UK
[f] The Wellcome Trust Centre for Human Genetics, Oxford, UK

ARTICLE INFO

Article history:
Received 14 June 2017
Received in revised form 29 August 2017
Accepted 7 September 2017
Available online 12 September 2017

Keywords:
ESC
iPSC
Cardiovascular development
Subtype differentiation
System-based data analysis
Nodal cells
Pacemaker

ABSTRACT

Adult cardiomyocytes (CMs) possess a highly restricted intrinsic regenerative potential — a major barrier to the effective treatment of a range of chronic degenerative cardiac disorders characterized by cellular loss and/or irreversible dysfunction and which underlies the majority of deaths in developed countries. Both stem cell programming and direct cell reprogramming hold promise as novel, potentially curative approaches to address this therapeutic challenge. The advent of induced pluripotent stem cells (iPSCs) has introduced a second pluripotent stem cell source besides embryonic stem cells (ESCs), enabling even autologous cardiomyocyte production. In addition, the recent achievement of directly reprogramming somatic cells into cardiomyocytes is likely to become of great importance. In either case, different clinical scenarios will require the generation of highly pure, specific cardiac cellular-subtypes. In this review, we discuss these themes as related to the cardiovascular stem cell and programming field, including a focus on the emergent topic of pacemaker cell generation for the development of biological pacemakers and *in vitro* drug testing.

Contents

Abbreviations: ADSC, adipose tissue-derived mesenchymal stem cell; Alcam, activated leukocyte cell adhesion molecule; AMI, acute myocardial infarction; ANF, natriuretic factor; AP, action potential; ASC, adult stem cell; AV, atrioventricular; AVB, atrioventricular bundle; AVN, atrioventricular node; BB, bundle branch; BCT, bioartificial cardiac tissue; bHLH, basic helix-loop-helix; Bry, Brachyury; CABG, coronary artery bypass graft; $Ca_v1.3$, calcium voltage-gated channel subunit alpha1 D; $Ca_v3.1$, calcium voltage-gated channel subunit alpha1 G; CCS, cardiac conduction system; CF, cardiac fibroblast; CHD, congenital heart disease; CM, cardiomyocyte; CMPC, cardiomyocyte progenitor cell; CMVEC, cardiac microvascular endothelial cell; CPC, cardiac progenitor cell; CS, conduction system; CV, cardiovascular; CVD, cardiovascular disease; Cx30.2, connexin30.2; Cx40, connexin40; Cx43, connexin43; Cx45, connexin45; ECG, electrocardiogram; EMILIN2, elastin microfibril interface 2; EPC, endothelial progenitor cell; EPCS, electric-pulse current stimulation; ESC, embryonic stem cell; FDA, Food and Drug Administration; FGF, fibroblast growth factor; FHF, first (primary) heart field; GF, growth factor; GFP, green fluorescence protein; GO, Gene Ontology; HCN4, hyperpolarization-activated cyclic nucleotide-gated cation channel 4; hPSCreg, Human Pluripotent Stem Cell registry; HTS, high-throughput sequencing; iCM, induced cardiomyocyte; iPSC, induced pluripotent stem cell; iSAB, induced sino-atrial body; Isl1, ISL LIM homeobox 1; JNK, c-Jun N-terminal kinase; LVEF, left ventricular ejection fraction; MAPK, mitogen-activated protein kinase; MB, molecular beacons; MEA, multi-electrode-array; Mlc2v, myosin, light polypeptide 2, regulatory, cardiac, slow; MSC, mesenchymal stem cell; Myh6, myosin, heavy chain 6, cardiac muscle, alpha; Myh7, myosin, heavy polypeptide 7, cardiac muscle, beta; $Na_v1.5$, sodium voltage-gated channel alpha subunit 5; Nkx2-5, NK2 homeobox 5; NPPA, natriuretic peptide A; ODE, ordinary differential equation; PA, polyacrylate; PDMS, polydimethylsiloxane; PLGA, polylactide-co-glycolide; PMC, pacemaker cell; PPT, protein-protein interaction; PSC, pluripotent stem cell; Rarg, retinoic acid receptor, gamma; ROCK, rho-associated, coiled-coil containing protein kinase; Rxra, retinoid X receptor, alpha; SA, sino-atrial; SAN, sinoatrial node; SCD, sudden cardiac death; SCN5A, sodium channel, voltage-gated, type V, alpha subunit; SHF, second heart field; Shox2, short stature homeobox 2; SIRPA, signal-reduced protein alpha; SSS, sick sinus syndrome; Tbx18, T-box 18; Tbx3, T-box 3; TF, transcription factor; THF, tertiary heart field; VCAM1, vascular cell adhesion molecule 1; VCS, ventricular conduction system; VEGF, vascular endothelial growth factor; wt, wild-type.

☆ This review is part of the *Advanced Drug Delivery Reviews* theme issue on "Advances in Stem Cell-Based Therapies".

* Corresponding author at: RTC, Department of Cardiac Surgery, Rostock University Medical Center, Schillingallee 69, 18057, Rostock, Germany.
E-mail address: robert.david@med.uni-rostock.de (R. David).

[1] Shared first authors.

http://dx.doi.org/10.1016/j.addr.2017.09.005
0169-409X/

1. Introduction

The advent of regenerative medicine has opened up new perspectives for so far insoluble clinical problems. Recent progress in understanding the biology of stem cell pluripotency and endogenous repair mechanisms has fostered a deeper understanding of its remarkable therapeutic potential for tissue repair or replacement. Such novel approaches are urgently required to effectively treat the growing burden of disorders characterized by irreversibly damaged or diseased tissue resulting in loss of organ/tissue function associated with a rapidly ageing population. Furthermore, through the production of autologous pluripotent stem cells, regenerative strategies hold promise in providing truly patient-specific therapies for structural and functional repair in disease.

Cardiovascular disease (CVD) is the leading cause of death worldwide (accounting for 31.3% in 2015) and is projected to rise further (WHO 2017). CVD encompasses a range of chronic disease states, including ischemic, rheumatic and hypertensive heart disease, in addition to extra-cardiac disorders such as stroke. Heart failure represents the final common phenotype resulting from a diverse range of inherited and acquired cardiac insults and affects ~26 million individuals worldwide [1]. Individuals with severe heart failure have a dismal prognosis with a worse 5-year adjusted mortality than many cancers [2]. To date, allogeneic heart transplantation remains the only available treatment option for patients with end-stage heart failure who are symptomatic despite optimal medical and device (cardiac resynchronization) therapy [3,4]. Despite advances in surgical technique, perioperative management and immunomodulation, a major limitation to its wider application is donor organ scarcity: in Europe in 2015, only 604 donor organs were successfully engrafted, while 1140 patients are on the active Eurotransplant waiting list [5]. An additional 209 recipients died before they could undergo heart transplantation [5]. Even for those transplanted, while symptomatic improvement and survival are in general markedly improved, outcomes (median ~11 year survival) are limited by long-term complications, in part associated with immunosuppression, including malignancy, infection, renal dysfunction and allograft vasculopathy [6]. In view of such limitations, highly innovative approaches are under exploration with the ultimate goal of establishing safe, durable cellular replacement and repair of injured or diseased myocardium, in addition to *in vitro* disease modeling and drug development applications [7–9]. A key requirement for these approaches is to ensure highly reliable and robust generation of fully functional cardiomyocytes with physiological properties as close as possible to their natural counterparts. Partially or terminally differentiated cells offer a relevant alternative to somatic stem cell transplantation, given that the latter are still a matter of controversial debate regarding their moderate therapeutic outcomes [10,11]. Pluripotent stem cells (PSC) and their derivatives offer an attractive source for both cell replacement and studying key cellular and molecular processes involved in cardiovascular disease. Equally, resident cells (*e.g.* fibroblasts) may also represent a readily accessible source of cells to study cell fate transition not only within, but even across, germ layers.

2. Tissue regeneration and repair for cardiovascular disease

Normal cardiac function and physiological homeostasis is achieved through the complex interaction of a diverse range of cell types broadly constituting myocyte, vascular and stromal compartments. Even among specific cell types, such as cardiomyocytes (CM), there exist different phenotypes (*e.g.* sinoatrial, atrial, nodal, Purkinje and ventricular). Disease processes do not affect all these cell types uniformly, with relatively greater impact on specific tissue components such as fibrosis or vascular insufficiency.

The human heart does exhibit some regenerative potential, albeit very low, with an annual cardiomyocyte turnover rate of 1% at age 25 years, reducing further to 0.45% by 75 years [12]. As a corollary, adult human cardiomyocytes are long-lived cells, such that <50% will be replaced over a life-span of 75 years. In contrast, the proportion of CM situated in mitosis and cytokinesis is highest in infancy and contributes to developmental growth, suggesting significant cardiac regenerative potential in children and adolescents [13]. Other studies, including data from animal models, have highlighted that CMs, upon transition from the mononucleate to a mature binucleate state, exit the cell cycle and lose their proliferative potential during a short postnatal period [14–16]. In the setting of common CVD such as acute myocardial infarction (MI), leading to the abrupt loss of up to ~1 billion CM, this intrinsic regeneration potential is vastly inadequate, resulting in structural (*i.e.* scar) rather than functional (*i.e.* contractile) repair, and potentially to progressive deleterious ventricular remodeling and post-MI heart failure. However, the identification of adult CM repopulation raises the possibility that either normally resident cell populations such as cardiac progenitor cells (CPCs), or pre-existing CM may represent sources for myocardial repair post-injury [13,17,18].

Accordingly, development of experimental protocols to robustly generate distinct cardiac cell types and define their specific clinical/preclinical applications is required. We will address the progress made recently with attempts at stem cell and somatic cell-based programming, detailing their therapeutic potential and current stage of development. A major contribution to these has been provided by applying insights gained from the study of cardiovascular developmental biology to which we turn our attention next.

3. Cardiogenesis during development and its regulation

Cardiac development occurs during the early stages of the embryonic phase, and is crucial to ensure adequate nutrient and oxygen supply to, as well as removal of waste from, the growing organism. The mature mammalian heart is highly complex in structure, divided macroscopically into four chambers macroscopically and constituting specific muscle and non-muscle cell types, including left and right atrial CM, left and right ventricular CM, and cells forming the conduction system, sinoatrial pacemaker, vascular smooth muscle, endo- and epicardium [19–23]. The generation of such developmentally diverse cell fates are achieved *via* spatiotemporally stringent molecular regulation, with clear evidence that myocardial cells derive from Brachyury$^+$ (Bry$^+$) mesodermal progenitor cells of the primitive streak during gastrulation through the impact of Wnt signaling [24–26].

Thereafter, two crucial transcription factors (TF) are regarded as cardiovascular fate-determining factors: the bHLH TF MesP1 (mesoderm posterior 1) [27–30] and the surface molecule Flk1 (also known as VEGFR2: vascular endothelial growth factor receptor 2) [31,32]. Further development is achieved from multipotent cardiac progenitor cells [33] and can be distinguished mainly in two origins: i) the first (primary) heart field (FHF) demarcating an Nkx2-5$^+$/Hcn4$^+$ cell population which forms the cardiac crescent [34–41], and ii) the second heart field (SHF) demarcating a Nkx2-5$^+$/Isl1$^+$ cell population derived from the pharyngeal mesoderm and lying medially and posterior to the FHF [41–46]. In avians, a decisive role for the tertiary heart field (THF) in pacemaker development of the sino-atrial (SA) node has also been reported [47,48]. Primary heart field progenitor cells will yield the myocardium of the left ventricle as well as a limited portion of the right ventricle, the right and left atria and large parts of the conduction system (CS), such as the atrioventricular (AV) node and the ventricular CS [20,42]. Multipotent progenitor cells of the SHF will yield myocardium of the right and left atria, the right ventricle and the outflow tract, as well as cardiac vascular smooth muscle and the endocardium [31,46,49]. In addition to these, epicardial progenitor cells give rise to cardiac fibroblasts, vascular smooth muscle, atrial and venous endothelial cells [20,50–52]. Moreover, pro-cardiogenic factors and signaling pathways play a decisive role during development and are distributed from the surrounding endoderm and mesoderm. These include bone morphogenetic proteins [53–57], notch [58], nodal and fibroblast growth factors [59–61], in addition to canonical and non-canonical Wnt/JNK [62–66].

A highly coordinated signaling network determines early cardiac progenitor as well as late specific cell fates, whose disruption can lead to abnormal embryonic development and congenital heart disease (CHD) characterized by malformation of specific cardiac structures [67]. CHD is the most common major congenital defect worldwide with a birth prevalence of between 0.58 and 0.9% [68,69]. Thus, dysregulation of TFs (*e.g.* Nkx2-5 [70–78], Gata4 [77,79–82] or members of the forkhead family [83]) is associated with various abnormalities including atrioventricular block, septal defects or pulmonary stenosis [84–86]. Exemplifying this, smoking-associated cardiac defects have been linked to promotor DNA hypermethylation of Tbx5 and Gata4 caused by maternal nicotine exposure [87]. In contrast, mutations in genes encoding cardiac ion channels are largely associated with phenotypes associated with sudden cardiac death (SCD) resulting from lethal arrhythmias [88].

4. Programming strategies for cardiovascular lineages

Insights into cardiac development and the potential serious sequelae arising from its disruption are a critical prerequisite for furthering disease modeling, drug development and cell replacement strategies.

In this chapter, we address approaches to enhance cardiovascular cell differentiation from adult stem cells (ASCs) ("directed differentiation"; Subsection 4.1) and from ESCs and iPSCs ("forward programming"; Subsection 4.2), in addition to discussing attempts to convert terminally differentiated somatic cell types into cardiac cells ("direct reprogramming"; Subsection 4.3) (Fig. 1).

The overriding aim of all programming strategies should be the generation of cells as physiologically close as possible to their natural counterparts. Moreover, to enable technology transfer from bench to bedside, procedures will have to be xeno-, serum-, feeder- and DNA-free. Accordingly, the approaches described below address potential options to overcome hurdles associated with the purity, yield and safety of physiologically functional CM subtypes (Fig. 1). As an exemplar of this, we describe approaches aimed at the generation of biological pacemaker cells for both therapeutic replacement and *in vitro* drug testing (Subsection 4.4), including our own recent progress in global

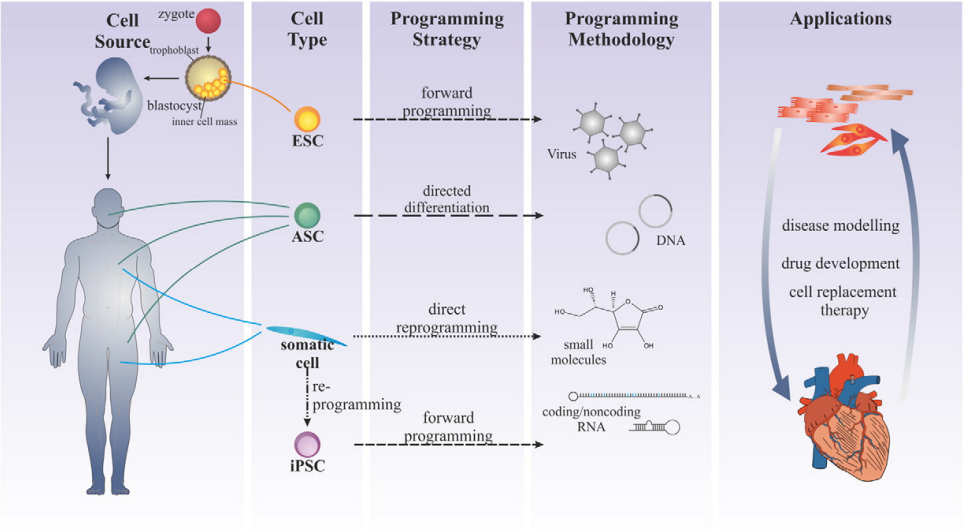


Fig. 1. Common programming strategies for cardiovascular lineages.

Table 1
Overview of recently published programming strategies of ASCs towards diverse cardiovascular subtypes.

Cell origin	Host	Delivery system	*In vivo/ in vitro*	Factor/substances	Target cell type	Special features	Literature
CMPC	Human	Synthetic nucleoside protein	*In vitro*	5 μmol/L 5-AZA following 1 ng/mL TGF-β1	CM-like	Spontaneously beating myocytes; gap-junctional communication and action potentials of maturing cardiomyocytes	Goumans et al., 2007 [117]
ADSC	Rat	Protein	*In vitro*	10 ng/mL TGF-β1	CM-like	Actin, cMhc	Gwak et al., 2009 [107]
ADSC	Human	Synthetic nucleoside	*In vitro*	10 μmol/L 5-AZA or 100 ng/mL TSA or co-culture with RNCM or modified cardiomyogenic medium	CM-like	Highest expression in direct contact co-culture with RNCM: Actin, Gata4, Nkx2-5, cTnT Spontaneous contractions Synchronous Ca^{2+} transient	Choi et al., 2010 [109]
CMPC	Human	miRNA	*In vitro*	miR-1 and miR-499	CM-like	Repression of HDAC4 and Sox6 Enhanced cardiomyogenesis	Sluijter et al., 2010 [118]
CMPC	Human	miRNA	*In vitro/in vivo*	miR-499	CM-like	Repression of Sox6 and Rod1 Enhanced cardiomyogenesis	Hosoda et al., 2011 [116]
ADSC	Rat	Synthetic nucleoside	*In vitro*	Planat-Bérnard or 5 μmol/L 5-AZA	CM-like	No spontaneous contraction Actn2, Cx-43	Carvalho et al., 2012 [105]
BMSC	Mouse	miRNA (Lentiviral)	*In vitro*	miR-1	CM-like	Downregulation of Hes-1 Expression of: Nkx2-5, GATA-4, cTnT, and Cx43	Huang et al., 2013 [102]
BMSC	Human	Synthetic nucleoside	*In vitro*	6 μmol/L 5-AZA or 10 ng/mL TGF-β1	CM-like	Expression of: GATA-4, Nkx2-5, Mlc-2a, actin Higher expression in AZA-group to control: Mlc-2a, Mlc-2v, cTnT	Mohanty et al., 2013 [100]
EPC	Human	Synthetic nucleoside	*In vitro*	5-AZA	CM-like	Expression of: Actn2, cTnT, cTnI and desmin	López-Ruiz et al., 2014 [112]
BMSC	Pig	Synthetic nucleoside lentiviral	*In vitro*	10 μmol/L 5-AZA IGF-1	CM-like	Expression of: GATA-4, Nkx2-5, β-MHC and MEF2c	Li et al., 2015 [99]
BMSC	Rabbit	DNA plasmid	*In vitro/in vivo*	GATA4, Nkx2-5 Extracellular environment co-culture with RNCM	CM-like	In combination with co-culture: significantly effective and enhance the ability to repair MI	Li and Zhang, 2015 [98]
BMSC	Mouse	Specific culture and substrate conditions	*In vitro*	0.3 mm-thick hcECM	CM-like	No evidence of CM differentiation	Oberwallner et al., 2015 [119]
BAT	Mouse	Specific culture conditions	*In vitro/in vivo*	1% methylcellulose/ Iscove's Modified Dulbecco's Medium containing hematopoietic cytokines	CCS-like	Regular beating Expression of: Nkx2-5, GATA6, Mef2c, ANF, α-MHC, β-MHC, MLC2a, MLC2v, but not GATA4	Takahashi et al., 2015 [110]
BMSC	Rat	Synthetic nucleoside	*In vitro/in vivo*	10 μmol/L 5-AZA	CM-like	Expression of: desmin, actin and cTnT	Yang et al., 2015 [97]
BMSC	Canine	Lentiviral	*In vitro*	Shox2 Co-culture with RNCMs	SAN-like	High levels of: Tbx3, HCN4, Cx45 Low levels of: Nkx2-5, Cx43 Able to pace RNCMs with a faster rate	Feng et al., 2016 [101]
SC-ADSC VL-ADSC CA-ADSC SS-ADSC	Mouse	Specific culture conditions	*In vitro/in vivo*	Medium suppl. with: - for vascular smooth muscle cell differentiation: TGF-β - for endothelial differentiation: hFGF, hVEGF, hIGF, AA, hEGF - for cardio-myocyte differentiation: PMA	CM-like, endothelial cells, vascular smooth muscle cells	Highest potential with: CA-ADSC	Nagata et al., 2016 [108]
DFC	Human	Small molecule	*In vitro/in vivo*	10 μM SAHA in ADMEM media; following continuous culture in media containing 1 μM of SAHA	CM-like	*In vitro*: expression of: α-SMA, TnnT2, desmin, Actc1 *In vivo* homing: 5.6 ± 1.0% heart 3.6 ± 1.1% liver 11.6 ± 2.7% kidney With differences in IL-2 and IL-10	Sung et al., 2016 [111]
E-ADSC P-ADSC O-ADSC	Human	Synthetic nucleoside retroviral	*In vitro*	10 μmol/L 5-AZA or ESRRG, GATA4, MEF2C, MESP1, MYOCD, TBX5, ZFPM2	CM-like	5-AZA: No increased expression of Actn2 or cTnT 7-factor-group: E-ADSC: increased Actn2 and cTnT	Wystrychowski et al., 2016 [106]

(continued on next page)

Table 1 (*continued*)

Cell origin	Host	Delivery system	*In vivo*/ *in vitro*	Factor/substances	Target cell type	Special features	Literature
BMSC	Human	Synthetic nucleoside	*In vitro*	10 μmol/L 5-AZA	CM-like	Upregulation of: Notch1, Gata4, Nkx2-5, α-actin, cTnT	Yu et al., 2016 [103]
BMSC	Mouse	lncRNA	*In vitro*	10 μmol/L 5-AZA Braveheart hypoxia/reoxygenation treatment	CM-like	Expressions of: α-actin, cTnT, Nkx2-5, Gata4, Gata6, Isl-1, EMT-associated genes (Snail, Twist, N-cadherin)	Hou et al., 2017 [104]

AA: ascorbic acid; Actc1: cardiac muscle alpha actin; Actn2: sarcomeric alpha-actinin; ADSC: adipose tissue-derived mesenchymal stem cells (E: epicardium, P: pericardium, O: omentum, SC: subcutaneous white adipose tissue, VL: visceral white adipose tissue; CA: cardiac brown adipose tissue, SS: subscapular brown adipose tissue); α-SMA: alpha-smooth muscle actin; BAT: brown adipose tissue derived stem cells from interscapular area; BMSC: bone marrow mesenchymal stem cells; CCS: cardiac conduction system; CM: cardiomyocyte; CMPC: cardiomyocyte progenitor cells; CPC: cardiac progenitor cells; cTnT: cardiac Troponin T; Cx43/45: Connexin43/45; DFC: dental follicle-derived mesenchymal stem cells; hcECM: human cardiac extracellular matrix; HDAC4: histone deacetylase 4; hEGF: human epidermal growth factor; hFGF: human fibroblast growth factor; hIGF: human insulin-like growth factor; hVEGF: human vascular endothelial growth factor; IL-2/10: interleukin-2/10; lncRNA: long noncoding RNA; miR: microRNA; Mlc: myosin light chain; PMA: phorbol myristate acetate; RNCM: rat neonatale cardiomyocytes; SAHA: suberoylanilide hydroxamic acid; TGF-β: transforming growth factor-β; TnnT2: cardiac muscle troponin T; TSA: trichostatin A; 5-AZA: 5-azacytidine.

transcriptome network analysis of "induced sino-atrial bodies (iSABs)" [89–91].

4.1. Forward programming of multipotent stem cells

4.1.1. Cardiac programming of adult stem cells

Early attempts at introducing adult (marrow) stem cells for transplantation therapy were performed in 1957 by E. Donnall Thomas between identical twins, with the recipient suffering from leukemia. This ushered in a phase of experimental work and clinical trials in hematopoietic transplantation which ultimately led to the award of Nobel Prize in Physiology or Medicine in 1990, for Thomas together with Joseph Murray "for their discoveries concerning organ and cell transplantation in the treatment of human disease" (http://www.nobelprize.org/nobel_prizes/medicine/laureates/1990/).

While the use of stem cell transplantation for cardiovascular disorders is a promising concept approaching clinical translation [10,92,93], based on neovascularization and improved endothelial function [94], its clinical efficacy has been the subject of repeated debate in view of the modest outcomes [10]. A small study of intramyocardial delivery of purified $CD133^+$ bone marrow SC demonstrated an encouraging but marginal improvement in left ventricular ejection fraction (LVEF) by ~6% after coronary artery bypass graft (CABG) surgery: CABG-only — preoperative 37.9 ± 10.3% to 41.3 ± 9.1% after 6 months; CABG with $CD133^+$ cell injection — preoperative 37.4% ± 8.4% to 47.1% ± 8.3% after 6 months [95]. Comparable results have been reported in studies of patients following acute myocardial infarction (AMI). Given the challenges in patient recruitment to cardiac regenerative trials [10], meta-analysis has been used to better discern the size of the potential therapeutic effect. A recent systematic review of 8 prospective randomized clinical trials containing 449 participants found no overall significant improvement in LVEF (1.47%, CI −4.5 to 7.45) in the setting of AMI following mesenchymal SC (MSC) transplantation [96]. However, exploratory subgroup analysis revealed a significant improvement in LVEF in those transplanted in the first week and also dependent upon cell dose administered (up to 3.3%).

(Re-)programming of ASCs may enable greater benefit from these multipotent cells. The sources of adult SC for potential cell fate alteration include MSC from bone marrow [97–104], adipose-tissue [105–110] or dental follicles [111], in addition to endothelial progenitor cells (EPCs) isolated from peripheral blood of patients with AMI or umbilical cord blood [112]. At least partial cardiogenic differentiation of these can be induced using a variety of exogenous manipulation strategies (Table 1) including: i) treatment with methylation inhibitors, such as 5-azacytidine and histone deacetylase inhibitors, such as trichostatin A [97,99,100,103–106,109,112]; ii) co-culture with isolated neonatal cardiomyocytes [98,101,109]; iii) forced exogenous overexpression of either TFs, such as Shox2 [101] and Gata4, Nkx2-5 [98], or use of a TF-cocktail [106], or non-coding RNAs, including miRNAs [102] and lncRNAs [104]; or iv) stimulation *via* media supplemented with various growth factors [100,107,108] or small molecules such as ascorbic acid [108] or suberoylanilide hydroxamic acid [111].

Recently published reports on cardiomyogenic differentiation suggest that cardiac marker expression of MSC- and $CD34^+$ progenitor cell-derivatives are actually based on fusion with endogenous cardiomyocytes of the recipient rather than on trans-differentiation *in vivo* [113,114]. To clarify the relative roles of secreted factors *versus* direct cell-cell contact, indirect-co-culture using cell culture inserts has been tested, identifying that direct cell-cell contact improves results and can even lead to human adipose tissue-derived mesenchymal stem cell (ADSC)-originating spontaneously beating CM-like cells [109]. Future studies should define the cardiogenic differentiation potential of ASC by entirely excluding cell fusion as a mechanism, *e.g. via* specific labeling of the neonatal CM used, or employment of different species. However, if cell fusion is consistently demonstrated to exert a positive influence on myogenic and functional regeneration of the affected tissue, this approach should not be abandoned. The large number of studies using the epigenetic modifier 5-azacytidine highlights the potential for harnessing epigenetic modulation to modify cell fate decisions to drive regeneration. While the outcomes reported are highly variable, expression of specific cardiac markers such as desmin, cardiac actin and Troponin has been demonstrated, in association with Notch signaling. The benefits of MSC pre-conditioning modification such as improved cell survival and proliferation, stimulation of paracrine factor secretion and increased angiogenesis – thereby promoting cardiac repair – have been described in detail elsewhere [115].

4.1.2. Forward programming of CM progenitor cells

Another promising approach is the support of pre-existing precursor cells: cardiomyocyte progenitor cells (CMPC) can be efficiently isolated from fetal hearts, as either $c\text{-}kit^+$ [116] or $Sca\text{-}1^+$ [117,118] cell populations. Overexpression of miR-1 and miR-499 in such cells has been shown to enhance their differentiation into cardiomyocytes *via* repression of histone deacetylase 4 and Sox6 [118], confirmed with overexpression of miR-499 in another CMPC population resulting in similar effects (repression of Sox6 and Rod1) by Hosoda and colleagues [116]. The cardiogenic potential of CMPC can also be efficiently enhanced using the methylation inhibitor 5-azacytidine in combination with TGF-β [117].

4.2. Programming of pluripotent stem cells

Pluripotent stem cells, encompassing embryonic and induced pluripotent stem cells, represent an attractive platform to study key cellular and molecular programs of early heart development. The Human Pluripotent Stem Cell registry (hPSCreg) listed a total of 1281 cell lines in

August 2017 (hESC: 707, hiPSC: 574) (http://hpscreg.eu/). Worldwide, the majority of hESC lines are recorded in the U.S.A., while in Europe the highest numbers are in the U.K.. The number of hiPSC lines is constantly increasing, with a dramatic increase from 120 lines in January 2016.

Since their first successful differentiation towards cardiomyocytic phenotypes was demonstrated in 1991, ESCs have grown to become an invaluable *in vitro* model to study cardiac development [120]. Currently, the number of publications relying on murine and human ESCs is increasing daily — a selected portion of these published over the past five years is shown in Table 2. Importantly, gene expression analysis has revealed a unique profile for each individual hESC line [121,122], which obviously results in variable self-renewal behavior and differentiation preferences [123,124]. Moreover, even the *in vitro* "micro-environment" of each laboratory can exert a strong impact on the cell line's gene expression signature [125].

To overcome ethical concerns as well as the poor accessibility of ESCs, iPSCs have become a major focus of interest since their seminal description a decade ago [126,127]. Representative recent studies using murine and human iPSCs for cardiogenic differentiation are outlined in Table 3.

The starting cell types for iPSC generation are available on a large scale from various easily accessible sources. Furthermore, autologous material enables the production of patient-specific iPSCs which can be expected to become highly relevant for personalized therapy as well as *in vitro* drug testing. The retained epigenetic memory of such cells, demonstrated by the incomplete reprogramming of non-CG methylation as well as differences in CG methylation and histone modifications [128] are important considerations. They represent a drawback on the one hand, by leading to intra-line variability within clones from a single subject, but can also be construed as advantageous with respect to the cells' enhanced ability to differentiate preferentially into their cell type of origin [129,130]. Thus, iPSCs derived from murine neonatal ventricular myocytes display a higher propensity towards spontaneous differentiation into beating CM compared to iPSCs derived from other somatic cells (*e.g.* tail-tip fibroblasts) [131]. In addition, the re-programming strategy towards the iPS cell stage itself has both substantial influence on programming efficiency, and affects the genetic profile of iPSCs themselves, resulting in high line-to-line variability. To date, the application of synthetic modified mRNA has the highest programming efficiency (~4.4%) using the TFs Oct4, Sox2, Klf4, c-Myc, Lin28 in combination with valproic acid [132]. Moreover, another study claimed "foot-print free" non-integrative mRNA-based reprogramming of somatic cells and subsequent effective differentiation towards a CM-like phenotype including sarcomeric marker expression and appropriate specific responses to pharmacological modulation [133]. In this regard, it has been shown that iPS-derived CMs can exhibit residual transgene expression of Oct4 and Nanog after lentiviral-mediated transduction [134], which has the potential to lead to tumor formation.

The multifaceted powerful potential of iPSC-derived cells has helped foster numerous preclinical studies, with the therapeutic effect of their derivatives largely ascribed to their paracrine effects in supporting ischemic tissue [135]. With regard to ESC, at present only a single phase I clinical trial using human ESC-derived $CD15^+Isl1^+$ progenitors for transplantation in severe heart failure is actively recruiting (ESCORT).

Apart from the cellular origin, concepts regarding differentiation of various PSCs are highly dependent on insights gained from study of natural embryonic development, to enable the driving of cell fate alongside time-, space- and signaling-dependent patterns in order to overcome hurdles associated with species-specification and inter-personal variations. The direct application of PSCs as purely undifferentiated cells is unfeasible due to their high teratogenic potential *in vivo* [136,137]. Despite experimental progress with PSC-derived CMs, achieving functional maturation of these cells has so far proved elusive *in vitro* [138]. This is likely to reflect the current infeasibility of precisely mimicking the entire natural microenvironment, including topographical, electrical, adhesive, mechanical, biochemical, and cell–cell interaction cues [139].

In this chapter we will discuss the two main emerging strategies for efficient forward programming towards CM-like cells: i) molecular programming using forced exogenous overexpression of lineage-specific TFs and miRNAs (Subsection 4.2.1); and ii) targeted differentiation *via* provision of optimized culture conditions (Subsection 4.2.2). Furthermore, specific concepts will be considered regarding selection (Subsection 4.2.3) – as well as maturation – (Subsection 4.2.4) strategies to enhance the quality of the final cell product.

4.2.1. Molecular programming

Early molecular programming studies using ESCs have demonstrated MesP1 to be an essential cardiac fate determinant [140–143]. Recent *in vivo* studies have confirmed the positive effect of MesP1 CPCs in promoting cardiovascular repair of murine hearts [144]. Moreover, gain-and-loss-of function experiments have disclosed a major role for miR-322/-503 in a MesP1 progenitor population in regulating early cardiac fate decision by negatively effecting neuroectoderm differentiation [145]. Thus, $MesP1^+$ mesodermal progenitors represent a heterogeneous population bearing a context-dependent potential to differentiate into cardiac, hematopoietic and skeletal myogenic progenitors [146–148], which is also the case for $Flk1^+$ mesodermal cells [149]. Numerous factors affect MesP1 and subsequent cardiac TF expression, such as Bry^+ [141], Cited2 [150], CIBZ (BTB domain-containing zinc finger protein) [151] and Fndc5 (Fibronectin type III domain-containing 5 protein (also known as: peroxisomal protein (PEP)) [152]. Another group of RNAs, the so called long noncoding RNAs (lncRNAs) play a still incompletely understood role during cardiac development; for instance, the lncRNA Braveheart has been demonstrated to act as an epigenetic modulator upstream of MesP1 using multiple ESCs [153].

Forced overexpression of key fate-determining factors has also been applied to the generation of SAN-like cells from ESCs, using the TFs Tbx3 [89], Shox2 [154] and Isl1 [155], thereby directing cell fate towards a pacemaker-like phenotype. This topic is discussed in more detail below (Subsection 4.4).

Data on molecular programming using forced exogenous overexpression of lineage-specific TFs are so far restricted to ESCs — to date; no attempts have been described in the literature for iPSCs.

4.2.2. Targeted differentiation

There are currently numerous direct programming protocols using PSCs which include a focus on continual optimization and refinement of the targeted differentiation process with respect to time- and dose-dependent application of cardiogenic modulators (Fig. 2). This may allow precise manipulation of PSCs through activation or inhibition of diverse implicated molecular pathways. While the protocols schematically presented in Fig. 2 have some similarities in common such as respective pathway activation or inhibition, they differ with respect to specific timeframes and overall duration of cell culture required.

Such approaches start, after preliminary ROCK inhibition [156–160], with initial activation of Wnt signaling using *via* Wnt activators such as CHIR99021 or direct application of Wnt3 in order to induce a mesodermal CPC population [144,156,157,160–162], together with the addition of Activin A and BMPs (bone morphogenetic proteins) [144,158,159,161,163–166]. Substances administered in the second step depend upon the desired lineage specification to either a working myocardial or a conduction system cellular phenotype. CM-like induction requires the inhibition of Wnt signaling *via* Wnt inhibitors [156,159–161], such as IWR1 and IWP2, as well as the addition of various growth factors (GFs), such as FGF (fibroblast GF) and VEGF (vascular endothelial GF) [157,159,161,163,165]. In addition, modulation of MAPK signaling, using SB203580 [133] and PD98059 [167] (both MAPK inhibitors), or Rho-kinase inhibitors (H1152) [158] is utilized. In contrast, differentiation towards SAN-like cells mandates inhibition of GF-, including Activin- and Nodal-signaling using inhibitors such as PD 173074 (FGF

Table 2
Overview of recently published programming strategies of ESCs towards diverse cardiovascular subtypes.

Cell origin	Host	Delivery system	*In vivo/in vitro*	Factor/substances	Target cell type	Special features	Literature
ESC (HES-2, H1, H9)	Human	Specific culture conditions	*In vitro*	Day 0–1: 0.5 ng/mL of BMP-4 Day 1–4: 10 ng/mL BMP-4, 5 ng/mL human bFGF, and 6 ng/mL Activin A Day 4–8: basal medium containing 10 ng/mL VEGF, 150 ng/mL Dkk-1 Day 8-end: basal medium with 10 ng/mL VEGF, 10 ng/mL human bFGF MEF-free and serum-free hESC adherent culture under cGMP and cGLP conditions	CM-like	27% $cTnT^+$ Expression of: sMHC, βMHC, Isl-1, Nkx2-5, MYH6, Tnnt2, Myl2, and Myl7	Chen et al., 2012 [163]
ESC (H7)	Human	Specific culture conditions	*In vitro*	1) MEF and SNL feeder cell layers + conventional SC culture medium containing ko-SR 2) bFGF 3) Matrigel matrix + commercial mTeSR1 medium	CM-like	Most efficient protocol: MEF and SNL feeder cell layers + conventional SC culture medium containing ko-SR Least efficient protocol: Matrigel matrix + commercial mTeSR1 medium; neural lineage induction	Ojala et al., 2012 [190]
ESC (m: J1; h:H1)	Mouse/human	Specific purification method	*In vitro/in vivo*	CM-specific MBs	CM-like	*In vitro*: myosin heavy chain-MB: ~97% $cTnT^+$ cells *In vivo*: improved cardiac function, without tumor formation after 4 weeks	Ban et al., 2013 [174]
ESC	Mouse	Specific culture conditions	*In vitro*	GSK3 inhibitor P38 MAPK inhibitors CaMKII inhibitors ERK activators	CM-like	ERK activators, CaMKII inhibitors: proliferative effects only on CMs in early developmental stage GSK3 inhibitor (BIO, CHIR), ERK activator (5 μM SU1498), CaMKII inhibitor (5 μM KN93): induced cell cycle progression in CM, resulting in CM proliferation	Uosaki et al., 2013 [168]
ESC	Mouse	Specific culture conditions	*In vitro*	44 cytokines/signaling molecules on day 3 of diff	CPC ($Nkx2\text{-}5^+$)	IGF1, IGF2, insulin, Wnt3a: significantly increase CPC formation IGF, insulin: promote Bry^+ mesodermal cell proliferation Activin A, BMP2 or BMP4: decrease CPC formation	Engels et al., 2014 [162]
ESC	Mouse	Specific culture conditions	*In vitro*	2% O_2 preconditioning (3 passages) of ESCs Diff as EBs in 20% O_2	CM-like	Significant increased expression of early differentiation markers FGF5, Eomes Increased gene expression of Eomes, Goosecoid, Bry, AFP, Sox17, FoxA2, and protein expression of Bry, Eomes, Sox17, FoxA2 — diff into mesodermal and endodermal lineages	Fynes et al., 2014 [185]
ESC	Human	Specific culture conditions	*In vitro*	2% O_2 preconditioning (3 passages) of ESCs Diff as EBs in 20% O_2	CM-like	Decreased expression of early differentiation markers FGF5, Eomes increased gene expression Nestin, β3-tubulin — diff into ectodermal lineage	Fynes et al., 2014 [185]
ESC (H9)	Human	Specific substrate mechanics	*In vitro*	1) TCPS 2) PA hydrogel substrate	CM-like	Intermediate stiffness of PA hydrogel yielded slightly higher $cTnT^+$ cells without significant difference to TCPS	Hazeltine et al., 2014 [188]
ESC (GSES)	Mouse	Dna-Plasmid specific purification method	*In vitro*	Tbx3 Myh6-promoter-based antibiotic selection	SAN-like	>80% physiologically and pharmacologically functional pacemaker cells with highly increased beating rates (300–400 bpm)	Jung et al., 2014 [89] Rimmbach et al., 2015 [90]
ESC	Mouse	Lentiviral	*In vitro*	Fndc5	CM-like	Sign. expression of: Flk1, Isl1, Nkx2-5, Gata4, Mef2C, α-MHC, cTnT, α-actinin, SM22α, α-SMA	Rabiee et al., 2014 [152]
ESC (H7, ESI-017)	Human	Specific culture and substrate conditions	*In vitro*	Matrix-free, scalable, and GMP-compliant process Culture: including first CHIR and second IWP-4 induction	CM-like	>90% pure CM; 1.5 to 2 × 10^9 CM/L (up to 1 L spinner flasks) ESI-017: 6 μM CHIR — ~91% $cTnT^+$ H7: 12 μM CHIR — ~92% $cTnT^+$	Chen et al., 2015 [157]
ESC	Mouse	Lentiviral	*In vitro/in vivo*	Nkx2-5 Isl1	CM-like SAN-like	Overexpression of Nkx2-5: inhibition of Isl1 expression Overexpression of Isl1: enhanced specification of cardiac progenitors, earlier cardiac differentiation, and increased cardiomyocyte number, upregulation of nodal-specific genes (*e.g.* Hcn4), downregulation of transcripts of working myocardium	Dorn et al., 2015 [155]
ESC (R1)	Mouse	Adenoviral	*In vitro*	Shox2	SAN-like	Increase in Cx45, decrease in Cx43, Nkx2-5 SHOX2-EBs beat spontaneously (83 ± 7% *versus* 15 ± 6%) Pacemaker-like AP profile (62%)	Ionta et al., 2015 [154]
ESC (CGR8, αPIG44)	Mouse	Specific culture conditions	*In vitro*	100 μM AA	CM-like	AA application from day 0 to 2 increases cardiogenesis 2–4-fold Day 5: increased expression of genes associated with angiogenesis, blood vessel development, hematopoiesis/erythropoiesis, Bry, Mef2c, Myl7	Ivanyuk et al., 2015 [169]
ESC (H7)	Human	Specific purification method	*In vitro*	CM-specific MBs	CM-like	NPPA-MB: ~92% $\alpha\text{-actinin}^+$ cells	Jha et al., 2015 [175]
ESC (α-PIG)	Mouse	Specific culture and substrate conditions	*In vitro*	0.3 mm-thick hcECM	CM-like	hcECM supported proliferation Significantly increased expression of: Myh6, Tnnt2, Nkx2-5 Matrigel or Geltrex use did not induce cardiac-specific markers	Oberwallner et al., 2015 [119]

ESC (derived from C57BL/6 mouse strain)	Mouse	Specific culture conditions	*In vitro*	10 μM resveratrol	CM-like	Promotes mESC differentiation towards CM Enhanced beating properties of EBs Significantly higher expression of: Nkx2-5, Mef2c, Tbx5, dHand2, αMHC, Cx43, cTnC1	Ding et al., 2016 [202]
ESC	Mouse	DNA-Plasmid	*In vitro*	CBIZ	CM-like	CBIZ depletion: induced expression of Bry, Mesp1, Gata6, Sox17 CBIZ overexpression: decreased expression of Bry, Mesp1, Flk1 Isl1, Gata4, Mlc., cTnI:	Kostka et al., 2016 [151]
ESC (lin–house and E14tg2a)	Mouse	Specific culture conditions	*In vitro*	Monolayer culture without feeder cells Day 0–1: IMDM/Ham's F12, N2 supplement B27 supplement, 0.5 mM AA, 4.5×10^{-4} M MTG Day 1–3: 8 ng/mL Activin A, 0.5 ng mL BMP4, 5 ng/mL hVEGF day 3–13: Stemfro-34 SF medium, 0.5 mM AA, 5 ng/mL hVEGF, 10 ng/mL bFGF, 50 ng/mL hFGF10	CM-like	Significant suppression of beating EBs 60% CM 35–40% cTnT$^+$/MF20$^+$ E14tg2a : more efficient CM-like cell yield with 5 ng/mL Activin A	Kokkinopoulos et al., 2016 [165]
ESC	Mouse	Specific culture conditions	*In vitro/In vivo*	25 ng/mL Activin, 20 ng/mL BMP2, 20 ng/mL BMP4, 100 ng/mL DLL1, 10 ng/mL bFGF, 10 ng/mL FGF8, 20 ng/mL Tgfb, 100 ng/mL Wnt3a, 5 μM IWR1, 5 μM SB431542	MesP1-CPC	BMP4 exposure: day 0–4 highest improvement of MesP1$^+$ cells (29.6%) BMP4 + IWR1: Mes1$^+$ cells (13.8%), however differentiate more efficiently into cardiac myocytes *In vivo*: injection of day 5 MesP1–CPCs led to improved survival of MI mice and decreased scar formation	Liu et al., 2016 [146]
ESC (E14T)	Mouse	Specific purification method	*In vitro/In vivo*	Selection based on VE-cadherin promoter	CEDP	Differentiation into: ~47% cTnT$^+$ and ~28% VE-cadherin$^+$ cells	Mahabe et al., 2016 [173]
ESC (H7 and H9)	Human	Specific culture conditions	*In vitro*	1) Day 0–1.5: BMP4, FGF2, Activin A, LY294002 Day 1.5–5: BMP4, FGF2, RA, IWR1/IWP2 Day 5–9: BMP4, FGF2 2) Day 0–2: CHIR99021, day 2–4: IWP2	CM-like	WNT3, WNT8: regulation of Bry expression and mesoderm induction (via FZD7 + canonical Wnt signaling) WNT5A/5B: regulation of MesP1 expression and cardiovascular development (via ROR2 + noncanonical Wnt signaling) WNT2, WNT5A/5B, WNT11: regulation of late functional CM diff (via FZD4, FZD5 + noncanonical Wnt signaling)	Mazzotta et al., 2016 [161]
ESC (5)	Mouse	DNA-Plasmid	*In vitro*	CITED2	CM-like	Cited2 depletion: significantly decreased expression of Brachyury, Mesp1, Isl1, Gata4, Tbx5 Cited2 overexpression: stimulation of Brachyury, Mesp1, Isl1, Gata4, Tbx5, MyH6, cTnT; protein interaction with Isl1	Pacheco-Leyva et al., 2016 [150]
ESC	Mouse	miRNA lentiviral	*In vitro*	miR-322/-503	CM-like	Highest enriched miRNA in Mesp1 lineage (miR-322/-503) Cdc191 is a direct target of miR-322/-503	Shen et al., 2016 [145]
ESC	Human	Specific culture conditions	*In vitro*	Monolayer-directed differentiation protocol Different concentration s of Activin A and BMP4	CM-like EC-like	miR-322/-503 selectively inhibits neuroectoderm differentiation Generation of distinct CVP populations following derivation of cardiogenic versus hemogenic mesoderm	Palpant et al., 2017 [164]
ESC (HES3-Nkx2-5$^{eGFP/w}$, HES2)	Human	Specific culture conditions	*In vitro/In vivo*	EB formation, 5% O_2 (d0–12) Generation of VLCMs: Day 0–3: 10 ng/mL rhBMP4, 6 ng/mL rhActivinA, 5 ng/mL rhbFGF Day 3–5: 0.5 μM IWP2, 10 ng/mL rhVEGF Day 5–12: 5 ng/mL rhVEGF Day 12–20: without additional factors Day 20: FACS sorting (NKX2-5:GFP$^+$ SIRPA$^+$CD90$^-$) Generation of SANLPCs: Day 0–3: 3 ng/mL rhBMP4, 2 ng/mL rhActivinA, 5 ng/mL rhbFGF Day 3–6: 2.5 ng/mL rhBMP4, 5.4 μM SB-431542, 0.25 μM Retinoic Acid (HES3: 480–960 nM PD 173074 at day 4, HES2 at day 3) Day 6–20: 5 ng/mL rhVEGF Day 20: FACS sorting (NKX2-5:GFP$^-$ SIRPA$^+$CD90$^-$)	CM-like SAN-like	FGF pathway blocks the development of NKX2-5$^+$ CM Earlier expression of the SAN lineage (TBX18, SHOX2, TBX3), typical pacemaker action potentials (90%), ion current profiles and chronotropic response	Protze et al., 2017 [159]

AA: ascorbic acid; βMHC: β myosin heavy chain; bFGF: basic fibroblast growth factor; BMP: bone morphogenetic protein; bpm: beats per minute; Bry: Brachyury; CAMKII: Ca^{2+}/calmodulin-dependent protein kinase II; CEDP: cardiac and endothelial dual-progenitor population; CHIR: CHIR99021, Wnt activator; CIBZ: BTB domain-containing zinc finger protein; CM: cardiomyocytes; CPC: cardiac progenitor cell; cTnC1: cardiac troponin C1; CVP: cardiovascular progenitor cell; diff: differentiation; EB: embryoid bodies ; EC: endothelial cell; ERK: extracellular signal-regulated kinase; ESC: embryonic stem cells; FndC5: fibronectin type III domain-containing 5 protein (also known as: peroxisomal protein (PEP)); GSK3: glycogen synthase kinase-3; hcECM: human cardiac extracellular matrix; IWP: WNT inhibitor; ko-SR: knockout serum replacement; LY294002: phosphoinositide 3-kinase inhibitor; MAPK: p38 mitogen-activated protein kinase; MB: molecular beacon; MTG: monothioglycerol; PA: polyacrylamide; PD 173074: FGF signaling inhibitor; SAN: sinoatrial node; SB-431542: Activin/Nodal/(TGFβ) signaling inhibitor; αMHC: sarcomeric myosin heavy chain; TCPS: tissue culture polystyrene; Tnnt2: cardiac troponin T2; VEGF: vascular endothelial growth factor; (X): number of cell lines used.

iPSC (from activated T cells, hPBMC)	Human	Retroviral Specific culture conditions Metabolic selection (lactate)	*In vitro*	Day 1–13: 5% O_2 and 7% CO_2 Day 1–2: 50% mTeSR, 45% DMEM (low glucose), 5% FBS, 1 μM H1152, 100 ng/mL bFGF Day 2: 2/3 medium change to: DMEM (low glucose), 10% FBS, 50 ng/mL bFGF Day 3–7: DMEM (low glucose), 10% FBS, 50 ng/mL bFGF, 6 ng/mL Activin A, 10 ng/mL BMP4 Day 8–13: DMEM (low glucose), 10% FBS Day 14–end: 20% O_2 and 7% CO_2 Day 16–18: DMEM (no glucose), 1% sodium DL-lactate solution	CM-like	Before selection: 30–45% CM, expression of cTnT and MLC2v After lactate selection: 90% CM, electrophysiological characterization: ventricular, atrial and nodal-like action potentials	Fuerstenau-Sharp et al., 2015 [158]
iPSC	Mouse	Specific culture conditions	*In vitro/in vivo*	αMHC-GFP EB formation Diff. after EB: 1 week treatment with 1% DMSO and/or 100 ng/mL NRG-1β	CM-like	Significantly higher expression of: Gata4, Gata6, αMhc, Myhl7, Myl3, cTnnc1, Ryr2, Serca2a Ventricular-like AP Preserves cardiac function in mouse model	Iglesias-García et al., 2015 [207]
iPSC (TiB7.4, AT25)	Mouse	Specific culture conditions	*In vitro*	100 μM AA	CM-like	AA application from day 0 to 2 increases cardiogenesis 2–4-fold	Ivanyuk et al., 2015 [169]
iPSC (DF19-9–11T.H)	Human	RNA	*In vitro*	EB-formation Diff after EB: transfection with CIR RNA	CM-like	Spindle-shaped cells with characteristic Expression of: cTnT, tropomyosin, α-actinin	Kochegarov et al., 2015 [208]
iPSC (AT25)	Mouse	Specific culture and substrate conditions	*In vitro*	0.3 mm-thick hcECM	CM-like	hcECM supported the proliferation Significantly increased expression of: Myh6, Tnnt2, Nkx2-5	Oberwallner et al., 2015 [119]
iPSC	Mouse	Lentiviral	*In vitro*	Telomerase-competent cell lines (TRF1 expression) AA	CM-like	Matrigel, Geltrex: did not induce cardiac-specific markers $iPSC^{highTRF1}$: differentiate earlier and more efficiently into CMs $iPSC^{lowTRF1}$: differentiate very efficiently to the ectoderm lineage AA: increase CM yield with $iPSC^{highTRF1}$	Aguado et al., 2016 [170]
iPSC (Duanqing Pei, Chinese Academy of Sciences)	Mouse	Specific selection method	*In vitro*	0.001–0.2 μmol/L EPI 10 μmol/L phentolamine 5 μmol/L propranolol 10 μmol/L phenylephrine 10 μmol/L clonidine 1 μmol/L isoproterenol 10 μmol/L PD98059	CM-like	EPI and activation of α1-AR: enhancement of CM differentiation *via* MEK-ERK1/2 signaling	Li et al., 2016 [167]
iPSC (from: MEFs)	Mouse	Retroviral Specific culture conditions	*In vitro*	EB formation Medium containing: DMEM, 15% fetal bovine serum, 0.2 mmol/L L-glutamine, 0.1 mmol/L nonessential amino acids, and 0.1 mmol/L β-mercaptoethanol	CM-like	After EB formation: expression of Flk1, α-actinin, α-MHC, cTnT, Cx43, Nkx2-5 Active beating: 360 bpm Incompletely differentiated iPS cells: teratoma formation after transplantation into a SCID mouse model of MI	Wang et al., 2016 [209]
iPSC (from foreskin tissue)	Human	Specific culture conditions Metabolic selection (lactate)	*In vitro*	Chemical-defined and albumin-free medium Day 0–1: S12 without insulin medium, 4–8 μM CHIR Day 1–3: S12 without insulin medium Day 3–5: S12 without insulin medium, 5 μM IWR-1 Day 5-end: S12 with insulin medium, atrial diff: 1 μM RA, ventricular diff: 1 μM BMS493	CM-like Atrial-like	Highly homogenous atrial and ventricular Myocytes in a scalable fashion with normal electrophysiological properties	Pei et al., 2017 [160]

AA: ascorbic acid; AP: action potential; AR: adrenergic receptor; BCT: bioartificial cardiac tissue; bpm: beats per minute; BMS493: RA inhibitor; Bry: Brachyury; CHIR: CHIR99021 (GSK-3 inhibitor, Wnt activator); CIR: specific fetal cardiac-inducing RNA (fragment of *N*-sulfoglucosaminesulfohydrolase and the caspase recruitment domain family member 14 precursor); Cx43: connexin 43; clonidine: selective α2-AR agonist; cTnT: cardiac troponin T; DMSO: dimethyl sulfoxide; EPI: epinephrine; hcECM: human cardiac extracellular matrix; hPBMC: human peripheral blood mononuclear cells; H1152: Rho-kinase (ROCK) inhibitor; IRX4: iroquois-class homeodomain protein IRX-4; isoproterenol: β-AR agonist; IWP-4: WNT inhibitor; IWR-1: WNT inhibitor; MEF: mouse embryonic fibroblast; MLC2a: myosin regulatory light chain 2, atrial isoform; MLC2v: myosin regulatory light chain 2, ventricular/cardiac muscle isoform; NRG-1β: Neuregulin-1β; PC: progenitor cells; PD98059: MEK1/2 inhibitor; MI: myocardial infarction; phenylephrine: selective α1-AR agonist; phentolamine: α-AR antagonist; propranolol: β-AR antagonist; RA: retinoic acid; SB203580: MAPK signaling inhibitor; SCID: severe combined immunodeficiency; TNNI3: cardiac troponin I; Tnnt2: cardiac troponin T2; TRF1: shelterin-complex protein; T3: Tri-iodo-L-thyronine.

signaling inhibitor) and SB-431542 (Activin/Nodal/TGFβ signaling inhibitor) [159].

The final cellular product varies among protocols with respect to marker gene expression, however, each approach results in expression profiles partly specific for CM-like or SAN-like cells, including a reported >90% cTnT$^+$ cells [157], >90% CD31$^+$/VE-cadherin$^+$ ECs [164] or 35–40% double positive cTnT$^+$/MF20$^+$ cells [165]. In addition, targeted differentiation-derived pacemaker-like cells exhibit specific action potential profiles [159]. Enhanced proliferation (up to 14-fold) of ESC-derived CM-like cells has been obtained through the addition of specific GSK3 and CaMKII inhibitors, as well as ERK activators [168]. Furthermore, epigenetic modulators have been used such as ascorbic acid [169–171], which, when applied during a specific time-frame from day 0 to day 2 of differentiation, leads to a 2–4-fold increase in cardiogenesis [169].

Such *in vitro* differentiation protocols approximate the highly sensitive and finely-tuned interplay of signaling pathways required for healthy embryonic development. Their success is critically dependent upon careful, albeit protracted, step-by-step protocol optimization.

4.2.3. Selection-strategies

Notwithstanding the existence of *in vitro* approaches described above, purity of PSC-CMs is still a major issue, with the choice of suitable and highly reliable surface marker for use in standard flow cytometry or magnetic isolation procedures for differentiating CMs still under debate. Accordingly, purification strategies based on cardiac specific intra-cellular marker expression remain an important approach.

Stable PSC lines containing: i) αMHC (99% MHC$^+$ cells) [171,172] or VE-cadherin (47% cTnT$^+$ cells) [173] promotor-linked antibiotic resistance genes have been used to efficiently select αMHC$^+$ cells or a cardiac and endothelial dual-progenitor population; ii) αMHC promotor-linked green fluorescence protein (GFP) gene [168], enabling GFP$^+$ flow cytometric selection. Both methods have the disadvantage of requiring stable exogenous DNA introduction. iii) Another approach, applied by Bao's group, enables the purification of CMs through specific molecular beacons (MBs) which target mRNAs, such as MBs targeting myosin heavy chain 6/7 (97% cTnT$^+$ cells) [174] or NPPA (92% α-actinin$^+$ cells) [175] mRNA. Additional approaches used include the selection of CM-like cells *via*: iv) mitochondria-specific fluorescent dyes (99% α-actinin$^+$ cells) [176]; or v) antibodies against partially cardiomyocyte-specific markers, *e.g.* SIRPA (signal-reduced protein alpha) (98% cTnT$^+$ cells) [177], EMILIN2 (elastin microfibril interface 2) (no qualitative statement about α-actinin$^+$ or cTnT$^+$ cell yield) [178], or VCAM1 (vascular cell adhesion molecule 1) (95% cTnT$^+$ cells) [179]. A consequence of the increasing mitochondrial-to-cell volume ratio during CM development is the metabolic substrate shift from glucose and lactate in early developmental stages to the primary reliance on fatty acid oxidation characterizing adult mature CMs [180]. These changes have been applied in non-genetic metabolic purification strategies (98% α-actinin$^+$ cells) [158,160,181,182]. Initial approaches used glucose-deplete/lactate-enriched media, achieving quite homogenous populations and increased purity [158,160]. In this regard, Kuppusamy et al. demonstrated a positive impact on cardiac maturation exerted by the let-7 family of microRNAs which are associated with metabolic energetics in maturing CM [183]. Overall, at present, other than standard promotor-linked selection approaches, the remaining novel selection strategies still require further evidence of reproducibility in independent laboratories using different PS cell lines.

An important consideration is that current PSC-derived CMs represent a mixture of nodal-, atrial-, ventricular-like and early-intermediate immature phenotypes, as evident from electrophysiological and pharmacological studies [89,134]. Transplantation of such CM mixtures could induce significant arrhythmia [184], thus further purification will be an indispensable prerequisite for clinical translation.

4.2.4. Maturation-strategies

In general, all published reports have – to a greater or lesser extent – generated an immature and physiologically incomplete CM phenotype, reflecting the heterogeneity of PSCs used as well as the distinction between human and murine cell lines [138]. A variety of substrates and protocol modifications are under close scrutiny with the goal of obtaining improvements in cardiogenic differentiation during culture to enable safe bench-to-bedside implementation.

Differential effects can arise from the cultivation procedure itself, including an important influence of cultured temperature and distinctions between use of 2D monolayer or 3D EB formation. Low oxygen preconditioning (2% O_2) can impact on lineage commitment [185], with murine ESCs displaying significantly increased expression of the early differentiation markers FGF5 and Eomes consistent with preferred differentiation towards mesodermal and endodermal lineages. In contrast, culturing human PSCs under low oxygen tension prior to spontaneous differentiation in EBs primes commitment to an ectodermal lineage, indicated by significant induction of β3-tubulin and Nestin [185].

An important but contentious issue is the use of either diverse co-culture systems or an adjusted material surface. PSC-derived CMs cultured on fibronectin-coated micro-grooved polydimethylsiloxane (PDMS) scaffolds exhibit a more organized sarcomeric structure, together with a more homogenous alignment and improved sarcoplasmic reticulum-based Ca^{2+} cycling [186]. Angelo's group introduced a biomimetic aligned nanofibrous cardiac patch which resembles the extracellular matrix of decellularized myocardium from rats [187]. The material used, namely polylactide-co-glycolide (PLGA), is an FDA approved therapeutic device due to its biodegradability and biocompatibility and has great potential to form the basis of an implantable cardiac patch. Moreover, this anisotropic environment additionally results in symmetric alignment of iPSC-derived CMs. ESC differentiation on intermediate stiffness polyacrylate (PA) hydrogel substrate resulted in only a slight enhancement of differentiation in comparison to common (rigid) polystyrene tissue culture [188]. Promising approaches using matrix-free, GMP-compliant culture protocols have yielded 94% cTnT$^+$ cells [157], which may facilitate advance towards clinical use.

hESCs co-cultured with AKT-activated endothelial cells led to an improvement in Nkx2-5$^+$ cell yield as well as faster beating frequencies compared to hESCs cultured on Matrigel alone [189]. These findings concur with observations of other groups exploiting the benefits of co-culture [119,171,190]. Co-culture with MEF or SNL feeder cells has yielded better results than Matrigel for the majority of cell lines investigated [190]. Other approaches rely on matrix-cell-composites, such as BCTs (bioartificial cardiac tissue: cells plus liquid collagen type I plus Matrigel) [171], or the application of cardiac extracellular matrix [119]. In order to mimic endogenous tissue with blood capillary networks, Akashi's group have developed a vascularized 3D-iPSC-CM tissue, which may provide more comprehensive data in the field of drug screening [191]. Another technique relies on the co-culture of human cardiac microvascular endothelial cells (hCMVECs) and hMSCs in combination with induced pluripotent stem cell-derived embryonic cardiac myocytes (hiPSC-ECMs) which also aims at generating *in vitro* vascularized cardiac tissue scaffolds [192]. The considerable potential importance of culturing hiPSC-CMs as human 3D heart tissues to overcome species-dependent discrepancies of CM behavior has also been highlighted, particularly with respect to preclinical drug screening [193].

To ensure valid comparisons between techniques, a standard characterization procedure is essential. However, to date the expression of a multitude of TF or surface markers have been interrogated at quite diverse time points. However, consistency with respect to both time and marker information is very important, substantiated by recently published data revealing time-dependent morphological and electrophysiological alterations of iPSC-derived CMs [194]. Reliance on 2D morphological analysis alone to determine CM growth and maturation has been suggested to be insufficient, with clear volume differences

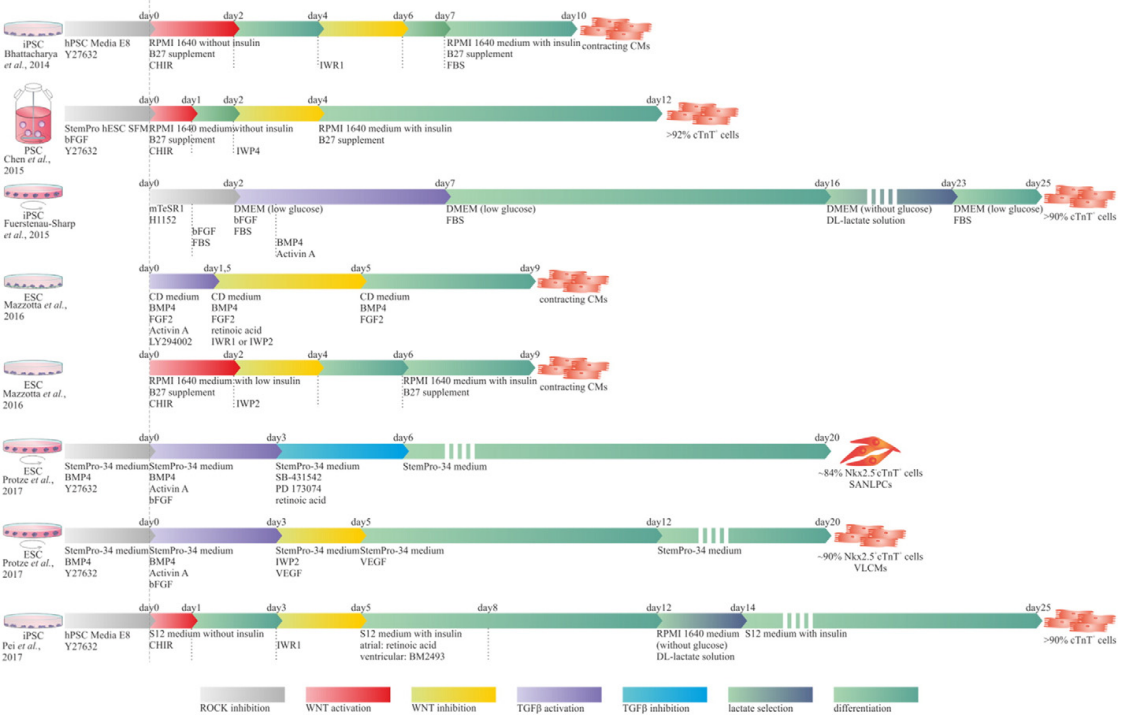


Fig. 2. Schematic timescale for targeted differentiation of human PSCs. This cartoon summarizes various strategies and protocols of pluripotent stem cell (ESC, iPSC) differentiation towards cardiomyocytic cells, relying on time- and dose-dependent application of cardiogenic modulators without any nucleic acid introduction or factor overexpression.

apparent with hypertrophic stimulation or long-term culture using 3D z-stack myofilament analysis [195]. In addition, the success of engraftment will be dependent on the state of the applied CMs, as has been demonstrated using an ischemic heart model in immuno-deficient mice [196].

To minimize risk of off-target effects attributable to incomplete knowledge of iPSC-derived CM behavior, their response to pharmacological manipulation is of central importance. Measurement on Multi-Electrode-Arrays (MEAs) are increasingly used for cell characterization and drug screening [197–199]. The use of voltage-sensitive fluorescent dyes (VSDs), such as di-4-ANEPPS, will also likely facilitate the investigation of action potential characteristics as has been demonstrated for commercially available hiPSC-derived cardiomyocytes ("iCells" and "Cor.4U") post-substance administration [200]. Monitoring of iPSC-CMs under conditions simulating ischemia is a prerequisite to correctly evaluate the impact of hypoxia and nutrient deprivation on future cell replacement therapies. In this regard, Brodarac et al. demonstrated poorer tolerance of murine iPSC-CMs to hypoxia and nutrient deprivation compared to neonatal murine cardiomyocytes, with a significantly higher proportion of poly-caspase-active, 7-aminoactinomycin D-positive and TUNEL-positive cells [201].

In summary, much remains to be done before safe cardiac cell replacement therapy based on PSC strategies becomes a clinical reality, in particular the introduction of GMP-compliant standards underlying precise DNA-, viral- and xeno-free protocols for the generation of mature functional CMs.

4.3. Direct reprogramming of somatic cells

Another ambitious approach to cell replacement therapy aims to avoid the relatively uncontainable pluripotent state, instead choosing to elicit a cell fate switch through the direct conversion of terminally differentiated somatic cells towards mature cell types of interest representing the same germ layer, or even across germ layers. Such an approach has a number of prerequisites for efficient and feasible reprogramming, including epigenetic modulation and lineage-specific intervention.

As early as 1987, murine fibroblasts were successfully converted into skeletal muscle cells using only a single key TF, namely MyoD [210]. Subsequently, numerous publications have reported direct conversion – with one or a number of TFs combined – into several somatic cell types, such as insulin$^+$ β-cells, hepatocytes, osteoblasts, hematopoietic lineage cells, neurons and cardiomyocytes (summarized in [211]). However, the quality of the cells obtained was highly variable. While, thus far, no single master regulator has been discovered to efficiently induce the switch of fully differentiated somatic cells towards a mature induced cardiomyocyte (iCM) lineage, promising results in the field of direct reprogramming have been made and will be discussed in this section.

The concept of patient-specific lineage-conversion holds enormous potential for future clinical applications including: i) elucidating individual disease pathogenesis; ii) lower risk of tumorigenesis and inflammation after cell transplantation compared with PSCs; iii) avoiding the need for transplantation through the possibility to directly convert resident cells; and iv) avoidance of ethical concerns regarding cellular source. Despite the many advantages, the actual consequences of massive fibroblast-to-myocyte programming *in situ* remain unknown and may potentially be detrimental to cardiac function [212].Contemporary research efforts focus on readily available murine cell types, with murine neonatal cardiac fibroblasts (CFs) currently the most efficient somatic cell source for direct reprogramming based on their heterogeneity and plasticity, as well as their resistance to the hypoxic environment of the injured myocardium [213–215]. Another important

source are murine embryonic fibroblasts [216–220]. However, neither of these two cell types are readily accessible from humans with few results so far reported using such human cell types [221,222]. In contrast to iPSC strategies, a cell fate switch is largely achieved through forced exogenous overexpression of lineage-specific TFs, with Gata4, Tbx5 and Mef2c the most frequently used TFs in combination [219, 223–229], or at least a critical part of a more complex composition [216,217,220,221,230,231] (Table 4). However, the reported marker gene expression patterns obtained vary widely between different laboratories and starting material; *e.g.* 30% [227] or 35% [223] cTnT$^+$ cells, 10–15% iCM [224], 3% [225] or 20% [229] αMHC$^+$ cells. Moreover, the cells display only marginal similarity to mature CM based on their molecular and electrophysiological phenotype [223]. Such low efficiencies may reflect insufficient construct design with respect to expression stoichiometry, with a tailored ratio of protein expression – constituting higher protein levels of Mef2c in combination with lower levels of Gata4 and Tbx5 – shown to enhance programming efficiency [228]. Several approaches now aim to improve reprogramming efficiency as well as maturation of iCMs by adding further TFs of the cardiac lineage such as Hand2 [216,217,219,220,222], together with signaling modulators such as inhibitors of TGF-β (A83-01, SB431542) [217,230], WNT (XAV939) [230] or ROCK (SR-3677, Thiazovivin, Y-27632) [217].

A number of pro-cardiogenic microRNAs have been identified [232–234] and applied in combination with TFs [217,222] or as sole modulators [235–237] for direct re-programming. MicroRNAs, for example microRNA-1, can interact with myogenic TFs such as SRF (serum response factor), Mef2c, MyoD or Nkx2-5 in a regulatory loop as repressors and cooperators [238–240]. MicroRNA-1 negatively impacts the Notch signaling pathway *via* direct repression of Dll1 [241] and its downstream factor Hes1 [102], resulting in expression of Gata4, Nkx2-5 and Myogenin. A combination of miR-1/-122/-208/499 and JAK inhibitor I has been demonstrated to induce a cell fate switch towards a cardiac-like phenotype with 28% αMHC$^+$ cells *in vitro* and improved cardiac outcomes *in vivo* [235–237].

One approach focuses on the use of a chemical cocktail to convert murine embryonic fibroblasts to iCMs with spindle, rod and round shaped morphologies [218]. The cells generated manifest action potentials of atrial- and ventricular-like cells. Notwithstanding the heterogeneity of the cells obtained, this method offers a potential alternative to genome integrative methods which may prove safer. However, the effective time windows for each chemical modulator, which include signaling pathway activators/inhibitors or epigenetic regulators, has to be clearly defined to achieve optimal results [220].

While progress has been made, so far the iCMs obtained are still immature in phenotype and inhomogeneous as a population, lacking the terminal structural and electrophysiological characteristics of authentic adult CMs.

4.4. Programming of cardiac conduction system cells

As outlined earlier, the availability of highly specific cardiomyocyte subtypes is critical for future tailored cell therapy of cardiovascular disease. The treatment of rhythm disorders by cardiac cell therapy will require a highly pure population of cells belonging to the cardiac conduction system [89,244–247]. As a corollary, while a mixture of cardiomyocyte subtypes is capable of coupling to working myocardium and setting the pace *in vivo* in a porcine model of complete atrioventricular block, the resulting rhythms are neither stable nor reliably exceed junctional escape rhythm rates [248].

4.4.1. Composition of the cardiac conduction system

Automaticity of the heart beat is crucial for life: the heart's regular contractile activity results from electrical impulses initiated and conducted by highly specialized cells within the heart which form the cardiac conduction system (CCS) [249–251].

The initial impulse is generated in a small number (~10,000) of highly specialized pacemaker myocytes which form the sinoatrial node (SAN) [242]. The SAN is located at the junction of the right atrium and the superior vena cava [249,250]. SAN cells differ from working myocardial cells in their content of ion channels and gap junction proteins [252, 253]. In particular they are rich in hyperpolarization-activated cyclic nucleotide-gated cation channel 4 (HCN4) and t-type calcium channel ($Ca_v3.1$). SAN cells lack natriuretic factor (ANF) and the gap junction proteins connexin43 (Cx43, predominant in ventricular and atrial cells) and connexin40 (Cx40, expressed in atrial working myocardium) [249,252,253]. In contrast, the main gap junction protein in the SAN is connexin45 (Cx45). The membrane voltage clock (cyclic activation and deactivation of membrane ion channels, resulting in part from f-channel conductance of the "funny current") and the subsarcolemmal Ca^{2+} clock (resulting from rhythmic spontaneous sarcoplasmic reticulum Ca^{2+} release) function synergistically to generate SAN automaticity [251]. An illustration of the major differences between these cells is that the upstroke of the action potential (AP) in SAN results from Ca^{2+}-channels as opposed to voltage-gated Na^+-channels typical of working myocardial cells [253].

The impulse that emerges then propagates rapidly and anisotropically through atrial myocardium until it reaches the atrioventricular node (AVN) where the signal is delayed. Atrial and ventricular myocardium are electrically isolated from each other, allowing the atria to contract first, a physiological requirement for optimal ventricular filling prior to ventricular systole [249,250].

The electrical impulse then proceeds through the ventricular conduction system (VCS). The latter consists of a fast-conducting atrioventricular bundle (AVB) (or His bundle), left and right bundle branches (BBs) and the Purkinje fiber network. The AVB commences at the AVN and proceeds through the ventricular septum, where it subdivides into the right and left BBs. The BBs themselves divide into Purkinje fibers, which are spread over the entirety of the left and right ventricles, enabling synchronous biventricular contraction [249,250,254].

While AVN cells resemble those of the SAN, cells of the VCS differ significantly from both SAN and ventricular working myocardial cells. In particular, while depolarization involves voltage-gated Na^+-channels, the "funny current" is still present in the VCS [253]. Similarly, the gap junction proteins are different, with Cx40 required in VCS cells to facilitate high conduction velocity from Purkinje fibers to ventricular cells [252].

Overall, a high degree of coordination is required between all these different cell types to ensure normal stability of cardiac rhythm and rate. One manifestation of a primary disturbance in cardiac pacemaker function is the so called "sick sinus syndrome" (SSS) [255].

4.4.2. Manifestations of sick sinus syndrome

SSS, or sinoatrial disease, refers to a chronic clinical syndrome reflecting SAN dysfunction which has a variety of causes [255]. It results in a variety of abnormalities, including sinus bradycardia, sinus pauses, sinus arrest and sinoatrial exit block and may result in chronotropic incompetence, *i.e.* insufficient augmentation of heart rate to meet physiological requirements during exercise or other stress [256]. In up to half of cases, SSS may be accompanied by paroxysmal atrial tachycardia and AVN conduction disturbance, as part of the tachycardia-bradycardia syndrome [255–259]. The rhythm disturbances can result in a variety of symptoms, including palpitations, lightheadedness, shortness of breath, exercise intolerance, fatigue and frank syncope [255].

While SSS occurs predominantly in the elderly, it is prevalent at all ages [255,257–259]. In young adults and children, SSS commonly results from post-operative atrial trauma or inherited disease. In the elderly, SSS may reflect progressive age-related attrition in SAN cell number [258], or result from a distinct disease process such as atherosclerosis. The sinus node is supported by the right coronary artery whose compromise transiently with ischaemia or through actual infarction can result in permanent SAN dysfunction [255,259]. Familial SSS

has been described to arise from genetic mutations, characteristically originating from alterations in one of three genes [255]: two of these, HCN4 and SCN5A (sodium channel, voltage-gated, type V, alpha subunit), are crucial for transmembrane ion exchange and thus highly relevant for action potential generation. The third gene implicated, MYH6 (myosin, heavy chain 6, cardiac muscle, alpha) plays an essential role in myosin formation to support cardiomyocyte contractility [255].

At present, symptomatic SSS is treated by implantation of an electrical pacemaker device, regardless of underlying etiology, and constitutes one of the major indications for permanent pacemaker implantation globally (30–50% of all cases) [255,257–260].

While the advent of implantable permanent pacemakers has revolutionized the management of life-threatening or highly symptomatic bradycardias, dramatically improving symptoms, quality of life and, in specific cases, prolonging survival [261] it bears some limitations. These can include the possibility of device infection (which can necessitate removal of the entire pacing system), limited battery lifespan (necessitating intermittent generator changes in those who are pacemaker dependent at specific intervals), lead damage or vessel thrombosis, dyssynchronous electromechanical activation and incomplete recapitulation of physiological heart rate increments and interference with external electromagnetic devices [262]. While many of these are theoretical or of small consequence at the individual patient level, their occurrence (particularly device infection) can be serious and can pose a relevant cumulative risk in certain groups, *e.g.* the pediatric pacing-dependent population.

To circumvent the concerns inherent to implantable electronic devices, a number of quite promising approaches towards engineering a biological pacemaker have recently appeared. These employ two principle strategies [260]: virus-based gene transfer aimed at converting resident cardiac cells into cells with pacemaker properties [242,243,263]; or, a cell-based strategy, in which *in-vitro* pre-processed cells are transplanted into the heart as pacemakers [89,159,154,264]. Of note, when generated from patient-derived iPSCs, such cells may become of great importance for personalized *in vitro* drug testing.

4.4.3. Modification and direct reprogramming of working myocardial cells

Generation of CCS cells follows similar principles to those used for working cardiomyocyte (re-)programming. However, the programming factors need to be carefully selected for this highly specific purpose. TFs such as T-box 3 (Tbx3), T-box 18 (Tbx18), short stature homeobox 2 (Shox2) or ISL LIM homeobox 1 (Isl1) play critical roles in the intrinsic development of PCs, but are absent or highly downregulated in the other cardiomyocyte subtypes [155,265–268]. The expression of ion channel components such as HCN4, calcium voltage-gated channel subunit alpha1 D (Cacna1d, $Ca_v1.3$), or the calcium voltage-gated channel subunit alpha1 G (Cagna1g, $Ca_v3.1$) also differs between PCs and the working myocardium, as do gap junction proteins such as Cx45 and Connexin30.2 (Cx30.2), both expressed instead of Cx43 [269–276].

The TF Tbx3 functions as transcriptional repressor during embryonic development, preventing expression of TFs typical for the working myocardium and thereby formation of working myocardial CMs. While Tbx3 is not solely responsible for SAN formation, it imposes a pacemaker gene program [268]. Accordingly, use of Tbx3 for direct reprogramming of resident myocardium cells appears attractive and has been evaluated in two different studies of Tbx3 overexpression in murine hearts. However, neither Tbx3 expression in atrial myocardium [266], nor tamoxifen-induced expression of Tbx3 in whole working myocardium [263] leads to fully functional PCs. While both reports demonstrate that Tbx3 can partially induce a number of pacemaker-related genes, the resulting cells still differ significantly from native pacemaker cells (PMCs) with regard to their overall expression patterns. In the first study, SAN specific markers were found to be upregulated such as Hcn4, Cx30.2 and Lbh. Similarly, atrial specific markers were downregulated, including natriuretic peptide A (Nppa), Cx40, Cx43 and sodium voltage-gated channel alpha subunit 5 (Scn5a, Nav1.5) [266]. The second study recapitulated the expression of some of these genes in isolated atrial cells (*e.g.* Lbh, Nppa, Cx43 and Cx40), but others reacted differently, such as Hcn4 which exhibited an unaltered expression pattern. Further, in the atrium another member of the cyclic nucleotide gated potassium channel family (Hcn1) was upregulated, while Hcn4 was actually downregulated in ventricular cells. Remarkably, the expression level of the TF NK2 homeobox 5 (Nkx2-5), which plays a pivotal role in working myocardium, was not modified by Tbx3 expression in these cells [263]. Overall, these findings suggest that Tbx3 alone is insufficient to convert working myocardium into PCs.

A related SAN-specification factor, Tbx18, has been tested after initial experiments with rat neonatal ventricular myocytes [242]. In contrast to other TFs tested (Shox2, Tbx3, Tbx5 and Tbx20), only Tbx18 transduction has been shown to significantly increase the number of spontaneously beating cultures. The resulting cells exhibit a more pacemaker-like morphology, as well as enhanced HCN4 expression and pacemaker-like cellular automaticity. Although the beating frequency of the transduced cells was double that of control cells, it was still far lower than that observed *in vivo* in rat hearts (95 bpm *vs.* 350 bpm) [242]. Importantly, the effect of Tbx18 has not only been demonstrated *in vitro*, but also reported in a large-animal model *in vivo* using adenoviral gene transfer. After pilot experiments in guinea pig hearts [242], a consecutive study evaluated the ability of Tbx18-expressing adenovirus injected into the interventricular septum of pigs to rescue induced complete heart block [243]. Examination of both animal models revealed evidence of partial transformation into pacemaker-like cells after transduction, as reflected by upregulation of Hcn4 and downregulation of working myocardial genes (Cx43 and Nkx2-5). In addition, ventricular ectopic beats were induced in both guinea pig and pigs hearts [242,243], with automaticity of the pig heart largely independent of the backup implanted electronic pacemaker for the short duration of the study [243]. Single cell analysis of the transducted guinea pig heart indicated the effect of Tbx18-expression to be only very transient: after 6 weeks, less than one third of transduced cells retained their pacemaker-like-morphology. However, due to limitations with respect to the recovery of single cells from the pig heart, insights into *in situ* reprogramming that could be gained from single cell analysis were lacking. Consistent with the temporal time course of expression of adenoviral vectors, SAN functional testing using electronic burst ventricular pacing revealed a rapid recovery in Tbx18-transduced animals at day 8, which increased to levels comparable with the control group (percutaneous GFP injected) by 2 weeks [243]. While the study describes the first partially successful *in situ* reprogramming towards a biological pacemaker in a clinically relevant large animal model, the long term effects of Tbx18-based reprogramming of working myocardium remain to be determined [242,243]. In this regard, a recent study used two independent Cre/loxP-mediated conditional transgenic mouse models to express Tbx18 in the atrial and ventricular myocardium during fetal development to further investigate the ability of Tbx18 to convert working myocardium into pacemaker cells [277]. Ectopic expression of Tbx18 was discernible from E12.5 or E14.5 and caused right ventricular hypoplasia, atrial dilatation and ventricular septal defects. In contrast to the experiments of the Marban group, no upregulation in expression of SAN-related genes was found in working myocardium, despite downregulation in chamber specific genes such as Cx40 and Nav1.5. Notably, Tbx18 expression also induced ectopic expression of atrial and ventricular marker genes, including Nppa in the ventricles and myosin, light polypeptide 2, regulatory, cardiac, slow (Myl2, Mlc2v) and myosin, heavy polypeptide 7, cardiac muscle, beta (Myh7) in the atria [277]. It remains to be seen whether the contrasting outcomes of the two studies arise from application of Tbx18 in different species [243,277], or from differing expression time points in the heart (fetal [277] *vs.* adult [243]). Such considerations will need to be carefully addressed to obviate potential side

Table 4
Overview of recently published direct reprogramming strategies of somatic cells towards diverse cardiovascular subtypes.

Cell origin	Host	Delivery system	*In vivo/in vitro*	Factor/substances	Target cell type	Special features	Literature
Neonatal cardiac fibroblasts	Mouse	Retroviral or lentiviral	*In vitro/in vivo*	Gata4, Mef2c, Tbx5	CM-like	30% cTnT^{+} cells (to a lesser extent in tail-tip fibroblasts) Upregulation of: Myh6, Actc1, Actn2, Nppa	Ieda et al., 2010 [227]
Neonatal cardiac fibroblasts	Mouse	Lentiviral MicroRNA specific culture conditions	*In vitro/in vivo*	miR-1/-122/-208/499 JAK inhibitor I	CM-like	Day 3: 2- to 3-fold upregulation of Mef2c, Tbx5, Hand2, Nkx2-5, Gata4 Day 6: expression of cTnT, cTnI, sarcomeric actinin Reprogramming efficiency: 1.13–5.28% in non-JAK inhibitor I-treated cells With JAK inhibitor I: 28% αMHC^{+} cells Enhanced cardiac function in mouse model	Jayawardena et al., 2012 and 2014 [235,236]
Fibroblasts (from hESCs; H9)	Human	Retroviral	*In vitro*	EGFP, ESRRG, GATA4, MEF2C, MESP1, TBX5, MYOCD, ZFPM2 SIS3	CM-like	20% functional Ca^{2+} transients Cardiac marker expression of: cTnT, α-Actinin, ACTC1, ACTN2, MYH6, MYL2, MYL7, TNNT2, NPPA, PLN, and RYR2 SIS3 significantly decreases αMHC^{+} cells	Fu et al., 2013 [221]
Tail tip and embryonic fibroblasts (B6;129S4)	Mouse	Retroviral	*In vitro*	M3 domain of mouse MyoD fused on carboxy-terminus of Mef2c, Gata4, Hand2, Tbx5 GSK126 (day 1–4), UNC0638 (day 3–7)	CM-like	Reprogramming efficiency: MM_3-GHT: 3.5% (> 15-fold increase) MM_3-GHT + GSK126: further increase to control 2.1-fold (most efficient combination) MM_3-GHT + UNC0638: further increase to control 2-fold	Hirai et al., 2013 [219] and 2014 [220]
NRVM	Rat	Adenoviral	*In vitro/in vivo*	Tbx18	SAN-like	Downregulation of Cx43 pacemaker-like AP profile (9.2%)	Kapoor et al., 2013 [242]
Neonatal foreskin and adult fibroblasts	Human	Retroviral	*In vitro*	Gata4, Hand2, Tbx5, myocardin, miR-1/-133 Culture time: 4–11 weeks	CM-like	~35% tropomyosin^{+} cells ~20% cTnT^{+} cells	Nam et al., 2013 [222]
CM	Pig	Adenoviral	*In vitro/in vivo*	Tbx18	SAN-like	Mean HR was higher in TBX18-transduced animals Sympathetic predominance in the TBX18-transduced group TBX18-transduced animals had persistent and stable activity	Hu et al., 2014 [243]

Embryonic fibroblasts (C57BL/6)	Mouse	Chemical cocktail	*In vitro*	On Matrigel 2-Stage protocol: Day 0–16: CRM (knockout DMEM, 15% FBS, and 5% KSR, 0.5% N2, 2% B27, 1% Glutamax, 1% NEAA, 0.1 mM β-mercaptoethanol, 50 μg/mL AA, 100 units/mL penicillin, 100 μg/mL streptomycin) + CRFVPT (10 μM CHIR (C), 10 μM RepSox (R), 50 μM Forskolin (F), 0.5 mM VPA (V), 5 μM Parnate, (P), 1 μM TTNPB (T)) Day 17–end: CMM (DMEM medium, 15% FBS, 2i (3 μM CHIR, 1 μM PD0325901), 1000 units/mL LIF, 50 μg/mL AA, and 1 μg/mL insulin)	CM-like	Morphology: spindle shape, rod shape or round shape Spontaneously beating activity: increases from day 8 Cardiac marker expression of: Mef2c, α-Actinin, Gata4, cTnT, Nkx2-5, α-MHC, N-cadherin, Cx43, cTnI Action potential of atrial- and ventricular-like CMs	Fu et al., 2015 [218]
Cardiac fibroblasts	Mouse	Retroviral Antibiotic selection	*In vitro*	Gata4, Mef2c, Tbx5	CM-like	Stoichiometry of G, M, T protein expression influences reprogramming efficiency High Mef2c and low Gata4, Tbx5 most efficient	Wang et al., 2015 [228]
Embryonic fibroblasts	Mouse	Retroviral	*In vitro*	Gata4, Hand2, Mef2c, Tbx5, miR-1/-133, Y-27632, Thiazovivin, SR-3677, A83-01	CM-like	Spontaneously beating activity without signaling inhibitors: GHMT > day 21, GHMT + miR-1/-133 > day 8 ROCK inhibitors enhance reprogramming of MEFs TGF-β inhibitors enhance reprogramming of MEFs most efficiently	Zhao et al., 2015 [217]
Embryonic, adult cardiac, tail tip	Mouse	Retroviral	*In vitro*	Gata4, Hand2, Mef2c, Tbx5, Akt1	CM-like	Spontaneously beating activity: MEFs > day 7 (50% > day 21), CFs > day 14, TTFs > day 21; Responsive to β-adrenoreceptor pharmacologic modulation, polynucleated, and hypertrophic	Zhou et al., 2015 [216]
Neonatal cardiac fibroblasts	Mouse	Retroviral (TFs) and lentiviral (shRNA)	*In vitro*	Gata4, Mef2c, Tbx5 shRNA of 35 selected components of chromatin modifying or remodeling complexes	CM-like	Bm1 downregulation significantly enhanced CM generation	Zhou et al., 2016 [231]
Neonatal cardiac fibroblasts	Mouse	TFs Specific culture conditions	*In vitro/in vivo*	Gata4, Mef2c, Tbx5 SB431542 XAV939	CM-like	8-Fold increased reprogramming efficiency Beating cells 1 week after reprogramming Enhanced cardiac function in mouse model	Mohamed et al., 2017 [230]

AA: ascorbic acid; Actc1: cardiac α-actin; Actn2: actinin α2; AP: action potential; Akt1: Akt1/protein kinase B; A83-01: TGF-β inhibitor; CHIR: CHIR99021 (GSK-3 inhibitor, Wnt activator); CM: cardiomyocytes; GSK126: Enhancer of Zeste Homolog 2 (Ezh2) inhibitor; KSR: knockout serum replacement; Myh6: α-myosin heavy chain; NEAA: non-essential amino acid; Nppa: natriuretic peptide precursor type A; NRVM: neonatal rat ventricular myocytes; PD0325901: MEK1/2 inhibitor; RepSox: TGF-β1 inhibitor; SAN: sino-atrial-nodal cells; SB431542: TGF-β inhibitor; shRNA: small hairpin RNA; SIS3: SMAD3 inhibitor (activated downstream of TGFβ signaling); SR-3677: ROCK inhibitor; TF: transcription factor; Thiazovivin: ROCK inhibitor; TTNPB: analog of retinoic acid; UNC0638: G9a and GLP inhibitor; VPA: valproic acid (histone deacetylase inhibitor); XAV939: WNT inhibitor; Y-27632: ROCK inhibitor.

effects before Tbx18 overexpression can be used to generate a biological pacemaker in patients.

Based on the experience with directly reprogramming fibroblasts into spontaneously beating cells, it seems likely that such an approach can be further developed to generate distinct cardiomyocyte subtypes, including pacemaker cells [278]. To identify possible TFs for such reprogramming, a study examined an initial group of 20 candidates by transducing them into embryonic fibroblasts of a mouse line expressing GFP under the control of Hcn4 regulatory promoter regions. An iterative process of successive omission of candidates expendable for EGFP expression led to the definition of the smallest group of factors promoting reporter expression, namely: Tbx5, Tbx3, Gata6 and either Retinoic acid receptor, gamma (Rarg) or Retinoid X receptor, alpha (Rxra). However, induction of significant Hcn4 expression alone does not seem sufficient to generate functional pacemaker cells as no spontaneous beating activity was observed, nor were the cells excitable *via* depolarization stimuli [278]. Moreover, in further experiments fibroblast transduction experiments using the "classical" cardiomyocyte reprogramming factors Gata4, Hand2, Mef2C and Tbx5, some of the reprogrammed cells expressed Hcn4-GFP. More detailed analysis revealed that this approach yielded multiple potential cardiac cell types: atrial-like, pacemaker-like and ventricular-like. Together, these data suggest that reprogramming of fibroblasts into PMC may be feasible in principle, but is still currently far from being reliably established.

4.4.4. Nodal cell programming of adult stem cells

Several studies have described diverse modifications of adult stem cells for pacemaker cell generation. These have primarily employed mesenchymal stem cells [101,279–290] derived from canine [101, 280–283], rat [284,291], rabbit [288–290] or human [279,285,286] tissue. Additional reports describe using adipose tissue-derived stem cells [287,292]. Most groups using ASCs chose to overexpress an Hcn-family member [279,281–286,288–290] to drive cell fate towards a nodal phenotype, while others used TFs such as Shox2 [101,280]. These have been combined with other treatments such as 5-Azacytidine [287] or electric-pulse current stimulation (EPCS) [280,281]. While utilizing divergent experimental setups, the scientific findings are quite comparable. Depending on the respective experiment, the resulting cells have displayed some nodal cell properties, for example measurable funny current (If) which could be enhanced with EPCS or with isoproterenol and blocked with cesium [280–282,286,289]. Moreover, expression of typical pacemaker genes like Cx45, Hcn4, Tbx3 are observed to increase, while genes associated with working myocardium, such as Cx43 and Nkx2-5, are downregulated [101,280,281]. Furthermore, a change in cell morphology towards a more pacemaker-like phenotype has been reported [101,280,287]. Co-culture of the modified ASCs with myocytes from newborn mice, regardless of origin, has led to increased beating frequency of the neonatal myocytes when compared to co-cultures with unmodified ASCs [101,286,289].

For *in vivo* testing, cells have been transplanted preferentially into canine hearts [282,286]. After induction of heart block, two publications describe the appearance of ventricular escape rhythms observed *via* ECG recordings [282,286]. However, this required vagal stimulation to induce sinus arrest to be apparent [282], and was associated with higher escape frequencies than control cell transplantation [286]. In all of these studies the lack of resulting cellular autonomous activity is strikingly consistent [101,279–291].

One report describes spontaneous activity of transformed ASCs derived from brown fat: interestingly, this phenomenon seems to reflect a reaction of the cells to the cultivation media as no genetic modification was applied [292]. While analysis of ultrastructural, proteomic, electrophysiological and pharmacological parameters of the resulting beating cells indicated the presence of some pacemaker-like features, further investigation is required, particularly over more prolonged culture periods [292].

In summary, significant further efforts are required to enable reprogramming of true nodal cells from ASCs.

4.4.5. Forward programming of pluripotent stem cells into nodal cells

PSCs represent a suitable source for any desired distinct cell type based on their unlimited differentiation potential. Concomitantly, this same unlimited differentiation potential represents a major obstacle to obtaining only a particular cell type. Thus a key challenge for the field is how to force PSCs exclusively towards a desired lineage.

The spontaneous differentiation rate of nodal cells from murine PSCs typically does not exceed ~1%. While great improvements have been made recently with regard to the differentiation of human PSCs into cardiac cell phenotypes using specific culture conditions [161,188,190,293, 294], the typical proportion of rare nodal cell types elicited still needs to be accurately defined.

Since the first description of mouse embryonic stem cells being differentiated into cardiomyocytes for the first time in 1991 [120], studies have examined the composition of different beating cells and attempted to specifically direct fate during differentiation. Besides considering cellular morphology and canonical marker expression patterns, the importance of recapitulating electrophysiological properties has been increasingly recognized. If one uses classical random differentiation protocols, the cells obtained represent a variety of cardiomyocyte cell types: nodal, atrial, ventricular and immature cardiomyocytes [245]. Accordingly, a key focus of interest has been how to reliably influence cell fate during differentiation. At present, there are three main strategies to enhance the proportion of nodal cell types within culture: stimulation *via* intrinsic culture conditions, enrichment *via* selection and forced overexpression of specific TFs.

With respect to the first strategy, the small molecule compound EBIO (1-ethyl-2-benzimidazolinone), a small-/intermediate-conductance Ca^{2+}-activated potassium channel modulator, has been postulated to increase the formation of nodal cells from murine ES cells. While application of EBIO to ES cells has been reported to lead to induction of sino-atrial and reduction in chamber-specific myocardial programs, the resultant cells had low beating frequencies, with no confirmation of ability to pace myocardium, or electrophysiological discrimination between mature pacemaker cells and those of an early/intermediate cell type which also spontaneously contract [295]. Interestingly, a recent study addressed the influence of EBIO on human PSC differentiation, shedding light on a mechanism of action *via* lineage-specific effects. While addition of EBIO resulted in dose-dependent enrichment of cardiomyocytes, with increased nodal- and atrial-like phenotypes, the effect was mainly attributable to a EBIO-induced severe reduction in cell survival, thereby favoring cardiac progenitor cell preservation [296].

Recently, a promising study described the generation of hPSC-derived pacemaker cells using a specific differentiation protocol combined with surface marker selection based on SIRPA (signal-regulatory protein alpha) [159]. SIRPA represents a cell-surface marker suitable to isolate populations of cardiomyocytes from hPSCs [177]. In combination with a transgene-independent differentiation protocol [159], the resulting hPSC-derived SAN-like cardiomyocyte cells fulfill a number of typical pacemaker features; in particular they are capable of pacing host tissue post-transplantation into the apex of rat hearts. However, further points will need to be addressed to better understand the cell generated: early/intermediate cell types were not taken into account despite the fact that the funny channel densities reported resemble those of immature cells, with action potential curves revealing clear plateau phases; an investigation of the characteristic Ca^{2+} release from sarcoplasmic reticulum; an examination of the specific morphology of single cells, specifically for typical spindle- or spider-shaped cells. An additional concern is the observed atypical expression of Cx43 and Cx40 [159].

A separate study [297] based on surface marker purification to isolate SAN progenitors used the activated leukocyte cell adhesion molecule (Alcam, CD166 antigen) during murine ESC differentiation. While the cells obtained display some pacemaker characteristics, selection

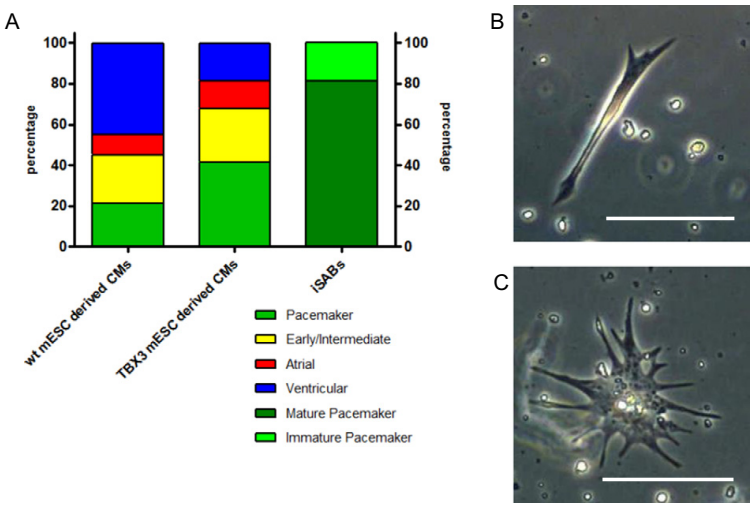


Fig. 3. Properties of iSAB derived single cells. (A) Cardiac subtype distribution based on single cell patch clamp analysis and funny channel density measurements. Morphology of iSAB derived representative (B) spindle and (C) spider cells; scale bar 100 μm.

via Alcam appears to be extremely dependent on time point and species. Thus, cell sorting at different differentiation time points yielded different degrees of sarcomeric α-actinin expression, a marker of cardiac-committed cells [297]. An earlier study using ALCAM as an expression marker in human embryonic stem cells revealed that the enriched cells manifested an embryonic cardiomyocyte phenotype [298]. Notably, only about 10% of Alcam-selected cells retained Hcn4 expression after 3 weeks in culture, a finding which may reflect a maturation process of initially Hcn4 positive early/intermediate CMs towards principally working myocardial cells over this timeframe [297].

In a further setting, the transcription factor Shox2 has been used with the goal of generating nodal cells from murine ESCs. Hashem and Claycomb transfected cells with a plasmid bearing a neomycin resistance gene controlled *via* the Shox2 promoter. Subsequent neomycin application during differentiation led to an almost pure population of Shox2 positive cells. Analysis of expression signatures revealed intrinsic nodal characteristics (Tbx3, HCN4, Cx45, $Ca_v1.3$, $Ca_v3.1$) as well as expression of other cardiac marker genes (Tbx5, Mlc2v, Cx43). however, while the cells were spontaneously active, no additional functional data were shown [299]. In a similar approach, Ionta et al. used Shox2 overexpression in mouse embryonic stem cells to force cells into a nodal cell lineage [154]. However, while the resulting cells were spontaneously active, the beating frequencies were below 80 bpm and therefore did not exceed those of WT-ES cell derived CM [154]. Consequently, it is unclear whether these cells represent functional pacemaker cells.

Our group has combined overexpression of the highly conserved key nodal cell inducer, Tbx3, with a neomycin resistance gene under controlled of the well-established αMHC-promoter. This approach leads to small aggregates consisting of ~300–500 cells, which we term "induced sino-atrial bodies" (iSABs). iSABs exhibit high beating frequencies of between 400 and 500 bpm *in vitro*, thereby for the first time truly corresponding to those of a murine heart and even exceeding *in vitro* cultivated nodal cells derived from mouse SAN. Evidence from extensive analysis of these cells, including confocal laser scanning microscopy, FACS, single-cell patch clamping (Fig. 3A), funny channel density measurements and Ca^{2+} imaging reveals that iSABs consistently represent over 80% mature functional nodal cells, with the remainder constituting immature nodal cells. Additional single cell analysis identifies characteristic spindle (Fig. 3B) and spider (Fig. 3C) cell morphology. To further address the pacing potential of iSABs, we employed the *ex vivo* model system of cultivated mouse ventricular slices. Remarkably, iSABs were capable of integrating into these slices, retaining their spontaneous activity and pacing the heart slices to result in robust contraction. We confirmed functional coupling to the slices using calcium-transient analysis, which revealed synchronization between iSABs and slices. Therefore, we have introduced, for the first time, highly pure PSC-derived nodal tissue which is functional on the physiological level *in vitro*, as well as in an *ex vivo* model [89]. Recently, we utilized iSABs as an *in vitro* model system to help decipher the role of the γ2 subunit of AMP-activated protein kinase (AMPK) in the regulation of SAN biology and infer a role for AMPK in control of mammalian intrinsic heart rate [300]. An important next step will be to determine the ability of iSABs to pace cardiac tissue *in vivo* and to prove whether this approach can be extended to human PSCs.

4.4.6. Systems-based network approaches to enhance nodal cell programming

Further improvement of nodal cell programing will require an in depth understanding of underlying gene regulatory mechanisms. Given the complexity of the cardiovascular system and cardiovascular diseases, systems-based approaches play an increasingly important role in elucidating interactions between underlying traits and processes by using multiple 'omics' layers [301]. Such systems-based approaches are global analyses in which the different molecular levels are investigated and then integrated into qualitative and quantitative mathematical models (*e.g.* Boolean models, ordinary differential equations, network analysis concepts, *etc.*), providing an additional layer of understanding for cardiomyocyte dynamics [302]. Hence we used our recently developed iSABs, representing the first highly pure, stem cell programming-derived nodal cell tissue, to define the pacemaker transcriptome [91]. We highlight up-to-date systems-based approaches and compare our findings with the existing literature on a multi-level scale to verify the overall quality of the iSAB model system and address its potential transferability towards the human SAN (Fig. 4).

The first step in such a systems-based data analysis procedure is to define the model system and available experimental input data (*e.g.* different SC-derived cardiomyocyte subtypes characterized by RNA-Seq data). Recent advances in high-throughput sequencing (HTS) have emphasized the important and versatile roles of coding and non-coding RNAs, including quantification of splice variants or identification of novel ncRNAs, during cardiac development. We used our RNA-Seq

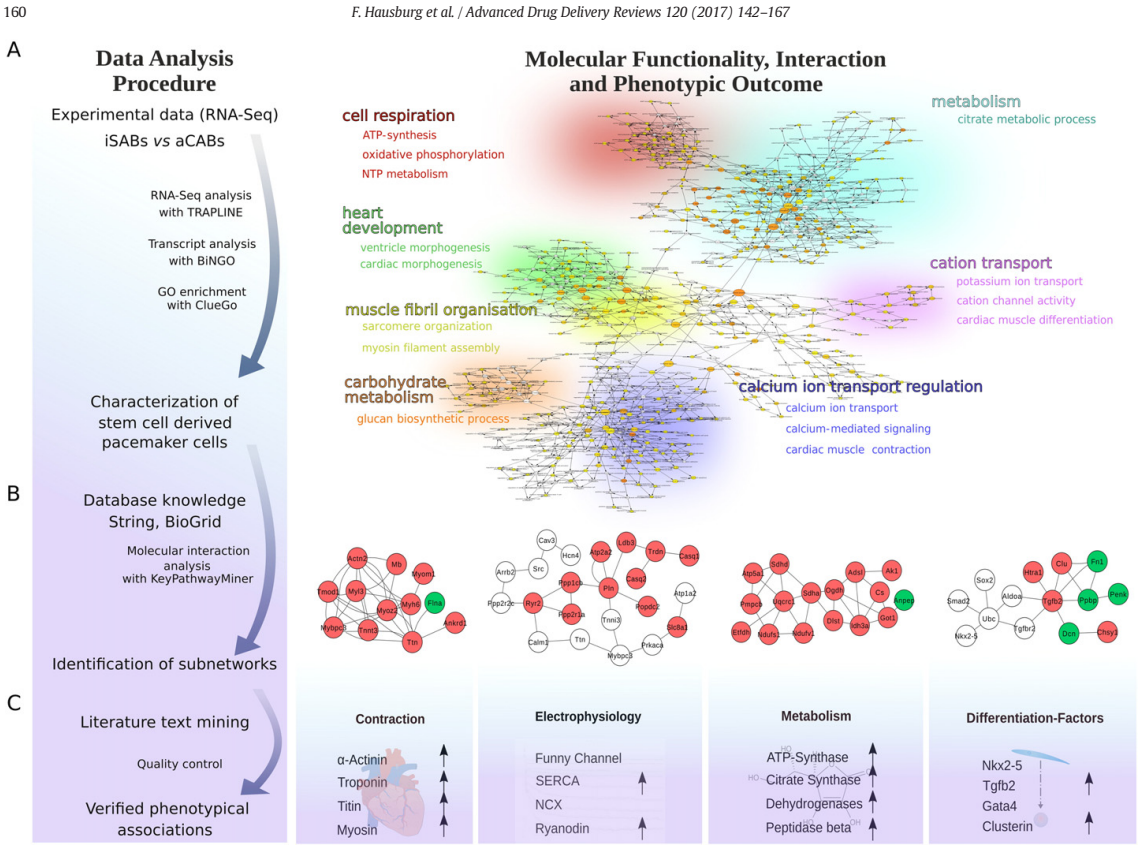


Fig. 4. Systems-based data analysis procedure for the identification of molecular functionalities, interactions and phenotypic associations applied to stem cell-derived cardiac cell types by using RNA-Seq data. (A) Calculation of overrepresented GO terms using the Cytoscape applications BiNGO and ClueGo. (B) Identified subnetworks obtained after KeyPathwayMiner analysis of the former constructed interactome network. Red represents the upregulated transcripts within iSABs and green represents the downregulated transcripts. The edges (lines between encircled genes) are experimentally verified interactions obtained from String and BioGrid. (C) Summary of the upregulated factors identified in the data and the literature for processes within contraction, electrophysiology, metabolism and differentiation.

workflow, called *TRAPLINE*, to define the differences in transcriptomes between iSABs and randomly distributed mESC-derived cardiomyocyte subtypes [89,91]. Using the Galaxy framework, the numerous embedded data analysis workflows guarantees simple access, easy extension and flexible adaption of computational tools to individual needs, as well as sophisticated analyses that do not require in-depth command-line knowledge [303,304]. Data analysis with *TRAPLINE* results in a set of differentially expressed genes, their corresponding protein-protein interactions, splice variants, promoter activity and predicted miRNA-target interactions [91].

A central concept in systems biology is that networks, rather than classic linear pathways, underlie biological processes. The concept of biological networks arose when classic signaling pathways were represented as graphs in which the components (*i.e.* expressed gene transcripts) were termed nodes and their interactions (*i.e.* genes encoding transcription factors or protein-protein interactions) were called links or edges [305]. Demonstrating the general cellular differences between the two cell types through analysis of the transcriptome, we used the Biological Networks Gene Ontology tool (*BiNGO*) to determine the Gene Ontology (GO) terms significantly overrepresented in a set of significantly upregulated transcripts in iSABs (Fig. 4A) [306]. These GO terms represent the accumulated biological processes of the significantly overexpressed transcripts. While the underlying statistical *p*-value, false discovery rate and family-wise error rate provide a good first impression of a specific functional category, it is important to also check the functional categories in the entire GO hierarchy. Thus it is highly likely that when a whole branch of the GO hierarchy is highlighted as being significantly overrepresented, the most intensely colored nodes furthest down the hierarchy can be expected to be the most biologically relevant ones.

The second step of the analysis characterizes the relevant components of the system (*e.g.* the actual set of significant differentially expressed gene transcripts driving the enhanced cardiac rhythm). A prominent method is that of gene set enrichment analysis (GWAS), performed by *ClueGo* [307]. The gene sets are analyzed on the basis of prior biological knowledge, such as use of GO or signaling pathways such as *Wikipathways*. Using statistical tests like the Fisher exact test, one can then ask whether the genes are enriched within a collection of pathways. Such analyses are highly dependent on using current versions of curated sets of GO annotations. Only annotated transcripts and ncRNAs can be integrated through GWAS analyses. Newly discovered transcripts or ncRNAs have to be characterized beforehand by other experiments or *in silico* prediction simulations [308,309]. One approach to overcome this limitation and independently link coding and non-coding transcripts without using annotations is the use of weighted correlation network analysis to identify clusters of highly correlated genes or an intramodular hub gene [310]. As described earlier, in using iSABs to define a function for AMPK in intrinsic heart rate regulation [300], we constructed a gene co-expression network from the iSAB transcriptome and identified

Prkag2 in a module highly interconnected with known pacemaker-relevant genes. Hierarchical clustering and classical multi-dimensional scaling revealed, in common with the SAN transcriptome, that this Prkag2-containing module signified an important signaling hub with significant connectivity to genes vital for normal SAN function [300].

The third analysis step is to determine how the identified gene transcripts interact with each other or regulate other relevant interaction partners. This can be done by using data mining approaches [311] and databases such as BioGrid and String to incorporate resulting protein-protein interactions (PPTs) into the network, which is subsequently investigated for underlying dynamics and the molecular enrichment among genes within the network mathematically (*i.e.* how it responds to various perturbations and interconnects with other data layers). Based on the combined use of the BioGrid and String databases, we have obtained a network with 8120 nodes and 55,720 edges, representing the interactome. After applying the tool *KeyPathwayMiner* [312], we were able to identify the most important subnetworks within the constructed interactome model to demonstrate known molecular interactions between significantly upregulated genes (Fig. 4B). Once such a network is developed and available for the researcher, more sophisticated mathematical models can be applied to the network. Based on the input datatype available, one can employ ordinary differential equations (ODE), discrete modeling or hybrid modeling (composed of ODE and logic sub-modules) as a strategy to handle large scale, non-linear biochemical networks [313]. It has been shown that these kinds of network approaches are able to identify a specific regulatory core within a large gene network and, moreover, to predict receptor signatures associated with certain diseases [314]. Nevertheless, such *in silico* simulations and subsequent signature predictions still need experimental validation.

Ultimately, the information obtained at each analysis step is combined to draw conclusions about the complex behavior of the stem cell-derived cardiac pacemaker model and compared to current knowledge about the human SAN (Fig. 4C). The results we have obtained for phenotypic associations such as contraction, electrophysiology, metabolism and differentiation factors, are in line with the current literature about the SAN. The knowledge gained from such systems-based analyses will be crucial for further optimization of cell programming and purification [91].

Acknowledgements

This work was supported by the Federal Ministry of Education and Research Germany (FKZ 0312138A, FKZ 031L0106C, FKZ 02NUK043C and FKZ 316159), the State Mecklenburg-Western Pomerania with EU Structural Funds (ESF/IVWM-B34-0030/10 and ESF/IVBM-B35-0010/12), and the DFG (DA1296/2-1) and the German Heart Foundation (F/01/12). A.Y. is supported by the Wellcome Trust (204442/Z/16/Z); the Academy of Medical Sciences (Clinical Lecturer Starter Grant); and the National Institute of Health Research in the form of an Academic Clinical Lectureship. In addition, F.H. and R.D. are supported by the FORUN Program of Rostock University Medical Centre (889001) and the DAMP Foundation. R.D. is further funded by the BMBF (VIP + 00240).

References

[1] A.P. Ambrosy, G.C. Fonarow, J. Butler, O. Chioncel, S.J. Greene, M. Vaduganathan, S. Nodari, C.S.P. Lam, N. Sato, A.N. Shah, M. Gheorghiade, The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries, J. Am. Coll. Cardiol. 63 (12) (2014) 1123–1133.
[2] A.L. Bui, T.B. Horwich, G.C. Fonarow, Epidemiology and risk profile of heart failure, Nat. Rev. Cardiol. 8 (1) (2011) 30–41.
[3] A. Seki, M.C. Fishbein, Predicting the development of cardiac allograft vasculopathy, Cardiovasc. Pathol. 23 (5) (2014) 253–260.
[4] J.A. Kobashigawa, The search for a gold standard to detect rejection in heart transplant patients: are we there yet? Circulation 135 (10) (2017) 936–938.
[5] Eurotransplant International Foundation, Annual Report 2015(Leiden) 2015.
[6] M. Tonsho, S. Michel, Z. Ahmed, A. Alessandrini, J.C. Madsen, Heart transplantation: challenges facing the field, Cold Spring Harb. Perspect. Med. 4 (5) (2014).
[7] A. Jain, R. Bansal, Applications of regenerative medicine in organ transplantation, J. Pharm. Bioallied Sci. 7 (3) (2015) 188–194.
[8] A. Heidary Rouchi, M. Mahdavi-Mazdeh, Regenerative medicine in organ and tissue transplantation: shortly and practically achievable? Int. J. Organ Transplant. Med. 6 (3) (2015) 93–98.
[9] G. Orlando, S. Soker, R.J. Stratta, A. Atala, Will regenerative medicine replace transplantation? Cold Spring Harb. Perspect. Med. 3 (8) (2013).
[10] N. Pavo, S. Charwat, N. Nyolczas, A. Jakab, Z. Murlasits, J. Bergler-Klein, M. Nikfardjam, I. Benedek, T. Benedek, I.J. Pavo, B.J. Gersh, K. Huber, G. Maurer, M. Gyöngyösi, Cell therapy for human ischemic heart diseases: critical review and summary of the clinical experiences, J. Mol. Cell. Cardiol. 75 (2014) 12–24.
[11] S.A. Fisher, C. Doree, A. Mathur, D.P. Taggart, E. Martin-Rendon, Stem cell therapy for chronic ischaemic heart disease and congestive heart failure, Cochrane Database Syst. Rev. 12 (2016), CD007888.
[12] O. Bergmann, R.D. Bhardwaj, S. Bernard, S. Zdunek, F. Barnabe-Heider, S. Walsh, J. Zupicich, K. Alkass, B.A. Buchholz, H. Druid, S. Jovinge, J. Frisen, Evidence for cardiomyocyte renewal in humans, Science 324 (5923) (2009) 98–102.
[13] M. Mollova, K. Bersell, S. Walsh, J. Savla, L.T. Das, S.-Y. Park, L.E. Silberstein, Cristobal G. Dos Remedios, D. Graham, S. Colan, B. Kühn, Cardiomyocyte proliferation contributes to heart growth in young humans, Proc. Natl. Acad. Sci. U. S. A. 110 (4) (2013) 1446–1451.
[14] A.N. Paradis, M.S. Gay, L. Zhang, Binucleation of cardiomyocytes: the transition from a proliferative to a terminally differentiated state, Drug Discov. Today 19 (5) (2014) 602–609.
[15] F. Li, X. Wang, J.M. Capasso, A.M. Gerdes, Rapid transition of cardiac myocytes from hyperplasia to hypertrophy during postnatal development, J. Mol. Cell. Cardiol. 28 (8) (1996) 1737–1746.
[16] Z. Liu, S. Yue, X. Chen, T. Kubin, T. Braun, Regulation of cardiomyocyte polyploidy and multinucleation by CyclinG1, Circ. Res. 106 (9) (2010) 1498–1506.
[17] J.C. Garbern, R.T. Lee, Cardiac stem cell therapy and the promise of heart regeneration, Cell Stem Cell 12 (6) (2013) 689–698.
[18] S.E. Senyo, M.L. Steinhauser, C.L. Pizzimenti, V.K. Yang, L. Cai, M. Wang, T.-D. Wu, J.-L. Guerquin-Kern, C.P. Lechene, R.T. Lee, Mammalian heart renewal by pre-existing cardiomyocytes, Nature 493 (7432) (2013) 433–436.
[19] M. Sahara, F. Santoro, K.R. Chien, Programming and reprogramming a human heart cell, EMBO J. 34 (6) (2015) 710–738.
[20] S.D. Vincent, M.E. Buckingham, How to make a heart: the origin and regulation of cardiac progenitor cells, Organogenesis in Development, Elsevier 2010, pp. 1–41.
[21] M. Xin, E.N. Olson, R. Bassel-Duby, Mending broken hearts: cardiac development as a basis for adult heart regeneration and repair, Nat. Rev. Mol. Cell Biol. 14 (8) (2013) 529–541.
[22] S.L. Paige, K. Plonowska, A. Xu, S.M. Wu, Molecular regulation of cardiomyocyte differentiation, Circ. Res. 116 (2) (2015) 341–353.
[23] S.M. Meilhac, F. Lescroart, C. Blanpain, M.E. Buckingham, Cardiac cell lineages that form the heart, Cold Spring Harb. Perspect. Med. 4 (9) (2014), a013888.
[24] V. Garcia-Martinez, G.C. Schoenwolf, Primitive-streak origin of the cardiovascular system in avian embryos, Dev. Biol. 159 (2) (1993) 706–719.
[25] D.A. Turner, P. Rue, J.P. Mackenzie, E. Davies, A.A. Martinez, Brachyury cooperates with Wnt/beta-catenin signalling to elicit primitive-streak-like behaviour in differentiating mouse embryonic stem cells, BMC Biol. 12 (2014) 63.
[26] P.P. Tam, M. Parameswaran, S.J. Kinder, R.P. Weinberger, The allocation of epiblast cells to the embryonic heart and other mesodermal lineages: the role of ingression and tissue movement during gastrulation, Development 124 (9) (1997) 1631–1642.
[27] G. Chiapparo, X. Lin, F. Lescroart, S. Chabab, C. Paulissen, L. Pitisci, A. Bondue, C. Blanpain, Mesp1 controls the speed, polarity, and directionality of cardiovascular progenitor migration, J. Cell Biol. 213 (4) (2016) 463–477.
[28] Y. Saga, Mesp1 expression is the earliest sign of cardiovascular development, Trends Cardiovasc. Med. 10 (8) (2000) 345–352.
[29] Y. Saga, S. Miyagawa-Tomita, A. Takagi, S. Kitajima, J.I. Miyazaki, T. Inoue, MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube, Development 126 (15) (1999) 3437–3447.
[30] Q. Liang, C. Xu, X. Chen, X. Li, C. Lu, P. Zhou, L. Yin, R. Qian, S. Chen, Z. Ling, N. Sun, The roles of Mesp family proteins: functional diversity and redundancy in differentiation of pluripotent stem cells and mammalian mesodermal development, Protein Cell 6 (8) (2015) 553–561.
[31] S.J. Kattman, T.L. Huber, G.M. Keller, Multipotent flk-1 + cardiovascular progenitor cells give rise to the cardiomyocyte, endothelial, and vascular smooth muscle lineages, Dev. Cell 11 (5) (2006) 723–732.
[32] Z. He, M. Grunewald, Y. Dor, E. Keshet, VEGF regulates relative allocation of Isl1 + cardiac progenitors to myocardial and endocardial lineages, Mech. Dev. 142 (2016) 40–49.
[33] K. Musunuru, I.J. Domian, K.R. Chien, Stem cell models of cardiac development and disease, Annu. Rev. Cell Dev. Biol. 26 (2010) 667–687.
[34] X. Liang, G. Wang, L. Lin, J. Lowe, Q. Zhang, L. Bu, Y. Chen, J. Chen, Y. Sun, S.M. Evans, HCN4 dynamically marks the first heart field and conduction system precursors, Circ. Res. 113 (4) (2013) 399–407.
[35] S.M. Stevens, W.T. Pu, HCN4 charges up the first heart field, Circ. Res. 113 (4) (2013) 350–351.
[36] T. Brade, L.S. Pane, A. Moretti, K.R. Chien, K.-L. Laugwitz, Embryonic heart progenitors and cardiogenesis, Cold Spring Harb. Perspect. Med. 3 (10) (2013), a013847.
[37] D. Später, M.K. Abramczuk, K. Buac, L. Zangi, M.W. Stachel, J. Clarke, M. Sahara, A. Ludwig, K.R. Chien, A HCN4 + cardiomyogenic progenitor derived from the first heart field and human pluripotent stem cells, Nat. Cell Biol. 15 (9) (2013) 1098–1106.

