An integrative workflow to study large-scale biochemical networks

Dissertation

zur

Erlangung des akademischen Grades

Doktor-Ingenieur (Dr.-Ing.)

Promotionsgebiet Systembiologie und Bioinformatik

Fakultät für Informatik und Elektrotechnik

Universität Rostock

vorgelegt von

Faiz M. Khan, geboren am 01. May 1985 in Peshawar-Pawaka

wohnhaft in Rostock-Südstadt, 6. November 2019

https://doi.org/10.18453/rosdok_id00002554

Erstgutachter:	Prof. Olaf Wolkenhauer	University of Rostock
Gutachter:	Prof. Jan Hasenauer	University of Bonn
	Prof. Georg Füllen	University of Rostock

Tag der Verteidigung: 21.06.2019

Acknowledgements

I express my gratitude to **Prof. Olaf Wolkenhauer** for giving me the opportunity to do a PhD under his supervision. I still remember his first Systems Biology lecture in MSc., which introduced and motivated me to the field. I learned a lot from his guidance and discussions.

I am especially thankful to **Julio Vera** who brought me to this research field by offering me a master thesis project followed by a PhD position in his group. Along with Julio Vera I am also thankful to **Ulf Schmitz** and **Shailendra Gupta** with whom I worked in a team for developing research strategies and their implementations, and hypotheses generation. I enjoyed many fruitful and helpful discussions with them.

I would like to extend my thanks to our experimental collaborators **Brigitte Pützer** and her team members especially **Stephan Marquardt** and **David Engelmann** from the Institute of Experimental Gene Therapy and Cancer Research, University of Rostock, Rostock Germany.

I am grateful to present and former **SBI Rostock** group members. I'm especially thankful to **Markus Wolfien** and Ulf Schmitz for reading this thesis and giving me valuable suggestions.

My gratitude to **Rieke Scholz** for all the good time you gave me. I would extend my thanks to everyone here in Rostock who gave me company and spent beautiful moments of life. Thanks to my dance partners **Claudia Wittmann**, **Melanie Blau**, and dance teachers **Salsera Chrisina**, **Alf Derno** and **Réne Glasgow**. Thanks to **Mukhtar Ullah**, **Zahid Hassan**, **M Riaz**, **Yasin Bhai**, and **Saif Butt**.

A big thank you to my parents **Ghulam Hussain Wakeel** and **Nageena** for their unconditional-love, encouragement and support. I miss you Daji (Dad). I also thank to all my siblings, uncles, aunts and cousins for their love and support.

Abstract

The recent development in quantitative measurements and access to interaction databases facilitate the construction of detailed molecular interaction maps of cellular processes. Such networks serve as a knowledge-base and being machine readable are amenable to computational analysis. Studying biochemical networks as a non-linear dynamical system is challenging due to a large number of components and complex network structures including feedback/feedforward loops.

In this thesis, I propose an integrative workflow to study large-scale biochemical networks by combining techniques from bioinformatics and systems biology. It integrates heterogeneous sources of biological information with network structure and dynamical systems analysis to unravel mechanisms underlying diseases. To this end, we constructed a comprehensive interaction map for the transcription factor E2F1 to understand its functional role in different traits of cancer, such as drug resistance and epithelial-mesenchymal transition (EMT). The network contains 1,015 nodes and 4,180 interactions. Further, the network structural analysis revealed a large number of feedback and feedforward loops. To make such a large, complex network suitable for dynamical analysis, I developed a flexible and extendible workflow based on multi-objective optimization function. It combines network structural properties with high-throughput and biomedical data to identify smaller modules which I refer to as "core-regulatory networks" that are amenable for analysis with dynamical systems theory. Using the proposed workflow, I identified core-regulatory networks from the E2F1 molecular interaction map underlying EMT in bladder and breast cancer. I carried out dynamical analyses of the core-regulatory networks using logic-based modeling. Using *in silico* stimulus-response and perturbation experiments, I detected molecular signatures and potential drug targets for each cancer type. The *in silico* predictions were validated with patient data and through *in vitro* experiments.

Moreover, I developed a hybrid modeling framework that combines ordinary differential equation models with logic-based models as a strategy to analyze the dynamics of large-scale non-linear biological systems. Using the proposed hybrid modeling strategy, I simulated the known dynamical features of the E2F1-p73/DNp73-miR205 network in drug resistance for different concentrations of the transcription factor E2F1 and receptor molecules. Further, the results of hybrid model analyses suggest that cancer cells might become independent of growth factors when E2F1 is highly expressed.

This thesis is a contribution to interdisciplinary cancer research, providing a methodology for the analysis of large-scale networks in molecular and cell biology.

Zusammenfassung

Die jüngste Entwicklung quantitativer Messungen und der Zugang zu Interaktionsdatenbanken ermöglichen die Erstellung detaillierter molekularer Interaktionskarten zellulärer Prozesse. Solche Netzwerke dienen als Wissensbasis und sind maschinenlesbar und können rechnerisch analysiert werden. Die Untersuchung biochemischer Netzwerke als nichtlineares dynamisches System ist aufgrund einer großen Anzahl von Komponenten und komplexen Netzwerkstrukturen, einschließlich Rückkopplungs-/Vorwärtskopplungsschleifen, eine Herausforderung.

In dieser Arbeit schlage ich einen integrativen Workflow vor, um biochemische Netzwerke in großem Maßstab zu untersuchen, indem Techniken aus der Bioinformatik und der Systembiologie kombiniert werden. Es integriert heterogene biologische Informationsquellen in die Netzwerkstruktur und die Analyse dynamischer Systeme, um Mechanismen aufzudecken, die Krankheiten zugrunde liegen. Zu diesem Zweck haben wir eine umfassende Interaktionskarte für den Transkriptionsfaktor E2F1 erstellt, um seine funktionelle Rolle bei verschiedenen Krebsmerkmalen wie Arzneimittelresistenz und epithelial-mesenchymal transition (EMT) zu verstehen. Das Netzwerk enthält 1.015 Knoten und 4.180 Interaktionen. Ferner ergab die Netzwerkstrukturanalyse eine große Anzahl von Rückkopplungs- und Vorwärtskopplungsschleifen. Um ein so großes, komplexes Netzwerk für dynamische Analysen geeignet zu machen, habe ich einen flexiblen und erweiterbaren Workflow entwickelt, der auf der Funktion zur Optimierung mehrerer Ziele basiert. Es kombiniert Netzwerkstruktureigenschaften mit Hochdurchsatzdaten und biomedizinischen Daten, um kleinere Module zu identifizieren, die ich als "Kernregulierungsnetzwerke" bezeichne, die für die Analyse mit der Theorie dynamischer Systeme geeignet sind. Unter Verwendung des vorgeschlagenen Workflows identifizierte ich aus der molekularen E2F1-Wechselwirkungskarte, die der EMT bei Blasen- und Brustkrebs zugrunde liegt, zentrale regulatorische Netzwerke. Ich habe dynamische Analysen der Kernregulierungsnetzwerke mit logikbasierter Modellierung durchgeführt. In silico Stimulus-Response- und Störungsexperimenten entdeckte ich molekulare Signaturen und potenzielle Wirkstofftargets für jeden Krebstyp. Die in silico Vorhersagen wurden mit Patientendaten und durch In-vitro-Experimente validiert.

Darüber hinaus entwickelte ich ein hybrides Modellierungsgerüst, das gewöhnliche Differentialgleichungsmodelle mit logikbasierten Modellen kombiniert, um die Dynamik nichtlinearer biologischer Großsysteme zu analysieren. Unter Verwendung der vorgeschlagenen Hybridmodellierungsstrategie simulierte ich die bekannten dynamischen Merkmale des E2F1-p73 / DNp73-miR205-Netzwerks hinsichtlich der Arzneimittelresistenz für verschiedene Konzentrationen des Transkriptionsfaktors E2F1 und der Rezeptormoleküle. Ferner legen die Ergebnisse von Hybridmodellanalysen nahe, dass Krebszellen von Wachstumsfaktoren unabhängig werden könnten, wenn E2F1 stark exprimiert wird.

Diese Dissertation ist ein Beitrag zur interdisziplinären Krebsforschung und liefert eine Methodik zur Analyse von großräumigen Netzwerken in der Molekular- und Zellbiologie.

Abbreviations

ODE	ordinary differential equation
EMT	epithelial-mesenchymal transition
GRNs	gene regulatory networks
PPI	protein-protein interaction
miRNAs	microRNAs
TF	transcription factors
MIM	molecular interaction map
SBGN	systems biology graphical notation
ER	entity relationship
PD	process description
AF	activity flow
SBML	systems biology markup language
ND	node degree
BC	betweenness centrality
PD	Parkinson's disease
IAV	influenza A virus
VEGF-C	vascular endothelial growth factor-C
EGFR	epidermal growth factor receptor
FBLs	feedback loops
FFLs	feedforward loops
GP	gene prioritization
DP	disease pathway
FC	fold-change
BF	boolean function
MF	multi-valued functions
LSS	logical steady state
RARA	retinoic acid receptor alpha
ProMoT	process modeling tool
CXCR1	chemokine (C-X-C motif) receptor 1
HMMR	hyaluronan-mediated motility receptor
IL1R1	interleukin 1 receptor type I
THRB	thyroid hormone receptor beta
EGF	epidermal growth factor
FGF2	fibroblast growth factor 2
TGFB1	transforming growth factor beta-1
MM	Michaelis-Menten
cFFL	coherent feedforward loop
iFFL	in-coherent feedforward loop
CNA	CellNetAnalyzer

Table of contents

Chapter 1

Introduction

Large parts of this chapter were published in a book chapter:

- **Khan F.M.**, Sadeghi M., Gupta S.K., Wolkenhauer O. (2018) A Network-Based Integrative Workflow to Unravel Mechanisms Underlying Disease Progression. In: Bizzarri M. (eds) Systems Biology. Methods in Molecular Biology, vol 1702. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-7456-6_12.

Synopsis
In this chapter, I describe the prospects and challenges of analyzing large-scale biochemical networks to discover and characterize key regulatory mechanisms underlying complex diseases (e.g., cancer). The large number and the nonlinear nature of interactions hamper the analysis of biochemical networks. Here, I discuss methods for structural and dynamical analysis of biochemical networks and propose an integrative workflow that combines these analyses with high throughput and clinical data to unravel mechanisms underlying diseases. Finally, I outline the structure of this thesis, the objective of each chapter, and summarize the contribution of my work to the field.

1.1. Motivation and objectives

In order to investigate the complex behavior of cancers, such as tumor invasion and drug resistance, interdisciplinary collaborations often start with the gathering of information from the literature and databases, summarizing components and their interactions relevant for the processes under consideration. The machine-readable representation of this information typically leads to interaction maps composed of a large number of components that are involved in signaling, gene regulation and metabolism. These interaction maps serve as a knowledge base and are amenable for computational analysis to generate new hypotheses. Analyzing the structure of biochemical networks provides useful information including network hubs [1], regulatory motifs [2], and often global features like small world organization [3] of the system. The analysis of regulatory motifs, especially feedback and feedforward loops, can provide

novel insights into the regulatory mechanisms controlling the normal and disease states of the cell [2, 4]. Their nonlinear dynamics, however, not only challenge human intuition but also limit the application of conventional data analysis tools [5, 6]. Moreover, for large scale biochemical networks a dynamical analysis is particularly difficult with mechanistic (*e.g.,* ordinary differential equation (ODE) based) approaches from the theory of dynamical systems. In order to exploit the advantages of large-scale biochemical networks in combination with mechanistic modeling, we need integrative approaches and computational workflows that combine different systems biology and bioinformatics tools to integrate heterogeneous sources of biological information. This can allow us to identify disease specific small regulatory modules that can be subjected to a more detailed analyses, followed by the prediction of molecular signatures and therapeutic targets.

In this thesis, I developed a new integrative workflow to study large-scale biochemical networks by combining techniques from bioinformatics and systems biology, in combination with the integration of experimental and clinical data to understand and unravel mechanisms underlying diseases. In this context, I constructed a comprehensive interaction map of the transcription factor E2F1 to understand its functional role in drug resistance and epithelial-mesenchymal transition (EMT). To make such large networks suitable for dynamical analysis, I developed a flexible and extendible method that combines network structural analysis with high throughput data and biomedical data to identify smaller 'modules' for specific disease phenotypes. I hereafter refer to these as "core-regulatory networks" [7, 8] that are amenable for an analysis with dynamical systems theory. A mathematical model (*e.g.,* ODE-based, logic-based or hybrid model) of the identified smaller modules can be constructed to give mechanistic insights into diseases and propose new hypotheses, which are subjected to experimental validation. Using the proposed workflow, I analyzed the large-scale molecular interaction network of E2F1 and identified core-regulatory networks for EMT processes in bladder and breast cancers. Further, I developed logic-based models of the core-networks and using *in silico* stimulus response analyses, I detected molecular signatures that render non-invasive cancer into invasive. Furthermore, I carried out *in silico* perturbation experiments and predicted therapeutic targets that can reduce the invasive behavior of cancer cells. The model predictions were validated experimentally and through patient data.

Systems biology approach offers mathematical modeling formalisms suitable for biological systems at different abstraction levels depending on the size, structure and type of data or information available. ODE-based kinetic models are well established approaches for detailed quantitative analyses of small, nonlinear systems, while logic-based models are well appreciated for coarse-grained qualitative analyses of large systems. I developed a hybrid modeling framework that combines kinetic models with logic-based models as a strategy to model large-scale, nonlinear biochemical networks [9]. The hybrid model characterized the nonlinear dynamics of the

E2F1-p73/DNp3-miR205 network in drug resistance and revealed that cancer cells might get independent of growth factors when the transcription factor E2F1 is highly expressed.

Ultimately, the proposed integrative workflow provides a powerful foundation to decipher the mechanisms underlying complex diseases that would lead to a better prognosis and an effective therapeutic intervention.
The specific objectives of this thesis are:

- Construction of an E2F1 interaction map (Chapter 2)
- Identification of tumor-specific core regulatory networks (Chapter 2)
- Logic-based modeling of tumor invasion (Chapter 3)
- ODE-based and hybrid modeling of drug resistance (Chapter 4)

1.2. The Systems Biology Approach

Biological processes are complex, composed of a large variety and number of components that interact in a nonlinear fashion in space and time. Non-linear processes in particular induce counter-intuitive behavior, which is impossible to understand by unaided human observations. The systems biology approach combines experiments with computational tools and methods to understand such complex processes [10, 11]. We consider the systems biology approach as an interdisciplinary collaboration that realizes an iterative cycle of data-driven modeling and model-driven experimentation (Figure 1.1). A research project taking a systems biology approach, often starts by gathering information about a biological process from literature and databases. The information is organized and represented in a form of machine-readable network which is a formalized representation of a large number of individual experimental results. The network is then computationally analyzed, for example modeled (mathematically) with a certain modeling formalism to perform dynamical system analyses. After calibration with experimental data, the model should mimic the biological phenomena under consideration, and could perform stimulus response and perturbations analyses, which are used to formulate/refine hypotheses and therefore, supports the design of new experiments. If experiments justify the hypothesis, it will generate new knowledge, which may be used for developing another model. In this way, the systems biology approach cycles between data-driven modeling and model-driven experimentation. More specifically, we can divide this approach into four main stages:
(i) Setting the context: This step starts with the formulation of a biological question that is to be investigated. For example, a general question could be: "What are the regulatory mechanism(s) underlying breast tumor metastasis or drug resistance?". This defines the project boundaries and gives directions to collect information from literature and databases. The collected information is then converted into a machine readable

format (*i.e.,* in the form of a network) for computational analysis. With the help of domain experts or using computational methods, project-specific network components or modules are then chosen.

(ii) Representation of a sub-/system: Mapping out the interactions among biochemical entities, like genes, proteins, or miRNAs, as a network provides a platform for structural and dynamical analysis of the system. A large set of network visualization tools (*e.g.,* CellDesigner [12], Cytoscape [13], VANTED [14] and SBGN-ED [15]) are used to construct these networks, representing biological processes from abstract to more detailed levels depending on the requirements of the biological question and available knowledge. Based on available knowledge and the domain expert's opinion the directions of the interactions (*e.g.,* activation or inactivation) among molecular entities are defined or hypothesized. Depending on network size and kinetic details, a suitable modeling formalism is chosen to analyze the dynamics of a system for stimulus response behavior and different perturbations.

(iii) Model construction: This step starts with the detailed description of the molecular interactions providing the biochemical and biophysical information. For example, biochemical interactions characterize the activation/inactivation in terms of phosphorylation/dephosphorylation and biophysical interactions describe what enzyme or catalyst regulates the reaction. For dynamical system analyses the interactions can be represented by: (i) a system of mathematical equations, which we refer to as the mathematical model *e.g.,* system of ODEs, or (ii) a system of computational entities forming computational model *e.g.,* Petri Nets [16]. Each equation in mathematical model contains variables and parameters. Variables are the quantity of interest in a model, which changes over time *e.g.,* concentrations of proteins in a cell. Parameters characterize the specific quantitative behavior of a model and are fixed quantities for a given computational experiment. Model parameter values are identified and characterized by model calibration with data [17, 18], and from available biological information and databases like SABIO-RK [19] and BioModels [20].

(iv) Validation and experimentation: After the identification of model parameter values, analytical tools (*e.g.,* bifurcation and sensitivity analysis) are used to identify/investigate the model's dynamical behavior. Model simulations reproduce the known aspects of biological reality and develop new hypotheses, which help to unravel unknown regulatory mechanisms underlying complex processes. Hypotheses made by model simulations need to be validated by designing new experiments. If the model predictions are validated by experiments it will provide a reasonable explanation of the underlying mechanisms and sharpen our understanding of the complex processes.

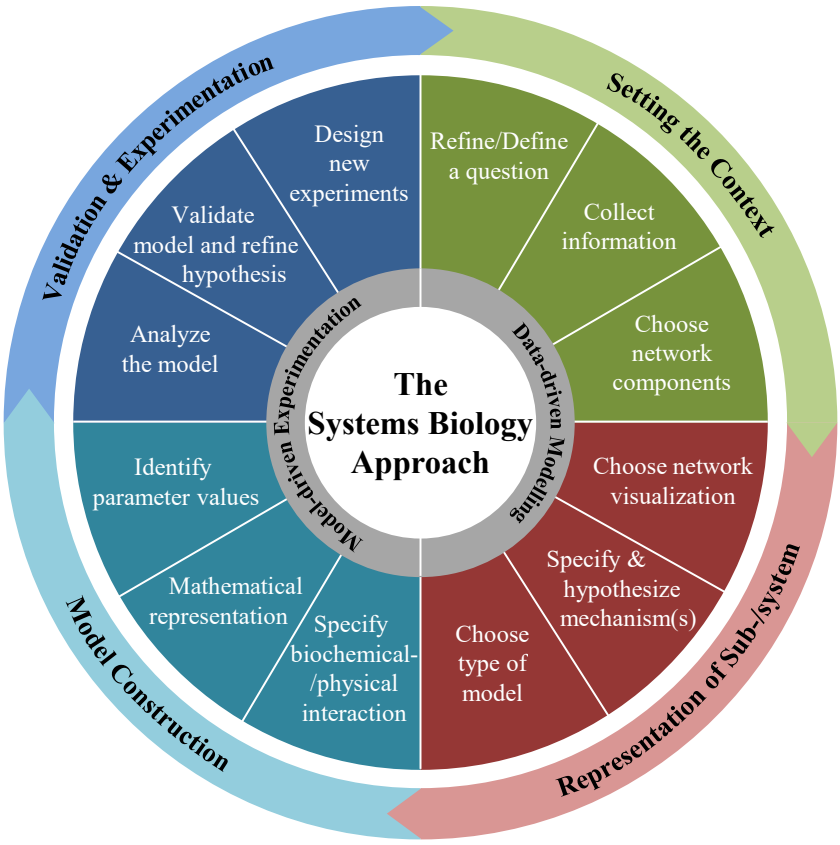


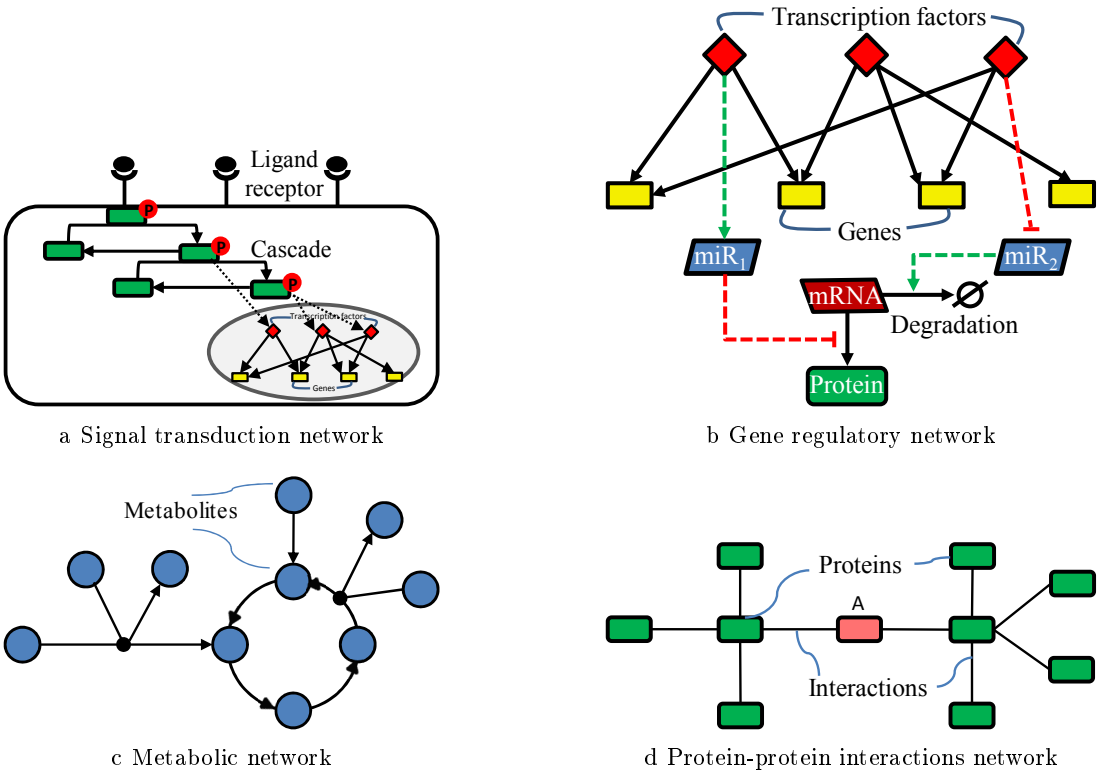
Figure 1.1. The systems biology approach. An iterative process of data-driven modeling and model-driven experimentation.The figure is taken from Khan et al., 2018 [21].

1.3. Biochemical networks

To elucidate the structure and functional roles of interacting components in a given cellular (mal)function, it is essential to have a blueprint of these interactions in the form of a network, which is a collection of nodes (genes, proteins, metabolites etc.) and edges (reactions, physical relation, conversion, transport step etc.). These networks serve as knowledge base and provide a foundation for computational analysis and, furthermore, give valuable insights into the interactions (mechanisms) among components that realize a certain functionality. Various types of interaction networks (including protein-protein interactions, metabolic, signaling, transcription and gene-regulatory networks) can be constructed by gathering the relationship data (*i.e.,* interactions) from individual studies and high-throughput screening (Figure 1.2).

Signal transduction networks are composed of nodes representing signaling proteins that generate and process information in response to external changes or stimuli (Figure 1.2a). The edges are chemical reactions that are often reversible modifications

where the signals are internally propagated in the form of, for example, protein phosphorylation or dephosphorylation cascades, which ultimately results in the activation or deactivation of a transcription factor. Numerous databases (*e.g.*, PID [22], BioCarta [23], SPIKE [24], WikiPathways [25], CST Signaling Pathways [26], The Cell Collective [27], iHOP [28], SignaLink [29] and NetPath [30]) are available that collect and organize data about signal transduction. These can be used to construct a signal transduction network.



a Signal transduction network

b Gene regulatory network

c Metabolic network

d Protein-protein interactions network

Figure 1.2. Visualization of biochemical networks. Various types of biochemical networks can be constructed based on physical or causal interactions among the interacting components. The figure are adopted from Khan et al., 2018 [21].

Gene regulatory networks (GRNs) are modeled as directed graphs where nodes are either transcription factors or genes and edges are the interactions between transcription factors and the genes that they regulate (Figure 1.2b). A GRN is a structured set of information that specifies how the transcription machinery responds to biological signals that accordingly change the expression of genes, allowing the cell to make the proteins they need at appropriate time and amount. This transcriptional regulation also works in a cascade fashion; an initial stimulus can transcribe a gene which triggers the expression of large gene sets in turn. For a GRN construction one can derive information from databases such as KEGG [31], TRANSFAC [32], TRRUST [33], HTRIdb [34], TRED [35] and TransPath [36]. Further, GRN also contains microRNAs (miRNAs), small RNA molecules that regulate the expression of certain genes by interacting with their mRNA targets and are involved in the modulation of various cellular functions [37]. miRNAs

interact with their targets (miRNA-target gene interactions) and the transcription factors (TF-miRNA interactions). miRNA-target gene interactions repress the translation of mRNAs into proteins or degrade the mRNA expression of the target. In TF-miRNA interactions, miRNA's are up or down-regulated by different transcription factors. For each type of interaction a number of public databases (*e.g.,* for miRNA-target gene interactions: miRecords [38], TarBase [39], miRTarBase [40], miRWalk [41], miRGen [42]; and for TF-miRNA interactions: TransmiR [43] and UCSC browser [44]) are available to provide such information.

Metabolic networks are composed of metabolites that are connected through biochemical reactions (*i.e.,* conversion between the metabolites catalyzed by enzyme, special type of proteins that regulate chemical reactions) (Figure 1.2c). Usually, metabolic reactions yield the production of energy exchange (ATP production/consumption) and synthesis of the biomolecules such as amino acids, sugars, and lipids which the cell needs for growth, reproduction, self-maintenance or communication with other cells. Some important databases that accumulate data for metabolic network reconstruction are KEGG [31], BioCyc [45], MetaCyc [46], BRENDA [47], BiGG [48], metaTIGER [49] etc.

Protein-protein interaction (PPI) networks represent the physical interaction among proteins, if they bind with each other (Figure 1.2d). It is a common practice to infer PPIs from experimental data [50]. The main experimental techniques to infer these networks are yeast two-hybrid screening and co-immunoprecipitation. There are plenty of databases (including BioGrid [51], HPRD [52], STRING [53], IntAct [54], PID [22], and MIPS [55]) that provide genome scale information about PPIs of various species. These databases provide undirected PPIs. In directed PPI networks, the interactions are labeled as either positive or negative based on the causal relationship of interacting proteins. A positive interaction indicates one of the interacting proteins being activated, while a negative interaction corresponds to an inactivated state. Recently, Vinayagam *et al.* developed a systematic method by integrating PPI networks with genetic screens to predict the 'signs' for large-scale PPIs[1] [56]. Further, a databases called SIGNOR[2] is also available to retrieve directed PPIs. Characterizing activation-inhibition relationships between interacting proteins can be used as a scaffold for understanding the mechanisms underlying complex biological processes.

All the described networks are interdependent and overlap with each other; and are likewise collectively responsible for cellular behavior. For example, the external signal activates the signaling pathway by ligand binding to receptors and the signal propagates subsequently from surface to nucleus by means of protein phosphorylation/dephosphorylation in the form of cascades. These cascades may activate or inhibit transcription factors that regulate the expression of genes to produce mRNA molecules. Certain sets of proteins can act as enzymes to catalyze the metabolic

[1] http://www.flyrnai.org/SignedPPI/
[2] https://signor.uniroma2.it/

reactions; and some of the *de novo* metabolite can in turn interact with transcription factors to regulate gene expression.

Efforts to represent interaction maps in an unambiguous and standard format, together with access to expression data have led to a common practice of developing comprehensive interaction maps that collect and organize molecular details to study complex cellular processes in normal and disease states [57–59]. These maps serve as knowledge-base and being machine-readable are amenable for computational analysis to formulate new hypotheses that can subsequently be tested by experiments.

1.4. Graphical representations of biochemical networks

Biochemical networks are not just drawings but a formalized representation of numerous experimental studies carried out in labs around the world that are amenable to analyze through computational tools. They are large and complex containing a variety of biochemical entities having large numbers of interactions. In order to uniformly organize and visualize such information, it requires a standardized encoding scheme. Several network representation and visualization formalisms have been proposed [60–64]. Among them the Molecular Interaction Map (MIM) named as 'Kohn Map' [64] and the Kitano's 'process diagram' [60] are the most popular ones. A MIM is a diagram convection for the representation of complex biochemical networks, containing multi-protein complexes, protein modifications, and enzymes that are substrates of other enzymes. MIMs provide an organized knowledge-base for all possible types of molecular interactions but due to graphical complexity and lack of software support their use is quite scarce. Kitano's proposed graphical notation made several improvements to the Kohn Map and provides a well defined and clearly visible notations system, powered by the 'process diagram', which helps to understand the temporal sequences of reactions. It is worth mentioning that the Kohn Map and the improvements made by Kitano provide the foundation for standard graphical representations of biological systems.

The Systems Biology Graphical Notation (SBGN) is adopted as a standard graphical representation which facilitates an unambiguous representation and easy exchange of complex biological knowledge in the form of standardized pathway maps [65]. SBGN consist of three visual languages: (i) Entity Relationship (ER), (ii) Process Description (PD), and (iii) Activity Flow (AF). The SBGN ER is largely inspired by MIM diagram convention to represent all the relationships in which an entity participates regardless of the temporal aspect. Kitano's Process Diagrams [66] and CellDesigner [12] are tools that adopted the SBGN Process Description (SBGN PD) to represent mechanistic and temporal relationships among entities involved in a biological process. It contains detailed process information and can easily be converted into standardized machine readable code such as Systems Biology Markup Language (SBML) [67] for computational analysis, including computer simulations. The SBGN Activity Flow (SBGN AF) represents the

causal relationship (e.g., stimulation, inhibition, state transitions) between molecules in an abstract manner. Including CellDesigner, many pathways tools (see the list at[3]) are already using the SBGN notation (Figure 1.3) for graphical representation. In this study, we used CellDesigner for the construction of the E2F1 regulatory map.

[3] http://sbgn.github.io/sbgn/software_support/

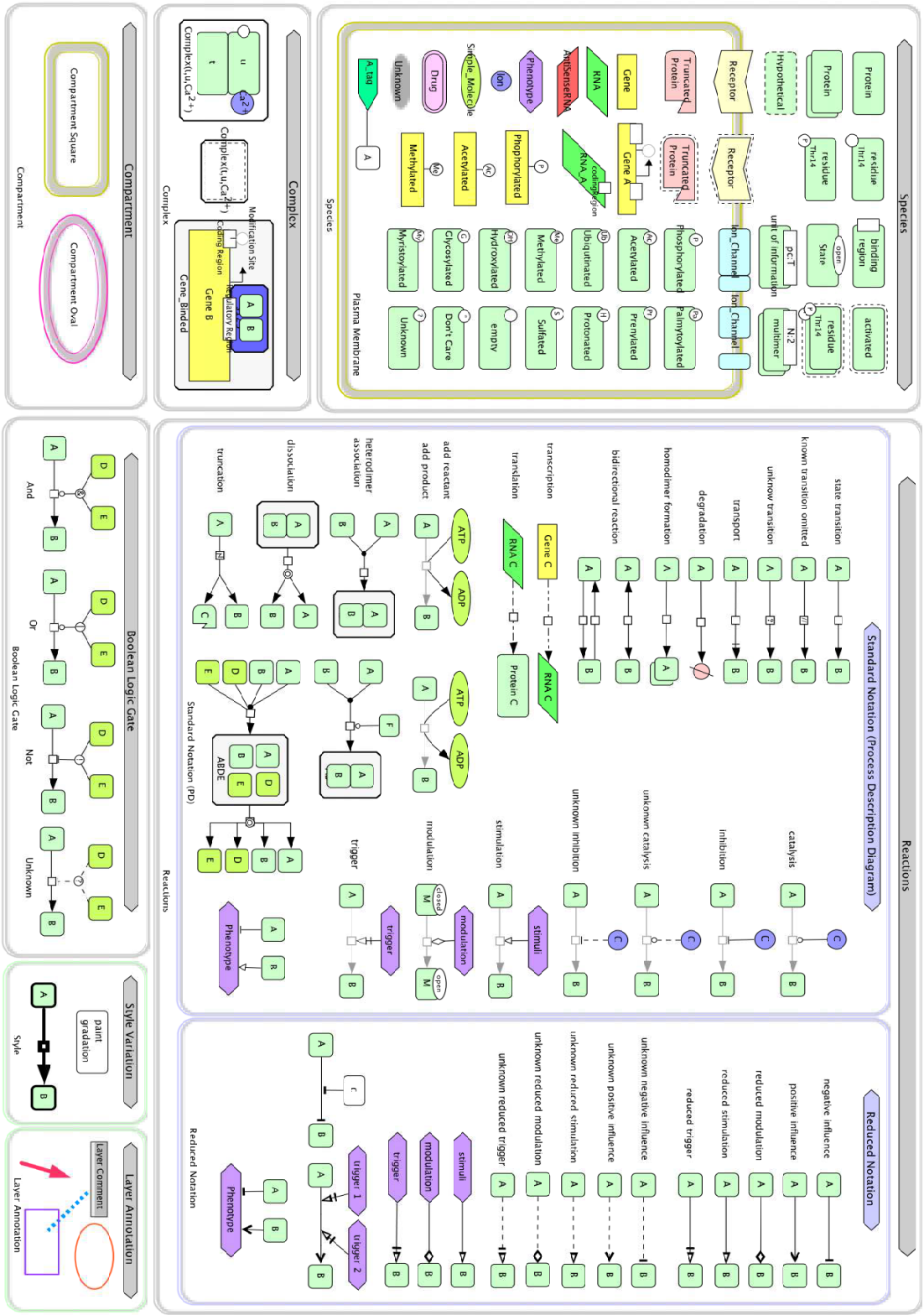


Figure 1.3. Glyphs of the graphical notation used by CellDesigner. The diagram illustrates the main graphical symbols used by CellDesinger (Version 4.4) for the representation of network components. The size and color of each symbol are configurable. Source http://celldesigner.org/documents/StartupGuide41.pdf.

1.5. Network analysis

After gathering and managing information regarding biological systems in the form of interactions networks, structural and dynamical analyses provide useful information about the network architecture and dynamics. The structural analysis of networks allows the identification of functional modules, regulatory motifs (including feedback and feed-forward loops) and node properties (including node degree (ND) and betweenness centrality (BC)). The node degree is the number of edges (or links) connected to a node. For example, in Figure 1.4 node 'B' has six edges, thus its degree is six (*i.e.,* $ND(B) = 6$). The betweenness centrality of a node is the number of shortest paths from all nodes to all others that pass through that node. In a network, a node which occurs in many of the shortest paths has a high betweenness centrality. In biochemical networks, nodes can be proteins, genes, miRNAs or metabolites etc.

Node properties have a significant role in the network topology [68–71]. For example, networks with a node degree distribution follows a power law $P(ND) = ND^{-r}$ are called scale-free networks, where r is an approximated parameter whose value ranges $2 < r < 3$. There are two important characteristics of scale-free networks; first, they contain 'hubs', nodes comprising many more connections than others; secondly, these networks are considered robust against single random perturbations [72, 73]. Similarly, betweenness centrality indicates the centrality of a node in a network. A node with a high BC serves as a gate keeper in the communication between different parts of a network. For example in Figure 1.4 node 'A' connects the top and bottom part of a network; it has the highest BC value [72].

Further, it is established that biological networks contains modules. A module is an aggregations of densely interconnected neighboring nodes as shown by the green box in Figure 1.4. It was shown that functionally akin nodes are located in close proximity and, thus, form functional modules [74, 75]. These functionally related genes can be associated with the same biological pathway and can have similar effects on certain disease phenotypes and may be targeted by structurally similar drugs [76, 77]. Further, biological networks are enriched in recurring structural patterns called network motifs [2]. Feedback and feed-forward loops are important network motifs which induces non-intuitive behavior and play a crucial role in system dynamics [2, 78]. I describe network motifs in detail in Section 2.1.4 on Page 26.

Network structure analysis revealed that the functioning and regulation of a network are governed by a certain set of organizing principles [68]. To understand the mechanisms of these organizing principles, mechanistic dynamical models are used to analyze the systems. The dynamical properties of a network characterize the temporal behavior of a network under certain conditions, which can help to explain the nature of regulation of interacting components in response to stimuli/perturbations that affect the functionality of cells, and ultimately their consequences on the cellular phenotype.

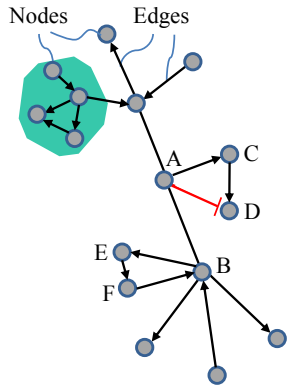


Figure 1.4. Properties of a molecular interaction network. In figure the circles represent nodes of a network connected by edges. Node 'A' has highest the betweenness centrality, most of the shortest parts that connect each pair of nodes in this network passes through node 'A'. Node 'B' has highest degree (*i.e.,* the large number of edges going in and out of this node). The green box represents a module, which is a cluster of highly inter-connected nodes. If these nodes are involved in the regulation of a particular cellular function it is called functional module.

1.6. Selection of mathematical modeling formalisms

Mathematical models are a well-established tool to elucidate the counter-intuitive nonlinear dynamical features of complex biochemical networks in living cells. Models are an abstract representation of the reality, and thereby their validity and usefulness depend strongly on the context and the assumptions being made. Modelling is the art of making appropriate assumptions. A variety of modeling formalisms have been proposed [6, 79] and are applied depending on the availability of data, the type of research question as well as the size and the structure of the system (Figure 1.5). Two commonly used modeling formalisms are: (i) detailed quantitative models of (small) functional modules using, for instance, ordinary differential equations (ODEs) [4, 80, 81]; and (ii) coarse grained qualitative models of (large) sub-cellular processes which can be encoded through logic-based formalisms [82].

ODE-based modeling: If the detailed kinetic information of relevant components in a network are largely known, for a small-scale network a number of modeling formalisms are used, such as, ODE-based models, stochastic models or agented-based models [79, 83]. Among them, ODE-based models are widely used to analyze the functional role of nonlinear biochemical networks [4, 80, 81]. In such models the reactions are represented by a set of differential equations describing the change in quantity (such as concentration) of reactants to products and *vice versa* in case of reversible reactions. Based on reaction rates and kinetic parameters such models usually yield high quality predictions of the system's dynamics with quantitative information about molecular concentrations. We used ODE-based models to explain the nonlinear dynamics induced by the E2F1-p73/DNp73-miR205 network of drug

resistance in melanoma (for details see Section 4.3 on Page 64). These models however, require accurate kinetic parameters, which is often infeasible for large networks, therefore, ODE-based models of large biochemical systems are very difficult if not impossible to obtain/generate. In such cases a Boolean/logic-based models are a suitable option [84, 85].

Logic-based modeling: Logic-based modeling is a popular approach to describe the qualitative temporal behavior of large systems of interactions where experimental data are sparse (not all can be measured, few time points) and uncertain (lack of replicates and precision) [86]. Logic-based models are qualitative and do not require detailed quantitative parameters which make them suitable for large-scale biochemical networks [6, 82, 84, 87–92]. In this modeling formalism, a network is represented as a graph with nodes and edges, where a node represents any molecular species and edges depict the type of effect that one species exert on the state of another in terms of activation/inactivation. The state of each species is determined by a logic-based function that links the incoming effect to a state. Logic-based models can provide predictive testable hypotheses, which are especially valuable in poorly understood large-scale systems [93, 94]. I used a logic-based model to analyze the stimulus-response behavior of tumor-specific EMT regulatory networks in bladder and breast cancer. The *in silico* analysis identified molecular signatures for each cancer type which were validated by patient data and through shRNA-based experiments (for details see Section 3.2 on Page 43). Logic-based models are qualitative and can not capture the network dynamics where small variations can produce an entirely different state/behavior. Therefore, the choice for an appropriate modeling formalism for biochemical networks which are large and have complex structures composed of multiple regulatory loops, is challenging. I proposed a modeling strategy, which I here refer to as "hybrid models" [9], that combines ODE-based and logic-based models to accommodate a large-scale, nonlinear system of interactions.

Hybrid models: Hybrid models combine different modeling formalisms to handle systems that contain multiple aspects; for example discrete and continuous, linear and non-linear dynamics. The biological system of the cell cycle is an appealing example that contains discrete and continuous aspects [95, 96]. Alfieri *et al.* modeled the cell cycle as a hybrid system using hybrid automata where the R-point (a point in G_1 phase at which the cell committed to the cell cycle) transition was modeled as a discrete event while the mitogenic stimulation of the system was realized as a continuous state by ODEs [96]. In Khan *et al.*, I proposed a hybrid modeling formalism that combines the features of ODEs and logic-based formalisms to provide an efficient solution to model large-scale, non-linear biochemical networks [9]. Using such a hybrid modeling approach, the network is organized and divided into different parts with distinctive regulatory features and each part is modeled with the suitable modeling formalism (Figure 1.5). For instance sub-networks that are enriched with feedback and feed-forward loops, and are therefore expected to display a highly nonlinear behavior are modeled using ODEs, whereas target

gene modules that represent an activation or inactivation regulation of dozens to hundreds of genes are modeled using a logic-based formalism. Using a hybrid model, I performed the stimulus-response analysis of the E2F1-p73/DNp73-miR205 network. It accurately captured the effect of nonlinear dynamics of regulatory motifs on a set of genes involved in drug resistance in melanoma. Further, model simulations revealed that with high expression of the E2F1 and EGFR receptors, tumor cells get independent of ligands (for details see Section 4.4, Page 66). A hybrid model provides a good compromise between quantitative/qualitative accuracy and scalability when considering large networks.

The standard representation of these models in the form of the Systems Biology Markup Language (SBML), SBML *qual* package in case of logic-based models, and the Systems Biology Graphical Notation (SBGN) allows for re-usability and exchange. Databases such as the BioModels Database [20], and the CellML [97] and SABIO-RK [19] repositories are used for model storage and exchange.

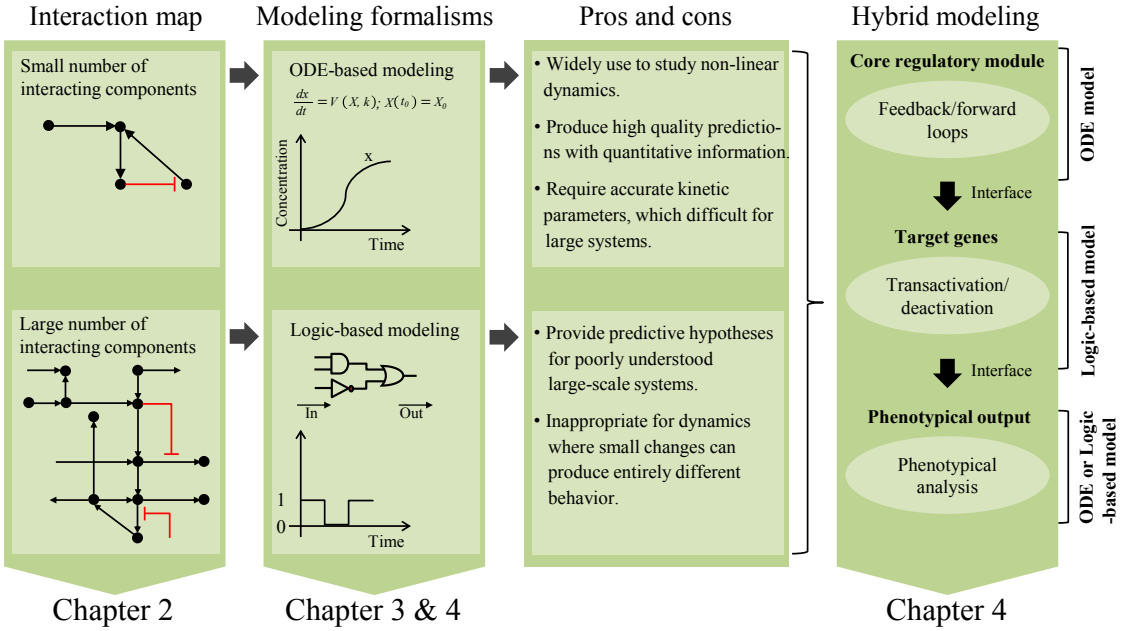


Figure 1.5. Mathematical modeling formalisms. The core regulatory network is modeled using ordinary differential equations (ODEs), the target genes module is modeled using discrete logic-based modeling (Boolean or multi-valued logic), while the phenotypical read-outs are encoded in ODEs or discrete logic-based modeling (multi-valued logic). The figure is adopted from Khan et al., 2018 [21].

1.7. An integrative workflow to unravel mechanisms underlying diseases

Recent technological advancements in biology yield remarkable amounts of data on biological organisms but the mechanisms behind diseases remain largely elusive, preventing the development of efficient diagnostic and therapeutic measures. Towards

mechanistic understanding, the systems biology and bioinformatics proposed many computational methods and tools to process data in the functional context of a biological system; but the inherent complexity associated with interacting components, for example regulatory circuits and cross-talks among pathways, and massive network size pose a methodological challenge to these methods. For example: (i) Transcriptomic data analysis tools identify associated gene expression patterns for normal to disease states; but due to genetic heterogeneity it can not explain the causes that induce such patterns; (ii) graph theoretical analyzes of biological systems that are translated into biochemical networks identifies important hub nodes and small recurring regulatory motifs but their utility is confined only to static analysis; and (iii) dynamical systems theory offers a large number of modeling formalisms to simulate the dynamics of biological systems for different stimuli and perturbations but each modeling formalism has its own limitations (see Figure 1.5). To overcome these limitations, we need strategies that combine the strength of different methods or tools, here named as "integrative workflows", with the goal to effectively analyze massive and complex networks to understand the functionality of biological systems. The wide applicability of high-throughput data and network-based analyses (*i.e.,* network structure and dynamics analysis) enables workflows to address biological and biomedical questions where individual conventional methods are not performing well.

These integrative methods in systems biology are scarce but highly demanded. A number of integrative methods are proposed that successfully identify key network elements in diseases. Some are based on integrating data and networks to identify regulatory modules in diseases [98–100]. Others integrate network structural properties with expression data to identify potential regulators for a disease [101–103]. The recent reviews by Mitra *et al.* [75] and Gustafsson [104] discuss integrative approaches, in particular networks and data, to identify modules in regulatory networks underlying disease.