[38] S.M. Wu, Y. Fujiwara, S.M. Cibulsky, D.E. Clapham, C.L. Lien, T.M. Schultheiss, S.H. Orkin, Developmental origin of a bipotential myocardial and smooth muscle cell precursor in the mammalian heart, Cell 127 (6) (2006) 1137–1150.
[39] A.F. Moorman, V.M. Christoffels, R.H. Anderson, M.J. van den Hoff, The heart-forming fields: one or multiple? Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 362 (1484) (2007) 1257–1265.
[40] A. Sizarov, J. Ya, R.A. de Boer, W.H. Lamers, V.M. Christoffels, A.F. Moorman, Formation of the building plano fthe human heart: morphogenesis, growth, and differentiation, Circulation 123 (10) (2011) 1125–1135.
[41] V. George, S. Colombo, K.L. Targoff, An early requirement for nkx2-5 ensures the first and second heart field ventricular identity and cardiac function into adulthood, Dev. Biol. 400 (1) (2015) 10–22.
[42] M. Buckingham, S. Meilhac, S. Zaffran, Building the mammalian heart from two sources of myocardial cells, Nat. Rev. Genet. 6 (11) (2005) 826–835.
[43] L.A. Dyer, M.L. Kirby, The role of secondary heart field in cardiac development, Dev. Biol. 336 (2) (2009) 137–144.
[44] F. Rochais, K. Mesbah, R.G. Kelly, Signaling pathways controlling second heart field development, Circ. Res. 104 (8) (2009) 933–942.
[45] P. Pandur, I.O. Sirbu, S.J. Kühl, M. Philipp, M. Kühl, Islet1-expressing cardiac progenitor cells: a comparison across species, Dev. Genes Evol. 223 (1–2) (2013) 117–129.
[46] A. Moretti, L. Caron, A. Nakano, J.T. Lam, A. Bernshausen, Y. Chen, Y. Qyang, L. Bu, M. Sasaki, S. Martin-Puig, Y. Sun, S.M. Evans, K.-L. Laugwitz, K.R. Chien, Multipotent embryonic isl1 + progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification, Cell 127 (6) (2006) 1151–1165.
[47] M. Bressan, G. Liu, T. Mikawa, Early mesodermal cues assign avian cardiac pacemaker fate potential in a tertiary heart field, Science (New York, N.Y.) 340 (6133) (2013) 744–748.
[48] M. Bressan, G. Liu, J.D. Louie, T. Mikawa, Cardiac Pacemaker Development from a Tertiary Heart Field, 2016, 281–288.
[49] A.M. Misfeldt, S.C. Boyle, K.L. Tompkins, V.L. Bautch, P.A. Labosky, H.S. Baldwin, Endocardial cells are a distinct endothelial lineage derived from Flk1 + multipotent cardiovascular progenitors, Dev. Biol. 333 (1) (2009) 78–89.
[50] J. Schlueter, T. Brand, Epicardial progenitor cells in cardiac development and regeneration, J. Cardiovasc. Transl. Res. 5 (5) (2012) 641–653.
[51] T.C. Katz, M.K. Singh, K. Degenhardt, J. Rivera-Feliciano, R.L. Johnson, J.A. Epstein, C.J. Tabin, Distinct compartments of the proepicardial organ give rise to coronary vascular endothelial cells, Dev. Cell 22 (3) (2012) 639–650.
[52] B. Zhou, Q. Ma, S. Rajagopal, S.M. Wu, I. Domian, J. Rivera-Feliciano, D. Jiang, A. von Gise, S. Ikeda, K.R. Chien, W.T. Pu, Epicardial progenitors contribute to the cardiomyocyte lineage in the developing heart, Nature 454 (7200) (2008) 109–113.
[53] J. Banach, W. Gilewski, A. Słomka, K. Buszko, J. Błazejewski, D. Karasek, D. Rogowicz, E. Zekanowska, W. Sinkiewicz, Bone morphogenetic protein 6—a possible new player in pathophysiology of heart failure, Clin. Exp. Pharmacol. Physiol. 43 (12) (2016) 1247–1250.
[54] F.-F. Li, X. Deng, J. Zhou, P. Yan, E.-Y. Zhao, S.-L. Liu, Characterization of human bone morphogenetic protein gene variants for possible roles in congenital heart disease, Mol. Med. Rep. 14 (2) (2016) 1459–1464.
[55] T.M. Schultheiss, J.B. Burch, A.B. Lassar, A role for bone morphogenetic proteins in the induction of cardiac myogenesis, Genes Dev. 11 (4) (1997) 451–462.
[56] B. Andrée, D. Duprez, B. Vorbusch, H.-H. Arnold, T. Brand, BMP-2 induces ectopic expression of cardiac lineage markers and interferes with somite formation in chicken embryos, Mech. Dev. 70 (1–2) (1998) 119–131.
[57] T. Schlange, B. Andrée, H.-H. Arnold, T. Brand, BMP2 is required for early heart development during a distinct time period, Mech. Dev. 91 (1–2) (2000) 259–270.
[58] P. Han, J. Bloomekatz, J. Ren, R. Zhang, J.D. Grinstein, L. Zhao, C.G. Burns, C.E. Burns, R.M. Anderson, N.C. Chi, Coordinating cardiomyocyte interactions to direct ventricular chamber morphogenesis, Nature 534 (7609) (2016) 700–704.
[59] H.Y. Lin, D.C. Lee, H.D. Wang, Y.H. Chi, I.M. Chiu, Activation of FGF1B promoter and FGF1 are involved in cardiogenesis through the signaling of PKC, but not MAPK, Stem Cells Dev. 24 (24) (2015) 2853–2863.
[60] L.J. Samuel, B.V. Latinkic, Early activation of FGF and nodal pathways mediates cardiac specification independently of Wnt/beta-catenin signaling, PLoS One 4 (10) (2009), e7650.
[61] Y. Watanabe, S. Zaffran, A. Kuroiwa, H. Higuchi, T. Ogura, R.P. Harvey, R.G. Kelly, M. Buckingham, Fibroblast growth factor 10 gene regulation in the second heart field by Tbx1, Nkx2-5, and Islet1 reveals a genetic switch for down-regulation in the myocardium, Proc. Natl. Acad. Sci. U. S. A. 109 (45) (2012) 18273–18,280.
[62] P. Pandur, M. Lasche, L.M. Eisenberg, M. Kuhl, Wnt-11 activation of a non-canonical Wnt signalling pathway is required for cardiogenesis, Nature 418 (6898) (2002) 636–641.
[63] S.M. Ahmad, P. Bhattacharyya, N. Jeffries, S.S. Gisselbrecht, A.M. Michelson, Two Forkhead transcription factors regulate cardiac progenitor specification by controlling the expression of receptors of the fibroblast growth factor and Wnt signaling pathways, Development 143 (2) (2016) 306–317.
[64] J. Martin, B.A. Afouda, S. Hoppler, Wnt/beta-catenin signalling regulates cardiomyogenesis via GATA transcription factors, J. Anat. 216 (1) (2010) 92–107.
[65] T. Nakamura, M. Sano, Z. Songyang, M.D. Schneider, A Wnt- and beta-catenin-dependent pathway for mammalian cardiac myogenesis, Proc. Natl. Acad. Sci. U. S. A. 100 (10) (2003) 5834–5839.
[66] C. Kwon, J. Arnold, E.C. Hsiao, M.M. Taketo, B.R. Conklin, D. Srivastava, Canonical Wnt signaling is a positive regulator of mammalian cardiac progenitors, Proc. Natl. Acad. Sci. U. S. A. 104 (26) (2007) 10894–10,899.
[67] T. Nakanishi, R.R. Markwald, H.S. Baldwin, B.B. Keller, D. Srivastava, H. Yamagishi, Etiology and Morphogenesis of Congenital Heart Disease: From Gene Function and Cellular Interaction to Morphology, Springer Open, Japan, 2016.
[68] A.J. Marelli, A.S. Mackie, R. Ionescu-Ittu, E. Rahme, L. Pilote, Congenital heart disease in the general population: changing prevalence and age distribution, Circulation 115 (2) (2007) 163–172.
[69] D. van der Linde, E.E. Konings, M.A. Slager, M. Witsenburg, W.A. Helbing, J.J. Takkenberg, J.W. Roos-Hesselink, Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis, J. Am. Coll. Cardiol. 58 (21) (2011) 2241–2247.
[70] J.J. Schott, D.W. Benson, C.T. Basson, W. Pease, G.M. Silverbach, J.P. Moak, B.J. Maron, C.E. Seidman, J.G. Seidman, Congenital heart disease caused by mutations in the transcription factor NKX2-5, Science (New York, N.Y.) 281 (5373) (1998) 108–111.
[71] D.W. Benson, G.M. Silverbach, A. Kavanaugh-McHugh, C. Cottrill, Y. Zhang, S. Riggs, O. Smalls, M.C. Johnson, M.S. Watson, J.G. Seidman, C.E. Seidman, J. Plowden, J.D. Kugler, Mutations in the cardiac transcription factor NKX2-5 affect diverse cardiac developmental pathways, J. Clin. Invest. 104 (11) (1999) 1567–1573.
[72] D.B. McElhinney, E. Geiger, J. Blinder, D. Woodrow Benson, E. Goldmuntz, NKX2-5 mutations in patients with congenital heart disease, J. Am. Coll. Cardiol. 42 (9) (2003) 1650–1655.
[73] H. Ashraf, L. Pradhan, E.L. Chang, R. Terada, N.J. Ryan, L.E. Briggs, R. Chowdhury, M.A. Zarate, Y. Sugi, H.J. Nam, D.W. Benson, R.H. Anderson, H. Kasahara, A mouse model of human congenital heart disease: high incidence of diverse cardiac anomalies and ventricular noncompaction produced by heterozygous Nkx2-5 homeodomain missense mutation, Circ. Cardiovasc. Genet. 7 (4) (2014) 423–433.
[74] L. Gioli-Pereira, A.C. Pereira, S.M. Mesquita, J. Xavier-Neto, A.A. Lopes, J.E. Krieger, NKX2-5 mutations in patients with non-syndromic congenital heart disease, Int. J. Cardiol. 138 (3) (2010) 261–265.
[75] M.M. Balci, R. Akdemir, NKX2-5 mutations and congenital heart disease: is it a marker of cardiac anomalies? Int. J. Cardiol. 147 (3) (2011) e44–5.
[76] P. Ouyang, E. Saarel, Y. Bai, C. Luo, Q. Lv, Y. Xu, F. Wang, C. Fan, A. Younoszai, Q. Chen, X. Tu, Q.K. Wang, A de novo mutation in NKX2-5 associated with atrial septal defects, ventricular noncompaction, syncope and sudden death, Clin. Chim. Acta 412 (1–2) (2011) 170–175.
[77] M. Salazar, F. Consoli, V. Villegas, V. Caicedo, V. Maddaloni, P. Daniele, G. Calaniello, S. Pachon, F. Nunez, G. Limongelli, G. Pacileo, B. Marino, J.E. Bernal, A. de Luca, B. Dallapiccola, Search of somatic GATA4 and NKX2-5 gene mutations in sporadic septal heart defects, Eur. J. Med. Genet. 54 (3) (2011) 306–309.
[78] X. Xie, X. Shi, X. Xun, L. Rao, Associations of NKX2-5 genetic polymorphisms with the risk of congenital heart disease: a meta-analysis, Pediatr. Cardiol. (2016).
[79] V. Garg, I.S. Kathiriya, R. Barnes, M.K. Schluterman, I.N. King, C.A. Butler, C.R. Rothrock, R.S. Eapen, K. Hirayama-Yamada, K. Joo, R. Matsuoka, J.C. Cohen, D. Srivastava, GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5, Nature 424 (6947) (2003) 443–447.
[80] S.K. Rajagopal, Q. Ma, D. Obler, J. Shen, A. Manichaikul, A. Tomita-Mitchell, K. Boardman, C. Briggs, V. Garg, D. Srivastava, E. Goldmuntz, K.W. Broman, D.W. Benson, L.B. Smoot, W.T. Pu, Spectrum of heart disease associated with murine and human GATA4 mutation, J. Mol. Cell. Cardiol. 43 (6) (2007) 677–685.
[81] W. Zhang, X. Li, A. Shen, W. Jiao, X. Guan, Z. Li, GATA4 mutations in 486 Chinese patients with congenital heart disease, Eur. J. Med. Genet. 51 (6) (2008) 527–535.
[82] Y. Chen, J. Mao, Y. Sun, Q. Zhang, H.B. Cheng, W.H. Yan, K.W. Choy, H. Li, A novel mutation of GATA4 in a familial atrial septal defect, Clin. Chim. Acta 411 (21—22) (2010) 1741–1745.
[83] H. Zhu, Forkhead box transcription factors in embryonic heart development and congenital heart disease, Life Sci. 144 (2016) 194–201.
[84] M.A. Chaix, G. Andelfinger, P. Khairy, Genetic testing in congenital heart disease: a clinical approach, World J. Cardiol. 8 (2) (2016) 180–191.
[85] M. Liu, L. Zhao, J. Yuan, Establishment of relational model of congenital heart disease markers and GO functional analysis of the association between its serum markers and susceptibility genes, Comput. Math. Methods Med. 2016 (2016), 9506829.
[86] J.R. Cowan, S.M. Ware, Genetics and genetic testing in congenital heart disease, Clin. Perinatol. 42 (2) (2015) 373–393 (ix).
[87] X.-Y. Jiang, Y.-L. Feng, L.-T. Ye, X.-H. Li, J. Feng, M.-Z. Zhang, H.S. Shelat, M. Wassler, Y. Li, Y.-J. Geng, X.-Y. Yu, Inhibition of Gata4 and Tbx5 by nicotine-mediated DNA methylation in myocardial differentiation, Stem Cell Rep. 8 (2) (2017) 290–304.
[88] D. Wei, L. Taq, M. Huang, Genetic variations involved in sudden cardiac death and their associations and interactions, Heart Fail. Rev. 21 (4) (2016) 401–414.
[89] J.J. Jung, B. Husse, C. Rimmbach, S. Krebs, J. Stieber, G. Steinhoff, A. Dendorfer, W.-M. Franz, R. David, Programming and isolation of highly pure physiologically and pharmacologically functional sinus-nodal bodies from pluripotent stem cells, Stem Cell Rep. 2 (5) (2014) 592–605.
[90] C. Rimmbach, J.J. Jung, R. David, Generation of murine cardiac pacemaker cell aggregates based on ES-cell-programming in combination with Myh6-promoter-selection, J. Vis. Exp. (96) (2015), e52465.
[91] M. Wolfien, C. Rimmbach, U. Schmitz, J.J. Jung, S. Krebs, G. Steinhoff, R. David, O. Wolkenhauer, TRAPLINE: a standardized and automated pipeline for RNA sequencing data analysis, evaluation and annotation, BMC Bioinformatics 17 (2016) 21.
[92] M.R. Rosen, R.J. Myerburg, D.P. Francis, G.D. Cole, E. Marbán, Translating stem cell research to cardiac disease therapies: pitfalls and prospects for improvement, J. Am. Coll. Cardiol. 64 (9) (2014) 922–937.
[93] A.A. Matar, J.J. Chong, Stem cell therapy for cardiac dysfunction, Springerplus 3 (2014) 440.
[94] K.E. Hatzistergos, J.M. Hare, Cell therapy: targeting endogenous repair versus remuscularization, Circ. Res. 117 (8) (2015) 659–661.
[95] C. Stamm, H.-D. Kleine, Y.-H. Choi, S. Dunkelmann, J.-A. Lauffs, B. Lorenzen, A. David, A. Liebold, C. Nienaber, D. Zurakowski, M. Freund, G. Steinhoff,

Intramyocardial delivery of CD133+ bone marrow cells and coronary artery bypass grafting for chronic ischemic heart disease: safety and efficacy studies, J. Thorac. Cardiovasc. Surg. 133 (3) (2007) 717–725.
[96] Z. Wang, L. Wang, X. Su, J. Pu, M. Jiang, B. He, Rational transplant timing and dose of mesenchymal stromal cells in patients with acute myocardial infarction: a meta-analysis of randomized controlled trials, Stem Cell Res Ther 8 (1) (2017) 21.
[97] W. Yang, H. Zheng, Y. Wang, F. Lian, Z. Hu, S. Xue, Nesprin-1 has key roles in the process of mesenchymal stem cell differentiation into cardiomyocyte-like cells in vivo and in vitro, Mol. Med. Rep. 11 (1) (2015) 133–142.
[98] P. Li, L. Zhang, Exogenous Nkx2-5- or GATA-4-transfected rabbit bone marrow mesenchymal stem cells and myocardial cell co-culture on the treatment of myocardial infarction in rabbits, Mol. Med. Rep. 12 (2) (2015) 2607–2621.
[99] J. Li, K. Zhu, Y. Wang, J. Zheng, C. Guo, H. Lai, C. Wang, Combination of IGF1 gene manipulation and 5AZA treatment promotes differentiation of mesenchymal stem cells into cardiomyocyte-like cells, Mol. Med. Rep. 11 (2) (2015) 815–820.
[100] S. Mohanty, S. Bose, K.G. Jain, B. Bhargava, B. Airan, TGFβ1 contributes to cardiomyogenic-like differentiation of human bone marrow mesenchymal stem cells, Int. J. Cardiol. 163 (1) (2013) 93–99.
[101] Y. Feng, P. Yang, S. Luo, Z. Zhang, H. Li, P. Zhu, Z. Song, Shox2 influences mesenchymal stem cell fate in a co-culture model in vitro, Mol. Med. Rep. 14 (1) (2016) 637–642.
[102] F. Huang, L. Tang, Z.-f. Fang, X.-q. Hu, J.-y. Pan, S.-h. Zhou, miR-1-mediated induction of cardiogenesis in mesenchymal stem cells via downregulation of Hes-1, Biomed. Res. Int. 2013 (2013), 216286.
[103] Z. Yu, Y. Zou, J. Fan, C. Li, L. Ma, Notch1 is associated with the differentiation of human bone marrow-derived mesenchymal stem cells to cardiomyocytes, Mol. Med. Rep. 14 (6) (2016) 5065–5071.
[104] J. Hou, H. Long, C. Zhou, S. Zheng, H. Wu, T. Guo, Q. Wu, T. Zhong, T. Wang, Long noncoding RNA Braveheart promotes cardiogenic differentiation of mesenchymal stem cells in vitro, Stem Cell Res Ther 8 (1) (2017) 4.
[105] P.H. Carvalho, A.P.F. Daibert, B.S. Monteiro, B.S. Okano, J.L. Carvalho, Daise Nunes Queiroz da Cunha, L.S.C. Favarato, V.G. Pereira, L.E.F. Augusto, R.J.D. Carlo, Diferenciação de células-tronco mesenquimais derivadas do tecido adiposo em cardiomiócitos, Arq. Bras. Cardiol. 100 (1) (2013) 82–89.
[106] W. Wystrychowski, B. Patlolla, Y. Zhuge, E. Neofytou, R.C. Robbins, R.E. Beygui, Multipotency and cardiomyogenic potential of human adipose-derived stem cells from epicardium, pericardium, and omentum, Stem Cell Res Ther 7 (1) (2016) 84.
[107] S.-J. Gwak, S.H. Bhang, H.S. Yang, S.-S. Kim, D.-H. Lee, S.-H. Lee, B.-S. Kim, In vitro cardiomyogenic differentiation of adipose-derived stromal cells using transforming growth factor-beta1, Cell Biochem. Funct. 27 (3) (2009) 148–154.
[108] H. Nagata, M. Ii, E. Kohbayashi, M. Hoshiga, T. Hanafusa, M. Asahi, Cardiac adipose-derived stem cells exhibit high differentiation potential to cardiovascular cells in C57BL/6 mice, Stem Cells Transl. Med. 5 (2) (2016) 141–151.
[109] Y.S. Choi, G.J. Dusting, S. Stubbs, S. Arunothayaraj, X.L. Han, P. Collas, W.A. Morrison, R.J. Dilley, Differentiation of human adipose-derived stem cells into beating cardiomyocytes, J. Cell. Mol. Med. 14 (4) (2010) 878–889.
[110] T. Takahashi, T. Nagai, M. Kanda, M.-L. Liu, N. Kondo, A.T. Naito, T. Ogura, H. Nakaya, J.-K. Lee, I. Komuro, Y. Kobayashi, Regeneration of the cardiac conduction system by adipose tissue-derived stem cells, Circ. J. 79 (12) (2015) 2703–2712.
[111] I.-Y. Sung, H.-N. Son, I. Ullah, D. Bharti, J.-M. Park, Y.-C. Cho, J.-H. Byun, Y.-H. Kang, S.-J. Sung, J.-W. Kim, G.-J. Rho, B.-W. Park, Cardiomyogenic differentiation of human dental follicle-derived stem cells by suberoylanilide hydroxamic acid and their in vivo homing property, Int. J. Med. Sci. 13 (11) (2016) 841–852.
[112] E. Lopez-Ruiz, M. Peran, M. Picon-Ruiz, M.A. Garcia, E. Carrillo, M. Jimenez-Navarro, M.C. Hernandez, I. Prat, E. de Teresa, J.A. Marchal, Cardiomyogenic differentiation potential of human endothelial progenitor cells isolated from patients with myocardial infarction, Cytotherapy 16 (9) (2014) 1229–1237.
[113] D. Avitabile, A. Crespi, C. Brioschi, V. Parente, G. Toietta, P. Devanna, M. Baruscotti, S. Truffa, A. Scavone, F. Rusconi, A. Biondi, Y. D'Alessandra, E. Vigna, D. Difrancesco, M. Pesce, M.C. Capogrossi, A. Barbuti, Human cord blood CD34+ progenitor cells acquire functional cardiac properties through a cell fusion process, Am. J. Physiol. Heart Circ. Physiol. 300 (5) (2011) H1875–84.
[114] B.T. Freeman, N.A. Kouris, B.M. Ogle, Tracking fusion of human mesenchymal stem cells after transplantation to the heart, Stem Cells Transl. Med. 4 (6) (2015) 685–694.
[115] A.A. Karpov, D.V. Udalova, M.G. Pliss, M.M. Galagudza, Can the outcomes of mesenchymal stem cell-based therapy for myocardial infarction be improved? Providing weapons and armour to cells, Cell Prolif. (2016).
[116] T. Hosoda, H. Zheng, M. Cabral-da-Silva, F. Sanada, N. Ide-Iwata, B. Ogórek, J. Ferreira-Martins, C. Arranto, D. D'Amario, F. del Monte, K. Urbanek, D.A. D'Alessandro, R.E. Michler, P. Anversa, M. Rota, J. Kajstura, A. Leri, Human cardiac stem cell differentiation is regulated by a mircrine mechanism, Circulation 123 (12) (2011) 1287–1296.
[117] M.-J. Goumans, T.P. de Boer, A.M. Smits, L.W. van Laake, P. van Vliet, C.H.G. Metz, T.H. Korfage, K.P. Kats, R. Hochstenbach, G. Pasterkamp, M.C. Verhaar, M.A.G. van der Heyden, D. de Kleijn, C.L. Mummery, T.A.B. van Veen, J.P.G. Sluijter, P.A. Doevendans, TGF-beta1 induces efficient differentiation of human cardiomyocyte progenitor cells into functional cardiomyocytes in vitro, Stem Cell Res. 1 (2) (2007) 138–149.
[118] Joost P.G. Sluijter, A. van Mil, P. van Vliet, Corina H.G. Metz, J. Liu, P.A. Doevendans, M.-J. Goumans, MicroRNA-1 and -499 regulate differentiation and proliferation in human-derived cardiomyocyte progenitor cells, Arterioscler. Thromb. Vasc. Biol. 30 (4) (2010) 859–868.
[119] B. Oberwallner, A. Brodarac, P. Anic, T. Saric, K. Wassilew, K. Neef, Y.-H. Choi, C. Stamm, Human cardiac extracellular matrix supports myocardial lineage commitment of pluripotent stem cells, Eur. J. Cardiothorac. Surg. 47 (3) (2015) 416–425 (discussion 425).
[120] A.M. Wobus, G. Wallukat, J. Hescheler, Pluripotent mouse embryonic stem cells are able to differentiate into cardiomyocytes expressing chronotropic responses to adrenergic and cholinergic agents and Ca2+ channel blockers, Differentiation 48 (3) (1991) 173–182.
[121] H. Skottman, M. Mikkola, K. Lundin, C. Olsson, A.-M. Stromberg, T. Tuuri, T. Otonkoski, O. Hovatta, R. Lahesmaa, Gene expression signatures of seven individual human embryonic stem cell lines, Stem Cells 23 (9) (2005) 1343–1356.
[122] M.J. Abeyta, A.T. Clark, R.T. Rodriguez, M.S. Bodnar, R.A.R. Pera, M.T. Firpo, Unique gene expression signatures of independently-derived human embryonic stem cell lines, Hum. Mol. Genet. 13 (6) (2004) 601–608.
[123] T. Tavakoli, X. Xu, E. Derby, Y. Serebryakova, Y. Reid, M.S. Rao, M.P. Mattson, W. Ma, Self-renewal and differentiation capabilities are variable between human embryonic stem cell lines I3, I6 and BG01V, BMC Cell Biol. 10 (2009) 44.
[124] L.C. Laurent, I. Ulitsky, I. Slavin, H. Tran, A. Schork, R. Morey, C. Lynch, J.V. Harness, S. Lee, M.J. Barrero, S. Ku, M. Martynova, R. Semechkin, V. Galat, J. Gottesfeld, Izpisua Belmonte, Juan Carlos, C. Murry, H.S. Keirstead, H.-S. Park, U. Schmidt, A.L. Laslett, F.-J. Muller, C.M. Nievergelt, R. Shamir, J.F. Loring, Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture, Cell Stem Cell 8 (1) (2011) 106–118.
[125] A.M. Newman, J.B. Cooper, Lab-specific gene expression signatures in pluripotent stem cells, Cell Stem Cell 7 (2) (2010) 258–262.
[126] K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors, Cell 126 (4) (2006) 663–676.
[127] K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, S. Yamanaka, Induction of pluripotent stem cells from adult human fibroblasts by defined factors, Cell 131 (5) (2007) 861–872.
[128] R. Lister, M. Pelizzola, Y.S. Kida, R.D. Hawkins, J.R. Nery, G. Hon, J. Antosiewicz-Bourget, R. O'Malley, R. Castanon, S. Klugman, M. Downes, R. Yu, R. Stewart, B. Ren, J.A. Thomson, R.M. Evans, J.R. Ecker, Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells, Nature 471 (7336) (2011) 68–73.
[129] O. Bar-Nur, H.A. Russ, S. Efrat, N. Benvenisty, Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells, Cell Stem Cell 9 (1) (2011) 17–23.
[130] V. Sanchez-Freire, A.S. Lee, S. Hu, O.J. Abilez, P. Liang, F. Lan, B.C. Huber, S.-G. Ong, W.X. Hong, M. Huang, J.C. Wu, Effect of human donor cell source on differentiation and function of cardiac induced pluripotent stem cells, J. Am. Coll. Cardiol. 64 (5) (2014) 436–448.
[131] H. Xu, B.A. Yi, H. Wu, C. Bock, H. Gu, K.O. Lui, J.-H.C. Park, Y. Shao, A.K. Riley, I.J. Domian, E. Hu, R. Willette, J. Lepore, A. Meissner, Z. Wang, K.R. Chien, Highly efficient derivation of ventricular cardiomyocytes from induced pluripotent stem cells with a distinct epigenetic signature, Cell Res. 22 (1) (2012) 142–154.
[132] L. Warren, P.D. Manos, T. Ahfeldt, Y.-H. Loh, H. Li, F. Lau, W. Ebina, P.K. Mandal, Z.D. Smith, A. Meissner, G.Q. Daley, A.S. Brack, J.J. Collins, C. Cowan, T.M. Schlaeger, D.J. Rossi, Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA, Cell Stem Cell 7 (5) (2010) 618–630.
[133] A. Mehta, V. Verma, M. Nandihalli, C.J.A. Ramachandra, G.L. Sequiera, Y. Sudibyo, Y. Chung, W. Sun, W. Shim, A systemic evaluation of cardiac differentiation from mRNA reprogrammed human induced pluripotent stem cells, PLoS One 9 (7) (2014), e103485.
[134] J. Zhang, G.F. Wilson, A.G. Soerens, C.H. Koonce, J. Yu, S.P. Palecek, J.A. Thomson, T.J. Kamp, Functional cardiomyocytes derived from human induced pluripotent stem cells, Circ. Res. 104 (4) (2009) e30–41.
[135] J.-H. Jung, X. Fu, P.C. Yang, Exosomes generated from iPSC-derivatives: new direction for stem cell therapy in human heart diseases, Circ. Res. 120 (2) (2017) 407–417.
[136] V.F.M. Segers, R.T. Lee, Stem-cell therapy for cardiac disease, Nature 451 (7181) (2008) 937–942.
[137] H. Okano, M. Nakamura, K. Yoshida, Y. Okada, O. Tsuji, S. Nori, E. Ikeda, S. Yamanaka, K. Miura, Steps toward safe cell therapy using induced pluripotent stem cells, Circ. Res. 112 (3) (2013) 523–533.
[138] C.C. Veerman, G. Kosmidis, C.L. Mummery, S. Casini, A.O. Verkerk, M. Bellin, Immaturity of human stem-cell-derived cardiomyocytes in culture: fatal flaw or soluble problem? Stem Cells Dev. 24 (9) (2015) 1035–1052.
[139] X. Yang, L. Pabon, C.E. Murry, Engineering adolescence: maturation of human pluripotent stem cell-derived cardiomyocytes, Circ. Res. 114 (3) (2014) 511–523.
[140] R. David, C. Brenner, J. Stieber, F. Schwarz, S. Brunner, M. Vollmer, E. Mentele, J. Müller-Höcker, S. Kitajima, H. Lickert, R. Rupp, W.-M. Franz, MesP1 drives vertebrate cardiovascular differentiation through Dkk-1-mediated blockade of Wnt-signalling, Nat. Cell Biol. 10 (3) (2008) 338–345.
[141] R. David, V.B. Jarsch, F. Schwarz, P. Nathan, M. Gegg, H. Lickert, W.-M. Franz, Induction of MesP1 by Brachyury(T) generates the common multipotent cardiovascular stem cell, Cardiovasc. Res. 92 (1) (2011) 115–122.
[142] A. Bondue, G. Lapouge, C. Paulissen, C. Semeraro, M. Iacovino, M. Kyba, C. Blanpain, Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification, Cell Stem Cell 3 (1) (2008) 69–84.
[143] A. Bondue, S. Tannler, G. Chiapparo, S. Chabab, M. Ramialison, C. Paulissen, B. Beck, R. Harvey, C. Blanpain, Defining the earliest step of cardiovascular progenitor specification during embryonic stem cell differentiation, J. Cell Biol. 192 (5) (2011) 751–765.
[144] Y. Liu, L. Chen, A.D. Diaz, A. Benham, X. Xu, C.S. Wijaya, F. Fa'ak, W. Luo, B. Soibam, A. Azares, W. Yu, Q. Lyu, M.D. Stewart, P. Gunaratne, A. Cooney, B.K. McConnell, R.J.

Schwartz, Mesp1 marked cardiac progenitor cells repair infarcted mouse hearts, Sci. Rep. 6 (2016) 31457.
[145] X. Shen, B. Soibam, A. Benham, X. Xu, M. Chopra, X. Peng, W. Yu, W. Bao, R. Liang, A. Azares, P. Liu, P.H. Gunaratne, M. Mercola, A.J. Cooney, R.J. Schwartz, Y. Liu, miR-322/-503 cluster is expressed in the earliest cardiac progenitor cells and drives cardiomyocyte specification, Proc. Natl. Acad. Sci. U. S. A. 113 (34) (2016) 9551–9556.
[146] S.C. den Hartogh, K. Wolstencroft, C.L. Mummery, R. Passier, A comprehensive gene expression analysis at sequential stages of in vitro cardiac differentiation from isolated MESP1-expressing-mesoderm progenitors, Sci. Rep. 6 (2016) 19386.
[147] S.S.-K. Chan, X. Shi, A. Toyama, R.W. Arpke, A. Dandapat, M. Iacovino, J. Kang, G. Le, H.R. Hagen, D.J. Garry, M. Kyba, Mesp1 patterns mesoderm into cardiac, hematopoietic, or skeletal myogenic progenitors in a context-dependent manner, Cell Stem Cell 12 (5) (2013) 587–601.
[148] S.S.-K. Chan, H.H.W. Chan, M. Kyba, Heterogeneity of Mesp1 + mesoderm revealed by single-cell RNA-seq, Biochem. Biophys. Res. Commun. 474 (3) (2016) 469–475.
[149] A. Okada, K. Tashiro, T. Yamaguchi, K. Kawabata, Selective differentiation into hematopoietic and cardiac cells from pluripotent stem cells based on the expression of cell surface markers, Methods Mol. Biol. 1341 (2016) 181–195.
[150] I. Pacheco-Leyva, A.C. Matias, D.V. Oliveira, J.M.A. Santos, R. Nascimento, E. Guerreiro, A.C. Michell, A.M. van de Vrugt, G. Machado-Oliveira, G. Ferreira, I. Domian, J. Braganca, CITED2 cooperates with ISL1 and promotes cardiac differentiation of mouse embryonic stem cells, Stem Cell Rep. 7 (6) (2016) 1037–1049.
[151] T. Kotoku, K. Kosaka, M. Nishio, Y. Ishida, M. Kawaichi, E. Matsuda, CIBZ regulates mesodermal and cardiac differentiation of by suppressing T and Mesp1 expression in mouse embryonic stem cells, Sci. Rep. 6 (2016) 34188.
[152] F. Rabiee, M. Forouzanfar, F. Ghazvini Zadegan, S. Tanhaei, K. Ghaedi, M. Motovali Bashi, H. Baharvand, M.H. Nasr-Esfahani, Induced expression of Fndc5 significantly increased cardiomyocyte differentiation rate of mouse embryonic stem cells, Gene 551 (2) (2014) 127–137.
[153] C.A. Klattenhoff, J.C. Scheuermann, L.E. Surface, R.K. Bradley, P.A. Fields, M.L. Steinhauser, H. Ding, V.L. Butty, L. Torrey, S. Haas, R. Abo, M. Tabebordbar, R.T. Lee, C.B. Burge, L.A. Boyer, Braveheart, a long noncoding RNA required for cardiovascular lineage commitment, Cell 152 (3) (2013) 570–583.
[154] V. Ionta, W. Liang, E.H. Kim, R. Rafie, A. Giacomello, E. Marban, H.C. Cho, SHOX2 overexpression favors differentiation of embryonic stem cells into cardiac pacemaker cells, improving biological pacing ability, Stem Cell Rep. 4 (1) (2015) 129–142.
[155] T. Dorn, A. Goedel, J.T. Lam, J. Haas, Q. Tian, F. Herrmann, K. Bundschu, G. Dobreva, M. Schiemann, R. Dirschinger, Y. Guo, S.J. Kuhl, D. Sinnecker, P. Lipp, K.-L. Laugwitz, M. Kuhl, A. Moretti, Direct nkx2-5 transcriptional repression of isl1 controls cardiomyocyte subtype identity, Stem Cells 33 (4) (2015) 1113–1129.
[156] S. Bhattacharya, P.W. Burridge, E.M. Kropp, S.L. Chuppa, W.-M. Kwok, J.C. Wu, K.R. Boheler, R.L. Gundry, High efficiency differentiation of human pluripotent stem cells to cardiomyocytes and characterization by flow cytometry, J. Vis. Exp. (91) (2014), 52010.
[157] V.C. Chen, J. Ye, P. Shukla, G. Hua, D. Chen, Z. Lin, J.-c. Liu, J. Chai, J. Gold, J. Wu, D. Hsu, L.A. Couture, Development of a scalable suspension culture for cardiac differentiation from human pluripotent stem cells, Stem Cell Res. 15 (2) (2015) 365–375.
[158] M. Fuerstenau-Sharp, M.E. Zimmermann, K. Stark, N. Jentsch, M. Klingenstein, M. Drzymalski, S. Wagner, L.S. Maier, U. Hehr, A. Baessler, M. Fischer, C. Hengstenberg, Generation of highly purified human cardiomyocytes from peripheral blood mononuclear cell-derived induced pluripotent stem cells, PLoS One 10 (5) (2015), e0126596.
[159] S.I. Protze, J. Liu, U. Nussinovitch, L. Ohana, P.H. Backx, L. Gepstein, G.M. Keller, Sinoatrial node cardiomyocytes derived from human pluripotent cells function as a biological pacemaker, Nat. Biotechnol. 35 (1) (2017) 56–68.
[160] F. Pei, J. Jiang, S. Bai, H. Cao, L. Tian, Y. Zhao, C. Yang, H. Dong, Y. Ma, Chemical-defined and albumin-free generation of human atrial and ventricular myocytes from human pluripotent stem cells, Stem Cell Res. 19 (2017) 94–103.
[161] S. Mazzotta, C. Neves, R.J. Bonner, A.S. Bernardo, K. Docherty, S. Hoppler, Distinctive roles of canonical and noncanonical Wnt signaling in human embryonic cardiomyocyte development, Stem Cell Rep. 7 (4) (2016) 764–776.
[162] M.C. Engels, K. Rajarajan, R. Feistritzer, A. Sharma, U.B. Nielsen, M.J. Schalij, A.A.F. de Vries, D.A. Pijnappels, S.M. Wu, Insulin-like growth factor promotes cardiac lineage induction in vitro by selective expansion of early mesoderm, Stem Cells 32 (6) (2014) 1493–1502.
[163] V.C. Chen, S.M. Couture, J. Ye, Z. Lin, G. Hua, H.-I.P. Huang, J. Wu, D. Hsu, M.K. Carpenter, L.A. Couture, Scalable GMP compliant suspension culture system for human ES cells, Stem Cell Res. 8 (3) (2012) 388–402.
[164] N.J. Palpant, L. Pabon, C.E. Friedman, M. Roberts, B. Hadland, R.J. Zaunbrecher, I. Bernstein, Y. Zheng, C.E. Murry, Generating high-purity cardiac and endothelial derivatives from patterned mesoderm using human pluripotent stem cells, Nat. Protoc. 12 (1) (2017) 15–31.
[165] I. Kokkinopoulos, H. Ishida, R. Saba, S. Coppen, K. Suzuki, K. Yashiro, Cardiomyocyte differentiation from mouse embryonic stem cells using a simple and defined protocol, Dev. Dyn. 245 (2) (2016) 157–165.
[166] S.J. Kattman, A.D. Witty, M. Gagliardi, N.C. Dubois, M. Niapour, A. Hotta, J. Ellis, G. Keller, Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines, Cell Stem Cell 8 (2) (2011) 228–240.
[167] X.-L. Li, D.I. Zeng, Y. Chen, L. Ding, W.-J. Li, T. Wei, D.-B. Ou, S. Yan, B. Wang, Q.-S. Zheng, Role of alpha- and beta-adrenergic receptors in cardiomyocyte differentiation from murine-induced pluripotent stem cells, Cell Prolif. 50 (1) (2017).
[168] H. Uosaki, A. Magadum, K. Seo, H. Fukushima, A. Takeuchi, Y. Nakagawa, K.W. Moyes, G. Narazaki, K. Kuwahara, M. Laflamme, S. Matsuoka, N. Nakatsuji, K. Nakao, C. Kwon, D.A. Kass, F.B. Engel, J.K. Yamashita, Identification of chemicals inducing cardiomyocyte proliferation in developmental stage-specific manner with pluripotent stem cells, Circ. Cardiovasc. Genet. 6 (6) (2013) 624–633.
[169] D. Ivanyuk, G. Budash, Y. Zheng, J.A. Gaspar, U. Chaudhari, A. Fatima, S. Bahmanpour, V.K. Grin, A.G. Popandopulo, A. Sachinidis, J. Hescheler, T. Saric, Ascorbic acid-induced cardiac differentiation of murine pluripotent stem cells: transcriptional profiling and effect of a small molecule synergist of Wnt/beta-catenin signaling pathway, Cell. Physiol. Biochem. 36 (2) (2015) 810–830.
[170] T. Aguado, F.J. Gutierrez, E. Aix, R.P. Schneider, G. Giovinnazo, M.A. Blasco, I. Flores, Telomere length defines the cardiomyocyte differentiation potency of mouse induced pluripotent stem cells, Stem Cells 35 (2) (2016) 362–373.
[171] G. Kensah, A. Roa Lara, J. Dahlmann, R. Zweigerdt, K. Schwanke, J. Hegermann, D. Skvorc, A. Gawol, A. Azizian, S. Wagner, L.S. Maier, A. Krause, G. Drager, M. Ochs, A. Haverich, I. Gruh, U. Martin, Murine and human pluripotent stem cell-derived cardiac bodies form contractile myocardial tissue in vitro, Eur. Heart J. 34 (15) (2013) 1134–1146.
[172] M.G. Klug, M.H. Soonpaa, G.Y. Koh, L.J. Field, Genetically selected cardiomyocytes from differentiating embronic stem cells form stable intracardiac grafts, J. Clin. Invest. 98 (1) (1996) 216–224.
[173] V.A. Maltabe, E. Barka, M. Kontonika, D. Florou, M. Kouvara-Pritsouli, M. Roumpi, S. Agathopoulos, T.M. Kolettis, P. Kouklis, Isolation of an ES-derived cardiovascular multipotent cell population based on VE-cadherin promoter activity, Stem Cells Int. 2016 (2016) 8305624.
[174] K. Ban, B. Wile, S. Kim, H.-J. Park, J. Byun, K.-W. Cho, T. Saafir, M.-K. Song, S.P. Yu, M. Wagner, G. Bao, Y.-S. Yoon, Purification of cardiomyocytes from differentiating pluripotent stem cells using molecular beacons that target cardiomyocyte-specific mRNA, Circulation 128 (17) (2013) 1897–1909.
[175] R. Jha, B. Wile, Q. Wu, A.H. Morris, K.O. Maher, M.B. Wagner, G. Bao, C. Xu, Molecular beacon-based detection and isolation of working-type cardiomyocytes derived from human pluripotent stem cells, Biomaterials 50 (2015) 176–185.
[176] F. Hattori, H. Chen, H. Yamashita, S. Tohyama, Y.-S. Satoh, S. Yuasa, W. Li, H. Yamakawa, T. Tanaka, T. Onitsuka, K. Shimoji, Y. Ohno, T. Egashira, R. Kaneda, M. Murata, K. Hidaka, T. Morisaki, E. Sasaki, T. Suzuki, M. Sano, S. Makino, S. Oikawa, K. Fukuda, Nongenetic method for purifying stem cell-derived cardiomyocytes, Nat. Methods 7 (1) (2010) 61–66.
[177] N.C. Dubois, A.M. Craft, P. Sharma, D.A. Elliott, E.G. Stanley, A.G. Elefanty, A. Gramolini, G. Keller, SIRPA is a specific cell-surface marker for isolating cardiomyocytes derived from human pluripotent stem cells, Nat. Biotechnol. 29 (11) (2011) 1011–1018.
[178] D. van Hoof, W. Dormeyer, S.R. Braam, R. Passier, J. Monshouwer-Kloots, D. Ward-van Oostwaard, Albert J.R. Heck, J. Krijgsveld, C.L. Mummery, Identification of cell surface proteins for antibody-based selection of human embryonic stem cell-derived cardiomyocytes, J. Proteome Res. 9 (3) (2010) 1610–1618.
[179] H. Uosaki, H. Fukushima, A. Takeuchi, S. Matsuoka, N. Nakatsuji, S. Yamanaka, J.K. Yamashita, Efficient and scalable purification of cardiomyocytes from human embryonic and induced pluripotent stem cells by VCAM1 surface expression, PLoS One 6 (8) (2011), e23657.
[180] G.D. Lopaschuk, J.S. Jaswal, Energy metabolic phenotype of the cardiomyocyte during development, differentiation, and postnatal maturation, J. Cardiovasc. Pharmacol. 56 (2) (2010) 130–140.
[181] S. Tohyama, F. Hattori, M. Sano, T. Hishiki, Y. Nagahata, T. Matsuura, H. Hashimoto, T. Suzuki, H. Yamashita, Y. Satoh, T. Egashira, T. Seki, N. Muraoka, H. Yamakawa, Y. Ohgino, T. Tanaka, M. Yoichi, S. Yuasa, M. Murata, M. Suematsu, K. Fukuda, Distinct metabolic flow enables large-scale purification of mouse and human pluripotent stem cell-derived cardiomyocytes, Cell Stem Cell 12 (1) (2013) 127–137.
[182] N. Hemmi, S. Tohyama, K. Nakajima, H. Kanazawa, T. Suzuki, F. Hattori, T. Seki, Y. Kishino, A. Hirano, M. Okada, R. Tabei, R. Ohno, C. Fujita, T. Haruna, S. Yuasa, M. Sano, J. Fujita, K. Fukuda, A massive suspension culture system with metabolic purification for human pluripotent stem cell-derived cardiomyocytes, Stem Cells Transl. Med. 3 (12) (2014) 1473–1483.
[183] K.T. Kuppusamy, D.C. Jones, H. Sperber, A. Madan, K.A. Fischer, M.L. Rodriguez, L. Pabon, W.-Z. Zhu, N.L. Tulloch, X. Yang, N.J. Sniadecki, M.A. Laflamme, W.L. Ruzzo, C.E. Murry, H. Ruohola-Baker, Let-7 family of microRNA is required for maturation and adult-like metabolism in stem cell-derived cardiomyocytes, Proc. Natl. Acad. Sci. U. S. A. 112 (21) (2015) E2785–94.
[184] James J.H. Chong, X. Yang, C.W. Don, E. Minami, Y.-W. Liu, J.J. Weyers, W.M. Mahoney, B. van Biber, S.M. Cook, N.J. Palpant, J.A. Gantz, J.A. Fugate, V. Muskheli, G.M. Gough, K.W. Vogel, C.A. Astley, C.E. Hotchkiss, A. Baldessari, L. Pabon, H. Reinecke, E.A. Gill, V. Nelson, H.-P. Kiem, M.A. Laflamme, C.E. Murry, Human embryonic-stem-cell-derived cardiomyocytes regenerate non-human primate hearts, Nature 510 (7504) (2014) 273–277.
[185] K. Fynes, R. Tostoes, L. Ruban, B. Weil, C. Mason, F.S. Veraitch, The differential effects of 2% oxygen preconditioning on the subsequent differentiation of mouse and human pluripotent stem cells, Stem Cells Dev. 23 (16) (2014) 1910–1922.
[186] C. Rao, T. Prodromakis, L. Kolker, U.A.R. Chaudhry, T. Trantidou, A. Sridhar, C. Weekes, P. Camelliti, S.E. Harding, A. Darzi, M.H. Yacoub, T. Athanasiou, C.M. Terracciano, The effect of microgrooved culture substrates on calcium cycling of cardiac myocytes derived from human induced pluripotent stem cells, Biomaterials 34 (10) (2013) 2399–2411.
[187] M. Khan, Y. Xu, S. Hua, J. Johnson, A. Belevych, P.M.L. Janssen, S. Gyorke, J. Guan, M.G. Angelos, Evaluation of changes in morphology and function of human induced pluripotent stem cell derived cardiomyocytes (HiPSC-CMs) cultured on an aligned-nanofiber cardiac patch, PLoS One 10 (5) (2015), e0126338.

[188] L.B. Hazeltine, M.G. Badur, X. Lian, A. Das, W. Han, S.P. Palecek, Temporal impact of substrate mechanics on differentiation of human embryonic stem cells to cardiomyocytes, Acta Biomater. 10 (2) (2014) 604–612.

[189] J. Pasquier, R. Gupta, D. Rioult, J. Hoarau-Vechot, R. Courjaret, K. Machaca, J. Al Suwaidi, E.G. Stanley, S. Rafii, D.A. Elliott, C. Abi Khalil, A. Rafii, Coculturing with endothelial cells promotes in vitro maturation and electrical coupling of human embryonic stem cell-derived cardiomyocytes, J. Heart Lung Transplant. 36 (6) (2017) 684–693.

[190] M. Ojala, K. Rajala, M. Pekkanen-Mattila, M. Miettinen, H. Huhtala, K. Aalto-Setala, Culture conditions affect cardiac differentiation potential of human pluripotent stem cells, PLoS One 7 (10) (2012), e48659.

[191] Y. Amano, A. Nishiguchi, M. Matsusaki, H. Iseoka, S. Miyagawa, Y. Sawa, M. Seo, T. Yamaguchi, M. Akashi, Development of vascularized iPSC derived 3D-cardiomyocyte tissues by filtration layer-by-layer technique and their application for pharmaceutical assays, Acta Biomater. 33 (2016) 110–121.

[192] M.T. Valarmathi, J.W. Fuseler, J.M. Davis, R.L. Price, A novel human tissue-engineered 3-D functional vascularized cardiac muscle construct, Front. Cell Dev. Biol. 5 (2017) 2.

[193] A. Eder, I. Vollert, A. Hansen, T. Eschenhagen, Human engineered heart tissue as a model system for drug testing, Adv. Drug Deliv. Rev. 96 (2016) 214–224.

[194] C.Y. Ivashchenko, G.C. Pipes, I.M. Lozinskaya, Z. Lin, X. Xiaoping, S. Needle, E.T. Grygielko, E. Hu, J.R. Toomey, J.J. Lepore, R.N. Willette, Human-induced pluripotent stem cell-derived cardiomyocytes exhibit temporal changes in phenotype, Am. J. Physiol. Heart Circ. Physiol. 305 (6) (2013) H913–22.

[195] C.E. Rupert, H.H. Chang, K.L.K. Coulombe, Hypertrophy changes 3D shape of hiPSC-cardiomyocytes: implications for cellular maturation in regenerative medicine, Cell. Mol. Bioeng. 10 (1) (2017) 54–62.

[196] S. Funakoshi, K. Miki, T. Takaki, C. Okubo, T. Hatani, K. Chonabayashi, M. Nishikawa, I. Takei, A. Oishi, M. Narita, M. Hoshijima, T. Kimura, S. Yamanaka, Y. Yoshida, Enhanced engraftment, proliferation, and therapeutic potential in heart using optimized human iPSC-derived cardiomyocytes, Sci. Rep. 6 (2016) 19111.

[197] T. Kitaguchi, Y. Moriyama, T. Taniguchi, A. Ojima, H. Ando, T. Uda, K. Otabe, M. Oguchi, S. Shimizu, H. Saito, M. Morita, A. Toratani, M. Asayama, W. Yamamoto, E. Matsumoto, D. Saji, H. Ohnaka, K. Tanaka, I. Washio, N. Miyamoto, CSAHi study: evaluation of multi-electrode array in combination with human iPS cell-derived cardiomyocytes to predict drug-induced QT prolongation and arrhythmia—effects of 7 reference compounds at 10 facilities, J. Pharmacol. Toxicol. Methods 78 (2016) 93–102.

[198] T. Kitaguchi, Y. Moriyama, T. Taniguchi, S. Maeda, H. Ando, T. Uda, K. Otabe, M. Oguchi, S. Shimizu, H. Saito, A. Toratani, M. Asayama, W. Yamamoto, E. Matsumoto, D. Saji, H. Ohnaka, N. Miyamoto, CSAHi study: detection of drug-induced ion channel/receptor responses, QT prolongation, and arrhythmia using multi-electrode arrays in combination with human induced pluripotent stem cell-derived cardiomyocytes, J. Pharmacol. Toxicol. Methods 85 (2017) 73–81.

[199] L.G.J. Tertoolen, S.R. Braam, B.J. van Meer, R. Passier, C.L. Mummery, Interpretation of field potentials measured on a multi electrode array in pharmacological toxicity screening on primary and human pluripotent stem cell-derived cardiomyocytes, Biochem. Biophys. Res. Commun. S0006-291X (17) (2017) 30219-X.

[200] M.P. Hortigon-Vinagre, V. Zamora, F.L. Burton, J. Green, G.A. Gintant, G.L. Smith, The use of ratiometric fluorescence measurements of the voltage sensitive dye Di-4-ANEPPS to examine action potential characteristics and drug effects on human induced pluripotent stem cell-derived cardiomyocytes, Toxicol. Sci. 154 (2) (2016) 320–331.

[201] A. Brodarac, T. Saric, B. Oberwallner, S. Mahmoodzadeh, K. Neef, J. Albrecht, K. Burkert, M. Oliverio, F. Nguemo, Y.-H. Choi, W.F. Neiss, I. Morano, J. Hescheler, C. Stamm, Susceptibility of murine induced pluripotent stem cell-derived cardiomyocytes to hypoxia and nutrient deprivation, Stem Cell Res Ther 6 (2015) 83.

[202] H. Ding, X. Xu, X. Qin, C. Yang, Q. Feng, Resveratrol promotes differentiation of mouse embryonic stem cells to cardiomyocytes, Cardiovasc. Ther. 34 (4) (2016) 283–289.

[203] N. Cao, Z. Liu, Z. Chen, J. Wang, T. Chen, X. Zhao, Y. Ma, L. Qin, J. Kang, B. Wei, L. Wang, Y. Jin, H.-T. Yang, Ascorbic acid enhances the cardiac differentiation of induced pluripotent stem cells through promoting the proliferation of cardiac progenitor cells, Cell Res. 22 (1) (2012) 219–236.

[204] T. Kamakura, T. Makiyama, K. Sasaki, Y. Yoshida, Y. Wuriyanghai, J. Chen, T. Hattori, S. Ohno, T. Kita, M. Horie, S. Yamanaka, T. Kimura, Ultrastructural maturation of human-induced pluripotent stem cell-derived cardiomyocytes in a long-term culture, Circ. J. 77 (5) (2013) 1307–1314.

[205] T.L. Medley, M. Furtado, N.T. Lam, R. Idrizi, D. Williams, P.J. Verma, M. Costa, D.M. Kaye, Effect of oxygen on cardiac differentiation in mouse iPS cells: role of hypoxia inducible factor-1 and Wnt/beta-catenin signaling, PLoS One 8 (11) (2013), e80280.

[206] X. Yang, M. Rodriguez, L. Pabon, K.A. Fischer, H. Reinecke, M. Regnier, N.J. Sniadecki, H. Ruohola-Baker, C.E. Murry, Tri-iodo-L-thyronine promotes the maturation of human cardiomyocytes-derived from induced pluripotent stem cells, J. Mol. Cell. Cardiol. 72 (2014) 296–304.

[207] O. Iglesias-Garcia, S. Baumgartner, L. Macri-Pellizzeri, J.R. Rodriguez-Madoz, G. Abizanda, E. Guruceaga, E. Albiasu, D. Corbacho, C. Benavides-Vallve, M. Soriano-Navarro, S. Gonzalez-Granero, J.J. Gavira, B. Krausgrill, M. Rodriguez-Manero, J.M. Garcia-Verdugo, C. Ortiz-de-Solorzano, M. Halbach, J. Hescheler, B. Pelacho, F. Prosper, Neuregulin-1beta induces mature ventricular cardiac differentiation from induced pluripotent stem cells contributing to cardiac tissue repair, Stem Cells Dev. 24 (4) (2015) 484–496.

[208] A. Kochegarov, A. Moses-Arms, L.F. Lemanski, A fetal human heart cardiac-inducing RNA (CIR) promotes the differentiation of stem cells into cardiomyocytes, In Vitro Cell. Dev. Biol. Anim. 51 (7) (2015) 739–748.

[209] H. Wang, Y. Xi, Y. Zheng, X. Wang, A.J. Cooney, Generation of electrophysiologically functional cardiomyocytes from mouse induced pluripotent stem cells, Stem Cell Res. 16 (2) (2016) 522–530.

[210] R.L. Davis, H. Weintraub, A.B. Lassar, Expression of a single transfected cDNA converts fibroblasts to myoblasts, Cell 51 (6) (1987) 987–1000.

[211] F. Hausburg, R. David, Cell programming for future regenerative medicine, in: G. Steinhoff (Ed.), Regenerative Medicine — From Protocol to Patient: 2. Stem Cell Science and Technology, 3rd ed.Springer International Publishing, Cham, s.l. 2016, pp. 389–424.

[212] R.S. Nagalingam, H.A. Safi, M.P. Czubryt, Gaining myocytes or losing fibroblasts: challenges in cardiac fibroblast reprogramming for infarct repair, J. Mol. Cell. Cardiol. 93 (2016) 108–114.

[213] J.K. Lighthouse, E.M. Small, Transcriptional control of cardiac fibroblast plasticity, J. Mol. Cell. Cardiol. 91 (2016) 52–60.

[214] Y. Gao, M. Chu, J. Hong, J. Shang, D.I. Xu, Hypoxia induces cardiac fibroblast proliferation and phenotypic switch: a role for caveolae and caveolin-1/PTEN mediated pathway, J. Thorac. Dis. 6 (10) (2014) 1458–1468.

[215] T. Moore-Morris, P. Cattaneo, M. Puceat, S.M. Evans, Origins of cardiac fibroblasts, J. Mol. Cell. Cardiol. 91 (2016) 1–5.

[216] H. Zhou, M.E. Dickson, M.S. Kim, R. Bassel-Duby, E.N. Olson, Akt1/protein kinase B enhances transcriptional reprogramming of fibroblasts to functional cardiomyocytes, Proc. Natl. Acad. Sci. U. S. A. 112 (38) (2015) 11864–11,869.

[217] Y. Zhao, P. Londono, Y. Cao, E.J. Sharpe, C. Proenza, R. O'Rourke, K.L. Jones, M.Y. Jeong, L.A. Walker, P.M. Buttrick, T.A. McKinsey, K. Song, High-efficiency reprogramming of fibroblasts into cardiomyocytes requires suppression of pro-fibrotic signalling, Nat. Commun. 6 (2015) 8243.

[218] Y. Fu, C. Huang, X. Xu, H. Gu, Y. Ye, C. Jiang, Z. Qiu, X. Xie, Direct reprogramming of mouse fibroblasts into cardiomyocytes with chemical cocktails, Cell Res. 25 (9) (2015) 1013–1024.

[219] H. Hirai, N. Katoku-Kikyo, S.A. Keirstead, N. Kikyo, Accelerated direct reprogramming of fibroblasts into cardiomyocyte-like cells with the MyoD transactivation domain, Cardiovasc. Res. 100 (1) (2013) 105–113.

[220] H. Hirai, N. Kikyo, Inhibitors of suppressive histone modification promote direct reprogramming of fibroblasts to cardiomyocyte-like cells, Cardiovasc. Res. 102 (1) (2014) 188–190.

[221] J.-D. Fu, N.R. Stone, L. Liu, C.I. Spencer, L. Qian, Y. Hayashi, P. Delgado-Olguin, S. Ding, B.G. Bruneau, D. Srivastava, Direct reprogramming of human fibroblasts toward a cardiomyocyte-like state, Stem Cell Rep. 1 (3) (2013) 235–247.

[222] Y.-J. Nam, K. Song, X. Luo, E. Daniel, K. Lambeth, K. West, J.A. Hill, J.M. DiMaio, L.A. Baker, R. Bassel-Duby, E.N. Olson, Reprogramming of human fibroblasts toward a cardiac fate, Proc. Natl. Acad. Sci. U. S. A. 110 (14) (2013) 5588–5593.

[223] J.X. Chen, M. Krane, M.-A. Deutsch, L. Wang, M. Rav-Acha, S. Gregoire, M.C. Engels, K. Rajarajan, R. Karra, E.D. Abel, J.C. Wu, D. Milan, S.M. Wu, Inefficient reprogramming of fibroblasts into cardiomyocytes using Gata4, Mef2c, and Tbx5, Circ. Res. 111 (1) (2012) 50–55.

[224] L. Qian, Y. Huang, C.I. Spencer, A. Foley, V. Vedantham, L. Liu, S.J. Conway, J.-D. Fu, D. Srivastava, In vivo reprogramming of murine cardiac fibroblasts into induced cardiomyocytes, Nature 485 (7400) (2012) 593–598.

[225] K. Inagawa, K. Miyamoto, H. Yamakawa, N. Muraoka, T. Sadahiro, T. Umei, R. Wada, Y. Katsumata, R. Kaneda, K. Nakade, C. Kurihara, Y. Obata, K. Miyake, K. Fukuda, M. Ieda, Induction of cardiomyocyte-like cells in infarct hearts by gene transfer of Gata4, Mef2c, and Tbx5, Circ. Res. 111 (9) (2012) 1147–1156.

[226] D. Srivastava, Making or breaking the heart: from lineage determination to morphogenesis, Cell 126 (6) (2006) 1037–1048.

[227] M. Ieda, J.-D. Fu, P. Delgado-Olguin, V. Vedantham, Y. Hayashi, B.G. Bruneau, D. Srivastava, Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors, Cell 142 (3) (2010) 375–386.

[228] L. Wang, Z. Liu, C. Yin, H. Asfour, O. Chen, Y. Li, N. Bursac, J. Liu, L. Qian, Stoichiometry of Gata4, Mef2c, and Tbx5 influences the efficiency and quality of induced cardiac myocyte reprogramming, Circ. Res. 116 (2) (2015) 237–244.

[229] L. Qian, E.C. Berry, J.-D. Fu, M. Ieda, D. Srivastava, Reprogramming of mouse fibroblasts into cardiomyocyte-like cells in vitro, Nat. Protoc. 8 (6) (2013) 1204–1215.

[230] T.M.A. Mohamed, N.R. Stone, E.C. Berry, E. Radzinsky, Y. Huang, K. Pratt, Y.-S. Ang, P. Yu, H. Wang, S. Tang, S. Magnitsky, S. Ding, K.N. Ivey, D. Srivastava, Chemical enhancement of in vitro and in vivo direct cardiac reprogramming, Circulation 135 (10) (2017) 978–995.

[231] Y. Zhou, L. Wang, H.R. Vaseghi, Z. Liu, R. Lu, S. Alimohamadi, C. Yin, J.-D. Fu, G.G. Wang, J. Liu, L. Qian, Bmi1 is a key epigenetic barrier to direct cardiac reprogramming, Cell Stem Cell 18 (3) (2016) 382–395.

[232] J.A. Kamps, G. Krenning, Micromanaging cardiac regeneration: targeted delivery of microRNAs for cardiac repair and regeneration, World J. Cardiol. 8 (2) (2016) 163–179.

[233] J.-D. Fu, S.N. Rushing, D.K. Lieu, C.W. Chan, C.-W. Kong, L. Geng, K.D. Wilson, N. Chiamvimonvat, K.R. Boheler, J.C. Wu, G. Keller, R.J. Hajjar, R.A. Li, Distinct roles of microRNA-1 and -499 in ventricular specification and functional maturation of human embryonic stem cell-derived cardiomyocytes, PLoS One 6 (11) (2011), e27417.

[234] K.D. Wilson, S. Hu, S. Venkatasubrahmanyam, J.-D. Fu, N. Sun, O.J. Abilez, Joshua J.A. Baugh, F. Jia, Z. Ghosh, R.A. Li, A.J. Butte, J.C. Wu, Dynamic microRNA expression programs during cardiac differentiation of human embryonic stem cells: role for miR-499, Circ. Cardiovasc. Genet. 3 (5) (2010) 426–435.

[235] T.M. Jayawardena, B. Egemnazarov, E.A. Finch, L. Zhang, J.A. Payne, K. Pandya, Z. Zhang, P. Rosenberg, M. Mirotsou, V.J. Dzau, MicroRNA-mediated in vitro and in vivo direct reprogramming of cardiac fibroblasts to cardiomyocytes, Circ. Res. 110 (11) (2012) 1465–1473.

[236] T. Jayawardena, M. Mirotsou, V.J. Dzau, Direct reprogramming of cardiac fibroblasts to cardiomyocytes using microRNAs, Methods Mol. Biol. 1150 (2014) 263–272.
[237] T.M. Jayawardena, E.A. Finch, L. Zhang, H. Zhang, C.P. Hodgkinson, R.E. Pratt, P.B. Rosenberg, M. Mirotsou, V.J. Dzau, MicroRNA induced cardiac reprogramming in vivo: evidence for mature cardiac myocytes and improved cardiac function, Circ. Res. 116 (3) (2015) 418–424.
[238] Y. Zhao, E. Samal, D. Srivastava, Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis, Nature 436 (7048) (2005) 214–220.
[239] N. Liu, A.H. Williams, Y. Kim, J. McAnally, S. Bezprozvannaya, L.B. Sutherland, J.A. Richardson, R. Bassel-Duby, E.N. Olson, An intragenic MEF2-dependent enhancer directs muscle-specific expression of microRNAs 1 and 133, Proc. Natl. Acad. Sci. U. S. A. 104 (52) (2007) 20844–20,849.
[240] L. Qian, J.D. Wythe, J. Liu, J. Cartry, G. Vogler, B. Mohapatra, R.T. Otway, Y. Huang, I.N. King, M. Maillet, Y. Zheng, T. Crawley, O. Taghli-Lamallem, C. Semsarian, S. Dunwoodie, D. Winlaw, R.P. Harvey, D. Fatkin, J.A. Towbin, J.D. Molkentin, D. Srivastava, K. Ocorr, B.G. Bruneau, R. Bodmer, Tinman/Nkx2-5 acts via miR-1 and upstream of Cdc42 to regulate heart function across species, J. Cell Biol. 193 (7) (2011) 1181–1196.
[241] K.N. Ivey, A. Muth, J. Arnold, F.W. King, R.-F. Yeh, J.E. Fish, E.C. Hsiao, R.J. Schwartz, B.R. Conklin, H.S. Bernstein, D. Srivastava, MicroRNA regulation of cell lineages in mouse and human embryonic stem cells, Cell Stem Cell 2 (3) (2008) 219–229.
[242] N. Kapoor, W. Liang, E. Marban, H.C. Cho, Direct conversion of quiescent cardiomyocytes to pacemaker cells by expression of Tbx18, Nat. Biotechnol. 31 (1) (2013) 54–62.
[243] Y.-F. Hu, J.F. Dawkins, H.C. Cho, E. Marban, E. Cingolani, Biological pacemaker created by minimally invasive somatic reprogramming in pigs with complete heart block, Sci. Transl. Med. 6 (245) (2014), 245ra94.
[244] E. Bardot, D. Calderon, F. Santoriello, S. Han, K. Cheung, B. Jadhav, I. Burtscher, S. Artap, R. Jain, J. Epstein, H. Lickert, V. Gouon-Evans, A.J. Sharp, N.C. Dubois, Foxa2 identifies a cardiac progenitor population with ventricular differentiation potential, Nat. Commun. 8 (2017) 14428.
[245] R. David, J. Stieber, E. Fischer, S. Brunner, C. Brenner, S. Pfeiler, F. Schwarz, W.-M. Franz, Forward programming of pluripotent stem cells towards distinct cardiovascular cell types, Cardiovasc. Res. 84 (2) (2009) 263–272.
[246] K. Maass, A. Shekhar, J. Lu, G. Kang, F. See, E.E. Kim, C. Delgado, S. Shen, L. Cohen, G.I. Fishman, Isolation and characterization of embryonic stem cell-derived cardiac Purkinje cells, Stem Cells 33 (4) (2015) 1102–1112.
[247] Q. Zhang, J. Jiang, P. Han, Q. Yuan, J. Zhang, X. Zhang, Y. Xu, H. Cao, Q. Meng, L. Chen, T. Tian, X. Wang, P. Li, J. Hescheler, G. Ji, Y. Ma, Direct differentiation of atrial and ventricular myocytes from human embryonic stem cells by alternating retinoid signals, Cell Res. 21 (4) (2011) 579–587.
[248] I. Kehat, L. Khimovich, O. Caspi, A. Gepstein, R. Shofti, G. Arbel, I. Huber, J. Satin, J. Itskovitz-Eldor, L. Gepstein, Electromechanical integration of cardiomyocytes derived from human embryonic stem cells, Nat. Biotechnol. 22 (10) (2004) 1282–1289.
[249] J.H. van Weerd, V.M. Christoffels, The formation and function of the cardiac conduction system, Development 143 (2) (2016) 197–210.
[250] A. Kennedy, D.D. Finlay, D. Guldenring, R. Bond, K. Moran, J. McLaughlin, The cardiac conduction system: generation and conduction of the cardiac impulse, Crit. Care Nurs. Clin. North Am. 28 (3) (2016) 269–279.
[251] B. Joung, M. Ogawa, S.-F. Lin, P.-S. Chen, The calcium and voltage clocks in sinoatrial node automaticity, Korean Circ. J. 39 (6) (2009) 217–222.
[252] I.P. Temple, S. Inada, H. Dobrzynski, M.R. Boyett, Connexins and the atrioventricular node, Heart Rhythm. 10 (2) (2013) 297–304.
[253] D.C. Bartos, E. Grandi, C.M. Ripplinger, Ion channels in the heart, Compr. Physiol. 5 (3) (2015) 1423–1464.
[254] J.P. Moore, J.A. Aboulhosn, Introduction to the congenital heart defects: anatomy of the conduction system, Card. Electrophysiol. Clin. 9 (2) (2017) 167–175.
[255] C. Walsh-Irwin, G.B. Hannibal, Sick sinus syndrome, AACN Adv. Crit. Care 26 (4) (2015) 376–380.
[256] R.M. John, S. Kumar, Sinus node and atrial arrhythmias, Circulation 133 (19) (2016) 1892–1900.
[257] H. Dobrzynski, M.R. Boyett, R.H. Anderson, New insights into pacemaker activity: promoting understanding of sick sinus syndrome, Circulation 115 (14) (2007) 1921–1932.
[258] G.A. Ewy, Sick sinus syndrome: synopsis, J. Am. Coll. Cardiol. 64 (6) (2014) 539–540.
[259] M. Semelka, J. Gera, S. Usman, Sick sinus syndrome: a review, Am. Fam. Physician 87 (10) (2013) 691–696.
[260] G. Tse, T. Liu, K.H. Li, V. Laxton, A.O. Wong, Y.W. Chan, W. Keung, C.W. Chan, R.A. Li, Tachycardia-bradycardia syndrome: electrophysiological mechanisms and future therapeutic approaches (review), Int. J. Mol. Med. 39 (3) (2017) 519–526.
[261] Indications and recommendations for pacemaker therapy, Am. Fam. Physician 71 (8) (2005) 1563–1570.
[262] M.R. Rosen, R.B. Robinson, P.R. Brink, I.S. Cohen, The road to biological pacing, Nat. Rev. Cardiol. 8 (11) (2011) 656–666.
[263] M.L. Bakker, G.J. Boink, B.J. Boukens, A.O. Verkerk, M. van den Boogaard, A.D. den Haan, W.M. Hoogaars, H.P. Buermans, J.M. de Bakker, J. Seppen, H.L. Tan, A.F. Moorman, P.A. t Hoen, V.M. Christoffels, T-box transcription factor TBX3 reprogrammes mature cardiac myocytes into pacemaker-like cells, Cardiovasc. Res. 94 (3) (2012) 439–449.
[264] B. Lown, Electrical reversion of cardiac arrhythmias, Br. Heart J. 29 (4) (1967) 469–489.
[265] D.U. Frank, K.L. Carter, K.R. Thomas, R.M. Burr, M.L. Bakker, W.A. Coetzee, M. Tristani-Firouzi, M.J. Bamshad, V.M. Christoffels, A.M. Moon, Lethal arrhythmias in Tbx3-deficient mice reveal extreme dosage sensitivity of cardiac conduction system function and homeostasis, Proc. Natl. Acad. Sci. U. S. A. 109 (3) (2012) E154–63.
[266] W.M.H. Hoogaars, A. Engel, J.F. Brons, A.O. Verkerk, F.J. de Lange, L.Y.E. Wong, M.L. Bakker, D.E. Clout, V. Wakker, P. Barnett, J.H. Ravesloot, A.F.M. Moorman, E.E. Verheijck, V.M. Christoffels, Tbx3 controls the sinoatrial node gene program and imposes pacemaker function on the atria, Genes Dev. 21 (9) (2007) 1098–1112.
[267] V. Vedantham, G. Galang, M. Evangelista, R.C. Deo, D. Srivastava, RNA sequencing of mouse sinoatrial node reveals an upstream regulatory role for Islet-1 in cardiac pacemaker cells, Circ. Res. 116 (5) (2015) 797–803.
[268] C. Wiese, T. Grieskamp, R. Airik, M.T.M. Mommersteeg, A. Gardiwal, C. de Gier-de Vries, K. Schuster-Gossler, A.F.M. Moorman, A. Kispert, V.M. Christoffels, Formation of the sinus node head and differentiation of sinus node myocardium are independently regulated by Tbx18 and Tbx3, Circ. Res. 104 (3) (2009) 388–397.
[269] M.R. Boyett, S. Inada, S. Yoo, J. Li, J. Liu, J. Tellez, I.D. Greener, H. Honjo, R. Billeter, M. Lei, H. Zhang, I.R. Efimov, H. Dobrzynski, Connexins in the sinoatrial and atrioventricular nodes, Adv. Cardiol. 42 (2006) 175–197.
[270] D. Eckardt, M. Theis, J. Degen, T. Ott, H.V.M. van Rijen, S. Kirchhoff, J.-S. Kim, J.M.T. de Bakker, K. Willecke, Functional role of connexin43 gap junction channels in adult mouse heart assessed by inducible gene deletion, J. Mol. Cell. Cardiol. 36 (1) (2004) 101–110.
[271] M.M. Kreuzberg, J.W. Schrickel, A. Ghanem, J.-S. Kim, J. Degen, U. Janssen-Bienhold, T. Lewalter, K. Tiemann, K. Willecke, Connexin30.2 containing gap junction channels decelerate impulse propagation through the atrioventricular node, Proc. Natl. Acad. Sci. U. S. A. 103 (15) (2006) 5959–5964.
[272] M.M. Kreuzberg, G. Sohl, J.-S. Kim, V.K. Verselis, K. Willecke, F.F. Bukauskas, Functional properties of mouse connexin30.2 expressed in the conduction system of the heart, Circ. Res. 96 (11) (2005) 1169–1177.
[273] L. Marger, P. Mesirca, J. Alig, A. Torrente, S. Dubel, B. Engeland, S. Kanani, P. Fontanaud, J. Striessnig, H.-S. Shin, D. Isbrandt, H. Ehmke, J. Nargeot, M.E. Mangoni, Functional roles of Ca(v)1.3, Ca(v)3.1 and HCN channels in automaticity of mouse atrioventricular cells: insights into the atrioventricular pacemaker mechanism, Channels (Austin) 5 (3) (2011) 251–261.
[274] J. Stieber, S. Herrmann, S. Feil, J. Loster, R. Feil, M. Biel, F. Hofmann, A. Ludwig, The hyperpolarization-activated channel HCN4 is required for the generation of pacemaker action potentials in the embryonic heart, Proc. Natl. Acad. Sci. U. S. A. 100 (25) (2003) 15235–15,240.
[275] A.G. Torrente, P. Mesirca, P. Neco, R. Rizzetto, S. Dubel, C. Barrere, M. Sinegger-Brauns, J. Striessnig, S. Richard, J. Nargeot, A.M. Gomez, M.E. Mangoni, L-type Cav1.3 channels regulate ryanodine receptor-dependent Ca2 + release during sino-atrial node pacemaker activity, Cardiovasc. Res. 109 (3) (2016) 451–461.
[276] M. Yamamoto, H. Dobrzynski, J. Tellez, R. Niwa, R. Billeter, H. Honjo, I. Kodama, M.R. Boyett, Extended atrial conduction system characterised by the expression of the HCN4 channel and connexin45, Cardiovasc. Res. 72 (2) (2006) 271–281.
[277] F. Greulich, M.-O. Trowe, A. Leffler, C. Stoetzer, H.F. Farin, A. Kispert, Misexpression of Tbx18 in cardiac chambers of fetal mice interferes with chamber-specific developmental programs but does not induce a pacemaker-like gene signature, J. Mol. Cell. Cardiol. 97 (2016) 140–149.
[278] Y.-J. Nam, C. Lubczyk, M. Bhakta, T. Zang, A. Fernandez-Perez, J. McAnally, R. Bassel-Duby, E.N. Olson, N.V. Munshi, Induction of diverse cardiac cell types by reprogramming fibroblasts with cardiac transcription factors, Development 141 (22) (2014) 4267–4278.
[279] I. Bruzauskaite, D. Bironaite, E. Bagdonas, V.A. Skeberdis, J. Denkovskij, T. Tamulevicius, V. Uvarovas, E. Bernotiene, Relevance of HCN2-expressing human mesenchymal stem cells for the generation of biological pacemakers, Stem Cell Res Ther 7 (1) (2016) 67.
[280] Y. Feng, S. Luo, S. Tong, L. Zhong, C. Zhang, P. Yang, Z. Song, Electric-pulse current stimulation increases if current in mShox2 genetically modified canine mesenchymal stem cells, Cardiology 132 (1) (2015) 49–57.
[281] Y. Feng, S. Luo, P. Yang, Z. Song, Electric pulse current stimulation increases electrophysiological properties of if current reconstructed in mHCN4-transfected canine mesenchymal stem cells, Exp. Ther. Med. 11 (4) (2016) 1323–1329.
[282] C. Jun, Z. Zhihui, W. Lu, N. Yaoming, W. Lei, Q. Yao, S. Zhiyuan, Canine bone marrow mesenchymal stromal cells with lentiviral mHCN4 gene transfer create cardiac pacemakers, Cytotherapy 14 (5) (2012) 529–539.
[283] W. Lu, N. Yaoming, R. Boli, C. Jun, Z. Changhai, Z. Yang, S. Zhiyuan, mHCN4 genetically modified canine mesenchymal stem cells provide biological pacemaking function in complete dogs with atrioventricular block, Pacing Clin. Electrophysiol. 36 (9) (2013) 1138–1149.
[284] J. Ma, C. Zhang, S. Huang, G. Wang, X. Quan, Use of rats mesenchymal stem cells modified with mHCN2 gene to create biologic pacemakers, J. Huazhong Univ. Sci. Technol. Med. Sci. 30 (4) (2010) 447–452.
[285] A.N. Plotnikov, I. Shlapakova, M.J. Szabolcs, P. Danilo, B.H. Lorell, I.A. Potapova, Z. Lu, A.B. Rosen, R.T. Mathias, P.R. Brink, R.B. Robinson, I.S. Cohen, M.R. Rosen, Xenografted adult human mesenchymal stem cells provide a platform for sustained biological pacemaker function in canine heart, Circulation 116 (7) (2007) 706–713.
[286] I. Potapova, A. Plotnikov, Z. Lu, P. Danilo, V. Valiunas, J. Qu, S. Doronin, J. Zuckerman, I.N. Shlapakova, J. Gao, Z. Pan, A.J. Herron, R.B. Robinson, P.R. Brink, M.R. Rosen, I.S. Cohen, Human mesenchymal stem cells as a gene delivery system to create cardiac pacemakers, Circ. Res. 94 (7) (2004) 952–959.
[287] J. Yang, T. Song, P. Wu, Y. Chen, X. Fan, H. Chen, J. Zhang, C. Huang, Differentiation potential of human mesenchymal stem cells derived from adipose tissue and bone marrow to sinus node-like cells, Mol. Med. Rep. 5 (1) (2012) 108–113.
[288] X.-J. Yang, Y.-F. Zhou, H.-X. Li, L.-H. Han, W.-P. Jiang, Mesenchymal stem cells as a gene delivery system to create biological pacemaker cells in vitro, J. Int. Med. Res. 36 (5) (2008) 1049–1055.

[289] Y.-F. Zhou, X.-J. Yang, H.-X. Li, L.-H. Han, W.-P. Jiang, Genetically-engineered mesenchymal stem cells transfected with human HCN1 gene to create cardiac pacemaker cells, J. Int. Med. Res. 41 (5) (2013) 1570–1576.
[290] Y.-F. Zhou, X.-J. Yang, H.-X. Li, L.-H. Han, W.-P. Jiang, Mesenchymal stem cells transfected with HCN2 genes by LentiV can be modified to be cardiac pacemaker cells, Med. Hypotheses 69 (5) (2007) 1093–1097.
[291] M. Tong, X.-J. Yang, B.-y. Geng, L.-H. Han, Y.-F. Zhou, X. Zhao, H.-X. Li, Overexpression of connexin 45 in rat mesenchymal stem cells improves the function as cardiac biological pacemakers, Chin. Med. J. 123 (12) (2010) 1571–1576.
[292] L. Chen, Z.-J. Deng, J.-S. Zhou, R.-J. Ji, X. Zhang, C.-S. Zhang, Y.-Q. Li, X.-Q. Yang, Tbx18-dependent differentiation of brown adipose tissue-derived stem cells toward cardiac pacemaker cells, Mol. Cell. Biochem. (2017).
[293] P.W. Burridge, E. Matsa, P. Shukla, Z.C. Lin, J.M. Churko, A.D. Ebert, F. Lan, S. Diecke, B. Huber, N.M. Mordwinkin, J.R. Plews, O.J. Abilez, B. Cui, J.D. Gold, J.C. Wu, Chemically defined generation of human cardiomyocytes, Nat. Methods 11 (8) (2014) 855–860.
[294] X. Lian, C. Hsiao, G. Wilson, K. Zhu, L.B. Hazeltine, S.M. Azarin, K.K. Raval, J. Zhang, T.J. Kamp, S.P. Palecek, Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling, Proc. Natl. Acad. Sci. U. S. A. 109 (27) (2012) E1848–57.
[295] A. Kleger, T. Seufferlein, D. Malan, M. Tischendorf, A. Storch, A. Wolheim, S. Latz, S. Protze, M. Porzner, C. Proepper, C. Brunner, S.-F. Katz, G. Varma Pusapati, L. Bullinger, W.-M. Franz, R. Koehntop, K. Giehl, A. Spyrantis, O. Wittekindt, Q. Lin, M. Zenke, B.K. Fleischmann, M. Wartenberg, A.M. Wobus, T.M. Boeckers, S. Liebau, Modulation of calcium-activated potassium channels induces cardiogenesis of pluripotent stem cells and enrichment of pacemaker-like cells, Circulation 122 (18) (2010) 1823–1836.
[296] M. Jara-Avaca, H. Kempf, M. Ruckert, D. Robles-Diaz, A. Franke, J. de La Roche, M. Fischer, D. Malan, P. Sasse, W. Solodenko, G. Drager, A. Kirschning, U. Martin, R. Zweigerdt, EBIO does not induce cardiomyogenesis in human pluripotent stem cells but modulates cardiac subtype enrichment by lineage-selective survival, Stem Cell Rep. 8 (2) (2017) 305–317.
[297] A. Scavone, D. Capilupo, N. Mazzocchi, A. Crespi, S. Zoia, G. Campostrini, A. Bucchi, R. Milanesi, M. Baruscotti, S. Benedetti, S. Antonini, G. Messina, D. DiFrancesco, A. Barbuti, Embryonic stem cell-derived CD166+ precursors develop into fully functional sinoatrial-like cells, Circ. Res. 113 (4) (2013) 389–398.
[298] W. Rust, T. Balakrishnan, R. Zweigerdt, Cardiomyocyte enrichment from human embryonic stem cell cultures by selection of ALCAM surface expression, Regen. Med. 4 (2) (2009) 225–237.
[299] S.I. Hashem, W.C. Claycomb, Genetic isolation of stem cell-derived pacemaker-nodal cardiac myocytes, Mol. Cell. Biochem. 383 (1–2) (2013) 161–171.
[300] Arash Yavari, Mohamed Bellahcene, Annalisa Bucchi, Syevda Sirenko, Katalin Pinter, Neil Herring, Julia J. Jung, Kirill Tarasov, Gabor Czibik, Violetta Steeples, Sahar Ghaffari, Chinh Nguyen, Alexander Stockenhuber, Emily J. Sharpe, Joshua R. St. Clair, Christian Rimmbach, Markus Wolfien, Yosuke Okamoto, Mingyi Wang, Bruce D. Ziman, Jack M. Moen, Dongmei Yang, Daniel Riordon, Christopher Ramirez, Manuel Paina, Joonho Lee, Jing Zhang, Ismayil Ahmet, Michael G. Matt, Yelena S. Tarasova, Dilair Baban, Natasha Sahgal, Helen Lockstone, Rathi Puliyadi, Joseph de Bono, John Gomes, Hannah Muskett, Mahon L. Maguire, Matthew Kelly, Pedro P.N. dos Santos, Nicola J. Bright, Angela Woods, Katja Gehmlich, Henrik Isackson, Gillian Douglas, David J.P. Ferguson, Jürgen E. Schneider, Andrew Tinker, Olaf Wolkenhauer, Keith M. Channon, Eduardo B. Sternick, David J. Paterson, Charles S. Redwood, David Carling, Catherine Proenza, Robert David, Mirko Baruscotti, Dario DiFrancesco, Edward G. Lakatta, Hugh Watkins, Houman Ashrafian, The γ2 subunit of AMP-activated protein kinase regulates mammalian heart rate, Nat. Commun. (2017) (accepted for publication).
[301] R.-S. Wang, B.A. Maron, J. Loscalzo, Systems medicine: evolution of systems biology from bench to bedside, Wiley Interdiscip. Rev. Syst. Biol. Med. 7 (4) (2015) 141–161.
[302] W.R. MacLellan, Y. Wang, A.J. Lusis, Systems-based approaches to cardiovascular disease, Nat. Rev. Cardiol. 9 (3) (2012) 172–184.
[303] S.C. Lott, M. Wolfien, K. Riege, A. Bagnacani, O. Wolkenhauer, S. Hoffmann, W.R. Hess, Customized workflow development and data modularization concepts for RNA-sequencing and metatranscriptome experiments, J. Biotechnol. S0168-1656 (17) (2017) 31499–31502.
[304] B.A. Grüning, J. Fallmann, D. Yusuf, S. Will, A. Erxleben, F. Eggenhofer, T. Houwaart, B. Batut, P. Videm, A. Bagnacani, M. Wolfien, S.C. Lott, Y. Hoogstrate, W.R. Hess, O. Wolkenhauer, S. Hoffmann, A. Akalin, U. Ohler, P.F. Stadler, R. Backofen, The RNA workbench: best practices for RNA and high-throughput sequencing bioinformatics in Galaxy, Nucleic Acids Res. (2017).
[305] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (11) (2003) 2498–2504.
[306] S. Maere, K. Heymans, M. Kuiper, BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks, Bioinformatics (Oxford, England) 21 (16) (2005) 3448–3449.
[307] G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.-H. Fridman, F. Pages, Z. Trajanoski, J. Galon, ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks, Bioinformatics (Oxford, England) 25 (8) (2009) 1091–1093.
[308] A.R. Paschoal, V. Maracaja-Coutinho, J.C. Setubal, Z.L.P. Simões, S. Verjovski-Almeida, A.M. Durham, Non-coding transcription characterization and annotation: a guide and web resource for non-coding RNA databases, RNA Biol. 9 (3) (2012) 274–282.
[309] L. Wang, X. Ma, X. Xu, Y. Zhang, Systematic identification and characterization of cardiac long intergenic noncoding RNAs in zebrafish, Sci. Rep. 7 (1) (2017) 1250.
[310] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinformatics 9 (2008) 559.
[311] M. Shouman, T. Turner, R. Stocker, Using Data Mining Techniques in Heart Disease Diagnosis and Treatment, 2012, 173–177.
[312] N. Alcaraz, J. Pauling, R. Batra, E. Barbosa, A. Junge, A.G.L. Christensen, V. Azevedo, H.J. Ditzel, J. Baumbach, KeyPathwayMiner 4.0: condition-specific pathway analysis by combining multiple omics studies and networks with Cytoscape, BMC Syst. Biol. 8 (2014) 99.
[313] F.M. Khan, U. Schmitz, S. Nikolov, D. Engelmann, B.M. Pützer, O. Wolkenhauer, J. Vera, Hybrid modeling of the crosstalk between signaling and transcriptional networks using ordinary differential equations and multi-valued logic, Biochim. Biophys. Acta 1844 (1 Pt B) (2014) 289–298.
[314] F.M. Khan, S. Marquardt, S.K. Gupta, S. Knoll, U. Schmitz, A. Spitschak, D. Engelmann, J. Vera, O. Wolkenhauer, B.M. Pützer, Unraveling a tumor type-specific regulatory core underlying E2F1-mediated epithelial-mesenchymal transition to predict receptor protein signatures, Nat. Commun. 8 (1) (2017) 198.

2.2.2 Comparison of gene expression for reprogrammed cardiac cell types

Müller P., **Wolfien M.**, Ekat K., Lang C.I., Koczan D., Wolkenhauer O., Hahn O., Peters K., Lang H., David R., Lemcke H. (2020).
RNA-Based Strategies for Cardiac Reprogramming of Human Mesenchymal Stromal Cells.
Cells. IF: 5.656, Citations (December 14, 2020): 0

The differentiation of adult mesenchymal stromal cells (MSCs) into a cardiac lineage is challenging compared to embryonic stem cells or induced pluripotent stem cells because they either bear tumorigenic risk due to genome modification or provoke ethical concerns. Here, it was investigated if MSCs derived from different sources, including bone marrow, dental follicle, and subcutaneous adipose tissue, can be driven into a cardiac lineage. The applied transient cell reprogramming strategy, which is based on miRNA and mRNA transfection for the generation of patient-specific cardiomyocytes, can be in turn used for regenerative medicine, cardiovascular research, and pharmacological studies.

In this manuscript, I analyzed and visualized the gene expression microarray data and performed the gene set and pathway enrichment analyses for the investigated cell types. It was discovered that adipose tissue-derived MSCs are the most susceptible cell type for the investigated reprogramming approaches, as confirmed by an enhanced expression of cardiac markers. Transcriptomic pathway analyses revealed an enrichment in cardiac progenitor development and heart development after mRNA treatment.

In agreement with published studies about MSCs overexpressing transcription factors, our results demonstrate the general feasibility of mRNA-based cardiac reprogramming of MSCs. Although we did not observe the formation of fully mature cardiomyocytes (*i.e.*, sarcomere formation), our data suggest that adult MSCs have the capability to acquire a cardiac-like phenotype after receiving a mRNA treatment coding for transcription factors that are known to regualte regulate heart development (*Gata4*, *Mesp1*, *Mef2c*, *Tbx5*). Yet, further optimization of the reprogramming process is mandatory to increase the reprogramming efficiency.

cells

MDPI

Article

RNA-Based Strategies for Cardiac Reprogramming of Human Mesenchymal Stromal Cells

Paula Mueller [1,2], Markus Wolfien [3], Katharina Ekat [4], Cajetan Immanuel Lang [5], Dirk Koczan [6], Olaf Wolkenhauer [3,7], Olga Hahn [4], Kirsten Peters [4], Hermann Lang [8], Robert David [1,2,*] and Heiko Lemcke [1,2]

1 Department of Cardiac Surgery, Reference and Translation Center for Cardiac Stem Cell Therapy (RTC), Rostock University Medical Center, 18057 Rostock, Germany; Paula.Mueller@uni-rostock.de (P.M.); Heiko.Lemcke@med.uni-rostock.de (H.L.)
2 Faculty of Interdisciplinary Research, Department Life, Light & Matter, University Rostock, 18059 Rostock, Germany
3 Institute of Computer Science, Department of Systems Biology and Bioinformatics, University of Rostock, 18057 Rostock, Germany; Markus.Wolfien@uni-rostock.de (M.W.); Olaf.wolkenhauer@uni-rostock.de (O.W.)
4 Department of Cell Biology, Rostock University Medical Center, 18057 Rostock, Germany; Katharina.Ekat@med.uni-rostock.de (K.E.); olga.hahn@med.uni-rostock.de (O.H.); Kirsten.Peters@med.uni-rostock.de (K.P.)
5 Department of Cardiology, Rostock University Medical Center, 18057 Rostock, Germany; Cajetan.lang@med.uni-rostock.de
6 Institute of Immunology, Rostock University Medical Center, 18057 Rostock, Germany; Dirk.Koczan@med.uni-rostock.de
7 Stellenbosch Institute of Advanced Study, Wallenberg Research Centre, Stellenbosch University, 7602 Stellenbosch, South Africa
8 Department of Operative Dentistry and Periodontology, Rostock University Medical Center, 18057 Rostock, Germany; Herman.Lang@med.uni-rostock.de
* Correspondence: Robert.David@med.uni-rostock.de; Tel.: +49-381-498-8973

Received: 6 December 2019; Accepted: 17 February 2020; Published: 22 February 2020

Abstract: Multipotent adult mesenchymal stromal cells (MSCs) could represent an elegant source for the generation of patient-specific cardiomyocytes needed for regenerative medicine, cardiovascular research, and pharmacological studies. However, the differentiation of adult MSC into a cardiac lineage is challenging compared to embryonic stem cells or induced pluripotent stem cells. Here we used non-integrative methods, including microRNA and mRNA, for cardiac reprogramming of adult MSC derived from bone marrow, dental follicle, and adipose tissue. We found that MSC derived from adipose tissue can partly be reprogrammed into the cardiac lineage by transient overexpression of GATA4, TBX5, MEF2C, and MESP1, while cells isolated from bone marrow, and dental follicle exhibit only weak reprogramming efficiency. qRT-PCR and transcriptomic analysis revealed activation of a cardiac-specific gene program and up-regulation of genes known to promote cardiac development. Although we did not observe the formation of fully mature cardiomyocytes, our data suggests that adult MSC have the capability to acquire a cardiac-like phenotype when treated with mRNA coding for transcription factors that regulate heart development. Yet, further optimization of the reprogramming process is mandatory to increase the reprogramming efficiency.

Keywords: mesenchymal stromal cells (MSC); mRNA; miRNA; cardiac reprogramming; cardiac differentiation

1. Introduction

Mesenchymal stromal cells (MSC) represent a multipotent cell population capable to differentiate into different cell types [1]. They are an easily-accessible cell source as they can be isolated at high yields from various kinds of human tissue, such as umbilical cord, bone marrow, dental pulp, adipose tissue, placenta, etc. [1]. The common mesenchymal cell types that emanate from MSC are osteocytes, chondrocytes, and adipocytes [2]. Due to their plasticity, MSC are considered as one of the most important cell types for the application in regenerative medicine as demonstrated by a huge number of pre-clinical studies and several clinical trials [3,4]. In addition, MSC mediate immunomodulatory and immunosuppressive effects that promote wound healing and tissue repair, while showing no teratoma formation post transplantation [5]. Nowadays, it is commonly accepted that the observed therapeutic impact induced by MSCs is mainly based on the secretion of paracrine factors rather than on the differentiation into cardiomyocytes.

In recent years, MSC have also been utilized for the generation of mesenchymal as well as non-mesenchymal cell lineages, including neuron-like, hepatocyte-like, and cardiac-like cells [6–10]. Despite these promising results, the differentiation of human MSC into fully mature cardiomyocytes bearing all their respective phenotypical and functional characteristics is difficult [11–15]. As MSC are located in various tissues, they represent a heterogeneous progenitor cell population dependent on the tissue source and the individual donor [16]. This heterogeneity could explain the variety in differentiation characteristics [17–19]. Therefore, it remains to be investigated which type of MSC favorably undergoes cardiac trans-differentiation, thus, is a suitable candidate for cardiac reprogramming strategies. Detailed knowledge about the cardiac differentiation potential of specific MSC populations is even more important as some studies showed enhanced therapeutic effects following cardiovascular lineage commitment of MSC [12].

The development of an approach to efficiently control the cardiac differentiation of MSC would be a crucial step for the production of patient-derived cardiomyocytes without any ethical concerns. As such, they can also serve as a model system, beneficial for basic cardiovascular research, drug screening, and translational applications. Currently, several re/programming strategies exist to guide the mesenchymal and non-mesenchymal differentiation of MSC, such as treatment with small molecules and cytokines, exposure to metabolic stress, co-culture experiments, or overexpression of regulatory proteins [20–24]. For the potential clinical use, transient, non-integrative reprogramming approaches are preferred to prevent permanent alterations of the genome and to reduce tumorigenic risk. Small non-coding RNAs, like microRNAs (miRNA) and chemically modified messenger RNA (mRNAs) allow the manipulation of cell behavior for a limited period of time, e.g., triggering (trans)-differentiation by activation of lineage-specific molecular pathways. Some studies have already shown that alteration of gene expression using selected miRNAs can induce cardiac differentiation of MSC to a small extent [15,25,26], while data about mRNA-based cardiac reprogramming is still lacking.

Unlike multipotent MSCs, pluripotent stem cells (PSCs) have been demonstrated to efficiently differentiate into cardiomyocytes, characterized by a profound sarcomere organization and spontaneous beating behavior [27]. Yet, these PSC-derived cardiomyocytes typically still represent an immature cell type, resembling a neonatal cell stage rather than an adult phenotype [28,29]. The common cardiac programming approaches used to guide cardiac differentiation of PSCs mainly relies on the application of cytokines and small molecules [30,31]. However, PSCs bear tumorigenic risk due to genome modification (induced pluripotent stem cells, iPSC) and provoke ethical concerns (embryonic stem cells, ESC). Therefore, increasing the efficiency of cardiac programming of MSC would be beneficial for cardiovascular research, including their therapeutic use.

Here, we examined whether MSC derived from different sources, including bone marrow (BM), dental follicle and subcutaneous adipose tissue can be driven towards a cardiac lineage using a transient reprogramming strategy based on miRNA and mRNA transfection. According to our results, adipose tissue-derived MSC (adMSC) were found to be the most susceptible cell type for this reprogramming

approach, as shown by enhanced expression of cardiac markers. At the same time, we observed the activation of transcriptome pathways involved in cardiac development following mRNA treatment.

2. Material and Methods

2.1. Cell Culture

BM-derived MSC (BM MSC) were obtained by sternal aspiration from donors undergone coronary bypass graft surgery. Anticoagulation was achieved by heparinization with 250 i.E./mL sodium heparin (Ratiopharm, Ulm, Germany). Mononuclear cells were isolated by density gradient centrifugation on 1077 Lymphocyte Separation Medium (LSM; PAA Laboratories, Pasching, Germany). MSC were enriched by plastic adherence and sub-cultured in MSC basal medium supplemented with SingleQuot (all Lonza, Cologne, Germany) and 1% Zellshield (Biochrom, Berlin, Germany).

Isolation of adMSC was performed by liposuction of healthy individuals. The extracted tissue was treated with collagenase for 30 min, followed by several filtrations and washing steps. The detailed process of adMSC isolation has been already described previously [32].

Dental follicle stem cells (DFSCs) were isolated from dental follicles of extracted wisdom teeth before tooth eruption. Following tooth removal, the follicle was removed and subjected to enzymatic treatment as presented earlier [33]. Upon tissue digestion, cells were seeded on tissue flasks and obtained by plastic adherence. DFSCs were maintained in DMEM-F12 (Thermo Fisher, Waltham, USA) supplemented with 10% FCS and 1% Zellshield.

All three types of stromal cells were maintained at 37 °C and 5% CO_2 humidified atmosphere. Medium was changed every 2–3 days. Sub-cultivation was performed when cells reached a confluency of ~80–90%.

All donors have given their written consent for the donation of their tissue according to the Declaration of Helsinki. The study was approved by the ethical committee of the Medical Faculty of the University of Rostock (registration number: bone marrow A2010-23; renewal in 2015; adipose tissue: A2013-0112, renewal in 2019, dental tissue: A 2017-0158).

2.2. Fluorescence-Activated Cell Sorting

The expression of cell surface markers was quantified by flow cytometric analysis. Stromal cells were labelled with antibodies CD29-APC, CD44-PerCP-Cy5.5, CD45-V500, CD73-PE, CD117-PE-Cy7, PerCP-Cy5.5 CD90 (BD Biosciences, San Jose, USA), and CD105-AlexaFluor488 (AbD Serotec, Oxford, UK). Respective isotype antibodies served as negative controls. A measurement of 3×10^4 events was carried out using BD FACS LSRII flow cytometer (BD Biosciences).

To evaluate miRNA and mRNA uptake efficiency, cells were treated with different amounts of Cy3-labeled Pre-miRNA Negative Control #1 (AM17120, Thermo Fisher) or GFP-mRNA (Trilink, San Diego, USA) and analyzed by flow cytometry 24 h post transfection. To detect cytotoxicity, cells were labelled with Near-IR LIVE/DEAD fixable dead cell stain kit (Molecular Probes, Eugene, USA). Analysis of flow cytometry data, including gating, was conducted with the FACSDiva software, Version 8. (Becton Dickinson).

2.3. Cardiac Reprogramming

For cardiac reprogramming, 1×10^5 cells/well were seeded on 0.1% gelatin-coated 6 well plates and cultured to 80% confluency. We transfected 40 pmol of each miRNA (Pre-miR™ hsa-miR-1, Pre-miR™hsa-miR-499a-5p, Pre-miR™hsa-miR-208a-3p, Pre-miR™hsa-miR-133a-3p, all Thermo Fisher) with Lipofectamine® 2000 according to the manufacturer's instructions (Thermo Fisher). Transfection of custom-made mRNA (Trilink) was performed with Viromer Red® transfection reagent (Lipocalyx, Halle, Germany). Cells were either transfected with 2 µg MESP1 or with a combination of 1 µg GATA4, 1 µg MEF2C and 1 µg TBX5. One day after transfection of miRNA or mRNA, cells were subjected to two different medium conditions. For cardiac induction medium I (card ind. I), cells were incubated in

RPMI, supplemented with B27 without insulin (Thermo fisher) for 7 days, followed by incubation in RPMI containing B27 +insulin/- vitamin A (Thermo Fisher) for another 21 days. Additionally, culture medium was supplemented with ascorbic acid (Sigma Aldrich, St. Louis, USA) and Wnt pathway targeting small molecules, including 6 μM CHIR99021 (days 1–2), and 5 μM IWP-2 (days 4–5) (both Stemcell Technologies). For cardiac induction II (card ind II), a commercially available cardiomyocyte differentiation kit was used according to the instructions given by the manufacturer (Thermo Fisher, A2921201).

2.4. IF Staining and Calcium Imaging

To verify multipotency, BM-MSC, DFSC and adMSC were subjected to in vitro differentiation towards osteogenic, chondrogenic and adipogenic lineages using the Mesenchymal Stem Cell Functional Identification Kit (R & D). Differentiation was induced by maintaining cells under different culture conditions according to the manufacturer instructions for 20 days. Subsequently, cells were fluorescently labelled to detect fatty acid-binding protein 4 (FABP4), Aggrecan and Osteocalcin to visualize successful differentiation into adipocytes, chondrocytes, and osteocytes.

For labelling of cardiac markers, cells were seeded on coverslips and fixed with 4% PFA. Antibody staining was performed as described elsewhere [34]. Cells were labelled with anti sarcomeric α-actinin (abcam, ab9465), anti-NKX2.5 (Santa Cruz, sc-8697), anti-TBX5 (abcam, ab137833) and anti-MEF2C (Santa Cruz, sc-313).

To visualize intracellular calcium, cells were cultured on 8 well chamberslides (Ibidi). Three days after seeding, cells were incubated with the calcium sensitive dye Cal520 (AATBioquest) for one hour at 37 °C and subjected to fluorescence microscopy. All fluorescence images were acquired using Zeiss ELYRA LSM 780 (Zeiss, Oberkochen, Germany).

2.5. RNA Isolation and Quantitative Real-Time Polymerase Chain Reaction

Isolation of cellular RNA was performed using the NucleoSpin® RNA isolation kit (Macherey-Nagel, Düren, Germany) according to the manufacturer instructions. The concentration and purity of isolated RNA was assessed with NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific). Subsequently, cDNA synthesis was performed with a High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific). The reverse transcription reaction was conducted using the MJ Mini™ thermal cycler (Bio-Rad).

Quantitative real-time PCR for cardiac marker genes was carried out using the StepOnePlus™ Real-Time PCR System (Applied Biosystems, Foster City, USA) with following reaction parameters (StepOne™ Software Version 2, Applied Biosystems, Germany): start at 50 °C for 2 min, initial denaturation at 95 °C for 10 min, denaturation at 95 °C for 15 s and annealing/elongation at 60 °C for 1 min with 40 cycles. A qPCR reaction contained: TaqMan® Universal PCR Master Mix (Thermo Fisher), respective TaqMan® Gene Expression Assay, UltraPure™ DNase/RNase-Free Distilled Water (Thermo Fisher), and 30 ng of the respective cDNA. The following target gene assays were used: ACTN2 (Hs00153809_m1); MYH6 (Hs01101425_m1) TBX5 (Hs00361155_m1); TNNI3 (Hs00165957_m1), GJA1 (Hs00748445_s1); HPRT (HS01003267_m1) (all Thermo Fisher). Obtained CT values were normalized to HPRT and data were calculated as fold-change expression, related to untreated control cells.

2.6. Microarray Analysis

RNA integrity was analyzed using the Agilent Bioanalyzer 2100 with the RNA Pico chip kit (Agilent Technologies). 200 ng of isolated RNAs were subjected to microarray hybridization as described in [35]. Hybridization was performed on Affymetrix Clariom™ D Arrays according to the manufacturer's instructions (Thermo Fisher).

Analysis of the microarray data was conducted with the provided Transcriptome Analysis Console Software from Thermo Fisher (Version 4.0.1, Waltham, USA). The analysis included quality control,

data normalization, and statistical testing for differential expression (Limma). Transcripts were considered as significantly differentially expressed with a fold change (FC) higher than 2 or smaller −2, false discovery rate (FDR) < 0.05, and $p < 0.05$. The pathway analyses were conducted based on a gene set enrichment analysis using Fisher's Exact Test (GSEA) on the Wiki-Pathways database. Only significant pathways have been selected.

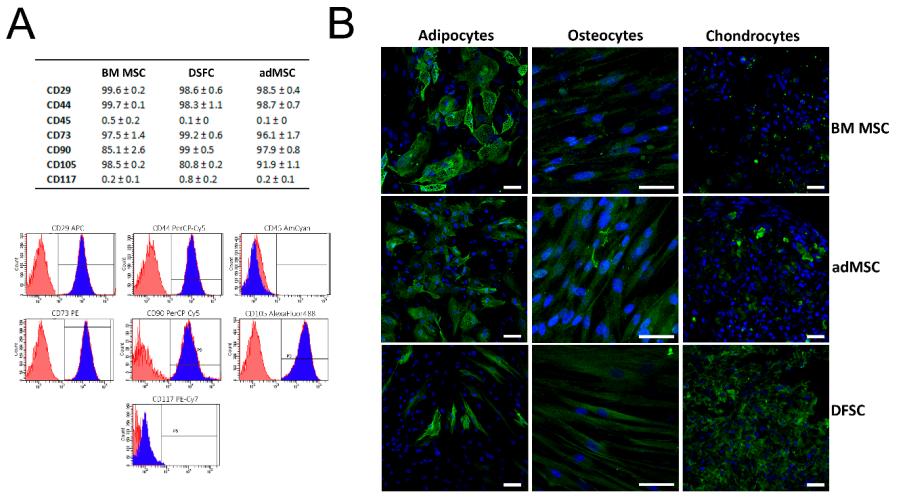
2.7. Statistical Analysis

Data are presented as mean ± SEM, obtained from three patients for each MSC type. Preparation of graphs and statistical analysis was performed using SIGMA Plot software (Systat Software GmbH, Erkrath, Germany). Statistical significance was considered as * $p \leq 0.5$, ** $p \leq 0.05$, *** $p \leq 0.001$.

3. Results

3.1. Characterization of Isolated MSC

Initially we performed flow cytometric analysis to investigate the presence of common mesenchymal surface markers in isolated MSC. The obtained data indicated a high expression of CD29, CD44, CD73, CD105 and CD90, while very low levels were detected for CD117 and CD45, indicating that stem cells possess properties of MSC (Figure 1A,B).



	BM MSC	DSFC	adMSC
CD29	99.6 ± 0.2	98.6 ± 0.6	98.5 ± 0.4
CD44	99.7 ± 0.1	98.3 ± 1.1	98.7 ± 0.7
CD45	0.5 ± 0.2	0.1 ± 0	0.1 ± 0
CD73	97.5 ± 1.4	99.2 ± 0.6	96.1 ± 1.7
CD90	85.1 ± 2.6	99 ± 0.5	97.9 ± 0.8
CD105	98.5 ± 0.2	80.8 ± 0.2	91.9 ± 1.1
CD117	0.2 ± 0.1	0.8 ± 0.2	0.2 ± 0.1

Figure 1. Phenotype-related and functional characterization of mesenchymal stromal cells (MSC): (**A**) Flow cytometric measurements revealed a high expression of common MSC surface markers (CD29, CD44, CD73, CD90, CD105), while very low levels were found for hematopoietic surface markers (CD45 and CD117). Representative flow cytometry charts of adipose tissue-derived MSC (adMSC) demonstrate the expression level of surface markers. Blue histograms represent measurement of CD surface marker with corresponding isotype control, shown in red. (**B**) Multi-lineage differentiation assay indicated adipogenic, osteogenic, and chondrogenic differentiation of MSC. Detection of adipocytes was performed by labelling of FABP4, while osteocytes and chondrocytes were identified by fluorescence staining of osteocalcein and aggrecan, respectively. Scale bar: 50 µm. Results in (**A**) are shown as mean ± SEM, obtained by analysis of three different donors for each MSC cell type.

MSC characteristics were further confirmed by a functional assay that demonstrated the multilineage differentiation capability of all three cell types. Upon incubation in lineage-specific induction medium, the cells were capable to differentiate into adipocytes, chondrocytes, and osteocytes, as shown by fluorescence labelling of specific differentiation markers (Figure 1B). As expected,

adMSC were found to profoundly express FABP4, if compared to osteocalcin and aggrecan labelling. In contrast, DFSCs favored chondrogenic differentiation indicated by strong fluorescence intensity of aggrecan staining.

Next, we compared the different MSC by analyzing their gene expression profiles using a microarray platform. The obtained data allowed us to compare the transcription profile among both, individual donors and MSC derived from different tissue. Boxplots of signal intensity distributions for each performed microarray are shown in Figure 2, indicating good data quality prior (blue) and after (red) normalization of the gene expression data (Figure 2A). A principal component analysis (PCA) was performed to show the common clustering of the triplicates (Figure 2B, blue, red and purple) as well as the differences of tested cell types, each represented by three different donors. We found that stromal cells from BM, adipose, and dental tissue are clearly distinct with respect to their transcriptomic profile. Interestingly, we detected a high donor-dependent variety of the gene expression for MSC derived from human BM (Figure 2B), suggesting a potential donor-specific impact on the efficacy of cardiac programming. A total of 1685 differentially expressed genes were detected, while 13 genes were shared by all MSC populations (Figure 2C). Most differentially expressed transcripts (679) have been found between MSCs obtained from BM and adipose tissue, suggesting a higher gene profile related diversity within these two MSC populations (Figure 2D). A list of differentially expressed genes between all MSC types is given in Table S1.

Next, we compared the different MSC by analyzing their gene expression profiles using a microarray platform. The obtained data allowed us to compare the transcription profile among both, individual donors and MSC derived from different tissue. Boxplots of signal intensity distributions for each performed microarray are shown in Figure 2, indicating good data quality prior (blue) and after (red) normalization of the gene expression data (Figure 2A). A principal component analysis (PCA) was performed to show the common clustering of the triplicates (Figure 2B, blue, red and purple) as well as the differences of tested cell types, each represented by three different donors. We found that stromal cells from BM, adipose, and dental tissue are clearly distinct with respect to their transcriptomic profile. Interestingly, we detected a high donor-dependent variety of the gene expression for MSC derived from human BM (Figure 2B), suggesting a potential donor-specific impact on the efficacy of cardiac programming. A total of 1685 differentially expressed genes were detected, while 13 genes were shared by all MSC populations (Figure 2C). Most differentially expressed transcripts (679) have been found between MSCs obtained from BM and adipose tissue, suggesting a higher gene profile related diversity within these two MSC populations (Figure 2D). A list of differentially expressed genes between all MSC types is given in Table S1.

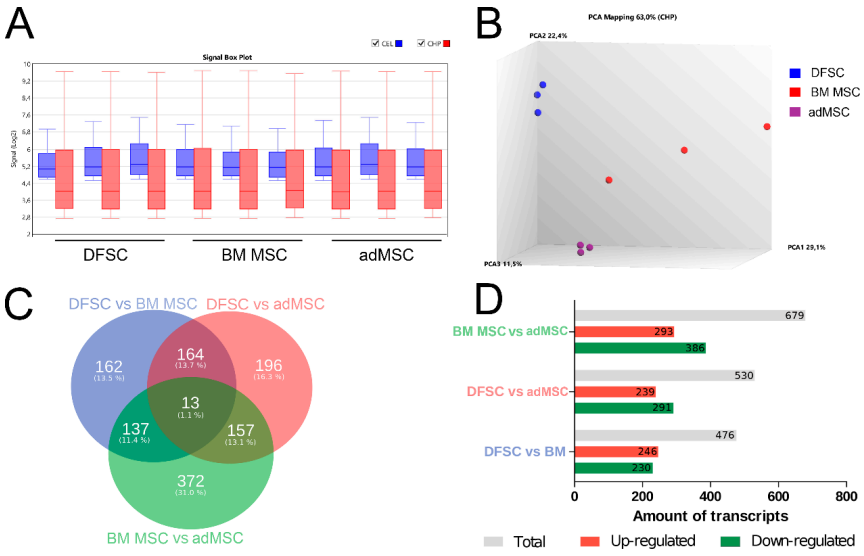


Figure 2. Comparative microarray analysis of undifferentiated dental follicle stem cells (DFSCs), bone marrow (BM) MSC, and adMSC. (**A**) Comparison of signal intensity for .cel files (blue) and .chp files (red) after normalization demonstrates sufficient data quality. (**B**) MSC from different sources are clearly distinct in regard to their transcription profile. A high patient-dependent variety was found for BM MSC, while adMSC and DFSCs demonstrate a more homogenous distribution. (**C**) Venn diagram visualizes expressed genes overlapping between different MSC cell types. (**D**) The numbers of up- and down-regulated transcripts is significantly differentially expressed in all three cell types.

Figure 2. Comparative microarray analysis of undifferentiated dental follicle stem cells (DFSCs), bone marrow (BM) MSC, and adMSC. (**A**) Comparison of signal intensity for .cel files (blue) and .chp files (red) after normalization demonstrates sufficient data quality. (**B**) MSC from different sources are clearly distinct in regard to their transcription profile. A high patient-dependent variety was found for BM MSC, while adMSC and DFSCs demonstrate a more homogenous distribution. (**C**) Venn diagram visualizes expressed genes overlapping between different MSC cell types. (**D**) The numbers of up- and down-regulated transcripts is significantly differentially expressed in all three cell types.

3.2. Reprogramming of MSC Using miRNA and Cardiac Induction Cell Culture Conditions

In order to induce cardiac reprogramming, cells were cultured under two different medium conditions (see Section 2.3), separately or in combination with myocardial miRNAs (myomiRNAs), that have been previously shown to induce cardiac differentiation in fibroblasts (miR-1, miR-499a, miR-208a, and miR-133a) [36]. As the efficiency of miRNA-based reprogramming largely depends on

In order to induce cardiac reprogramming, cells were cultured under two different medium conditions (see Section 2.3), separately or in combination with myocardial miRNAs (myomiRNAs), that have been previously shown to induce cardiac differentiation in fibroblasts (miR-1, miR-499a, miR-208a, and miR-133a) [36]. As the efficiency of miRNA-based reprogramming largely depends on

proper intracellular miRNA uptake, we evaluated miRNA transfection conditions using Cy3-labelled miRNA. Depending on the amount of transfected miRNA, uptake efficiencies of ~80–95% were achieved in all three cell types tested (Figure 3A). Importantly, only minimal cytotoxic effects were observed following transfection of miRNA (Figure 3B).

The success of myo-miRNA-based cardiac reprogramming was determined by qRT-PCR analysis of cardiac specific marker genes four weeks post transfection. Compared to control cells, cardiac induction medium II was found to be the most effective treatment leading to an induction of α-actinin, TBX5, GJA1, and cardiac Troponin I. While the level of α-actinin mRNA was strongly increased in all three cell types, a less pronounced effect was observed for cardiac Troponin I (Figure 3C). Notably, adMSCs showed the highest expression levels of cardiac marker genes after the treatment with cardiac induction medium II, when compared to MSCs obtained from dental follicle as well as BM, and therefore have been identified as the preferred candidate for our cardiac programming approach.

Cells 2020, 9, x FOR PEER REVIEW 7 of 19

proper intracellular miRNA uptake, we evaluated miRNA transfection conditions using Cy3-labelled miRNA. Depending on the amount of transfected miRNA, uptake efficiencies of ~80–95% were achieved in all three cell types tested (Figure 3A). Importantly, only minimal cytotoxic effects were observed following transfection of miRNA (Figure 3B).

The success of myo-miRNA-based cardiac reprogramming was determined by qRT-PCR analysis of cardiac specific marker genes four weeks post transfection. Compared to control cells, cardiac induction medium II was found to be the most effective treatment leading to an induction of α-actinin, TBX5, GJA1, and cardiac Troponin I. While the level of α-actinin mRNA was strongly increased in all three cell types, a less pronounced effect was observed for cardiac Troponin I (Figure 3C). Notably, adMSCs showed the highest expression levels of cardiac marker genes after the treatment with cardiac induction medium II, when compared to MSCs obtained from dental follicle as well as BM, and therefore have been identified as the preferred candidate for our cardiac programming approach.

Surprisingly, our data also revealed that transfection with myo-miRNAs did not provoke an additional, beneficial effect on the expression of cardiac markers. Likewise, the cardiac induction medium containing RPMI and small molecules (Figure 3C, card induction I) did not promote the cardiac differentiation of MSC.

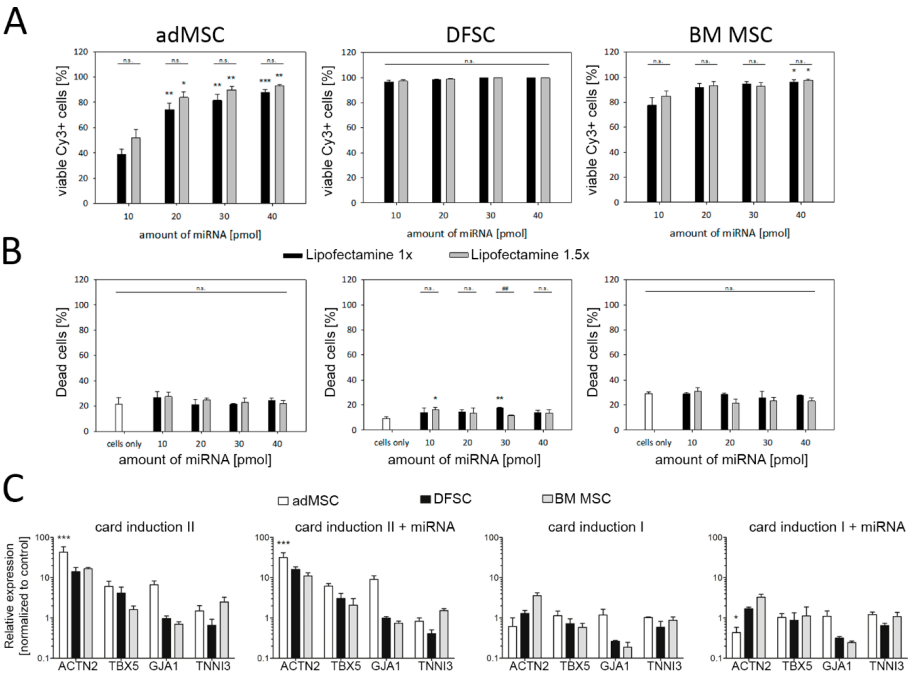


Figure 3. miRNA transfection and programming efficiency in MSC. (**A**) Uptake of miRNA was determined using Cy3-labelled miRNA and flow cytometry. (**B**) Detection of dead cells revealed low cytotoxicity induced by miRNA transfection. (**C**) Relative expression of cardiac marker genes among all tested cell types, four weeks after transfection and cultivation under different culture conditions. Reprogramming efficiency with cardiac induction medium II and myo-miRNAs (miR-1, miR-499, miR-208, miR-133) resulted in an up-regulation of cardiac specific markers in all types of MSC, while most profound up-regulation was found for cardiac induction medium II. Among tested MSC, the strongest increase of cardiac gene expression was observed for adMSC. Note, no beneficial effects on cardiac programming were observed following myo-miRNA transfection. Data are shown as mean ± SEM, obtained from three donors for each MSC type. Statistical analysis was performed using ANOVA test, followed by Bonferroni post-hoc analysis. * $p \leq 0.5$, ** $p \leq 0.05$, *** $p \leq 0.001$.

Figure 3. miRNA transfection and programming efficiency in MSC. (**A**) Uptake of miRNA was determined using Cy3-labelled miRNA and flow cytometry. (**B**) Detection of dead cells revealed low cytotoxicity induced by miRNA transfection. (**C**) Relative expression of cardiac marker genes among all tested cell types, four weeks after transfection and cultivation under different culture conditions. Reprogramming efficiency with cardiac induction medium II and myo-miRNAs (miR-1, miR-499, miR-208, miR-133) resulted in an up-regulation of cardiac specific markers in all types of MSC, while most profound up-regulation was found for cardiac induction medium II. Among tested MSC, the strongest increase of cardiac gene expression was observed for adMSC. Note, no beneficial effects on cardiac programming were observed following myo-miRNA transfection. Data are shown as mean ± SEM, obtained from three donors for each MSC type. Statistical analysis was performed using ANOVA test, followed by Bonferroni post-hoc analysis. * $p \leq 0.5$, ** $p \leq 0.05$, *** $p \leq 0.001$.

medium containing RPMI and small molecules (Figure 3C, card induction I) did not promote the cardiac differentiation of MSC.

3.3. mRNA-Based Reprogramming of adMSC

As the transfection of miRNA did not further improve cardiac differentiation, we asked whether the application of modified mRNAs might boost the reprogramming efficiency in adMSC, which had been found to be the most promising cell type for the differentiation towards the cardiac lineage (Figure 3).

For mRNA-based programming of adMSC, cells were either transfected with single MESP1 mRNA or with a combination of GATA4, MEF2C, and TBX5 mRNA (GMT). First, mRNA transfection and translation efficiency were determined with mRNA encoding GFP to evaluate the optimal amount of mRNA showing strong expression while causing minimal cytotoxic effects. As demonstrated by flow cytometry and fluorescence microscopy, approximately 80% of cells express the GFP protein 24 h post transfection with 1 µg of mRNA (Figure 4A–C). Considering the increasing cytotoxicity when higher amounts of mRNA are transfected, reprogramming experiments were performed with 1–2 µg of individual mRNA (Figure 4D).

Analysis by qRT-PCR showed that both MESP1 and GMT transfection resulted in elevated levels of selected cardiac marker genes, compared to untreated control cells (Figure 4E). The most prominent incline of gene expression was observed for α-actinin, which was confirmed on the protein level by immunostaining showing a faint signal in cells treated with MESP1 and GMT mRNAs (Figure 4F). Additional antibody staining of early cardiac transcription factors demonstrated the expression of MEF2C and NKX2.5 on the protein level in GMT treated cells (Figure 4G and Figure S1). Interestingly, a profound increase of the expression level was also found for TBX5 that has been used for mRNA transfection in the GMT-treated group, verified by fluorescence microscopy (Figure S1).

Moreover, we observed differences of the intracellular Ca^{2+} concentration between treated groups. Following labelling of intracellular Ca^{2+}, GMT transfected cells demonstrated a more intensive fluorescence signal than observed for MESP1 treated cells and the control group (Figure S2).

To obtain a deeper understanding of the mRNA-induced effects on the gene expression profile of treated adMSC, we conducted a microarray analysis of cells that underwent cardiac reprogramming. The signal intensity values detected on each microarray had a similar spread after normalization, indicating a well-suited data quality for further downstream data analysis (Figure 5A). The PCA plot visualizes the differences in gene expression among treated groups, showing that control cells (blue) share a high similarity regarding their transcription profile (Figure 5B). In contrast, reprogramming with cardiac induction medium II (red), MESP1 (green), and GMT (purple) mRNA induced a strong donor-dependent alteration of gene levels, however, the treatment specific groups remain distinguishable from each other.

The numbers of significant total up-regulated and down-regulated transcripts are represented in Figure 5C, indicating a distinct change of gene expression following cardiac reprogramming. The highest number of genes differentially expressed was found in MESP1 (6669 transcripts) and GMT (5649) treated cells. Interestingly, more transcripts are down-regulated than up-regulated in most of the comparisons.

The corresponding Venn diagram (Figure 5D) compares the significantly expressed genes of the three different reprogramming approaches related to untreated control cells. The largest amount of transcripts (2828 transcripts, 33.6%) was found to be commonly regulated by all three treatments. The second largest proportion of differentially expressed genes is shared by GMT vs. Control and MESP1 vs. Control (1816 transcripts, 21.6%). Notably, the largest unique set of transcripts was found in cells transfected with MESP1 mRNA (1660 transcripts, 19.7%). A detailed comparison of up-regulated (Figure 5E, red) and down-regulated (Figure 5E, green) genes among these three reprogrammed groups indicates that the differences between MESP1 and GMT treatment vs. cardiac induction medium II are more profound (189 up-regulated, 276 down-regulated transcripts), while MESP1 and GMT only showed one differentially up-regulated transcript that was not previously up-regulated in other

comparisons (Figure 5E). A detailed list of differentially expressed genes found in all reprogrammed groups is shown in in Table S1.

Cells **2020**, *9*, x FOR PEER REVIEW 9 of 19

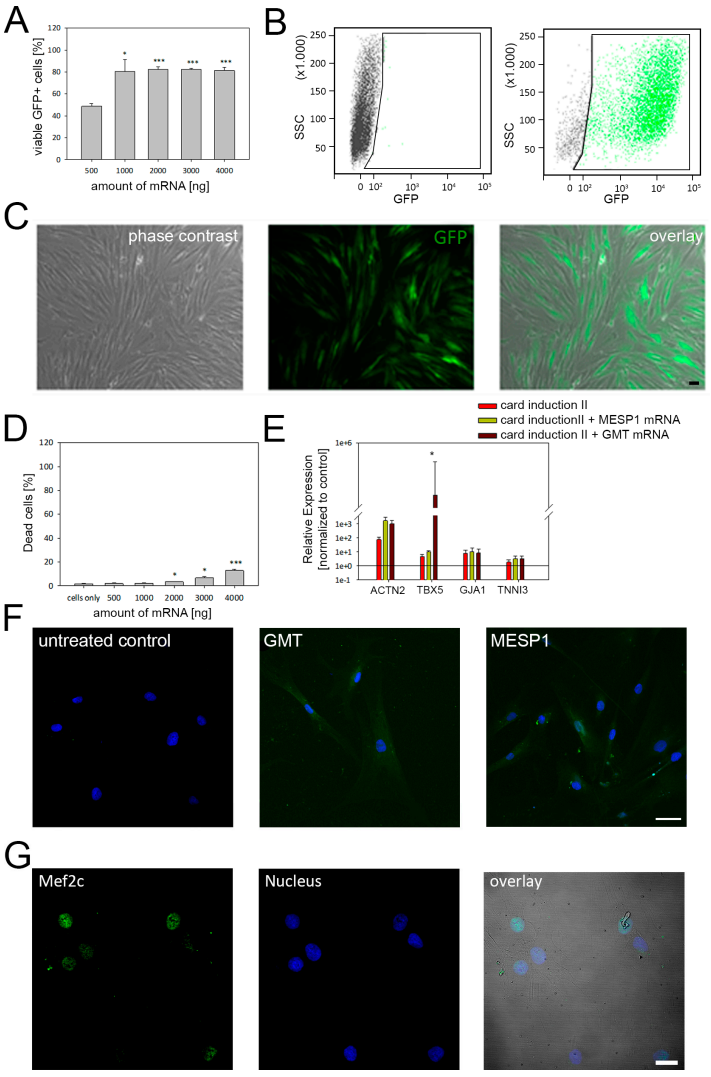


Figure 4. mRNA-based cardiac programming of adMSC. (**A**) Concentration-dependent expression of transfected mRNAs was evaluated with mRNA coding GFP. The quantitative flow cytometry analysis demonstrated maximum transfection efficiency of ~80% when ≤ 1000 ng mRNA were applied. (**B**) Representative scatterplots of control cells (left) and cells transfected with GFP mRNA (right). (**C**) Corresponding microscopy images of cells expressing GFP following mRNA treatment. (**D**) Cytotoxic effects were only induced when mRNA amounts higher than 1000 ng were used for transfection. (**E**) Compared to untreated control cells, higher gene expression levels of selected cardiac markers were detected for all reprogramming conditions, in particular for α-actinin. (**F**) Immunolabeling of cells using anti α-actinin antibody results in a faint fluorescence signal in cells transfected with MESP1 and GATA4, MEF2C, and TBX5 (GMT) mRNAs, Scale Bar: 25 µm. (**G**) Moreover, GMT treated cells also demonstrated protein expression of MEF2C, an early cardiac transcription factor. Flow cytometry and qRT-PCR data are shown as mean ± SEM, obtained from three different donors. Statistical analysis was performed using one-way ANOVA. * $p \leq 0.5$, ** $p \leq 0.05$, *** $p \leq 0.001$.

Figure 4. mRNA-based cardiac programming of adMSC. (**A**) Concentration-dependent expression of transfected mRNAs was evaluated with mRNA coding GFP. The quantitative flow cytometry analysis demonstrated maximum transfection efficiency of ~80% when ≤ 1000 ng mRNA were applied. (**B**) Representative scatterplots of control cells (left) and cells transfected with GFP mRNA (right). (**C**) Corresponding microscopy images of cells expressing GFP following mRNA treatment. (**D**) Cytotoxic effects were only induced when mRNA amounts higher than 1000 ng were used for transfection. (**E**) Compared to untreated control cells, higher gene expression levels of selected cardiac markers were detected for all reprogramming conditions, in particular for α-actinin. (**F**) Immunolabeling of cells using anti α-actinin antibody results in a faint fluorescence signal in cells transfected with MESP1 and GATA4, MEF2C, and TBX5 (GMT) mRNAs, Scale Bar: 25 µm. (**G**) Moreover, GMT treated cells also demonstrated protein expression of MEF2C, an early cardiac transcription factor. Flow cytometry and qRT-PCR data are shown as mean ± SEM, obtained from three different donors. Statistical analysis was performed using one-way ANOVA. * $p \leq 0.5$, ** $p \leq 0.05$, *** $p \leq 0.001$.

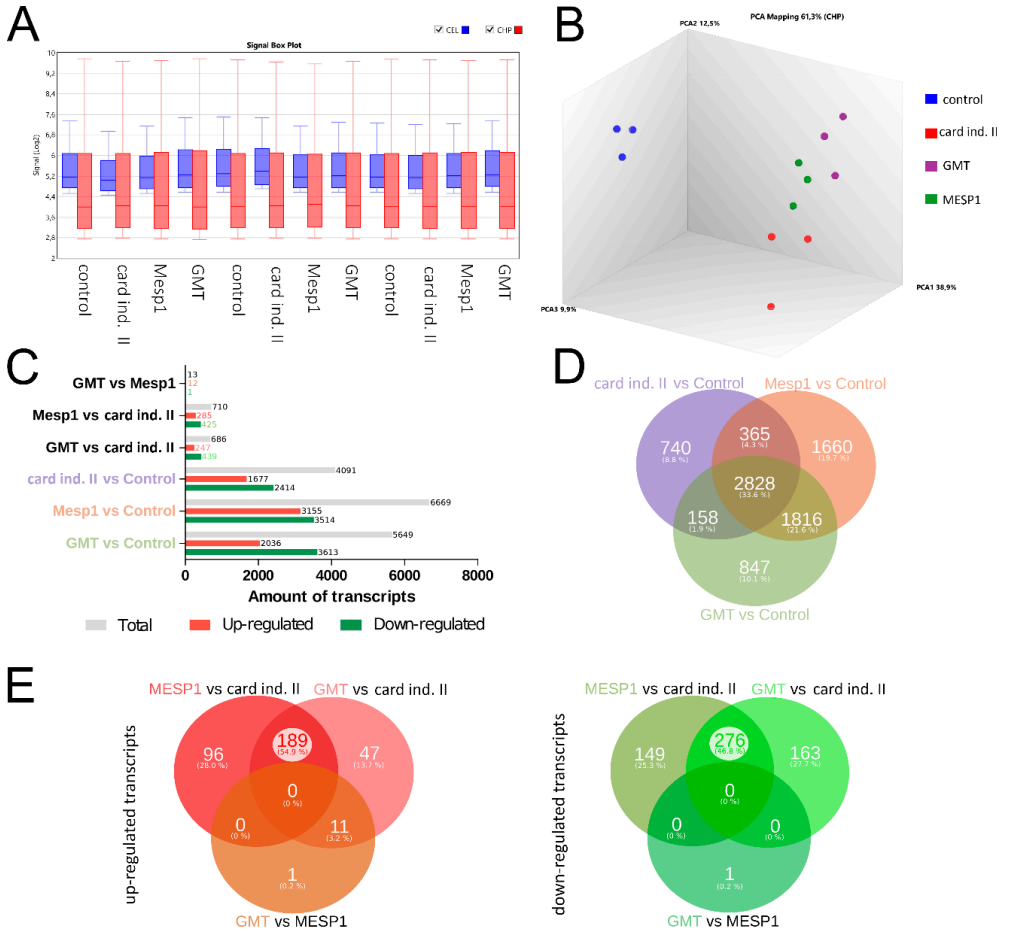


Figure 5. Transcriptome based comparison of reprogrammed adMSC. (**A**) Quality control of microarray data. Box plot of signal intensity of performed microarrays on .cel (blue) and .chp files normalization (red) confirm good data quality. (**B**) Principal component analysis (PCA) demonstrates clustering of treated groups, clearly showing the impact of respective reprogramming conditions on the transcriptomic profile compared to control cells (blue). Yet, cells subjected to MESP1 (green), GMT (purple) or cardiac induction medium II solely (red) remain distinguishable. (**C**) Up-and down-regulated transcripts and corresponding Venn diagram (**D**) showing the impact of reprogrammed cells compared to control. Most differentially expressed transcripts were regulated by all three reprogramming treatments (2828 genes), while 1816 transcripts are shared by GMT vs. control and MESP1 vs. control. (**E**) Detailed comparison of common and distinct up-regulated (red) and down-regulated (green) transcripts among the three reprogrammed groups. The differences found for optimized medium vs. MESP1 and GMT transfections are much more prominent than the differences between MESP1 and GMT.

These data indicate a strong change of gene expression when cells are subjected to cardiac induction medium II, with more distinct effects induced by mRNA transfections.

To evaluate the influence of the differentially expressed genes on important cardiac development pathways, we integrated our microarray gene expression data into the WikiPathways database and identified significantly enriched pathways for "heart development" (Figure 6A) and "cardiac progenitor differentiation" (Figure 6B). The pathway visualization indicates proteins mainly involved in cardiac development, while up-regulated and down-regulated transcripts of respective programming treatments are labelled in red and green, respectively. As shown in Figure 6, cardiac induction medium II as well as mRNA programming by MESP1 and GMT influence the gene expression profile of several key transcription factors and signaling molecules involved in cardiac differentiation, such as IGF, VEGF, TBX5, GATA4 and HAND2 (Figure 6A,B). Most changes on

treatment were labelled in red or green, respectively (Figure 6). As shown by MESP1 (60%) and induction medium II as well as 2nd. Mi programming by MESP1 and GMT influence the gene expression profile of several key transcription factors and signaling molecules involved in cardiac differentiation, such as IGF, VEGF, TBX5, GATA4 and HAND2 (Figure 6A,B). Most changes on pathway genes were induced by GMT treatment (92%), followed by MESP1 (60%) and cardiac induction medium (52%). Additional immunofluorescence labelling of GMT treated cells, confirmed the expression of early cardiac transcription factors, including NKX2.5, TBX5 and MEF2C (Figures 4G and S1).

Taken together, the results obtained by microarray analysis clearly indicate that reprogramming with cardiac induction medium II and mRNA induced a strong alteration of the transcription patterns, with high similarity in mRNA transfected cells compared to cells cultured in cardiac induction medium solely.

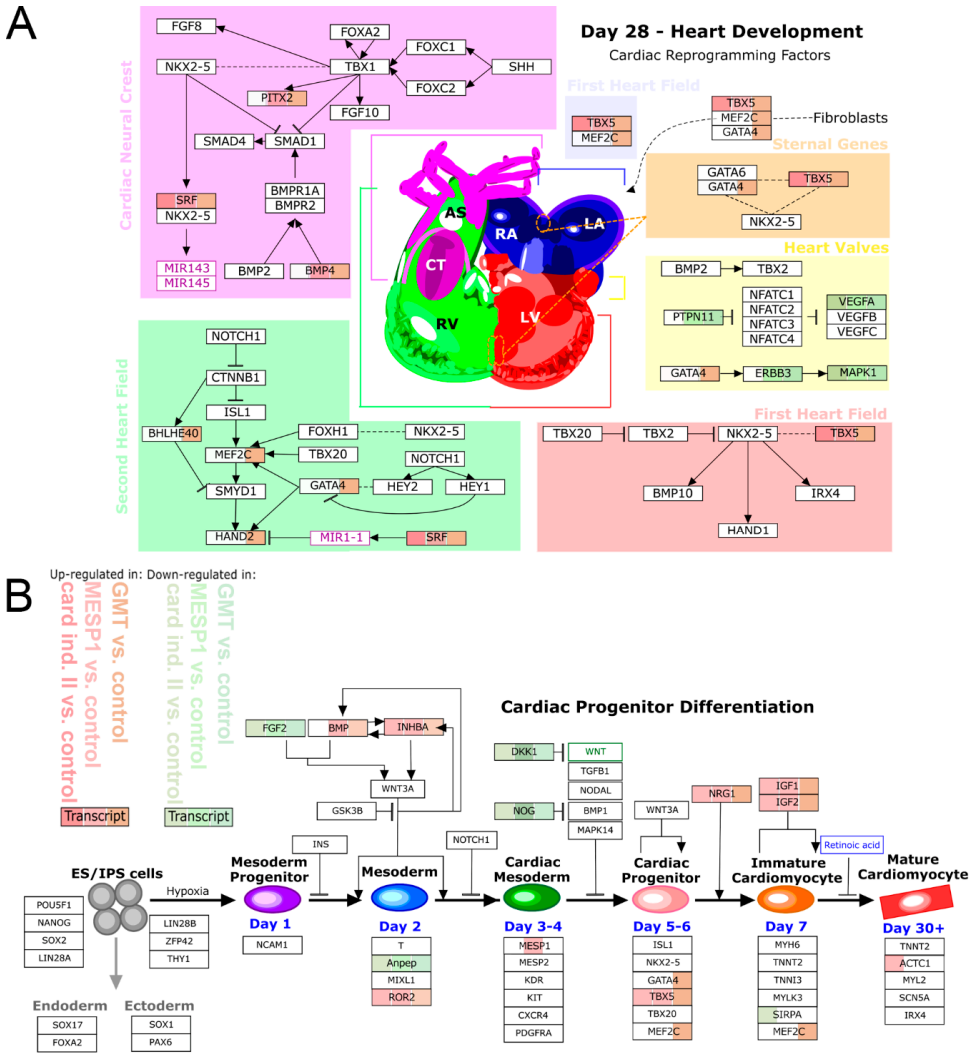


Figure 6. The impact of reprogramming on cardiac-differentiation pathways. Up-regulated and down-regulated transcripts of respective programming conditions are labelled in red or green color. (**A**, **B**) Strongest up-regulation of transcripts involved in cardiac development ((**A**) heart development, (**B**) cardiac progenitor differentiation) was mainly found in GMT reprogrammed cells, followed by MESP1 treatment and cardiac induction medium II. Key cardiac transcription factors and signaling molecules were significantly up-regulated, including TBX5, GATA4, MEF2C, HAND2, BMP4, and IGF.

Taken together, the results obtained by microarray analysis clearly indicate that reprogramming with cardiac induction medium II and mRNA induced a strong alteration of the transcription patterns with high similarity in mRNA transfected cells compared to cells cultured in cardiac induction medium solely.

4. Discussion

In vitro generated cardiomyocytes are an important tool for cardiovascular research, as they can be utilized for disease modelling or for the development of drug screening assays to assess the cardiac toxic risk of established or newly synthesized drugs [37–39]. Moreover, promising preclinical data suggests the therapeutic potential of generated cardiomyocytes for the treatment of cardiac diseases to overall improve heart regeneration and function [40,41]. Although several stem cell types are available to produce cardiac cells, the ideal source of stem cells remains elusive as each has its own advantages and drawbacks. Adult MSC can be easily isolated from human donors in large quantities, possess immunomodulatory properties and can be propagated in vitro [12]. Further, they can overcome certain limitations that have been attributed to PSCs, including ESC and iPSC. In contrast to ESC, MSC do not provoke any ethical concerns [12,37,38]. Moreover, pre-clinical studies demonstrated a tumorigenic potential of ESC and iPSC-derived cell products that has not been observed for MSC to date [42–45]. However, other pre-clinical and clinical trial data showed that the transplantation of iPSCs-derived cardiomyocytes did not result in teratoma formation [46–48]. These different outcomes might be associated with the transplantation of residual undifferentiated cells along with the PSC product that increases the possibility of tumorigenesis. In this regard, the therapeutic use of PSC requires the establishment of differentiation protocols allowing the generation of highly pure PSC-derived cell types, e.g., cardiomyocytes [49]. The major advantage in comparison to adult stem cells is the cardiac differentiation potential of ESCs and iPSCs. So far, PSC have been found to be the only stem cell type capable to differentiate into functional, premature cardiomyocytes showing pronounced sarcomere organization, contraction capacity, and subtype specific ion channel composition [50,51]. Thus, for the generation of cardiomyocytes applied in regenerative medicine PSC are currently superior to MSC as no efficient cardiac reprogramming strategies have been developed for adult stem cells yet.

The successful cardiac differentiation of human MSC into fully mature cardiomyocytes is by far more challenging. Adult cardiomyocytes are characterized by a specific cell shape, structural organization, ion channel composition and mechanical properties; important features that need to be addressed when generating stem cell-derived cardiac cells [52]. Former reports led to contradictory results about the programming efficiency of MSC. While some reports described spontaneous beating associated with the formation of sarcomeric protein structures, other studies failed to generate cardiac-like cells from adult MSC [53–57].

One reason for this might be attributed to the fact that MSC may represent a heterogeneous stem cell population with different functional and phenotype-related properties as well as varying therapeutic potential [58]. A notion that is supported by our microarray data, indicating a high diversity of the expressed transcripts among MSC obtained from BM, dental pulp and adipose tissue (Figure 2). Likewise, our functional data revealed cell type-dependent differentiation capacity of tested MSC (Figure 1). Previous studies have also reported distinct characteristics between MSC from different sources regarding surface marker expression, proliferation rate, and differentiation potency [17,19,58,59]. For example, adMSC were observed to favor osteogenic differentiation and demonstrate higher proliferation when compared with DSFCs [18,60]. Moreover, our results suggest that these different biological characteristics of MSC could have an impact on the selected strategy and efficiency of cardiac programming as adMSC demonstrated a more pronounced incline of cardiac marker expression than BM MSC and DFSCs (Figure 3). In line with these data, Kakkar et al. recently described human adMSC to be a better choice for cardiac programming using a combination of small molecules and cytokines. Compared to BM MSC, adMSC exhibited a higher expression of α-actinin, troponin and connexin43 following cardiac induction with 5-Azacytidine and TGF-β1 [61].

Similarly, a comparative study revealed that adMSC expressed significantly more cardiomyocyte specific biomarkers as DFSCs following cardiac programming with cytokine supplemented culture medium [11]. The impact of MSC origin on programming capability was also shown for non-cardiac cell lineages like hepatocytes and smooth muscle cells [59,62].

Myo-miRNA based programming has been successfully applied for the conversion of cardiac fibroblasts, into cardiomyocytes [36]. For MSCs, cardiac induction by miRNA is less efficient as shown by different groups [25,63,64]. For example, it was demonstrated that transfection with miRNA-1-2 promote the expression of GATA4, NKX2.5 and cardiac Troponin in BM MSCs [15]. Similarly, miR-149 and miR-1 were found to slightly trigger myocardial differentiation, albeit without formation of sarcomere structures or beating activity [25,65]. We did not observe any additional effects on the expression of selected cardiac marker genes following miRNA treatment. This might be attributed to the fact that the miRNA concentrations used in this study are not sufficient to significantly increase the expression level of cardiac-specific genes, although uptake efficiency for miRNA was about 80%. In this regard, some studies have used viral vectors to ensure constitutive overexpression of miRNA [25,64]. Given that miRNAs have a very short half live, transient transfection approaches, as used in our study, might be less effective.

Proper cardiac development requires the activation and inhibition of many different pathways modulated by several transcription factors [66]. MESP1 was shown to drive cardiovascular fate of stem cells during embryonic development, while the combination of GATA4, MEF2C and TBX5 was described to induce the cardiac differentiation of murine and human fibroblasts, leading to spontaneously contracting cells with cardiomyocyte-like expression profile [67–70]. Therefore, we have concluded that this approach might be applicable to reprogram human adMSC. Using an mRNA-based setting we induced the overexpression of GATA4, MEF2C, and TBX5 as well as MESP1, which provoked an incline of genes involved in cardiac differentiation (Figure 4). To our knowledge this combination of transcription factors has not been applied before to induce cardiac differentiation of human adMSC. In contrast to our strategy, most of the previous studies performed overexpression of transcription factors by application of retro- or lentiviral systems. For example, in a study by Wystrychowski et al., adMSC from cardiac tissue were treated with seven transcription factors, including GATA4, MEF2C, MESP1, and TBX5, that resulted in an elevated number of cells positive for α-actinin and troponin [71]. However, no clear sarcomere structures have been observed, suggesting a premature cardiac progenitor state. Similarly, forced expression of another factor of the T-box family, TBX20, provokes an up-regulation of sarcomeric proteins, without cardiomyocyte specific sarcomere organization [72]. These data are in line with our observations as we could also detect a moderate signal for α-actinin, albeit without the presence of sarcomere structures (Figure 4).

Yet, our programming approach leads to a strong induction of the key cardiac transcription factors GATA4, MEF2C, MESP1 and TBX5, which corresponds to the transfected mRNAs used for programming. However, it is known that mRNAs underlie fast turnover, suggesting that mRNA transfection activated the expression of its endogenous counterparts [73,74]. At the same time, the current study demonstrates that mRNA transfection boosts the cardiac programming effects induced by culture conditions targeting important signaling pathways such as the WNT cascade.

The manipulation of signaling pathways by cytokines and small molecules is the most common methodology to generate large amounts of PSC-derived functional cardiomyocytes [30,31]. In addition, the overexpression of transcription factors, like Tbx3 and MESP1, can influence cell fate decision in PSCs [75,76]. While these techniques allow highly efficient programming of ESCs and iPSCs, we observed significantly less programming efficiency for MSCs in the current study. However, the comparison of programming protocols used for PSCs and multipotent stem cells is difficult due to their different developmental stages and resulting culture conditions prerequisites. Yet, it was shown that cytokines like BMP4, IL and TGF improve cardiac development of human and non-human MSCs [57,77]. However, the cardiomyocyte-like cells derived from these programmed MSCs lack profound sarcomere formation, beating activity and physiological maturation [78,79]. This is in

accordance to our data indicating that mRNA transfection could promote the expression of early cardiac proteins, while differentiation efficiency and elaboration of a terminal cardiac phenotype is profoundly limited when compared to PSC differentiation protocols [27,31].

Together with previous studies of adMSC overexpressing transcription factors, our results demonstrate the feasibility of mRNA-based cardiac reprogramming of MSC. However, the absence of sarcomere structures and spontaneous cell beating suggests a yet quite incomplete reprogramming, leading to an immature cardiac cell type. Hence, there is an urgent need for further optimization. Since mRNAs are degraded over time, multiple transfection steps might increase the reprogramming efficiency, a strategy that is already applied for the generation of iPSCs from adult cells [74,80]. Moreover, proportions of GATA4, MEF2C, and TBX5 protein expression has been described to play a crucial role for the quality of cardiac reprogramming [81], thus, different ratios of transfected mRNA could positively influence the outcome of reprogrammed adMSC. This will have to be addressed in future studies as the impact of mRNA ratios and mRNA concentration on cardiac programming might be affected in a donor specific manner. Former data already demonstrated donor-to-donor variability of MSC functional potential, including differentiation capacity [82,83]. Beside age and gender, underlying diseases are known to influence cellular properties of MSCs [82]. This is supported by our microarray results, showing a large variety of the transcription profile of BM MSCs that have been obtained from patients suffering from cardiovascular diseases. On the contrary, adMSCs and DFSCs derived from healthy donors shared similar transcription patterns, suggesting same programming conditions required to induce cardiac development. Nevertheless, it is recommended to adapt mRNA conditions for each individual patient to obtain maximum programming efficiency.

In addition, more comparative studies are required to identify and characterize MSC subtypes most susceptible for specific transdifferentiation towards the respective desired target cells, including non-mesodermal and mesodermal cell types such as cardiomyocytes.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4409/9/2/504/s1.

Author Contributions: P.M., H.L. (Hermann Lang) and R.D. performed the study design. P.M. carried out cell culture experiments, RNA isolation, flow cytometry, qRT-PCR and respective data analysis. M.W. supported analysis of microarray data, subfigure preparation and corrected the manuscript. K.E. isolated and pre-cultured the DFSC. K.P. and O.H. isolated, characterized and pre-cultured the adMSC. D.K. carried out microarray experiments, including RNA quality measurement. K.P., H.L. (Heiko Lemcke), C.I.L., O.W., and R.D. proofread and revised the manuscript. H.L. (Hermann Lang) collected microscopy data, conceptualized and wrote the manuscript with contribution from P.M. and R.D. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the EU structural Fund (ESF/14-BM-A55-0024/18). In addition, R.D and P.M. are supported by the DFG (DA1296/6-1). R.D. is further supported by the DAMP foundation, the German Heart Foundation (F/01/12) and the BMBF (VIP+ 00240). In addition, H.L. is supported by the FORUN Program of Rostock University Medical Centre (889001 and 889003) and the Josef and Käthe Klinz Foundation (T319/29737/2017).

Conflicts of Interest: The authors declare no conflict of interest. The funders were not involved in study design, data collection and interpretation, and manuscript preparation.

References

1. Rajabzadeh, N.; Fathi, E.; Farahzadi, R. Stem cell-based regenerative medicine. *Stem Cell Investig.* **2019**, *6*, 19. [CrossRef] [PubMed]
2. Samsonraj, R.M.; Raghunath, M.; Nurcombe, V.; Hui, J.H.; van Wijnen, A.J.; Cool, S.M. Concise Review: Multifaceted Characterization of Human Mesenchymal Stem Cells for Use in Regenerative Medicine. *Stem Cells Transl. Med.* **2017**, *6*, 2173–2185. [CrossRef] [PubMed]
3. Squillaro, T.; Peluso, G.; Galderisi, U. Clinical trials with mesenchymal stem cells: An update. *Cell Transplant.* **2016**, *25*, 829–848. [CrossRef] [PubMed]
4. Collichia, M.; Jones, D.A.; Beirne, A.-M.; Hussain, M.; Weeraman, D.; Rathod, K.; Veerapen, J.; Lowdell, M.; Mathur, A. Umbilical cord-derived mesenchymal stromal cells in cardiovascular disease: review of preclinical and clinical data. *Cytotherapy* **2019**, *21*, 1007–1018. [CrossRef]

5. Guerrouahen, B.S.; Sidahmed, H.; Al Sulaiti, A.; Al Khulaifi, M.; Cugno, C. Enhancing Mesenchymal Stromal Cell Immunomodulation for Treating Conditions Influenced by the Immune System. *Stem Cells Int.* **2019**, *2019*, 7219297. [CrossRef]
6. Aguilera-Castrejon, A.; Pasantes-Morales, H.; Montesinos, J.J.; Cortés-Medina, L.V.; Castro-Manrreza, M.E.; Mayani, H.; Ramos-Mandujano, G. Improved Proliferative Capacity of NP-Like Cells Derived from Human Mesenchymal Stromal Cells and Neuronal Transdifferentiation by Small Molecules. *Neurochem. Res.* **2017**, *42*, 415–427. [CrossRef]
7. Tsai, W.-L.; Yeh, P.-H.; Tsai, C.-Y.; Ting, C.-T.; Chiu, Y.-H.; Tao, M.-H.; Li, W.-C.; Hung, S.-C. Efficient programming of human mesenchymal stem cell-derived hepatocytes by epigenetic regulations. *J. Gastroenterol. Hepatol.* **2017**, *32*, 261–269. [CrossRef]
8. Papadimou, E.; Morigi, M.; Iatropoulos, P.; Xinaris, C.; Tomasoni, S.; Benedetti, V.; Longaretti, L.; Rota, C.; Todeschini, M.; Rizzo, P.; et al. Direct Reprogramming of Human Bone Marrow Stromal Cells into Functional Renal Cells Using Cell-free Extracts. *Stem Cell Reports* **2015**, *4*, 685–698. [CrossRef]
9. Cai, B.; Li, J.; Wang, J.; Luo, X.; Ai, J.; Liu, Y.; Wang, N.; Liang, H.; Zhang, M.; Chen, N.; et al. microRNA-124 Regulates Cardiomyocyte Differentiation of Bone Marrow-Derived Mesenchymal Stem Cells Via Targeting STAT3 Signaling. *Stem Cells* **2012**, *30*, 1746–1755. [CrossRef]
10. Li, J.; Zhu, K.; Wang, Y.; Zheng, J.; Guo, C.; Lai, H.; Wang, C. Combination of IGF-1 gene manipulation and 5-AZA treatment promotes differentiation of mesenchymal stem cells into cardiomyocyte-like cells. *Mol. Med. Rep.* **2015**, *11*, 815–820. [CrossRef]
11. Loo, Z.X.; Kunasekaran, W.; Govindasamy, V.; Musa, S.; Abu Kasim, N.H. Comparative analysis of cardiovascular development related genes in stem cells isolated from deciduous pulp and adipose tissue. *Sci. World J.* **2014**, *2014*, 186508. [CrossRef] [PubMed]
12. Müller, P.; Lemcke, H.; David, R. Stem Cell Therapy in Heart Diseases—Cell Types, Mechanisms and Improvement Strategies. *Cell. Physiol. Biochem.* **2018**, *48*, 2607–2655. [CrossRef] [PubMed]
13. Szaraz, P.; Gratch, Y.S.; Iqbal, F.; Librach, C.L. In Vitro Differentiation of Human Mesenchymal Stem Cells into Functional Cardiomyocyte-like Cells. *J. Vis. Exp.* **2017**, *9*, 55757. [CrossRef] [PubMed]
14. Markmee, R.; Aungsuchawan, S.; Narakornsak, S.; Tancharoen, W.; Bumrungkit, K.; Pangchaidee, N.; Pothacharoen, P.; Puaninta, C. Differentiation of mesenchymal stem cells from human amniotic fluid to cardiomyocyte-like cells. *Mol. Med. Rep.* **2017**, *16*, 6068–6076. [CrossRef]
15. Shen, X.; Pan, B.; Zhou, H.; Liu, L.; Lv, T.; Zhu, J.; Huang, X.; Tian, J. Differentiation of mesenchymal stem cells into cardiomyocytes is regulated by miRNA-1-2 via WNT signaling pathway. *J. Biomed. Sci.* **2017**, *24*, 29. [CrossRef]
16. O'Connor, K.C. Molecular Profiles of Cell-to-Cell Variation in the Regenerative Potential of Mesenchymal Stromal Cells. *Stem Cells Int.* **2019**, *2019*, 1–14. [CrossRef]
17. Elahi, K.C.; Klein, G.; Avci-Adali, M.; Sievert, K.D.; MacNeil, S.; Aicher, W.K. Human Mesenchymal Stromal Cells from Different Sources Diverge in Their Expression of Cell Surface Proteins and Display Distinct Differentiation Patterns. *Stem Cells Int.* **2016**, *2016*, 1–9. [CrossRef]
18. D'Alimonte, I.; Mastrangelo, F.; Giuliani, P.; Pierdomenico, L.; Marchisio, M.; Zuccarini, M.; Di Iorio, P.; Quaresima, R.; Caciagli, F.; Ciccarelli, R. Osteogenic Differentiation of Mesenchymal Stromal Cells: A Comparative Analysis Between Human Subcutaneous Adipose Tissue and Dental Pulp. *Stem Cells Dev.* **2017**, *26*, 843–855. [CrossRef]
19. Kwon, A.; Kim, Y.; Kim, M.; Kim, J.; Choi, H.; Jekarl, D.W.; Lee, S.; Kim, J.M.; Shin, J.-C.; Park, I.Y. Tissue-specific Differentiation Potency of Mesenchymal Stromal Cells from Perinatal Tissues. *Sci. Rep.* **2016**, *6*, 23544. [CrossRef]
20. Leijten, J.; Georgi, N.; Moreira Teixeira, L.; van Blitterswijk, C.A.; Post, J.N.; Karperien, M. Metabolic programming of mesenchymal stromal cells by oxygen tension directs chondrogenic cell fate. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 13954–13959. [CrossRef]
21. Occhetta, P.; Pigeot, S.; Rasponi, M.; Dasen, B.; Mehrkens, A.; Ullrich, T.; Kramer, I.; Guth-Gundel, S.; Barbero, A.; Martin, I. Developmentally inspired programming of adult human mesenchymal stromal cells toward stable chondrogenesis. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 4625–4630. [CrossRef] [PubMed]
22. Yannarelli, G.; Pacienza, N.; Montanari, S.; Santa-Cruz, D.; Viswanathan, S.; Keating, A. OCT4 expression mediates partial cardiomyocyte reprogramming of mesenchymal stromal cells. *PLoS ONE* **2017**, *12*, e0189131. [CrossRef] [PubMed]

23. Lemcke, H.; Gaebel, R.; Skorska, A.; Voronina, N.; Lux, C.A.; Petters, J.; Sasse, S.; Zarniko, N.; Steinhoff, G.; David, R. Mechanisms of stem cell based cardiac repair-gap junctional signaling promotes the cardiac lineage specification of mesenchymal stem cells. *Sci. Rep.* **2017**, *7*, 1–17. [CrossRef] [PubMed]
24. Li, L.; Xia, Y. Study of adipose tissue-derived mesenchymal stem cells transplantation for rats with dilated cardiomyopathy. *Ann. Thorac. Cardiovasc. Surg.* **2014**, *20*, 398–406. [CrossRef]
25. Zhao, X.-L.; Yang, B.; Ma, L.-N.; Dong, Y.-H. MicroRNA-1 effectively induces differentiation of myocardial cells from mouse bone marrow mesenchymal stem cells. *Artif. Cells Nanomed. Biotechnol.* **2015**, *44*, 1665–1670. [CrossRef]
26. Dai, F.; Du, P.; Chang, Y.; Ji, E.; Xu, Y.; Wei, C.; Li, J. Downregulation of MiR-199b-5p inducing differentiation of bone-marrow mesenchymal stem cells (BMSCs) toward cardiomyocyte-like cells via HSF1/HSP70 pathway. *Med. Sci. Monit.* **2018**, *24*, 2700–2710. [CrossRef]
27. Burridge, P.W.; Matsa, E.; Shukla, P.; Lin, Z.C.; Churko, J.M.; Ebert, A.D.; Lan, F.; Diecke, S.; Huber, B.; Mordwinkin, N.M.; et al. Chemically defned generation of human cardiomyocytes. *Nat. Methods* **2014**, *11*, 855–860. [CrossRef]
28. Jiang, Y.; Park, P.; Hong, S.M.; Ban, K. Maturation of cardiomyocytes derived from human pluripotent stem cells: Current strategies and limitations. *Mol. Cells* **2018**, *41*, 613–621.
29. Chen, R.; He, J.; Wang, Y.; Guo, Y.; Zhang, J.; Peng, L.; Wang, D.; Lin, Q.; Zhang, J.; Guo, Z.; et al. Qualitative transcriptional signatures for evaluating the maturity degree of pluripotent stem cell-derived cardiomyocytes. *Stem Cell Res. Ther.* **2019**, *10*, 113. [CrossRef]
30. D'Antonio-Chronowska, A.; Donovan, M.K.R.; Young Greenwald, W.W.; Nguyen, J.P.; Fujita, K.; Hashem, S.; Matsui, H.; Soncin, F.; Parast, M.; Ward, M.C.; et al. Association of Human iPSC Gene Signatures and X Chromosome Dosage with Two Distinct Cardiac Differentiation Trajectories. *Stem Cell Rep.* **2019**, *13*, 924–938. [CrossRef]
31. Lian, X.; Zhang, J.; Azarin, S.M.; Zhu, K.; Hazeltine, L.B.; Bao, X.; Hsiao, C.; Kamp, T.J.; Palecek, S.P. Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/β-catenin signaling under fully defined conditions. *Nat. Protoc.* **2013**, *8*, 162–175. [CrossRef] [PubMed]
32. Meyer, J.; Salamon, A.; Herzmann, N.; Adam, S.; Kleine, H.-D.; Matthiesen, I.; Ueberreiter, K.; Peters, K. Isolation and Differentiation Potential of Human Mesenchymal Stem Cells From Adipose Tissue Harvested by Water Jet-Assisted Liposuction. *Aesthetic Surg. J.* **2015**, *35*, 1030–1039. [CrossRef] [PubMed]
33. Müller, P.; Ekat, K.; Brosemann, A.; Köntges, A.; David, R.; Lang, H. Isolation, characterization and microRNA-based genetic modification of human dental follicle stem cells. *J. Vis. Exp.* **2018**, *2018*, e58089.
34. Thiele, F.; Voelkner, C.; Krebs, V.; Müller, P.; Jung, J.J.; Rimmbach, C.; Steinhoff, G.; Noack, T.; David, R.; Lemcke, H. Nkx2.5 Based Ventricular Programming of Murine ESC-Derived Cardiomyocytes. *Cell. Physiol. Biochem.* **2019**, *53*, 337–354.
35. Koczan, D.; Fitzner, B.; Zettl, U.K.; Hecker, M. Microarray data of transcriptome shifts in blood cell subsets during S1P receptor modulator therapy. *Sci. Data* **2018**, *5*, 180145. [CrossRef]
36. Jayawardena, T.M.; Egemnazarov, B.; Finch, E.A.; Zhang, L.; Payne, J.A.; Pandya, K.; Zhang, Z.; Rosenberg, P.; Mirotsou, M.; Dzau, V.J. MicroRNA-mediated in vitro and in vivo direct reprogramming of cardiac fibroblasts to cardiomyocytes. *Circ. Res.* **2012**, *110*, 1465–1473. [CrossRef]
37. Sala, L.; Gnecchi, M.; Schwartz, P.J. Long QT Syndrome Modelling with Cardiomyocytes Derived from Human-induced Pluripotent Stem Cells. *Arrhythmia Electrophysiol. Rev.* **2019**, *8*, 105. [CrossRef]
38. Brodehl, A.; Ebbinghaus, H.; Deutsch, M.-A.; Gummert, J.; Gärtner, A.; Ratnavadivel, S.; Milting, H. Human Induced Pluripotent Stem-Cell-Derived Cardiomyocytes as Models for Genetic Cardiomyopathies. *Int. J. Mol. Sci.* **2019**, *20*, 4381. [CrossRef]
39. Protze, S.I.; Lee, J.H.; Keller, G.M. Human Pluripotent Stem Cell-Derived Cardiovascular Cells: From Developmental Biology to Therapeutic Applications. *Cell Stem Cell* **2019**, *25*, 311–327. [CrossRef]
40. Rikhtegar, R.; Pezeshkian, M.; Dolati, S.; Safaie, N.; Afrasiabi Rad, A.; Mahdipour, M.; Nouri, M.; Jodati, A.R.; Yousefi, M. Stem cells as therapy for heart disease: iPSCs, ESCs, CSCs, and skeletal myoblasts. *Biomed. Pharmacother.* **2019**, *109*, 304–313. [CrossRef]
41. Jackson, A.O.; Tang, H.; Yin, K. HiPS-Cardiac Trilineage Cell Generation and Transplantation: a Novel Therapy for Myocardial Infarction. *J. Cardiovasc. Transl. Res.* **2019**, *13*, 110–119. [CrossRef] [PubMed]

42. Hentze, H.; Soong, P.L.; Wang, S.T.; Phillips, B.W.; Putti, T.C.; Dunn, N.R. Teratoma formation by human embryonic stem cells: Evaluation of essential parameters for future safety studies. *Stem Cell Res.* **2009**, *2*, 198–210. [CrossRef] [PubMed]
43. Yong, K.W.; Choi, J.R.; Dolbashid, A.S.; Wan Safwani, W.K.Z. Biosafety and bioefficacy assessment of human mesenchymal stem cells: What do we know so far? *Regen. Med.* **2018**, *13*, 219–232. [CrossRef]
44. Duinsbergen, D.; Salvatori, D.; Eriksson, M.; Mikkers, H. Tumors Originating from Induced Pluripotent Stem Cells and Methods for Their Prevention. *Ann. N. Y. Acad. Sci.* **2009**, *1176*, 197–204. [CrossRef] [PubMed]
45. Seminatore, C.; Polentes, J.; Ellman, D.; Kozubenko, N.; Itier, V.; Tine, S.; Tritschler, L.; Brenot, M.; Guidou, E.; Blondeau, J.; et al. The postischemic environment differentially impacts teratoma or tumor formation after transplantation of human embryonic stem cell-derived neural progenitors. *Stroke* **2010**, *41*, 153–159. [CrossRef] [PubMed]
46. Menasché, P.; Vanneaux, V.; Hagège, A.; Bel, A.; Cholley, B.; Parouchev, A.; Cacciapuoti, I.; Al-Daccak, R.; Benhamouda, N.; Blons, H.; et al. Transplantation of Human Embryonic Stem Cell–Derived Cardiovascular Progenitors for Severe Ischemic Left Ventricular Dysfunction. *J. Am. Coll. Cardiol.* **2018**, *71*, 429–438. [CrossRef] [PubMed]
47. Funakoshi, S.; Miki, K.; Takaki, T.; Okubo, C.; Hatani, T.; Chonabayashi, K.; Nishikawa, M.; Takei, I.; Oishi, A.; Narita, M.; et al. Enhanced engraftment, proliferation, and therapeutic potential in heart using optimized human iPSC-derived cardiomyocytes. *Sci. Rep.* **2016**, *6*, 1–14. [CrossRef]
48. Liu, Y.W.; Chen, B.; Yang, X.; Fugate, J.A.; Kalucki, F.A.; Futakuchi-Tsuchida, A.; Couture, L.; Vogel, K.W.; Astley, C.A.; Baldessari, A.; et al. Human embryonic stem cell-derived cardiomyocytes restore function in infarcted hearts of non-human primates. *Nat. Biotechnol.* **2018**, *36*, 597–605. [CrossRef]
49. Ito, E.; Miyagawa, S.; Takeda, M.; Kawamura, A.; Harada, A.; Iseoka, H.; Yajima, S.; Sougawa, N.; Mochizuki-Oda, N.; Yasuda, S.; et al. Tumorigenicity assay essential for facilitating safety studies of hiPSC-derived cardiomyocytes for clinical application. *Sci. Rep.* **2019**, *9*, 1–10. [CrossRef]
50. Oikonomopoulos, A.; Kitani, T.; Wu, J.C. Pluripotent Stem Cell-Derived Cardiomyocytes as a Platform for Cell Therapy Applications: Progress and Hurdles for Clinical Translation. *Mol. Ther.* **2018**, *26*, 1624–1634. [CrossRef] [PubMed]
51. Tan, S.H.; Ye, L. Maturation of Pluripotent Stem Cell-Derived Cardiomyocytes: A Critical Step for Drug Development and Cell Therapy. *J. Cardiovasc. Transl. Res.* **2018**, *11*, 375–392. [CrossRef] [PubMed]
52. Scuderi, G.J.; Butcher, J. Naturally Engineered Maturation of Cardiomyocytes. *Front. Cell Dev. Biol.* **2017**, *5*, 50. [CrossRef] [PubMed]
53. Rose, R.A.; Jiang, H.; Wang, X.; Helke, S.; Tsoporis, J.N.; Gong, N.; Keating, S.C.J.; Parker, T.G.; Backx, P.H.; Keating, A. Bone marrow-derived mesenchymal stromal cells express cardiac-specific markers, retain the stromal phenotype, and do not become functional cardiomyocytes in vitro. *Stem Cells* **2008**, *26*, 2884–2892. [CrossRef] [PubMed]
54. Shim, W.S.N.; Jiang, S.; Wong, P.; Tan, J.; Chua, Y.L.; Seng Tan, Y.; Sin, Y.K.; Lim, C.H.; Chua, T.; Teh, M.; et al. Ex vivo differentiation of human adult bone marrow stem cells into cardiomyocyte-like cells. *Biochem. Biophys. Res. Commun.* **2004**, *324*, 481–488. [CrossRef] [PubMed]
55. Martin-Rendon, E.; Sweeney, D.; Lu, F.; Girdlestone, J.; Navarrete, C.; Watt, S.M. 5-Azacytidine-treated human mesenchymal stem/progenitor cells derived from umbilical cord, cord blood and bone marrow do not generate cardiomyocytes in vitro at high frequencies. *Vox Sang.* **2008**, *95*, 137–148. [CrossRef]
56. Ramkisoensing, A.A.; Pijnappels, D.A.; Askar, S.F.A.; Passier, R.; Swildens, J.; Goumans, M.J.; Schutte, C.I.; de Vries, A.A.F.; Scherjon, S.; Mummery, C.L.; et al. Human embryonic and fetal Mesenchymal stem cells differentiate toward three different cardiac lineages in contrast to their adult counterparts. *PLoS ONE* **2011**, *6*, e24164. [CrossRef]
57. Shi, S.; Wu, X.; Wang, X.; Hao, W.; Miao, H.; Zhen, L.; Nie, S. Differentiation of Bone Marrow Mesenchymal Stem Cells to Cardiomyocyte-Like Cells Is Regulated by the Combined Low Dose Treatment of Transforming Growth Factor-? 1 and 5-Azacytidine. *Stem Cells Int.* **2016**, *2016*, 11. [CrossRef]
58. Jin, H.J.; Bae, Y.K.; Kim, M.; Kwon, S.J.; Jeon, H.B.; Choi, S.J.; Kim, S.W.; Yang, Y.S.; Oh, W.; Chang, J.W. Comparative analysis of human mesenchymal stem cells from bone marrow, adipose tissue, and umbilical cord blood as sources of cell therapy. *Int. J. Mol. Sci.* **2013**, *14*, 17986–18001. [CrossRef]

59. Li, J.; Xu, S.; Zhao, Y.; Yu, S.; Ge, L.; Xu, B.; Yu, S.; Yu, S.; Ge, L.; Ge, L.; et al. Comparison of the biological characteristics of human mesenchymal stem cells derived from exfoliated deciduous teeth, bone marrow, gingival tissue, and umbilical cord. *Mol. Med. Rep.* **2018**, *18*, 4969–4977. [CrossRef]
60. Mohamed-Ahmed, S.; Fristad, I.; Lie, S.A.; Suliman, S.; Mustafa, K.; Vindenes, H.; Idris, S.B. Adipose-derived and bone marrow mesenchymal stem cells: a donor-matched comparison. *Stem Cell Res. Ther.* **2018**, *9*, 168. [CrossRef]
61. Kakkar, A.; Nandy, S.B.; Gupta, S.; Bharagava, B.; Airan, B.; Mohanty, S. Adipose tissue derived mesenchymal stem cells are better respondents to TGFβ1 for in vitro generation of cardiomyocyte-like cells. *Mol. Cell. Biochem.* **2019**, *460*, 53–66. [CrossRef] [PubMed]
62. Bajek, A.; Olkowska, J.; Walentowicz-Sadłecka, M.; Sadłecki, P.; Grabiec, M.; Porowińska, D.; Drewa, T.; Roszkowski, K. Human adipose-derived and amniotic fluid-derived stem cells: A preliminary in vitro study comparing myogenic differentiation capability. *Med. Sci. Monit.* **2018**, *24*, 1733–1741. [CrossRef] [PubMed]
63. Guo, X.; Bai, Y.; Zhang, L.; Zhang, B.; Zagidullin, N.; Carvalho, K.; Du, Z.; Cai, B. Cardiomyocyte differentiation of mesenchymal stem cells from bone marrow: New regulators and its implications. *Stem Cell Res. Ther.* **2018**, *9*, 44. [CrossRef] [PubMed]
64. Neshati, V.; Mollazadeh, S.; Fazly Bazzaz, B.S.; de Vries, A.A.F.; Mojarrad, M.; Naderi-Meshkin, H.; Neshati, Z.; Mirahmadi, M.; Kerachian, M.A. MicroRNA-499a-5p Promotes Differentiation of Human Bone Marrow-Derived Mesenchymal Stem Cells to Cardiomyocytes. *Appl. Biochem. Biotechnol.* **2018**, *186*, 245–255. [CrossRef] [PubMed]
65. Lu, M.; Xu, L.; Wang, M.; Guo, T.; Luo, F.; Su, N.; Yi, S.; Chen, T. MiR-149 promotes the myocardial differentiation of mouse bone marrow stem cells by targeting Dab2. *Mol. Med. Rep.* **2018**, *17*, 8502–8509. [CrossRef] [PubMed]
66. Fujita, J.; Tohyama, S.; Kishino, Y.; Okada, M.; Morita, Y. Concise Review: Genetic and Epigenetic Regulation of Cardiac Differentiation from Human Pluripotent Stem Cells. *Stem Cells* **2019**, *37*, 992–1002. [CrossRef] [PubMed]
67. Ieda, M.; Fu, J.-D.; Delgado-Olguin, P.; Vedantham, V.; Hayashi, Y.; Bruneau, B.G.; Srivastava, D. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **2010**, *142*, 375–386. [CrossRef]
68. Chen, J.X.; Krane, M.; Deutsch, M.A.; Wang, L.; Rav-Acha, M.; Gregoire, S.; Engels, M.C.; Rajarajan, K.; Karra, R.; Abel, E.D.; et al. Inefficient reprogramming of fibroblasts into cardiomyocytes using Gata4, Mef2c, and Tbx5. *Circ. Res.* **2012**, *111*, 50–55. [CrossRef]
69. Fu, J.D.; Stone, N.R.; Liu, L.; Spencer, C.I.; Qian, L.; Hayashi, Y.; Delgado-Olguin, P.; Ding, S.; Bruneau, B.G.; Srivastava, D. Direct reprogramming of human fibroblasts toward a cardiomyocyte-like state. *Stem Cell Reports* **2013**, *1*, 235–247. [CrossRef]
70. David, R.; Brenner, C.; Stieber, J.; Schwarz, F.; Brunner, S.; Vollmer, M.; Mentele, E.; Müller-Höcker, J.; Kitajima, S.; Lickert, H.; et al. MesP1 drives vertebrate cardiovascular differentiation through Dkk-1-mediated blockade of Wnt-signalling. *Nat. Cell Biol.* **2008**, *10*, 338–345. [CrossRef]
71. Wystrychowski, W.; Patlolla, B.; Zhuge, Y.; Neofytou, E.; Robbins, R.C.; Beygui, R.E. Multipotency and cardiomyogenic potential of human adipose-derived stem cells from epicardium, pericardium, and omentum. *Stem Cell Res. Ther.* **2016**, *7*, 84. [CrossRef] [PubMed]
72. Neshati, V.; Mollazadeh, S.; Fazly Bazzaz, B.S.; de Vries, A.A.; Mojarrad, M.; Naderi-Meshkin, H.; Neshati, Z.; Kerachian, M.A. Cardiomyogenic differentiation of human adipose-derived mesenchymal stem cells transduced with Tbx20-encoding lentiviral vectors. *J. Cell. Biochem.* **2018**, *119*, 6146–6153. [CrossRef] [PubMed]
73. Chen, Y.-H.; Coller, J. A Universal Code for mRNA Stability? *Trends Genet.* **2016**, *32*, 687–688. [CrossRef] [PubMed]
74. Warren, L.; Lin, C. mRNA-Based Genetic Reprogramming. *Mol. Ther.* **2019**, *27*, 729–734. [CrossRef] [PubMed]
75. Weidgang, C.E.; Russell, R.; Tata, P.R.; Kühl, S.J.; Illing, A.; Müller, M.; Lin, Q.; Brunner, C.; Boeckers, T.M.; Bauer, K.; et al. TBX3 directs cell-fate decision toward mesendoderm. *Stem Cell Reports* **2013**, *1*, 248–265. [CrossRef] [PubMed]
76. Chan, S.S.K.; Shi, X.; Toyama, A.; Arpke, R.W.; Dandapat, A.; Iacovino, M.; Kang, J.; Le, G.; Hagen, H.R.; Garry, D.J.; et al. Mesp1 patterns mesoderm into cardiac, hematopoietic, or skeletal myogenic progenitors in a context-dependent manner. *Cell Stem Cell* **2013**, *12*, 587–601. [CrossRef]

77. Lv, Y.; Gao, C.-W.; Liu, B.; Wang, H.-Y.; Wang, H.-P. BMP-2 combined with salvianolic acid B promotes cardiomyocyte differentiation of rat bone marrow mesenchymal stem cells. *Kaohsiung J. Med. Sci.* **2017**, *33*, 477–485. [CrossRef]
78. Bhuvanalakshmi, G.; Arfuso, F.; Kumar, A.P.; Dharmarajan, A.; Warrier, S. Epigenetic reprogramming converts human Wharton's jelly mesenchymal stem cells into functional cardiomyocytes by differential regulation of Wnt mediators. *Stem Cell Res. Ther.* **2017**, *8*, 185. [CrossRef]
79. Ibarra-Ibarra, B.R.; Franco, M.; Paez, A.; López, E.V.; Massó, F. Improved efficiency of cardiomyocyte-like cell differentiation from rat adipose tissue-derived mesenchymal stem cells with a directed differentiation protocol. *Stem Cells Int.* **2019**, *2019*, 8940365. [CrossRef]
80. Steinle, H.; Weber, M.; Behring, A.; Mau-Holzmann, U.; Schlensak, C.; Wendel, H.P.; Avci-Adali, M. Generation of iPSCs by Nonintegrative RNA-Based Reprogramming Techniques: Benefits of Self-Replicating RNA versus Synthetic mRNA. *Stem Cells Int.* **2019**, *2019*, 1–16. [CrossRef]
81. Wang, L.; Liu, Z.; Yin, C.; Asfour, H.; Chen, O.; Li, Y.; Bursac, N.; Liu, J.; Qian, L. Stoichiometry of Gata4, Mef2c, and Tbx5 influences the efficiency and quality of induced cardiac myocyte reprogramming. *Circ. Res.* **2015**, *116*, 237–244. [CrossRef] [PubMed]
82. Qayed, M.; Copland, I.; Galipeau, J. Allogeneic Versus Autologous Mesenchymal Stromal Cells and Donor-to-Donor Variability. In *Mesenchymal Stromal Cells*; Elsevier: Amsterdam, The Netherlands, 2017; pp. 97–120.
83. McLeod, C.M.; Mauck, R.L. On the origin and impact of mesenchymal stem cell heterogeneity: new insights and emerging tools for single cell analysis. *Eur. Cell. Mater.* **2017**, *34*, 217–231. [CrossRef] [PubMed]

2.2.3 RNA co-expression analysis supports findings about AMPK

Yavari, A., ..., **Wolfien, M.**, ..., Wolkenhauer, O., ..., and Ashrafian, H. (2017).
Mammalian 2 AMPK regulates intrinsic heart rate.
Nature Communications. IF: 11.880, Citations (December 14, 2020): 10

The adenosine monophosphate activated protein kinase (AMPK) is a conserved serine/threonine kinase, whose activity maintains cellular energy homeostasis. Eukaryotic AMPK exists as complexes, in which the regulatory subunit confers an energy sensor function by binding adenine nucleotides. Humans bearing activating mutations in the 2 subunit exhibit a phenotype including unexplained slowing of heart rate (bradycardia). Here, it is shown that 2 AMPK activation downregulates fundamental SA cell pacemaker mechanisms to lower heart rate, including sarcolemmal hyperpolarization activated current (I_f) and ryanodine receptor-derived diastolic local subsarcolemmal Ca^{2+} release. In contrast, loss of 2 AMPK induces a reciprocal phenotype of increased heart rate and prevents the adaptive intrinsic bradycardia of endurance training.

In this study, I applied TRAPLINE on murine SA cells and visualized the differentially expressed transcripts as a network. The outcome of this investigation was cross-validated with cardiac qPCR and a SA-node microarray. Furthermore to investigate the co-expression of 2 AMPK, I conducted a WGCN analysis, which included hierarchical clustering, construction of the topological overlap matrix via the soft-thresholding method, multi-dimension analysis, and a network-screening analysis to distinguish between signal and noise transcripts. The identified co-expression cluster, including 2 AMPK, was subsequently checked for hub genes and characterized via GO and pathway enrichment analyses.

In summary, our results reveal that in mammals, for which heart rate is a key determinant of cardiac energy demand, AMPK functions in an organ-specific manner to maintain cardiac energy homeostasis and determines cardiac physiological adaptation to exercise by modulating intrinsic SA cell behavior.

nature COMMUNICATIONS

OPEN

Correspondence and requests for materials should be addressed to A.Y. (email: arash.yavari@well.ox.ac.uk) or to H.A. (email: houman.ashrafian@cardiov.ox.ac.uk)
#A full list of authors and their affliations appears at the end of the paper

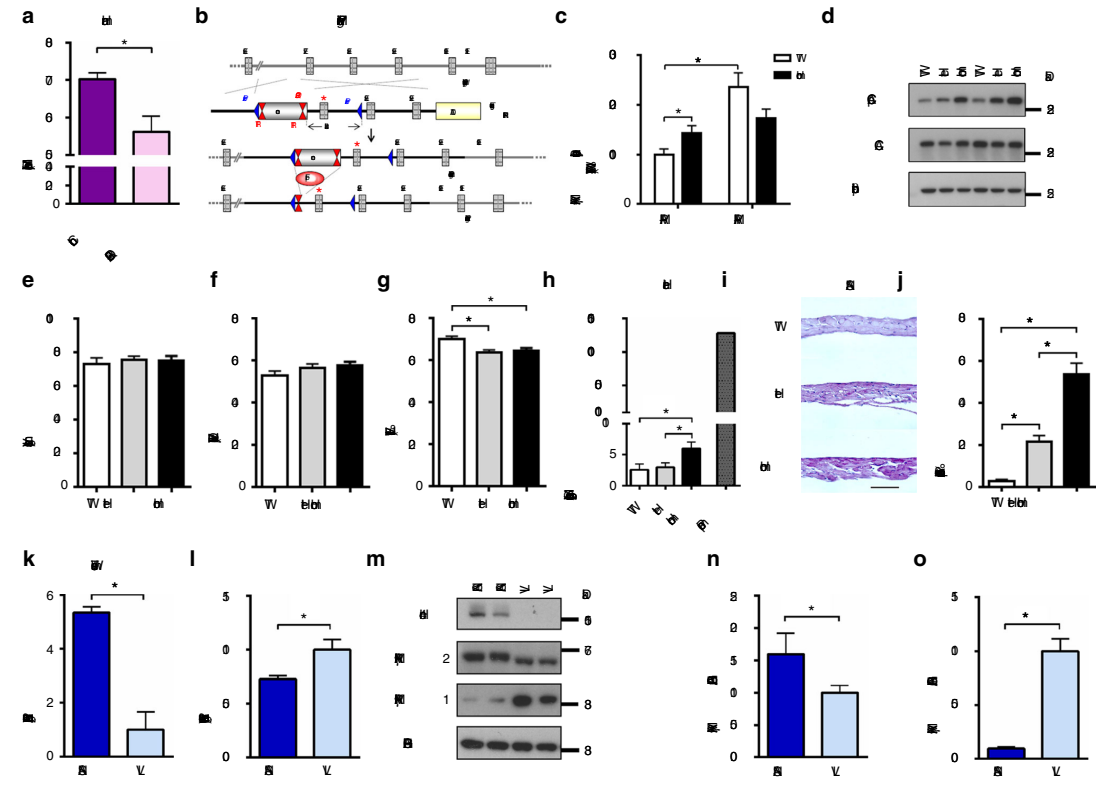
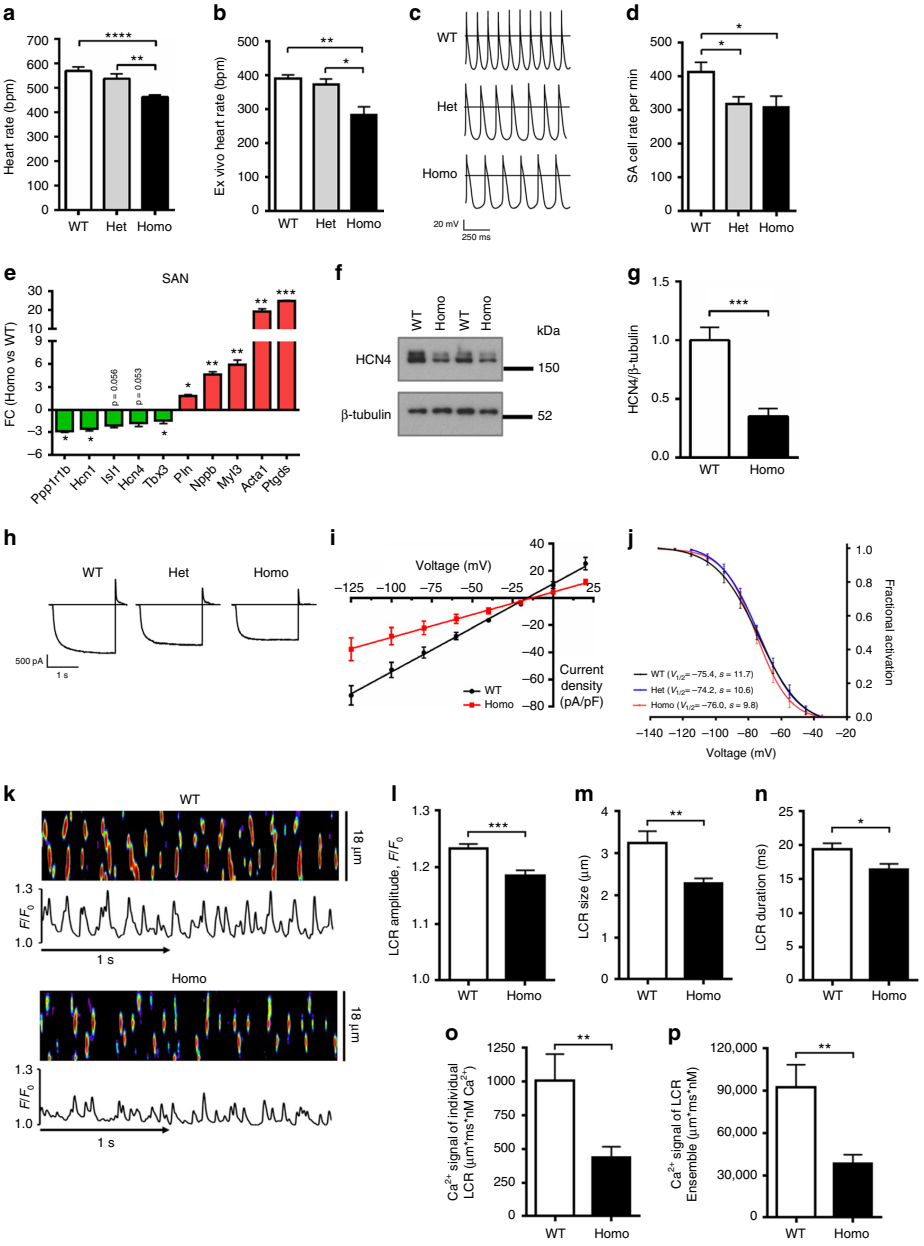


Fig. 1 Generation of the R299Q γ2 AMPK knock-in mouse and enrichment of γ2 AMPK in WT SA nodes. **a** Mean 24-h heart rate (HR) of human heterozygous R302Q γ2 mutation carriers (age 41.2 ± 2.8 years) vs genotype-negative sibling controls (age 38.9 ± 2.3 years) ($n = 10$–15). All subjects had any anti-arrhythmic drugs or β-adrenoceptor blockers discontinued for 5 days prior to ECG and none were on amiodarone. **b** Schematic of gene-targeting strategy to generate the R299Q γ2 AMPK knock-in. Neo, neomycin selection cassette; FRT, Flp recombinase recognition target; red asterisk denotes mutation in exon 7 of *Prkag2*. **c** γ2 AMPK-specific activity of freeze-clamped ex vivo perfused hearts measured by SAMS peptide phophorylation assay in the absence or presence of AMP ($n = 18$–22). **d** Representative western blot of whole heart tissue from R299Q γ2 and WT mice for phospho-(p) ACC ($n = 11$–15). **e**–**g** Cine MRI analysis of left ventricular (LV) mass (**e**), end-diastolic volume (EDV) (**f**) and ejection fraction (EF) (**g**) in R299Q γ2 and WT mice aged 2 months ($n = 8$–19). **h** Cardiac tissue glycogen content from 12 month R299Q γ2 and WT mice together with a positive control heart from a homozygous *Gaa* (encoding acid α-glucosidase) knockout mouse ($n = 12$–15). **i**, **j** Periodic acid-Schiff (PAS) staining (**i**) (scale bar, 5 µm) and quantification of glycogen content (**j**) (as %PAS-positive cells/field) of SA node (SAN) sections ($n = 12$). **k**, **l** Quantitative real-time PCR (qRT-PCR) of γ2 and γ1 AMPK isoform relative gene expression levels (normalized to β-actin) from normal murine SA node and LV (SAN, $n = 4$; LV, $n = 10$). **m**–**o** Western blot (**m**) and densitometry (**n**, **o**) of γ2 and γ1 AMPK in normal murine SA node and LV, together with SA node positive (HCN4) and loading (GAPDH) controls ($n = 6$–8). Uncropped western blots are shown in Supplementary Fig. 10. **a**, **k**, **l**, **n**, **o** Student's *t*-test was performed. **c** Kruskal–Wallis test followed by Dunn's multiple comparisons test was performed. **e**–**h**, **j** One-way analysis of variance (ANOVA) followed by Holm–Sidak's multiple comparisons test was performed. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. RE relative expression, AU arbitrary units. **a**, **c**, **e**–**h**, **j**–**l**, **n**, **o** Data are shown as means ± s.e.m.

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-017-01342-5 ARTICLE

of hypothalamic AMPK to the central effect of hormones influencing feeding behavior, such as ghrelin[6], leptin[7], and thyroid hormone[8]. In both mice and humans, activating mutations in γ2 AMPK that alter hypothalamic orexigenic neuronal excitability and drive caloric surfeit[9] suggest that AMPK activity, influenced by the nucleotide sensing γ subunit, has adapted in these highly specialized cells to function as a systemic energy sensor, defending the metabolic requirements of the entire organism. However, a role for regulation of organ-specific caloric accounting by AMPK activity in cell populations beyond such central neuronal circuits has not been reported.

The heart exhibits the highest organ-specific resting metabolic rate of any organ[10] and remarkable energetic stewardship, with the highest work increment of any organ, achieving dynamic workload of 5000–50,000 mmHg beats/min, typically at very high heart rates (HRs), without any increase in free ADP levels. HR scales linearly with myocardial oxygen consumption, with the latter further increased by the enhanced contractility



accompanying an increased HR[11]. As the background HR is set by the intrinsic automaticity of the cardiac pacemaker—a population of highly specialized sinoatrial (SA) cells—these observations suggest a direct link between SA cell firing rate and cardiac energy homeostasis.

In humans, activating mutations in the gene encoding the γ2 subunit of AMPK ( ) result in an autosomal dominant disorder whose heterogeneous phenotypic spectrum includes left ventricular hypertrophy (LVH) and prominent electrophysiological disturbances[12,13]. Cardiac-restricted transgenic mouse models overexpressing human mutations recapitulate major aspects of this spectrum, including severe LVH, ventricular pre-excitation, and propensity to sudden death[14,15]. Histological analyses of hearts from these models and human mutation carriers have identified cardiomyocyte glycogen accumulation and suggested a direct role for glycogen in the pathogenesis of pre-excitation[14]. However, the prominent sinus bradycardia, which contributes to the substantial requirement for early pacemaker implantation, remains poorly understood[13,16,17].

Here, we use a combination of genetic, electrophysiological, transcriptomic, and cellular approaches applied to genetic models of altered AMPK function to examine its role in the regulation of the mammalian cardiac pacemaker. Our results reveal an important organ-specific function for γ2 AMPK in the regulation of intrinsic SA cell firing rate in health and disease, linking this conserved cellular energy sensor to the control of mammalian SA node and thereby myocardial energy homeostasis through its influence on HR.

Results

γ2 . The majority of cardiomyopathy-causing mutations are missense substitutions of highly conserved residues within or in close proximity to the CBS motifs of the γ2 subunit of AMPK[18]. Studies of transgenic mice and acute viral transduction experiments[19] suggest that the primary effect of these γ2 mutations to be basal activation of the enzyme complex, likely due to a failure to adequately sense inhibitory ATP. Furthermore, the relative proportions of the different γ subunits appear to be important; for example, even overexpression of wild-type (WT) γ2 has been associated with a cardiac phenotype in mice, including LVH and glycogen excess[14]. This may reflect altered γ isoform stoichiometry (i.e., the γ1/γ2 ratio, with γ1 representing the physiologically predominant cardiac isoform)[20].

We observed that humans carrying the R302Q mutation in (the most frequently described) exhibit sinus bradycardia with a significantly lower resting HR compared with genotype-negative sibling controls (Fig. 1a). To gain insights into the pathogenesis of -related sinus bradycardia free of confounders inherent to overexpression transgenesis, we used gene-targeted mice with the R299Q mutation (orthologous to R302Q in humans) introduced into murine , permitting the expression and regulation of mutant protein under endogenous control mechanisms (Fig. 1b)[9]. Mice heterozygous (Het) for the R299Q γ2 mutation were interbred to generate WT and homozygous (Homo) mice. Competitive multiplex PCR confirmed expression of R299Q γ2 transcripts in mutant but not WT mice (Supplementary Fig. 1a). Western blotting confirmed comparable cardiac γ2 and γ1 expression across genotypes (Supplementary Fig. 1b–d). Cardiac γ2-specific basal AMPK activity was increased in R299Q γ2 mice compared with WT (Fig. 1c). Consistent with its proximity to the nucleotide-binding site[2] and previous reports[18,21], activation of R299Q γ2 AMPK complexes by AMP was limited compared with WT (Fig. 1c). In R299Q γ2 hearts, a corresponding increase in the phosphorylation of acetyl-CoA carboxylase at Ser79, a canonical AMPK target, was observed (Fig. 1d; Supplementary Fig. 1e), consistent with a basal gain-of-function of γ2 AMPK complexes in vivo.

Cine MRI revealed no evidence for LVH or LV dilatation, but R299Q γ2 mice exhibited a subtle reduction in contractile performance at 2 months of age (Fig. 1e–g), with no progression at 10 months (Supplementary Fig. 1f–i). Cardiac energetics, as determined by in vivo ^{31}P-MRS measurement of the phosphocreatine/γ-ATP ratio, was unaltered at 2 months (Supplementary Fig. 1j). Cardiac histology and ultrastructure of R299Q γ2 mice appeared indistinguishable from WT mice (Supplementary Fig. 2a, c). We found a small increase in biochemical cardiac glycogen content in homozygous R299Q γ2 mice at 12 months (Fig. 1h), associated with upregulation of several genes involved in glucose transport (,) and glycogen metabolism (,) (Supplementary Fig. 2b).

In contrast to findings in whole heart, detailed regional histological analysis of SA node sections revealed a striking excess of glycogen in R299Q γ2 mice (Fig. 1i, j; Supplementary Fig. 2h–j), with increased maximal SA node thickness (Supplementary Fig. 2e) but otherwise unremarkable histological appearances (Supplementary Fig. 2d, f, g), including no evidence of apoptosis on TUNEL staining, suggesting correspondingly greater AMPK activation[22] in the SA node. Accordingly, we assessed γ isoform transcript expression in normal murine SA nodes compared to left ventricles (LVs) and found , but not , to be enriched in SA nodes (Fig. 1k, l). We observed corresponding SA node enrichment of γ2 protein, but a striking paucity of γ1, suggesting that γ2 is the predominant γ isoform in this tissue (Fig. 1m–o; Supplementary Fig. 1k–m). We also observed significantly lower expression of α2 AMPK in the SA node compared to the LV (Supplementary Fig. 1n, o).

Reminiscent of the sinus bradycardia of human R302Q mutation carriers (Fig. 1a), invasive electrophysiology studies performed under isoflurane general anesthesia revealed a reduction in sinus HR of homozygous R299Q γ2 mice in vivo

Fig. 2 γ2 AMPK activation lowers intrinsic HR by downregulating SA cell I_f and Ca^{2+} clock pacemaker mechanisms. **a** HR in beats per minute (bpm) of R299Q γ2 and WT mice under anesthesia ($n = 7$–12). **b** HR during ex vivo-isolated cardiac perfusion ($n = 6$–11). **c** Representative action potentials from SA cells isolated from R299Q γ2 and WT mice. **d** Mean beating rate of SA cells from groups illustrated in **c** ($n = 17$ cells). **e** qRT-PCR validation of differentially expressed genes on SA node microarray ($n = 3$). FC fold-change. **f**, **g** Representative western blot (**f**) and analysis (**g**) of HCN4 levels in SA nodes from R299Q γ2 and WT mice. **h** Representative SA cell I_f traces during steps to 125 mV. **i** Mean fully activated *I*/*V* curves (I_f density plotted against membrane voltage) recorded in WT and R299Q γ2 SA cells. Linear data fitting yielded significant differences ($P < 0.0001$) in I_f slope conductance (648 and 333 pS/pF for WT and homozygous R299Q γ2 SA cells, respectively) ($n = 8$–10 cells/2–6 mice). **j** Mean voltage dependence of I_f activation of WT and R299Q γ2 SA cells ($n = 6$ per genotype). Half-activation voltages ($V_{1/2}$, mV) and inverse-slope factors (*s*, mV) depicted. **k** Representative confocal line-scan images and Ca^{2+} transients of isolated, single, permeabilized WT, and homozygous R299Q γ2 SA node cells bathed in 50 nmol/L free [Ca^{2+}]. **l**–**n** Mean spontaneous local Ca^{2+} release (LCR) amplitude (**l**) expressed as peak value (*F*) normalized to minimal (F_0) fluorescence, size (**m**), and duration (**n**). **o**, **p** Ca^{2+} signal of individual LCR (**o**) and LCR ensembl (**p**) ($n = 15$–17 cells/3 mice per genotype). Uncropped western blots are shown in Supplementary Fig. 10. **a**, **b**, **d** One-way ANOVA followed by Holm–Sidak's multiple comparisons test was performed. **i** Comparison of the slopes of linear regression lines was performed. **e**, **g**, **l**–**p** Student's *t*-test was performed. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$. **a**, **b**, **d**, **e**, **g**, **l**–**p** Data are shown as means ± s.e.m.

(Fig. 2a). The PR interval and anterograde atrioventricular conduction parameters were not significantly different to WT, with no evidence of ventricular pre-excitation either at baseline or with programmed atrial stimulation (Supplementary Table 1). Ambulatory HR recordings confirmed invasive findings, with marked sinus bradycardia in homozygous R299Q γ2 compared with WT mice (416 ± 13 bpm vs 549 ± 15 bpm, respectively, $P < 0.0001$, Student's *t*-test; Supplementary Fig. 3a, b).

γ2 AMPK activation reduces SA cell automaticity. Assessment of isolated perfused hearts from R299Q γ2 mice demonstrated a lower intrinsic HR (Fig. 2b). As a corollary, isolated SA cells from R299Q γ2 mice exhibited reduced basal firing rate but unaltered maximum diastolic potential (MDP) (Fig. 2c, d; Supplementary Fig. 3c). We next measured SA cell rate responses to catecholamine and muscarinic stimulation. We observed significant increases in SA cell firing rates in response to the β-adrenergic receptor agonist isoproterenol in homozygous R299Q γ2 mice, with a magnitude of change from baseline comparable to WT mice, thereby reaching a marginally lower maximal rate (Supplementary Fig. 3e). Both genotypes exhibited profound reductions in SA cell firing rates in response to the endogenous muscarinic receptor agonist acetylcholine (ACh) (Supplementary Fig. 3f). These data indicate that the activating R299Q γ2 AMPK mutation, in the context of a broadly healthy SA node with chronotropic competence, induces intrinsic sinus bradycardia in mice by reducing the basal SA cell firing rate while retaining high responsivity to catecholamine and cholinergic rate modulation.

γ2 AMPK activation reprograms the SA node transcriptome. To delineate the molecular mechanisms contributing to the intrinsic sinus bradycardia of R299Q γ2 mice, we obtained SA node gene expression profiles and identified significant differences in gene expression (Supplementary Fig. 4; Supplementary Tables 2 and 3). These included: changes suggestive of a transition of the SA node to a less nodal phenotype (upregulation in *Myl2*, *Myl3*, *Nppb*, and *Tnnt3*); downregulation of transcriptional regulators critical to SA node development and function (*Isl1* and *Tbx3*)[23,24]; alterations in constituents of the sarcolemmal membrane voltage clock (upregulation in *Kcna5* and *Kcnv2*); genes involved in Ca^{2+} homeostasis or regulation of the sarcoplasmic reticulum (SR) Ca^{2+} clock (upregulation in *Parv*, *Pln*, and downregulation in *Atp2a1*, *Calr*, *Casq1*, and *Ppp1r1b*); and genes related to AMPK's canonical function (i.e., altered expression of genes involved in glucose homeostasis and glycogen metabolism with upregulation in *Fbp2*, *Ganc*, *Pfkfb2*, *Phkb*, *Pgm3*, and downregulation in *Pygm* and *Hk1*). These genes were observed to cluster around and interact with (Supplementary Figs. 5–7) a network of key regulators of pacemaker clock function, including genes encoding cAMP-activated protein kinase (*Pka*) (Supplementary Fig. 5, network 2; Supplementary Fig. 6, network 5), short stature homeobox 2 (*Shox2*) (Supplementary Fig. 6, network 2), the cardiac ryanodine receptor (*Ryr2*), and the calcium and calmodulin-dependent protein kinase II (*CamkII*) (Supplementary Fig. 6, network 6). Quantitative real-time PCR (qRT-PCR) of SA nodes confirmed many of these changes (Fig. 2e), suggesting that γ2 AMPK activation has a transcriptional influence to remodel the coupled-clock and accounting for the observed changes in SA node function.

γ2 AMPK activation downregulates SA I_f and spontaneous LCRs. The transcription factors ISL1 and TBX3 critically promote the SA cell-specific gene program[23,24], including expression of *Hcn4*. HCN4, highly expressed in the mammalian SA node, is one of four hyperpolarization-activated cyclic nucleotide-gated channel isoforms constituting f-channels which are responsible for the sarcolemmal hyperpolarization-activated "funny" current, I_f, critically contributing to the spontaneous depolarization of pacemaker cells, and whose reduced expression is associated with sinus bradycardia[25]. We measured SA node HCN4 protein expression and found a marked reduction in homozygous R299Q γ2 mice (Fig. 2f, g; Supplementary Fig. 3d). Patch-clamping of isolated homozygous R299Q γ2 SA cells revealed a reduction in I_f density with a substantial and significant reduction in whole-cell I_f conductance compared with WT (Fig. 2h, i), but no alteration in the I_f voltage-dependence of activation (Fig. 2j), supporting the contribution of lower f-channel density to reduced SA node I_f. SA cells from both genotypes exhibited similar shifts in the I_f activation curve upon β-adrenoceptor or muscarinic stimulation using isoproterenol (Iso) or acetylcholine (ACh), respectively, suggesting unperturbed I_f modulation by these agonists (Supplementary Fig. 3g, h).

Spontaneous rhythmic SR local Ca^{2+} release (LCR) also crucially influences SA cell automaticity[26]. Given the gene expression profile findings, we undertook confocal imaging of LCRs in individual permeabilized WT and R299Q γ2 SA cells (in which the impact of HCN4 and other sarcolemmal electrogenic molecules constituting the membrane clock are uncoupled from the Ca^{2+} clock) bathed in fixed physiologic free [Ca^{2+}]. We found significantly lower mean LCR amplitude, size, and duration in R299Q γ2 vs WT SA cells (Fig. 2k–n), resulting in a >50% lower spontaneous Ca^{2+} signal of individual and ensemble LCRs (Fig. 2o, p), which activate the Na^+/Ca^{2+} exchanger current (I_{ncx}) in intact SA cells during diastolic depolarization. Consistent with the transcriptome data, immunohistochemistry of isolated SA cells from homozygous R299Q γ2 mice revealed signals for greater phospholamban (PLN) protein expression (Supplementary Fig. 3i), a key negative modulator of LCR periodicity via its inhibitory effects on the SR Ca^{2+} uptake pump, sarco(endo) plasmic reticulum Ca^{2+}-ATPase (SERCA)[26]. We verified this finding quantitatively using western blotting, confirming increased phospholamban content in homozygous R299Q γ2 SA nodes compared with WT SA nodes, both in absolute terms and when expressed relative to its cognate protein, SERCA (Supplementary Fig. 3j, l), consistent with reduced SR Ca^{2+} replenishment. However, we identified no significant effect of the R299Q γ2 mutation on levels of other key constituents of spontaneous intracellular Ca^{2+} cycling contributing to pacemaker function, including SERCA itself, calsequestrin (CASQ1), the cardiac ryanodine receptor (RYR2), the Na^+/Ca^{2+}exchanger (NCX1), or L-type Ca^{2+} channels (LTCC) (Supplementary Fig. 3k, m–p). Altogether, these data indicate that γ2 AMPK activation co-ordinately reduces two fundamental components of the SA cell intrinsic pacemaker mechanism: I_f and LCRs.

WGCN analysis links *Prkag2* to a network of pacemaker genes. To confirm AMPK's ability to acutely and reversibly alter intrinsic SA nodal automaticity free from the potential confounding of a constitutive transgenic setting, we first examined the role of AMPK and its modulation using induced murine pacemaker cell aggregates: terminally differentiated induced sinoatrial bodies (iSABs). These are highly pure, spontaneously contracting aggregates consisting substantially of physiologically functional pacemaker cells derived through forward programming with the nodal inducer TBX3 and *Myh6*-promoter-based antibiotic selection of murine pluripotent stem cells[27,28]. Sequencing (RNASeq) of iSABs' gene expression profiles, when compared to control antibiotic-selected cardiac bodies (aCaBs, a heterogeneous mixture of cardiomyocyte subtypes)[27], revealed increased expression of γ2, but not γ1 transcript, and significant

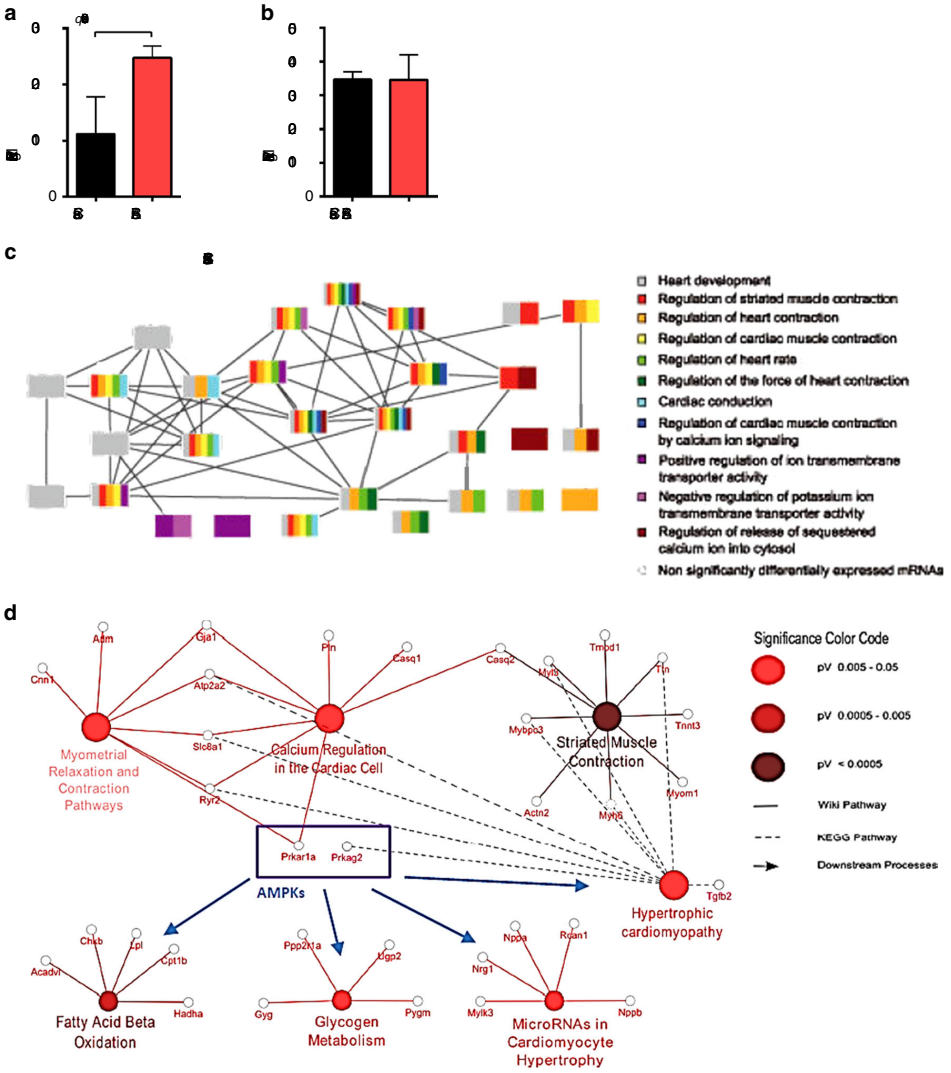
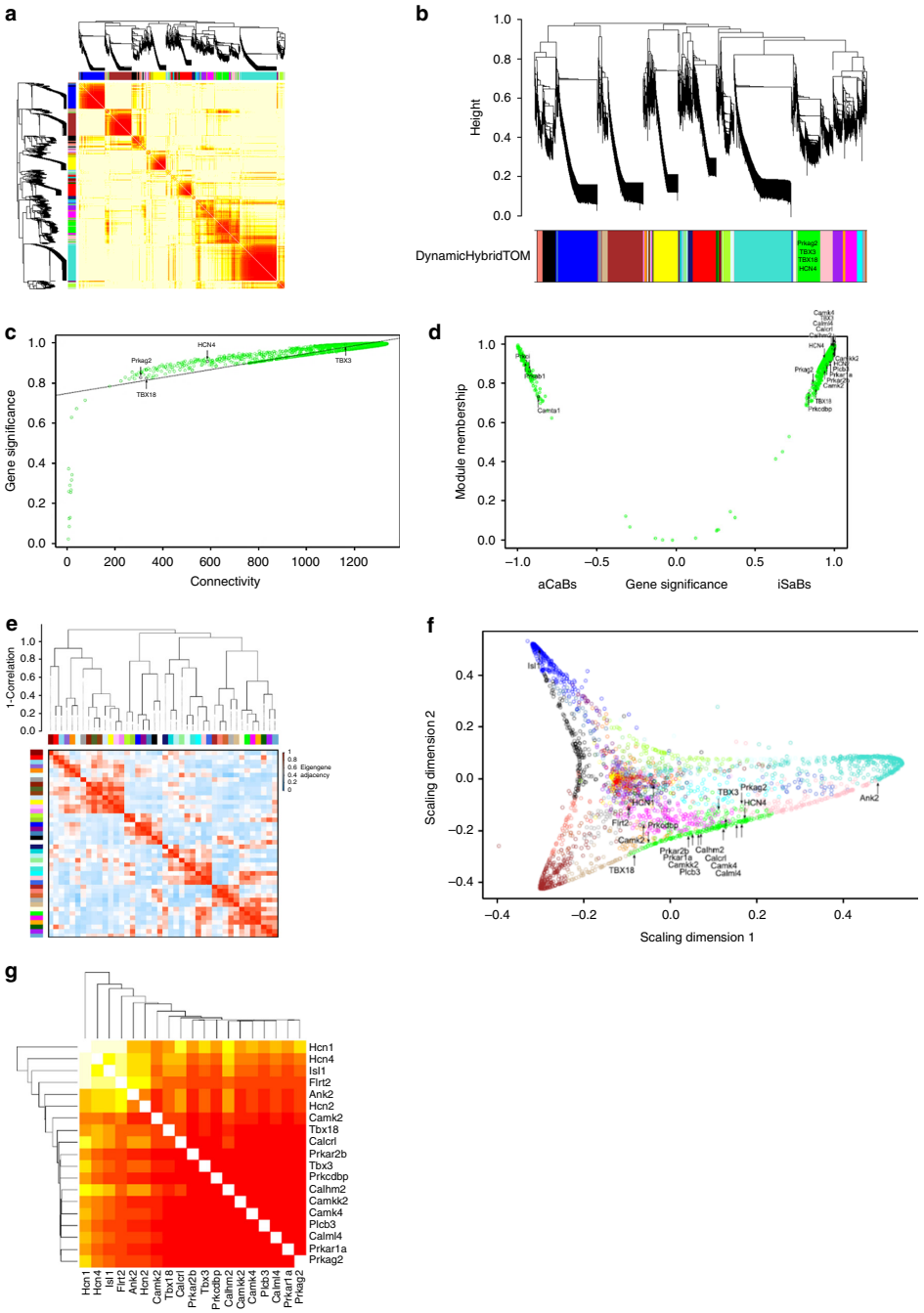


Fig. 3 RNA-Seq-derived expression levels of γ AMPK isoforms and gene ontology analysis of iSABs vs aCaBs. **a**, **b** *Prkag2* (**a**) and *Prkag1* (**b**) gene expression in iSABs (induced sinoatrial bodies) vs aCaBs (antibiotic-selected cardiac bodies—a mixture of cardiomyocyte subtypes) by RNA-Seq ($n = 3$). **c**, **d** Functional annotation with gene ontology (GO) analysis of iSAB gene expression identifying enrichment of GO terms associated with pacemaking (**c**) and significant enrichment of AMPK-dependent downstream targets and ontological processes (**d**). **a**, **b** Data are shown as means ± s.e.m.

enrichment of gene ontologies related to AMPK-mediated and Ca^{2+}-mediated signaling, the type 1A regulatory subunit of PKA (*Prkar1a*), striated muscle contraction, fatty acid β-oxidation, and glycogen metabolism (Fig. 3a–d). We then constructed a weighted gene co-expression network (WGCN) and identified *Prkag2* in a module (green) of the most highly interconnected genes (corr = 0.74), including *Tbx3*, *Tbx18*, *Hcn4*, *Prkar1a*, *Prkar2b*, *Camk2*, *Camkk2*, *Camk4*, and *Calml4* (Fig. 4a–c). Comparison of iSABs and aCaBs within the module (1,500 and 200 genes, respectively) confirmed that genes expected to be co-expressed in endogenous pacemaker cells were found only in iSABs (Fig. 4d). Hierarchical clustering and multi-dimensional scaling revealed, in commonality with the SA node transcriptome, that the *Prkag2*-containing module represents a major signaling hub with significant connectivity to genes critical to SA node pacemaker function, including *Tbx3*, *Isl1*, and *Hcn4* (Fig. 4e–g). Having identified significant co-expression and correlation of *Prkag2* with known pacemaker-relevant factors in iSABs (Supplementary Table 4), we tested whether pharmacological activation of AMPK (Fig. 5a) could lower iSAB-beating frequency. We observed a reversible, dose–response reduction in beating rate using both an AMP-mimetic agent (AICAR) and a small-molecule cyclic benzimidazole derivative (compound 991), the latter binding specifically to the β subunit of AMPK to cause direct allosteric activation[29,30] (Fig. 5b–e; Supplementary Movies 1 and 2).

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-017-01342-5 ARTICLE

Adenovirus-mediated γ2 AMPK gain-of-function reduces SA cell firing. We next tested whether acute γ2 AMPK-specific activation could alter the firing rate of individual fully differentiated mammalian WT SA cells. Given the absence of γ isoform-selective AMPK activators, we used adenoviral gene transfer to acutely overexpress R299Q γ2 (Ad-R299Q γ2), WT γ2 (Ad-WT γ2), or empty vector (Ad-mCherry) in primary SA cells isolated from WT (adult C57BL/6J) mice[31]. Whole-cell current-clamp recordings from cultured C57BL/6J SA cells infected with Ad-WT γ2 revealed no detectable effect on spontaneous SA cell beating rate of WT γ2 overexpression (300 ± 15 bpm) over that of Ad-mCherry (293 ± 16 bpm) (Fig. 6a). In contrast, C57BL/6J SA



cells infected with adenovirus carrying the activating R299Q γ2 mutation displayed a significantly slower (>30%) beating rate (192 ± 22 bpm, $P < 0.01$ compared with Ad-mCherry or Ad-WT γ2, one-way ANOVA; Fig. 6a). We then asked whether these findings were consistent in a larger mammalian species with electrophysiological properties closer to humans. Adult rabbit SA cells are recognized as excellent models for studying pacemaker mechanisms and exhibit action potentials with significantly closer fidelity to human than rodent species[32]. Congruent with our observations in WT murine SA cells, adenoviral transduction of stable cultured (72 h) adult rabbit SA cells with Ad-R299Q γ2 significantly reduced spontaneous cell beating rate to ~50% of that observed with either empty vector or Ad-WT γ2 ($P < 0.0001$, one-way ANOVA; Fig. 6b, c).

We next applied a similar gene transfer approach in an attempt to rescue the reduced beating rate phenotype of SA cells isolated from homozygous R299Q γ2 mice. Transfection with Ad-WT γ2 vector, through competing out of the mutant allele, completely restored the reduced firing rate of homozygous R299Q γ2 SA cells to that of primary C57BL/6J SA cells treated with empty vector (Fig. 6d).

Collectively, these data establish that (i) acute specific activation of γ2 AMPK is sufficient to substantively reduce the intrinsic firing rate of WT SA cells from two distinct mammalian species, and (ii) the phenotype of reduced SA cell automaticity observed in R299Q γ2 mice can be fully reversed postnatally with short-term gene transfer, arguing against abnormal developmental SA node patterning as a substantial driver of the phenotype in vivo.

γ2 AMPK has a physiological role in limiting resting HR. In view of the finding of lower sinus rate associated with the activating R299Q γ2 mutation, we investigated whether tonic γ2 AMPK activation has a physiological role in limiting HR. To address this, we developed a γ2 knockout model by crossing the R299Q γ2 line with Sox2cre transgenic mice to allow global embryonic deletion of the floxed-mutated exon 7 of *Prkag2* (Supplementary Fig. 8a). We confirmed absence of R299Q γ2 transcript, and loss of γ2 AMPK protein and activity, without significant effect on γ1 in these mice (Homo fl Cre+) (Fig. 7a; Supplementary Fig. 8b–f). We also observed no differences in gross cardiac structural or functional phenotype compared with WT Cre+ controls (Supplementary Fig. 8g–j). In contrast to R299Q γ2 mice, and supporting a physiological role for γ2 AMPK activation in regulating HR, we found that γ2 loss led to small but significant increases in HR, both in vivo under anesthesia (Fig. 7b) and during ambulatory telemetric recordings (Supplementary Fig. 9a, b), as well as ex vivo (Fig. 7c). Consistent with greater intrinsic HR, SA cells from Homo fl Cre+ mice displayed enhanced automaticity (Fig. 7d, e), but equivalent MDP and cell capacitance to WT Cre+ (Supplementary Fig. 9c, d). Homo fl Cre + mice displayed greater cardiac *Hcn1* and *Hcn4* expression than WT Cre+ (Fig. 7f, g); however, we did not identify a statistically significant increase in I_f density from Homo fl Cre+ isolated SA cells or changes in fractional activation (Fig. 7h–j). In response to isoproterenol, SA cells from Homo fl Cre+ mice reached a similar maximal rate to WT Cre+ but, having starting from a higher baseline rate, reflected a smaller proportional increase from baseline (Supplementary Fig. 9e). SA cells from both genotypes exhibited marked reductions in beating rate in response to acetylcholine, and similar mean shifts in the I_f activation curve following isoproterenol or acetylcholine stimulation (Supplementary Fig. 9f–h). Measurement of SA cell LCR revealed non-significant trends to greater spontaneous Ca^{2+} signals in Homo fl Cre+ mice compared with WT Cre+ (Supplementary Fig. 9i–m).

Loss of γ2 AMPK rescues bradycardia in FNIP1-deficient mice. Given the relatively subtle increment in sinus HR observed in the γ2 AMPK knockout, we considered whether the impact of γ2 loss may be more clearly substantiated in a model characterized by severe sinus bradycardia that was likely to be AMPK-dependent. FNIP1 (encoding folliculin-interacting protein 1) is a negative modulator of AMPK. As such it represents an alternative genetic strategy to powerfully activate AMPK. FNIP1 homozygous null mice, through activating AMPK, manifest marked sinus bradycardia[33]. We generated mice deficient in both FNIP1 and γ2 AMPK by crossing FNIP1 null mice with Sox2cre-driven γ2 knockout mice and found that loss of γ2 AMPK was sufficient to rescue FNIP1-deficient bradycardia (Supplementary Fig. 9n). This observation reinforces the conclusions that AMPK activation is sufficient to cause sinus bradycardia, and that the HR effect is directly attributable to the γ2 subunit.

γ2 AMPK is required to develop intrinsic resting bradycardia of endurance exercise. To determine whether γ2 AMPK has a broader role in physiological HR regulation, we investigated its involvement in the widely recognized phenomenon of intrinsic resting bradycardia, which follows endurance exercise training[34]. Exercise is known to activate AMPK in skeletal muscle[35,36], with AMPK in turn having a major role in this tissue's adaptive response. We examined the effect of 10 weeks of voluntary wheel running exercise (Ex), sufficient to activate cardiac AMPK in resting mice after exercise training (Fig. 8a), on intrinsic HR in comparison to sedentary controls (S). WT Cre+ and Homo fl Cre+ mice ran comparable distances (6.30±0.51 vs 6.72±0.64 km/24 h, $P = 0.58$), durations, and average speeds (Fig. 8b–d). We determined intrinsic HR using ex vivo intact SA node/atrial preparations[37] and found those of Ex WT Cre+ to display a significantly lower spontaneous beating rate than those from S WT Cre+ mice (364 vs 412 per min, $P < 0.01$, one-way ANOVA)—consistent with training-induced intrinsic resting bradycardia; however, critically, no corresponding training effect was observed on the intrinsic atrial rate of Homo fl Cre+ mice (Fig. 8e). We confirmed these findings in isolated SA cells, observing reduced automaticity of SA cells from trained WT Cre+ but not from trained Homo fl Cre+ mice (Fig. 8f, g). Consistent

Fig. 4 WGCN analysis identifies *Prkag2* in a central hub of pacemaker regulating genes. **a** Weighted gene co-expression network (WGCN) analysis-derived visualization of the iSAB–aCaB gene network by heat map plot. The heat map shows the topological overlap matrix (TOM) among all genes in the analysis. Light yellow represents low overlap and darker red represents higher overlap. Blocks along the diagonal are modules. Dendrograms and module color assignments are shown at the top and along the left side, respectively. **b** Refinement of the gene modules showing the gene dendrogram (average linkage) and module color assignment based on dynamic hybrid TOM clustering. **c** Plot of gene significance and intra-modular connectivity illustrating high correlation within the green module containing *Prkag2*. **d** Plot of co-expressed genes in the green module vs gene significance subdividing iSABs and aCaBs. **e** Further investigation of the relationship and connectivity among the investigated modules illustrated by (upper portion) a hierarchical clustering dendrogram (average linkage) and (lower portion) eigenvalue adjacency heatmap. **f** Multi-dimensional scaling plot identifying the green module as a major signaling hub connecting multiple genes critical to pacemaker functionality. **g** Heat map illustrating the TOM among genes depicted in **e**. Each column and row refers to a single gene. Light yellow represents low overlap and darker red represents higher overlap

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-017-01342-5 ARTICLE

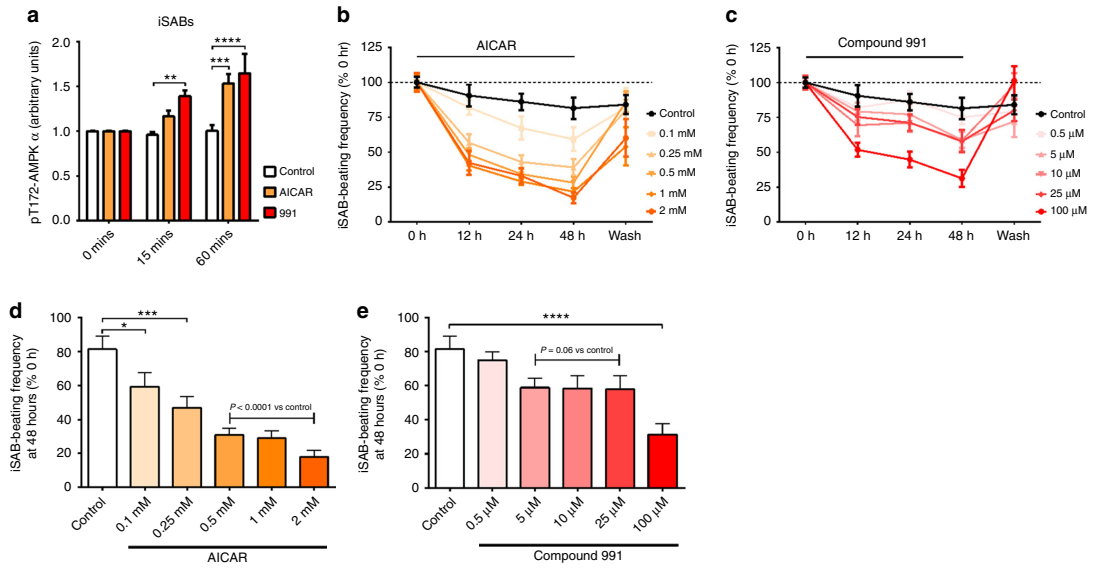


Fig. 5 Pharmacological activation of AMPK reduces the spontaneous beating rate of iSABs. **a** ELISA analysis of α AMPK Thr172 phosphorylation in iSABs treated with the AMPK activator AICAR- or the small-molecule AMPK activator compound 991 ($n = 4$). **b** Effect of incubation with variable doses of AICAR or control on iSAB-beating rate. **c** Effect of incubation with variable doses of compound 991 or control on iSAB-beating rate. **d** Bar chart representation of AICAR dose-response effect data shown in **b** specifically for the 48 h incubation time point. **e** Bar chart representation of compound 991 dose-response effect data shown in **c** specifically for the 48 h incubation time point. **a** Two-way ANOVA was performed. **d**, **e** One-way ANOVA followed by Holm–Sidak's multiple comparisons test was performed. $^{*}P < 0.05$, $^{**}P < 0.01$,$^{***}P < 0.001$, $^{****}P < 0.0001$. **a**–**e** Data are shown as means ± s.e.m.

with a previous report[38], we observed a clear reduction in SA whole-cell I_f density in trained WT Cre+ mice, but no effect of training on Homo fl Cre+ mice or fractional activation (Fig. 8h–j). Altogether, these findings indicate a critical gatekeeper function for γ2 AMPK activation to enable the intrinsic bradycardic adaptation to endurance exercise.

Discussion

By characterizing a murine model of a human AMPK-dependent monogenic bradycardic disorder, we identify a crucial function for γ2 AMPK, traditionally regarded as a minority AMPK subunit, as a major SA isoform with a role in regulating SA node automaticity, and thereby resting HR. This effect is mediated through influencing the major signaling networks of SA cell-autonomous factors regulating pacemaker functionality (e.g., TBX3 and ISL1) and core sarcolemmal (I_f) and subcellular (SR-derived LCRs)-coupled pacemaker mechanisms (Fig. 8k). We observe an opposite HR phenotype resulting from the loss of γ2 AMPK, and describe an indispensable role for this energy sensor in the genesis of intrinsic endurance bradycardia, implicating a non-redundant function for AMPK in mammalian physiological HR regulation and exercise adaptation. The relatively subtle impact of γ2 AMPK loss at baseline, but its obligatory requirement to develop intrinsic endurance bradycardia, exemplify AMPK's primary function as a sensor that is quiescent at rest but exquisitely responsive to stress.

The R299Q γ2 AMPK gene-targeted mouse model is notable for its relatively restricted cardiac phenotype—contrasting with both the uniformly malignant phenotype of transgenic mice overexpressing mutant human γ2 under a powerful cardiac-restricted promoter and the full expression of the human phenotype—specifically the lack of ventricular pre-excitation or significant LVH. Human PRKAG2 cardiomyopathy is now recognized to be highly heterogeneous, variably penetrant, and generally milder than initially reported, an observation typical of how our understanding of monogenic disorders evolves. We have previously evaluated a series of 20 patients with the R302Q *PRKAG2* mutation, orthologous to the knock-in mutation carried by R299Q γ2 knock-in mice. Although 18 of these patients (90%) had sinus bradycardia, only 2 (10%) had LVH and none had WPW syndrome. Other clinical reports document the absence of LVH with this or other *PRKAG2* mutations[39,40], suggesting that neither LVH nor pre-excitation are universal features of human PRKAG2 cardiomyopathy.

The phenotype of transgenic mice often differ from humans. In addition to recognized functional differences between mice and humans in terms of allometric scaling, prominent in the cardiovascular system and predisposing humans to a more marked bradycardia *per* se, an additional difference accounting for the subtlety of the heterozygote (and indeed homozygote) murine phenotype is likely to be the relative difference in the cardiac expression of γ2 in mouse vs human. Use of more penetrant mutations and an overexpression transgenic approach are likely to be required to consistently generate the more extreme end of the phenotypic spectrum in mice. Substantiating these gene dosage and stoichiometric considerations, other recently generated gene-targeted γ2 AMPK-mutant mice bearing mutations with severe biochemical consequences exhibit a remarkably consistent sinus bradycardia but otherwise subtle cardiac phenotype[41].

AMPK, by virtue of being at the intersection of systemic energy sensing and caloric regulation[9], appears well placed to determine the established coupling between basal metabolic rate and HR, and is likely to contribute to various allometric scaling phenomena, which seem to have empirical validity, albeit imperfect, over a diverse range of organisms[42]. Activation of

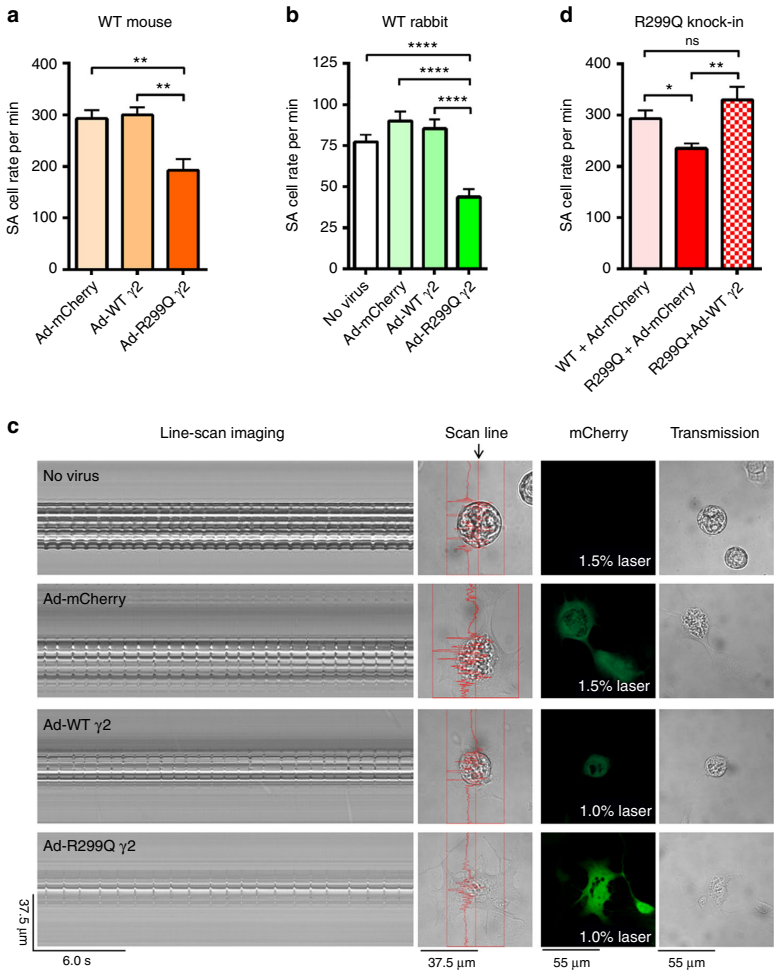


Fig. 6 Adenovirus-mediated γ2 AMPK gain-of-function reduces intrinsic firing rate of WT mammalian SA cells. **a** Mean firing rates of SA cells isolated from WT C57BL/6J mice following infection with adenoviruses carrying control (Ad-mCherry), WT γ2 (Ad-WT γ2), or R299Q γ2 (Ad-R299Q γ2) constructs ($n = 7$–9). **b** Mean spontaneous beating rate of WT rabbit SA cells following adenoviral infection ($n = 29$–36). **c** Representative line-scan images of spontaneous contraction (column 1), scan line (column 2), mCherry density (column 3), and corresponding transmission images (column 4) of WT rabbit SA cells following adenoviral infection. **d** Mean firing rates of SA cells isolated from homozygous R299Q γ2 mice following adenoviral infection with control (Ad-mCherry) or WT γ2 constructs ($n = 5$–9). Mean firing rate of WT SA cells following infection with Ad-mCherry (bar identical to that in **a**) also depicted for comparison. **a**, **b**, **d** One-way ANOVA followed by Holm–Sidak's multiple comparisons test was performed. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$; ns not significant. **a**, **b**, **d** Data are shown as means ± s.e.m

γ2 AMPK complexes remodels SA cell gene expression and electrophysiology, reducing intrinsic resting HR to diminish myocardial work. A central inference of our study is that, with its co-option at the metazoan divergence, AMPK has transitioned from being a purely cell-autonomous regulator of energy charge to co-ordinating organ-specific and systemic-caloric accounting. An example of this broader influence is its indispensable role in the regulation of exercise-related changes in HR.

Akin to the similarly highly conserved clock genes that regulate circadian biology at multiple hierarchical levels, ranging from cell-autonomous regulation, through entrainment of organ-specific time cues at different developmental stages, to the systemic control of both the persistence and periodicity of circadian rhythms[43], γ2 AMPK regulates the firing frequency of individual pacemaker cells, thereby gating basal cardiac contractile rate to ensure medium- to long-term myocardial energy homeostasis and to influence whole-mammal energy expenditure[9].

Athletes' hearts need to remain parsimonious at rest, yet retain the capacity to perform optimally during vigorous physical activity. In the context of an increased stroke volume due to cardiac chamber enlargement, a reduced intrinsic sinus rate—mediated by intermittent physiologic SA node γ2 AMPK activation—maintains basal cardiac energy expenditure, yet leaves substantial chronotropic reserve to accommodate peak activity demands; loss of γ2 AMPK specifically abrogates this adaptation.

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-017-01342-5 ARTICLE

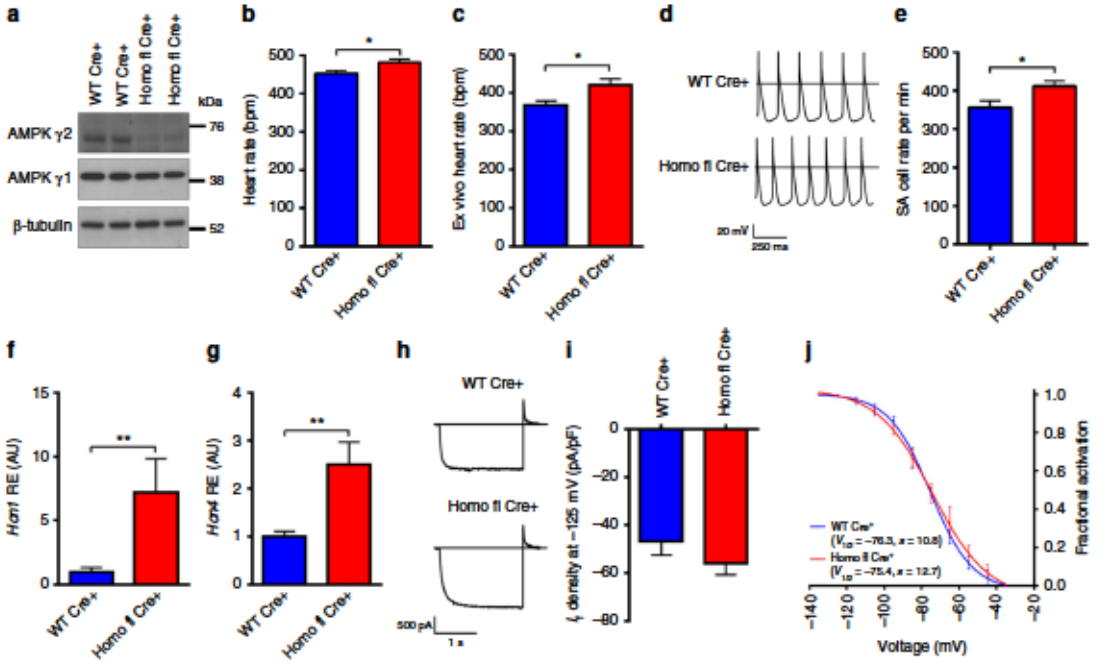


Fig. 7 Loss of γ2 AMPK increases resting heart rate and SA cell automaticity. **a** Cardiac western blot for γ AMPK isoforms in Sox2cre-driven γ2 knockout mice (Homo fl Cre+) and WT Sox2cre carrying controls (WT Cre+). **b** HR of Homo fl Cre+ and WT Cre+ mice under anesthesia ($n = 7$–12). **c** HR during ex vivo-isolated cardiac perfusion ($n = 8$–9). **d** Representative action potentials recorded from isolated SA cells. **e** Mean beating rate of SA cells from genotypes illustrated in **d** ($n = 18$–20 cells/3–5 mice). **f**, **g** Relative gene expression (by qRT-PCR) of *Hcn1* and *Hcn4* from whole hearts ($n = 5$–6). **h** Representative I_f traces during steps to −125 mV. (**i**) Mean I_f density at −125 mV ($n = 28$–31 cells/7–8 mice). **j** Mean voltage dependence of I_f activation of isolated SA cells ($n = 7$–9). Uncropped western blots are shown in Supplementary Fig. 10. **b**, **c**, **e**–**g**, **i** Student's *t*-test was performed. $^*P < 0.05$, $^{**}P < 0.01$. **b**, **c**, **e**–**g**, **i** Data are shown as means ± s.e.m.

In normal physiology, SA node γ2 AMPK activation acts as a brake to chronically increased HR, mitigating against substantial cardiac energy expenditure. Conversely, pathological *PRKAG2* mutations result in inappropriate and persistent elevation in SA node γ2 AMPK activity, effectively tonically activating the signal driving the intrinsic resting bradycardic response to endurance exercise. In a bidirectional way, therefore, our findings explain the deleterious SA node pathology observed in *PRKAG2* mutation carriers and provide a molecular substrate for increasingly recognized, albeit infrequent, potential long-term sequelae of the athletic heart, which can include increased risk of symptomatic SA node disease with need for pacemaker implantation in later life[44]. The specific ability of γ2-containing AMPK complexes to regulate intrinsic SA node firing and HR raise the possibility that its selective modulation may hold therapeutic potential in both states.

Methods

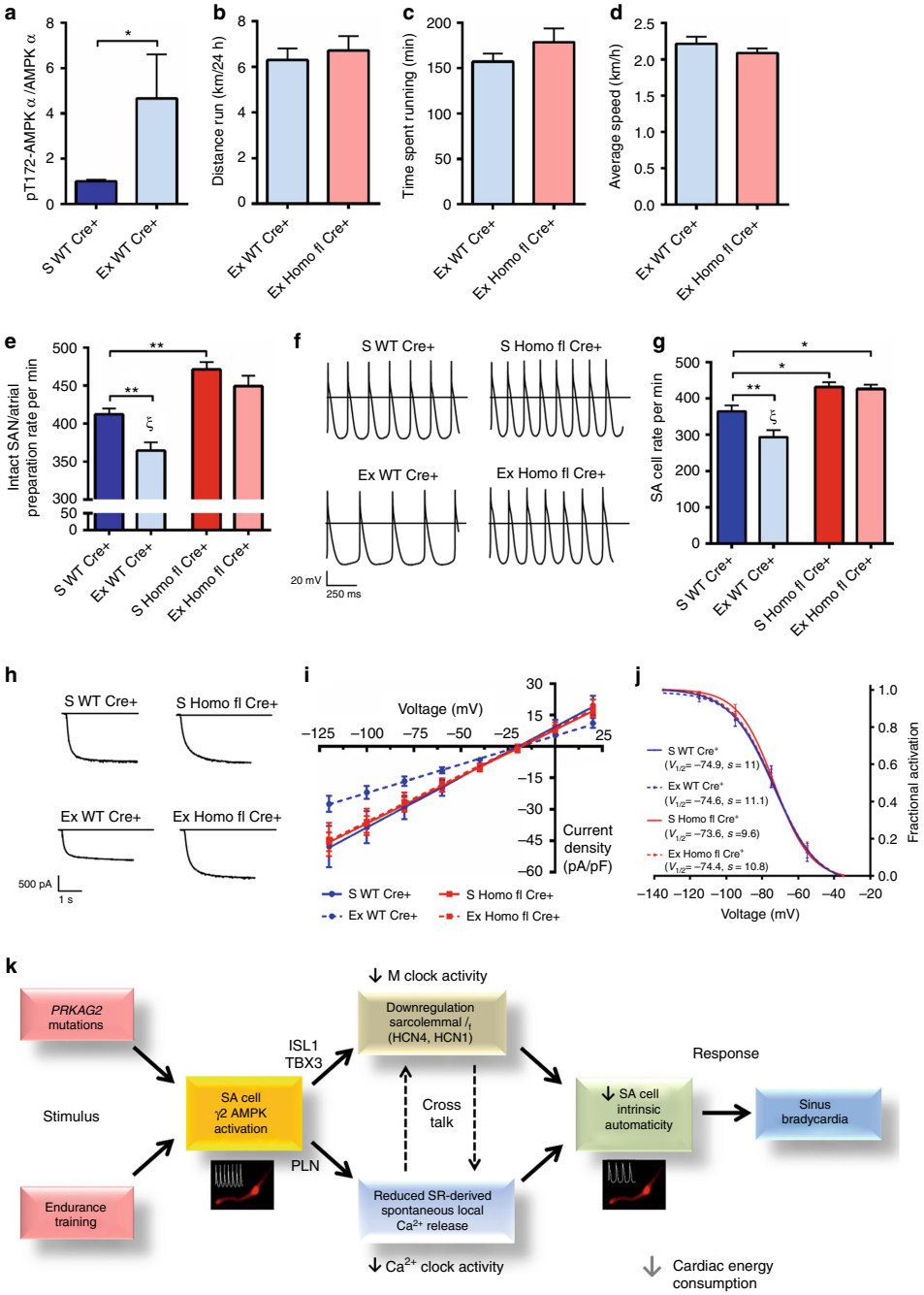
Human R302Q γ2 heterozygous carriers and HR measurement. Assessment of HR in human subjects was approved by the local Research Ethics Committee (Comitê de Ética em Pesquisa, Faculdade Ciências Médicas, Minas Gerais, Brazil). All study subjects provided written informed consent prior to participation. All subjects underwent genotyping for the R302Q γ2 mutation by PCR amplification and fluorescent dideoxy sequencing of exon 7 of *PRKAG2*. Mean HR was obtained from 15 heterozygous R302Q γ2 carriers and 10 genotype-negative sibling controls (41.2 ± 2.8 vs 38.9 ± 2.3 years, mean ± s.e.m.) using 24-h HR monitors (DMS Cardioscan Premier 11 Recorder DMS 300-8). In subjects with indwelling permanent pacemakers (6 R302Q γ2 carriers and no controls), HR was assessed by programming the generator to VVI mode and measuring the HR after a waiting period of 10 min. Antiarrhythmic drugs (including β-blockers) were discontinued for at least 5 days before HR assessment. No subject had atrial fibrillation or was on amiodarone.

Animals. Animal studies were performed in accordance with the 1986 British Home Office Animals Scientific Procedures Act (UK) incorporating European Directive 2010/63/EU, the European Directive (86/609/CEE) on the care and use of laboratory animals, and the Guide for the Care and Use of Laboratory Animals published by the National Institutes of Health (NIH Publication no. 85–23, revised 1996). All experimental protocols involving animals were assessed and approved by the local ethical review committee: University of Oxford Animal Welfare and Ethical Review Body; Animal Welfare committee of the University of Milan and the Italian Minister of Health (Italian D.lgs 116/92 and D. Lgs no. 2014/26); the NIH Institutional Animal Care and Use Committee; and the University of Colorado Denver—Anschutz Medical Campus Institutional Animal Care and Use Committee (protocol number 84814(06)1E). Experimental animal work was undertaken blind to genotype.

R299Q γ2 mice have been previously described[9]. Gene targeting was used to introduce the R299Q point mutation (equivalent to the human R302Q mutation) into exon 7 of murine *Prkag2*. The gene-targeting vector contained a shorter 5′ homology arm (in intron 6 and exon 7) amplified by PCR from genomic DNA isolated from 129Sv embryonic stem (ES) cells. The point mutation was introduced by PCR and the positive selection cassette (*neo*), flanked by FRT (Flp recombinase recognition target) sites to enable excision of the neo cassette, was inserted in front of exon 7. The targeting strategy included insertion of *loxP* sites within intron 6 upstream and within intron 7 downstream of the mutation, respectively. The longer 3′ homology arm was obtained from a BAC clone (Sanger Institute, Cambridge, UK), with a Diphtheria Toxin A (DTA) cassette attached to the 3′ homology arm for negative selection. The homology arms with the mutation were cloned into a suitable targeting vector in our laboratory; selection cassettes and genomic DNA were provided by GenOway (Lyon, France). Transfection of 129Sv embryonic stem (ES) cells, selection, isolation and confirmation of positive clones by Southern blot analysis, injection of positive clones into C57BL/6 embryos and in vivo excision of the positive selection cassette were performed by GenOway. Correct homologous recombination in the positive ES clones was confirmed by

Southern blot analysis. The presence of the point mutation and the distal *loxP* site were validated by sequencing. Positive clones were injected into C57BL/6 embryos. Highly chimeric males were then bred with Flp-expressing mice and the *FRT*-flanked neo cassette deleted, resulting in a floxed knockin *Prkag2* allele. Mice heterozygous for the R299Q γ2 knockin mutation were initially on a mixed C57BL/6/129/Ola genetic background and subsequently backcrossed to C57BL/6 for at least seven generations. R299Q γ2 mice were genotyped by PCR from earnotch tissue-derived gDNA using primers (Supplementary Table 5) hybridizing either side of the *loxP-FRT* sequence in intron 6 to distinguish the WT *Prkag2* locus from the recombined, Flp-excised *Prkag2* allele.

Sox2cre transgenic mice[45] that had been backcrossed for six generations onto a C57BL/6 genetic background were crossed with R299Q γ2 mice to achieve global γ2 deletion via the conditional deletion of the floxed exon 7 of *Prkag2*. Mice deficient in both alleles, representing a global knockout of γ2 (termed Homo fl Cre+), were compared with control mice hemizygous for the Sox2cre transgene but wild-type for *Prkag2* (termed WT Cre+). Sox2cre γ2 knockout mice were



NATURE COMMUNICATIONS | DOI: 10.1038/s41467-017-01342-5 ARTICLE

genotyped from gDNA using separate PCRs to assess for the R299Q γ2 mutation (as above), the Sox2cre transgene, and detection of intact and excised exon 7 from *Prkag2* (primer sequences detailed in Supplementary Table 5).

The generation and phenotype of *Pnip1* null mutants is as previously described[33] (MGI ID:5806459).

Allelic discrimination. Competitive multiplex PCR of cardiac cDNA for specific detection of R299Q γ2 transcript was undertaken using TaqMan MGB fluorogenic probes specific for the wild-type (5′-FAM-AGTCCGTGCAGCGC-MGB-3′) or mutant (5′-VIC-AGTCCAAGCAGCGC-MGB-3′) *Prkag2* sequence and common exon-spanning primers (Supplementary Table 5) flanking the site of the mutation on exon 7. Primers were designed (Primer Express 3.0) and reactions were undertaken in accordance with published guidance[46] on a StepOne Real-Time PCR system (Applied Biosystems). Data analysis and visualization were with StepOne software (v2, Applied Biosystems).

Western blotting. Protein extraction and western blotting were undertaken largely as previously described[47]. Briefly, snap-frozen cardiac tissue aliquots were ground in liquid nitrogen and homogenized in ice-cold buffer comprising 50 mM Tris base, 250 mM sucrose, 1 mM EDTA, protease and phosphatase inhibitor cocktail tablets (Roche, West Sussex, UK), 50 mM NaF, 5 mM sodium pyrophosphate, 1 mM dithiothreitol (DTT), 1 mM benzamidine, 0.1 mM phenylmethylsulphonyl fluoride (PMSF), and 1 mM sodium orthovanadate. Extracts were then centrifuged at 13,000×*g* for 15 min at 4 °C. Protein concentration was determined from diluted aliquots of the soluble fraction by BCA protein assay (Thermo Fisher Scientific) with samples then diluted in fresh lysis buffer to yield equivalent final protein concentrations. Lysates were mixed with lithium dodecyl sulfate sample buffer with DTT (50 mM) (nuPAGE, Invitrogen) and boiled at 95 °C for 5 min. For western blotting of SA nodes, each sample was pooled from three individual SA nodes and homogenized and lysed in RIPA buffer (Thermo Fisher Scientific: 25 mM Tris-HCl (pH 7.6), 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS) supplemented with Halt protease inhibitor cocktail (Thermo Fisher Scientific), Halt phosphatase inhibitor cocktail (Thermo Fisher Scientific) and 1 mM phenylmethyl sulfonyl fluoride, using a Precellys24 homogenizer (Bertin Instruments) with tissue homogenization kit at 4 °C. Loading controls were run on the same blot.

SDS-PAGE was undertaken on pre-cast polyacrylamide gels (Nupage 4–12% Bis Tris gel, Novex, Invitrogen) and transfered onto polyvinylidene difluoride membranes (Immun-Blot, Bio-Rad) using an electrophoretic transfer cell (Mini Trans-Blot, Bio-Rad). Blocked membranes (5% milk/tris-buffered saline with Tween-20, TBST) were incubated with primary antibody, followed by TBST washes and secondary horseradish peroxidase (HRP)-conjugated antibody detection. Bands were visualized using ECL reagents (GE Healthcare, Buckinghamshire, UK), and films scanned with subsequent analysis of digital images using ImageJ (NIH). Uncropped western blots accompanied by the location of molecular weight markers are shown in Supplementary Fig. 10.

The following antibodies were used: anti-phospho-ACC (#3661) at 1:1,000 working concentration, anti-ACC (#3676) at 1:1,000, anti-phospho-AMPK Thr172 (#2535) at 1:1,000 from Cell Signaling (New England BioLabs, Hertfordshire, UK); anti-γ1 AMPK (ab32508) at 1:1,000 and anti-β-tubulin (ab6046) at 1:4,000 from Abcam (Cambridge, UK); anti-γ2 AMPK (sc-19141) at 1:500 and anti-α2 AMPK (sc-19131) from Santa Cruz Biotechnology (TX, USA); anti-HCN4 (APC-052) at 1:200 from Alomone Labs (Israel); anti-GAPDH (MAB374) at 1:4,000 from Merck Millipore (Hertfordshire, UK); anti-PLN (ab86930) at 1:2,000, anti-SERCA2 ATPase (ab91032) at 1:2,000, anti-NCX1 (ab177952) at 1:2,000 from Abcam (Cambridge, MA, USA); anti-CASQ at 1:2,000 (PA1-913), anti-RYR2 at 1:1,000 (MA3-916) and anti-DHPR1 alpha (for LTCC, PAS-23010) at 1:1,000 from Thermo Fisher Scientific (Waltham, MA, USA). HRP-conjugated secondary antibodies used were anti-rabbit IgG (NA934) from GE Healthcare and anti-goat IgG (ab6741) from Abcam.

AMPK activity assay. Cardiac AMPK activity was measured from immunoprecipitated AMPK complexes by SAMS peptide phosphorylation assay essentially as previously described[15]. In brief, protein extracts were prepared as per samples for western blotting, including addition of phosphatase inhibitors. AMPK γ subunit isoform-specific antibody was pre-bound to a 50% protein G-sepharose bead slurry on an orbital shaker (IKA Vibrax VXR) at 4 °C for 2 h. These were then gently centrifuged and washed in ice-cold PBS/1% triton, followed by a further ice-cold PBS wash. Tissue lysate was added to pre-bound protein G-antibody mix in ice-cold 1% triton/HBA buffer (50 mM HEPES, 50 mM sodium fluoride, 5 mM sodium pyrophosphate, 1 mM EDTA, 10% glycerol [v/v], 1 mM DTT, 1 mM benzamidine, 0.1 mM PMSF, supplemented with a protease inhibitor cocktail tablet (Roche), pH to 7.4 at room temperature). Immunoprecipitation (IP) was performed on an orbital shaker at 4 °C for 2–3 h (typically IP 30 µL protein G-Sepharose/antibody slurry and 250 µg of sample lysate, made up to 500 µL in HBA/1% triton with fresh protease inhibitors).

In-house antibodies were used for immunoprecipitation of γ2 (rabbit polyclonal, C-terminus directed) and γ1 (rabbit polyclonal) AMPK. AMPK activity from immune complexes was determined by measuring the incorporation of [γ-^{32}P]-ATP into the SAMS synthetic peptide substrate, with/without 0.2 mM AMP, by scintillation counting (Tri-Carb 2800TR, PerkinElmer, UK).

In vivo cine magnetic resonance imaging. High-resolution in vivo cine MRI was performed on a cohort of R299Q γ2 mice and WT littermate controls at 2 and 10 months of age, to accurately assess left ventricular volumes, function, and mass with high spatial resolution. Anesthesia was induced in an anesthetic chamber using 4% isoflurane in 100% oxygen. Electrodes were positioned subcutaneously, and mice were positioned prone on a dedicated mouse cradle and maintained at 1.5–2% isoflurane at 2 L/min oxygen flow. Temperature was maintained at ~37 °C using a homeostatically controlled warm air blanket. Cardiac and respiratory signals were continuously monitored and used for combined ECG gating and respiratory gating. Eyes were protected with a petroleum-based ophthalmic ointment. Cine MRI experiments were carried out using a horizontal 210 mm bore 9.4 T magnet with VNMRS DirectDrive console and 60 mm i.d. 1,000 mT/m actively shielded gradient system (Agilent Technologies, USA). A 33 mm internal diameter, quadrature-driven birdcage resonator (Rapid Biomedical, Germany) was used for signal transmission/reception. Cine imaging was carried out as described previously[48]. Multi-frame left ventricular-short axis slices were acquired (7–10 contiguous slices, 1 mm thickness, 18–32 frames per cardiac cycle) covering apex to base. Images were reconstructed off-line as TIFF files using custom-written software. End-diastolic and end-systolic frames were selected for each slice according to maximal and minimal ventricular cavity size and semi-automated image segmentation performed by a single operator using AMIRA software (Visage Imaging) blind to mouse ID/genotype.

In vivo cardiac ^{31}P magnetic resonance spectroscopy. MR spectroscopy experiments were carried out using a 9.4 T magnet as above with 120 mm i.d. 600 mT/m actively shielded gradient system (Agilent Technologies, USA). An actively decoupled variable tune/match 14 mm diameter ^{31}P surface coil was purpose built in-house and used in conjunction with a double tuned ^{1}H/^{31}P volume resonator (Rapid Biomedical, Germany) for acquisition. Animals were prepared as described above. Shimming and scouting were carried out using the ^{1}H channel of the volume coil. A removable 4 mm point sphere filled with 15 M H_3PO_4 was placed outside the animal cradle to allow for accurate and rapid pulse calibration using an unlocalized pulse-acquire experiment. 2D acquisition weighted (Hanning) CSI (chemical shift imaging) data were acquired for WT, heterozygote R299Q γ2, and homozygote R299Q γ2 male mice aged ~10 weeks (n = 7, 18, 9, respectively) from a 5 mm thick mid-ventricular short axis slice (in-plane voxel size of 2.31 × 2.31 mm, 13 × 13 PE steps, 30 × 30 mm FOV, 5 mm slice thickness, 8191 scans, TR = 0.87 ms). Acquisitions were cardiac gated and a TR of ~250 ms (two cardiac cycles) was used with a 30° flip angle. Total scan time for the experiment was ~35 min. Multi-slice ^{1}H anatomical images covering the field of view of the CSI experiment were acquired to confirm the position and tissue content of the CSI voxels. Fully sampled data were zerofilled to 64 × 64 PE steps, and reconstructed as described previously[49].

For each mouse, a 3 × 3 grid of spectra from voxels located at the septum of the heart was fitted using a Voigt lineshape (in-house software), and the PCr, γ-ATP, and 2,3-diphosphoglycerate (DPG) signal amplitudes estimated. The spectrum

Fig. 8 γ2 AMPK is critically required for the intrinsic bradycardic adaptation to endurance exercise. **a** Results of western blot analysis of α AMPK Thr172 phosphorylation in whole heart tissue from sedentary (S) and exercised (Ex, 10 weeks of voluntary wheel running) WT Cre+ mice (n = 8–10). **b–d** Average daily distance (**b**), time (**c**), and speed (**d**) of voluntary wheel running during a 10-week training period of WT Cre+ and Homo fl Cre+ mice (n = 17–26). **e** Spontaneous beating rate of isolated intact SA node/atrial preparations from S and Ex WT Cre+ and Homo fl Cre+ mice (n = 10–22). **f** Representative action potentials recorded from isolated SA cells. **g** Mean beating rate of isolated SA cells from S and Ex groups (n = 12–27 cells). **h** Representative SA cell I_f traces during steps to −125 mV. **i** Mean fully activated *I/V* curves recorded in SA cells. Linear data fitting yielded statistically significant differences ($P < 0.0001$) in I_f slope conductance of SA cells from exercised WT Cre+ mice only, with conductance values of 481 (S WT Cre+), 447 (S Homo fl Cre+), 272 (Ex WT Cre+) and 447 pS/pF (Ex Homo fl Cre+) (n = 6–14 cells/4–8 mice per group). **j** Mean voltage-dependence of I_f activation of SA cells from both S and Ex groups (n = 6–15). **k** Schematic depicting the central function of SA cell γ2 AMPK in overall cardiac energy accounting. **a–d** Student's *t*-test was performed; **e**, **g**, one-way ANOVA followed by (**e**) Holm–Sidak's multiple comparisons test or (**g**) Fisher's least significant difference test was performed. *$P < 0.05$, **$P < 0.01$, $^{\S}P < 0.0001$ for both Ex WT Cre+ vs S Homo fl Cre+ and Ex WT Cre+ vs Ex Homo fl Cre+ comparisons. **a–e**, **g** Data are shown as means ± s.e.m.

with the lowest γ-ATP/DPG ratio was selected as the nominal blood spectrum, and its PCr and γ-ATP amplitudes, normalized to DPG amplitude, were subtracted from those of the remaining eight myocardial spectra. Finally, T1 saturation correction for residual PCr and γ-ATP amplitudes was carried out using the mean TR for the CSI experiment, and published T1 values[50]. The mean PCr/γ-ATP ratio was then calculated for the myocardium.

Cardiac histology. Hearts were rinsed in ice-cold PBS and immerse-fixed in 10% neutral buffered formalin (VWR, Leicestershire, UK) for 24 h, then processed in an automated tissue processor (Bavimed Histomaster, Germany) overnight. Cross-sections (5–7 µm) were obtained using a microtome (Leica RM 2155), spread onto polysine-coated glass slides (VWR), and allowed to dry. Sections were stained with hematoxylin and eosin (Sigma-Aldrich, Dorset, UK), periodic acid-Schiff (Sigma-Aldrich) or Picrosirius red (Polysciences, Germany). Images were acquired with a Nikon light microscope (Nikon Eclipse TE2000U) coupled to a digital camera (Nikon Digital Sight DS-5M).

Transmission electron microscopy. Hearts were extracted, cut finely into small (~1 mm^3) blocks and fixed overnight in 4% glutaraldehyde in 100 mM phosphate buffer, followed by post-fixation in 2% osmium tetroxide in 100 mM phosphate buffer. Specimens then underwent en bloc treatment with uranyl acetate, dehydration in ethanol, and transferred to propylene oxide, prior to embedding. Ultra-thin sections (50–70 nm) were cut and stained with uranyl acetate and lead citrate, and examined in a JEOL 1200EX electron microscope.

Biochemical glycogen content. Cardiac glycogen was quantified in R299Q γ2 and WT mice aged 12 months essentially as described[15]. In brief, snap-frozen tissue samples ground under dry ice were homogenized in lysis buffer (50 mM Tris, 0.25 M mannitol, 1 mM DTT, 0.1 mM PMSF adjusted to pH 7.4 and supplemented with protease inhibitor tablets [Roche]) and sonicated prior to syringe-and-needle homogenization. A total of 2 M KOH at 70 °C was used to solubilize glycogen followed by amyloglucosidase digestion overnight to release glucose. Glucose content was then determined using an enzyme coupled assay (Roche) to generate glucose-6-phosphate (G6P) by hexokinase, followed by oxidation of G6P by G6P dehydrogenase in the presence of $NADP^+$, with spectrophotometric measurement of NADPH after 30 min.

Cardiac qPCR. RNA extraction, cDNA synthesis, and RT-PCR were performed as previously described[47]. qRT-PCR using inventoried TaqMan gene expression assays was used to determine relative gene expression of the following: γ2 and γ1 AMPK in wild-type C57BL/6 murine SA nodes and LVs (*Prkag2*, Mm00513977_m1; *Prkag1*, Mm00450298_g1); and cardiac expression of *Slc2a1* (Mm00441480_m1), *Slc2a4* (Mm00436615_m1), *Gyg* (Mm00516516_m1), and *Pygm* (Mm00478582_m1) in R299Q γ2 and WT mice. Relative expression of the target gene was quantified using the $2^{-\Delta\Delta Ct}$ method using *Actb* (Mm00607939_s1) as the endogenous control (where ΔCt is the difference in cycle threshold value between the target transcript and the endogenous control). Samples were analyzed in at least duplicate, and samples minus reverse transcriptase enzyme and samples minus cDNA template were used as negative controls (to exclude gDNA contamination or reagent cross-contamination, respectively) and checked to ensure they failed to reach threshold by 40 cycles.

SA node histology. SA nodes were harvested using an operating microscope and 12 consecutive 5 µm paraffin-embedded sections obtained. SA node sections were stained with the following: Masson's Trichrome (Masson's Trichrome stain kit, American MasterTech, Lodi, CA); reticulum plus picric red stains to determine collagen and outline atrial myocytes (Chandler's precision reticulum stain kit, American MasterTech); and Periodic acid-Schiff (PAS kit, American MasterTech). High-resolution digital images of cross-sections were acquired using a Leica microscope. Morphometric evaluation was performed using a computerized imaging analysis system (Metamorph, University Imaging) blind to genotype according to a modified prior method[51]. Quantification of PAS-stained areas and cells (clusters of small, round, homogeneous PAS-positive granular structures) was taken from images at ×400 magnification.

Invasive electrophysiology. Conduction parameters were assessed in 3-month-old R299Q γ2 and WT littermate controls based on modifications of a published EPS protocol for mice[52]. Under 1.5% isoflurane anesthesia, a 1.1F electrophysiology catheter (EPR-801, Millar Instruments) was introduced into the right atrium and ventricle via the right internal jugular vein. High-fidelity intracardiac electrograms were filtered at 0.5–500 Hz and continuously recorded using PowerLab and LabChart software (ADInstruments, Oxford). After obtaining a baseline surface ECG, programmed stimulation protocols were performed as described[53] using a pacing stimulator with output set at twice diastolic threshold (s88 Grass stimulator, Grass technologies, USA). Sinus cycle length (SCL) and sinus node recovery time (SNRT) were determined by delivering an atrial pacing train at a cycle length of 100 ms for 15 s. SNRT was calculated as the maximum interval from the last-paced atrial complex to the first spontaneous complex after cessation of pacing. All measurements were repeated in triplicate. Atrioventricular (AV) Wenckebach cycle length (AVW) and AV 2:1 cycle length were determined using burst atrial pacing by progressively reducing pacing cycle length from baseline in decrements of 10 ms (incremental atrial pacing) until 1:1 AV conduction was reliably lost (Wenckebach cycle length), or resulted in failure of conduction of every other pacing stimulus to the ventricle (2:1 cycle length). AV nodal effective refractory period (AVNERP) was determined using programmed atrial extrastimulus pacing with an eight stimuli atrial drive train (S1) delivered at 100 ms, followed by a single premature stimulus (S2) progressively decremented until ventricular conduction was lost. AVNERP was defined as the longest S1–S2 coupling interval where S2 failed to capture the ventricle. Right atrial programmed electrical stimulation was used in an attempt to induce an atrioventricular re-entrant tachycardia (AVRT) as evidence for the existence of an accessory pathway. An atrial double extrastimulus technique was used with an eight stimuli primary drive train (S1) delivered at a fixed cycle length of 100 ms, followed by premature stimuli (S2 and S3) coupled at 80 ms and decremented to refractoriness. Further provocation for AVRT included atrial-burst pacing at fixed cycle length, including following β-adrenoceptor stimulation by isoproterenol administration (1 ng/g IV). All recordings were analyzed off-line, blind to genotype, using LabChart software (ADInstruments).

ECG biotelemetry. Data were recorded using biotelemetry sensors (HDX-11 or ETA-F20 from Data Sciences International, St. Paul, MN) recorded at 1000 Hz. All recordings were taken during the light cycle with mice held in a standard 12 h light–dark cycle with constant temperature and humidity.

Ex vivo cardiac perfusion and HR. Ex vivo cardiac perfusion was performed as previously described[47] with minor modifications. Briefly, following pentobarbitone anesthesia (140 mg/kg IP) and systemic heparinization (150 IU), hearts were removed, cannulated, and retrogradely perfused at 37 °C in Langendorff mode at a constant perfusion pressure of 80 mmHg with a modified Krebs–Henseleit buffer that contained the following (in mmol/L): NaCl, 118.5; $NaHCO_3$, 25; KCl, 4.75; KH_2PO_4, 1.18; $MgSO_4$, 1.19; $CaCl_2$, 1.41; D-glucose, 11; pyruvate, 5; pH 7.4; saturated with 95% O_2 and 5% CO_2. Inclusion criteria used were: time interval to aortic cannulation of under 3 min and absence of persistent arrhythmias during stabilization. For measurement of AMPK activity in the context of preserved cellular energetic status, hearts were freeze-clamped after a 25 min stabilization period of perfusion as previously described[21]. For ex vivo HR, we used Krebs–Henseleit buffer without supplemental pyruvate. A fluid-filled polyvinylchloride film balloon was introduced into the LV and connected to a pressure transducer (ADInstruments). The reading was digitally processed to provide a ventricular pressure recording from which HR was subsequently derived blinded to genotype. Intrinsic HR was recorded as the average measured rate over 30 s, 5 min after cannulation.

SA node isolation and RNA extraction. The heart with lungs was quickly removed and immersed at 4 °C to wash out the blood in an external solution that contained the following (in mmol/L): NaCl, 137; KCl, 4.9; NaH_2PO_4, 1.2; glucose, 15; HEPES, 20; $MgCl_2$, 1; pH 7.4. The heart–lung block was pinned to the tissue bath to excise the right atrium (RA) and SA node under a stereomicroscope. The tissue bath was perfused with the external solution at a rate of 10 mL/min. After removal of both ventricles and the left atrium, the RA was opened to expose the crista terminalis, the intercaval area and the interatrial septum. This preparation was pinned by small stainless steel pins to the chamber with the endocardial side exposed up and trimmed carefully to extract the SA node region correctly. The SA node region was delimited by the borders of the crista terminalis, the interatrial septum, superior and inferior vena cava. All tissues were snap-frozen in liquid nitrogen and stored at −80 °C for subsequent RNA extraction. Four mouse SA node tissues were pooled and processed together for the extraction of one sample of total RNA. RNA was extracted with an RNeasy Mini Kit (Qiagen) using DNase on-column digestion according to the manufacturer's protocol.

SA node gene expression profiling and microarray analysis. R299Q γ2 and WT littermate controls (four homozygote R299Q γ2, four heterozygote R299Q γ2, and three WT) were hybridized to Illumina MouseRef-8 v2 bead chips. BeadArrays were scanned by the Illumina BeadStation 500X. All data were log_2-transformed and normalized by Robust Spline Normalization using the lumi software package in Bioconductor[54]. Overall, 18,185 nucleotide probes were filtered from 25,697 total using an Illumina detection P value of $\alpha = 0.05$. Significant genes were selected by a one-way ANOVA or t-test (FDR = 0.05) for further analysis. GSEA software[55] was used to select differentially expressed (DE) genes by calculating out a "score" using the Kolmogorov–Smirnov (KS) test. The score indicates how a gene relates with genotype, where positive values relate to upregulation and negative values relate to downregulation. Only transcripts with a KS score > |0.15| were considered DE. Transcripts demonstrating greater than 1.2-fold change in expression were processed using Ingenuity Pathway Analysis software (Ingenuity Systems Inc., CA, USA) to identify networks and canonical pathways overrepresented in enriched genes.

SA node microarray validation. Total RNA (850 ng) were used for cDNA preparation in a 40 µL reaction volume with MMLV reverse transcriptase (RT, Life

Technologies, CA, USA) with random hexamers. All cDNA synthesis reactions were accompanied by no template controls for the detection of possible contamination and no RT controls to detect potential genomic DNA. qRT-PCR was performed using an ABI Prism 7900HT Sequence Detection System (Applied Biosystems) with a 384-well platform. Reactions were performed with FastStart Universal SYBR Green Master mix with ROX (Roche) using manufacturer recommended conditions. Forward and reverse primers used are detailed in Supplementary Table 5.

Preliminary reactions were run to determine amplification efficiency. The size of the amplicon and its specificity were assessed by agarose gel analysis and a post-amplification dissociation curve, respectively. Each well contained 0.5 µL of cDNA solution and 10 µL of reaction mixture. Samples were assessed in quadruplicate and relative expression calculated using the $2^{-\Delta\Delta Ct}$ method using *Hprt* as endogenous control (where ΔCt is the difference in cycle threshold value between the target transcript and the endogenous control). Statistical analysis was undertaken using a one-tailed Student's *t*-test.

Murine SA cell electrophysiology and I_f measurement. Hearts were extracted, and the SA node region dissected and used for isolation of SA cells by an enzymatic and mechanical dissociation procedure as previously described[56]. Briefly, hearts were rapidly removed and placed in a pre-warmed (37 °C) Tyrode solution (in mmol/L: NaCl, 140; KCl, 5.4; CaCl2, 1.8; MgCl2, 1; D-glucose, 5.5; HEPES-NaOH, 5; pH 7.4) containing 1,000 U heparin. The SA node region was surgically exposed, isolated, and cut into small strips, which were placed in an enzymatic solution containing the following: collagenase (224 U/ml, Worthington), elastase (1.42 U/mL, Sigma-Aldrich), and protease (0.45 U/mL, Sigma-Aldrich) to loosen intercellular connections. The enzymatic solution was then removed and SA strips placed in a high-K^+, low-Na^+, Ca^{2+}, and Mg^{2+}-free solution. Cells were then fully dispersed and separated by manual agitation of SA strips using a glass pipette with a small tip (~2–3 mm diameter). Cells were finally reacclimatized to physiological concentrations of NaCl, KCl, $MgCl_2$, and $CaCl_2$ by adding the necessary amounts of a solution containing 1.8 mM $CaCl_2$ and 10 mM NaCl, and normal Tyrode with BSA 1 mg/mL. Isolated single cells were kept at 4 °C for the day of the experiment and patch-clamp experiments were performed in the whole-cell configuration at 35 ± 0.5 °C. The pipette solution contained (in mmol/L): K-aspartate, 130; NaCl, 10; EGTA-KOH, 5; $CaCl_2$, 2; $MgCl_2$, 2; ATP (Na-salt), 2; creatine phosphate, 5; GTP (Na-salt), 0.1; pH 7.2. Action potentials were recorded from spontaneously beating SA node myocytes or small uniformly beating aggregates of pacing cells (2–5 cells) superfused with normal Tyrode solution and the rate and MDP values measured with customized software. I_f was recorded from single cells superfused with Tyrode solution containing $BaCl_2$ (1 mmol/L) and $MnCl_2$ (2 mmol/L). Steady-state current amplitudes were calculated at the end of a 3 s step to −125 mV (holding potential, −35 mV). A two-step protocol was used to assess the voltage dependence of the current, with a first step to a test voltage in the range −35/−135 mV and a second step to −120 mV. Test step durations varied from 10 s at −35 mV to 5/7.5 s at −135 mV to allow full current activation at each voltage. The Boltzmann distribution ($y = 1/(1+\exp((V-V_{1/2})/s)$: V, voltage; y, fractional activation; $V_{1/2}$, half-activation voltage; s, inverse-slope factor) was used to fit experimental data-points. Fully activated I_f current–voltage (IV) relationships were obtained according to a previously published protocol[57]. Shifts of the I_f activation curve induced by isoproterenol (30 nM) or acetylcholine (30 nM) were measured near the midpoint of the activation curve (−75 mV) as previously reported[58].

Local Ca^{2+} releases in permeabilized SA node cells. SA node pacemaker cells were isolated from SA node tissue of 3-month-old mice and permeabilized using 0.01% saponin. After saponin washout, the solution was changed to a recording solution that contained the following (in mmol/L): fluo-4 pentapotassium salt, 0.03; $CaCl_2$, 0.099 (free $[Ca^{2+}] \sim 50$ nM); $C_4H_6NO_4K$ (DL-aspartic acid potassium salt), 100; KCl, 25; NaCl, 10; MgATP, 3; $MgCl_2$, 0.81 (~1 mM free Mg^{2+}); Hepes, 20; EGTA, 0.5; phosphocreatine, 10; creatine phosphokinase (5 U/mL); pH 7.2. The cytosolic free Ca^{2+} at given total Ca^{2+}, Mg^{2+}, ATP, and EGTA concentrations was calculated using a computer program (WinMAXC 2.50, Stanford University). Spontaneous characteristics of Local Ca^{2+} Releases (LCR) were measured by confocal microscopy in fixed free $[Ca^{2+}]$ as previously described[59]. The amplitude of individual LCRs was expressed as peak value (F) normalized to minimal fluorescence (F_0). LCR spatial size (FWHM) was indexed as the full width at half-maximum amplitude. LCR duration (FDHM) was characterized as the full duration at half-maximum amplitude. The number of LCRs was normalized per 100 µm of the line-scan image and during a 1 s time interval. The Ca^{2+} signal of an individual LCR was estimated as previously described[59]. $M = FWHM \times FDHM \times (\Delta Ca^{2+}$, nmol/L)/2. The Ca^{2+} signal of the LCR ensemble was estimated by integrating the Ca^{2+} signal produced by each LCR and normalized per 100 µm of the line-scan image and during a 1 s time interval[59].

SA cell immunohistochemistry. To localize the expression of HCN4 in situ, 5 µm paraffin-embedded SA node sections were stained with a monoclonal anti-HCN4 antibody (1:800, ab85023, Abcam) employing the Dako EnVision+ System-HRP (DAB) kit (#K4006, CA, USA). The 3,3′-diaminobenzidine (DAB) substrate-chromogen reaction (brown color) was visualized using a Leica microscope. High-resolution/high-magnification (×400) digital images of representative cross-sections of SA node bodies were acquired using a computerized imaging analysis system (Metamorph, University Imaging).

To visualize the expression of phospholamban (PLN), freshly isolated SA cells were fixed, blocked, and permeabilized using the Image-iT Fixation/Permeabilisation kit (Life Technologies). Immunohistochemistry was undertaken using 1:200 primary mouse monoclonal anti-PLN (A010-14, Badrilla, Leeds, UK) and 1:1000 secondary Alexa Fluor 488 donkey anti-mouse (A212202, Life Technologies) antibodies. Nuclear counterstaining was with 4′,6-diamidino-2-phenylindole (DAPI). Fluorescence images were visualized using a Zeiss LSM710 confocal scanning microscope with a 63 × 1.4 NA immersion oil objective and images recorded with ZEN 2 acquisition software (Zeiss, Germany).

iSAB and aCaB generation. Generation of iSABs (induced sinoatrial bodies) and aCaBs (antibiotic-selected cardiac bodies) was performed as previously described[27,28]. In brief, murine cell lines[27] were used to generate iSABs and aCaBs and grown in high-glucose DMEM with stable glutamine (GIBCO) containing the following: 10% FBS Superior (Biochrom), 100 µM non-essential amino acids (GIBCO), 1% penicillin/streptomycin (GIBCO), and 100 µM β-mercaptoethanol (Sigma) in the presence of 1,000 U/mL of leukemia inhibitory factor (LIF, Millipore). Differentiation was performed by hanging drop culture for 2 days using 1,000 cells as starting material for one EB (embryoid body) in Iscove's basal medium (Biochrom) containing the following: 10% FBS (Biochrom), 100 µM non-essential amino acids (GIBCO), 1% penicillin/streptomycin (GIBCO), and 450 µM 1-thioglycerol. Cells were differentiated for an additional 4 days in suspension culture, and at day 6 of differentiation 15 EBs were seeded onto one well of a 24-well-plate. At day 8 post-seeding, antibiotic selection was initiated using 400 µg/mL G418 (Biochrom). Four days thereafter, aCaBs and iSABs were isolated by treatment with 6,000 U/mL Collagenase IV (GIBCO) for 30 min. Single cells were obtained by further dissociation of the bodies using 100% Accutase (Affymetrix) for 15 min. Potential mycoplasma contamination was routinely controlled for twice a week using the PCR based MycoSPY kit system (Biontex). The iSABs were generated according to Rimmbach et al.[28]

ELISA assessment of α AMPK Thr172 phosphorylation was performed using the PathScan Phospho-AMPKα (Thr172) Sandwich ELISA Kit (Cell Signaling Technology Inc., USA) according to the manufacturer's protocol. For this, 60 µg/mL protein was isolated from iSABs at 0, 15 and 60 min following treatment with 100 µM Compound 991, 0.5 mM AICAR or control and subjected to ELISA. Experiments were performed with four biological replicates, each of which was analyzed using two technical replicates.

iSAB RNA sequencing and data analysis. For library generation and sequencing, cultured adherent cells were drained from the culture medium, washed, and directly lysed by addition of lysis buffer[60]. A total of 1 µL of this lysate was used for cDNA synthesis and amplification with the SMARTer kit (Clontech, Mountain View CA, USA) according to the manufacturer's instructions. In brief, cDNA synthesis was initiated by annealing a polyA-specific primer and adding a reverse transcriptase with terminal transferase activity. The newly synthesized first strand cDNA was then tailed first with a homopolymer stretch by terminal transferase and then with a specific amplification tag by template switching. The resulting double-tagged cDNA was amplified by PCR, fragmented by sonication (Bioruptor, Diagenode, Liege, Belgium; 25 cycles 30 s on/30 s off) and converted to barcoded Illumina sequencing libraries using the NEBnext Ultra DNA library preparation kit (New England BioLabs, MA, USA). After PCR enrichment the libraries were purified with AmpureXP magnetic beads (Beckman-Coulter, CA, USA) and quantified on a Bioanalyzer 2100 (Agilent, CA, USA). Libraries were pooled at equimolar amounts and sequenced on an Illumina GenomeAnalyzer IIx in single-read mode with a read-length of 78 nucleotides and a depth of 21–32 million raw reads per replicate.

We performed adapter clipping and quality trimming procedures for data pre-processing[61]. We aligned the reads to the mm9 genome with the aid of TopHat2[62]. Differential expression analysis was performed using Cufflinks2/Cuffdiff2[63,64]. We considered genes with >2-fold change and a q value < 0.05 as significantly differentially expressed. The gene annotation, including functional annotation clustering and functional classification, was performed with DAVID and based on gene ontology terms[65]. We used our openly available RNA sequencing pipeline (TRAPLINE) for data analysis[66].