Here, I propose an integrative workflow that combines techniques from systems biology and bioinformatics, in combination with the integration of experimental and clinical data to analyze large-scale biochemical networks with the aim to investigate complex mechanisms underlying diseases (Figure 1.6). The workflow combines network structural analysis with high-throughput and biomedical data to identify tumor-type specific “core-regulatory networks” that are amenable for analysis with dynamical systems theory. Using suitable modeling formalism, the core networks can be analyzed to detect disease-specific molecular signatures (*i.e.,* sets of network derived diagnostic/prognostic biomarkers). The proposed molecular signatures are validated by clinical data which are further subjected to *in vitro* experiments. The proposed workflow was applied to our reconstructed E2F1 molecular interaction map and was able to identify core-regulatory networks in invasive bladder and breast cancer (see Chapter 2). Using a logic-based modeling formalism, I analyzed the stimulus response behavior of the core networks to identify distinctive molecular signatures for EMT

regulation in both cancer types. Further, I performed *in silico* perturbations of the core network and determined possible drug targets (see Table 3.3, Page 49), which can render high invasiveness in cancer cells (for details see Chapter 3). The model predictions were validated with patient data which were further verified by *in vitro* experiments.

The integrative workflow cannot provide an exact representation of cellular events but it guides the formulation of hypotheses and their validation in experiments.

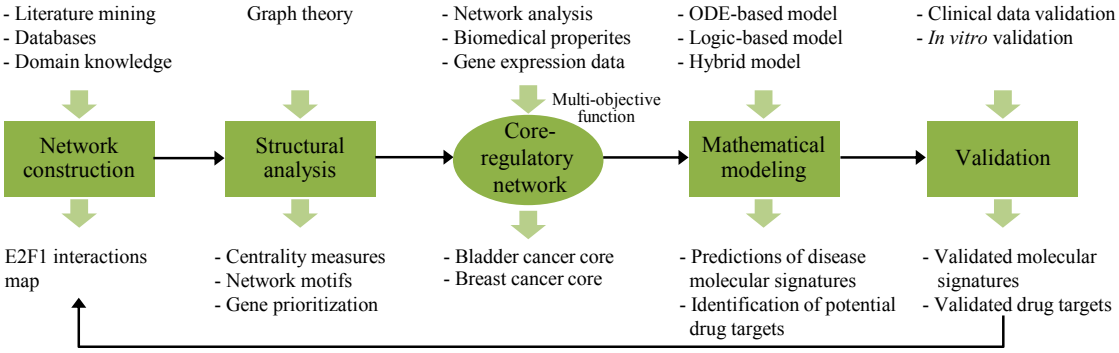


Figure 1.6. Integrative workflow. An integrative workflow to study large-scale biological networks to unravel regulatory mechanisms underlying diseases. The workflow combines network analysis with high-throughput transcriptomic data and biomedical information to identify disease specific core-regulatory networks that are amenable for analysis with dynamical systems theory. Using a suitable modeling formalism, the core-networks can be analyzed for different input-stimuli and perturbations to identify molecular signatures for diseases that are subject to experimental validation. The figure is adopted from Khan et al., 2018 [21].

1.8. Outline of the thesis

The present work is a development of an integrative workflow to study large-scale biochemical networks to understand and unravel mechanisms underlying complex diseases, such as cancer. Figure 1.7 on Page 18 illustrates an outline of the thesis, which is organized in the following chapters:

Chapter 2 is divided into two parts. In the first part (**Part A**) I describe the steps to collect and organize the disperse knowledge for a certain biological phenomenon in a format amenable to computer simulations. I list numerous databases that can be used to extract information for different types of interactions *e.g.,* protein-protein interactions. We construct a comprehensive interaction map of transcription factor E2F1 by deriving and incorporating information from literature and databases. The map is constructed in the pathway diagram editor CellDesigner and visualized in SBGN format. The second part (**Part B**) of the chapter describes the design of a methodology that integrates network structural properties, specifically centrality measures and feedback/feed forward loops, with expression data and other biomedical

knowledge to identify a tumor-specific core regulatory network. Using the proposed methodology, I derive tumor specific core-networks from the large-scale E2F1 interaction map for epithelial-mesenchymal transition (EMT) regulation in bladder and breast cancer. The core-networks contain critical molecular interactions and regulatory motifs that may drive a cancerous phenotype.

Chapter 3 & 4 focus on the methods of dynamical analysis, namely ODE-based kinetic models, logic-based models and hybrid models, to provide mechanistic insights into disease progression. Specifically, in **Chapter 3**, I describe the development of logic-based models to analyze the dynamics of core-regulatory networks, previously identified in **Chapter 2**. The models are calibrated with expression data and the *in silico* simulations predict molecular signatures for EMT regulation in bladder and breast cancer. Further, model analysis based on *in silico* perturbations identified potential therapeutic targets for each cancer type. The model predictions were validated using TCGA patient data and *in vitro* by shRNA-based experiments.

Chapter 4 contains ODE-based and hybrid models of chemoresistance. In this chapter, I provide a motivation for hybrid models, which combine ODE-based and logic-based modeling. Section 4.2 describes ODE-based models, the law of mass-action kinetics, and the type of analysis that can be carried out by ODE models. In Section 4.3, I describe the ODE model-based detection of molecular signatures for chemoresistance in a network around E2F1. Model analyses identified a signature that includes the transcription factor E2F1, two isoforms of the transcription factor TP73 (P73 and DNp73) and the microRNA miR-205. An unbalance in the E2F1-P73/DNp73-miRNA205 circuitry can mediate chemoresistance in melanoma cells.

Further, in Section 4.4, I describe the development of a hybrid model that combines ODEs and logic-based models as a strategy to model large-scale, nonlinear biochemical networks. As a proof of principle, I derive a hybrid model of the E2F1 chemoresistance. Model simulations recapitulate chemosensitivity and chemoresistance of tumor cells depending on the concentrations of growth factors, E2F1 and TGFB, as well as the level of induced genotoxic-stress. For high expression levels of E2F1 and TGFB-1 alone or in combination, the *in silico* experiments show resistance against anti-cancer drugs, which are in consensus with our kinetic model and other biological evidences in the literature.

Chapter 5 concludes and discusses how the methods developed during the course of this thesis can be used in the field for other studies. Finally, I preview possible future directions and open questions that could be addressed to improve the effectiveness of similar integrative workflows in computational biology.

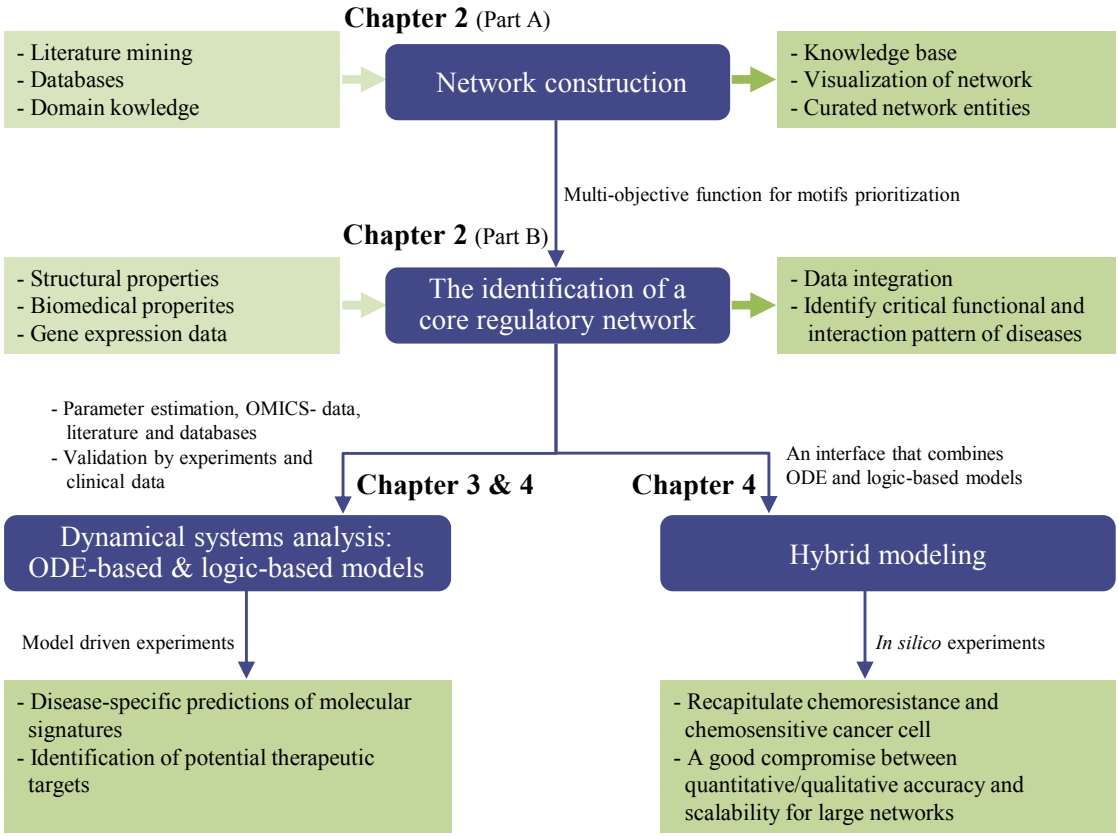


Figure 1.7. Thesis outline. Studies presented in this thesis are organized in the following chapters: (i) **Chapter 2** has two parts where Part A illustrates the construction of biochemical networks; and Part B describes the methodology of tumor specific core-regulatory network identification, (ii) **Chapter 3** includes logic-based modeling as a dynamical systems analysis tool to perform *in silico* experiments of the core-regulatory networks identified in **Chapter 2** to unravel mechanisms (*e.g.* identify molecular signatures and therapeutic targets) underlying EMT regulation in bladder and breast cancer, and (iii) **Chapter 4** describes the hybrid modeling approach as a strategy to simulate large-scale nonlinear network as dynamical systems. The blue rounded rectangles represent the methods developed, the light green rectangles list the possible inputs to respective methods to produce the output, listed in dark green rectangles.

Chapter 2

Construction and analysis of the E2F1 molecular interaction map (MIM)

The methodology for the network construction and subsequent analysis is published in the following publications:

- **Khan FM**, Marquardt S, Gupta SK, Knoll S, Schmitz U, Spitschak A, Engelmann D, Vera J, Wolkenhauer O, Pützer BM. Unraveling a tumor type-specific regulatory core underlying E2F1-mediated epithelial-mesenchymal transition to predict receptor protein signatures. Nat Commun. 2017 Aug 4;8(1):198. https://doi.org/10.1038/s41467-017-00268-2.
- **Khan FM**, Sadeghi M, Gupta SK, Wolkenhauer O. (2018) A Network-Based Integrative Workflow to Unravel Mechanisms Underlying Disease Progression. In: Bizzarri M. (eds) Systems Biology. Methods in Molecular Biology, vol 1702. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-7456-6_12.
- **Khan FM**, Gupta SK, Wolkenhauer O. Integrative workflows for network analysis. Essays in biochemistry. 2018 Oct 26;62(4):549-61. https://doi.org/10.1042/EBC20180005.

Synopsis

Processes in living cells are carried out by a large number of interactions whose dysregulations result in complex diseases like cancer. Networks are the blue print representation of these interactions and their analyses provide useful insights into diseases. The massive size and complex nature of interactions hamper conventional computational tools to decipher networks for disease mechanisms. In this chapter, I propose an integrative workflow to analyze large networks. This chapter is divided into two parts: (A) Construction of the E2F1 molecular interaction map; and (B) The identification of tumor type-specific core-regulatory networks. The first part includes the list of databases and repositories used for retrieving information about molecular interactions. Brief descriptions of the modules in the map, and the annotation of molecular species and interactions are provided. The second part contains the development of a workflow that integrates network structural properties with high-throughput and biomedical data through a multi-objective optimization function to

identify tumor type-specific core-regulatory network, which are amenable for analysis using dynamical systems theory.

2.1. Part A: Construction of the E2F1 MIM

2.1.1. Background and motivation

Cellular processes are regulated by a large number of interactions among biological entities such as genes, proteins, RNAs, mRNAs, enzymes, transcription factors and other molecules. To understand the mechanisms behind normal and malfunctioned (linked to diseases) execution of these processes, it is essential to initially organize data and knowledge about them in a network representation (*i.e.,* vertices connected by edges). Such networks (*e.g.,* molecular interaction maps) are a formalized representation of information that can subsequently be analyzed with computational algorithms. They serve as knowledge-bases, which help in: (i) gathering disperse information about complex biological systems in one place, (ii) managing and organizing information in a standard pathway diagram format that is helpful to conceptually analyze and intuitively visualize the network components, (iii) provide information about the interactions to develop hypotheses that can experimentally be tested, and (iv) provide a foundation to derive mathematical models to analyze the dynamics of interacting components.

The construction of networks includes phases of manual curation, validation and encoding. Various approaches are used to map out molecular interactions underlying certain biological processes. A number of computational techniques have been developed, to infer biochemical networks from experimental data [105] and high throughput data [106, 107]. Despite their importance, they face many challenges, for example false positive and false negative interactions. More detailed and highly focused networks, centered around a particular disease [57–59] or cellular process [108–110], can be constructed by domain expert knowledge (functional and structural information), diligent manual search of published literature and publically available databases (*e.g.,* KEGG [31], BioCyc [45], HPRD [52], IntAct [54], DIP [22] and miRTarBase [40]). To avoid the laborious manual curation for network construction, some methods were developed to automatically reconstruct networks by retrieving interactions or sub-networks from existing maps and models [111, 112], and some Cytoscape 'Network Generation' plugins including Bisogenet [113], OmniPath [114] etc. Combining automatic network reconstruction with domain knowledge, manual literature search and data retrieval from public repositories would provide a reasonable strategy to construct detailed and fully annotated large-scale biochemical networks (see Figure 2.1). Further, for the construction of a interaction network, one can introduce more regulatory layers, which contain, for example, transcriptional or post-transcriptional interactions. Databases providing information about transcriptional interactions include TRANSFAC [32], TRRUST [33] and HTRIdb [34]. Resources for post-transcriptional

interactions include miRTarBase [40], miRWalk [41] and TriplexRNA [115]).

Network-based studies are becoming a popular approach to collect and organize molecular details and get insights into the mechanisms underlying dysregulated processes in diseases. For example, the Parkinson's disease (PD) map by Fujita *et al.* containing detailed molecular information on various aspects of PD pathogenesis [57]. Matsuoka *et al.* created a large detailed pathway map of the influenza A virus (IAV) replication cycle [58]. The map is annotated with around 500 scientific articles, includes information from previous influenza maps, and incorporates pathway information from KEGG, PANTHER and Reactom databases. This study was intended to develop a broader picture of the functional mechanism of IAV and its associated host response. Further the map was used for *in silico* analysis to identify several critical targets in the IAV life cycle. Calzone *et al.* constructed a comprehensive map of RB/E2F pathway interactions in the regulation of the cell cycle [109]. They identified different structural modules in the map based on clusters of regulatory cycles.

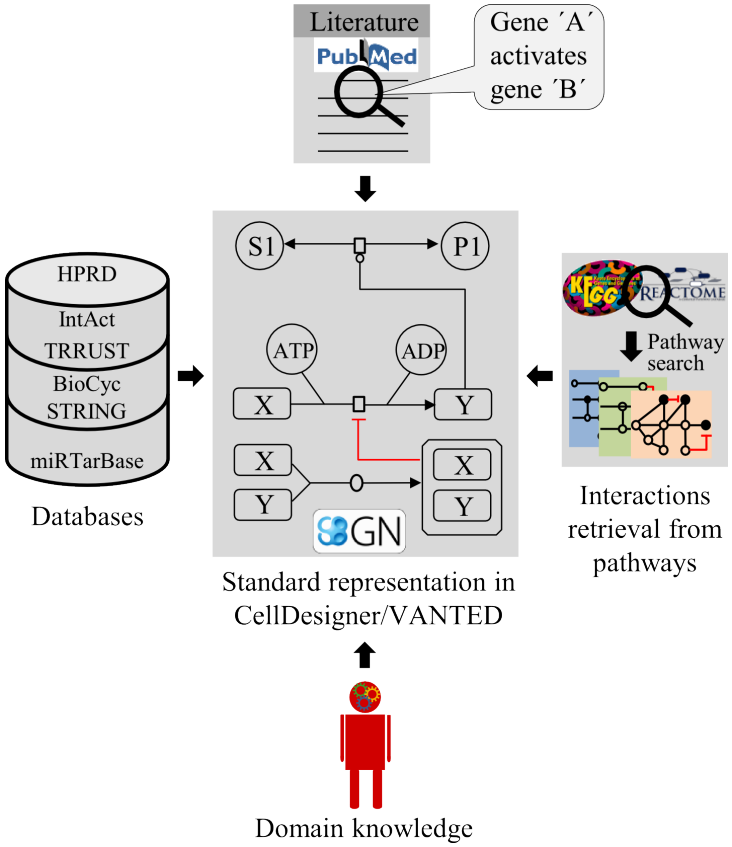


Figure 2.1. Scheme for biochemical network construction. More detailed and highly focused networks, centered around a particular disease or process, can be constructed using expert domain knowledge (functional and structural information), manual search and curation of published literature and publically available databases (*e.g.,* HPRD, IntAct, DIP, BioCyc, KEGG, REACTOME and miRTarBase). The figure is taken from Khan et al., 2018 [21].

To mechanistically understand the role of E2F, a family of transcription factors, in cancer, more specifically in drug resistance, tumor invasion and metastasis, we constructed a molecular interaction map (MIM) by retrieving information from literature and interaction databases. The map contains state-of-the-art knowledge on signaling, transcriptional, post-transcriptional, and protein-protein interactions around the E2F family. All the interactions are carefully curated with references and validated by domain experts.

2.1.2. E2F1 interaction map

E2F1 is the most prominent member of the E2F family, involved in a number of essential cancer-related cellular processes such as proliferation, apoptosis and differentiation [116]. E2F1 is a remarkable example of a network hub as this protein interacts with many genes, proteins, and other transcription factors through a variety of regulatory mechanisms. In the context of solid tumors, unbalanced E2F1 regulation can lead to the emergence of aggressive tumor cell phenotypes, which drives cancer progression, resistance to anti-cancer drugs and the rise of metastatic lesions [117–121]. To understand how E2F1 interacts with different molecules and how it mediates cancer related processes upon activation of different receptors, we constructed a detailed molecular interaction map (Figure 2.2) based on information retrieved from published literature and databases as well as E2F1 cofactors [122] recently identified by our colleagues from the Institute of Experimental Gene Therapy and Cancer Research (IEGT)[1] headed by Prof. Brigitte M. Pützer. More specifically, we retrieved data on protein-protein interactions from STRING [53] (v9.1), IntAct [54], and HPRD [52] (release 9). Transcription factors and their target genes were retrieved from TRANSFAC [32] and relevant literature. Moreover, we included miRNA-target interactions, which were extracted from the miRTarBase [40] database (release 4.5), a database of validated miRNA-target interactions. Transcription factors of miRNAs have been extracted from the TransmiR [43] database (v1.2). Additionally, we searched PubMed for publications about validated E2F transcription factors, their post-translational modifications, molecular interactions and connections to certain diseases, especially to cancer. Furthermore, we manually curated interactions (for assigning directions to the interactions, *i.e.,* activation/inhibition, and relevant references) that were retrieved semi-automatically from databases like STRING, HPRD, IntAct, KEGG, or the EMBL-EBI search engine PSICQUIC [123] and the text mining tool iHop [28]. In order to assure the accuracy of the network, we randomly selected 10 percent of the interactions and asked independent domain experts to cross-validate them. Over 98% of the interactions were derived correctly.

The E2F1 regulation and activity map constructed in this thesis is based on manual exploration of over 800 publications related to E2F1 and E2F family proteins, as well as

[1] http://www.iegt-rostock.de/en/

connected pathways playing a role in cancer-related cellular processes. The map, shown in Figure 2.2, contains 873 nodes (including miRNAs, which are not visualized in the Figure 2.2) and 2315 interactions. In order to improve the visualization, we modularized the map into different modules: (i) Extra-/intracellular receptor signaling ($n = 113$); (ii) Post-translational modifiers of E2F1 ($n = 24$); (iii) Regulators of E2F1 transcriptional activities ($n = 66$), where n the number of factors in each layer. Furthermore, there are seven functional modules: (i) Cell cycle ($n = 145$); (ii) Quiescence ($n = 29$); (iii) DNA repair ($n = 33$); (iv) Metabolism ($n = 11$); (v) Apoptosis ($n = 89$); (vi) Survival ($n = 52$); and (vii) EMT/Invasion/Angiogenesis ($n = 69$). Where n is the number of nodes in each module.

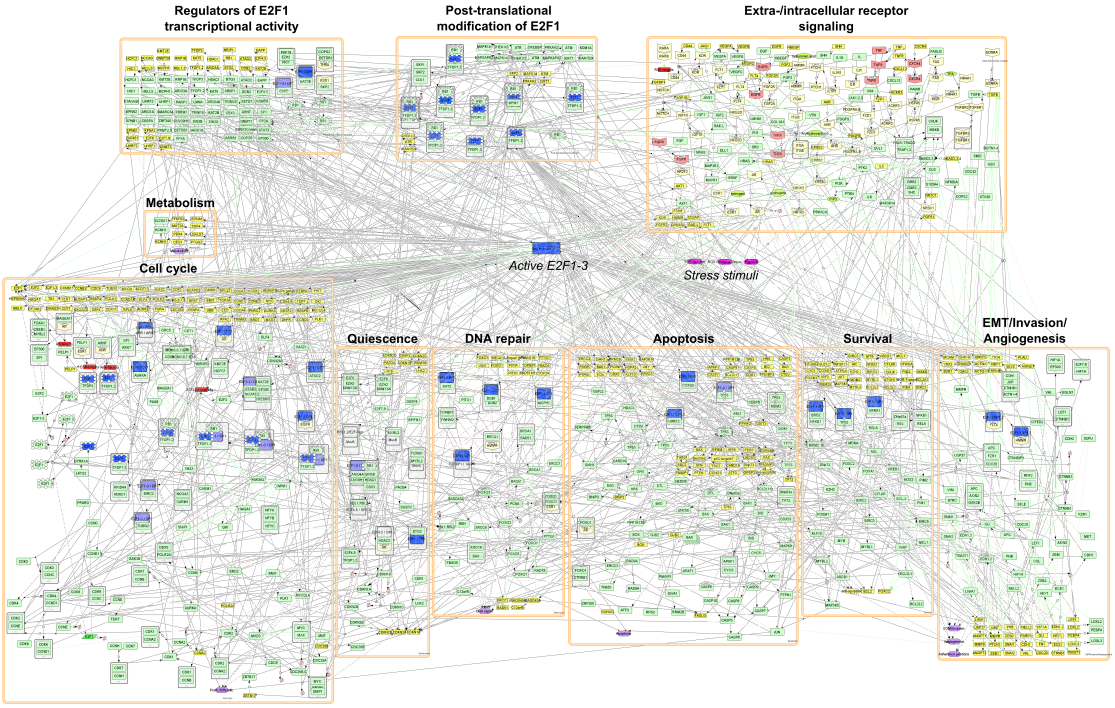


Figure 2.2. A comprehensive map of E2F1 interactions in tumor progression and metastasis. The map is organized into different functional and regulatory modules. This figure is taken from Khan et al., 2017 [124]. For better visualization, transcription and translation processes are condensed into one reaction directly leading from gene (yellow boxes) to protein (green or blue boxes with round edges). In the map we pool proteins of the same family into one node (*e.g.,* FGFR for FGFR1-4, FGF for FGF1-23, ITGA and ITGB for alpha and beta integrins). The miRNA layer is not included in this CellDesigner diagram, but for network analysis, the Cytoscape version of map, includes all the family members of place holders and miRNAs (Figure A.3 in Appendix A). The interactive E2F1 interaction map can be accessed at https://navicell.curie.fr/pages/maps_e2f1.html.

2.1.3. Graphical representation

The E2F1 interaction map (in Figure 2.2) was built with the process diagram editor CellDesigner [12] (v4.3) and visualized as a SBGN (Systems Biology Graphical Notation) compliant diagram [65]. We created different regulatory and functional parts based on the role of molecules in influencing E2F1 activities and determining cell fates. Next, I incorporated into the map detailed text-based annotations including Official name, HGNC ID, Entrez ID, and UniProt ID for nodes, and PubMed reference IDs for reactions. HGNC ID is an identifier assigned by HUGO Gene Nomenclature Committee [125]. Entrez ID is a unique identifier assigned by National Center for Biotechnology Information (NCBI), which provides detailed information about every gene, such as genomic location, gene products and their attributes, markerf, phenotypes and links to citations [126]. UniProt ID is a unique identifier for each protein assigned by UniProt database, a resource for protein sequence and functional information [127]. All this information is stored in the Notes section of CellDesigner. We used complex formation and dissociation information to accommodate activities of protein monomers, dimers and oligomers. By assigning web links to the annotations in the map, we turned it into an interactive resource. A complete list of the E2F1 CellDesigner map interactions with corresponding PubMed references can be downloaded at[2].

Web version: I created web version of the map using a pathway browser platform called NaviCell [128], which can be accessed at[3], where users can browse its contents. NaviCell uses the graphical representation of maps created in CellDesigner. It provides 4 levels of zooming for visualization and facilitates an easy navigation of the network components. If annotations are provided in the CellDesigner .xml file, users can see it by clicking on the biological entities. It also provides web-blog, which allow users to contribute in refinement and update of the map.

I created four images of different zoom levels: (i) detailed zoom level, it is the original representation of map, (ii) hidden-details zoom level, which hides the entity modification, complex names and reaction numbers, (iii) pruned zoom level, shows only major routes in the map, and (iv) top-level zoom view, which is an abstract representation of the map. All these zoom levels are separately created by JavaScript provided in the NaviCell user guide[4]. In order to create the NaviCell web map, zoomed images and the CellDesigner .xml files are converted into html and JavaScript files using NaviCell factory, which is a java-based software embedded into Cytoscape plugin BiNoM [129].

2.1.4. Network structural analysis

To evaluate the structural properties of the E2F1 interaction map, we converted the E2F1 map into a format suitable for analytical tools such as Cytoscape [13]. Towards this, all types of reactions were categorized into activation; inhibition and neutral interaction

[2] https://sourceforge.net/projects/khan-phd/files/Supplementary-file-1.xlsx/download
[3] https://navicell.curie.fr/pages/maps_e2f1.html
[4] https://navicell.curie.fr/doc/usersguide.pdf

(a complete list of interactions in a format suitable for analysis in Cytoscape can be downloaded at[5]). Moreover, complexes that take part in a reaction were dissected and separate reactions were established for their components, for example: a reaction for complex 'AB' that activates 'C' was split into two separate interactions: (i) 'A' activates 'C'; and (ii) 'B' activates 'C'. The purpose to dissect the complexes into separate reactions was to add the post-transcriptional regulatory layer (miRNAs) and map expression data on the map. While in the CellDesigner map (Figure 2.2) proteins of the same family are pooled in one node, the Cytoscape map (see Figure A.3 in Appendix A) contains all respective family members and interactions among them, thereby allowing the use of existing tools for data integration and network analysis. The Cytoscape file of the map can be accessed at[6].

Topological properties of the E2F1 map

The E2F1 molecular interaction map in Cytoscape format contains 1,015 nodes and 4,180 interactions. I determined topological properties (see Table 2.1) of the network using the Cytoscape plugin 'NetworkAnalyzer' [130]. The average number of neighbors for each node in the network is 7.89, which indicates that the network is well-connected. I calculated the average clustering coefficient of the network ($\overline{C} = 0.226$) and the network diameter ($D = 8$). The clustering coefficient indicates the density of connections among the neighbors of a node [131]. The comparably large value of $\overline{C}$ and large diameter of the network indicate the modular organization of nodes in the network [70, 132]. Furthermore, I fitted a power law of the form $y = a + x^{-r}$ to the clustering coefficient distribution (Figure A.2 in Appendix A). The results $a = 0.906$; $r = 0.710$ indicate a hierarchical structure of the network [72, 132]. The small average characteristic path length ($l\_G = 3.258$) and large average clustering coefficient indicate that the network has a small world architecture [3, 72, 131]. The small world property reveals that signals can propagate very fast through the whole network. The values of important topological properties for each node are listed in Supplementary-file-2 at[7].

Some of the node properties (node degree and betweenness centrality) of the E2F1 interaction map were mapped to visual properties in Figure A.3 in Appendix A. In this representation the node size is determined by its degree, *i.e.,* number of edges connected to the node and node color denotes the betweenness centrality (green: low; red: high), *i.e.,* the amount of control that a node exerts over the interactions of other nodes in the network. Since the network was constructed by focusing on interactions around E2F1, it is no surprise that E2F1 represents the largest node followed by other members of the E2F family (E2F2/3). Other nodes with very high node degree are TP53 and MYC, which are known for their role in tumorigenesis. E2F1 has also the highest betweenness centrality value ($C_b(E2F1) = 0.4222$) indicating that E2F1 plays a central role for the

[5] https://sourceforge.net/projects/khan-phd/files/Supplementary-file-2.xlsx/download
[6] https://sourceforge.net/projects/khan-phd/files/E2F1-map-in-Cytoscape.xml/download
[7] https://sourceforge.net/projects/khan-phd/files/Supplementary-file-2.xlsx/download

signal flow in this network. By determining these topological properties one can identify important nodes as potential candidates for therapy design [72, 133].

Further, I fitted a power law of the form $y = a + x^{-b}$ to the degree distribution of the network nodes (Figure A.1 in Appendix A), where y indicates the number of nodes that share a particular degree x ($a = 320.15$ and $b = 1.249$). From this result, I conclude that the network has a scale-free topology, which is consistent with the fact that the network contains few high-degree nodes also known as hubs. Networks containing few hubs are generally heterogeneous in terms of node degree and are considered to be robust against single random perturbations [72, 73].

Topological parameters	Values
Number of nodes	1015
Number of edges	4174
Clustering coefficient	0.226
Network diameter	8
Network radius	4
Characteristic path length	3.258
Avg. number of neighbors	7.892

Table 2.1. Topological parameter values of the E2F1 regulatory network [124].

Network motifs

Network motifs are the interaction patterns that occur significantly more often than in random networks [2]. These motifs are a sort of small molecular circuitry that cells use to process information. Moreover, network motifs govern dynamical responses of the cell to external or internal fluctuations [78, 134]. Feedback loops (FBLs) and feedforward loops (FFLs) are important regulatory network motifs. In a feedback loop, a node is directly or indirectly regulated by it's own target. Figures 2.3a and 2.3b, for example, illustrate the indirect activation/inhibition of a node 'X' by its own target. FBLs can either be positive or negative depending whether they have an even or odd number of negative links in the loop. Negative FBLs have an odd number of negative links (Figure 2.3b) and positive FBLs have an even number of negative links (Figure 2.3a). A negative FBL exerts negative effects on the evolution of their components, while a positive FBL exerts positive effects [135]. FFLs are characterized by interactions in which a node is a mutual target of another node and its target. In Figures 2.3c and 2.3d, for example, 'X' regulates 'Y', and then 'X' and 'Y' mutually target 'Z'. FFLs can either be coherent or in-coherent depending on the parity of negative links in the loop. A FFL is coherent (Figure 2.3c) if the parity is even and it is in-coherent (Figure 2.3d) when the parity is odd.

Using the Cytoscape plugin NetDS [136] (v3.0), I identified a large set of three-nodes FBLs ($n = 444$; 213 positive, 228 negative and 3 neutral) in the E2F1 interaction map

(Supplementary-file-3 at[8]). I considered only FBLs of three nodes due to the fact that larger sized network loops are typically composed of one or more small size loops [78]. Moreover, I have also identified feedforward loops (n = 11904) of path length 2 (*i.e.*, the maximum path from source to target is 2). In intracellular regulatory networks, FBLs provide stability and robustness against intrinsic and extrinsic noise, homeostasis or even all-or-nothing patterns of activation [2, 78, 134]. Very often, these network motifs are disrupted or abnormally regulated in cancer and, therefore their computational analysis can provide important clues to the emergence of cancer phenotypes. However, the identification of important feedback/feedforward loops, which are more relevant in certain biological context of a highly connected network, is a methodological challenge.

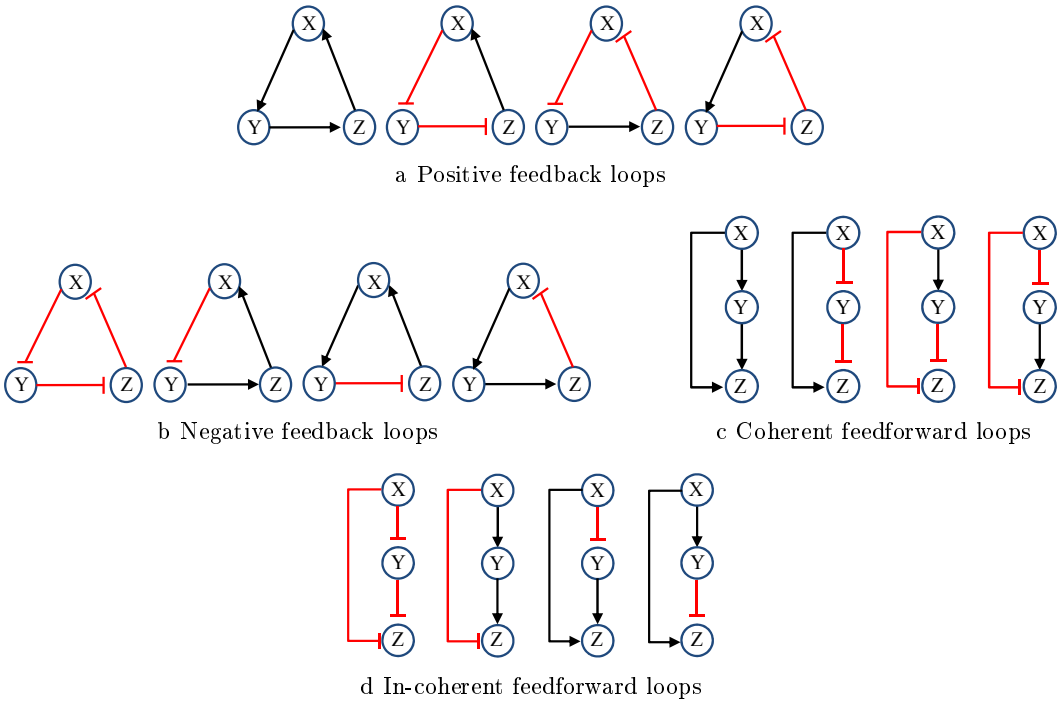


Figure 2.3. Feedback and feedforward loops. **a-b** All possible 3-node feedback loops. **c-d** All possible of path length 2 (or 3-node) feedforward loops. The figures are adopted from Khan et al., 2018 [21].

[8]https://sourceforge.net/projects/khan-phd/files/Supplementary-file-3.xlsx/download

2.2. Part B: Identification of a tumor-specific core regulatory network

2.2.1. Background and motivation

The set of mutated, deregulated or epigenetically modified cancer genes is highly tumor-type specific. More importantly, these genes are integrated in a small set of modules and regulatory pathways [137]. Furthermore, it has been found that such modules and pathways are enriched in regulatory motifs to shape and fine-tune basic cellular phenotypes that are subverted in diseases like cancer [78]. The advance in sequencing and omics technologies provides us with data that can be used to identify and characterize cancer and tumor-specific regulatory motifs. The analysis of these motifs has the potential to give insights into various aspects of carcinogenesis, tumor progression and metastasis [138–140]. Regulatory motifs, including feedback and feedforward loops are a source for nonlinear regulatory behavior [2, 68], therefore, for their analysis dynamical system theory is inevitable [5, 6]. But large-scale biochemical networks contain a large number of regulatory loops, which make a dynamical analysis difficult, particularly with mechanistic (*e.g.,* ODE-based) approaches.

In order to exploit the advantages of large-scale biochemical networks, in combination with mechanistic modeling for disease understanding, we require integrative approaches and computational workflows to identify disease-specific small regulatory/functional modules that can be subjected to a more detailed analysis, followed by the prediction of molecular signatures and therapeutic targets. Towards this, I have developed a flexible and extendible integrative workflow based on a multi-objective optimization function that combines network structural analysis with high-throughput and biomedical data to identify smaller 'modules' for specific disease phenotypes. I hereafter refer to these as "core-regulatory networks", which are amenable for analysis with dynamical systems theory [4]. For the validation of my workflow, I applied it to the E2F1 molecular interaction map (shown in Figure A.1 in Appendix A) and identified core-regulatory networks for the regulation of the epithelial-mesenchymal transition (EMT) in bladder and breast cancer.

2.2.2. Methodology for the identification of a tumor-specific core network

I have developed a novel method to prioritize recurring structures in a network (*i.e.,* motifs including feedback/feedforward loops) with respect to the disease phenotype under investigation (Figure 2.4). More specifically, the method incorporates: (a) topological properties, node degree (ND) and betweenness centrality (BC), (b) the gene prioritization (GP) scores, (c) relatedness to disease pathway (DP), and (d) gene expression fold-change (FC) data. GP is inspired by the well-established 'guilt-by-association' concept [141], which assumes that genes which are physically or functionally close to each other tend to be involved in the same phenotype/disease. GP

ranks important genes for a certain disease based on the network structure and known disease markers using network propagation methods, *e.g.,* random walk with restart. (DP) accounts the association of motifs with a certain disease by checking if any of the nodes in a motif is part of the pathway:

$$n(M_j \cap DP) \tag{2.1}$$

where n is the number of nodes in a motif M_j present in a disease pathway, and its value is $(0 \leq n \leq length\ of\ a\ motif)$, and $j = 1, \ldots, m$, for m number of motifs M_j. Associating network components with DP can provide a foundation to explain genetic heterogeneity across patients [142]. (FC) accounts for data from a normal to disease phenotype which makes the methodology specific to a disease or phenotype.

Here, I refer to $(ND + BC)$, GP, DP, and FC as four prioritization sub-functions (objective functions) $(f_l, l = 1, \ldots, 4)$, which are used in the multi-objective optimization function [143] to optimize the ranking score (F_M) of a motif for a given weight w_k:

$$F_M = \sum_{l=1}^{4} w_{lk} \cdot \sigma f_l \tag{2.2}$$

where w_l, is the weight assigned to sub-function (f_l) which is normalized by σ (equal to maximum value of f_l); and k is the number of user defined weighting scenarios. One can iteratively modify the values of the weighting factors w_k to optimize the function (F) for one or multiple sub-functions. In this case no single solution exists that simultaneously optimizes the function (F). Thus, one can have a number of Pareto optimum solutions. A solution is called Pareto optimal, if none of the optimized functions (F) can be improved in value for a given objective functions without degrading some of the other objective functions. For example, to optimize (maximize) F for objective function f_1 and f_2:

$$F = w_1 \cdot f_l + w_2 \cdot f_2 \tag{2.3}$$

the values of w_1 and w_2 can be iteratively modified to maximize the functions (F) for respective objective function. If one wants to maximize F for f_1 by assigning a higher value to w_1 than w_2, it will reduce the value for f_2, and vice versa. The Pareto set is obtained by merging all the non-identical solutions. From Pareto sets, top ranked motifs are selected based on score greater than or equal to certain cut-off $(F_M >= cut - off)$. The core-regulatory network is obtained by taking the union of the top ranked motifs.

Using this workflow, I identified core-regulatory networks from the large-scale E2F1 molecular interaction map (MIM) driving the epithelial-mesenchymal transition (EMT) regulation in bladder and breast cancer. The workflow involves the following steps: (i) Network analysis, (ii) Motif identification, (iii) Gene prioritization score, (iv) Association of network motif with disease pathway, (v) Fold change expression profile, (vi) Motif ranking using multi-objective optimization function, and (vii) Derivation of the regulatory core (see Figure 2.4).

(i) Network analysis: I performed network analyses to identify the structure and node properties, which help to identify important nodes for network organization and information flow. Topological and node properties were determined using the Cytoscape plugin NetworkAnalyzer [130]. In particular, I first calculated for each node the *ND* and *BC*. The average *ND* and *BC* of each feedback loop were calculated.
(ii) Network motif identification: I identified a large set of 3-node feedback loops ($m = 444$; 213 positive, 228 negative and 3 neutral) from the E2F1 MIM using the Cytoscape plugin NetDS [136] (v3.0). The complete lists of FBLs are provided at[9].
(iii) Gene prioritization score: *GP* is a method to identify important genes or proteins in a network based on the network structure and a set of known disease markers using a global network propagation method. For the *GP*, I first selected the known EMT markers (including VIM, ZEB1, ZEB2, SNAI1, SNAI2, CDH1, TWIST1, TWIST2) in our map as proposed in Lanouille *et al.* [144] and then calculated the score for all the nodes using a random walk with restart algorithm implemented in the Cytoscape plugin GPEC [145]. In diffusion methods, random walk with restart is the most popular disease gene prediction method [146, 147], where nodes that are most frequently traversed by random walks are assumed to be important, *i.e.,* may be associated with a disease or a phenotype [76]. I calculated the *GP* score of nodes based on these EMT markers. Based on *GP* score of each node, the average *GP* score for every motif is calculated, a list of high *GP* scores nodes is available at[10].
(iv) Association with disease pathway: I chose association of a motif with *DP* as one of the parameter for motif ranking. The idea is to determine the association of motifs with a certain disease pathway by checking if any of the nodes in a motif is part of the pathway. (*DP*) values range from zero to the total number of nodes in a motif. For example, if none of the nodes in a 3-node motif is present in the disease pathway the value of this parameter is zero. Conversely, if all the nodes are present then the value of the parameter is 3. For this study, I used KEGG's cancer disease pathway (KEGG: hsa05200) to find the association of a motif with cancer and based on each node association with the *DP* (the average *DP* score for every motif is calculated).
(v) Fold change expression profile: I integrated gene expression profiles of two cancer cell lines for E2F1-driven highly aggressive tumors in bladder and breast. We used the ArrayExpress database (ArrayExpress accession number: E-MTAB-2706) [148] to find suitable gene expression data in non-invasive and invasive bladder and breast cancer cell lines. In particular, we used RT-4 as non-invasive and UM-UC-3 as invasive cell lines in bladder cancer, while MCF-7 as non-invasive and MDA-MB231 as invasive cell lines in breast cancer. Differential expression analysis was performed using the DEseq R-package with $method = `blind'$ and $fitType = `local'$ (v1.22.1) [149]. Furthermore, I calculated the absolute average fold change for a motif based on the change in expression values of each node in non-invasive to invasive phenotype.

[9] https://sourceforge.net/projects/khan-phd/files/Supplementary-file-3.xlsx/download
[10] https://sourceforge.net/projects/khan-phd/files/Supplementary-file-2.xlsx/download

(vi) Motif ranking: Using a multi-objective optimization function (Equation 2.2), I designed an algorithm, which calculates scores for each motif based on based on which they are ranked. The scores are calculated based on objective functions (*i.e., ND*, *BC*, *DP*, *GP* and *FC*) and weights w_k:

$$F_{M_{jk}} = \frac{w_{1k}}{2} \cdot \frac{\langle ND \rangle_j}{\max(ND)} + \frac{w_{1k}}{2} \cdot \frac{\langle BC \rangle_j}{\max(BC)} + w_{2k} \cdot \frac{\langle DP \rangle_j}{\max(DP)} + w_{3k} \cdot \frac{\langle GP \rangle_j}{\max(GP)} + w_{4k} \cdot \frac{\langle |FC| \rangle_j}{\max(FC)} \quad (2.4)$$

Here $F_{M_{jk}}$ is the score of each motif ($M_j, j = 1, \ldots, m$) for different weighting scenarios ($w_k, k = 1, \ldots, 13$) as given in Table 2.2. w_{1k} to w_{4k} are weighting factors pounding the importance of given objective function (*ND*, *BC*, *DP*, *GP* and *FC*).

In order to have the uniform range of values for each property, the objective functions are normalized to their maximum values. For example, for $\langle ND \rangle_j$, is normalized by dividing each value by the maximum value of ND ($\max(ND)$) in the motif list. In order not to over emphasize topological properties in motif prioritization, I assigned half of the weighting factor to $\langle ND \rangle$ and $\langle BC \rangle$. Further, to approximate the Pareto set of all non-dominated motif rankings, I iteratively modified the values of the weighting factors w_{lk} and computed the score $F_{M_{jk}}$ for each motif.

A complete list of the selected motifs along with structural and biomedical parameters, as well as their scores are provided in Supplementary-file-3 at[11]. Further, I selected the top ten scored motifs from each Pareto set. In this way, I obtained 32 and 28 non-redundant motifs associated with an invasive phenotype in bladder and breast cancer, respectively.

(vii) Derivation of the core-regulatory network: The core-regulatory network is obtained by merging the sets of top ranked motifs resulted in three disjoint sub-networks in both tumor types, which I connected by reviving interactions among the nodes from the E2F1 regulatory network. In this way, tumor-specific (bladder and breast cancer) core-regulatory networks are obtained as shown in Figure 2.5. Both core-regulatory networks include the transcription factors E2F1-3 and the cell cycle regulators RB1, MYC, CDKN2A, TP53/MDM2, SP1, FOXA1, FOXO3 and AKT1. Furthermore, both contain the enzyme SIRT1, which modifies targets like E2F1, TP53 and histones to silence their function. In addition, both core networks contain the CDH1 regulators SNAI1/2 and TWIST1 and interaction partners of CTNNB1 (AXIN2, LEF1).

In bladder cancer, FGFR1 and its downstream regulatory subunits of the protein phosphatase 2 (PP2; inhibitor of cell growth and division), and the pro-proliferative inhibitor of PP2, KIAA1524 (CIP2A) are part of the core-regulatory network, as well as the TGFBR1/2 downstream signaling molecules SMAD2-4 and their regulator ZEB1. The breast cancer core network includes FN1, which is associated with migration; FLT4, involved in angiogenesis; GSK3B, an anti-proliferative enzyme; KPNA2, a nucleopore transporter, and the EP300 associated transcriptional activator NCOA3.