WGCN analysis. Weighted gene coexpression network analysis was performed by applying the R package "WGCNA" to RNASeq data[67]. We first constructed the topological overlap matrix (TOM) of all investigated transcripts (~30,000) using the soft thresholding method. We calculated the eigenvalues of the transcripts and evaluated adjacency based on distance. We subjected transcripts to hierarchical clustering (average linkage) and assigned transcripts with the dynamic hybrid method into groups. We computed connectivity based on the interaction partners (k) and evaluated gene significance, representing module membership. Finally, we computed a network screening analysis using the WGCNA package to distinguish between true positive results and noise.

Gene ontology analysis. Networks were built using several applications in Cytoscape[68]. ClueGo was used to visualize and cluster the gene annotation terms into groups[69]. The KEGG and Panther pathway databases were used to obtain specific gene annotations[70–72]. The network interaction graph was built with the aid of enhanced Graphics and integrates fold change values (http://apps.cytoscape.org/apps/enhancedGraphics). Interactions between mRNAs were identified with "Agilent literature search" and are based on validated publications (http://apps.cytoscape.org/apps/agilentliteraturesearch).

Isolation and culture of primary murine SA cells. SA cells were isolated as previously described[31] from 2–4-month-old male C57BL/6J (Jackson Laboratories) or homozygous R299Q γ2 mice. Mice were anesthetized by inhalation of isoflurane prior to killing. Hearts were rapidly excised, the ventricles and left atria removed, and the SA node dissected from the remaining right atrial tissue at 35 °C in a heparinized Tyrode's solution, which consisted of the following (in mM): 140 KCl, 5.4 KCl, 1.2 KH_2PO_4, 5 HEPES, 5.55 D-Glucose, 1 $MgCl_2$, 1.8 $CaCl_2$, with a pH adjusted to 7.4 with NaOH. SA node tissue was enzymatically digested by 4.75 U elastase (Worthington Biochemical), and 3.75 mg Liberase TM (Roche) for 15 min at 35 °C in a modified Tyrode's solution, containing (in mM) 140 NaCl, 5.4 KCl, 1.2 KH_2PO_4, 5 HEPES, 18.5 D-glucose, 0.066 $CaCl_2$, 50 taurine, 1 mg/mL bovine serum albumin (BSA), with pH adjusted to 6.9 with NaOH. Following digestion, tissue was transferred to a modified KB solution (in mM: 100 potassium glutamate, 10 potassium aspartate, 25 KCl, 10 KH_2PO_4, 2 $MgSO_4$, 20 taurine, 5 creatine, 0.5 EGTA, 20 glucose, 5 HEPES, and 0.1% BSA; pH adjusted to 7.2 with KOH) at 35 °C and SA cells dissociated by mechanical trituration with a fire-polished glass pipette. The calcium concentration of the cell suspension was gradually increased to 1.8 mM.

Following calcium re-adaptation, SA cells were pelleted (at ~3000 RPM) and the supernatant carefully aspirated. For plating, SA cells were suspended in plating media, which contained Media199 (#M4530, Sigma) supplemented with 10 mM 2,3-butanedione monoxime (BDM), 10,000 U penicillin/10 mg streptomycin, and 5% (v/v) FBS. Glass coverslips were prepared by coating 12-mm diameter coverslips for ~1 h at 37 °C with 100 ng/mL mouse laminin (BD Biosciences, San Jose, CA, USA) diluted in phosphate-buffered saline (PBS). Excess laminin/Tyrode's was removed immediately before cell plating. SA cells were plated such that the cells from one mouse SA node were seeded onto one 12-mm round coverslip. SA cells were allowed to settle and adhere to the coverslip for 4–6 h at 37 °C in an atmosphere of 95% air/5% CO_2 before changing the media to culture media, which consisted of Media199 supplemented with 0.1 mg/mL bovine serum albumin (BSA; Sigma), 10 mM BDM, 10 μg/mL insulin, 5.5 μg/mL transferrin, 5 ng/mL selenium (ITS; Sigma), and 10,000 U penicillin/10 mg streptomycin. Culture media was exchanged every 24 h.

Adenoviral transduction of primary murine SA cell cultures. SA cells on each coverslip were counted immediately prior to viral delivery. Adenoviral infections were performed on the same day as isolation in serum-free culture media at a multiplicity of infection (MOI) of 100 (100 infectious agents per target cell) as described[31]. Adenoviruses Ad-mCherry, Ad-mCherry-mPrkag2 (WT)-FLAG (i.e. Ad-WT γ2) and Ad-mCherry-mPrkag2(R299Q)-FLAG (i.e. Ad-R299Q γ2) were constructed, amplified and purified by Vector BioLabs (Philadelphia, PA, USA). SA cells were incubated with virus-containing media overnight (~12–14 h) and replaced with fresh culture media the following morning.

A fragment of a coverslip bearing SA cells was transferred to the recording chamber of an inverted microscope. During all experiments, cells were constantly perfused (1–2 mL/min) with extracellular solution at 35 ± 1 °C. SA cells were identified by their characteristic morphology, small size, and generation of spontaneous action potentials. Patch clamp recordings used borosilicate glass pipettes with resistances of 1.5–3 MΩ. Data were acquired at 5–20 kHz and low-pass filtered at 1 kHz using an Axopatch 1D or 200B amplifier, Digidata 1322a or 1440a A/D converter and ClampEx software (Molecular Devices). The fast component of pipette capacitance was minimized in all recordings using the patch-clamp amplifier. Membrane capacitance was estimated from responses to 10 mV test pulses using the membrane test function in ClampEx. Spontaneous beating rates were recorded from SA cells in the whole-cell configuration in current-clamp mode without injected current. Cells were constantly perfused with normal Tyrode's solution (in mM: 140 NaCl, 5.4 KCl, 1.2 KH_2PO_4, 5 HEPES, 5.55 glucose, 1 $MgCl_2$, 1.8 $CaCl_2$; pH adjusted to 7.4 with NaOH). The intracellular (pipette) solution was composed of the following (in mM): 140 K-Aspartate, 10 HEPES, 1.8 $MgCl_2$, 10 NaCl, 0.1 EGTA, 0.02 $CaCl_2$; pH adjusted to 7.2 with KOH. Recordings were only made from infected cells, identified by mCherry fluorescence, in each culture. Viral transduction efficiency was essentially 100%, with no difference in efficiency between the different constructs and with each cell in the dataset infected. Beating rates were determined from averages of the instantaneous frequency during 15–30 s recording windows in the presence of 1 nM Isoproterenol (ISO; Calbiochem) in the bath.

Primary rabbit SA cell culture and adenoviral transduction. Adult rabbits were treated in accordance with the NIH Guide for the Care and Use of Laboratory Animals (animal protocol number: 034LCS2016). Single, spindle-shaped, spontaneously beating SA cells were isolated from the hearts of New Zealand rabbits (Charles River Laboratories, Wilmington, MA, USA) as described previously[73]. To generate cultured SA cells (c-SANC), freshly isolated SA cells were diluted 20 times with serum-containing medium, and centrifuged for 10 min at 500 rpm. Following aspiration of the supernatant, cells were plated at a density of 0.5×10^4 per cm^2 on laminin pre-coated (20 μg/mL, Sigma-Aldrich) glass-bottom dishes for culture. The serum-containing medium[74] contained a 73% salt solution (in mmol/L: NaCl 116, KCl 5.4, $MgCl_2$ 0.8, NaH_2PO_4 0.9, D-Glucose 5.6, Hepes 20, $CaCl_2$ 1.8, $NaHCO_3$ 26), 20% M199 (Sigma-Aldrich), in the presence of (in mmol/L) creatine 5, taurine 5, insulin–transferrin–selenium-X 0.1%, 4% fetal bovine serum, 2% horse serum, and 1% penicillin and streptomycin (pH = 7.4 at 37 °C).

Cells were incubated in serum-containing medium for the first 24 h, and then cultured in serum-free medium for adenoviral infection. Adenoviruses Ad-mCherry-mPrkag2 (WT)-FLAG and Ad-mCherry-mPrkag2(R299Q)-FLAG were introduced into c-SANC by an acute adenoviral gene-transfer technique using a MOI of 1,000 for 48 h. In addition to no adenoviral culture control, Ad-mCherry was employed as an adenoviral vector control. All functional and immune-labeling experiments were performed with cells cultured for 72 h.

Measurement of spontaneous beating rate and immuno-labeling of cultured rabbit SA cells post-adenoviral transduction. In the three adenoviral-treated groups, the density of mCherry was employed as a guide to successful infection visualized via laser 543 nm (1 or 1.5% laser power). The bath superfusion solution contained (in mmol/L): NaCl 140, KCl 5.4, $MgCl_2$ 1, HEPES 5, $CaCl_2$ 1.8, and Glucose 5.5 (pH = 7.4). All functional measurements were performed at 34 ± 0.5 °C. Spontaneous contraction was recorded via the line-scan of transmission image using a confocal microscope (Zeiss LSM510, Carl Zeiss Inc., Germany) and the spontaneous beating rate calculated from the average duration between peak onsets[74].

After each functional measurement, c-SANC were fixed with 4% Paraformaldehyde (10 min) for immuno-labeling. After permeabilization (1% Triton X-100 in PBS, 15 min) and blocking (PBS with 2% IgG-free BSA, 5% donkey serum, 0.02% NaN_3, and 0.1% Triton X-100, 4 h), c-SANC were incubated with primary anti-Flag M2 (1:100, Sigma) at 4 °C overnight and then stained with Cy5-conjugated donkey anti-mouse secondary antibody (1:1000, Jackson ImmunoResearch laboratories, USA) for 1 h[74]. In the negative control group, only secondary antibody was applied. A 633 nm or 543 nm (at 10% power) laser was employed to excite the fluorophore Cy5 or mCherry via a confocal microscope (Zeiss LSM510, Carl Zeiss Inc., Germany). For semi-quantification, the average density of Flag and mCherry were measured using ImageJ (1.48v, National Institute of Health, USA), with the nuclear area excluded.

Echocardiography. Transthoracic echocardiography was undertaken as previously described[47]. In brief, Sox2cre γ2 AMPK knockout and WT Cre+ mice aged ~2 months underwent general anesthesia with isoflurane (3–4% induction and 1–1.5% maintenance) in oxygen administered via nosecone. Images were acquired on a heated platform using a Vevo 2100 Imaging System (VisualSonics, Toronto) and analyzed off-line blind to genotype using Vevo software.

Surface ECG recordings. Anesthetized ECG recordings were obtained under light isoflurane (1.25%) anesthesia. Four fine needle electrodes were positioned subcutaneously and a 5 min period of equilibration undertaken, prior to a 10 min ECG recording with filtering and amplification of signals by Bio Amplifier (ADInstruments, Oxford, UK). ECG waveforms were then analyzed off-line using LabChart software (ADInstruments) blind to genotype.

Voluntary wheel running and intact SA node/atrial preparation rate measurement. Sox2cre γ2 AMPK knockout (Homo fl Cre+) and WT Cre+ mice aged 3 months were singly housed in cages containing a freely rotating, angled running track (Lillico, UK), with wheel rotations monitored by use of a reed switch connected to a computerized exercise monitoring system (Micro 1401, CED, Cambridge) as described[75]. Mice were allowed to acclimatize to single housing for several days and then data was continuously recorded for 10 weeks. During this period, cage disturbance was kept to a minimum, with ad libitum access to the running wheel. Data acquisition and analysis blind to genotype were carried out using Spike2 software (CED). For determination of intrinsic rate as part of these experiments, atria were rapidly dissected and transferred to a 2 mL organ bath containing Tyrode solution at 37 ± 0.5 °C, where they were allowed to equilibrate as previously described[37].

Statistical analysis. Data are presented as means ± s.e.m. Numbers of mice were determined by power calculations using in-house and available published data. Unless otherwise stated, data were analyzed with Student's *t*-test or one-way analysis of variance (ANOVA) followed by Holm–Sidak's multiple comparisons test. Non-parametric data were analyzed by Kruskal–Wallis test followed by Dunn's multiple comparisons test. Statistical analysis was performed with GraphPad Prism Software (v6.0, CA, USA) with $P < 0.05$ considered statistically significant.

Data availability. Microarray and RNASeq datasets generated for this study have been deposited in the Gene Expression Omnibus (Accession Number GSE73047) and the Sequence Read Archive (Accession Number SRS1064711), respectively.

Received: 12 June 2017 Accepted: 8 September 2017
Published online: 02 November 2017

References

1. Oakhill, J. S. et al. AMPK is a direct adenylate charge-regulated protein kinase. *Science* **332**, 1433–1435 (2011).
2. Xiao, B. et al. Structural basis for AMP binding to mammalian AMP-activated protein kinase. *Nature* **449**, 496–500 (2007).
3. Bungard, D. et al. Signaling kinase AMPK activates stress-promoted transcription via histone H2B phosphorylation. *Science* **329**, 1201–1205 (2010).
4. Chantranupong, L., Wolfson, R. L. & Sabatini, D. M. Nutrient-sensing mechanisms across evolution. *Cell* **161**, 67–83 (2015).
5. Kahn, B. B., Alquier, T., Carling, D. & Hardie, D. G. AMP-activated protein kinase: ancient energy gauge provides clues to modern understanding of metabolism. *Cell Metab.* **1**, 15–25 (2005).
6. Andrews, Z. B. et al. UCP2 mediates ghrelin's action on NPY/AgRP neurons by lowering free radicals. *Nature* **454**, 846–851 (2008).
7. Minokoshi, Y. et al. AMP-kinase regulates food intake by responding to hormonal and nutrient signals in the hypothalamus. *Nature* **428**, 569–574 (2004).
8. Lopez, M. et al. Hypothalamic AMPK and fatty acid metabolism mediate thyroid regulation of energy balance. *Nat. Med.* **16**, 1001–1008 (2010).
9. Yavari, A. et al. Chronic activation of gamma2 AMPK induces obesity and reduces beta cell function. *Cell Metab.* **23**, 821–836 (2016).
10. Wang, Z., O'Connor, T. P., Heshka, S. & Heymsfield, S. B. The reconstruction of Kleiber's law at the organ-tissue level. *J. Nutr.* **131**, 2967–2970 (2001).
11. Boerth, R. C., Covell, J. W., Pool, P. E. & Ross, J. Jr. Increased myocardial oxygen consumption and contractile state associated with increased heart rate in dogs. *Circ. Res.* **24**, 725–734 (1969).
12. Blair, E. et al. Mutations in the gamma(2) subunit of AMP-activated protein kinase cause familial hypertrophic cardiomyopathy: evidence for the central role of energy compromise in disease pathogenesis. *Hum. Mol. Genet.* **10**, 1215–1220 (2001).
13. Gollob, M. H. et al. Identification of a gene responsible for familial Wolff-Parkinson-White syndrome. *N. Engl. J. Med.* **344**, 1823–1831 (2001).
14. Arad, M. et al. Transgenic mice overexpressing mutant PRKAG2 define the cause of Wolff-Parkinson-White syndrome in glycogen storage cardiomyopathy. *Circulation* **107**, 2850–2856 (2003).
15. Davies, J. K. et al. Characterization of the role of gamma2 R531G mutation in AMP-activated protein kinase in cardiac hypertrophy and Wolff-Parkinson-White syndrome. *Am. J. Physiol. Heart Circ. Physiol.* **290**, H1942–H1951 (2006).
16. Murphy, R. T. et al. Adenosine monophosphate-activated protein kinase disease mimicks hypertrophic cardiomyopathy and Wolff-Parkinson-White syndrome: natural history. *J. Am. Coll. Cardiol.* **45**, 922–930 (2005).
17. Sternick, E. B. et al. Clinical, electrocardiographic, and electrophysiologic characteristics of patients with a fasciculoventricular pathway: the role of PRKAG2 mutation. *Heart Rhythm* **8**, 58–64 (2011).
18. Scott, J. W. et al. CBS domains form energy-sensing modules whose binding of adenosine ligands is disrupted by disease mutations. *J. Clin. Invest.* **113**, 274–284 (2004).
19. Folmes, K. D. et al. Distinct early signaling events resulting from the expression of the PRKAG2 R302Q mutant of AMPK contribute to increased myocardial glycogen. *Circ. Cardiovasc. Genet.* **2**, 457–466 (2009).
20. Cheung, P. C., Salt, I. P., Davies, S. P., Hardie, D. G. & Carling, D. Characterization of AMP-activated protein kinase gamma-subunit isoforms and their role in AMP binding. *Biochem. J.* **346**(Pt 3), 659–669 (2000).
21. Zou, L. et al. N488I mutation of the gamma2-subunit results in bidirectional changes in AMP-activated protein kinase activity. *Circ. Res.* **97**, 323–328 (2005).
22. Hunter, R. W., Treebak, J. T., Wojtaszewski, J. F. & Sakamoto, K. Molecular mechanism by which AMP-activated protein kinase activation promotes glycogen accumulation in muscle. *Diabetes* **60**, 766–774 (2011).
23. Liang, X. et al. Transcription factor ISL1 is essential for pacemaker development and function. *J Clin Invest.* **125**, 3256–3268 (2015).
24. Hoogaars, W. M. et al. Tbx3 controls the sinoatrial node gene program and imposes pacemaker function on the atria. *Genes Dev.* **21**, 1098–1112 (2007).
25. DiFrancesco, D. The role of the funny current in pacemaker activity. *Circ. Res.* **106**, 434–446 (2010).
26. Lakatta, E. G., Maltsev, V. A. & Vinogradova, T. M. A coupled SYSTEM of intracellular Ca2+ clocks and surface membrane voltage clocks controls the timekeeping mechanism of the heart's pacemaker. *Circ. Res.* **106**, 659–673 (2010).
27. Jung, J. J. et al. Programming and isolation of highly pure physiologically and pharmacologically functional sinus-nodal bodies from pluripotent stem cells. *Stem Cell Rep.* **2**, 592–605 (2014).
28. Rimmbach, C., Jung, J. J. & David, R. Generation of murine cardiac pacemaker cell aggregates based on ES-cell-programming in combination with Myh6-promoter-selection. *J. Vis. Exp.* **96**, e52465 (2015).
29. Xiao, B. et al. Structural basis of AMPK regulation by small molecule activators. *Nat. Commun.* **4**, 3017 (2013).
30. Bultot, L. et al. A benzimidazole derivative small molecule 991 enhances AMPK activity and glucose uptake induced by AICAR or contraction in skeletal muscle. *Am. J. Physiol. Endocrinol. Metab.* **311**, E706–E719 (2016).
31. St Clair, J. R., Sharpe, E. J. & Proenza, C. Culture and adenoviral infection of sinoatrial node myocytes from adult mice. *Am. J. Physiol. Heart. Circ. Physiol.* **309**, H490–H498 (2015).
32. Yang, D., Lyashkov, A. E., Li, Y., Ziman, B. D. & Lakatta, E. G. RGS2 overexpression or G(i) inhibition rescues the impaired PKA signaling and slow AP firing of cultured adult rabbit pacemaker cells. *J. Mol. Cell Cardiol.* **53**, 687–694 (2012).
33. Siggs, O. M. et al. Mutation of Fnip1 is associated with B-cell deficiency, cardiomyopathy, and elevated AMPK activity. *Proc. Natl Acad. Sci. USA* **113**, E3706–E3715 (2016).
34. Stein, R., Medeiros, C. M., Rosito, G. A., Zimerman, L. I. & Ribeiro, J. P. Intrinsic sinus and atrioventricular node electrophysiologic adaptations in endurance athletes. *J. Am. Coll. Cardiol.* **39**, 1033–1038 (2002).
35. Wojtaszewski, J. F., Nielsen, P., Hansen, B. F., Richter, E. A. & Kiens, B. Isoform-specific and exercise intensity-dependent activation of 5'-AMP-activated protein kinase in human skeletal muscle. *J. Physiol.* **528**, 221–226 (2000).
36. Winder, W. W. & Hardie, D. G. Inactivation of acetyl-CoA carboxylase and activation of AMP-activated protein kinase in muscle during exercise. *Am. J. Physiol.* **270**, E299–E304 (1996).
37. Danson, E. J. & Paterson, D. J. Enhanced neuronal nitric oxide synthase expression is central to cardiac vagal phenotype in exercise-trained mice. *J. Physiol.* **546**, 225–232 (2003).
38. D'Souza, A. et al. Exercise training reduces resting heart rate via downregulation of the funny channel HCN4. *Nat. Commun.* **5**, 3775 (2014).
39. Gollob, M. H. et al. Novel PRKAG2 mutation responsible for the genetic syndrome of ventricular preexcitation and conduction system disease with childhood onset and absence of cardiac hypertrophy. *Circulation* **104**, 3030–3033 (2001).
40. Govindan, M., Ward, D. & Behr, E. A rare connection: fasciculoventricular pathway in PRKAG2 disease. *J. Cardiovasc. Electrophysiol.* **21**, 329–332 (2010).
41. Yang, X. et al. Physiological expression of AMPKgamma2RG mutation causes Wolff-Parkinson-White syndrome and induces kidney injury in mice. *J. Biol. Chem.* **291**, 23428–23439 (2016).
42. West, G. B., Woodruff, W. H. & Brown, J. H. Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals. *Proc. Natl Acad. Sci. USA* **99**(Suppl 1), 2473–2478 (2002).
43. Reppert, S. M. & Weaver, D. R. Coordination of circadian timing in mammals. *Nature* **418**, 935–941 (2002).
44. Baldesberger, S. et al. Sinus node disease and arrhythmias in the long-term follow-up of former professional cyclists. *Eur. Heart J.* **29**, 71–78 (2008).
45. Hayashi, S., Lewis, P., Pevny, L. & McMahon, A. P. Efficient gene modulation in mouse epiblast using a Sox2Cre transgenic mouse strain. *Mech. Dev.* **119** (Suppl 1), S97–S101 (2002).
46. Livak, K. J. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet. Anal.* **14**, 143–149 (1999).
47. Ashrafian, H. et al. Fumarate is cardioprotective via activation of the Nrf2 antioxidant pathway. *Cell Metab.* **15**, 361–371 (2012).
48. Schneider, J. E. et al. Fast, high-resolution in vivo cine magnetic resonance imaging in normal and failing mouse hearts on a vertical 11.7 T system. *J. Magn. Reson. Imaging* **18**, 691–701 (2003).
49. Maguire, M. L., Geethanath, S., Lygate, C. A., Kodibagkar, V. D. & Schneider, J. E. Compressed sensing to accelerate magnetic resonance spectroscopic imaging: evaluation and application to 23Na-imaging of mouse hearts. *J. Cardiovasc. Magn. Reson.* **17**, 45 (2015).
50. Flogel, U., Jacoby, C., Godecke, A. & Schrader, J. In vivo 2D mapping of impaired murine cardiac energetics in NO-induced heart failure. *Magn. Reson. Med.* **57**, 50–58 (2007).
51. Mattison, J. A. et al. Resveratrol prevents high fat/sucrose diet-induced central arterial wall inflammation and stiffening in nonhuman primates. *Cell Metab.* **20**, 183–190 (2014).
52. Berul, C. I., Aronovitz, M. J., Wang, P. J. & Mendelsohn, M. E. In vivo cardiac electrophysiology studies in the mouse. *Circulation* **94**, 2641–2648 (1996).
53. Gomes, J. et al. Electrophysiological abnormalities precede overt structural changes in arrhythmogenic right ventricular cardiomyopathy due to mutations

in desmoplakin-A combined murine and human study. *Eur. Heart J.* **33**, 1942–1953 (2012).
54. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547–1548 (2008).
55. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
56. Baruscotti, M. et al. Deep bradycardia and heart block caused by inducible cardiac-specific knockout of the pacemaker channel gene Hcn4. *Proc. Natl Acad. Sci. USA* **108**, 1705–1710 (2011).
57. DiFrancesco, D. A study of the ionic nature of the pace-maker current in calf Purkinje fibres. *J. Physiol.* **314**, 377–393 (1981).
58. Bois, P., Renaudon, B., Baruscotti, M., Lenfant, J. & DiFrancesco, D. Activation of f-channels by cAMP analogues in macropatches from rabbit sino-atrial node myocytes. *J. Physiol.* **501**(Pt 3), 565–571 (1997).
59. Sirenko, S. et al. Sarcoplasmic reticulum Ca2+ cycling protein phosphorylation in a physiologic Ca2+ milieu unleashes a high-power, rhythmic Ca2+ clock in ventricular myocytes: relevance to arrhythmias and bio-pacemaker design. *J. Mol. Cell Cardiol.* **66**, 106–115 (2014).
60. Ramskold, D. et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
61. Blankenberg, D. et al. Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26**, 1783–1785 (2010).
62. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
63. Trapnell, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46 (2013).
64. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
65. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
66. Wolfien, M. et al. TRAPLINE: a standardized and automated pipeline for RNA sequencing data analysis, evaluation and annotation. *BMC Bioinformatics* **17**, 21 (2016).
67. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
68. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
69. Bindea, G. et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
70. Ogata, H. et al. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
71. Pico, A. R. et al. WikiPathways: pathway editing for the people. *PLoS Biol.* **6**, e184 (2008).
72. Thomas, P. D. et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141 (2003).
73. Vinogradova, T. M. Z. Y. et al. Sinoatrial node pacemaker activity requires Ca2+/calmodulin-dependent protein kinase II activation. *Circ. Res.* **87**, 760–767 (2000).
74. Yang, D., Lyashkov, A. E., Li, Y., Ziman, B. D. & Lakatta, E. G. RGS2 overexpression or G_i inhibition rescues the impaired PKA signaling and slow AP firing of cultured adult rabbit pacemaker cells. *J. Mol. Cell Cardiol.* **53**, 687–694 (2012).
75. De Bono, J. P., Adlam, D., Paterson, D. J. & Channon, K. M. Novel quantitative phenotypes of exercise training in mouse models. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **290**, R926–R934 (2006).

Acknowledgements

We thank GenOway (Lyon, France) for generating R299Q γ2 mice; Rachel Hagen and Lee-Anne Stork for technical assistance in the British Heart Foundation (BHF) Experimental MRI Unit (Oxford, UK); Karen McGuire, Kate Thomson, and Jessica Woodley for help with R302Q genotyping (Oxford Medical Genetics Laboratories); Michael Shaw, Ashley Hale, and Craig Lygate (Oxford) for support with phenotyping; James Brown and Phil Townsend (Oxford) for general laboratory support; Shannon Marshall (NIA/NIH) for technical support; Paul Bastian and Elin Lehrmann of the Laboratory of Genetics and Genomics (NIA/NIH) for performing microarray experiments, analysis, and GEO submission. This work was supported by a BHF Intermediate Clinical Fellowship (FS/15/8/31155 to N.H.); BHF Senior Basic Science Research Fellowship (FS/11/50/29038 to J.E. S.); MRC/EPSRC Grant (G0600829, H.M. and M.L.M.); the Federal Ministry of Education and Research Germany (FKZ 0312138A and FKZ 316159), the State Mecklenburg-Western Pomerania with EU Structural Funds (ESF/IV-WM-B34-0030/10 and ESF/IV-BM-B35-0010/12), the DFG (DA 1296-1), the German Heart Foundation (F/01/12), the FORUN Program of Rostock University Medical Centre (889001), the EU funded CaSyM project (Grant Agreement #305033), the DAMP Foundation and the BMBF (VIP+ 00240) (to J.J.J., C.Ri., M.W., O.W., and R.D.); Ministero dell'Istruzione, dell'Università e della Ricerca (PRIN 2010BWY8E9) and the EU (LSHM-CT-2006-018676 NORMACOR) (D.D.); Fondazione Cariplo (CLARIFY, 2014-0822, M.B., and ACROSS, 2014-0728, D. D.); Intramural Research Program of the National Institutes of Health, National Institute on Aging (E.G.L); the Wellcome Trust (Research Training Fellowship, 086632/Z/08/Z), the Academy of Medical Sciences (Clinical Lecturer Starter Grant), and the National Institute for Health Research in the form of an Academic Clinical Lectureship (A.Y.). A.Y. (RE/08/004), H.W., and H.A. acknowledge support from the BHF Centre of Research Excellence, Oxford.

Author contributions

A.Y., K.P., C.R., H.W., and H.A. conceived the project. A.Y. designed and performed experiments, undertook data analysis/interpretation, and wrote the manuscript. K.P. developed the gene-targeting strategy and made the R299Q γ2 vector. M.Be., A.B., S.S., K.P., N.H., J.J.J., K.V.T., E.J.S., M.W., G.C., V.S., S.G., C.N., A.S., J.R.St.C., C.Ri., Y.O., D.Y., M.W., B.D.Z., J.M.M., D.R.R., C.Ra., M.P., J.L., J.Z., I.A., M.G.M., Y.S.T., D.B., N.S., H.L., R.P., J.d.B., O.M.S., J.G., H.M., M.L.M., Y.B., M.K., P.P.N.d.S., N.J.B., A.W., K.G., H.I., G.D., D.J.P.F., J.E.S., A.T., O.W., K.M.C., R.J.C., E.B.S., D.C., and M.B. performed experiments, analyzed and/or interpreted data, or provided reagents. A.B., N.H., E.B.S., D.J.P., C.S.R., D.C., C.P., R.D., M.B., D.DiF., E.G.L., H.W. and H.A. designed experiments and interpreted results. H.W. and H.A. supervised the project and co-wrote the manuscript with comments from co-authors.

Additional information

Supplementary Information accompanies this paper at doi:10.1038/s41467-017-01342-5.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Arash Yavari[1,2,3], Mohamed Bellahcene[2,3], Annalisa Bucchi[4,5], Syevda Sirenko[6], Katalin Pinter[2,3], Neil Herring[7], Julia J. Jung[8,9], Kirill V. Tarasov[6], Emily J. Sharpe[10], Markus Wolfien[11], Gabor Czibik[2,3], Violetta Steeples[2,3], Sahar Ghaffari[2,3], Chinh Nguyen[2,3], Alexander Stockenhuber[2,3], Joshua R.St. Clair[10], Christian Rimmbach[8,9], Yosuke Okamoto[6], Dongmei Yang[6], Mingyi Wang[6], Bruce D. Ziman[6], Jack M. Moen[6], Daniel R. Riordon[6],

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-017-01342-5 ARTICLE

Christopher Ramirez[6], Manuel Paina[4,5], Joonho Lee[6], Jing Zhang[6], Ismayil Ahmet[6], Michael G. Matt[6], Yelena S. Tarasova[6], Dilair Baban[3], Natasha Sahgal[3], Helen Lockstone[3], Rathi Puliyadi[2,3], Joseph de Bono[2,3], Owen M. Siggs[3,12], John Gomes[13], Hannah Muskett[2,3], Mahon L. Maguire[2,3], Youlia Beglov[2,3], Matthew Kelly[2,3], Pedro P.N. dos Santos[14], Nicola J. Bright[15], Angela Woods[15], Katja Gehmlich[2,3], Henrik Isackson[2], Gillian Douglas[2,3], David J.P. Ferguson[16], Jürgen E. Schneider[2,3], Andrew Tinker[13,17], Olaf Wolkenhauer[11,18], Keith M. Channon[2,3], Richard J. Cornall[3,12], Eduardo B. Sternick[14], David J. Paterson[7], Charles S. Redwood[2], David Carling[15], Catherine Proenza[10], Robert David[8,9], Mirko Baruscotti[4,5], Dario DiFrancesco[4,5], Edward G. Lakatta[6], Hugh Watkins[2,3] & Houman Ashrafian[1,2,3]

[1]Experimental Therapeutics, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DU, UK. [2]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DU, UK. [3]The Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, UK. [4]Department of Biosciences, Università degli Studi di Milano, Milan, 20133, Italy. [5]Centro Interuniversitario di Medicina Molecolare e Biofisica Applicata, University of Milano, Milan, 20133, Italy. [6]Laboratory of Cardiovascular Science, Intramural Research Program, National Institute on Aging, NIH, Baltimore, MD 21224, USA. [7]Burdon Sanderson Cardiac Science Centre, Department of Physiology, Anatomy & Genetics, University of Oxford, Oxford OX1 3PT, UK. [8]Department of Cardiac Surgery, Rostock University Medical Centre, 18057 Rostock Germany. [9]Department Life, Light and Matter, Interdisciplinary Faculty, Rostock University, 18059 Rostock Germany. [10]Department of Physiology and Biophysics, University of Colorado School of Medicine, Aurora, CO 80045, USA. [11]Department of Systems Biology and Bioinformatics, University of Rostock, Rostock 18051, Germany. [12]MRC Human Immunology Unit, Weatherall Institute for Molecular Medicine, Nuffield Department of Medicine, University of Oxford, Oxford OX3 9DS, UK. [13]Department of Medicine, BHF Laboratories, The Rayne Institute, University College London, London WC1E 6JJ, UK. [14]Instituto de Pós-Graduação, Faculdade de Ciências Médicas de Minas Gerais, Belo Horizonte 30.130-110, Brazil. [15]Cellular Stress Group, MRC London Institute of Medical Sciences, Imperial College London, London W12 0NN, UK. [16]Nuffield Department of Clinical Laboratory Science, University of Oxford, Oxford OX3 9DU, UK. [17]The Heart Centre, William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, London EC1M 6BQ, UK. [18]Stellenbosch Institute of Advanced Study (STIAS), Wallenberg Research Centre at Stellenbosch University, Stellenbosch 7602, South Africa. Hugh Watkins and Houman Ashrafian contributed equally to this work.

2.2.4 Community standards and software for whole-cell modeling

Waltemath, D., ..., **Wolfien, M.**, ..., and Schreiber, F. (2016).
Toward Community Standards and Software for Whole-Cell Modeling.
IEEE Transactions on Biomedical Engineering. IF: 4.491, Citations (December 14, 2020): 25

Whole-cell (WC) modeling is a promising tool for biological research, bioengineering, and medicine. However, substantial work remains to create accurate, comprehensive models of complex cells. We organized the 2015 Whole-Cell Modeling Summer School to teach WC modeling and evaluate the need for new WC modeling standards and software by recoding a published WC model in the Systems Biology Markup Language (SBML).

I was involved in understanding and processing the transcription submodel of *M. genitalium*, in which current open source modeling standards like the SBML and the Simulation Experiment Description Markup Language (SED-ML) were applied. The processing included the submodel encoding and improvement, as well as the development of an interface for the integration submodules that was related to the transcription part.

In summary, our analysis revealed several challenges to representing WC models using the current standards. We,therefore, propose several new WC modeling standards, software, and databases. We anticipate that these new standards and software will enable more comprehensive models.

IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 63, NO. 10, OCTOBER 2016 2007

Toward Community Standards and Software for Whole-Cell Modeling

Dagmar Waltemath*, Jonathan R. Karr, Frank T. Bergmann, Vijayalakshmi Chelliah, Michael Hucka, Marcus Krantz, Wolfram Liebermeister, Pedro Mendes, Chris J. Myers, *Fellow, IEEE*, Pinar Pir, Begum Alaybeyoglu, Naveen K Aranganathan, Kambiz Baghalian, Arne T. Bittig, Paulo E. Pinto Burke, Matteo Cantarelli, Yin Hoon Chew, Rafael S. Costa, Joseph Cursons, Tobias Czauderna, Arthur P. Goldberg, Harold F. Gómez, Jens Hahn, Tuure Hameri, Daniel F. Hernandez Gardiol, Denis Kazakiewicz, Ilya Kiselev, Vincent Knight-Schrijver, Christian Knüpfer, Matthias König, Daewon Lee, Audald Lloret-Villas, Nikita Mandrik, J. Kyle Medley, Bertrand Moreau, Hojjat Naderi-Meshkin, Sucheendra K. Palaniappan, Daniel Priego-Espinosa, Martin Scharm, Mahesh Sharma, Kieran Smallbone, Natalie J. Stanford, Je-Hoon Song, Tom Theile, Milenko Tokic, Namrata Tomar, Vasundra Touré, Jannis Uhlendorf, Thawfeek M Varusai, Leandro H. Watanabe, Florian Wendland, Markus Wolfien, James T. Yurkovich, Yan Zhu, Argyris Zardilis, Anna Zhukova, and Falk Schreiber

Manuscript received August 3, 2015; accepted April 18, 2016. Date of publication June 10, 2016; date of current version September 16, 2016. The Rostock and Utah meetings were supported by the Volkswagen Foundation (Grant 88495 to D. Waltemath and F. Schreiber). The work of J. R. Karr was supported by the James S. McDonnell Foundation Postdoctoral Fellowship Award in Studying Complex Systems and the National Science Foundation under Grant 1548123. The work of J. Cursons was supported by the Australian Research Council Centre of Excellence in Convergent Bio-Nano Science and Technology through Project CE140100036. *Asterisk indicates corresponding authors*.

*D. Waltemath is with the Institute of Computer Science, University of Rostock, 18051, Rostock, Germany (e-mail: dagmar.waltemath@uni-rostock.de).

J. R. Karr and A. P. Goldberg are with the Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai.

F. T. Bergmann is with BioQuant, University of Heidelberg.

V. Chelliah, N. K. Aranganathan, and A. Lloret-Villas are with the European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory.

M. Hucka is with the Department of Computing and Mathematical Sciences, California Institute of Technology.

M. Krantz, J. Hahn, and J. Uhlendorf are with the Department of Biology, Humboldt University of Berlin.

W. Liebermeister is with the Institute of Biochemistry, University Medicine Charité Berlin.

P. Mendes is with the Manchester Institute of Biotechnology and the School of Computer Science, University of Manchester, and also with the Center for Quantitative Medicine and the Department of Cell Biology, University of Connecticut Health Center.

C. J. Myers and L. H. Watanabe are with the Department of Electrical and Computer Engineering, University of Utah.

P. Pir is with Gebze Technical University.

B. Alaybeyoglu is with the Department of Chemical Engineering, Boğaziçi University.

K. Baghalian is with the Department of Plant Sciences, University of Oxford.

A. T. Bittig, M. Scharm, T. Theile, V. Touré, F. Wendland, and M. Wolfien are with the Institute of Computer Science, University of Rostock.

P. E. Pinto Burke is with the Institute of Science and Technology, Federal University of São Paulo.

M. Cantarelli is with OpenWorm.

Y. H. Chew is with the Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, and also with the Centre for Synthetic and Systems Biology, University of Edinburgh.

R. S. Costa is with the Centre of Intelligent Systems-IDMEC, Instituto Superior Técnico, University of Lisbon.

J. Cursons is with the Systems Biology Laboratory, Melbourne School of Engineering, University of Melbourne, and also with the ARC Centre of Excellence in Convergent Bio-Nano Science and Technology, Melbourne School of Engineering, University of Melbourne.

T. Czauderna is with the Faculty of Information Technology, Monash University.

H. F. Gómez is with the Department of Biosystems Science and Engineering, ETH Zürich.

T. Hameri and D. F. Hernandez Gardiol are with the Laboratory of Computational Systems Biotechnology, Swiss Federal Institute of Technology.

D. Kazakiewicz is with the Center for Statistics, Universiteit Hasselt, and also with the Center for Innovative Research, Medical University of Białystok.

I. Kiselev is with the Design Technological Institute of Digital Techniques, Siberian Branch of the Russian Academy of Sciences.

V. Knight-Schrijver is with the Babraham Institute.

C. Knüpfer is with the Institut für Informatik.

M. König is with the Institute of Biochemistry, Humboldt-University Berlin.

D. Lee and J.-H. Song are with the Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology.

N. Mandrik is with the Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences.

J. K. Medley is with the Department of Bioengineering, University of Washington.

B. Moreau is with CoSMo Company.

H. Naderi-Meshkin is with the Stem Cell and Regenerative Medicine Research Department, Iranian Academic Center for Education, Culture Research (ACECR).

S. K. Palaniappan is with the Rennes—Bretagne Atlantique Research Centre, Institute for Research in Computer Science and Automation.

D. Priego-Espinosa is with the Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México.

M. Sharma is with the Department of Pharmacoinformatics, National Institute of Pharmaceutical Education and Research.

K. Smallbone and N. J. Stanford are with the Manchester Centre for Integrative Systems Biology, University of Manchester.

M. Tokic is with the Laboratory of Computational Systems Biotechnology, Swiss Federal Institute of Technology, and also with the Swiss Institute of Bioinformatics.

N. Tomar is with the Department of Dermatology, University Medicine, Friedrich-Alexander University of Erlangen-Nürnberg.

T. M. Varusai is with the Department of Systems Biology Ireland, University College Dublin.

J. T. Yurkovich is with the Department of Bioengineering, University of California.

Y. Zhu is with the Monash Institute of Pharmaceutical Sciences, Monash University.

A. Zardilis is with the Centre for Synthetic and Systems Biology, University of Edinburgh.

A. Zhukova is with the Institut de Biochimie et Génétique Cellulaires, National Center for Scientific Research, and also with the University of Bordeaux.

F. Schreiber is with the Faculty of Information Technology, Monash University, and also with the Department of Computer and Information Science, University of Konstanz.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2016.2560762

***Abstract*— *Objective:* Whole-cell (WC) modeling is a promising tool for biological research, bioengineering, and medicine. However, substantial work remains to create accurate comprehensive models of complex cells. *Methods:* We organized the 2015 Whole-Cell Modeling Summer School to teach WC modeling and evaluate the need for new WC modeling standards and software by recoding a recently published WC model in the Systems Biology Markup Language. *Results:* Our analysis revealed several challenges to representing WC models using the current standards. *Conclusion:* We, therefore, propose several new WC modeling standards, software, and databases. *Significance:* We anticipate that these new standards and software will enable more comprehensive models.**

***Index Terms*—Computational biology, education, simulation, standards, systems biology, whole-cell (WC) modeling.**

I. Introduction

COMPUTATIONAL modeling is a powerful tool for biological research, bioengineering, and medicine to understand complex systems. It has been used to identify gene functions [1], engineer metabolic pathways [2], and identify drug targets [3]. Computational models also have the potential to help bioengineers design new microorganisms that can synthesize high-value chemicals, sense toxins, and decontaminate waste, as well as help clinicians interpret individual omics profiles and personalize medical therapy [4]. Realizing this potential requires more comprehensive models that can predict phenotype from genotype. In turn, this requires improved modeling and simulation standards and software [5]–[10].

Recently, Karr *et al.* developed the first whole-cell (WC) model which represents every individual gene function [11]. The model represents the life cycle of a single *Mycoplasma genitalium* bacterial cell and predicts the dynamics of every molecular species. The model is composed of 28 pathway submodels that are represented using multiple mathematical formalisms including stochastic simulation, ordinary differential equations (ODEs), flux balance analysis (FBA), and Boolean rules. The model was implemented in MATLAB.

The *M. genitalium* model has been used to gain novel insights into nongenetic cell cycle regulation mechanisms [11], learn unknown kinetic rate parameters from phenotypic data [12], calculate the metabolic costs of synthetic circuits [13], and repurpose antibiotics [14].

Karr *et al.* extensively documented the model, developed the WholeCellKB [15], WholeCellSimDB [16], and WholeCellViz [17] software tools to provide user-friendly interfaces to the model, and published the model open source. This has enabled other researchers to reuse the model [12]–[14].

However, significant domain expertise is still needed to reuse the model or to develop new WC models. The multialgorithm modeling methodology is complex. The model is difficult to understand, reuse, and extend because it is described directly in terms of its numerical simulation rather than in a software-independent format. The model code is difficult to learn and reuse because it is large, complex, and intertwined with the details of the *M. genitalium* model. The simulation code is also slow. Furthermore, the simulation code requires the proprietary MATLAB software package.

New standards and software tools are needed to help researchers build and simulate WC models. They would help researchers reuse, reproduce, and compare models, as well as share models through repositories such as BioModels [18].

Several systems biology standards have been developed by the *COmputational Modeling in BIology NEtwork* (COMBINE) [8], including the *Systems Biology Markup Language* (SBML) [19], CellML [20], the *Simulation Experiment Description Markup Language* (SED-ML) [21], and the *Systems Biology Graphical Notation* (SBGN) [22] (see Table I). SBML and CellML are formats for representing mathematical models. CellML describes the mathematics, whereas SBML describes biological processes. Both support several modeling formalisms including ODEs and FBA. SED-ML describes and enables researchers to reproduce computational experiments. SBGN is a visual notation for describing biological processes. However, none of these standards have been used for WC modeling.

We organized the 2015 Whole-Cell Modeling Summer School to train students in WC modeling and to evaluate the need for new WC modeling standards and software. The school focused on creating a reusable WC model by recoding the *M. genitalium* model in SBML. We focused on SBML because SBML is the most widely used systems biology standard and there was insufficient time to evaluate multiple standards. The school also aimed to improve numerous details of the model, visualize the model with SBGN, and describe model simulations with SED-ML. The SBML-encoded submodels and SBGN diagrams are available at https://github.com/whole-cell-tutors/wholecell/releases/tag/meeting-report.

Most importantly, the school generated extensive community discussion on how to best build and simulate WC models. This report describes the outcome of these discussions, including our recommendations for new standards and software to accelerate WC modeling. We also describe our progress toward recoding the *M. genitalium* model in SBML and the lessons that we learned about organizing research-based schools.

II. 2015 Whole-Cell Modeling Summer School

The school was held March 9–13, 2015, at the University of Rostock, Germany. It was organized by D. Waltemath and F. Schreiber and funded by the Volkswagen Foundation. 43 students and nine instructors participated in the school. A follow up meeting involving 15 of the original and six additional participants was held October 10–11, 2015, at the University of Utah, USA. All of the materials for the school are available at http://sites.google.com/site/vwwholecellsummerschool.

We advertised the school through community mailing lists, conference calendars, and websites. Applicants were asked to describe their experience and interest in WC modeling. We chose 43 participants from 118 applicants based on three criteria.

1) We identified the most qualified and enthusiastic applicants.
2) We gave preference to students, female applicants, and applicants from developing countries.

TABLE I
SYSTEMS BIOLOGY STANDARDS AND STANDARDIZATION EFFORTS

Acronym	Name	Type	Description	Ref.
CellML	CellML	Standard	Describes models in terms of mathematical relationships	20
COMBINE	COmputational Modeling in BIology NEtwork	Community	Develops computational biology standards and software	8
SBGN	Systems Biology Graphical Notation	Standard	Describes biochemical pathway diagrams	23
SBML	Systems Biology Markup Language	Standard	Describes models in terms of biochemical processes	24
SBML Arrays	SBML Package: Arrays	Standard	Describes arrays	25
SBML Comp	SBML Package: Hierarchical Model Composition	Standard	Describes how model are composed from other models	26
SBML Distrib	SBML Package: Distributions	Standard	Describes random distributions	27
SBML FBC	SBML Package: Flux Balance Constraints	Standard	Describes constraint-based models	28
SBML Multi	SBML Package: Multistate and Multicomponent Species	Standard	Supports rule-based modeling	25
SBML Spatial	SBML Package: Spatial Processes	Standard	Describes spatially-resolved models	29
SED-ML	Simulation Experiment Description Markup Language	Standard	Describes computational experiments	21

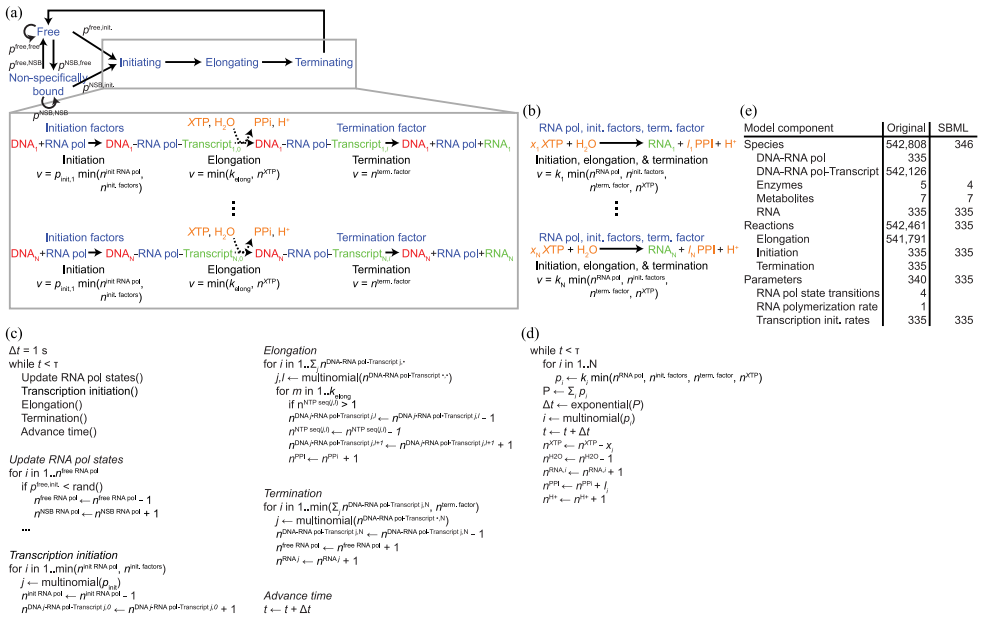


Fig. 1. Comparison of the original and SBML transcription submodels. (a) Original transcription submodel included two subsubmodels: (1) a Markov model that describes how RNA polymerase switches among freely diffusing, nonspecifically bound, and initiating states and (2) an ad hoc stochastic model that describes how RNA polymerase initiates transcription, elongates individual bases by walking along DNA, and terminates transcripts. (b) We created the SBML transcription submodel by simplifying the original submodel. The SBML submodel only represents transcription initiation, elongation, and termination; lumps the initiation, elongation, and termination of each RNA species into a single reaction; and does not explicitly represent DNA-protein binding. (c) Equivalent population-based ad hoc stochastic simulation algorithm for the original submodel. The original submodel was implemented using a more efficient particle-based algorithm. To facilitate comparison with the population-based SBML version, we have described an equivalent population-based algorithm. (d) We also improved the SBML submodel by replacing the ad hoc stochastic simulation algorithm with the Gillespie algorithm. (e) Statistics of the original and improved transcription submodels in population-based representations.

3) We selected participants to represent a broad range of scientific disciplines.

We used the same criteria to select instructors.

The school began with introductory lectures on WC modeling and the existing systems biology standards by J. Karr and M. Hucka and introductory discussions on model composition, state representation, and stochastic modeling. Most of the school was devoted to active learning sessions in which the students and instructors were divided into 11 groups and challenged to use SBML to recode the *M. genitalium* model, use SBGN to visualize the model, and use SED-ML to simulate the model. Groups 1–8 encoded submodels. Group 9 developed a submodel integration scheme. Group 10 annotated and visualized the model. Group 11 helped all of the other groups understand, encode, and improve the model. Table SI, available in the online supplementary material, lists the groups and participants of both meetings. Each day concluded with community discussions. In addition, the school included a poster session and networking activities.

The students learned about state-of-the-art WC modeling; the open challenges to building more complex models; open-

source modeling software; the importance of reproducibility; and the SBML, SED-ML, and SBGN standards. The students also expanded their professional networks. Several of the students reported that the skills and knowledge they gained from the school would enhance their research.

We learned several lessons about organizing research-based schools.

1) Students enjoy working on research problems more than solving prescribed exercises. This engages students in the field, challenges them, and helps them build practical skills.
2) Research-based schools should have clear background knowledge expectations, learning objectives, and research goals. This helps students decide whether to participate, prepare, and learn efficiently.
3) Research-based schools should have a flexible schedule, multidisciplinary participants, and a high teacher-to-student ratio.

This allows students to engage in impromptu discussions, draw on multiple perspectives, and get feedback and iterate quickly.

III. Toward an Improved SBML-Encoded WC Model

In addition to teaching students about WC modeling and the systems biology standards, the school aimed to improve the *M. genitalium* model and to encode the model in SBML.

A. Submodel Encoding

We pursued several strategies to encode submodels in SBML. Several groups encoded submodels by reading the original documentation of the model; drawing pathway diagrams using software tools such as CellDesigner [30] and VANTED [31], and writing scripts to generate SBML from the diagrams. Other groups used model design tools such as Antimony [32], BioUML [33], COBRApy [34], COPASI [35], iBioSim [36], and libRoadRunner [37] to recode submodels based on the original documentation. A few of the groups encoded submodels by converting the MATLAB code to SBML. As an example, Fig. 1 and File S1 illustrate how we recoded the transcription submodel.

We encountered several challenges to encoding the submodels in SBML. First, understanding the submodels was time-consuming because many students were not familiar with the modeled biology, many of the submodel details are described only in the MATLAB code, and the model documentation only summarizes the model. For these reasons, J. Karr, one of the authors of the original model, helped all of the groups understand the modeled biology and mathematics. Dr. Karr also helped several groups simplify their encoding tasks by recommending that they recode only the most important model components. For example, Dr. Karr suggested that the transcription group represent the transcription of each RNA species as a single lumped reaction rather than hundreds of thousands of individual base elongation reactions. It would have been challenging to recode the model without Dr. Karr. The essentiality of Dr. Karr's guidance underscores the need for improved WC modeling methods and standards.

Second, it was difficult to encode the original serial and randomized algorithms into SBML because SBML does not explicitly represent sequential operations and plain SBML does not support random number generation. We overcame these problems by formalizing submodels as Gillespie algorithm stochastic simulations [38].

Third, in many cases, we had to either enumerate the particle-based state representations used by the original model or approximate the original model. For example, the translation group approximated the original model by lumping all of the elongation reactions for each protein into a single reaction. The replication group used indicator variables to enumerate the particle-based chromosome representation from the original model. However, this enumerated representation requires millions of variables, which is prohibitively expensive, and makes it difficult to represent the exclusion of multiple proteins from binding the same base. Furthermore, it is impractical to edit this verbose enumerated representation.

Fourth, we had to enumerate all of the arrays used by the original model because few SBML simulators support arrays. This created verbose SBML files that are difficult to interpret and maintain and slow to simulate.

In summary, we concluded that it is currently difficult to encode WC models in SBML. WC modeling would be accelerated by expanded software support for model composition, rule-based modeling, arrays, and random number generation.

B. Submodel Improvement

We also improved several aspects of the original model. As described above, we replaced the ad hoc stochastic simulation algorithms and rate laws used by the original submodels with the Gillespie algorithm and mass action kinetics. As an example, Fig. 1 and File S1 compare the original and SBML versions of the transcription submodel. We anticipate that these changes will improve the biological accuracy of WC models. The original model used these ad hoc algorithms and rate laws to achieve sufficient performance. Going forward, a high-performance parallel simulator is needed to achieve adequate performance of the Gillespie algorithm.

C. Model Integration

The integration group created a scheme for combining the submodels. First, they defined the global species as the union of all submodel species. Second, they standardized the species names to create consistent submodel-global species interfaces.

Third, the group designed a new multialgorithm simulation strategy to overcome the limitations of the original simulation algorithm. In particular, the group sought to correctly implement the arrow of time by integrating submodels within the same time step based on the same input state. The integration group also sought to develop an algorithm that has a variable time step that can be optimized to balance accuracy and performance.

1) The group considered sequentially integrating the submodels within each time step and setting the time step small enough that only one submodel would advance the cell state within each time step. However, this strategy is prohibitively expensive.

TABLE II
NEW STANDARDS AND SOFTWARE NEEDED TO ACCELERATE WC MODELING

Type	Description
Database	Expanded molecular biological databases such as ChEBI [39]
Software	Data curation tools for aggregating the data to build models
Software	Pathway/genome database to organize model training data
Standard	Sequence- and rule-based multialgorithmic modeling language
Software	Model design tools that generate models from pathway/genome databases
Software	Distributed parameter estimation tools
Software	Frameworks for systematically verifying model
Software	High-performance, parallel, rule-based multialgorithm simulator
Standard	Extended SBGN standard for hybrid maps containing Process Description, Entity Relationship, and Activity Flow nodes
Software	Visualization software that supports contextual zooming

2) The group considered generalizing the original algorithm by dividing each of the global species pools into multiple, independent subspecies pools for each submodel; integrating the submodels in parallel; and merging the subspecies to update the global species. However, it is difficult to apply this strategy to coupled variables such as those that represent the protein occupancy of the chromosome.
3) The group decided to interpret the species changes predicted by each submodel as requests and implement a central controller that accepts or rejects these changes at the end of each time step to update the global species.

This strategy is computationally efficient and generalizable.

Finally, the group explored implementing this algorithm using both the SBML hierarchical model composition package [26] and SED-ML shared variables. The group concluded that both implementations are feasible. The group used iBioSim to test these strategies because iBioSim is one of the only SBML-compatible simulators that supports model composition.

D. Annotation, Documentation, and Visualization

The documentation group was responsible for annotating the model. The group aimed to define every model element independently from external databases and to provide cross references to databases where possible to help users interpret the model. For example, they used InChI [40] to define small molecule species in terms of structures. They defined DNA, RNA, and protein species as polymers of small molecules. The group wrote scripts to identify cross references for each model entity. However, many entities are not represented by any database. The group contributed the missing metabolite structures to ChEBI [39] and concluded that the biological databases must be expanded to help aggregate data for models.

The group also helped the other groups visualize submodels by providing advice on SBGN and diagramming tools such as SBGN-ED [41], a VANTED add-on for creating, editing, and validating SBGN diagrams. The main visualization problem encountered by the group was that WC models require large intuitive diagrams that are difficult to lay out automatically.

E. Progress and Future Work

We produced draft SBML and SBGN versions of the submodels. However, significant work remains to combine, identify, and verify the submodels. Using the lessons learned, a subgroup of the participants are continuing to recode the submodels and integrate the submodels into a single model. We expect that the final model will be more scalable, extensible, and easy to use than the original model. We also plan to build an SBML-compatible multialgorithm simulator by expanding analysis tools, such as iBioSim and BioUML.

After recoding the model, we plan to identify and validate the new model. We will validate the model in two steps.

1) We will use the experimental data that was used to validate the original model.
2) To more rigorously validate the new model, we will compare the model to newly published single-gene deletion strain growth rates [12] that were not available when the original model was developed.

We aim to publish the SBML-encoded model to BioModels, along with SED-ML tests, SBGN diagrams, and textual documentation. Publication in BioModels will make the model searchable, retrievable, and reusable. We believe this valuable community resource will demonstrate how to describe WC models in standard formats, and it will help other researchers build upon the model.

IV. TOWARD SBML-, SED-ML-, AND SBGN-BASED STANDARDS FOR WC MODELING

The school was the first attempt to encode a WC model using standards. Thus, we were not surprised to learn that the current standards and community software do not easily support WC modeling. Importantly, the school generated ideas for new WC modeling standards and software that will enable researchers to build vastly more comprehensive models.

A. New Standards

Two new standards are needed to facilitate WC modeling. A new SBML package should be created to support DNA, RNA, and protein sequence-based reaction patterns. This would enable researchers to easily model sequence-dependent reactions such as the methylation or protein binding of specific DNA motifs. This package would also help integrate genomics and bioinformatics with systems modeling.

SBGN must also be expanded to support: 1) hybrid diagrams that contain process description, entity relationship, and activ-

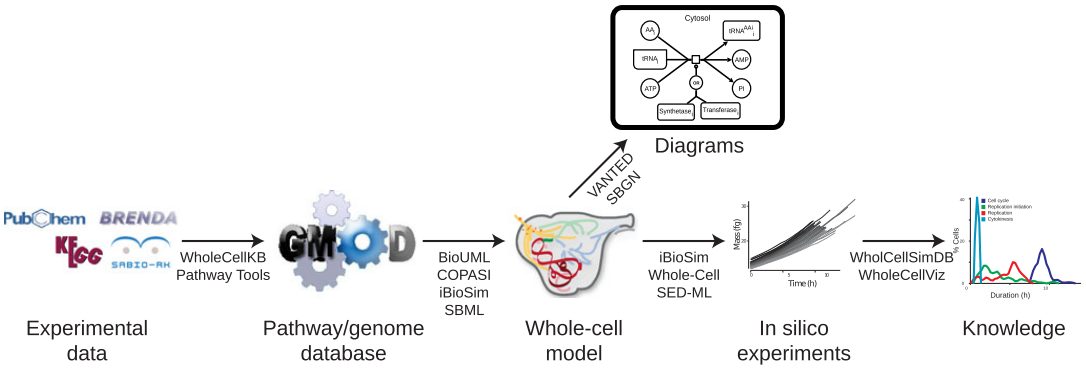


Fig. 2. WC modeling workflow. Researchers will (1) assemble data into pathway/genome databases, (2) use these databases to construct models, (3) identify and verify models, (4) use multialgorithm simulators to conduct *in silico* experiments, and (5) analyze these experiments to discover biology.

ity flow elements; and 2) visualizations at multiple levels of granularity.

B. New Software Tools and Databases

Several new software tools and databases are needed to accelerate WC modeling (see Table II). A high-performance simulator must be developed. This simulator should be parallelized to enable the simulation of vastly larger models that require more computing and memory than are available on a single machine. This requires research to determine how to concurrently integrate mathematically heterogeneous submodels that share state. The simulator should leverage recent advances in parallel discrete event simulation [42].

The simulator must also implement the SBML Multistate and Multicomponent Species package [43] to support rule-based modeling. This will enable more succinct model descriptions, making models easier to understand and edit. For example, translation could be described using a single reaction pattern parameterized by mRNA-specific translation initiation rates rather than by enumerating each individual reaction. By separating mathematical descriptions from parameter values, reaction patterns will also clarify the connection between dynamical models and their underlying data. Implementing this package would also enable modelers to efficiently simulate models with combinatorial state spaces, which, in turn, will enable the encoding of more complex models.

Ultimately, to accurately predict phenotypes, WC models must also represent spatially-dependent processes. Currently, researchers are independently pursuing WC and spatial modeling. For example, the *M. genitalium* model only represents three compartments, and the most advanced spatial models only represent individual pathways. WC and spatial modeling should be combined by adding support for the SBML Spatial Processes package [29] to the new WC simulator.

New model design software must be developed to help researchers quickly build WC models. This software should help researchers systematically build WC models from experimental data organized into pathway/genome database. In turn, this software will help researchers build bigger models.

New data curation tools are needed to aggregate data to build more comprehensive models. The software should automatically aggregate data from public databases, as well as accelerate manual curation from individual publications. This software will also make WC models more reproducible by automatically recording each data source. Natural language processing [44], crowdsourcing [45], and machine learning should also be explored to accelerate data curation.

New pathway/genome database software is needed to organize the data required to build WC models. To clarify the connection between computational models and their underlying experimental data, this software should use semantic annotations to describe how experimental data are used to build computational models.

New model parameter estimation and model verification tools are also needed to identify and verify computationally expensive WC models. To better estimate WC models, we must generalize our model reduction methods and adopt distributed numerical optimization techniques [46]. To more systematically verify WC models, we should adopt formal probabilistic verification techniques from electrical engineering [47].

New algorithms are needed to automatically create intuitive visualizations of large networks and the SBGN viewers should utilize contextual zooming to display diagrams at multiple levels of granularity.

In addition, biological databases, such as ChEBI, must be expanded to help researchers annotate WC models in terms of external entities.

C. Systematic WC Modeling Pipeline

The new standards and software tools will enable a five step approach to WC model-driven discovery (see Fig. 2).

1) Researchers will use data curation tools to aggregate heterogeneous data into pathway/genome databases. These

databases will use semantic annotations to describe the connection between models and their underlying data.
2) Researchers will use design tools to build WC models from pathway/genome databases. These tools will export models to software-independent formats such as SBML.
3) Model identification and verification tools will be used to estimate parameters and test models.
4) A multialgorithm simulator will be used to conduct *in silico* experiments.
5) Simulation databases and visualization software such as WholeCellSimDB and WholeCellViz will be used to discover new biology by visualizing and analyzing *in silico* experiments.

Together, this pipeline will enable more researchers to more easily build, manage, simulate, and reproduce WC models. These new tools will also enable researchers to build more comprehensive models of more complex eukaryotic cells. Ultimately, this will enable WC modeling to support synthetic biology and personalized medicine.

V. Conclusion

The 2015 Whole-Cell Modeling Summer School trained young scientists in WC modeling and standards by challenging them to recode a WC model in SBML. Additional courses are needed to provide theoretical training in multialgorithm modeling, model reduction, and parameter estimation, as well as practical training in WC model building.

We made significant strides toward recoding the model in SBML. We also improved the model by replacing the ad hoc algorithms and rate laws used by the original model with the Gillespie algorithm and mass action kinetics. We designed an improved multialgorithm simulation metaalgorithm. Through validating the model by comparison to quantitative growth rate measurements, we anticipate that we will also discover and add several unknown parallel pathways to the model. We have produced preliminary SBML versions of all of the submodels of the *M. genitalium* model, and we are working to develop a software program to simulate the combined model. We plan to publish the new SBML-encoded model to BioModels.

Most importantly, our community discussions generated clear goals for new WC modeling software and standards. We recommend that researchers develop a new SBML-compatible simulator that supports both model composition and sequence- and rule-based modeling, as well as develop new model design, parameter estimation, model testing, and visualization tools. We also recommend expanding the biological databases to facilitate model building and annotation. Furthermore, we believe that SBGN should be extended to support hybrid diagrams, advanced graph layout, and contextual zooming. Finally, we recommend evaluating CellML as another potential WC modeling standard.

In summary, we believe that WC modeling will be an important tool for biological science, bioengineering, and medicine. Achieving this potential requires new WC modeling software and standards. In turn, this requires expanding the WC modeling field, including training young researchers.

References

[1] J. L. Reed *et al.*, "Systems approach to refining genome annotation," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 46, pp. 17480–17484, 2006.
[2] J. W. Lee *et al.*, "Systems metabolic engineering of microorganisms for natural and non-natural chemicals," *Nature Chem. Biol.*, vol. 8, no. 6, pp. 536–546, 2012.
[3] D. S. Lee *et al.*, "Comparative genome-scale metabolic reconstruction and flux balance analysis of multiple *Staphylococcus aureus* genomes identify novel antimicrobial drug targets," *J. Bacteriol.*, vol. 191, no. 12, pp. 4015–4024, 2009.
[4] J. Carrera and M. W. Covert, "Why build whole-cell models?" *Trends Cell Biol*, vol. 25, no. 12, pp. 719–722, 2015.
[5] D. N. Macklin *et al.*, "The future of whole-cell modeling," *Curr. Opin. Biotechnol.*, vol. 28, pp. 111–115, 2014.
[6] J. R. Karr *et al.*, "The principles of whole-cell modeling," *Curr. Opin. Microbiol.*, vol. 27, pp. 18–24, 2015.
[7] J. R. Karr *et al.*, "Summary of the DREAM8 parameter estimation challenge: Toward parameter identification for whole-cell models," *PLoS Comput. Biol.*, vol. 11, no. 5, p. e1004096, 2015.
[8] M. Hucka *et al.*, "Promoting coordinated development of community-based information standards for modeling in biology: The COMBINE initiative," *Frontiers Bioeng. Biotechnol.*, vol. 3, p. 19, 2015.
[9] E. Klipp *et al.*, "Systems biology standards—The community speaks," *Nat. Biotechnol.*, vol. 25, no. 4, pp. 390–391, 2007.
[10] F. Büchel *et al.*, "Path2Models: Large-scale generation of computational models from biochemical pathway maps," *BMC Syst. Biol.*, vol. 7, no. 1, p. 116, 2013.
[11] J. R. Karr *et al.*, "A whole-cell computational model predicts phenotype from genotype," *Cell*, vol. 150, no. 2, pp. 389–401, 2012.
[12] J. C. Sanghvi *et al.*, "Accelerated discovery via a whole-cell model," *Nature Methods*, vol. 10, no. 12, pp. 1192–1195, 2013.
[13] O. Purcell *et al.*, "Towards a whole-cell modeling approach for synthetic biology," *Chaos*, vol. 23, no. 2, p. 025112, 2013.
[14] D. Kazakiewicz *et al.*, "A combined systems and structural modeling approach repositions antibiotics for mycoplasma genitalium," *Comput. Biol. Chem.*, vol. 59, pp. 91–97, 2015.
[15] J. R. Karr *et al.*, "WholeCellKB: Model organism databases for comprehensive whole-cell models," *Nucleic Acids Res.*, vol. 41, pp. D787–D792, 2013.
[16] J. R. Karr *et al.*, "WholeCellSimDB: A hybrid relational/HDF database for whole-cell model predictions," *Database*, vol. 2014, p. bau095, 2014.
[17] R. Lee *et al.*, "WholeCellViz: Data visualization for whole-cell models," *BMC Bioinformat.*, vol. 14, p. 253, 2013.
[18] V. Chelliah *et al.*, "BioModels: Ten-year anniversary," *Nucleic Acids Res.*, vol. 43, no. D1, pp. D542–D548, 2015.
[19] M. Hucka *et al.*, "The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models," *Bioinformatics*, vol. 19, no. 4, pp. 524–531, 2003.
[20] W. J. Hedley *et al.*, "A short introduction to CellML," *Philos. Trans. R. Soc. London A*, vol. 359, pp. 1073–1089, 2001.
[21] D. Waltemath *et al.*, "Reproducible computational biology experiments with SED-ML—The simulation experiment description markup language," *BMC Syst. Biol.*, vol. 5, no. 1, p. 198, 2011.
[22] N. Le Novère *et al.*, "The systems biology graphical notation," *Nature Biotechnol.*, vol. 27, pp. 735–741, 2009.
[23] N. Le Novère *et al.*, "The systems biology graphical notation," *Nature Biotechnol.*, vol. 27, no. 8, pp. 735–741, 2009.
[24] M. Hucka *et al.*, "The systems biology markup language (SBML): Language specification for level 3 version 1 core." *J. Integrative Bioinformat.*, vol. 12, no. 2, p. 266, 2015.
[25] D. Waltemath *et al.*, "Meeting report from the fourth meeting of the computational modeling in biology network (COMBINE)," *Standards Genomic Sci.*, vol. 9, no. 3, pp. 1285–1301, 2014.
[26] L. P. Smith *et al.*, "SBML level 3 package: Hierarchical model composition, version 1 release 3," *J. Integrative Bioinformat.*, vol. 12, no. 2, p. 268, 2015.
[27] S. L. Moodie *et al.*, "The distributions package for SBML level 3," 2015. [Online]. Available: http://sourceforge.net/p/sbml/code/HEAD/tree/trunk/specifications/sbml-level-3/version-1/distrib/sbml-level-3-distrib-package-proposal.pdf?format=raw. Accessed on: Feb 26, 2016.

[28] B. G. Olivier and F. T. Bergmann, "The systems biology markup language (SBML) level 3 package: Flux balance constraints," *J. Integrative Bioinformat.*, vol. 12, no. 2, p. 269, 2015.
[29] J. C. Schaff *et al.*, "SBML level 3 package: Spatial processes," 2015.
[30] A. Funahashi *et al.*, "CellDesigner 3.5: A versatile modeling tool for biochemical networks," *Proc. IEEE*, vol. 96, no. 8, pp. 1254–1265, Aug. 2008.
[31] H. Rohn *et al.*, "VANTED v2: A framework for systems biology applications," *BMC Syst. Biol.*, vol. 6, p. 139, 2012.
[32] L. P. Smith *et al.*, "Antimony: A modular model definition language," *Bioinformatics*, vol. 25, no. 18, pp. 2452–2454, 2009.
[33] F. Kolpakov, "BioUML: Visual modeling, automated code generation and simulation of biological systems," *Proc. 5th Int. Conf. Bioinformat. Genome Regulation Struct.*, vol. 3, pp. 281–285, 2006.
[34] A. Ebrahim *et al.*, "COBRApy: Constraints-based reconstruction and analysis for python," *BMC Syst. Biol.*, vol. 7, no. 1, p. 74, 2013.
[35] S. Hoops *et al.*, "COPASI—A COmplex PAthway SImulator," *Bioinformatics*, vol. 22, pp. 3067–3074, 2006.
[36] C. Madsen *et al.*, "Design and test of genetic circuits using iBioSim," *IEEE Des. Test Comput.*, vol. 29, no. 3, pp. 32–39, Oct. 2012.
[37] E. T. Somogyi *et al.*, "libRoadRunner: A high performance SBML simulation and analysis library," *Bioinformatics*, vol. 31, no. 20, Oct 2015.
[38] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *J. Phys. Chem.*, vol. 81, no. 25, pp. 2340–2361, 1977.
[39] J. Hastings *et al.*, "ChEBI in 2016: Improved services and an expanding collection of metabolites," *Nucleic Acids Res.*, vol. 44, no. D1, pp. D1214–D1219, 2016.
[40] S. R. Heller *et al.*, "InCHI—The worldwide chemical structure identifier standard," *J. Cheminformat.*, vol. 5, no. 1, p. 7, 2013.
[41] T. Czauderna *et al.*, "Editing, validating and translating of SBGN maps," *Bioinformatics*, vol. 26, no. 18, pp. 2340–2341, 2010.
[42] A. P. Goldberg *et al.*, "Toward scalable whole-cell modeling of human cells," in *Proc. Annu. ACM Conf. SIGSIM Principles Adv. Discrete Simul.*, 2016, pp. 259–262.
[43] F. Zhang and M. Meier-Schellersheim, "SBML Level 3 Package Specification: Multistate/Multicomponent Species (Version 1, Release 0.1 Draft 369)," 2015. [Online]. Available: http://sbml.org/Documents/Specifications/SBML_Level_3/Packages/multi. Accessed on: May 25, 2015.
[44] P. M. Nadkarni *et al.*, "Natural language processing: An introduction," *J. Amer. Med. Inform. Assoc.*, vol. 18, no. 5, pp. 544–551, 2011.
[45] B. M. Good and A. I. Su, "Crowdsourcing for bioinformatics," *Bioinformatics*, vol. 2013, p. btt333, Jun. 2013.
[46] A. F. Villaverde *et al.*, "A cooperative strategy for parameter estimation in large scale systems biology models," *BMC Syst. Biol.*, vol. 6, no. 1, p. 1, 2012.
[47] M. Kwiatkowska *et al.*, "PRISM 4.0: Verification of probabilistic real-time systems," in *Comput Aided Verification*. Berlin, Germany: Springer, 2011, pp. 585–591.

The 2015 Whole-Cell Modeling Summer School in Rostock included the 54 participants listed in Table SI. Photo: University of Rostock IT and Media Center.

2.3 Integration of heterogeneous data in clinical stem-cell therapy

Systems medicine emerged as an invaluable tool to investigate ***complex diseases*** *by integrating multidimensional datasets and numerous mathematical approaches with data from pre-clinical and clinical studies. Stem cell-based regenerative therapies for the treatment of ischemic myocardium are currently a subject of intensive investigation. A variety of cell populations have been demonstrated to be safe and to exert some positive effects in human Phase I and II clinical trials; however, conclusive evidence of efficacy is still lacking. While the relevance of animal models for appropriate pre-clinical safety and efficacy testing with regard to application in Phase III studies continues to increase, concerns have been expressed regarding the validity of the mouse model to predict clinical results. In this Section, I introduce current developments in stem cell applications in cardiac regeneration and facilitate the basic understanding of stem cell efficiency via computational methods. In particular, a literature-based meta-analysis of 21 studies was conducted to investigate the effect of regenerative cell therapies after cardiac infarction in mice. Furthermore, data from 82 patients of a Phase III clinical trial was analyzed by means of ML and network approaches to determine a therapy response stratification and biomarker signature identification.*

2.3.1 Regeneration of heart diseases by means of stem cell applications

Steinhoff, G., Nesteruk, J., **Wolfien, M.**, Große, J., Ruch, U., Vasudevan, P., and Müller, P. (2017).
Stem cells and heart disease - Brake or accelerator?
Advanced Drug Delivery Reviews. IF: 16.361, Citations (December 14, 2020): 21

After two decades of intensive research and attempts of clinical translation, stem cell-based therapies for cardiac diseases are not getting closer to clinical success. This review tries to unravel the obstacles and focuses on underlying mechanisms, as the target for regenerative therapies. At present, the principal outcome in clinical therapy does not reflect the experimental evidence. It seems that the scientific obstacle is a lack of knowledge integration obtained from tissue repair and disease mechanisms.

In this article, I discussed potential standards for a proper, reproducible data analysis strategy in clinical research and hospital settings. Data mining, data management, NGS processing, network approaches, and ML are major computational approaches that must be integrated into sustainable clinical workflows because they yield a high potential, especially in the field of stem cell research. By using bioinformatics and systems medicine approaches, I proposed a semi-automated and self-adapting processing cyclus to evolve regenerative therapies as an interative process.

In summary, recent findings from clinical trials delineate mechanisms of stem cell dysfunction and gene defects in repair mechanisms as a cause of atherosclerosis and heart disease. These findings require a redirection of current practice of stem cell therapy and a reset using more detailed analyses of stem cell function interfering with disease mechanisms. To accelerate scientific development we suggest intensifying unified computational data analyses and shared-data knowledge by using controlled open-access data platforms.

Advanced Drug Delivery Reviews 120 (2017) 2–24

Contents lists available at ScienceDirect

Advanced Drug Delivery Reviews

journal homepage: www.elsevier.com/locate/addr

Stem cells and heart disease - Brake or accelerator?☆

Gustav Steinhoff [a,*], Julia Nesteruk [a], Markus Wolfien [b], Jana Große [a], Ulrike Ruch [a], Praveen Vasudevan [a], Paula Müller [a]

[a] University Medicine Rostock, Department of Cardiac Surgery, Reference and Translation Center for Cardiac Stem Cell Therapy, University Medical Center Rostock, Schillingallee 35, 18055 Rostock, Germany
[b] University Rostock, Institute of Computer Science, Department of Systems Biology and Bioinformatics, Ulmenstraße 69, 18057 Rostock, Germany

ARTICLE INFO

Article history:
Received 28 July 2017
Received in revised form 12 October 2017
Accepted 13 October 2017
Available online 18 October 2017

Keywords:
Stem cell
Cardiac disease
Tissue repair
HSC
EPC
MSC
SH2B3
Quality management
Systems medicine

ABSTRACT

After two decades of intensive research and attempts of clinical translation, stem cell based therapies for cardiac diseases are not getting closer to clinical success. This review tries to unravel the obstacles and focuses on underlying mechanisms as the target for regenerative therapies. At present, the principal outcome in clinical therapy does not reflect experimental evidence. It seems that the scientific obstacle is a lack of integration of knowledge from tissue repair and disease mechanisms. Recent insights from clinical trials delineate mechanisms of stem cell dysfunction and gene defects in repair mechanisms as cause of atherosclerosis and heart disease. These findings require a redirection of current practice of stem cell therapy and a reset using more detailed analysis of stem cell function interfering with disease mechanisms. To accelerate scientific development the authors suggest intensifying unified computational data analysis and shared data knowledge by using open-access data platforms.

Contents

Abbreviations: ACC, American College of Cardiology; acLDL, acetylated-low density lipoprotein; AHA, American Heart Association; ALL, acute lymphoblastic leukemia; ANG1, angiopoietin 1; ATMP, Advanced Therapy Medicinal Product; BM, bone marrow; BMSC, bone marrow-derived stem cell; B-reg, regulatory B-cells; CABG, coronary artery bypass graft; CDC, cardiosphere-derived cell; CCL2, C-C motif chemokine ligand 2; CCTRN, Cardiovascular Cell Therapy Research Network; CPC, cardiac progenitor cell; CSC, cardiac stem cell; CEC, circulating endothelial cell; CFU, colony-forming unit; CSF-1, colony-stimulating factor 1; CVD, cardiovascular disease; CXCL12, C-X-C motif chemokine 12; CXCR4, C-X-C chemokine receptor 4; EMA, European Medicines Agency; EPC, endothelial progenitor cell; EPO, erythropoietin; ESC, embryonic stem cell; FAIR, findable, accessible, interoperable and reusable; FDA, US Food and Drug Administration; GCP, Good Clinical Practice; GLP, Good Laboratory Practice; GMP, Good Manufacturing Practice; G-CSF, granulocyte-colony stimulating factor; GSP, Good Scientific Practice; GVP, Good Vigilance Practice; GWAS, genome-wide association studies; GxP, good practice; HGF, hepatocyte growth factor; HIF-1 alpha, hypoxia-inducible factor 1 alpha; HLA-G5, human leukocyte antigen class I molecule G5; HSC, hematopoietic stem cell; IDO, indoleamine 2,3-dioxygenase; IFN, interferon; IGF, insulin-like growth factor; IGFBP, insulin-like growth factor-binding protein; IgM, immunoglobulin M; IL, interleukin; iPSC, induced pluripotent stem cell; IVUS, intravascular ultrasound; lncRNAs, long non-coding RNAs; IP-10/CXCL10, interferon gamma-induced protein 10/C-X-C motif chemokine 10; LVEF, left ventricular ejection fraction; LVESV, left ventricular end-systolic volume; MI, myocardial infarction; ML, machine learning; MACE, major adverse cardiac events; MMP, matrix metalloproteinase; MNC, mononuclear cell; MRI, magnetic resonance imaging; MSC, mesenchymal stem cell; mTOR, mechanistic target of rapamycin kinase; mTORC1, mechanistic target of rapamycin kinase complex 1; NGS, Next Generation Sequencing; NK cells, natural killer cells; NO, nitric oxide; NOS, nitric oxide synthase; NR, non-responder; OCTGT, Office of Cellular, Tissue and Gene Therapeutics; PB, peripheral blood; PEI, Paul-Ehrlich Institute; PGE2, prostaglandin E2; QC, quality control; R, responder; RCV, retrograde coronary sinus; ROS, reactive oxygen species; SAE, serious adverse event; SC, stem cell; SCF, stem cell factor; SDF-1 alpha, stromal cell-derived factor 1 alpha; SH2, Src homology region 2; SH2B3/LNK, SH2B adapter protein 3/lymphocyte adaptor protein; SNP, single nucleotide polymorphism; TBX5, T-box transcription factor 5; TGF-β1, transforming growth factor beta 1; TLR, toll like receptor; TNF alpha, tumor necrosis factor alpha; T-reg, regulatory T-cells; t-SNE, t-distributed stochastic neighboring embedding; VE-cadherin, vascular endothelial cadherin; VEGF, vascular endothelial growth factor; VEGFR, vascular endothelial growth factor receptor; WGCNA, Weighted Gene Co-expression Network Analysis.

☆ This review is part of the *Advanced Drug Delivery Reviews* theme issue on "Advances in Stem Cell-Based Therapies".

* Corresponding author.
E-mail addresses: gustav.steinhoff@med.uni-rostock.de (G. Steinhoff), iuliia.nesteruk@med.uni-rostock.de (J. Nesteruk), markus.wolfien@uni-rostock.de (M. Wolfien), jana.grosse@med.uni-rostock.de (J. Große), ulrike.ruch@med.uni-rostock.de (U. Ruch), praveen.vasudevan@med.uni-rostock.de (P. Vasudevan), paula.mueller@med.uni-rostock.de (P. Müller).

https://doi.org/10.1016/j.addr.2017.10.007

1. Introduction

This review addresses the current state of development of cardiac stem cell (SC) therapy. Although the first clinical cardiac SC phase I trials having started in 2001 [1] still licensed and standardized therapies are not realized after 16 years of clinical development. This requires a critical analysis of scientific evidence for diagnosis, genetic control, clinical indication, and treatment approach. The review comprises knowledge on SC function for cardiac homeostasis and repair as well as strategies for treatment and diagnostic development deducted from the recently published results of the phase III PERFECT trial [2]. The five main topics in this review listed in Table 1 are focusing on (1) SC function and delivery, (2) disease pathomechanism, treatment options and biomarkers, (3) standardization and quality control (QC) for best practice, (4) data analysis and (5) comprehensive treatment centers.

2. Stem cell function and delivery in heart disease

2.1. Function

The evolving detection of SC functions in tissue development, regeneration and repair have advanced medical sciences in the last two decades leading to the understanding of tissue development and regeneration during aging as well as repair (Fig. 1). The observation that an injury in different organs, such as muscle, liver, and brain, triggers bone marrow (BM)-derived cells to the area of damage where they contribute to regenerative processes has provided the basis for bone marrow-derived stem cell (BMSC) therapy [3–6]. In human heart transplants, the immigration and differentiation of recipient myofibroblast was observed in 1989 as well as different kinetics of cell replacement [7,8]. In animal models of myocardial infarction (MI), intramyocardial injections of BMSCs preserved left ventricular contractile function and reduced fibrosis formation [9,10].

Hematopoeitic stem cells (HSCs) have been well characterized by membrane markers like $CD133^{+}$ as well as distinct progenitor subtypes differentiating into blood and immune lineages [11]. In 1997, **endothelial progenitor cells (EPCs)** were first described by Asahara et al. [12] as a main component of cardiovascular and tissue regeneration [13,14]. Cardiovascular lineage EPCs can be generally identified by their capability to express endothelial phenotypical markers like CD133, CD34, CD117, CD184, vascular endothelial growth factor receptor 2 (VEGFR2, KDR, Flk1), and vascular endothelial cadherin (VE-cadherin). They also possess some endothelial cell functional characteristics in vitro and in vivo, like acetylated-low density lipoprotein (acLDL) uptake and the

Table 1
Main topics of the review.

1. Stem cell function and delivery in heart disease
2. Disease pathomechanism, therapeutic options, diagnostic biomarkers
3. Definition of quality standard and best practice
4. International standard of data analysis
5. Comprehensive centers for R&D integrated disease treatment

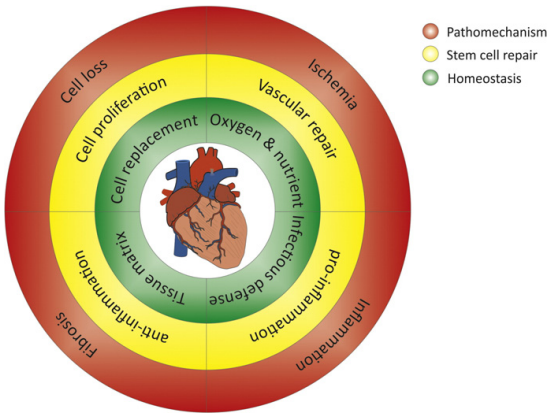


Fig. 1. Stem cell functions adapting to tissue homeostasis, repair, and dysfunction.

formation of endothelial colony-forming units (CFUs). Under steady state conditions, the concentration of $CD34^+$ EPCs in peripheral blood (PB) is much lower than in BM [15]. Interestingly, it was shown in clinical studies that the number of circulating EPCs significantly increased during the early phase of acute MI, suggesting that these cells may contribute to healing processes [16–19].

Mesenchymal stem cells (MSCs) are part of the BMSC pool [20] as well as a basic component for perivascular tissue throughout all tissues and organs [21]. MSCs are characterized by fibroblast clonal potency and multicellular differentiation into osteocytes, adipocytes and chondrocytes. Furthermore, their role for induction and down-regulation of T-lymphocyte response has been observed in graft-versus-host disease [22]. Intramyocardial or intravenous application of MSCs has been reported to reduce post MI damage in animal studies and clinical trials [23,24].

Cardiac stem cells (CSC) were described for the first time in 2003 by Beltrami et al. with the presence of self-renewing c-kit^+ cells in the adult heart, which differentiated into cardiomyocytes, smooth muscle cells as well as endothelial cells and regenerated functional myocardium [25–27]. To date, various additional populations of putative endogenous cardiac stem and progenitor cells have been identified in the heart, including Isl-1^+ cells [28–30], Sca-1^+ cells [31–33], cardiosphere-derived cells (CDCs) [34–37] and cardiac side population cells [38–40]. Recent evidence from genetic fate mapping studies further confirmed that in addition to proliferating pre-existing cardiomyocytes also differentiating resident cardiac stem and progenitor cells can contribute to post-natal cardiomyogenesis [41–43].

Embryonic stem cells (ESCs) have the ability to differentiate into derivatives of all three germ layers [44]. Several protocols have been successfully developed to induce cardiomyocytes from ESCs in vitro [45]. Although the generation of fully mature cardiomyocytes in large yields and with high purity is still unfeasible [46], these studies demonstrated a strong cardiogenic potential of ESCs. However, clinical translation of these cells has been hampered by significant obstacles, including ethical concerns [47], the risk of immune rejection [48,49], genetic instability [50,51] and tumorigenic potential [44]. In fact, some early studies hypothesize that the cardiac environment is sufficient to induce the differentiation of ESCs into cardiomyocytes [52,53]. Nevertheless, this suggestion has been refuted, since the formation of teratomas was detected after intramyocardial injection of undifferentiated ESC [48,54,55]. Consequently, more attention has been given to the identification, generation and purification of ESC-derived cardiac progenitor cells. Numerous preclinical studies demonstrated stable electromechanical integration of ESC-derived cardiomyocytes into the host myocardium leading to reduced scar size, decreased cardiac remodeling and improved cardiac function without teratoma formation in small [54,56–61] and large [55,62,63] animal models. These encouraging data have paved the way for the first phase I clinical trial by Menasché in 2013 (ESCORT) using human ESC-derived cardiac progenitors (*Isl-1*$^+$ and SSEA-1^+) embedded in a fibrin scaffold in six patients with severely impaired cardiac function scheduled for CABG. Although initial results from the first patient are promising [64], it is still too early to assess the safety as well as the therapeutic benefit of these cells in humans.

Induced pluripotent stem cells (iPSCs) are pluripotent SCs generated directly from somatic cells through a reprogramming process. In 2006, Takahashi and Yamanaka published for the first time the generation of iPSCs from mouse fibroblasts by retroviral transduction of four different transcription factors (Oct3/4, Sox2, c-Myc, and Klf4) [65]. One year later, the same group could transfer this technology to human fibroblasts [66]. It was shown that generated iPSCs not only express ESC markers and exhibit morphology, proliferation as well as tumorigenic properties similar to ESCs, but also bear comparable

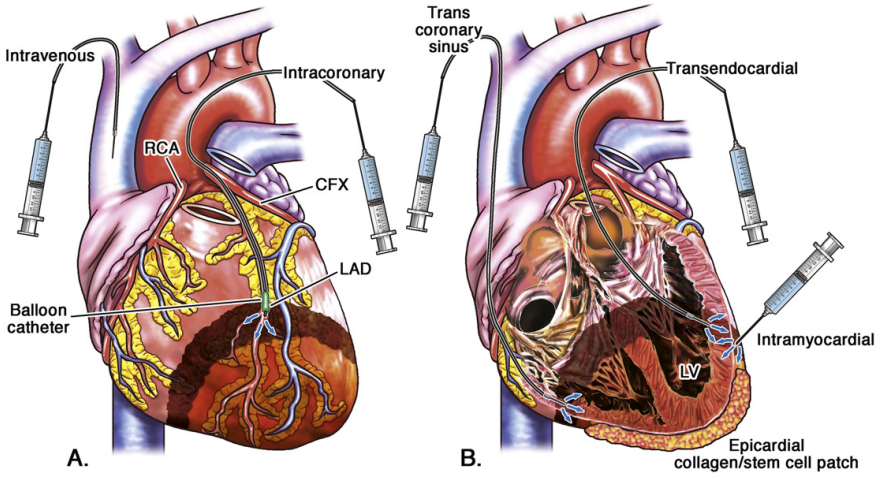


Fig. 2. Intravascular (A) and intramyocardial (B) delivery techniques of stem cells for cardiac disease treatment (reprint from: [104]).

cardiogenic potential [65,67]. Although terminal differentiation of iPSCs into fully mature cardiomyocytes in vitro is still an unreached goal [68–72], functional cardiomyocytes have been generated from both mouse [73–75] and human [76–79] iPSCs. Moreover, promising in vivo experiments in MI models showed the engraftment as well as improved cardiac performance, reduced infarction size and attenuated cardiac remodeling after iPSC-derived cardiomyocyte transplantation [80–86]. Importantly, the iPSC-generation technology enables the creation of patient-specific pluripotent stem cells [87] that can be used for genetic repair. However, first enthusiasm has been tempered by the investigation that syngeneic mouse iPSCs were immunogenic and rejected following transplantation [88,89]. This observation was contrary to the initial suggestion that autologous iPSCs would not cause immune response in the host due to the so-called "self-tolerance" [90]. Since then, significant genomic instabilities in iPSC lines including epigenetic memory, aberrant methylation patterns and mutations have been reported resulting from variations in parenteral somatic cells or occurring during the reprogramming process and culturing time [91–95]. Recently, new protocols have been developed improving the efficiency and safety for the generating of iPSCs by using virus-free and non-integrative approaches e.g. proteins [96,97], mRNAs [98], microRNAs [99,100], and small molecules [101–103]. Nevertheless, iPSC-derived cardiomyocytes did not reach clinical grade, yet.

2.2. Delivery techniques

Delivery techniques could be conventionally divided into three categories: intracoronary delivery, intramyocardial application (including endoventricular, transepicardial and transvascular injections) and intravenous delivery (including retrograde coronary sinus (RCV) and peripheral intravenous infusions). The choice of techniques is not regulated and depends on preferences of SC centers, their equipment opportunities and experience. The delivery routes and techniques are summarized in Fig. 2 (reprinted from: [104]).

2.3. Intracoronary application

Selective intracoronary delivery route was used with the intention to minimize shedding to non-targeted organs [105]. To facilitate transendothelial passage and migration into the infarcted zone, an angioplasty balloon was inflated in the proximal segment of coronary arteries and cells were infused in the culprit vessel (stop-flow technique). The transmigration process is facilitated in injured and ischemic viable tissue [106], as local myocardial ischemia is a potent stimulus for chemokinesis of SCs due to stromal cell-derived factor 1 (SDF-1) and C-X-C chemokine receptor 4 (CXCR4) signaling [107–109]. Therefore, the ischemia-producing stimulus may actually promote the cell adhesion and extravasation into the myocardial tissue [110,111]. Delivery of SCs to the injured or failing myocardium by a simple intracoronary administration seemed to be sufficient to promote myocardial repair [112,113]. The intracoronary approach, however, should be reserved for the smaller, mononuclear BMSCs, since intracoronary application of cultured cell types, like MSCs or skeletal myoblasts, was associated with microembolisation and significant microvascular obstruction [114].

2.4. Intramyocardial application

2.4.1. Endoventricular intramyocardial injection

The intramyocardial injection of SCs results in the direct delivery into the myocardial target area, without dependence on vascular access or sufficient cell migration across the endothelial barrier. Percutaneous endoventricular injection of cell preparations are generally guided using routine fluoroscopic ventriculography or by electromechanical mapping using the NogaStar mapping catheter (Biologics Delivery Systems, Diamond Bar, CA). The percutaneous transendocardial delivery catheters Helix® and C-Cath® use direct fluoroscopic imaging [115–117]. The myocardial cell retention of these catheters typically varies between the 20–35%. The Myostar® injection catheter is interfaced with the NOGA© 3-dimensional electromechanical endoventricular mapping system. Based on the structural reconstruction of unipolar voltage mapping in combination with the mapping of linear local shortening, the operator can distinguish in real time viable myocardium from non-viable, hibernating or scar tissue. Clinical studies have established safety and feasibility of the transendocardial intramyocardial injection in the setting of chronic heart failure [118], refractory angina [119], as well as in subacute MI [120]. However, cell loss and shedding to non-targeted organs is still considerable for all percutaneous injection catheters.

2.4.2. Transepicardial application

SC implantation during open heart surgery is generally performed into well exposed epicardial ischemic areas allowing for multiple injections within, and principally, around the infarct area with a thin needle [120,121]. This procedure is limited to certain areas of the left ventricle, and cannot be used for the septal myocardial segments. Moreover, the therapeutic effect of intramyocardial SC therapy is difficult to be unequivocally interpreted when performed together with a revascularization procedure. Therefore, recent reports about surgical "stand alone" SC therapy are of particular interest [122,123].

Epicardial delivery of bioengineered composite sheets harboring SCs is a new method of SC delivery. SC sheets adhere to the epicardial surface spontaneously, or as collagen-based patches [124,125]. Adipose-derived stromal cell sheets resulted in significant improved survival and left ventricular remodeling in an infarct model in rats compared to intramyocardial injections [126].

2.4.3. Transvascular delivery

The transvascular delivery of SCs occurs via transvenous or transarterial delivery under intravascular ultrasound (IVUS) imaging. The success transvenous intramyocardial delivery of a cell-hydrogel by the TransAccess® delivery catheter (Medtronic Vascular, Santa Rosa, CA, USA) was described [127,128]. The Mercator Cricket® and Bullfrog® perpendicular-positioned microneedles (Mercator MedSystems, Emeryville, CA) penetrate the coronary artery wall and allows the user to inject SCs directly into the perivascular space (tunica adventitia of the coronary artery) [129–131].

2.5. Intravenous delivery

2.5.1. Retrograde coronary sinus infusion

RCV infusion is performed by femoral venous access and the positioning of a conventional angioplasty balloon into the mid portion of the coronary sinus followed by SC infusion. Clinical and preclinical studies demonstrated safety, efficacy and high cell retention of delivering via the coronary sinus [132–134]. Coronary sinus delivery could be recommended in cases of severe aortic stenosis, severe peripheral artery disease or intraventricular thrombus formations which precludes percutaneous endoventricular injection. The method is also possible for patients with a resynchronization device. Potential complications are coronary sinus rupture and embolisation [135,136].

2.5.2. Peripheral intravenous infusion

Intravenous infusion is the most safe and cost-effective SC delivery method. Its safety and feasibility has been shown in swine model of MI [137] and later in a phase I clinical study after delivery of allogeneic MSCs [138]. The study resulted in a significant improvement of the ejection fraction in the treated group versus placebo at 12 month follow-up, although the myocardial retention following intravenous injections is mere 0.5% [139].

3. Disease pathomechanism, therapeutic options, diagnostic biomarkers

3.1. Disease pathomechanism

3.1.1. Ischemia

The contractile workload of heart tissue is dependent on sufficient oxygen supply to maintain homeostasis of viability and function. The aerobic metabolism of the heart at rest needs about 8–15 mL O_2/min/100 g tissue – more than double the consumption of brain tissue with 3 mL O_2/min/100 g tissue [140]. With exercise and increased hypertensive workload this need for oxygen can increase up to >70 mL O_2/min/100 g. Unrestricted blood supply and oxygen delivery from capillaries to cardiomyocyte cell membrane are crucial factors for maintaining sufficient pump function. The surveillance relies mainly on the capillary density surrounding and in contact to cardiomyocyte fibrils [140]. Oxygen shortage leads to activation of hypoxia-inducible factor 1 alpha (HIF-1 alpha) by cardiomyocytes, which induces local vascular endothelial growth factor (VEGF) release to the circulation by endothelial cells as well as induced endothelial expression of VEGFR2 (CD309) and induce SDF1 alpha as a ligand for circulating EPCs. Circulating angiogenetic capacity of EPCs is enhanced by combined release of VEGF and erythropoeitin (EPO) by endothelial cells, pericytes, and MSCs in tissue. By this pathway the hypoxic state of every single cardiomyocyte can activate local angiogenesis of capillary sprouting on demand [141]. This mechanism is dependent on the co-activation of neighboring endothelial cells, pericytes, macrophages, and MSCs for systemic proliferation/release of EPCs and circulating endothelial cell (CECs) as well as expression of local homing receptors for targeted angiogenesis sprouting.

3.1.2. Cardiomyocyte crosstalk with the regulatory BMSC niche for vascularization

Tissue hypoxia mediators are released by ischemic cardiomyocytes and are directly activating local endothelial cells and BM-derived EPCs. By this mechanism, they induce blood driven angiogenesis at the ischemic endothelial spot by targeting circulating EPCs and VEGF, SDF-1 or Vitronectin receptors to the endothelial surface [142]. For the local angiogenesis induction, at least local endothelial expression of VEGFR2 is required for the induction of sprouting from blood vessels by local adherent EPCs [143,144]. In the BMSC niche, angiogenic chemokine stimulation enhances EPC or HSC proliferation and release, but precise stimulatory pathways have to be unraveled [142,145]. EPC proliferation is influenced by insulin-like growth factor (IGF)-1 and VEGF [146]. EPO production is attributed to kidney tissue interstitial cells, most likely pericytes [147]. However, also pericytes in brain and liver are producing EPO upon local ischemic activation by HIF-1 alpha [143]. It is conceivable that pericytes in other cardiovascular tissues and heart may be inducible to produce EPO upon hypoxia.

Progenitor cell trafficking is regulated by hypoxic gradients through HIF-1 alpha induction of C-X-C motif chemokine 12 (CXCL12)/SDF-1 [148]. Moreover, in vitro hypoxia induced HIF-1 alpha induces mTOR complex 1 (mTORC1) axis in human umbilical cord blood MSCs leading to cell-cycle and F-actin modulation with increased proliferation and migration [149]. The mTOR pathway leading to proliferation is also induced by hypoxia mediator in BM niche cells [150]. This local process involves tissue macrophages capable to influence the cell cycle of vascular endothelial cells through the paracrine wnt ligand WNT7b [151] and in conjunction with pericytes, which secrete angiopoietin 1 (ANG1) to regulate angiogenesis leading to AKT-mTOR activation [152]. At the same time, monocyte derived dendritic cells seem to regulate angiogenesis during development through the production of VEGF and transforming growth factor beta 1 (TGF-β1) [153]. VEGF recruited numerous BM-derived macrophages to the tissue through signaling by CXCR4, the ligand being expressed by pericytes [154]. A similar mechanism involving CXCL12 was shown to guide these BM-derived progenitor cells to sites of vascular expansion in the embryo. In addition, CXCR4 and CXCL12/SDF-1 were up-regulated in response to hypoxia in ischemic tissue, which resulted in the recruitment of similar myeloid cells [148]. In conclusion, hypoxic cardiomyocyte are able to induce a local and systemic SC proliferation response for adaptive capillary angiogenesis.

3.1.3. Early and late inflammation

The cell-cell interactions between immune cells, such as macrophages and T-cells, with $CD34^+$ SCs and MSCs are critically important for the development of the further inflammatory repair process. Enhancing factors for repair are IGF-1 and activin. The shortage of oxygen supply to the cardiac tissue leads to necrosis, apoptosis [155] and release of various factors like interleukin (IL)-1, reactive oxygen species (ROS), nitric oxide (NO) and immunoglobulin M (IgM) by the stressed cells. Toll like receptors (TLRs) are of essential importance in the activation of the immune response after MI [156]. The primary response to these paracrine signals is mounted by the cardiac resident macrophages. They have different functions and roles in comparison to the peripheral BM-derived macrophages [157] whose development is regulated by colony-stimulating factor 1 (CSF-1) [158]. These activated tissue resident macrophages secrete high levels of pro-inflammatory molecules like IL-1, IL-6, IL-8, NO and TNF alpha [159,160]. TLR3 has been identified as a polarizing signaling effect on pro- or anti-inflammatory activation of MSCs and immune cells. Down-regulation mechanisms of MSCs and macrophages at the site of injury can aid cardiac repair quality [161]. In this context the role of TLRs as modulator of MSC expression remains remarkable and may need further investigation [162].

In recent years, immune modulation into pro- or anti-inflammatory states have received major attention for interventional immune therapy. While the mentioned early inflammatory process is mediated mostly by components of the innate immune system, the chronic inflammatory phase is effected mainly by the adaptive immune system. The lymphocytes are activated by antigen presenting cells [163] and lead to destruction of cells bearing these antigens by cytotoxic T-cells while B-cells amplify this response by producing antibodies. However, an important component of the whole inflammatory process and its successful resolution depends on a particular subset of T-helper cells called regulatory T-cells (T-reg) and regulatory B-cells (B-reg) [164–166].

3.1.4. Inflammatory cells and SC interaction

The early inflammatory phase that occurs after the onset of ischemia and mediated by the host immune cells and SCs is mostly complete by the time SCs or cardiomyocytes are transplanted in a clinical setting. Therefore, it is important to understand the interaction of these cells with the host immune and SC pools in the ischemic heart at certain/different time points to improve the efficacy and timing of cell therapy as well as their ability to replace the cardiomyocytes lost during ischemia. This would help in deciding the best outcome expected from these cells and their appropriate interactions with the host repair response. It could also help in selecting suitable response signatures to a particular cell type. The ideal scenario where a particular cell type induces cardiomyocyte proliferation would necessitate for these cells to eventually resolve the inflammatory process and not induce further long-term secretion of anti-inflammatory molecules that could increase scar formation.

Currently, there are no studies available on the interaction of transplanted HSCs, EPCs or cardiomyocytes with the inflammatory pathway under ischemic conditions. This is a fundamental gap in SC therapy studies considering the importance of inflammation on the overall outcome. However, the relationship between these SCs and the inflammatory components is not exactly known. HSC proliferation and differentiation is regulated by pro-inflammatory cytokines such as TNF, IL-1, IL-6, IL-8 and interferons (IFNs) [167]. Long-term stimulation of TLR4 leads to impaired long-term self-renewal of HSCs [168]. Also, BM-resident macrophages govern the retention of HSCs within the BM [169]. It would be interesting to speculate the proliferation and

retention of HSCs in the cardiac tissue where the macrophages are recruited. The mechanism of interaction of EPCs directly with the immune cells is not yet clear, but, EPCs have been shown to secrete endothelial nitric oxide synthase (NOS), inducible NOS, VEGF-A, SDF-1, IGF-1 and hepatocyte growth factor (HGF) under ischemic conditions [170]. NOS has been shown to be a versatile player in the immune system, being able to alter the functions of macrophages, T-cells, eosinophils and neutrophils [171]. SDF-1 has been shown to play a role in the recruitment of lymphocytes, monocytes and driving macrophage differentiation [172,173].

The immunomodulatory functions of MSCs have thus far been more broadly investigated than other SCs. MSCs have been shown to inhibit T-cell proliferation by the release of TGF-β [174]. This activity is however mediated by prostaglandin E2 (PGE2) [175] and dependent on levels of inducible NOS in mice and indoleamine 2,3-dioxygenase (IDO), and soluble human leukocyte antigen class I molecule G5 (HLA-G5) in humans [176,177]. MSCs modulate B-cells [178] through soluble factors like C-C motif chemokine ligand 2 (CCL2) [179]. Similar to T-cell inhibition, inflammatory stimulation of MSCs is required for their B-cell inhibitory activity [180]. MSCs also modify dendritic cells with the addition of PGE2 [181] and IL-6 [182]. They have been also shown to inhibit natural killer cells (NK cells), but only at high MSC/NK cell ratios [183] and in the presence of IDO or PGE2 [184]. Recently, TLR3 or TLR4 pathways were identified to induce MSC suppression and T-reg induction [185]. TLR3 displays a protective role in mouse models of atherosclerosis [186], and activation of TLR3 signaling is associated with ischemic preconditioning-induced protection against brain ischemia and attenuation of reactive astrogliosis [187,188]. In addition, TLR3 activators show effects on human vascular cells [186]. Unlike their inhibitory effects on the other immune cells, BM-derived MSCs shifted the macrophages to a more anti-inflammatory phenotype through the release of PGE2 and cell-contact mediated signaling [189].

3.1.5. Fibrosis

The failure of the terminally differentiated cardiac myocytes to proliferate for repair of the infarct myocardium leads to impaired wound healing and ultimately, the formation of a scar. The initial inflammatory phase as described above involving cells such as monocytes and macrophages are responsible for the migration of local fibroblasts and their subsequent conversion to myofibroblasts in order to stabilize the heart after infarction. The origin of these cells is however contentious since both cardiac MSCs [190] and circulating monocytes [191] have been shown to contribute to fibrosis. The clinically relevant problem is the formation of a mature scar where type III collagen is replaced by type I collagen due to the secretion of various matrix metalloproteinases (MMPs). This can also have major effects over time since distal regions of the heart undergo gradual fibrosis leading to increased global cardiac function deterioration and arrhythmogenesis [192]. Therefore, SC therapies should also be determined based on their ability to replace these myofibroblasts, thereby resolving the scar. However, this capability is often deemphasized when choosing a particular cell type for transplantation over other competencies like angiogenic potential.

3.1.6. Cell renewal

Starting with initial transplantation studies of HSCs into cardiac tissue in 2001 and 2002, BMSCs were thought to contribute to the functional recovery of damaged myocardial tissue by coupling electromechanically with the recipient myocardium after acquiring a cardiomyogenic fate [9,193]. However, the trans-differentiation of these cells into cardiomyocytes within the heart tissue remains inconclusive up to date, with studies both supporting [194–197] and refuting [198–200] this notion. Others indicated fusion of BM-derived SCs with endogenous cardiomyocytes as a predominant mechanism for the transformation of BM-derived SCs into cardiomyocytes [201–203]. Similarly, both of these potential mechanisms have been attributed to the generation of cardiomyocytes from PB- and adipose tissue-derived stem and progenitor cells [204–207]. Interestingly, it was demonstrated that adult cardiomyocytes re-enter the cell cycle after fusion with hematopoietic and mesenchymal stem and progenitor cells in vitro and in vivo [208,209]. However, the detected fractions of grafted cells converting along the cardiac lineage were very small [201,210,211]. The main functions of tissue immigrating or transplanted HSCs, which are $CD117^+$ (c-kit or SCF-receptor) and/or $CD34^+$, as well as resident or migrating MSC, which are $CD105^+$ (endoglin) and/or $CD271^+$, still have to be defined depending on tissue specific homeostasis and disease mechanism. Altogether, adult MSCs/CSCs and HSCs may potentially be transformed into cardiomyocytes through (trans)differentiation and/or fusion with pre-existing cardiomyocytes. This, however, can be considered only as a rare event in cardiac tissue regeneration response. This is reflected in low survival and engraftment [34,211–214] as well as limited generation of cardiomyocytes from injected SCs implying that direct re-muscularization has only a limited contribution to beneficial SC effects.

Conflicting results have been obtained with respect to the cardiomyogenic differentiation potential of cardiac stem and progenitor cells. On the one hand, in vitro studies observed the expression of cardiac markers and structural proteins several days after cell cultivation under certain conditions [25,26,37,38,40] and in co-culture with neonatal or adult cardiomyocytes [28,32,38,39,215]. In line with these findings, various in vivo experiments demonstrated that transplanted cardiac stem and progenitor cells give rise to newly generated cardiomyocyte-like cells through direct differentiation [25–27,216–219]. On the other hand, it was shown that fusion of injected cells with pre-existing cardiomyocytes contributes equally to the generation of cardiomyocytes from injected cells [33]. Moreover, the maturity of engrafted cardiac stem and progenitor cells remains highly controversial, with studies reporting the contractile function [25] and the full maturation of newly-formed cardiomyocytes within two months [27], whereas others did not observe the phenotype of mature cardiomyocytes one year after cell transplantation [212]. Likewise, quantified fractions of engrafted cells acquiring a cardiomyogenic lineage vary considerably from modest to substantial [32,40,212,220]. Some lineage tracing studies also refuted the significant myocardial potential of putative resident cardiac progenitor cells after injury [221–223], while others indicated their intrinsic regenerative capacity by replacing lost cardiomyocytes [27]. Only limited cardiomyocyte turnover occurs in vivo [224,225]. In a meta-analysis, including 80 animal studies, cardiac derived precursor cell therapy was shown to significantly improve LVEF by 10.7% compared with placebo controls [226]. This was not different from results of extracardiac precursor and SCs [227]. Interestingly, CSCs had a significantly greater beneficial effect in small animal models compared with large animal models (~12% vs. 5% improved LVEF), while cell source, comorbidities, use of immunosuppression and disease models did not influence the effects on cardiac function. In 2011, Bolli et al. published the first report of cardiac SC therapy in humans [228]. In this phase I clinical study (SCIPIO) no mortality or CSC-related adverse events were observed following the intracoronary infusion of autologous $c\text{-}kit^+$ cardiac SC in patients with ischemic cardiomyopathy. In addition, cardiac magnetic resonance results in an improved global as well as regional left ventricular function, a reduced infarct size as well as an increase in viable tissue four and twelve months after SC application [229]. In 2012, the CADUCEUS trial demonstrated the safety and feasibility of intracoronary infused CDCs grown from endomyocardial biopsy specimens in patients with left ventricular dysfunction after MI [144]. After CSC transplantation, analyses demonstrated significant reduction in the size of the infarct as well as an increase in the amount of viable myocardium, regional contractility and regional systolic wall thickening compared with controls, whereas left ventricular function and volumes did not differ between groups [144,230]. However, as these early phase I clinical studies should show the safety of the therapy they were not powered to determine the efficacy of CSCs.

3.2. Therapeutic options

3.2.1. Enhancing the circulating EPC pool

BM-derived mononuclear cells (MNCs) can be isolated from BM aspirates through density gradient centrifugation. Notably, the overall composition of BM-derived MNCs is primarily that of predominantly differentiated blood cells with a low percentage of early committed cells at various maturation stages, with only little amounts comprising HSCs, EPCs and MSCs [231]. In 2001, Strauer et al. demonstrated for the first time that intracoronary application of autologous BM-derived MNCs is feasible under clinical conditions and results in modified cardiac tissue response (e.g. scar regeneration) after MI [232]. In the first controlled study, the application of BM-MNCs in patients with acute MI significantly improved local contractility and perfusion, reduced left ventricular end-systolic volume (LVESV), and decreased infarction site compared with the standard therapy group [105]. Furthermore, the authors postulated that therapeutic effects were associated with myocardial regeneration and neovascularization. Since then, large numbers of clinical trials investigated the potential of BM-derived MNCs for the treatment of ischemic and non-ischemic heart diseases. However, therapeutic benefits continued to remain controversial. In fact, several randomized, controlled studies demonstrated a significantly improved cardiac function after intravenous or intracoronary BM-derived MNC transplantation [233–241], while others could not prove cell-based benefits [242–251]. This heterogeneity is also reflected by recent meta-analyses revealing not only an overall mild (2–5%) improvement of the global heart performance and a possible attenuation of adverse cardiac remodeling [252–256], but also failed to detected therapeutic effects of BM-derived MNC on left ventricular function [257].

To date, similar findings have been made in trials employing the systemic administration of granulocyte-colony stimulating factor (G-CSF) used to stimulate the mobilization of stem and progenitor cells from BM [258]. However, meta-analyses demonstrated that the G-CSF treatment as a stand-alone therapy has no beneficial effects on myocardial regeneration [259–261]. The intramyocardial and intracoronary transplantation of G-CSF mobilized HSCs/EPCs also yielded controversial results. In particular, various clinical studies demonstrated the safety and an improved cardiac performance following the injection of PB-derived MNCs [262–265] and $CD34^+$ cells [119,266–274], while some failed to detect additional reliable and significant therapeutic effects compared to control groups [248,275,276].

Notwithstanding, the feasibility and safety of intravascular BM-derived MNC transfusion was demonstrated in all of the clinical trials while disease treatment efficacy has not been proven in pivotal phase III trials. In fact, several randomized, controlled phase II studies demonstrated a significantly improved cardiac function after intramyocardial or intracoronary BM-derived MNC transplantation [242–249], while others could not prove cell-based benefits [250–259].

Based on these inconclusive clinical results, the strategy may have to be reset to understand the underlying disease mechanism. This can be seen in the recently intensifying interest in clinical research programs on the mechanism of action of angiogenetic or immune response modulation by the US - Cardiovascular Cell Therapy Research Network (CCTRN) [277] and the TACTICS EU-network [278].

3.2.2. Intramyocardial application

The initial observation in the first-in-man phase I trial for intramyocardial transplantation of purified $CD133^+$ BMSCs performed by our group in Rostock in June 2001 revealed promising results with induction of cardiac regeneration by >10% left ventricular ejection fraction (LVEF) increase [1]. The phase II trial confirmed the finding in BMSC treated patients vs. coronary artery bypass graft (CABG) controls [121]. Similar findings were reported by Patel et al. [279]. However, in placebo controlled trials the BMSC induced improvement in heart function was not different from placebo controls [280]. Recently, in the randomized double blinded placebo controlled multicenter phase III trial – PERFECT; an improvement in the heart function (LVEF >5%) was observed in 60% of placebo as well as SC treated patients [2]. Interestingly, the large gain in LVEF was not different in the SC and placebo CABG treatment groups: ΔLVEF (Placebo +8.8% vs. $CD133^+$ BMSCs +10·4%, Δ$CD133^+$ vs. placebo +2.58%, $p = 0.414$). In fact, the LVEF increase in placebo-treated patients undergoing BM harvest was remarkable (+8.8% vs. +3.5%) in comparison with earlier results [121]. This finding of 60% LVEF improvement (categorized as 'responders') was also reported in CABG surgery for patients with reduced pump function [281]. In the clinical setup of chronic ischemic heart failure of the PERFECT trial, 40% of patients were 'non-responders' to induction of cardiac regeneration irrespective their treatment with placebo or $CD133^+$ BMSCs. Induction of circulating EPCs as well as angiogenesis and heart function improvement was significantly reduced in non-responders [2]. Cardiac function improvement by purified intramyocardial BMSCs was effective only in patients with angiogenesis response by circulating EPCs [2].

3.3. Diagnostic biomarkers

3.3.1. Monitoring of angiogenesis response in the PERFECT trial

The interest in diagnostic use of angiogenesis factors and cytokines in blood is rising, since different levels of biomarkers across a spectrum of pathophysiological processes of different diseases were revealed. Monitoring of biomarker concentrations in blood can not only provide the clinician information about the diagnosis, but can improve prognostication and treatment strategies at the same time [282,283].

The advantage of the PERFECT phase III trial was a study design that included blood harvesting and storage before, 24 and 72 h as well as 10 days after CABG with or without SC application [2]. Resulting from this, the dynamics of 13 angiogenesis factors and cytokines (VEGF, stem cell factor (SCF), SDF-1, IGF-1, insulin-like growth factor-binding protein (IGFBP)-2 and -3, interferon gamma-induced protein 10 (IP-10), tumor necrosis factor alpha (TNF alpha), IL-6,8,10, EPO, vitronectin) were obtained in time. The study revealed that responder and non-responder but not $CD133^+$ SC and control groups differed in angiogenesis factors before operation. Responders were defined as having a ΔLVEF at 6 months versus baseline higher than 5%. It was shown, that non-responders display basically elevated angiogenesis stimulating factors as VEGF and EPO, as well as pro-inflammatory factor IP-10 and decreased level of IGFBP-3 accompanied by reduced amount of EPCs in PB. The VEGF level did not change in time, in contrast, a low baseline level of

Table 2
Functions of SH2B3 in hematopoietic, vascular and interstitial cells.

	References
Cellular pathway downregulation	
Cytokine pathways: IL-7, IL-11	[304,369]
Growth factors in HSC and MSC: EpoR, SCR, Jagged1	[370]
Integrin and actin signaling	[306]
VCAM-1	[371]
PDGF Rec	[372]
Inhibition of cellular function	
Stem cell proliferation (HSC, EPC, MSC?)	[373,374]
Lymphocyte proliferation (B-lymphocytes)	[309]
Endothelial activation	[306]
Blood MNC and thrombocyte proliferation	[303,375]
Systemic inhibitory effect	
Blood, immune and cardiovascular proliferation	[315]
Angiogenesis and vascular repair	[307,314,376]
Endothelial activation	[298,316]
Immune cell	

VEGF in the responder group doubled in 10 days after CABG procedure [2].

3.3.2. Ischemia and angiogenesis - a failure of stem cells?

The non-responders in the PERFECT trial show typically lowered $CD133^+/CD34^+/CD117^+$ EPC and thrombocyte counts and elevated angiogenesis stimulating factors such as VEGF and EPO in the PB [2]. In contrast, responders display basically elevated EPCs and thrombocytes also in the absence of angiogenesis stimulating factors. The underlying mechanism for a lack of response to induction of cardiac regeneration may be a failure of vascular repair by reduced circulating EPCs. This mechanism has been observed already twelve years ago to be associated with progression in atherosclerosis and coronary artery disease [284] as well as in-stent restenosis [285–288]. Low EPC titers were correlated to untreated hypercholesterolemia, whereas HMGcoA reductase (statin) therapy has been associated with EPC recruitment, activation and improved survival, and improved vascular repair following injury. The e-HEALING, a post-marketing registry of 5000 'all-comers' coronary artery disease patients indeed suggested that the EPC capturing Genous stent was associated with reduced clinical major adverse cardiac events (MACE) and late stent thrombosis [289].

Recent clinical findings on the association of BM failure and cardiovascular disease are setting the spotlight on BMSC function as the main disease pathomechanism. Altered BM subpopulations have been associated with response to SC therapy in CCTRN trial analysis [277,290]. Recently, Jaiswal et al. have found the association of BM-derived EPCs or altered clonal hematopoiesis and atherosclerotic cardiovascular disease [291, 292]. Reduction in hematopoietic clonal capacity have been shown to be relevant for post MI heart failure in mouse models with increased HSC apoptosis [293]. Recent research results from cardiac SC therapy trial PERFECT show the impact of BMSC failure on cardiac regeneration as well as a pivotal role of a defined responder or non-responder status for the induction of cardiac tissue regeneration [2]. Genetic control of cardiovascular disease on SC level may be associated with gene dysfunction of SH2B adaptor protein 3 (SH2B3) [2]. First results assume that an increased SH2B3 expression in PB is associated with non-response to cardiac function improvement [2]. Moreover, this pattern of heart failure is associated with reduced EPCs, non-response to angiogenic stimulation, and reduced angiogenesis in the heart [2].

3.3.3. SH2B3 adaptor protein regulates EPC and SC response in cardiovascular disease

SH2B3, also known as lymphocyte adaptor protein (LNK), is a frequent cause of diseases resulting from genetic variations, acquired or inherited in nonmalignant and malignant hematological diseases [294, 295]. Susceptibility to celiac disease type 13, rheumatism, insulin-dependent diabetes mellitus, and other autoimmune diseases have been demonstrated. Most interesting is the association with hypertensive, arteriosclerotic and coronary disease that has been recently described [296–298]. It can be envisaged that increased inflammatory activation in atherosclerosis and hypertension are associated with lowered SH2B3 expression levels [297,298]. This may also be the basis of increased cancer inflammation associated with SH2B3 [299,300].

The SH2B3 adaptor protein was first described in lymphocytes [301]. Alternatively spliced transcript variants encoding different isoforms have been identified for this gene. Transcription produces seven different mRNAs, six alternatively spliced variants and one unspliced form [302].

The encoded protein is a key negative regulator for cell proliferation and cell activation in the blood, immune, and cardiovascular system [303]. Intracellular pathway modification involves cytokine signaling like IL-7 and IL-11 in B-cell progenitors [304] and control of hematopoiesis by abrogation of growth factor signaling [305].

The protein is expressed mainly in blood cells and immune cells [298]. The tissue expression on endothelia is of importance with respect to regulation of immune activation [298]. As an example the gene transfer of SH2B3 does prevent endothelial cell activation and apoptosis [306]. Recently, in mouse models the relevance SH2B3 gene expression for EPC proliferation, release and restitution of angiogenetic and wound healing capacity has been demonstrated [307].

Mutations in this gene have been associated with hematological disease based on HSC dysfunction in myelodysplasia, erythrocytosis, anemia or myeloproliferative neoplasms including leukemia and lymphoma [305]. Furthermore, mutations have been found to be associated with a variety of autoimmune, cancer and cardiovascular diseases (Table 2). The lessons learned from altered gene function by inherited mutational variants can enable us to associate genotype and phenotype of altered function in patients with the corresponding disease. This further allows the establishment of novel diagnostic and therapeutic strategies in treating these diseases.

The initial description in the PERFECT outcome analysis of assumed SH2B3 gene expression enhancement in the PB of non-responders suggests a potential regulatory role of SH2B3 with respect to suppression of the BM response [2]. Moreover, association with hematological traits, coronary artery disease [296], and arteriosclerosis have been found for point mutations of SH2B3 promotor regions as well as influence of SH2B3 single nucleotide polymorphism (SNP) on human longevity [308].

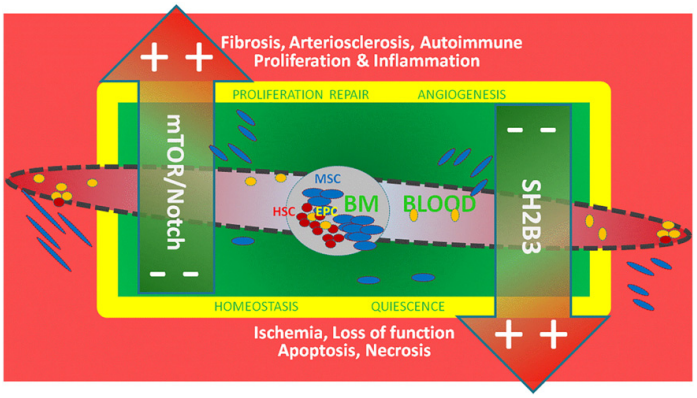


Fig. 3. Stem cell switch hypothesis: Homeostasis and stem cell mediated disease. Stem cell and bone marrow dysfunction in cardiovascular disease (red area). Proliferation and functional control of HSCs/EPCs/MSCs by SH2B3 vs. Notch/mTOR in homeostasis (green), repair (yellow) and disease pathomechanism (red).

Fig. 4. Development of a medicinal product. The scheme illustrates the old developmental process (fragmented). Feedback and quality testing are incomplete. (GCP - Good Clinical Practice, GLP - Good Laboratory Practice, GMP - Good Manufacturing Practice, GVP - Good Vigilance Practice).

Enhanced proliferation in T-cell leukemia has been demonstrated to be driven by Notch1 stimulated mTOR1/Pld signaling. Loss of SH2B3 activity has been found to induce acute lymphoblastic leukemia (ALL) proliferation in Notch1-transgenic mice and was also found in a subpopulation of ALL patients in acute leukemic cells [309]. Homozygous loss of SH2B3 expression was found to be associated with induced Notch1 driven proliferation in relapsing T-ALL [310,311]. The alterations included nonsense and missense mutations affecting the pleckstrin homology and/or the Src homology region 2 (SH2) domains. Mutations in SH2B3 have also been identified in lymphoid malignancies including ALL as both germline and somatic events [312,313].

It has been shown in mouse models that the SH2B3 interference induces EPC proliferation, peripheral EPC release and enhanced angiogenesis [307,314,315]. However, further clinical evaluations of SH2B3 expression are needed to unravel the precise mechanism in humans [2,316]. These may lead to clarification on proposed pivotal regulation of HSCs, EPCs, and MSCs by SH2B3 (Fig. 3). The authors pose this observation as a "stem cell switch hypothesis" to define the role and interplay of SH2B3 vs. Notch/mTOR for homeostatic function and/or dysregulation equilibrium of SCs (HSCs, EPCs, MSCs) in disease (Fig. 3). Further studies have to unravel this question in clinical and experimental setups.

4. Definition of quality standard and best practice

It is conceivable that the continuous presence of all scientific expertise and critical analysis has to be implied in every step of development as well as the decision for reaching milestones to exit to the following step. To avoid risks of failure scientifically based quality management is mandatory. The obvious obstacles in development require a new approach to achieve best practice. The advance of scientific knowledge to molecular disease diagnosis is not yet a solid basis for molecular and cellular interventions in disease pathways. Therefore, to the author's opinion, it is mandatory to change the quality management approach in general. The identification of obstacles for the development and treatment of heart disease as well as new technical solutions have to be discussed.

4.1. Good practice - a classical quality standard is not enough

The development of medicinal products is structured from historical experience by a stepwise interval development divided in basic research, preclinical development and clinical studies. The basic research approach (according to Good Scientific Practice (GSP)) has been drawn from pharmaceutical chemistry and is aiming for a therapeutical substance. For cell products this very often has limited applicability because of complex reactions and tissue products. Moreover, the use of autologous and allogenic cells is more similar to regulation in transfusion or transplantation medicine and interferes with patient dependent rights and ethical aspects.

Scientific verification is needed in each step of the development. The outcome of preclinical experiments (safety, quality and efficacy) is a crucial point and determines whether clinical studies are reasonable. It depends on the clinical result if the (investigational) product will be approved by a national or international authority and thereby will enter the market to be administered routinely to patients (Fig. 4). Only in some cases (e.g. for Advanced Therapy Medicinal Products (ATMPs)) structured and lifelong patient vigilance monitoring after the approval of the medical SC product are performed to observe its long-term-behavior.

Besides the scientific and clinical development it is important to take early considerations in account which are dealing with aspects such as the social view on the new product, the medical need, the accessibility for patients and the mode of payment/reimbursement. A suspended marketing authorization of an implant to repair cartilage (MACI®; EMA/282918/2013, EMEA/H/C/002522) [317], that was not profitable, illustrates the importance of economically issues: e.g. an economically viable manufacturing process or suitable distribution concepts are besides the proven safety and efficacy of the new product also worth to consider. This current developmental chain concept, which has been applied in the initial phase of SC therapies has to be considered as inefficient and contains high risks of failure in the developmental process (Fig. 4).

For nearly two decades cell therapeutics to treat heart diseases are developed and several clinical trials were done. Nevertheless, no approved cell therapy is available for heart failure patients so far (approx. 200 clinical trials (phase I-III)); and there is no product legally in the market [318]. A slightly more optimistic situation exists for cell therapies of other indications (e.g. cartilage repair Spherox® [319] or cornea repair Holoclar® [320]) or other types of ATMPs (gene therapeutics: Glybera® [321], Imlygic® [322], Strimvelis® [323], Zalmoxis® [324]). A discussion about reasons for these low numbers of legally available regenerative therapies recently accelerates in the scientific community. A well-known point is their novelty as a new class of medicinal products with a high complexity regarding their development, manufacturing, characterization and especially in cardiac therapies their administration [325]. Indeed regulators and developers are now in a process of adapting the regulatory framework related to cellular therapies to possible and proven risks and opportunities [326]. Due to the continually growing progress in cell therapy knowledge and experience, there are numerous ways of support provided by the authorities (e.g. European Medicines Agency (EMA), US Food and Drug Administration (FDA)). They created some tools such as ATMP classification, certification, scientific advice, adoptive pathways pilot program and classification of ATMPs. Moreover, all involved parties are invited to profit from concentrated knowledge at EMA's Academia or FDA's Office of Cellular, Tissue and Gene Therapeutics (OCTGT). The most important move, however, came from Japan introducing a new regulatory Regenerative Medicine Act in 2014 aiming to achieve both intensified scientific quality control as well as earlier translation to clinical trials [327].

These recently enhanced regulatory adaptions combined with increasing political awareness raise some hope for reduced obstacles on the way to market authorization for more cell therapies [328]. Although

Table 3
Proposals for improved development.

1. Diagnose disease mechanism and pathway signature
2. Classify (CD, Rseq) and standardize (stem) cell product
3. Establish specific molecular and/or biopsy imaging
4. Monitor treatment effect by biomarkers
5. Use disease specific model validation
6. Establish lifelong quality management for all clinical diagnostics and procedures
7. Establish biobank for tissue, blood, and cell product
8. Use machine learning/Systems medicine for data analysis validation
9. Individualize therapy to target disease and tissue repair mechanism
10. Integrate expert knowledge, but validate by computational approaches like machine learning

improvement is still needed at this point it is not only the restrictive regulations that impede or even prevent innovations.

As basic research and pre-clinical development mostly take place in academically environment, it is important for researchers to know, that dealing with critical aspects of cell therapy development at an early stage is essential [329]. As mentioned above, authorities are willing to support and collaborate. So every research team/developer has the chance to find its own particular way due of the uniqueness of their new product. General guidance cannot easily be translated in specific requirements for a certain cell product [325]. Based on the author's experience, it is very helpful to be early, continuing and actively in contact with the relevant authorities and equally important patient organizations to develop cell therapies successfully. Additional to the frequently complained regulatory hurdles, the attitude of scientific community itself has a major impact on the successful development of all new medicinal products. The competition factor that could cause intratransparency regarding data access and quality is just one example.

Collaboration and sharing of knowledge of publicity-funded research and pharmaceutical industry, in turn, could be the key to enhanced development of cell therapies [328]. Furthermore, Foley and Whitaker suggested in 2012 that at least cell therapies, which demand therapeutic procedures like all currently known cell therapy approaches for heart failure, would benefit from an early collaboration of developers and clinicians [330], because a suitable way of distribution, preparation or administration is also an important issue to widespread a therapy.

Besides all regulatory or economic issues there is apparently a discrepancy between the large amount of preclinical data on the one hand and the outcome of clinical trials with cell therapies for ischemic heart diseases on the other [331]. Because of their role of being responsible for public/patient safety no agency would have agreed to test a particular cell product in humans without promising in vivo results. According to published data, early as well as late stage clinical trials with heart failure related products could show their safety and feasibility during the observation period [278,332]. These are important and valuable results but verification of safety is only the first part of the developmental process for medicinal products. Clinical trials also need to address the question of efficacy: the new cell or another medicinal product must benefit the patient in a measurable way (endpoints), which could not be confirmed for any cardiac cell therapy so far. For all interested parties – patients as well as authorities and developers – it is a frustrating situation when enormous scientific and monetary research efforts as well as patient's involvement (e.g. frequent consultations, follow-up interviews and treatment with poorly conceived therapies) lead to no advantageous cell products [333]. Following consequences from this are intentions to spend resources (money and time) for more promising and "fancy" approaches. The worst effect will be a further growing grey market of unproven therapies (e.g. SC tourism) in under-regulated countries as well as in Europe or the U.S. [334]. where no regulations are given to protect patients from poor investigated products.

It is obvious that the current approach of medicinal product translation from academic basic science to commercial SC product development lacks large scale knowledge integration and has a high risk of failure to reach standard therapy. Attempts have been made in US and recently in Europe to form large scale task force projects like CCTRN [277], TACTICS [278] or PERFECT [2] with longterm setup to face the challenge of complex continuous diagnostic and therapy development. The uncertain situation of the whole endeavor of development of a regenerative cardiovascular medicine based on SCs at the moment requires an international consensus program lead by governmental research councils in collaboration with regulatory authorities. For this shortcomings program based approaches to improve development (Table 3) have to be defined and tested.