[11] https://sourceforge.net/projects/khan-phd/files/Supplementary-file-3.xlsx/download

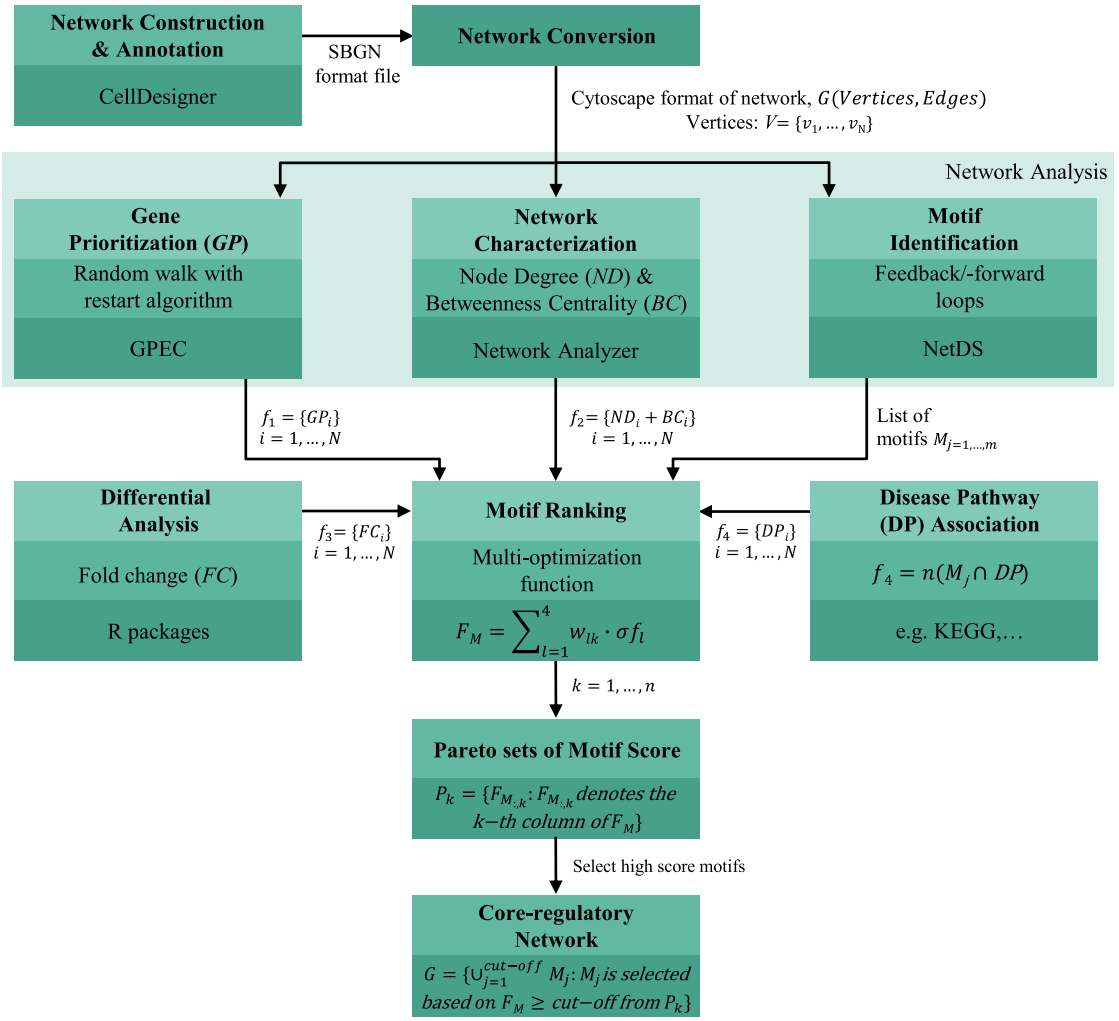


Figure 2.4. Methodology for the derivation of a core-regulatory network. First, a molecular interaction network is constructed in a machine-readable format (SBGN) by integrating and curating data from public repositories and the literature. The network is then converted into a format suitable for analysis using computational tool (Cytoscape). Network analysis is used to determine node properties (ND and BC), disease genes prioritization (GP), and identified network motifs.Theses network properties are combined with gene expression fold changes (FC) and association with disease pathways (DP) for ranking network motifs. This is done using multi-objective optimization function. Using different weighting (w_{lk}) scenarios, the ranking scores for each motif are maximized for objective functions individually and in combinations, which results in a Pareto set of motifs. The top 10 high scored motifs are selected from each Pareto set. A core-regulatory network is constructed from the union of top ranked motifs. The methodology is published in Khan et al., 2017 [124] and the figure is adopted from Khan et al., 2018 [150].

Additionally, we observed feedback loops concerning NFKB1 regulation (CHUK, NFKBIA; related to cell survival) in bladder cancer, and anti-apoptotic BIRC2/3 and pro-apoptotic TRAF1 factors in breast cancer, respectively. Interestingly, both loops are related to each other, since NFKB1 and BIRC2/3 are survival molecules and

NFKB1 activates BIRC transcription. The regulatory cores, which we consider as the drivers of invasive phenotypes contain 41 nodes and 107 interactions in bladder cancer and 35 nodes and 86 interactions in breast cancer.

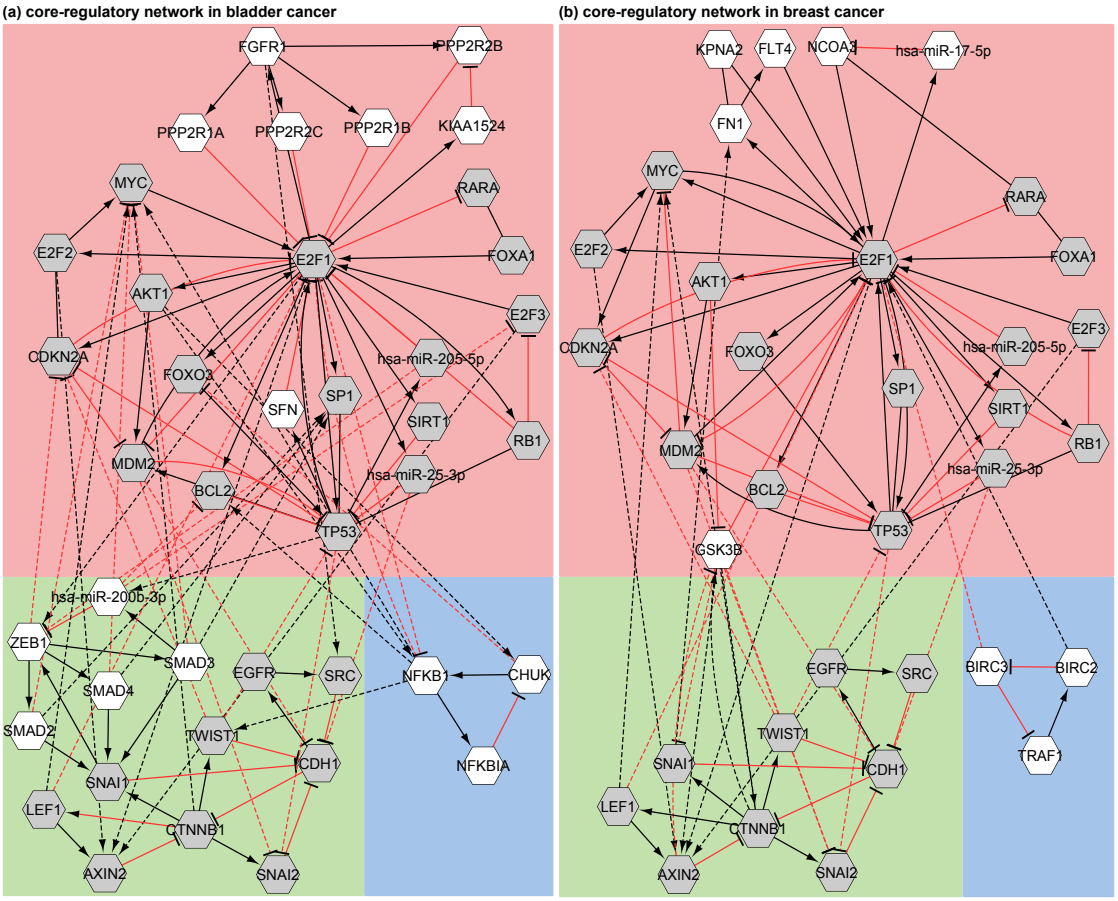


Figure 2.5. Core-regulatory networks that drive tumor invasiveness in (a) bladder and (b) breast cancer. The regulatory cores contain three disjoint subnetworks (shown in pink, green and blue background colors). Blue dotted lines represent direct interactions extracted from the comprehensive E2F1 regulatory network to interconnect these subnetworks. The networks were generated using Cytoscape plugin NetDS. Nodes in gray color are common in both regulatory core networks, whereas those in white color are unique. The figures are adopted from Khan et al., 2017 [124].

2.2.3. Weighting schemes for motif ranking

I chose five different sets of 13 weighting scenarios (*i.e.*, $k = 1, \ldots, 13$), each giving more importance to one or another objective function in the Equation 2.4. The weighting scenarios are shown in Table 2.2. In the first set, only one objective function was given importance for ranking. In the sets 2-4, I considered two, three and four objective functions respectively and applied consistently higher weights to the absolute expression fold change of the motif to identify tumor type/process-specific top ranked motifs. In

the last set, we assigned equal weights to all the objective functions considered.

Sets	w_1	w_2	w_3	w_4
Set 1	1	0	0	0
	0	0	1	0
	0	0	0	1
Set 2	1/4	0	0	3/4
	0	1/4	1	3/4
	0	0	1/4	3/4
Set 3	1/8	1/8	0	3/4
	1/8	0	1/8	3/4
	0	1/8	1/8	3/4
Set 4	1/16	1/16	1/8	3/4
	1/16	1/8	1/16	3/4
	1/8	1/16	1/16	3/4
Set 5	1/4	1/4	1/4	1/4

Table 2.2. Weighting scenarios for motif ranking [124].

2.3. Summary of results and discussion

Using the proposed workflow, I identified E2F1-mediated core-regulatory networks for the regulation of epithelial-mesenchymal transition in bladder and breast cancer. The core-regulatory networks are comparatively small, which makes them amenable for analysis using dynamical systems theory to provide deeper insights into the network behavior in response to different stimuli and perturbations.

In the context of our work, Calzone and coauthors [109] reconstructed a comprehensive map of the E2F transcription factor family. Their work focused on the differing roles of E2F family members in the cell cycle, reflecting the complex interplay between the E2Fs, RB1, its homologs RBL1 and RBL2, the cyclins/cyclin dependent kinases, and cell cycle arresters. In contrast, our map sets the main focus on the newly discovered role of activating members of the E2F family (E2F1-3) in cancer development and progression, with an emphasis on pro- and anti-apoptotic (survival), angiogenic as well as EMT relevant functions. We included additional key players connected to E2F1 directly or through its neighbors along with a post-transcriptional layer of miRNAs in the context of cancer. Interestingly, the majority of the components in the map by Calzone and co-workers are included in our map.

The underlying idea of my integrative workflow is that the structural and data-driven analysis of biochemical networks allows the identification of key functional modules, here named "core-regulatory network", composed of regulatory motifs and critical molecular interactions, which can have significant impact on the regulation of cellular processes, who's dysregulation results in complex diseases. This work presents a precedent of combining network-based high-throughput data analysis, network

reduction and Boolean modeling. Existing work either focuses on network-based analysis [140, 151] or Boolean network construction and simulation [152, 153]. Further, my methodology includes an innovative element in terms of network reduction, namely the use of an algorithm employing multi-objective optimization concepts to rank and select key regulatory motifs, based on network-topological features, biomedical information and expression profiles.

Motif identification methods have been previously used to recognize key network regulators. For example, Zhang and coworkers ranked network motifs using gene expression data to detect breast cancer susceptible genes [103] and Koschützki and coauthors used motifs with various network topological parameters to identify important nodes in a biochemical network [154]. I here introduced a new motif ranking scheme using a weighted multi-objective optimization function that integrates topological (*e.g.,* node degree and betweenness centrality) and non-topological (*e.g.,* gene expression, gene prioritization) properties. Topological properties account for the structural importance of the nodes, and non-topological properties for their cancer-type and context-specific relevance. I used multiple weighting scenarios in the multi-objective optimization function to provide motif ranking unbiased as much as possible regarding the properties assessed in the function. The proposed multi-objective optimization function and ranking scheme can easily be extended to add new information layers in the workflow. Thus, I think the method proposed can be used for investigating other cancer networks besides those focused on the E2F family discussed here.

Chapter 3

Logic-based dynamical systems analysis of the E2F1 tumor invasion network

The models described in this chapter are the ones published in the following publication:

- **Khan FM**, Marquardt S, Gupta SK, Knoll S, Schmitz U, Spitschak A, Engelmann D, Vera J, Wolkenhauer O, Pützer BM. Unraveling a tumor type-specific regulatory core underlying E2F1-mediated epithelial-mesenchymal transition to predict receptor protein signatures. Nat Commun. 2017 Aug 4;8(1):198. https://doi.org/10.1038/s41467-017-00268-2.

Synopsis
Mathematical models are a well-established tool to elucidate the dynamics of complex processes in living cells. Depending on the available information, the structure and size of a network behind a system and the purpose of modeling, a number of modeling formalisms are available. For a large system of interactions with sparse data, the coarse grained qualitative models are suitable tool, which can be encoded using a logic-based formalism. In this chapter, I present the development, calibration and analyses of a logic-based model (in particular Boolean models and its multi-valued variant) of epithelial-mesenchymal transition (EMT) regulation in bladder and breast cancer. In silico simulations of stimulus-response behavior identified molecular EMT signatures for both cancers. Further, I performed in silico perturbation experiments to predict potential drug targets. The predicted signatures and drug targets have been validated through patient data and by in vitro experiments.

3.1. Background and motivation

The molecular details yielded by omics technologies and the availability of databases that collect, organize and share information about interactions among proteins, genes and miRNAs, enable researchers to model cellular processes as molecular interaction networks. These networks encoded in a computer-readable format (*e.g.,* SBGN and

SBML), using tools like Cytoscape and CellDesigner, turn them into knowledge-base that provide useful information, such as regulatory motifs and hub nodes [58, 59, 109, 110, 155]. However, their utility is confined to a static analysis, which does not help to explain "cause and effect" relationships, the ubiquitous phenomenon in biological systems [156]. Moreover, only from the network connectivity it is impossible to dynamically characterize complex network structures, like feedback/feedforward loops and cross-talks in network modules, which affect the regulatory behavior of a system. Nevertheless, networks are an important step towards understanding a system and providing a foundation for the development of dynamical models [5, 6, 79, 157]. The dynamical analysis helps in understanding the behavior of a system upon different stimuli; and supports the formulation of new hypotheses about the unknown aspect of a system upon internal or external perturbations [156, 158]. Using a systems biology approach (Figure 1.1), the iterative cycle of data-driven modeling and model-driven experimentations refine formulated hypotheses until they are validated.

Models are abstract representations of the reality, which provide a sense (*i.e.,* understanding the behavior) of the original system, depending on available information and the purpose of modeling [159]. Modeling, is the process of the creation and use of a model [159]. Mathematical models describe certain aspects of the reality (*i.e.,* processes in a cell) in terms of functions or equations, which contain variables and parameters (Equation 3.1):

$$v(X_1, \ldots, X_N;\ k_1, \ldots, k_N) \tag{3.1}$$

where v is the function that evaluates, *e.g.* the temporal behavior of systems having state variables X_1,...,X_N and parameters k_1,...,k_N. Variables are the quantity of interest in model analysis, which typically change over time, *e.g.* the concentration of proteins, RNAs or metabolites in a cell. Parameters are quantities, which are fixed for a given computational experiment to characterize a specific quantitative behavior of a model. Parameter values are typically derived from literature, databases such as BioModels [20], BRENDA [160] and SABIO-RK [19], or can be estimated from experimental data by calibrating the model to recapitulate the biological processes under investigation [17, 18]. After calibrating the model with experimental data, one can perform large sets of repetitive *in silico* experiments for many different conditions, which may be time-consuming and expensive with wet-lab experiments.

Models can be created at different levels of abstraction, ranging from coarse grained qualitative models of (large) sub-cellular processes to detailed quantitative models of a (small) functional module. Logic-based models are popular to qualitatively analyze biological systems [91, 92, 161–163], where experimental data is frequently sparse (not everything can be measured, few time points) and uncertain (lack of replicates and precision), which make the identification of kinetic parameters difficult [6, 82, 87–89, 163]. They are qualitative and do not require detailed

quantitative parameters [6, 82, 87–92, 163]. Moreover, they provide predictive testable hypotheses, which are especially valuable in poorly understood large-scale systems [93, 94]. Logic-based formalisms make simplifying assumptions and assign discrete states to components of a system based on a set of logical rules that are derived from their regulatory interactions. The simplest logic-based model is the Boolean model popularized by Kauffmann (1969) [164], where network components can be assigned two possible state values (*e.g.,* 1 or 0; representing active or inactive; ON or OFF states respectively).

Being qualitative in nature, logic-based models are conceptually and computationally simple, and are easier to calibrate against data, yet they provide the ability to capture cause-effect relationships [165, 166]. Therefore, they have received significant appreciation in modeling large-scale biochemical networks [86, 93, 94]. They have been successfully used to generate hypotheses and find explanations for various biological processes, ranging from the cell cycle [167–169] to cell differentiation [170, 171], cell death [91, 172, 173], and cell migration [152, 174]. Further, logic-based models have been successfully applied to identify possible alterations in pathways [162, 175] and proposed different combinations of drug targets through *in silico* perturbations [84, 161, 176]. Such models can be integrated with patient-specific data to derive a personalized model, which can recapitulate and analyze the gain-of-function or loss-of-function of mutated genes in specific disease pathways. As a result these models can aid clinical decision-making or improve therapy towards a personalized design [177, 178]. Like other modeling formalisms, logic-based modeling consists of three main steps: (i) model construction; (ii) model calibration; and (iii) model analysis.

3.1.1. Model construction

In logic-based modeling a network is represented as a graph containing nodes and edges, where nodes represent molecular species and edges depict the type of effect that one species exerts on the state of another species (*e.g.,* activation or inactivation) see Figure 3.1**a**. The state of each species is determined by a logical function that links the incoming effect to its state. In Boolean models, nodes ($X_{1,\ldots,n}$) of a network correspond to Boolean variables that can have values of either 1 or 0, and edges define the type of interactions (*e.g.,* activation or inhibition). The state of a node at the succeeding time step $X(t+1)$ is determined by a Boolean function (BF) of the current state $X(t)$ of nodes regulating it (Equation 3.2):

$$X_i(t+1) = BF(X_1(t), X_2(t), \ldots, X_n(t)) \tag{3.2}$$

Boolean functions determine the states of a node using Boolean operators/gates: 'NOT' (denoted by !), 'OR' (denoted by $\vee$) and 'AND' (denoted by $\wedge$). Where, the 'NOT' gate is used to model inhibition. For example, the first function (BF1) in

Figure 3.1**b**, BF_1 encodes the negative effect of X_3 on X_1 activity. In case of an independent regulation of more than one molecular species on a target species, we connected them with an 'OR' gate. For example, in BF_3 and BF_4, X_1 and X_2 can independently activate X_3 and X_4. Finally, an 'AND' gate represents the interaction, where more than one species collectively regulates the expression level of a mutual target species. BF_2, for example, encodes the collective influence of X_1 and X_3 on X_2.

Multi-valued logic is used as an alternative to encode the behavior of networks, which allows to represent the state of a node in more than two ordinal levels [91, 155]. Illustrated in Figure 3.1**c**), I show how functions ($MF_{1\text{-}4}$) following a multi-valued logic can be used to model ordinal levels of a phenotypic outcome of the interaction network. The phenotype is modeled in four ordinal levels (from 0 to 3) based on the Boolean state of factors regulating it. Representing a phenotype in multiple ordinal levels helps to assess the aggregated effect of various network components (*i.e.*, X_2, X_3, X_4) on it.

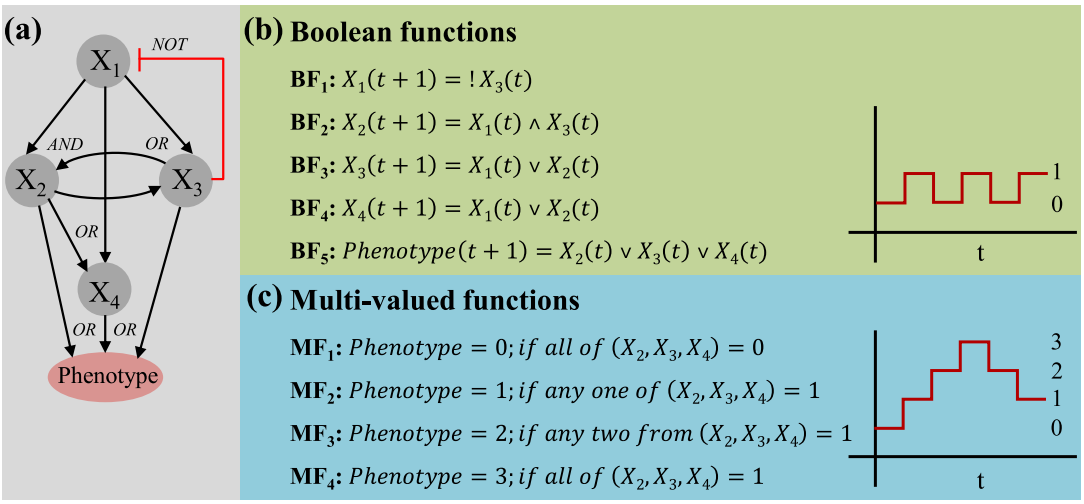


Figure 3.1. Logic-based representation of a biochemical network. (**a**) The network consists of four nodes (X_{1-4}), the interactions of which are linked to a certain phenotype. (**b**) Boolean functions ($BF_{1\text{-}5}$) of nodes that derive their subsequent state based on the present state of the nodes regulating them using logic operators (NOT (denoted by !), OR (denoted by $\vee$), and AND (denoted by $\wedge$). (**c**) Multi-valued logic functions represent the phenotype in more than two ordinal levels. The figure is adopted from Khan et al., 2017 [124].

3.1.2. Model calibration

Once a biochemical system has been converted into a Boolean model using logical operators, the next step is the model calibration against available data. Model calibration makes the simulations match as good as possible to experimental observations. Calibration of logic-based modeling mainly decide about the use of 'NOT', 'OR' or 'AND' operator in a logical function. They can be estimated either based on prior knowledge about biological process or by training the model with

experimental data. A number of tools and methods are available that automatically generate and train logical models from data and prior knowledge encoded by networks [166, 179, 180]. For example, when there is more than one activator for a molecular species, one has to specify whether the species is active in the presence of either regulator or only if both are present. Suppose, we have expression level changes of genes X_1, X_2, X_3 and X_4, in a cancer versus normal condition (Table 3.1). If we train the model shown in Figure 3.1 with this data, the Boolean function of $X_1(t+1)$, $X_2(t+1)$ and $X_4(t+1)$ will remain the same as shown in Figure 3.1**b**, while that of species $X_3(t+1)$ will change from an 'OR' operator to an 'AND'. Figure 3.1 shows that species X_3 is regulated (activated) by X_1 and X_2. The data in Table 3.1 shows that the expression levels of regulator X_1 is up and X_2 down, while X_3 is also down, which means for the activation of X_3 both the regulators must be active. Such type of causal relationship is encoded by a logic operator 'AND', therefore, the 'OR' operator in BF_3 will be replaced by 'AND' *i.e.*, BF_3: $X_3(t+1) = X_1(t) \wedge X_2(t)$, which will improve model fitting to the data.

Calibrating the model against data does not only enrich the biological context of the model, but also significantly increase its predictive power [179]. Using this approach, I calibrated logic-based models of EMT in bladder and breast cancer against differential expression data from non-invasive to invasive cancer cell lines. The model calibration substantially increased the model fit to the data; and their simulations successfully identified molecular signatures for disease-specific phenotypes, which were subsequently validated by *in vitro* experiments (see Section 3.2).

Species	**Expression level**
X_1	High ▲
X_2	Low ▼
X_3	Low ▼
X_4	High ▲

Table 3.1. Hypothetical expression levels

3.1.3. Model analysis

After model construction and calibration, the next step is to dynamically analyze the behavior of the biological system for different stimuli or perturbations using *in silico* simulations. Stability analysis is a widely used method to analyze the dynamical behavior of the biological system, especially the input-output or the cause-effect behavior [135, 181]. For a given set of inputs or initial conditions and perturbations, it describes the long term behavior of a system over time. Such analyses are well suited for model simulations

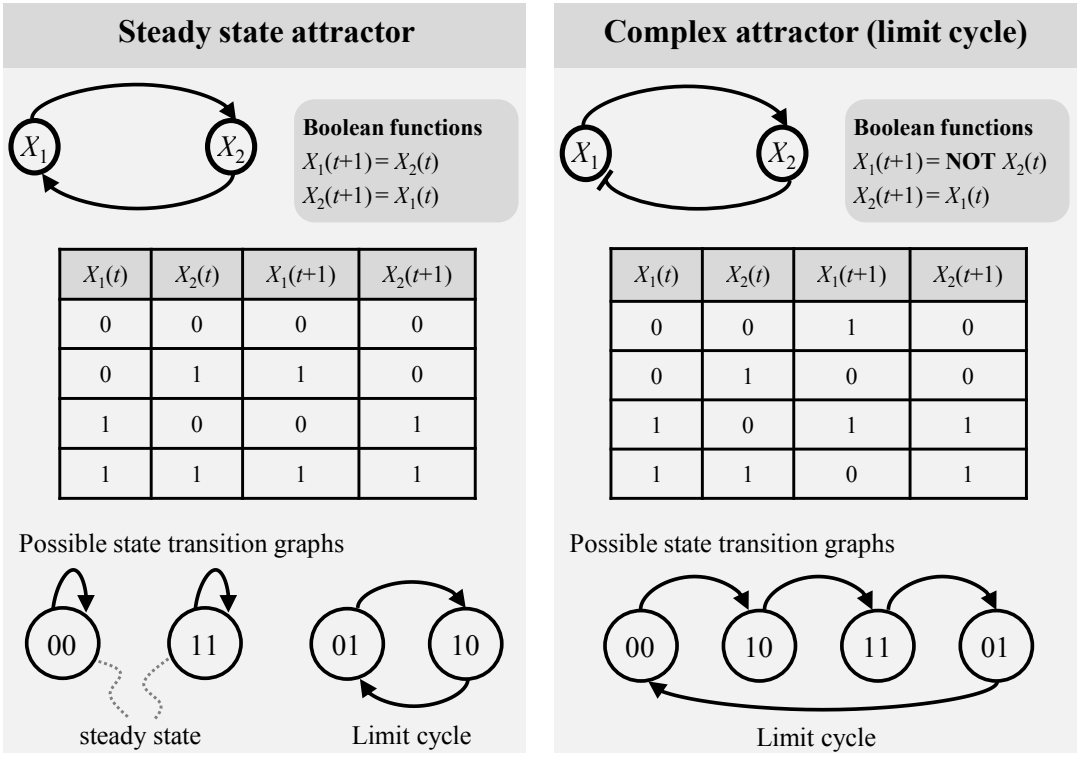
to recapitulate the transition from one cellular state to another [92, 173, 174, 177, 182], which is the output of many complex processes in a biological system.

In logic-based modeling the system dynamics can be represented in the form of a *state transition graph*, which shows the state of each component in the network and the possible transitioning among them using either a synchronous or an asynchronous update scheme. In a synchronous update, the states of all nodes update from the current state '$X(t)$' to the next state '$X(t+1)$', simultaneously. The transition between the states is deterministic and each system state has at most one subsequent state [135] (see Figure 3.2). Whereas, in the asynchronous scheme the states of a system are updated according to the timescale of individual biological events and systems can have more than one subsequent state (see Appendix C). Asynchronous updates can be non-deterministic (*i.e.,* stochastic scheme, which choses random events for a system update) or deterministic where the timescale of every event is known (which in many cases is not possible) [181].

In stability analysis, the model starts from an initial condition and evolves over time by transitions from one state to another until eventually reaching a fixed point (steady state) or complex attractor. An attractor is a state/condition towards which a system tends to evolve, which in biological systems is usually correspond to steady state activation of cellular components associated with cellular phenotypes (such as proliferation or apoptosis) or oscillatory behavior (such as the cell cycle or circadian rhythms) [183–185]. In fixed point attractors (Figure 3.2a), the states of the system get stuck at one point and do not change over time; while in complex attractors (limit cycle) (Figure 3.2b), the system oscillates among a set of states. René Thomas proposed important conjectures about the structure of networks that generate either multiple steady states or oscillation: (i) a positive circuit in the interaction graph is the necessary condition for multiple stable states; (ii) a negative circuit in the interaction graph is the necessary condition for oscillatory behavior [186]. The existence of a steady state attractor does not depend on the update scheme, while complex attractors are highly dependent on the update scheme [82, 135, 181].

In Boolean models, a network with n nodes has 2^n possible states. In a review article, Wang *et al.* (2012) provides a survey of applications, challenges and their possible solutions of attractor analysis Boolean models [82]. Attractor analysis of a system enables to study the behavior of a system in response to different stimuli and perturbation, which helps to predict the activity states/levels of components in a certain phenotypical trait [173, 187]. For example, logic-based model of attractor analysis was used to reveal the mechanism of cell fate decision in response to cytokine stimuli, such as, TNF and FASL [173]. The attractors were correspond to survival, necrosis and apoptotic phenotypes, which were nicely characterized by the activity levels of specific components (*i.e.,* genes) leading to the respective phenotypes. In another study, a Boolean model of the p53 network used attractor analysis to investigate the network response to DNA damage [187]. Attractors represented the p53 regulated cell fate, such as cell cycle arrest and cell death, and identified particular

combination states of molecules that corresponded to a specific cellular outcome. Furthermore, targeted perturbations revealed that the MDM inhibitor nutlin-3 triggers p53 oscillation, which induced cell cycle arrest, while combining nutlin-3 with inhibition of Wip1 resulted in sustained increase in p53 activity and triggered cell death, which were subsequently validated in MCF7 breast cancer cell line. For large biochemical networks, the identification of all possible 2^n attractors is difficult (the number of attractors increases exponentially) and intractable; and asynchronous update schemes tends to make the analysis even more complicated [82]. Many methods are proposed to tackle this problem, for example, one solution is to use network reduction techniques to simplify the network while preserving the essential dynamics [82].



a Steady state attractor

b Cyclic attractor

Figure 3.2. Attractor analysis for (a) a steady state attractor and (b) a complex attractor. (a) The interaction graph represents a positive circuit where molecular species X_1 activates species X_2, and X_2 activates X_1. The logical relation for each species is represented by a Boolean function. For the synchronous state update, the state transitions from $X(t)$ to $X(t+1)$ are shown in the state table, and also visualized in state transition graph. State transition graphs show that the systems has two stable states (00 and 11) which upon state updates never change. (b) The interaction graph represents a negative feedback loop where species X_1 activates species X_2 while X_2 represses the expression of X_1. The logical relation for each species is represented by a Boolean function. Starting from the initial state 00, for synchronous update, the sequence of states is visualized in the state transition graph. The state transition graph depicts a limit cycle behavior of the system dynamics.

In addition to attractor analysis, logical steady state (LSS) analysis has been proposed [188], where steady states of the components are determined by propagating a sustained input signal to the output layer until no further state updates occur. The LSS of each network element is consistent with the value of its associated logical function. Once the system reaches a LSS, no state of a node switches anymore and the system remains in this state. The LSS of a dynamical system is equivalent for both synchronous and asynchronous state updates. For a given set of input nodes, LSS determines unique steady state values of nodes for large-scale biochemical networks. LSS can be used to study how the signal propagates through the network, which helps to explore the interplay of pathways involved in normal cellular processes, disease states, and upon perturbations allows to determine important players for phenotype traits [90, 174, 189]. Saez-Rodriguez *et al.* (2007) developed a large-scale Boolean model for T-cell receptor signaling activation and using LSS analysis, they deciphered the unforeseen signaling events for CD28 and Fyn perturbations, which were subsequently validated by experiments [90]. Further, LSS analysis can explain stimulus-response behavior of large networks [84, 91]. For example, using LSS analysis for a given set of input stimuli, large-scale qualitative model of EGFR/ErbB receptors predicted the necessary conditions for JNK and p38 cascade activation in response to PI3K dependent/independent signaling pathway in primary hepatocytes and HepG2 [84].

3.2. Logic-based model of E2F1 in tumor invasion

Cancer cells have the ability to desciminate from primary tumor, move to nearby and distant organs, and establish new colonies [190]. Invasion and metastasis development are highly complex, and their exact mechanisms and molecular determinants are largely unknown. The transcription factor E2F1 has recently been identified as a key regulator of processes involved in tumor invasiveness and metastasis [119, 122, 191]. Cells gain migratory and invasive properties through an extremely complex process called epithelial-mesenchymal transition (EMT) in response to various internal and external signals [192–194]. To understand the functional relatedness of E2F1 with EMT, I developed a logic-based model of the core-regulatory networks identified in Chapter 2. I carried out logical steady state analysis for different expression levels of E2F1 and receptor molecules to identify signatures for EMT in bladder and breast cancer. Furthermore, I performed *in silico* perturbations to identify potential therapeutic targets, which can be manipulated to counteract invasion.

3.2.1. Model construction

I developed logic-based models of the core-regulatory networks (shown in Figure 2.5) of bladder and breast cancer to evaluate their input-output relationship for different expression levels of E2F1, receptor stimuli, and different knock-in/out perturbations.

I established logical rules based on the network structure and the inspection of the available literature about the interactions using logic-based modeling formalism described in Section 3.1.1. The obtained model contains three layers: (i) an input layer, (ii) a regulatory layer, and (iii) an output layer representing the phenotype Figure 3.3 and 3.4. The input layer of the logic-based models contain E2F1 and all receptors present in the regulatory core. In bladder and breast cancer, two common receptors are part of the input layer: (i) EGFR and (ii) the Retinoic Acid Receptor Alpha (RARA). Additionally, we found FGFR1 only in the regulatory core of bladder cancer. The models were constructed in the Process Modeling Tool (ProMoT) [195] and the layouts were arranged in yEd graphical editor[1].

In order to analyze the effects of all possible input signals on the core-regulatory networks, I expanded the input layers by including additional receptors present in the E2F1 map (Figure 2.2). I took the receptors that are directly connected to nodes constituting the regulatory cores. Thus, I included TGFBR, which is connected to SMADs, and the Chemokine (C-X-C motif) Receptor 1 (CXCR1) connected to ZEB1 and SNAI1 in the bladder cancer model. Similarly, I expanded the breast cancer input layer with the Hyaluronan-Mediated Motility Receptor (HMMR) connected to FN1; the TGFBR connected to SNAI1 and SNAI2; the InterLeukin 1 Receptor type I (IL1R1) and the Thyroid Hormone Receptor Beta (THRB) connected to TRAF1 and MYC, respectively. For bladder cancer, I selected TGFBR1 and for breast TGFBR2 due to their tissue-specific expression profiles. I derived Boolean functions for the input signals and their propagation through the nodes constituting the regulatory layer (see Appendix D and E).

The output layer of the models comprises a unique node that represents the EMT process as the driver of the invasive phenotype, which is modeled by a multi-valued logic function involving the EMT markers present in the regulatory core. Here, the multi-valued function accepts four ordinal levels, ranging from 0 (no EMT) to 3 (high EMT).

3.2.2. Model calibration with expression data

I used qualitative information based on expression fold changes, from invasive (UM-UC-3; MDA-MB231) to non-invasive (RT-4, MCF-7) cell lines of bladder and breast cancer, to approximate the activation level of nodes having multiple regulators [166, 179, 180]. For example, in our bladder cancer model, CDH1 is inhibited by multiple molecules including SNAI1, SNAI2, TWIST1, SRC, miR-25 and MDM2. Differential expression analysis in bladder cancer showed that the expressions of SNAI2, SRC, miR-25 and MDM2 were down-regulated in the invasive cell line, while SNAI1 and TWIST1 were up-regulated. It is well established that CDH1 is down-regulated in most invasive cancer phenotypes [144, 194, 196], therefore, I considered SNAI1 and TWIST1 more relevant for the regulation of CDH1 than others. To encode the decisive role of SNAI1 and TWIST1

[1] https://www.yworks.com/products/yed

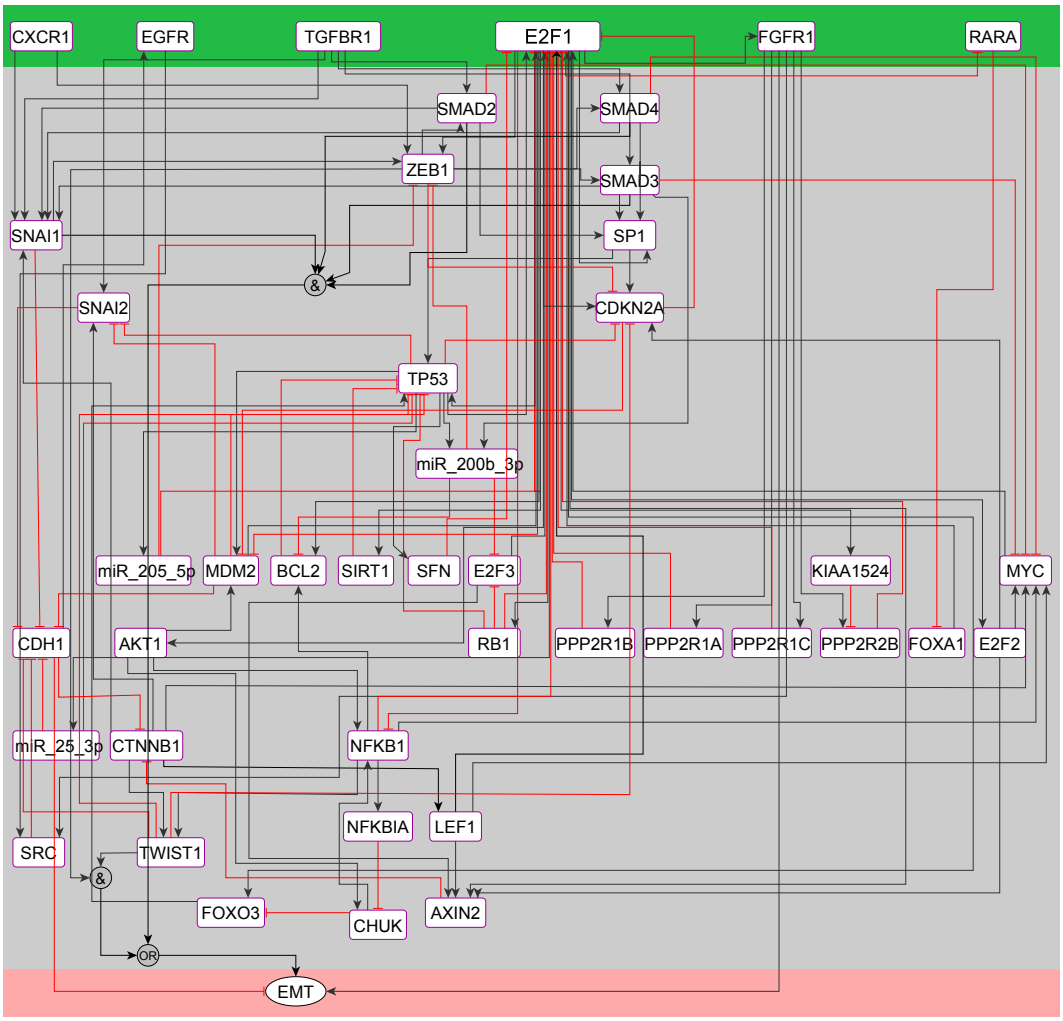


Figure 3.3. Logic-based model of the core network regulating invasive phenotype in bladder cancer. The black and red lines represent the type of interactions (*i.e.,* activation by black and inhibition by red) among the interacting components. The microRNAs, receptors and proteins are represented by rectangular boxes, while the EMT phenotype is represented by an ellipse in the output layer. The model is divided into three layers, *i.e.,* the input layer (green), the regulatory layer (gray) and the output layer (pink). The figure is adopted from Khan et al., 2017 [124].

in the regulation of CDH1, I used an 'AND' gate to connect them with the rest of the factors, see Equation 3.3. For logical rules of both models, see Appendix D and E.

$$CDH1 = !(SRC \wedge SNAI2 \wedge miR{-}25 \wedge MDM2) \vee (!\,SNAI1) \vee (!\,TWIST1) \quad (3.3)$$

Further, I modeled the EMT phenotype by multi-valued logic functions, representing its four ordinal levels (from 0, accounting for inactive EMT, to 3, accounting for full activation) based on the sum of Boolean states of factor 1: [SMAD2/3/4, SNAI1, ZEB1, TWIST1]; factor 2: CDH1; and factor 3: FGFR1, with the following structure:

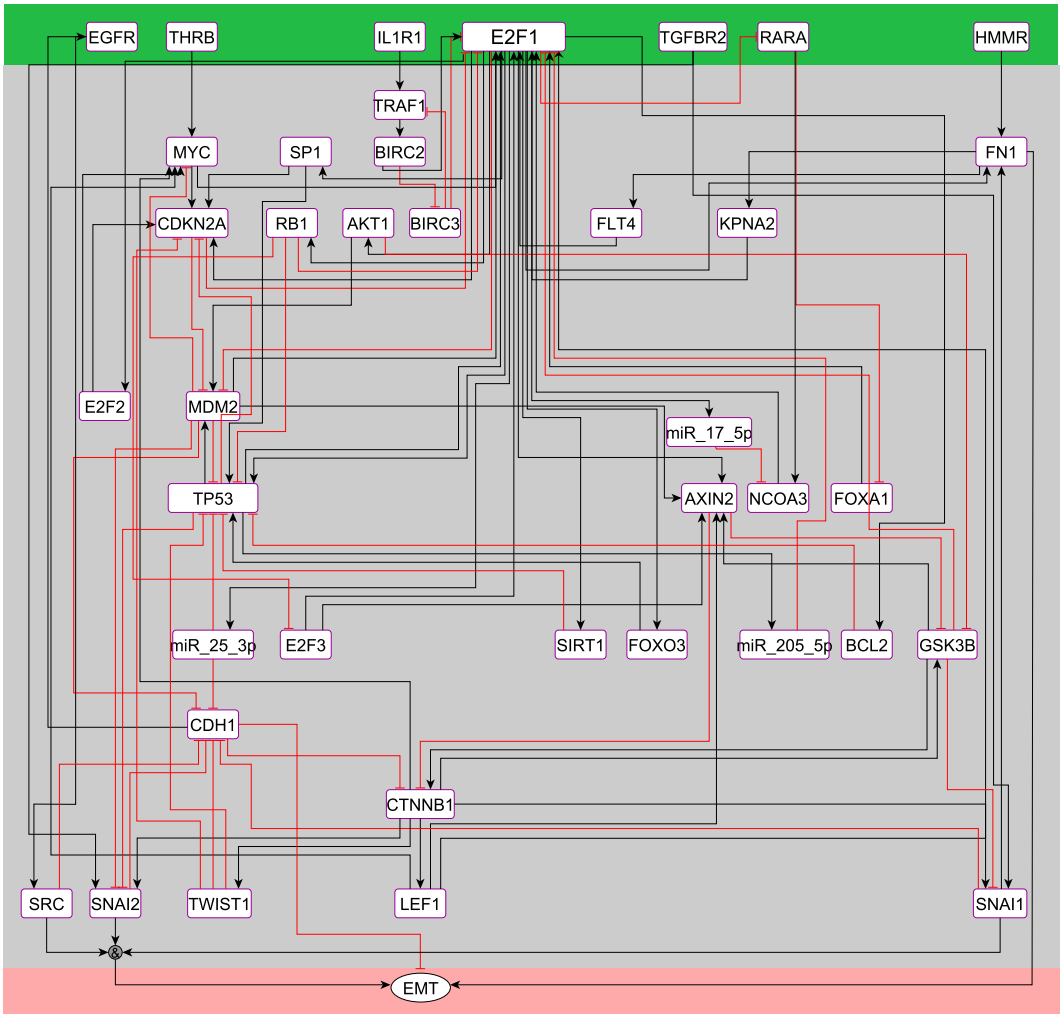


Figure 3.4. Logic-based model of the core networks regulating invasive phenotype in breast cancer. The black and red lines represent the type of interactions (*i.e.,* activation by black and inhibition by red) among the interacting components. The microRNAs, receptors and proteins are represented by rectangular boxes, while the EMT phenotype is represented by an ellipse in the output layer. The model is divided into three layers, *i.e.,* the input layer (green), the regulatory layer (gray) and the output layer (pink). The figure is adopted from Khan et al., 2017 [124].

$$EMT = [(SMAD2/3/4 \wedge SNAI1) \vee (ZEB1 \wedge TWIST1)] \wedge (!\, CDH1) \wedge FGFR1 \tag{3.4}$$

The motivation to select these factors as drivers of EMT was due to the fact that CDH1 is a widely accepted driver of EMT together with SMAD2/3/4, SNAI1, ZEB1 and TWIST1 [144]. I also considered receptor proteins as decisive factors determining the EMT phenotype, if they are present in the regulatory core, highly overexpressed in the invasive phenotype and not connected to any of the EMT markers (*e.g.,* FGFR1 in bladder cancer) [197, 198].

3.2.3. Predictive model simulations

I determined the LSS of each variable in the model for different initial values of the input nodes using the software tool CellNetAnalyzer [199]. I consider two sets of scenarios for simulations characterized by: (i) high expression of E2F1 (*i.e.,* $E2F1 = 1$); and (ii) low expression (*i.e.,* $E2F1 = 0$) in all possible Boolean combinations of receptor molecules in the input layer. For all possible Boolean combinations, I obtained 64 input vectors for bladder cancer and 128 for breast cancer. For these inputs, I simulated the models to evaluate their impact on the level of EMT (Table 3.2). The simulation results suggest that when E2F1, TGFBR1 and FGFR1 are simultaneously active, bladder cancer cells become highly invasive (*i.e.,* $EMT = 3$). A similar effect was observed in breast cancer when E2F1, TGFBR2 and EGFR are simultaneously active.

Furthermore, I carried out *in silico* perturbation experiments to identify important nodes that can be exploited for therapeutic interventions. Perturbation experiments were performed for a highly invasive phenotype ($EMT = 3$) by changing the Boolean state of each node in the 'regulatory layer' to reduce the invasiveness. To identify potential drug targets (*i.e.,* single target or combinations with minimum number of targets), I used single and double perturbations iteratively and observed the most prominent reduction of EMT in the latter case (simulation results are provided in an Excel file at[2]). The results of the perturbation experiments suggest that in bladder cancer: (i) double knockout of ZEB1 in combination with either SNAI1, TWIST1 or NFKB1; (ii) knockout of ZEB1 and activation of CDH1; or (iii) knockout of SMAD2/3/4 in combination with TWIST1 or NFKB1 reduces EMT to 1. In case of breast cancer double perturbation by silencing SRC, FN1, SNAI1, SNAI2 or activation of CDH1 in any of the combinations reduces EMT to 1 (see Table 3.3).

3.2.4. Validation of *in silico* predictions in cancer cell lines

The *in silico* simulations revealed a common effect of E2F1 and TGFB1 signaling on tumor invasiveness in both cancer types. More specifically, TGFBR1 and FGFR1 in combination with highly expressed E2F1 induce the most invasive phenotype in bladder cancer, while in breast cancer TGFBR2, EGFR and E2F1 triggers high levels of invasiveness. In order to validate the predicted influence of the receptors and E2F1 on the invasive phenotype, our collaborators (from the Institute of Experimental Gene Therapy and Cancer Research (IEGT)[3], University Rostock Medical Center) used chemical inhibitors and a shRNA-based approach to target these key players. Although, receptors EGFR and FGFR1 are highly expressed in both cancer types (Fig. 5b on page 160), but *in silico* simulations suggest their distinctive role in each cancer type, which was confirmed by *in vitro* experiments, see Fig. 5a on page 160. It clearly shows that inhibition of EGFR in UM-UC-3 (bladder cancer cells) and FGFR1 in

[2]https://sourceforge.net/projects/khan-phd/files/Boolean_in_Silico_perturbations.xlsx/download
[3]https://www.iegt-rostock.de/en/

(a) Bladder cancer

E2F1	TGFBR1	FGFR1	EGFR	CXCR1	RARA	EMT
0	0	0	0/1	0/1	0/1	0
0	0	1	0/1	0/1	0/1	1
0	1	0	0/1	0/1	0/1	1
0	1	1	0/1	0/1	0/1	2
1	0	0	0/1	0/1	0/1	1
1	0	1	0/1	0/1	0/1	2
1	1	0	0/1	0/1	0/1	2
1	1	1	0/1	0/1	0/1	3

(b) Breast cancer

E2F1	TGFBR2	EGFR	HMMR	THRB	IL1R1	RARA	EMT
0	0	0	0/1	0/1	0/1	0/1	0
0	0	1	0/1	0/1	0/1	0/1	1
0	1	0	0/1	0/1	0/1	0/1	1
0	1	1	0/1	0/1	0/1	0/1	2
1	0	0	0/1	0/1	0/1	0/1	1
1	0	1	0/1	0/1	0/1	0/1	2
1	1	0	0/1	0/1	0/1	0/1	2
1	1	1	0/1	0/1	0/1	0/1	3

Table 3.2. The effect of E2F1 and receptor molecules in different combinations on the EMT phenotype in (a) bladder and (b) breast cancer model. Table (a) is the summary of 64 *in silico* simulations of bladder cancer. Each row represents the result of eight simulations; where for the given Boolean state of E2F1, TGFBR1 and FGFR1, all the eight combinations of EGFR, CXCR1 and RARA result in the same phenotypical output. Table (b) is the summary of 128 *in silico* simulations of breast cancer. Each row represents the result of 16 simulations; where for the given Boolean state of E2F1, TGFBR2 and EGFR, all the 16 combinations of HMMR, THRB, IL1R1 and RARA results in the same phenotypical output. The results are published in Khan et al., 2017 [124].