4.2. The role of good practice

What is the reason for this disappointing setting in cell therapies although the developmental process should be performed according to good practice (GxP)? GxP in particularly with regard to medicinal products includes the regulated international quality standards Good

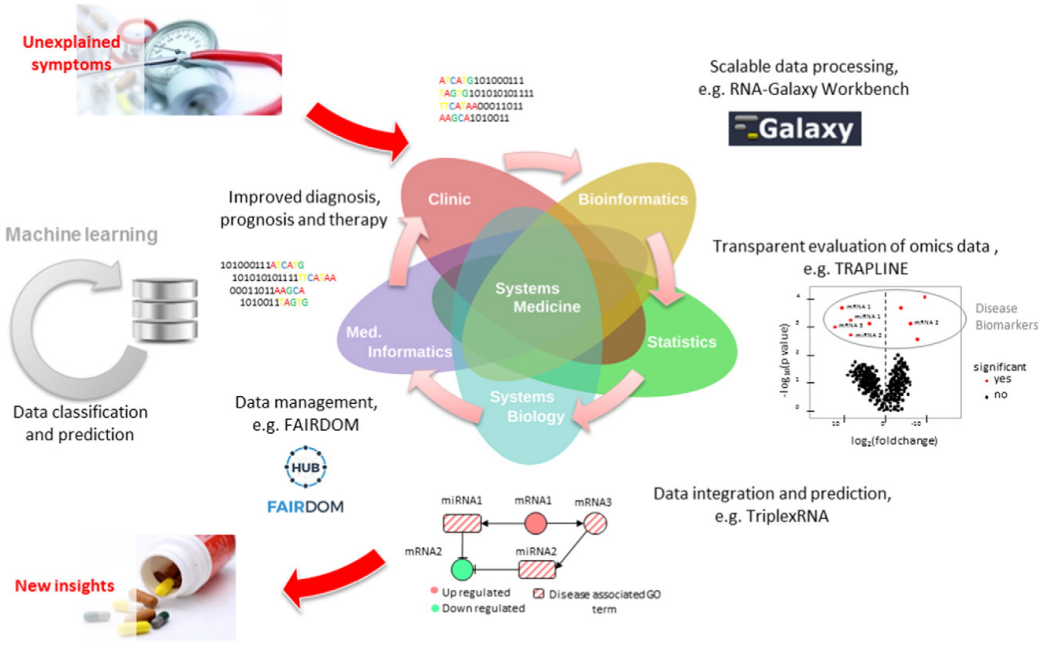


Fig. 5. Integration of a Systems Medicine approach within a stem cell therapy development scenario.

Laboratory Practice (GLP) for preclinical research, Good Manufacturing Practice (GMP) for the manufacturing of the investigational product and Good Clinical Practice (GCP) for clinical research (Fig. 4). In the author's opinion, there is a massive misunderstanding in expecting a certain quality per se from these standards, but in contrast GxP includes only formal specifications on how data in preclinical and clinical settings are recorded (and stored). GxP does not assure the usage of appropriate (= scientifically relevant) methods, thus in general GxP alone cannot guarantee a scientific significance of analyses or examinations.

In order to ensure only reproducibility and correctness of data, preclinical investigations should be performed according to GLP, a worldwide accepted standard that determines the way how non-clinical test results are generated and documented. In health and environmental relevant safety testing of industrial chemicals well established and approved test methods (e.g. in vivo skin irritation and corrosion data (OECD TG404 for toxicity testing)) already exist. But especially for new approaches like cell products GLP does not state, if certain tests are scientifically suitable to address specific questions. This way still needs to be gone for cellular products and their characterization. Therefore, the development, testing and definition of specific outcome parameters for efficacy and safety testing are mandatory for the field of SC therapy.

GMP is a second standard, which is the prerequisite in manufacturing cell or other medicinal products intended for (pre)clinical testing or routine care. For one part of GMP, QC of the medicinal product, there are special pharmaceutical rules for quality and analytics (methods) that are officially recognized and under permanent revision (pharmacopoeia), which have verifiable to be validated on every QC-site by the holder of a manufacturing permission. Because of the novelty of cellular therapies, there are only few quality related methods for testing cellular therapeutic agents (cell number, sterility). Like in GLP there is also a need for significant tests in GMP that prove efficacy (in terms of stable and effective quality) and safety of a certain cell product in relation to heart failure or other medical indication. This lack of functional characterization of cell products beyond cell numbers, few extracellular markers and sterility has to be removed by scientific investigations (basic and pre-clinical research) that have to be peer reviewed and discussed by the scientific community to establish common accepted standards for relevant markers, parameters and analytical procedures (comparable to Pharmacopoeia) in this field.

An example for this standardization process is the longstanding development of the cluster determinant membrane epitope characterization or the genecard database (e.g. https://www.ncbi.nlm.nih.gov/gene). A similar profiling has to be requested for an international SC molecular sequencing profile register. This could be used to characterize genetic or disease variations in the phenotype and genotype of SC.

GCP is an ethic and scientific standard, used only how to plan and perform clinical trials in humans and to document and report the outcome. Also GCP is characterized through specialized formal documents (e.g. Declaration of Helsinki, ICH guideline, etc.), which aim to protect human rights and quality of the produced data. But, as mentioned above, the main focus of attention of GCP is again just on documentation of data. However, this is only one side of the coin – the explanatory power of data, on the other side, depends on applicability of the performed method in terms of validation. Biological or technical variance or inconsistence of diagnostic findings can lead to non-evaluation of endpoints and can either be caused by methodical differences – which become even more relevant in multi-center trials – or due to using simply inappropriate methods. The first issue can easily be solved by the validation of a method (e.g. magnetic resonance imaging (MRI) assessment) for every study site. The second issue needs more commitment, because a comprehensive evaluation of safety and efficacy for cell products in clinical studies needs at first the definition of relevant parameters by intense research and an open discussion in the scientific community.

5. International standard of data analysis

Systems medicine emerged as an inevitable tool to investigate complex diseases by the integration of multidimensional datasets and numerous mathematical approaches with data from pre-clinical and clinical studies. The iterative cycle of data-driven modeling and model-driven experimentation, in which alternative hypotheses are postulated and refined until they are validated, helps in identifying new mechanistic details of cell-biological processes and previously unidentified regulatory interactions in the system [335]. However, systems approaches are widely perceived as basic research, so that a main current challenge is to shift from the “need” of translating basic finding into clinical research towards the integration between non-clinical and clinical data. Cardiovascular diseases (CVDs), being multifactorial, may be a potential field test for Systems Medicine [336]. Moreover, it has been shown that current CVD risk scores could be improved in accuracy by computational approaches that identify disease risk and predict the maximum personalized treatment benefit [337]. In this section, we highlight the opportunities and hurdles of data mining, novel sequencing approaches, network methodologies and machine learning (ML) for cardiac research (Fig. 5).

5.1. Data mining and management

Accessing and retrieving high quality omics datasets is the first great challenge to overcome while working in the field of cardiac diseases. Analysis of high-throughput experimental data together with patient phenotypic information has led to the identification of sets of candidate genes, proteins and pathways that may be implicated in many disease conditions. In order to build a higher level picture of the underlying processes involved in the disease pathology, it is necessary to integrate various classes of heterogeneous information and to explore the complex relationships between entities such as diseases, candidate genes, proteins, interactions and pathways [338]. In contrast to monolithic databases, graph databases provide a powerful framework for a combined storage, querying and envisioning of such complex biological datasets. For example the graph database platform Bio4j integrates popular databases like GO, RefSeq, NCBI Taxonomy and Expasy Enzyme DB and allows for intrinsic and extrinsic semantic feature implementation to enhance their respective relationships and importance towards a common biological perspective [339]. Another aspect is the need for an environment that allows the management and sharing of generated heterogeneous datasets and computational models in the context of the experiments, which created them. The FAIRDOMHub is a framework for publishing FAIR (findable, accessible, interoperable and reusable) Data, operating procedures and models for the Systems Biology community that enables researchers to organize, share and publish data, models and protocols for an enhanced reproducibility and reusability of research results [340] (Fig. 5).

5.2. Next generation sequencing

The actual generation of Next Generation Sequencing (NGS) data is steadily increasing, especially the numerous data being generated for genome-wide association studies (GWAS) have uncovered numerous genetic variants (SNPs) and alternative splicing forms that are associated with blood pressure [341] as well as human heart development [342,343]. Especially SH2B3 emerged as a powerful switch for the influence on blood pressure by using GWAS, meta- and network analyses [344,345]. An extended identification and characterization of additional T-box transcription factor 5 (TBX5) mutations and SNPs was also achieved, which hold promise for a therapeutic strategy targeting TBX5 associated developmental abnormalities and diseases [346]. These novel sequencing technologies have also resulted in the discovery of previously unannotated long non-coding RNAs (lncRNAs), which are under

further investigation for the amelioration of CVDs [347]. Furthermore, it has been reported that the interplay of chromatin modifications and non-coding RNAs in the heart also plays a bigger role than previously expected [348]. The list of examples for sequencing success stories could be continued, but we want to put more emphasize on the current bottleneck of this emerging technology – transparent, reproducible and proper data analysis strategies [349]. With respect to the number of data analysis steps, the complexity of decisions on tool selection is increasing likewise, hence calling for systematic workflow development and management frameworks [350]. For this reason, Galaxy [351] and the Galaxy-RNA-Workbench [352] are providing a general framework that makes advanced computational tools accessible without the need of prior extensive training. Galaxy seeks to make data-intensive research more accessible, transparent and reproducible by providing a web-based environment in which users can perform computational analyses and have all of the details automatically tracked for later inspection, publication, or reuse. Data analysis pipelines within such data analysis platforms can be easily used for the QC, complete data processing and advanced predictive analyses and evaluations, as it has been recently shown for RNA sequencing datasets [353]. Increasing the ease of use and comprehensiveness of tools and computational methodologies an interactive environment framework for Galaxy was designed to combine Galaxy's tools and workflows with popular advanced computational environments such as Jupyter [354]. This development tremendously simplifies the daily routine of tool developers and non-computational end users. Taken together, tailor-made and expert-driven scientifically developed computational workflows instead of rigid data analyses are mandatory for an appropriate preclinical and clinical data analysis [349].

5.3. Network approaches

Network approaches are another central concept in Systems Medicine, because they combine the existing knowledge about classic linear pathways with experimental data. Biological networks occur on many different levels such as genes, transcripts, proteins, metabolites, organelles, cells, organs, organisms, and social systems. In general, they appear to exhibit an architecture described mathematically as "scale free," in which most nodes have few links but a small fraction of nodes (called hubs) are highly interconnected [355]. Those hub-genes are assumed to be biologically relevant, because they represent mediators to interconnect and regulate different processes that might play a critical key role within the underlying network of involved pathways. In addition to broadly applied single pathway analyses, differential network detection provides enhanced explanatory insights while taking into account the changing interplay of pathways, e.g. during disease progression [356]. Nowadays, predicting molecular commonalities between phenotypically related diseases, even if they do not share primary disease genes is possible and it can be assumed that network-based approaches, relying on an increasingly accurate interactome, are poised to become unavoidable in interpreting disease-associated genome variations [357]. Furthermore, gene co-expression network based approaches such as Weighted Gene Co-expression Network Analysis (WGCNA), which is one of the most powerful approaches, have been widely used in analyzing microarray and RNA sequencing data, especially for identifying functional modules and hub-like genes [358]. However, it has to be taken into consideration that there might be major topological difference between RNA-seq and microarray co-expression in the form of low overlaps between hub-like genes from each network

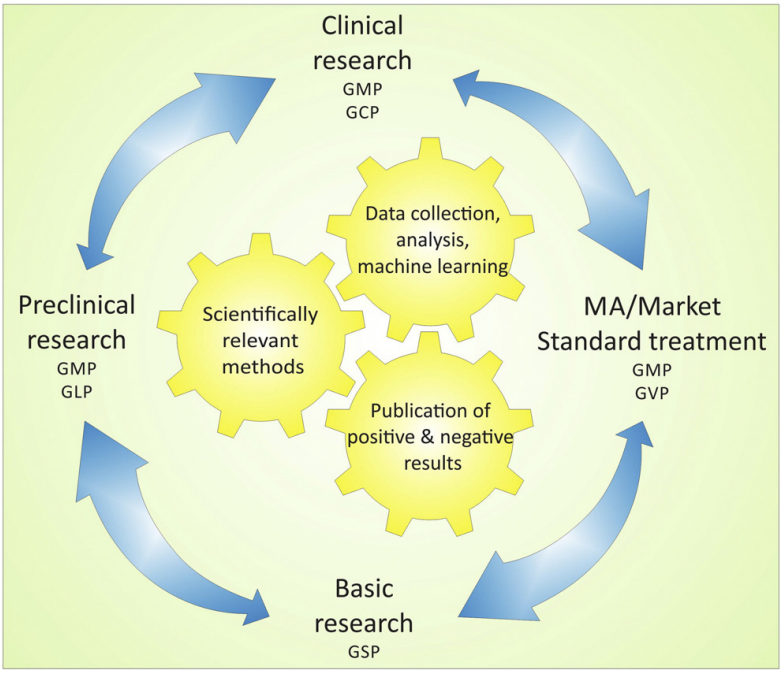


Fig. 6. Development of a regenerative medicine therapy as an integrated process. Every stage of research is a component of a circular process that is based on four central approaches and spins in both directions. Each result, if positive or negative, affects other parts of the development. This approach leads to enhanced knowledge about the (cell) product as well as the affected and underlying mechanisms in humans. Quality standards and risk-benefit-evaluation are the main elements of this approach. GCP - Good Clinical Practice, GLP - Good Laboratory Practice, GMP - Good Manufacturing Practice, GSP - Good Scientific Practice, GVP - Good Vigilance Practice, MA - Market authorization.

due to changes in the correlation of expression noise within different technologies [359]. Disease maps are another novel expert-approved pathway-based reconstruction of a network customizing a particular disease or being used as a graphical review on the molecular mechanisms of a disease. It is a collection of interconnected signaling, metabolic and gene regulatory pathways stored in standard Systems Biology formats (e.g. SBGN, SBML, BioPAX) [338]. As interdisciplinary projects continue to generate large amounts of heterogeneous datasets, the network approaches presented here, may offer useful solutions for knowledge integration.

5.4. Artificial intelligence and machine learning

As previously described, established approaches to CVD risk assessment, such as the recommendations by the American Heart Association/American College of Cardiology (AHA/ACC), predict the prognostic risk of CVD based on common risk factors like cholesterol, age, smoking, and diabetes [360]. However, there are numerous patients remaining that fail to be identified by these classical linear prediction models and some patients are unnecessarily treated [361]. These models may thus oversimplify complex high-dimensional datasets by using too few parameters and not considering non-linear interactions among the measured parameters. With the rise of highly efficient ML algorithms, alternative approaches to classical linear prediction models have been developed that have the potential to use available "Big data" for better prognosis and diagnosis [362]. The artificial intelligence relies on algorithms to learn in a supervised or unsupervised manner the provided input data by minimizing the error between predicted and observed outcomes (supervised) and, finally, unravelling the complex and non-linear interactions between the parameters [363]. ML significantly improves the accuracy of cardiovascular risk prediction, increases the number of patients identified who could benefit from a preventive treatment and likewise avoids unnecessary treatment of others [361]. Our group recently used clinical and advanced subgroup measurements in the phase III clinical trial to enable a therapy responsive patient classification for a potential intervention with an applied ML model and obtained a prediction accuracy of above 90% [2]. In addition to pure classification of disease and healthy states, supervised ML can be used to uncover unexpected parameters to be important for the choice between these states that can be subsequently used to further investigate the underlying mechanism or a particular molecule. In addition to supervised ML, unsupervised ML approaches need no specific ground truth for training the actual model, but are based on their non-linearity dimension reduction less effective to identify a specific set of important parameters. These unsupervised statistical learning analyses assume that there are naturally occurring subclasses within patients that behave differently yet reproducibly across a number of populations and varying scenarios (e.g. treatments, ethnologies, environments). Thus, the first part of a study emphasizes finding intrinsic structure within patient phenotypic data, which can then be evaluated retrospectively and prospectively for predicting treatment outcomes and guiding clinical trial design [364]. By applying non-linear approaches like t-distributed stochastic neighbor embedding (t-SNE) for dimensional reduction, our group identified two distinct groups, respectively with responder and non-responder characteristics, within the data and, thus, could independently confirm our response biomarker signature hypothesis [2]. The results obtained can subsequently be used for supervised learning, e.g. for an estimation and prediction of distinct classification groups as well as the underlying important features. Taken together, ML is becoming an invaluable asset to test and evaluate novel classification hypotheses of a disease or clinical syndrome and should be a mandatory analysis for a robust and independent validation state.

6. Comprehensive centers for R&D integrated disease treatment

The current health care concept experiences an economic dominance of business models based on medicinal products and health care service to be purchased by patients for gaining health. Principally, this system needs supervision for outcome control and should be integrated in society priority to improve disease burden. The unique chance of SC based therapies, however, is to realize a basic repair of disease. This enormous task can only be realized by supervised quality management in the developmental process of therapies as well as in standard care.

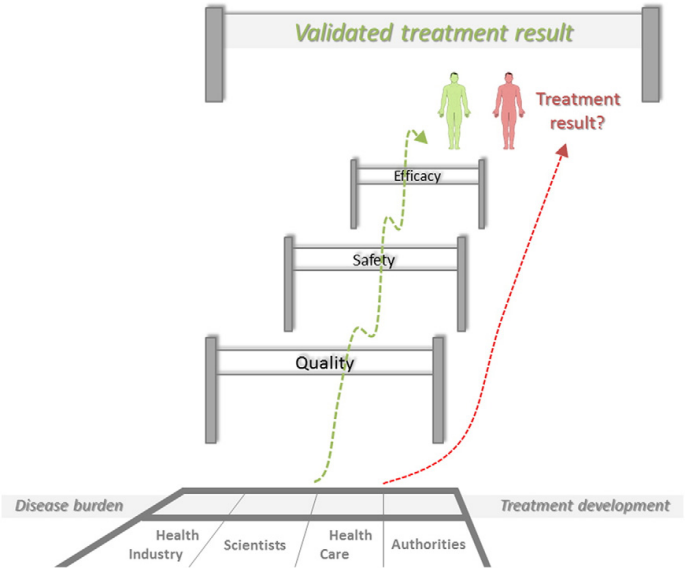


Fig. 7. Disease burden and treatment development: Who benefits? At least three hurdles need to be passed during the development of new cell therapies for patients: quality, efficacy and safety have to be proven for market authorization. Nevertheless there are ways to bypass these steps to reach patients. Here in the grey market zone, patients get treated and the outcome can be very dangerous (in red). Reasons for the grey market are diverse: scientific prestige, sales volume etc. But also "over-regulation" by laws – not adapted to current situations – could result in a decelerated development. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The model of comprehensive treatment centers integrating research, development, licensing, clinical development and highest quality lifelong patient care could be the masterfile for the development of cardiac stem cell therapies in the context of adequately specialized heart disease centers.

As initially described, the classical flow of information leads in one direction: from basic and preclinical research via clinical analysis to authorization and in best case beyond (vigilance; Fig. 4). This one-sidedness limits the generation and multiplication of knowledge because unexpected results from pre- and clinical research are apparently blocking continuous improvement. Recently, in a white paper of TACTICS-members a new translational axis from basic research via preclinical research involving several different stages of animal models to clinical trials with feedback to all previous stages was suggested [278].

In contrast to this proposal, the authors state that translation includes not only a directed process with backwards oriented feedback (Fig. 4), but has to be routed in an integrative process management, where the complete interdisciplinary team (scientists, clinicians, engineers, etc.) is involved in each and every step of the development to maintain complex analyses. The road picture as depicted in Fig. 6 will be a circling knowledge process like a circling road fitting solutions to basic, preclinical and clinical exits. Especially the availability of special diagnostic and imaging technology has to bridge over all developmental steps. In some cases there is a chance for products failed in clinical trials to be reinvestigated in basic research considering the actual state-of-the-art. Apparently, the first licensed heart related regenerative and cell therapy has to pass this developmental circle repetitively before reaching the status of a highly standardized therapy. All clinical results need to be investigated again in preclinical settings in more detail to clarify the affected disease mechanisms. Moreover, granting of the marketing authorization is not the end, but the beginning of cell therapy standardization.

Given a continuous vigilance under GCP the possibility for conditioned early approval of therapy would have a safer prospect. The authors postulate that all ATMP therapies should be monitored lifelong like in comprehensive cancer center concept for tumor therapy or organ transplantation medicine according to Good Vigilance Practice (GVP). IT-based concepts have been suggested for improved patient care in diabetes [365].

Only a limited amount of specialized patients (defined by age, sex, comorbidities, etc.) can be analyzed before regulatory approval. The most important stage to demonstrate safety and long-term behavior of new products is therefore after approval and market authorization (= pharmacovigilance). Because of this pharmacovigilance needs to be obligatory for every treating, clinic and manufacturer must be specified and controlled by competent authorities. Possible long-time effects or other outcome can be taken into consideration for the risk-benefit-balance evaluation that is performed annually by EMA. In Germany, the competent authority (Paul-Ehrlich-Institute (PEI)) coordinates and supervises pharmacovigilance systems for four product classes (ATMP, in vitro-diagnostics, tissues, blood components) where side effects and severe events are reported and evaluated. These systems must serve as an accessible database for every patient and physician.

The authors assume further, that an intensive monitoring of patients including bio-banking should be employed to give access for later analysis e.g. marker analysis [2]. The current situation would benefit from a standardized approach to gather preclinical and clinical data as well as biological samples. Therefore, all information related to a certain therapy/cell product can be made available for the scientific and clinical community. Recently a first step was done by EMA with its freely available "clinical data" website (www.clinicaltrialsregister.eu). As a result theoretical simulations ("high-dimensional data analysis") and experimental setups could be adapted more easily. This also includes the characterization of patient conditions compared to healthy individuals. Another example for pattern recognition and thereof arising possibilities is the work of Sengupta et al. [366] where a cognitive ML algorithm could be generated for cardiac imaging. Another Systems Medicinal approach was used to detect differential connections between diseases associated networks [357]. In summary this new way of large scale analyzing methods require harmonized and standardized datasets for research and clinical treatment.

Until now, there is at least in the field of heart failure treatment no consensus about which kind of clinical endpoints are crucial and accepted by the scientific community. This leads to a non-defined clinical outcome (Fig. 7). Among other aspects this point was also already addressed by Fernández-Avilés et al. who additionally assumed that multidisciplinary collaborations being aware of these limitations could not overcome them until now [278]. Therefore, the knowledge has to become even broader: proving efficacy goes along with understanding the underlying mechanisms. Thus the scientific discussion of the before mentioned topics needs to be forced and would also benefit from an open-minded culture of publishing "negative" results to reduce repetition of failed approaches. Such publication bias could lead to unjustified transition or stop of approaches in the developmental process [367]. Likewise, also successful approaches have to be questioned, as Nowbar et al. showed discrepancies in the enhancement of ejection fraction in trials investigating autologous BMSCs [368]. Finally the debate should end in accepted scientific standards that will elevate GxP to a higher level for cell therapies comparable to the area of chemical medicinal products (small molecules). The knowledge-driven discourse is the essential part during the development of cell therapies: ideally as long as knowledge grows, standards need to be adapted to present circumstances.

Considering the findings of the PERFECT trial [2] – every patient has a responder (or a non-responder) biomarker signature – it is worth to analyze recently published clinical data with no significant outcome again. Concomitant research programs, additionally to the clinical trial protocols, could benefit a detailed and comprehensive analysis. With a more nuanced view supposed negative results could turn out to be more specific than initially assumed. Furthermore, the scientific strategy in planning future investigations in terms of suitable tests and animal models need to be optimized. Like exemplified in part 3.1 using Systems Medicine approaches and accurate data analysis settings, the outcome of preclinical and clinical studies can be enhanced.

Linked to all previously discussed issues are the questions regarding regulatory hurdles and the financing of research for small and middle sized companies as well as for universities. Well performed research takes well skilled scientists and money over a long period of time. Just the concerted efforts of all involved parties (authorities, industry, patient organizations, researchers, scientific publication machinery, etc.) can result in an increased input-output-ratio for cell therapy products.

Based on this knowledge, it will be mandatory to intensify data exchange and establish interdisciplinary standards of cooperation. There is a great need for a central clinical database for CVD mechanisms and regenerative approaches, unanimous quality management, classification of SC products by gene regulation, cluster determinant platforms for (stem) cell differentiation, diagnostic fostering of biomarker and imaging applications (Table 3). Ultimately, applying artificial intelligence by means of ML methodologies may be a crucial step towards deeper insights into the complex mechanisms of cardiovascular regeneration.

Funding

This work was supported by the Federal Ministry of Education and Research Germany (FKZ 0312138A and FKZ 316159), the State Mecklenburg-Western Pomerania with EU Structural Funds (ESF/IVWM-B34-0030/10 and ESF/IVBM-B35-0010/12).

Declaration of interests

All authors declare no competing interests.

Contributors

GS performed the conception, design, drafting, and writing supervision of the manuscript. JN, MW, JG, UR, PV, PM performed extensive literature search and wrote the manuscript. All authors read and approved the final version for submission.

References

[1] C. Stamm, B. Westphal, H.-D. Kleine, M. Petzsch, C. Kittner, H. Klinge, C. Schümichen, C.A. Nienaber, M. Freund, G. Steinhoff, Autologous bone-marrow stem-cell transplantation for myocardial regeneration, Lancet 361 (9351) (2003) 45–46.

[2] G. Steinhoff, J. Nesteruk, M. Wolfien, G. Kundt, J. Börgermann, R. David, J. Garbade, J. Große, A. Haverich, H. Hennig, A. Kaminski, J. Lotz, F.-W. Mohr, P. Müller, R. Oostendorp, U. Ruch, S. Sarikouch, A. Skorska, C. Stamm, G. Tiedemann, F.M. Wagner, O. Wolkenhauer, Cardiac function improvement and bone marrow response -: outcome analysis of the randomized PERFECT phase III clinical trial of intramyocardial CD133(+) application after myocardial infarction, EBioMedicine 22 (2017) 208–224.

[3] M.A. Eglitis, E. Mezey, Hematopoietic cells differentiate into both microglia and macroglia in the brains of adult mice, Proc. Natl. Acad. Sci. U. S. A. 94 (8) (1997) 4080–4085.

[4] N.D. Theise, M. Nimmakayalu, R. Gardner, P.B. Illei, G. Morgan, L. Teperman, O. Henegariu, D.S. Krause, Liver from bone marrow in humans, Hepatology 32 (1) (2000) 11–16.

[5] T.R. Brazelton, F.M. Rossi, G.I. Keshet, H.M. Blau, From marrow to brain: expression of neuronal phenotypes in adult mice, Science (New York, N.Y.) 290 (5497) (2000) 1775–1779.

[6] R.E. Bittner, C. Schofer, K. Weipoltshammer, S. Ivanova, B. Streubel, E. Hauser, M. Freilinger, H. Hoger, A. Elbe-Burger, F. Wachtler, Recruitment of bone-marrow-derived cells by skeletal and cardiac muscle in adult dystrophic mdx mice, Anat. Embryol. 199 (5) (1999) 391–396.

[7] G. Steinhoff, K. Wonigeit, H.J. Schäfers, A. Haverich, Expression of monomorphic and polymorphic major histocompatibility complex determinants in human heart grafts, Transplant. Proc. 20 (1 Suppl 1) (1988) 67–71.

[8] G. Steinhoff, K. Wonigeit, H.J. Schäfers, A. Haverich, Sequential analysis of monomorphic and polymorphic major histocompatibility complex antigen expression in human heart allograft biopsy specimens, J. Heart Transplant. 8 (5) (1989) 360–370.

[9] D. Orlic, J. Kajstura, S. Chimenti, I. Jakoniuk, S.M. Anderson, B. Li, J. Pickel, R. McKay, B. Nadal-Ginard, D.M. Bodine, A. Leri, P. Anversa, Bone marrow cells regenerate infarcted myocardium, Nature 410 (6829) (2001) 701–705.

[10] A.A. Kocher, M.D. Schuster, M.J. Szabolcs, S. Takuma, D. Burkhoff, J. Wang, S. Homma, N.M. Edwards, S. Itescu, Neovascularization of ischemic myocardium by human bone-marrow-derived angioblasts prevents cardiomyocyte apoptosis, reduces remodeling and improves cardiac function, Nat. Med. 7 (4) (2001) 430–436.

[11] C.J. Eaves, Hematopoietic stem cells: concepts, definitions, and the new reality, Blood 125 (17) (2015) 2605–2613.

[12] T. Asahara, Isolation of putative progenitor endothelial cells for angiogenesis, Science (New York, N.Y.) 275 (5302) (1997) 964–966.

[13] M.C. Yoder, Human endothelial progenitor cells, Cold Spring Harb. Perspect. Med. 2 (7) (2012) a006692.

[14] S. Takizawa, E. Nagata, T. Nakayama, H. Masuda, T. Asahara, Recent progress in endothelial progenitor cell culture systems: potential for stroke therapy, Neurol. Med. Chir. 56 (6) (2016) 302–309.

[15] M. Korbling, Peripheral blood stem cell versus bone marrow allotransplantation: does the source of hematopoietic stem cells matter? Blood 98 (10) (2001) 2900–2908.

[16] W. Wojakowski, M. Tendera, A. Michalowska, M. Majka, M. Kucia, K. Maslankiewicz, R. Wyderka, A. Ochala, M.Z. Ratajczak, Mobilization of CD34/CXCR4+, CD34/CD117+, c-met+ stem cells, and mononuclear cells expressing early cardiac, muscle, and endothelial markers into peripheral blood in patients with acute myocardial infarction, Circulation 110 (20) (2004) 3213–3220.

[17] W. Wojakowski, M. Tendera, A. Zebzda, A. Michalowska, M. Majka, M. Kucia, K. Maslankiewicz, R. Wyderka, M. Krol, A. Ochala, K. Kozakiewicz, M.Z. Ratajczak, Mobilization of CD34(+), CD117(+), CXCR4(+), c-met(+) stem cells is correlated with left ventricular ejection fraction and plasma NT-proBNP levels in patients with acute myocardial infarction, Eur. Heart J. 27 (3) (2006) 283–289.

[18] M. Massa, V. Rosti, M. Ferrario, R. Campanelli, I. Ramajoli, R. Rosso, G.M. de Ferrari, M. Ferlini, L. Goffredo, A. Bertoletti, C. Klersy, A. Pecci, R. Moratti, L. Tavazzi, Increased circulating hematopoietic and endothelial progenitor cells in the early phase of acute myocardial infarction, Blood 105 (1) (2005) 199–206.

[19] A.M. Leone, S. Rutella, G. Bonanno, A. Abbate, A.G. Rebuzzi, S. Giovannini, M. Lombardi, L. Galiuto, G. Liuzzo, F. Andreotti, G.A. Lanza, A.M. Contemi, G. Leone, F. Crea, Mobilization of bone marrow-derived stem cells after myocardial infarction and left ventricular function, Eur. Heart J. 26 (12) (2005) 1196–1204.

[20] L. da Silva Meirelles, A.I. Caplan, N.B. Nardi, In search of the in vivo identity of mesenchymal stem cells, Stem Cells 26 (9) (2008) 2287–2299.

[21] L. da Silva Meirelles, P.C. Chagastelles, N.B. Nardi, Mesenchymal stem cells reside in virtually all post-natal organs and tissues, J. Cell Sci. 119 (Pt 11) (2006) 2204–2213.

[22] K. Le Blanc, I. Rasmusson, B. Sundberg, C. Götherström, M. Hassan, M. Uzunel, O. Ringdén, Treatment of severe acute graft-versus-host disease with third party haploidentical mesenchymal stem cells, Lancet 363 (9419) (2004) 1439–1441.

[23] H.C. Quevedo, K.E. Hatzistergos, B.N. Oskouei, G.S. Feigenbaum, J.E. Rodriguez, D. Valdes, P.M. Pattany, J.P. Zambrano, Q. Hu, I. McNiece, A.W. Heldman, J.M. Hare, Allogeneic mesenchymal stem cells restore cardiac function in chronic ischemic cardiomyopathy via trilineage differentiating capacity, Proc. Natl. Acad. Sci. U. S. A. 106 (33) (2009) 14022–14027.

[24] A.R. Williams, K.E. Hatzistergos, B. Addicott, F. McCall, D. Carvalho, V. Suncion, A.R. Morales, J. Da Silva, M.A. Sussman, A.W. Heldman, J.M. Hare, Enhanced effect of combining human cardiac stem cells and bone marrow mesenchymal stem cells to reduce infarct size and to restore cardiac function after myocardial infarction, Circulation 127 (2) (2013) 213–223.

[25] A.P. Beltrami, L. Barlucchi, D. Torella, M. Baker, F. Limana, S. Chimenti, H. Kasahara, M. Rota, E. Musso, K. Urbanek, A. Leri, J. Kajstura, B. Nadal-Ginard, P. Anversa, Adult cardiac stem cells are multipotent and support myocardial regeneration, Cell 114 (6) (2003) 763–776.

[26] C. Bearzi, M. Rota, T. Hosoda, J. Tillmanns, A. Nascimbene, A. de Angelis, S. Yasuzawa-Amano, I. Trofimova, R.W. Siggins, N. Lecapitaine, S. Cascapera, A.P. Beltrami, D.A. D'Alessandro, E. Zias, F. Quaini, K. Urbanek, R.E. Michler, R. Bolli, J. Kajstura, A. Leri, P. Anversa, Human cardiac stem cells, Proc. Natl. Acad. Sci. U. S. A. 104 (35) (2007) 14068–14073.

[27] G.M. Ellison, C. Vicinanza, A.J. Smith, I. Aquila, A. Leone, C.D. Waring, B.J. Henning, G.G. Stirparo, R. Papait, M. Scarfo, V. Agosti, G. Viglietto, G. Condorelli, C. Indolfi, S. Ottolenghi, D. Torella, B. Nadal-Ginard, Adult c-kit(pos) cardiac stem cells are necessary and sufficient for functional cardiac regeneration and repair, Cell 154 (4) (2013) 827–842.

[28] K.-L. Laugwitz, A. Moretti, J. Lam, P. Gruber, Y. Chen, S. Woodard, L.-Z. Lin, C.-L. Cai, M.M. Lu, M. Reth, O. Platoshyn, J.X.-J. Yuan, S. Evans, K.R. Chien, Postnatal isl1+ cardioblasts enter fully differentiated cardiomyocyte lineages, Nature 433 (7026) (2005) 647–653.

[29] R. Genead, C. Danielsson, A.B. Andersson, M. Corbascio, A. Franco-Cereceda, C. Sylven, K.-H. Grinnemo, Islet-1 cells are cardiac progenitors present during the entire lifespan: from the embryonic stage to adulthood, Stem Cells Dev. 19 (10) (2010) 1601–1615.

[30] C.-L. Cai, X. Liang, Y. Shi, P.-H. Chu, S.L. Pfaff, J. Chen, S. Evans, Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart, Dev. Cell 5 (6) (2003) 877–889.

[31] S. Uchida, P. de Gaspari, S. Kostin, K. Jenniches, A. Kilic, Y. Izumiya, I. Shiojima, K. Grosse Kreymborg, H. Renz, K. Walsh, T. Braun, Sca1-derived cells are a source of myocardial renewal in the murine adult heart, Stem Cell Rep. 1 (5) (2013) 397–410.

[32] X. Wang, Q. Hu, Y. Nakamura, J. Lee, G. Zhang, A.H.L. From, J. Zhang, The role of the sca-1+/CD31- cardiac progenitor cell population in postinfarction left ventricular remodeling, Stem Cells 24 (7) (2006) 1779–1788.

[33] H. Oh, S.B. Bradfute, T.D. Gallardo, T. Nakamura, V. Gaussin, Y. Mishina, J. Pocius, L.H. Michael, R.R. Behringer, D.J. Garry, M.L. Entman, M.D. Schneider, Cardiac progenitor cells from adult myocardium: homing, differentiation, and fusion after infarction, Proc. Natl. Acad. Sci. U. S. A. 100 (21) (2003) 12313–12318.

[34] K. Malliaras, T.-S. Li, D. Luthringer, J. Terrovitis, K. Cheng, T. Chakravarty, G. Galang, Y. Zhang, F. Schoenhoff, J. van Eyk, L. Marban, E. Marban, Safety and efficacy of allogeneic cell therapy in infarcted rats transplanted with mismatched cardiosphere-derived cells, Circulation 125 (1) (2012) 100–112.

[35] I. Chimenti, R.R. Smith, T.-S. Li, G. Gerstenblith, E. Messina, A. Giacomello, E. Marban, Relative roles of direct regeneration versus paracrine effects of human cardiosphere-derived cells transplanted into infarcted mice, Circ. Res. 106 (5) (2010) 971–980.

[36] A.J. White, R.R. Smith, S. Matsushita, T. Chakravarty, L.S.C. Czer, K. Burton, E.R. Schwarz, D.R. Davis, Q. Wang, N.L. Reinsmoen, J.S. Forrester, E. Marban, R. Makkar, Intrinsic cardiac origin of human cardiosphere-derived cells, Eur. Heart J. 34 (1) (2013) 68–75.

[37] E. Messina, L. de Angelis, G. Frati, S. Morrone, S. Chimenti, F. Fiordaliso, M. Salio, M. Battaglia, M.V.G. Latronico, M. Coletta, E. Vivarelli, L. Frati, G. Cossu, A. Giacomello, Isolation and expansion of adult cardiac stem cells from human and murine heart, Circ. Res. 95 (9) (2004) 911–921.

[38] O. Pfister, F. Mouquet, M. Jain, R. Summer, M. Helmes, A. Fine, W.S. Colucci, R. Liao, CD31- but not CD31+ cardiac side population cells exhibit functional cardiomyogenic differentiation, Circ. Res. 97 (1) (2005) 52–61.

[39] C.M. Martin, A.P. Meeson, S.M. Robertson, T.J. Hawke, J.A. Richardson, S. Bates, S.C. Goetsch, T.D. Gallardo, D.J. Garry, Persistent expression of the ATP-binding cassette transporter, Abcg2, identifies cardiac SP cells in the developing and adult heart, Dev. Biol. 265 (1) (2004) 262–275.

[40] T. Oyama, T. Nagai, H. Wada, A.T. Naito, K. Matsuura, K. Iwanaga, T. Takahashi, M. Goto, Y. Mikami, N. Yasuda, H. Akazawa, A. Uezumi, S. Takeda, I. Komuro, Cardiac side population cells have a potential to migrate and differentiate into cardiomyocytes in vitro and in vivo, J. Cell Biol. 176 (3) (2007) 329–341.

[41] P.C.H. Hsieh, V.F.M. Segers, M.E. Davis, C. MacGillivray, J. Gannon, J.D. Molkentin, J. Robbins, R.T. Lee, Evidence from a genetic fate-mapping study that stem cells refresh adult mammalian cardiomyocytes after injury, Nat. Med. 13 (8) (2007) 970–974.

[42] K. Malliaras, Y. Zhang, J. Seinfeld, G. Galang, E. Tseliou, K. Cheng, B. Sun, M. Aminzadeh, E. Marban, Cardiomyocyte proliferation and progenitor cell recruitment underlie therapeutic regeneration after myocardial infarction in the adult mouse heart, EMBO Mol. Med. 5 (2) (2013) 191–209.

[43] S.E. Senyo, M.L. Steinhauser, C.L. Pizzimenti, V.K. Yang, L. Cai, M. Wang, T.-D. Wu, J.-L. Guerquin-Kern, C.P. Lechene, R.T. Lee, Mammalian heart renewal by pre-existing cardiomyocytes, Nature 493 (7432) (2013) 433–436.

[44] J.A. Thomson, Embryonic stem cell lines derived from human blastocysts, Science 282 (5391) (1998) 1145–1147.

[45] M. Talkhabi, N. Aghdami, H. Baharvand, Human cardiomyocyte generation from pluripotent stem cells: a state-of-art, Life Sci. 145 (2016) 98–113.
[46] C.C. Veerman, G. Kosmidis, C.L. Mummery, S. Casini, A.O. Verkerk, M. Bellin, Immaturity of human stem-cell-derived cardiomyocytes in culture: fatal flaw or soluble problem? Stem Cells Dev. 24 (9) (2015) 1035–1052.
[47] J.A. Robertson, Human embryonic stem cell research: ethical and legal issues, Nat. Rev. Genet. 2 (1) (2001) 74–78.
[48] J. Nussbaum, E. Minami, M.A. Laflamme, J.A.I. Virag, C.B. Ware, A. Masino, V. Muskheli, L. Pabon, H. Reinecke, C.E. Murry, Transplantation of undifferentiated murine embryonic stem cells in the heart: teratoma formation and immune response, FASEB J. 21 (7) (2007) 1345–1357.
[49] R.-J. Swijnenburg, M. Tanaka, H. Vogel, J. Baker, T. Kofidis, F. Gunawan, D.R. Lebl, A.D. Caffarelli, J.L. de Bruin, E.V. Fedoseyeva, R.C. Robbins, Embryonic stem cell immunogenicity increases upon differentiation after transplantation into ischemic myocardium, Circulation 112 (9 Suppl) (2005) I166–72.
[50] T. Nouspikel, Genetic instability in human embryonic stem cells: prospects and caveats, Future Oncol. 9 (6) (2013) 867–877.
[51] D. Ilic, C. Ogilvie, Concise review: human embryonic stem cells-what have we done? What are we doing? Where are we going? Stem Cells 35 (1) (2017) 17–25.
[52] J.-Y. Min, Y. Yang, K.L. Converso, L. Liu, Q. Huang, J.P. Morgan, Y.-F. Xiao, Transplantation of embryonic stem cells improves cardiac function in postinfarcted rats, J. Appl. Physiol. 92 (1) (2002) 288–296.
[53] A. Behfar, L.V. Zingman, D.M. Hodgson, J.-M. Rauzier, G.C. Kane, A. Terzic, M. Puceat, Stem cell differentiation requires a paracrine pathway in the heart, FASEB J. 16 (12) (2002) 1558–1566.
[54] O. Caspi, I. Huber, I. Kehat, M. Habib, G. Arbel, A. Gepstein, L. Yankelson, D. Aronson, R. Beyar, L. Gepstein, Transplantation of human embryonic stem cell-derived cardiomyocytes improves myocardial performance in infarcted rat hearts, J. Am. Coll. Cardiol. 50 (19) (2007) 1884–1893.
[55] G. Blin, D. Nury, S. Stefanovic, T. Neri, O. Guillevic, B. Brinon, V. Bellamy, C. Rucker-Martin, P. Barbry, A. Bel, P. Bruneval, C. Cowan, J. Pouly, S. Mitalipov, E. Gouadon, P. Binder, A. Hagege, M. Desnos, J.-F. Renaud, P. Menasche, M. Puceat, A purified population of multipotent cardiovascular progenitors derived from primate pluripotent stem cells engrafts in postmyocardial infarcted nonhuman primates, J. Clin. Invest. 120 (4) (2010) 1125–1139.
[56] Y. Shiba, S. Fernandes, W.-Z. Zhu, D. Filice, V. Muskheli, J. Kim, N.J. Palpant, J. Gantz, K.W. Moyes, H. Reinecke, B. van Biber, T. Dardas, J.L. Mignone, A. Izawa, R. Hanna, M. Viswanathan, J.D. Gold, M.I. Kotlikoff, N. Sarvazyan, M.W. Kay, C.E. Murry, M.A. Laflamme, Human ES-cell-derived cardiomyocytes electrically couple and suppress arrhythmias in injured hearts, Nature 489 (7415) (2012) 322–325.
[57] M.A. Laflamme, J. Gold, C. Xu, M. Hassanipour, E. Rosler, S. Police, V. Muskheli, C.E. Murry, Formation of human myocardium in the rat heart from human embryonic stem cells, Am. J. Pathol. 167 (3) (2005) 663–671.
[58] L.W. van Laake, R. Passier, J. Monshouwer-Kloots, A.J. Verkleij, D.J. Lips, C. Freund, K. den Ouden, D. Ward-van Oostwaard, J. Korving, L.G. Tertoolen, C.J. van Echteld, P.A. Doevendans, C.L. Mummery, Human embryonic stem cell-derived cardiomyocytes survive and mature in the mouse heart and transiently improve function after myocardial infarction, Stem Cell Res. 1 (1) (2007) 9–24.
[59] M.A. Laflamme, K.Y. Chen, A.V. Naumova, V. Muskheli, J.A. Fugate, S.K. Dupras, H. Reinecke, C. Xu, M. Hassanipour, S. Police, C. O'Sullivan, L. Collins, Y. Chen, E. Minami, E.A. Gill, S. Ueno, C. Yuan, J. Gold, C.E. Murry, Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts, Nat. Biotechnol. 25 (9) (2007) 1015–1024.
[60] Y. Yeghiazarians, M. Gaur, Y. Zhang, R.E. Sievers, C. Ritner, M. Prasad, A. Boyle, H.S. Bernstein, Myocardial improvement with human embryonic stem cell-derived cardiomyocytes enriched by p38MAPK inhibition, Cytotherapy 14 (2) (2012) 223–231.
[61] A. Tomescot, J. Leschik, V. Bellamy, G. Dubois, E. Messas, P. Bruneval, M. Desnos, A.A. Hagege, M. Amit, J. Itskovitz, P. Menasche, M. Puceat, Differentiation in vivo of cardiac committed human embryonic stem cells in postmyocardial infarcted rats, Stem Cells 25 (9) (2007) 2200–2205.
[62] J.J.H. Chong, X. Yang, C.W. Don, E. Minami, Y.-W. Liu, J.J. Weyers, W.M. Mahoney, B. van Biber, S.M. Cook, N.J. Palpant, J.A. Gantz, J.A. Fugate, V. Muskheli, G.M. Gough, K.W. Vogel, C.A. Astley, C.E. Hotchkiss, A. Baldessari, L. Pabon, H. Reinecke, E.A. Gill, V. Nelson, H.-P. Kiem, M.A. Laflamme, C.E. Murry, Human embryonic-stem-cell-derived cardiomyocytes regenerate non-human primate hearts, Nature 510 (7504) (2014) 273–277.
[63] C. Ménard, A.A. Hagège, O. Agbulut, M. Barro, M.C. Morichetti, C. Brasselet, A. Bel, E. Messas, A. Bissery, P. Bruneval, M. Desnos, M. Pucéat, P. Menasché, Transplantation of cardiac-committed mouse embryonic stem cells to infarcted sheep myocardium: a preclinical study, Lancet 366 (9490) (2005) 1005–1012.
[64] P. Menasche, V. Vanneaux, A. Hagege, A. Bel, B. Cholley, I. Cacciapuoti, A. Parouchev, N. Benhamouda, G. Tachdjian, L. Tosca, J.-H. Trouvin, J.-R. Fabreguettes, V. Bellamy, R. Guillemain, C. Suberbielle Boissel, E. Tartour, M. Desnos, J. Larghero, Human embryonic stem cell-derived cardiac progenitors for severe heart failure treatment: first clinical case report, Eur. Heart J. 36 (30) (2015) 2011–2017.
[65] K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors, Cell 126 (4) (2006) 663–676.
[66] K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, S. Yamanaka, Induction of pluripotent stem cells from adult human fibroblasts by defined factors, Cell 131 (5) (2007) 861–872.
[67] K. Pfannkuche, H. Liang, T. Hannes, J. Xi, A. Fatima, F. Nguemo, M. Matzkies, M. Wernig, R. Jaenisch, F. Pillekamp, M. Halbach, H. Schunkert, T. Saric, J. Hescheler, M. Reppel, Cardiac myocytes derived from murine reprogrammed fibroblasts: intact hormonal regulation, cardiac ion channel expression and development of contractility, Cell. Physiol. Biochem. 24 (1–2) (2009) 73–86.
[68] C.-W. Kong, F.G. Akar, R.A. Li, Translational potential of human embryonic and induced pluripotent stem cells for myocardial repair: insights from experimental models, Thromb. Haemost. 104 (1) (2010) 30–38.
[69] E. Poon, C.-W. Kong, R.A. Li, Human pluripotent stem cell-based approaches for myocardial repair: from the electrophysiological perspective, Mol. Pharm. 8 (5) (2011) 1495–1504.
[70] X. Yang, L. Pabon, C.E. Murry, Engineering adolescence: maturation of human pluripotent stem cell-derived cardiomyocytes, Circ. Res. 114 (3) (2014) 511–523.
[71] M.E. Hartman, D.-F. Dai, M.A. Laflamme, Human pluripotent stem cells: prospects and challenges as a source of cardiomyocytes for in vitro modeling and cell-based cardiac repair, Adv. Drug Deliv. Rev. 96 (2016) 3–17.
[72] A. Barbuti, P. Benzoni, G. Campostrini, P. Dell'Era, Human derived cardiomyocytes: a decade of knowledge after the discovery of induced pluripotent stem cells, Dev. Dyn. 245 (12) (2016) 1145–1158.
[73] G. Narazaki, H. Uosaki, M. Teranishi, K. Okita, B. Kim, S. Matsuoka, S. Yamanaka, J.K. Yamashita, Directed and systematic differentiation of cardiovascular cells from mouse induced pluripotent stem cells, Circulation 118 (5) (2008) 498–506.
[74] C. Mauritz, K. Schwanke, M. Reppel, S. Neef, K. Katsirntaki, L.S. Maier, F. Nguemo, S. Menke, M. Haustein, J. Hescheler, G. Hasenfuss, U. Martin, Generation of functional murine cardiac myocytes from induced pluripotent stem cells, Circulation 118 (5) (2008) 507–517.
[75] A. Kuzmenkin, H. Liang, G. Xu, K. Pfannkuche, H. Eichhorn, A. Fatima, H. Luo, T. Saric, M. Wernig, R. Jaenisch, J. Hescheler, Functional characterization of cardiomyocytes derived from murine induced pluripotent stem cells in vitro, FASEB J. 23 (12) (2009) 4168–4180.
[76] J. Zhang, G.F. Wilson, A.G. Soerens, C.H. Koonce, J. Yu, S.P. Palecek, J.A. Thomson, T.J. Kamp, Functional cardiomyocytes derived from human induced pluripotent stem cells, Circ. Res. 104 (4) (2009) e30–41.
[77] J. Ma, L. Guo, S.J. Fiene, B.D. Anson, J.A. Thomson, T.J. Kamp, K.L. Kolaja, B.J. Swanson, C.T. January, High purity human-induced pluripotent stem cell-derived cardiomyocytes: electrophysiological properties of action potentials and ionic currents, Am. J. Physiol. Heart Circ. Physiol. 301 (5) (2011) H2006–17.
[78] L. Zwi, O. Caspi, G. Arbel, I. Huber, A. Gepstein, I.-H. Park, L. Gepstein, Cardiomyocyte differentiation of human induced pluripotent stem cells, Circulation 120 (15) (2009) 1513–1523.
[79] P.W. Burridge, E.T. Zambidis, Highly efficient directed differentiation of human induced pluripotent stem cells into cardiomyocytes, Methods Mol. Biol. 997 (2013) 149–161.
[80] M. Kawamura, S. Miyagawa, K. Miki, A. Saito, S. Fukushima, T. Higuchi, T. Kawamura, T. Kuratani, T. Daimon, T. Shimizu, T. Okano, Y. Sawa, Feasibility, safety, and therapeutic efficacy of human induced pluripotent stem cell-derived cardiomyocyte sheets in a porcine ischemic cardiomyopathy model, Circulation 126 (11 Suppl 1) (2012) S29–37.
[81] L. Citro, S. Naidu, F. Hassan, M.L. Kuppusamy, P. Kuppusamy, M.G. Angelos, M. Khan, Comparison of human induced pluripotent stem-cell derived cardiomyocytes with human mesenchymal stem cells following acute myocardial infarction, PLoS One 9 (12) (2014), e116281.
[82] L. Zhang, J. Guo, P. Zhang, Q. Xiong, S.C. Wu, L. Xia, S.S. Roy, J. Tolar, T.D. O'Connell, M. Kyba, K. Liao, J. Zhang, Derivation and high engraftment of patient-specific cardiomyocyte sheet using induced pluripotent stem cells generated from adult cardiac fibroblast, Circ. Heart Fail. 8 (1) (2015) 156–166.
[83] L. Ye, Y.-H. Chang, Q. Xiong, P. Zhang, L. Zhang, P. Somasundaram, M. Lepley, C. Swingen, L. Su, J.S. Wendel, J. Guo, A. Jang, D. Rosenbush, L. Greder, J.R. Dutton, J. Zhang, T.J. Kamp, D.S. Kaufman, Y. Ge, J. Zhang, Cardiac repair in a porcine model of acute myocardial infarction with human induced pluripotent stem cell-derived cardiovascular cells, Cell Stem Cell 15 (6) (2014) 750–761.
[84] L. Carpenter, C. Carr, C.T. Yang, D.J. Stuckey, K. Clarke, S.M. Watt, Efficient differentiation of human induced pluripotent stem cells generates cardiac cells that provide protection following myocardial infarction in the rat, Stem Cells Dev. 21 (6) (2012) 977–986.
[85] C. Mauritz, A. Martens, S.V. Rojas, T. Schnick, C. Rathert, N. Schecker, S. Menke, S. Glage, R. Zweigerdt, A. Haverich, U. Martin, I. Kutschka, Induced pluripotent stem cell (iPSC)-derived Flk-1 progenitor cells engraft, differentiate, and improve heart function in a mouse model of acute myocardial infarction, Eur. Heart J. 32 (21) (2011) 2634–2641.
[86] A. Martens, G. Kensah, S. Rojas, A. Rotärmel, H. Baraki, A. Haverich, U. Martin, I. Gruh, I. Kutschka, Induced pluripotent stem cell (iPSC)-derived cardiomyocytes engraft and improve heart function in a mouse model of acute myocardial infarction, Thorac. Cardiovasc. Surg. 60 (S 01) (2012).
[87] V.K. Singh, N. Kumar, M. Kalsan, A. Saini, R. Chandra, Mechanism of induction: induced pluripotent stem cells (iPSCs), J. Stem Cells 10 (1) (2015) 43–62.
[88] T. Zhao, Z.-N. Zhang, Z. Rong, Y. Xu, Immunogenicity of induced pluripotent stem cells, Nature 474 (7350) (2011) 212–215.
[89] A.S. Boyd, N.P. Rodrigues, K.O. Lui, X. Fu, Y. Xu, Concise review: immune recognition of induced pluripotent stem cells, Stem Cells 30 (5) (2012) 797–803.
[90] S. Kaneko, S. Yamanaka, To be immunogenic, or not to be: that's the iPSC question, Cell Stem Cell 12 (4) (2013) 385–386.
[91] J.M. Polo, S. Liu, M.E. Figueroa, W. Kulalert, S. Eminli, K.Y. Tan, E. Apostolou, M. Stadtfeld, Y. Li, T. Shioda, S. Natesan, A.J. Wagers, A. Melnick, T. Evans, K. Hochedlinger, Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells, Nat. Biotechnol. 28 (8) (2010) 848–855.

[92] A. Doi, I.-H. Park, B. Wen, P. Murakami, M.J. Aryee, R. Irizarry, B. Herb, C. Ladd-Acosta, J. Rho, S. Loewer, J. Miller, T. Schlaeger, G.Q. Daley, A.P. Feinberg, Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts, Nat. Genet. 41 (12) (2009) 1350–1353.

[93] R. Lister, M. Pelizzola, Y.S. Kida, R.D. Hawkins, J.R. Nery, G. Hon, J. Antosiewicz-Bourget, R. O'Malley, R. Castanon, S. Klugman, M. Downes, R. Yu, R. Stewart, B. Ren, J.A. Thomson, R.M. Evans, J.R. Ecker, Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells, Nature 471 (7336) (2011) 68–73.

[94] A. Gore, Z. Li, H.-L. Fung, J.E. Young, S. Agarwal, J. Antosiewicz-Bourget, I. Canto, A. Giorgetti, M.A. Israel, E. Kiskinis, J.-H. Lee, Y.-H. Loh, P.D. Manos, N. Montserrat, A.D. Panopoulos, S. Ruiz, M.L. Wilbert, J. Yu, E.F. Kirkness, J.C. Izpisua Belmonte, D.J. Rossi, J.A. Thomson, K. Eggan, G.Q. Daley, L.S.B. Goldstein, K. Zhang, Somatic coding mutations in human induced pluripotent stem cells, Nature 471 (7336) (2011) 63–67.

[95] M. Yoshihara, Y. Hayashizaki, Y. Murakawa, Genomic instability of iPSCs: challenges towards their clinical applications, Stem Cell Rev. (2016).

[96] D. Kim, C.-H. Kim, J.-I. Moon, Y.-G. Chung, M.-Y. Chang, B.-S. Han, S. Ko, E. Yang, K.Y. Cha, R. Lanza, K.-S. Kim, Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins, Cell Stem Cell 4 (6) (2009) 472–476.

[97] H. Zhou, S. Wu, J.Y. Joo, S. Zhu, D.W. Han, T. Lin, S. Trauger, G. Bien, S. Yao, Y. Zhu, G. Siuzdak, H.R. Scholer, L. Duan, S. Ding, Generation of induced pluripotent stem cells using recombinant proteins, Cell Stem Cell 4 (5) (2009) 381–384.

[98] L. Warren, P.D. Manos, T. Ahfeldt, Y.-H. Loh, H. Li, F. Lau, W. Ebina, P.K. Mandal, Z.D. Smith, A. Meissner, G.Q. Daley, A.S. Brack, J.J. Collins, C. Cowan, T.M. Schlaeger, D.J. Rossi, Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA, Cell Stem Cell 7 (5) (2010) 618–630.

[99] N. Miyoshi, H. Ishii, H. Nagano, N. Haraguchi, D.L. Dewi, Y. Kano, S. Nishikawa, M. Tanemura, K. Mimori, F. Tanaka, T. Saito, J. Nishimura, I. Takemasa, T. Mizushima, M. Ikeda, H. Yamamoto, M. Sekimoto, Y. Doki, M. Mori, Reprogramming of mouse and human cells to pluripotency using mature microRNAs, Cell Stem Cell 8 (6) (2011) 633–638.

[100] F. Anokye-Danso, C.M. Trivedi, D. Juhr, M. Gupta, Z. Cui, Y. Tian, Y. Zhang, W. Yang, P.J. Gruber, J.A. Epstein, E.E. Morrisey, Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency, Cell Stem Cell 8 (4) (2011) 376–388.

[101] P. Hou, Y. Li, X. Zhang, C. Liu, J. Guan, H. Li, T. Zhao, J. Ye, W. Yang, K. Liu, J. Ge, J. Xu, Q. Zhang, Y. Zhao, H. Deng, Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds, Science (New York, N.Y.) 341 (6146) (2013) 651–654.

[102] P.J. Kang, J.-H. Moon, B.S. Yoon, S. Hyeon, E.K. Jun, G. Park, W. Yun, J. Park, M. Park, A. Kim, K.Y. Whang, G.Y. Koh, S. Oh, S. You, Reprogramming of mouse somatic cells into pluripotent stem-like cells using a combination of small molecules, Biomaterials 35 (26) (2014) 7336–7345.

[103] Y. Zhao, T. Zhao, J. Guan, X. Zhang, Y. Fu, J. Ye, J. Zhu, G. Meng, J. Ge, S. Yang, L. Cheng, Y. Du, C. Zhao, T. Wang, L. Su, W. Yang, H. Deng, A XEN-like state bridges somatic cells to pluripotency during chemical reprogramming, Cell 163 (7) (2015) 1678–1691.

[104] B.-E. Strauer, G. Steinhoff, 10 years of intracoronary and intramyocardial bone marrow stem cell therapy of the heart: from the methodological origin to clinical practice, J. Am. Coll. Cardiol. 58 (11) (2011) 1095–1104.

[105] B.E. Strauer, M. Brehm, T. Zeus, M. Kostering, A. Hernandez, R.V. Sorg, G. Kogler, P. Wernet, Repair of infarcted myocardium by autologous intracoronary mononuclear bone marrow cell transplantation in humans, Circulation 106 (15) (2002) 1913–1918.

[106] S.J. Szilvassy, M.J. Bass, G. van Zant, B. Grimes, Organ-selective homing defines engraftment kinetics of murine hematopoietic stem cells and is compromised by ex vivo expansion, Blood 93 (5) (1999) 1557–1566.

[107] T. Kamota, T.-S. Li, N. Morikage, M. Murakami, M. Ohshima, M. Kubo, T. Kobayashi, A. Mikamo, Y. Ikeda, M. Matsuzaki, K. Hamano, Ischemic pre-conditioning enhances the mobilization and recruitment of bone marrow stem cells to protect against ischemia/reperfusion injury in the late phase, J. Am. Coll. Cardiol. 53 (19) (2009) 1814–1822.

[108] I. Elmadbouh, H.K. Haider, S. Jiang, N.M. Idris, G. Lu, M. Ashraf, Ex vivo delivered stromal cell-derived factor-1alpha promotes stem cell homing and induces angiomyogenesis in the infarcted myocardium, J. Mol. Cell. Cardiol. 42 (4) (2007) 792–803.

[109] Q. Jiang, P. Song, E. Wang, J. Li, S. Hu, H. Zhang, Remote ischemic postconditioning enhances cell retention in the myocardium after intravenous administration of bone marrow mesenchymal stromal cells, J. Mol. Cell. Cardiol. 56 (2013) 1–7.

[110] M. Sussman, Cardiovascular biology. Hearts and bones, Nature 410 (6829) (2001) 640–641.

[111] F.S. Loffredo, M.L. Steinhauser, J. Gannon, R.T. Lee, Bone marrow-derived cell therapy stimulates endogenous cardiomyocyte progenitors and promotes cardiac repair, Cell Stem Cell 8 (4) (2011) 389–398.

[112] S. Fuchs, R. Baffour, Y.F. Zhou, M. Shou, A. Pierre, F.O. Tio, N.J. Weissman, M.B. Leon, S.E. Epstein, R. Kornowski, Transendocardial delivery of autologous bone marrow enhances collateral perfusion and regional function in pigs with chronic experimental myocardial ischemia, J. Am. Coll. Cardiol. 37 (6) (2001) 1726–1732.

[113] M. Galinanes, M. Loubani, J. Davies, D. Chin, J. Pasi, P.R. Bell, Autotransplantation of unmanipulated bone marrow into scarred myocardium is safe and enhances cardiac function in humans, Cell Transplant. 13 (1) (2004) 7–13.

[114] D. Furlani, M. Ugurlucan, L. Ong, K. Bieback, E. Pittermann, I. Westien, W. Wang, C. Yerebakan, W. Li, R. Gaebel, R.-K. Li, B. Vollmar, G. Steinhoff, N. Ma, Is the intravascular administration of mesenchymal stem cells safe? Mesenchymal stem cells and intravital microscopy, Microvasc. Res. 77 (3) (2009) 370–376.

[115] L.M. de La Fuente, S.H. Stertzer, J. Argentieri, E. Peñaloza, J. Miano, B. Koziner, C. Bilos, P.A. Altman, Transendocardial autologous bone marrow in chronic myocardial infarction using a helical needle catheter: 1-year follow-up in an open-label, nonrandomized, single-center pilot study (the TABMMI study), Am. Heart J. 154 (1) (2007) (79.e1-7).

[116] B. Trachtenberg, D.L. Velazquez, A.R. Williams, I. McNiece, J. Fishman, K. Nguyen, D. Rouy, P. Altman, R. Schwarz, A. Mendizabal, B. Oskouei, J. Byrnes, V. Soto, M. Tracy, J.P. Zambrano, A.W. Heldman, J.M. Hare, Rationale and design of the transendocardial injection of autologous human cells (bone marrow or mesenchymal) in chronic ischemic left ventricular dysfunction and heart failure secondary to myocardial infarction (TAC-HFT) trial: a randomized, double-blind, placebo-controlled study of safety and efficacy, Am. Heart J. 161 (3) (2011) 487–493.

[117] A. Behfar, J.-P. Latere, J. Bartunek, C. Homsy, D. Daro, R.J. Crespo-Diaz, P.G. Stalboerger, V. Steenwinckel, A. Seron, M.M. Redfield, A. Terzic, Optimized delivery system achieves enhanced endomyocardial stem cell retention, Circ. Cardiovasc. Interv. 6 (6) (2013) 710–718.

[118] E.C. Perin, H.F.R. Dohmann, R. Borojevic, S.A. Silva, A.L.S. Sousa, C.T. Mesquita, M.I.D. Rossi, A.C. Carvalho, H.S. Dutra, H.J.F. Dohmann, G.V. Silva, L. Belem, R. Vivacqua, F.O.D. Rangel, R. Esporcatte, Y.J. Geng, W.K. Vaughn, J.A.R. Assad, E.T. Mesquita, J.T. Willerson, Transendocardial, autologous bone marrow cell transplantation for severe, chronic ischemic heart failure, Circulation 107 (18) (2003) 2294–2302.

[119] D.W. Losordo, R.A. Schatz, C.J. White, J.E. Udelson, V. Veereshwarayya, M. Durgin, K.K. Poh, R. Weinstein, M. Kearney, M. Chaudhry, A. Burg, L. Eaton, L. Heyd, T. Thorne, L. Shturman, P. Hoffmeister, K. Story, V. Zak, D. Dowling, J.H. Traverse, R.E. Olson, J. Flanagan, D. Sodano, T. Murayama, A. Kawamoto, K.F. Kusano, J. Wollins, F. Welt, P. Shah, P. Soukas, T. Asahara, T.D. Henry, Intramyocardial transplantation of autologous CD34+ stem cells for intractable angina: a phase I/IIa double-blind, randomized controlled trial, Circulation 115 (25) (2007) 3165–3172.

[120] N. Dib, R.E. Michler, F.D. Pagani, S. Wright, D.J. Kereiakes, R. Lengerich, P. Binkley, D. Buchele, I. Anand, C. Swingen, M.F. Di Carli, J.D. Thomas, W.A. Jaber, S.R. Opie, A. Campbell, P. McCarthy, M. Yeager, V. Dilsizian, B.P. Griffith, R. Korn, S.K. Kreuger, M. Ghazoul, W.R. MacLellan, G. Fonarow, H.J. Eisen, J. Dinsmore, E. Diethrich, Safety and feasibility of autologous myoblast transplantation in patients with ischemic cardiomyopathy: four-year follow-up, Circulation 112 (12) (2005) 1748–1755.

[121] C. Stamm, H.-D. Kleine, Y.-H. Choi, S. Dunkelmann, J.-A. Lauffs, B. Lorenzen, A. David, A. Liebold, C. Nienaber, D. Zurakowski, M. Freund, G. Steinhoff, Intramyocardial delivery of CD133+ bone marrow cells and coronary artery bypass grafting for chronic ischemic heart disease: safety and efficacy studies, J. Thorac. Cardiovasc. Surg. 133 (3) (2007) 717–725.

[122] H.M. Klein, A. Ghodsizad, R. Marktanner, L. Poll, T. Voelkel, M.R. Mohammad Hasani, C. Piechaczek, N. Feifel, M. Stockschlaeder, E.R. Burchardt, B.J. Kar, I. Gregoric, E. Gams, Intramyocardial implantation of CD133+ stem cells improved cardiac function without bypass surgery, Heart Surg. Forum 10 (1) (2007) E66–9.

[123] G. Pompilio, G. Steinhoff, A. Liebold, M. Pesce, F. Alamanni, M.C. Capogrossi, P. Biglioli, Direct minimally invasive intramyocardial injection of bone marrow-derived AC133+ stem cells in patients with refractory ischemia: preliminary results, Thorac. Cardiovasc. Surg. 56 (2) (2008) 71–76.

[124] A. Bel, V. Planat-Bernard, A. Saito, L. Bonnevie, V. Bellamy, L. Sabbah, L. Bellabas, B. Brinon, V. Vanneaux, P. Pradeau, S. Peyrard, J. Larghero, J. Pouly, P. Binder, S. Garcia, T. Shimizu, Y. Sawa, T. Okano, P. Bruneval, M. Desnos, A.A. Hagège, L. Casteilla, M. Pucéat, P. Menasché, Composite cell sheets: a further step toward safe and effective myocardial regeneration by cardiac progenitors derived from embryonic stem cells, Circulation 122 (11 Suppl) (2010) S118–23.

[125] H. Hamdi, V. Planat-Benard, A. Bel, H. Neamatalla, L. Saccenti, D. Calderon, V. Bellamy, M. Bon, M.-C. Perrier, C. Mandet, P. Bruneval, L. Casteilla, A.A. Hagège, M. Pucéat, O. Agbulut, P. Menasché, Long-term functional benefits of epicardial patches as cell carriers, Cell Transplant. 23 (1) (2014) 87–96.

[126] H. Hamdi, V. Planat-Benard, A. Bel, E. Puymirat, R. Geha, L. Pidial, H. Nematalla, V. Bellamy, P. Bouaziz, S. Peyrard, L. Casteilla, P. Bruneval, A.A. Hagège, O. Agbulut, P. Menasché, Epicardial adipose stem cell sheets results in greater post-infarction survival than intramyocardial injections, Cardiovasc. Res. 91 (3) (2011) 483–491.

[127] C.A. Thompson, B.A. Nasseri, J. Makower, S. Houser, M. McGarry, T. Lamson, I. Pomerantseva, J.Y. Chang, H.K. Gold, J.P. Vacanti, S.N. Oesterle, Percutaneous transvenous cellular cardiomyoplasty. A novel nonsurgical approach for myocardial cell transplantation, J. Am. Coll. Cardiol. 41 (11) (2003) 1964–1971.

[128] J.C. George, J. Goldberg, M. Joseph, N. Abdulhameed, J. Crist, H. Das, V.J. Pompili, Transvenous intramyocardial cellular delivery increases retention in comparison to intracoronary delivery in a porcine model of acute myocardial infarction, J. Interv. Cardiol. 21 (5) (2008) 424–431.

[129] X. Wang, M.N. Jameel, Q. Li, A. Mansoor, X. Qiang, C. Swingen, C. Panetta, J. Zhang, Stem cells for myocardial repair with use of a transarterial catheter, Circulation 120 (11 Suppl) (2009) S238–46.

[130] S. Medicetty, D. Wiktor, N. Lehman, A. Raber, Z.B. Popovic, R. Deans, A.E. Ting, M.S. Penn, Percutaneous adventitial delivery of allogeneic bone marrow-derived stem cells via infarct-related artery improves long-term ventricular function in acute myocardial infarction, Cell Transplant. 21 (6) (2012) 1109–1120.

[131] T. Siminiak, P. Burchardt, M. Kurpisz, Postinfarction heart failure: surgical and trans-coronary-venous transplantation of autologous myoblasts, Nat. Clin. Pract. Cardiovasc. Med. 3 (Suppl. 1) (2006) S46–51.

[132] A.N. Patel, S. Mittal, G. Turan, A.A. Winters, T.D. Henry, H. Ince, N. Trehan, REVIVE trial: retrograde delivery of autologous bone marrow in patients with heart failure, Stem Cells Transl. Med. 4 (9) (2015) 1021–1027.

[133] J. Tuma, R. Fernández-Viña, A. Carrasco, J. Castillo, C. Cruz, A. Carrillo, J. Ercilla, C. Yarleque, J. Cunza, T.D. Henry, A.N. Patel, Safety and feasibility of percutaneous

retrograde coronary sinus delivery of autologous bone marrow mononuclear cell transplantation in patients with chronic refractory angina, J. Transl. Med. 9 (2011) 183.
[134] L. Zakharova, H. Nural-Guvener, L. Feehery, S. Popovic, J. Nimlos, M.A. Gaballa, Retrograde coronary vein infusion of cardiac explant-derived c-Kit+ cells improves function in ischemic heart failure, J Heart Lung Transplant 33 (6) (2014) 644–653.
[135] N. Dib, H. Khawaja, S. Varner, M. McCarthy, A. Campbell, Cell therapy for cardiovascular disease: a comparison of methods of delivery, J. Cardiovasc. Transl. Res. 4 (2) (2011) 177–181.
[136] S.J. Hong, D. Hou, T.J. Brinton, B. Johnstone, D. Feng, P. Rogers, W.F. Fearon, P. Yock, K.L. March, Intracoronary and retrograde coronary venous myocardial delivery of adipose-derived stem cells in swine infarction lead to transient myocardial trapping with predominant pulmonary redistribution, Catheter. Cardiovasc. Interv. 83 (1) (2014) E17–25.
[137] M.E. Halkos, Z.-Q. Zhao, F. Kerendi, N.-P. Wang, R. Jiang, L.S. Schmarkey, B.J. Martin, A.A. Quyyumi, W.L. Few, H. Kin, R.A. Guyton, J. Vinten-Johansen, Intravenous infusion of mesenchymal stem cells enhances regional perfusion and improves ventricular function in a porcine model of myocardial infarction, Basic Res. Cardiol. 103 (6) (2008) 525–536.
[138] J.M. Hare, J.H. Traverse, T.D. Henry, N. Dib, R.K. Strumpf, S.P. Schulman, G. Gerstenblith, A.N. DeMaria, A.E. Denktas, R.S. Gammon, J.B. Hermiller Jr., M.A. Reisman, G.L. Schaer, W. Sherman, A randomized, double-blind, placebo-controlled, dose-escalation study of intravenous adult human mesenchymal stem cells (prochymal) after acute myocardial infarction, J. Am. Coll. Cardiol. 54 (24) (2009) 2277–2286.
[139] W. Wang, Q. Jiang, H. Zhang, P. Jin, X. Yuan, Y. Wei, S. Hu, Intravenous administration of bone marrow mesenchymal stromal cells is safe for the lung in a chronic myocardial infarction model, Regen. Med. 6 (2) (2011) 179–190.
[140] F.J. Giordano, Oxygen, oxidative stress, hypoxia, and heart failure, J. Clin. Invest. 115 (3) (2005) 500–508.
[141] B.A. Rose, T. Yokota, V. Chintalgattu, S. Ren, L. Iruela-Arispe, A.Y. Khakoo, S. Minamisawa, Y. Wang, Cardiac myocyte p38α kinase regulates angiogenesis via myocyte-endothelial cell cross-talk during stress-induced remodeling in heart, J. Biol. Chem. (2017).
[142] P. Madeddu, C. Emanueli, Switching on reparative angiogenesis: essential role of the vascular erythropoietin receptor, Circ. Res. 100 (5) (2007) 599–601.
[143] M. Crisan, M. Corselli, W.C.W. Chen, B. Péault, Perivascular cells for regenerative medicine, J. Cell. Mol. Med. 16 (12) (2012) 2851–2860.
[144] R.R. Makkar, R.R. Smith, K. Cheng, K. Malliaras, L.E.J. Thomson, D. Berman, L.S.C. Czer, L. Marbán, A. Mendizabal, P.V. Johnston, S.D. Russell, K.H. Schuleri, A.C. Lardo, G. Gerstenblith, E. Marbán, Intracoronary cardiosphere-derived cells for heart regeneration after myocardial infarction (CADUCEUS): a prospective, randomised phase 1 trial, Lancet 379 (9819) (2012) 895–904.
[145] W. Qiao, W. Wang, E. Laurenti, A.L. Turinsky, S.J. Wodak, G.D. Bader, J.E. Dick, P.W. Zandstra, Intercellular network structure and regulatory motifs in the human hematopoietic system, Mol. Syst. Biol. 10 (2014) 741.
[146] R. Rohban, B. Prietl, T.R. Pieber, Crosstalk between stem and progenitor cellular mediators with special emphasis on vasculogenesis, Transfus. Med. Hemother. 44 (3) (2017) 174–182.
[147] R. Kramann, B.D. Humphreys, Kidney pericytes: roles in regeneration and fibrosis, Semin. Nephrol. 34 (4) (2014) 374–383.
[148] D.J. Ceradini, A.R. Kulkarni, M.J. Callaghan, O.M. Tepper, N. Bastidas, M.E. Kleinman, J.M. Capla, R.D. Galiano, J.P. Levine, G.C. Gurtner, Progenitor cell trafficking is regulated by hypoxic gradients through HIF-1 induction of SDF-1, Nat. Med. 10 (8) (2004) 858–864.
[149] H.J. Lee, J.M. Ryu, Y.H. Jung, S.Y. Oh, S.-J. Lee, H.J. Han, Novel pathway for hypoxia-induced proliferation and migration in human mesenchymal stem cells: involvement of HIF-1α, FASN, and mTORC1, Stem Cells 33 (7) (2015) 2182–2195.
[150] N. Urao, M. Ushio-Fukai, Redox regulation of stem/progenitor cells and bone marrow niche, Free Radic. Biol. Med. 54 (2013) 26–39.
[151] I.B. Lobov, S. Rao, T.J. Carroll, J.E. Vallance, M. Ito, J.K. Ondr, S. Kurup, D.A. Glass, M.S. Patel, W. Shu, E.E. Morrisey, A.P. McMahon, G. Karsenty, R.A. Lang, WNT7b mediates macrophage-induced programmed cell death in patterning of the vasculature, Nature 437 (7057) (2005) 417–421.
[152] H.E. Moon, K. Byun, H.W. Park, J.H. Kim, J. Hur, J.S. Park, J.K. Jun, H.-S. Kim, S.L. Paek, I.K. Kim, J.H. Hwang, J.W. Kim, D.G. Kim, Y.C. Sung, G.-Y. Koh, C.W. Song, B. Lee, S.H. Paek, COMP-Ang1 potentiates EPC treatment of ischemic brain injury by enhancing angiogenesis through activating AKT-mTOR pathway and promoting vascular migration through activating Tie2-FAK pathway, Exp. Neurobiol. 24 (1) (2015) 55–70.
[153] V. Plaks, T. Birnberg, T. Berkutzki, S. Sela, A. BenYashar, V. Kalchenko, G. Mor, E. Keshet, N. Dekel, M. Neeman, S. Jung, Uterine DCs are crucial for decidua formation during embryo implantation in mice, J. Clin. Invest. 118 (12) (2008) 3954–3965.
[154] J.W. Pollard, Trophic macrophages in development and disease, Nat. Rev. Immunol. 9 (4) (2009) 259–270.
[155] M.I.F.J. Oerlemans, S. Koudstaal, S.A. Chamuleau, D.P. de Kleijn, P.A. Doevendans, J.P.G. Sluijter, Targeting cell death in the reperfused heart: pharmacological approaches for cardioprotection, Int. J. Cardiol. 165 (3) (2013) 410–422.
[156] F. Arslan, M.B. Smeets, L.A.J. O'Neill, B. Keogh, P. McGuirk, L. Timmers, C. Tersteeg, I.E. Hoefer, P.A. Doevendans, G. Pasterkamp, D.P.V. de Kleijn, Myocardial ischemia/reperfusion injury is mediated by leukocytic toll-like receptor-2 and reduced by systemic administration of a novel anti-toll-like receptor-2 antibody, Circulation 121 (1) (2010) 80–90.
[157] S. Epelman, K.J. Lavine, A.E. Beaudin, D.K. Sojka, J.A. Carrero, B. Calderon, T. Brija, E.L. Gautier, S. Ivanov, A.T. Satpathy, J.D. Schilling, R. Schwendener, I. Sergin, B. Razani, E.C. Forsberg, W.M. Yokoyama, E.R. Unanue, M. Colonna, G.J. Randolph, D.L. Mann, Embryonic and adult-derived resident cardiac macrophages are maintained through distinct mechanisms at steady state and during inflammation, Immunity 40 (1) (2014) 91–104.
[158] J.A. Hamilton, Colony-stimulating factors in inflammation and autoimmunity, Nat. Rev. Immunol. 8 (7) (2008) 533–544.
[159] A. Christ, S. Bekkering, E. Latz, N.P. Riksen, Long-term activation of the innate immune system in atherosclerosis, Semin. Immunol. 28 (4) (2016) 384–393.
[160] D. Fairweather, S. Frisancho-Kiss, S.A. Yusung, M.A. Barrett, S.E. Davis, R.A. Steele, S.J.L. Gatewood, N.R. Rose, IL-12 protects against coxsackievirus B3-induced myocarditis by increasing IFN-gamma and macrophage and neutrophil populations in the heart, J. Immunol. 174 (1) (2005) 261–269.
[161] J.W. Godwin, A.R. Pinto, N.A. Rosenthal, Chasing the recipe for a pro-regenerative immune system, Semin. Cell Dev. Biol. 61 (2017) 71–79.
[162] O. Delarosa, W. Dalemans, E. Lombardo, Toll-like receptors as modulators of mesenchymal stem cells, Front. Immunol. 3 (2012) 182.
[163] A. Yilmaz, B. Dietel, I. Cicha, K. Schubert, R. Hausmann, W.G. Daniel, C.D. Garlichs, C. Stumpf, Emergence of dendritic cells in the myocardium after acute myocardial infarction - implications for inflammatory myocardial damage, Int. J. Biomed. Sci. IJBS 6 (1) (2010) 27–36.
[164] C. Mauri, A. Bosma, Immune regulatory function of B cells, Annu. Rev. Immunol. 30 (2012) 221–241.
[165] M. Franquesa, F.K. Mensah, R. Huizinga, T. Strini, L. Boon, E. Lombardo, O. DelaRosa, J.D. Laman, J.M. Grinyó, W. Weimar, M.G.H. Betjes, C.C. Baan, M.J. Hoogduijn, Human adipose tissue-derived mesenchymal stem cells abrogate plasmablast formation and induce regulatory B cells independently of T helper cells, Stem Cells 33 (3) (2015) 880–891.
[166] X. Yan, A. Anzai, Y. Katsumata, T. Matsuhashi, K. Ito, J. Endo, T. Yamamoto, A. Takeshima, K. Shinmura, W. Shen, K. Fukuda, M. Sano, Temporal dynamics of cardiac immune cell accumulation following acute myocardial infarction, J. Mol. Cell. Cardiol. 62 (2013) 24–35.
[167] K.Y. King, M.A. Goodell, Inflammatory modulation of HSCs: viewing the HSC as a foundation for the immune response, Nat. Rev. Immunol. 11 (10) (2011) 685–692.
[168] B.L. Esplin, T. Shimazu, R.S. Welner, K.P. Garrett, L. Nie, Q. Zhang, M.B. Humphrey, Q. Yang, L.A. Borghesi, P.W. Kincade, Chronic exposure to a TLR ligand injures hematopoietic stem cells, J. Immunol. 186 (9) (2011) 5367–5375.
[169] I.G. Winkler, N.A. Sims, A.R. Pettit, V. Barbier, B. Nowlan, F. Helwani, I.J. Poulton, N. van Rooijen, K.A. Alexander, L.J. Raggatt, J.-P. Lévesque, Bone marrow macrophages maintain hematopoietic stem cell (HSC) niches and their depletion mobilizes HSCs, Blood 116 (23) (2010) 4815–4828.
[170] M. Ii, H. Nishimura, A. Iwakura, A. Wecker, E. Eaton, T. Asahara, D.W. Losordo, Endothelial progenitor cells are rapidly recruited to myocardium and mediate protective effect of ischemic preconditioning via "imported" nitric oxide synthase activity, Circulation 111 (9) (2005) 1114–1120.
[171] C. Bogdan, Nitric oxide and the immune response, Nat. Immunol. 2 (10) (2001) 907–916.
[172] C.C. Bleul, R.C. Fuhlbrigge, J.M. Casasnovas, A. Aiuti, T.A. Springer, A highly efficacious lymphocyte chemoattractant, stromal cell-derived factor 1 (SDF-1), J. Exp. Med. 184 (3) (1996) 1101–1109.
[173] X. Wan, W. Xia, Y. Gendoo, W. Chen, W. Sun, D. Sun, C. Cao, Upregulation of stromal cell-derived factor 1 (SDF-1) is associated with macrophage infiltration in renal ischemia-reperfusion injury, PLoS One 9 (12) (2014), e114564.
[174] T.J. Smith, Insulin-like growth factor-I regulation of immune function: a potential therapeutic target in autoimmune diseases? Pharmacol. Rev. 62 (2) (2010) 199–236.
[175] M. Matysiak, W. Orlowski, M. Fortak-Michalska, A. Jurewicz, K. Selmaj, Immunoregulatory function of bone marrow mesenchymal stem cells in EAE depends on their differentiation state and secretion of PGE2, J. Neuroimmunol. 233 (1–2) (2011) 106–111.
[176] J. Su, X. Chen, Y. Huang, W. Li, J. Li, K. Cao, G. Cao, L. Zhang, F. Li, A.I. Roberts, H. Kang, P. Yu, G. Ren, W. Ji, Y. Wang, Y. Shi, Phylogenetic distinction of iNOS and IDO function in mesenchymal stem cell-mediated immunosuppression in mammalian species, Cell Death Differ. 21 (3) (2014) 388–396.
[177] Z. Selmani, A. Naji, I. Zidi, B. Favier, E. Gaiffe, L. Obert, C. Borg, P. Saas, P. Tiberghien, N. Rouas-Freiss, E.D. Carosella, F. Deschaseaux, Human leukocyte antigen-G5 secretion by human mesenchymal stem cells is required to suppress T lymphocyte and natural killer function and to induce CD4+CD25highFOXP3+ regulatory T cells, Stem Cells 26 (1) (2008) 212–222.
[178] A. Corcione, F. Benvenuto, E. Ferretti, D. Giunti, V. Cappiello, F. Cazzanti, M. Risso, F. Gualandi, G.L. Mancardi, V. Pistoia, A. Uccelli, Human mesenchymal stem cells modulate B-cell functions, Blood 107 (1) (2006) 367–372.
[179] M. Rafei, J. Hsieh, S. Fortier, M. Li, S. Yuan, E. Birman, K. Forner, M.-N. Boivin, K. Doody, M. Tremblay, B. Annabi, J. Galipeau, Mesenchymal stromal cell-derived CCL2 suppresses plasma cell immunoglobulin production via STAT3 inactivation and PAX5 induction, Blood 112 (13) (2008) 4991–4998.
[180] F. Schena, C. Gambini, A. Gregorio, M. Mosconi, D. Reverberi, M. Gattorno, S. Casazza, A. Uccelli, L. Moretta, A. Martini, E. Traggiai, Interferon-γ-dependent inhibition of B cell activation by bone marrow-derived mesenchymal stem cells in a murine model of systemic lupus erythematosus, Arthritis Rheum. 62 (9) (2010) 2776–2786.
[181] G.M. Spaggiari, H. Abdelrazik, F. Becchetti, L. Moretta, MSCs inhibit monocyte-derived DC maturation and function by selectively interfering with the generation of immature DCs: central role of MSC-derived prostaglandin E2, Blood 113 (26) (2009) 6576–6583.
[182] X. Liu, X. Qu, Y. Chen, L. Liao, K. Cheng, C. Shao, M. Zenke, A. Keating, R.C.H. Zhao, Mesenchymal stem/stromal cells induce the generation of novel IL-10-dependent regulatory dendritic cells by SOCS3 activation, J. Immunol. 189 (3) (2012) 1182–1192.

[183] P.A. Sotiropoulou, S.A. Perez, A.D. Gritzapis, C.N. Baxevanis, M. Papamichail, Interactions between human mesenchymal stem cells and natural killer cells, Stem Cells 24 (1) (2006) 74–85.

[184] G.M. Spaggiari, A. Capobianco, H. Abdelrazik, F. Becchetti, M.C. Mingari, L. Moretta, Mesenchymal stem cells inhibit natural killer-cell proliferation, cytotoxicity, and cytokine production: role of indoleamine 2,3-dioxygenase and prostaglandin E2, Blood 111 (3) (2008) 1327–1333.

[185] I. Rashedi, A. Gómez-Aristizábal, X.-H. Wang, S. Viswanathan, A. Keating, TLR3 or TLR4 activation enhances mesenchymal stromal cell-mediated Treg induction via notch signaling, Stem Cells 35 (1) (2017) 265–275.

[186] J.E. Cole, T.J. Navin, A.J. Cross, M.E. Goddard, L. Alexopoulou, A.T. Mitra, A.H. Davies, R.A. Flavell, M. Feldmann, C. Monaco, Unexpected protective role for toll-like receptor 3 in the arterial wall, Proc. Natl. Acad. Sci. U. S. A. 108 (6) (2011) 2372–2377.

[187] L.-n. Pan, W. Zhu, Y. Li, X.-l. Xu, L.-j. Guo, Q. Lu, J. Wang, Astrocytic toll-like receptor 3 is associated with ischemic preconditioning-induced protection against brain ischemia in rodents, PLoS One 9 (6) (2014), e99526.

[188] Y. Li, X.-l. Xu, D. Zhao, L.-n. Pan, C.-W. Huang, L.-j. Guo, Q. Lu, J. Wang, TLR3 ligand poly IC attenuates reactive astrogliosis and improves recovery of rats after focal cerebral ischemia, CNS Neurosci. Ther. 21 (11) (2015) 905–913.

[189] K. Németh, A. Leelahavanichkul, P.S.T. Yuen, B. Mayer, A. Parmelee, K. Doi, P.G. Robey, K. Leelahavanichkul, B.H. Koller, J.M. Brown, X. Hu, I. Jelinek, R.A. Star, E. Mezey, Bone marrow stromal cells attenuate sepsis via prostaglandin E(2)-dependent reprogramming of host macrophages to increase their interleukin-10 production, Nat. Med. 15 (1) (2009) 42–49.

[190] S. Carlson, J. Trial, C. Soeller, M.L. Entman, Cardiac mesenchymal stem cells contribute to scar formation after myocardial infarction, Cardiovasc. Res. 91 (1) (2011) 99–107.

[191] K.A. Cieslik, G.E. Taffet, S. Carlson, J. Hermosillo, J. Trial, M.L. Entman, Immune-inflammatory dysregulation modulates the incidence of progressive fibrosis and diastolic stiffness in the aging heart, J. Mol. Cell. Cardiol. 50 (1) (2011) 248–256.

[192] M. Miragoli, N. Salvarani, S. Rohr, Myofibroblasts induce ectopic activity in cardiac tissue, Circ. Res. 101 (8) (2007) 755–758.

[193] C. Toma, M.F. Pittenger, K.S. Cahill, B.J. Byrne, P.D. Kessler, Human mesenchymal stem cells differentiate to a cardiomyocyte phenotype in the adult murine heart, Circulation 105 (1) (2002) 93–98.

[194] M. Rota, J. Kajstura, T. Hosoda, C. Bearzi, S. Vitale, G. Esposito, G. Iaffaldano, M.E. Padin-Iruegas, A. Gonzalez, R. Rizzi, N. Small, J. Muraski, R. Alvarez, X. Chen, K. Urbanek, R. Bolli, S.R. Houser, A. Leri, M.A. Sussman, P. Anversa, Bone marrow cells adopt the cardiomyogenic fate in vivo, Proc. Natl. Acad. Sci. U. S. A. 104 (45) (2007) 17783–17788.

[195] J. Kajstura, M. Rota, B. Whang, S. Cascapera, T. Hosoda, C. Bearzi, D. Nurzynska, H. Kasahara, E. Zias, M. Bonafe, B. Nadal-Ginard, D. Torella, A. Nascimbene, F. Quaini, K. Urbanek, A. Leri, P. Anversa, Bone marrow cells differentiate in cardiac cell lineages after infarction independently of cell fusion, Circ. Res. 96 (1) (2005) 127–137.

[196] S. Murasawa, A. Kawamoto, M. Horii, S. Nakamori, T. Asahara, Niche-dependent translineage commitment of endothelial progenitor cells, not cell fusion in general, into myocardial lineage cells, Arterioscler. Thromb. Vasc. Biol. 25 (7) (2005) 1388–1394.

[197] Y.-s. Yoon, A. Wecker, L. Heyd, J.-S. Park, T. Tkebuchava, K. Kusano, A. Hanley, H. Scadova, G. Qin, D.-H. Cha, K.L. Johnson, R. Aikawa, T. Asahara, D.W. Losordo, Clonally expanded novel multipotent stem cells from human bone marrow regenerate myocardium after myocardial infarction, J. Clin. Invest. 115 (2) (2005) 326–338.

[198] C.E. Murry, M.H. Soonpaa, H. Reinecke, H. Nakajima, H.O. Nakajima, M. Rubart, K.B.S. Pasumarthi, J.I. Virag, S.H. Bartelmez, V. Poppa, G. Bradford, J.D. Dowell, D.A. Williams, L.J. Field, Haematopoietic stem cells do not transdifferentiate into cardiac myocytes in myocardial infarcts, Nature 428 (6983) (2004) 664–668.

[199] L.B. Balsam, A.J. Wagers, J.L. Christensen, T. Kofidis, I.L. Weissman, R.C. Robbins, Haematopoietic stem cells adopt mature haematopoietic fates in ischaemic myocardium, Nature 428 (6983) (2004) 668–673.

[200] T. Yoshioka, N. Ageyama, H. Shibata, T. Yasu, Y. Misawa, K. Takeuchi, K. Matsui, K. Yamamoto, K. Terao, K. Shimada, U. Ikeda, K. Ozawa, Y. Hanazono, Repair of infarcted myocardium mediated by transplanted bone marrow-derived CD34+ stem cells in a nonhuman primate model, Stem Cells 23 (3) (2005) 355–364.

[201] J.M. Nygren, S. Jovinge, M. Breitbach, P. Sawen, W. Roll, J. Hescheler, J. Taneera, B.K. Fleischmann, S.E.W. Jacobsen, Bone marrow-derived hematopoietic cells generate cardiomyocytes at a low frequency through cell fusion, but not transdifferentiation, Nat. Med. 10 (5) (2004) 494–501.

[202] J. Andrade, J.T. Lam, M. Zamora, C. Huang, D. Franco, N. Sevilla, P.J. Gruber, J.T. Lu, P. Ruiz-Lozano, Predominant fusion of bone marrow-derived cardiomyocytes, Cardiovasc. Res. 68 (3) (2005) 387–393.

[203] M. Alvarez-Dolado, R. Pardal, J.M. Garcia-Verdugo, J.R. Fike, H.O. Lee, K. Pfeffer, C. Lois, S.J. Morrison, A. Alvarez-Buylla, Fusion of bone-marrow-derived cells with Purkinje neurons, cardiomyocytes and hepatocytes, Nature 425 (6961) (2003) 968–973.

[204] S. Zhang, D. Wang, Z. Estrov, S. Raj, J.T. Willerson, E.T.H. Yeh, Both cell fusion and transdifferentiation account for the transformation of human peripheral blood CD34-positive cells into cardiomyocytes in vivo, Circulation 110 (25) (2004) 3803–3807.

[205] H. Iwasaki, A. Kawamoto, M. Ishikawa, A. Oyamada, S. Nakamori, H. Nishimura, K. Sadamoto, M. Horii, T. Matsumoto, S. Murasawa, T. Shibata, S. Suehiro, T. Asahara, Dose-dependent contribution of CD34-positive cell transplantation to concurrent vasculogenesis and cardiomyogenesis for functional regenerative recovery after myocardial infarction, Circulation 113 (10) (2006) 1311–1325.

[206] E.T.H. Yeh, S. Zhang, H.D. Wu, M. Korbling, J.T. Willerson, Z. Estrov, Transdifferentiation of human peripheral blood CD34+-enriched cell population into cardiomyocytes, endothelial cells, and smooth muscle cells in vivo, Circulation 108 (17) (2003) 2070–2073.

[207] X. Bai, Y. Yan, Y.-H. Song, M. Seidensticker, B. Rabinovich, R. Metzele, J.A. Bankson, D. Vykoukal, E. Alt, Both cultured and freshly isolated adipose tissue-derived stem cells enhance cardiac function after acute myocardial infarction, Eur. Heart J. 31 (4) (2010) 489–501.

[208] S. Zhang, E. Shpall, J.T. Willerson, E.T.H. Yeh, Fusion of human hematopoietic progenitor cells and murine cardiomyocytes is mediated by alpha 4 beta 1 integrin/vascular cell adhesion molecule-1 interaction, Circ. Res. 100 (5) (2007) 693–702.

[209] A. Acquistapace, T. Bru, P.-F. Lesault, F. Figeac, A.E. Coudert, O. Le Coz, C. Christov, X. Baudin, F. Auber, R. Yiou, J.-L. Dubois-Rande, A.-M. Rodriguez, Human mesenchymal stem cells reprogram adult cardiomyocytes toward a progenitor-like state through partial cell fusion and mitochondria transfer, Stem Cells 29 (5) (2011) 812–824.

[210] L. Wang, J. Deng, W. Tian, B. Xiang, T. Yang, G. Li, J. Wang, M. Gruwel, T. Kashour, J. Rendell, M. Glogowski, B. Tomanek, D. Freed, R. Deslauriers, R.C. Arora, G. Tian, Adipose-derived stem cells are an effective cell candidate for treatment of heart failure: an MR imaging study of rat hearts, Am. J. Physiol. Heart Circ. Physiol. 297 (3) (2009) H1020–31.

[211] L. Zeng, Q. Hu, X. Wang, A. Mansoor, J. Lee, J. Feygin, G. Zhang, P. Suntharalingam, S. Boozer, A. Mhashilkar, C.J. Panetta, C. Swingen, R. Deans, A.H.L. From, R.J. Bache, C.M. Verfaillie, J. Zhang, Bioenergetic and functional consequences of bone marrow-derived multipotent progenitor cell transplantation in hearts with postinfarction left ventricular remodeling, Circulation 115 (14) (2007) 1866–1875.

[212] X.-L. Tang, Q. Li, G. Rokosh, S.K. Sanganalmath, N. Chen, Q. Ou, H. Stowers, G. Hunt, R. Bolli, Long-term outcome of administration of c-kit(POS) cardiac progenitor cells after acute myocardial infarction: transplanted cells do not become cardiomyocytes, but structural and functional improvement and proliferation of endogenous cells persist for at least one year, Circ. Res. 118 (7) (2016) 1091–1105.

[213] K.U. Hong, Q.-H. Li, Y. Guo, N.S. Patton, A. Moktar, A. Bhatnagar, R. Bolli, A highly sensitive and accurate method to quantify absolute numbers of c-kit+ cardiac stem cells following transplantation in mice, Basic Res. Cardiol. 108 (3) (2013) 346.

[214] M. Penicka, P. Widimsky, P. Kobylka, T. Kozak, O. Lang, Images in cardiovascular medicine. Early tissue distribution of bone marrow mononuclear cells after transcoronary transplantation in a patient with acute myocardial infarction, Circulation 112 (4) (2005) e63–5.

[215] R.R. Smith, L. Barile, H.C. Cho, M.K. Leppo, J.M. Hare, E. Messina, A. Giacomello, M.R. Abraham, E. Marban, Regenerative potential of cardiosphere-derived cells expanded from percutaneous endomyocardial biopsy specimens, Circulation 115 (7) (2007) 896–908.

[216] B. Dawn, A.B. Stein, K. Urbanek, M. Rota, B. Whang, R. Rastaldo, D. Torella, X.-L. Tang, A. Rezazadeh, J. Kajstura, A. Leri, G. Hunt, J. Varma, S.D. Prabhu, P. Anversa, R. Bolli, Cardiac stem cells delivered intravascularly traverse the vessel barrier, regenerate infarcted myocardium, and improve cardiac function, Proc. Natl. Acad. Sci. U. S. A. 102 (10) (2005) 3766–3771.

[217] A.M. Smits, L.W. van Laake, K. den Ouden, C. Schreurs, K. Szuhai, C.J. van Echteld, C.L. Mummery, P.A. Doevendans, M.-J. Goumans, Human cardiomyocyte progenitor cell transplantation preserves long-term function of the infarcted mouse myocardium, Cardiovasc. Res. 83 (3) (2009) 527–535.

[218] D.A. D'Alessandro, J. Kajstura, T. Hosoda, A. Gatti, R. Bello, F. Mosna, S. Bardelli, H. Zheng, D. D'Amario, M.E. Padin-Iruegas, A.B. Carvalho, M. Rota, M.O. Zembala, D. Stern, O. Rimoldi, K. Urbanek, R.E. Michler, A. Leri, P. Anversa, Progenitor cells from the explanted heart generate immunocompatible myocardium within the transplanted donor heart, Circ. Res. 105 (11) (2009) 1128–1140.

[219] M. Rota, M.E. Padin-Iruegas, Y. Misao, A. de Angelis, S. Maestroni, J. Ferreira-Martins, E. Fiumana, R. Rastaldo, M.L. Arcarese, T.S. Mitchell, A. Boni, R. Bolli, K. Urbanek, T. Hosoda, P. Anversa, A. Leri, J. Kajstura, Local activation or implantation of cardiac progenitor cells rescues scarred infarcted myocardium improving cardiac function, Circ. Res. 103 (1) (2008) 107–116.

[220] X.-L. Tang, G. Rokosh, S.K. Sanganalmath, F. Yuan, H. Sato, J. Mu, S. Dai, C. Li, N. Chen, Y. Peng, B. Dawn, G. Hunt, A. Leri, J. Kajstura, S. Tiwari, G. Shirk, P. Anversa, R. Bolli, Intracoronary administration of cardiac progenitor cells alleviates left ventricular dysfunction in rats with a 30-day-old infarction, Circulation 121 (2) (2010) 293–305.

[221] J.H. van Berlo, O. Kanisicak, M. Maillet, R.J. Vagnozzi, J. Karch, S.-C.J. Lin, R.C. Middleton, E. Marban, J.D. Molkentin, c-kit+ cells minimally contribute cardiomyocytes to the heart, Nature 509 (7500) (2014) 337–341.

[222] N. Sultana, L. Zhang, J. Yan, J. Chen, W. Cai, S. Razzaque, D. Jeong, W. Sheng, L. Bu, M. Xu, G.-Y. Huang, R.J. Hajjar, B. Zhou, A. Moon, C.-L. Cai, Resident c-kit(+) cells in the heart are not cardiac stem cells, Nat. Commun. 6 (2015) 8701.

[223] Q. Liu, R. Yang, X. Huang, H. Zhang, L. He, L. Zhang, X. Tian, Y. Nie, S. Hu, Y. Yan, L. Zhang, Z. Qiao, Q.-D. Wang, K.O. Lui, B. Zhou, Genetic lineage tracing identifies in situ Kit-expressing cardiomyocytes, Cell Res. 26 (1) (2016) 119–130.

[224] O. Bergmann, R.D. Bhardwaj, S. Bernard, S. Zdunek, F. Barnabe-Heider, S. Walsh, J. Zupicich, K. Alkass, B.A. Buchholz, H. Druid, S. Jovinge, J. Frisen, Evidence for cardiomyocyte renewal in humans, Science (New York, N.Y.) 324 (5923) (2009) 98–102.

[225] J. Kajstura, K. Urbanek, S. Perl, T. Hosoda, H. Zheng, B. Ogorek, J. Ferreira-Martins, P. Goichberg, C. Rondon-Clavo, F. Sanada, D. D'Amario, M. Rota, F. Del Monte, D. Orlic, J. Tisdale, A. Leri, P. Anversa, Cardiomyogenesis in the adult human heart, Circ. Res. 107 (2) (2010) 305–315.

[226] P.P. Zwetsloot, A.M.D. Vegh, S.J. Jansen of Lorkeers, G.P.J. van Hout, G.L. Currie, E.S. Sena, H. Gremmels, J.W. Buikema, M.-J. Goumans, M.R. Macleod, P.A. Doevendans, S.A.J. Chamuleau, J.P.G. Sluijter, Cardiac stem cell treatment in myocardial infarction: a systematic review and meta-analysis of preclinical studies, Circ. Res. 118 (8) (2016) 1223–1232.

[227] C.I. Lang, M. Wolfien, A. Langenbach, P. Müller, O. Wolkenhauer, A. Yavari, H. Ince, G. Steinhoff, B.J. Krause, R. David, Ä. Glass, Cardiac cell therapies for the treatment

of acute myocardial infarction: a meta-analysis from mouse studies, Cell. Physiol. Biochem. 42 (1) (2017) 254–268.
[228] R. Bolli, A.R. Chugh, D. D'Amario, J.H. Loughran, M.F. Stoddard, S. Ikram, G.M. Beache, S.G. Wagner, A. Leri, T. Hosoda, F. Sanada, J.B. Elmore, P. Goichberg, D. Cappetta, N.K. Solankhi, I. Fahsah, D.G. Rokosh, M.S. Slaughter, J. Kajstura, P. Anversa, Cardiac stem cells in patients with ischaemic cardiomyopathy (SCIPIO): initial results of a randomised phase 1 trial, Lancet 378 (9806) (2011) 1847–1857.
[229] A.R. Chugh, G.M. Beache, J.H. Loughran, N. Mewton, J.B. Elmore, J. Kajstura, P. Pappas, A. Tatooles, M.F. Stoddard, J.A.C. Lima, M.S. Slaughter, P. Anversa, R. Bolli, Administration of cardiac stem cells in patients with ischemic cardiomyopathy: the SCIPIO trial: surgical aspects and interim analysis of myocardial function and viability by magnetic resonance, Circulation 126 (11 Suppl 1) (2012) S54–64.
[230] K. Malliaras, R.R. Makkar, R.R. Smith, K. Cheng, E. Wu, R.O. Bonow, L. Marban, A. Mendizabal, E. Cingolani, P.V. Johnston, G. Gerstenblith, K.H. Schuleri, A.C. Lardo, E. Marban, Intracoronary cardiosphere-derived cells after myocardial infarction: evidence of therapeutic regeneration in the final 1-year results of the CADUCEUS trial (CArdiosphere-derived aUtologous stem CElls to reverse ventricULar dySfunction), J. Am. Coll. Cardiol. 63 (2) (2014) 110–122.
[231] S. Dimmeler, A.M. Zeiher, Cell therapy of acute myocardial infarction: open questions, Cardiology 113 (3) (2009) 155–160.
[232] B.E. Strauer, M. Brehm, T. Zeus, N. Gattermann, A. Hernandez, R.V. Sorg, G. Kogler, P. Wernet, Intrakoronare, humane autologe Stammzelltransplantation zur Myokardregeneration nach Herzinfarkt, Dtsch. Med. Wschr. 126 (34–35) (2001) 932–938.
[233] T. Choudhury, A. Mozid, S. Hamshere, C. Yeo, C. Pellaton, S. Arnous, N. Saunders, P. Brookman, A. Jain, D. Locca, A. Archbold, C. Knight, A. Wragg, C. Davies, P. Mills, M. Parmar, M. Rothman, F. Choudry, D.A. Jones, S. Agrawal, J. Martin, A. Mathur, An exploratory randomized control study of combination cytokine and adult autologous bone marrow progenitor cell administration in patients with ischaemic cardiomyopathy: the REGENERATE-IHD clinical trial, Eur. J. Heart Fail. 19 (1) (2017) 138–147.
[234] V. Schachinger, S. Erbs, A. Elsasser, W. Haberbosch, R. Hambrecht, H. Holschermann, J. Yu, R. Corti, D.G. Mathey, C.W. Hamm, T. Suselbeck, B. Assmus, T. Tonn, S. Dimmeler, A.M. Zeiher, Intracoronary bone marrow-derived progenitor cells in acute myocardial infarction, N. Engl. J. Med. 355 (12) (2006) 1210–1221.
[235] K.C. Wollert, G.P. Meyer, J. Lotz, S. Ringes Lichtenberg, P. Lippolt, C. Breidenbach, S. Fichtner, T. Korte, B. Hornig, D. Messinger, L. Arseniev, B. Hertenstein, A. Ganser, H. Drexler, Intracoronary autologous bone-marrow cell transfer after myocardial infarction: the BOOST randomised controlled clinical trial, Lancet 364 (9429) (2004) 141–148.
[236] H.V. Huikuri, K. Kervinen, M. Niemela, K. Ylitalo, M. Saily, P. Koistinen, E.-R. Savolainen, H. Ukkonen, M. Pietila, J.K.E. Airaksinen, J. Knuuti, T.H. Makikallio, Effects of intracoronary injection of mononuclear bone marrow cells on left ventricular function, arrhythmia risk profile, and restenosis after thrombolytic therapy of acute myocardial infarction, Eur. Heart J. 29 (22) (2008) 2723–2732.
[237] B. Assmus, D.H. Walter, F.H. Seeger, D.M. Leistner, J. Steiner, I. Ziegler, A. Lutz, W. Khaled, J. Klotsche, T. Tonn, S. Dimmeler, A.M. Zeiher, Effect of shock wave-facilitated intracoronary cell therapy on LVEF in patients with chronic heart failure: the CELLWAVE randomized clinical trial, JAMA 309 (15) (2013) 1622–1631.
[238] E. Pokushalov, A. Romanov, A. Chernyavsky, P. Larionov, I. Terekhov, S. Artyomenko, O. Poveshenko, E. Kliver, N. Shirokova, A. Karaskov, N. Dib, Efficiency of intramyocardial injections of autologous bone marrow mononuclear cells in patients with ischemic heart failure: a randomized study, J. Cardiovasc. Transl. Res. 3 (2) (2010) 160–168.
[239] Q. Zhao, Y. Sun, L. Xia, A. Chen, Z. Wang, Randomized study of mononuclear bone marrow cell transplantation in patients with coronary surgery, Ann. Thorac. Surg. 86 (6) (2008) 1833–1840.
[240] S. Hu, S. Liu, Z. Zheng, X. Yuan, L. Li, M. Lu, R. Shen, F. Duan, X. Zhang, J. Li, X. Liu, Y. Song, W. Wang, S. Zhao, Z. He, H. Zhang, K. Yang, W. Feng, X. Wang, Isolated coronary artery bypass graft combined with bone marrow mononuclear cells delivered through a graft vessel for patients with previous myocardial infarction and chronic heart failure: a single-center, randomized, double-blind, placebo-controlled clinical trial, J. Am. Coll. Cardiol. 57 (24) (2011) 2409–2415.
[241] M. Lu, S. Liu, Z. Zheng, G. Yin, L. Song, H. Chen, X. Chen, Q. Chen, S. Jiang, L. Tian, Z. He, S. Hu, S. Zhao, A pilot trial of autologous bone marrow mononuclear cell transplantation through grafting artery: a sub-study focused on segmental left ventricular function recovery and scar reduction, Int. J. Cardiol. 168 (3) (2013) 2221–2227.
[242] J.H. Traverse, T.D. Henry, S.G. Ellis, C.J. Pepine, J.T. Willerson, D.X.M. Zhao, J.R. Forder, B.J. Byrne, A.K. Hatzopoulos, M.S. Penn, E.C. Perin, K.W. Baran, J. Chambers, C. Lambert, G. Raveendran, D.I. Simon, D.E. Vaughan, L.M. Simpson, A.P. Gee, D.A. Taylor, C.R. Cogle, J.D. Thomas, G.V. Silva, B.C. Jorgenson, R.E. Olson, S. Bowman, J. Francescon, C. Geither, E. Handberg, D.X. Smith, S. Baraniuk, L.B. Piller, C. Loghin, D. Aguilar, S. Richman, C. Zierold, J. Bettencourt, S.L. Sayre, R.W. Vojvodic, S.I. Skarlatos, D.J. Gordon, R.F. Ebert, M. Kwak, L.A. Moye, R.D. Simari, Effect of intracoronary delivery of autologous bone marrow mononuclear cells 2 to 3 weeks following acute myocardial infarction on left ventricular function: the LateTIME randomized trial, JAMA 306 (19) (2011) 2110–2119.
[243] E.C. Perin, J.T. Willerson, C.J. Pepine, T.D. Henry, S.G. Ellis, D.X.M. Zhao, G.V. Silva, D. Lai, J.D. Thomas, M.W. Kronenberg, A.D. Martin, R.D. Anderson, J.H. Traverse, M.S. Penn, S. Anwaruddin, A.K. Hatzopoulos, A.P. Gee, D.A. Taylor, C.R. Cogle, D. Smith, L. Westbrook, J. Chen, E. Handberg, R.E. Olson, C. Geither, S. Bowman, J. Francescon, S. Baraniuk, L.B. Piller, L.M. Simpson, C. Loghin, D. Aguilar, S. Richman, C. Zierold, J. Bettencourt, S.L. Sayre, R.W. Vojvodic, S.I. Skarlatos, D.J. Gordon, R.F. Ebert, M. Kwak, L.A. Moye, R.D. Simari, Effect of transendocardial delivery of autologous bone marrow mononuclear cells on functional capacity, left ventricular function, and perfusion in chronic heart failure: the FOCUS-CCTRN trial, JAMA 307 (16) (2012) 1717–1726.
[244] T. Patila, M. Lehtinen, A. Vento, J. Schildt, J. Sinisalo, M. Laine, P. Hammainen, A. Nihtinen, R. Alitalo, P. Nikkinen, A. Ahonen, M. Holmstrom, K. Lauerma, R. Poyhia, M. Kupari, E. Kankuri, A. Harjula, Autologous bone marrow mononuclear cell transplantation in ischemic heart failure: a prospective, controlled, randomized, double-blind study of cell transplantation combined with coronary bypass, J Heart Lung Transplant 33 (6) (2014) 567–574.
[245] R.T. Sant'Anna, J. Fracasso, F.H. Valle, I. Castro, N.B. Nardi, J.R.M. Sant'Anna, I.A. Nesralla, R.A.K. Kalil, Direct intramyocardial transthoracic transplantation of bone marrow mononuclear cells for non-ischemic dilated cardiomyopathy: INTRACELL, a prospective randomized controlled trial, Rev. Bras. Cir. Cardiovasc. 29 (3) (2014) 437–447.
[246] K.-L. Ang, D. Chin, F. Leyva, P. Foley, C. Kubal, S. Chalil, L. Srinivasan, L. Bernhardt, S. Stevens, L.T. Shenje, M. Galinanes, Randomized, controlled trial of intramuscular or intracoronary injection of autologous bone marrow cells into scarred myocardium during CABG versus CABG alone, Nat. Clin. Pract. Cardiovasc. Med. 5 (10) (2008) 663–670.
[247] K. Lunde, S. Solheim, S. Aakhus, H. Arnesen, M. Abdelnoor, T. Egeland, K. Endresen, A. Ilebekk, A. Mangschau, J.G. Fjeld, H.J. Smith, E. Taraldsrud, H.K. Grogaard, R. Bjornerheim, M. Brekke, C. Muller, E. Hopp, A. Ragnarsson, J.E. Brinchmann, K. Forfang, Intracoronary injection of mononuclear bone marrow cells in acute myocardial infarction, N. Engl. J. Med. 355 (12) (2006) 1199–1209.
[248] A. Hirsch, R. Nijveldt, P.A. van der Vleuten, J.G.P. Tijssen, W.J. van der Giessen, R.A. Tio, J. Waltenberger, J.M. ten Berg, P.A. Doevendans, W.R.M. Aengevaeren, J.J. Zwaginga, B.J. Biemond, A.C. van Rossum, J.J. Piek, F. Zijlstra, Intracoronary infusion of mononuclear cells from bone marrow or peripheral blood compared with standard therapy in patients after acute myocardial infarction treated by primary percutaneous coronary intervention: results of the randomized controlled HEBE trial, Eur. Heart J. 32 (14) (2011) 1736–1747.
[249] D. Surder, R. Manka, V. Lo Cicero, T. Moccetti, K. Rufibach, S. Soncin, L. Turchetto, M. Radrizzani, G. Astori, J. Schwitter, P. Erne, M. Zuber, C. Auf der Maur, P. Jamshidi, O. Gaemperli, S. Windecker, A. Moschovitis, A. Wahl, I. Buhler, C. Wyss, S. Kozerke, U. Landmesser, T.F. Luscher, R. Corti, Intracoronary injection of bone marrow-derived mononuclear cells early or late after acute myocardial infarction: effects on global left ventricular function, Circulation 127 (19) (2013) 1968–1979.
[250] A.W. Heldman, D.L. DiFede, J.E. Fishman, J.P. Zambrano, B.H. Trachtenberg, V. Karantalis, M. Mushtaq, A.R. Williams, V.Y. Suncion, I.K. McNiece, E. Ghersin, V. Soto, G. Lopera, R. Miki, H. Willens, R. Hendel, R. Mitrani, P. Pattany, G. Feigenbaum, B. Oskouei, J. Byrnes, M.H. Lowery, J. Sierra, M.V. Pujol, C. Delgado, P.J. Gonzalez, J.E. Rodriguez, L.L. Bagno, D. Rouy, P. Altman, C.W.P. Foo, J. Da Silva, E. Anderson, R. Schwarz, A. Mendizabal, J.M. Hare, Transendocardial mesenchymal stem cells and mononuclear bone marrow cells for ischemic cardiomyopathy: the TAC-HFT randomized trial, JAMA 311 (1) (2014) 62–73.
[251] T. Santoso, C.-W. Siu, C. Irawan, W.-S. Chan, I. Alwi, K.-H. Yiu, A. Aziz, Y.-L. Kwong, H.-F. Tse, Endomyocardial implantation of autologous bone marrow mononuclear cells in advanced ischemic heart failure: a randomized placebo-controlled trial (END-HF), J. Cardiovasc. Transl. Res. 7 (6) (2014) 545–552.
[252] K. Sadat, S. Ather, W. Aljaroudi, J. Heo, A.E. Iskandrian, F.G. Hage, The effect of bone marrow mononuclear stem cell therapy on left ventricular function and myocardial perfusion, J. Nucl. Cardiol. 21 (2) (2014) 351–367.
[253] Y. Wen, B. Chen, C. Wang, X. Ma, Q. Gao, Bone marrow-derived mononuclear cell therapy for patients with ischemic heart disease and ischemic heart failure, Expert. Opin. Biol. Ther. 12 (12) (2012) 1563–1573.
[254] Z. Ye, B.-L. Zhang, X.-X. Zhao, Y.-W. Qin, H. Wu, J. Cao, J.-L. Zhang, J.-Q. Hu, X. Zheng, R.-L. Xu, Intracoronary infusion of bone marrow-derived mononuclear cells contributes to longstanding improvements of left ventricular performance and remodelling after acute myocardial infarction: a meta-analysis, Heart Lung Circ. 21 (11) (2012) 725–733.
[255] R. de Jong, J.H. Houtgraaf, S. Samiei, E. Boersma, H.J. Duckers, Intracoronary stem cell infusion after acute myocardial infarction: a meta-analysis and update on clinical trials, Circ. Cardiovasc. Interv. 7 (2) (2014) 156–167.
[256] V. Jeevanantham, M. Butler, A. Saad, A. Abdel-Latif, E.K. Zuba-Surma, B. Dawn, Adult bone marrow cell therapy improves survival and induces long-term improvement in cardiac parameters: a systematic review and meta-analysis, Circulation 126 (5) (2012) 551–568.
[257] J.H. Traverse, T.D. Henry, L.A. Moye, Is the measurement of left ventricular ejection fraction the proper end point for cell therapy trials? An analysis of the effect of bone marrow mononuclear stem cell administration on left ventricular ejection fraction after ST-segment elevation myocardial infarction when evaluated by cardiac magnetic resonance imaging, Am. Heart J. 162 (4) (2011) 671–677.
[258] W.P. Sheridan, C.G. Begley, C.A. Juttner, J. Szer, L.B. To, D. Maher, K.M. McGrath, G. Morstyn, R.M. Fox, Effect of peripheral-blood progenitor cells mobilised by filgrastim (G-CSF) on platelet recovery after high-dose chemotherapy, Lancet 339 (8794) (1992) 640–644.
[259] K. Moazzami, A. Roohi, B. Moazzami, Granulocyte colony stimulating factor therapy for acute myocardial infarction, Cochrane Database Syst. Rev. (5) (2013), CD008844.
[260] D. Zohlnhofer, A. Dibra, T. Koppara, A. de Waha, R.S. Ripa, J. Kastrup, M. Valgimigli, A. Schomig, A. Kastrati, Stem cell mobilization by granulocyte colony-stimulating factor for myocardial recovery after acute myocardial infarction: a meta-analysis, J. Am. Coll. Cardiol. 51 (15) (2008) 1429–1437.
[261] A. Abdel-Latif, R. Bolli, E.K. Zuba-Surma, I.M. Tleyjeh, C.A. Hornung, B. Dawn, Granulocyte colony-stimulating factor therapy for cardiac repair after acute myocardial infarction: a systematic review and meta-analysis of randomized controlled trials, Am. Heart J. 156 (2) (2008) 216–226 (.e9).

[262] H.-J. Kang, H.-S. Kim, S.-Y. Zhang, K.-W. Park, H.-J. Cho, B.-K. Koo, Y.-J. Kim, D.S. Lee, D.-W. Sohn, K.-S. Han, B.-H. Oh, M.-M. Lee, Y.-B. Park, Effects of intracoronary infusion of peripheral blood stem-cells mobilised with granulocyte-colony stimulating factor on left ventricular systolic function and restenosis after coronary stenting in myocardial infarction: the MAGIC cell randomised clinical trial, Lancet 363 (9411) (2004) 751–756.

[263] H.-J. Kang, H.-Y. Lee, S.-H. Na, S.-A. Chang, K.-W. Park, H.-K. Kim, S.-Y. Kim, H.-J. Chang, W. Lee, W.J. Kang, B.-K. Koo, Y.-J. Kim, D.S. Lee, D.-W. Sohn, K.-S. Han, B.-H. Oh, Y.-B. Park, H.-S. Kim, Differential effect of intracoronary infusion of mobilized peripheral blood stem cells by granulocyte colony-stimulating factor on left ventricular function and remodeling in patients with acute myocardial infarction versus old myocardial infarction: the MAGIC Cell-3-DES randomized, controlled trial, Circulation 114 (1 Suppl) (2006) I145–51.

[264] Z.-q. Li, M. Zhang, Y.-z. Jing, W.-w. Zhang, Y. Liu, L.-j. Cui, L. Yuan, X.-z. Liu, X. Yu, T.-s. Hu, The clinical study of autologous peripheral blood stem cell transplantation by intracoronary infusion in patients with acute myocardial infarction (AMI), Int. J. Cardiol. 115 (1) (2007) 52–56.

[265] T. Tatsumi, E. Ashihara, T. Yasui, S. Matsunaga, A. Kido, Y. Sasada, S. Nishikawa, M. Hadase, M. Koide, R. Nakamura, H. Irie, K. Ito, A. Matsui, H. Matsui, M. Katamura, S. Kusuoka, S. Matoba, S. Okayama, M. Horii, S. Uemura, C. Shimazaki, H. Tsuji, Y. Saito, H. Matsubara, Intracoronary transplantation of non-expanded peripheral blood-derived mononuclear cells promotes improvement of cardiac function in patients with acute myocardial infarction, Circ. J. 71 (8) (2007) 1199–1207.

[266] B. Vrtovec, M. Sever, M. Jensterle, G. Poglajen, A. Janez, N. Kravos, G. Zemljic, M. Cukjati, P. Cernelc, F. Haddad, J.C. Wu, U.P. Jorde, Efficacy of CD34+ stem cell therapy in nonischemic dilated cardiomyopathy is absent in patients with diabetes but preserved in patients with insulin resistance, Stem Cells Transl. Med. 5 (5) (2016) 632–638.

[267] L. Lezaic, A. Socan, G. Poglajen, P.K. Peitl, M. Sever, M. Cukjati, P. Cernelc, J.C. Wu, F. Haddad, B. Vrtovec, Intracoronary transplantation of CD34(+) cells is associated with improved myocardial perfusion in patients with nonischemic dilated cardiomyopathy, J. Card. Fail. 21 (2) (2015) 145–152.

[268] G. Poglajen, M. Sever, M. Cukjati, P. Cernelc, I. Knezevic, G. Zemljic, F. Haddad, J.C. Wu, B. Vrtovec, Effects of transendocardial CD34+ cell transplantation in patients with ischemic cardiomyopathy, Circ. Cardiovasc. Interv. 7 (4) (2014) 552–559.

[269] D.W. Losordo, T.D. Henry, C. Davidson, J. Sup Lee, M.A. Costa, T. Bass, F. Mendelsohn, F.D. Fortuin, C.J. Pepine, J.H. Traverse, D. Amrani, B.M. Ewenstein, N. Riedel, K. Story, K. Barker, T.J. Povsic, R.A. Harrington, R.A. Schatz, Intramyocardial, autologous CD34+ cell therapy for refractory angina, Circ. Res. 109 (4) (2011) 428–436.

[270] A. Kawamoto, H. Iwasaki, K. Kusano, T. Murayama, A. Oyamada, M. Silver, C. Hulbert, M. Gavin, A. Hanley, H. Ma, M. Kearney, V. Zak, T. Asahara, D.W. Losordo, CD34-positive cells exhibit increased potency and safety for therapeutic neovascularization after myocardial infarction compared with total mononuclear cells, Circulation 114 (20) (2006) 2163–2169.

[271] B. Vrtovec, G. Poglajen, L. Lezaic, M. Sever, A. Socan, D. Domanovic, P. Cernelc, G. Torre-Amione, F. Haddad, J.C. Wu, Comparison of transendocardial and intracoronary CD34+ cell transplantation in patients with nonischemic dilated cardiomyopathy, Circulation 128 (11 Suppl 1) (2013) S42–9.

[272] B. Vrtovec, G. Poglajen, L. Lezaic, M. Sever, D. Domanovic, P. Cernelc, A. Socan, S. Schrepfer, G. Torre-Amione, F. Haddad, J.C. Wu, Effects of intracoronary CD34+ stem cell transplantation in nonischemic dilated cardiomyopathy patients: 5-year follow-up, Circ. Res. 112 (1) (2013) 165–173.

[273] S. Pasquet, H. Sovalat, P. Henon, N. Bischoff, Y. Arkam, M. Ojeda-Uribe, R. Le Bouar, V. Rimelen, I. Brink, R. Dallemand, J.-P. Monassier, Long-term benefit of intracardiac delivery of autologous granulocyte-colony-stimulating factor-mobilized blood CD34+ cells containing cardiac progenitors on regional heart structure and function after myocardial infarct, Cytotherapy 11 (8) (2009) 1002–1015.

[274] F.-Y. Lee, Y.-L. Chen, P.-H. Sung, M.-C. Ma, S.-N. Pei, C.-J. Wu, C.-H. Yang, M. Fu, S.-F. Ko, S. Leu, H.-K. Yip, Intracoronary transfusion of circulation-derived CD34+ cells improves left ventricular function in patients with end-stage diffuse coronary artery disease unsuitable for coronary intervention, Crit. Care Med. 43 (10) (2015) 2117–2132.

[275] T.J. Povsic, T.D. Henry, J.H. Traverse, F.D. Fortuin, G.L. Schaer, D.J. Kereiakes, R.A. Schatz, A.M. Zeiher, C.J. White, D.J. Stewart, E.M. Jolicoeur, T. Bass, D.A. Henderson, P. Dignacco, Z. Gu, H.R. Al-Khalidi, C. Junge, A. Nada, A.S. Hunt, D.W. Losordo, The RENEW trial: efficacy and safety of intramyocardial autologous CD34(+) cell administration in patients with refractory angina, JACC Cardiovasc. Interv. 9 (15) (2016) 1576–1585.

[276] J.-H. Choi, J. Choi, W.-S. Lee, I. Rhee, S.-C. Lee, H.-C. Gwon, S.H. Lee, Y.H. Choe, D.W. Kim, W. Suh, D.-K. Kim, E.-S. Jeon, Lack of additional benefit of intracoronary transplantation of autologous peripheral blood stem cell in patients with acute myocardial infarction, Circ. J. 71 (4) (2007) 486–494.

[277] D.A. Taylor, E.C. Perin, J.T. Willerson, C. Zierold, M. Resende, M. Carlson, B. Nestor, E. Wise, A. Orozco, C.J. Pepine, T.D. Henry, S.G. Ellis, D.X.M. Zhao, J.H. Traverse, J.P. Cooke, R.C. Schutt, A. Bhatnagar, M.B. Grant, D. Lai, B.H. Johnstone, S.L. Sayre, L. Moyé, R.F. Ebert, R. Bolli, R.D. Simari, C.R. Cogle, Identification of bone marrow cell subpopulations associated with improved functional outcomes in patients with chronic left ventricular dysfunction: an embedded cohort evaluation of the FOCUS-CCTRN trial, Cell Transplant. 25 (9) (2016) 1675–1687.

[278] F. Fernández-Avilés, R. Sanz-Ruiz, A.M. Climent, L. Badimon, R. Bolli, D. Charron, V. Fuster, S. Janssens, J. Kastrup, H.-S. Kim, T.F. Lüscher, J.F. Martin, P. Menasché, R.D. Simari, G.W. Stone, A. Terzic, J.T. Willerson, J.C. Wu, Global position paper on cardiovascular regenerative medicine: scientific statement of the transnational alliance for regenerative therapies in cardiovascular syndromes (TACTICS) international group for the comprehensive cardiovascular application of regenerative medicinal products, Eur. Heart J. 38 (33) (2017) 2532–2546.

[279] A.N. Patel, L. Geffner, R.F. Vina, J. Saslavsky, H.C. Urschel JR, R. Kormos, F. Benetti, Surgical treatment for congestive heart failure with autologous adult stem cell transplantation: a prospective randomized study, J. Thorac. Cardiovasc. Surg. 130 (6) (2005) 1631–1638.

[280] B.A. Nasseri, W. Ebell, M. Dandel, M. Kukucka, R. Gebker, A. Doltra, C. Knosalla, Y.-H. Choi, R. Hetzer, C. Stamm, Autologous CD133+ bone marrow cells and bypass grafting for regeneration of ischaemic myocardium: the Cardio133 trial, Eur. Heart J. 35 (19) (2014) 1263–1274.

[281] K. Vakil, V. Florea, R. Koene, J.V. Kealhofer, I. Anand, S. Adabag, Effect of coronary artery bypass grafting on left ventricular ejection fraction in men eligible for implantable cardioverter-defibrillator, Am. J. Cardiol. 117 (6) (2016) 957–960.

[282] A. Sharma, B.G. Demissei, J. Tromp, H.L. Hillege, J.G. Cleland, C.M. O'Connor, M. Metra, P. Ponikowski, J.R. Teerlink, B.A. Davison, M.M. Givertz, D.M. Bloomfield, H. Dittrich, D.J. van Veldhuisen, G. Cotter, J.A. Ezekowitz, M.A.F. Khan, A.A. Voors, A network analysis to compare biomarker profiles in patients with and without diabetes mellitus in acute heart failure, Eur. J. Heart Fail. 10 (2017) 1310–1320.

[283] S.L. Chow, A.S. Maisel, I. Anand, B. Bozkurt, R.A. de Boer, G.M. Felker, G.C. Fonarow, B. Greenberg, J.L. Januzzi, M.S. Kiernan, P.P. Liu, T.J. Wang, C.W. Yancy, M.R. Zile, Role of biomarkers for the prevention, assessment, and management of heart failure: a scientific statement from the American Heart Association, Circulation 135 (22) (2017) e1054–e1091.

[284] N. Werner, S. Kosiol, T. Schiegl, P. Ahlers, K. Walenta, A. Link, M. Böhm, G. Nickenig, Circulating endothelial progenitor cells and cardiovascular outcomes, N. Engl. J. Med. 353 (10) (2005) 999–1007.

[285] H.J. Duckers, S. Silber, R. de Winter, P. den Heijer, B. Rensing, M. Rau, H. Mudra, E. Benit, S. Verheye, W. Wijns, P.W. Serruys, Circulating endothelial progenitor cells predict angiographic and intravascular ultrasound outcome following percutaneous coronary interventions in the HEALING-II trial: evaluation of an endothelial progenitor cell capturing stent, EuroIntervention 3 (1) (2007) 67–75.

[286] J. Aoki, P.W. Serruys, H. van Beusekom, A.T.L. Ong, E.P. McFadden, G. Sianos, Willem J. van der Giessen, E. Regar, P.J. de Feyter, H.R. Davis, S. Rowland, M.J.B. Kutryk, Endothelial progenitor cell capture by stents coated with antibody against CD34: the HEALING-FIM (healthy endothelial accelerated lining inhibits neointimal growth-first in man) registry, J. Am. Coll. Cardiol. 45 (10) (2005) 1574–1579.

[287] C. Real, F. Caiado, S. Dias, Endothelial progenitors in vascular repair and angiogenesis: how many are needed and what to do? Cardiovasc. Hematol. Disord. Drug Targets 8 (3) (2008) 185–193.

[288] A.F. Low, C.-H. Lee, S.-G. Teo, M.Y. Chan, E. Tay, Y.-P. Lee, E. Chong, M. Co, E. Tin Hay, Y.-T. Lim, H.-C. Tan, Effectiveness and safety of the genous endothelial progenitor cell-capture stent in acute ST-elevation myocardial infarction, Am. J. Cardiol. 108 (2) (2011) 202–205.

[289] S. Silber, P. Damman, M. Klomp, M.A. Beijk, M. Grisold, E.E. Ribeiro, H. Suryapranata, J. Wójcik, K. Hian Sim, J.G.P. Tijssen, R.J. de Winter, Clinical results after coronary stenting with the Genous™ bio-engineered R stent™: 12-month outcomes of the e-HEALING (healthy endothelial accelerated lining inhibits neointimal growth) worldwide registry, EuroIntervention 6 (7) (2011) 819–825.

[290] A. Bhatnagar, R. Bolli, B.H. Johnstone, J.H. Traverse, T.D. Henry, C.J. Pepine, J.T. Willerson, E.C. Perin, S.G. Ellis, D.X.M. Zhao, P.C. Yang, J.P. Cooke, R.C. Schutt, B.H. Trachtenberg, A. Orozco, M. Resende, R.F. Ebert, S.L. Sayre, R.D. Simari, L. Moyé, C.R. Cogle, D.A. Taylor, Bone marrow cell characteristics associated with patient profile and cardiac performance outcomes in the LateTIME-cardiovascular cell therapy research network (CCTRN) trial, Am. Heart J. 179 (2016) 142–150.

[291] S. Jaiswal, P. Fontanillas, J. Flannick, A. Manning, P.V. Grauman, B.G. Mar, R.C. Lindsley, C.H. Mermel, N. Burtt, A. Chavez, J.M. Higgins, V. Moltchanov, F.C. Kuo, M.J. Kluk, B. Henderson, L. Kinnunen, H.A. Koistinen, C. Ladenvall, G. Getz, A. Correa, B.F. Banahan, S. Gabriel, S. Kathiresan, H.M. Stringham, M.I. McCarthy, M. Boehnke, J. Tuomilehto, C. Haiman, L. Groop, G. Atzmon, J.G. Wilson, D. Neuberg, D. Altshuler, B.L. Ebert, Age-related clonal hematopoiesis associated with adverse outcomes, N. Engl. J. Med. 371 (26) (2014) 2488–2498.