MDA-MB231 (breast cancer cells) had a minor influence on invasiveness. Furthermore, the *in silico* predictive signatures for invasion were confirmed by inhibiting E2F1, TGFBR1/2 and FGFR1 in UM-UC-3 cells and E2F1, TGFBR1/2 and EGFR in MDA-MB231 cells, which exhibit a profound impact on the invasive behavior in the respective cancer type with the highest effect upon combined inhibition, see Fig. 5c on page 160. Moreover, our experimental partners examined the overexpression of E2F1 and receptor molecules EGFR, FGFR, TGFBR1/2 by stimulated their lingand to induce invasion in RT-4 (bladder) and MCF-7 (breast) epithelial cells. Figures 5d on page 160, show that overexpression of E2F1 or FGFR1 or TGFB1 (alone or in combinations) enhance the invasiveness in RT-4, and overexpression of E2F1 or EGFR

(a) Bladder cancer									
Signature			Double *in silico* perturbations in regulatory layer						Output
E2F1	TGFBR1	FGFR1	ZEB1	TWIST1	SNAI1	NFKB1	SMAD2,3,4	CDH1	EMT
1	1	1	1	1	1	1	1	0	3
1	1	1	0	0	1	1	0	1	1
1	1	1	0	1	0	1	0	1	1
1	1	1	0	0	1	0	0	1	1
1	1	1	0	1	1	1	0	1	1
1	1	1	1	0	1	1	0	1	1
1	1	1	1	0	1	0	0	1	1
(b) Breast cancer									
Signature			Double *in silico* perturbations in regulatory layer						Output
E2F1	TGFBR2	EGFR	SRC	FN1	SNAI1	SNAI2	CDH1		EMT
1	1	1	1	1	1	1	0		3
1	1	1	0	1	1	1	1		1
1	1	1	0	0	1	1	0		1
1	1	1	0	1	0	1	1		1
1	1	1	0	1	1	0	1		1
1	1	1	1	0	1	1	1		1
1	1	1	1	1	0	1	1		1
1	1	1	1	1	1	0	1		1
1	1	1	1	0	0	1	0		1
1	1	1	1	0	1	0	0		1

Table 3.3. ***In silico*** **simulations of tandem perturbations of highly invasive ($EMT = 3$) phenotype in (a) bladder and (b) breast cancer models.** The first rows in both cancer models represent the predicted molecular signatures for highest EMT level. The brown boxes represent the perturbed state of genes and their effect on the EMT phenotype is shown in the last column. The table is adopted from Khan et al., 2017 [124].

or TGFB1 (alone or in combinations) raise the invasive potential of MCF-7. It also confirmed that EGFR and FGFR1 have less impact on RT-4 and MCF-7 invasive potential respectively.

To validate the predictions of my *in silico* perturbation simulations (Table 3.3), our experimental partners decided to knockdown NFKB1 and SMAD3 in UM-UC-3 cells and SRC and FN1 in MDA-MB231 cells, instead of modulating the other well known EMT markers, such as SNAI1/2, TWIST1, ZEB1 or CDH1. They performed single and double knockdown of these genes and measured both, the transcriptional and EMT/MET response. Whereas removal of single genes resulted in a clear reversal of the EMT phenotype (reduced invasion) in both cell lines. The strongest effect was observed after double knockdown, as shown by their lowest invasive capacity, see Fig. 6 on page 161.

3.2.5. Validation of model predictions with patient data

To further validate the molecular signatures that were predicted, by *in silico* simulations, to regulate invasiveness in bladder cancer, we used data from a patient cohort ($n = 165$) [191], in which a correlation between E2F1 expression and superficial to invasive progression was observed (GEO id: GSE13507). We grouped the patients into high and low expression profiles of E2F1, TGFBR1 and FGFR1 from their respective median expression values. For each group, the progression-free survival probability was calculated and found that the survival probability was higher in the patient group with low expression of each molecule individually. Furthermore, we identified the subgroups of patients with high vs low expression of: (i) E2F1-FGFR1; (ii) E2F1-TGFBR1; and (iii) E2F1-FGFR1-TGFBR1. The progression-free survival probability of each subgroup revealed that patients with high expression of E2F1-FGFR1 have the lowest mean survival time (33.79 months), while those with low expression of E2F1-FGFR1-TGFBR1 have the best prognosis (93.35 months) among all the subgroups analyzed (see Kaplan-Meier plots in Figure 3.5a-c). Our analyses indicate that patients with a low expression of the molecular signatures survive more than twice as long as the patient subgroup with high expression.

To validate the molecular signature from the regulatory core in breast cancer, we used data from the TRANSBIG network (GEO id: GSE7390; $n = 198$) generated by Desmedt and colleagues (2007) [200]. Similar to bladder cancer patients, we observed that the progression-free survival probability of breast cancer patients was low for high expression levels of E2F1, EGFR and TGFBR2 in different combinations (Figure 3.5d-f). Interestingly, we observed the highest mean survival time (97.29 months) in the patient subgroup with low expression of E2F1-EGFR-TGFBR2. Similar to the bladder cancer analyses, the patient subgroup with high expression of all three components had nearly half the mean survival time (50.11 months) compared to the subgroup with low expression.

We further validated molecular signatures in large patient cohorts of TCGA bladder cancer (BLCA; $n = 426$) and TCGA breast cancer (BRCA; $n = 1218$) accessible through UCSC Xena http://xena.ucsc.edu. Based on the molecular signatures identified through model simulations, we successfully stratified bladder and breast cancer patients into early and advance stages of disease (p-value < 0.005) (Figure 3.6a,b). In order to assess the capability of our workflow to predict significant molecular signatures associated with invasive phenotypes, we generated 30 random signatures of three nodes from each of the regulatory cores and arbitrarily assigned high or low expression values. We observed that the molecular signature predicted for bladder cancer is the only one that distinguishes between the early and advanced stage of cancer (Figure 3.6c). In case of breast cancer, in addition to the predicted signature, some of the random signatures were also able to distinguish between aggressive and less-aggressive cancer types (Figure 3.6d). This might be due to the highly heterogeneous nature of breast cancers. Overall, our analysis

reveals that an invasive tumor phenotype in bladder cancer is driven by E2F1, TGFBR1 and FGFR1, while in case of breast cancer it is driven by E2F1, TGFBR2 and EGFR.

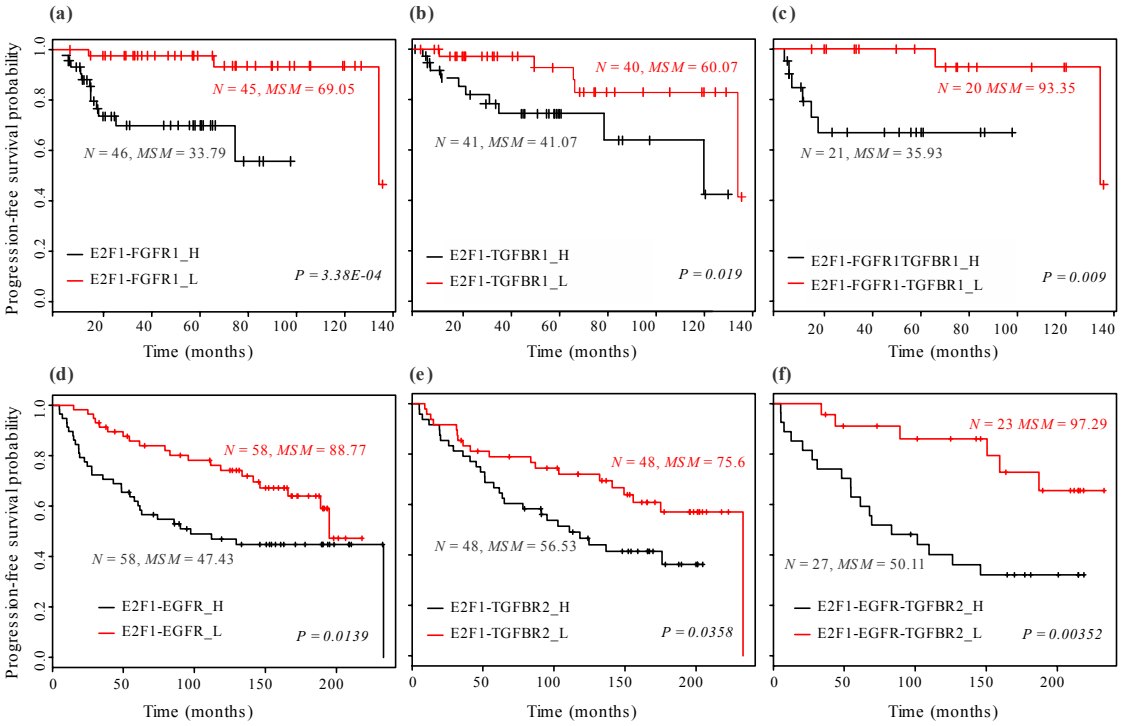


Figure 3.5. Kaplan-Meier plots of progression-free survival of patients with bladder and breast cancer. Plots **a-c** show the survival curves for patients with high and low expression of combined signatures (E2F1-FGFR1; E2F1-TGFBR1; and E2F1-FGFR1-TGFBR1) in bladder cancer patients, while plots **d-f** are for the combined signatures of E2F1-EGFR; E2F1-TGFBR2; and E2F1-EGFR-TGFBR2 in breast cancer patients. In both cases, patients with low expression of the *in silico* predicted signatures have high mean survival times and *vice versa*. High expression of molecular signature(s) is represented as '_H' (black curve) and low expression as '_L' (red curve). 'N' is the number of patients observed with high/low expression of molecular signatures and 'MSM' is the mean survival month from the patient group. p-values shown in the figures are from a log rank test. The figure is taken from Khan et al., 2017 [124].

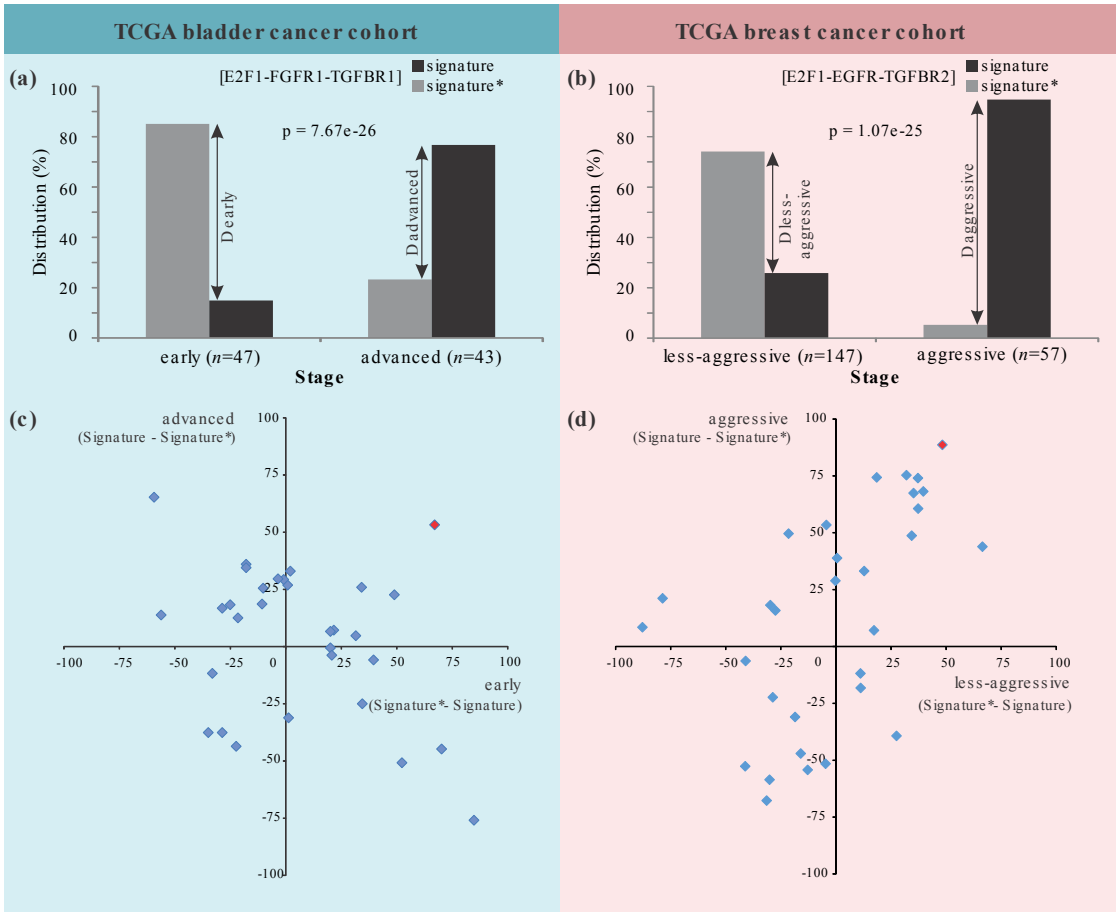


Figure 3.6. Validation of molecular signatures in (a) TCGA bladder and (b) TCGA breast cancer cohorts. In the bladder cancer cohort, first the patients are filtered who are diagnosed with early (stage II) and advanced (stage IV) stages. Further in both stages, we identified subsets of patients where expression of genes of the molecular signature (*i.e.,* E2F1, TGFBR1 and FGFR1) was above (signature) or below (signature*) the respective mean expression value. In case of the breast cancer cohort, we classified patients into aggressive (basal and Her2) vs. less-aggressive (luminal A and B) molecular subtypes based on PAM50 stages provided. Afterwards, we identified those subsets of patients where expression of genes of the molecular signature (*i.e.,* E2F1, TGFBR2 and EGFR) was above (signature) or below (signature*) the respective mean expression value. The distribution difference (%) of patients was calculated within pathological stages in both cancer types. Similarly, distribution differences for 30 random signatures vs. signatures* derived from (c) bladder and (d) breast cancer regulatory cores are shown as scatter plots. Signatures in the upper right or lower left corners nicely distinguish patient phenotypes. The molecular signatures predicted from my proposed workflow are shown in red. p-values are calculated using Pearson's χ^2 test. The figure is taken from Khan et al., 2017 [124].

3.2.6. Summary of results

Mathematical models, in combination with experimental data, are powerful tools to predict and understand the effect that different stimuli have on the behavior of a system.

Such systems biology approach further supports the formulation of new hypotheses about the effect of specific internal or external perturbations [156, 158]. To understand the molecular sources of EMT regulation in bladder and breast cancer, I developed logic-based models of the core-regulatory networks identified in Chapter 2. I divided each of the core-networks into three layers: (i) input layer, (ii) regulatory layer, and (iii) output layer. The input layer contains E2F1 and the receptor molecules. The regulatory layers consist of all components of the core networks except E2F1 and receptor molecules. The output layer comprises one node that represents the EMT process as the driver of the invasive phenotype. The input layer and regulatory layers were encoded with Boolean logic, while the phenotypical output of the network was modeled with multi-valued logics. Multi-valued logic allows us to model several activity levels of the phenotype, which helps in assessing the aggregated effect of various network components on the phenotype [155]. However, the use of multi-valued logic increases the complexity of the model, therefore, I applied it only to the phenotypical output. In logic-based models, when a molecule has multiple regulator, the choice for 'OR' and 'AND' gate is challenging. Towards this, I used qualitative information based on fold-change expression data to derive the Boolean functions.

I simulated the model for different combinations of input components, and their analysis revealed that high levels of the E2F1-TGFBR1-FGFR1 in bladder cancer and the E2F1-TGFBR2-EGFR in breast cancer constitute the molecular signatures that represent the most aggressive phenotype in the respective cancer type. Surprisingly, the other receptors that are part of the input layers in the models had no effect on the EMT process. The simulation results are in agreement with previous experimental findings in bladder cancer studies where high levels of E2F1 [201], TGFBR1 [202] and FGFR1 [197] were independently associated with tumor invasion. Similarly, in breast cancer studies, high expression of E2F1 [203], TGFBR2 [204] and EGFR [205] was separately observed to regulate invasive tumor phenotypes. Further, we validated the role of predicted signatures using bladder and breast cancer patient survival data from independent studies, see in Figure 3.5 and 3.6. For all our predicted signatures, high expression of the constituent molecules mapped to low patient survival and *vice versa*. These correlations prove that my approach successfully identified tumor-specific molecular signatures regulating EMT and driving invasive phenotypes.

Further, the *in silico* predicted signatures were validated by *in vitro* experiments. Inhibition of the predicted signatures in invasive bladder (UM-UC-3) and breast (MDA-MB231) cancer cell lines shows profound impact on the invasive behavior of the respective cell line and the highest effect observed upon combined inhibition Fig. 5c on page 160. Moreover, model simulations reveal that receptor molecules EGFR, CXCR1 and RAEA had no effect on the EMT in bladder cancer, and similarly in case of breast cancer the receptors molecules FGFR, HMMR, THRB, IL1R1 and RARA had no effect. The minimal effect of EGFR in bladder cancer and that of FGFR in breast cancer were confirmed experimentally (Fig. 5a on page 160).

Furthermore, I performed *in silico* perturbations to identify potential therapeutic candidates in cells, which have overexpression of the model-based identified molecular signatures (*i.e.,* E2F1-TGFBR1-FGFR1 in bladder and E2F1-TGFBR2-EGFR in breast cancer). From the list of identified candidates, our experimental partners, validated SMAD3-NFKB1 in UM-UC-3 bladder cancer and SRC-FN1 in MDA-MB231 in breast cancer by knock down using shRNA-based experiments. The experimental results are in consensus with the model predictions, which shows a significant reduction in cell invasion by inhibiting the therapeutic targets alone or in combination (Fig. 6 on page 161).

Overall, model-based treatment recommendations of E2F1-driven tumor, such as advanced bladder or breast cancer, have the potential to support cohort-specific treatment of patients in order to avoid therapy resistance and cope with aggressive cancers. Finally, the workflow that I proposed (combining network structure analysis, high-throughput data analysis and a dynamical model) and our comprehensive E2F1 interaction map can be applied to other cancer types in which E2F1 plays a similar role, as well as to characterize mechanisms leading to other phenotypes, like chemoresistance or angiogenesis, related to this transcription factor.

Chapter 4

Hybrid modeling of large-scale biochemical networks

The models, the methodology for hybrid modeling and their subsequent analyses are published in the following publications:

- Vera J, Schmitz U, Lai X, Engelmann D, **Khan FM**, Wolkenhauer O, Pützer BM. Kinetic modeling-based detection of genetic signatures that provide chemoresistance via the E2F1-p73/DNp73-miR-205 network. Cancer research. 2013 Feb 27. https://doi.org/10.1158/0008-5472.CAN-12-4095.

- **Khan FM**, Schmitz U, Nikolov S, Engelmann D, Pützer BM, Wolkenhauer O, Vera J. Hybrid modeling of the crosstalk between signaling and transcriptional networks using ordinary differential equations and multi-valued logic. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics. 2014 Jan 1;1844(1):289-98. https://doi.org/10.1016/j.bbapap.2013.05.007.

Synopsis

Mathematical models can be created at different levels of abstraction, ranging from coarse-grained qualitative (e.g., logic-based) models of (large) sub-cellular processes to detailed quantitative (e.g., ODEs based) models of (small) highly nonlinear processes. In systems biology, it is an emerging dilemma to conciliate models of massive networks and the adequate description of nonlinear dynamics with a single modeling framework. In this chapter I present a hybrid modeling framework that combines ODEs and logic-based models as a tool to dynamically analyze large-scale, nonlinear biochemical networks. As a proof of concept, I illustrate the construction and analysis of a hybrid model for regulatory network centered around the E2F1, a transcription factor involve in cancer. The network is organized and divided into different parts with distinctive regulatory features and each part is modeled with the suitable modeling formalism. A hybrid model provides a good compromise between quantitative/qualitative accuracy and scalability when considering large networks.

4.1. Background and motivation for hybrid modeling

In cells, biological processes are driven by complex networks integrating genes, transcripts, like mRNAs and microRNAs (miRNAs), proteins and small molecules. Concentrations and activity of these biomolecules are continuously changing in a concerted manner in response to internal and external cell signals. Those changes are regulated by multiple nested biological circuits that may contain feedback and feedforward loops. These nonlinear biological circuits give rise to important cell features like robustness against noise, adaptation despite environmental changes or hysteretic responses with multistability [206]. In the last decade it has been found that some biochemical networks are extremely large and complex [58, 59, 108–110]: they are commonly composed of hundreds of compounds and are enriched in nonlinear motifs like feedback and feedforward loops [2, 134]. Under these conditions, the analysis of biochemical networks evades human intuition. However, mathematical modeling is an appropriate tool to help in understanding those networks [158, 207, 208]. Depending on the features of the biochemical network under consideration and the available experimental data, a variety of frameworks for mathematical modeling is available. These range from simple and abstract approaches to biologically detailed ones, from deterministic to probabilistic or from spatio-temporal continuous to discrete ones [209].

Models in ordinary differential equations (ODEs), accounting for the time dependent variation of variables representing the concentration or activation state of biochemical molecules, have been used to describe biological networks for decades [210]. These ODE models allow the quantitative simulation of the concentration changes or activity profiles of proteins, genes, RNAs and other molecules over time. These features allow comparisons of model predictions with most of the standard experimental measurements. When biological compartments are considered, they can also account for a qualitative description of spatial features [211]. Finally, ODE models are the most widely used modeling framework when investigating the features of biochemical networks containing regulatory circuits like feedback- and/or feedforward loops [4]. However, a complete characterization of the model requires to specify the values of several kinds of model parameters (for example, initial protein concentration and rate constants), which in biochemical mid-size networks becomes a computationally intensive task and requires large amounts of quantitative experimental data and sophisticated optimization algorithms. For larger networks parameter estimation becomes cumbersome and difficulties to identify unique values for model parameters emerge [212–214].

An alternative to the ODE models is the discrete modeling approach. It is a qualitative approach that depends only on the network structure (*i.e.,* parameter free) with the simplifying assumption that network nodes, accounting for the expression level or activation state of biological molecules, exist only in a well-defined set of possible discrete numerical values [213]. Discrete models allow the dynamical analysis of large

biological systems of interactions among proteins, genes and other biomolecules to understand their behavior upon different input-stimuli and perturbations, and make predictions on biologically relevant scenarios [164, 199, 215]. Boolean models are the simplest discrete models, in which each element of the network can have one of two possible states (1: ON, expressed or active; 0: OFF, non-expressed or inactive) at any time point. This assumption is supported by biological evidences, *e.g.,* when genes/proteins exhibit ON/OFF switch like behavior. This can be adopted by assuming a reduced set of biologically meaningful values for the model variables [175, 213, 216]. A number of recent papers illustrates how discrete logic models can capture the behavior of large systems where the interactions are modeled with simple Boolean assumptions [90, 91, 217]. However, discrete logic models do not reproduce some of the time-dependent features associated to nonlinear circuits, like those containing nested feedback/feedforward loops where small changes in the variable produce entirely different dynamics [2, 134].

A single modeling formalism is not sufficient to accommodate the inherent complexity of biochemical networks (*i.e.,* large number of components and nonlinear network structure) [218]. Therefore, it is necessary to develop suitable modeling strategies that realize a compromise between the ability to simulate the behavior of large biochemical networks, the complexity associated to the existence of multiple regulatory loops in them and the diversity of sources of experimental data. Towards this, I propose a hybrid modeling framework, composed of ODE and logic sub-modules, as a strategy to handle large-scale, nonlinear biochemical networks associated to cancer and other complex diseases. I illustrate the construction of this kind of model using a toy model of the regulation of the anti-apoptotic molecule BCL-w in Section 4.4 on Page 70. Further, as a proof of concept I applied the hybrid modeling strategy to a regulatory network centered around E2F1, to characterize its role along with other receptor molecules in drug resistance. The model reproduced known features of the network regulated the drug resistance phenotype and hypothesized that for high expression of the E2F1 the tumor cell become independent of growth factors, which is one of the hallmark of cancer [190, 219].

Hybrid models are mathematical and computational constructs that combine interdependent variables that distributed over discrete/continuous or deterministic/stochastic domains [220]. They account in an integrative manner for different spatio-temporal scales of the same biological phenomenon, which are described using different interconnected modeling frameworks [96, 221]. Further, hybrid modeling *i.e.,* combination of different modeling formalisms, can be used to simulate large systems of interacting molecules involving signaling, transcription and metabolic layers [218, 222, 223]. The hybrid modeling approach seems especially valuable when dealing with the multifactorial and multi-level nature of biological phenomena like cancer emergence and progression [224, 225].

In the following text, I will first describe ODE-based models, different laws of mass

action kinetics, and type of dynamical analysis one can perform with ODEs. Second, I will present an ODE-based model of the E2F1-mediated drug resistance. Then, I will present in detail the hybrid model of the drug resistance, its construction, implementation and analysis.

4.2. ODE-based kinetic models

Kinetic-based models have successfully established their role in explaining the non-intuitive behavior of complex cellular processes [6, 83, 226], and providing an understanding of the molecular basis of cell functions in healthy and disease conditions [227, 228]. In a recent review Magi et al. (2017) evaluate the current status of mathematical models in the pursuit of a mechanistic understanding of complex cancer phenotypes [228]. They provide a good overview of how mathematical models unravel the non-intuitive dynamics of sub-/networks regulating the hallmarks of cancer: cell cycle regulation abnormalities, resistance to cell death, angiogenesis, invasion and metastasis, dysregulation of immune responses, and metabolism. Kinetic-based models are more suitable for the investigation of nonlinear biological systems where the number of interactions is small and sufficient quantitative data is available for parameter characterization. In systems biology, mathematical models are developed in the following steps: (i) construction of the model structure, (ii) derivation of mathematical equations, (iii) identification of parameters and initial conditions, and (iv) performing model simulations that can explain the nonlinear behavior and unravel mechanisms, which lead to testable hypotheses.

In kinetic models, the biological species are represented by variables (such as, metabolite concentration or amount of mRNA), which are characterized by rate laws that are modulated by various parameters. A general ODE system of equations is written as:

$$\frac{dX_i}{dt} = v_i(X_1, \ldots, X_N;\ k_1, \ldots, k_N) \qquad i=1,\ldots,N \tag{4.1}$$

where dX_i/dt represents the rate of change of species X_i at time t. $X_1, \ldots, X_N$ is a set of N number of variables representing the concentration of species. The temporal behavior of the system is evaluated by functions v_i (kinetic laws) depending on variables $X_1, \ldots, X_N$ and parameters $k_1, \ldots, k_N$. Variables are the quantities of interest in model analysis, which typically change over time. Parameters are quantities, which are fixed for a given computational experiment to characterize specific quantitative behavior of a model. Parameter values are typically derived from either literature, or databases, such as BioModels [20], Brenda [160] and SABIO-RK [19], and can be estimated from experimental data by calibrating the model to recapitulate the biological process as good as possible [17, 18]. The parameterization of the model presented in this chapter is based on: (i) published information, (ii) estimation by terning them to fit published quantitative and qualitative data, and (iii) fixing some parameter values to normalize variables to the

basal, non-stressed level of mRNA and protein (more details are provided in Appendix F Section F.2). In the following sub-sections, I will briefly describe different laws of rate kinetics, their general representation and applications.

4.2.1. Mass action kinetics

To evaluate the dynamics of a system, one needs to determine what the function v_i (*i.e.,* reaction rates) actually is. In systems biology practice, it can be described by appropriate rate kinetics such as 'mass action kinetics'. Mass action kinetics state that the rate of a reaction is proportional to the probability of a collision of the reactants [229]. This probability is in turn proportional to the product of concentrations of the reacting species to the power of molecularity, which is the number in which the molecular species enter the reaction [229]. The general representation of reaction rate in mass action kinetics can be written as:

$$v_i = k_i \cdot \prod_{j=1}^{l} x_j^{g_{ij}} \;\; ; \;\; \forall g_{ij} = m_{ij} \in \{0, 1, 2\} \tag{4.2}$$

where v_i $(i = 1, \ldots, N)$ is reaction rate of i^{th} reaction. Every reaction rate v_i is described as a product of a rate constant (k_i) and biochemical species $(x_j, j = 1, \ldots, l)$ participating in the reaction. g_{ij} denotes the kinetic orders which is equal to the number of molecules of x_j involved in a reaction. When g_{ij} is zero the specie is not involve in a reaction, when it is 1 the species participates, and for a value of 2 it represents the dimerization.

For example, using law of mass action the rate of a reaction,

$$S_1 + S_2 \underset{k_{-2}}{\overset{k_1}{\rightleftharpoons}} 2P \tag{4.3}$$

where one molecule of substrate S_1 and S_2 are converted into two molecules of product P and *vice versa* in a reversible reaction, can be written as:

$$v_1 = k_1 S_1 \cdot S_2; \quad v_2 = k_{-2} P^2$$

$$v = v_1 - v_2 = k_1 S_1 \cdot S_2 - k_{-2} P^2 \tag{4.4}$$

where v_1 (the product of kinetic constant k_1 with substrates S_1 and S_2) and v_2 (product of kinetic constant K_{-2} with power of 2 (molecularity of P in the reaction) of product P) are the rates of forward and backward reactions respectively. The net rate v, here represents the rate of formation of product P, is equal to the production rate (v_1) minus depletion rate (v_2). Law of mass action is the precise and detailed mathematical description for chemical reaction kinetics using ordinary differential equations for which detailed structure of the system must be known, but this is difficult to achieve when many interacting components are involved. However, assuming some conditions (*e.g.,* steady state), more simplified rate kinetics (such as Michaelis-Menten and Hill functions) can be derived to model a system of chemical reactions.

4.2.2. Michaelis-Menten kinetics

For a quasi-steady state assumption, for example when a reaction is much faster than other reactions, Michaelis-Menten (MM) is the algebraic substitution of differential equations in mass action kinetics [229]. The general structure of MM reactions contains an enzyme E, which binds to a substrate S to form an enzyme-substrate complex ES which in turn releases product P and the original enzyme E, represented by the following reaction:

$$E + S \underset{k_{-1}}{\overset{k1}{\rightleftharpoons}} ES \xrightarrow{k2} P + E \tag{4.5}$$

The Michaelis-Menten approximation for the reaction rate is:

$$v = \frac{v_{max} S}{k_M + S} \tag{4.6}$$

where v_{max} is the maximal rate, and k_M is the Michaelis-Menten constant, which is half of the maximum rate Figure 4.1. The Michaelis-Menten approximation holds, if the concentration of the substrate in large excess of the enzyme and the quasi-steady state assumption is fulfilled [229]. This means the elementary reaction produces short-lived intermediates (such as the enzyme-substrate (ES) complex) and the reversible conversion of an intermediate complex back into to the substrate is much faster than its conversion into product. For detailed derivation and discussion of MM rate law, I would refer the reader to the books written by Edda Klipp [229] and Brian Ingalls [157].

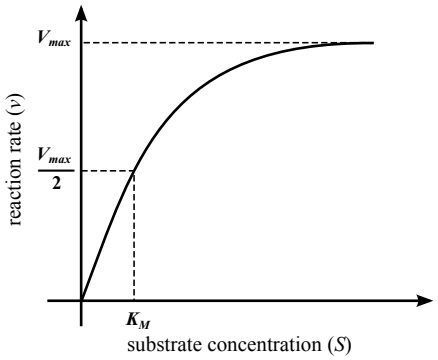


Figure 4.1. Michaelis-Menten. The reaction rate approaches to maximum rate v_{max} as the concentration of the substrate increases. The Michaelis-Menten constant is the substrate concentration at which the reaction rate equal to half of the maximum rate v_{max}.

4.2.3. Hill function

Besides MM kinetics, the Hill function is also used to replace the differential term in mass action kinetics by an algebraic function. It is used to describe the level of cooperativity of a ligand binding to an enzyme or receptor [157, 229]. This can be generalized for

example to model biological processes involving cooperative interactions, which can result in switch-like behavior. The general form of Hill function is the following:

$$v = \frac{S^n}{(S^n + K^n)} \tag{4.7}$$

where K is Hill constant which is equal to half-saturation concentration of ligand and n is the Hill coefficient, describing the cooperative effect on process (*i.e.,* the steepness of curve to reach from minimum to maximum rate and *vice versa*) see Figure 4.2. For $n = 1$, there is no cooperative effect and the shape of the curve is hyperbolic, while bigger values ($n > 1$) induce a cooperative effect, which produce a sigmoidal (S-shaped) curve. For detailed derivation and discussion of the Hill function, the reader is referred to the book written by Brian Ingalls [157].

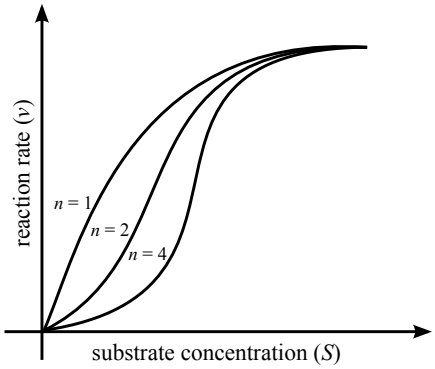


Figure 4.2. Hill function. For Hill coefficient $n = 1$, the curve is hyperbolic. When n increases, the signoidal cure becomes more switch-like. The figure is adopted from [157].

4.2.4. Power-law models

The power-law model is another convenient representation for mass action kinetics [210, 230]. It describes the reaction rate as a product of a rate constant and all the reactants raised to power of non-integer values. Compared to mass action kinetics shown in Equation (4.2), in which the kinetic orders g_{ij} are positive integers, power-law models allow non-integer values (*i.e.,* $g_{ij} \in \mathbb{R}$) Figure 4.3. The general form of the power-law model is:

$$v_i = c_i \cdot k_i \cdot \prod_{j=1}^{l} x_j^{g_{ij}} \quad ; \quad \forall g_{ij} \in \mathbb{R} \tag{4.8}$$

Using a power-law model, the reaction rates v_1 and v_2 for Reaction 4.3 can be written in the following form:

$$v_1 = k_1[S_1 \cdot S_2]^g; \quad v_2 = 2k_{-2}([P]^g)^2 \tag{4.9}$$

where g is the power-law exponent, which can be any non-integer value. Based on available knowledge about the cellular environment in which the reaction takes place, two type of formulations exist for power-law models: simplified and detailed power-law models [207, 231]. When little knowledge about the structure is available, the simplified power-law model is well suited with the assumption of a homogeneous and well mixed environment, in which both positive and negative (representing inhibition) real values are applied for kinetic order [232]. When sufficient knowledge is available, the detailed power-law model is used, which takes only positive real values for kinetic orders and provides a more realistic characterization of inhomogeneity and molecular crowding of the cellular environment [231].

We used power-law models for the construction of the E2F1-mediated chemoresistance model, described in this chapter, due to the following remarkable features: (i) a power-law model can represent the dynamics of inhibitory, activating and cooperative processes by keeping the mathematical structure of the system intact and just modifying the value of kinetic order [207, 232]; (ii) it allows to model different strengths of activation/inhibition of reactants in certain biological processes; and (iii) it makes detailed mathematical models possible with little prior knowledge of the structure of the network.

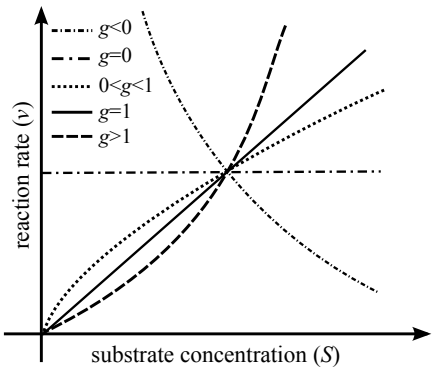


Figure 4.3. Power-law model. Power-laws can model different dynamics by choosing appropriate values of kinetic orders g [207]. Power-law model with negative rate constant (*i.e.*, $g < 0$) represent inhibition. A zero value for g indicates the substrate has no effect on the reaction rate. The value of rate constant in the range greater than zero and less than 1 (*i.e.*, $0 < g > 1$) describes the saturation-like behavior. When the value of g is equal to 1 ($g = 1$), it represent the linear dynamics; and finally rate constant greater than 1 ($g > 1$) reproducing the cooperative processes.

4.2.5. ODE-based model analysis

After constructing the model structure, derivation of mathematical equations and identification of parameter values, model simulation and analysis begin. Model

simulations are performed to study the dynamic behavior, the temporal evaluation of components, of a system for a given condition (*i.e.,* for given initial values and a set of parameters). In case of ODEs, model simulations are carried out by numerical integration using ODE solvers (*e.g.,* ode23 from MATLAB). Model simulations facilitate a temporal evaluation of states, and thus the behavior of a systems. If small variations in the conditions produce large changes in the solution, we call the system unstable. If small variations in the conditions produced infinitesimal small changes in the solution, then the systems is stable. Such type of simulations are called the **stability analysis**, which can be used to study the behavior of the systems for different input stimuli and perturbations. For input stimuli one can study, for example, how a system responds to signals from external environment. For perturbations one can evaluate, for example, how a system behaves when variations are made in the dependent variables, which can mimic mutations in the biological entities; or transient input signal (*e.g.,* some sort of stress) is apply to the system, which can be drug intervention. Depending on the network structure (*i.e.,* feedback or feedforward loops) of the system, these changes in initial conditions and perturbations in parameters induce interesting dynamical features on the state variables [2, 134]. In the following text, I will describe the basic concept of stability (steady state and transient) analysis of feedforward loops, which we later use to elucidate the mechanisms of E2F1-p73/Dnp73-miR-205 in chemoresistance.

Stability analysis

In dynamical systems theory, the time dependent behavior of a system is characterized by its states, which is the state of its components at a given point in time [229]. In stability analysis, model simulations usually start from an initial condition and evolve over time until eventually reaching a steady state or oscillation. The time-course from initial state of a system to the long-time (or asymptotic) steady state is called transient behavior Figure 4.4 [157]. On the other hand, in a steady state, variables remain constant in time, *i.e.,* the concentration of the species reach a steady level and exhibit no variation over time [135, 157, 229]. Steady state of a system reflects its long term behavior. To study the steady state, those states in which, the rate of change reaches to zero are considered:

$$\frac{dX_i}{dt} = 0; \quad i = 1, \ldots, N \tag{4.10}$$

If a system is at steady state, it should remain there until an external perturbation occurs. Depending on the system structure, perturbations can have the following effects: (i) stable: the systems returns to steady state; (ii) unstable: the system leaves the steady state; and (iii) the system behavior is indifferent [229]. If the system remains in any of the states for sufficiently long time ($t \to \infty$) that behavior is called asymptotic stable.

Stability analysis is highly influenced by the structure of a system, for example,

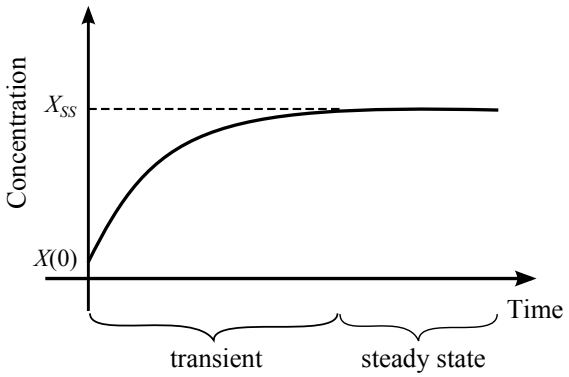


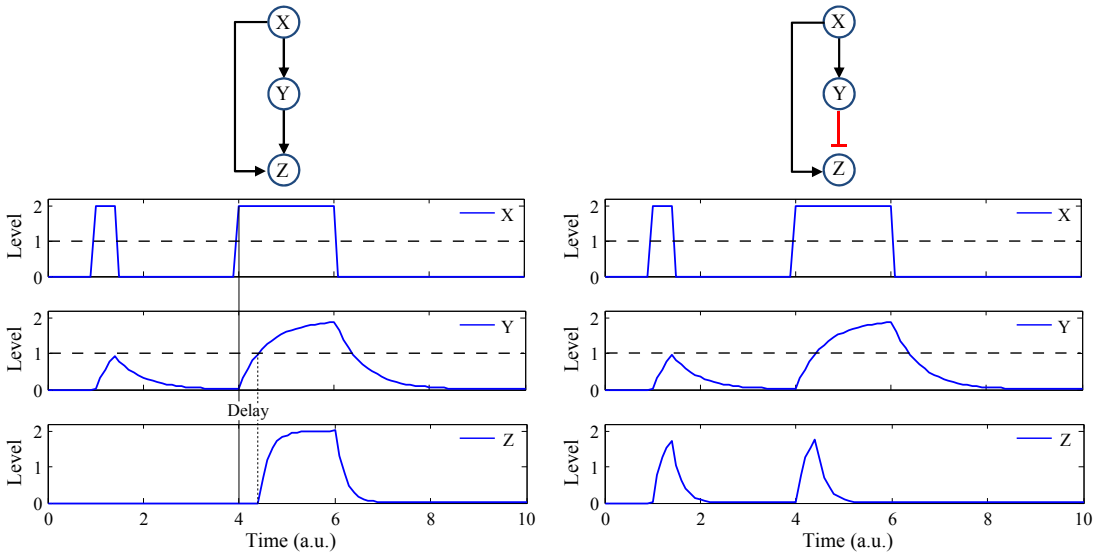
Figure 4.4. Transient and steady-state behavior. The transient behavior is the time course from initial value $X(0)$ till the system settles into a steady state at point X_{ss}.

feedback and feedforward loops [2, 134], which induce nonlinear dynamics in the system's behavior. A feedback loop induces a switch like behavior and provides a mechanism for a memory effect to a transient input signal. On the other hand, feedforward loops speed up or delay the system response to transient input signal. There are eight types of feedforward loops, which are subdivided into two categories: (i) coherent feedforward loop (cFFL) (see Figure 2.3c); and (ii) in-coherent feedforward loop (iFFL) (see Figure 2.3d). The coherent type-1 and incoherent type-1 feedforward loop are the most abundant [229, 233]. Coherent type-1 FFLs act as a filter for weak pulses and cause a delayed response see Figure 4.5a, while incoherent type-1 FFLs can act as a pulse generator with fixed maximal length and can accelerate the system's response to an input signal (see Figure 4.5b). Futher, incoherent FFLs induce non-monototic behavior in system dynamics [234]. In a non-monotonic response, the output of the systems first increase with the input signal, but later decrease when the input signal is high (see Figure F.3). Transient analysis of network motifs can help to understand the dynamics of biochemical networks in response to time varying stimuli/perturbations, which can provide valuable insights into mechanisms underlying disease [235].

Figure 4.5 shows the dynamical behavior induced by coherent/incoherent feedforward loop. In the loops, 'X' regulates 'Y' and then 'X' and 'Y' regulate 'Z'. The two regulators of 'Z' can be processed by logical *AND* or *OR* functions.

4.3. ODE model-based detection of molecular signatures in chemoresistance

Resistance to drugs is a major cause of cancer therapy failure. E2F1 is a transcription factor that is involved in cell cycle progression, apoptosis, metastasis and drug



a Coherent FFL (*AND* logic) b Incoherent FFL (*AND* logic)

Figure 4.5. Dynamical behavior of (a) coherent FFL and (b) incoherent FFL. In the loops, 'X' regulates 'Y' and then 'X' and 'Y' regulate 'Z'. The two regulators of 'Z' can be processed by logical *AND* or *OR* functions. **(a)** Coherent FFL type 1 with AND logic. The structure of the coherent type 1 FFL (first row), and the curves show the transient input X (second row), intermediate node Y (third row) and output Z (fourth row). The curve of Z shows that short pulses are filtered out, the response to the longer pulse is delayed, but the response to the end of pulse is immediate. **(b)** Incoherent FFL type 1 with AND logic. Each onset pulse in the input leads to a pulse in Z with a fixed maximal length. I reproduced the plots by model simulations from the book written by Edda Klipp [229] for FFL with parameter values $a_y = 1.7$, $b_y = 1.7$, $a_z = 10$, and $b_z = 5$.

resistance through an intricate signaling and gene regulatory network, which includes other transcription factors like p73 [236], and cancer-related microRNAs like miR-205 [119]. To investigate mechanistically the emergence of drug resistance, we carried out dynamical system analyses of the regulatory network summarizing knowledge of E2F1 and its interplay with p73/DNp73 and miR-205 in cancer drug responses (see Figure 4.6**a**).

The structure of the E2F1-centered network that mediates the potential resistance against anticancer therapies is a complex system, *i.e.,* it is enriched with multiple network motifs, including feedback and feedforward loops. Furthermore, these motifs are well interconnected, overlapping, and cross-talk with other cancer-related signaling pathways [120, 237]. Networks containing several of these motifs often induce non-intuitive regulatory patterns, which require the use of mathematical modeling to understand their function and regulation [238, 239].

We developed an ODE-based kinetic model of the E2F1-p73/DNp73-miRNA-205 network to get insights into the mechanisms underlying drug resistance. The model represents the network response to certain genotoxic (DNA damage causing agents can

lead to cell death) and cytostatic (agents toxic to cells may stop active growth and induce cell death) drugs. The prototypical simulations (see Figure 4.6**b**) show that our model is able to describe the cascade of regulatory events after drug administration including stress signals (top), downstream intracellular responses including E2F1-p73/DNp73-miR205, the regulation of pro- and antiapoptotic signals (for genotoxic drugs) or proliferative signals (for cytostatic drugs) (middle), and finally the global effect on tumor cell population at the bottom. In Appendix F, I provided the detailed description of the ODE-based model construction, parameterization and model simulations.

Taken together, our *in silico* simulations suggest a relevant role for E2F1, DNp73, and miR-205 in the regulation of resistance to several anticancer drugs. To this end, we conducted simulations in which we iteratively perturbed the values of the parameters accounting for E2F1 and miR-205 induction and computed the percentage of surviving cells 48 hours after genotoxic drug administration. As shown in Figure 4.7**a**, *in silico* cell lines with high levels of E2F1 display a high survival rate in response to chemotherapy. The simulation results are in agreement with the experimental data produced by our collaborators from the Institute of Experimental Gene Therapy and Cancer Research (IEGT)[1], University Rostock Medical Center, on melanoma SK-Mel-147 cells (Figure 4.7**b**) [119]. SK-Mel-147 cells were treated with cisplatin and a high percentage of cells underwent apoptosis (Scenario 1 in Figure 4.7**b**). Inhibition of miR-205 with antagomir-205 reduces the percentage of apoptotic cells after treatment (Scenario 2), whereas selective knockdown of endogenously high E2F1 leads to a pronounced increase in apoptosis (Scenario 3). Finally, repression of both E2F1 and miR-205 reduces the apoptotic effect of cisplatin (Scenario 4). These experimental results confirm the model predictions displayed in Figure 4.7**a** for the scenarios named with the same numbers.

In addition, we found that a genetic signature composed of high E2F1, low miR-205, and high ERBB3 can render tumor cells insensitive to both cytostatic and genotoxic drugs (see Appendix F Figure F.2). Our model simulations also suggest that conventional genotoxic drug treatment favors selection of chemoresistant cells in genetically heterogeneous tumors, in a manner requiring dysregulation of incoherent feedforward loops that involve E2F1, p73/DNp73, and miR-205 (see Appendix F Figure F.3).

4.4. Hybrid modeling: combining ODE- and logic-based models

Here, I address the problem of modeling large biochemical networks involving dozens to hundreds of receptors, kinase proteins, transcription factors, mRNAs and microRNAs.

[1] http://www.iegt-rostock.de/en/

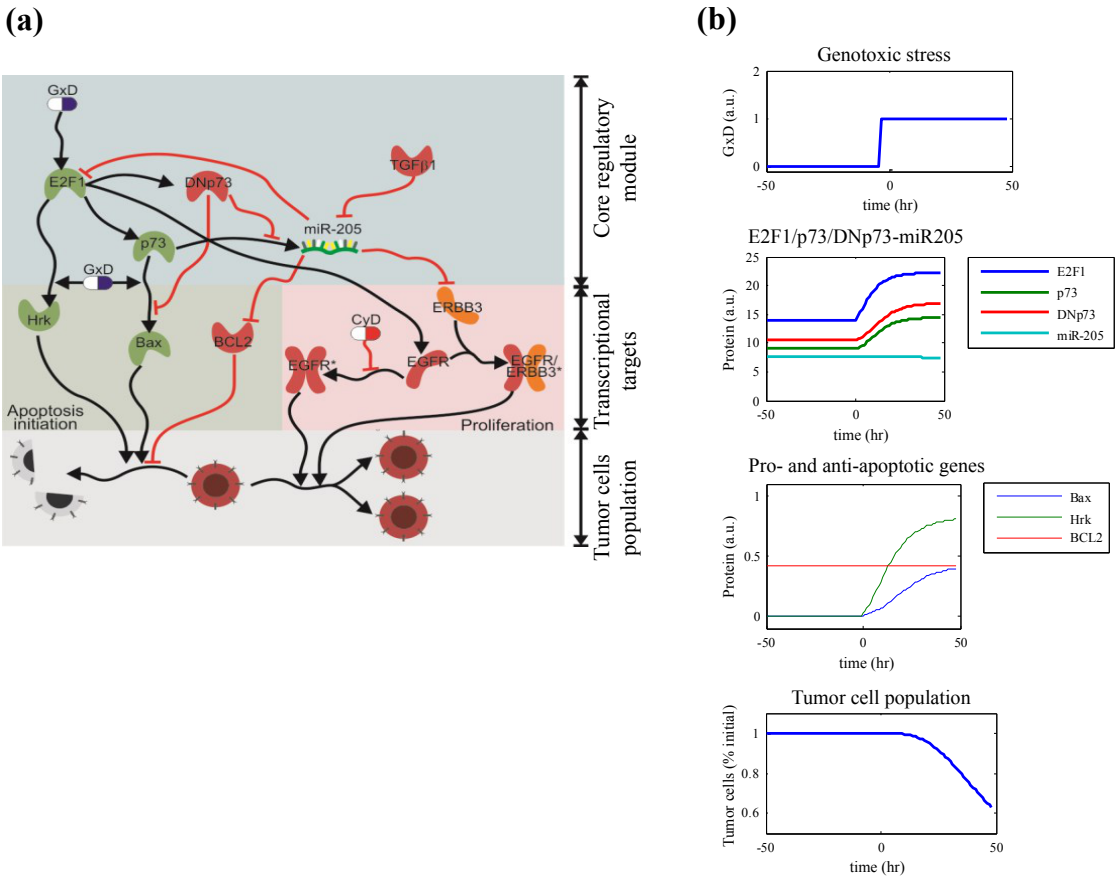


Figure 4.6. Kinetic model for the E2F1-p73/DNp73-miR205 network. **(a)** The model is composed of four modules: The core regulatory module (blue) describes the dynamics of E2F1, p73, DNp73, and miR-205. The transcriptional targets module accounts for representative apoptosis and proliferation-associated targets of E2F1, p73, and miR-205, which may play a role in drug response. The tumor cell population module connects the biochemical network to the fate of a population of tumor cells. GxD accounts for genotoxic and CyD for cytostatic drug; red-colored symbols for oncogenes and green for tumor suppressors. **(b)** Prototypical simulation of the model showing the cascade of regulatory events after drug administration. The figures are adopted from Vera et al., [121].

Their expression and activity are typically regulated by multiple, cross-talking pathways that may even contain overlapping regulatory loops. The strategy is to organize and divide the network into three parts with distinctive regulatory features (Figure 4.8**a**): **(i) The core regulatory module,** a highly interconnected subnetwork that contains the signaling and transcriptional pathways enriched in feedback and feedforward loops, and therefore is expected to display a highly nonlinear behavior; **(ii) The target gene module,** which accounts for dozens to hundreds of genes whose expression is directly regulated by the proteins and transcription factors included in the core regulatory module and can be experimentally quantified using high-throughput transcriptomics and proteomics techniques; and **(iii) A set of phenotypical**

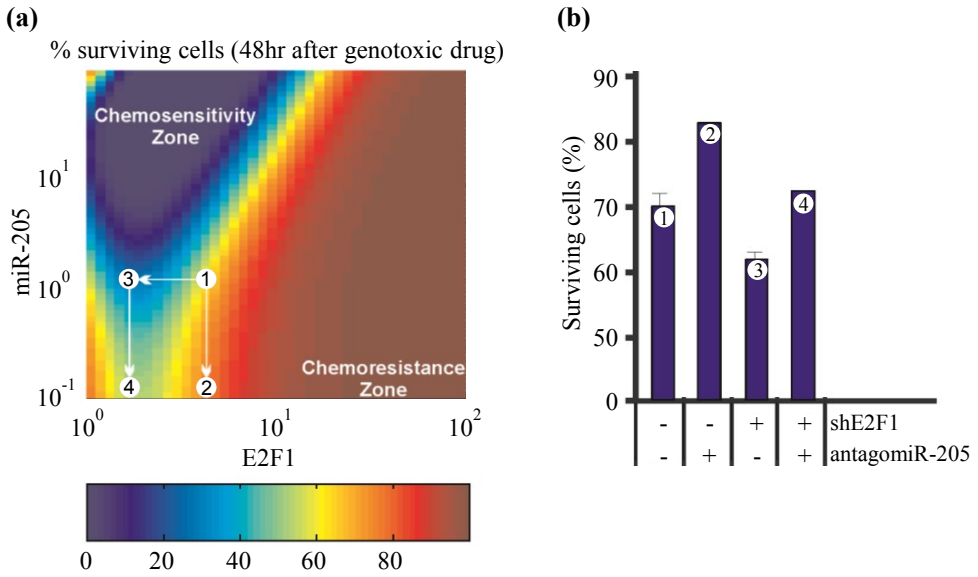


Figure 4.7. The interplay between E2F1 and miR-205 promotes resistance to genotoxic drugs. (a) Model simulations for different values of the synthesis rates for E2F1 and miR-205 (k_{E2F1}, $k_{mir-205} \in [10^{-1},10^{1}]$) [121]. We determined the percentage of surviving cells 48 hours after genotoxic stress. **(b)** Validation experiments with SK-Mel-147 cells. The percentage of surviving cells was measured 48 hours after cisplatin treatment for: (1) control; (2) after miR-205 knockdown; (3) after E2F1 knockdown; and (4) after miR-205 and E2F1 knockdown. The scenarios named with the same numbers. Data were extracted and processed from Alla and colleagues [119].

read-outs, phenomenological variables accounting for relevant phenotypes triggered upon activation of groups of genes in the target gene module. Input signals external to the networks are assumed to regulate or interact with key compounds of the core regulatory module; although in some cases they could also be direct regulators of the target genes.

In this approach, I modeled the core regulatory network using ODEs while the target gene module and the phenotypical read-outs are encoded using discrete logic modeling (Boolean or multi-valued logic). Both parts of the model are connected using a discretization interface. Given the set of continuous time-dependent variables of the core regulatory module ($X_{i,C}$, where $i = 1, 2, \ldots, k$), I defined a set of auxiliary discrete variables ($X_{i,D}$, where $i = 1, 2, \ldots, k$) for the discretization interface. Furthermore, I discretized the time by considering a discrete and finite set of p time points within the duration of the simulation (t_j, where $j = 1, 2, \ldots, p$). For every time point considered, the interface assigns values to the auxiliary discrete variables, $X_{i,D}(t_j)$, following a set of discretization rules that depend on the current values of the continuous variables, $X_{i,C}(t_j)$, and n_i threshold values $TH_{i,m}$, where $m = 1, 2, \ldots, n_i$. The number of thresholds for each continuous variable defines the number of ordinal states for the auxiliary discrete variables. The structure of the discretization rules is displayed in Table 4.1.

for $i = 1, 2, \ldots, k$
if $X_{i,C}(t_j) < TH_{i,1} \rightarrow X_{i,D}(t_j) = 0$
if $TH_{i,1} \leq X_{i,C}(t_j) \leq TH_{i,2} \rightarrow X_{i,D}(t_j) = 1$
if $TH_{i,2} \leq X_{i,C}(t_j) \leq TH_{i,3} \rightarrow X_{i,D}(t_j) = 2$
$\vdots$
if $TH_{i,n-1} \leq X_{i,C}(t_j) \leq TH_{i,n} \rightarrow X_{i,D}(t_j) = n_i - 1$

Table 4.1. Definition of the discretization rules. For the each of the k variables of the core regulatory module ($X_{i,C}$, where $i = 1, 2, \ldots, k$) at the considered time point t_j, I assign a value $X_{i,C}(t_j)$ using these discretization rules.

To simulate the whole model, first the core module is initialized and simulated with a set of values defined for the model inputs (Figure 4.8). At each of the p time points defined, the values of the auxiliary discrete variables, $X_{i,D}(t_j)$, are calculated following the discretization rules. Those discrete values are passed as inputs to the logical module accounting for the target genes and are used to update the values of its discrete variables following a logical steady-state analysis. Subsequently, the updated values for the model variables in the transcriptional circuitry module are used to update the phenotypical read-outs. In this manner and for every discrete time point (t_j), the model simulation generates: (i) a set of continuous values for the variables of the regulatory core module, $X_{i,C}(t_j)$; (ii) a set of discrete values for the auxiliary variables, $X_{i,D}(t_j)$; (iii) values for the discrete variables accounting for the expression of the target genes, $Tg(t_j)$; and (vi) Boolean values for the phenotypic read-outs (Figure 4.8**c**).

This approach has the advantage that the complex and highly regulated part of the network, which is enriched in regulatory loops is encoded with a modeling approach that provides many computational and analytical tools to investigate its properties [211, 240]. On the other hand, the use of logic modeling to describe the associated target genes allows the simulation of massive transcriptional networks with dozens to hundreds of biologically relevant genes with low computational burden [84, 166]. In this line, the data produced in high-throughput transcriptomics and proteomics experiments can be discretized in the described way to substantiate these large-scale transcriptional networks.

4.4.1. Software implementation

For the implementation of the hybrid modeling approach, I used built-in ODE solvers from MATLAB to simulate the dynamics of the core regulatory module (MathWorks, MA) and CellNetAnalyzer (CNA) [199] to simulate the logical part of the target genes and phenotypical read-outs see Figure 4.8**b**. To connect both subparts, I developed a

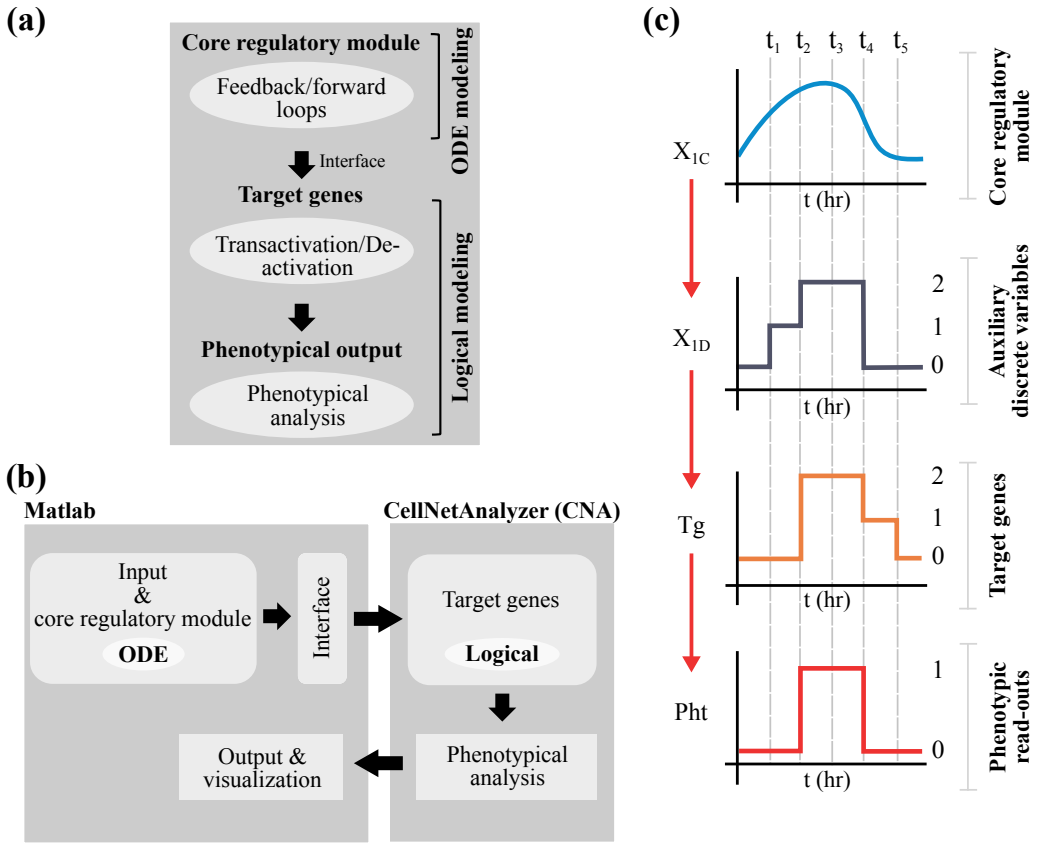


Figure 4.8. Hybrid modeling methodology. **(a)** Structure of the model. **(b)** Data workflow in the hybrid model during simulations. **(c)** Sketch of a simulation, with continuous and discrete variable sets, where X_{1C} and X_{1D} are continuous and discrete variables respectively, Tg represents target gene sets and Pht stands for the phenotype. The figures are adopted from Khan et al., [9].

MATLAB script that discretize the values of critical transcription factors and miRNAs in the core regulatory module; and used these values as an input vector for the simulations in CNA. Discretization is carried out using predefined thresholds.

Using the input vector, I applied the following CNA command to perform logical steady state (LSS) analysis:

$[spec\_lss] = CNAcomputeLSS(cnap, [input\_vector])$;

In this command, the vector [spec_lss] contains LSS values of the species in the target gene module 'cnap' which are based on the input_vector calculated in the core module. The output of the model simulations (continuous and discrete variable values) can be stored in a matrix, whose content is visualized using color coded surface plots.

4.4.2. Toy model demonstration of hybrid modeling

To illustrate the functioning of the proposed hybrid modeling approach, I derived a small model describing the regulation of the tumor suppressor miRNA-205 by the transcription factor p73, its isoform DNp73, and the subsequent regulation of the anti-apoptotic targets

like BCL-w (Figure 4.9**a**) [119, 121]. I have simulated this regulatory module using a model in ordinary differential equations (Figure 4.9**c**) and the hybrid modeling approach (Figure 4.9**d**). The ODE model was extracted from our previous publication [121]. To construct the hybrid version of the model, I preserved the ODEs describing the miR-205 regulation and generated a multi-valued logic model to describe the regulation of BCL-w by miR-205 (Figure 4.9**b**). Furthermore, the continuous ODE values of miR-205 ($miR205_{con}$) were discretized ($miR205_{disc}$) into three possible values using the rule described in Figure 4.9**b**. The two surface plots generated in the model simulations clearly show that the hybrid model preserves the nonlinear nature of the BCL-w regulation for different expression values of p73 and DNp73 as displayed by the ODE model [121]. However, BCL-w expression is now represented by three ordinal states (0: no expression, 1: low expression and 2: high expression). MATLAB code for the toy model is given in Appendix I.

4.5. Case study: hybrid model of chemoresistance

4.5.1. Background

The protein E2F1 is a transcription factor. Upon stimulation with different signals E2F1 can promote the activation and subsequent expression of different sets of genes involved in DNA replication and repair, cell cycle progression, proliferation and/or apoptosis [241, 242]. Activation of E2F1 by growth factor associated signals induces expression of target genes promoting cell cycle progression and proliferation [119, 243]. In addition, after genotoxic stress E2F1 regulates the expression and transcriptional activity of the tumor suppressors p53 and p73, promoting the expression of multiple genes related to the regulation of apoptosis and cell cycle arrest [120, 237, 244]. Having the ability to regulate cell proliferation and apoptosis, it is not surprising that E2F1 has been found to be deeply involved in tumor progression and the resistance against anti-cancer drugs [119, 121, 245].

Rather than isolated, E2F1 is part of a vast and complex biochemical network involving regulation events and a transcriptional circuitry (Figure 2.2 on page 23). The network is enriched in nonlinear network motifs, including feedback and feedforward loops. This makes the use of mathematical modeling for the investigation of E2F1 deregulation in cancer mandatory. Several mathematical models have been developed to investigate the regulation of E2F1 in different biological contexts. For example, a number of papers elucidation the role of E2F1 in the regulation of the mammalian G1-S cell cycle transition by means of ODE modeling and bifurcation analysis [246, 247]. Others have investigated the mechanisms by which cellular stress (*e.g.,* DNA damage) shifts E2F1 activity to induce apoptosis, and how this mechanism is distorted in cancer cells [248–250]. The role of miRNAs in the regulation of E2F1 has also recently been investigated via mathematical modeling [251]. Previously, we developed a mathematical model to investigate the E2F1-p73/DNp73-miR-205 regulatory circuit in the emergence

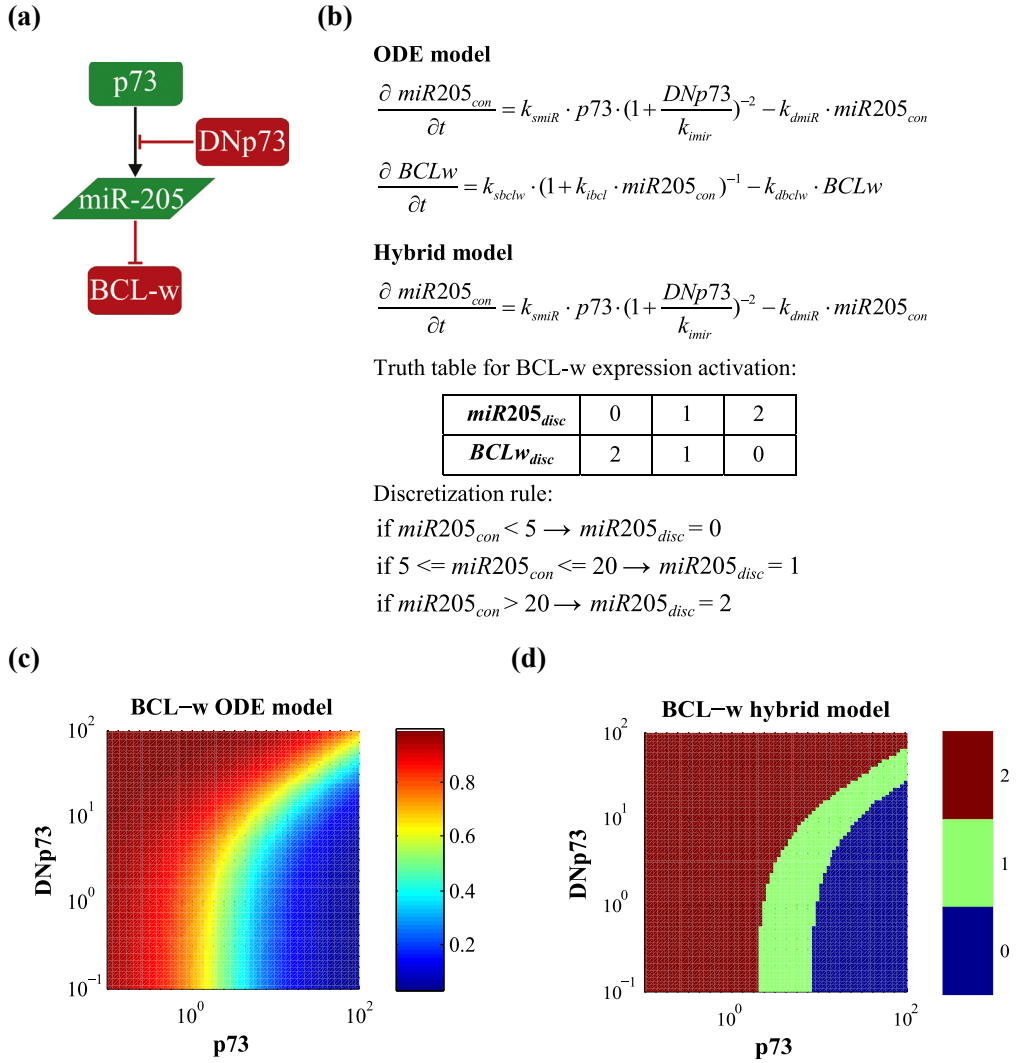


Figure 4.9. Toy model of hybrid modeling methodology. **(a)** Graphical representation of transcriptional regulation of the oncogene BCL-w by the transcription factor p73 and its isoform DNp73. **(b)** ODE model [121] and hybrid model representations. **(c)** ODE model simulations. **(d)** Hybrid model simulations. In the simulations, different expression levels for p73 and DNp73 were considered and the steady-state expression levels of miR-205 and BCL-w were computed using the two versions of the model. ODE model parameters were: $k_{dmiR} = 0.029$; $k_{smiR} = 0.1010$; $k_{imir} = 10.2672$; $k_{sbclw} = 0.1$; $k_{dbclw} = 0.1$; $k_{iblc} = 0.18$; $p73, DNp73$ $[10^{-1}, 10^{2}]$. The figures are adopted from Khan et al., [9].

of resistance to anti-cancer drugs [121]. Taken together, these studies investigate the regulation of E2F1 in different biological contexts, but none of them has addressed the question of how these events are modulating E2F1 expression and activity affecting the expression of large transcriptional networks downstream of E2F1. The model described here is the first approach to answer this question with a focus on the role of E2F1 in chemoresistance.

4.5.2. Model structure

The derived model accounts for the critical elements of the signaling and transcriptional network regulating the activity of E2F1 in tumor and normal cells (Figure 4.10). The network can be divided into three parts: (i) the core regulatory module, a signaling/transcriptional subnetwork accounting for the regulation of the expression and activity of E2F1 by proliferative and genotoxic stress signals, (ii) the target genes, a network accounting for the E2F1 modulated expression of critical genes involved in the regulation of cell proliferation, genotoxic drug response and apoptosis; and (iii) the phenotypical read-outs, a set of phenomenological variables accounting for the activation of relevant phenotypes (*e.g.,* proliferation, apoptosis, chemoresistance) upon the expression of given genes in the target genes module.

The core regulatory module has been extended to account for the regulation of p73, DNp73 and miR-205, which form together with E2F1 a regulatory circuit in the context of the genotoxic stress response and apoptosis initiation [119, 121]. Furthermore, I included a sub-module accounting for the activation of E2F1 via EGFR mediated proliferative signals and the E2F1 mediated transcriptional regulation of EGFR [252]. Both sub-modules constitute experimentally confirmed cases of feedback loops, a negative one in the case of E2F1-p73/DNp73-miR-205 [119] and a positive one in the case of E2F1/EGFR [121, 252]. In addition, input variables regulating the dynamics of the core module account for: (a) the concentration of growth factors activating EGF receptor (GF), (b) the cancer-related changes in the expression rate of E2F1 (FS), (c) the expression level for the oncogene TGFβ-1 (TGFB1), and (d) the dose of a genotoxic agent administered as anti-cancer drug (GxS). TGFβ-1 exemplifies a set of proteins whose dynamics is external to the core module, but regulate the expression and activation of the core regulatory module compounds.

On the other hand, genes considered in the target gene module account for: (i) E2F1-regulated genes involved in cancer-related abnormal cell proliferation (*e.g.,* GAB2, CCND1, ANCCA); (ii) E2F1-modulated genes involved in DNA-damage response and apoptosis initiation (*e.g.,* BIM, DUSP1-3, HRK, NOXA); (iii) p73-regulated genes involved in apoptosis initiation (*e.g.,* BAX, PUMA); and (iv) miR-205-repressed genes involved in apoptosis repression and chemoresistance (*e.g.,* BCL2, ABCA2). These genes are used here as an example and represent a wider set of genes which undergo similar regulation.

Following the approach proposed, I derived an ODE model to describe the dynamics of the core regulatory module, which was extracted and adapted from our previous publication [121] described in Section 4.3. For the target genes, I constructed a discrete multi-valued logic model, which was connected to phenotypical read-outs using also multi-valued logic. The core regulatory module and the target gene module were connected using a computational interface that discretizes the values of the model variables in the regulatory module that control the expression of the genes in the target

genes, namely, active E2F1 (E2F1p), p73, DNp73, and miR-205. The discretization rules were established using manual training [121], where minimum expression level is discretized to 0 and maximum to 1 or 2 in case of multi-valued logic 4.5.3. Taken together, the model is composed by nine ODEs (see Equations (4.11) to (4.19) on page 75) and 35 multi-valued logic functions (Table 4.3 on page 78).

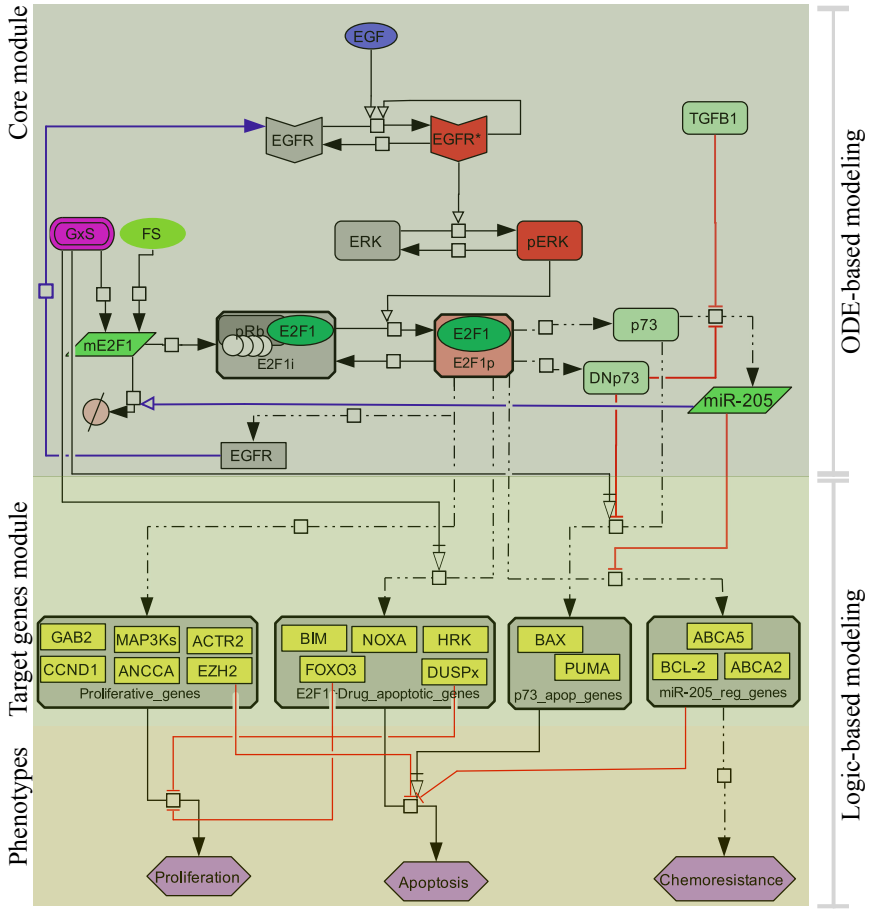


Figure 4.10. Structure of the hybrid model of E2F1 for chemoresistance. The figure displays a CellDesigner map of the system and consists of three parts: (i) the core regulatory module, a transcriptional subnetwork accounting for the regulation of the expression and activity of E2F1 by proliferative and genotoxic stress signals, (ii) the target gene module, a network accounting for the E2F1 modulated expression of critical genes involved in the regulation of cell proliferation, genotoxic drug response and apoptosis, and (iii) the phenotypical read-outs (purple hexagons), a set of phenomenological variables that connect the transcriptional circuitry with phenotypic responses. The figure is adopted from Khan et al., [9].

4.5.3. Model construction

In order to illustrate the hybrid modeling approach, I adopted a kinetic model of the E2F1-p73/DNp73-miR-205 circuitry [121] using ordinary differential equations (ODEs) as a core module and extended with a feedback loop analysis of E2F1-EGFR. The kinetic

model accounts for the temporal evolution of system variables over time under basal and genotoxic stress conditions. The interactive components of the core module are: EGF, EGFR, ERK, E2F1, p73/DNp73 and miR205. The target gene module contains E2F1, p73/DNp73 and miR205 targets that account for different cellular functions like proliferation, apoptosis, and chemoresistance. The target gene module is modeled using a logical approach, which depends on the input vector [E2F1 p73 DNp73 miR-205]. For calculating the input vector, I used thresholds representing a drug-induced genotoxic stress level that is sufficient to induce high expression of E2F1, so that the inhibitory role of DNp73 dominates the activator role of p73 in miR-205 synthesis and the expression level of miR-205 reaches less than value of 1 [121].

(i) Core module: The core module is composed of nine ODEs accounting for the evolution in time of: inactive and active epidermal growth factor receptor (represented by EGFR and EGFR* respectively) (4.11), (4.12); extracellular-signal-regulated kinase (ERK) (4.13); transcription factor E2F1-mRNA (4.14); un-phosphorylated E2F1 (E2F1i) (4.15); phosphorylated E2F1 (E2F1p) (4.16); both isoforms of p73 (respectively, p73 and DNp73) (4.17), (4.18); and miR-205 (4.19).

$$\frac{dEGFR}{dt} = -k_1 \cdot (GF + k_2 \cdot EGFR^*) \cdot EGFR - \left(k_3 \cdot \frac{E2F1}{k_x + E2F1}\right) - k_4 \cdot EGFR \tag{4.11}$$

$$\frac{dEGFR^*}{dt} = k_1 \cdot (GF + k_2 \cdot EGFR^*) \cdot EGFR - k_{xx} \cdot EGFR^* \tag{4.12}$$

$$\frac{dpERK}{dt} = k_5 \cdot EGFR \cdot (ERKt - ERK) - k_6 \cdot pERK \tag{4.13}$$

$$\frac{dmE2F1}{dt} = GxS \cdot k_7 + FS \cdot k_8 - k_9 \cdot (1 + k_{10} \cdot miR205) \cdot mE2F1 \tag{4.14}$$

$$\frac{dE2F1i}{dt} = k_{11} \cdot mE2F1 - k_{12} \cdot E2F1i \cdot pERK + k_{13} \cdot E2F1p - k_{14} \cdot E2F1i \tag{4.15}$$

$$\frac{dE2F1p}{dt} = k_{15} \cdot E2F1i \cdot pERK - k_{16} \cdot E2F1p \tag{4.16}$$

$$\frac{dp73}{dt} = k_{17} \cdot E2F1p - k_{18} \cdot p73 \tag{4.17}$$

$$\frac{dDNp73}{dt} = k_{19} \cdot E2F1p - k_{20} \cdot DNp73 \tag{4.18}$$

$$\frac{dmiR205}{dt} = k_{21} \cdot TGFB^{-1} \cdot p73 \cdot \left(1 + \frac{DNp73}{k_{22}}\right) - k_{23} \cdot miR205 \tag{4.19}$$

The MM-type equation (4.11) accounts for the E2F1 mediated synthesis of **EGFR** (k_3) and its degradation (k_4) [121, 252, 253]. Equation (4.12) accounts for the external ligand EGF and auto regulation (k_1, k_2) of **active EGFR** (*i.e.,* $EGFR^*$) and its degradation (k_{xx}) [121, 254]. Equation (4.13) accounts for the EGFR-mediated phosphorylation (k_5) of **ERK** to represent mitogenic signaling pathways and its degradation (k_6) [255]. Equation (4.14) accounts for basal and genotoxic stress-mediated synthesis of **E2F1 mRNA**, characterized by parameter k_8 and k_7, respectively [121]. GxS is an input variable accounting for genotoxic stress. Furthermore, the core module contains a term for the miR-205 mediated E2F1

degradation (k_{10}) [256] and its basal degradation (k_9) [257]. Equation (4.15) accounts for the E2F1 mRNA-mediated synthesis (k_{11}) of un-phosphorylated E2F1 (here represented by E2F1i) [121, 258]. It also contains an expression for mitogenesis signal-mediated and basal degradation (k_{12}, k_{14}) of E2F1i [259]. Equation (4.16) accounts for the mitosis-mediated phosphorylation (k_{15}) of **E2F1 protein** (represented by E2F1p) and for its basal degradation (k_{16}) [259]. Equations (4.17) and (4.18) account for the E2F1-mediated synthesis of **p73** and **DNp73** (k_{17} and k_{19} respectively) [260], and basal degradation (k_{16} and k_{18}). In addition, equation (4.19) accounts for miR-205 synthesis and degradation (k_{21}, k_{23}) in which p73 acts as activator of **miR-205**, while DNp73 plays an inhibitory role [261, 262]. Here, we use a reduced power law term [207]. According to the observation reported for several tumor types, we assumed that when both isoforms of p73 are overexpressed the inhibitory effect of DNp73 dominates and cancels out p73 activation of miR-205 [263]. Besides, we included a term for the repression of miR-205 by the oncogenic signal from the TGFβ-1 pathway (TGFB1) [121, 264].

(ii) Discretization rules: Here I have chosen threshold values for discretization using manual training of the model variables E2F1, p73, DNp73 and miR205 [121]. If the continuous values of these variables lie below or above a certain limit, it is assigned a discrete value of '0', '1' or '2'. These discrete values represent different activity levels of those variables depending on Boolean or multi-valued logic. In case of Boolean logic, discrete value '0' represents low expression, which is not sufficient to activate the downstream target, while '1' represents high expression, which is sufficient to activate the downstream target. In multi-valued logic '0' represents low expression, '1' represents a medium level, and '2' represents a high level of expression. Using the system, I discretized E2F1 for multi-valued logic *i.e.,* '0', '1' and '2' representing low, medium and high expression respectively. p73, DNp73 and miR205 are discretized as '0' and '1' representing low and high expression.

(iii) Target gene and phenotypical module: In the proposed hybrid methodology, I modeled the target gene module in a multi-valued logic and phenotypical module in simple Boolean logic. The set of rules, listed in Table 4.3, describes how E2F1, p73, DNp73 and miR-205 activate or inhibit the target genes involved in proliferation, apoptosis, and chemoresistance. All reactions are supported by the literature (PubMed references are provided). The rules are written in CellNetAnayzer (CNA) notation for performing logical steady state analysis [199]: $\langle coefficient \rangle reactant = \langle coefficient \rangle product$. The activity levels of reactants and products are denoted by coefficient in front of them.

4.5.4. Hybrid model simulations of E2F1 network in drug resistance

I simulated cancer and non-cancer like cells with different input vectors (*i.e.,* initial values) by modulating the values of the model inputs GF, FS, TGFB1, and GxS Table 4.4. I considered two sets of scenarios, characterized by: (i) the presence of a sufficient amount

if $E2F1p_c < 1.15 \rightarrow E2F1p_D = 0$
if $1.15 \leq E2F1p_c(t_j) < 12 \rightarrow E2F1p_D = 1$
if $E2F1p_c \geq 12 \rightarrow E2F1p_D = 2$
if $miR205_c < 5 \rightarrow miR205_D = 0$
if $miR205_c \geq 5 \rightarrow miR205_D = 1$
if $p73_c < 1 \rightarrow p73_D = 0$
if $p73_c \geq 1 \rightarrow p73_D = 1$
if $DNp73_c < 10 \rightarrow DNp73_D = 0$
if $DNp73_c \geq 10 \rightarrow DNp73_D = 1$

Table 4.2. Definition of the discretization rules for E2F1, miR205, p73 and DNp73. $E2F1p_c$, $miR205_c$, $p73_c$, $DNp73_c$ represent continuous value of expression and $E2F1p_d$, $miR205_d$, $p73_d$, $DNp73_d$ represent their discrete state.

of growth factor ($GF = 1$) to trigger proliferation, and (ii) the lack of growth factor stimulation ($GF = 0$). Next, I modified sequentially and systematically the values of the other input variables, obtaining all possible biological scenarios accounting for the effect of genotoxic drug administration on cancer and non-cancer cells with/without overexpression of E2F1 and TGFβ-1. The hybrid model was used to compute the steady-state values of the model variables. The effect of genotoxic drug administration (GxS) in these different biological scenarios is substantiated in the model by the emerging gene expression pattern and triggering one of the phenotypes considered in the model. The computation of the complete set of simulations took approximately 1 second on a conventional workstation.

The results obtained are in accordance with those previously derived from an ODE model and with other experimental evidence in the literature [119, 121, 252]. However, our hybrid model delivered additional interesting results. For example, our model simulations for conditions of non-growth factor stimulation ($GF = 0$) indicate that in the case of sufficient overexpression of E2F1 tumor cells may become self-sufficient in growth signals [190, 219]. We link this simulation result to the existence of a positive feedback loop between E2F1 and EGFR [252]. Furthermore, sufficient overexpression of either E2F1 ($E2F1 = 10$) or TGFβ-1 ($TGFB1 = 10$) can promote resistance to anti-cancer drugs by the repression of miR-205. This conclusion is in agreement to the results of our previous publication [121]. On the other hand, another version of the model, completely constructed using multi-valued logic, fails to reproduce the features associated to the existence of the indicated feedback loops. I provided the details of the exclusive logical representation of the model in Appendix K.

Furthermore, I modified the E2F1 synthesis parameter (FS) and TGFβ-1 expression level (TGFB1) to investigate their combined effect on the emergence of a resistant phenotype [121]. To this end, the values of the input variables FS and TGFB1 were iteratively modified and simulations were performed to observe whether resistance to apoptosis initiation (chemoresistance) after stimulation with a genotoxic drug emerged

No.	Reactions	Functionality	Type of Rec.	References (PubMed)
1a	E2F1p = Gab2	Proliferation, Metastasis	Monotone	15574337, 17310989, 19342374
1b	2 E2F1p = 2 Gab2	Proliferation, Metastasis	Monotone	15574337, 17310989, 19342374
2a	E2F1p = CCND1	Proliferation, Metastasis	Monotone	17637753, 23278726
2b	2 E2F1p = 2 CCND1	Proliferation, Metastasis	Monotone	17637753, 23278726
3a	E2F1p = ANCCA	Proliferation, Metastasis	Monotone	19843847
3b	2 E2F1p = 2 ANCCA	Proliferation, Metastasis	Monotone	19843847
4a	E2F1p = MAP3Ks	Proliferation, Metastasis	Monotone	20026813
4b	2 E2F1p = 2 MAP3Ks	Proliferation, Metastasis	Monotone	20026813
5a	E2F1p = ACTR2	Proliferation, Metastasis	Monotone	20124470, 16648476
5b	2 E2F1p = 2 ACTR2	Proliferation, Metastasis	Monotone	20124470, 16648476
6a	E2F1p + GxD = EZH2	Proliferation, Metastasis	Monotone	19893569, 20675184
6b	E2F1p + 2 GxD = 2 EZH2	Proliferation, Metastasis	Monotone	19893569, 20675184
7	E2F1p + 2 !EZH2 = BIM	Apoptosis	Non-monotone	19893569, 20675184
8	E2F1p + GxD = FOXO3	Apoptosis	Non-monotone	20013022, 17482685
9	E2F1p + GxD = DUSPx	Apoptosis	Non-monotone	20013022
10	E2F1p + GxD = NOXA	Apoptosis	Non-monotone	17146434, 14684737
11	E2F1p + GxD = HRK	Apoptosis	Non-monotone	14684737, 9130713
12	P73 + !DNp73 = BAX	Apoptosis	Non-monotone	14634023
13	P73 + !DNp73 = PUMA	Apoptosis	Non-monotone	14684737, 14634023
14	!miR-205 + E2F1p = Bcl-2	Chemoresistance, Aggressiveness	Monotone	22871739, 21159167, 11559537
15	!miR-205 + E2F1p = ABCA2	Chemoresistance, Aggressiveness	Monotone	22871739, 21159167
16	!miR-205 + E2F1p = ABCA5	Chemoresistance, Aggressiveness	Monotone	22871739, 21159167
17	Proliferative_complex + !DUSPx + !FOXO3 = Proliferation		Monotone	20806045, 22107151
18	E2F1_Drug_apoptotic_comp + p73_apop_genes + 2 !EZH2 + !miR-205_reg_genes = Apoptosis		Monotone	19893569, 14634023, 22871739, 21159167
19	miR-205_reg_genes = Chemoresistance		Monotone	22871739

Table 4.3. Multi-valued logic-based rules of target gene and phenotypical module. This set of rules encodes how the E2F1-p73/DNp73-miR-205 circuitry regulates target genes involved in proliferation, apoptosis, and chemoresistance [9]. The orange color rows in table contain equations for target gene module and violet color rows for phenotypical module. The "+" symbol represents a logical 'AND' and the "!" symbol represents a logical 'NOT'. In CellNetAnalyzer the monotone type of reaction represent that the logical level of the product is activated by an equal or greater logical level of the reactant, while in non-monotone reactions the product level strictly dependents on the given logical level of reactant [199]. Here I modeled EZH2 as non-monotone which means it will inhibit apoptosis only when it is highly expressed (*i.e.,* logical level 2).

(Table 4.4). Our model simulations suggest that drug resistance emerges with lower combined expression level of E2F1 and TGFβ-1 compared to their individual expression levels. In contrast to Boolean models, the use of a hybrid model is sufficient to determine precise expression levels of E2F1 and TGFβ-1 shifting the phenotype from chemosensitive (blue area) to chemoresistant (red area) (Figure 4.11).

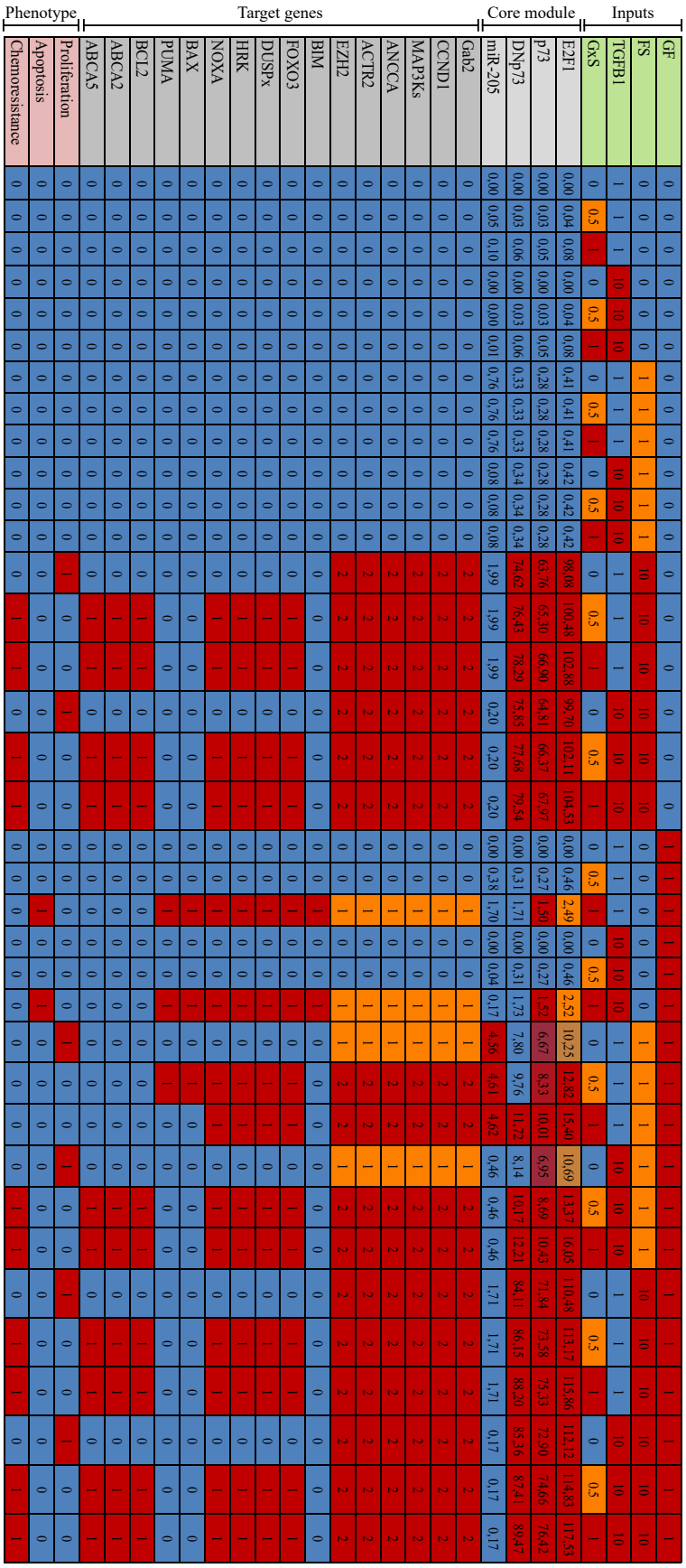


Table 4.4. Model simulations for the effect of genotoxic drug administration on cancer and non-cancer cells with/without overexpression of E2F1 and TGFβ-1, and with/without growth factor stimulations [9].