[292] S. Jaiswal, P. Natarajan, A.J. Silver, C.J. Gibson, A.G. Bick, E. Shvartz, M. McConkey, N. Gupta, S. Gabriel, D. Ardissino, U. Baber, R. Mehran, V. Fuster, J. Danesh, P. Frossard, D. Saleheen, O. Melander, G.K. Sukhova, D. Neuberg, P. Libby, S. Kathiresan, B.L. Ebert, Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease, N. Engl. J. Med. 377 (2) (2017) 111–121.

[293] P.O. Iversen, P.R. Woldbaek, T. Tønnessen, G. Christensen, Decreased hematopoiesis in bone marrow of mice with congestive heart failure, Am. J. Physiol. Regul. Integr. Comp. Physiol. 282 (1) (2002) R166–72.

[294] Wellcome Trust Case Control Consortium, Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls, Nature 447 (7145) (2007) 661–678.

[295] T.L. Lasho, A. Pardanani, A. Tefferi, LNK mutations in JAK2 mutation-negative erythrocytosis, N. Engl. J. Med. 363 (12) (2010) 1189–1190.

[296] H. Schunkert, I.R. König, S. Kathiresan, M.P. Reilly, T.L. Assimes, H. Holm, M. Preuss, A.F.R. Stewart, M. Barbalic, C. Gieger, D. Absher, Z. Aherrahrou, H. Allayee, D. Altshuler, S.S. Anand, K. Andersen, J.L. Anderson, D. Ardissino, S.G. Ball, A.J. Balmforth, T.A. Barnes, D.M. Becker, L.C. Becker, K. Berger, J.C. Bis, S.M. Boekholdt, E. Boerwinkle, P.S. Braund, M.J. Brown, M.S. Burnett, I. Buysschaert, J.F. Carlquist, L. Chen, S. Cichon, V. Codd, R.W. Davies, G. Dedoussis, A. Dehghan, S. Demissie, J.M. Devaney, P. Diemert, R. Do, A. Doering, S. Eifert, N.E.E. Mokhtari, S.G. Ellis, R. Elosua, J.C. Engert, S.E. Epstein, U. de Faire, M. Fischer, A.R. Folsom, J. Freyer, B. Gigante, D. Girelli, S. Gretarsdottir, V. Gudnason, J.R. Gulcher, E. Halperin, N. Hammond, S.L. Hazen, A. Hofman, B.D. Horne, T. Illig, C. Iribarren, G.T. Jones, J.W. Jukema, M.A. Kaiser, L.M. Kaplan, J.J.P. Kastelein, K.-T. Khaw, J.W. Knowles, G. Kolovou, A. Kong, R. Laaksonen, D. Lambrechts, K. Leander, G. Lettre, M. Li, W. Lieb, C. Loley, A.J. Lotery, P.M. Mannucci, S. Maouche, N. Martinelli, P.P. McKeown, C. Meisinger, T. Meitinger, O. Melander, P.A. Merlini, V. Mooser, T. Morgan, T.W.

Mühleisen, J.B. Muhlestein, T. Münzel, K. Musunuru, J. Nahrstaedt, C.P. Nelson, M.M. Nöthen, O. Olivieri, R.S. Patel, C.C. Patterson, A. Peters, F. Peyvandi, L. Qu, A.A. Quyyumi, D.J. Rader, L.S. Rallidis, C. Rice, F.R. Rosendaal, D. Rubin, V. Salomaa, M.L. Sampietro, M.S. Sandhu, E. Schadt, A. Schäfer, A. Schillert, S. Schreiber, J. Schrezenmeir, S.M. Schwartz, D.S. Siscovick, M. Sivananthan, S. Sivapalaratnam, A. Smith, T.B. Smith, J.D. Snoep, N. Soranzo, J.A. Spertus, K. Stark, K. Stirrups, M. Stoll, W.H.W. Tang, S. Tennstedt, G. Thorgeirsson, G. Thorleifsson, M. Tomaszewski, A.G. Uitterlinden, A.M. van Rij, B.F. Voight, N.J. Wareham, G.A. Wells, H.-E. Wichmann, P.S. Wild, C. Willenborg, J.C.M. Witteman, B.J. Wright, S. Ye, T. Zeller, A. Ziegler, F. Cambien, A.H. Goodall, L.A. Cupples, T. Quertermous, W. März, C. Hengstenberg, S. Blankenberg, W.H. Ouwehand, A.S. Hall, P. Deloukas, J.R. Thompson, K. Stefansson, R. Roberts, U. Thorsteinsdottir, C.J. O'Donnell, R. McPherson, J. Erdmann, N.J. Samani, Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease, Nat. Genet. 43 (4) (2011) 333–338.

[297] W. Wang, Y. Tang, Y. Wang, L. Tascau, J. Balcerek, W. Tong, R.L. Levine, C. Welch, A.R. Tall, N. Wang, LNK/SH2B3 loss of function promotes atherosclerosis and thrombosis, Circ. Res. 119 (6) (2016) e91–e103.

[298] B.L. Dale, M.S. Madhur, Linking inflammation and hypertension via LNK/SH2B3, Curr. Opin. Nephrol. Hypertens. 25 (2) (2016) 87–93.

[299] R.J. Hung, C.M. Ulrich, E.L. Goode, Y. Brhane, K. Muir, A.T. Chan, L. Le Marchand, J. Schildkraut, J.S. Witte, R. Eeles, P. Boffetta, M.R. Spitz, J.G. Poirier, D.N. Rider, B.L. Fridley, Z. Chen, C. Haiman, F. Schumacher, D.F. Easton, M.T. Landi, P. Brennan, R. Houlston, D.C. Christiani, J.K. Field, H. Bickeböller, A. Risch, Z. Kote-Jarai, F. Wiklund, H. Grönberg, S. Chanock, S.I. Berndt, P. Kraft, S. Lindström, A.A. Al Olama, H. Song, C. Phelan, N. Wentzensen, U. Peters, M.L. Slattery, T.A. Sellers, G. Casey, S.B. Gruber, D.J. Hunter, C.I. Amos, B. Henderson, Cross cancer genomic investigation of inflammation pathway for five common cancers: lung, ovary, prostate, breast, and colorectal cancer, J. Natl. Cancer Inst. 107 (11) (2015).

[300] L.-W. Ding, Q.-Y. Sun, D.-C. Lin, W. Chien, N. Hattori, X.-M. Dong, S. Gery, M. Garg, N.B. Doan, J.W. Said, J.-F. Xiao, H. Yang, L.-Z. Liu, X. Meng, R.Y.-J. Huang, K. Tang, H.P. Koeffler, LNK (SH2B3): paradoxical effects in ovarian cancer, Oncogene 34 (11) (2015) 1463–1474.

[301] S. Takaki, J.D. Watts, K.A. Forbush, N.T. Nguyen, J. Hayashi, J. Alberola-Ila, R. Aebersold, R.M. Perlmutter, Characterization of Lnk. An adaptor protein expressed in lymphocytes, J. Biol. Chem. 272 (23) (1997) 14562–14570.

[302] https://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?exdb=AceView&db=36a&term=SH2B3.

[303] H. Takizawa, K. Eto, A. Yoshikawa, H. Nakauchi, K. Takatsu, S. Takaki, Growth and maturation of megakaryocytes is regulated by Lnk/Sh2b3 adaptor protein through crosstalk between cytokine- and integrin-mediated signals, Exp. Hematol. 36 (7) (2008) 897–906.

[304] Y. Cheng, K. Chikwava, C. Wu, H. Zhang, A. Bhagat, D. Pei, J.K. Choi, W. Tong, LNK/SH2B3 regulates IL-7 receptor signaling in normal and malignant B-progenitors, J. Clin. Invest. 126 (4) (2016) 1267–1281.

[305] S. Takaki, H. Morita, Y. Tezuka, K. Takatsu, Enhanced hematopoiesis by hematopoietic progenitor cells lacking intracellular adaptor protein, Lnk, J. Exp. Med. 195 (2) (2002) 151–160.

[306] M. Chatelais, J. Devallière, C. Galli, B. Charreau, Gene transfer of the adaptor Lnk (SH2B3) prevents porcine endothelial cell activation and apoptosis: implication for xenograft's cytoprotection, Xenotransplantation 18 (2) (2011) 108–120.

[307] J.H. Lee, S.T. Ji, J. Kim, S. Takaki, T. Asahara, Y.-J. Hong, S.-M. Kwon, Specific disruption of Lnk in murine endothelial progenitor cells promotes dermal wound healing via enhanced vasculogenesis, activation of myofibroblasts, and suppression of inflammatory cell recruitment, Stem Cell Res Ther 7 (1) (2016) 158.

[308] K. Fortney, E. Dobriban, P. Garagnani, C. Pirazzini, D. Monti, D. Mari, G. Atzmon, N. Barzilai, C. Franceschi, A.B. Owen, S.K. Kim, Genome-wide scan informed by age-related disease identifies loci for exceptional human longevity, PLoS Genet. 11 (12) (2015), e1005728.

[309] A. Perez-Garcia, A. Ambesi-Impiombato, M. Hadler, I. Rigo, C.A. LeDuc, K. Kelly, C. Jalas, E. Paietta, J. Racevskis, J.M. Rowe, M.S. Tallman, M. Paganin, G. Basso, W. Tong, W.K. Chung, A.A. Ferrando, Genetic loss of SH2B3 in acute lymphoblastic leukemia, Blood 122 (14) (2013) 2425–2432.

[310] C. Caliceti, P. Nigro, P. Rizzo, R. Ferrari, ROS, Notch, and Wnt signaling pathways: crosstalk between three major regulators of cardiovascular biology, Biomed. Res. Int. 2014 (2014) 318714.

[311] S.T. Oh, When the brakes are lost: LNK dysfunction in mice, men, and myeloproliferative neoplasms, Ther. Adv. Hematol. 2 (1) (2011) 11–19.

[312] C.M. Lindqvist, A. Lundmark, J. Nordlund, E. Freyhult, D. Ekman, J. Carlsson Almlöf, A. Raine, E. Övernäs, J. Abrahamsson, B.-M. Frost, D. Grandér, M. Heyman, J. Palle, E. Forestier, G. Lönnerholm, E.C. Berglund, A.-C. Syvänen, Deep targeted sequencing in pediatric acute lymphoblastic leukemia unveils distinct mutational patterns between genetic subtypes and novel relapse-associated genes, Oncotarget 7 (39) (2016) 64071–64088.

[313] E.C. Hales, J.W. Taub, L.H. Matherly, New insights into Notch1 regulation of the PI3K-AKT-mTOR1 signaling axis: targeted therapy of γ-secretase inhibitor resistant T-cell acute lymphoblastic leukemia, Cell. Signal. 26 (1) (2014) 149–161.

[314] S.H. Lee, K.B. Lee, J.H. Lee, S. Kang, H.G. Kim, T. Asahara, S.M. Kwon, Selective interference targeting of Lnk in umbilical cord-derived late endothelial progenitor cells improves vascular repair, following hind limb ischemic injury, via regulation of JAK2/STAT3 signaling, Stem Cells 33 (5) (2015) 1490–1500.

[315] S.-M. Kwon, Y.-K. Lee, A. Yokoyama, S.-Y. Jung, H. Masuda, A. Kawamoto, Y.M. Lee, T. Asahara, Differential activity of bone marrow hematopoietic stem cell subpopulations for EPC development and ischemic neovascularization, J. Mol. Cell. Cardiol. 51 (3) (2011) 308–317.

[316] H. Kim, S. Kim, S.H. Baek, S.-M. Kwon, Pivotal cytoprotective mediators and promising therapeutic strategies for endothelial progenitor cell-based cardiovascular regeneration, Stem Cells Int. 2016 (2016) 8340257.

[317] CHMP, MACI - Matrix applied characterised autologous cultured chondrocytes, 2013.

[318] E. Martin-Rendon, Meta-analyses of human cell-based cardiac regeneration therapies: what can systematic reviews tell us about cell therapies for ischemic heart disease? Circ. Res. 118 (8) (2016) 1264–1272.

[319] CHMP, Spherox: spheroids of human autologous matrix-associated chondrocytes. (EMA/CHMP/304021/2017), 2017.

[320] CHMP, Holoclar, 2015.

[321] CHMP, Glybera, INN - alipogene tiparvovec, 2012.

[322] CHMP, Imlygic, INN - talimogene laherparepvec, 2016.

[323] CHMP, Strimvelis, Common name - autologous CD34+ enriched cell fraction that contains CD34+ cells transduced with retroviral vector that encodes for the human ADA cDNA sequence, 2016.

[324] CHMP, Zalmoxis, common name: allogeneic T cells genetically modified with a retroviral vector encoding for a truncated form of the human low affinity nerve growth factor receptor (ΔLNGFR) and the herpes simplex I virus thymidine kinase (HSV-TK Mut2), 2016.

[325] H.T. Vestergaard, L.D. Apote, C.K. Schneider, C. Herberts, The evolution of nonclinical regulatory science: advanced therapy medicinal products as a paradigm, Mol. Ther. 21 (9) (2013) 1644–1648.

[326] European Medicines Agency, Advanced Therapy Medicines: Exploring Solutions to Foster Development and Expand Patient Access in Europe (EMA/345874), 2016.

[327] Y. Fujita, A. Kawamoto, Regenerative medicine legislation in Japan for fast provision of cell therapy products, Clin. Pharmacol. Ther. 99 (1) (2016) 26–29.

[328] M.S. Corbett, A. Webster, R. Hawkins, N. Woolacott, Innovative regenerative medicines in the EU: a better future in evidence? BMC Med. 15 (1) (2017) 49.

[329] R. Maciulaitis, L. D'Apote, A. Buchanan, L. Pioppo, C.K. Schneider, Clinical development of advanced therapy medicinal products in Europe: evidence that regulators must be proactive, Mol. Ther. 20 (3) (2012) 479–482.

[330] L. Foley, M. Whitaker, Concise review: cell therapies: the route to widespread adoption, Stem Cells Transl. Med. 1 (5) (2012) 438–447.

[331] F. Trindade, A. Leite-Moreira, J. Ferreira-Martins, R. Ferreira, I. Falcao-Pires, R. Vitorino, Towards the standardization of stem cell therapy studies for ischemic heart diseases: bridging the gap between animal models and the clinical setting, Int. J. Cardiol. 228 (2017) 465–480.

[332] K.C. Wollert, G.P. Meyer, J. Müller-Ehmsen, C. Tschöpe, V. Bonarjee, A.I. Larsen, A.E. May, K. Empen, E. Chorianopoulos, U. Tebbe, J. Waltenberger, H. Mahrholdt, B. Ritter, J. Pirr, D. Fischer, M. Korf-Klingebiel, L. Arseniev, H.-G. Heuft, J.E. Brinchmann, D. Messinger, B. Hertenstein, A. Ganser, H.A. Katus, S.B. Felix, M.P. Gawaz, K. Dickstein, H.-P. Schultheiss, D. Ladage, S. Greulich, J. Bauersachs, Intracoronary autologous bone marrow cell transfer after myocardial infarction: the BOOST-2 randomised placebo-controlled clinical trial, Eur. Heart J. (2017, Apr 19) https://doi.org/10.1093/eurheartj/ehx188.

[333] Editorial, A futile cycle in cell therapy, Nat. Biotechnol. 35 (4) (2017) 291.

[334] R. Connolly, T. O'Brien, G. Flaherty, Stem cell tourism—a web-based analysis of clinical services available to international travellers, Travel Med. Infect. Dis. 12 (6 Pt B) (2014) 695–701.

[335] O. Wolkenhauer, Why model? Front. Physiol. 5 (2014) 21.

[336] F. Montecucco, F. Carbone, F.L. Dini, M. Fiuza, F.J. Pinto, A. Martelli, D. Palombo, G. Sambuceti, F. Mach, R. de Caterina, Implementation strategies of systems medicine in clinical research and home care for cardiovascular disease patients, Eur. J. Intern. Med. 25 (9) (2014) 785–794.

[337] E. Björnson, J. Borén, A. Mardinoglu, Personalized Cardiovascular Disease Prediction and Treatment-A Review of Existing Strategies and Novel Systems Medicine Tools, 2016.

[338] A. Lysenko, I.A. Roznovăț, M. Saqi, A. Mazein, C.J. Rawlings, C. Auffray, Representing and querying disease networks using graph databases, BioData Min. 9 (2016) 23.

[339] P. Pareja-Tobes, R. Tobes, M. Manrique, E. Pareja, E. Pareja-Tobes, Bio4j: A High-Performance Cloud-Enabled Graph-Based Data Platform, bioRxiv, 2015.

[340] K. Wolstencroft, S. Owen, O. Krebs, Q. Nguyen, N.J. Stanford, M. Golebiewski, A. Weidemann, M. Bittkowski, L. An, D. Shockley, J.L. Snoep, W. Mueller, C. Goble, SEEK: a systems biology data and model management platform, BMC Syst. Biol. 9 (2015) 33.

[341] G.B. Ehret, P.B. Munroe, K.M. Rice, M. Bochud, A.D. Johnson, D.I. Chasman, A.V. Smith, M.D. Tobin, G.C. Verwoert, S.-J. Hwang, V. Pihur, P. Vollenweider, P.F. O'Reilly, N. Amin, J.L. Bragg-Gresham, A. Teumer, N.L. Glazer, L. Launer, J.H. Zhao, Y. Aulchenko, S. Heath, S. Sõber, A. Parsa, J. Luan, P. Arora, A. Dehghan, F. Zhang, G. Lucas, A.A. Hicks, A.U. Jackson, J.F. Peden, T. Tanaka, S.H. Wild, I. Rudan, W. Igl, Y. Milaneschi, A.N. Parker, C. Fava, J.C. Chambers, E.R. Fox, M. Kumari, M.J. Go, P. van der Harst, W.H.L. Kao, M. Sjögren, D.G. Vinay, M. Alexander, Y. Tabara, S. Shaw-Hawkins, P.H. Whincup, Y. Liu, G. Shi, J. Kuusisto, B. Tayo, M. Seielstad, X. Sim, K.-D.H. Nguyen, T. Lehtimäki, G. Matullo, Y. Wu, T.R. Gaunt, N.C. Onland-Moret, M.N. Cooper, C.G.P. Platou, E. Org, R. Hardy, S. Dahgam, J. Palmen, V. Vitart, P.S. Braund, T. Kuznetsova, C.S.P.M. Uiterwaal, A. Adeyemo, W. Palmas, H. Campbell, B. Ludwig, M. Tomaszewski, I. Tzoulaki, N.D. Palmer, T. Aspelund, M. Garcia, Y.-P.C. Chang, J.R. O'Connell, N.I. Steinle, D.E. Grobbee, D.E. Arking, S.L. Kardia, A.C. Morrison, D. Hernandez, S. Najjar, W.L. McArdle, D. Hadley, M.J. Brown, J.M. Connell, A.D. Hingorani, I.N.M. Day, D.A. Lawlor, J.P. Beilby, R.W. Lawrence, R. Clarke, J.C. Hopewell, H. Ongen, A.W. Dreisbach, Y. Li, J.H. Young, J.C. Bis, M. Kähönen, J. Viikari, L.S. Adair, N.R. Lee, M.-H. Chen, M. Olden, C. Pattaro, J.A.H. Bolton, A. Köttgen, S. Bergmann, V. Mooser, N. Chaturvedi, T.M. Frayling, M. Islam, T.H. Jafar, J. Erdmann, S.R. Kulkarni, S.R. Bornstein, J. Grässler, L. Groop, B.F. Voight, J. Kettunen, P. Howard, A. Taylor, S. Guarrera, F. Ricceri, V. Emilsson, A.

Plump, I. Barroso, K.-T. Khaw, A.B. Weder, S.C. Hunt, Y.V. Sun, R.N. Bergman, F.S. Collins, L.L. Bonnycastle, L.J. Scott, H.M. Stringham, L. Peltonen, M. Perola, E. Vartiainen, S.-M. Brand, J.A. Staessen, T.J. Wang, P.R. Burton, M. Soler Artigas, Y. Dong, H. Snieder, X. Wang, H. Zhu, K.K. Lohman, M.E. Rudock, S.R. Heckbert, N.L. Smith, K.L. Wiggins, A. Doumatey, D. Shriner, G. Veldre, M. Viigimaa, S. Kinra, D. Prabhakaran, V. Tripathy, C.D. Langefeld, A. Rosengren, D.S. Thelle, A.M. Corsi, A. Singleton, T. Forrester, G. Hilton, C.A. McKenzie, T. Salako, N. Iwai, Y. Kita, T. Ogihara, T. Ohkubo, T. Okamura, H. Ueshima, S. Umemura, S. Eyheramendy, T. Meitinger, H.-E. Wichmann, Y.S. Cho, H.-L. Kim, J.-Y. Lee, J. Scott, J.S. Sehmi, W. Zhang, B. Hedblad, P. Nilsson, G.D. Smith, A. Wong, N. Narisu, A. Stančáková, L.J. Raffel, J. Yao, S. Kathiresan, C.J. O'Donnell, S.M. Schwartz, M.A. Ikram, W.T. Longstreth, T.H. Mosley, S. Seshadri, N.R.G. Shrine, L.V. Wain, M.A. Morken, A.J. Swift, J. Laitinen, I. Prokopenko, P. Zitting, J.A. Cooper, S.E. Humphries, J. Danesh, A. Rasheed, A. Goel, A. Hamsten, H. Watkins, S.J.L. Bakker, W.H. van Gilst, C.S. Janipalli, K.R. Mani, C.S. Yajnik, A. Hofman, F.U.S. Mattace-Raso, B.A. Oostra, A. Demirkan, A. Isaacs, F. Rivadeneira, E.G. Lakatta, M. Orru, A. Scuteri, M. Ala-Korpela, A.J. Kangas, L.-P. Lyytikäinen, P. Soininen, T. Tukiainen, P. Würtz, R.T.-H. Ong, M. Dörr, H.K. Kroemer, U. Völker, H. Völzke, P. Galan, S. Hercberg, M. Lathrop, D. Zelenika, P. Deloukas, M. Mangino, T.D. Spector, G. Zhai, J.F. Meschia, M.A. Nalls, P. Sharma, J. Terzic, M.V.K. Kumar, M. Denniff, E. Zukowska-Szczechowska, L.E. Wagenknecht, F.G.R. Fowkes, F.J. Charchar, P.E.H. Schwarz, C. Hayward, X. Guo, C. Rotimi, M.L. Bots, E. Brand, N.J. Samani, O. Polasek, P.J. Talmud, F. Nyberg, D. Kuh, M. Laan, K. Hveem, L.J. Palmer, Y.T. van der Schouw, J.P. Casas, K.L. Mohlke, P. Vineis, O. Raitakari, S.K. Ganesh, T.Y. Wong, E.S. Tai, R.S. Cooper, M. Laakso, D.C. Rao, T.B. Harris, R.W. Morris, A.F. Dominiczak, M. Kivimaki, M.G. Marmot, T. Miki, D. Saleheen, G.R. Chandak, J. Coresh, G. Navis, V. Salomaa, B.-G. Han, X. Zhu, J.S. Kooner, O. Melander, P.M. Ridker, S. Bandinelli, U.B. Gyllensten, A.F. Wright, J.F. Wilson, L. Ferrucci, M. Farrall, J. Tuomilehto, P.P. Pramstaller, R. Elosua, N. Soranzo, E.J.G. Sijbrands, D. Altshuler, R.J.F. Loos, A.R. Shuldiner, C. Gieger, P. Meneton, A.G. Uitterlinden, N.J. Wareham, V. Gudnason, J.I. Rotter, R. Rettig, M. Uda, D.P. Strachan, J.C.M. Witteman, A.-L. Hartikainen, J.S. Beckmann, E. Boerwinkle, R.S. Vasan, M. Boehnke, M.G. Larson, M.-R. Järvelin, B.M. Psaty, G.R. Abecasis, A. Chakravarti, P. Elliott, C.M. van Duijn, C. Newton-Cheh, D. Levy, M.J. Caulfield, T. Johnson, Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk, Nature 478 (7367) (2011) 103–109.

[342] H. Wang, Y. Chen, X. Li, G. Chen, L. Zhong, G. Chen, Y. Liao, W. Liao, J. Bin, Genome-wide analysis of alternative splicing during human heart development, Sci Rep 6 (1) (2016) 35520.

[343] K. Chen, X. Dai, J. Wu, Alternative splicing: an important mechanism in stem cell biology, World J. Stem Cells 7 (1) (2015) 1–10.

[344] T. Huan, T. Esko, M.J. Peters, L.C. Pilling, K. Schramm, C. Schurmann, B.H. Chen, C. Liu, R. Joehanes, A.D. Johnson, C. Yao, S.-X. Ying, P. Courchesne, L. Milani, N. Raghavachari, R. Wang, P. Liu, E. Reinmaa, A. Dehghan, A. Hofman, A.G. Uitterlinden, D.G. Hernandez, S. Bandinelli, A. Singleton, D. Melzer, A. Metspalu, M. Carstensen, H. Grallert, C. Herder, T. Meitinger, A. Peters, M. Roden, M. Waldenberger, M. Dörr, S.B. Felix, T. Zeller, R. Vasan, C.J. O'Donnell, P.J. Munson, X. Yang, H. Prokisch, U. Völker, Joyce B.J. van Meurs, L. Ferrucci, D. Levy, M.I. McCarthy, A meta-analysis of gene expression signatures of blood pressure and hypertension, PLoS Genet. 11 (3) (2015), e1005035.

[345] T. Huan, Q. Meng, M.A. Saleh, A.E. Norlander, R. Joehanes, J. Zhu, B.H. Chen, B. Zhang, A.D. Johnson, S. Ying, P. Courchesne, N. Raghavachari, R. Wang, P. Liu, International Consortium for Blood Pressure GWAS, C.J. O'Donnell, R. Vasan, P.J. Munson, M.S. Madhur, D.G. Harrison, X. Yang, D. Levy, Integrative network analysis reveals molecular mechanisms of blood pressure regulation, Mol. Syst. Biol. 11 (1) (2015) 799.

[346] T. Zhu, L. Qiao, Q. Wang, R. Mi, J. Chen, Y. Lu, J. Gu, Q. Zheng, T-box family of transcription factor-TBX5, insights in development and disease, Am. J. Transl. Res. 9 (2) (2017) 442–453.

[347] S. Ounzain, R. Micheletti, T. Beckmann, B. Schroen, M. Alexanian, I. Pezzuto, S. Crippa, M. Nemir, A. Sarre, R. Johnson, J. Dauvillier, F. Burdet, M. Ibberson, R. Guigó, I. Xenarios, S. Heymans, T. Pedrazzini, Genome-wide profiling of the cardiac transcriptome after myocardial infarction identifies novel heart-specific long non-coding RNAs, Eur. Heart J. 36 (6) (2015) (353-68a).

[348] P. Mathiyalagan, S.T. Keating, X.-J. Du, A. El-Osta, Interplay of chromatin modifications and non-coding RNAs in the heart, Epigenetics 9 (1) (2014) 101–112.

[349] S.C. Lott, M. Wolfien, K. Riege, A. Bagnacani, O. Wolkenhauer, S. Hoffmann, W.R. Hess, Customized workflow development and data modularization concepts for RNA-sequencing and metatranscriptome experiments, J. Biotechnol. (2017).

[350] S. Lampa, M. Dahlö, P.I. Olason, J. Hagberg, O. Spjuth, Lessons learned from implementing a national infrastructure in Sweden for storage and analysis of next-generation sequencing data, GigaScience 2 (1) (2013) 9.

[351] E. Afgan, D. Baker, M. van den Beek, D. Blankenberg, D. Bouvier, M. Čech, J. Chilton, D. Clements, N. Coraor, C. Eberhard, B. Grüning, A. Guerler, J. Hillman-Jackson, G. Von Kuster, E. Rasche, N. Soranzo, N. Turaga, J. Taylor, A. Nekrutenko, J. Goecks, The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update, Nucleic Acids Res. 44 (W1) (2016) W3–W10.

[352] B.A. Grüning, J. Fallmann, D. Yusuf, S. Will, A. Erxleben, F. Eggenhofer, T. Houwaart, B. Batut, P. Videm, A. Bagnacani, M. Wolfien, S.C. Lott, Y. Hoogstrate, W.R. Hess, O. Wolkenhauer, S. Hoffmann, A. Akalin, U. Ohler, P.F. Stadler, R. Backofen, The RNA workbench: best practices for RNA and high-throughput sequencing bioinformatics in galaxy, Nucleic Acids Res. 45 (2017) D626–D634.

[353] M. Wolfien, C. Rimmbach, U. Schmitz, J.J. Jung, S. Krebs, G. Steinhoff, R. David, O. Wolkenhauer, TRAPLINE: a standardized and automated pipeline for RNA sequencing data analysis, evaluation and annotation, BMC Bioinformatics 17 (2016) 21.

[354] B.A. Grüning, E. Rasche, B. Rebolledo-Jaramillo, C. Eberhard, T. Houwaart, J. Chilton, N. Coraor, R. Backofen, J. Taylor, A. Nekrutenko, Jupyter and galaxy: easing entry barriers into complex data analyses for biomedical researchers, PLoS Comput. Biol. 13 (5) (2017), e1005425.

[355] A.J. Lusis, J.N. Weiss, Cardiovascular networks: systems-based approaches to cardiovascular disease, Circulation 121 (1) (2010) 157–170.

[356] G. Gambardella, M.N. Moretti, R. de Cegli, L. Cardone, A. Peron, D. Di Bernardo, Differential network analysis for the identification of condition-specific pathway activity and regulation, Bioinformatics 29 (14) (2013) 1776–1785.

[357] J. Menche, A. Sharma, M. Kitsak, S.D. Ghiassian, M. Vidal, J. Loscalzo, A.-L. Barabási, Disease networks. Uncovering disease-disease relationships through the incomplete interactome, Science (New York, N.Y.) 347 (6224) (2015) (1257601).

[358] J. Liu, L. Jing, X. Tu, Weighted gene co-expression network analysis identifies specific modules and hub genes related to coronary artery disease, BMC Cardiovasc. Disord. 16 (2016) 54.

[359] S. Ballouz, W. Verleyen, J. Gillis, Guidance for RNA-seq co-expression network construction and analysis: safety in numbers, Bioinformatics 31 (13) (2015) 2123–2130.

[360] D.C. Goff, D.M. Lloyd-Jones, G. Bennett, S. Coady, R.B. D'Agostino, R. Gibbons, P. Greenland, D.T. Lackland, D. Levy, C.J. O'Donnell, J.G. Robinson, J.S. Schwartz, S.T. Shero, S.C. Smith, P. Sorlie, N.J. Stone, P.W.F. Wilson, 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines, J. Am. Coll. Cardiol. 63 (25 Pt B) (2014) 2935–2959.

[361] S.F. Weng, J. Reps, J. Kai, J.M. Garibaldi, N. Qureshi, H.G. Leufkens, B. Liu, Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One 12 (4) (2017), e0174944.

[362] Z. Obermeyer, E.J. Emanuel, Predicting the future — big data, machine learning, and clinical medicine, N. Engl. J. Med. 375 (13) (2016) 1216–1219.

[363] S. Dreiseitl, L. Ohno-Machado, Logistic regression and artificial neural network classification models: a methodology review, J. Biomed. Inform. 35 (5–6) (2002) 352–359.

[364] A.M. Shah, B. Claggett, N.K. Sweitzer, S.J. Shah, I.S. Anand, L. Liu, B. Pitt, M.A. Pfeffer, S.D. Solomon, Prognostic importance of impaired systolic function in heart failure with preserved ejection fraction and the impact of spironolactone, Circulation 132 (5) (2015) 402–414.

[365] H.J. Vrijhoef, A.G. de Belvis, M. de La Calle, M.S. de Sabata, B. Hauck, S. Montante, A. Moritz, D. Pelizzola, M. Saraheimo, N.A. Guldemond, IT-supported integrated care pathways for diabetes: a compilation and review of good practices, Int. J. Care Coord. 20 (1–2) (2017) 26–40.

[366] P.P. Sengupta, Y.-M. Huang, M. Bansal, A. Ashrafi, M. Fisher, K. Shameer, W. Gall, J.T. Dudley, Cognitive machine-learning algorithm for cardiac imaging: a pilot study for differentiating constrictive pericarditis from restrictive cardiomyopathy, Circ. Cardiovasc. Imaging 9 (6) (2016).

[367] Sanne J. Jansen of Lorkeers, P.A. Doevendans, S.A.J. Chamuleau, All preclinical trials should be registered in advance in an online registry, Eur. J. Clin. Investig. 44 (9) (2014) 891–892.

[368] A.N. Nowbar, M. Mielewczik, M. Karavassilis, H.-M. Dehbi, M.J. Shun-Shin, S. Jones, J.P. Howard, G.D. Cole, D.P. Francis, Discrepancies in autologous bone marrow stem cell trials and enhancement of ejection fraction (DAMASCENE): weighted regression and meta-analysis, BMJ 348 (2014) g2688.

[369] I. Louria-Hayon, C. Frelin, J. Ruston, G. Gish, J. Jin, M.M. Kofler, J.-P. Lambert, H.A. Adissu, M. Milyavsky, R. Herrington, M.D. Minden, J.E. Dick, A.-C. Gingras, N.N. Iscove, T. Pawson, Lnk adaptor suppresses radiation resistance and radiation-induced B-cell malignancies by inhibiting IL-11 signaling, Proc. Natl. Acad. Sci. U. S. A. 110 (51) (2013) 20599–20604.

[370] M. Ishige-Wada, S.-M. Kwon, M. Eguchi, K. Hozumi, H. Iwaguro, T. Matsumoto, N. Fukuda, H. Mugishima, H. Masuda, T. Asahara, Jagged-1 signaling in the bone marrow microenvironment promotes endothelial progenitor cell expansion and commitment of CD133+ human cord blood cells for postnatal Vasculogenesis, PLoS One 11 (11) (2016), e0166660.

[371] J. Fitau, G. Boulday, F. Coulon, T. Quillard, B. Charreau, The adaptor molecule Lnk negatively regulates tumor necrosis factor-alpha-dependent VCAM-1 expression in endothelial cells through inhibition of the ERK1 and -2 pathways, J. Biol. Chem. 281 (29) (2006) 20148–20159.

[372] S. Gueller, S. Hehn, V. Nowak, S. Gery, H. Serve, C.H. Brandts, H.P. Koeffler, Adaptor protein Lnk binds to PDGF receptor and inhibits PDGF-dependent signaling, Exp. Hematol. 39 (5) (2011) 591–600.

[373] P.L. Auer, A. Teumer, U. Schick, A. O'Shaughnessy, K.S. Lo, N. Chami, C. Carlson, S. de Denus, M.-P. Dubé, J. Haessler, R.D. Jackson, C. Kooperberg, L.-P.L. Perreault, M. Nauck, U. Peters, J.D. Rioux, F. Schmidt, V. Turcot, U. Völker, H. Völzke, A. Greinacher, L. Hsu, J.-C. Tardif, G.A. Diaz, A.P. Reiner, G. Lettre, Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits, Nat. Genet. 46 (6) (2014) 629–634.

[374] R. McPherson, A. Tybjaerg-Hansen, Genetics of coronary artery disease, Circ. Res. 118 (4) (2016) 564–578.

[375] E.Y. Lavrikova, A.G. Nikitin, T.L. Kuraeva, V.A. Peterkova, N.M. Tsitlidze, D.A. Chistiakov, V.V. Nosikov, The carriage of the type 1 diabetes-associated R262W variant of human LNK correlates with increased proliferation of peripheral blood monocytes in diabetic patients, Pediatr. Diabetes 12 (2) (2011) 127–132.

[376] S.-M. Kwon, T. Suzuki, A. Kawamoto, M. Ii, M. Eguchi, H. Akimaru, M. Wada, T. Matsumoto, H. Masuda, Y. Nakagawa, H. Nishimura, K. Kawai, S. Takaki, T. Asahara, Pivotal role of lnk adaptor protein in endothelial progenitor cell biology for vascular regeneration, Circ. Res. 104 (8) (2009) 969–977.

2.3.2 Evaluation of cell therapies for the treatment of cardiac infarction

Lang, C.I., **Wolfien, M.**, ..., Wolkenhauer, O., ..., and Glass, Ä. (2017).
Cardiac Cell Therapies for the Treatment of Acute Myocardial Infarction:
A Meta-Analysis from Mouse Studies.
Cellular Physiology and Biochemistry. IF: 5.5, Citations (December 14, 2020): 18

Stem cell-based regenerative therapies for the treatment of ischemic myocardium are currently a subject of intensive investigation. A variety of cell populations have been demonstrated to be safe and to exert positive effects in human Phase I and II clinical trials; however, conclusive evidence of efficacy is not yet proven. While the relevance of animal models for appropriate pre-clinical safety and efficacy testing, including applications in Phase III studies, continues to increase, concerns are expressed regarding the validity of the mouse model to predict clinical results. Hundreds of preclinical studies for cardiac repair have assessed the efficacy of different cell preparations, including pluripotent stem cells, we undertook a systematic re-evaluation of the data from the mouse model, which initially paved the way for the first clinical trials in this field.

I developed a meta-analysis compliant workflow and conducted the statistical analyses. First, an estimator comparison was done based on the study variance τ^2 and inconsistency I^2. Furthermore, I checked for study biases via funnel-plot visualizations and Egger's regression testing. In the following, I utilized random-effects and fixed-effects module analyses to summarize study outcomes within a forest-plot. Finally, meta-regression analyses of different subgroups have been performed to identify moderators that have a significant influence to the left ventricular ejection fraction (LVEF) improvement.

In summary, we see the mouse as a valid model to evaluate the efficacy of cell-based therapies for the treatment of ischemic myocardial damage. However, further studies are required to understand the mechanisms underlying stem cell-based improvement of cardiac function after ischemia.

Cellular Physiology and Biochemistry

Cell Physiol Biochem 2017;42:254-268
DOI: 10.1159/000477324
Published online: May 19, 2017
Accepted: March 02, 2017

Original Paper

Cardiac Cell Therapies for the Treatment of Acute Myocardial Infarction: A Meta-Analysis from Mouse Studies

Cajetan Immanuel Lang[a,b] Markus Wolfien[c] Anne Langenbach[k] Paula Müller[b] Olaf Wolkenhauer[c,d] Arash Yavari[e,f,g] Hüseyin Ince[a] Gustav Steinhoff[b,j] Bernd Joachim Krause[h] Robert David[b,j] Änne Glass[i]

[a]Department of Cardiology, University Hospital Rostock, [b]Reference and Translation Center for Cardiac Stem Cell Therapy, University of Rostock, Rostock, [c]Department of Systems Biology and Bioinformatics, University of Rostock, Rostock, Germany; [d]Stellenbosch Institute of Advanced Study, Wallenberg Research Centre at Stellenbosch University, Stellenbosch, South Africa; [e]Experimental Therapeutics, Radcliffe Department of Medicine, University of Oxford, Oxford, [f]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, [g]The Wellcome Trust Centre for Human Genetics, Oxford, UK; [h]Department of Nuclear Medicine, University Hospital Rostock, Rostock, [i]Institute for Biostatistics and Informatics in Medicine and Ageing Research, University Hospital Rostock, Rostock, [j]Department Life, Light and Matter of the Interdisciplinary Faculty at Rostock University Rostock, Rostock, [k]Department of Anaesthesia and Intensive Care Medicine, University Hospital Rostock Germany

Key Words
Heart • Stem cell therapy • Mouse • Meta-analysis

Abstract
Aims: Stem cell-based regenerative therapies for the treatment of ischemic myocardium are currently a subject of intensive investigation. A variety of cell populations have been demonstrated to be safe and to exert some positive effects in human Phase I and II clinical trials, however conclusive evidence of efficacy is still lacking. While the relevance of animal models for appropriate pre-clinical safety and efficacy testing with regard to application in Phase III studies continues to increase, concerns have been expressed regarding the validity of the mouse model to predict clinical results. Against the background that hundreds of preclinical studies have assessed the efficacy of numerous kinds of cell preparations - including pluripotent stem cells - for cardiac repair, we undertook a systematic re-evaluation of data from the mouse model, which initially paved the way for the first clinical trials in this field. ***Methods and Results:*** A systematic literature screen was performed to identify publications reporting results of cardiac stem cell therapies for the treatment of myocardial ischemia in the mouse model. Only peer-reviewed and placebo-controlled studies using magnet resonance imaging (MRI) for left ventricular ejection fraction (LVEF) assessment were included. Experimental data from 21 studies involving 583 animals demonstrate a significant improvement in LVEF of 8.59%

C.I. Lang, M. Wolfien, R. David and Änne Glass contributed equally to this work.

Prof. Dr. Robert David
Schillingallee, 68, 18057 Rostock (Germany)
Tel. +49 381-4 94 6049, Fax +49 381-4 94 61 02, E-Mail robert.david@med.uni-rostock.de

KARGER

Cellular Physiology and Biochemistry
Cell Physiol Biochem 2017;42:254-268
DOI: 10.1159/000477324
Published online: May 19, 2017

+/- 2.36; p=.012 (95% CI, 3.7–13.8) compared with control animals. ***Conclusion:*** The mouse is a valid model to evaluate the efficacy of cell-based advanced therapies for the treatment of ischemic myocardial damage. Further studies are required to understand the mechanisms underlying stem cell based improvement of cardiac function after ischemia.

Introduction

Despite rapid advancements in both pharmacological and interventional treatment options, coronary heart disease remains the most common cause of death in Europe - accounting for 1.8 million deaths each year [1]. The quest for new therapeutic approaches to prevent adverse myocardial remodelling post-infarction and limit the subsequent development of irreversible heart failure gave rise to the field of cardiac stem cell therapy. Pivotal trials rapidly pushed stem cell therapies from bench to bedside, with small animal models in particular – namely mice and rats – serving as the basis for safety and efficacy testing [2-4].

Since the first patient was treated with intracoronary infusion of bone marrow stem cells in 2001 [2], numerous Phase I and Phase II studies have repeatedly shown the safety and feasibility of various cardiac stem cell therapies [5]. While these studies proved the excellent safety profile of the tested cell products, there still remains a paucity of data on efficacy because of the small numbers of patients included and the lack of statistical power. Accordingly, several groups are currently recruiting patients for Phase III clinical trials aiming to robustly address the issue of clinical efficacy of stem cell therapy for myocardial repair. [6] (*NCT01768702*, *NCT02059512*, *NCT01569178*)

Until now, clinical trials have largely utilised cell types from the bone marrow that are readily available. These do not necessarily reflect stem cell populations with high potential to regenerate myocardium, however [7]. Accordingly, pluripotent stem cells have been intensively investigated as a source for the generation of cardiomyocytes [8], cells of the conduction system [9] or cardiovascular progenitors [10]. Clinical translation of these highly advanced cell products requires new methods for appropriate safety and efficacy testing with regard to application in patients. The role of animal models to meet these requirements and ensure a full understanding of the biology of stem cell-based therapies is substantially and continually increasing [11].

Due to similarities in heart rate, anatomical and physiological parameters, large animal models have been advocated as superior to rodents in their ability to predict the results of clinical studies in cardiac regeneration [11, 12]. To our knowledge, however, there is no conclusive evidence supporting the contention that large animal models are superior to rodents – particularly mice – for efficacy testing of cardiac stem cell therapies.

Against the background that the mouse model is cost-effective, readily genetically modified and that over 25 years of experience in the field of murine embryonic stem cell research exists, we performed a meta-analysis to assess the validity of mouse models to predict improvement in left ventricular ejection fraction (LVEF) in clinical trials of regenerative stem cell therapy. To ensure comparability, we included only controlled studies, which assessed LVEF as a surrogate parameter for efficacy using magnetic resonance imaging (MRI), representing the gold standard for assessment of LVEF in humans.

Materials and Methods

Search strategy

Articles published on Medline between January 1980 and October 2015 were searched via PubMed using the following search terms: TERM A: "(Mouse) AND (stem cells OR progenitor cells OR bone marrow OR mesenchymal OR hematopoietic) AND (myocardial infarction OR cardiac repair OR myocardial regeneration)" and TERM B: "(Mouse) AND (stem cells OR progenitor cells OR bone marrow OR mesenchymal

Cellular Physiology and Biochemistry
Cell Physiol Biochem 2017;42:254-268
DOI: 10.1159/000477324
Published online: May 19, 2017

OR hematopoietic) AND (myocardial infarction OR cardiac repair OR myocardial regeneration) AND (MRI)". Only English, peer-reviewed and published reports were included. The retrieved studies were carefully examined to exclude potentially duplicate or overlapping data.

Eligibility criteria: Inclusion/exclusion of articles

The abstracts of all studies retrieved by the above mentioned search terms were reviewed. Whenever the respective abstract did not provide enough data for a decision based on our predefined eligibility criteria (Fig. 1), the material and methods section was carefully studied.

Data abstraction

The following information was extracted from complete manuscripts of eligible studies: basal characteristics of the study and LVEF. If necessary, data were estimated from graphics or recalculated by available data [13]. Standard deviations were determined or recalculated from standard errors and *vice versa*. In the final analysis, only studies using MRI for the assessment of LVEF were included. Data derived by echocardiography, nuclear imaging, or pressure-volume loops were excluded. In cases of missing data, corresponding authors were contacted, with two authors from six separate manuscripts responding [13].

Data analysis

For the first time, we performed a random effects meta-analysis and fixed effects meta-regression analysis that included all available data on Medline regarding cardiac stem cell therapies post-acute myocardial infarction (MI) or in the setting of chronic myocardial ischemia in mice meeting our eligibility criteria. Our primary effect size was the difference in mean LVEF (reported in %) at follow up between control and treated animals. Both groups underwent MI induction. In the case of multiple measurements over time, data measured at the longest duration of follow up were used for analysis. Only data within the range of one to six weeks after cell application were included.

We have compared numerous statistical models and chosen the maximum-likelihood (ML) estimator [14] to conduct our random effects meta-analysis model. The subsequently obtained continuous variables are reported as weighted mean differences (calculated via the weighted least square algorithm), together with 95% confidence interval (CI), between cell-treated mice and control groups. Our choice to use the ML estimator rather than the restricted maximum-likelihood (REML) is based on the credibility interval of ML which is the same as the REML interval, but where ML covers the effects of nuisance parameters [15].

Overall homogeneity of differences in mean LVEF of single studies was evaluated based on Cochrane's chi-squared test and the estimator τ^2. As described in Higgins et al. [16] heterogeneity was considered significant at p<.1. Inconsistency was estimated by using the I^2 statistic; values of 25%, 50%, and 75% were considered low, moderate, and high inconsistency, known respectively as "Higgins rule of Thumb" [16].

Using meta-regression analysis, the following subgroup analyses were performed: cell type (embryonic stem cells (ESC), mesenchymal stem cells (MSC), adult mesodermal cells (MC) and cardiovascular cells (CVC)), cell origin (autologous, heterologous, xenogeneic), number of injected cells ($1\text{-}4*10^5$; $0.5\text{-}3*10^6$), number of injections (1-2; 3-4), measurement time points (1-3 weeks; >3 weeks), gender of recipient (female or male) and the injected cell types (human, mice).

Funnel plots, Egger's weighed regression and "the trim and fill method" of Duval et al. [17] were used to detect potential publication bias [18].

All analyses were performed with *R* (Version 3.2.0). In particular, the meta-analysis was performed with the *metafor* package [19] and the statistical tests and power analysis were computed by using the *pwr* package [20].

Results

Electronic searching identified 1,394 publications. After review of the respective abstracts – and where necessary the material and methods section – 1,226 articles were excluded. Assessment of the remaining 228 papers resulted in 38 articles for detailed evaluation, including supplemental material. A final 21 studies involving 583 animals were identified to meet pre-specified inclusion criteria (Fig. 1). Our restrictive selection strategy

Cellular Physiology and Biochemistry
Cell Physiol Biochem 2017;42:254-268
DOI: 10.1159/000477324
Published online: May 19, 2017

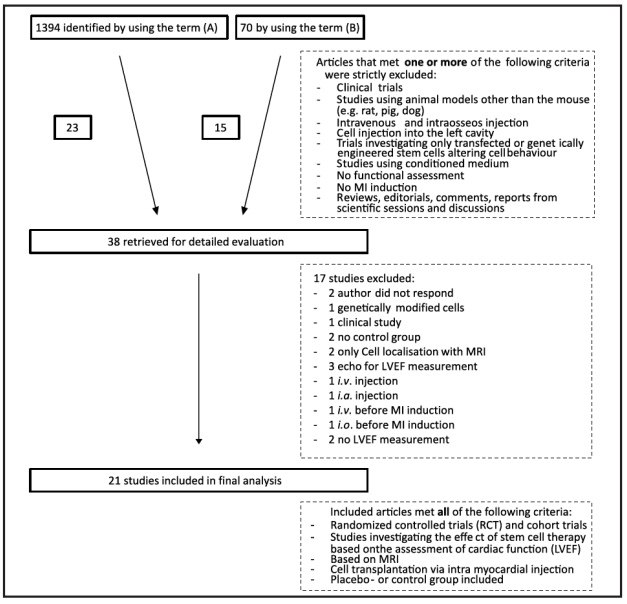


Fig. 1. Flow-chart illustrating the selection process for included studies. Only studies using unmodified stem cell injections into the heart of mice and which employed MRI for LVEF assessment were considered. *i.a. – intra arterial*, *i.v. – intravenous*, *i.o. – intraosseous.*

aimed to identify a comparable group of controlled studies all using the same modality for LVEF assessment, a key consideration given the small size of the murine heart.

Study characteristics

A total of 583 mice from 21 studies containing 34 groups for comparison of the primary endpoint (LVEF) were included in the meta-analysis. All studies were published between 2006 and 2015. Only young adult animals, aged 8 to 16 weeks, were used. Myocardial infarction was induced by surgical permanent occlusion of the left anterior descending artery (LAD) in 28 groups and by surgical cryo-injury in one group. 100,000 to 3,000,000 cells suspended in saline, PBS or medium were injected directly into the myocardium using a syringe, constituting 5 - 30µl applied via 1 - 5 single injections. More details are provided in Table 1. Treated animals received cell suspensions, whereas control animals received saline, PBS or medium (suspending agent) alone. LVEF served as a surrogate measure of effect in all studies. The follow-up period lasted from one to six weeks. Survival curves and mortality data were only provided by four studies and, thus, not included in the meta-analysis. Data on left ventricular end-diastolic volume and left ventricular end-systolic volume were not reported in seven groups and thus not considered. None of the studies included reported safety end points such as "Major Cardiovascular Adverse Events" (MACE). In 22 of the 34 included groups, immunodeficient mice were used, yet tumor – more precisely teratoma – formation in animals receiving murine PSCs were reported in two of these groups only [21, 22].

Meta-analysis

The random effects meta-analysis model revealed that cardiac stem cell therapies, including both adult and pluripotent stem cells, significantly improve left ventricular systolic function after ischemic damage in mice: cell injection leads to an increase in LVEF of 8.59% +/- 2.36 (95% CI 3.7 – 13.8; p=.012, Fig. 2) in treated animals in comparison to control animals receiving suspending agent only. This improvement is referred to as the "overall" effect in the following section.

The model was chosen after performing the test for homogeneity (p=.083, *Fig. 3*) and calculating the between-study variance (τ^2=.01, Fig. 3), as well as inconsistency (I^2=33.39%,

Cellular Physiology and Biochemistry
Cell Physiol Biochem 2017;42:254-268
DOI: 10.1159/000477324
Published online: May 19, 2017

Author	n_c	n_t	Cell type	Donor	Host	Sex	Number of Cells	Number of Injections	Volume	Suspending Agent	Follow-up (weeks)
Arai et al. [57]	5	7	ESC	129Sv/J	129Sv/J	f	250000	1	25	medium	4
Au et al. [58]	5	5	ESC	129Sv/J	CD-1 (ICR)	n/a	300000	1	10	medium	1
Bai et al. [59]	7	7	fASC	human	SCID	m	500000	2	30	PBS	4
		7	ASC	human	SCID	m	500000	2	30	PBS	4
denHaan et al. [60]	12	11	CMPC	human	NOD/SCID	m	200000	5	20	medium	2
Drey et al. [61]	5	5	MSC	C57BL/6	C57BL/6	m	100000	1	10	PBS	4
Grauss et al. [62]	14	12	MSC	human	NOD/SCID	m	200000	5	20	medium	2
Grauss et al. [63]	12	10	MSC	human	NOD/SCID	m	200000	5	20	medium	2
Hendry et al. [21]	6	23	ESC	129/Sv	Scid/beige	f	250000	n/a	25	saline	1
		17	EF	n/a	Scid/beige	f	250000	n/a	25	saline	1
Huber et al. [64]	7	8	ESC-EC	human	FVB	n/a	2000000	2	30	PBS	4
Kim et al.[22]	9	9	AMC	human	Scid/beige	m	250000	n/a	20	matrigel	4
		8	AMC kit	human	Scid/beige	m	250000	n/a	20	matrigel	4
		9	iPSC	human	Scid/beige	m	250000	n/a	20	matrigel	4
Kofidis et al. [65]	5	5	ESC-CM	human	Scid/beige	f	1000000	n/a	25	medium	3
Liao et al. [66]	8	8	ESC-CM	129/Sv	CD-1 (ICR)	f	300000	1	10	medium	3
		8	ESC	129/Sv	CD-1 (ICR)	f	300000	1	10	medium	3
Liu et al.[67]	10	37	CPC	human	Scid/beige	f	1000000	3	20	PBS	4
Mauritz et al. [10]	13	14	iPSC Flk-	mouse	Scid/beige	n/a	500000	1	15	PBS	2
		11	iPSC Flk+	mouse	Scid/beige	n/a	500000	1	15	PBS	2
Ong et al. [68]	8	8	iPSC-CM	human	NOD/SCID	f	2000000	n/a	20	n/a	5
Paulis et al. [69]	8	5	eCM	CD-1	CD-1	m	$1\text{-}2x10^5$	n/a	5	medium	2
		7	SM	CD-1	CD-1	m	$1\text{-}2x10^5$	n/a	5	medium	2
		7	MSC	CD-1	CD-1	m	$1\text{-}2x10^5$	n/a	5	medium	2
Rojas et al. [70]	10	9	iPSC	mouse	Scid/beige	n/a	1000000	n/a	15	fibrinogen	2
Smits et al. [71]	9	11	CMPC	human	NOD/SCID	m	500000	2	10	PBS	4
		10	CMPC-CM	human	NOD/SCID	m	500000	2	10	PBS	4
vanLaake et al. [72]	13	12	ESC-CM	human	NOD/SCID	m	3000000	3	15	medium	4
		14	ESC-CM	human	NOD/SCID	m	1000000	1	15	medium	4
		12	ESC-non CM	human	NOD/SCID	m	1000000	1	15	medium	4
Wang et al. [73]	6	6	MSC	human	SCID	f	500000	n/a	25 - 30	medium	4
Winter et al.[74]	17	20	EPDC	human	NOD/SCID	n/a	400000	n/a	n/a	medium	6
		13	CMPC	human	NOD/SCID	n/a	400000	n/a	n/a	medium	6
		14	CMPC+EPDC	human	NOD/SCID	n/a	400000	n/a	n/a	medium	6

Table 1. Characteristics of included studies. ESC, embryonic stem cell; iPSC, induced pluripotent stem cell; fASC, fresh adipose tissue-derived stem cell; CMPC, cardiomyocyte progenitor cell; MSC, mesenchymal stem cell; EF, embryonic fibroblast; ESC-CM, ESC-derived cardiomyocytes; iPSC, induced pluripotent stem cell; eCM, embryonic cardiomyocyte; SM, skeletal myoblast; CMPC-CM, CMPC-derived cardiomyocytes; EPDC, epicardium derived cell; PBS, phosphate buffered saline; F, female; M, male; n/a, not applicable; 129Sv/J, CD-1, C57/BL6, 129/Sv: mouse strains; SCID, NOD/SCID SCID-beige: immune-deficient murine strains; n, number of animals (n_c, control; n_t, treated)

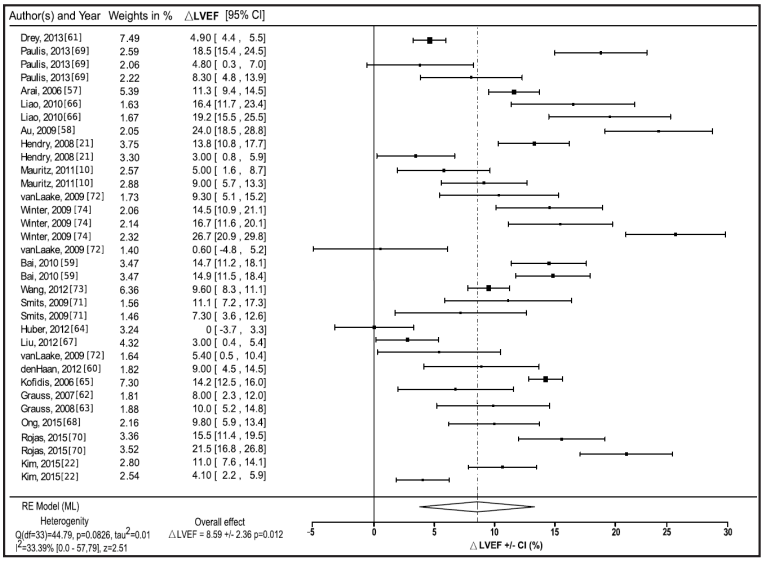


Fig. 2. Results of the meta-analysis visualized in a forest plot. Forest plot based on random effects model (maximum likelihood estimator; weights calculated by weighted least square algorithm) and difference in LVEF (reported in %) and the corresponding 95% confidence intervals.

Cellular Physiology and Biochemistry
Cell Physiol Biochem 2017;42:254-268
DOI: 10.1159/000477324
Published online: May 19, 2017
© 2017 The Author(s). Published by S. Karger AG, Basel
www.karger.com/cpb
Lang et al.: Meta-Analysis: Cardiac Stem Cell Therapies in Mice

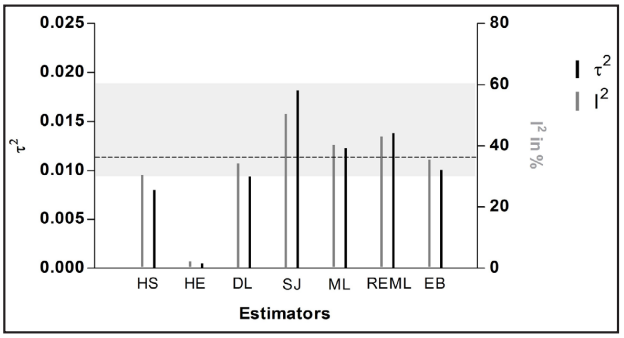


Fig. 3. Characterization of the random effect model estimator. Between study variance τ^2 and inconsistency I^2 to assess the heterogeneity of studies regarding Hunter-Schmidt estimator (HS), DerSimonian-Laird estimator (DL), Sidik-Jonkman estimator (SJ), the maximum-likelihood estimator (ML), restricted maximum-likelihood estimator (REML) and empirical Bayes estimator (EB). τ^2 is considered to be significant (p=.01) according to Salenti et al. [16]. I^2 evaluation according to Higgins et al. [56]: low (< 25%), moderate (25% - 75%) and high (>75%) inconsistency.

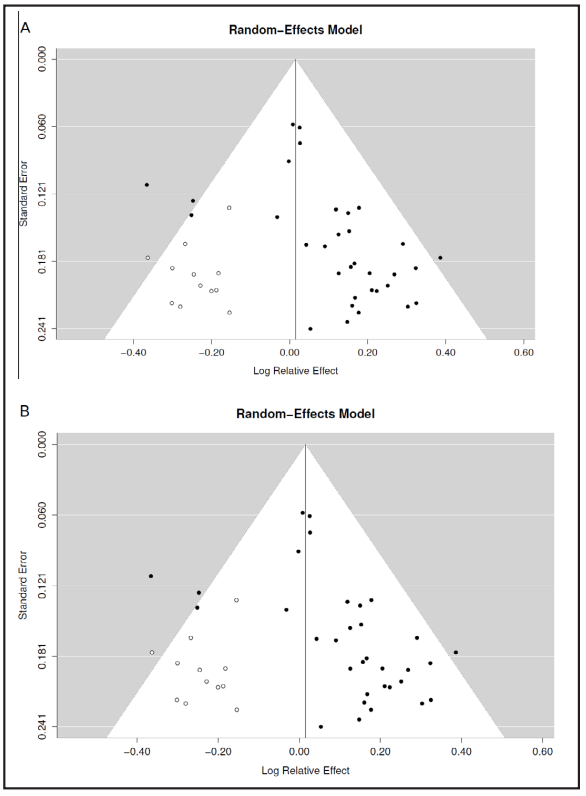


Fig. 4. Funnel plots assessing for publication bias. The funnel plots show observed study outcomes (black dots) and predicted missing studies (white dots) on the x-axis plotted against their corresponding standard errors on the y-axis. A vertical line indicates the estimate based on A, the random effects, or B the fixed-effects model. A pseudo confidence interval region is drawn around this value with bounds equal to ±1.96*SE, where SE is the mean standard error value from the y-axis.

Fig. 3). Furthermore, no publication bias could be detected (Fig. 4)*:* funnel plots demonstrated symmetric behaviour and Egger's regression test for symmetry was not significant (p=.4).

Subgroup analysis

In addition to assessing the overall efficacy of cardiac cell therapies in mice, we also performed a subgroup analysis using meta-regression to explore the impact of different cell types, cell origin, cell number, recipient gender, follow-up time and number of injections on LVEF (Table 2).

Cellular Physiology and Biochemistry
Cell Physiol Biochem 2017;42:254-268
DOI: 10.1159/000477324
Published online: May 19, 2017

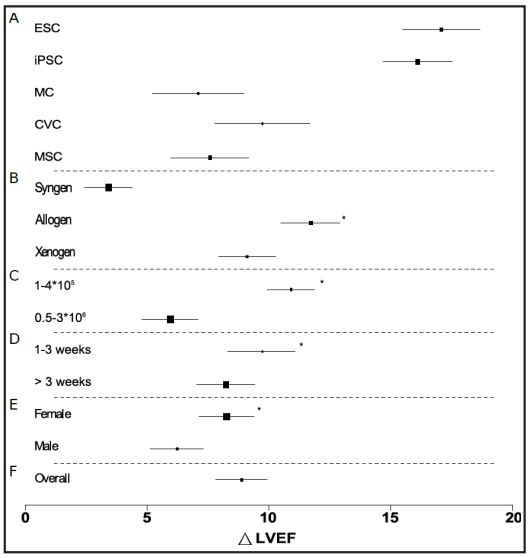


Fig. 5. Subgroup analysis to identify significant moderators. Meta-regression analysis of subgroups revealed factors that significantly influence the magnitude of the functional improvement afforded by cell therapies - using LVEF improvement as a surrogate marker for efficacy. A Cell type has no significant effect on the magnitude of LVEF improvement (p<.48). B Cell origin has an impact on efficacy: allogeneic cells are most effective (12.9%; p=.046). C Less than 500,000 cells are more effective than higher numbers (p=.013). D The highest increase in LVEF can be measured up to 3 weeks post-transplantation (p=.004). E Females benefit more from cardiac stem cell therapies than male mice (p=.003). F The overall effect of all investigated studies. *Marked as significant according to regression coefficient of the respective fixed-effects model.

Table 2. Summary of meta-regression analyses for different factors and their respective impact on the effect (ΔLVEF) of cardiac stem cell therapies. ESC, embryonic stem cells; iPSC, induced pluripotent stem cell; MC, mesodermal cells; CVC, cardiovascular cells and progenitors; MSC, mesenchymal stem cells. *Marked as significant according to regression coefficient of the respective fixed-effects model

Variable	Subgroups	Difference in LVEF (%)	p-value
	Overall	8.59	
Cell types	ESC	17.1	0.483
	iPSC	15.9	
	MC	7.2	
	CVC	9.9	
	MSC	7.5	
Cell origin	Syngen	3.45	0.046*
	Allogen	11.89	
	Xenogen	9.2	
Gender	Female	8.7	0.003*
	Male	6.5	
Cell amount	1-4*10⁵	10.78	0.013*
	0.5-3*10⁶	6.3	
Follow up	1-3 weeks	10.6	0.004*
	> 3 weeks	5.3	
Injections	1-2	8.7	0.14
	3-5	5.5	

Which cell type is most effective? Applied cells were divided into four subgroups: ESC (embryonic stem cells), iPSC (induced pluripotent stem cells), MSC (mesenchymal stem cells), CVC (cardiovascular cells) and MC (adult mesodermal cells). These led to an improvement in LVEF of 17.1% (ESC), 15.9% (iPSC), 9.9% (CVC), 7.5% (MSC) and 7.2% (MC) compared to the respective control group (Fig. 5 A). However, there were no significant differences in efficacy between cell type (p=.48).

Impact of cell origin: Syngeneic, allogeneic and xenogeneic cells had significantly different impacts on LVEF improvement (3.45%, 11.89%, 9.2%; p=.046), favoring transplantation of allogeneic cells (Fig. 5 B).

Cellular Physiology and Biochemistry
Cell Physiol Biochem 2017;42:254-268
DOI: 10.1159/000477324
Published online: May 19, 2017

Number of injected cells: The injection of less than 500,000 cells resulted in significantly higher LVEF improvements than the injection of 500,000 cells or more (LVEF improvement of 10.8% vs. 6.3%; p=.01; Fig. 5 C).

Follow up time: LVEF improvements measured 1 to 3 weeks following cell injection were significantly higher compared with longer follow-up (> 3 weeks); (LVEF improvement of 10.6% vs. 5.3%; p=.004; Fig. 5 D).

Female vs. Male: Female mice benefited significantly more from cardiac cell therapies than their male counterparts (LVEF improvement of 8.7% vs. 6.5%; p=.003; Fig. 5 E).

Impact of other factors: No significant differences were observed between groups for the number of injections (Table 2).

Power-Analysis

Based on our results, we used the Δ LVEF improvement as a surrogate measure of effect to perform a *post hoc* power analysis. Using our parameters (n=583, Δ LVEF=8.59%, α=.05, one-tailed *t* test) we obtained a power of .81.

Discussion

This meta-analysis of 22 studies, including a total of 583 animals, was performed to assess the efficacy of cardiac cell therapies in mitigating post-MI contractile dysfunction in mice. Our analysis shows that intra-myocardial cell injection increases LVEF by 8.59% as measured by MRI. Moreover, meta-regression analyses performed to identify moderators responsible for the LVEF improvement indicate that:

i) Significantly greater increases in LVEF are associated with:
   a. Application of allogenic cells (compared to syngeneic and xenogeneic);
   b. Female animal recipients;
   c. Less than 500,000 injected cells;
   d. Follow-up times up to three weeks post-cell application.
ii) Cell type and number of cell injections have no significant impact on the magnitude of LVEF improvement.

These findings are discussed in detail, below.

Relevance of the mouse model for efficacy testing of cardiac stem cell therapies

Studies using rodents have contributed significantly to recent advances in cardiovascular biology and provided proof-of-concept for the development of novel therapeutics [23]. However, the ability of small rodents - particularly the mouse model - to predict the results of human stem cell-based myocardial regenerative trials has been vigorously challenged [11]. This assumption is based on the premise that the anatomical and physiological differences between mice and humans are sufficient to lead to variant results between preclinical models and clinical trials. However, objective data corroborating this assumption are lacking. The most frequently quoted differences between human and murine hearts relate to divergences in calcium handling properties [24], ventricular expression of motor proteins [25], organ size and beating frequency, and coronary architecture [26]. That these fundamental differences can be clinically germane is exemplified by the finding that mice with heart failure benefit from phospholamban (PLN) ablation, whereas humans lacking PLN develop lethal dilated cardiomyopathy [24].

Yet, it would be wrong to conclude from this observation that mice are inappropriate models to predict results of clinical stem cell therapy trials. Rather, such differences between mice and men highlight that an understanding of the mode of action of therapies is indispensable for the development of novel therapies using rodent models. As a corollary, the appropriateness of the mouse model to predict the clinical results of cardiac stem cell therapies depends fundamentally on the alleged targeted biological mechanism. Aiming

Cellular Physiology and Biochemistry
Cell Physiol Biochem 2017;42:254-268
DOI: 10.1159/000477324
Published online: May 19, 2017

at functional replacement of beating cardiomyocytes alone, fundamental interspecies physiological and anatomical differences might limit the validity of rodent results to predict clinical outcome [27]. However, when highly conserved biological pathways and processes are targeted, we believe the mouse to be a valid model for the prediction of clinical results of regenerative therapy. Beneficial effects of cardiac cell therapies have been mainly attributed to modulation of apoptosis [28, 29], inflammation and angiogenesis [30-33]. Intriguingly, these mechanisms are highly conserved biological pathways in mammals [34-36]. This corroborates our hypothesis that the mouse is a valid model to predict the magnitude of LVEF improvement in clinical trials of cell therapies for the treatment of ischemic myocardium. The mouse model is cheap, has a short reproductive cycle and can be easily genetically modified [26], providing an excellent tool for studies of mode of action, proof-of-concept studies and biological safety testing [37].

Our meta-analysis reveals that application of cardiac cell therapies for treatment of ischaemic myocardium in mice results in an increase in LVEF by ~8.6%. This value is in striking agreement with results from the largest meta-analysis undertaken in large animal models (including 1,415 animals) which reported an increase in LVEF by 8.3% following stem cell application [38].

At the point of having identified an attractive cell type for cardiac repair, the ultimate application of the respective cell preparation has to be made in animal models faithfully recapitulating the clinical setting. In this context, large animals with similar anatomical and physiological properties, like pigs, will play a predominant role in translational research. This applies particularly to points in the final steps from bench to bedside, such as the exact clinical scenario treated, route of application, the time point of cell administration, the choice of device for cell injection, safety of these tools and the biodistribution of cells in a model with similar anatomical proportions.

In this context, large animal models with anatomical and physiological properties similar to patients, enable the use of clinical relevant endpoints such as mortality, major adverse cardiovascular events (MACE), cardiac dimensions and hemodynamics.

Impact of cell source (allogeneic)

To the best of our knowledge, neither preclinical nor clinical studies have compared the efficacy of allogeneic *versus* autologous stem cells as regenerative therapy post-MI. Hare et al. [39] showed that both autologous and allogeneic MSCs are safe in the treatment of ischemic cardiomyopathy in patients. Furthermore, data from a meta-analysis in large animal models suggest that allogeneic MSCs are as efficacious as autologous MSCs in improving myocardial pump function [38].

Our meta-analysis shows that use of allogeneic cells leads to significantly greater LVEF improvement than syngeneic cells. Even though syngeneic cell transplantation into in-bred mice strains is not directly comparable to autologous cell transplantation, our results support the postulate that allogeneic cell application may provide an attractive alternative to autologous cell-based therapies. The allogeneic approach may allow for 'off the shelf' stem cell therapies and the use of highly potent cell preparations from young healthy donors.

Gender

To our knowledge, this meta-analysis is the second one to provide evidence that female sex is associated with greater responsivity to cardiac cell therapies. Our results are in agreement with a meta-analysis investigating the influence of patient characteristics on study results by meta-regression. In the latter study, male individuals benefitted less than females from intracoronary infused bone marrow stem cells for the treatment of acute myocardial infarction [40]. Interestingly, the relevance of gender specific approaches in the field of cardiovascular medicine has increased over the last few years. Both clinical and preclinical studies indicate that female sex favourably influences the remodelling and adaptive response to myocardial infarction [41]. Furthermore, a greater resistance of female myocardium to ischemia/reperfusion injury has been demonstrated in several animal

Cellular Physiology and Biochemistry
Cell Physiol Biochem 2017;42:254-268
DOI: 10.1159/000477324
Published online: May 19, 2017

models [41]. Increased angiogenesis, which might be mediated by estradiol-dependent pathways, has been suggested as a potential mechanism underlying these effects [42, 43]. Further evidence is required to elucidate the impact of sex on responsiveness to cardiac cell therapy.

Cell number

Different strategies have been proposed to improve the very low engraftment rates following intramyocardial cell injection [44-46], a reasonable consideration when targeting functional replacement of deceased cardiomyocytes. With a primary goal of functional tissue replacement it seems logical that the effects of cell-based therapies will depend on the number of cells administered [47]. However, other modes of action underlying the beneficial effect of cardiac cell therapies beyond direct regeneration, such as paracrine effects, are currently assumed to predominate [32]. Notwithstanding this, little attention has been paid to date to dose-response relationships in the field of cardiac cell therapies. Meta-analyses of clinical studies have reported inconsistent results concerning the effect of administered cell number on LVEF improvement [5, 48-51].

Intriguingly, in our study meta-regression analysis revealed that - in mice - injection of cell numbers lower than 500,000 lead to significantly higher LVEF improvements. This suggests that paracrine mechanisms, as opposed to simple functional tissue replacement, are a key contributor to the observed improvement in LVEF.

Follow up

Both clinical and preclinical studies have reported that the positive effects of cardiac cell therapies on LVEF fade away during long-term follow up [13, 52]. Peak LVEF improvement has been assessed after six months in the clinical setting [52] and at up to one to two weeks in large animal models [13]. This is in line with the results from our meta-regression analysis showing that LVEF improvement is significantly higher within the first 3 weeks after cell application compared to longer follow-up times.

Number of cell injections

The number of cell injections did not have a significant effect on LVEF improvement. This correlation has not been directly investigated yet - either in clinical or preclinical studies. It has been speculated that intramyocardial injections disrupt tissue architecture and lead to inhomogeneous cell distribution within the infarcted area [47]. In order to provide robust data on both the optimal number of cells and individual injections, preclinical and clinical studies systematically addressing these questions for a specific cell type are necessary.

"Regenerative potential" of ESCs

The preclinical data obtained from our meta-analysis indicates that ESCs have a high potential to improve cardiac function following myocardial infarction (Fig. 5 A). Early reports have suggested "guided" differentiation into cardiomyocytes of ESCs transplanted into healthy and ischemic myocardium [53]. This hypothesis has been refuted by Nussbaum et al. who showed that neither healthy nor ischemic myocardium guides differentiation of ESCs into cardiomyocytes [54]. In fact, undifferentiated ESCs form teratomas in both syngeneic and allogeneic recipients [54], thereby strictly excluding them from therapeutic approaches.

While efficacy has been demonstrated for ESCs, their incapability to form cardiomyocytes *in vivo* suggests that modes of action beyond direct regeneration underlie the reported beneficial effect of intramyocardially transplanted ESCs.

Burt et al. showed that mitotically inactivated ESCs improve cardiac function although do not survive long-term, thus circumventing adverse effects such as tumour formation [55]. The proposed mode of action was transient function as an *in vivo* feeder layer that nurses damaged myocardium.

While further investigation is necessary to understand the mechanisms underlying improved cardiac function other than functional tissue replacement, our data demonstrate

Cellular Physiology and Biochemistry
Cell Physiol Biochem 2017;42:254-268
DOI: 10.1159/000477324
Published online: May 19, 2017

that the mouse is a valid model to address the efficacy of cell-based therapy post myocardial infarction.

Conclusion

To our knowledge, this is the first systematic review and meta-analysis to assess the effect of cell therapies in murine models of acute myocardial infarction.

In contrast to previous meta-analyses addressing large animals [13, 38] and humans [49], pluripotent stem cells and their derivatives have been included.

Furthermore, the magnitude of LVEF improvement is strikingly similar to results obtained from the most extensive meta-analysis of large animal models [13, 38]. This emphasizes the high relevance and reliability of the mouse model for evaluating the effect of new cell types for cardiac repair.

Funding

This work has been funded by the Federal Ministry of Education and Research Germany (FKZ 0312138A, FKZ 316159 and FKZ 02NUK043C) and the State Mecklenburg-Western Pomerania with EU Structural Funds (ESF/IV-WM-B34-0030/10 and ESF/IV-BM-B35-0010/12), by the BMBF (VIP+-Project 03VP00241), DFG (DA 1296-1), the German Heart Research Foundation (F/34/15) and by the FORUN Program of Rostock University Medical Centre (889001) and the EU funded CaSyM project (grant agreement #305033).

Disclosure Statement

The authors report no relationships that could be constructed as a conflict of interest.

References

1 Nichols M TN, Luengo-Fernandez R, Leal J, Gray A, Scarborough P, Rayner M: European Cardiovascular Disease Statistics 2012. European Heart Network, Brussels, European Society of Cardiology, Sophia Antipolis 2012;
2 Strauer BE, Brehm M, Zeus T, Gattermann N, Hernandez A, Sorg RV, Kogler G, Wernet P: [Intracoronary, human autologous stem cell transplantation for myocardial regeneration following myocardial infarction]. Dtsch Med Wochenschr 2001;126:932-938.
3 Stamm C, Westphal B, Kleine HD, Petzsch M, Kittner C, Klinge H, Schumichen C, Nienaber CA, Freund M, Steinhoff G: Autologous bone-marrow stem-cell transplantation for myocardial regeneration. Lancet 2003;361:45-46.
4 Menasche P, Vanneaux V, Fabreguettes JR, Bel A, Tosca L, Garcia S, Bellamy V, Farouz Y, Pouly J, Damour O, Perier MC, Desnos M, Hagege A, Agbulut O, Bruneval P, Tachdjian G, Trouvin JH, Larghero J: Towards a clinical use of human embryonic stem cell-derived cardiac progenitors: a translational experience. Eur Heart J 2014;10.1093/eurheartj/ehu192
5 Abdel-Latif A, Bolli R, Tleyjeh IM, Montori VM, Perin EC, Hornung CA, Zuba-Surma EK, Al-Mallah M, Dawn B: Adult bone marrow-derived cells for cardiac repair: a systematic review and meta-analysis. Arch Intern Med 2007;167:989-997.
6 Donndorf P, Kaminski A, Tiedemann G, Kundt G, Steinhoff G: Validating intramyocardial bone marrow stem cell therapy in combination with coronary artery bypass grafting, the PERFECT Phase III randomized multicenter trial: study protocol for a randomized controlled trial. Trials 2012;13:99.
7 Segers VF, Lee RT: Stem-cell therapy for cardiac disease. Nature 2008;451:937-942.
8 David R, Franz WM: From pluripotency to distinct cardiomyocyte subtypes. Physiology (Bethesda) 2012;27:119-129.

Cellular Physiology and Biochemistry
Cell Physiol Biochem 2017;42:254-268
DOI: 10.1159/000477324
Published online: May 19, 2017

9 Jung JJ, Husse B, Rimmbach C, Krebs S, Stieber J, Steinhoff G, Dendorfer A, Franz WM, David R: Programming and isolation of highly pure physiologically and pharmacologically functional sinus-nodal bodies from pluripotent stem cells. Stem Cell Reports 2014;2:592-605.
10 Mauritz C, Martens A, Rojas SV, Schnick T, Rathert C, Schecker N, Menke S, Glage S, Zweigerdt R, Haverich A, Martin U, Kutschka I: Induced pluripotent stem cell (iPSC)-derived Flk-1 progenitor cells engraft, differentiate, and improve heart function in a mouse model of acute myocardial infarction. Eur Heart J 2011;32:2634-2641.
11 Harding J, Roberts RM, Mirochnitchenko O: Large animal models for stem cell therapy. Stem Cell Res Ther 2013;4:23.
12 Cibelli J, Emborg ME, Prockop DJ, Roberts M, Schatten G, Rao M, Harding J, Mirochnitchenko O: Strategies for improving animal models for regenerative medicine. Cell Stem Cell 2013;12:271-274.
13 van der Spoel TI, Jansen of Lorkeers SJ, Agostoni P, van Belle E, Gyongyosi M, Sluijter JP, Cramer MJ, Doevendans PA, Chamuleau SA: Human relevance of pre-clinical studies in stem cell therapy: systematic review and meta-analysis of large animal models of ischaemic heart disease. Cardiovasc Res 2011;91:649-658.
14 Viechtbauer W: Bias and efficiency of meta-analytic variance estimators in the random-effects model. Journal of Educational and Behavioral Statistics 2005;30:261-293.
15 Dodge Y, International Statistical Institute.: The Oxford dictionary of statistical terms, [6th, Oxford University Press, Oxford, 2003.
16 Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP: Evaluating the quality of evidence from a network meta-analysis. PLoS One 2014;9:e99682.
17 Duval S, Tweedie R: A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. J Am Stat Assoc 2000;95:89-98.
18 Egger M, Davey Smith G, Schneider M, Minder C: Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629-634.
19 Viechtbauer W: Conducting Meta-Analyses in R with the metafor Package. J Stat Software 2010;36:1-48.
20 Champely S: Basic Functions for Power Analysis. 2015.
21 Hendry SL, 2nd, van der Bogt KE, Sheikh AY, Arai T, Dylla SJ, Drukker M, McConnell MV, Kutschka I, Hoyt G, Cao F, Weissman IL, Connolly AJ, Pelletier MP, Wu JC, Robbins RC, Yang PC: Multimodal evaluation of in vivo magnetic resonance imaging of myocardial restoration by mouse embryonic stem cells. J Thorac Cardiovasc Surg 2008;136:1028-1037 e1021.
22 Kim PJ, Mahmoudi M, Ge X, Matsuura Y, Toma I, Metzler S, Kooreman NG, Ramunas J, Holbrook C, McConnell MV, Blau H, Harnish P, Rulifson E, Yang PC: Direct evaluation of myocardial viability and stem cell engraftment demonstrates salvage of the injured myocardium. Circ Res 2015;116:e40-50.
23 Chong JJ, Murry CE: Cardiac regeneration using pluripotent stem cells--progression to large animal models. Stem Cell Res 2014;13:654-665.
24 Haghighi K, Kolokathis F, Pater L, Lynch RA, Asahi M, Gramolini AO, Fan GC, Tsiapras D, Hahn HS, Adamopoulos S, Liggett SB, Dorn GW, 2nd, MacLennan DH, Kremastinos DT, Kranias EG: Human phospholamban null results in lethal dilated cardiomyopathy revealing a critical difference between mouse and human. J Clin Invest 2003;111:869-876.
25 James J, Zhang Y, Wright K, Witt S, Glascock E, Osinska H, Klevitsky R, Martin L, Yager K, Sanbe A, Robbins J: Transgenic rabbits expressing mutant essential light chain do not develop hypertrophic cardiomyopathy. J Mol Cell Cardiol 2002;34:873-882.
26 Zaragoza C, Gomez-Guerrero C, Martin-Ventura JL, Blanco-Colio L, Lavin B, Mallavia B, Tarin C, Mas S, Ortiz A, Egido J: Animal models of cardiovascular diseases. J Biomed Biotechnol 2011;2011:497841.
27 Chong JJ, Yang X, Don CW, Minami E, Liu YW, Weyers JJ, Mahoney WM, Van Biber B, Cook SM, Palpant NJ, Gantz JA, Fugate JA, Muskheli V, Gough GM, Vogel KW, Astley CA, Hotchkiss CE, Baldessari A, Pabon L, Reinecke H, Gill EA, Nelson V, Kiem HP, Laflamme MA, Murry CE: Human embryonic-stem-cell-derived cardiomyocytes regenerate non-human primate hearts. Nature 2014;510:273-277.
28 Zhang J, He Z, Xiao W, Na Q, Wu T, Su K, Cui X: Overexpression of BAG3 Attenuates Hypoxia-Induced Cardiomyocyte Apoptosis by Inducing Autophagy. Cell Physiol Biochem 2016;39:491-500.
29 Ludwig M, Tolk A, Skorska A, Maschmeier C, Gaebel R, Lux CA, Steinhoff G, David R: Exploiting AT2R to Improve CD117 Stem Cell Function In Vitro and In Vivo--Perspectives for Cardiac Stem Cell Therapy. Cell Physiol Biochem 2015;37:77-93.

Cellular Physiology and Biochemistry
Cell Physiol Biochem 2017;42:254-268
DOI: 10.1159/000477324
Published online: May 19, 2017

30 Zhang J, Wu Y, Chen A, Zhao Q: Mesenchymal stem cells promote cardiac muscle repair via enhanced neovascularization. Cell Physiol Biochem 2015;35:1219-1229.
31 Liu J, Wu P, Wang H, Wang Y, Du Y, Cheng W, Xu Z, Zhou N, Wang L, Yang Z: Necroptosis Induced by Ad-HGF Activates Endogenous C-Kit+ Cardiac Stem Cells and Promotes Cardiomyocyte Proliferation and Angiogenesis in the Infarcted Aged Heart. Cell Physiol Biochem 2016;40:847-860.
32 Gnecchi M, Zhang Z, Ni A, Dzau VJ: Paracrine mechanisms in adult stem cell signaling and therapy. Circ Res 2008;103:1204-1219.
33 Teng X, Chen L, Chen W, Yang J, Yang Z, Shen Z: Mesenchymal Stem Cell-Derived Exosomes Improve the Microenvironment of Infarcted Myocardium Contributing to Angiogenesis and Anti-Inflammation. Cell Physiol Biochem 2015;37:2415-2424.
34 Holmes DI, Zachary I: The vascular endothelial growth factor (VEGF) family: angiogenic factors in health and disease. Genome Biol 2005;6:209.
35 Reed JC, Doctor K, Rojas A, Zapata JM, Stehlik C, Fiorentino L, Damiano J, Roth W, Matsuzawa S, Newman R, Takayama S, Marusawa H, Xu F, Salvesen G, Godzik A, Group RG, Members GSL: Comparative analysis of apoptosis and inflammation genes of mice and humans. Genome Res 2003;13:1376-1388.
36 Takao K, Miyakawa T: Genomic responses in mouse models greatly mimic human inflammatory diseases. Proc Natl Acad Sci U S A 2015;112:1167-1172.
37 Patten RD, Hall-Porter MR: Small animal models of heart failure: development of novel therapies, past and present. Circ Heart Fail 2009;2:138-144.
38 Jansen Of Lorkeers SJ, Eding JE, Vesterinen HM, van der Spoel TI, Sena ES, Duckers HJ, Doevendans PA, Macleod MR, Chamuleau SA: Similar effect of autologous and allogeneic cell therapy for ischemic heart disease: systematic review and meta-analysis of large animal studies. Circ Res 2015;116:80-86.
39 Hare JM, Fishman JE, Gerstenblith G, DiFede Velazquez DL, Zambrano JP, Suncion VY, Tracy M, Ghersin E, Johnston PV, Brinker JA, Breton E, Davis-Sproul J, Schulman IH, Byrnes J, Mendizabal AM, Lowery MH, Rouy D, Altman P, Wong Po Foo C, Ruiz P, Amador A, Da Silva J, McNiece IK, Heldman AW, George R, Lardo A: Comparison of allogeneic vs autologous bone marrow-derived mesenchymal stem cells delivered by transendocardial injection in patients with ischemic cardiomyopathy: the POSEIDON randomized trial. JAMA 2012;308:2369-2379.
40 Bai Y, Sun T, Ye P: Age, gender and diabetic status are associated with effects of bone marrow cell therapy on recovery of left ventricular function after acute myocardial infarction: a systematic review and meta-analysis. Ageing Res Rev 2010;9:418-423.
41 Ostadal B, Ostadal P: Sex-based differences in cardiac ischaemic injury and protection: therapeutic implications. Br J Pharmacol 2014;171:541-554.
42 Barnabas O, Wang H, Gao XM: Role of estrogen in angiogenesis in cardiovascular diseases. J Geriatr Cardiol 2013;10:377-382.
43 Masuda H, Kalka C, Takahashi T, Yoshida M, Wada M, Kobori M, Itoh R, Iwaguro H, Eguchi M, Iwami Y, Tanaka R, Nakagawa Y, Sugimoto A, Ninomiya S, Hayashi S, Kato S, Asahara T: Estrogen-mediated endothelial progenitor cell biology and kinetics for physiological postnatal vasculogenesis. Circ Res 2007;101:598-606.
44 Terrovitis J, Lautamaki R, Bonios M, Fox J, Engles JM, Yu J, Leppo MK, Pomper MG, Wahl RL, Seidel J, Tsui BM, Bengel FM, Abraham MR, Marban E: Noninvasive quantification and optimization of acute cell retention by in vivo positron emission tomography after intramyocardial cardiac-derived stem cell delivery. J Am Coll Cardiol 2009;54:1619-1626.
45 Cheng K, Li TS, Malliaras K, Davis DR, Zhang Y, Marban E: Magnetic targeting enhances engraftment and functional benefit of iron-labeled cardiosphere-derived cells in myocardial infarction. Circ Res 2010;106:1570-1581.
46 Zimmermann WH, Melnychenko I, Wasmeier G, Didie M, Naito H, Nixdorff U, Hess A, Budinsky L, Brune K, Michaelis B, Dhein S, Schwoerer A, Ehmke H, Eschenhagen T: Engineered heart tissue grafts improve systolic and diastolic function in infarcted rat hearts. Nat Med 2006;12:452-458.
47 Sanganalmath SK, Bolli R: Cell therapy for heart failure: a comprehensive overview of experimental and clinical studies, current challenges, and future directions. Circ Res 2013;113:810-834.

Cellular Physiology and Biochemistry
Cell Physiol Biochem 2017;42:254-268
DOI: 10.1159/000477324
Published online: May 19, 2017

48 Lipinski MJ, Biondi-Zoccai GG, Abbate A, Khianey R, Sheiban I, Bartunek J, Vanderheyden M, Kim HS, Kang HJ, Strauer BE, Vetrovec GW: Impact of intracoronary cell therapy on left ventricular function in the setting of acute myocardial infarction: a collaborative systematic review and meta-analysis of controlled clinical trials. J Am Coll Cardiol 2007;50:1761-1767.
49 de Jong R, Houtgraaf JH, Samiei S, Boersma E, Duckers HJ: Intracoronary stem cell infusion after acute myocardial infarction: a meta-analysis and update on clinical trials. Circ Cardiovasc Interv 2014;7:156-167.
50 Gyongyosi M, Wojakowski W, Lemarchand P, Lunde K, Tendera M, Bartunek J, Marban E, Assmus B, Henry TD, Traverse JH, Moye LA, Surder D, Corti R, Huikuri H, Miettinen J, Wohrle J, Obradovic S, Roncalli J, Malliaras K, Pokushalov E, Romanov A, Kastrup J, Bergmann MW, Atsma DE, Diederichsen A, Edes I, Benedek I, Benedek T, Pejkov H, Nyolczas N, Pavo N, Bergler-Klein J, Pavo IJ, Sylven C, Berti S, Navarese EP, Maurer G, Investigators A: Meta-Analysis of Cell-based CaRdiac stUdiEs (ACCRUE) in patients with acute myocardial infarction based on individual patient data. Circ Res 2015;116:1346-1360.
51 Jeevanantham V, Butler M, Saad A, Abdel-Latif A, Zuba-Surma EK, Dawn B: Adult Bone Marrow Cell Therapy Improves Survival and Induces Long-Term Improvement in Cardiac Parameters A Systematic Review and Meta-Analysis. Circulation 2012;126:551-+.
52 Schaefer A, Zwadlo C, Fuchs M, Meyer GP, Lippolt P, Wollert KC, Drexler H: Long-term effects of intracoronary bone marrow cell transfer on diastolic function in patients after acute myocardial infarction: 5-year results from the randomized-controlled BOOST trial--an echocardiographic study. Eur J Echocardiogr 2010;11:165-171.
53 Singla DK, Hacker TA, Ma L, Douglas PS, Sullivan R, Lyons GE, Kamp TJ: Transplantation of embryonic stem cells into the infarcted mouse heart: formation of multiple cell types. J Mol Cell Cardiol 2006;40:195-200.
54 Nussbaum J, Minami E, Laflamme MA, Virag JA, Ware CB, Masino A, Muskheli V, Pabon L, Reinecke H, Murry CE: Transplantation of undifferentiated murine embryonic stem cells in the heart: teratoma formation and immune response. FASEB J 2007;21:1345-1357.
55 Burt RK, Chen YH, Verda L, Lucena C, Navale S, Johnson J, Han X, Lomasney J, Baker JM, Ngai KL, Kino A, Carr J, Kajstura J, Anversa P: Mitotically inactivated embryonic stem cells can be used as an in vivo feeder layer to nurse damaged myocardium after acute myocardial infarction: a preclinical study. Circ Res 2012;111:1286-1296.
56 Higgins JP, Thompson SG, Deeks JJ, Altman DG: Measuring inconsistency in meta-analyses. BMJ 2003;327:557-560.
57 Arai T, Kofidis T, Bulte JW, de Bruin J, Venook RD, Berry GJ, McConnell MV, Quertermous T, Robbins RC, Yang PC: Dual in vivo magnetic resonance evaluation of magnetically labeled mouse embryonic stem cells and cardiac function at 1.5 t. Magn Reson Med 2006;55:203-209.
58 Au KW, Liao SY, Lee YK, Lai WH, Ng KM, Chan YC, Yip MC, Ho CY, Wu EX, Li RA, Siu CW, Tse HF: Effects of iron oxide nanoparticles on cardiac differentiation of embryonic stem cells. Biochem Biophys Res Commun 2009;379:898-903.
59 Bai X, Yan Y, Song YH, Seidensticker M, Rabinovich B, Metzele R, Bankson JA, Vykoukal D, Alt E: Both cultured and freshly isolated adipose tissue-derived stem cells enhance cardiac function after acute myocardial infarction. Eur Heart J 2010;31:489-501.
60 den Haan MC, Grauss RW, Smits AM, Winter EM, van Tuyn J, Pijnappels DA, Steendijk P, Gittenberger-De Groot AC, van der Laarse A, Fibbe WE, de Vries AA, Schalij MJ, Doevendans PA, Goumans MJ, Atsma DE: Cardiomyogenic differentiation-independent improvement of cardiac function by human cardiomyocyte progenitor cell injection in ischaemic mouse hearts. J Cell Mol Med 2012;16:1508-1521.
61 Drey F, Choi YH, Neef K, Ewert B, Tenbrock A, Treskes P, Bovenschulte H, Liakopoulos OJ, Brenkmann M, Stamm C, Wittwer T, Wahlers T: Noninvasive in vivo tracking of mesenchymal stem cells and evaluation of cell therapeutic effects in a murine model using a clinical 3.0 T MRI. Cell Transplant 2013;22:1971-1980.
62 Grauss RW, Winter EM, van Tuyn J, Pijnappels DA, Steijn RV, Hogers B, van der Geest RJ, de Vries AA, Steendijk P, van der Laarse A, Gittenberger-de Groot AC, Schalij MJ, Atsma DE: Mesenchymal stem cells from ischemic heart disease patients improve left ventricular function after acute myocardial infarction. Am J Physiol Heart Circ Physiol 2007;293:H2438-2447.
63 Grauss RW, van Tuyn J, Steendijk P, Winter EM, Pijnappels DA, Hogers B, Gittenberger-De Groot AC, van der Geest R, van der Laarse A, de Vries AA, Schalij MJ, Atsma DE: Forced myocardin expression enhances the therapeutic effect of human mesenchymal stem cells after transplantation in ischemic mouse hearts. Stem Cells 2008;26:1083-1093.

Cellular Physiology and Biochemistry
Cell Physiol Biochem 2017;42:254-268
DOI: 10.1159/000477324
Published online: May 19, 2017

64 Huber BC, Ransohoff JD, Ransohoff KJ, Riegler J, Ebert A, Kodo K, Gong Y, Sanchez-Freire V, Dey D, Kooreman NG, Diecke S, Zhang WY, Odegaard J, Hu S, Gold JD, Robbins RC, Wu JC: Costimulation-adhesion blockade is superior to cyclosporine A and prednisone immunosuppressive therapy for preventing rejection of differentiated human embryonic stem cells following transplantation. Stem Cells 2013;31:2354-2363.
65 Kofidis T, Lebl DR, Swijnenburg RJ, Greeve JM, Klima U, Robbins RC: Allopurinol/uricase and ibuprofen enhance engraftment of cardiomyocyte-enriched human embryonic stem cells and improve cardiac function following myocardial injury. Eur J Cardiothorac Surg 2006;29:50-55.
66 Liao SY, Liu Y, Siu CW, Zhang Y, Lai WH, Au KW, Lee YK, Chan YC, Yip PM, Wu EX, Wu Y, Lau CP, Li RA, Tse HF: Proarrhythmic risk of embryonic stem cell-derived cardiomyocyte transplantation in infarcted myocardium. Heart Rhythm 2010;7:1852-1859.
67 Liu J, Narsinh KH, Lan F, Wang L, Nguyen PK, Hu S, Lee A, Han L, Gong Y, Huang M, Nag D, Rosenberg J, Chouldechova A, Robbins RC, Wu JC: Early stem cell engraftment predicts late cardiac functional recovery: preclinical insights from molecular imaging. Circ Cardiovasc Imaging 2012;5:481-490.
68 Ong SG, Huber BC, Lee WH, Kodo K, Ebert AD, Ma Y, Nguyen PK, Diecke S, Chen WY, Wu JC: Microfluidic Single-Cell Analysis of Transplanted Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes After Acute Myocardial Infarction. Circulation 2015;132:762-771.
69 Paulis LE, Klein AM, Ghanem A, Geelen T, Coolen BF, Breitbach M, Zimmermann K, Nicolay K, Fleischmann BK, Roell W, Strijkers GJ: Embryonic cardiomyocyte, but not autologous stem cell transplantation, restricts infarct expansion, enhances ventricular function, and improves long-term survival. PLoS One 2013;8:e61510.
70 Rojas SV, Martens A, Zweigerdt R, Baraki H, Rathert C, Schecker N, Rojas-Hernandez S, Schwanke K, Martin U, Haverich A, Kutschka I: Transplantation Effectiveness of Induced Pluripotent Stem Cells Is Improved by a Fibrinogen Biomatrix in an Experimental Model of Ischemic Heart Failure. Tissue Eng Part A 2015;21:1991-2000.
71 Smits AM, van Laake LW, den Ouden K, Schreurs C, Szuhai K, van Echteld CJ, Mummery CL, Doevendans PA, Goumans MJ: Human cardiomyocyte progenitor cell transplantation preserves long-term function of the infarcted mouse myocardium. Cardiovasc Res 2009;83:527-535.
72 van Laake LW, Passier R, den Ouden K, Schreurs C, Monshouwer-Kloots J, Ward-van Oostwaard D, van Echteld CJ, Doevendans PA, Mummery CL: Improvement of mouse cardiac function by hESC-derived cardiomyocytes correlates with vascularity but not graft size. Stem Cell Res 2009;3:106-112.
73 Wang J, Najjar A, Zhang S, Rabinovich B, Willerson JT, Gelovani JG, Yeh ET: Molecular imaging of mesenchymal stem cell: mechanistic insight into cardiac repair after experimental myocardial infarction. Circ Cardiovasc Imaging 2012;5:94-101.
74 Winter EM, van Oorschot AA, Hogers B, van der Graaf LM, Doevendans PA, Poelmann RE, Atsma DE, Gittenberger-de Groot AC, Goumans MJ: A new direction for cardiac regeneration therapy: application of synergistically acting epicardium-derived cells and cardiomyocyte progenitor cells. Circ Heart Fail 2009;2:643-653.

2.3.3 ML-assisted outcome analysis of a Phase III clinical trial

Steinhoff, G., Nesteruk, J., **Wolfien, M.**, ..., and Wolkenhauer, O. (2017).
Cardiac Function Improvement of the Randomized PERFECT Phase III Clinical Trial of Intramyocardial CD133$^+$ Application After Myocardial Infarction
EBioMedicine. IF: 6.680, Citations (December 14, 2020): 23

The Phase III clinical trial PERFECT (NCT00950274)[17] was designed to assess clinical safety and efficacy of an intramyocardial CD133$^+$ bone marrow stem cell treatment combined with coronary artery bypass graft (CABG) surgery for the induction of improved cardiac repair. This was a multicentric, double-blinded, and randomized placebo controlled trial. The study was conducted across six centres in Germany between October 2009 through March 2016 and was stopped due slow recruitment after positive interim analysis in March 2015. The inclusion criteria were post-infarction patients with chronic ischemia and reduced LVEF (25-50%). 82 patients were randomized in two groups receiving an intramyocardial application of 5 ml placebo or a suspension of 0.5-5 million CD133$^+$ cells.

In this study, I proposed and utilized the computational analysis strategy, in which I applied supervised and unsupervised ML to clinical routine measurements and accompanying research parameters with respect to learning on small datasets. Unsupervised ML techniques included t-SNE dimensional reduction to a three dimensional space, which was afterwards fitted to two dimensions for an improved interpretability. I also compared and combined several classical ML techniques (e.g., SVM, Random forest, Boosting) for patient stratification and most important feature selection. Hereby, peripheral blood markers were identified that show a distinct therapeutic-response profile and further characteristic baseline processes related to inflammation, proliferation, and vascularization. These signature was identified by using patient samples obtained prior therapy application and is currently refined as a predictive model for future therapies. The signature has been patented for the prospective use of responsive patient classification prior to therapy.

In summary, the PERFECT trial outcome analysis demonstrates that the regulation of induced cardiac repair is linked to the circulating pool of CD133$^+$ endothelial progenitor cells, thrombocytes, and associated with SH2B3 gene expression. Based on these findings, responders to cardiac functional improvement may be identified prior a surgery by a peripheral blood biomarker signature.

[17] https://clinicaltrials.gov/ct2/show/NCT00950274

EBioMedicine 22 (2017) 208–224

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

Cardiac Function Improvement and Bone Marrow Response – Outcome Analysis of the Randomized PERFECT Phase III Clinical Trial of Intramyocardial CD133$^+$ Application After Myocardial Infarction

Gustav Steinhoff [a,*], Julia Nesteruk [a], Markus Wolfien [b], Günther Kundt [c], The PERFECT Trial Investigators Group [1:], Jochen Börgermann [d,1], Robert David [a,1], Jens Garbade [e,1], Jana Große [a,1], Axel Haverich [f,1], Holger Hennig [b,g,1], Alexander Kaminski [a,1], Joachim Lotz [h,1], Friedrich-Wilhelm Mohr [e,1], Paula Müller [a,1], Robert Oostendorp [i,1], Ulrike Ruch [a,1], Samir Sarikouch [f,1], Anna Skorska [a,1], Christof Stamm [j,1], Gudrun Tiedemann [a,1], Florian Mathias Wagner [k,1], Olaf Wolkenhauer [b,1]

[a] Department of Cardiac Surgery, Reference and Translation Center for Cardiac Stem Cell Therapy, University Medicine Rostock, Schillingallee 35, 18055 Rostock, Germany
[b] University Rostock, Institute of Computer Science, Department of Systems Biology and Bioinformatics, Ulmenstraße 69, 18057 Rostock, Germany
[c] University Medicine Rostock, Department of Medical Informatics and Biostatistics, Ernst-Heydemann-Straße 8, 18055 Rostock, Germany
[d] Heart and Diabetes Center North Rhine Westfalia, University Hospital of the Ruhr, University Bochum, Georgstraße 11, 32545 Bad Oeynhausen, Germany
[e] Department of Cardiac Surgery, Heart Center University Medicine Leipzig, Strümpellstraße 39, 04289 Leipzig, Germany
[f] Medical School Hannover, Department of Heart-, Thoracic-, and Vascular Surgery, Carl-Neuberg- Straße 1, 30625 Hannover, Germany
[g] Broad Institute of Harvard and MIT, Imaging Platform, 415 Main St, Cambridge, MA 02142, USA
[h] University Medicine Goettingen, Department of Diagnostic Radiology, Robert-Koch-Straße 40, 37075 Göttingen, Germany
[i] Department of Medicine III, Technical University Munich, Klinikum rechts der Isar, Ismaninger Straße 22, 81675 München, Germany
[j] German Heart Center Berlin, Department of Heart-, Thoracic-, and Vascular Surgery, Augustenburger Platz 1, 13353 Berlin, Germany
[k] Department of Cardiac and Vascular Surgery, University Heart Center Hamburg, Martinistraße 52, 20246 Hamburg, Germany

ARTICLE INFO

Article history:
Received 8 June 2017
Received in revised form 18 July 2017
Accepted 24 July 2017
Available online 29 July 2017

Keywords:
Randomised double-blinded phase III multicentre trial
CD133$^+$
CD34$^+$
Endothelial progenitor cell (EPC)
SH2B3
Lnk adaptor
Cardiac repair
Cardiac stem cell therapy
Angiogenesis

ABSTRACT

Objective: The phase III clinical trial PERFECT was designed to assess clinical safety and efficacy of intramyocardial CD133$^+$ bone marrow stem cell treatment combined with CABG for induction of cardiac repair.
Design: Multicentre, double-blinded, randomised placebo controlled trial.
Setting: The study was conducted across six centres in Germany October 2009 through March 2016 and stopped due slow recruitment after positive interim analysis in March 2015.
Participants: Post-infarction patients with chronic ischemia and reduced LVEF (25–50%). Interventions: Eighty-two patients were randomised to two groups receiving intramyocardial application of 5 ml placebo or a suspension of 0.5–5 × 10^6 CD133$^+$.
Outcome: Primary endpoint was delta (Δ) LVEF at 180 days (d) compared to baseline measured in MRI.
Findings (prespecified): Safety ($n = 77$): 180 d survival was 100%, MACE $n = 2$, SAE $n = 49$, without difference between placebo and CD133$^+$. Efficacy ($n = 58$): The LVEF improved from baseline LVEF 33.5% by +9.6% at 180 d, $p = 0.001$ ($n = 58$). Treatment groups were not different in ΔLVEF (ANCOVA: Placebo +8.8% vs. CD133$^+$ +10.4%, ΔCD133$^+$ vs placebo +2.6%, $p = 0.4$).
Findings (post hoc): Responders (R) classified by ΔLVEF ≥ 5% after 180 d were 60% of the patients (35/58) in both treatment groups. ΔLVEF in ANCOVA was +17.1% in (R) vs. non-responders (NR) (ΔLVEF 0%, $n = 23$). NR were characterized by a preoperative response signature in peripheral blood with reduced CD133$^+$ EPC (RvsNR: $p = 0.005$) and thrombocytes ($p = 0.004$) in contrast to increased Erythropoeitin ($p = 0.02$), and SH2B3 mRNA expression ($p = 0.073$). Actuarial computed mean survival time was 76.9 ± 3.32 months (R) vs. +72.3 ± 5.0 months (NR), HR 0.3 [CI 0.07–1.2]; $p = 0.067$.Using a machine learning 20 biomarker response parameters were identified allowing preoperative discrimination with an accuracy of 80% (R) and 84% (NR) after 10-fold cross-validation.

* Corresponding author.
E-mail addresses: gustav.steinhoff@med.uni-rostock.de (G. Steinhoff), iuliia.nesteruk@med.uni-rostock.de (J. Nesteruk), markus.wolfien@uni-rostock.de (M. Wolfien), guenther.kundt@med.uni-rostock.de (G. Kundt), jboergermann@hdz-nrw.de (J. Börgermann), robert.david@med.uni-rostock.de (R. David), jens.garbade@helios-kliniken.de (J. Garbade), jana.grosse@med.uni-rostock.de (J. Große), haverich.axel@mh-hannover.de (A. Haverich), holger.hennig@uni-rostock.de, holgerh@broadinstitute.org (H. Hennig), alexander.kaminski@med.uni-rostock.de (A. Kaminski), joachim.lotz@med.uni-goettingen.de (J. Lotz), chir@herzzentrum-leipzig.de (F.-W. Mohr), paula.mueller@med.uni-rostock.de (P. Müller), robert.oostendorp@tum.de (R. Oostendorp), ulrike.ruch@med.uni-rostock.de (U. Ruch), anna.skorska@med.uni-rostock.de, sarikouch.samir@mh-hannover.de (A. Skorska), stamm@DHZB.de (C. Stamm), gudrun.tiedemann@web.de (G. Tiedemann), fl.wagner@uke.de (F.M. Wagner), olaf.wolkenhauer@uni-rostock.de (O. Wolkenhauer).
[1] Shared last author.

http://dx.doi.org/10.1016/j.ebiom.2017.07.022

Interpretation: The PERFECT trial analysis demonstrates that the regulation of induced cardiac repair is linked to the circulating pool of CD133 + EPC and thrombocytes, associated with SH2B3 gene expression. Based on these findings, responders to cardiac functional improvement may be identified by a peripheral blood biomarker signature. TRIAL REGISTRATION: ClinicalTrials.gov NCT00950274.

Research in Context

Evidence Before This Study

Intramyocardial $CD133^+$ purified autologous bone marrow stem cell (BMSC) transplantation has been investigated as an adjunctive strategy to coronary artery bypass graft (CABG) revascularization in order to improve left ventricular heart function following deterioration of left ventricular ejection fraction (LVEF) after acute myocardial ST-segment elevation infarction (STEMI), and coronary artery 3-vessel disease sequentially treated by acute PCI and secondary CABG revascularization. Previous safety and efficacy (phase I, IIa, IIb) trials have demonstrated clinical safety and some evidence of therapeutic efficacy of adjunctive $CD133^+$ BMSC treatment adjunctive to CABG coronary revascularization. The randomised double-blinded placebo controlled PERFECT-trial was designed to assess clinical safety and efficacy in a, ICH-GCP complaint study setting. Post hoc biomarker and subgroup analyses were performed to identify $CD133^+$ bone marrow stem cell related cardiac repair mechanisms related to interventional $CD133^+$ BMSC transplantation.

Added Value of This Study

The study demonstrates the central regulatory importance of $CD^{133,34+}$ EPC response for angiogenesis, suppression of response by SH2B3, impact for cardiac tissue repair, selection of responding patients, and monitoring of angiogenesis response by combined diagnostic factors using machine learning.

Implications of All the Available Evidence

The described mechanism of suppression bone marrow $CD133^+$ angiogenesis response may have a pivotal role in cardiovascular tissue repair. Selection of patients by specific diagnostic peripheral blood biomarkers appears to be feasible and may lead to tailored therapy in cardiovascular disease. The lack of vascular repair by reduced blood angiogenesis may be a decisive determinant for cardiovascular disease and impaired tissue repair.

1. Introduction

Reparative therapies using stem cells for the repair of heart tissue have been at the forefront of preclinical and clinical development during the past 16 years (Fisher et al., 2016). Among the different approaches, the direct implantation of bone marrow- derived cells into heart tissue still attracts the most dedicated clinical developmental attention since the first-in-man application in 2001 and several promising clinical pilot trials (Stamm et al., 2003; Tse et al., 2003; Stamm et al., 2007). Yet, in these trials, clinically relevant improvements of LVEF as well as non-responsive patients were observed both in treatment and placebo groups (Henry et al., 2016; Nasseri et al., 2014; Bartunek et al., 2016). This has raised the question of induction of reparative mechanisms independent of stem cell application and potential suppressive factors of vascular repair associated with $CD34^+$ Endothelial Progenitor Cells (EPC) (Werner et al., 2005; Taylor et al., 2016; Bhatnagar et al., 2016; Contreras et al., 2017).

In light of this uncertainty, we have attempted to investigate the mechanism of cardiac repair and the role of bone marrow CD133 + EPC regulated angiogenesis using the results of the clinical PERFECT trial and its data recorded (Donndorf et al., 2012). Extensive additional laboratory analyses was carried out to delineate the underlying mechanisms and to develop diagnostic approaches for identifying patient (non)responsiveness to stem cell therapies by analyzing the following clinical features: 1. Baseline characteristics of treatment responders vs. non-responders; 2. Mechanism of action for cardiac regeneration and diagnostic access; 3. Relevance of LVEF endpoint for long term survival.

2. Methods

2.1. Trial Design

The PERFECT trial was a randomised, multicenter, placebo-controlled, double-blinded phase III study investigating the effects of intramyocardial $CD133^+$ BMSC treatment in combination with coronary artery bypass graft revascularization (CABG) for post infarction myocardial ischemia (Donndorf et al., 2012). The trial performed according to ICH-GCP was listed under the EudraCT number 2006-006404-11, DRKS number DRKS00000213, and approved by the committee of the University Medicine Rostock (FK 2007-07) and all trial sites in Germany (Supplement Appendix 1). Regulatory approval was given by the Paul-Ehrlich-Institute, Langen, Germany. The trial was registered at ClinicalTrials.gov identifier: NCT00950274. Characteristics of trial design, changes to trial design, outcomes, interim analysis, and recruitment period are depicted in Appendix 2 (Supplement) and the Clinical Trial Report (Appendix 1).

Inclusion criteria of the PERFECT trial (Supplement Appendices 1 and 2) were (a) coronary artery disease after myocardial infarction with the indication for CABG surgery, (b) reduced LVEF (25–50%) and (c) presence of a localized kinetic/hypokinetic/hypoperfused area of LV myocardium defining the SC target area (Supplement Fig. 1). According to the trial flow chart (Supplement Fig. 2) assessments were performed preoperative and at days 1, 3, 10, 90, and 180 post operation. In addition, safety (MACE) follow up was performed at 24 months post-treatment.

2.2. Participants and Study Settings

A total of 119 patients were screened in 6 centres in Germany (Fig. 1). All patients signed the informed consent form and were included in the study. Eighty-two (82) patients were randomised to active treatment or placebo. The allocation of patients to the different analysis sets is shown in Fig. 1. Initially, we evaluated the basic patient characteristics of the randomised patient groups for safety set (SAS) analysis ($n = 77$) and per-protocol set (PPS) efficacy analysis ($n = 58$) respectively for subanalysis of MRI early/late, primary endpoint responder/non-responder, biomarkers, preoperative cardiac disease state, age, sex, concomitant diseases, taking medications, operative procedures and postoperative course (Table 1).

2.3. Cell Preparation and Manufacturing

All patients enrolled in the study underwent bone marrow aspiration (mean 166 ± 20 ml) and withdrawal of 20 ml blood one to two

days before CABG surgery. To ensure consistent quality and individual safety of the cell product, central manufacturing according to GMP standard was performed at Seracell GmbH, Rostock. CD133$^+$ cells were selected from the bone marrow aspirate of each patient and individuals in the active group received autologous CD133$^+$ cells suspended in physiological saline + 10% autologous serum. Patients of the control group received the placebo preparation with saline + 10% autologous serum; their CD133$^+$ cells were stored by the cell product manufacturing site. In the CD133$^+$ group the recovery percentage of CD133$^+$ cells was 23.7 ± 10.4%, non-target cell depletion efficiency was >99.2% and the final dose of CD133$^+$ cells administered was 2.29 × 10^6 ± 1.42. Cell counts were determined by FACS using single platform analysis. The final preparation dose was 0.5 × 10^6–5 × 10^6 CD133$^+$ cells suspended in 5 ml of saline supplemented with 10% autologous serum, drawn into 5 × 1 ml syringes.

2.4. Randomisation and Masking

Randomisation to study treatment was done after all screening procedures had been performed, eligibility for the study confirmed and after bone-marrow aspiration. We used permuted block randomisation, randomly varying block sizes, stratified by study site (Rosenberger and Lachin, 2003). Patients were randomised on a 1:1 basis to receive CD133$^+$ cells or placebo (Fig. 1). The study was performed in a double blind manner up to final data closure in 4/2016. Only the cell preparation team at the contract GMP manufacturer was unblinded for

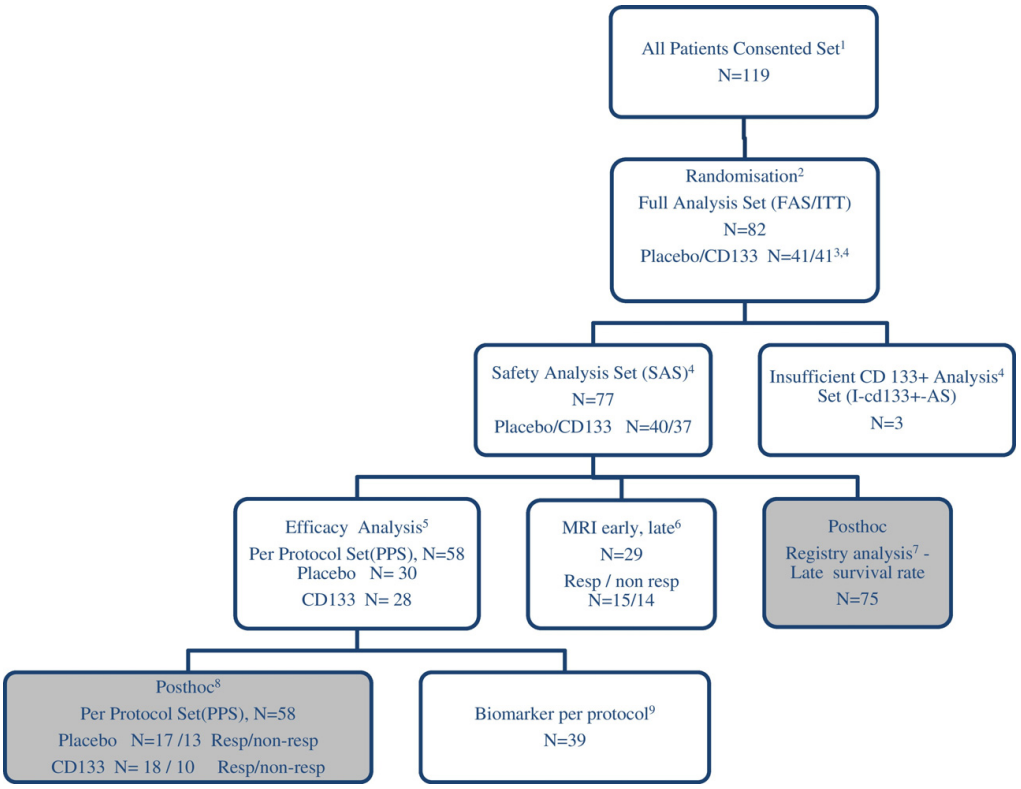


Fig. 1. PERFECT Trial flowchart and prespecified or post hoc analysis sets. The randomised multicentre trial was performed double-blinded placebo controlled through six heart centres in Germany according to ICH-GCP and is depicted according to CONSORT and STARD guidelines:1 A total of 119 patients were screened in 6 centres in Germany from Sept. 2009 through June 2015. All patients signed the informed consent form and were included in the study. Thirty-seven participants were excluded before randomisation due to newly identified exclusion criteria such as severe arrhythmia. 2 Eighty-two (82) patients were randomised to active treatment or placebo. Two (Stamm et al., 2003) patients were randomised but not treated because the CD 133 + preparation did not comply with the release criteria for GMP. 3 Forty (48.8%) patients received an injection of CD133 + cells and 40 (48·8%) received an injection of placebo. 4 Three patients were excluded because of insufficient CD133 + cell count below minimum dosis resulting in the safety-analysis-population ($n = 77$). 5 After a careful review of the blinded data in a blind data review meeting conducted on the 20 May 2016 a total of 19 patients were excluded from the full analysis population due to protocol violations with incomplete MRI follow-up data leading to the Per Protocol Set (PPS) efficacy-analysis-population ($n = 58$). Patient distribution for PPS efficacy population by study centres: German Heart Center Berlin 8%, Medical School Hannover 28%, University Medicine Rostock 38%, Heart and Diabetes Center Bad Oeynhausen 5%, Heart Center Leipzig 13%, University Medicine Hamburg 10%. 6 Additional MRI at day 10 postoperative for subanalysis of early and late postoperative changes for subgroup analysis early and late postoperative changes. 7 Post hoc analysis for actuarial survival was performed in registry analysis 7 years after FPI on Nov. 1, 2016. 8 Post hoc analysis was additionally performed in the efficacy group ($n = 58$) to unravel contributing non CD133$^+$ injection related factors of late improvement. Patients were grouped in the efficacy analysis set according to effective response in primary endpoint as responder or non-responder (Δ LVEF 180 d vs.0 responder ≥ 5% vs. non-responder < 5%). According to this post hoc analysis 35 patients from 58 (60.3%) were responders to treatment. This Responder/non-responder (R/NR) ratio was similar respectively in the placebo group 56·5% (R/NR 17/30 pt.) and in the CD133 + group 64% (R/NR 18/28 pt.) (Placebo vs. CD133 +; $p = 0{\cdot}373$). Responder (35/58) and non-responder (23/58) analysis was performed in efficacy group ($n = 58$). 9 Biomarkers were studied in 39 patients of the efficacy group ($n = 58$) independent on placebo/CD133 + or responder/non-responder group. All laboratory tests were realized in patients located in the Rostock centre ($n = 31$), where immediate laboratory analysis of FACS and CFU was guaranteed. Additional patients from other centres (8/58) were evaluated also in the Biomarker cohort according to realized parameters. Biobank at time point (pre- and postoperative day) −2, −1, +1, +3, +10, +180: Peripheral blood MNC/FACS (CD 133, 34, 117, 184, 309, 45, 31, 14), CFU-Hill, serum analysis angiogenesis factors and cytokines; Bone-marrow MNC, Isolated CD133 + FACS (CD133, 34, 117, 184, 309, 45, 31, 14), CFU-EC, RNA-seq.

Table 1
Patient characteristics and randomisation analysis sets.