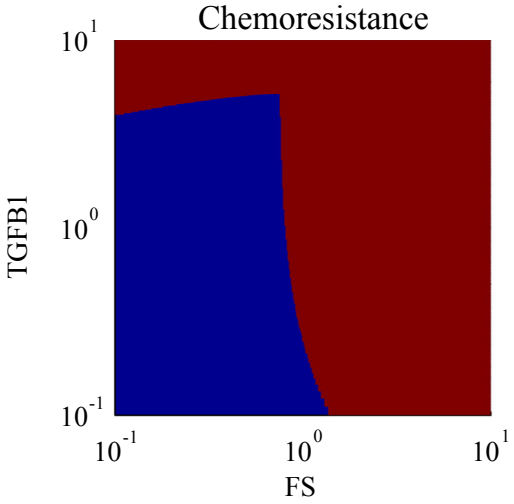


Figure 4.11. E2F1-TGFB1-Chemoresistance relation. Emergence of chemoresistance upon deregulation of E2F1 and TGFβ-1 expressions [9]. The values of the input variables accounting for E2F1 synthesis rate (FS) and TGFβ-1 expression (TGFB1) were iteratively modified in an interval of physiologically feasible values (values were normalized to the ones in the wild-type chemosensitive cell line described in [119, 121]). I simulated the system under genotoxic drug administration and analyzed for each pair of values for FS and TGFB1, respectively, whether apoptosis initiation was triggered (blue area) or the cell displays chemoresistance (red area).

MATLAB code for the hybrid model is given in Appendix J. Logic-based model (in CNA representation) of the target gene and phenotypical module can be access at[2].

4.6. Summary of results and discussion

In the present chapter, I proposed and illustrated the construction of hybrid models, composed of ODEs and logic sub-modules, as a strategy to handle large-scale, nonlinear biochemical networks. In this methodogy, I organized the network into three parts: (i) the core regulatory module, a subnetwork of signaling and transcriptional processes that displays high interconnectivity and is enriched in nonlinear regulatory loops, (ii) the target gene module, accounting for genes whose expression is directly regulated by the signaling proteins and transcription factors included in the core module, and (iii) the phenotypical read-outs. The core module is modeled using ODEs, while the target genes and the phenotypical read-outs are encoded using discrete logic modeling. Furthermore, an interface is designed that links both parts by discretizing the values of variables in the core module. Although the approach proposed seems simple, a number of questions emerge when trying to generalize the method and make it more accurate.

In this case study, the network construction relied on the manual retrieval of biomedical information and integration into a regulatory map. Real case study networks can be much more complex and large, especially in terms of the target genes,

[2]https://sourceforge.net/projects/khan-phd/files/Target_Gene_Phenotype_CNA.rar/download

which makes it hard to manually retrieve the information. One approach to construct these networks could be the systematic use of existing databases of protein-protein interactions and transcriptional regulation [121, 124, 265, 266]. Further, the determination of a core regulatory module is typically not straightforward. It is required, to determine in a systematic manner, which parts of the network display structural complexity and are enriched in nonlinear regulatory loops. Towards this, I developed a network-based integrative workflow (Chapter 2 Section 2.2 on Page 28) that combines network analysis, with expression data and biomedical information to identify disease-specific core-regulatory networks [124].

In addition, in some cases there may be feedback between the core regulatory module and the target gene module. In our case study, for example, EGFR is a transcriptional target of E2F1, in turn EGFR regulates E2F1 activity in a system that constitutes a positive feedback loop [252]. Given that this motif affects the regulation of the system, the dynamics of the transcriptional target EGFR were included in the nonlinear core regulatory module. This idea could be used to account for this kind of motifs in other models. In a more advanced setup of the methodology, network biology methods would be used to detect feedback loops between different parts of the network. These motifs would then be included in the nonlinear core regulatory module. In addition, the use of logical modeling would be reserved for linear transcriptional regulation.

When modeling the core regulatory module with ODEs, two fundamental tasks have to be addressed: first, the derivation of the model equations and second, the estimation of model parameters. The construction of model equations may benefit from using canonical modeling frameworks like classical kinetic or power-law modeling to obtain more consistent, homogeneous and easy to derive equations [207, 210, 232]. Secondly, the characterization of the ODE model will require adequate quantitative experimental data (dose-response experiments, quantitative time series, etc; see [207, 211] for further details), but also adequate methodologies for parameter estimation [214, 267].

A critical element of the proposed approach is to determine the discretization rules for the core regulatory variables that control the target genes. These rules are the basis of the interface between the ODE and logic modules. In this case study, the discretization rules were derived by manual training [211], but in a real case with a larger set of genes involved in the target gene module a more systematic approach will be necessary. This approach will require transcriptomics and proteomics data from experiments, in which expression and activity of the proteins and transcription factors controlling transcriptional programs are modulated in a dose-response fashion. Furthermore, computational methodologies will be required to determine precise discretization rules and gene-dependent activation thresholds for the transcriptional factors [268, 269].

In Table 4.5 we compared the performance of ODE, discrete logic and hybrid models with respect to a number of biological and computational features, which are important when choosing a strategy to construct a mathematical model for a biochemical system. The table indicates that among them there is not a perfect modeling framework, able

to deal better than the others in all the features considered. This is a conclusion that can be extended to a wider pool of modeling frameworks and emerges after inspecting decades of scientific publications on modeling of biochemical models [5, 6, 270, 271]. Some modeling frameworks are clearly better in some features but have a poor performance for other features. When choosing the right modeling framework one has to think which of those features that are more important in the biochemical system investigated. I think that hybrid models like the ones proposed are a good compromise between quantitative/qualitative accuracy and scalability when considering large biochemical networks with a small core of highly interconnected processes and a large set of transcriptionally and post-transcriptionally regulated genes.

	ODE models	Logic models	Hybrid models
Biophysical realism			
Accuracy			
Interpretability			
Encoding of non-linearity			
Calibration techniques			
Integrability of quantitative data			
Scalability			
Computability			

Legend	Good	Acceptable	Poor

Table 4.5. Comparison of ODE, discrete logic and hybrid models for the analysis of biochemical networks [9].

Chapter 5

Summary of results and discussion

Unraveling mechanisms underlying diseases for the prediction of diagnostic and therapeutic markers has motivated the development of various systems biology and bioinformatics approaches to process data. For example, (i) methods to identify expression patterns for normal and diseased states; (ii) graph theoretical analyses of biochemical networks to identify important hub nodes and small recurring regulatory motifs; and (iii) dynamical systems theory to simulate the dynamics of biological systems. Individually these approaches cannot capture the entire biological complexity associated with regulatory processes involved in diseases. For example, the expression patterns identified by data analysis tools, like machine learning, cannot explain the underlying mechanisms and causes that induce such patterns; the utility of graph theory is confined only to static analysis; and each modeling formalism offered by dynamical systems theory has its own limitations due to the fact that biological systems are enormous in size and complexity (*i.e.,* large number of components interacting in nonlinear fashion). The integration of multiple systems biology and bioinformatics methods, here named as "integrative workflows", is a promising approach to analyze large-scale, complex biochemical networks to provide insights into diseases. Further, these integrative workflows in systems biology are scarce but highly demanded [75].

Towards this, I developed an **integrative workflow** that combines heterogeneous sources of information to analyze large-scale biochemical networks in complex diseases (Figure 2.4). More specifically, the workflow integrates network structural properties with omics data, biomedical information and dynamical analysis to understand mechanisms underlying diseases. Using the proposed workflow, I analyzed a large-scale molecular interaction map of E2F1, a transcription factor involved in cancer, to understand its role in the regulation of epithelial-mesenchymal transition (EMT). First, we constructed a **molecular interaction map** from the literature and databases around E2F1 in cancer. The network is presented in standard SBGN format containing annotation for nodes and interactions. I performed network structural analysis using Cytoscape plugins to determine the centrality measures and identifying network motifs.

Centrality measures including degree centrality and betweenness centrality provide useful information *e.g.,* these measurements can be used to detect hub nodes (nodes with large degree) and gate keeper nodes (nodes with high betweenness centrality), which play an essential role in network organization. Network motifs are repeated small structural patterns and are considered to be the reservoir for network dynamics [2]. In biology, sets of genes/proteins/miRNAs involved in certain processes/phenotypes are highly interconnected and regulate each other through regulatory loops (motifs) including feedback and feedforward loops [2, 134]. Very often, these network motifs are disrupted or abnormally regulated in cancer, and the analysis of their differential regulation can provide important information on the emergence of disease phenotypes [134]. However, the identification of important feedback loops in a highly connected network is a methodological challenge [103, 154].

I here introduced a new motif ranking scheme using a weighted multi-objective function that combines the topological properties (*e.g.,* node degree and betweenness centrality) with non-topological properties (*e.g.,* gene expression), which enables to identify **core-regulatory networks** in bladder and breast cancer. We consider core-regulatory networks the important driver for the epithelial-mesenchymal transition (EMT). Topological properties account for the structural importance of the nodes, and non-topological properties for their disease-type and context-specific relevance. The core networks are amenable for dynamical modeling using a suitable mathematical formalism. Using logic-based models, the *in silico* analysis of core-networks identified molecular signatures and therapeutic targets for each cancer type. Model-based predictions were validated experimentally and through patient data. The workflow presented in this dissertation realizes a complete cycle of the system biology approach (presented in Figure 1.1).

I used a multi-objective function $F_M = \sum_{l=1}^{l=4} w_{lk} \cdot \sigma f_l$; where w_l is weight assigned to sub-function f_l which is normalized by σ) with multiple weighting scenarios ($w_k; k = 1, \ldots, n$) to provide motif ranking as unbiased as possible regarding the assessed objective functions. Here, I refer to $(ND + BC)$, GP, DP, and FC as four objective functions, where ND is the node degree, BC is betweenness centrality, GP is gene prioritization, DP is disease pathway association, and FC is expression fold change. Since we do not have any reference criteria for ranking, I manually chose 13 weighting scenarios, each giving weight (*i.e.,* importance) to one or another objective function. In this way I obtained the Pareto sets of ranking scores for each feedback loop. Pareto sets contain solutions where none of the optimized solution (F) can be improve in value for a given objective function without degrading some of the other objective functions. Top-ranked loops from each scenarios were selected and merged using Cytoscape to a obtain **core-regulatory network**.

I used the linear summation of multi-objective function that can highlight the solutions where sub-functions have uneven values *i.e.,* some sub-functions have minimal and other has maximal values. In this case the optimized solution is better for some sub-functions

and worse for others. The alternatives to linear summation can be, for example, the nonlinear product or power law function (*i.e.,* $F = \prod_j (\sigma f_j)^{w_j}$), which will highlight only those solutions where sub-functions has non-zero even values (*i.e.,* the solution is better for all sub-functions), which is practically hard if not impossible to achieve.

A major bottleneck to our workflow is the preparation of a directed regulatory network. This requires huge manual efforts as many state-of-art interaction databases either do not provide directions for interactions or they contain false positive/negative interactions complied by text-mining algorithms from published literature. I believe that new methods are needed for defining regulatory directions to interacting molecules. This may be possible by extracting regulatory information from databases (*e.g.,* BioModels database), literature and/or already published molecular interaction maps. Towards this, some efforts are already initiated, for example, the SIGNOR database [272] containing directed interactions; and the INDRA tool [273], which extracts molecular mechanisms in term of activation, inactivation and complex formation from the literature via text mining.

Using **logic-based modeling**, I analyzed the core-networks for: their stimulus-response behavior to identify molecular signatures relevant for disease progression; and perturbations to determine possible therapeutic targets. I divided the model into three layers: (i) input, (ii) regulatory, and (iii) output layer. The input layer contains E2F1 and the receptor molecules, and the regulatory layer contains all components of core network except E2F1 and receptor molecules. The output layer comprises only one node that represents the EMT process as the driver of invasive phenotype. The input layer and regulatory layers are encoded with Boolean logic. The phenotypical output is modeled with multi-valued logics, which accepts four ordinal levels ranging from 0 (no EMT) to 3 (high EMT). Here, multi-valued logic is used to assess the aggregated effect of various network components on the phenotype [155]. However, the use of multi-valued logic increases the complexity of model, therefore, I applied it only to the phenotypical output. Further, I used qualitative information based on fold-change expression data to derive the Boolean functions to approximate the activation level of molecules having multiple regulators.

For different combinations of receptor molecules and E2F1, the **model simulations revealed molecular signatures**, E2F1-TGFBR1-FGFR1 in bladder cancer and E2F1-TGFBR2-EGFR in breast cancer, that render cells from non-invasive to invasive cancer when highly expressed. Surprisingly, the other receptors that are part of the input layers in the models had no effect on the EMT process. Model predictions are in agreement with previous experimental findings in bladder and breast cancer studies, where high expression of these molecules were separately observed to regulate tumor invasion [203–205]. Further, to **confirm** the role of predicted signatures on the regulation of tumor invasion, we used bladder and breast cancer patient survival data from independent studies. For all our predicted signatures, high expression of the constituent molecules mapped to low patient survival and *vice versa* (Figure 3.5). Further, compare to random signatures, the model predicted signatures more distinctly

classify patients into early and advanced stages of cancer (Figure 3.6). Furthermore, the **model predictions were validated** by shRNA-based *in vitro* experiments (Fig. 5 on page 160). Inhibition of the predicted signatures in invasive bladder (UM-UC-3) and breast (MDA-MB231) cancer cell lines shows profound impact on the invasive behavior of the respective cell line and the highest effect (drastic reduction in invasion) observed upon combined inhibition (Fig. 5c on page 160). The role of other receptors, for example, EGFR in bladder cancer not constituting molecular signatures had no effect on tumor invasion, were also experimentally validated. Based on these results, one can propose that EGFR targeted therapies might be ineffective in invasive bladder cancer exhibiting elevated levels of E2F1. Moreover, the *in silico* perturbation experiments **identified potential therapeutic targets** for invasive cancer cells. From the list of identified candidates, provided in Table 3.3, our experimental partner validated SMAD3-NFKB1 in UM-UC-3 bladder cancer and SRC-FN1 in MDA-MB231 in breast cancer. The experimental results are in consensus with model prediction, which show significant reduction in cell invasion by inhibiting the therapeutic targets alone or in combination (more profound effect) (Fig. 6 on page 161).

Integrative workflow-based treatment recommendations for diseases, such as advanced bladder or breast cancer, have the potential to support cohort-specific patient treatment in order to avoid therapy resistance and cope with aggressive cancers. Further, I think the method proposed can be used to investigate other disease networks besides those focused on the E2F family discussed here.

Dynamical analysis of large-scale biochemical networks suffer from the scarcity of detailed quantitative data to characterize model parameters. Typically, such type of data is available for a small number of molecules, which allows the construction of a detailed quantitative model (*i.e.,* ODE-based models) for small networks. On the other hand, high throughput omics data can be used for coarse grained dynamical analysis (*e.g.,* using logic-based models) of large-scale networks. It is an emerging dilemma to consolidate models of large-scale biochemical networks and adequate description of nonlinear dynamics. To address this challenge, I developed a **hybrid modeling method**, which combines ODE-based models with logic-based models as a strategy to simulate large-scale, nonlinear biochemical networks as a dynamical system.

I illustrated the methodology by constructing a hybrid model of a biochemical network around the transcription factor E2F1 in drug resistance (Figure 4.10). The network is organized and divided into three modules with distinctive regulatory features: (i) Regulatory module, highly interconnected subnetwork containing feedback and feedforward loops, which induces nonlinearity in systems dynamics, (ii) Transcription activation of gene module, containing different sets of genes involved in certain phenotypes, which are directly regulated by molecular factors in the regulatory module, and (iii) Phenotypical read-outs module. The regulatory module was encoded by an ODE-based formalism, which is widely used to capture the nonlinear dynamics induced by regulatory motifs. The transcription gene module was modeled with

logic-based formalism, which showed how genes are activated/inactivated based on the expression level of variables in the regulatory module. I also encoded the phenotypical read-outs by logical formalism, which indicates how cell switch from one phenotype to other. The simulations of hybrid model are in consensus with those we previously derived from an ODE-based models, especially the effect of incoherent feedforward loops between E2F1-p73/DNp73-miR205 on drug resistance. Further, the hybrid model simulations derived additional interesting results, for example, in case of sufficient overexpression of E2F1 the tumor cell may become independent of growth signals (Table 4.4). We attribute this nonlinear behavior due to the feedback loop exist between E2F1 and EGFR (Figure 4.10).

A critical issue for hybrid models relates to the timing in the model simulations. ODE models are continuous-time models, in which time is, at least theoretically, a continuous variable, while in discrete logic-models the time is discrete *per se*. A more advanced version of the computational interface between the ODE and logic submodules will require rules for time discretization. Therefore, a criterion has to be established to re-compute iteratively in defined time-intervals, the discrete values for the regulators of the target genes (Table G.1 in Appendix G). This strategy could even be enhanced by integrating time-dependent gene expression and phenotypic data in the analysis.

An interesting option, when using a modular representation of regulatory networks connected to phenotypical readouts, is the investigation of phenotype probability distributions in heterogeneous cell populations. This kind of approach can be used to connect single-cell and cell-population models. Toward this alternative to our method, other approaches based on continuous variables may be considered (*e.g.,* partial differential equations) [274]. In the current study, our approach was used to investigate the behavior of the signaling and transcriptional machinery of single cells belonging to homogenous populations, and their connection to the emergence of given phenotypes. Thus, the discrete representation of the phenotypical readouts remains valid. Furthermore, the hybrid approach proposed could be integrated in more advanced methodologies accounting for the dynamics of cell populations, like for example agent-based models of cell populations [275]. In this case, our hybrid model approach would account for the regulatory machinery of each individual cell, which would be connected to the phenotypic parameters controlling, in the agent-based model, the dynamics of the modeled cell. For models of mid-size cell populations, the use of our hybrid approach for describing the internal regulatory machinery of each cell reduces the computational effort necessary for simulations of the agent-based model.

Exploring large-scale nonlinear dynamical networks will remain an art form. What I am aiming for here is a rational approach to what is effectively guesswork, forced upon us by the wonderful complexity found in living systems. The computational analyses of such networks can improve our understanding of disease processes in a mechanistic way. Ultimately, this shall provide the ability to manipulate and optimize processes towards treatment.

Bibliography

[1] Xionglei He and Jianzhi Zhang. Why do hubs tend to be essential in protein networks? *PLoS Genet*, 2(6):e88, 2006.

[2] Uri Alon. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8(6):450–461, 2007.

[3] David A Fell and Andreas Wagner. The small world of metabolism. *Nature Biotechnology*, 18(11):1121–1122, 2000.

[4] John J Tyson, Katherine C Chen, and Bela Novak. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Current Opinion in Cell Biology*, 15(2):221–231, 2003.

[5] Olaf Wolkenhauer. Why model? *Frontiers in Physiology*, 5:21, 2014.

[6] Nicolas Le Novère. Quantitative and logic modelling of molecular and gene networks. *Nature Reviews Genetics*, 16(3):146–158, 2015.

[7] Man-Sun Kim, Dongsan Kim, Nam Sook Kang, and Jeong-Rae Kim. The core regulatory network in human cells. *Biochemical and Biophysical Research Communications*, 484(2):348–353, 2017.

[8] Dongsan Kim, Man-Sun Kim, and Kwang-Hyun Cho. The core regulation module of stress-responsive regulatory networks in yeast. *Nucleic Acids Research*, 40(18): 8793–8802, 2012.

[9] Faiz M Khan, Ulf Schmitz, Svetoslav Nikolov, David Engelmann, Brigitte M Pützer, Olaf Wolkenhauer, and Julio Vera. Hybrid modeling of the crosstalk between signaling and transcriptional networks using ordinary differential equations and multi-valued logic. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1844(1):289–298, 2014.

[10] Trey Ideker, Timothy Galitski, and Leroy Hood. A new approach to decoding life: systems biology. *Annual Review of Genomics and Human Genetics*, 2(1):343–372, 2001.

[11] Hiroaki Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, 2002.

[12] Akira Funahashi, Mineo Morohashi, Hiroaki Kitano, and Naoki Tanimura. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *Biosilico*, 1(5):159–162, 2003.

[13] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–2504, 2003.

[14] Björn H Junker, Christian Klukas, and Falk Schreiber. VANTED: a system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics*, 7(1):109, 2006.

[15] Tobias Czauderna, Christian Klukas, and Falk Schreiber. Editing, validating and translating of SBGN maps. *Bioinformatics*, 26(18):2340–2341, 2010.

[16] Ina Koch. Petri nets in systems biology. *Software & Systems Modeling*, 14(2): 703–710, 2015.

[17] Maksat Ashyraliyev, Yves Fomekong-Nanfack, Jaap A Kaandorp, and Joke G Blom. Systems biology: parameter estimation for biochemical models. *Febs Journal*, 276(4):886–902, 2009.

[18] Xin Lai. *A Systems Biology Approach to Unravel the Cellular Function of MicroRNAs*. PhD thesis, 2012.

[19] Ulrike Wittig, Renate Kania, Martin Golebiewski, Maja Rey, Lei Shi, Lenneke Jong, Enkhjargal Algaa, Andreas Weidemann, Heidrun Sauer-Danzwith, Saqib Mir, et al. SABIO–RK–database for biochemical reaction kinetics. *Nucleic Acids Research*, 40(D1):D790–D796, 2012.

[20] Chen Li, Marco Donizelli, Nicolas Rodriguez, Harish Dharuri, Lukas Endler, Vijayalakshmi Chelliah, Lu Li, Enuo He, Arnaud Henry, Melanie I Stefan, et al. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Systems Biology*, 4(1):92, 2010.

[21] Faiz M Khan, Mehdi Sadeghi, Shailendra K Gupta, and Olaf Wolkenhauer. A Network-Based Integrative Workflow to Unravel Mechanisms Underlying Disease Progression. In *Systems Biology*, pages 247–276. Springer, 2018.

[22] Carl F Schaefer, Kira Anthony, Shiva Krupa, Jeffrey Buchoff, Matthew Day, Timo Hannay, and Kenneth H Buetow. PID: the pathway interaction database. *Nucleic Acids Research*, 37(suppl 1):D674–D679, 2009.

[23] Darryl Nishimura. BioCarta. *Biotech Software & Internet Report: The Computer Software Journal for Scient*, 2(3):117–120, 2001.

[24] Arnon Paz, Zippora Brownstein, Yaara Ber, Shani Bialik, Eyal David, Dorit Sagir, Igor Ulitsky, Ran Elkon, Adi Kimchi, Karen B Avraham, et al. SPIKE: a database of highly curated human signaling pathways. *Nucleic acids research*, 39(suppl_1): D793–D799, 2010.

[25] Alexander R Pico, Thomas Kelder, Martijn Pvan Iersel, Kristina Hanspers, Bruce R Conklin, and Chris Evelo. WikiPathways: pathway editing for the people. *PLoS Biology*, 6(7), 2008.

[26] *CST-Cell Signaling Technology Pathway Database*, 1999 (accessed February 3, 2016). http://www.cellsignal.com/.

[27] Tomáš Helikar, Bryan Kowal, Sean McClenathan, Mitchell Bruckner, Thaine Rowley, Alex Madrahimov, Ben Wicks, Manish Shrestha, Kahani Limbu, and Jim A Rogers. The cell collective: toward an open and collaborative approach to systems biology. *BMC Systems Biology*, 6(1):96, 2012.

[28] Robert Hoffmann and Alfonso Valencia. A gene network for navigating the literature. *Nature Genetics*, 36(7):664–664, 2004.

[29] Dávid Fazekas, Mihály Koltai, Dénes Türei, Dezső Módos, Máté Pálfy, Zoltán Dúl, Lilian Zsákai, Máté Szalay-Bekő, Katalin Lenti, Illés J Farkas, et al. SignaLink 2–a signaling pathway resource with multi-layered regulatory networks. *BMC Systems Biology*, 7(1):1, 2013.

[30] Kumaran Kandasamy, Sujatha S Mohan, Rajesh Raju, Shivakumar Keerthikumar, Ghantasala S Sameer Kumar, Abhilash K Venugopal, Deepthi Telikicherla, Daniel J Navarro, Suresh Mathivanan, Christian Pecquet, et al. NetPath: a public resource of curated signal transduction pathways. *Genome Biology*, 11(1):1–9, 2010.

[31] Minoru Kanehisa and Susumu Goto. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30, 2000.

[32] Edgar Wingender, Xin Chen, R Hehl, Holger Karas, Ines Liebich, V Matys, T Meinhardt, M Prüß, Ingmar Reuter, and Frank Schacherer. TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Research*, 28(1): 316–319, 2000.

[33] Heonjong Han, Jae-Won Cho, Sangyoung Lee, Ayoung Yun, Hyojin Kim, Dasom Bae, Sunmo Yang, Chan Yeong Kim, Muyoung Lee, Eunbeen Kim, et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Research*, 46(D1):D380–D386, 2017.

[34] Luiz A Bovolenta, Marcio L Acencio, and Ney Lemke. HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics*, 13(1):405, 2012.

[35] C Jiang, Zhenyu Xuan, Fang Zhao, and Michael Q Zhang. TRED: a transcriptional regulatory element database, new entries and other development. *Nucleic Acids Research*, 35(suppl 1):D137–D140, 2007.

[36] Mathias Krull, Susanne Pistor, Nico Voss, Alexander Kel, Ingmar Reuter, Deborah Kronenberg, Holger Michael, Knut Schwarzer, Anatolij Potapov, Claudia Choi, et al. TRANSPATH®: an information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acids Research*, 34(suppl 1): D546–D551, 2006.

[37] Natascha Bushati and Stephen M Cohen. microRNA functions. *Annu. Rev. Cell Dev. Biol.*, 23:175–205, 2007.

[38] Feifei Xiao, Zhixiang Zuo, Guoshuai Cai, Shuli Kang, Xiaolian Gao, and Tongbin Li. miRecords: an integrated resource for microRNA–target interactions. *Nucleic Acids Research*, 37(suppl 1):D105–D110, 2009.

[39] Praveen Sethupathy, Benoit Corda, and Artemis G Hatzigeorgiou. TarBase: A comprehensive database of experimentally supported animal microRNA targets. *RNA*, 12(2):192–197, 2006.

[40] Sheng-Da Hsu, Feng-Mao Lin, Wei-Yun Wu, Chao Liang, Wei-Chih Huang, Wen-Ling Chan, Wen-Ting Tsai, Goun-Zhou Chen, Chia-Jung Lee, Chih-Min Chiu, et al. miRTarBase: a database curates experimentally validated microRNA–target interactions. *Nucleic acids research*, 39(suppl_1):D163–D169, 2010.

[41] Harsh Dweep, Carsten Sticht, Priyanka Pandey, and Norbert Gretz. miRWalk–database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. *Journal of Biomedical Informatics*, 44(5):839–847, 2011.

[42] Panagiotis Alexiou, Thanasis Vergoulis, Martin Gleditzsch, George Prekas, Theodore Dalamagas, Molly Megraw, Ivo Grosse, Timos Sellis, and Artemis G Hatzigeorgiou. miRGen 2.0: a database of microRNA genomic information and regulation. *Nucleic acids research*, 38(suppl_1):D137–D141, 2009.

[43] Juan Wang, Ming Lu, Chengxiang Qiu, and Qinghua Cui. TransmiR: a transcription factor–microRNA regulation database. *Nucleic Acids Research*, 38 (suppl 1):D119–D122, 2010.

[44] Donna Karolchik, Robert Baertsch, Mark Diekhans, Terrence S Furey, Angie Hinrichs, YT Lu, Krishna M Roskin, Matthias Schwartz, Charles W Sugnet,

Daryl J Thomas, et al. The UCSC genome browser database. *Nucleic Acids Research*, 31(1):51–54, 2003.

[45] Peter D Karp, Christos A Ouzounis, Caroline Moore-Kochlacs, Leon Goldovsky, Pallavi Kaipa, Dag Ahrén, Sophia Tsoka, Nikos Darzentas, Victor Kunin, and Núria López-Bigas. Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Research*, 33(19):6083–6089, 2005.

[46] Ron Caspi, Hartmut Foerster, Carol A Fulcher, Rebecca Hopkinson, John Ingraham, Pallavi Kaipa, Markus Krummenacker, Suzanne Paley, John Pick, Seung Y Rhee, et al. MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Research*, 34(suppl 1):D511–D516, 2006.

[47] Antje Chang, Maurice Scheer, Andreas Grote, Ida Schomburg, and Dietmar Schomburg. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Research*, 37(suppl 1):D588–D592, 2009.

[48] Jan Schellenberger, Junyoung O Park, Tom M Conrad, and Bernhard Ø Palsson. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, 11(1):213, 2010.

[49] John W Whitaker, Ivica Letunic, Glenn A McConkey, and David R Westhead. metaTIGER: a metabolic evolution resource. *Nucleic Acids Research*, 37(suppl 1): D531–D538, 2009.

[50] Benjamin A Shoemaker and Anna R Panchenko. Deciphering protein–protein interactions. Part II. Computational methods to predict protein and domain interaction partners. *PLoS Comput Biol*, 3(4):e43, 2007.

[51] Chris Stark, Bobby-Joe Breitkreutz, Teresa Reguly, Lorrie Boucher, Ashton Breitkreutz, and Mike Tyers. BioGRID: a general repository for interaction datasets. *Nucleic Acids Research*, 34(suppl 1):D535–D539, 2006.

[52] TS Keshava Prasad, Renu Goel, Kumaran Kandasamy, Shivakumar Keerthikumar, Sameer Kumar, Suresh Mathivanan, Deepthi Telikicherla, Rajesh Raju, Beema Shafreen, Abhilash Venugopal, et al. Human protein reference database–2009 update. *Nucleic Acids Research*, 37(suppl 1):D767–D772, 2009.

[53] Damian Szklarczyk, Andrea Franceschini, Michael Kuhn, Milan Simonovic, Alexander Roth, Pablo Minguez, Tobias Doerks, Manuel Stark, Jean Muller, Peer Bork, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Research*, 39(suppl 1): D561–D568, 2011.

[54] Sandra Orchard, Mais Ammari, Bruno Aranda, Lionel Breuza, Leonardo Briganti, Fiona Broackes-Carter, Nancy H Campbell, Gayatri Chavali, Carol Chen, Noemi Del-Toro, et al. The MIntAct project-IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Research*, 42(D1):D358–D363, 2013.

[55] Hans-Werner Mewes, Dmitrij Frishman, Ulrich Güldener, Gertrud Mannhaupt, K Mayer, Martin Mokrejs, Burkhard Morgenstern, Martin Münsterkötter, Stephen Rudd, and B Weil. MIPS: a database for genomes and protein sequences. *Nucleic Acids Research*, 30(1):31–34, 2002.

[56] Arunachalam Vinayagam, Jonathan Zirin, Charles Roesel, Yanhui Hu, Bahar Yilmazel, Anastasia A Samsonova, Ralph A Neumüller, Stephanie E Mohr, and Norbert Perrimon. Integrating protein-protein interaction networks with phenotypes reveals signs of interactions. *Nature Methods*, 11(1):94–99, 2014.

[57] Kazuhiro A Fujita, Marek Ostaszewski, Yukiko Matsuoka, Samik Ghosh, Enrico Glaab, Christophe Trefois, Isaac Crespo, Thanneer M Perumal, Wiktor Jurkowski, Paul MA Antony, et al. Integrating pathways of Parkinson's disease in a molecular interaction map. *Molecular Neurobiology*, 49(1):88–102, 2014.

[58] Yukiko Matsuoka, Hiromi Matsumae, Manami Katoh, Amie J Eisfeld, Gabriele Neumann, Takeshi Hase, Samik Ghosh, Jason E Shoemaker, Tiago JS Lopes, Tokiko Watanabe, et al. A comprehensive map of the influenza A virus replication cycle. *BMC Systems Biology*, 7(1):1, 2013.

[59] Gang Wu, Lisha Zhu, Jennifer E Dent, and Christine Nardini. A comprehensive molecular interaction map for rheumatoid arthritis. *PloS One*, 5(4):e10137, 2010.

[60] Hiroaki Kitano. A graphical notation for biochemical networks. *Biosilico*, 1(5): 169–176, 2003.

[61] Isabelle Pirson, Nathalie Fortemaison, Christine Jacobs, Sarah Dremier, Jacques E Dumont, and Carine Maenhaut. The visual display of regulatory information and networks. *Trends in Cell Biology*, 10(10):404–408, 2000.

[62] Daniel L Cook, Joel F Farley, and Stephen J Tapscott. A basis for a visual language for describing, archiving and analyzing functional models of complex biological systems. *Genome Biology*, 2(4):research0012–1, 2001.

[63] Ron Maimon and Sam Browning. Diagrammatic notation and computational structure of gene networks. In *Proceedings of the Second International Conference on Systems Biology*, volume 1, pages 311–317, 2001.

[64] Kurt W Kohn. Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Molecular Biology of the Cell*, 10(8):2703–2734, 1999.

[65] Nicolas Le Novere, Michael Hucka, Huaiyu Mi, Stuart Moodie, Falk Schreiber, Anatoly Sorokin, Emek Demir, Katja Wegner, Mirit I Aladjem, Sarala M Wimalaratne, et al. The systems biology graphical notation. *Nature Biotechnology*, 27(8):735–741, 2009.

[66] Hiroaki Kitano, Akira Funahashi, Yukiko Matsuoka, and Kanae Oda. Using process diagrams for the graphical representation of biological networks. *Nature Biotechnology*, 23(8):961–966, 2005.

[67] Michael Hucka, Andrew Finney, Herbert M Sauro, Hamid Bolouri, John C Doyle, Hiroaki Kitano, Adam P Arkin, Benjamin J Bornstein, Dennis Bray, Athel Cornish-Bowden, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19 (4):524–531, 2003.

[68] Hiroaki Kitano. A robustness-based approach to systems-oriented drug design. *Nature Reviews Drug Discovery*, 5(3):202–210, 2007.

[69] Attila Gursoy, Ozlem Keskin, and Ruth Nussinov. Topological properties of protein interaction networks from a structural perspective. *Biochemical Society Transactions*, 36(6):1398–1403, 2008.

[70] Zhihua Zhang and Jianzhi Zhang. A big world inside small-world networks. *PloS One*, 4(5):e5686, 2009.

[71] Hawoong Jeong, Sean P Mason, A-L Barabási, and Zoltan N Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–42, 2001.

[72] Albert-Laszlo Barabasi and Zoltan N Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004.

[73] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

[74] Max Kotlyar, Kristen Fortney, and Igor Jurisica. Network-based characterization of drug-regulated genes, drug targets, and toxicity. *Methods*, 57(4):499–507, 2012.

[75] Koyel Mitra, Anne-Ruxandra Carvunis, Sanath Kumar Ramesh, and Trey Ideker. Integrative approaches for finding modular structure in biological networks. *Nature Reviews Genetics*, 14(10):719–732, 2013.

[76] Xiujuan Wang, Natali Gulbahce, and Haiyuan Yu. Network-based methods for human disease gene prediction. *Briefings in Functional Genomics*, 10(5):280–293, 2011.

[77] Martin Oti and Han G Brunner. The modular nature of genetic diseases. *Clinical Genetics*, 71(1):1–11, 2007.

[78] Esti Yeger-Lotem, Shmuel Sattath, Nadav Kashtan, Shalev Itzkovitz, Ron Milo, Ron Y Pinter, Uri Alon, and Hanah Margalit. Network motifs in integrated cellular networks of transcription–regulation and protein–protein interaction. *Proceedings of the National Academy of Sciences of the United States of America*, 101(16): 5934–5939, 2004.

[79] Ezio Bartocci and Pietro Lió. Computational modeling, formal analysis, and tools for systems biology. *PLoS Comput Biol*, 12(1):e1004591, 2016.

[80] Valentina Raia, Marcel Schilling, Martin Böhm, Bettina Hahn, Andreas Kowarsch, Andreas Raue, Carsten Sticht, Sebastian Bohl, Maria Saile, Peter Möller, et al. Dynamic mathematical modeling of IL13-induced signaling in Hodgkin and primary mediastinal B-cell lymphoma allows prediction of therapeutic targets. *Cancer Research*, 71(3):693–704, 2011.

[81] Zhike Zi and Edda Klipp. Constraint-based modeling and kinetic analysis of the Smad dependent TGF-β signaling pathway. *PloS One*, 2(9):e936, 2007.

[82] Rui-Sheng Wang, Assieh Saadatpour, and Reka Albert. Boolean modeling in systems biology: an overview of methodology and applications. *Physical Biology*, 9(5):055001, 2012.

[83] Zhiwei Ji, Ke Yan, Wenyang Li, Haigen Hu, and Xiaoliang Zhu. Mathematical and Computational Modeling in Complex Biological Systems. *BioMed Research International*, 2017, 2017.

[84] Regina Samaga, Julio Saez-Rodriguez, Leonidas G Alexopoulos, Peter K Sorger, and Steffen Klamt. The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. *PLoS Comput Biol*, 5(8):e1000438, 2009.

[85] Rebekka Schlatter, Nicole Philippi, Gaby Wangorsch, Robert Pick, Oliver Sawodny, Christoph Borner, Jens Timmer, Michael Ederer, and Thomas Dandekar. Integration of Boolean models exemplified on hepatocyte signal transduction. *Briefings in Bioinformatics*, 13(3):365–376, 2012.

[86] Stefan Bornholdt. Less is more in modeling large genetic networks. *Science*, 310 (5747):449–451, 2005.

[87] Assieh Saadatpour and Réka Albert. Boolean modeling of biological regulatory networks: a methodology tutorial. *Methods*, 62(1):3–12, 2013.

[88] Thomas Handorf and Edda Klipp. Modeling mechanistic biological networks: an advanced Boolean approach. *Bioinformatics*, 28(4):557–563, 2012.

[89] Steven Watterson and Peter Ghazal. Use of logic theory in understanding regulatory pathway signaling in response to infection. *Future Microbiology*, 5(2): 163–176, 2010.

[90] Julio Saez-Rodriguez, Luca Simeoni, Jonathan A Lindquist, Rebecca Hemenway, Ursula Bommhardt, Boerge Arndt, Utz-Uwe Haus, Robert Weismantel, Ernst D Gilles, Steffen Klamt, et al. A logical model provides insights into T cell receptor signaling. *PLoS Comput Biol*, 3(8):e163, 2007.

[91] Rebekka Schlatter, Kathrin Schmich, Ima Avalos Vizcarra, Peter Scheurich, Thomas Sauter, Christoph Borner, Michael Ederer, Irmgard Merfort, and Oliver Sawodny. ON/OFF and beyond-a boolean model of apoptosis. *PLoS Comput Biol*, 5(12):e1000595, 2009.

[92] Assieh Saadatpour, Rui-Sheng Wang, Aijun Liao, Xin Liu, Thomas P Loughran, István Albert, and Réka Albert. Dynamical and structural analysis of a T cell survival network identifies novel candidate therapeutic targets for large granular lymphocyte leukemia. *PLoS Comput Biol*, 7(11):e1002267, 2011.

[93] Saikat Chowdhury, Rachana N Pradhan, and Ram Rup Sarkar. Structural and logical analysis of a comprehensive hedgehog signaling pathway to identify alternative drug targets for glioma, colon and pancreatic cancer. *PloS One*, 8(7): e69132, 2013.

[94] Sarah M Assmann and Réka Albert. Discrete dynamic modeling with asynchronous update, or how to model complex systems in the absence of quantitative information. *Plant Systems Biology*, pages 207–225, 2009.

[95] Vincent Noël, Sergey Vakulenko, and Ovidiu Radulescu. A hybrid mammalian cell cycle model. *arXiv preprint arXiv:1309.0870*, 2013.

[96] Roberta Alfieri, Ezio Bartocci, Emanuela Merelli, and Luciano Milanesi. Modeling the cell cycle: From deterministic models to hybrid systems. *Biosystems*, 105(1): 34–40, 2011.

[97] Andrew K Miller, Justin Marsh, Adam Reeve, Alan Garny, Randall Britten, Matt Halstead, Jonathan Cooper, David P Nickerson, and Poul F Nielsen. An overview of the CellML API and its implementation. *BMC Bioinformatics*, 11(1):178, 2010.

[98] Yanyan Ping, Yulan Deng, Li Wang, Hongyi Zhang, Yong Zhang, Chaohan Xu, Hongying Zhao, Huihui Fan, Fulong Yu, Yun Xiao, et al. Identifying core gene modules in glioblastoma based on multilayer factor-mediated dysfunctional regulatory networks through integrating multi-dimensional genomic data. *Nucleic Acids Research*, 43(4):1997–2007, 2015.

[99] Mohamed Hamed, Christian Spaniol, Alexander Zapp, and Volkhard Helms. Integrative network-based approach identifies key genetic elements in breast invasive carcinoma. *BMC Genomics*, 16(5):S2, 2015.

[100] Wenting Li, Rui Wang, Linfu Bai, Zhangming Yan, and Zhirong Sun. Cancer core modules identification through genomic and transcriptomic changes correlation detection at network level. *BMC Systems Biology*, 6(1):1, 2012.

[101] Pei Wang, Jinhu Lü, and Xinghuo Yu. Identification of important nodes in directed biological networks: A network motif approach. *PloS One*, 9(8):e106132, 2014.

[102] William T Budd, Sarah Seashols, Danielle Weaver, Cyriac Joseph, and Zendra E Zehner. A networks method for ranking microRNA dysregulation in cancer. *BMC Systems Biology*, 7(5):1, 2013.

[103] Yuji Zhang, Jianhua Xuan, G Benilo, Robert Clarke, and Habtom W Ressom. Network motif-based identification of breast cancer susceptibility genes. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5696–5699. IEEE, 2008.

[104] Mika Gustafsson, Colm E Nestor, Huan Zhang, Albert-László Barabási, Sergio Baranzini, Sören Brunak, Kian Fan Chung, Howard J Federoff, Anne-Claude Gavin, Richard R Meehan, et al. Modules, networks and systems medicine for understanding disease and aiding diagnosis. *Genome Medicine*, 6(10):82, 2014.

[105] Adrien Rougny, Pauline Gloaguen, Nathalie Langonné, Eric Reiter, Pascale Crépieux, Anne Poupon, and Christine Froidevaux. A logic-based method to build signaling networks and propose experimental plans. *Scientific Reports*, 8(1):7830, 2018.

[106] Javad Zahiri, Joseph Hannon Bozorgmehr, and Ali Masoudi-Nejad. Computational prediction of protein–protein interaction networks: algorithms and resources. *Current Genomics*, 14(6):397–414, 2013.

[107] Wei-Po Lee and Wen-Shyong Tzou. Computational methods for discovering gene networks from expression data. *Briefings in Bioinformatics*, 10(4):408–423, 2009.

[108] Etienne Caron, Samik Ghosh, Yukiko Matsuoka, Dariel Ashton-Beaucage, Marc Therrien, Sébastien Lemieux, Claude Perreault, Philippe P Roux, and Hiroaki Kitano. A comprehensive map of the mTOR signaling network. *Molecular Systems Biology*, 6(1):453, 2010.

[109] Laurence Calzone, Amélie Gelay, Andrei Zinovyev, François Radvanyi, and Emmanuel Barillot. A comprehensive modular map of molecular interactions in RB/E2F pathway. *Molecular Systems Biology*, 4(1):0174, 2008.

[110] Kanae Oda, Yukiko Matsuoka, Akira Funahashi, and Hiroaki Kitano. A comprehensive pathway map of epidermal growth factor receptor signaling. *Molecular Systems Biology*, 1(1), 2005.

[111] Anna Ritz, Christopher L Poirel, Allison N Tegge, Nicholas Sharp, Kelsey Simmons, Allison Powell, Shiv D Kale, and TM Murali. Pathways on demand: automated reconstruction of human signaling networks. *npj Systems Biology and Applications*, 2:16002, 2016.

[112] Jochen Supper, Lucía Spangenberg, Hannes Planatscher, Andreas Dräger, Adrian Schröder, and Andreas Zell. BowTieBuilder: modeling signal transduction pathways. *BMC systems biology*, 3(1):67, 2009.

[113] Alexander Martin, Maria E Ochagavia, Laya C Rabasa, Jamilet Miranda, Jorge Fernandez-de Cossio, and Ricardo Bringas. BisoGenet: a new tool for gene network building, visualization and analysis. *BMC Bioinformatics*, 11(1):91, 2010.

[114] Dénes Türei, Tamás Korcsmáros, and Julio Saez-Rodriguez. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nature Methods*, 13(12):966, 2016.

[115] Ulf Schmitz, Xin Lai, Felix Winter, Olaf Wolkenhauer, Julio Vera, and Shailendra K Gupta. Cooperative gene regulation by microRNA pairs and their identification using a computational workflow. *Nucleic Acids Research*, 42(12):7539–7552, 2014.

[116] Shirley Polager and Doron Ginsberg. E2F–at the crossroads of life and death. *Trends in Cell Biology*, 18(11):528–535, 2008.

[117] Vijay Alla, David Engelmann, Annett Niemetz, Jens Pahnke, Anke Schmidt, Manfred Kunz, Stephan Emmrich, Marc Steder, Dirk Koczan, and Brigitte M Pützer. E2F1 in melanoma progression and metastasis. *Journal of the National Cancer Institute*, 2009.

[118] Brigitte M Pützer, Marc Steder, and Vijay Alla. Predicting and preventing melanoma invasiveness: advances in clarifying E2F1 function. *Expert Review of Anticancer Therapy*, 10(11):1707–1720, 2010.

[119] Vijay Alla, Bhavani S Kowtharapu, David Engelmann, Stephan Emmrich, Ulf Schmitz, Marc Steder, and Brigitte M Pützer. E2F1 confers anticancer drug resistance by targeting ABC transporter family members and Bcl-2 via the p73/DNp73-miR-205 circuitry. *Cell Cycle*, 11(16):3067–3078, 2012.

[120] David Engelmann and Brigitte M Pützer. The dark side of E2F1: in transit beyond apoptosis. *Cancer Research*, 72(3):571–575, 2012.

[121] Julio Vera, Ulf Schmitz, Xin Lai, David Engelmann, Faiz M Khan, Olaf Wolkenhauer, and Brigitte M Pützer. Kinetic Modeling–Based Detection of Genetic Signatures That Provide Chemoresistance via the E2F1-p73/DNp73-miR-205 Network. *Cancer Research*, 73(12):3511–3524, 2013.

[122] Claudia Meier, Alf Spitschak, Kerstin Abshagen, Shailendra Gupta, Joel M Mor, Olaf Wolkenhauer, Jörg Haier, Brigitte Vollmar, Vijay Alla, and Brigitte M Pützer. Association of RHAMM with E2F1 promotes tumour cell extravasation by transcriptional up-regulation of fibronectin. *The Journal of Pathology*, 234(3): 351–364, 2014.

[123] Bruno Aranda, Hagen Blankenburg, Samuel Kerrien, Fiona SL Brinkman, Arnaud Ceol, Emilie Chautard, Jose M Dana, Javier De Las Rivas, Marine Dumousseau, Eugenia Galeota, et al. PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nature Methods*, 8(7):528–529, 2011.

[124] Faiz M Khan, Stephan Marquardt, Shailendra K Gupta, Susanne Knoll, Ulf Schmitz, Alf Spitschak, David Engelmann, Julio Vera, Olaf Wolkenhauer, and Brigitte M Pützer. Unraveling a tumor type-specific regulatory core underlying E2F1-mediated epithelial-mesenchymal transition to predict receptor protein signatures. *Nature Communications*, 8(1):198, 2017.

[125] Sue Povey, Ruth Lovering, Elspeth Bruford, Mathew Wright, Michael Lush, and Hester Wain. The HUGO gene nomenclature committee (HGNC). *Human Genetics*, 109(6):678–680, 2001.

[126] Donna Maglott, Jim Ostell, Kim D Pruitt, and Tatiana Tatusova. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Research*, 33(suppl_1):D54–D58, 2005.

[127] UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Research*, 43(D1):D204–D212, 2014.

[128] Inna Kuperstein, David PA Cohen, Stuart Pook, Eric Viara, Laurence Calzone, Emmanuel Barillot, and Andrei Zinovyev. NaviCell: a web-based environment for navigation, curation and maintenance of large molecular interaction maps. *BMC Systems Biology*, 7(1):100, 2013.

[129] Andrei Zinovyev, Eric Viara, Laurence Calzone, and Emmanuel Barillot. BiNoM: a Cytoscape plugin for manipulating and analyzing biological networks. *Bioinformatics*, 24(6):876–877, 2007.

[130] Yassen Assenov, Fidel Ramírez, Sven-Eric Schelhorn, Thomas Lengauer, and Mario Albrecht. Computing topological parameters of biological networks. *Bioinformatics*, 24(2):282–284, 2008.

[131] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[132] Erzsébet Ravasz, Anna Lisa Somera, Dale A Mongru, Zoltán N Oltvai, and A-L Barabási. Hierarchical organization of modularity in metabolic networks. *Science*, 297(5586):1551–1555, 2002.

[133] Qian Peng and Nicholas J Schork. Utility of network integrity methods in therapeutic target identification. *Frontiers in Genetics*, 5, 2014.

[134] John J Tyson and Béla Novák. Functional motifs in biochemical reaction networks. *Annual Review of Physical Chemistry*, 61:219–240, 2010.

[135] René Thomas and Richard d'Ari. *Biological feedback*. CRC press, 1990.

[136] Duc-Hau Le and Yung-Keun Kwon. NetDS: a Cytoscape plugin to analyze the robustness of dynamics and feedforward/feedback loop structures of biological networks. *Bioinformatics*, 27(19):2767–2768, 2011.

[137] Nevan J Krogan, Scott Lippman, David A Agard, Alan Ashworth, and Trey Ideker. The cancer cell map initiative: defining the hallmark networks of cancer. *Molecular Cell*, 58(4):690–698, 2015.

[138] Shinichiro Wachi, Ken Yoneda, and Reen Wu. Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics*, 21(23):4205–4208, 2005.

[139] Pall F Jonsson and Paul A Bates. Global topological features of cancer proteins in the human interactome. *Bioinformatics*, 22(18):2291–2297, 2006.

[140] Matan Hofree, John P Shen, Hannah Carter, Andrew Gross, and Trey Ideker. Network-based stratification of tumor mutations. *Nature Methods*, 10(11):1108–1115, 2013.

[141] Stephen Oliver. Proteomics: guilt-by-association goes global. *Nature*, 403(6770):601–603, 2000.

[142] Hyeonjeong Lee and Miyoung Shin. Mining pathway associations for disease-related pathway activity analysis based on gene expression and methylation data. *BioData Mining*, 10(1):3, 2017.

[143] Julio Vera, Carlos González-Alcón, Alberto Marín-Sanguino, and Néstor Torres. Optimization of biochemical systems through mathematical programming: Methods and applications. *Computers & Operations Research*, 37(8):1427–1438, 2010.

[144] Samy Lamouille, Jian Xu, and Rik Derynck. Molecular mechanisms of epithelial–mesenchymal transition. *Nature Reviews Molecular Cell Biology*, 15(3):178–196, 2014.

[145] Duc-Hau Le and Yung-Keun Kwon. GPEC: a Cytoscape plug-in for random walk-based gene prioritization and biomedical evidence collection. *Computational Biology and Chemistry*, 37:17–23, 2012.

[146] Sebastian Köhler, Sebastian Bauer, Denise Horn, and Peter N Robinson. Walking the interactome for prioritization of candidate disease genes. *The American Journal of Human Genetics*, 82(4):949–958, 2008.

[147] Saket Navlakha and Carl Kingsford. The power of protein interaction networks for associating genes with diseases. *Bioinformatics*, 26(8):1057–1063, 2010.

[148] Christiaan Klijn, Steffen Durinck, Eric W Stawiski, Peter M Haverty, Zhaoshi Jiang, Hanbin Liu, Jeremiah Degenhardt, Oleg Mayba, Florian Gnad, Jinfeng Liu, et al. A comprehensive transcriptional portrait of human cancer cell lines. *Nature Biotechnology*, 33(3):306–312, 2015.

[149] Simon Anders and Wolfgang Huber. Differential expression analysis for sequence count data. *Genome Biology*, 11(10):R106, 2010.

[150] Faiz M Khan, Shailendra K Gupta, and Olaf Wolkenhauer. Integrative workflows for network analysis. *Essays in Biochemistry*, 62(4):549–561, 2018.

[151] Mariano J Alvarez, Yao Shen, Federico M Giorgi, Alexander Lachmann, B Belinda Ding, B Hilda Ye, and Andrea Califano. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nature Genetics*, 2016.

[152] Steven Nathaniel Steinway, Jorge GT Zanudo, Wei Ding, Carl Bart Rountree, D Feith, TP Loughran, and R Albert. Network modeling of TGFb signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint sonic hedgehog and Wnt pathway activation. *Cancer Research*, 21:5963–977, 2014.

[153] Junyan Lu, Hanlin Zeng, Zhongjie Liang, Limin Chen, Liyi Zhang, Hao Zhang, Hong Liu, Hualiang Jiang, Bairong Shen, Ming Huang, et al. Network modelling reveals the mechanism underlying colitis-associated colon cancer and identifies novel combinatorial anti-cancer targets. *Scientific Reports*, 5, 2015.

[154] Dirk Koschützki, Henning Schwöbbermeyer, and Falk Schreiber. Ranking of network elements based on functional substructures. *Journal of Theoretical Biology*, 248(3):471–479, 2007.

[155] Daniel V Guebel, Ulf Schmitz, Olaf Wolkenhauer, and Julio Vera. Analysis of cell adhesion during early stages of colon cancer based on an extended multi-valued logic approach. *Molecular BioSystems*, 8(4):1230–1242, 2012.

[156] Eberhard O Voit. *The inner workings of life: Vignettes in systems biology*. Cambridge University Press, 2016.

[157] Brian P Ingalls. *Mathematical modeling in systems biology: an introduction*. MIT press, 2013.

[158] Olaf Wolkenhauer and Mihajlo Mesarović. Feedback dynamics and cell function: Why systems biology is called Systems Biology. *Molecular BioSystems*, 1(1):14–16, 2005.

[159] Boris P Bezruchko and Dmitry A Smirnov. *Extracting knowledge from time series: An introduction to nonlinear empirical modeling.* Springer Science & Business Media, 2010.

[160] Antje Chang, Maurice Scheer, Andreas Grote, Ida Schomburg, and Dietmar Schomburg. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Research*, 37(suppl_1):D588–D592, 2008.

[161] Peter Bloomingdale, Van Anh Nguyen, Jin Niu, and Donald E Mager. Boolean network modeling in systems pharmacology. *Journal of Pharmacokinetics and Pharmacodynamics*, pages 1–22, 2018.

[162] Jorge GT Zañudo, Steven N Steinway, and Réka Albert. Discrete dynamic network modeling of oncogenic signaling: mechanistic insights for personalized treatment of cancer. *Current Opinion in Systems Biology*, 2018.

[163] Wassim Abou-Jaoudé, Pauline Traynard, Pedro T Monteiro, Julio Saez-Rodriguez, Tomáš Helikar, Denis Thieffry, and Claudine Chaouiya. Logical modeling and dynamical analysis of cellular networks. *Frontiers in Genetics*, 7:94, 2016.

[164] Stuart A Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of theoretical biology*, 22(3):437–467, 1969.

[165] Steven Watterson, Stephen Marshall, and Peter Ghazal. Logic models of pathway biology. *Drug Discovery Today*, 13(9):447–456, 2008.

[166] Julio Saez-Rodriguez, Leonidas G Alexopoulos, Jonathan Epperlein, Regina Samaga, Douglas A Lauffenburger, Steffen Klamt, and Peter K Sorger. Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Molecular Systems Biology*, 5(1), 2009.

[167] Adrien Fauré, Aurélien Naldi, Claudine Chaouiya, and Denis Thieffry. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics*, 22(14):e124–e131, 2006.

[168] Maria I Davidich and Stefan Bornholdt. Boolean network model predicts cell cycle sequence of fission yeast. *PloS One*, 3(2):e1672, 2008.

[169] José CM Mombach, Cristhian A Bugs, and Claudine Chaouiya. Modelling the onset of senescence at the G1/S cell cycle checkpoint. *BMC Genomics*, 15(7):S7, 2014.

[170] Mariana Esther Martinez-Sanchez, Luis Mendoza, Carlos Villarreal, and Elena R Alvarez-Buylla. A minimal regulatory network of extrinsic and intrinsic factors recovers observed patterns of CD4+ T cell differentiation and plasticity. *PLoS Comput Biol*, 11(6):e1004324, 2015.

[171] Luis Mendoza and Ioannis Xenarios. A method for the generation of standardized qualitative dynamical systems of regulatory networks. *Theoretical Biology and Medical Modelling*, 3(1):13, 2006.

[172] Laleh Kazemzadeh, Marija Cvijovic, and Dina Petranovic. Boolean model of yeast apoptosis as a tool to study yeast and human apoptotic regulations. *Frontiers in Physiology*, 3:446, 2012.

[173] Laurence Calzone, Laurent Tournier, Simon Fourquet, Denis Thieffry, Boris Zhivotovsky, Emmanuel Barillot, and Andrei Zinovyev. Mathematical modelling of cell-fate decision in response to death receptor engagement. *PLoS Comput Biol*, 6(3):e1000702, 2010.

[174] David PA Cohen, Loredana Martignetti, Sylvie Robine, Emmanuel Barillot, Andrei Zinovyev, and Laurence Calzone. Mathematical modelling of molecular pathways enabling tumour cell invasion and migration. *PLoS Comput Biol*, 11(11):e1004571, 2015.

[175] Ritwik Layek, Aniruddha Datta, Michael Bittner, and Edward R Dougherty. Cancer therapy design based on pathway logic. *Bioinformatics*, 27(4):548–555, 2011.

[176] Mahesh Kumar Padwal, Uddipan Sarma, and Bhaskar Saha. Comprehensive logic based analyses of Toll-like receptor 4 signal transduction pathway. *PloS One*, 9(4): e92481, 2014.

[177] Akshata R Udyavar, David J Wooten, Megan Hoeksema, Mukesh Bansal, Andrea Califano, Lourdes Estrada, Santiago Schnell, Jonathan M Irish, Pierre P Massion, and Vito Quaranta. Novel hybrid phenotype revealed in small cell lung cancer by a transcription factor network model that can explain tumor heterogeneity. *Cancer Research*, 77(5):1063–1074, 2017.

[178] M Leire Ruiz-Cerdá, Itziar Irurzun-Arana, Ignacio González-Garcia, Chuanpu Hu, Honghui Zhou, An Vermeulen, Iñaki F Trocóniz, and José David Gómez-Mantilla. Towards patient stratification and treatment in the autoimmune disease lupus erythematosus using a systems pharmacology approach. *European Journal of Pharmaceutical Sciences*, 94:46–58, 2016.

[179] Camille Terfve, Thomas Cokelaer, David Henriques, Aidan MacNamara, Emanuel Goncalves, Melody K Morris, Martijn van Iersel, Douglas A Lauffenburger, and

Julio Saez-Rodriguez. CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms. *BMC Systems Biology*, 6(1): 133, 2012.

[180] Julio Saez-Rodriguez, Arthur Goldsipe, Jeremy Muhlich, Leonidas G Alexopoulos, Bjorn Millard, Douglas A Lauffenburger, and Peter K Sorger. Flexible informatics for linking experimental data to mathematical models via DataRail. *Bioinformatics*, 24(6):840–847, 2008.

[181] Assieh Saadatpour, István Albert, and Réka Albert. Attractor analysis of asynchronous Boolean models of signal transduction networks. *Journal of Theoretical Biology*, 266(4):641–656, 2010.

[182] Sui Huang and Donald E Ingber. Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks. *Experimental Cell Research*, 261(1):91–103, 2000.

[183] Reka Albert and Juilee Thakar. Boolean modeling: a logic-based dynamic approach for understanding signaling and regulatory networks and for making useful predictions. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 6(5):353–369, 2014.

[184] Martin Hopfensitz, Christoph Müssel, Markus Maucher, and Hans A Kestler. Attractors in Boolean networks: a tutorial. *Computational Statistics*, 28(1):19–36, 2013.

[185] Stuart A Kauffman. *The origins of order: Self-organization and selection in evolution*. Oxford University Press, USA, 1993.

[186] René Thomas. On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. *Springer Series in Synergetics*, 9:180–193, 1981.

[187] Minsoo Choi, Jue Shi, Sung Hoon Jung, Xi Chen, and Kwang-Hyun Cho. Attractor landscape analysis reveals feedback loops in the p53 network that control the cellular response to DNA damage. *Sci. Signal.*, 5(251):ra83–ra83, 2012.

[188] Steffen Klamt, Julio Saez-Rodriguez, Jonathan A Lindquist, Luca Simeoni, and Ernst D Gilles. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics*, 7(1):56, 2006.

[189] Oyebode J Oyeyemi, Oluwafemi Davies, David L Robertson, and Jean-Marc Schwartz. A logical model of HIV-1 interactions with the T-cell activation signalling pathway. *Bioinformatics*, 31(7):1075–1083, 2014.

[190] Douglas Hanahan and Robert A Weinberg. The hallmarks of cancer. *Cell*, 100(1): 57–70, 2000.

[191] Ju-Seog Lee, Sun-Hee Leem, Sang-Yeop Lee, Sang-Cheol Kim, Eun-Sung Park, Sang-Bae Kim, Seon-Kyu Kim, Yong-June Kim, Wun-Jae Kim, and In-Sun Chu. Expression signature of E2F1 and its associated genes predict superficial to invasive progression of bladder tumors. *Journal of Clinical Oncology*, 28(16):2660–2667, 2010.

[192] Sarah Heerboth, Genevieve Housman, Meghan Leary, Mckenna Longacre, Shannon Byler, Karolina Lapinska, Amber Willbanks, and Sibaji Sarkar. EMT and tumor metastasis. *Clinical and Translational Medicine*, 4(1):6, 2015.

[193] David M Gonzalez and Damian Medici. Signaling mechanisms of the epithelial-mesenchymal transition. *Sci. Signal.*, 7(344):re8–re8, 2014.

[194] Yifan Wang and Binhua P Zhou. Epithelial-mesenchymal transition-a hallmark of breast cancer metastasis. *Cancer Hallmarks*, 1(1):38–49, 2013.

[195] Sebastian Mirschel, Katrin Steinmetz, Michael Rempel, Martin Ginkel, and Ernst Dieter Gilles. PROMOT: modular modeling for systems biology. *Bioinformatics*, 25(5):687–689, 2009.

[196] Geert Berx, Karl-Friedrich Becker, Heinz Höfler, and Frans Van Roy. Mutations of the human E-cadherin (CDH1) gene. *Human Mutation*, 12(4):226–237, 1998.

[197] Darren C Tomlinson, Euan W Baxter, Paul M Loadman, Mark A Hull, and Margaret A Knowles. FGFR1-induced epithelial to mesenchymal transition through MAPK/PLCγ/COX-2-mediated mechanisms. *PloS One*, 7(6):e38972, 2012.

[198] Min Hye Jang, Eun Joo Kim, Yoomi Choi, Hee Eun Lee, Yu Jung Kim, Jee Hyun Kim, Eunyoung Kang, Sung-Won Kim, In Ah Kim, and So Yeon Park. FGFR1 is amplified during the progression of in situ to invasive breast carcinoma. *Breast Cancer Research*, 14(4):R115, 2012.

[199] Steffen Klamt, Julio Saez-Rodriguez, and Ernst D Gilles. Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Systems Biology*, 1(1):2, 2007.

[200] Christine Desmedt, Fanny Piette, Sherene Loi, Yixin Wang, Françoise Lallemand, Benjamin Haibe-Kains, Giuseppe Viale, Mauro Delorenzi, Yi Zhang, Mahasti Saghatchian d'Assignies, et al. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clinical Cancer Research*, 13(11):3207–3214, 2007.

[201] Ankur Sharma, Wen-Shuz Yeow, Adam Ertel, Ilsa Coleman, Nigel Clegg, Chellappagounder Thangavel, Colm Morrissey, Xiaotun Zhang, Clay ES Comstock,

Agnieszka K Witkiewicz, et al. The retinoblastoma tumor suppressor controls androgen signaling and human prostate cancer progression. *The Journal of Clinical Investigation*, 120(12):4478–4492, 2010.

[202] Damian Medici, Elizabeth D Hay, and Daniel A Goodenough. Cooperation between snail and LEF-1 transcription factors is essential for TGF-β1-induced epithelial-mesenchymal transition. *Molecular Biology of the Cell*, 17(4):1871–1879, 2006.

[203] Wasia Rizwani, Courtney Schaal, Sateesh Kunigal, Domenico Coppola, and Srikumar Chellappan. Mammalian lysine histone demethylase KDM2A regulates E2F1-mediated gene transcription in breast cancer cells. *PLoS One*, 9(7):e100888, 2014.

[204] Archana Dhasarathy, Dhiral Phadke, Deepak Mav, Ruchir R Shah, and Paul A Wade. The transcription factors Snail and Slug activate the transforming growth factor-beta signaling pathway in breast cancer. *PloS One*, 6(10):e26514, 2011.

[205] Rohit Bhargava, William L Gerald, Allan R Li, Qiulu Pan, Priti Lal, Marc Ladanyi, and Beiyun Chen. EGFR gene amplification in breast cancer: correlation with epidermal growth factor receptor mRNA and protein expression and HER-2 status and absence of EGFR-activating mutations. *Modern Pathology*, 18(8):1027, 2005.

[206] Ronojoy Ghosh and Claire Tomlin. Lateral inhibition through delta-notch signaling: A piecewise affine hybrid model. *Hybrid Systems: Computation and Control*, pages 232–246, 2001.

[207] Julio Vera, Eva Balsa-Canto, Peter Wellstead, Julio R Banga, and Olaf Wolkenhauer. Power-law models of signal transduction pathways. *Cellular Signalling*, 19(7):1531–1541, 2007.

[208] Olaf Wolkenhauer, Charles Auffray, Simone Baltrusch, Nils Blüthgen, Helen Byrne, Marta Cascante, Andrea Ciliberto, Trevor Dale, Dirk Drasdo, David Fell, et al. Systems biologists seek fuller integration of systems biology approaches in new cancer research programs, 2010.

[209] Hidde De Jong. Modeling and simulation of genetic regulatory systems: a literature review. *Journal of Computational Biology*, 9(1):67–103, 2002.

[210] Michael A Savageau. Biochemical systems analysis: I. Some mathematical properties of the rate law for the component enzymatic reactions. *Journal of Theoretical Biology*, 25(3):365–369, 1969.

[211] Julio Vera and Olaf Wolkenhauer. Mathematical tools in cancer signalling systems biology. In *Cancer Systems Biology, Bioinformatics and Medicine*, pages 185–212. Springer, 2011.

[212] Melody K Morris, Julio Saez-Rodriguez, Peter K Sorger, and Douglas A Lauffenburger. Logic-based models for the analysis of cell signaling networks. *Biochemistry*, 49(15):3216–3224, 2010.

[213] Tomáš Helikar, Naomi Kochi, John Konvalina, and Jim A Rogers. Boolean modeling of biochemical networks. *The Open Bioinformatics Journal*, 5:16–25, 2011.

[214] Chih-Lung Ko, Eberhard O Voit, and Feng-Sheng Wang. Estimating parameters for generalized mass action models with connectivity information. *BMC Bioinformatics*, 10(1):140, 2009.

[215] Leon Glass and Stuart A Kauffman. The logical analysis of continuous, non-linear biochemical control networks. *Journal of Theoretical Biology*, 39(1):103–129, 1973.

[216] François Jacob and Jacques Monod. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3(3):318–356, 1961.

[217] Amit Singh, Juliana M Nascimento, Silke Kowar, Hauke Busch, and Melanie Boerries. Boolean approach to signalling pathway modelling in HGF-induced keratinocyte migration. *Bioinformatics*, 28(18):i495–i501, 2012.

[218] Jonathan R Karr, Jayodita C Sanghvi, Derek N Macklin, Miriam V Gutschow, Jared M Jacobs, Benjamin Bolival, Nacyra Assad-Garcia, John I Glass, and Markus W Covert. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2):389–401, 2012.

[219] Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5):646–674, 2011.

[220] Rajeev Alur, Calin Belta, Franjo Ivančić, Vijay Kumar, Max Mintz, George J Pappas, Harvey Rubin, and Jonathan Schug. Hybrid modeling and simulation of biomolecular networks. In *International workshop on hybrid systems: Computation and control*, pages 19–32. Springer, 2001.

[221] Jasmin Fisher and Thomas A Henzinger. Executable cell biology. *Nature Biotechnology*, 25(11):1239–1249, 2007.

[222] Emanuel Gonçalves, Joachim Bucher, Anke Ryll, Jens Niklas, Klaus Mauch, Steffen Klamt, Miguel Rocha, and Julio Saez-Rodriguez. Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models. *Molecular BioSystems*, 9(7):1576–1583, 2013.

[223] A Ryll, J Bucher, A Bonin, Sophia Bongard, E Gonçalves, Julio Saez-Rodriguez, J Niklas, and Steffen Klamt. A model integration approach linking signalling and gene-regulatory logic with kinetic metabolic models. *Biosystems*, 124:26–38, 2014.

[224] Adelinde M Uhrmacher, Daniela Degenring, and Bernard Zeigler. Discrete event multi-level models for systems biology. In *Transactions on computational systems biology I*, pages 66–89. Springer, 2005.

[225] Katarzyna A Rejniak and Alexander RA Anderson. Hybrid models of tumor growth. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 3(1): 115–125, 2011.

[226] Edda Klipp and Wolfram Liebermeister. Mathematical modeling of intracellular signaling pathways. *BMC Neuroscience*, 7(1):S10, 2006.

[227] H Frederik Nijhout, Janet A Best, and Michael C Reed. Using mathematical models to understand metabolism, genes, and disease. *BMC Biology*, 13(1):79, 2015.

[228] Shigeyuki Magi, Kazunari Iwamoto, and Mariko Okada-Hatakeyama. Current Status of Mathematical Modeling of Cancer–From the Viewpoint of Cancer Hallmarks. *Current Opinion in Systems Biology*, 2017.

[229] Edda Klipp, Wolfram Liebermeister, Christoph Wierling, Axel Kowald, and Ralf Herwig. *Systems biology: a textbook*. John Wiley & Sons, 2016.

[230] Michael A Savageau. Biochemical systems analysis: II. The steady-state solutions for an n-pool system using a power-law approximation. *Journal of Theoretical Biology*, 25(3):370–379, 1969.

[231] Michael A Savageau. Development of fractal kinetic theory for enzyme-catalysed reactions and implications for the design of biochemical pathways. *Biosystems*, 47 (1):9–36, 1998.

[232] Eberhard O Voit. *Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists*. Cambridge University Press, 2000.

[233] Shmoolik Mangan and Uri Alon. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences*, 100(21):11980–11985, 2003.

[234] Shai Kaplan, Anat Bren, Erez Dekel, and Uri Alon. The incoherent feed-forward loop can generate non-monotonic input functions for genes. *Molecular Systems Biology*, 4(1):203, 2008.

[235] Bree B Aldridge, George Haller, Peter K Sorger, and Douglas A Lauffenburger. Direct Lyapunov exponent analysis enables parametric study of transient signalling governing cell behaviour. *IEE Proceedings-Systems Biology*, 153(6):425–432, 2006.

[236] Thorsten Stiewe and Brigitte M Pützer. Role of the p53-homologue p73 in E2F1-induced apoptosis. *Nature Genetics*, 26(4):464, 2000.

[237] David Engelmann and Brigitte M Pützer. Translating DNA damage into cancer cell death-A roadmap for E2F1 apoptotic signalling and opportunities for new drug combinations to overcome chemoresistance. *Drug Resistance Updates*, 13(4):119–131, 2010.

[238] Julio Vera and Olaf Wolkenhauer. A system biology approach to understand functional activity of cell communication systems. *Methods in Cell Biology*, 90: 399–415, 2008.

[239] Julio Vera, Oliver Rath, Eva Balsa-Canto, Julio R Banga, Walter Kolch, and Olaf Wolkenhauer. Investigating dynamics of inhibitory and feedback loops in ERK signalling using power-law models. *Molecular BioSystems*, 6(11):2174–2191, 2010.

[240] Julio Vera, Xin Lai, Ulf Schmitz, and Olaf Wolkenhauer. MicroRNA-regulated networks: the perfect storm for classical molecular biology, the ideal scenario for systems biology. In *MicroRNA Cancer Regulation*, pages 55–76. Springer, 2013.

[241] James DeGregori. The genetics of the E2F family of transcription factors: shared functions and unique roles. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, 1602(2):131–150, 2002.

[242] Heiko Müller, Adrian P Bracken, Richard Vernell, M Cristina Moroni, Fred Christians, Emanuela Grassilli, Elena Prosperini, Elena Vigo, Jonathan D Oliner, and Kristian Helin. E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. *Genes & Development*, 15(3):267–285, 2001.

[243] Shirley Polager and Doron Ginsberg. p53 and E2F: partners in life and death. *Nature Reviews Cancer*, 9(10):738, 2009.

[244] David Engelmann, Susanne Knoll, Daniel Ewerth, Marc Steder, Anja Stoll, and Brigitte M Pützer. Functional interplay between E2F1 and chemotherapeutic drugs defines immediate E2F1 target genes crucial for cancer cell death. *Cellular and Molecular Life Sciences*, 67(6):931–948, 2010.

[245] Brigitte M Pützer and David Engelmann. E2F1 apoptosis counterattacked: evil strikes back. *Trends in Molecular Medicine*, 19(2):89–98, 2013.

[246] Maciej Swat, Alexander Kel, and Hanspeter Herzel. Bifurcation analysis of the regulatory modules of the mammalian G1/S transition. *Bioinformatics*, 20(10): 1506–1511, 2004.

[247] Guang Yao, Tae Jun Lee, Seiichi Mori, Joseph R Nevins, and Lingchong You. A bistable Rb–E2F switch underlies the restriction point. *Nature Cell Biology*, 10(4): 476–482, 2008.

[248] Gheorghe Craciun, Baltazar Aguda, and Avner Friedman. Mathematical analysis of a modular network coordinating the cell cycle and apoptosis. *Math Biosci Eng*, 2:473–485, 2005.

[249] Baltazar D Aguda and Christopher K Algar. A structural analysis of the qualitative networks regulating the cell cycle and apoptosis. *Cell Cycle*, 2(6):538–543, 2003.

[250] Xiao-Peng Zhang, Feng Liu, and Wei Wang. Coordination between cell cycle progression and cell fate decision by the p53 and E2F1 pathways in response to DNA damage. *Journal of Biological Chemistry*, 285(41):31571–31580, 2010.

[251] Baltazar D Aguda, Yangjin Kim, Melissa G Piper-Hunter, Avner Friedman, and Clay B Marsh. MicroRNA regulation of a cancer network: consequences of the feedback loops involving miR-17-92, E2F, and Myc. *Proceedings of the National Academy of Sciences*, 105(50):19678–19683, 2008.

[252] Vijay Alla, David Engelmann, Annett Niemetz, Jens Pahnke, Anke Schmidt, Manfred Kunz, Stephan Emmrich, Marc Steder, Dirk Koczan, and Brigitte M Pützer. E2F1 in melanoma progression and metastasis. *Journal of the National Cancer Institute*, 102(2):127–133, 2010.

[253] Sanchita Mukherjee, Jianbo Dong, Carrie Heincelman, Melanie Lenhart, Angela Welford, and Angela Wandinger-Ness. Functional analyses and interaction of the XAPC7 proteasome subunit with Rab7. *Methods in Enzymology*, 403:650–663, 2005.

[254] AM Honegger, A Schmidt, A Ullrich, and J Schlessinger. Separate endocytic pathways of kinase-defective and-active EGF receptor mutants expressed in same cells. *The Journal of Cell Biology*, 110(5):1541–1548, 1990.

[255] Alex Von Kriegsheim, Daniela Baiocchi, Marc Birtwistle, David Sumpton, Willy Bienvenut, Nicholas Morrice, Kayo Yamada, Angus Lamond, Gabriella Kalna, Richard Orton, et al. Cell fate decisions are specified by the dynamic ERK interactome. *Nature Cell Biology*, 11(12):1458–1464, 2009.

[256] Altaf A Dar, Shahana Majid, David de Semir, Mehdi Nosrati, Vladimir Bezrookove, and Mohammed Kashani-Sabet. miRNA-205 suppresses melanoma cell proliferation and induces senescence via regulation of E2F1 protein. *Journal of Biological Chemistry*, 286(19):16606–16614, 2011.

[257] Lioudmila V Sharova, Alexei A Sharov, Timur Nedorezov, Yulan Piao, Nabeebi Shaik, and Minoru SH Ko. Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Research*, 16(1):45–58, 2008.

[258] G Chiorino and M Lupi. Variability in the timing of G 1/S transition. *Mathematical Biosciences*, 177:85–101, 2002.

[259] Fabio Martelli and David M Livingston. Regulation of endogenous E2F1 stability by the retinoblastoma family proteins. *Proceedings of the National Academy of Sciences*, 96(6):2858–2863, 1999.

[260] Meredith Irwin, Maria Carmen Marin, Andrew C Phillips, Ratnam S Seelan, David I Smith, Wanguo Liu, Elsa R Flores, Kenneth Y Tsai, Tyler Jacks, Karen H Vousden, et al. Role for the p53 homologue p73 in E2F-1-induced apoptosis. *Nature*, 407(6804):645–648, 2000.

[261] Paolo Gandellini, Marco Folini, Nicole Longoni, Marzia Pennati, Mara Binda, Maurizio Colecchia, Roberto Salvioni, Rosanna Supino, Roberta Moretti, Patrizia Limonta, et al. miR-205 exerts tumor-suppressive functions in human prostate through down-regulation of protein kinase Cε. *Cancer Research*, 69(6):2287–2295, 2009.

[262] Takahito Nakagawa, Masato Takahashi, Toshinori Ozaki, Ken-ichi Watanabe, Satoru Todo, Hiroyuki Mizuguchi, Takao Hayakawa, and Akira Nakagawara. Autoinhibitory regulation of p73 by ΔNp73 to modulate cell survival and death through a p73-specific target element within the ΔNp73 promoter. *Molecular and Cellular Biology*, 22(8):2575–2585, 2002.

[263] MP Deyoung and LW Ellisen. p63 and p73 in human cancer: defining the network. *Oncogene*, 26(36):5169, 2007.

[264] Yuriko Katoh and Masaru Katoh. Hedgehog signaling, epithelial-to-mesenchymal transition and miRNA. *International Journal of Molecular Medicine*, 22(3):271–275, 2008.

[265] J Vera, S Nikolov, X Lai, A Singh, and O Wolkenhauer. Model-based investigation of the transcriptional activity of p53 and its feedback loop regulation via 14-3-3σ. *IET Systems Biology*, 5(5):293–307, 2011.

[266] Xin Lai, Ulf Schmitz, Shailendra K Gupta, Animesh Bhattacharya, Manfred Kunz, Olaf Wolkenhauer, and Julio Vera. Computational analysis of target hub gene repression regulated by multiple and cooperative miRNAs. *Nucleic Acids Research*, 40(18):8818–8834, 2012.

[267] I-Chun Chou and Eberhard O Voit. Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Mathematical Biosciences*, 219(2):57–83, 2009.

[268] Federica Eduati, Javier De Las Rivas, Barbara Di Camillo, Gianna Toffolo, and Julio Saez-Rodriguez. Integrating literature-constrained and data-driven inference of signalling networks. *Bioinformatics*, 28(18):2311–2317, 2012.

[269] Katja Rateitschak and Olaf Wolkenhauer. Thresholds in transient dynamics of signal transduction pathways. *Journal of Theoretical Biology*, 264(2):334–346, 2010.

[270] Nuno Tenazinha and Susana Vinga. A survey on methods for modeling and analyzing integrated biological networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 8(4):943–958, 2011.

[271] Guy Karlebach and Ron Shamir. Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, 9(10):770, 2008.

[272] Livia Perfetto, Leonardo Briganti, Alberto Calderone, Andrea Cerquone Perpetuini, Marta Iannuccelli, Francesca Langone, Luana Licata, Milica Marinkovic, Anna Mattioni, Theodora Pavlidou, et al. SIGNOR: a database of causal relationships between biological entities. *Nucleic Acids Research*, 44(D1):D548–D554, 2015.

[273] Benjamin M Gyori, John A Bachman, Kartik Subramanian, Jeremy L Muhlich, Lucian Galescu, and Peter K Sorger. From word models to executable models of signaling networks using automated assembly. *Molecular Systems Biology*, 13(11): 954, 2017.

[274] Jean Clairambault, Stéphane Gaubert, and Thomas Lepoutre. Circadian rhythm and cell population growth. *Mathematical and Computer Modelling*, 53(7):1558–1567, 2011.

[275] Thomas Sütterlin, Simone Huber, Hartmut Dickhaus, and Niels Grabe. Modeling multi-cellular behavior in epidermal tissue homeostasis via finite state machines in multi-agent systems. *Bioinformatics*, 25(16):2057–2063, 2009.

[276] Laura Rosanò, Roberta Cianfrocca, Francesca Spinella, Valeriana Di Castro, Maria Rita Nicotra, Alessandro Lucidi, Gabriella Ferrandina, Pier Giorgio Natali, and Anna Bagnato. Acquisition of chemoresistance and EMT phenotype is linked with activation of the endothelin A receptor pathway in ovarian carcinoma cells. *Clinical Cancer Research*, 17(8):2350–2360, 2011.

[277] Payman Samavarchi-Tehrani, Azadeh Golipour, Laurent David, Hoon-ki Sung, Tobias A Beyer, Alessandro Datti, Knut Woltjen, Andras Nagy, and Jeffrey L Wrana. Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell*, 7(1): 64–77, 2010.

[278] Erik D Wiklund, Jesper B Bramsen, Toby Hulf, Lars Dyrskjøt, Ramshanker Ramanathan, Thomas B Hansen, Sune B Villadsen, Shan Gao, Marie S Ostenfeld, Michael Borre, et al. Coordinated epigenetic repression of the miR-200 family and miR-205 in invasive bladder cancer. *International Journal of Cancer*, 128(6): 1327–1334, 2011.

[279] Albert Sorribas, Benito Hernández-Bermejo, Ester Vilaprinyo, and Rui Alves. Cooperativity and saturation in biochemical networks: a saturable formalism using Taylor series approximations. *Biotechnology and Bioengineering*, 97(5):1259–1277, 2007.

[280] Gerry Melino, Francesca Bernassola, Marco Ranalli, Karen Yee, Wei Xing Zong, Marco Corazzari, Richard A Knight, Doug R Green, Craig Thompson, and Karen H Vousden. p73 Induces apoptosis via PUMA transactivation and Bax mitochondrial translocation. *Journal of Biological Chemistry*, 279(9):8076–8083, 2004.

[281] Antonio Costanzo, Paola Merlo, Natalia Pediconi, Marcella Fulco, Vittorio Sartorelli, Philip A Cole, Giulia Fontemaggi, Maurizio Fanciulli, Louis Schiltz, Giovanni Blandino, et al. DNA damage-dependent acetylation of p73 dictates the selective activation of apoptotic target genes. *Molecular Cell*, 9(1):175–186, 2002.

[282] Natalia Pediconi, Alessandra Ianari, Antonio Costanzo, Laura Belloni, Rita Gallo, Letizia Cimino, Antonio Porcellini, Isabella Screpanti, Clara Balsano, Edoardo Alesse, et al. Differential regulation of E2F1 apoptotic target genes in response to DNA damage. *Nature Cell Biology*, 5(6):552, 2003.

[283] Hailong Wu, Shoumin Zhu, and Yin-Yuan Mo. Suppression of cell growth and invasion by miR-205 in breast cancer. *Cell Research*, 19(4):439, 2009.

[284] Shoumin Zhu, Abbes Belkhiri, and Wael El-Rifai. DARPP-32 increases interactions between epidermal growth factor receptor and ERBB3 to promote tumor resistance to gefitinib. *Gastroenterology*, 141(5):1738–1748, 2011.

[285] Christine Blattner, Alison Sparks, and David Lane. Transcription factor E2F-1 is upregulated in response to DNA damage in a manner analogous to that of p53. *Molecular and Cellular Biology*, 19(5):3704–3713, 1999.

[286] Eva Balsa-Canto, Antonio A Alonso, and Julio R Banga. An iterative identification procedure for dynamic modeling of biochemical networks. *BMC Systems Biology*, 4(1):11, 2010.

[287] Weg M Ongkeko, Xiao Qi Wang, Wai Yi Siu, Anita WS Lau, Katsumi Yamashita, Adrian L Harris, Lynne S Cox, and Randy YC Poon. MDM2 and MDMX bind and stabilize the p53-related protein p73. *Current Biology*, 9(15):829–832, 1999.

[288] Iqbal Dulloo, Ganesan Gopalan, Gerry Melino, and Kanaga Sabapathy. The antiapoptotic DeltaNp73 is degraded in a c-Jun–dependent manner upon genotoxic stress through the antizyme-mediated pathway. *Proceedings of the National Academy of Sciences*, 107(11):4902–4907, 2010.

[289] Zoya S Kai and Amy E Pasquinelli. MicroRNA assassins: factors that regulate the disappearance of miRNAs. *Nature Structural & Molecular Biology*, 17(1):5–10, 2010.

[290] Meiguo Xin and Xingming Deng. Nicotine inactivation of the proapoptotic function of Bax through phosphorylation. *Journal of Biological Chemistry*, 280(11):10781–10789, 2005.

[291] Cristina Sanz, Adalberto Benito, Naohiro Inohara, Daryoush Ekhterae, Gabriel Nunez, and Jose Luis Fernandez-Luna. Specific and rapid induction of the proapoptotic protein Hrk after growth factor withdrawal in hematopoietic progenitor cells. *Blood*, 95(9):2742–2747, 2000.

[292] Michael T Beck, Susan K Peirce, and Wen Y Chen. Regulation of bcl-2gene expression in human breast cancer cells by prolactin and its antagonist, hPRL-G129R. *Oncogene*, 21(33):5047–5047, 2002.

[293] Malgorzata Magdalena Sak, Kamilla Breen, Sissel Beate Rønning, Nina Marie Pedersen, Vibeke Bertelsen, Espen Stang, and Inger Helene Madshus. The oncoprotein ErbB3 is endocytosed in the absence of added ligand in a clathrin-dependent manner. *Carcinogenesis*, 33(5):1031–1039, 2012.

[294] Andrey Frolov, Kyle Schuller, Ching-Wei D Tzeng, Emily E Cannon, Brandon C Ku, J Harrison Howard, Selwyn M Vickers, Martin J Heslin, Donald J Buchsbaum, and J Pablo Arnoletti. ErbB3 expression and dimerization with EGFR influence pancreatic cancer cell sensitivity to erlotinib. *Cancer Biology & Therapy*, 6(4): 548–554, 2007.

Appendix A

Topological properties of the E2F1 map

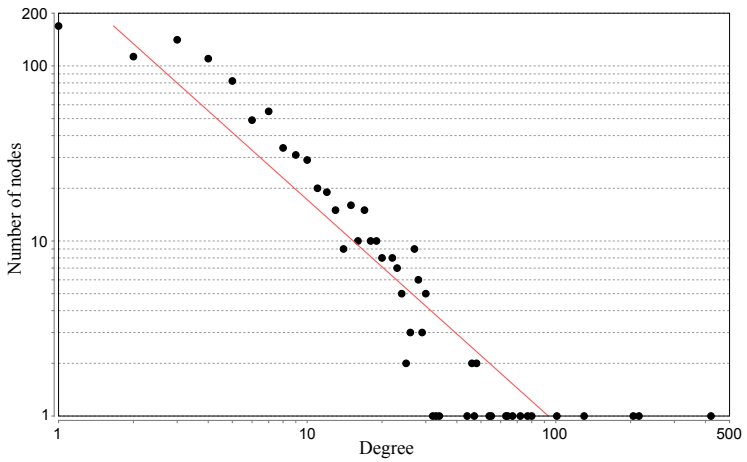


Figure A.1. Node degree distribution of E2F1 interaction map. The red line indicates that node degree distribution follows a power law, which indicates a scale-free topology of the network. The figure is adopted from Khan et al., 2017 [124].

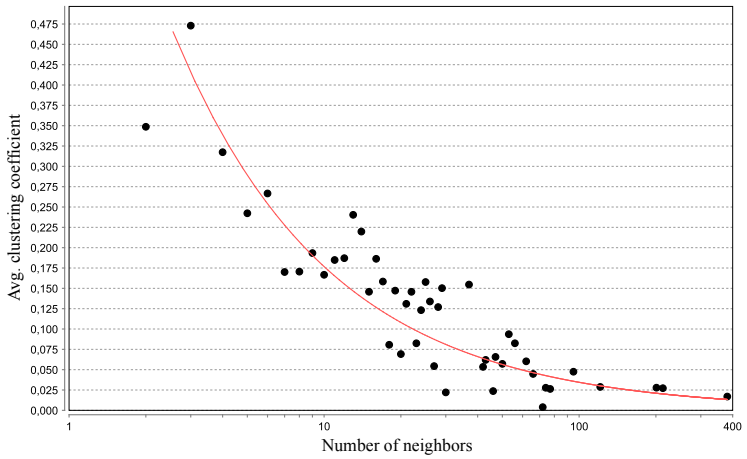


Figure A.2. Clustering coefficient distribution of E2F1 interaction map.

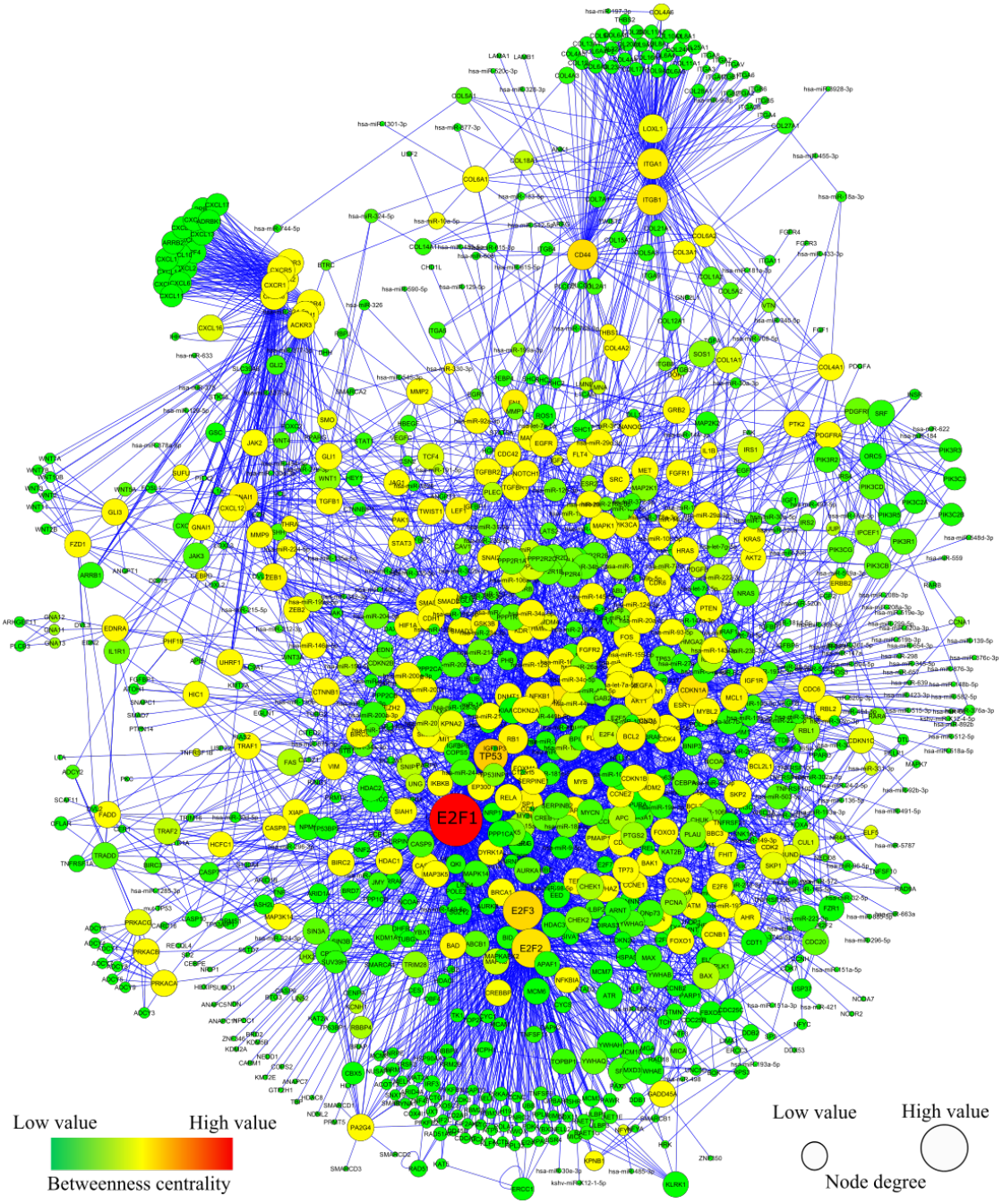


Figure A.3. Cytoscape view of E2F1 interaction map. The size of the nodes represents the value of the node degree. The node color ranges from green (low betweenness centrality) to red (high betweenness centrality). The figure is adopted from Khan et al., 2017 [124].

Appendix B

Algorithm for multi-objective function

Below is the MATLAB code for multi-objection optimization function used to calculate scores for network motifs.