Patient characteristics and randomisation analysis sets I												
	Safety (SAS) all	Safety (SAS) placebo	CD133 +	P	Efficacy placebo/CD133 +	CD133 +	P	Efficacy (PPS) Resp	NonResp	p	MRI early/late Plac/CD133	Biomarker
N	77 (mean, SD, min-max, median)	40 (mean, SD, min-max, median)	37 (mean, SD, min-max, median)		30	28		35	23		29	39
Basic data												
Age (y)	63.2 ± 8.37 34–79 Median: 63.0	62.9 ± 8.50 35–79 Median: 61.5	63.5 ± 8.34 34–78 Median: 65.0	0.751[a]	63.6 ± 7.75 53–79 Median: 62	64.0 ± 7.20 50–78 Median: 65	0.853[a]	62.9 ± 7.21 50–78 Median: 61	65.3 ± 7.68 53–79 Median: 66.0	0.231[a]	62.9 ± 6.86 51–79 Median: 62	64.8 ± 7.46 53–79 Median: 66.0
Sex/male%	67 (88.2)	34 (85)	33 (89.2)	0.739[c]	26 (86.7)	26 (92.9)	0.671[c]	31 (88.6)	21 (91.3)	1.000[c]	27 (93.1)	33 (84.6)
Body mass index (kg/m^2)	76, 28.8 ± 4.12 19.4–38.6 Median: 28.3	39, 29.1 ± 4.28 19.4–38.6 Median: 28.6	28.5 ± 3.98 19.6–38.0 Median: 28.1	0.575[a]	29.0 ± 3.81 19.4–37.2 Median: 29.4	28.8 ± 4.11 19.6–38.0 Median: 28.4	0.804[a]	28.7 ± 4.14 19.4–38.0 Median: 29.1	29.1 ± 3.64 22.9–35.2 Median: 28.4	0.723[a]	28.7 ± 3.87 22.5–38.0 Median: 27.8	29.5 ± 3.90 19.6–38 Median: 30.2:
Last myocardial infarction	47, ≤6 months: 21 (44.7) 7–12 months: 6 (12.8) >12 months: 20 (42.6)	21, ≤6 months: 10 (47.6) 7–12 months: 3 (14.3) >12 months: 20 (38.1)	26, ≤6 months: 11 (42.3) 7–12 months: 3 (11.5) >12 months: 12 (46.2)	0.631[b]	15, ≤6 months: 4 (26.7) 7–12 months: 3 (20.0) <12 months: 8 (53.3)	19, ≤6 months: 8 (42.1) 7–12 months: 2 (10.5) >12 months: 9 (47.4)	0.584[b]	19, ≤6 months: 8 (42.31 7–12 months: 2 (10.5) >12 months: 9 (47.4)	15, ≤6 months: 4 (26.7) 7–12 months: 3 (20.0) >12 months: 8 (53.3)	0.341[b]	19, ≤6 months: 5 (26.3) 7–12 months: 5 (26.3) >12 months: 9 (47.4)	22, ≤6 months: 8 (36.4) 7–12 months: 4 (18.2) >12 months: 10 (45.5)
PCI prior to CABG, n%	20 (26.0)	12 (30.0)	8 (21.6)	0.445[c]	9 (30.0)	6 (21.4)	0.554[c]	5 (14.7)	10 (41.7)	0.074[c]	8 (27.6)	9 (23.1)
Diabetes(%	42.9	50.0	35.1	0.198[c]	46.7	42.9	0.893	38.2	56.5	0.190	37.9	30.8
Hypert. (%)	83.1	85.0	81.1	0.464[c]	90.0	89.3	0.828	91.2	91.3	1.000	96.6	100
Hyperlipidemia (%)	61.0	65.0	56.8	0.491[c]	60.0	64.3	0.791	65.7	56.5	0.583	86.2	84.6
Laboratory parameters												
LDL cholesterol, mg/dl	65, 2.88 ± 0.86 0.80–5.0 Median: 2.70	34, 2.92 ± 0.748 1.6–5.0 Median: 2.75	31, 2.84 ± 0.98 0.80–4.80 Median: 2.70	0.695[a]	26, 2.91 ± 0.75 1.6–5.0 Median: 2.75	24, 2.84 ± 0.915 1.60–4.80 Median: 2.70	0.767[a]	30, 2.90 ± 0.911 1.60–5.0 Median: 2.65	20, 2.84 ± 0.698 1.6–4.1 Median: 2.70	0.788[a]	26, 2.83 ± 0.82 1.6–5.0 Median: 2.65	35, 2.98 ± 0.92 1.60–5.0 Median: 2.80
HDL cholesterol, mg/dl	65, 1.12 ± 0.293 0.60–1.98 Median: 1.10	34, 1.12 ± 0.237 0.70–1.60 Median: 1.10	31, 1.12 ± 0.35 0.60–1.98 Median: 1.00	0.476[b]	26, 1.13 ± 0.218 0.80–1.5 Median: 1.10	24, 1.04 ± 0.286 0.60–1.70 Median: 0.900	0.114[b]	30, 1.05 ± 0.252 0.60–1.70 Median: 0.95	20, 1.15 ± 0.252 0.80–1.60 Median: 1.10	0.177[b]	26, 1.06 ± 0.239 0.60–1.50 Median: 1.05	35, 1.11 ± 0.253 0.80–1.70 Median: 1.10
Triglycerides, mol/dl	68, 1.81 ± 0.98 0.60–6.40 Median: 1.60	38, 1.99 ± 1.14 0.80–6.40 Median: 1.60	30, 1.59 ± 0.70 0.60–3.40 Median: 1.50	0.166[b]	28, 2.06 ± 1.26 0.90–6.4 Median: 1.60	24, 1.70 ± 0.72 0.70–3.40 Median: 1.65	0.451[b]	31, 1.86 ± 1.13 0.80–6.4 Median: 1.60	21, 1.94 ± 0.951 0.70–4.50 Median: 1.80	0.495[b]	26, 1.83 ± 1.14 0.70–6.40 Median: 1.60	36, 2.03 ± 1.17 0.70–6.4 Median: 1.75
CRP (mg/l)	76, 0.565 ± 0.846 0.10–7.0 Median: 0.400	0.635 ± 1.11 0.10–7.00 Median: 0.400	36, 0.486 ± 0.383 0.10–1.70 Median: 0.400	0.983[b]	0.403 ± 0.275 0.10–1.20 Median: 0.400	0.511 ± 0.42 0.10–1.70 Median: 0.35	0.504[b]	0.469 ± 0.3471 0.10–1.40 Median: 0.400	0.435 ± 0.370 0.10–1.70 Median: 0.30	0.641[b]	0.483 ± 0.433 0.10–1.70 Median:0.300	0.505 ± 0.389 0.10–1.70 Median: 0.400
Creatinine (μmol/l)	90.0 ± 22.8 48–160 Median: 87	90.4 ± 21.4 53–152 Median: 86	91.4 ± 24.4 48–160 Median: 87	0.992[b]	91.3 ± 23.2 53–152 Median: 85.5	92.6 ± 25.4 48–160 Median: 89	0.913[b]	92.0 ± 26.2 48–160 Median: 88.0	91.8 ± 21.1 53–132 Median: 87.0	0.611[b]	89.4 ± 23.8 57–160 Median: 82.0	95.6 ± 25.2 48–160 Median:
Leucocytes (10^9/l)	76, 8.05 ± 1.78 5.0–11.9 Med.: 7.90	8.06 ± 1.75 5.0–11.9 Median: 8.00	36, 8.04 ± 1.83 5.1–11.7 Med.: 7.90	0.975[a]	8.03 ± 1.78 5–11.8 Median: 8.00	7.91 ± 1.94 5.1–11.7 Median: 7.70	0.807[a]	7.99 ± 1.86 5.0–11.7 Median: 7.70	7.94 ± 1.86 5.1–11.8 Median: 7.80	0.921[a]	8.07 ± 1.65 5.1–11.7 Median: 7.70	8.25 ± 1.91 5.1–11.8 Median: 8.00
Thrombocytes (10^9/l)	242 ± 78.2 73.620 Median:231	252 ± 91.4 144–620 Median: 231	232 ± 60.2 73–351 Median: 232	0.714[b]	246 ± 82.3 144–620 Median: 231	229 ± 65.7 73–351 Median: 229	0.709[b]	257 ± 81.5 123–620 Median: 238	208 ± 51.2 73–311 Median: 220	0.004[b]	228 ± 63.8 73–351 Median: 223	239 ± 85.8 73–620 Median: 234
NT Pro-BNP (pg/ml)	1468 ± 1947 108–12,735 Median: 803	1474 ± 2378 108–12,735 Median: 646	1560 ± 1370 225–7230 Median: 1028	0.065[b]	1551 ± 2647 108–12,735 Median: 681	1560 ± 1527 225–7230 Median: 1048	0.079[b]	1266 ± 1469 137–8444 Median: 688	1925 ± 2903 108–12,735 Median: 1025	0.861[b]	28, 1753 ± 2796 108–12,735 Med.: 687	1757 ± 2382 108–12,735 Median: 1063
Medication												
Aspirin (%)	97.4	97.5	97.3	1.000[c]	96.7	100.0	1.000	100.0	95.7	0.397	29 (100)	39 (100)
Statin (%)	97.4	95.0	100.0	0.494[c]	96.7	100.0	1.000	97.1	100.0	1.000	26 (89.7)	33 (84.6)
β-blocker (%)	98.7	97.5	100.0	1.000[c]	100.0	100.0	n/a	100.0	100.0	1.000	24 (82.8)	35 (89,7)
ACE inh. (%)	81.8	82.5	81.1	1.000[c]	83.3	82.1	1.000	85.7	78.3	0.496	21 (72.4)	26 (66.7)
AT1 rec. Antag. (%)	32.5	32.5	32.4	1.000[c]	36.7	28.6	0.583	28.6	39.1	0.568	4 (13.8)	6 (15.4)

(continued on next page)

Table 1 (*continued*)

Patient characteristics and randomisation analysis sets I												
	Safety (SAS) all	Safety (SAS) placebo	CD133+	P	Efficacy placebo/CD133+	CD133+	P	Efficacy (PPS) Resp	NonResp	p	MRI early/late Plac/CD133	Biomarker
Aldosteron Antag.(%)	55.8	57.5	54.1	0.821[c]	63.3	64.3	1.000	60.0	69.6	0.579	10 (34.5)	8 (20.5)
Diuretic (%)	93.5	92.5	94.6	1.000[c]	93.3	96.4	1.000	97.1	91.3	0.557	20 (69.0)	28 (71.8)
Ca-antag. (%)	37.7	32.5	43.2	0.356[c]	36.7	42.9	0.789	40.0	39.1	1.000	2 (6.9)	5 (12.8)
Anti-arrh. (%)	7.8	7.5	8.1	1.000[c]	3.3	7.1	0.605	5.7	4.3	1.000	1 (3.4)	1 (2.6)
Risk factors and status												
Smoking (previous) N (%)	35 (45.5)	20 (50)	15 (40.5)	0.494[c]	13 (43.3)	12 (42.9)	1.000[c]	18 (51.4)	7 (30.4)	0.175[c]	14 (48.3)	18 (46.2)
Smoking (actual) N (%)	20 (26.0)	8 (20)	12 (32.4)	0.299[c]	7 (23.3)	8 (28.6)	0.767[c]	10 (28.6)	5 (21.7)	0.760[c]	8 (27.6)	11 (28.2)
EuroScore	4.33 ± 3.44 0.13–17.1 Median: 2.98	3.98 ± 3.64 0.13–17.1 Median: 2.60	4.69 ± 3.21 0.88–11.9 Median: 3.57	0.170[b]	3.94 ± 3.24 1.33–16.3 Median: 2.59	4.80 ± 3.35 1.33–11.9 Median: 3.66	0.246[b]	3.97 ± 2.64 1.33–11.94 Median: 2.74	4.95 ± 4.09 1.33–16.3 Median: 3.22	0.583[b]	4.31 ± 3.26 1.33–16.3 Median: 3.22	4.77 ± 3.73 1.33–16.3 Median: 3.22
NYHA (class) N(%)	1: 10 (13.0) 2: 29 (37.7) 3: 36 (46.8) 4: 2 (2.6)	1: 4 (10.0) 2: 16 (40.0) 3: 19 (47.5) 4: 1 (2.5)	1: 6 (16.2) 2: 13 (35.1) 3: 17 (45.9) 4: 1 (2.7)	0.872[d]	1: 4 (13.3) 2: 9 (30.0) 3: 16 (53.3) 4: 1 (3.3)	1: 5 (17%9) 2: 10 (35.7) 3: 12 (42.9) 4: 1 (3.6)	0.881[d]	1: 4 (11.8) 2: 8 (23.5) 3: 21 (61.8) 4: 1 (2.9)	1: 5 (20.8) 2: 11 (45.8) 3: 7 (29.2) 4: 1 (4.2)	0.180[d]	1: 5 (17.2) 2: 10 (34.5) 3: 13 (44.8) 4: 1 (3.4)	1: 8 (20.5) 2: 12 (30.8) 3: 17 (43.6) 4: 2 (5.1)
CCS (class)	76, 1.46 ± 1.18 0–3 Median: 2	39, 1.62 ± 1.16 0–3 Median: 2	1.30 ± 1.20 0–3 Median: 1	0.259[b]	29, 1.59 ± 1.18 0–3 Median: 2	1.14 ± 1.24 0–3 Median:1	0.199[b]	34, 1.38 ± 1.21 0–3 Median: 1.5	1.35 ± 1.27 0–3 Median: 2	0.878[b]	28, 1.57 ± 1.23 0–3 Median: 2	38, 1.61 ± 1.22 0–3 Median: 2
6MWT-baseline (meter)	64, 372 ± 109 108–644 Median: 360	36, 376 ± 112 108–644 Median: 367.5	28, 367 ± 107 192–628 Median: 360	0.759[a]	27, 374 ± 114 108–644 Median: 350	20, 376 ± 92.0 206–570 Median: 361	0.967[a]	30, 368 ± 95.4 108–570 Median: 361	17 388 ± 119 206–644 Median: 360	0.530[a]	26, 388 ± 128 108–644 Median: 365	32, 383 ± 114 108–644 Median: 385
Patient characteristics and randomisation analysis set II												
	Safety (SAS) All	Safety (SAS) Placebo	CD133+	P	Efficacy (PPS) Placebo/CD133+	CD133+	P	Efficacy (PPS) Resp	NonResp	p	MRI early/late Placebo/CD133+	Biomarker
N	77 (mean,SD,min-max, median)	40 (mean,SD,min-max, median)	37 (mean,SD,min-max, median)		30	28		35	23		29	39
Myocardial function, perfusion and infarction												
Area of infarction Septal (segments 1,5,10,11)	58, 11 (19.0)	29, 8 (27.6)	29, 3 (10.3)	0.179[c]	22, 2 (9.1)	21, 3 (14.3)	0.664[c]	24, 5 (20.8)	19, 0 (0)	0.056[c]	19, 4 (21.1)	29, 3 (10.3)
Posterior (segments 2,6, 8,9,11)	58, 26 (44.8)	29, 13 (44.8)	29, 14 (48.3)	1.000[c]	22, 17 (77.3)	21, 12 (57.1)	0.203[c]	24, 9 (37.5)	19, 18 (94.7)	<0.001[c]	19, 11 (57.9)	29, 24 (82.8)
Anterior (segments 3,5,6,7,9,10,11)	58, 24 (41.4)	29, 13 (44.8)	29, 8 (27.6)	0.274[c]	22, 8 (36.4)	21, 11 (52.4)	0.364[c]	24, 11 (45.8)	19, 7 (36.8)	0.756[c]	19, 7 (36.8)	29, 12 (41.4)
Lateral (segments 4,7,8,9,10,11)	58, 17 (29.3)	29, 8 (27.6)	29, 9 (31.0)	1.000[c]	22, 8 (36.4)	21, 7 (33,3)	1.000[c]	24, 9 (37.5)	19, 6 (31.6)	0.755[c]	19, 7 (36.8)	29, 5 (17.2)
Combined (%) (score 5–11)	58, 24 (41.4)	29, 12 (41.4)	29 12 (41.4)	1.000[c]	22, 11 (50.0)	21, 9 (42.9)	0.763[c]	24, 11 (45.8)	19, 9 (47.4)	1.000[c]	19, 8 (42.1)	29, 16 (55.2)
Coronary artery stenosis > 50% LMCA N (%)	21 (27.3)	13 (35.1)	8 (20.0)	0.200[c]	7 (23.3)	10 (35.7)	0.390[c]	11 (32.4)	6 (25.0)	0.772[c]	12 (41.4)	11 (28.2)
Coronary artery stenosis > 50% RIVA N (%)	76, 66 (86.8)	33 (89.2)	39, 33 (84.6)	0.737[c]	29, 24 (82.8)	25 (89.3)	0.706[c]	30 (88.2)	19 (82.6)	0.697[c]	22 (75.9)	38, 35 (92.1)
Coronary artery stenosis > 50% RCX N (%)	76, 58 (76.3)	29 (78,4)	39, 29 (74.4)	0.790[c]	29, 20 (69.0)	25 (89.3)	0.103[c]	28 (82.4)	23, 17 (73.9)	1.000[c]	24 (82.8)	38, 31 (81.6)
Coronary artery stenosis > 50% RCA N (%)	76, 69 (90.8)	35 (94.6)	39, 34 (87.2)	0.432[c]	29, 25 (86.2)	27 (96.4)	0.352[c]	31 (91.2)	23, 21 (91.3)	1.000[c]	27 (93.1)	38, 35 (92.1)
Scar size (MRI)-baseline (g)	70, 31.3 ± 15.7 2–89 Median: 29.5	37, 32.2 ± 12.6 2–56 Median: 32	33, 30.2 ± 18.8 4–89 Median: 29	0.573[a]	28, 30.4 ± 12.3 2–49	25, 31.9 ± 20.8 4–89	0.755[a]	33, 27.5 ± 14.6 2–59 Median: 25	20, 37.1 ± 18.5 14–89 Median: 36	0.042[a]	27, 31.1 ± 17.3 6–89 Median:	36, 30.9 ± 15.9 6–89 Median: 27.5
Non-viable tissue (MRI) – baseline (g)	69, 24.8 ± 16.2 0–70 Median: 22	36, 25.0 ± 14.2 0–56 Median: 25.5	33, 24.5 ± 18.5 0–70 Median: 20	0.897[a]	27, 23.0 ± 13.9 0–50 Median: 22.0	25, 25.9 ± 20.4 0–70 Median: 19.0	0.551[a]	33, 21.5 ± 16.2 0–62 Median: 18.0	19, 29.6 ± 18.1 7–70 Median: 25.0	0.102[a]	26, 23.0 ± 17.6 4–70 Median: 19.5	36, 23.0 ± 15.7 3–70 Median: 18.5
LV mass (MRI) (g)	75, 182 ± 43.0 101–287 Median: 180	39, 179 ± 42.5 104–287 Median: 178	36, 185 ± 43.9 101–274 Median: 187	0.608[a]	184 ± 44.4 104–287 Median: 183	183 ± 36.7 122–270 Median: 186	0.933[a]	182 ± 38.5 104–270 Median: 186	186 ± 44.1 111–287 Median: 185	0.711[a]	178 ± 47.6 104–287 Median: 178	188 ± 44.4 104–287 Median: 187

LVEF (MRI) – baseline (%)	76, 34.3 ± 6.42 25–49 Median: 34	39, 35.6 ± 6.67 25–49 Median: 36	32.8 ± 5.89 25–48 Median: 32	0.056[b]	34.4 ± 6.46 25–49 Median: 35.0	32.5 ± 5.89 25–48 Median: 32.0	0.249[b]	32.8 ± 5.42 25–48 Median: 32	34.6 ± 7.35 25–49 Median: 35.0	0.285[a]	34.9 ± 6.34 26–49 Median: 35.0	34.1 ± 6.40 25–48 Median: 34.0
LVEDV index (MRI) – baseline (ml)	76, 109 ± 29.4 41–194 Median: 107.5	39, 106 ± 26.2 45–176 Median: 107	112 ± 32.5 41–194 Median: 109	0.432[a]	107 ± 26.4 45–176 Median: 109	107 ± 32.6 41–194 Median: 104	0.941[a]	101 ± 27.9 41–162 Median: 101	117 ± 29.1 76–194 Median: 110	0.033[a]	110 ± 35.5 41–194 Median: 109	100 ± 29.6 41–194 Median: 101
LVESV index (MRI) – baseline (ml)	76, 71.3 ± 22.4 21–141 Median: 71	39, 68.7 ± 19.2 31–110 Median: 69	74.0 ± 25.4 21–141 Median: 75	0.308[a]	71.2 ± 19.8 31–110 Median: 71.0	70.4 ± 25.3 21–141 Median: 73.0	0.893[a]	67.0 ± 20.8 21–110 Median: 69	76.7 ± 24.0 40–141 Median: 77.0	0.109[a]	72.9 ± 25.8 21–141 Median: 69.0	65.9 ± 23.2 21–141 Median: 69.0
Stress Perfusion score (mean Segment 1–17) (MRI)	58, 0.84 ± 0.39 0–1.6 Median 0.88	28, 0.83 ± 0.38 0–1.6 Median 0.84	29, 0.84 ± 0.4 0–1.6 Median 1.0	0.774[b]	24, 0.81 ± 0.36 0.2–1.6 Median 0.81	27, 0.87 ± 0.40 0–1.6 Median 1	0.330[b]	32, 0.78 ± 0.38 0–1.4 Median 0.84	19, 0.94 ± 0.38 0.4–1.6 Median 1	0.172[b]	27, 0.72 ± 0.43 0–1.6 Median 0.72	39, 0.86 ± 0.39 0.2–1.6 Median 0.86
Patient characteristics and randomisation analysis set III												
	Safety (SAS) All	Safety (SAS) Placebo	CD133 +	P	Efficacy (PPS) Placebo/CD133 +	CD133 +	P	Efficacy (PPS) Resp	NonResp	p	MRI early/late Placebo/CD133 +	Biomarker
N	77	40	37		30	28		35	23		29	39
Operative procedure and postoperative course												
CD133 + BMSC treated infarct area (% LV Segments)												
Segment 1 (%)	13 (16.9)	7 (17.5)	6 (16.2)	0.542[c]	5 (16.7)	3 (10.7)	0.707[c]	6 (17.1)	2 (8.7)	0.458[c]	4 (13.8)	4 (10.3)
Segment 2 (%)	1 (1.3)	1 (2.5)	0 (0)	1.000[c]	1 (3.3)	0 (0)	1.000[c]	1 (2.9)	0 (0)	1.000[c]	0 (0)	1 (2.6)
Segment 3 (%)	14 (18.2)	9 (22.5)	5 (13.5)	0.382[c]	8 (26.7)	5 (17.9)	0.534[c]	9 (25.7)	4 (17.4)	0.534[c]	8 (27.6)	7 (17.9)
Segment 4 (%)	44 (57.1)	24 (60.0)	20 (54.1)	0.650[c]	18 (60.0)	15 (53.6)	0.791[c]	21 (60.0)	12 (52.2)	0.597[c]	18 (62.1)	21 (53.8)
Segment 5 (%)	51 (66.2)	27 (67.5)	24 (64.9)	0.815[c]	19 (63.3)	17 (60.7)	1.000[c]	24 (68.6)	12 (52.2)	0.272[c]	20 (69.0)	20 (51.3)
Segment 6 (%)	34 (44.2)	19 (47.5)	15 (40.5)	0.647[c]	14 (46.7)	10 (35.7)	0.435[c]	15 (42.9)	9 (39.1)	1.000[c]	14 (48.3)	12 (30.8)
Segment 7 (%)	26 (33.8)	14 (35.0)	12 (32.4)	1.000[c]	10 (33.3)	10 (35.7)	1.000[c]	12 (34.3)	8 (34.8)	1.000[c]	11 (37.9)	13 (33.3)
Segment 8 (%)	6 (7.8)	4 (10.0)	2 (5.4)	0.676[c]	3 (10.0)	2 (7.1)	1.000[c]	4 (11.4)	1 (4.3)	0.639[c]	3 (10.3)	5 (12.8)
Segment 9 (%)	17 (22.1)	9 (22.5)	8 (21.6)	1.000[c]	6 (20.0)	6 (21.4)	1.000[c]	9 (25.7)	3 (13.0)	0.329[c]	6 (20.7)	7 (17.9)
Segment 10 (%)	59 (76.6)	30 (75.0)	29 (78.4)	0.792[c]	22 (73.3)	25 (89.3)	0.182[c]	29 (82.9)	18 (78.3)	0.738[c]	24 (82.8)	33 (84.6)
Segment 11 (%)	65 (84.4)	31 (77.5)	34 (91.9)	0.117[c]	22 (73.3)	27 (96.4)	0.026[c]	32 (91.4)	17 (73.9)	0.135[c]	26 (89.7)	32 (82.1)
Segment 12 (%)	54 (70.1)	27 (67.5)	27 (73.0)	0.628[c]	19 (63.3)	21 (75.0)	0.402[c]	26 (74.3)	14 (60.9)	0.385[c]	21 (/2.4)	26 (66.7)
Segment 13 (%)	37 (48.1)	20 (50.0)	17 (45.9)	0.821[c]	15 (50.0)	12 (42.9)	0.610[c]	18 (51.4)	9 (39.1)	0.426[c]	12 (41.4)	19 (48.7)
Segment 14 (%)	22 (28.6)	12 (30.0)	10 (27.0)	0.806[c]	10 (33.3)	9 (32.1)	1.000[c]	13 (37.1)	6 (26.1)	0.410[c]	8 (27.6)	14 (35.9)
Segment 15 (%)	56 (72.7)	27 (67.5)	29 (78.4)	0.317[c]	19 (63.3)	25 (89.3)	0.031[c]	29 (82.9)	15 (65.2)	0.209[c]	22 (75.9)	29 (74.4)
Segment 16(%)	67 (87.0)	33 (82.5)	34 (91.9)	0.314[c]	23 (76.7)	26 (92.9)	0.147[c]	30 (85.7)	19 (82.6)	1.00[c]	25 (86.2)	33 (84.6)
Segment 17 (%)	43 (55.8)	22 (55.0)	21 (56.8)	1.000[c]	18 (60.0)	16 (57.1)	1.000[c]	21 (60.0)	13 (56.5)	1.000[c]	13 (44.8)	26 (66.7)
Distal CABG-anastomoses	3.44 ± 0.90	3.35 ± 0.95	3.54 ± 0.84	0.426[b]	3.4 ± 0.97	3.64 ± 0.87	0.351[b]	3.57 ± 0.884	3.43 ± 0.992	0.542[b]	3.48 ± 0.871	3.49 ± 0.914
N	2–5 Median: 3	2–5 Median: 3	2–5 Median: 3		2–5 Median: 3	2–5 Median 4		2–5 Median: 4	2–5 Median: 3		2–5 Median: 3	2–5 Median: 3
Aortic clamping time (min)	65.9 ± 21.6 24–154 Median: 62	64.0 ± 18.4 24–110 Median: 63	68.0 ± 24.7 37–154 Median: 60	0.422[a]	63.5 ± 18.8 24–110 Median: 63.0	68.4 ± 26.9 37–154 Median: 59.5	0.429[a]	67.7 ± 23.4 37–154 Median: 65.0	63.1 ± 22.4 24–131 Median: 59	0.460[a]	59.6 ± 17.7 24–97 Median: 55	61.8 ± 16.2 24–97 Median 60
ECC time (min)	106 ± 34.8 38–236 Median: 102	100 ± 27.5 38–161 Median: 102	112 ± 40.9 53–236 Median: 106	0.155[a]	102 ± 23.9 38–161 Median: 102	113 ± 44.7 53–236 Median: 103	0.248[a]	109 ± 39.0 53–236 Median: 102	105 ± 30.4 38–185 Median: 102	0.644[a]	99.9 ± 37.6 38–236 Median: 93	106 ± 32.0 38–236 Median: 102
Postoperative												
CK max (U/l)	1299 ± 2525 192–16,584 Median: 583	1565 ± 2955 192–16,584 Median: 692	1012 ± 1959 200–12,062 Median: 547	0.119[b]	1262 ± 1885 192–10,116 Median: 711	752 ± 664 200–3263 Median: 561	0.205[b]	913 ± 805 203–4056 Median: 672	1171 ± 2086 192–10,116 Median: 547	0.369[b]	701 ± 552 192–2800 Median: 562	1229 ± 1717 198–10,116 Median: 677
CK-MB max (U/l)	60.0 ± 118 4–892 Median: 37.0	57.6 ± 93.9 10–611 Median: 31	62.5 ± 141 4–892 Median: 41	0.642[b]	60.6 ± 107 24–611 Median: 30	39.8 ± 15.7 4–79 Median: 41.5	0.575[b]	60.1 ± 97.9 17–611 Median: 42	36.2 ± 20.5 4–107 Median: 28	0.028[b]	36.1 ± 15.8 17–82 Median: 30.0	42.8 ± 23.1 21–115 Median: 31.0

Data are n (%), mean ± SD, minimum-maximum, median (interquartile range). BMSC bone marrow stem cell, CABG coronary artery bypass surgery, PCI, percutaneous coronary intervention; AMI, acute myocardial infarction; CAD, coronary artery disease; LAD, left anterior descending coronary artery; LCX, left circumflex coronary artery; RCA, right coronary artery; CK, creatine kinase; MRI, magnetic resonance imaging; ACE, angiotensin-converting enzyme; ARB, angiotensin receptor blocker; MRA, mineralocorticoid receptor antagonist. Perfusion score: 0- normal perfusion, 1 – hypoperfusion, 2 – strong reduced perfusion.

[a] *t*-Test for independent samples.

[b] *U* test Mann-Whitney.

[c] Fisher's exact test.

[d] Chi-square test.

production of placebo or CD133$^+$. The appearance of the final placebo and cellular product was indistinguishable to the investigators. In the event of a medical emergency, and necessity for breaking the code, an emergency envelope was available 24 h a day, 7 days a week for a member of the treatment team responsible for patient recruitment and clinical assessment, bone marrow harvest and performing the treatment.

2.5. Magnetic Resonance Imaging

Cardiac MRI was performed in the participating study centres according to an identical standard protocol. Each centre provided test MRI scans to ensure image quality and adherence to the protocol before recruiting patients into the study. Patients were scanned in the supine position in 1·5 T scanners with dedicated cardiac software, using retrospective ECG gating and a phased array receiver coil. Standard imaging protocol included morphologic images of the whole thorax, functional measurements of the heart for LV-volumes and function, perfusion-MRI with adenosine for detection of ischemia, and gadolinium late enhancement measurement for the assessment of LV viability. LV volumes were measured based on a series of breath-hold SSFP-CINE sequences. An end-diastolic, four-chamber view of the left ventricle at end-expiration provided the reference image on which a series of contiguous short axis slices was positioned to cover the entire left ventricle. Infarct volume was assessed on late-gadolinium enhancement MRI images in short axis orientation and vertical long axis. All MRI analyses were performed in a core lab at the University Hospital Göttingen, Department of Diagnostic and Interventional Radiology, whose group members were unaware of treatment assignments. Core lab MRI readings were used to evaluate patient eligibility for the trial. Images were analysed with QMass MR 7.6 software (Medis Medical Imaging Systems).

2.6. Interventions

Placebo (5 ml saline + 10% autologous serum) or CD133$^+$ stem cell (5 ml purified CD133$^+$ BMSC in saline + 10% autologous serum) were administered intramyocardially into the infarction border zone (penumbra) during the cardiac surgical procedure. The procedure was performed with extracorporeal circulatory support, aortic cross clamping and cardioplegic arrest. The injections were done before cross-clamp release. The 5 ml suspensions were distributed in 15–20 injections applied within 3 min in the region of interest (infarct border zone) according to the affected left ventricular segments (see Supplement Fig. 1) at the end of bypass surgery. Not more than one injection per square centimetre was performed. During the whole duration of the study, patients were treated per the standards of the centres and the American Heart Association (AHA) guidelines.

2.7. Outcomes

2.7.1. Prespecified Primary Outcome

Delta (Δ) LVEF at 180 d postoperatively versus baseline (Δ 180 d vs. 0), measured by MRI at rest.

2.7.2. Prespecified Secondary Outcome

Objectives were (Δ 6 m vs. 0) left ventricular dimensions (LVEDV, LVESV), classification of heart failure (NYHA, CCS), NT-proBNP, scar and non-viable tissue, 6-minute-walk-test, adverse events (AE), serious adverse events (SAE), major adverse cardiac events (MACE), Serious Unexpected Serious Adverse Reactions (SUSAR), and Quality-of-Life (QoL). MACE outcome analysis was performed at 24 months.

2.7.3. Post Hoc Analysis

Kaplan-Meier survival (long term vigilance registry approved by the ethics committee of the University Medicine Rostock: A 2017-0031).

2.8. Biomarkers

2.8.1. Prespecified

Distinct hematopoetic and endothelial CD133$^+$ EPC subpopulations and angiogenesis capacity were tested in a cohort of 39 patients in bone marrow (BM) and peripheral blood (PB) employing coexpression analysis using four-laser flow cytometric methods (LSR II, Becton Dickinson, Heidelberg, Germany) for costaining panel enumeration of EPC (Costaining panel CD133, 34, 117, 184, 309, 105, 45) and circulating endothelial cells (CEC) (Costaining panel: CD31, 146, 34, 45, 105, 184, 309) as well as in vitro CFU-EC, CFU-Hill and in vivo Matrigel plug assay. NT-proBNP as well as virus analysis were performed for EBV, CMV, and Parvovirus by IgG and antigen analysis in peripheral blood serum. Post hoc analysis before final data closure was performed for serum angiogenesis factors and cytokines.

2.8.2. Post Hoc Analysis

BM subpopulation analysis and SH2B3 mRNA RT-PCR in peripheral blood (PB): Methods and analysis of biomarkers studied in BM CD133$^+$ and PBMNCs samples using cytometric bead array (CBA) and enzyme-linked immunosorbent assay (ELISA) and RT-PCR are depicted in Supplement Appendix 3. Samples were taken from informed study patients who gave their written consent according to the Declaration of Helsinki. (approval by the Ethical committee, Rostock University Medical Center 2009; No. HV-2009-0012). Analyses and examinations were performed before unblinding of the trial and under careful adherence to the protection of data privacy (pseudonyms).

2.9. Statistical Analysis

The stratification of the primary analysis by centre was neglected in the sample size calculation. Instead of the analysis of covariance (ANCOVA) used in the primary analysis, the two-sample *t*-test scenario with equal variances was considered. Sample size was determined with the assumption of a two-sided type I error (α) at 5% and a type II error (β) at 10% (i.e. a power at 90%). The scenario of a difference in LVEF at month 6 post-operatively between the two treatment arms of 4 to 5% was considered as a clinically relevant difference. With a difference of 4.5 and a standard deviation of 7.5, at least $n = 60$ patients per group were considered necessary and, with an additional 15% drop-out rate, a total of at least 142 patients were to be randomised. Sample size was calculated using the commercial program nQuery Advisor 5.0, section 8, Table MTT0-1 (Hofmann et al., 2002). Computation was realized using central and non-central t-distribution where the non-centrality parameter is $\sqrt{n}\,\delta/\sqrt{2}$ and δ is defined as effect size $|\mu 1\text{-}\mu 2|/\sigma$ (O'Brien et al., 1993). The two-sided hypothesis for the continuous primary efficacy variable LVEF at 6 months (180 days) postoperatively will be assessed using analysis of covariance (ANCOVA) adjusting for baseline LVEF. Statistical analyses, final data set calculation, and preparation were performed by Koehler GmbH, Freiburg, and G.K., who was not involved in patient recruitment and follow-up.

Multivariance analysis included the ANCOVA, MANCOVA comparison of Placebo vs. CD133 + and for LVEF responders vs. non-responders group using all single parameters of CRF outcome dataset specified in Table 1 and biomarker analysis listed in Appendix 3. Given the complexity of variables additionally machine learning was applied for validation of parameter correlations.

2.10. Data Analysis With Machine Learning

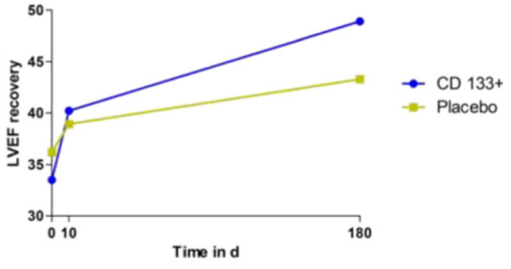
Identifying key features and classification of the comprehensive patient data was obtained by employing supervised and unsupervised machine learning (ML) algorithms (Kuhn, 2008). We preprocessed the data while removing features with low variance and high correlation for dimension reduction following best practices recommendations. Missing measurements were filled with zeros as frequently used in

Table 2
Overall results of the ANCOVA[a] for primary and secondary outcome parameters in Placebo vs. CD 133$^+$ BMSC (PPS; $n = 58$).

	Estimated Baseline	Estimate (at 180 days)	Standard-error	95% CI	*p*-Value
LVEF (%)					
Placebo[b] ($n_{evaluable} = 30$)	33.52	42.30	2.17	[38.0, 46.6]	<**0.001**
CD133+[b] ($n_{evaluable} = 28$)		43.93	2.33	[39.0, 48.5]	<**0.001**
ΔCD133+-Placebo[c]		2.58	3.13	[−3.7, 8.9]	0.414
LVEDV (index)					
Placebo[b] ($n_{evaluable} = 30$)	107.12	100.97	11.21	[79.0, 122.9]	0.113
CD133+[b] ($n_{evaluable} = 28$)		105.86	12.01	[82.3, 129.4]	0.882
ΔCD133+-Placebo[c]		5.80	7.40	[−9.1, 20.7]	0.437
LVESV (index)					
Placebo[b] ($n_{evaluable} = 30$)	71.52	58.87	8.90	[41.4, 76.3]	<**0.001**
CD133+[b] ($n_{evaluable} = 28$)		61.54	9.53	[42.8, 80.2]	0.053
ΔCD133+-Placebo[c]		2.51	6.04	[−9.6, 14.6]	0.680
Scar size (g)					
Placebo[b] ($n_{evaluable} = 27$)	31.48	34.52	3.36	[27.9, 41.1]	0.087
CD133+[b] ($n_{evaluable} = 23$)		28.13	3.94	[20.4, 35.9]	0.212
ΔCD133+-Placebo[c]		−7.53	3.19	[−14.0, −1.1]	**0.023**
Non-viable tissue (g)					
Placebo[b] ($n_{evaluable} = 27$)	25.20	27.78	3.73	[20.5, 35.1]	0.099
CD133+[b] ($n_{evaluable} = 23$)		21.57	4.38	[13.0, 30.1]	0.177
ΔCD133+-Placebo[c]		−7.71	3.13	[−14.0, −1.4]	**0.018**
LV mass (g)					
Placebo[b] ($n_{evaluable} = 30$)	183.93	173.87	15.78	[142.9, 204.8]	**0.025**
CD133+[b] ($n_{evaluable} = 28$)		171.00	16.91	[137.9, 204.1]	0.051
ΔCD133+-Placebo[c]		−3.23	6.83	[−16.9, 10.5]	0.638
6 MWT (meter)					
Placebo[b] ($n_{evaluable} = 25$)	384.73	434.80	21.14	[393.4, 476.2]	**0.039**
ΔCD133+[b] ($n_{evaluable} = 17$)		441.74	31.10	[380.8, 502.7]	0.058
CD133+-Placebo[c]		20.19	29.72	[−40.1, 80.5]	0.501
NT-proBNP					
Placebo[b] ($n_{evaluable} = 28$)	1489.83	766.36	655.89	[−519.2, 2051.9]	**0.037**
CD133+[b] ($n_{evaluable} = 26$)		1465.50	706.34	[81.1, 2849.9]	0.699
ΔCD133+-Placebo[c]		996.82	324.15	[344.7, 1648.9]	**0.004**

Source: P132_perfect - EFF02T.sas Data Extract: 15JUL2016 Generation Date: 10AUG2016 21:02.
Bold values indicate significance at $p < 0.05$.
[a] ANCOVA in final analysis (GK) For primary endpoint analysis in SAP-CTR (Appendix 1) an additional analysis was made using a mixed model analysis for repeat measures approach (MMRM) in order to compensate possible artefacts due to incomplete data groups. This was the approach used for the interim analysis as well.
[b] Average change from Baseline.
[c] Difference in Treatment Groups.

standard data imputation practices. We compared the following supervised algorithms: AdaBoost, Support Vector Machines (SVM) and Random Forest (RF) (Forman and Cohen, 2004). Small clinical datasets are often prone to overfitting. We employed classifiers that are suitable for training on small data sets for a comparison of features given little training and chose the most appropriate algorithm according to accuracy and robustness towards overfitting (Saeb and Al-Naqeb, 2016). Supervised ML models have been 10-fold cross-validated. We then applied feature selection from AdaBoost and RF to further reduce the number of features to <20. We employed t-distributed stochastic neighbor embedding (t-SNE) for unsupervised machine learning classification and nonlinear dimensionality reduction (Maaten and Hinton, 2008).



	before	10 days	180 days	p
CD 133$^+$ (N=14)	33.5	40.2	48.9	0.015/0.001
Placebo (N=15)	36.2	38.9	43.3	0.215/0.077

Fig. 2. Early and late recovery of LVEF in Placebo and CD133$^+$ groups. MRI analysis of LVEF (%) is depicted in 29 patients with intermediate MRI at day 10 postoperatively and at 180 days. *p value for delta LVEF at 10 days versus 0. #p value for delta LVEF at 6 months versus 10 days.

2.11. Role of the Funding

The funding had no role in study design, in the collection, analysis, interpretation of data, in the writing of the report, and in the decision to submit the manuscript for publication. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

3. Results

Patient baseline characteristics analysed in SAS and PPS patient populations are depicted in Table 1. Analysis follows the description of prespecified cohort analyses SAS ($n = 77$) and PPS ($n = 58$) placebo vs. CD133$^+$ (Fig. 1). Post hoc analysis was additionally performed to

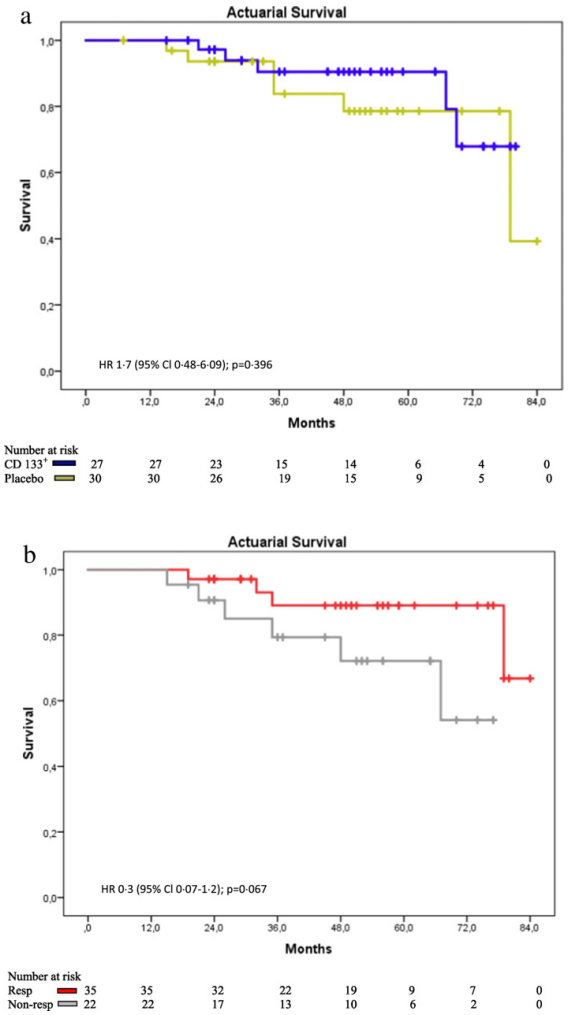


Fig. 3. a: Kaplan-Meier survival analysis in longterm follow-up: Placebo vs. CD133$^+$. b: Kaplan-Meier survival analysis in longterm follow-up: Responder vs. Non-responder

analyse factors influencing primary endpoint outcome. For this, patients were grouped as responders (increase in LVEF ≥ 5% at 180 days) or and non-responders (increase in LVEF < 5% at 180 d). According to this post hoc analysis 35/58 (60.3%) patients were responders and 23/58 (39.7%) did not improve in LVEF. This responder/non-responder (NR) ratio was similar in the placebo group 57/43% (R/NR: 17/13 pt.) and in the CD133$^+$ group 64/36% (R/NR: 18/10 pt.) respectively (placebo vs. CD133$^+$: $p = 0.373$).

3.1. Safety Outcome Analysis

Prespecified safety outcome ($n = 77$): Up to 180 d follow-up, two MACE-incidents occurred in 2.6% of the patients ($n = 2$), ventricular arrhythmias occurring in one patient in the placebo group and one in the CD133 + group (Supplement Table 2). During the main trial phase until 180 days 80 days there was a total of 49 SAE, 24 (15 subjects) in the placebo group and 25 (19 subjects) in the CD133$^+$ group (Supplement Table 3). There were no statistical differences observed between the placebo and the CD133$^+$ group neither overall nor in any of the system organ classes. The most common SAEs were cardiac disorders such as atrial fibrillation, ventricular arrhythmia and cardiac failure, as well as respiratory and wound infections (Supplement Table 3). Of these, 19 were classified as possibly related (placebo 13/68, CD133$^+$ 6/67; $p = 0.156$) (Supplement Table 4). There were no signs of related classifications of adverse events (Supplement Table 5) or unwanted tissue formation (data not shown) for CD133$^+$ treatment in the initial patient treatment follow-up to 180 days. Post hoc safety analysis in PPS ($n = 58$): NR revealed increase in lung infection ($p = 0.021$) (Supplement Table 6).

3.2. Efficacy Outcome Analysis

The PPS efficacy analysis group ($n = 58$) was characterized by reduced pump function post MI (measured in MRI at rest) with baseline LVEF 33.5%, SD ± 6.26% [Min-Max-25–49], $n = 58$.

Table 3
Overall results of the ANCOVA for primary and secondary parameters in Responder vs. Non-responder ($n = 58$).

	Estimated Baseline	Estimate (at 180 days)	Standard-error	95% CI	*p*-Value
LVEF (%)					
Responder[a] ($n_{evaluable} = 35$)	33.52	49.34	3.76	[42.0; 56.7]	<**0.001**
Non-responder[a] ($n_{evaluable} = 23$)		33.57	5.73	[22.3; 44.8]	0.287
Responder - Non-responder[b]		17.10	2.08	[12.9; 21.3]	<**0.001**
LVEDV (index)					
Responder[a] ($n_{evaluable} = 35$)	107.12	90.77	9.72	[71.7; 109.8]	**0.009**
Non-responder[a] ($n_{evaluable} = 23$)		122.43	14.80	[93.4; 151.4]	0.483
Responder - Non-responder[b]		−20.98	7.58	[−36.2; −5.8]	**0.008**
LVESV (index)					
Responder[a] ($n_{evaluable} = 35$)	71.52	46.66	8.99	[29.0; 64.3]	<**0.001**
Non-responder[a] ($n_{evaluable} = 23$)		80.70	13.69	[53.9; 107.5]	0.376
Responder - Non-responder[b]		−27.93	5.02	[−38.0; −17.8]	<**0.001**
Scar size (ml)					
Responder[a] ($n_{evaluable} = 31$)	31.48	27.48	2.86	[21.9; 33.1]	0.980
Non-responder[a] ($n_{evaluable} = 19$)		38.26	4.67	[29.1; 47.4]	0.934
Responder - Non-responder[b]		−8.19	3.50	[−15.2; −1.1]	**0.024**
Non-viable tissue (ml)					
Responder[a] ($n_{evaluable} = 31$)	25.1	20.81	3.12	[14.7; 26.9]	0.841
Non-responder[a] ($n_{evaluable} = 19$)		31.63	5.09	[21.7; 41.6]	0.981
Responder - Non-responder[b]		−8.55	3.56	[−15.7; −1.4]	**0.021**
LV mass (ml)					
Responder[a] ($n_{evaluable} = 35$)	183.93	168.71	12.89	[143.5; 194.0]	**0.032**
Non-responder[a] ($n_{evaluable} = 23$)		178.22	19.61	[139.8; 216.7]	0.092
Responder - Non-responder[b]		−6.01	7.01	[−20.1; 8.1]	0.396
6 Minute Walk Test (meter)					
Responder[a] ($n_{evaluable} = 27$)	384.73	430.57	18.23	[394.8; 466.3]	**0.016**
Non-responder[a] ($n_{evaluable} = 15$)		450.27	32.81	[386.0; 514.6]	0.141
Responder - Non-responder[b]		−7.19	29.82	[−67.7; 53.4]	0.811
NT-proBNP					
Responder[a] ($n_{evaluable} = 32$)	1489.83	588.41	561.48	[−512.1; 1689]	**0.005**
Non-responder[a] ($n_{evaluable} = 22$)		1851.45	816.69	[250.7; 3452]	0.867
Responder - Non-responder[b]		−1318.40	326.42	[−1975; −661.7]	**0.002**

Source: P132_perfect - EFF02T.sas Data Extract: 15JUL2016 Generation Date: 10AUG2016 21:02.
[a] Average change from Baseline.
[b] Difference in Treatment Groups, CI = Confidence Interval.

3.2.1. Prespecified Primary Endpoint

Six months post treatment the left ventricular function showed a considerable increase in LVEF of +9.6% ± SD 11.3% [Min-Max-13–42], $p < 0.001$ ($n = 58$). To discriminate early improvement of left ventricular function by CABG revascularization and late myocardial reverse remodeling, additional intermediate MRI analysis at hospital discharge was available in a subgroup of patients ($n = 29$). This revealed mainly late (day 10–180) increase of ΔLVEF by +6.5%, SD ± 7.92% [Min-Max-11–23], $p = 0.007$ ($n = 29$). In ANCOVA analysis of the primary endpoint the placebo group improved from baseline LVEF 33.5% to 42.3% at 180 days (ΔLVEF +8.8%, and the $CD133^+$ group LVEF was raised from 33.5% to 43.9% (ΔLVEF +10.4% (Table 2). Treatment group difference $CD133^+$ versus placebo with +2.58, $p = 0.414$ was not statistically significant in ANCOVA analysis (Table 2). $CD133^+$ stem cell group displayed ΔLVEF improvement mainly in the late phase (day 10–180 ΔLVEF) with +8.8%,SD ± 6.38% [Min-Max-4–10], $p = 0.001$ ($n = 14$) versus placebo controls (day 10–180 ΔLVEF) +4.3%, SD ± 8.8% [Min-Max-11–23], $p = 0.077$ ($n = 15$) (Fig. 2).

3.2.2. Prespecified Secondary Endpoint

The delta (Δ) change of ventricular dimensions between the $CD133^+$ versus placebo groups after 180 days was not significant in ANCOVA for LVESV index 2.51 ml/m^2, $p = 0.680$ and for LVEDV index +5.80 ml/m^2, $p = 0.437$ (Table 2). Increased reductions in scar size by −7.53 g, $p = 0.023$ and non-viable tissue by −7·71 g, $p = 0.018$ (Table 2) were detected in $CD133^+$ versus placebo.

Improvement (Δ) of segmental myocardial perfusion MRI at 180 days versus vs. baseline was observed for $CD133^+$ ($p = 0.006$), but not in placebo group ($p = 0.065$) (Supplement Table 1). Improvement (Δ) of hypoperfused LV-segments after stem cell/placebo injections under adenosine stress induction was present in $CD133^+$ group ($p = 0.006$) in comparison to non-injected segments ($p = 0.057$) as compared to placebo group (injected segments $p = 0.045$; non-injected segment $p = 0.140$) (Supplement Table 1). In contrast, the reduction (Δ) of NT-proBNP values was elevated in placebo versus $CD133^+$ ($p = 0.004$) (Table 2).

3.2.3. Prespecified Survival

100% at 180 days. Post hoc actuarial computed mean survival time was 70.1 ± 4.75 months ($CD133^+$) vs. 72.0 ± 3.46 months (placebo), and at 5 years follow-up 76.8% ($CD133^+$)/88.1% survival (placebo), HR 1.7 (95% CI 0.48–6.09); $p = 0.396$) (Fig. 3a).

3.3. Responder/Non-responder

In post hoc primary endpoint analysis treatment responders were defined as having a ΔLVEF at 180 days versus baseline higher than 5%. This results in dissemination of 35 responders in a cohort of 58 patients were characterized by an overall increase in ΔLVEF in ANCOVA at 180 d/0 of +17.1% (Table 3). LVEF increase was +19.1% in $CD133^+$ vs. +13.9% in placebo, $p = 0.099$, $n = 35$ (data

Table 4
Analysis of angiogenesis related biomarkers in blood.

Responder versus non-responder						
Biomarker (peripheral blood, unit)	Time point	Responder ($n = 15$)	P 10 days vs 0	Non-responder ($n = 8$)	P 10 days vs 0	P^A R vs NR
SH2B3 mRNA (ΔCT %)	0	−1.17 ± 0.28	…	−1.56 ± 0.51	…	0.073
CD34	0.	0.072 ± 0.05	0.197	0.039 ± 0.017	0.116	**0.027**
(% MNC) -EPC	10 d	0.059 ± 0.048		0.027 ± 0.01		**0.026**
CD133	0	0.048 ± 0.031	0.245	0.021 ± 0.011	0.932	**0.005**
(% MNC) – EPC	10 d	0.041 ± 0.039		0.021 ± −0.013		0.105
CD133,117	0	0.019 ± −0.016	0.421	0.007 ± 0.008	0.765	**0.024**
(% MNC) EPC	10 d	0.022 ± 0.024		0.006 ± 0.004		**0.024**
CD146	0	1.1 ± 0,57	…	2.2 ± 1.3	…	0.053
(% MNC) -CEC	10 d	1.72 ± 1.73		1.86 ± 1.53		0.853
IGFBP-3 (ng/ml)	0	2121.9 ± 487.1	0.115	1623.7 ± 651.4	0.257	0.089
	10 d	1753.6 ± 830.8		1378.4 ± 518.7		0.261
VEGF (pg/ml)	0	24.6 ± −36.6	0.015	39.6 ± 33.4	0.913	0.056
	10 d	51.2 ± 55.8		40.8 ± −44.5		0.528
IP-10 (pg/ml)	0	96.7 ± 42.6	0.04	157.6 ± 94.5	0.01	0.076
	10 d	63.3 ± 28.3		95.8 ± 85.2		0.324
EPO (mlU/ml)	0	5.9 ± 3.7	0.001	16.9 ± 14.1	0.006	**0.023**
	10	60.1 ± 27.7		42.1 ± 23.9		0.180
Placebo versus CD133 +						
Biomarker (peripheral blood, unit)	Time point	Stem cell ($n = 11$)	P	Control ($n = 13$)	P	P^A
SH2B3 mRNA (ΔCT %)	0	−1.35 ± 0.45	…	−1.29 ± 0.41	…	0.756
CD34 (% MNC) -EPC	0.	0.062 ± 0.037	0.128	0.064 ± 0.053	0.250	0.975
	10 d	0.041 ± 0.038		0.058 ± 0.047		0.363
CD133 (% MNC) – EPC	0	0.04 ± 0.03	0.338	0.04 ± 0.029	0.619	0.995
	10 d	0.032 ± 0.026		0.038 ± 0.032		0.637
CD133,117 (% MNC) – EPC	0	0.014 ± 0.013	0.902	0.016 ± 0.017	0.265	0.892
	10 d	0.015 ± 0.02		0.019 ± 0.022		0.626
CD146 (% MNC) -CEC	0	1.53 ± 1.33	…	1.48 ± 0.67	…	0.919
	10 d	1.64 ± 1.55		1.87 ± 1.74		0.750
IGFBP-3 (ng/ml)	0	1950.6 ± 689.9	0.139	1946.8 ± 507	0.231	0.972
	10 d	1561.6 ± 783.2		1679.4 ± 742.6		0.715
VEGF (pg/ml)	0	30.2 ± 29.1	0.142	29.6 ± 39.1	0.124	0.961
	10 d	55.8 ± −58.5		38.5 ± 44.7		0.293
IP-10 (pg/ml)	0	129.2 ± 96.7	0.011	102.9 ± 34.6	0.001	0.275
	10 d	83.2 ± 77.9		64.5 ± 22.7	…	0.457
EPO (mlU/ml)	0	7.7 ± 3.1	0.001	10.3 ± 12.6	0.001	0.561
	10 d	53.5 ± −30.6		56.4 ± 25.5		0.814

Responder versus non-responder and placebo versus $CD133^+$ groups were analysed for change in biomarkers of peripheral blood samples between preoperative (Assessment I) and day 10 postoperative (discharge). The data are derived from the Rostock cohort with complete analysis (per protocol clinical dataset and biomarker). In this cohort all samples were immediately processed to avoid any change of the samples due to storage or transport. Data are expressed as mean values ± Standard deviation, P-value between time point 0 and 10 days, P^A -value between responder/non-responder, stem cell/control in each time point, PB – peripheral blood, EPO - Erythropoeitin.

not shown). In contrast, non-responders showed a ΔLVEF at 180 d/0 by 0%, SE ±5·73% [CI 22.3; 44.8] $p = 0.287$ (placebo/NR +3·3%, $CD133^+$/NR-2·4%).

Post hoc secondary endpoint: Responders showed a significant reduction in LV-dimensions (LVEDV $p = 0.008$, LVESV $p = 0.0001$) and reduction in NT-pro-BNP, $p = 0.002$ compared to non-responders

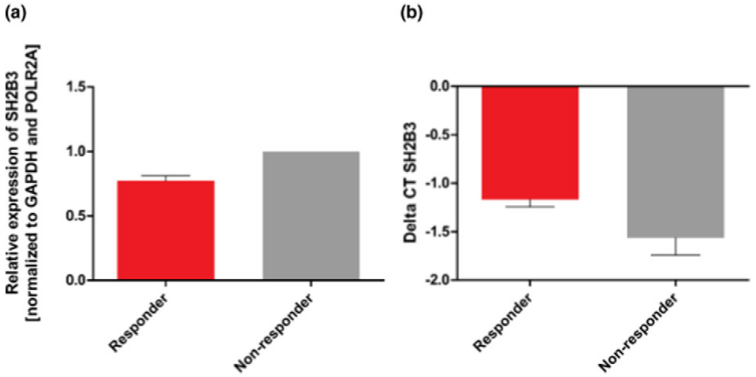


Fig. 4. SH2B3 expression analysis in peripheral blood of responder and non-responder. Whole blood samples were obtained from 21 patients before coronary artery bypass graft (CABG) revascularization. Relative expression of SH2B3 (a) and corresponding ΔCT values (b) were calculated using the $2^{-\Delta\Delta CT}$ method. All values are presented as mean ± SEM and normalized to GAPDH and POLR2A. $n = 13$ (responder); $n = 8$ (non-responder). ΔCT values: $p = 0.073$.

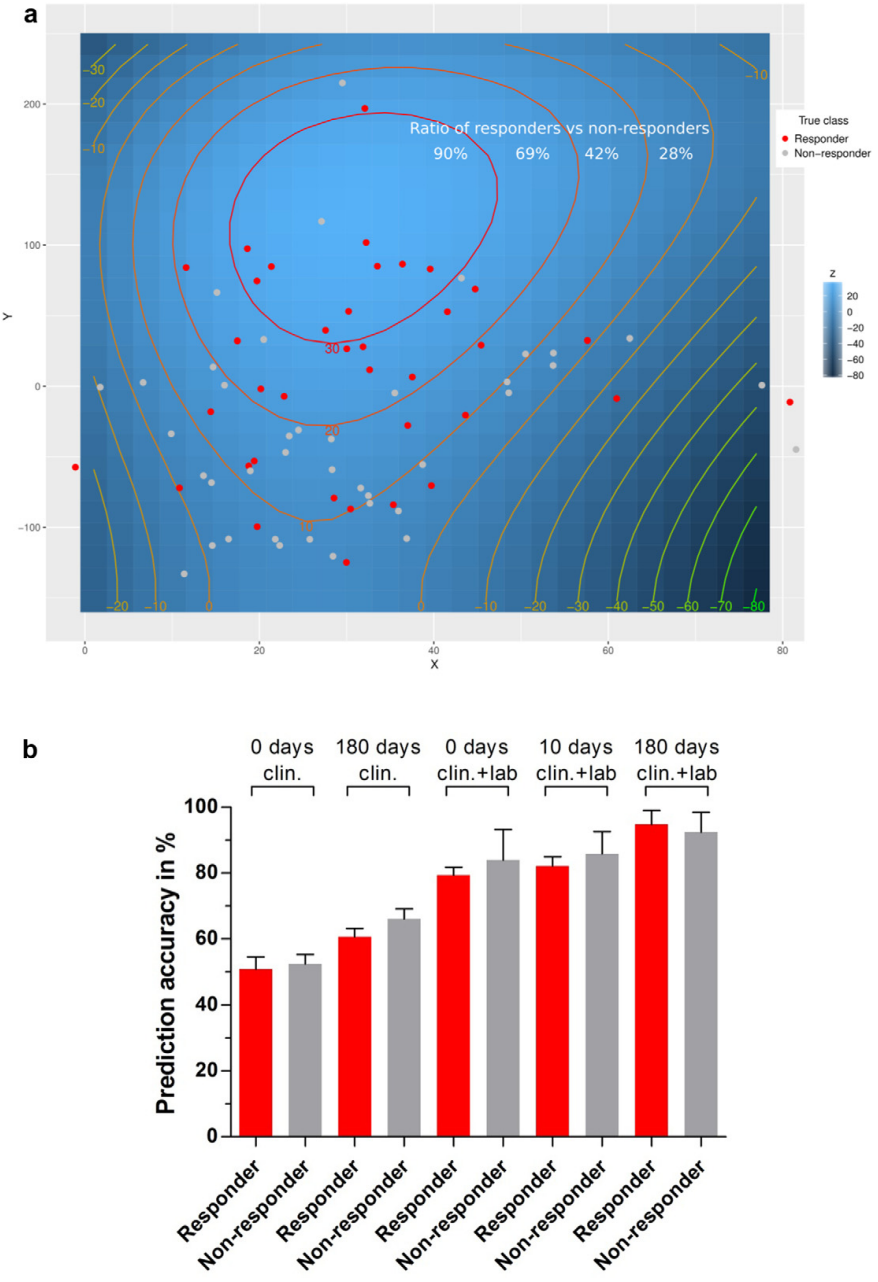


Fig. 5. a Three-dimensional t-SNE calculation of the Rostock subgroup. The variables x and y refer to the newly calculated features that are used to classify the patients into distinct groups. The model was subsequently fitted by a polynomial (n^3) equation to visualize the z-axis as a geographic profile. The respective colors for the responder (red dot) and non-responder (grey dot) patients have been added afterwards. The classified groups have been roughly summarized by a red and grey dashed line. Results are obtained after 3000 iterations. The calculation of the ratio between responder and non-responder is indicated for each circle. It is more likely for the non-responder group to be located at smaller z-values ($z < 20$, ratio < 42%). The responders tend to be enriched within the light blue areas ($z > 20$) including a ration > 69%. b Obtained supervised ML prediction results for pre- and postoperative time points (0 days to 180 days) of the clinical and clinical & laboratory dataset to distinguish between responder and non-responder. The graph shows the true positive prediction results of five independent feature selected ML models (AdaBoost for feature selection and RF for final prediction).The error bars indicate the respective accuracy standard deviation for the constructed models that have been obtained after 100 iterations. The 100 model iterations are significant different according to one-way ANOVA ($p < 0.001$).

Table 5
Machine-learning selected parameters for diagnostic discrimination of responders and non-responders.

Computationally selected features for the multi-centric clinical trial data subset (0–180 days) $N = 58$	Weights for the selected features	Computationally selected features for the clinical trial data and laboratory biomarker subset of the Rostock group (day 0 - preoperative) $N = 31$	Weights for the selected features
DeltaViable tissue 6 m/0	2.554	NT proBNP 0	9.718
Triglycerides 0	2.260	VEGF_I	7.810
Scarsize 6 months	2.159	Erythropoietin_I	4.262
DeltaScarsize 6 m/0	2.063	Vitronectin_I	3.898
Nonviable tissue 6 months	1.999	CFU_Hill_I	2.871
Body mass index 0	1.982	CD45Neg_EPC_I	2.186
6MWT 0	1.974	CD117_184_PB_EPC_IHG_I	2.146
DeltaEF 6 m/0	1.967	CD45_117_184_EPC_I	2.118
6MWT 10 days	1.920	CD45_133_146_PB_CEC_I	1.969
LVEF 0	1.890	Thrombocytes I	1.951
Bypasstime min	1.883	IGFBP-3_I	1.922
Euroscore 0	1.874	CD133 pro ml PB_I IHG	1.910
CKmax	1.857	CD146_PB_CEC_I	1.799
Scarsize 0	1.771	CD105_PB_CEC_I	1.793
NTproBNP 0	1.771	CD45_133_34_105_PB_CEC_I	1.489
Crossclamptime	1.675	MatrigelPlug_PB_31_I	1.475
Delta6MWT 6 m/0	1.673	CD45_133_34_117_309_EPC_I	1.420
Creatinine 0	1.645	Delta_CT_SH2B3_I	1.393
LVESV 0	1.604	Weight	1.363
Weight 0	1.389	LVESV I 0	1.352
Accuracy	63.35%	Accuracy	81.64%

Selected features of the AdaBoost ML algorithm showing the most informative selection criteria for the subsequently created ML models. The features are ordered due to their calculated weights in a decreasing manner. Accuracies are based on 100 independent predictions of 10-fold cross-validation calculations (Model has been built after AdaBoost feature selection and random forest feature learning).

(Table 3). This was not reflected by a similar improvement of 6 MWT ($p = 0.811$).

The intramyocardial tissue recovery was found in responders with improvement in scar size RvNR −8.19 g, $p = 0.0238$ (Table 3). $CD133^+$ treated NR also displayed reduction in scar size ($CD133^+$ NR Δscar size 180 d/0: −13·9 g, SD ±20·9 g placebo NR +11·9, SD ±16·7 g, $p = 0.008$, $n = 20$) and non viable tissue (Δnon viable tissue 180 d/0: $CD133^+$ NR -12.4 g, SD ±19.3 g vs. placebo NR +11.5 g, SD ±12.0 g, $p = 0.004$, $n = 19$) (data not presented). This tendency was not observed in responders: scar size ($CD133^+$ NR vs. placebo NR -1.9, SD ±16.0 g vs. placebo +2.5, SD ±13.2 g, $p = 0.398$, $n = 33$) and non viable tissue ($CD133^+$ NR vs. placebo NR −1.4, SD ±16.7 g vs. placebo +1.8, SD ±12.3 g, $p = 0.544$, $n = 32$). Improvement (Δ) of segmental myocardial perfusion MRI at 180 days versus vs. baseline was observed for R ($p = 0.004$), but not in NR group ($p = 0.101$) (Supplement Table 1). Improvement (Δ180 d/0) of hypoperfused LV-segments under adenosine stress induction was present in R group in injected segments ($p = 0.009$) as well as in non-injected segments ($p = 0.017$), whereas in NR only injected segments were improved (injected segments $p = 0.034$; non-injected segment $p = 0.383$) (Supplement Table 1). Long term survival: Actuarial computed mean survival time was 76.9 ± 3.32 months (R) vs. +72.3 ± 5.0 months (NR), HR 0.3 [CI 0.07–1.2]; $p = 0.067$ (Fig. 3b).

3.4. Peripheral Blood Biomarker Profile

Circulating EPC ($CD133^+/CD34^+/CD117^+$) in peripheral blood were found to be reduced by a factor of two in NR versus R before treatment. For $CD34^+$ MNC subpopulations preoperative blood levels were (R): $CD34^+$ 0.072%, SD ± 0.05% vs. (NR) 0.039%, SD ±0.017, RvsNR $p = 0.027$. Similar difference was found preoperatively for $CD133^+$ and $CD133^+CD117^+$ subpopulations (Table 4). This difference was not found for the comparison of placebo and $CD133^+$ (Table 4). In contrast, $CD146^+$ CEC showed higher preoperative levels in non-responders versus responders ($p = 0.053$) (Table 4).

Postoperatively, reduction of EPC in NR remained significant until discharge: peripheral blood $CD34^+$ (NR vs. R $p = 0.026$ preop and day 10) and $CD133^+$ $CD117^+$ (NR vs. R $p = 0.024$ preop and day 10) despite postoperative increased levels of EPO (NR: preop. 16.9 U/ml, SD ±14.1 U/ml; NR day 10: 42.1 U/ml, SD ±23.9 U/ml; $p = 0.006$ preop/day 10) and reduction of IP10/CXCL10 (NR preop: 157.6 pg/ml, SD ± 94.5 pg/ml; NRday 10: 95.8 pg/ml, SD ±85.2 pg/ml; $p = 0.01$ preop/day 10).

Treatment responders were characterized preoperatively by lower serum levels of pro-angiogenic factors such as VEGF ($p = 0.056$ R/NR), EPO ($p = 0.023$ R/NR), CXCL10/IP10 ($p = 0.076$ R/NR), higher levels of IGFBP-3 ($p = 0.089$ R/NR) (Table 4), as well as strong induction of

VEGF (+26.6 pg/ml, $p = 0.015$ preop/day 10) at day 10 after intervention versus non-responders (+1.2 pg/ml, $p = 0.913$ preop/day 10) (Table 4). The CFU-EC capacity of purified CD 133+ bone marrow cells was positive in all tested patients without difference between responders ($n = 13$; mean 63.133 ± SD 13.6) and non-responders ($n = 9$; mean 77,833 ± SD 15.81) $p = 0.177$. Matrigel plug assay in vivo was positive in responders and non-responders (Supplement Table 7).

Thrombocyte counts were preoperatively reduced in NR (208 × 10^9/L, SD ± 51.2 109/L [CI 73–311], $n = 23$) versus R (257 × 109/L, SD ± 81.5109/L [CI 123–620] $n = 35$) (NR vs R: $p = 0.004$, $n = 58$) before treatment. Suspecting bone marrow stem cell suppression by finding reduced PB thrombocyte and $CD133^+$ $CD34^+$ EPC count, we tested RT-PCR gene expression analysis of SH2B3 mRNA coding for the lnk adaptor protein SH2B3 which is associated with inhibition of hematopoietic stem cell response for EPC and megakaryocytes in immediately frozen blood samples. First analysis in 21 patients revealed a tendency of increased mRNA expression in peripheral blood with non-responders ($p = 0{\cdot}073$) (Fig. 4, Table 4).

To identify a diagnostic response signature for R/NR we used machine learning methods as a tool for the prediction of functional improvement after CD133+ BMDC therapy and CABG surgery. First analyses were performed to particularly exclude overfitting in small populations. Then, blinded patient data from the PERFECT clinical database (Table 1) was investigated by t-SNE unsupervised ML, which is able to cluster similar patients in close proximity and reveals distinct groups (Fig. 5). Investigating the underlying segmentation, the firstline supervised ML analysis was made for all time points to place patient characteristics into two distinct groups (Fig. 5). The calculation independently assigned patient characteristics according to ΔLVEF at 180 days confirming the preselection criteria of >5% (Table 5). Then we used machine learning algorithms to investigate the decisive parameters to a response signature. For this the underlying PERFECT clinical dataset and biomarker laboratory measurements (Table 1, Appendix 3) were combined and analysed to validate classification specificity of parameter profiles for responders and non-responders before and after the CABG procedure. In particular, we used discriminative primary and secondary endpoint parameters as well as thrombocyte and leukocyte counts. Using only the clinical parameters ($n = 160$) classification resulted in a specificity of responders assuming mean accuracy of 63.35% (180 days) (Table 5). Combination of preoperative clinical data ($n = 49$) and biomarker laboratory parameters ($n = 142$), however, revealed higher sensitivity of angiogenesis/EPC/CEC related parameters in peripheral blood already preoperative with respective assuming max. Accuracy of 81.64% ± SE 0.51% [CI 80.65–82.65] ($n = 31$) (Table 5). Interestingly, 17/20 relevant parameters were related to angiogenesis parameters, bone marrow EPC/CEC responses, NT-proBNP, and SH2B3 gene expression in peripheral blood (Table 5). Using both clinical and biomarker parameters preoperative prediction accuracy for responders was 79.35% ± SE 0.24% [CI 78.87–79.84] ($n = 31$) and for non-responders 83.95% ± SE 0.93% [CI 82.10–85.80] ($n = 31$). Postoperative evaluation at day 10 ($n = 382$) revealed a prediction accuracy of 82.12% ± SE 0.28% [CI 81.56–82.67] ($n = 31$) (R) and 85.89% ± SE 0.67% [CI 84.56–87.22] ($n = 31$) (NR) (Fig. 5b), while day 0–180 combined clinical and biomarker analysis ($n = 522$) allowed a prediction accuracy of 94.77% ± SE 0.43% [CI 93.92–95.63] ($n = 31$) (R) and 92.44% ± SE 0.60% [CI 91.24–93.64] ($n = 31$) (NR) (Fig. 5).

4. Discussion

4.1. Baseline Characteristics of Treatment Responders vs. Non-responders

Induction of cardiac repair in patients with heart failure after myocardial infarction and ischemic cardiomyopathy has been targeted using numerous approaches including cardiac stem cell therapy (Fisher et al., 2016). However, the lack of efficacy and the lack of

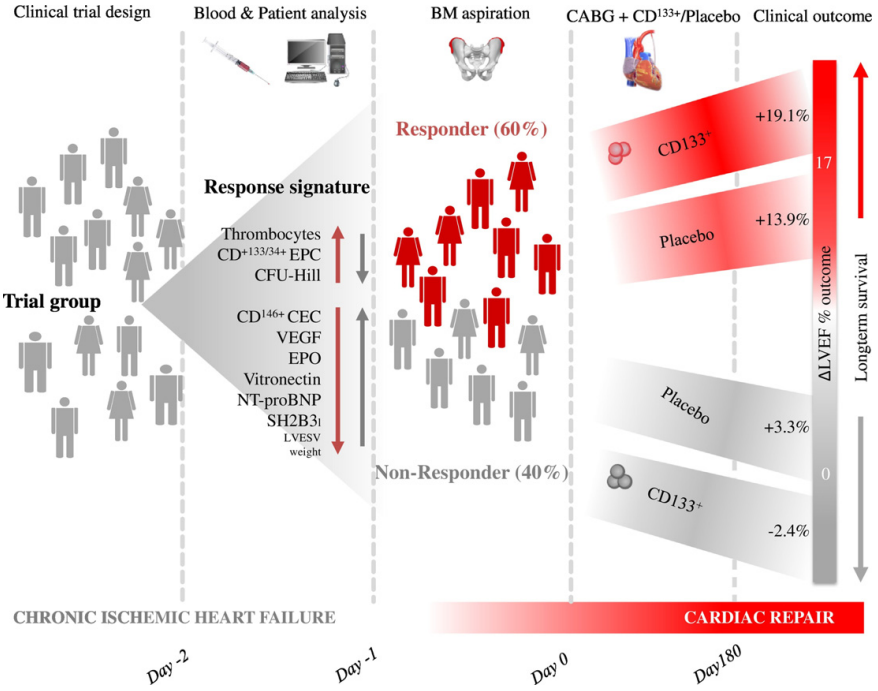


Fig. 6. Outcome results of the PERFECT trial.

response predictability have been the main obstacles for treatment standardization and success (Tian et al., 2014). Our approach employing CD133$^+$ autologous bone marrow derived cells intramyocardially in conjunction with CABG revascularization was promising in previous Phase I and Phase IIa trials and led to the multicentric placebo controlled phase III PERFECT trial investigation, which again confirmed the induction of cardiac repair (Stamm et al., 2003; Tse et al., 2003; Stamm et al., 2007; Nasseri et al., 2014). In the meantime, however, similar trials involving placebo-treated controls undergoing bone marrow harvest showed almost the same improvement of LVEF in the CD133 + group as in the placebo group (Nasseri et al., 2014; Bartunek et al., 2016). Similarly, the Chart-1 trial demonstrated a relevant functional recovery in only 60% of the patients, whereas 40% of both cell-treated and placebo-treated patients were non-responsive for unkown reasons (Bartunek et al., 2017). This significant non-responder rate was recently corroborated in CABG surgery for patients with reduced pump function (Vakil et al., 2016). In the clinical setting of the PERFECT trial, a nearly identical percentage of patients were non-responders to induction of cardiac repair, irrespective of their treatment with placebo or CD133$^+$ cells.

The underlying mechanism for a lack of response to induction of cardiac repair may be a failure of vascular repair by reduced circulating EPC. This mechanism was shown already 12 years ago to be associated with progression in atherosclerosis and coronary artery disease (Werner et al., 2005). Recently, the investigation of responders to cardiac repair in the CCTRN-trials obtained similar findings in bone marrow of BMDC treated non-responsive chronic ischemic heart failure patients (Bhatnagar et al., 2016; Contreras et al., 2017). In the PERFECT trial we found a striking difference in cardiac recovery between responders and non-responders. This was found for the first time to be associated with a specific signature composition of angiogenesis related biomarkers in peripheral blood. This was accompagnied by improved microvascular perfusion in the myocardium. Non-responsive patients did not exhibit any change in deteriorated left ventricular pump function both in placebo and CD133$^+$ groups. Only a minor effect on scar size and non-viable tissue repair was found in intramyocardial treated CD133$^+$ NR. In addition to numerous local tissue processes that have been shown to influence myocardial repair, such as fibrosis, inflammation, apoptosis, and potential endogenous cardiac stem cell niches, our data support the notion that blood and bone marrow components regeneration also play a key role.

4.2. Mechanism of Action for Cardiac Repair and Diagnostic Access

The typical blood components in non-responders are lowered CD133$^+$ CD34$^+$ CD117$^+$ EPC and thrombocytes counts in the peripheral blood and elevated angiogenesis stimulating factors as VEGF and EPO. In contrast, responders display basically elevated EPC and thrombocytes also in the absence of angiogenesis stimulating factors. We propose that the mechanism of impaired angiogenesis is caused by a dysfunctional bone marrow response. Potential mechanisms of impaired angiogenesis response may be either the anti-angiogenic interference of inflammatory cytokines, such as IP10, or NT-proBNP that may influence EPC proliferation or release mechanisms (Strieter et al., 1995; Stamm et al., 2003; Cesari et al., 2008). In this context, the first description of upregulated SH2B3 gene expression enhancement in the peripheral blood of non-responders associated with reduced EPC and thrombocyte counts suggests a potential regulatory role of SH2B3 with respect to suppression of the bone marrow response (Cesari et al., 2008; Kwon et al., 2009; Lee et al., 2016). Experimental models have depicted the potential importance and diagnostic or therapeutic relevance of SH2B3 gene expression and lnk adaptor protein SH2B3 for regulation of bone marrow responses and impairment of angiogenesic capacity (Ishige-Wada et al., 2016; Takizawa et al., 2008). Moreover, associations with hematological traits, coronary artery disease, and arteriosclerosis have been found for point mutations of SH2B3 promotor regions as well as influence of SH2B3 SNP on human longevity (Auer et al., 2014; McPherson and Tybjaerg-Hansen, 2016; Fortney et al., 2015). However, further clinical evaluation of SH2B3 expression is needed to unravel the precise mechanism in humans.

Feature selection based on our machine learning approach led to the identification of decisive factors for lack of response and the induction of cardiac repair, which can be used for diagnostic R/NR selection before and monitoring of during treatment. The core factors for laboratory diagnosis in peripheral blood were NT-proBNP, VEGF, Erythropoeitin, vitronectin, circulating EPC/CEC/Thrombocytes, SH2B3 mRNA expression, the CFU-Hill assay/Matrigel plug for peripheral blood, as well as weight and LVESV index. We found a statistical correlation of the identified factors and calculated their diagnostic use for the selection of responder and non-responder patients using repeated cross-validation (Fig. 6).

4.3. Relevance of LVEF Endpoint for Longterm Survival

The current analysis of longterm survival benefit in patients with induction of LVEF recovery after CABG/CD133$^+$ treatment suggests a clinical conversion of progressive heart failure by restitution of ventricular function. Moreover, considering the proposed underlying mechanism of impaired angiogenesis and vascular repair capacity of bone marrow, cardiac functional restitution may be dependent on bone marrow function. The current example of peripheral blood analysis focusing on angiogenesis factors and bone marrow derived cell subpopulations allows the definition of signature constellations defining normal or pathological stimulation/response patterns. The machine learning tool independently confirmed the response state as well as the angiogenesis factors involved in deficient response.

Long term deficit in vascular repair may result in progressive heart failure. CABG surgery can be considered as a potent intervention for the induction of cardiac repair most likely stimulated by bone marrow harvest prior to surgery as a preconditioning signal in responders (Blatt et al., 2016). Of utmost importance, however, is the further analysis of factors downregulating blood repair mechanisms in non-responders.

5. Conclusion

The PERFECT trial shows that cardiac tissue repair and restitution of left ventricular function can be successfully installed in ischemic heart disease by CABG surgery associated with presence of enhanced peripheral circulating CD133 + EPC level. In addition, dysfunctional left ventricular post-infarct tissue may be recruited by the local injection of purified CD133 + BMDC. The induction of cardiac repair, however, is correlated to CD133 + EPC release from bone marrow. Resistence of HSC/EPC to growth factor induction may be caused by elevated SH2B3 gene expression in non-responders. The diagnostic sensitivity of the responder vs. non-responder signature may be useful for diagnosis of deficient repair capacity in cardiovascular disease and for the preselection of patients for inductive stem cell therapy.

Limitations of the Study

Main limitations of the study are: 1. Preterm closure of recruitment resulting in limited patient number for efficacy analysis. 2. Non-significant CD133$^+$ effect on primary endpoint despite positive intermediate analysis. 3. Unknown mechanism of treatment unresponsiveness interfering with treatment intervention. 4. Need for further clinical evaluation of suspected blood/bone marrow suppression by SH2B3/lnk activator. 5. Predictive value of response signature in larger patient populations.

Declaration of Interests

All authors declare no competing interests.

Funding

German Ministry of Research and Education (BMBF): FKZ0312138A, EU ESF/IV-WM-B34-0011/08, ESF/IV-WM-B34-0030/10 and Miltenyi Biotec GmbH, Bergisch-Gladbach, Germany.

Acknowledgements

This study was funded by the German Ministry of Research and Education, Berlin, Germany, and Miltenyi Biotec GmbH, Bergisch-Gladbach, Germany. We would like to thank the PERFECT study group for their dedicated performance and support of the trial. We thank Guilio Pompilio (Milano, Italy), Francesco Siclari (Zuerich, Switzerland), Warren Sherman (New York City, USA), and Johannes Waltenberger (Münster, Germany) who served as members of the Data and Safety Monitoring Board. We thank the medical writers Dr. Claudia Frumento for the careful preparation of data in the CTR and James Hewlett for careful correction of the manuscript. The statisticians Uta Mehdorn and Horst Lorenz for careful control of data analysis.

Contributors

G Steinhoff contributed to study design, trial organization, medical controlling, enrolment and clinical follow-up of patients, research plan, analysis of clinical data, analysis of research data, data collection, data control, and drafted the manuscript.

A Haverich, S Sarikouch, FW Mohr, J Garbade, C Stamm, J Börgermann, F M Wagner, A Kaminski contributed to enrolment and clinical follow-up of patients, data collection and interpretation.
G Tiedemann contributed to study design, management and GxP control of the trial.
J Lotz analysed MRI data.
G Kundt performed statistical analyses.
J Nesteruk contributed to laboratory study design, follow-up analysis of patients, data collection, and statistical analyses.
P Oostendorp examined and interpreted data for final analysis.
P Mueller, J Große, A Skorska, U Ruch, R David performed laboratory analysis and examined data collection.
M Wolfien, H Hennig and O Wolkenhauer applied and investigated machine learning data analysis.

All authors contributed to final data interpretation, critically revised the manuscript, and approved the final version for submission.

PERFECT Study Group: Jana Große, Ulrike Ruch, Alexander Kaminski, Christian Klopsch, Peter Donndorf, Julia Nesteruk, Anna Skorska, Paula Müller, Robert David, Ralf Gaebel, Cornelia Lux, Peter Mark, Andreas Martens, Axel Haverich, Andreas-Matthaeus Bader, Michael Dandel, Christoph Knosalla, Elke Wenzel, T. Deuse, Anne-K. Funkat, Florian Wagner, Hermann Reichenspurner, Dirk Balshüsemann, Jens Brickwedel, Bernd Schröder, Dagmar Hartung, Günther Kundt, Heike Kurzidim, Heike Windhagen, Ilona Maeding, Ina Wagner, IPiroze Minoo Davierwala, J. Börgermann, Jan Kormann, Jan Martin Sohns, Jens Garbade, Joachim Lotz, Katharina Gawenda, Katharina-Wiebke Felke, Katja Schönefeld, Liane Preußner, Mandy Ludwig, Marcel Vollroth, Martin Fasshauer, Melanie Wittenberg-Marangione, Michiel Morshuis, Monika Teichmann, Murat Aktas, Nicole Schütz, Nicole Sprathof, Pascal Dohmen, Friedrich-Wilhelm Mohr, Roland Hetzer, Christof Stamm, S. Helms, S. Huebler, Samir Sarikouch, Sandra Bubritzki, Sebastian Rojas, Silke Holtkamp, Tanja Otto, Tatjana Alberg, Ulrike Hess, Ingo Kutschka, and Gustav Steinhoff.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ebiom.2017.07.022.

References

Auer, P.L., Teumer, A., Schick, U., et al., 2014. Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits. Nat. GenetNat. Genet. 46 (6):629–634. http://dx.doi.org/10.1038/ng.2962 (Epub 2014 Apr 28).

Bartunek, J., Terzic, A., Davison, B.A., et al., 2016. Cardiopoietic cell therapy for advanced ischemic heart failure: results at 39 weeks of the prospective, randomized, double blind, sham-controlled CHART-1 clinical trial. Eur. Heart J. Dec 23. pii: ehw543. http://dx.doi.org/10.1093/eurheartj/ehw543 (Epub ahead of print).

Bartunek, J., Terzic, A., Davison, B.A., et al., 2017. Cardiopoietic cell therapy for advanced ischaemic heart failure: results at 39 weeks of the prospective, randomized, double blind, sham-controlled CHART-1 clinical trial. Eur. Heart J. 38 (9):648–660. http://dx.doi.org/10.1093/eurheartj/ehw543.

Bhatnagar, A., Bolli, R., Johnstone, B.H., et al., 2016. Cardiovascular cell therapy research network (CCTRN). Bone marrow cell characteristics associated with patient profile and cardiac performance outcomes in the LateTIME-cardiovascular cell therapy research network (CCTRN) trial. Am. Heart J. 179, 142–150.

Blatt, A., Elbaz-Greener, G.A., Tuby, H., Maltz, L., et al., 2016. Low-level laser therapy to the bone marrow reduces scarring and improves heart function post-acute myocardial infarction in the pig. Photomed. Laser Surg. 34 (11), 516–524.

Cesari, F., Caporale, R., Marcucci, R., et al., 2008. NT-proBNP and the anti-inflammatory cytokines are correlated with endothelial progenitor cells' response to cardiac surgery. Atherosclerosis 199 (1), 138–146.

Contreras, A., Orozco, A.F., Resende, M., et al., 2017. Identification of cardiovascular risk factors associated with bone marrow cell subsets in patients with STEMI: a biorepository evaluation from the CCTRN TIME and LateTIME clinical trials. Basic Res. Cardiol. 112 (1), 3.

Donndorf, P., Kaminski, A., Tiedemann, G., Kundt, G., Steinhoff, G., 2012. Validating intramyocardial bone marrow stem cell therapy in combination with coronary artery bypass grafting, the PERFECT phase III randomized multicenter trial: study protocol for a randomized controlled trial. Trials 13:99. http://dx.doi.org/10.1186/1745-6215-13-99.

Fisher, S.A., Mathur, A., Taggart, D.P., Martin-Rendon, E., 2016. Stem cell therapy for chronic ischaemic heart disease and congestive heart failure. Cochrane Database Syst. Rev. 12, CD007888. http://dx.doi.org/10.1002/14651858.CD007888.pub3.

Forman, G., Cohen, I., 2004. Learning From Little: Comparison of Classifiers Given Little Training. http://dx.doi.org/10.1007/978-3-540-30116-5_17.

Fortney, K., Dobriban, E., Garagnani, P., et al., 2015. Genome-wide scan informed by age-related disease identifies loci for exceptional human longevity. PLoS Genet. 11 (12), e1005728.

Henry, Timothy D., Moyé, Lem, Traverse, Jay H., 2016. Consistently inconsistent—bone marrow mononuclear stem cell therapy following acute myocardial infarction - a decade later. Circ. Res. 119, 404–406.

Hofmann, W.K., de Vos, S., Elashoff, D., et al., 2002. Relation between resistance of Philadelphia-chromosome-positive acute lymphoblastic leukaemia to the tyrosine kinase inhibitor STI571 and gene-expression profiles: a gene-expression study. Lancet 359 (9305), 481–486.

Ishige-Wada, M., Kwon, S.M., Eguchi, M., et al., 2016. Jagged-1 signaling in the bone marrow microenvironment promotes endothelial progenitor cell expansion and commitment of CD133+ human cord blood cells for postnatal Vasculogenesis. PLoS One 11 (11):e0166660. http://dx.doi.org/10.1371/journal.pone.0166660.

Kuhn, M., 2008. Building Predictive Models in R using the caret package. J. Stat. Softw. 28 (5):1–26. http://dx.doi.org/10.18637/jss.v028.i05.

Kwon, S.M., Suzuki, T., Kawamoto, A., Ii, M., Eguchi, M., Akimaru, H., Wada, M., Matsumoto, T., Masuda, H., Nakagawa, Y., Nishimura, H., Kawai, K., Takaki, S., Asahara, T., 2009. Pivotal role of Ink adaptor protein in endothelial progenitor cell biology for vascular regeneration. Circ. Res. 104 (8), 969–977.

Lee, Jun Hee, Ji, Seung Taek, Kim, Jaeho, et al., 2016. Specific disruption of Lnk in murine endothelial progenitor cells promotes dermal wound healing via enhanced vasculogenesis, activation of myofibroblasts, and suppression of inflammatory cell recruitment. Stem Cell Res Ther 7 (1), 158.

Maaten, L.V.D., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9: 2579–2605. http://jmlr.org/papers/v9/vandermaaten08a.html.

McPherson, R., Tybjaerg-Hansen, A., 2016. Genetics of coronary artery disease. Circ. Res. 118 (4), 564–578 Feb 19.

Nasseri, B.A., Ebell, W., Dandel, M., et al., 2014. Autologous CD133+ bone marrow cells and bypass grafting for regeneration of ischaemic myocardium: the Cardio133 trial. Eur. Heart J. 35 (19), 1263–1274.

Signified power analysis for *t*-tests through multivariate hypotheses. In: O'Brien, R.G., Muller, K.E., Edward, L.K. (Eds.), Applied Analysis of Variance in Behavioral Science. Marcel Dekker, New York.

Rosenberger, William F., Lachin, John M., 2003. Randomization in Clinical Trials: Theory and Practice. Wiley publ, pp. 1–14.

Saeb, A.T., Al-Naqeb, D., 2016. The impact of evolutionary driving forces on human complex diseases: a population genetics approach. Scientifica (Cairo) 2016, 2079704. http://dx.doi.org/10.1155/2016/2079704 (Epub 2016 May 30).

Stamm, C., Westphal, B., Kleine, H.D., et al., 2003. Autologous bone-marrow stem-cell transplantation for myocardial regeneration. Lancet 361 (9351), 45–46.

Stamm, C., Kleine, H.D., Choi, Y.H., et al., 2007. Intramyocardial delivery of CD133+ bone marrow cells and coronary artery bypass grafting for chronic ischemic heart disease: safety and efficacy studies. J. Thorac. Cardiovasc. Surg. 133 (3), 717–725.
Strieter, R.M., Kunkel, S.L., Arenberg, D.A., Burdick, M.D., Polverini, P.J., 1995. Interferon gamma-inducible protein 10 (IP-10), a member of the C-X-C chemokine family, is an inhibitor of angiogenesis. Biochem. Biophys. Res. Commun. 210 (1), 51–57.
Takizawa, H., Eto, K., Yoshikawa, A., Nakauchi, H., Takatsu, K., Takaki, S., 2008 Jul. Growth and maturation of megakaryocytes is regulated by Lnk/Sh2b3 adaptor protein through crosstalk between cytokine- and integrin-mediated signals. Exp. Hematol. 36 (7), 897–906.
Taylor, D.A., Perin, E.C., Willerson, J.T., et al., 2016. Cardiovascular cell therapy research network (CCTRN). Identification of bone marrow cell subpopulations associated with improved functional outcomes in patients with chronic left ventricular dysfunction: an embedded cohort evaluation of the FOCUS-CCTRN trial. Cell Transplant. 25 (9), 1675–1687.
Tian, T., Chen, B., Xiao, Y., Yang, K., Zhou, X., Zhou, X., 2014. Intramyocardial autologous bone marrow cell transplantation for ischemic heart disease: a systematic review and meta-analysis of randomized controlled trials. Atherosclerosis 233 (2), 485–492.
Tse, H.F., Kwong, Y.L., Chan, J.K., Lo, G., Ho, C.L., Lau, C.P., 2003. Angiogenesis in ischaemic myocardium by intramyocardial autologous bone marrow mononuclear cell implantation. Lancet 361 (9351), 47–49.
Vakil, K., Florea, V., Koene, R., et al., 2016. Effect of coronary artery bypass grafting on left ventricular ejection fraction in men eligible for implantable cardioverter-defibrillator. Am. J. Cardiol. 117, 957–960.
Werner, N., Kosiol, S., Schiegl, T., et al., 2005. Circulating endothelial progenitor cells and cardiovascular outcomes. N. Engl. J. Med. 353 (10), 999–1007.

Nomenclature

AE: Adverse Event
AESI: Adverse Event of Special Interest
AHA: American Heart Association
ANCOVA: Analysis of covariance
BM: Bone marrow
BMSC: Bone marrow stem cells
BMDC: Bone marrow derived cells
CABG: Coronary Artery Bypass Graft
CAP-EPC: Concentrated Ambient Particels – Endothelial Progenitor Cells
CBA: Cytometric Bead Array
CCS: Canadian Cardiovascular Society
CCTRN: Cardiovascular Cell Therapy Research Network
CD: Cluster of Differentiation
CEC: Circulating endothelial cells, CEC panel, CDs measured in PB
CFU: Colony-forming unit
CI: Confidence interval
CMV: Cytomegalovirus
EA: Early Antigen
EBNA1: EBV-Nuclear Antigen 1
EBV: Epstein-Barr-Virus
EC: Endothelial Cells
ECG: Echocardiography
ELISA: Enzyme-Linked Immunosorbend Assay
EPC: Endothelial Progenitor Cells, EPC panel, CDs measured in PB
EPO: Erythropoietin
GMP: Good Manufacturing Practice
HR: Hazard ratio
HIF: Hypoxia-Inducible Factor, transcription factor
ICH GCP: Tripartite Guidelines Guideline for Good Clinical Practice
IGF-1: Insulin-like Growth Factor 1
IGFBP2/3: Insulin-like Growth Factor-Binding Protein 2/3
IHG: Analysis performed in accordance with ISHAGE guidelines
IL: Interleukin
IP-10: Interferon Gamma-induced Protein 10 also known as C-X-C motif chemokine 10 (CXCL10)
LMCA: Left Main Coronary Artery
LVEDV: Left Ventricular End Diastolic Volume
LVEF: Left Ventricular Ejection Fraction
LVESD: Left Ventricular End Systolic Dimension
MACE: Major Adverse Cardiovascular Events
ML: Machine learning
MNC: Mononuclear cells
MRI: Magentic Resonance Imaging
6MWT: 6-Minute Walk Test
NT-proBNP: B-type Brain Natriuretic Peptide
PB: Peripheral blood
PBMNC: mononuclear cells isolated from peripheral blood
PCI: Percutaneous Coronary Intervention
PEI: Paul-Ehrlich Institute
PPS: Group of patients for per-protocol set
SAE: Serious adverse event
SAS: Group of patients for safety set
SDF-1: Stromal Cell-derived Factor 1
SH2B3: Lnk [Src homology 2-B3 (SH2B3)] belongs to a family of SH2-containing proteins with important adaptor functions
SCF: Stem Cell Factor
STEMI: ST- segment Elevation Infarction
SUSAR: Suspected Unexpected Serious Adverse Reaction
TNF: Tumor Necrosis Factor
t-SNE: t-distributed neighbor embedding
VCA: Virus-Capsid-Antigen
VEGF: Vascular Endothelial Growth Factor
VEGF rec: Vascular Endothelial Growth Factor Receptor
VEGFR2/KDR: Vascular Endothelial Growth Factor Receptor 2/Kinase Insert Domain Receptor

3 Conclusion and outlook for customized workflow development in systems medicine

This section enhances the individual discussions of the manuscripts already presented. It summarizes the main aspects and evaluates the initial hypotheses about developing workflows in systems medicine and appraises the results obtained for the cardiac research field. The intensely debated topic of the social and ethical considerations of human sequence analysis in patients and AI analysis procedures in general is also addressed. Furthermore, a brief outlook for prospective work based on this thesis is provided, as is a conclusion about the overall computational and biological impact.

3.1 NGS and network analyses in preclinical and clinical research

This section recapitulates the contributions made towards supporting medical translational research by utilizing state-of-the-art computational methods and systems medicine approaches. This has been exemplified in numerous pre-clinical and clinical studies about cardiac regeneration. Here, the results obtained are compared to other related findings, and the validity of the initial hypotheses is examined.

At the beginning of the thesis, hypotheses about sequencing and network analyses were postulated, and here, these hypotheses are evaluated: i) workflow development facilitates the reusability of data analysis procedures in sequencing analyses; ii) Galaxy is a sustainable analysis framework for genomic and transcriptomic investigations in biomedical research; iii) single-nuclei RNA-Seq analyses can uncover in-depth information about cell type compositions and RNA kinetics in adult mammalian hearts; iv) signaling network analysis can support the evaluation of reprogrammed cardiac subtypes; v) co-expression analyses of RNA-Seq data can validate hub-genes responsible for heart rate influence; vi) the potential of cell therapies for cardiac regeneration can be investigated by means of preclinical studies; and vii) patient stratification for stem cell therapy after myocardial infarction is possible through an integrative dataset from human peripheral blood samples.

i) RNA-Seq workflow development

Tools that are no longer maintained or were not designed to address evolving RNA-Seq protocols and the rapidly increasing amount of available sequence data from first- (Sanger), second- (454, Solexa, Illumina), and third-generation sequencing approaches (IonTorrent, SOLiD, Nanopore, PacBio) become outdated over time (Lott et al., 2017). In addition, data analysis tools that are continuously maintained may change their behavior and parameters due to tool version changes. Therefore, the reuse of previously generated workflows is often not simple or even possible if the workflow was developed outside of a data analysis framework or computational container. Another major challenge is the comparison, benchmarking, selection, and integration of available tools (in their current version), which is time-consuming and requires computational domain expertise. Depending on the number of samples, the scale of time series, and sequencing depth, computations may require heavy computational resources, such as cluster, grid, and cloud computing solutions. Thus, an adaptive management of available computing resources by load balancers and queuing systems is often invaluable in creating analysis workflows (Lachmann et al., 2020; Lott et al., 2017). These obstacles are major barriers for non-computational users to apply

advanced bioinformatics workflows, which is why the development of easily accessible and applicable data analysis frameworks is essential for proper research in life science and at a clinical level.

State-of-the-art RNA-Seq methods are used for preprocessing, genome mapping (Bray et al., 2016; Kim et al., 2013), and the identification of DE genes (Love et al., 2014; Seyednasrollah et al., 2013). In accordance with Nookaew *et al.* (2012), more than one DE analysis tool or algorithm is used to obtain significantly differentially expressed genes of the given datasets because it is not yet possible to name a universal workflow for all types of sequencing experiments, applications, and datasets. In particular, workflow development as a procedure should be a comparison of different genome mapping and DE analysis tools, which are all based on different algorithms. The results obtained for one comparison in this thesis are shown in Fig. 2.2 and confirm the results of Seyednasrollah *et al.* (2013), which state that Cufflinks2 is the most conservative and secure (with respect to false positive genes) DE analysis method when using TopHat2 or BWA as an alignment tool. Additionally, the analysis regarding the idea of data pre-processing from Lamm *et al.* (2011) is extended by integrating several modules to improve mapping accuracy and validating their benefits. The underlying workflow TRAPLINE supports Sandve *et al.*'s (2013) idea of reproducible computational research and was the first fully integrated and published workflow on the Galaxy web-based platform, including the mentioned modules in Fig. 2.1 (Section 2.1). The workflow is freely available[1] and ready to use as a Docker container, which were some of the limiting factors of previously proposed workflows, e.g., those of Robinson *et al.* (2010) and Zhao *et al.* (2014).

Nevertheless, the workflow also has limitations. First, there is a problem with discarding FPKMs at a low level (between zero and one), which has been thoroughly discussed in the literature but not yet entirely solved. Here, FPKM values were discarded until they reached an FPKM threshold of one because smaller values could be artefacts (Mortazavi et al., 2008). In contrast, the more advanced method of Jiang *et al.* (2011) proposes using housekeeping gene spike-ins as a control, and there is also the more recent method of Chen *et al.* (2014) of using Gene Ontology (GO) terms. These methods can be compared functionally, such as utilizing DNA/protein standards in molecular biology (e.g., in PCR or western blots); however, because of unreliable data about housekeeping genes, adopting these methods were not applicable in the datasets used in this thesis. In Section 2.1, the need for bias correction (Fig. 2.2) was assessed. An even more advanced and promising possibility of reducing biases may be realized via the method of Finotello *et al.* (2014),

[1] http://bit.ly/rnaseqwolfien14

who used a maxcount approach for counting the reads instead of using raw counts, the "*Fragments per kilobase of exon model per million mapped reads*" (FPKM), or "*Transcripts per Million*" (TPM). The last of these was preferred throughout this thesis because all current quantification tools, such as FeatureCounts and Kallisto, use it as a standard output.

As noted above, the half-life time of computational tools is limited, especially in a rapidly growing field such as NGS data analysis. Today, halfway through 2020, the TRAPLINE tools used are already the subject of criticism and the developers of TopHat2 and Cufflinks no longer suggest using their tools for a comprehensive analysis. There are already available successor tools such as STAR and Kallisto for mapping, or Sleuth and DESeq2 for differential expression analyses that are more sensitive, accurate, and efficient. However, the tools used here still produced high numbers of top-rated journal publications in the past; therefore, the results are still reliable. Another important aspect is the amount of computational resources needed to run a specific tool because, once downloaded and installed, the tools provided in the TRAPLINE Docker container can be used offline on a standard personal computer (4GB RAM, 50GB diskspace). TRAPLINE was also designed as a two-step analysis workflow, which means that, unless the integrated NGS data processing tools are already obsolete, the independent second part of data annotation and transcript characterization can still be used for miRNA-target prediction and protein-protein interaction assignment. The modularity of Galaxy also allows for an easy integration of other current tools or workflows that can be utilized as an initial data processing step. The output from such workflows can also be highly customized and may include quality reports, calculations and predictions for novel transcripts, probabilities of differentially expressed transcripts, and transcript characterizations including annotations, such as GO, KEGG, Panther, WikiPathways, DisGeNet, and Reactome, as well as corresponding visualizations (e.g., volcano plots, heatmaps, PCAs, networks, sashimi plots). All the tools available on the Galaxy Tool Shed can be installed along with their automatically resolved dependencies with a single click in the Galaxy interface.

Independent of the analysis method used, RNA-Seq data analysis is still a computational approach with many mathematical assumptions to investigate the transcriptome and thus it is necessary to verify the translated proteins encoded by the mRNAs because a translation to a functional protein is not certain for every mRNA (Cooper, 2000). In addition, there are two major sources of errors prior NGS data analysis (Nagalakshmi et al., 2008). On the one hand, there are base assembly errors that occur in cDNA library generation after the resynthesis step of the DNA polymerase (reverse transcriptase). On the other hand, the

influence of the experimentalist on this standardized enzymatic procedure is low (Kircher and Kelso, 2010).

The way workflow management frameworks and cloud computing services bridge the gap between tool developers and end users is discussed above; however, the use of single workflows for specific tasks (e.g., the RNA-Seq analysis workflow presented) can be facilitated by assembling multiple workflows into a single *connective workflow*. Such universal connective workflows can be understood as collaboration efforts within bioinformatics because every analysis requires a certain type of expertise. Using and connecting numerous workflows to apply multilayered approaches can incorporate several independent algorithms to test or benchmark different workflows (Fig. 3.1). This in turn facilitates the certainty of the knowledge obtained because independent algorithms, such as RNA-Seq analysis, WGCNA, and ML/DL, can be combined for highly specialized research questions. The anticipated strategy for such connective workflows could of course be realized with a set of Galaxy workflows for a higher reusability or via dockerized R-Studio tools and scripts for software package persistence (Boettiger, 2015; Boettiger, Carl and Eddelbuettel, 2017). In addition, an interoperable standard of workflows, namely the common workflow language, joins command-line tools across multiple platforms with workflows and, likewise, offers a modular concept of functional workflows that are built around containerized software solutions (Documentation available at CWL[2]). In order to adapt tools and workflows over time and ensure reusability and sustainability, it is also recommended to remain up to date with the changes in the tools themselves with a registration in tool management platforms such as OMICtools[3] or bio.tools[4] (Ison et al., 2019), in which tools are described using the meta-descriptive EDAM Ontology (Ison et al., 2013).

In summary, the first hypothesis can be confirmed because the overall functionality and applicability of the workflow was shown and validated through different applications. Unless some of the tools are obsolete, the workflow was internally adapted and continuously refined over time for the detection of significantly differential genes and further downstream analyses.

ii) Galaxy as an analysis framework

The overall positive effect of computational workflows was presented in the introduction (Section 1.2.1). Galaxy, as one prominent workflow management framework, was used to develop and disseminate the computational analysis strategy. The framework's applicability

[2] https://www.commonwl.org/
[3] https://omictools.com/
[4] https://bio.tools/

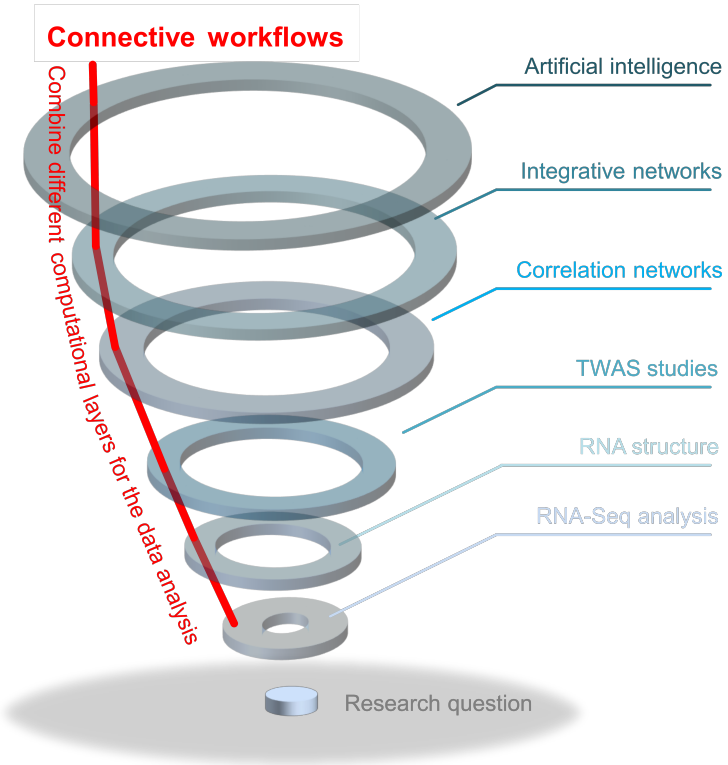


Figure 3.1: Impact of a connective workflow. Example of a connective workflow to solve a specific research question with different computational layers. The analysis starts with an RNA-Seq workflow and subsequently continues with the indicated advanced processes.

was also highlighted on smaller scale for single workflows (e.g., TRAPLINE; Wolfien et al., 2016), as well as for larger efforts such as the RNA workbench (Fallmann et al., 2019; Grüning et al., 2017). From the biological user perspective, Galaxy and similar data analysis frameworks offer easily accessible and applicable tools for complex computational investigations. Only a short period of time is needed to become familiar with the general functionalities, which can later be quickly adapted to other analysis approaches or Galaxy servers. However, more effort is needed for tool developers to embed and maintain their tools in Galaxy. A tool must be wrapped into an *.html*-compliant shell, e.g., through Planemo[5]. Often, the tools in Galaxy do not have the complete functionalities of their native unix counterparts. Once a tool is embedded within Galaxy or its Tool Shed, though, it can be combined in workflows with hundreds of different tools to increase its applicability (Bagnacani et al., 2019; Lott et al., 2017). Other bioinformaticians can also utilize this workflow modularity and can combine different tools into data analysis workflows and publish them for the Galaxy community, as was done with the TRAPLINE workflow (Section 2.1).

[5] https://planemo.readthedocs.io/en/latest/writing.html

The German Network for Bioinformatics Infrastructure (de.NBI) and the European Network ELIXIR support the establishment of a research oriented computing infrastructure and training scientists with respect to diverse bioinformatics questions. In particular, the workflow-centric de.STAIR project was inspired by experience with TRAPLINE; it also focuses on the needs of the experimental researchers for robust data analysis tools and develops tailor-made workflows for RNA-Seq experiments. After setting up a current data analysis workflow, a useful dissemination strategy is also key for an enhanced reuse (Grüning et al., 2017). For example, for the rapid dissemination of the RNA workbench and for an easy integration with other high-throughput sequencing analysis tasks, an implementation within the Galaxy framework was done. A major advantage of relying on Galaxy as the core framework is that it is possible to leverage its scalability, which enables the RNA workbench to run on single-CPU installations and on large, multi-node, high-performance computing environments. Furthermore, Galaxy provides researchers the means to reproduce their own workflow analyses, enabling them to rerun entire pipelines, or publish and share them with others. The RNA workbench is containerized, *i.e.*, administrators can deploy it via Docker. This makes it possible to have all the tool installation dependencies resolved in advance, while still keeping maintenance tasks to a minimum. The provided layer of virtualization also allows the handling of user-defined input data in a secure and compartmentalized way, a key requirement for researchers working on sensitive data (e.g., patient data in clinics). Running the containerized RNA workbench simply requires installing Docker and starting the Galaxy RNA workbench image. Furthermore, containerizing Galaxy enables a customized Galaxy instance with a selected subset of tools dedicated to specific data analysis tasks, while keeping deployment and installation simple.

Galaxy also provides the means to share interoperable workflow provenance because workflows and specific Docker containers for workflows can be commonly stored to fully reproduce the results of an analysis. The tool versions and dependencies used are preserved in the container. However, Galaxy is usually embedded as a whole within a Docker container, including a scheduler for monitoring and distributing Galaxy's work queue for "*WorkflowRequests*." Other data analysis frameworks, such as BioConda or Snakemake, are more flexible and lightweight in this way, while allowing single tools to be encapsulated in small computational containers (around 10 MB). In addition, the concept of a "*Common Workflow Language*" (CWL), which is an open standard for describing analytical workflows and tools in a way that makes them portable and scalable, uses the flexibility of single Docker containers that are combined into larger workflows (Amstutz et al., 2016). Galaxy only partially implements these concepts and mainly relies on monolithic containers.

However, these frameworks are in fact not easily applicable for non-computational experts and are not web-based, as Galaxy is.

In summary, the second hypotheses can upheld; without many alternative computational frameworks for data analysis, as a free and community-based platform, Galaxy enables a broad range of users to comprehensively analyze their data on their own with the help of extensive online training material.

iii) Single-nuclei RNA-Seq data to uncover cellular subpopulations

Single-nuclei technology was used rather than single-cell sequencing because the irregular cell shape and large size of the CM can disturb the overall cell capture (Wolfien et al., 2020a). Respecting the "*eleven grand challenges in single-cell data science*," (Lähnemann et al., 2020) the cellular composition of an adult mammalian heart was comprehensively shown for the first time. A transient state of CM, rather than a fixed turnover of cells, was shown; this was also supported by our differentiation trajectory reconstruction and cell fate probability quantification. We were the first to confirm a second differentiation lineage for CM on a single-nuclei level (Wolfien et al., 2020a,b). Further experimental evidence was given in previous studies, but here, we could clearly show a transient state of immature CM and endothelial cells to develop into mature CM (Wolfien et al., 2020a). The results also indicate that in-breed mice populations such as Bl6 might lack this particular cell type for as yet unknown reasons (Wolfien et al., 2020b). One reason might be the enhanced genetic robustness of the outbreed Fzt:DU mice strain or the enlarged heart size (Dietl et al., 2004).

Additional data integration of single-cell data of the *Tabula muris* project (Schaum et al., 2018) was achieved through a Seurat integration, and it enabled us to exclusively reveal proliferative CM (Galow et al., 2020), which have also not been detected before in a single-cell analysis of an entire heart. Previously, there was controversy regarding whether proliferative CM in adult hearts exist and where new CM originate. In future validation studies, the tissues investigated must be expanded to include more specific sub-regions, such as the SA node and the AV node (Goodyer et al., 2019). These two highly specialized regions of the heart are essential to understand; however, due to their small size, the quantity of captured cells is imited and requires a pooling of animals. Another aspect to consider is the analysis of time-series single-cell data to identify the turnover of proliferative CMs during embryogenesis (Asp et al., 2019).

The annotation of single-cell clusters is still a biased, manually curated effort. Only with domain expertise is it possible to identify specific subpopulations, which is why the research effort for an automated yet transparent computational solution continues. With the help of statistical classification, *i.e.*, ML and algorithms based on previously curated and annotated datasets, rare cell types can be easily and independently detected in single-cell data. First algorithms such as scCATCH (Shao et al., 2020), SCSA (Cao et al., 2020), or the further ML-based oversampling techniques of Bej *et al.* (2019) have already shown significant potential.

The gold standard, however, remains experimental validation with multi omics approaches to confirm novel subpopulations of cells (Grün, 2020). Nevertheless, it will quickly become challenging to experimentally identify and isolate very small amounts of specific cell types that are embedded and highly connected within larger cell clusters without noticeably different morphology. This might be one reason these cell subpopulations have not yet been investigated in detail.

In summary, we were able to identify numerous aspects of cellular compositions and could confirm previously known hypotheses on the single-cell level. We have further shown that the conclusions are also driven in large part by the different input datasets; thus, we suggested integrating more than a single dataset, if possible, from different mice strains, such as our Fzt:DU mice strain, for a more robust analysis.

iv) Network-based approaches for subpopulation characterization
During my PhD research, different cell populations have been characterized through network approaches: i) in Section 2.2.1, iSaBs were extensively evaluated towards antibiotic selected cardiac bodies (aCaBs), and ii) refer to the differentiation of adult mesenchymal stromal cells (MSC) into cardiomyocytes (Section 2.2.2).

The first comparison of undifferentiated pluripotent stem cells as control (GSES), Tbx3 transfected and antibiotic selected cells (Tbx3MHC), and antibiotic selected cells only (MHC) directly exhibits the differences between functional pacemaker cells and CM (Jung et al., 2014). Significantly differentially expressed genes were detected between the Tbx3MHC (pacemaker cells) *vs.* the GSES group (undifferentiated pluripotent stem cells). Transcript characterizations were obtained with GO (biological and molecular processes) and pathway enrichment analyses (Young et al., 2010). As expected, the resulting networks showed a greater enrichment and enhanced influence of mRNAs associated with cardiac processes in comparison to control cells because both transfections are known to

enhance cardiomyocyte development (Jung et al., 2014). The method was also adapted for another comparison of Tbx3MHC and MHC (cardiomyocytes). The results again showed a significant enrichment of identical GO terms that were seen previously. This can be considered proof of the positive effect of the Tbx3 and Myh6 transfection leading to beating, pacemaker-like cells. The interactions and GO terms between the mRNAs determined are illustrated in Fig. 2.33 (Figure 3 in Yavari *et al.* 2017), which provides an overview about the characteristics of significantly overexpressed mRNAs in the pacemaker cells.

The transcription factor **Tbx3** is permanently but slightly (not significantly differentially) expressed in all cell types. Bakker *et al.* (2012) and Wiese *et al.* (2009) have postulated that Tbx3 is a clear indicator of functionality in pacemaker-like cells with an SA function. Here, the data also shows a positive effect of Tbx3 overexpression in SA functionality. Tbx3 might be slightly expressed because it is highly regulated, especially through Tbx5 (van den Boogaard et al., 2012). **Tbx5** is a central molecule in the network and is associated with many connections to all relevant biologic processes. The **Myh6**-positive selected groups only partially overexpress the Myh6 mRNA (group of Tbx3MHC) in comparison to GSES cells. Myh6 is also associated as a key player in functional cardiac development (Granados-Riveron et al., 2010). Finally, **Tbx18**, a member of the same t-box transcription factor family, was reported to enable the reprogramming of chamber myocardium towards nodal cells. It is not significantly differentially expressed in the current analysis, which contradicts the results of Kapoor *et al.* (2013). The subsequent analysis of all RNA-Seq gene expression transcripts in the Tbx3MHC *vs.* MHC cell comparison via the Cytoscape application KeyPathwayminer also revealed different subnetworks related to contraction, electrophysiology, metabolism, and even differentiation factors; see Fig. 2.24 (Figure 4 in Hausburg *et al.* 2017).

The second comparison investigated the applicability of non-integrative methods, including microRNAs and mRNAs, for the cardiac reprogramming of adult MSC derived from bone marrow, dental follicle, and adipose tissue. Microarray experiments were used to confirm that MSCs derived from adipose tissue can partly be reprogrammed into the cardiac lineage by the transient overexpression of Gata4, Tbx5, Mef2c, and Mesp1, while cells isolated from bone marrow and dental follicle exhibit only weak reprogramming efficiency. Subsequently, three differentiation protocols were compared by pathway enrichment analysis and are highlighted in in Fig. 2.30 (Figure 6 in Müller *et al.* 2020). Together with previous studies of adult MSC overexpressing transcription factors, the results indicate the feasibility of mRNA-based cardiac reprogramming of MSCs.

In summary, the fourth hypothesis can be confirmed because both the enrichment and network analyses show a significant up-regulation of mRNAs with a cardiac impact and a significant enrichment of GO terms with relevance to the heart, especially in the double transgenic group of Tbx3MHC in comparison to the other groups. In addition, the second network analysis has contributed to the molecular characterization of CM differentiation.

v) The impact of ncRNAs and gene co-expression

The TRAPLINE workflow identified numerous differentially expressed mRNAs and ncRNAs, such as miRNAs, which are actually immature pri-miRNAs in this NGS experiment. To draw meaningful conclusions, the matured forms of the miRNAs have to be experimentally validated (e.g., via PCR). The target predictions are based on mathematical algorithms; therefore, they are only theoretical and are often also not experimentally validated interactions (Dweep et al., 2011). In addition, lncRNAs have received increasing attention in cardiac research. For example, the publication of Wang *et al.* (2014) shows the influences of lncRNAs in cardiac hypertrophy. In this thesis, lncRNAs were also identified, and these were previously annotated but not yet characterized in other tissues (Tanaka et al., 2000). **Snhg5** is significantly differentially expressed in Tbx3/Myh6 double transgenic cells only. Snhg5 was recently found to act as an miRNA sponge and, therefore, requires investigation and characterization in cardiac tissue, as well (Wang et al., 2018). Interestingly, Tbx5 binds to the promotor region of Snhg5, which supports the idea of a potential meaningful influence (Yang et al., 2013).

In the previously mentioned single-cell analysis, the expression of Hand2os1 (Anderson et al., 2016; Han et al., 2019) also a long non-coding RNA that orchestrates heart development by dampening Heart And Neural Crest Derivatives Expressed 2 (**Hand2**) expression, was identified as distinguishing immature CMs from fully differentiated populations (Section 2.1.6). In addtion to the mature, atrial, and ventricular CMs, another Hand2os1 high CM population has a 1.5-fold expression enrichment that apparently originates from the Hand2os1 low population. Based on the recent findings of de Soysa *et al.* (2019), who identified Hand2 as a specifier of outflow tract cells but not right ventricular cells during embryonal development, it can be assumed that the population identified represents cells of the outflow tract.

This thesis took the gene co-expression method into account as a GBA approach to cluster uncharacterized ncRNAs to already known mRNAs (Section 2.2.3). WGCNA was applied for this task and identified 2 AMPK as one central hub gene in iSaBs, along with ncRNAs such as miRNA1982 (Yavari et al., 2017). A review of Chakraborty *et*

al. (2020) has also identified 2 AMPK as essential for the expression and regulation of ion channels and ion transporters, including cytosolic Ca^{2+} handling proteins. AMPK activation is thought to be protective by preventing metabolic stress, favorably modulating membrane electrophysiology including cytosolic Ca^{2+} dynamics, preventing cellular growth, and hypertrophic remodeling (Chakraborty et al., 2020). 2 AMPK is also one of the clinical and ECG variables to predict the outcome of genetic testing in hypertrophic cardiomyopathy (Robyns et al., 2020).

The overall consistency of RNA-Seq data can be seen in the dendrograms illustrated in Fig. 2.34 (Figure 4 in Yavari *et al.* 2017), which clearly show similarities within the groups; therefore, they are clustered together by two different algorithms. Moreover, the differences between the groups are visualized with the help of the distance correlation matrix in Fig. 2.34. Both figures address the data reliability and, for this reason, the quantifiable influences on the cardiac cell differentiation of Tbx3MHC double transgenic clones in comparison to single transgenic MHC and GSES control cells. To further clarify the differences, the workflow that was developed and validated was applied to identify the significantly differentially expressed mRNAs. Here, mRNAs from RNA-Seq are favored because a comparison of NGS technologies with microarray and quantitative PCR approaches revealed that NGS data facilitates higher sensitivity and accuracy (Mortazavi et al., 2008). Moreover, the unspecific mRNA background signals of the transcriptome obtained by RNA-Seq are very low because the cDNA sequences can unambiguously be mapped to unique regions of the genome (Wang et al., 2009).

In summary, co-expression analyses have been successfully utilized to validate the hub gene 2 AMPK, which is relevant for the heart rate regulation. However, commonly associated ncRNAs have been identified via this GBA approach, but their specific interaction and role within the co-expression module has yet to be validated with further computational or experimental methods, as shown in Wolfien *et al.* (2019).

vi) The potential of cell therapies

Mice can be a valid model for cell therapies because, due to our meta-analysis, we showed that we can achieve comparable improvements as previously indicated in pigs and humans (Lang et al., 2017). Wang *et al.* (2019) have confirmed this in a patient-specific meta-analysis that included 14 randomized clinical trials and a total of 669 participants.

As other studies have done, we identified the gender as well as the cell origin as a source of important variety (Jansen of Lorkeers et al., 2015). Our meta-analysis shows that the

use of allogeneic cells can lead to significantly greater left ventricular ejection fraction (LVEF) improvement than syngeneic cells. Even though syngeneic cell transplantation into in-breed mice strains is not directly comparable to autologous cell transplantation, our results support the hypothesis that allogeneic cell applications can indeed provide an attractive alternative to autologous cell-based therapies. The gender-specific findings are in agreement with a meta-analysis of the influence of patient characteristics on study results by meta-regression (Bai et al., 2010). In that study, male individuals benefitted less than females from intracoronary-infused bone marrow stem cells for the treatment of acute myocardial infarction. Interestingly, the relevance of gender-specific approaches in the field of cardiovascular medicine has increased in recent years (Cadeddu Dessalvi et al., 2019; Romiti et al., 2019). Both clinical and preclinical studies indicate that the female sex favorably influences the remodeling and adaptive response to myocardial infarction (Fels and Manfredi, 2019; Ostadal and Ostadal, 2014).

Since the current study was limited to magnet resonance imaging (MRI) to measure LVEF, a further meta-analysis has to be conducted to involve additional clinically applied techniques, such as ultrasound or pressure-volume loop analysis. MRI is meant to be the gold standard in LVEF measurement, but it nevertheless depends on the user and technology (Nabeshima et al., 2019; Wood et al., 2014).

In contrast to this, Gyöngyösi *et al.* (2018) have reviewed current meta-analyses and concluded that the potential beneficial effect of cell therapies after heart failure is still inconclusive and statistically underpowered. The findings in this thesis suggest the beneficial impact of cell therapies in mice, but cannot entirely solve the complex puzzle of the regenerative response. In agreement with the mice data, our findings in patients show that cell therapies are beneficial in responsive patients (60% of patients investigated). Here, the responsiveness dependents on several parameters, such as inflammation and immune status, as well as the angiogenic potential and the amount of proliferating $CD133^{+}$ positive stem cells (Steinhoff et al., 2017a,b).

In summary, this hypothesis can only be partially accepted because there is still an ongoing debate to which this thesis has contributed three publications. Indeed, we can investigate the influence of cell therapies and moderator values via meta-analyses in mice as a valid model, but an ultimate conclusion about their effectiveness in all circumstances cannot be given if only murine data is taken into account.

vii) Stratification of patients

AI models have now been developed to be less opaque black boxes that lack interpretability and transparency; formerly, this was the most important reason patients and clinicians had a skeptical view of this technology, as it is understandable to mistrust unfamiliar interfaces and to hesitate when giving a machine or mathematical algorithm the responsibility of making life-critical decisions (Begoli et al., 2019; Rudin, 2019). A CDSS should be seen as an extended tool, such as a stethoscope, for patient diagnostics that a clinician can naturally utilize to make a therapeutic decision.

In oncology, the application of AI approaches, including its well-known branch of ML, already has a significant impact on enabling precision oncology based on the supervised classification of single-source omics (Azuaje, 2019). These new approaches were successfully transferred to the study setup of the PERFECT Phase III clinical trial (Section 2.3.3). Responsive *vs.* non-responsive patient subgroups have been classified for specific regenerative cell therapy after myocardial infarction. The input data used was mainly derived from peripheral blood. Peripheral blood is used increasingly often as an indicator system because it can serve as a systemic readout of the whole body and not only of specific organs (Hogan et al., 2019; Steinhoff et al., 2017b; Wilkinson, Meredyth G Ll et al., 2020). There is also an ongoing extension of our study that includes RNA-Seq data and derived mutational data obtained from RNA transcripts (Wolfien et al., 2020c). Overall, there is still a limited quantity of patients due to the broad and cost-intensive testing procedures, but this thesis contributes to refining the testing scenarios into a more focused panel, which will be tested in future clinical studies and applications.

Such a focused panel is in agreement with a current comparison of ML methods with traditional models for using administrative claims with electronic medical records to predict heart failure outcomes (Desai et al., 2020), in which approaches with traditional logistic regression were compared on predicting key outcomes in patients with heart failure; the added value of predictive models was evaluated with EHR data. There have been 9,502 patients (aged 65 years or older) in this study with at least one heart failure diagnosis; 6,113 of these patients were included in the training set, and 3,389 were used as the testing set. The study contains a large data set with clinical standard parameters, which have no clear superior predictive value when using ML for stratification. Since we initially also observed the limited predictive capacity in our study, we also used additional molecular data, which could improve the prediction accuracy of our model from 64% to 82%. This was another indication for us to use more specific molecular data, such as RNA-Seq. Stratification

results of a larger cohort show an increase of more than 10% in terms of the predictive accuracy and sensitivity (Wolfien et al., 2020c).

Another essential aspect of our study was the so-called "*expert-in-the-loop*" system (Girardi et al., 2016), which was not only used to improve the selection of the input datasets and evaluation of the predictive performance of the feature selection algorithms, but also to guide the learning process. The interdisciplinary use of ontology-originated transcript information, patient meta-information, and blood and protein data can help medical domain experts become familiar with applying advanced ML and network applications to a high degree. The ultimate aim would be to enable non-computational experts to apply these algorithms and find matching diagnostic support or similar cases, as has been proposed in other CDSSs within cardiac research (Groenhof et al., 2019). The meta-review of Groenhof *et al.* also concludes that the benefit of such a CDSS is highly dependent on the data input.

A last important consideration is the combination of AI approaches and traditional research-oriented mechanistic models (*in vivo* and *in vitro*) that are used not only predict the disease outcome, but also to identify the origin of a certain disease because, for reliable decisions, it is necessary to properly investigate the causes (Baker et al., 2018; Wolfien et al., 2020c). Cabitza *et al.* (2017) have noted that users and designers of a CDSS need to be aware of the inevitable, intrinsic uncertainties that are deeply embedded in medical sciences. Begoli *et al.* (2019), meanwhile, have even written about the need to develop a principled and formal uncertainty quantification discipline in medical AI, especially in its modern, data-rich DL guise.

In summary, based on the analysis of the PERFECT clinical trial, it can be concluded that patient stratification for stem-cell therapy after myocardial infarction is possible through an integrative dataset from human peripheral blood samples. The biomarker signature identified for the responsive patients has to be applied for further predictions and validated on an independent patient cohort.

3.2 Social and ethical considerations of AI and RNA-Seq in the clinic

This section emphasizes the social and ethical impact of AI and RNA-Seq technologies that are used with increasing frequency in clinics.

AI methods seek to solve individual problems within one specific task. While they may excel in interpreting image and contextual information, they are so far not able to make associations the way a human brain does, and they cannot replace doctors in all the tasks they perform (Visvikis et al., 2019). Visvikis *et al.* (2019) have also concluded that AI has not yet achieved the same level of performance as a human expert in all situations; therefore, a full artificial doctor still belongs to the domain of science fiction. However, the role of physicians is likely to evolve as these new techniques are integrated into their practice.

Changing doctor-patient relationships with AI-supported decisions

The major aim, of course, is that an integrative AI concept will allow doctors to spend more time on personal discussions with patients, while leaving time-consuming statistical calculations and predictions to the CDSS (Warraich et al., 2018). Having more time on the patient side would enable doctors to provide better care, which enhances patient trust, the foundation of the doctor-patient relationship (Lysaght et al., 2019). However, doctors also need to ensure that the AI-assisted CDSS does not obstruct the doctor-patient relationship; they must realize that the legal and moral responsibilities for decisions still lies with them. Thus, implementers may need to ensure that doctors are adequately trained on the benefits and pitfalls of AI-assisted CDSS and apply them in practice to *augment* rather than *replace* their clinical decision-making capabilities and duties to patients (Lysaght et al., 2019). To be successful and accepted, a full degree of information transparency should be provided to patients about the features, limitations, and suggestions involved in the AI systems that assist clinicians with their decision-making (Keane and Topol, 2018). This would be an extension of the current classic formulation of informed patient consent, which reflects a disclosure of all relevant information during the decision making process (e.g., information at hand to accept/reject a diagnosis and consent to a proposed therapy plan). However, using these benefits will require a free and rapid flow of information from the EHR to the CDSS platform into reportable outputs that can be validated and disseminated to others outside the doctor-patient relationship, as well. This will require fundamental trade-offs for the control and supervision that patients have regarding the information contained

within the EHR (Lysaght et al., 2019). To circumvent this, researchers and administrators could use aggregated, de-identified data to undertake their analysis. However, it must be noted that no data can be truly de-identified, especially in the era of high-quality imaging and molecular deep sequencing (Palanisamy and Thirunavukarasu, 2019).

AI and genetic information

Patients obtaining genetic testing for various diseases, ranging from cancer to cardiomyopathy, have steadily increased. In addition, today, RNA-Seq experiments can also reveal specific mutational information and may undergo the same guidelines used for DNA-Seq (Karamperis et al., 2020). In this growing, dynamic context, a grasp of the ethical principles and history underlying clinical genetics will provide clinicians with better tools to guide their practice and help patients navigate complex medical-psychosocial terrain (Braverman et al., 2018). However, incidental or secondary findings in the course of testing are an important aspect to consider; once a concerning finding is identified, should it be reported to the patient and the family, even if it was previously excluded? Addressing those concerns is an ongoing bioethical debate on the current disclosure action of such findings in clinical practice (Green et al., 2013; Jamuar et al., 2016). However, there is consensus in the medical community that secondary findings with actionable clinical significance should be given to patients. In particular, in light of data-intense ML algorithms using such genomic data, what these algorithms actually learn and utilize for patient stratification should be transparent.

Humans *plus* or *versus* machines?

The usability of AI-assisted approaches as well as possible perspectives and drawbacks give rise to the question of whether AI is exaggerated or is a gatekeeper of improved medicine (Fig. 3.2). Since the first AI diagnostic system for MRI heart interpretation (Arterys)[6] was approved by the FDA in 2017, more followed in accelerating numbers (in total, two in 2017 and twelve in 2018), which might indicate the increasing impact of AI in the near future of healthcare (Topol, 2019). Nevertheless, AI will not be a panacea, and if used improperly, these systems can replicate or even augment faulty practices rather than improve clinical decisions (Kelly et al., 2019). Each AI application that can potentially be used in the clinic must first be investigated, tested on multiple independent datasets, and thoroughly questioned before its release to the clinical area (Keane and Topol, 2018). One of the most pressing concerns is that AI might become more intelligent than humans, reaching a state called "*superintelligence*" (Mulgan, 2016). This may lead to accelerated technological advancement by surpassing human control with goals that may not match

[6] https://arterys.com/

current societal norms, which is why there is a growing branch of prevention for AI safety that focuses on AI containment (Hall, 2019). Another concerning ethical problem concerns the introduction of rules and data resources that guide AI decisions because ML/DL algorithms can work from vast repositories of information, and the final rationale for their decisions thus might not be completely clear in the end (Oravec, 2019; Vogel, 2017). However, one could argue that this accurately reflects our own human intuition. There is in fact ample useful information in our data that certainly affects our daily life in a positive manner if utilized appropriately. For this reason, current research and this thesis focus on supportive systems instead of systems that autonomously make decisions, such as self-driving cars. Thus, in sum, the result is the human *plus* the machine rather than versus the machine.

Citizen science and public outreach

One possibility to help to empower patients with such new technical developments is to inform them about and involve them in the current potential, limitations, applications, and future directions. Currently, it is still challenging to translate multi-omics techniques and ML into the healthcare system; even if multi-omics approaches have provided a large number of potential biomarkers and achieved resonable short-term benefits (e.g., more accurate treatments, better stratification; Lu and Zhan, 2018). Nevertheless, it will take a long time to fulfill the long-term benefits, such as sensitive, early diagnosis, and significantly improved overall survival (Lu and Zhan, 2018). Recent findings demonstrate that citizen science-based approaches play an important role in the public awareness, acceptance, and implementation of personalized medicine, *i.e.*, genomic testing (PGP-UK Consortium, 2018). Therefore, the presentation of research, including methods and results for non-experts and laypeople, requires a higher level of attention from researchers. Accordingly, the results of this thesis were presented as a Science Slam[7] in Bremen 2018, and the thesis has contributed to a publicly available video, "A Modeler's Tale"[8], about using standards in modeling.

3.3 What has been achieved from a biological perspective?

This section summarizes the most important biological insights and presents ongoing as well as further planned experimental perspectives that were discovered during this thesis.

[7] https://youtu.be/fy0CD7bIeI0
[8] https://figshare.com/articles/A_Modeler_s_Tale/3423371

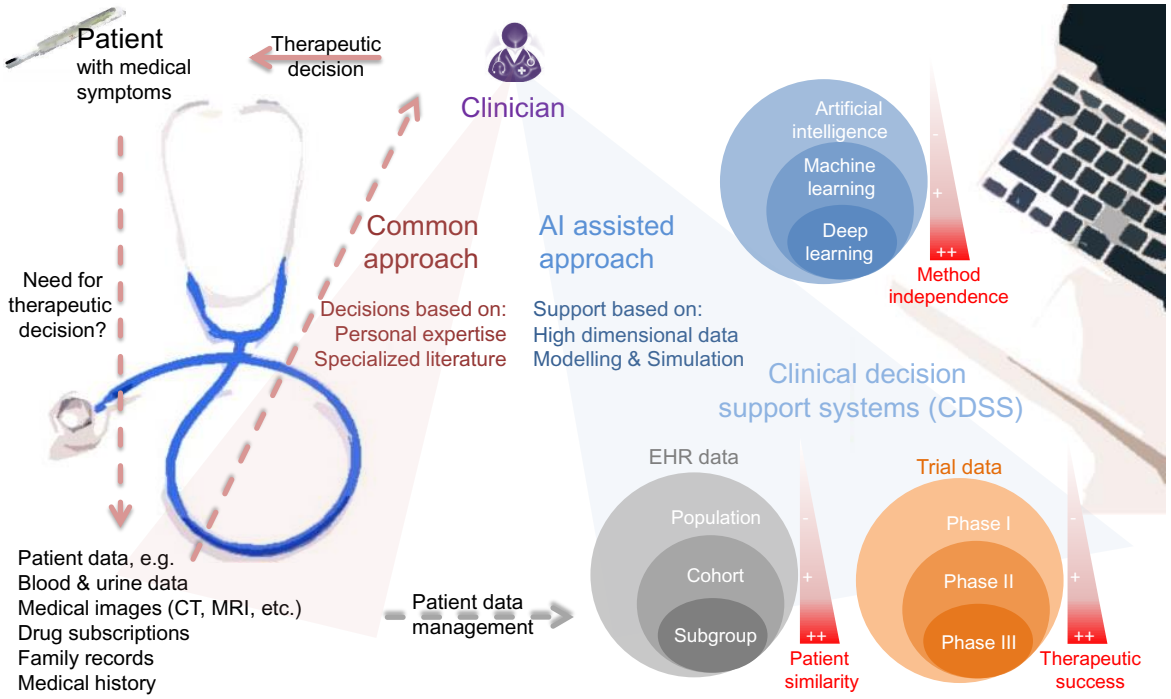


Figure 3.2: Comparison of common and AI-assisted approaches for therapeutic decisions. The clinician is still in the center of the therapeutic decision but in addition to the common approach it will be possible to obtain additional information based on AI-assisted approaches for an enhanced guidance based on electronic health records (EHR), clinical trial data, and different AI model combinations.

The findings of this thesis improved the molecular characterization (significant mRNAs and ncRNAs) of the iSaBs and CM in general (Wolfien et al., 2016). Network analyses revealed specifically enriched subnetworks for iSaBs (Hausburg2017) and contributed to an estimation of the differentiation potential of cardiomyocytes after reprogramming (Müller et al., 2020). Using this molecular iSaBs data, it was possible to validate experimental findings around 2 AMPK, and its role as a potential hub gene and heart rate regulator was confirmed via co-expression network analysis (Yavari et al., 2017). These results in turn contributed to a successful research proposal, and the generated pacemaker cells are now applied as a drug testing system to reduce the quantity of mice experiments (iRhythmics).[9]

[9] https://irhythmics.med.uni-rostock.de/

In terms of single-cell RNA-Seq experiments, this thesis investigated the cellular composition of an entire mammalian heart and contributed to the understanding of the cellular kinetics between the different cell clusters identified. Furthermore, it was confirmed that cardiomyocytes can also be derived from an endothelial cell lineage and not only from a cardiac lineage (Galow et al., 2020; Wolfien et al., 2020a,b). In addition, it would be useful to investigate different vertebrate species to evaluate the cellular composition of different vertebrate lineages, which could give further evidence about the overall composition of the evolutionary development of this highly complex organ. In addition, novel spatial sequencing techniques could be applied to investigate the molecular profile and cellular morphology, as well (Asp et al., 2019).

The AI-assisted clinical analyses utilized in this thesis greatly improved the preoperative prediction accuracy of responsive patients in comparison to traditional statistical models. In line with previous studies, it was shown that the measured routine blood parameters currently in use are not sufficient to stratify patients with the complex disease mechanism. The extension of non-clinical measurements, such as specialized FACS for stem cell surface markers, cell viability assays, and extensive inflammation marker testing, improved the prediction accuracy by more than 20% (Steinhoff et al., 2017b). However, an even higher accuracy is achieved with the integration of RNA-Seq data within the ML feature selection analysis (Wolfien et al., 2020c). Based on the small patient cohort, the molecular mechanism of the responsive patients could not be characterized in-depth. It was only indicated that **Sh2b3** likely plays an important role in responsive patients. A larger patient cohort and time-series gene expression data would be of high value to reveal a specific responsive pattern. Additional experiments in different mice models addressing the role of Sh2b3 show already promising results (Wolfien et al., 2020c).

In conclusion, this thesis provides new insights into mRNAs, miRNAs, and lncRNAs as essential molecules with respect to the differences between various cardiomyocyte cell types, especially functional pacemaker cells. The findings can be considered a significant contribution towards the understanding of this cell type and as a verification of the previous results of Jung *et al.* (2014). Nevertheless, the analyses described are *in silico* predictions, which are a useful starting point for further experiments but must be validated experimentally. In general, the results obtained are essential to advancing the production of *de novo*, highly enriched, stem cell-derived applications for therapy after myocardial infarction or improved characteristics of mature SA cells.

3.4 Conclusions derived from a computer science perspective

This section summarizes the most important computational insights and presents future developments.

The computational Galaxy workflow TRAPLINE that was developed was the basis of several biological applications and has already been reused, extended, and modified for further research projects (more than 50 overall citations) (Grüning et al., 2017; Hausburg et al., 2017; Wolfien et al., 2016; Yavari et al., 2017). The experience gained during the development of this Galaxy workflow was used to establish a BMBF-funded national service for RNA-Seq analyses and workflow development within de.NBI. Together with the University of Freiburg and the Leibniz Institute on Aging in Jena, flexible workflows were built and integrated into the Galaxy framework (de.STAIR). The initial workflow and network analyses were also extended and are currently applied within the BMBF-funded GB-XMAP project, in which RNA-Seq data from ulcerative colitis and schizophrenia patients are analyzed with network approaches to investigate the joint gut-brain relation.

Empowering non-computational users through training

A unique, community-driven set of training materials was jointly developed with more than 150 experts to empower non-computational experts to individually analyze complex datasets (Batut et al., 2018).[10] To facilitate education about computational analysis, the material generated was taught online and at international on-site trainings to graduates and PhD students, post-docs, and professors. A list of training sessions that have been given in light of de.NBI can be found in the Appendix or at de.STAIR training.[11] To self-empower the user, thorough documentation and easily applicable trainings are essential. Traditional formats, e.g., tool descriptions and plain 'README' files, are insufficient for a complex, rapidly changing topic like NGS analyses, which is why the Galaxy Training Material uses i) introductory slides; ii) step-by-step, hands-on guides, including explanatory images; and iii) input files that are ready to use via Zenodo, as well as dedicated Galaxy instances (provided as Docker containers) and Galaxy tours (Batut et al., 2018).

Improved quality of diagnostics and therapies through CDSS

By using the feature selection methods of classical ML approaches, model explainability was achieved in this research project. Interpreting a model on a technical level is still

[10] https://training.galaxyproject.org
[11] https://github.com/destairdenbi/training-material

very distinct from interpreting its decision about the underlying biology, which is why the "*doctor-in-the-loop*" approach was applied; it is increasingly important for clinicians and patients to find explanations and gain trust in AI model predictions, which are almost inevitable. The next phase of the PERFECT study includes ongoing research to validate the findings of the first analysis in 2017 (Steinhoff et al., 2017b; Wolfien et al., 2020c).

In terms of AI model applications in a real-world scenario, there are also plans to use the model predictions on patients with current therapeutic needs because it is not sufficient to validate the models on existing data; they must also be validated and refined in a real-world clinical setting. Thus, the quality of a CDSS should not only be grounded in the current performance metrics, but it should also be required to proof the clinically important benefits in relevant outcomes compared with standard care, along with the satisfaction of patients and doctors.

As was noted in the beginning, healthcare involves rapid decisions in which both machines and humans can generate errors. Thus, iterative systems are needed to detect, prevent, and correct such errors. The interplay of humans and smart algorithms therefore provides a valuable opportunity for such a system. Everyone should be willing to question and change thier own practices, improve standardization, and adhere to guidelines where possible (Steinhoff et al., 2017a). It is inevitable that healthcare professionals and patients will become familiar with upcoming and already existing AI technologies to ensure that CDSS and medical "*traffic-light-systems*" are used appropriately and that the decisions made are as transparent as possible.

In conclusion, this thesis facilitates the transferability and applicability of computational approaches via workflows and contributes to their overall development. The workflows generated have been applied in different use cases in scientific research projects and have already been reused by other research groups. The computational resources provided at Docker-Hub, and Zenodo also had high download rates. Moreover, the *in silico* findings derived have been already experimentally validated and have led to newly funded projects.

Bibliography

S. P. Adam, S.-A. N. Alexandropoulos, P. M. Pardalos, and M. N. Vrahatis. No free lunch theorem: a review. *Approximation and Optimization*, pages 57–82, 2019. URL https://doi.org/10.1007/978-3-030-12767-1_5.

X. Adiconis, D. Borges-Rivera, R. Satija, D. S. DeLuca, M. A. Busby, et al. Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nature Methods*, 10(7):623–629, may 2013. URL https://doi.org/10.1038/nmeth.2483.

E. Afgan, D. Baker, M. van den Beek, D. Blankenberg, D. Bouvier, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research*, 44(W1):W3–W10, jul 2016. URL https://doi.org/10.1093/nar/gkw343.

E. Afgan, D. Baker, B. Batut, M. van den Beek, D. Bouvier, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, 46(W1):W537–W544, jul 2018. doi: 10.1093/nar/gky379. URL https://doi.org/10.1093/nar/gky379.

T. Aittokallio and B. Schwikowski. Graph-based methods for analysing networks in cell biology. *Briefings in Bioinformatics*, 7(3):243–255, may 2006. URL https://doi.org/10.1093/bib/bbl022.

K. M. Akat, D. Moore-McGriff, P. Morozov, M. Brown, T. Gogakos, et al. Comparative RNA-sequencing analysis of myocardial and circulating small RNAs in human heart failure and their utility as biomarkers. *PNAS*, 111(30):11151–11156, jul 2014. URL https://doi.org/10.1073/pnas.1401724111.

P. P. Amaral, M. B. Clark, D. K. Gascoigne, M. E. Dinger, and J. S. Mattick. lncrnadb: a reference database for long noncoding rnas. *Nucleic Acids Res*, 39(Database issue): D146–51, 2011. URL https://doi.org/10.1093/nar/gkq1138.

P. Amstutz, B. Chapman, J. Chilton, M. Heuer, S. Stojanovic, et al. Common Workflow Language, v1.0 Common Workflow Language (CWL) Command Line Tool Description, v1.0. *Figshare online Material*, 2016. URL https://doi.org/10.6084/m9.figshare.3115156.v2.

K. M. Anderson, D. M. Anderson, J. R. McAnally, J. M. Shelton, R. Bassel-Duby, and E. N. Olson. Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. *Nature*, 539(7629):433–436, 2016. URL https://doi.org/10.1038/nature20128.

L. M. Andre, C. R. M. Ausems, D. G. Wansink, and B. Wieringa. Abnormalities in skeletal muscle myogenesis, growth, and regeneration in myotonic dystrophy. *Frontiers in Neurology*, 9:368, 2018. URL https://doi.org/10.3389/fneur.2018.00368.

R. Apweiler, T. Beissbarth, M. R. Berthold, N. Blüthgen, Y. Burmeister, et al. Whither systems medicine? *Experimental & Molecular Medicine*, 50(3):e453, mar 2018. URL https://doi.org/10.1038/emm.2017.290.

M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, et al. Gene ontology: tool for the unification of biology. *Nat Genet*, 25(1):25–9, 2000. URL https://doi.org/10.1038/75556.

M. Asp, S. Giacomello, L. Larsson, C. Wu, D. Fürth, et al. A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell*, 179(7): 1647–1660.e19, dec 2019. URL https://doi.org/10.1016/j.cell.2019.11.025.

F. Azuaje. Artificial intelligence for precision oncology: beyond patient stratification. *npj Precision Oncology*, 3(1):1–5, dec 2019. URL https://doi.org/10.1038/s41698-019-0078-1.

A. Bagnacani, M. Wolfien, and O. Wolkenhauer. Tools for Understanding miRNA?mRNA Interactions for Reproducible RNA Analysis. *Methods in Molecular Biology, book series*, pages 199–214, 2019. URL https://doi.org/10.1007/978-1-4939-8982-9_8.

E. M. Bahassi and P. J. Stambrook. Next-generation sequencing technologies: breaking the sound barrier of human genetics. *Mutagenesis*, 29(5):303–310, sep 2014. doi: 10.1093/mutage/geu031. URL https://doi.org/10.1093/mutage/geu031.

Y. Bai, T. Sun, and P. Ye. Age, gender and diabetic status are associated with effects of bone marrow cell therapy on recovery of left ventricular function after acute myocardial infarction: A systematic review and meta-analysis. *Ageing Research Reviews*, 9(4): 418–423, oct 2010. URL https://doi.org/10.1016/j.arr.2010.05.001.

M. Baker. 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604):452–454, may 2016. URL https://doi.org/10.1038/533452a.

R. E. Baker, J.-M. Peña, J. Jayamohan, and A. Jérusalem. Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biology Letters*, 14(5):20170660, 2018. URL https://doi.org/10.1098/rsbl.2017.0660.

M. L. Bakker, G. J. Boink, B. J. Boukens, A. O. Verkerk, M. van den Boogaard, A. D. den Haan, et al. T-box transcription factor tbx3 reprogrammes mature cardiac myocytes into pacemaker-like cells. *Cardiovasc Res*, 94(3):439–49, 2012. URL https://doi.org/10.1093/cvr/cvs120.

S. Ballouz, W. Verleyen, and J. Gillis. Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. *Bioinformatics*, 31(13):2123–2130, 02 2015. ISSN 1367-4803. doi: 10.1093/bioinformatics/btv118. URL https://doi.org/10.1093/bioinformatics/btv118.

D. C. Bartos, E. Grandi, and C. M. Ripplinger. Ion Channels in the Heart. *Comprehensive Physiology*, 5(3):1423–1464, jun 2015. URL http://doi.wiley.com/10.1002/cphy.c140069.

B. Batut, S. Hiltemann, A. Bagnacani, D. Baker, V. Bhardwaj, et al. Community-driven data analysis training for biology. *Cell Systems*, 6(6):752 – 758.e1, 2018. URL https://doi.org/10.1016/j.cels.2018.05.012.

B. K. Beaulieu-Jones and C. S. Greene. Reproducibility of computational workflows is automated using continuous analysis. *Nature Biotechnology*, 35(4):342–346, apr 2017. URL https://doi.org/10.1038/nbt.3780.

P. E. Beeler, D. W. Bates, and B. L. Hug. Clinical decision support systems. *Swiss Medical Weekly*, 144, 2014. URL https://doi.org/10.4414/smw.2014.14073.

E. Begoli, T. Bhattacharya, and D. Kusnezov. The need for uncertainty quantification in machine-assisted medical decision making. *Nature Machine Intelligence*, 1(1):20–23, jan 2019. URL https://doi.org/10.1038/s42256-018-0004-1.

S. Bej, N. Davtyan, M. Wolfien, M. Nassar, and O. Wolkenhauer. Loras: An oversampling approach for imbalanced datasets. *arXiv*, 2019. URL https://arxiv.org/abs/1908.08346v3.

S. M. Bello, M. Shimoyama, E. Mitraka, S. J. F. Laulederkind, C. L. Smith, et al. Disease Ontology: improving and unifying disease annotations across species. *Disease models & mechanisms*, 11(3), 2018. URL https://doi.org/10.1242/dmm.032839.

E. J. Benjamin, S. S. Virani, C. W. Callaway, A. M. Chamberlain, A. R. Chang, et al. Heart Disease and Stroke Statistics 2018 Update: A Report From the American Heart Association. *Circulation*, 137(12), mar 2018. URL https://doi.org/10.1161/CIR.0000000000000558.

D. R. Bentley, S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456 (7218):53–9, 2008. URL https://doi.org/10.1038/nature07517.

K. Bersell, S. Arab, B. Haring, and B. Kühn. Neuregulin1/ErbB4 Signaling Induces Cardiomyocyte Proliferation and Repair of Heart Injury. *Cell*, 138(2):257–270, jul 2009. URL https://doi.org/10.1016/j.cell.2009.04.060.

G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, 25(8):1091–1093, 02 2009. URL https://doi.org/10.1093/bioinformatics/btp101.

J. S. Bloom, Z. Khan, L. Kruglyak, M. Singh, and A. A. Caudy. Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel gene expression microarrays. *BMC Genomics*, 10:221, 2009. URL https://doi.org/10.1186/1471-2164-10-221.

C. Boettiger. An introduction to Docker for reproducible research. *ACM SIGOPS Operating Systems Review*, 49(1):71–79, jan 2015. URL https://doi.org/10.1145/2723872.2723882.

Boettiger, Carl and D. Eddelbuettel. An Introduction to Rocker: Docker Containers for R. *arXiv*, 2017. URL https://arxiv.org/abs/1710.03675.

M. Boyett, H. Honjo, and I. Kodama. The sinoatrial node, a heterogeneous pacemaker structure. *Cardiovascular Research*, 47(4):658–687, 09 2000. URL https://doi.org/10.1016/S0008-6363(00)00135-8.

M. R. Boyett, H. Dobrzynski, M. K. Lancaster, S. A. Jones, H. Honjo, et al. Sophisticated architecture is required for the sinoatrial node to perform its normal pacemaker function. *J Cardiovasc Electrophysiol*, 14(1):104–6, 2003. URL https://doi.org/10.1046/j.1540-8167.2003.02307.x.

G. Braverman, Z. E. Shapiro, and J. A. Bernstein. Ethical Issues in Contemporary Clinical Genetics. *Mayo Clinic Proceedings: Innovations, Quality & Outcomes*, 2(2):81–90, jun 2018. URL https://doi.org/10.1016/j.mayocpiqo.2018.03.005.

N. L. Bray, H. Pimentel, P. Melsted, and L. Pachter. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, 34(5):525–527, may 2016. URL https://doi.org/10.1038/nbt.3519.

M. Bressan, G. Liu, and T. Mikawa. Early mesodermal cues assign avian cardiac pacemaker fate potential in a tertiary heart field. *Science*, 340(6133):744–748, may 2013. URL https://doi.org/10.1126/science.1232877.

K. M. Broughton, B. J. Wang, F. Firouzi, F. Khalafalla, S. Dimmeler, et al. Mechanisms of Cardiac Repair and Regeneration. *Circulation research*, 122(8):1151–1163, apr 2018. URL https://doi.org/10.1161/CIRCRESAHA.117.312586.

F. Cabitza, R. Rasoini, and G. F. Gensini. Unintended Consequences of Machine Learning in Medicine. *JAMA*, 318(6):517–518, 08 2017. URL https://doi.org/10.1001/jama.2017.7797.

C. Cadeddu Dessalvi, A. Pepe, C. Penna, A. Gimelli, R. Madonna, et al. Sex differences in anthracycline-induced cardiotoxicity: the benefits of estrogens. *Heart Failure Reviews*, 24(6):915–925, nov 2019. URL https://doi.org/10.1007/s10741-019-09820-2.

X. Cai, C. H. Hagedorn, and B. R. Cullen. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*, 10(12):1957–1966, dec 2004. URL http://www.rnajournal.org/cgi/doi/10.1261/rna.7135204.

J. Cal-Gonzalez, I. Rausch, L. K. Shiyam Sundar, M. L. Lassen, O. Muzik, et al. Hybrid imaging: Instrumentation and data processing. *Frontiers in Physics*, 6:47, 2018. URL https://doi.org/10.3389/fphy.2018.00047.

Y. Cao, X. Wang, and G. Peng. SCSA: A Cell Type Annotation Tool for Single-Cell RNA-seq Data. *Frontiers in Genetics*, 11:490, may 2020. URL https://doi.org/10.3389/fgene.2020.00490.

P. Carninci, T. Kasukawa, S. Katayama, J. Gough, M. C. Frith, et al. The transcriptional landscape of the mammalian genome. 309(5740):1559–1563, 2005. URL https://doi.org/10.1126/science.1112014.

P. Chakraborty, S. Nattel, and K. Nanthakumar. Linking cellular energy state to atrial fibrillation pathogenesis: Potential role of adenosine monophosphate?activated protein kinase. *Heart Rhythm*, apr 2020. URL https://doi.org/10.1016/j.hrthm.2020.03.025.

F. Chaudhry, J. Isherwood, T. Bawa, D. Patel, K. Gurdziel, et al. Single-Cell RNA Sequencing of the Cardiovascular System: New Looks for Old Diseases. *Frontiers in Cardiovascular Medicine*, 6, dec 2019. URL https://doi.org/10.3389/fcvm.2019.00173.

C. M. Chen, Y. L. Lu, C. P. Sio, G. C. Wu, W. S. Tzou, et al. Gene ontology based housekeeping gene selection for rna-seq normalization. *Methods*, 67(3):354–63, 2014. URL https://doi.org/10.1016/j.ymeth.2014.01.019.

E. Y. Chen, C. M. Tan, Y. Kou, Q. Duan, Z. Wang, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*, 14:128, apr 2013. URL https://doi.org/10.1186/1471-2105-14-128.

G. Chen, M. Lu, Z. Shi, S. Xia, Y. Ren, et al. Development and validation of machine learning prediction model based on computed tomography angiography-derived hemodynamics for rupture status of intracranial aneurysms: a Chinese multicenter study. *European Radiology*, 2020. URL https://doi.org/10.1007/s00330-020-06886-7.

G. Chiapparo, X. Lin, F. Lescroart, S. Chabab, C. Paulissen, et al. Mesp1 controls the speed, polarity, and directionality of cardiovascular progenitor migration. *Journal of Cell Biology*, 213(4):463–477, may 2016. URL https://doi.org/10.1083/jcb.201505082.

M. Civelek and A. J. Lusis. Systems genetics approaches to understand complex traits. *Nature Reviews Genetics*, 15(1):34–48, jan 2014. URL https://dx.doi.org/10.1038%2Fnrg3575.

N. Cloonan, Q. Xu, G. J. Faulkner, D. F. Taylor, D. T. Tang, et al. Rna-mate: a recursive mapping strategy for high-throughput rna-sequencing data. *Bioinformatics*, 25(19): 2615–6, 2009. URL https://dx.doi.org/10.1093%2Fbioinformatics%2Fbtp459.

A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, et al. A survey of best practices for RNA-seq data analysis. *Genome biology*, 17:13, jan 2016. URL https://doi.org/10.1186/s13059-016-0881-8.

G. M. Cooper. The cell: A molecular approach. 2nd edition. *ISSN: 0878931066*, 2000.

G. Currie and C. Delles. Precision Medicine and Personalized Medicine in Cardiovascular Disease. *Advances in Experimental Medicine and Biology*, 1065:589–605, 2018. URL https://doi.org/10.1007/978-3-319-77932-4_36.

da Costa, BR and Juni, P. Systematic reviews and meta-analyses of randomized trials: principles and pitfalls. *European Heart Journal*, 35:3336–3345, dec 2014. URL https://doi.org/10.1093/eurheartj/ehu424.

L. M. Davis, M. E. Rodefeld, K. Green, E. C. Beyer, and J. E. Saffitz. Gap junction protein phenotypes of the human heart and conduction system. *Journal of Cardiovascular Electrophysiology*, 6(10):813–822, 1995. URL https://doi.org/10.1111/j.1540-8167.1995.tb00357.x.

L. de la Garza, J. Veit, A. Szolek, M. Röttig, S. Aiche, et al. From the desktop to the grid: scalable bioinformatics via workflow conversion. *BMC bioinformatics*, 17:127, mar 2016. URL https://doi.org/10.1186/s12859-016-0978-9.

T. Y. de Soysa, S. S. Ranade, S. Okawa, S. Ravichandran, et al. Single-cell analysis of cardiogenesis reveals basis for organ-level developmental defects. *Nature*, 572(7767): 120–124, aug 2019. URL https://doi.org/10.1038/s41586-019-1414-x.

A. Del Mazo-Barbara, V. Nieto, C. Mirabel, B. Reyes, J. García-López, et al. Streamlining the qualification of computerized systems in GxP-compliant academic cell therapy facilities. *Cytotherapy*, 18(9):1237–1239, sep 2016. URL https://doi.org/10.1016/j.jcyt.2016.06.003.

R. J. Desai, S. V. Wang, M. Vaduganathan, T. Evers, and S. Schneeweiss. Comparison of Machine Learning Methods With Traditional Models for Use of Administrative Claims With Electronic Medical Records to Predict Heart Failure Outcomes. *JAMA network open*, 3(1):e1918962, jan 2020. URL https://doi.org/10.1001/jamanetworkopen.2019.18962.

P. Di Tommaso, E. Palumbo, M. Chatzou, P. Prieto, M. L. Heuer, and C. Notredame. The impact of Docker containers on the performance of genomic pipelines. *PeerJ*, 3, 2015. URL https://doi.org/10.7717/peerj.1273.

G. Dietl, M. Langhammer, and U. Renne. Model simulations for genetic random drift in the outbred strain Fzt:DU. *Archives Animal Breeding*, 47(6):595–604, oct 2004. URL https://doi.org/10.5194/aab-47-595-2004.

S. E. Dilsizian and E. L. Siegel. Artificial Intelligence in Medicine and Cardiac Imaging: Harnessing Big Data and Advanced Computing to Provide Personalized Medical Diagnosis and Treatment. *Current Cardiology Reports*, 16(1):441, jan 2014. URL https://doi.org/10.1007/s11886-013-0441-8.

H. Dobrzynski, M. R. Boyett, and R. H. Anderson. New insights into pacemaker activity. *Circulation*, 115(14):1921–1932, 2007. URL https://doi.org/10.1161/CIRCULATIONAHA.106.616011.

K. G. Dowell, A. K. Simons, H. Bai, B. Kell, Z. Z. Wang, et al. Novel insights into embryonic stem cell self-renewal revealed through comparative human and mouse systems biology networks. *Stem Cells*, 32(5):1161–1172, 2014. URL https://doi.org/10.1002/stem.1612.

S. Dreiseitl and L. Ohno-Machado. Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics*, 35(5): 352 – 359, 2002. URL https://doi.org/10.1016/S1532-0464(03)00034-0.

H. Dweep, C. Sticht, P. Pandey, and N. Gretz. mirwalk–database: prediction of possible mirna binding sites by "walking" the genes of three genomes. *J Biomed Inform*, 44(5): 839–47, 2011. URL https://doi.org/10.1016/j.jbi.2011.05.002.

J. Fallmann, P. Videm, A. Bagnacani, B. Batut, M. A. Doyle, et al. The RNA workbench 2.0: next generation RNA data analysis. *Nucleic Acids Research*, 47(W1):W511–W515, 05 2019. ISSN 0305-1048. doi: 10.1093/nar/gkz353. URL https://doi.org/10.1093/nar/gkz353.

M. Farahbod and P. Pavlidis. Differential coexpression in human tissues and the confounding effect of mean expression levels. *Bioinformatics*, 35(1):55–61, jan 2019. URL https://doi.org/10.1093/bioinformatics/bty538.

J. A. Fels and G. Manfredi. Sex Differences in Ischemia/Reperfusion Injury: The Role of Mitochondrial Permeability Transition. *Neurochemical Research*, 44(10):2336–2345, oct 2019. URL https://doi.org/10.1007/s11064-019-02769-6.

F. Finotello, E. Lavezzo, L. Bianco, L. Barzon, P. Mazzon, et al. Reducing bias in rna sequencing data: a novel approach to compute counts. *BMC Bioinformatics*, 15 Suppl 1:S7, 2014. URL https://doi.org/10.1186/1471-2105-15-s1-s7.

M. J. Foglia and K. D. Poss. Building and re-building the heart by cardiomyocyte proliferation. *Development*, 143(5):729–740, 2016. URL https://doi.org/10.1242/dev.132910.

L. J. Frey. Data integration strategies for predictive analytics in precision medicine. *Personalized Medicine*, 15(6):543–551, nov 2018. URL https://doi.org/10.2217/pme-2018-0035.

A.-M. Galow, M. Wolfien, P. Müller, M. Bartsch, R. M. Brunner, et al. Integrative cluster analysis of whole hearts reveals proliferative cardiomyocytes in adult mice. *Cells*, 9(5): 1144, May 2020. URL http://dx.doi.org/10.3390/cells9051144.

R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, et al. Chipster: user-friendly analysis software for microarray and other high-throughput data. *Genome Biology*, 5(10):R80, 2004. URL https://doi.org/10.1186/gb-2004-5-10-r80.

V. George, S. Colombo, and K. L. Targoff. An early requirement for nkx2.5 ensures the first and second heart field ventricular identity and cardiac function into adulthood. *Developmental Biology*, 400(1):10–22, apr 2015. URL https://doi.org/10.1016/j.ydbio.2014.12.019.

D. Girardi, J. Küng, R. Kleiser, M. Sonnberger, D. Csillag, et al. Interactive knowledge discovery with the doctor-in-the-loop: a practical example of cerebral aneurysms research. *Brain Informatics*, 3(3):133–143, sep 2016. URL https://doi.org/10.1007/s40708-016-0038-2.

V. Gligorijević and N. Pržulj. Methods for biological data integration: perspectives and challenges. *Journal of the Royal Society, Interface / the Royal Society*, 12(112): 20150571–, 2015. URL https://doi.org/10.1098/rsif.2015.0571.

W. R. Goodyer, B. M. Beyersdorf, D. T. Paik, L. Tian, G. Li, et al. Transcriptomic Profiling of the Developing Cardiac Conduction System at Single-Cell Resolution. *Circulation Research*, 125(4):379–397, aug 2019. URL https://doi.org/10.1161/circresaha.118.314578.

P. Granados Moreno, Y. Joly, and B. M. Knoppers. Public?Private Partnerships in Cloud-Computing Services in the Context of Genomic Research. *Frontiers in Medicine*, 4:3, jan 2017. URL https://doi.org/10.3389/fmed.2017.00003.

J. T. Granados-Riveron, T. K. Ghosh, M. Pope, F. Bu'Lock, C. Thornborough, et al. Alpha-cardiac myosin heavy chain (myh6) mutations affecting myofibril formation are associated with congenital heart defects. *Hum Mol Genet*, 19(20):4007–16, 2010. URL https://doi.org/10.1093/hmg/ddq315.

R. C. Green, J. S. Berg, W. W. Grody, S. S. Kalia, B. R. Korf, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genetics in Medicine*, 15(7):565–574, 2013. URL https://doi.org/10.1038/gim.2013.73.

L. Griebel, H. U. Prokosch, F. Köpcke, D. Toddenroth, J. Christoph, et al. A scoping review of cloud computing in healthcare, dec 2015. URL https://doi.org/10.1186/s12911-015-0145-7.

T. K. J. Groenhof, F. W. Asselbergs, R. H. Groenwold, D. E. Grobbee, F. L. Visseren, and M. L. Bots. The effect of computerized decision support systems on cardiovascular risk factors: A systematic review and meta-analysis. *BMC Medical Informatics and Decision Making*, 19(1):108, jun 2019. URL https://doi.org/10.1186/s12911-019-0824-x.

D. Grün. Deciphering Cell Fate Decision by Integrated Single-Cell Sequencing Analysis. *Annual Reviews*, 2020. doi: 10.1146/annurev-biodatasci-111419. URL https://doi.org/10.1146/annurev-biodatasci-111419-.

B. A. Grüning, J. Fallmann, D. Yusuf, S. Will, A. Erxleben, F. Eggenhofer, et al. The RNA workbench: best practices for RNA and high-throughput sequencing bioinformatics in Galaxy. *Nucleic Acids Research*, 45:D626–D634, jun 2017. URL https://doi.org/10.1093/nar/gkx409.

T. Gutschner, M. Haemmerle, and S. Diederichs. Malat1 - a paradigm for long noncoding rna function in cancer. *Journal of Molecular Medicine*, 91, 03 2013. URL https://doi.org/10.1007/s00109-013-1028-y.

M. Gyöngyösi, P. M. Haller, D. J. Blake, and E. Martin Rendon. Meta-Analysis of Cell Therapy Studies in Heart Failure and Acute Myocardial Infarction. *Circulation Research*, 123(2):301–308, jul 2018. URL https://doi.org/10.1161/circresaha.117.311302.

R. Haas, A. Zelezniak, J. Iacovacci, S. Kamrad, S. J. Townsend, and M. Ralser. Designing and interpreting 'multi-omic' experiments that may change our understanding of biology. *Current Opinion in Systems Biology*, 6:37–45, dec 2017. URL https://doi.org/10.1016/j.coisb.2017.08.009.

A. B. Haidich. Meta-analysis in medical research. *Hippokratia*, pages 29–37, dec 2010.

L. Hakes, J. W. Pinney, D. L. Robertson, and S. C. Lovell. Protein-protein interaction networks and biology?what's the connection? *Nature Biotechnology*, 26(1):69–72, jan 2008. URL https://doi.org/10.1038/nbt0108-69.

M. Hall. Artificial intelligence and nuclear medicine. *Nuclear medicine communications*, 40(1):1, 2019. URL https://dx.doi.org/10.1097%2FMNM.0000000000000937.

X. Han, J. Zhang, Y. Liu, X. Fan, S. Ai, et al. The lncRNA Hand2os1/Uph locus orchestrates heart development through regulation of precise expression of Hand2. *Development (Cambridge)*, 146(13), jul 2019. URL https://doi.org/10.1242/dev.176198.

S. Harrer, P. Shah, B. Antony, and J. Hu. Artificial Intelligence for Clinical Trial Design. *Cell*, 2019. URL https://doi.org/10.1016/j.tips.2019.05.005.

Y. Hasin, M. Seldin, and A. Lusis. Multi-omics approaches to disease. *Genome biology*, 18 (1):83, 2017. URL https://doi.org/10.1186/s13059-017-1215-1.

F. Hausburg, J. Jung, M. Hoch, M. Wolfien, A. Yavari, et al. (Re-)programming of subtype specific cardiomyocytes. *Advanced Drug Delivery Reviews*, 2017. URL https://doi.org/10.1016/j.addr.2017.09.005.

Z. He, M. Grunewald, Y. Dor, and E. Keshet. VEGF regulates relative allocation of Isl1+ cardiac progenitors to myocardial and endocardial lineages. *Mechanisms of Development*, 142:40–49, nov 2016. URL https://doi.org/10.1016/j.mod.2016.10.004.

J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. W. on evidence-based medicine) Li, M. J. Page, V. A. Welch, and Cochrane Collaboration. Cochrane handbook for systematic reviews of interventions. *Cochrane Collaboration*, 2019. ISSN 9781119536628.

S. A. Hogan, A. Courtier, P. F. Cheng, N. F. Jaberg-Bentele, S. M. Goldinger, et al. Peripheral blood TCR repertoire profiling may facilitate patient stratification for immunotherapy against melanoma. *Cancer Immunology Research*, 7(1):77–85, jan 2019. URL https://doi.org/10.1158/2326-6066.CIR-18-0136.

A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller. Causability and explainability of artificial intelligence in medicine. *WIREs Data Mining and Knowledge Discovery*,

9(4), 2019. URL https://doi.org/10.1002/widm.1312.

L. Hood, R. Balling, C. Auffray, J. Clairambault, A. Deutsch, et al. Revolutionizing medicine in the 21st century through systems approaches. *Biotechnology Journal*, 7(8): 992–1001, aug 2012. URL http://doi.wiley.com/10.1002/biot.201100306.

N. E. Ilott and C. P. Ponting. Predicting long non-coding RNAs using RNA sequencing. *Methods*, 63(1):50–59, sep 2013. URL https://doi.org/10.1016/j.ymeth.2013.03.019.

J. Ison, H. Ienasescu, P. Chmura, E. Rydza, H. Menager, et al. The bio.tools registry of software tools and data resources for the life sciences. *Genome Biology*, 20(1):164, aug 2019. URL https://doi.org/10.1186/s13059-019-1772-6.

J. C. Ison, M. Kalavs, I. Jonassen, D. M. Bolser, M. Uludag, et al. Edam: an ontology of bioinformatics operations, types of data and identifiers, topics and formats. *Bioinformatics*, 29:1325 – 1332, 2013. URL https://dx.doi.org/10.1093%2Fbioinformatics%2Fbtt113.

A. Jain and R. Bansal. Applications of regenerative medicine in organ transplantation. *Journal of pharmacy & bioallied sciences*, 7(3):188–94, 2015. URL https://doi.org/10.4103/0975-7406.160013.

S. S. Jamuar, J. L. Kuan, M. Brett, Z. Tiang, W. L. W. Tan, et al. Incidentalome from Genomic Sequencing: A Barrier to Personalized Medicine? *EBioMedicine*, 5:211–216, mar 2016. URL https://doi.org/10.1016/j.ebiom.2016.01.030.

S. J. Jansen of Lorkeers, J. E. C. Eding, H. M. Vesterinen, T. I. G. Van Der Spoel, et al. Similar effect of autologous and allogeneic cell therapy for ischemic heart disease: Systematic review and meta-analysis of large animal studies. *Circulation Research*, 116 (1):80–86, jan 2015. URL https://doi.org/10.1161/circresaha.116.304872.

L. Jiang, F. Schlesinger, C. A. Davis, Y. Zhang, R. Li, et al. Synthetic spike-in standards for rna-seq experiments. *Genome Res*, 21(9):1543–51, 2011. URL https://dx.doi.org/10.1101%2Fgr.121095.111.

R. M. John and S. Kumar. Sinus Node and Atrial Arrhythmias. *Circulation*, 133(19): 1892–1900, may 2016. URL https://doi.org/10.1161/CIRCULATIONAHA.116.018011.

B. Joung, L. Tang, M. Maruyama, S. Han, Z. Chen, et al. Intracellular calcium dynamics and acceleration of sinus rhythm by β-adrenergic stimulation. *Circulation*, 119(6): 788–796, feb 2009. URL https://doi.org/10.1161/circulationaha.108.817379.

J. J. Jung, B. Husse, C. Rimmbach, S. Krebs, J. Stieber, et al. Programming and isolation of highly pure physiologically and pharmacologically functional sinus-nodal bodies from pluripotent stem cells. *Stem Cell Reports*, 2(5):592–605, 2014. URL https://dx.doi.org/10.1016%2Fj.stemcr.2014.03.006.

N. Kapoor, W. Liang, E. Marban, and H. C. Cho. Direct conversion of quiescent cardiomyocytes to pacemaker cells by expression of tbx18. *Nat Biotechnol*, 31(1):54–62, 2013. URL https://doi.org/10.1038/nbt.2465.

K. Karamperis, S. Wadge, M. Koromina, and G. P. Patrinos. Chapter 10. Genetic Testing. *Applied Genomics and Public Health*, pages 189–208, jan 2020. URL http://dx.doi.org/10.1016/B978-0-12-813695-9.00010-8.

K. J. Karczewski and M. P. Snyder. Integrative omics for health and disease. *Nature reviews. Genetics*, 19(5):299–310, may 2018. URL https://doi.org/10.1038/nrg.2018.4.

P. A. Keane and E. J. Topol. With an eye to AI and autonomous diagnosis. *npj Digital Medicine*, 1(1):40, dec 2018. doi: 10.1038/s41746-018-0048-y. URL https://doi.org/10.1038/s41746-018-0048-y.

I. Kehat, A. Gepstein, A. Spira, J. Itskovitz-Eldor, and L. Gepstein. High-resolution electrophysiological assessment of human embryonic stem cell-derived cardiomyocytes. *Circulation Research*, 91(8):659–661, 2002. URL https://doi.org/10.1016/j.yjmcc.2019.09.015.

C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King. Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17(1):1–9, oct 2019. URL https://doi.org/10.1186/s12916-019-1426-2.

A. Kennedy, D. D. Finlay, D. Guldenring, R. Bond, K. Moran, and J. McLaughlin. The Cardiac Conduction System: Generation and Conduction of the Cardiac Impulse, sep 2016. URL https://doi.org/10.1016/j.cnc.2016.04.001.

A. Kikkawa. Random Matrix Analysis for Gene Interaction Networks in Cancer Cells. *Scientific Reports*, 8(1):10607, dec 2018. URL https://doi.org/10.1038/s41598-018-28954-1.

D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, et al. Tophat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, 14(4):R36, 2013. URL https://doi.org/10.1186/gb-2013-14-4-r36.

M. Kircher and J. Kelso. High-throughput dna sequencing–concepts and limitations. *Bioessays*, 32(6):524–36, 2010. URL https://dx.doi.org/10.1126%2Fscience.1158441.

J. A. Kobashigawa. The Search for a Gold Standard to Detect Rejection in Heart Transplant Patients. *Circulation*, 135(10):936–938, mar 2017. URL https://doi.org/10.1161/circulationaha.117.026752.

J. Koster and S. Rahmann. Snakemake–a scalable bioinformatics workflow engine. *Bioinformatics*, 28(19):2520–2522, oct 2012. URL https://doi.org/10.1093/bioinformatics/bts480.

K. R. Kukurba and S. B. Montgomery. RNA Sequencing and Analysis. *Cold Spring Harbor protocols*, 2015(11):951–69, apr 2015. URL https://doi.org/10.1101/pdb.top084970.

J. T. Kung, D. Colognori, and J. T. Lee. Long noncoding RNAs: Past, present, and future. *Genetics*, 193(3):651–669, 2013. doi: 10.1534/genetics.112.146704. URL https://doi.org/10.1534/genetics.112.146704.

J. M. Kwon, Y. Lee, Y. Lee, S. Lee, and J. Park. An algorithm based on deep learning for predicting in-hospital cardiac arrest. *Journal of the American Heart Association*, 2018. URL https://doi.org/10.1161/JAHA.118.008678.

G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, et al. RNA velocity of single cells. *Nature*, 560(7719):494–498, aug 2018. URL https://doi.org/10.1038/s41586-018-0414-6.

A. Lachmann, D. J. Clarke, D. Torre, Z. Xie, and A. Ma'ayan. Interoperable RNA-Seq analysis in the cloud. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1863(6):194521, jun 2020. URL https://doi.org/10.1016/j.bbagrm.2020.194521.

D. Lähnemann, J. Köster, E. Szczurek, D. J. McCarthy, S. C. Hicks, et al. Eleven grand challenges in single-cell data science. *Genome Biology*, 21(1):1–35, feb 2020. URL https://doi.org/10.1186/s13059-020-1926-6.

A. T. Lamm, M. R. Stadler, H. Zhang, J. I. Gent, and A. Z. Fire. Multimodal rna-seq using single-strand, double-strand, and circligase-based capture yields a refined and extended description of the c. elegans transcriptome. *Genome Res*, 21(2):265–75, 2011. URL https://doi.org/10.1101/gr.108845.110.

S. Lampa, M. Dahlö, P. I. Olason, J. Hagberg, and O. Spjuth. Lessons learned from implementing a national infrastructure in Sweden for storage and analysis of next-generation sequencing data. *GigaScience*, 2(1):9, dec 2013. URL https://doi.org/10.1186/2047-217X-2-9.

C. I. Lang, M. Wolfien, A. Langenbach, P. Müller, O. Wolkenhauer, et al. Cardiac Cell Therapies for the Treatment of Acute Myocardial Infarction: A Meta-Analysis from Mouse Studies. *Cellular Physiology and Biochemistry*, 42(1):254–268, jun 2017. URL https://doi.org/10.1159/000477324.

P. Langfelder and S. Horvath. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, 9(1):559, dec 2008. URL https://doi.org/10.1186/1471-2105-9-559.

Y. H. Lee. An overview of meta-analysis for clinicians. *The Korean journal of internal medicine*, 33(2):277–283, 2018. URL https://doi.org/10.3904/kjim.2016.195.

J. A. Leopold, B. A. Maron, and J. Loscalzo. The application of big data to cardiovascular disease: Paths to precision medicine. *Journal of Clinical Investigation*, 130(1):29–38, jan 2020. doi: 10.1172/JCI129203. URL https://doi.org/10.1172/jci129203.

L. Liu, Y. Li, S. Li, N. Hu, Y. He, et al. Comparison of next-generation sequencing systems. *J Biomed Biotechnol*, 2012:251364, 2012a. URL https://doi.org/10.1155/2012/251364.

Y. Liu, M. Koyutürk, J. S. Barnholtz-Sloan, and M. R. Chance. Gene interaction enrichment and network analysis to identify dysregulated pathways and their interactions in complex diseases. *BMC systems biology*, 6:65, jun 2012b. URL https://doi.org/10.1186/1752-0509-6-65.

Y. Liu, H. Nie, H. Liu, and F. Lu. Poly(A) inclusive RNA isoform sequencing (PAIso?seq) reveals wide-spread non-adenosine residues within RNA poly(A) tails. *Nature Communications*, 10(1):1–13, dec 2019. URL https://doi.org/10.1038/s41467-019-13228-9.

S. C. Lott, M. Wolfien, K. Riege, A. Bagnacani, O. Wolkenhauer, et al. Customized workflow development and data modularization concepts for RNA-Sequencing and metatranscriptome experiments. *Journal of Biotechnology*, jul 2017. doi: 10.1016/j.jbiotec.2017.06.1203. URL https://doi.org/10.1016/j.jbiotec.2017.06.1203.

M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, dec 2014. URL https://doi.org/10.1186/s13059-014-0550-8.

M. Lu and X. Zhan. The crucial role of multiomic approach in cancer research and clinically relevant outcomes. *EPMA Journal*, 9(1):77–102, mar 2018. URL https://doi.org/10.1007/s13167-018-0128-8.

Y. Luo, X. Zhao, J. Zhou, J. Yang, Y. Zhang, et al. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nature Communications*, 8(1):1–13, dec 2017. URL https://doi.org/10.1038/s41467-017-00680-8.

A. J. Lusis and J. N. Weiss. Cardiovascular Networks. *Circulation*, 121(1):157–170, jan 2010. URL https://doi.org/10.1161/circulationaha.108.847699.

T. Lysaght, H. Y. Lim, V. Xafis, and K. Y. Ngiam. AI-Assisted Decision-making in Healthcare: The Application of an Ethics Framework for Big Data in Health and Research. *Asian Bioethics Review*, 11(3):299–314, sep 2019. URL https://doi.org/10.1007/s41649-019-00096-0.

A. Lysenko, I. A. Roznov, M. Saqi, A. Mazein, C. J. Rawlings, and C. Auffray. Representing and querying disease networks using graph databases. *BMC Bio Data Mining*, 2016.

URL https://doi.org/10.1186/s13040-016-0102-8.

F. H. Martini, J. L. Nath, and E. F. Bartholomew. Fundamentals of anatomy and physiology. *ISSN 0134396022*, 2020.

L. McInnes, J. Healy, and J. Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv*, feb 2018. URL http://arxiv.org/abs/1802.03426.

S. M. Meilhac, F. Lescroart, C. Blanpain, and M. E. Buckingham. Cardiac Cell Lineages that Form the Heart. *Cold Spring Harbor Perspectives in Medicine*, 4(9):a013888–a013888, sep 2014. URL https://doi.org/10.1101/cshperspect.a013888.

J. Menche, A. Sharma, M. Kitsak, S. D. Ghiassian, M. Vidal, et al. Uncovering disease-disease relationships through the incomplete interactome. *Science*, 347(6224), 2015. URL https://doi.org/10.1126/science.1257601.

B. Mirza, W. Wang, J. Wang, H. Choi, N. C. Chung, and P. Ping. Machine Learning and Integrative Analysis of Biomedical Big Data. *Genes*, 10(2):87, jan 2019. URL https://doi.org/10.3390/genes10020087.

S. Mohseni, N. Zarei, and E. D. Ragan. A survey of evaluation methods and measures for interpretable machine learning. *arXiv*, 2018. URL arXiv:1811.11839.

M. Mollova, K. Bersell, S. Walsh, J. Savla, L. T. Das, et al. Cardiomyocyte proliferation contributes to heart growth in young humans. *PNAS*, 110(4):1446–1451, jan 2013. URL https://doi.org/10.1073/pnas.1214608110.

A. F. Moorman and V. M. Christoffels. Cardiac chamber formation: development, genes, and evolution. *Physiol Rev*, 83(4):1223–67, 2003. URL https://doi.org/10.1152/physrev.00006.2003.

A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, and B. Wold. Mapping and quantifying mammalian transcriptomes by rna-seq. *Nat Methods*, 5(7):621–8, 2008. URL https://doi.org/10.1038/nmeth.1226.

T. Mulgan. Superintelligence: Paths, dangers, strategies. *The Philosophical Quarterly*, 2016. URL https://doi.org/10.1093/pq/pqv034.

P. Müller, M. Wolfien, K. Ekat, C. I. Lang, D. Koczan, et al. Rna-based strategies for cardiac reprogramming of human mesenchymal stromal cells. *Cells*, 9(2):504, Feb 2020. ISSN 2073-4409. URL http://dx.doi.org/10.3390/cells9020504.

U. Müller-Ruch, A. Skorska, H. Lemcke, G. Steinhoff, and R. David. GLP: A requirement in cell therapies - perspectives for the cardiovascular field. *Advanced Drug Delivery Reviews*, apr 2020. URL https://doi.org/10.1016/j.addr.2020.04.003.

K. Musunuru, I. J. Domian, and K. R. Chien. Stem Cell Models of Cardiac Development and Disease. *Annual Review of Cell and Developmental Biology*, 26(1):667–687, nov 2010. URL https://doi.org/10.1146/annurev-cellbio-100109-103948.

Y. Nabeshima, H. Namisaki, T. Teshima, Y. Kurashige, A. Kakio, et al. Impact of a training program incorporating cardiac magnetic resonance imaging on the accuracy and reproducibility of two-dimensional echocardiographic measurements of left ventricular volumes and ejection fraction. *Cardiovascular Ultrasound*, 17(1), oct 2019. URL https://doi.org/10.1186/s12947-019-0173-z.

U. Nagalakshmi, Z. Wang, K. Waern, C. Shou, D. Raha, et al. The transcriptional landscape of the yeast genome defined by rna sequencing. *Science*, 320(5881):1344–9, 2008. URL https://dx.doi.org/10.1126%2Fscience.1158441.

D. Nishimura. Biocarta. *Biotech Software & Internet Report*, 2(3), 2001. URL https://doi.org/10.1089/152791601750294344.

I. Nookaew, M. Papini, N. Pornputtapong, G. Scalcinati, L. Fagerberg, et al. A comprehensive comparison of rna-seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in saccharomyces cerevisiae. *Nucleic Acids Res*, 40(20):10084–97, 2012. URL https://doi.org/10.1093/bib/bbt086.

Z. Obermeyer and E. J. Emanuel. Predicting the future-big data, machine learning, and clinical medicine. *New England Journal of Medicine*, 375(13):1216–1219, sep 2016. URL https://dx.doi.org/10.1056%2FNEJMp1606181.

N. A. O'Leary, M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids research*, 44(D1):D733–45, jan 2016. URL https://doi.org/10.1093/nar/gkv1189.

E. N. Olson and D. Srivastava. Molecular pathways controlling heart development. *Science*, 1996. URL https://doi.org/10.1126/science.272.5262.671.

J. A. Oravec. Artificial intelligence, automation, and social welfare: Some ethical and historical perspectives on technological overstatement and hyperbole. *Ethics and Social Welfare*, 13(1):18–32, 2019. URL https://doi.org/10.1080/17496535.2018.1512142.

B. Ostadal and P. Ostadal. Sex-based differences in cardiac ischaemic injury and protection: Therapeutic implications. *British Journal of Pharmacology*, 171(3):541–554, feb 2014. URL https://doi.org/10.1111/bph.12270.

F. Ozsolak, A. R. Platt, D. R. Jones, J. G. Reifenberger, L. E. Sass, et al. Direct rna sequencing. *Nature*, 461(7265):814–8, 2009. URL https://doi.org/10.1038/nature08390.

V. Palanisamy and R. Thirunavukarasu. Implications of big data analytics in developing healthcare frameworks ? a review. *Journal of King Saud University - Computer and Information Sciences*, 31(4):415 – 425, 2019. URL https://doi.org/10.1016/j.jksuci.2017.12.007.

L. Papp, C. P. Spielvogel, I. Rausch, M. Hacker, and T. Beyer. Personalizing medicine through hybrid imaging and medical big data analysis. *Frontiers in Physics*, 6:51, 2018. URL https://doi.org/10.3389/fphy.2018.00051.

A. N. Paradis, M. S. Gay, and L. Zhang. Binucleation of cardiomyocytes: The transition from a proliferative to a terminally differentiated state. *Drug Discovery Today*, 19(5):602–609, 2014. URL https://doi.org/10.1016/j.drudis.2013.10.019.

R. D. Peng. Reproducible research in computational science. *Science (New York, N.Y.)*, 334(6060):1226–7, dec 2011. URL https://doi.org/10.1126/science.1213847.

E. G. Peyster, A. Madabhushi, and K. B. Margulies. Advanced Morphologic Analysis for Diagnosing Allograft Rejection. *Transplantation*, 102(8):1230–1239, aug 2018. URL https://doi.org/10.1097/tp.0000000000002189.

P.-U. PGP-UK Consortium. Personal Genome Project UK (PGP-UK): a research and citizen science hybrid project in support of personalized medicine. *BMC Medical Genomics*, 11(1):108, dec 2018. URL https://doi.org/10.1186/s12920-018-0423-1.

J. Podnar, H. Deiderick, G. Huerta, and S. Hunicke-Smith. Next-Generation Sequencing RNA-Seq Library Construction. *Current Protocols in Molecular Biology*, 106:4.21.1–4.21.19, apr 2014. URL https://doi.org/10.1002/0471142727.mb0421s106.

A. Poplawski, F. Marini, M. Hess, T. Zeller, J. Mazur, and H. Binder. Systematically evaluating interfaces for RNA-seq analysis from a life scientist perspective. *Briefings in Bioinformatics*, 17(2):213–223, 06 2015. URL https://doi.org/10.1093/bib/bbv036.

M. A. Quail, M. Smith, P. Coupland, T. D. Otto, S. R. Harris, et al. A tale of three next generation sequencing platforms: comparison of ion torrent, pacific biosciences and illumina miseq sequencers. *BMC Genomics*, 13:341, 2012. URL https://doi.org/10.1186/1471-2164-13-341.

S. Raghunathan, J. F. Islas, B. Mistretta, D. Iyer, L. Shi, et al. Conversion of human cardiac progenitor cells into cardiac pacemaker-like cells. *Journal of Molecular and Cellular Cardiology*, 138:12–22, jan 2020. URL https://doi.org/10.1016/j.yjmcc.2019.09.015.

A. Rajkomar, J. Dean, and I. Kohane. Machine Learning in Medicine. *New England Journal of Medicine*, 380(14):1347–1358, apr 2019. URL https://doi.org/10.1056/nejmra1814259.

S. Ranganathan, M. R. Gribskov, K. Nakai, and C. Scho?nbach. Encyclopedia of bioinformatics and computational biology. *Elsevier*, 2019. ISSN 0128114142.

C. Rimmbach, J. J. Jung, and R. David. Generation of murine cardiac pacemaker cell aggregates based on ES-cell-programming in combination with Myh6-promoter-selection. *Journal of visualized experiments : JoVE*, page e52465, feb 2015. URL https://doi.org/10.3791/52465.

M. D. Robinson, D. J. McCarthy, and G. K. Smyth. edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1): 139–40, 2010. URL https://doi.org/10.1093/bioinformatics/btp616.

T. Robyns, J. Breckpot, D. Nuyens, B. Vandenberk, A. Corveleyn, et al. Clinical and ECG variables to predict the outcome of genetic testing in hypertrophic cardiomyopathy. *European Journal of Medical Genetics*, 63(3):103754, mar 2020. URL https://doi.org/10.1016/j.ejmg.2019.103754.

G. Romiti, R. Cangemi, F. Toriello, E. Ruscio, S. Sciomer, et al. Sex-Specific Cut-Offs for High-Sensitivity Cardiac Troponin: Is Less More? *Cardiovascular therapeutics*, 2019, 2019. URL https://doi.org/10.1155/2019/9546931.

C. Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019. URL https://doi.org/10.1038/s42256-019-0048-x.

P. S. Russo, G. R. Ferreira, L. E. Cardozo, M. C. Bürger, R. Arias-Carrasco, et al. CEMiTool: A Bioconductor package for performing comprehensive modular co-expression analyses. *BMC Bioinformatics*, 19(1):56, feb 2018. URL https://doi.org/10.1186/s12859-018-2053-1.

A. Saha, Y. Kim, A. D. H. Gewirtz, B. Jo, C. Gao, et al. Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome research*, 27(11): 1843–1858, 2017. URL https://doi.org/10.1101/gr.216721.116.

M. Sahara, F. Santoro, and K. R. Chien. Programming and reprogramming a human heart cell. *The EMBO Journal*, 34(6):710–738, mar 2015. URL https://doi.org/10.15252/embj.201490563.

G. K. Sandve, A. Nekrutenko, J. Taylor, and E. Hovig. Ten simple rules for reproducible computational research. *PLoS Comput Biol*, 9(10):e1003285, 2013. URL https://doi.org/10.1371/journal.pcbi.1003285.

M. Scharm, F. Wendland, M. Peters, M. Wolfien, T. Theile, and D. Waltemath. The CombineArchiveWeb application - A web based tool to handle files associated with

modelling results. *CEUR Workshop Proceedings*, 1320, 2014. ISSN 16130073. URL http://ceur-ws.org/Vol-1320/paper_19.pdf.

M. Scharm, O. Wolkenhauer, M. Jalili, and A. Salehzadeh-Yazdi. GEMtractor: extracting views into genome-scale metabolic models. *Bioinformatics*, 36(10):3281–3282, 01 2020. URL https://doi.org/10.1093/bioinformatics/btaa068.

N. Schaum, J. Karkanias, N. F. Neff, A. P. May, S. R. Quake, et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*, 562(7727):367–372, oct 2018. URL https://doi.org/10.1038/s41586-018-0590-4.

W. L. Schulz, T. Durant, A. J. Siddon, and R. Torres. Use of application containers and workflows for genomic data analysis. *Journal of Pathology Informatics*, 7(1):53, 2016. URL https://doi.org/10.4103/2153-3539.197197.

I. Semenov, R. Osenev, S. Gerasimov, G. Kopanitsa, D. Denisov, and Y. Andreychuk. Experience in Developing an FHIR Medical Data Management Platform to Provide Clinical Decision Support. *International Journal of Environmental Research and Public Health*, 17(1):73, dec 2019. URL https://doi.org/10.3390/ijerph17010073.

F. Seyednasrollah, A. Laiho, and L. L. Elo. Comparison of software packages for detecting differential expression in rna-seq studies. *Brief Bioinform*, 2013. URL https://doi.org/10.1093/bib/bbt086.

P. Shah, F. Kendall, S. Khozin, R. Goosen, J. Hu, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. *Nature Digital Medicine*, 2019. URL https://doi.org/10.1038/s41746-019-0148-3.

K. Shameer, M. A. Badgeley, R. Miotto, B. S. Glicksberg, J. W. Morgan, and J. T. Dudley. Translational bioinformatics in the era of real-time biomedical, health care and wellness data streams. *Briefings in bioinformatics*, 18(1):105–124, 2017. doi: 10.1093/bib/bbv118. URL https://doi.org/10.1093/bib/bbv118.

X. Shao, J. Liao, X. Lu, R. Xue, N. Ai, and X. Fan. scCATCH: Automatic Annotation on Cell Types of Clusters from Single-Cell RNA Sequencing Data. *iScience*, 23(3):100882, mar 2020. URL https://doi.org/10.1016/j.isci.2020.100882.

M. E. Silverman and A. Hollman. Discovery of the sinus node by keith and flack: on the centennial of their 1907 publication. *Heart*, 93(10):1184–1187, 2007. URL http://dx.doi.org/10.1136/hrt.2006.105049.

J. Simkin and A. W. Seifert. Concise review: Translating regenerative biology into clinically relevant therapies: Are we on the right path? *Stem Cells Translational Medicine*, 7(2):220–231, 2018. doi: 10.1002/sctm.17-0213. URL https://doi.org/10.1002/sctm.17-0213.

D. N. Slenter, M. Kutmon, K. Hanspers, A. Riutta, J. Windsor, et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic acids research*, 46(D1):D661–D667, jan 2018. URL https://doi.org/10.1093/nar/gkx1064.

E. J. Sontheimer and R. W. Carthew. Silence from within: Endogenous siRNAs and miRNAs. *Cell*, 122(1):9–12, jul 2005. URL https://doi.org/10.1016/j.cell.2005.06.030.

D. Später, M. K. Abramczuk, K. Buac, L. Zangi, M. W. Stachel, et al. A HCN4+ cardiomyogenic progenitor derived from the first heart field and human pluripotent stem cells. *Nature Cell Biology*, 15(9):1098–1106, sep 2013. URL https://doi.org/10.1038/ncb2824.

O. Spjuth, E. Bongcam-Rudloff, J. Dahlberg, M. Dahlö, A. Kallio, et al. Recommendations on e-infrastructures for next-generation sequencing. *GigaScience*, 5:26, jun 2016. URL https://doi.org/10.1186/s13742-016-0132-7.

G. Steinhoff, J. Nesteruk, M. Wolfien, J. Groβe, U. Ruch, et al. Stem cells and heart disease - Brake or accelerator? *Advanced Drug Delivery Reviews*, 2017a. URL https://doi.org/10.1016/j.addr.2017.10.007.

G. Steinhoff, J. Nesteruk, M. Wolfien, G. Kundt, J. Börgermann, et al. Cardiac Function Improvement and Bone Marrow Response -Outcome Analysis of the Randomized PERFECT Phase III Clinical Trial of Intramyocardial CD133+ Application After Myocardial Infarction. *EBioMedicine*, 2017b. URL https://doi.org/10.1016/j.ebiom.2017.07.022.

Z. D. Stephens, S. Y. Lee, F. Faghri, R. H. Campbell, C. Zhai, et al. Big Data: Astronomical or Genomical? *PLOS Biology*, 13(7), jul 2015. URL https://doi.org/10.1371/

journal.pbio.1002195.

M. Sultan, M. H. Schulz, H. Richard, A. Magen, A. Klingenhoff, et al. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*, 321(5891):956–60, 2008. URL https://doi.org/10.1126/science.1160342.

R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker. An overview of clinical decision support systems: benefits, risks, and strategies for success. *npj Digital Medicine*, 3(1):1–10, dec 2020. URL https://doi.org/10.1038/s41746-020-0221-y.

R. Tanaka, H. Satoh, M. Moriyama, K. Satoh, Y. Morishita, et al. Intronic u50 small-nucleolar-rna (snorna) host gene of no protein-coding potential is mapped at the chromosome breakpoint t(3;6)(q27;q15) of human b-cell lymphoma. *Genes Cells*, 5(4):277–87, 2000. URL https://doi.org/10.1046/j.1365-2443.2000.00325.x.

B. Tjaden, S. S. Goodwin, J. A. Opdyke, M. Guillier, D. X. Fu, et al. Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Research*, 34(9):2791–2802, 01 2006. URL https://doi.org/10.1093/nar/gkl356.

E. J. Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1):44–56, jan 2019. URL https://doi.org/10.1038/s41591-018-0300-7.

K. Trachana, R. Bargaje, G. Glusman, N. D. Price, S. Huang, and L. E. Hood. Taking systems medicine to heart. *Circulation Research*, 122(9):1276–1289, 2018. URL https://doi.org/10.1161/CIRCRESAHA.117.310999.

C. M. Tracy, A. E. Epstein, D. Darbar, J. P. DiMarco, S. B. Dunbar, et al. 2012 accf/aha/hrs focused update of the 2008 guidelines for device-based therapy of cardiac rhythm abnormalities. *Circulation*, 126(14):1784–1800, 2012. URL https://doi.org/10.1161/cir.0b013e3182618569.

A. Traister, R. Patel, A. Huang, S. Patel, J. Plakhotnik, et al. Cardiac regenerative capacity is age- and disease-dependent in childhood heart disease. *PLoS ONE*, 13(7), jul 2018. URL https://doi.org/10.1371/journal.pone.0200342.

A. K. Triantafyllidis and A. Tsanas. Applications of machine learning in real-life digital health interventions: Review of the literature. *J Med Internet Res*, 21(4):e12286, Apr 2019. URL https://doi.org/10.2196/12286.

F. Uellendahl-Werth, M. Wolfien, A. Franke, O. Wolkenhauer, and D. Ellinghaus. A benchmark of hemoglobin blocking during library preparation for mRNA-Sequencing of human blood samples. *Scientific Reports*, 10(1):1–10, dec 2020. URL https://doi.org/10.1038/s41598-020-62637-0.

M. van den Boogaard, L. Y. Wong, F. Tessadori, M. L. Bakker, L. K. Dreizehnter, et al. Genetic variation in t-box binding element functionally affects scn5a/scn10a enhancer. *J Clin Invest*, 122(7):2519–30, 2012. URL https://doi.org/10.1172/jci62613.

L. van der Maaten and G. Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008. ISSN 1533-7928.

K. Verma. Base of a Research: Good Clinical Practice in Clinical Trials. *J Clin Trials*, 3: 128, 2013. URL http://dx.doi.org/10.4172/2167-0870.1000128.

M. Vidal, M. E. Cusick, and A.-L. Barabási. Interactome networks and human disease. *Cell*, 144(6):986–98, mar 2011. URL https://doi.org/10.1016/j.cell.2011.02.016.

A. A. Vinks, R. W. Peck, M. Neely, and D. R. Mould. Development and Implementation of Electronic Health Record?Integrated Model?Informed Clinical Decision Support Tools for the Precision Dosing of Drugs. *Clinical Pharmacology & Therapeutics*, 107(1):129–135, jan 2020. URL https://doi.org/10.1002/cpt.1679.

D. Visvikis, C. Cheze Le Rest, V. Jaouen, and M. Hatt. Artificial intelligence, machine (deep) learning and radio(geno)mics: definitions and nuclear medicine imaging applications. *European Journal of Nuclear Medicine and Molecular Imaging*, 46(13):2630–2637, dec 2019. URL https://doi.org/10.1007/s00259-019-04373-w.

L. Vogel. Plan needed to capitalize on robots, ai in health care. *CMAJ*, 189(8):E329–E330, 2017. doi: 10.1503/cmaj.1095395. URL https://doi.org/10.1503/cmaj.1095395.

D. Waltemath, J. Karr, F. Bergmann, V. Chelliah, M. Hucka, et al. Toward Community Standards and Software for Whole-Cell Modeling. *IEEE Transactions on Biomedical*

Engineering, 63(10), 2016. URL https://doi.org/10.1109/tbme.2016.2560762.

K. Wang, F. Liu, L. Y. Zhou, B. Long, S. M. Yuan, et al. The long noncoding rna chrf regulates cardiac hypertrophy by targeting mir-489. *Circ Res*, 114(9):1377–88, 2014. URL https://doi.org/10.1161/circresaha.114.302476.

Y. Wang, F. Xu, J. Ma, J. Shi, S. Chen, et al. Effect of stem cell transplantation on patients with ischemic heart failure: A systematic review and meta-analysis of randomized controlled trials. *Stem Cell Research*, 10(1):125, apr 2019. URL https://doi.org/10.1186/s13287-019-1214-0.

Z. Wang, M. Gerstein, and M. Snyder. Rna-seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10(1):57–63, 2009. URL https://doi.org/10.1038/nrg2484.

Z. Wang, Z. Wang, J. Liu, and H. Yang. Long non-coding RNA SNHG5 sponges miR-26a to promote the tumorigenesis of osteosarcoma by targeting ROCK1. *Biomedicine and Pharmacotherapy*, 107:598–605, nov 2018. URL https://doi.org/10.1016/j.biopha.2018.08.025.

H. J. Warraich, R. M. Califf, and H. M. Krumholz. The digital transformation of medicine can revitalize the patient-clinician relationship. *Nature Digital Medicine*, 1(1), dec 2018. URL https://doi.org/10.1038/s41746-018-0060-2.

J. N. Weinstein, E. A. Collisson, G. B. Mills, K. R. Shaw, B. A. Ozenberger, et al. The cancer genome atlas pan-cancer analysis project. *Nature Genetics*, 45(10):1113–1120, oct 2013. URL https://doi.org/10.1038/ng.2764.

S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLOS ONE*, 12(4):1–14, 04 2017. URL https://doi.org/10.1371/journal.pone.0174944.

C. Wiese, T. Grieskamp, R. Airik, M. T. Mommersteeg, A. Gardiwal, et al. Formation of the sinus node head and differentiation of sinus node myocardium are independently regulated by tbx18 and tbx3. *Circ Res*, 104(3):388–97, 2009. URL https://doi.org/10.1161/circresaha.108.187062.

Wilkinson, Meredyth G Ll, Radziszewska, Anna, Wincup, Chris, Ioannou, Yiannis, Isenberg, David A, et al. Using peripheral blood immune signatures to stratify patients with adult and juvenile inflammatory myopathies. *Rheumatology*, 59(1):194–204, jan 2020. URL https://doi.org/10.1093/rheumatology/kez252.

A. Wolff, M. Bayerlová, J. Gaedcke, D. Kube, and T. Beißbarth. A comparative study of RNA-Seq and microarray data analysis on the two examples of rectal-cancer patients and Burkitt Lymphoma cells. *PLOS ONE*, 13(5), may 2018. URL https://doi.org/10.1371/journal.pone.0197162.

M. Wolfien, C. Rimmbach, U. Schmitz, J. Jung, S. Krebs, et al. TRAPLINE: A standardized and automated pipeline for RNA sequencing data analysis, evaluation and annotation. *BMC Bioinformatics*, 17(1), 2016. URL https://doi.org/10.1186/s12859-015-0873-9.

M. Wolfien, D. L. Brauer, A. Bagnacani, and O. Wolkenhauer. Workflow Development for the Functional Characterization of ncRNAs. *Methods in Molecular Biology, book series*, pages 111–132, 2019. URL https://doi.org/10.1007/978-1-4939-8982-9_5.

M. Wolfien, A.-M. Galow, P. Müller, M. Bartsch, R. M. Brunner, et al. Single-nucleus sequencing of an entire mammalian heart: Cell type composition and velocity. *Cells*, 9 (2):318, Jan 2020a. URL http://dx.doi.org/10.3390/cells9020318.

M. Wolfien, A.-M. Galow, P. Müller, M. Bartsch, R. M. Brunner, et al. Single nuclei sequencing of entire mammalian hearts: strain-dependent cell-type composition and velocity. *Cardiovascular Research*, 116(7):1249–1251, 04 2020b. URL https://doi.org/10.1093/cvr/cvaa054.

M. Wolfien, D. Klatt, A. A. Salybekov, M. Ii, M. Komatsu-Horii, et al. Hematopoietic Stem-Cell Senescence and Myocardial Repair: Coronary Artery Disease Genotype/Phenotype Analysis of Post-MI Myocardial Regeneration Response Induced by CABG/CD133+ Bone Marrow Hematopoietic Stem Cell Treatment in RCT PERFECT Phase 3. *EBioMedicine*, jun 2020c. URL https://doi.org/10.1016/j.ebiom.2020.102862.

O. Wolkenhauer. Why model? *Frontiers in physiology*, 5:21, 2014. URL https://doi.org/10.3389/fphys.2014.00021.

O. Wolkenhauer, C. Auffray, R. Jaster, G. Steinhoff, and O. Dammann. The road from systems biology to systems medicine. *Pediatric Research*, 73(4-2):502–507, apr 2013. URL https://doi.org/10.1038/pr.2013.4.

K. Wolstencroft, O. Krebs, J. L. Snoep, N. J. Stanford, F. Bacall, et al. FAIRDOMHub: a repository and collaboration environment for sharing systems biology research. *Nucleic Acids Research*, 45(D1):D404–D407, jan 2017. URL https://doi.org/10.1093/nar/gkw1032.

P. W. Wood, J. B. Choy, N. C. Nanda, and H. Becher. Left ventricular ejection fraction and volumes: It depends on the imaging method. *Echocardiography*, 31(1):87–100, jan 2014. URL https://doi.org/10.1111/echo.12331.

J. Xia, C. D. Fjell, M. L. Mayer, O. M. Pena, D. S. Wishart, and R. E. W. Hancock. INMEX?a web-based tool for integrative meta-analysis of expression data. *Nucleic Acids Research*, 41(W1):W63–W70, 06 2013. URL https://doi.org/10.1093/nar/gkt338.

J.-H. Yang, J.-H. Li, S. Jiang, H. Zhou, and L.-H. Qu. Chipbase: a database for decoding the transcriptional regulation of long non-coding rna and microrna genes from chip-seq data. *Nucleic Acids Research*, 41(Database-Issue):177–187, 2013. URL https://doi.org/10.1093/nar/gks1060.

K. C. Yang, K. A. Yamada, A. Y. Patel, V. K. Topkara, I. George, et al. Deep RNA sequencing reveals dynamic regulation of myocardial noncoding RNAs in failing human heart and remodeling with mechanical circulatory support. *Circulation*, 129(9):1009–1021, mar 2014. URL https://doi.org/10.1161/circulationaha.113.003863.

L. Yang, M. O. Duff, B. R. Graveley, G. G. Carmichael, and L. L. Chen. Genomewide characterization of non-polyadenylated RNAs. *Genome Biology*, 12(2):R16, feb 2011. URL https://doi.org/10.1186/gb-2011-12-2-r16.

A. Yavari, M. Bellahcene, A. Bucchi, S. Sirenko, K. Pinter, et al. Mammalian γ2 AMPK regulates intrinsic heart rate. *Nature Communications*, 8(1):1–19, dec 2017. URL https://doi.org/10.1038/s41467-017-01342-5.

M. D. Young, M. J. Wakefield, G. K. Smyth, and A. Oshlack. Gene ontology analysis for rna-seq: accounting for selection bias. *Genome Biol*, 11(2):R14, 2010. URL https:

//doi.org/10.1186/gb-2010-11-2-r14.

S. Zhao, W. P. Fung-Leung, A. Bittner, K. Ngo, and X. Liu. Comparison of rna-seq and microarray in transcriptome profiling of activated t cells. *PLoS One*, 9(1):e78644, 2014. URL https://doi.org/10.1371/journal.pone.0078644.

M. Zitnik, F. Nguyen, B. Wang, J. Leskovec, A. Goldenberg, and M. M. Hoffman. Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Information Fusion*, 50:71–91, oct 2019. URL https://doi.org/10.1016/j.inffus.2018.09.012.

List of Abbreviations

aCaB	Antibiotic-selected Cardiac Body
AI	Artificial Intelligence
AMPK	Adenosine Monophosphate Activated Protein Kinase
ANOVA	Analysis of Variance
AUC	Area Under the Curve
AV	Atrioventricular
BAM	Binary Format For Storing Sequence Data
BiNGO	Biological Networks Gene Ontology Tool
bp	Base Pair
bpm	Beats Per Minute
BP	Biological Process
CDSS	Clinical Decision Support System
CM	Cardiomyocytes
CVD	Cardiovascular Disease
DAVID	Database for Annotation, Visualization, and Integrated Discovery
DE	Differential Expression
de.NBI	German Network for Bioinformatics Infrastructure
DBG	De Bruijn Graph
DL	Deep Learning
DMEM	Dulbecco's Modified Eagle Medium
EB	Embryoid Body
ESC	Embryonic Stem Cells
EDAM	Embrace Data And Methods
EHR	Electronic Health Records
FBS	Fetal Bovine Serum
FC	Fold Change
FDR	False Discovery Rate
FHIR	Fast Healthcare Interoperability Resources

FPKM	Fragments per Kilobase of Exon Model per Million Mapped Reads
Gata4	GATA-Binding Protein 4
GBA	Guilt-by-Association
GEM	Genome-scale Metabolic Models
GSEA	Gene Set Enrichment Analysis
GSES	Mouse Embryonic Stem Cell
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
GxP	Good Practice
Hand2	Heart- and Neural Crest Derivatives-expressed Protein 2
Hand2os1	Hand2, Opposite Strand 1
Hcn4	Hyperpolarization Activated Cyclic Nucleotide Gated Potassium Channel 4
IF	Impact Factor
iSaB	Induced Sinoatrial Body
iPSC	Induced Pluripotent Stem Cells
LVEF	Left Ventricular Ejection Fraction
MDS	Multidimensional Scaling
Mef2c	Myocyte Enhancer Factor 2C
Mesp1	Mesoderm Posterior 1 Homolog
ML	Machine Learning
mM	Millimolar
MRI	Magnet Resonance Imaging
MSC	Mesenchymal Stromal Cells
Myh6	Myosin Heavy Chain 6
ncRNA	Non-coding RNA
NGS	Next Generation Sequencing
Nkx2.5	Homeobox Protein NK-2 Homolog E
PMID	PubMed Identifier
Q	Phred Quality Score
qRT-PCR	Quantitative Real Time Polymerase Chain Reaction
QT	Quality Trimming
RBC	RNA Bioinformatics Center
Rik	Gene Placeholder of Riken Consortium
RNA	Ribonucleic Acid
ROC	Receiver Operating Characteristics
RPKM	Reads per Kilobase of Exon Model per Million Mapped Reads

RTC	Reference and Translation Center for Cardiac Stem Cell Therapy
SA	Sinoatrial
SD	Standard Deviation
Sh2b3	Src homology 2 Adaptor Protein 3
Snhg5	Small Nucleolar RNA Host Gene 5
Star	Spliced Transcripts Alignment to a Reference
SVM	Support Vector Machine
t-SNE	t-distributed Stochastic Neighbor Embedding
Tbx3	T-box Transcription Factor 3
Tbx5	T-box Transcription Factor 5
Tbx18	T-box Transcription Factor 18
TF	Transcription factor
TMM	Trimmed Mean of M-values
TRAPLINE	Transparent Reproducile and Applicable Pipeline
TPM	Transcripts per Million
U	Units
UMAP	Uniform Mannifold Approximation and Projection
WGCNA	Weighted Gene Co-expression Network Analysis

List of Figures

List of Tables

Curriculum Vitae

Markus Wolfien, M.Sc.

Date of birth: 15 February 1989
Nationality: German
Place of birth: Haldensleben

Education

2014-present	PhD candidate, at the Faculty of Computer Science, University of Rostock PhD thesis: "*Customized workflow development and omics data integration concepts in Systems Medicine*".
2012-2014	Master of Science, Medical Biotechnology, University of Rostock Master's thesis: "*Next generation sequencing data analysis of stem cell derived cardiomyocyte cell types*". (mark: 1.2)
2008-2012	Bachelor of Science, Biosystems engineering, Otto von Guericke University Magdeburg Bachelor's thesis: "*Verification of the translocation for RelA in Helicobacter pylori infected cells through immunofluorescence.*" (mark: 1.7)
2007-2008	Civilian service at DRK sheltered accommodation Haldensleben
2001-2007	Abitur at "Gymnasium Haldensleben" (mark: 2.4)

Work experience

2014-present	Department of Systems Biology and Bioinformatics at University of Rostock PhD student and scientific employee
2013-2014	Department of Systems Biology and Bioinformatics at University of Rostock Master's student and scientific employee within the SEMS BMBF junior research group
2012	Institute of Experimental Internal Medicine at Otto von Guericke University Magdeburg Scientific employee within the group of Cellular Infection Biology
2011-2012	Institute of Experimental Internal Medicine at Otto von Guericke University Magdeburg Bachelor's student
2009	Institute of Medical Microbiology at Otto von Guericke University Magdeburg Trainee

Project experience

2018-2021	Funded by the ESF project iRhythmics: Programming pacemaker cells for *in vitro* drug testing
2017-2019	Co-working on assessing the risk of gut-brain-cross-diseases in GB-XMap
2016-2021	Co-working on the structured analysis and integration of RNA-Seq experiments in de.STAIR
2015-2018	Household position in the Faculty of Computer Science
2015-2018	Funded by the BMBF project Collar: Utilizing and developing NGS data analysis workflows
2014-2015	Curator of the Systems Medicine Web Hub

Miscellaneous activities

2018-2021	GMDS project group leader for "Data Processing Workflows"
2018-2019	Attendance of a Science Slam in Bremen and the Rostock's eleven
2017-present	Joint holder of two patents for medical applications
2016	Poster Award at "Interdisziplinärer Förder-Kongress Junge Wissenschaft und Praxis - Medizin 4.0" in Berlin (Charité) (doi: 10.6084/m9.figshare.4029069)
2015-present	de.NBI Trainer of under-graduate & graduate students, MDs, PhDs, and Professors
2015-present	Actively involved in BMBF, DFG, and ESF grant writing
2013-present	Member of the MINT excellence program

Contributions in Peer-reviewed Publications

ORCID: 0000-0002-1887-4772

Publications are peer-reviewed and statistics are based on Scopus
(Accessed at December 14, 2020): h-index: 7, citations: 171

Publications indicated with * refer to an equal first authorship.

1. **Wolfien, M.***, Rimmbach, C.*, Schmitz, U.*, *et al.* (2016). TRAPLINE: A standardized and automated pipeline for RNA sequencing data analysis, evaluation and annotation. *BMC Bioinformatics. IF: 2.970, Citations: 16*

   I critically compared, benchmarked, and evaluated bioinformatics approaches. In particular, I developed and implemented a comprehensive data analysis workflow in Galaxy that integrates the best-performing tools for data analysis, data evaluation, and annotation. I also compared and selected databases for protein-protein interactions and miRNA-mRNA interactions because an easy integration in the Galaxy framework was necessary. I wrote the initial version of the manuscript. I am the corresponding author.

2. Waltemath, D., ..., **Wolfien, M.**, *et al.* (2016). Toward Community Standards and Software for Whole-Cell Modeling. *IEEE Transactions on Biomedical Engineering. IF: 4.491, Citations: 25*

   This manuscript evolved from a summer school about whole-cell modelling and contains contributions of more than 50 authors. I was involved in understanding and processing the transcription submodel of *M. genitalium*. I participated in the planning of the hackathon and contributed to the manuscript writing.

3. Lott, S.C.*, **Wolfien, M.***, Riege, K.*, Bagnacani, A.*, *et al.* (2017). Customized workflow development and data modularization concepts for RNA-Sequencing and

metatranscriptome experiments. *Journal of Biotechnology. IF: 3.142, Citations: 6*

Here, I was responsible for Section 2 about the growing importance for well-suited, correctly developed scientific workflows. In addition, I worked on the part about developing better workflows as a bioinformatician for the end user, as well as the technical illustration of cloud computing frameworks. I drafted the first version(s) of Fig.1 and Fig.2 and critically revised the manuscript.

4. Gruening, B.A., ..., **Wolfien, M.**, *et al.* (2017). The RNA workbench: best practices for RNA and high-throughput sequencing bioinformatics in Galaxy. *Nucleic Acids Research. IF: 10.727, Citations: 21*

   I contributed to this article by incorporating my experience gained from the development of TRAPLINE. In particular, I revised and tested workflows for the RNA-Seq data analysis and the underlying training materials. I also contributed in revising the paper, figures, and reviewer comments.

5. Hausburg, F., Jung, J.J., Hoch, M., **Wolfien, M.**, *et al.* (2017). (Re-)programming of subtype specific cardiomyocytes. *Advanced Drug Delivery Reviews. IF: 16.361, Citations: 6*

   Here, I applied and transfered the RNA-Seq data analysis workflow TRAPLINE to a cardiac use case, in which different stem cell-derived cardiomyocyte subtypes are compared and characterized. I developed the systems-based data analysis workflow in Cytoscape to identify enriched subnetworks. I incorporated and discussed the subsequent results in Section 4.4.6 and designed Fig.4.

6. Yavari, A., ..., **Wolfien, M.**, *et al.* (2017). Mammalian 2 AMPK regulates intrinsic heart rate. *Nature Communications. IF: 11.880, Citations: 10*

   I applied TRAPLINE on RNA-Seq data of murine SA cells and visualized the differentially expressed transcripts as a network (Fig.3). To investigate the co-expression of 2 AMPK, I conducted a weighted gene co-expression network analysis, which included hierarchical clustering, construction of the topological overlap matrix, multi-dimension analysis, and a network screening analysis (Fig.4). The identified co-expression cluster, including 2 AMPK, was subsequently investigated for hub-genes and characterized via Gene Ontology and pathway enrichment analyses.

7. Steinhoff, G., Nesteruk, J., **Wolfien, M.**, *et al.* (2017). Stem cells and heart disease - Brake or accelerator? *Advanced Drug Delivery Reviews. IF: 16.361, Citations: 21*

   I discussed potential computational and best-practice standards for a proper,

reproducible data analysis strategy in clinical research and hospital settings (Section 5). I proposed a semi-automated and self-adapting processing cyclus to evolve regenerative therapies, used as an interative process of data mining from databases, next generation sequencing data processing, network approaches, and machine learning that have to be integrated into such sustainable clinical workflows (Fig.5). I critically revised the final manuscript.

8. Lang, C.I.*, **Wolfien, M.***, *et al.* (2017). Cardiac Cell Therapies for the Treatment of Acute Myocardial Infarction: A Meta-Analysis from Mouse Studies. *Cellular Physiology and Biochemistry. IF: 5.5, Citations: 18*

   I developed a univariate meta-analysis compliant workflow and conducted the statistical analyses with the *R*-package *metafor*. The statistical data analysis was solely my responsibility, in which I first did an estimator comparison on the study variance τ^2 and inconsistency I^2 (Fig.3). I also checked for study biases via funnel-plot visualizations and Egger's regression testing (Fig.4), and utilized random/fixed effects module analyses to summarize study outcomes in a forest-plot (Fig.5). Ultimately, I conducted meta-regression analyses of different subgroups to identify moderators that have a significant influence to the LVEF improvement (Tab.2). I critically revised the final manuscript.

9. Steinhoff, G., Nesteruk, J., **Wolfien, M.**, *et al.* (2017). Cardiac Function Improvement of the Randomized PERFECT Phase III Clinical Trial of Intramyocardial CD133$^+$ Application After Myocardial Infarction. *EBioMedicine. IF: 6.680, Citations: 23*

   In this study, I proposed and utilized the computational analysis strategy, in which I applied supervised and unsupervised machine learning (ML) approaches to clinical routine measurements and accompanying research parameters with respect to learning on small datasets. I applied the unsupervised ML techniques for dimensional reduction and patient clustering (Fig.5). The key result of this paper would not have been possible without my comparison and integration of several classical ML techniques for patient stratification and most important feature selection (Fig.4 and Tab.5). I also drafted the main figure for the discussion section (Fig.6) and critically revised the manuscript.

10. Batut, B., ..., **Wolfien, M.**, *et al.* (2018). Community-driven Data Analysis Training for Biology. *Cell Systems. IF: 8.640, Citations: 7*

    This manuscript evolved from a community effort to built an extensive

training material for computational data analysis for the Galaxy framework. I contributed to the Galaxy trainings about RNA-Seq data analyses, as well as "*Quality control*" and "*Mapping*". I contributed to the manuscript writing.

11. **Wolfien, M.**, Brauer, D. L., Bagnacani, A., and Wolkenhauer, O. (2019). Workflow development for the functional characterization of ncRNAs. *Methods in Molecular Biology. Downloads: 1,100; Citations: 2*

    In this book chapter, I developed an experimental and computational strategy to identify and functionally characterize ncRNAs based on RNA-Seq data, as well as further database information (Fig.2). This strategy includes analyses from transcriptome-wide association studies, guilt-by-association, molecular network analyses, and artificial intelligence guided predictions. I refer to such an integration of diverse tools as *connective workflow* development. I structured the text and designed all figures in the article. I am the corresponding author of the chapter.

12. Bagnacani, A., **Wolfien, M.**, and Wolkenhauer, O. (2019). Tools for Understanding miRNA-mRNA Interactions for reproducible RNA Analysis. *Methods in Molecular Biology. Downloads: 978, Citations: 0*

    In this work, I defined use cases for miRNA-mRNA interactions as well as selections of relevant tools or combinations of RNA-Seq analyses into workflows. For this reason, I graded different workflows based not only on the achieved results, but on its accessibility, ease-of-use, and applicabilty for an RNA-Seq test case scenario.

13. Salehzadeh-Yazdi, A., **Wolfien, M.**, and Wolkenhauer, O. (2019). Applications of Genome-Scale Metabolic Models and Data Integration in Systems Medicine. *Focus on Systems Theory and Research. Citations: 0*

    In this book chapter, I contributed reasons about the integration of transcriptomics data into Genome-Scale Metabolic Models (GEMs) by comparing different algorithms and classification criteria. I also provided examples of GEMs used in systems medicine approaches for the identification of biomarkers and drug targets in metabolism-related disorders (e.g., obesity, and aging). I contributed to Fig.1 and critically revised the manuscript.

14. **Wolfien M.***, Galow A.M.*, Müller P.*, *et al.* (2020). Single Nuclei Sequencing of an entire Mammalian Heart: Cell Type Composition and Velocity . *Cells. IF: 5.656, Citations: 1*

    In this publication, I developed the computational workflow for the analysis

of single-cell RNA-Seq data. In particular, a genomic index for the murine mm10 built was generated and uploaded on Zenodo (https://zenodo.org/record/3623148). Subsequently, the nuclei were clustered and the RNA-velocity calculation was applied to investigate the RNA kinetics. I generated Fig.1, Fig.2, as well as the online material at FairdomHub (https://doi.org/10.15490/fairdomhub.1.study.713.1). I drafted and critically revised the manuscript.

15. Hahn O., Ingwersen L.C., Soliman A., Hamed M., Fuellen G., **Wolfien M.**, *et al.* (2020). TGF-ß1 induces changes in the energy metabolism of white adipose tissue-derived human adult mesenchymal stem/stromal cells in vitro. *Metabolites. IF: 3.303, Citations: 0*

    In this work, I analyzed and visualized the gene expression microarray data in which several quality control and processing steps were performed (e.g., normalization comparison, PCA, and DE visualization via heatmaps). I generated Fig.4 as well as Fig.6 and critically revised the final manuscript.

16. Müller P., **Wolfien M.**, Ekat K., *et al.* (2020). RNA-Based Strategies for Cardiac Reprogramming of Human Mesenchymal Stromal Cells. *Cells. IF: 5.656, Citations: 0*

    In this manuscript, I analyzed and visualized the gene expression microarray data and performed gene set, as well as pathway enrichment analyses. I generated Fig.2, Fig.5, and Fig.6. I critically revised the final manuscript.

17. **Wolfien M.***, Galow A.M.*, Müller P.*, *et al.* (2020). Single Nuclei Sequencing of entire Mammalian Hearts: Strain-dependent Cell Type Composition and Velocity. *Cardiovascular Research. IF: 7.014, Citations: 0*

    In this manuscript, I evaluated and refined the computational workflow for the analysis of single-cell RNA-Seq data that was previously published in January 2020. In particular, I extended the input data sets and integrated data obtained from a different mice strain for an in-depth comparison, *i.e.*, Fzt:DU and BL6. I visualized Fig.1 and generated the online material at FairdomHub (https://doi.org/10.15490/fairdomhub.1.assay.1227.1). I drafted and critically revised the manuscript.

18. Uellendahl-Werth F., **Wolfien M.**, Franke A., Wolkenhauer O., Ellinghaus D. (2020). A benchmark of hemoglobin blocking during library preparation for mRNA-Sequencing of human blood samples. *Scientific Reports. IF: 4.011, Citations: 0*

In this work, I supported the bioinformatics analyses with regards to the RNA-Seq benchmark and underlying computational workflow development strategy. I helped interpreting the obtained data and critically revised the manuscript

19. Galow A.M.*, **Wolfien M.***, Müller P., *et al.* (2020). Integrative cluster analysis of whole hearts reveals proliferative cardiomyocytes in adult mice. *Cells. IF: 5.656, Citations: 0*

    Here, I provided the basic data analysis template and contributed to the visualization of all figures. I critically revised the manuscript.

20. **Wolfien M.**, Klatt D., Salybekov A.A., *et al.* (2020). Hematopoietic Stem-Cell Senescence and Myocardial Repair. *EBioMedicine. IF: 6.680, Citations: 0*

    In this publication, I was the main investigator of the RNA-Seq, ML, and correlation analyses for the murine and patient data. In addition, I performed the gene co-expression, mutational profiling analysis, and unsupervised clustering of the patient data. The final ML stratification and feature selection algorithm was optimized by me, which resulted in an improved prediction accuracy of about 96%. I contributed to all Figures and contributed in planning and critically revising this manuscript.

List of Filed Patents

1. **Methods for an optimized generation of iSaBs (2017)**
   Based on the significantly differential expressed ncRNAs, transcription, and surface factors that have been identified via the TRAPLINE workflow, we filed an international patent entitled "*Method for producing sinoatrial node cells (cardiac pacemaker cells) from stem cells, and method for purifying sinoatrial node cells produced from stem cells*" (WO2017108895A1).

2. **A device to stratify patients for a cardiac response therapy (2019)**
   Clinical routine measurements and molecular data of the Phase III study PERFECT were investigated via ML algorithms. Feature-selection identified a biomarker signature for responsive patients prior to stem-cell therapy and, thus, was filed as an international patent entitled "*Method for prediction of response to disease therapy*" (PCT/EP2019/077650).

List of Given Trainings

1. Bagnacani A., **Wolfien M.**, Wolkenhauer O.: *RNA-Seq data analysis with Galaxy for clinical applications.* (Conference tutorial) GMDS (2019), Dortmund

2. Bagnacani A., **Wolfien M.**, Lott S., Riege C., Hess W., Hoffmann S., Wolkenhauer O.: *Galaxy for linking bisulfite sequencing with RNA sequencing.* (Three-day training) de.NBI (2019), Rostock

3. **Wolfien M.**, Bagnacani A., Wolkenhauer O.: *RNA-Seq data analysis with Galaxy for clinical applications.* (Conference tutorial) GMDS (2018), Osnabrück

4. Riege C., Bagnacani A., **Wolfien M.**, Lott S., Hess W., Wolkenhauer O., Hoffmann S.: *A Primer for RNA-sequencing Processing, Interpreting, and Visualization.* (Three-day training) de.NBI (2018), Jena

5. **Wolfien M.**, Bagnacani A., Wolkenhauer O.: *Introduction to RNA-Seq analysis with Galaxy.* (One-day training) de.NBI (2018), Kiel

6. Lott S., Riege C., **Wolfien M.**, Bagnacani A., Wolkenhauer O., Hoffmann S., Hess W.: *A Primer for RNA-sequencing Processing, Interpreting, and Visualization.* (Three-day training) de.NBI (2017), Freiburg

7. **Wolfien M.**, Bagnacani A., Wolkenhauer O.: *Using Next Generation Sequencing Data Analysis in the Clinic.* (Conference tutorial) CASyM Winter School (2017), Ljubljana (Slowenien)

8. **Wolfien M.**, Wolkenhauer O.: *Processing and Usage of Next Generation Sequencing Data Analysis in the Clinic.* (Conference tutorial) First Conference of the European Association of Systems Medicine (2016), Berlin

List of Selected Talks

1. **Wolfien M.**: *Bioinformatics data analysis and computational workflows in cardiac regeneration research.* (Conference talk) 7th Graduate Meeting of the LL & M (2020), Rostock

2. **Wolfien M.**, Wolkenhauer O.: *Systems biology and bioinformatics approaches in iRhythmics.* (Project meeting) iRhythmics Kickoff meeting (2019), Dummerstorf

3. **Wolfien M.**: *Deep Learning - An introduction.* (Invited talk) Annual Conference of the Northern German Community for Nuclear Medicine (2019), Hamburg

4. **Wolfien M.**: *Artificial intelligence, the terminator, and hospitals.* (Invited talk) Rostock's eleven (2019), Rostock

5. **Wolfien M.**, Steinhoff G., Wolkenhauer O.: *Reprogramming the stem cell switch for cardiac regeneration.* (Invited talk) Joint meeting for cardiac regeneration (2019), Stettin, Poland

6. **Wolfien M.**: *Künstliche Intelligenz - vom Terminator ins Krankenhaus.* (Invited talk) BMBF Science Slam, video can be accessed at[1] (2018), Bremen

7. **Wolfien M.**, Steinhoff G., Wolkenhauer O.: *Outcome analysis of the PERFECT clinical trial.* (Invited talk) Joint meeting for cardiac regeneration (2018), Tokai University Isehara, Japan

8. **Wolfien M.**, Steinhoff G., Wolkenhauer O.: *Künstliche Intelligenz für unterstützende Vorhersagen in der Medizin.* (Conference talk) Nconf (2018), Rostock

[1] https://www.youtube.com/watch?v=fy0CD7bIeI0

9. **Wolfien M.**, Steinhoff G., Wolkenhauer O.: *Automatische Prozesse dank künstlicher Intelligenz.* (Conference talk) Baltic Logistics Conference (2018), Rostock

10. **Wolfien M.**, Bagnacani A., Wolkenhauer O.: *Structured Analysis and Integration of RNA-Seq Experiments - de.STAIR* (Conference talk) 2nd NGS workshop Dummerstorf (2018), Dummerstorf

11. **Wolfien M.**: *Our hearts in machines.* (Invited talk) Posterslam - Lange Nacht der Wissenschaft (2018), Rostock

12. **Wolfien M.**, Bagnacani A., Wolkenhauer O.: *Customized workflow development and omics data integration concepts in Systems Medicine.* (Conference talk) GMDS (2017), Oldenburg

13. **Wolfien M.**, Steinhoff G., Wolkenhauer O.: *Outcome analysis of the PERFECT clinical trial.* (Invited talk) MHH, Department of Cardiothoracic, Transplantation, and Vascular Surgery Hannover (2017), Hannover

14. **Wolfien M.**, Freiesleben S., Kriehuber R., Iliakis G., Wolkenhauer O.: *NGS based analysis for structural DNA variation with applications to radiation research* (Conference talk) GBS (2016), Erlangen

15. **Wolfien M.**, Schmitz U., Wolkenhauer O.: *TRAPLINE: A standardized and automated pipeline for RNA sequencing data analysis, evaluation and annotation.* (Invited talk) Friedrich-Alexander Universität (2016), Erlangen-Nürnberg

16. **Wolfien M.**, Bagnacani A., Wolkenhauer O.: *Workflow development and data integration - de.STAIR.* (Project meeting) 2nd RBC Kickoff Meeting (2016), University of Freiburg, RNA Bioinformatics Center

17. **Wolfien M.**, Bagnacani A., Gebhardt T., Scharm M.: *A Modeler's Tale* (Invited talk) 10th International CellML Workshop, video can be accessed at[2] (2016), Auckland, New Zealand

[2] https://figshare.com/articles/A_Modeler_s_Tale/3423371

List of Supervised Theses

1. Matti Hoch (B.Sc. in Medical Biotechnology from Rostock University Medicine, 2017): "*Identifizierung von aktivierten Signalwegen in kardialen Stammzelltypen mittels Netzwerkanalyse.*"

2. Mariam Nassar (M.Sc. in Computer Science from University of Rostock, 2018): "*A Machine Learning approach to identify White Blood Cells.*"

3. David Leon Brauer (B.Sc. in Medical Biotechnology from Rostock University Medicine, in 2018): "*Functional Characterization of Long NonCoding RNAs from Stem Cell Derived Cardiomyocyte Cell Types.*"

4. Florian Uellendahl-Werth (M.Sc. in Biochemistry and Molecular Biology from University of Kiel, 2019): "*Systematic Evaluation of Globinblock and RNA-Seq Tools for Whole Blood Samples.*"

5. Maximilian Hillemanns (M.Sc. in Medical Informatics from University of Lübeck, 2020): "*Comparison of Deep Learning algorithms for 3D image analysis of cardiomyocyte cell populations.*"

List of Selected Posters

Referenz- und Translationszentrum für kardiale Stammzelltherapie Universität Rostock

SYSTEMS BIOLOGY BIOINFORMATICS ROSTOCK

Centenary Institute life saving research
Cancer. Inflammation. Cardiovascular.

TRAPLINE: An Integrated Galaxy Pipeline for RNAseq Data Processing, Evaluation and Prediction

Next Generation Sequencing (NGS) enables researchers to acquire deeper insights into cellular functions. The lack of standardized and automated methodologies poses a challenge for the analysis and interpretation of RNA sequencing data. We present a freely available, state-of-the-art bioinformatics workflow that integrates the best performing data analyses and data evaluation methods (www.sbi.uni-rostock.de/RNAseqTRAPLINE).

Current Data Analysis

Numerous software packages and pipelines have already been introduced for RNAseq analyses. Nevertheless, these tools are mainly based on programming languages like *C*, *Python* or *R* and require advanced expertise in programming for proper implementation and analysis.

What is **TRAPLINE**?

Using Galaxy, we developed an easy to use, comprehensive, **T**ransparent, **R**eproducible and **A**utomated analysis **P**ipe**LINE** for RNAseq data processing and evaluation. TRAPLINE can be accessed via the public Galaxy page of TRAPLINE and is ready to use without any restrictions.

The Galaxy Platform

Galaxy[1] is a free web-based platform for omics research that addresses the following needs:

- Accessibility
- Reproducibility
- Transparency

Multi-step analyses can be computed as *workflows*.

Data Processing

Data Pre-Processing Alignment DE

RNAseq Data

FASTQ Files
Uploaded data (Illumina, SOLiD, Solexa) will be formatted by *FASTQ Groomer* [2]

FASTQclipper [2]

Clip Seq. Adapter
Removal of preliminary adapter sequences
Minimum remaining sequence length ≥ 15 bases

FASTQ Quality Trimmer [2]

Quality Trimming
Discarding low quality reads
Q ≤ 20 starting at 5'; window/step = 1
Evaluate Read Quality, GC-Content

TopHat2 [3]

Genome Mapping
Using single or paired end reads
Default alignment options and built-in genome

Picard Toolkit [4]

Correcting Reads
Investigation of duplicated reads for read correction and SNP detection
Maximum offset = 100
Multiple Correction
SNP Calling Ready

Cufflinks2, Cuffdiff2 [5]

Ref. Annotation

Gene Expression
Calculating RPKM values and identifying sign. diff. genes with FDR ≤ 0.05, p ≤ 0.05, FC ≥ 2
RPKM Normalization, Bias Correction

Data Evaluation and Prediction

Transcriptomic Analysis Function & Network

Significant DE Genes
Lists with up and down-regulated genes

Promoters & Spliced Genes
Lists with potential splicing sites and multi-promoter genes

DAVID [5] ***miRanda*** [6] ***BioGRID*** [7]

Functional Annotation
Gene annotation with DAVID link (e.g. Go terms, Interpro)
ID = Official Gene Symbol
List Type = Gene List

miRNA Targets

miRNA Target Prediction
Using predicted binding sites to identify target genes and evaluate their impact on diff.expr. genes
Conserverd and non-conserved miRNAs

Protein Interactions

Find Protein Interactions
Using Protein-Protein interactions from BioGRID
BioGRID Organism Version 3.3.122

Functional Classification ***Compare Datasets*** ***Compare Datasets***

Enriched Genes
Genes annotated by GO terms and clustered in functionally enriched groups
Default parameters

miRNA Target Genes
List of sign. regulated miRNAs and their predicted target mRNAs

Prot.–Prot. Interactions
List of interacting proteins, between up and down-regulated mRNAs

Join Datasets

Network construction
Using the information of all data evaluation modules to easily enable the user building networks in Cytoscape or further network analysis tools (e.g. ClueGO)

Benefits of using **TRAPLINE**

Our proposed pipeline includes all relevant RNA sequencing data processing modules and is furthermore, easily applicable and needs no time consuming installation processes. Experimentalists will be able to analyze their data on their own without learning programming skills or intense computational knowledge. Accessibility and sharing the data and results are worldwide possible. Gaining quickly in-depth insights into the biology underlying the investigated data, our work for the first time introduces an automated Galaxy workflow including detailed data processing, data evaluation and prediction modules.

References

[1] Blankenberg D, Hillman-Jackson J. Analysis of next-generation sequencing data using Galaxy. Methods Mol Biol. 2014;1150:21-43.
[2] Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A. Manipulation of FASTQ data with Galaxy. Bioinformatics. 2010:1783-5.
[3] Trapnell C, Roberts A, Goff L, Pertea G, Kim D, *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012:562-78.
[4] http://broadinstitute.github.io/picard
[5] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols. 2009;4(1):44-57.
[6] Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. Nucleic Acids Res. 2008: D149-53.
[7] Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, *et al.* The BioGRID interaction database: Nucleic Acids Res. 2015: D470-8.

Markus Wolfien
Ulf Schmitz
Robert David
Olaf Wolkenhauer

www.sbi.uni-rostock.de
www.centenary.org.au
www.cardiac-stemcell-therapy.com

bit.ly/TRAPLINE_RNAseq

CombineArchiveToolkit

facilitating the transfer of research results

Universität Rostock Traditio et Innovatio

SYSTEMS BIOLOGY BIOINFORMATICS ROSTOCK

SEMS simulation experiment management system

The COMBINE Idea

The 'COmputational Modeling in Biology' NEtwork (COMBINE) is an initiative to coordinate the development of the various community standards and formats for computational models (BioPax, SBGN, SBML, SED-ML, etc.) [1]. One of the major goals of COMBINE is to improve the interoperability of these standards, and to support fledging efforts aimed at filling gaps or new needs.

The steadily increasing size and complexity of models and derived data poses the challenge of sharing reproducible results. Today, these results typically consist of multiple model files, simulation descriptions, publications, and meta data. The question how to provide all relevant files and modelling results, in a reliable and reproducible manner, remains.

In 2011 the COMBINE community [2] proposed the COMBINE archive format [3] which is a container that bundles all files related to a project into a single file. Typically, it comprises the model files needed to run a particular set of experiments. In addition, it contains all associated files that are needed to reproduce the experiments such as simulation experiment descriptions (SED-ML), semantic annotations, or graphical representations in SBGN-ML. All files can be equipped with meta-information such as people attributions and details about the files inside the archive. Generally, a COMBINE archive is encoded using the Open Modelling EXchange format (OMEX).

What's the gap?

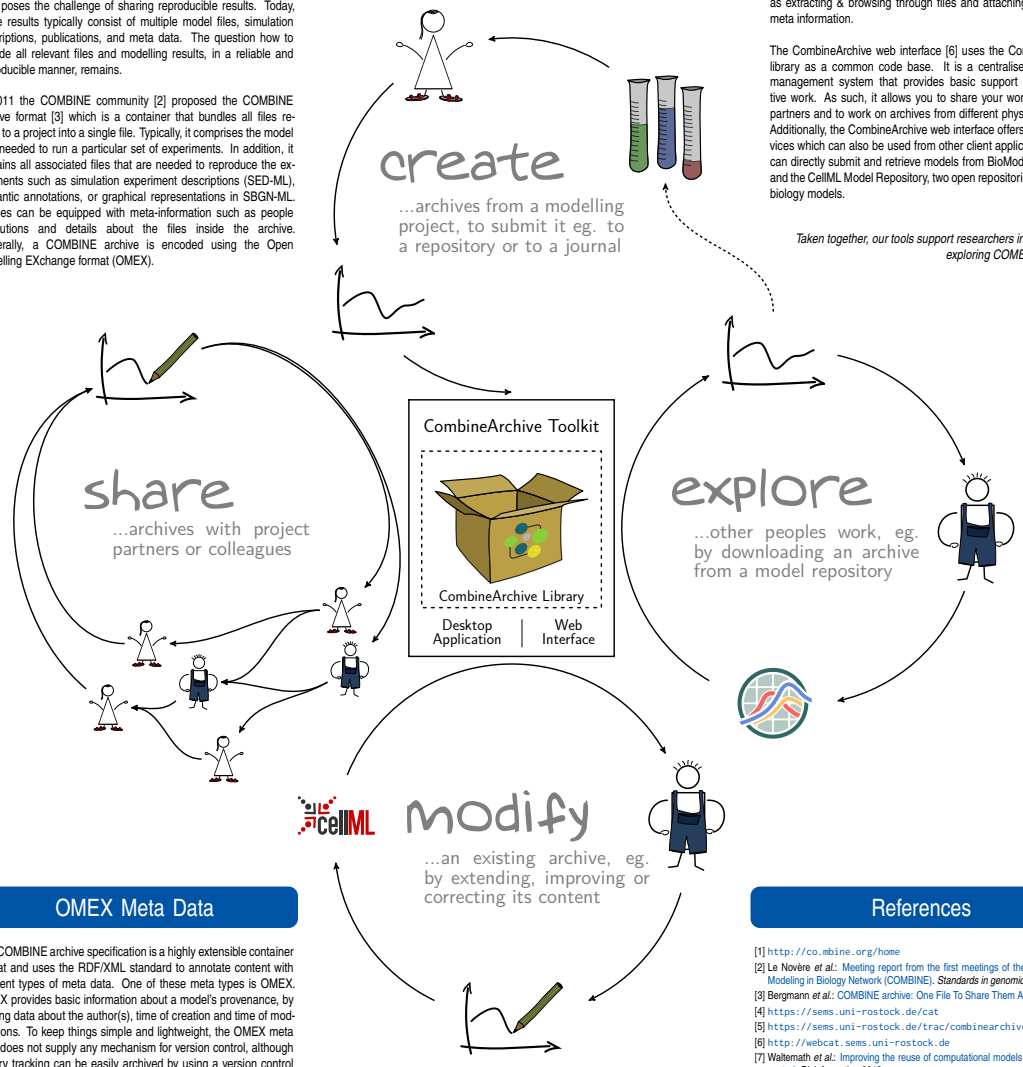
Manually handling COMBINE archives is tedious and error prone. Consequently, computational support is needed to undertake this task. Only then, it will become possible to exchange COMBINE archives seamlessly between different applications and repositories. Such a tool is constrained to provide mechanisms to create, explore and modify files and meta information in a COMBINE archive.

Our Approach

To provide the needed computational support, we developed the CombineArchiveToolkit [4]. It consists of a core library, a desktop application, and a web based interface. The CombineArchive-Library [5] was implemented using latest Java technologies.
It offers all necessary methods to handle COMBINE archives, such as extracting & browsing through files and attaching & retrieving meta information.

The CombineArchive web interface [6] uses the CombineArchive library as a common code base. It is a centralised cloud data management system that provides basic support for collaborative work. As such, it allows you to share your workspaces with partners and to work on archives from different physical locations. Additionally, the CombineArchive web interface offers RESTful services which can also be used from other client applications. Users can directly submit and retrieve models from BioModels Database and the CellML Model Repository, two open repositories of systems biology models.

Taken together, our tools support researchers in creating and exploring COMBINE archives.



OMEX Meta Data

The COMBINE archive specification is a highly extensible container format and uses the RDF/XML standard to annotate content with different types of meta data. One of these meta types is OMEX. OMEX provides basic information about a model's provenance, by holding data about the author(s), time of creation and time of modifications. To keep things simple and lightweight, the OMEX meta data does not supply any mechanism for version control, although history tracking can be easily archived by using a version control system [7].

References

[1] http://co.mbine.org/home

[2] Le Novère *et al.*: Meeting report from the first meetings of the Computational Modeling in Biology Network (COMBINE). *Standards in genomic sciences*, 2011.

[3] Bergmann *et al.*: COMBINE archive: One File To Share Them All. *arXiv*, 2014.

[4] https://sems.uni-rostock.de/cat

[5] https://sems.uni-rostock.de/trac/combinearchive

[6] http://webcat.sems.uni-rostock.de

[7] Waltemath *et al.*: Improving the reuse of computational models through version control. *Bioinformatics*, 2013.

Martin Scharm, Florian Wendland, Martin Peters, Markus Wolfien, Tom Theile, Dagmar Waltmath

http://www.sbi.uni-rostock.de

Bundesministerium für Bildung und Forschung

de.NBI
GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE

RNA Bioinformatics Center

SYSTEMS BIOLOGY BIOINFORMATICS ROSTOCK

de.NBI – German Network for Bioinformatics Infrastructure

deSTAIR

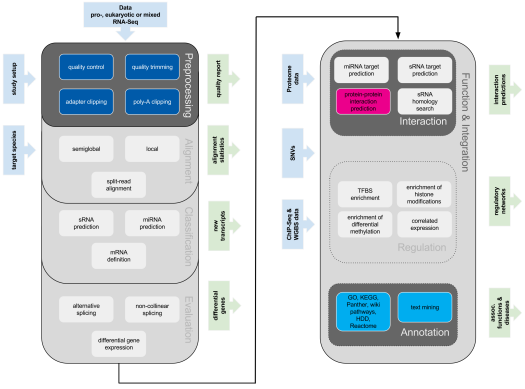
Structured Analysis and Integration of RNA-Seq Experiments

Markus Wolfien, Andrea Bagnacani and Olaf Wolkenhauer
Dept. of Systems Biology and Bioinformatics, University of Rostock Germany

Short description of the project

The variety of RNA-Seq study designs and characteristic properties of studied organisms greatly affect bioinformatics analysis. The partner project de.STAIR, a collaboration of the Universities Freiburg, Rostock and Leipzig that started on November 1st, will complement the workbench of the RNA bioinformatics center (RBC) with tools and workflows to enhance the analysis and integration of transcriptomic data from various sources. We provide workshops, training programs and screen casts for experts and non-experts to lower the barriers to RNA-Seq data analysis. In addition, the de.STAIR consortium will keep the software packages up to date and perform a customized adjustment for upcoming technologies. The Rostock site evaluates and implements customizable GALAXY workflows for the **pre-processing** of RNA-Seq data. Users will be enabled to quickly and reproducibly analyze data of various RNA-Seq protocols using **one-touch pipelines** (e.g. for mRNA-Seq, miRNA-Seq or dRNA-Seq) and to **annotate** their results with **integrated** data obtained from databases, **predictions** and **experiments**.

Project activities



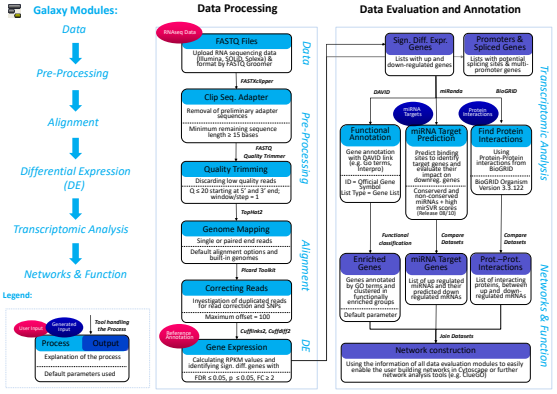
Publications

Wolfien M, Rimmbach R, Schmitz U, Jung JJ, Krebs S, Steinhoff G, David R, **Wolkenhauer O**. TRAPLINE: A standardized and automated pipeline for RNA sequencing data analysis, evaluation and annotation. *BMC Bioinformatics*. 2016. doi: 10.1186/s12859-015-0873-9.

Schmitz U, Lai X, Winter F, **Wolkenhauer O**, Vera J, Gupta S. Cooperative gene regulation by microRNA pairs and their identification using a computational workflow. *Nucleic Acid Research*, 2014. doi: 10.1093/nar/gku465.

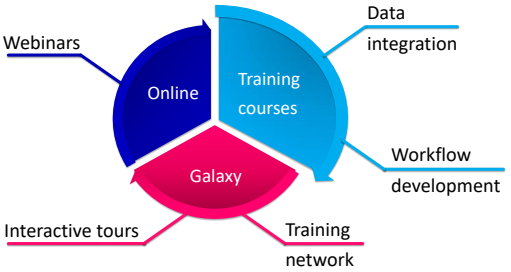
de.NBI services

- Guidance for the design of RNA-Seq experiments
- Workflow design and integration into the RBC workbench
- Providing workflow solutions, e.g. Docker/rkt containers
- Evaluation of preprocessing modules for NGS data analysis
- Facilitating the annotation and characterization of transcripts
- Data retrieval APIs for knowledge-base integration in Galaxy



de.NBI training and education

- Taking part in Galaxy online training material development – responsible curators for NGS quality control
- Training course for NGS data analysis and integration @ 1st European Association for Systems Medicine conference



sbi.uni-rostock.de/destair/

Bundesministerium für Bildung und Forschung
PTJ Projektträger Jülich Forschungszentrum Jülich
Förderkennzeichen Nr.: 031L0106C

http://www.denbi.de

Universität Rostock Traditio et Innovatio

SYSTEMS BIOLOGY BIOINFORMATICS ROSTOCK

Towards automating workflow analyses in Galaxy

Andrea Bagnacani[1], *Markus Wolfien*[1], *Martin Scharm*[1], *Olaf Wolkenhauer*[1]
[1]*Systems Biology and Bioinformatics, University of Rostock, Ulmenstr. 69, 18051 Rostock*

Introduction

The Galaxy community is promoting RNA-Seq protocols and best practices through the reuse of existing tools, and the consolidation of a Training Network to provide guidance to researchers through example datasets, tutorials, and interactive tours. However, the more tools and techniques are showcased, the more complex the options for tool chaining and parametrization become.

Objectives

- Assist researchers in carrying out their analyses
- Integrate the Galaxy framework with an interactive recommendation system leveraging on community consolidated best practices as well as EDAM annotated tools
- Promote the adoption of well established pipelines
- Allow room for experimental tools
- Consolidate protocols and reproducibility

Materials and Methods

Tool chaining and parametrization through Galaxy's interactive tours

Galaxy
galaxyproject.org

Tool operations and input / output formats through Elixir's bio.tools

EDAM
edamontology.org

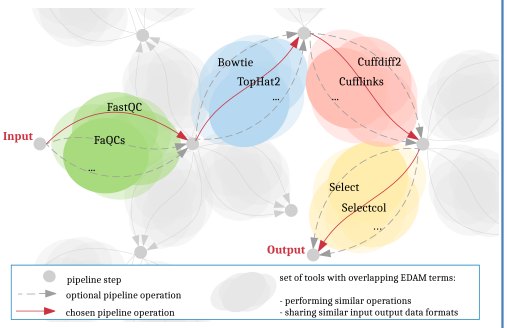
Best practices and user-tracked data of RBC's Galaxy instance

RBC
denbi.de/rbc

destairdenbi
destair.bioinf.uni-leipzig.de

deSTAIR

Results and discussion



- Galaxy tools are grouped by function
- Each tool function bridges two different states of data
- Tools are chained on their input / output data formats
- Galaxy tours recommends pertinent tools step by step
- Users decide which tool to select and parametrize

Conclusion

A recommendation system enhances the visibility of each Galaxy tool, relieving the user from browsing tool categories, or sticking to the usual analysis tools

Bioinformatics

Name (HTML)	Database search	Database search results (HTML)

Transcription factors and regulatory sites > Functional, regulatory, and non-coding RNA > RNA

RNA family identifier (Textual format)	Sequence assembly visualisation	Concentration (PNG)
Gene name (Textual format)	Data retrieval	Sequence composition plot (PNG, Dot-braket format, SVG)
Pathway or network name (Textual format)	Filtering	Database search results (CSV, JSON, HTML)
Organism name (Textual format)	Aggregation	Patway or network (PNG)
	Pathway or network visualisation	

Tool pertinence is inferred from manually curated EDAM annotations, therefore a tool's pertinence is as accurate as its bio.tools annotation

References

Lott SC et al. Customized workflow development and data modularization concepts for RNA-Sequencing and metatranscriptome experiments. Journal of Biotechnology, 2017. 10.1016/j.jbiotec.2017.06.1203

de.STAIR

Galaxy-based modular workflow generator for guided data analysis

Andrea Bagnacani[1], Konstantin Riege[2], Steffen C. Lott[3], Markus Wolfien[1], Olaf Wolkenhauer[1], Wolfgang R. Hess[3], Steve Hoffmann[2]

[1] Department of Systems Biology & Bioinformatics, University of Rostock, Rostock, Germany
[2] Computational Biology, Leibniz Institute on Aging, Jena, Germany
[3] Genetics and Experimental Bioinformatics, Faculty of Biology, University of Freiburg, Freiburg, Germany

FKZ 031L0106

The project

Shared Galaxy workflows promote the dissemination of best-practice approaches for computational Life Science analyses. Workflows are explained in their sequence of pre-selected tools, and illustrated through manually curated interactive tours, to support the adoption of well established formats, as well as provide a mean for the self-training of a new generation of data analysts. However, tool pre-selection misses to address the use of alternative computational approaches. Our proposed design overcomes this limitation, by defining alternative best-practice approaches for completing established workflow analyses, and lowering the curatorial effort needed to maintain the corresponding set of alternative interactive tours.

Progress

Workflow design

- Organization of established data analyses into modules
- Collection of modules as interchangeable blocks
- Curation of per-module interactive tours

case of differential gene expression analysis

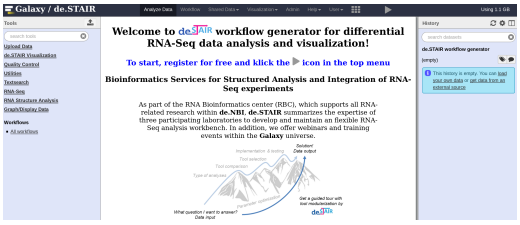
⟹ Lower curatorial effort:

Identified modules	Approaches per module	Design	Interactive tours to maintain
3	4, 4, 2	monolitic	$\prod_{i=1}^{modules} approach_i = 32$
		modular	$\sum_{i=1}^{modules} approach_i = 10$

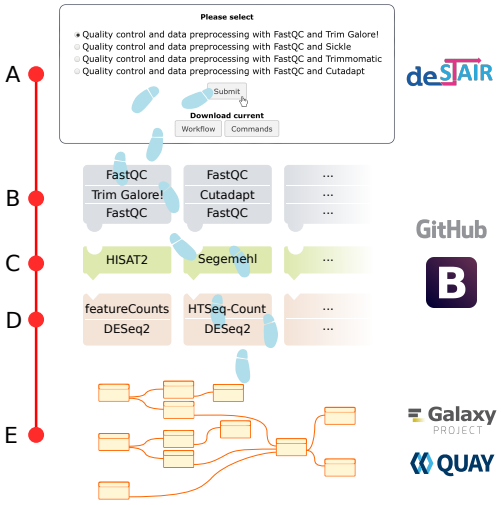
Reproducible workflows

- Development of a system to compose alternative modules
- Development of a Galaxy webhook to trigger each module
- Dockerization

⟹ Alternative best practice workflows & self-training:



Services



The system asks which approach to use (A) for the first analysis module. Upon completion, the user selects the preferred set of tools (B, C, D) to complete the analysis. The chosen route (blue trail) composes the final workflow (E), which can be exported and shared for downstream analysis.

Training and education

Support to students	Dummerstorf 29.09.2017 - 29.09.2017
A primer for RNA-Seq processing, interpreting and visualization	Freiburg 04.10.2017 - 06.10.2017
Support to students	Rostock 10.10.2017 - 12.10.2017
Introduction to RNA-Seq data analysis with Galaxy	Kiel 07.03.2018 - 07.03.2018
Support to students	Rostock 21.05.2018 - 25.05.2018
A primer for RNA-Seq processing, interpreting and visualization	Jena 27.06.2018 - 29.06.2018
RNA-Seq data analysis with Galaxy for clinical applications (GMDS 2018)	Osnabrück 04.09.2018 - 04.09.2018

About the project

Funded by de.NBI	Scientist	Andrea Bagnacani
	Scientist	Konstantin Riege
	Scientist	Steffen C. Lott
Other staff involved	Scientist	Markus Wolfien
	Scientist (PI)	Prof. Olaf Wolkenhauer
	Scientist (PI)	Prof. Wolfgang R. Hess
	Scientist (PI)	Prof. Steve Hoffmann

Publications

Lott S. C., Wolfien M., Riege K., Bagnacani A., *et al.* (2017). Customized workflow development and data modularization concepts for RNA-sequencing and metatranscriptome experiments. *Journal of biotechnology*, 261, 85–96

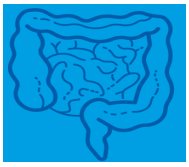
Afgan E., Baker D., Batut B., Van Den Beek M., *et al.* (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic acids research*, 46 (W1), W537–W544.

Batut B., Hiltemann S., Bagnacani A., Baker D., *et al.* (2018). Community-driven data analysis training for biology. *Cell Systems*, 6 (6), 752 – 758.e1.

GB-X Map: Causal Pathways and Motifs Common in IBD and Schizophrenia

David Ellinghaus[1], Saptarshi Bej[2], Carlo Maj[3], Markus Wolfien[2], Oleg Borisov[3], Sören Mucha[1], Per Hoffmann[4,5,6,7], Franziska Degenhardt[6,7], Andrea Bagnacani[2], Stefan Schreiber[1], Peter Michael Krawitz[3], Markus Nöthen[6,7], Olaf Wolkenhauer[2]

(1) Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany, (2) University Rostock, Institute of Computer Science, Department of Systems Biology and Bioinformatics, Ulmenstraße 69, 18057 Rostock, Germany, (3) Institute for Genomic Statistics and Bioinformatics, University Hospital Bonn, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany, (4) Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, Switzerland, (5) Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland, (6) Institute of Human Genetics, University of Bonn, Bonn, Germany, (7) Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany



Abstract: Psychiatric comorbidity in inflammatory bowel disease (IBD) is well known, with a higher incidence of schizophrenia (SCZ) in IBD cohorts compared with controls (incidence rate ratio [IRR] = 1.64) [1]. An overlap for common GWAS loci has been observed, in particular for the MHC complex on chromosome 6p21. The objective of the **GB-XMAP (Gut-brain cross-disease map)** Vernetzungsfonds project is to decipher the mechanisms of action of disease-predisposing GWAS loci shared between ulcerative colitis (UC) and SCZ. The e:Med consortia "Sysinflame" and "IntegraMent" have generated a wealth of genome-wide genotyping and gene expression data that are used for exploration.

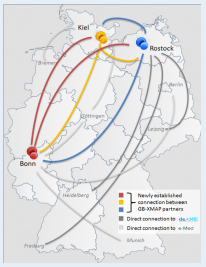
Method and data

Genome-wide association studies have identified >300 risk loci for IBD and SCZ. The GB-XMAP project, a new strategic alliance between two e:Med centres in Bonn and Kiel as well as one de.NBI node in Rostock, uses transcriptome (RNA-seq and array experiments) of whole-blood samples of 500 UC, 500 SCZ patients and 1,500 healthy controls in combination with GWAS SNP array data availabel from >20,000 UC and >30,000 SCZ patients and >79,000 healthy controls.

• Objective 1: We estimate the genetically regulated component of expression of potential risk genes of established UC and SCZ GWAS loci for a wide range of tissues by means of tissue specific cis-eQTL models followed by transcriptome wide association studies (TWAS).

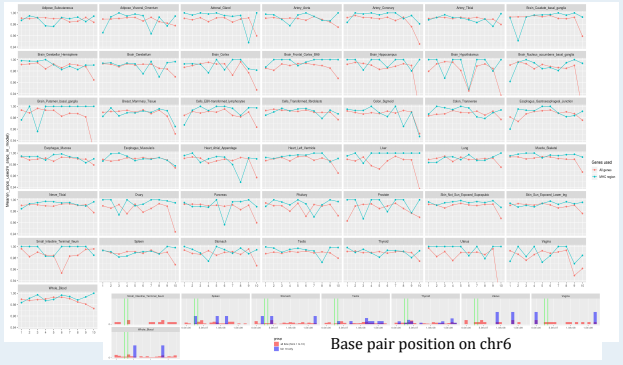
• Objective 2: We construct a single cross-disease interaction map, derive a common regulatory core, and predict disease gene signatures using in silico model simluation to create a multidimensional model.

	Genome-wide data		Transcriptome-wide data	
Technology	Immunochip	GWAS	RNA-Seq	HumanHT-12 v4
Ulcerative Colitis	14,513	6,945	500	-
Schizophrenia	1,000	>34,000	500	-
Cases total	15,513	>40,945	1000	-
Controls Kiel+Bonn	>34,000	>45,000	250+250	250+750



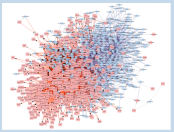
TWAS Benchmark

Expression imputation works for non-MHC/MHC genes

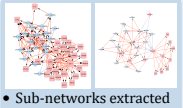


Base pair position on chr6

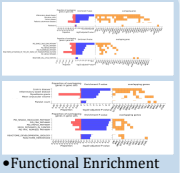
Network Analysis



• GWAS Network for IBD
• 970 Nodes
• 3802 Edges
• All feed forward loops and feedback loops involving at least two of the input genes included
•Tool used: Bisogenet [3]



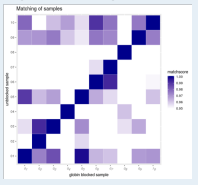
• Sub-networks extracted by MCODE Algorithm
• MCODE Algorithm extracts relatively dense motifs from a larger network
• After integrating the RNA-sequencing data with the GWAS data the networks will be updated



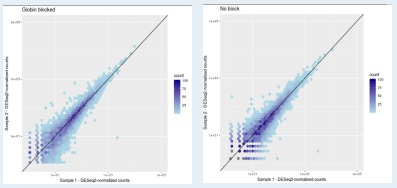
•Functional Enrichment of the motifs extracted by MCODE
• Tool used for Enrichment Analysis: FUMA-GWAS [2]

Improved RNA-Seq Analysis

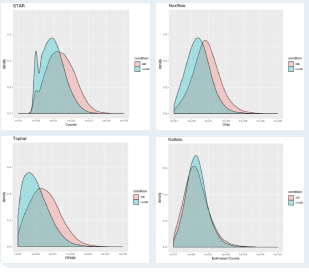
50-80% of seq. reads map to globin RNA in whole blood samples. Efficient gene-expression analysis of whole-blood samples with globin block (GB) oligos from Lexogen:



Heatmap of normalized Pearson's rho values (matchscore) shows that globin blocking (GB) RNA-seq has no impact on the correlation of "technical" replicates (GB versus non-GB). RNA-seq GB and non-GB samples from same individuals perfectly match, shown for a subset of 10 individuals.



Correlation plots showing the strong correlation of counts normalized by DESeq2 from two healthy individuals. GB does not significantly increase the variance (biological variability) or occurence of outliers in the data.



Most common RNA-seq alignment tools profit from reduced globin transcripts: More transcripts with increased counts for differential expression analysis. STAR: raw counts; Nextflow: STAR&stringTie; Tophat: FPKM (Fragments Per Kilobase Million); Kallisto: Pseudoalignment.

References:
[1] Bernstein et al. Increased Burden of Psychiatric Disorders in Inflammatory BoweDisease. Inflamm Bowel Dis. (2018). July 7 doi: 10.1093 ePub ahead
[2] Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nature Communications, 8(1), 1826
[3] Martin, A., Ochagavia, M. E., Rabasa, L. C., Miranda, J., Fernandez-de-Cossio, J., & Bringas, R. (2010). BisoGenet: a new tool for gene network building, visualization and analysis. BMC Bioinformatics, 11, 91. https://doi.org/10.1186/1471-2105-11-91

Universität Rostock Traditio et Innovatio

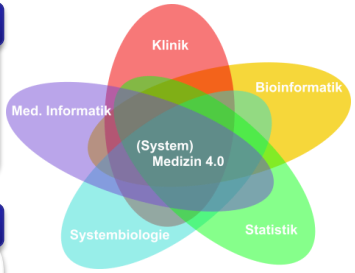
SYSTEMS BIOLOGY BIOINFORMATICS ROSTOCK

(System) Medizin 4.0 – Der direkte Nutzen interdisziplinärer Forschungsansätze

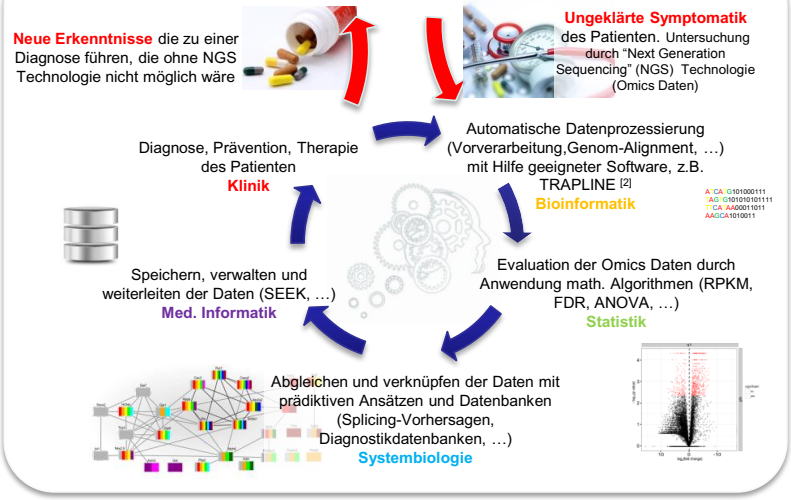
Praktische Beispiele für die katalytische Wirkung der Systemmedizin

Systemmedizin

Die Systemmedizin ist eine neue Herangehensweise, Krankheiten besser zu verstehen und zu behandeln. Dabei verknüpft sie neueste Methoden aus den Lebenswissenschaften mit Methoden aus den Informationswissenschaften und macht die resultierenden Ergebnisse für die Behandlung von Patienten in Kliniken und Arztpraxen nutzbar. Von zentraler Bedeutung sind hierbei methodische Ansätze innerhalb der Genomik, Transkriptomik und Proteomik, mit denen in Hochdurchsatz-Verfahren große Datenmengen (med. „Big Data“) und präzise Einblicke erhoben werden können, beispielsweise über das menschliche Genom [1].



„From bench to bedside” – Ein Fallbeispiel aus der Sytemmedizin



Was kann **TRAPLINE**?

Die Grundlage bildet Galaxy [3], eine freie, online basierte Plattform für Omics Analysen, die folgende Vorteile bietet:

- Verfügbarkeit
- Reproduzierbarkeit
- Annotation

Unter der Benutzung von Galaxy haben wir eine, intuitive, benutzerfreundliche, **T**ransparente, **R**eproduzierbare und **A**utomatisierte **P**ipe**LINE** für RNAseq Analysen und Auswertungen entwickelt [2]. Es sind Module für Vorverarbeitung, Genom-mapping, DE-Analysen, Splice- und Isoformdetektion, Vorhersage von miRNA & Protein-Protein Interaktionen und viele weitere Funktionen integriert. **TRAPLINE** kann u.a. über die öffentlich zugängliche Galaxywebseite gefunden und sofort benutzt werden.

bit.ly/TRAPLINE_RNAseq

Unsere Projekte in der Anwendung

Das „nachwachsende Herz“ ist zurzeit noch eine Vision. Es konnte gezeigt werden, dass sich embryonale Stammzellen von Mäusen und Menschen durch Zugabe von Wachstumsfaktoren in Herzmuskelzellen differenzieren und damit möglicherweise später einmal die abgestorbenen Herzmuskelzellen ersetzen können.

Referenz- und Translationszentrum für kardiale Stammzelltherapie Universität Rostock

http://www.cardiac-stemcell-therapy.com

Das bessere Verständnis von komplexen DNA Läsionen wird zu einer effektiveren Bewertung von Strahlenrisiken führen. Dieses Ziel soll durch die Identifikation von Parametern für die Komplexität von DNA Schäden erreicht werden, anhand derer sich Zellletalität und genomische Instabilität zuverlässig bestimmen lassen. Die Ergebnisse leisten einen direkten Beitrag hinsichtlich der individuellen Optimierung der Strahlentherapie.

Universitätsklinikum Essen

PTJ Projektträger Jülich Forschungszentrum Jülich

Europaweite Initiative

Das Bestreben, die Systemmedizin als interdisziplinäre Initiative zu etablieren und zu vernetzen, wird durch ein europäisches Konsortium, *i.e.* CaSyM, verwirklicht. Aktuelle Informationen über eine aktive Teilnahme, Jobs, Interviews, Events und laufende Projekte finden Sie auf der Onlineplattform Systems-medicine.net.

Systems Medicine.net
www.systems-medicine.net

CASYM EUROPE
www.casym.eu

Symbiose der (System) Medizin 4.0

Medizin 4.0, als globale Vernetzung der computergestützten Informationswelt mit dem patientenorientierten Gesundheitssystem, erzeugt komplexe Bedürfnisse die nur in interdisziplinären Teams, welche die Systemmedizin bereits involviert, gelöst werden können. Eine Verstärkung der Vernetzung zwischen Forschung, Patienten und Industrie sollte das gemeinsame Ziel einer besseren und persönlicheren Versorgung, Behandlung, Diagnose und Therapiemöglichkeit beinhalten. Eine schnelle Übertragung der Ergebnisse auf ein Produkt, sowie erhöhte Transparenz in der Forschung für den Patienten würde den partizipativen Charakter erhöhen. Zusätzlich würde eine bessere Ressourcenausnutzung der akademischen Strukturen stattfinden und größere Fortschritte in der Prävention und der Prädiktion von Krankheiten erzielt [4].

Literatur

[1] Schmitz U and Wolkenhauer O. Systems Medicine. *Springer*. 2016. ISBN: 978-1- 4939-3282-5

[2] Wolfien M, Rimmbach R, Schmitz U, Jung JJ, Krebs S, Steinhoff G, David R, Wolkenhauer O. TRAPLINE: A standardized and automated pipeline for RNA sequencing data analysis, evaluation and annotation. *BMC Bioinformatics*. 2016. doi: 10.1186/s12859-015-0873-9

[3] Blankenberg D, Hillman-Jackson J. Analysis of next-generation sequencing data using Galaxy. *Methods Mol Biol.* 2014;1150:21-43.

[4] Hood L and Friend SH. Predictive, personalized, preventive, participatory (P4) cancer medicine. 2011. *Nature Reviews Clinical Oncology*. doi:10.1038/nrclinonc.2010.227

Markus Wolfien,
Olaf Wolkenhauer
www.sbi.uni-rostock.de

markus.wolfien@uni-rostock.de
olaf.wolkenhauer@uni-rostock.de

Workflow available at: bit.ly/TRAPLINE_RNAseq

Universität Rostock Traditio et Innovatio

SYSTEMS BIOLOGY BIOINFORMATICS ROSTOCK

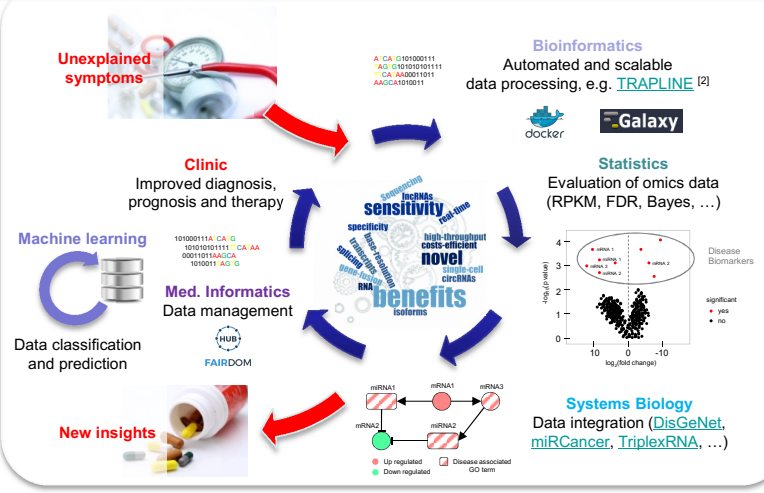
RTC

Customized workflow development and data integration concepts in Systems Medicine

Workflows: Symbiosis of research and tool integration for clinical application

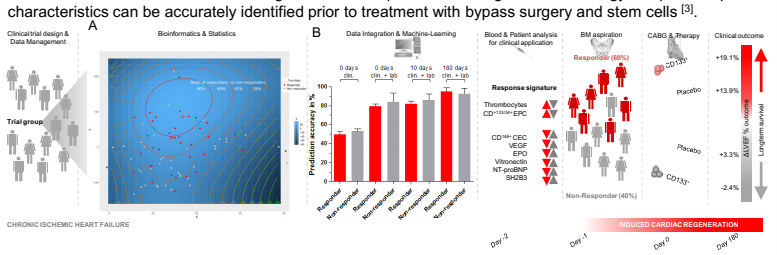
Workflow management frameworks and cloud computing services are bridging the gap between tool developers and end users, aiming towards an easy applicable and up-scalable computational data analysis. This in turn allows for an improved data reproducibility, process documentation, and monitoring of submitted jobs. Finally, workflows facilitate the use of state-of-the-art computational tools which would be hard to access for non-experts without graphical user interface frameworks. However, the use of workflows could be even more simplified for experimental and clinical researchers by strengthening the specific focus on the addressed research hypothesis and lessening the effort for the selection of the most appropriate tool [1].

Integrating data using a Systems Medicine approach



Application of the integrative workflow: The Phase III clinical trial PERFECT

Regenerative therapies using stem cells for the repair of heart tissue have been at the forefront of preclinical and clinical development during the past 16 years. To build upon this progress, the Phase III clinical trial PERFECT was designed to assess clinical safety and efficacy of intramyocardial CD133+ bone marrow stem cell treatment combined with coronary artery bypass graft (CABG) for induction of cardiac repair. The primary endpoint was delta left ventricular ejection fraction (ΔLVEF) at 180 days compared to baseline measured with magnet resonance imaging (MRI). Responders (R) classified by ΔLVEF ≥ 5% after 180 d were 60% of the patients (35/58) in both treatment groups (+17.1% in R vs. non-responders (NR). The PERFECT trial shows that cardiac tissue repair and restitution of left ventricular function can be successfully installed in ischemic heart disease by CABG surgery associated with presence of enhanced circulating CD133+ and CD34+ EPC level. Using this new computer-aided diagnostic technology, responsive patient characteristics can be accurately identified prior to treatment with bypass surgery and stem cells [3].



Outcome results of the PERFECT trial. A Three-dimensional t-SNE calculation of the Rostock subgroup. The variables x and y refer to the newly calculated features that are used to classify the patients into distinct groups. The model was subsequently fitted by a polynomial (n^5) equation to visualize the z-axis as a geographic profile. The respective colors for the responder (red dot) and non-responder (grey dot) patients have been added afterwards. The classified groups have been roughly summarized by a red and grey dashed line. Results are obtained after 3000 iterations. The calculation of the ratio between responder and non-responder is indicated for each circle. **B** Obtained supervised ML prediction results for pre- and postoperative time points (0 days to 180 days) of the clinical and clinical & laboratory dataset to distinguish between R and NR. The graph shows the true positive prediction results of feature selected ML models (AdaBoost for feature selection and RF for final prediction).The error bars indicate the respective accuracy standard deviation for the constructed models that have been obtained after 100 iterations. The 100 model iterations are significant different according to one-way ANOVA ($p < 0.001$).

Acknowledgements

de.NBI www.denbi.de

ELIXIR GERMANY elixir-europe.org

deSTAIR destair.bioinf.uni-leipzig.de/

Collar sbi.uni-rostock.de/collar

This work has been supported by de.NBI, KIT and the respective projects of de.STAIR and Collar to ensure high-quality bioinformatics in life sciences research and biomedicine. As a result, we organize training courses and summer schools on using computational tools, standards in modeling, workflow development in Galaxy and compute services with Docker.

What is **TRAPLINE**?

The basis is Galaxy [1], an open-source, online platform for omics data analyses that addresses the following needs:

- Accessibility
- Reusability
- Annotation

Using Galaxy, we developed a **T**ransparent, **R**eproducible & **A**utomated analysis **P**ipe**LINE**, named **TRAPLINE**, for RNA sequencing data processing, evaluation, annotation and prediction. The predictions are based on modules which are able to identify novel transcripts, protein-protein interactions, miRNA targets and alternatively splicing variants or promoter enriched sites. The obtained results can be visualized in a network. TRAPLINE can be accessed via the published Galaxy page or manuscript [2].

bit.ly/TRAPLINE_RNAseq

An emerging solution to deploy the workflows, including all necessary tools and dependencies, are software channels and containers like Bioconda, Docker or rkt. These containers allow the packaging of workflows in an isolated and self-contained system that simplifies the distribution. The technology combines areas from systems research, *esp.* OS virtualization, cross-platform portability, modular reusable elements, versioning, and a "DevOps" philosophy [1].

References

[1] Lott SC, **Wolfien M**, Riege K, Bagnacani A, Wolkenhauer O, Hoffmann S, Hess WR. Customized workflow development and data modularization concepts for RNA-Sequencing and metatranscriptome experiments. *Journal of Biotechnology. 2017* doi.org/10.1016/j.jbiotec.2017.06.1203

[2] **Wolfien M**, Rimmbach R, Schmitz U, Jung JJ, Krebs S, Steinhoff G, David R, Wolkenhauer O. TRAPLINE: A standardized and automated pipeline for RNA sequencing data analysis, evaluation and annotation. *BMC Bioinformatics.* 2016. doi: 10.1186/s12859-015-0873-9.

[3] Steinhoff G, Nesteruk J, **Wolfien M**, Kundt G, The PERFECT Trial Investigators. Cardiac Function Improvement and Bone Marrow Response Outcome Analysis of the Randomized Perfect Phase III Clinical Trial of Intramyocardial CD133 + Application After Myocardial Infarction. *EBioMedicine. 2017.* doi.org/10.1016/j.ebiom.2017.07.022

Markus Wolfien markus.wolfien@uni-rostock.de
Gustav Steinhoff gustav.steinhoff@med.uni-rostock.de
Olaf Wolkenhauer olaf.wolkenhauer@uni-rostock.de
www.sbi.uni-rostock.de

Get the poster at Figshare: goo.gl/fzeGu8

Bundesministerium für Bildung und Forschung