Matlab code for motifs ranking

```
%Read average node degree (ND), avereage Betweenness Centrality (BC), average Disease Pathway (DP), average Gene Prioritization (GP) and absolute average expression Fold Change (FG) of each feedback loop from Excel file
ND = xlsread('Path to Excel file\Excel file.xls',Sheet_number,'A2:A445');% read average Node Degree (ND)
BC = xlsread('Path to Excel file\Excel file.xls',Sheet_number,'B2:B445');% read average Betweenness Centrality (BC)
DP = xlsread('Path to Excel file\Excel file.xls',Sheet_number,'C2:C445');% read average Disease Pathway (DP)
GP = xlsread('Path to Excel file\Excel file.xls',Sheet_number,'D2:D445');% read average Gene Prioritization (GP)
FC = xlsread('Path to Excel file\Excel file.xls',Sheet_number,'E2:E445');% read absolute average expression Fold Change (FG) for breast cancer
ND_max = max(ND);
BC_max = max(BC);
DP_max = max(DP);
GP_max = max(GP);
FC_max = max(FC);
clear F_obj;

% assigning different weights to ND, BC, DP, GP and FC which represent 13 scenarios for multi - objective function
w = [1 0 0 0;0 0 1 0;0 0 0 1;0 0 1/4 3/4;0 1/4 0 3/4;1/4 0 0 3/4;1/16 1/16 1/8 3/4;1/16 1/8 1/16 3/4;1/8 1/16 1/16 3/4;1/4 1/4 1/4 1/4;0 1/8 1/8 3/4;1/8 1/8 0 3/4;1/8 0 1/8 3/4];

for i = 1:length(w)  % For loop for 13 weighting scenario
    for j = 1:length(ND) % For loop for all 444 feedback loops
F_obj(j,i) = w(i,1) /2 * ND(j,1)/ ND_max + w(i,1) /2 * BC(j,1)/ BC_max + w(i,2) * DP(j,1)/ DP_max + w(i,3) * GP(j,1)/ GP_max + w(i,4) * FC(j,1)/ FC_max; %Multi-objective function
    end
```

```
21 end
22 % Writing ranking score of all 444 feedback loops for 13 scenarios
23 SUCCESS = XLSWRITE('Path where to save the ranking score for feedback loops\
       FBLs_ranking_score_in_xyy_type_cancer.xls',F_obj)
```

Appendix C

Asynchronous states update of logic-based models

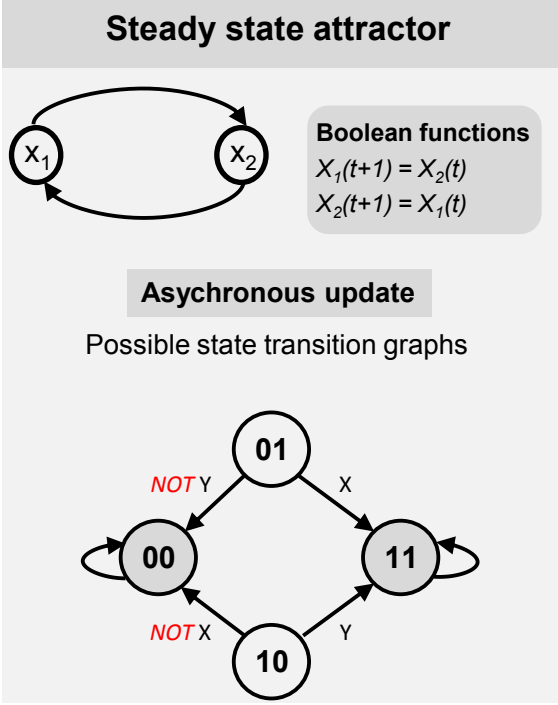


Figure C.1. Asynchronous update of the positive feedback loops. Network shown in the upper left corner is a positive feedback loop, where x_1 activates x_2 and x_2 activates x_1 which is represented by Boolean functions. State transition graph represents the asynchronous state update of the systems where the binary digits from left represent the state of nodes x_1 and x_2, respectively. In asynchronous update, starting from initial condition, a state can have more than one successor which means only one node changes in a state transition form x(t) to x(t+1), therefore, it is possible for a state to leave the cycle. In the state transition graph the gray circles represent steady states 00 and 11. The states 01 and 10 are unstable state which can be followed by either 00 or 11 depending on weather x or y changes first.

Appendix D

Logic-based rules for bladder cancer regulatory core

The reactions are represented in CellNetAnalyzer (CNA) format: $(Coeff.)\ Reactants = (Coeff.)\ Products$, where 'Coeff.' stands for coefficient which indicates the level of expreesion of reactants and products.

Logic-based rules for bladder cancer regulatory core simulations

(Coeff.) Reactants = (Coeff.) Products	References (Pubmed IDs)
1 !nfkb1 = 1 e2f1	19088195
1 !e2f1 + akt1 = 1 nfkb1	9368006; 10485711; Calibrated the equation with bladder cancer cell lines RT-4 (less-invasive) to UM-UC-3 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)
1 e2f1 = 1 axin2	15572025, 15766563
1 e2f1 = 1 zeb1	25466554
1 cdh1 = 1 egfr	10969083
1 egfr = 1 src	21303975;23066441
1 !src + 1 !snai2 + 1 !mir_25_3p + 1 !mdm2 = 1 cdh1	23906494;22450326,25189528; Calibrated the equation with bladder cancer cell lines RT-4 (less-invasive) to UM-UC-3 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 5485619)
1 fgfr1 = 1 ppp2r2c	17255109
1 fgfr1 = 1 ppp2r1a	17255109
1 fgfr1 = 1 ppp2r1b	17255109
1 fgfr1 = 1 ppp2r2b	17255109
1 !kiaa1524 = 1 ppp2r2b	17632056
1 !e2f1 = 1 mdm2	20812030
1 !mdm2 = 1 tp53	18485870
1 akt1 = 1 mdm2	15140942;15574337
1 !sirt1 = 1 tp53	16354677
1 !bcl2 = 1 tp53	12670866
1 !mir_25_3p = 1 tp53	20935678
1 tp53 = 1 e2f1	24076372
1 tp53 = 1 mir_205_5p	22578566
1 tp53 = 1 sfn	17546054;21625211
1 e2f2 = 1 myc	11564866
1 !rara = 1 foxa1	25303530
1 e2f1 = 1 cdkn2a	15175242;11034215
1 e2f2 = 1 cdkn2a	11883935
1 sp1 = 1 cdkn2a	21555589
1 tp53 = 1 mdm2	19106616
1 !rb1 = 1 tp53	17080083
1 foxo3 = 1 tp53	25241761
1 !sfn = 1 e2f1	20196847
1 !mir_205_5p = 1 e2f1	21454583;22578566
1 myc = 1 e2f1	18345030
1 !ppp2r2b = 1 e2f1	25483052; 24204027
1 !ppp2r1b = 1 e2f1	25483052; 24204027
1 !ppp2r1a = 1 e2f1	25483052; 24204027
1 !ppp2r2c = 1 e2f1	25483052; 24204027
1 e2f3 = 1 e2f1	10766737;14576826
1 foxa1 = 1 e2f1	21900400
1 !cdkn2a = 1 e2f1	21478909
1 !tp53 = 1 cdkn2a	25241761;22480684;20676136
1 !cdkn2a = 1 mdm2	17909018
1 e2f1 = 1 tp53	11739724
1 sp1 = 1 tp53	16740634
1 mdm2 = 1 e2f1	16170383
1 ctnnb1 + 1 lef1 = 1 snai2	25189528;14623871
1 ctnnb1 + 1 lef1 = 1 snai1	25189528;14623871
1 ctnnb1 = 1 twist1	25189528
1 !snai1 = 1 cdh1	21317430; Calibrated the equation with bladder cancer cell lines RT-4 (non-invasive) to UM-UC-3 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)
1 !twist1 = 1 cdh1	22581441;20519943; Calibrated the equation with bladder cancer cell lines RT-4 (non-invasive) to UM-UC-3 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)
1 ctnnb1 = 1 lef1	17980157
1 lef1 = 1 axin2	11809808
1 !cdh1 + 1 !axin2 = 1 ctnnb1	22007144;22935447
1 smad4 + 1 smad3 + 1 smad2 = 1 snai1	20519943

1 smad3 = 1 mir_200b_3p	22020340
1 !mir_200b_3p = 1 zeb1	21049046;18829540;18376396;19854497;19839049;18411277
1 !nfkbia + 1 akt1 = 1 chuk	9346485;10485710
1 chuk = 1 nfkb1	22833419
1 nfkb1 = 1 nfkbia	14743216
1 e2f1 = 1 rb1	15362224
1 e2f1 = 1 foxo3	17482685
1 e2f1 = 1 fgfr1	12543798
1 e2f1 = 1 kiaa1524	12766778
1 e2f1 = 1 sp1	8657142
1 e2f1 = 1 e2f2	15014447
1 !e2f1 = 1 rara	24608861
1 e2f1 = 1 mir_25_3p	19034270
1 e2f1 = 1 akt1	15574337
1 e2f1 = 1 sirt1	19188449
1 e2f1 = 1 bcl2	11559537;11704823
1 !rb1 = 1 e2f1	19401190
1 !rb1 = 1 e2f3	19249677
1 snai1 = 1 zeb1	18411277
1 lef1 = 1 e2f1	17980157
1 fgfr1 = 1 src	20154139
1 !mdm2 = 1 snai2	19448627
1 !mir_200b_3p = 1 e2f3	22144583;22139708
1 nfkb1 = 1 twist1	17332324
1 !zeb1 = 1 cdkn2a	24371144
1 lef1 = 1 myc	17466981
1 !mir_200b_3p = 1 bcl2	21993663
1 tp53 = 1 mir_200b_3p	21518799
1 smad4 + 1 smad3 + 1 smad2 = 1 sp1	15242331;16714330;10878024
1 lef1 + 1 e2f3 = 1 axin2	25189528;Calibrated the equation with bladder cancer cell lines RT-4 (non-invasive) to UM-UC-3 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)
1 !twist1 = 1 tp53	18504427
1 ctnnb1 = 1 myc	9727977
1 !tp53 = 1 snai2	16207734
1 !mir_205_5p = 1 zeb1	19502803;18376396
1 e2f2 = 1 axin2	15572025, 15766563
1 !twist1 = 1 cdkn2a	25368021
1 !smad4 + 1 !smad3 + 1 !smad2 = 1 myc	19022773;12150994
1 nfkb1 = 1 bcl2	12170775
1 nfkb1 = 1 myc	11114727
1 !chuk = 1 foxo3	16887827
1 tgfbr + 1 zeb1 = 1 smad2	25189528; 21593157;21717360
1 tgfbr + 1 zeb1 = 1 smad3	25189528; 21593157;21717360
1 tgfbr + 1 zeb1 = 1 smad4	25189528; 21593157;21717360
1 tgfbr = 1 snai1	25189528
1 tgfbr = 1 snai2	25189528
1 e2f1 + 1 cxcr1 = 1 zeb1	25466554;http://www.fedoa.unina.it/id/eprint/9883; Calibrated the equation with bladder cancer cell lines RT-4 (non-invasive) to UM-UC-3 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)
1 cxcr1 + 1 smad4 + 1 smad3 + 1 smad2 = 1 snai1	25189528;20519943; Calibrated the equation with bladder cancer cell lines RT-4 (non-invasive) to UM-UC-3 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)

Multi-valued logic based rules for EMT phenotype

1 !cdh1 = 1 emt	23461975;24556840
1 fgfr1 = 1 emt	22701738
(1 smad4 + 1 smad3 + 1 smad2 + snai1) OR (1 zeb1 + 1 twist1) = 1 emt	22945800; 20713713;21317430; Calibrated the equation with bladder cancer cell lines RT-4 (non-invasive) to UM-UC-3 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)
(1 smad4 + 1 smad3 + 1 smad2 + snai1) OR (1 zeb1 + 1 twist1)+ 1 !cdh1 = 2 emt	Calibrated the equation with bladder cancer cell lines RT-4 (non-invasive) to UM-UC-3 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)

1 fgfr1 + 1 !cdh1 = 2 emt	Calibrated the equation with bladder cancer cell lines RT-4 (non-invasive) to UM-UC-3 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)
1 fgfr1 + (1 smad4 + 1 smad3 + 1 smad2 + snai1) OR (1 zeb1 + 1 twist1) = 2 emt	Calibrated the equation with bladder cancer cell lines RT-4 (non-invasive) to UM-UC-3 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)
1 fgfr1 + (1 smad4 + 1 smad3 + 1 smad2 + snai1) OR (1 zeb1 + 1 twist1) + 1 !cdh1 = 3 emt	Calibrated the equation with bladder cancer cell lines RT-4 (non-invasive) to UM-UC-3 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)

Appendix E

Logic-based rules for breast cancer regulatory core

The reactions are represented in CellNetAnalyzer (CNA) format: $(Coeff.)\ Reactants = (Coeff.)\ Products$, where 'Coeff.' stands for coefficient which indicates the level of expreesion of reactants and products.

Logic-based rules for breast cancer regulatory core simulations

(Coeff.) Reactants = (Coeff.) Products	References (Pubmed IDs)
1 !akt1 = 1 gsk3b	10385618, 12459251
1 !axin2 + 1 ctnnb1 = 1 gsk3b	22935447
1 !bcl2 = 1 tp53	12670866
1 !birc2 = 1 birc3	11084335;22274400;19153467
1 !birc3 = 1 e2f1	15674331
1 !cdh1 + 1 !axin2 = 1 ctnnb1	15662113;22935447
1 !cdkn2a = 1 e2f1	21478909
1 !cdkn2a = 1 mdm2	17909018
1 !e2f1 = 1 mdm2	20812030
1 !e2f1 = 1 rara	24608861
1 !gsk3b = 1 e2f1	18367454
1 lef1 = 1 e2f1	17980157
1 !mdm2 = 1 myc	10541552
1 !mdm2 = 1 snai2	19448627
1 !mdm2 = 1 tp53	18485870
1 !mir_17_5p = 1 ncoa3	16940181
1 !mir_205_5p = 1 e2f1	21454583;22578566
1 !mir_25_3p = 1 tp53	20935678
1 !rara = 1 foxa1	25303530
1 !rb1 = 1 e2f1	19401190
1 !rb1 = 1 e2f3	19249677
1 !rb1 = 1 tp53	17080083
1 !sirt1 = 1 tp53	16354677
1 !src + 1 !snai1 = 1 cdh1	21317430; Calibrated the equation with breast cancer cell lines MCF-7 (non-invasive) to MDA-MB231 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)
1 !src + 1 !snai2 = 1 cdh1	21317430; Calibrated the equation with breast cancer cell lines MCF-7 (non-invasive) to MDA-MB231 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)
1 !tp53 = 1 cdkn2a	25241761;22480684;20676136
1 !tp53 = 1 snai2	16207734
1 !twist1 + 1 !mdm2 + 1 !mir_25_3p = 1 cdh1	22581441;20519943; Calibrated the equation with breast cancer cell lines MCF-7 (non-invasive) to MDA-MB231 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)
1 !twist1 = 1 cdkn2a	25368021
1 !twist1 = 1 tp53	18504427
1 akt1 = 1 mdm2	15140942;15574337
1 birc2 = 1 e2f1	21653699
1 cdh1 = 1 egfr	10969083
1 ctnnb1 + 1 !gsk3b = 1 snai1	25189528;14623871
1 ctnnb1 + 1 lef1 = 1 twist1	25189528
1 ctnnb1 = 1 lef1	17980157
1 ctnnb1 = 1 myc	9727977
1 e2f1 + 1 hmmr = 1 fn1	Calibrated the equation with breast cancer cell lines MCF-7 (non-invasive) to MDA-MB231 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619); 25042645
1 e2f1 + 1 snai1 = 1 fn1	Calibrated the equation with breast cancer cell lines MCF-7 (non-invasive) to MDA-MB231 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619); 25042645; 24360956
1 e2f1 = 1 akt1	15574337
1 e2f1 = 1 axin2	15572025, 15766563
1 e2f1 = 1 bcl2	11559537;11704823
1 e2f1 = 1 cdkn2a	15175242;11034215
1 e2f1 = 1 e2f2	15014447
1 e2f1 = 1 fn1	25042645
1 e2f1 = 1 foxo3	17482685
1 e2f1 = 1 mir_17_5p	19034270;19066217
1 e2f1 = 1 mir_25_3p	19034270
1 e2f1 = 1 rb1	15362224
1 e2f1 = 1 sirt1	19188449
1 e2f1 = 1 sp1	8657142
1 e2f1 = 1 tp53	11739724

1 e2f2 = 1 cdkn2a	11883935
1 e2f2 = 1 myc	11564866;90001421
1 e2f3 = 1 axin2	15572025, 15766563
1 e2f3 = 1 e2f1	10766737;14576826
1 egfr = 1 src	21303975;23066441
1 flt4 = 1 e2f1	24014887
1 fn1 = 1 flt4	
1 fn1 = 1 kpna2	21988832
1 foxa1 = 1 e2f1	21900400
1 foxo3 = 1 tp53	25241761
1 gsk3b = 1 axin2	12023307;23602568
1 gsk3b = 1 ctnnb1	22935447
1 il1r1 + 1 !birc3 = 1 traf1	9384571
1 kpna2 = 1 e2f1	22843992; 25109899
1 lef1 = 1 axin2	11809808
1 lef1 = 1 myc	17466981
1 mdm2 = 1 axin2	
1 mdm2 = 1 e2f1	16170383
1 myc = 1 cdkn2a	11494151
1 myc = 1 e2f1	18345030
1 ncoa3 = 1 e2f1	15169882
1 rara = 1 ncoa3	25303530
1 sp1 = 1 cdkn2a	21555589
1 sp1 = 1 tp53	16740634
1 ctnnb1 + 1 tgfbr = 1 snai1	25189528;14623871;Calibrated the equation with breast cancer cell lines MCF-7 (non-invasive) to MDA-MB231 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)
1 ctnnb1 + 1 tgfbr = 1 snai2	25189528;14623871;Calibrated the equation with breast cancer cell lines MCF-7 (non-invasive) to MDA-MB231 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)
1 tgfbr = 1 snai1	25189528
1 tgfbr = 1 snai2	25189528
1 thrb = 1 myc	
1 tp53 = 1 e2f1	24076372
1 tp53 = 1 mdm2	17875722;9724636
1 tp53 = 1 mir_205_5p	21518799
1 traf1 = 1 birc2	
1 vegfc = 1 flt4	24014887
Multi-valued logic based rules for EMT phenotype	
1 fn1 = 1 emt	25042645;12730682
1 !cdh1 = 1 emt	23461975;24556840
1 src + 1 snai1 + 1 snai2 = 1 emt	22945800;23913825;20713713;21317430
1 src + 1 fn1 + 1 snai1 + 1 snai2 = 2 emt	Calibrated the equation with breast cancer cell lines MCF-7 (non-invasive) to MDA-MB231 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)
1 src + 1 !cdh1 + 1 snai1 + 1 snai2 = 2 emt	Calibrated the equation with breast cancer cell lines MCF-7 (non-invasive) to MDA-MB231 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)
1 !cdh1 + 1 fn1 = 2 emt	Calibrated the equation with breast cancer cell lines MCF-7 (non-invasive) to MDA-MB231 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)
1 src + 1 !cdh1 + 1 fn1 + 1 snai1 + 1 snai2 = 3 emt	Calibrated the equation with breast cancer cell lines MCF-7 (non-invasive) to MDA-MB231 (invasive) from ArrayExpress database (accession no.:E-MTAB-2706, PMID: 25485619)

Appendix F

Construction and simulation of the drug resistance model

F.1. Model construction of E2F1 network in drug resistance

We set up and analyzed a kinetic model using ordinary differential equations of the regulatory map (shown in Figure 4.6) representing the network response to different anticancer drugs. The kinetic model is composed of twelve ordinary differential equations accounting for the evolution in time of: E2F1 mRNA (in the model represented with the variable mE2F1) and protein (E2F1), both isoforms of p73 (respectively, p73 and DNp73), miR-205 (miR205), E2F1 regulated pro-apoptotic genes (represented by Harakiri, Hrk), p73 regulated pro-apoptotic genes (Bax), miR-205 repressed anti-apoptotic genes (BCL2), E2F1 regulated and active EGFR (EGFR*), cytostatic drug inhibited EGFR (EGFR I), and ERBB3 (ERBB3). In addition, we considered another differential equation that accounts for the population size of tumor cells with the genetic background defined by our model (TC). The kinetic model is organized in four interconnected modules: (i) The core regulatory module, (ii & iii) two transcriptional target modules and; (iv) one for tumor cell population. A detailed description of the model structure and ODE equations is:

(i) The core regulatory module: The core regulatory module of the kinetic model is composed of E2F1, both isoforms of p73 (respectively, p73 and DNp73) and miR-205 and accounts for their temporal evolution under drug administration [119]. We include here the transcriptional regulation of miR-205 expression by the known oncogene TGFβ-1 (TGFB1), which exemplifies the set of regulators of miR-205 expression external to the core module. The inclusion of this variable allows for the investigation of cross-talk, via miR-205, between E2F1 signaling and other oncogenic signals involved in the epithelial-mesenchymal transition (EMT). EMT has been associated with the emergence

of chemoresistance [276].

$$\frac{dmE2F1}{dt} = GxD \cdot k_1 + FS \cdot k_2 - k_3 \cdot (1 + k_4 \cdot miR205) \cdot mE2F1 \tag{F.1}$$

$$\frac{dE2F1}{dt} = k_5 \cdot mE2F1 - k_6 \cdot E2F1 \tag{F.2}$$

$$\frac{dp73}{dt} = k_7 \cdot E2F1 - k_8 \cdot p73 \tag{F.3}$$

$$\frac{dDNp73}{dt} = k_9 \cdot E2F1 - k_{10} \cdot DNp73 \tag{F.4}$$

$$\frac{dmiR205}{dt} = k_{11} \cdot TGFB1^{-1} \cdot p73 \cdot \left(1 + \frac{DNp73}{k_{13}}\right)^{-2} - k_{12} \cdot miR205 \tag{F.5}$$

In case of E2F1 mRNA, Eq. F.1, the model contains mass-action kinetic rates accounting for its basal and genotoxic-stress mediated synthesis (characterized respectively by the parameters k_2 and FS, and k_1 ; GxD is an input variable accounting for genotoxic drug), as well as for its basal degradation (k_3). In addition, the model includes a term accounting for the miR-205 mediated E2F1 repression (k_4, [256]). We notice that in our qualitative kinetic model, mE2F1 represents the fraction of transcriptionally active messenger RNA in a similar manner that our previous publications [266].

In case of E2F1 protein expression, Eq. F.2, the model describes its mRNA-mediated synthesis (k_5) and basal degradation (k_3). In case of p73 and DNp73, Eq. F.3 and Eq. F.4, the model includes the E2F1-mediated synthesis (k_7 and k_9, respectively [236]) and basal degradation (k_8 and k_{10}) for both isoforms.

In case of miR-205, Eq. F.5, the model describes the regulation of its synthesis and degradation (k_{11}, k_{12}). In this equation, we assumed that p73 acts as activator of miR-205 synthesis, while DNp73 plays an inhibitory role [119, 261, 262]. According to the observations reported for several tumor types, we assumed that when both isoforms are overexpressed the inhibitory effect of DNp73 dominates and cancels out p73 activation of miR-205 [263]. Besides, we included the repression of miR-205 by the oncogenic signal from the TGFβ-1 pathway (TGFB1). This is to exemplify the set of regulators of miR-205 expression, which are external to the E2F1-p73/DNp73-miR-205 network [264, 277, 278]. To represent these features, we used reduced power-law terms [207, 279].

(ii) Apoptosis related module: This module includes pro- and antiapoptotic transcriptional targets, which are regulated by the network components in response to genotoxic stress. Precisely in this module the ODEs account: (i) for representing the dynamics of proapoptotic proteins whose expression is promoted by E2F1, we included a variable accounting for Harakiri (Hrk) expression [244]; (ii) similarly, for proapoptotic proteins whose expression is positively regulated by p73 and negatively regulated by DNp73, a variable accounting for Bax (Bax) expression is defined [280]; and finally (iii) for antiapoptotic proteins whose expression is repressed by miR-205, we chose BCL-2 ($BCL2$) as a representative and defined a corresponding variable [119]. These pro- and antiapoptotic proteins are used here as an example and represent a wider set of proteins which undergo similar regulation. This kind of simplification has been successfully applied to reduce the complexity of models of biochemical networks,

for example, in the study conducted by Aguda and colleagues [251].

$$\frac{dBax}{dt} = k_{14} \cdot DS \cdot \frac{p73^g}{k_{16}^g + p73^g} \cdot \left(1 + \frac{DNp73}{k_{13}}\right)^{-2} - k_{15} \cdot Bax \tag{F.6}$$

$$\frac{dHrk}{dt} = k_{17} \cdot DS \cdot \frac{E2F1^g}{k_{16}^g + E2F1^g} - k_{18} \cdot Hrk \tag{F.7}$$

$$\frac{dBCL2}{dt} = k_{19} \cdot \left(1 + \frac{miR205}{k_{20}}\right)^{-1} - k_{21} \cdot BCL2 \tag{F.8}$$

In order to model Bax protein expression, Eq. F.6, we introduced a rate equation that accounts for Bax synthesis and degradation (k_{14} and k_{15} respectively). For Bax synthesis and based on experimental results [281], we assumed that DNA damage signals trigger acetylation and activation of p73 which is required to promote Bax expression. Drug induced DNA damage signals are here encoded in the binary variable DS (genotoxic drug treatment: $DS = 1$; no treatment $DS = 0$). P73 mediated synthesis of Bax is represented by a Hill equation (k_{16}, g). This is to ensure that relevant levels of pro-apoptotic genes (here Bax) occur only in presence of sufficient p73 levels. We have used this strategy previously to model the transcriptional regulation of other DNA-damage dependent genes [265]. To account for the inhibitory effect of DNp73 on the expression of Bax, we adopted the same modeling strategy as for miR-205.

In case of Harakiri (Hrk) Eq. F.7, the model contains rates equations accounting for its synthesis and degradation (k_{17} and k_{18} respectively). Based on experimental results [282], we assumed that DNA damage signals are required for the acetylation and activation of E2F1, which triggers the Hrk expression. Therefore we made the synthesis dependent on a binary variable DS which accounts for genotoxic drug treatment ($DS = 1$) or no treatment ($DS = 0$). Like in the case of p73-regulated pro-apoptotic genes, a Hill equation describes the E2F1-regulated synthesis of Hrk.

In case of BCL-2 ($BCL2$), Eq. F.8, the model contains rates equations accounting for the synthesis and degradation (k_{19} and k_{21} respectively). We have previously shown that BCL-2 expression is post-transcriptionally regulated by miR-205 [119]. This is represented in our model by a reduced power-law term [207, 279] (k_{20}).

(iii) Proliferative related target module: Furthermore, the model includes a module for receptors whose activity is regulated in response to cytostatic drugs. This module includes variables accounting for the dynamics of EGFR, whose expression is promoted by E2F1 [252]; and ERBB3, which is repressed by miR-205 [261, 283]. In several tumors the heterodimerization of these receptors has been linked to abnormal cell proliferation and changes in the sensitivity of tumor cells to cytostatic drugs [284]. Additional input variables account for the effect of genotoxic (GxD) and cytostatic (CyD) drugs triggering DNA damage-E2F1-mediated apoptosis and abolishment of tumor cell proliferation, respectively [285].

The module contains ODEs for the *de novo* EGFR (EGFR), growth factor-activated EGFR ($EGFR^*$), cytostatic drug inhibited EGFR ($EGFR^i$), and miR-205 repressed

ERBB3 ($ERBB3$). The following three equations describe the EGFR dynamics:

$$\frac{dEGFR}{dt} = k_{36} \cdot \frac{E2F1^{g_2}}{k_{35}^{g_2} + E2F1^{g_2}} - k_{28} \cdot GF \cdot EGFR - k_{30} \cdot CyD \cdot EGFR - k_{40} \cdot EGFR$$
$$\frac{dEGFR^*}{dt} = k_{28} \cdot GF \cdot EGFR - k_{29} \cdot EGFR^*$$
$$\frac{dEGFR^i}{dt} = k_{30} \cdot CyD \cdot EGFR \cdot \left(1 + \frac{ERBB3}{ER^{TH}}\right)^{-1} - k_{31} \cdot EGFR^i$$

We defined a variable accounting for the total amount of EGFR, $EGFR^{Total}$:

$$EGFR^{Total} = EGFR + EGFR^* + EGFR^i$$

Thus, we can estimate the amount of de novo inactive EGFR as:

$$EGFR = EGFR^{Total} - EGFR^* - EGFR^i$$

According to previous results from our collaborators [117], we assume that the total amount of EGFR ($EGFR^T$) is regulated by E2F1. We modeled this with dependency with a Hill equation and therefore we can change the previous equation to:

$$EGFR = k_{36} \cdot \frac{E2F1^{g_2}}{k_{35}^{g_2} + E2F1^{g_2}} - EGFR^* - EGFR^i$$

When we apply this to the previous equations, the model reduces to:

$$\frac{dEGFR^*}{dt} = k_{28} \cdot GF \cdot \left(k_{36} \cdot \frac{E2F1^{g_2}}{k_{35}^{g_2} + E2F1^{g_2}} - EGFR^* - EGFR^i\right) - k_{29} \cdot EGFR^* \quad \text{(F.9)}$$
$$\frac{dEGFR^i}{dt} = k_{30} \cdot CyD \cdot \left(k_{36} \cdot \frac{E2F1^{g_2}}{k_{35}^{g_2} + E2F1^{g_2}} - EGFR^* - EGFR^i\right) - k_{31} \cdot EGFR^i \quad \text{(F.10)}$$
$$\frac{dERBB3}{dt} = k_{33} \cdot \left(1 + \frac{miR205}{k_{33}}\right)^{-1} - k_{34} \cdot ERBB3 \quad \text{(F.11)}$$

In case of $EGFR^*$, Eq. F.9, the model contains equations accounting for its growth factor (GF) mediated activation and its degradation (k_{28} and k_{29}, respectively). We notice that GF here is a binary input variable (stimulation with growth factor: $GF = 1$; no stimulation: $GF = 0$). In case of $EGFR^i$, Eq. F.10, the model contains an equation that accounts its cytostatic drug (CyD) mediated inhibition and its degradation (k_{30} and k_{31}, respectively). Given that interactions between EGFR and ERBB3 promote tumor resistance to certain cytostatic drugs [284], we included a reduced power-law term accounting for the effect of ERBB3 as inhibitor of cytostatic drugs regulating EGFR activity (ER^{TH}). For ERBB3, Eq. F.11, the model contains equations accounting for its synthesis and degradation (k_{32} and k_{34}, respectively). According to previous experimental results [261, 283], we assumed that ERBB3 expression is post-transcriptionally regulated by miR-205 and therefore used a reduced power-law term [207, 279] to account for the miR-205-mediated inhibition of ERBB3 (k_{33}).

(iv) Tumor cell population: Finally, the last module includes a phenomenological kinetic equation, which connects the components of the network to the dynamics of

tumor cell population. In this equation, the representatives for pro- and antiapoptotic proteins regulate the cell death rate after genotoxic stress, whereas EGFR and ERBB3 control the tumor cell proliferation rate. This module composed of a unique ordinary differential equation accounting for the size of a population of tumor cells (TC).

$$\frac{dTC^*}{dt} = k_{24} \cdot EGFR^* \cdot TC - k_{22} \cdot Hrk \cdot Bax \cdot \left(1 + \frac{BCL2}{k_{23}}\right)^{-1} \cdot TC \qquad \text{(F.12)}$$

In this equation, Eq. F.12, active EGFR connects the proliferation of the tumor cell population to proliferative signals [252] (k_{24}), while the representatives for pro- and anti-apoptotic proteins (Hrk, Bax, and BCL2, respectively) regulate the cell death rate after genotoxic stress (k_{22}, k_{23}). We notice that in some of the simulations we considered tumors to be composed of three subpopulations of tumor cells with different genetic signatures. Towards this end, we constructed three instances of the model, which differ in their genetic signature as explained below in subsection F.3.

F.2. Model parameterization of E2F1 in drug resistance

Values of the model parameters characterizing the reaction rates were assigned following a hybrid strategy [211, 286], which is composed of: (i) Extracting parameter values from published information (e.g. protein, mRNA and miRNA half-lives); (ii) Estimating a subset of parameter values by tuning them to fit published quantitative and qualitative data; (iii) fixing some parameter values to normalize variables to the basal, non-stressed levels of mRNA and protein. Finally, we assigned values within the region of feasible values to few parameters in a way our model reproduces the phenotype shown by chemosensitive tumor cells, according to published knowledge [119] . The parameters and their assigned values are listed in Table F.1. The plots in figure SM3 compare our model predictions with the available (own and published) data after estimating some of the model parameters.

F.3. Model simulations to detect genetic signatures for drug resistance

In model simulations, we considered two families of drugs: genotoxic drugs, like doxorubicin and cisplatin, inducing apoptosis; and cytostatic drugs, like erlotinib and lapatinib, repressing tumor cell proliferation. In some scenarios, it has been observed that these drugs loose efficacy when some of the network components are dysregulated [119, 294]. With the help of the kinetic model, we analyzed whether specific genetic signatures of the core module can be linked to a phenotype response of resistance to these anti-cancer drugs. We define a genetic signature as a group of genes in a tumor cell whose combined expression pattern is linked to a specific phenotype (*i.e.,* chemoresistant or chemosensitive). In line with this, we designed a nominal *in*

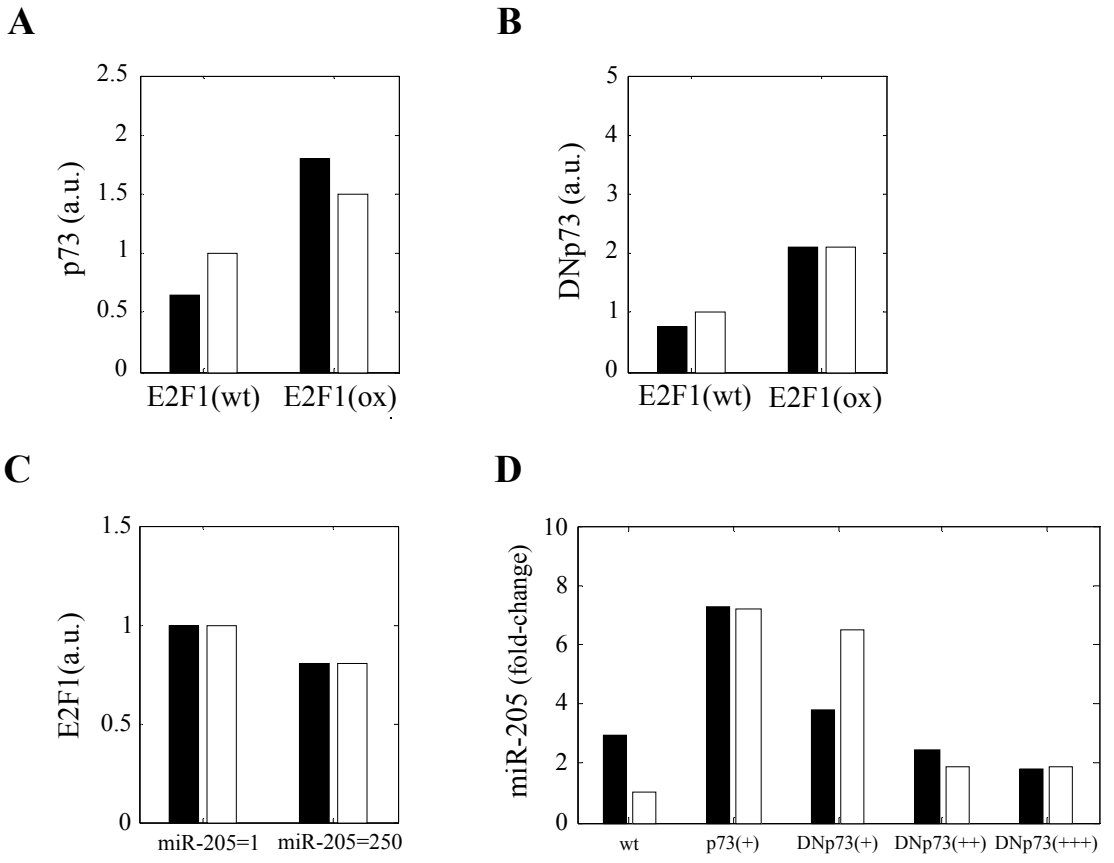


Figure F.1. Fitting of the model to the experimental data: The model parameters k_4, k_7, k_9, k_{11}, k_{13} were estimated by manual tuning to minimize the distance between model simulations (black bar) and experimental data (white bar). We show the comparison between both quantities for a chosen set of parameter values from Table F.1. Data displayed in A, B and D were produced and quantified by our collaborators and published in [117], while data in C was extracted from 1 (a.u.: arbitrary units; wt: wild-type E2F1; ox: overexpressed E2F1; p73(+): cells transfected with $1\mu g$ pcDNAp73; DNp73(+): $1\mu g$ pcDNAp73 + $0.1\mu g$ DNp73; DNp73(++): $1\mu g$ pcDNAp73 + $0.5\mu g$ DNp73; DNp73(+++): $1\mu g$ pcDNAp73 + $1\mu g$ DNp73). The figures are taken from Vera et al., [121].

silico chemosensitive genetic signature, inspired on that of SK-Mel-147 [119]. This signature accounts for the expression of the network components for which an *in silico* cell line holds the phenotype of a chemosensitive tumor cell; that is, abnormal proliferation but responsiveness to genotoxic drugs (via triggering of apoptosis) and cytostatic drugs (via inhibition of proliferation, see Figure F.2B).

To detect genetic signatures providing chemoresistance, we generated a population of 104 model configurations obtained by random perturbation of the model parameter values of the nominal chemosensitive genetic signature. In the following, we define these distinct configurations as *in silico* cell lines. For each one of them, we simulate the model and obtained the expression levels of the network components and the tumor cell population

Parameters	Description of associated process	Value	Comments
k_1	genotoxic-stress mediated synthesis of E2F1 mRNA	1.237 hr^{-1}	represents effective drug dose†
k_2	basal synthesis of E2F1 mRNA	2.085 hr^{-1}	assumed for normalization
k_3	basal degradation of E2F1 mRNA	0.139 hr^{-1}	ref. [257]
k_4	miR-205 mediated repression of E2F1	0.01 a.u.	estimated, figure F.1
k_5	mRNA-mediated synthesis of E2F1	0.231 hr^{-1}	assumed for normalization
k_6	basal degradation of E2F1	0.231 hr^{-1}	ref. [259]
k_7	E2F1 mediated synthesis of p73	0.150 hr^{-1}	estimated, figure F.1
k_8	basal degradation of p73	1.386 hr^{-1}	ref. [287]
k_9	E2F1 mediated synthesis of DNp73	0.1317 hr^{-1}	estimated, figure F.1
k_{10}	basal degradation of DNp73	0.173 hr^{-1}	ref. [288]
k_{11}	p73 mediated synthesis of miR-205	0.101 hr^{-1}	estimated, figure F.1
k_{12}	basal degradation of miR-205	0.029 hr^{-1}	ref. [289]
k_{13}	threshold effective DNp73 repression of miR-205	10.2672	estimated, figure F.1
k_{14}	p73 mediated synthesis of Bax	1 a.u.	assumed for normalization
k_{15}	degradation of Bax	0.1 hr^{-1}	ref. [290]
k_{16}	E2F2 level producing 1/2 of maximum expression level for Hrk	16.89 a.u.	estimated*
g	Hill-constant	5.61 a.u.	estimated*
k_{17}	E2F1 mediated synthesis of Hrk	0.1 a.u.	assumed for normalization
k_{18}	degradation of Hrk	0.1 hr^{-1}	ref. [291]
k_{19}	synthesis of BCL-2	0.03 a.u.	assumed for normalization
k_{20}	threshold miR-205 repression of BCL2	5.56 a.u.	assumed
k_{21}	degradation of BCL2	$\approx$ 0.1 hr^{-1}	ref. [292]
k_{22}	rate of apoptotic cells increase	1.0 a.u.	assumed
k_{23}	threshold for effective anti-apoptotic gene expression	0.028 a.u.	assumed for normalization
k_{24}	tumor size duplication time	0.0021 hr^{-1}	estimated¥ [252]
FS	factor for E2F1 expression modulation	1 a.u.	tunable
k_{28}	growth factor mediated EGFR activation	2.7726 a.u.	assumed
k_{29}	EGFR*deactivation/degradation	0.6931 hr^{-1}	ref. [254]
k_{30}	cytostastic drug mediated inhibition of EGFR activation	0.6931 a.u.	assumed
k_{31}	EGFR* deactivation/degradation	0.06931 hr^{-1}	ref. [253]
k_{32}	ERBB3 synthesis	0.2773 a.u.	assumed for normalization
k_{33}	threshold miR-205 repression of ERBB3	1 a.u.	assumed for normalization
k_{34}	ERBB3 degradation	0.2773 hr^{-1}	ref. [293]
k_{35}	E2F2 level producing 1/2 of maximum expression level for EGFR	7.0711 a.u.	estimated£
g_2	Hill-constant	1.5053	estimated£
ER^{TH}	threshold for effective ERBB3-mediated inhibition of cytostatic drug	0.1 hr^{-1}	assumed for normalization
GF	growth factor	[0,1] a.u.	tunable, binary
DS	genotoxic-drug mediated induction of pro-apoptotic genes	[0,1] a.u.	tunable, binary

Table F.1. Model definition and parameters: † In the current model parameterization, this parameter represents a drug-induced genotoxic stress level sufficient to induce high levels (60%) of apoptotic cells 48 hr after stress induction in a population of normal tumor cells, as described in the main text. * Those parameters were estimated to reproduce a sigmoid curve in the transcription of the pro-apoptotic genes, in which the normalized transcription function gets 0.05 for $E2F1 = 10$ (a.u.) and 0.9 for $E2F1 = 25$. The same applies to p73-promoted pro-apoptotic genes. ¥This value is equivalent to the duplication time for the tumor size (measured as total amount of tumor cells) of two weeks. £ Those parameters were estimated to reproduce a sigmoid curve in the transcription of the pro-apoptotic genes, in which the normalized transcription function gets 0.05 for $E2F1 = 1$ (a.u.) and 0.95 for $E2F1 = 50$.

in three cancer-relevant scenarios: (i) under nonstress conditions, at time zero, and after 120 hours; (ii) after genotoxic drug administration, at 48 hours; and (iii) after cytostatic drug administration, at 120 hours. The *in silico* cell lines were classified into the following groups: (i) cell lines that show abnormal proliferation in nonstress conditions, named as tumor cells; (ii) those resistant to genotoxic drugs; (iii) those resistant to cytostatic drugs; and (iv) those resistant to both genotoxic and cytostatic drugs. For each group, the data describing the expression levels of the network components at time zero were normalized with respect to the nominal chemosensitive values (Figure F.2C)..

Our results indicate that 30% of the *in silico* cell lines are resistant to genotoxic drugs (Figure F.2C). These cells display high basal levels (at time zero, nonstress conditions) of E2F1 and DNp73, whereas miR-205 is downregulated compared with the nominal chemosensitive tumor cell and the group of *in silico* tumor cells (Figure F.2A). In addition, approximately 1% of the cell lines display resistance to cytostatic drugs. These cells have a genetic signature composed of much higher basal E2F1, DNp73, and ERBB3 levels, whereas miR-205 appears strongly downregulated and TGFβ-1 is moderately upregulated. In both cases, EGFR is expressed at its maximum level, but our analysis indicates that this is a property already acquired by the group of tumor cells as a consequence of high levels of E2F1. Finally, a small fraction of the cell lines display a phenotype of double resistance and their genetic signature is very similar to that of cytostatic drug-resistant cells.

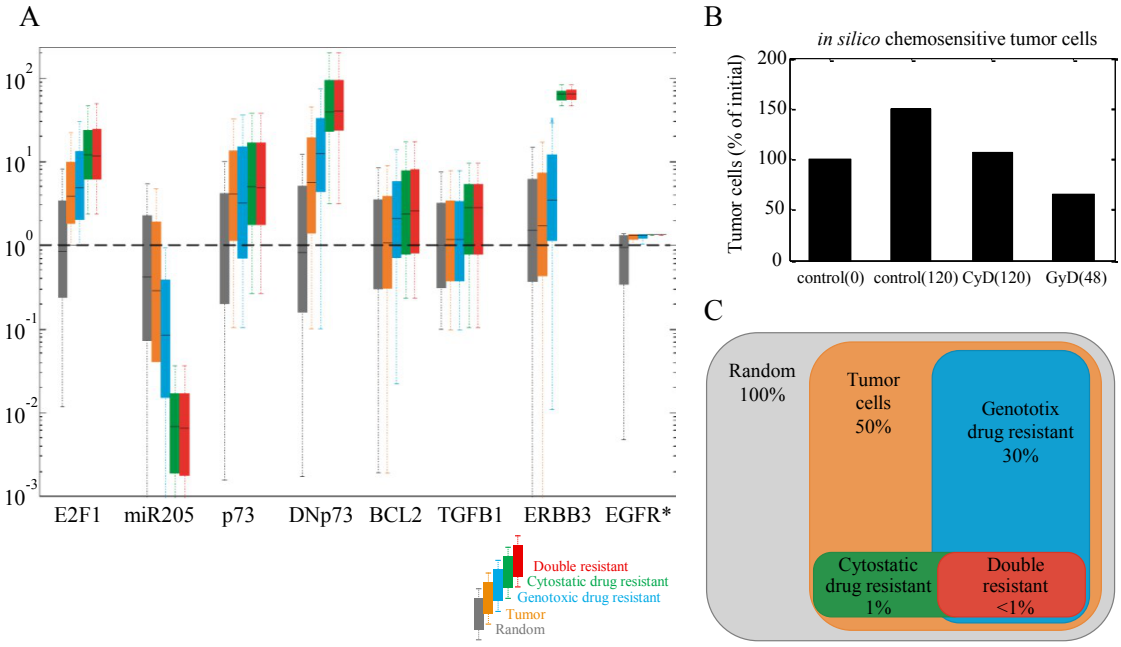


Figure F.2. (**A**) Genetic signatures of the E2F1-p73/DNp73-miR-205 network that confer drug resistance. Box-and-whisker plots of simulated non-stress levels for E2F1, p73, DNp73, miR-205, BCL2, TGFβ-1, ERBB3 and EGFR for the different subsets of *in silico* cells generated: (i) entire set of randomly generated solutions (grey); (ii) tumor cells (orange); (iii) genotoxic drug resistant (blue); (iv) cytostatic drug resistant (green); (v) double drug resistant (red). In each bar the central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles, the whiskers extend to the most extreme data points. The dashed horizontal line in the plot accounts for the values in the nominal chemosensitive genetic signature. (**B**) Phenotype of the *in silico* chemosensitive tumor cell line. Percentage of cells with respect to initial population. Legend: control(0): population size at 0hr, no drug injection; control(120): population at 120hr, no drug injection; CyD(120): 120 hr after cytostatic drug; GxD(48): 48 hr after genotoxic drug. (**C**) Distribution of the initial set of randomly generated 10^4 *in silico* cell lines in the different subpopulations. Percentage over the total is indicated. The figures are adopted from Vera et al., [121].

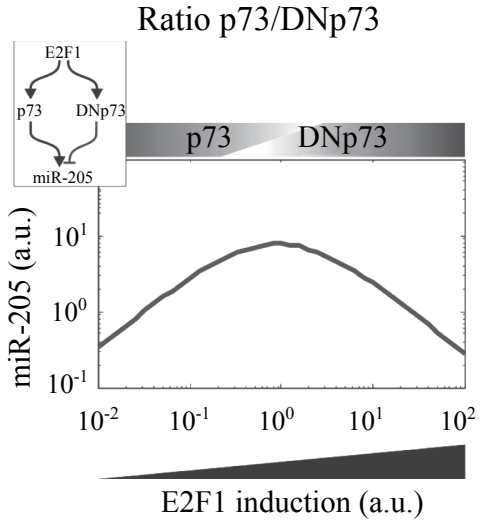


Figure F.3. Design principles underlying chemoresistance. E2F1-regulated shift of the p73/DNp73 ratio can induce non-monotonic, nonlinear regulation of miR-205 expression. The peak of miR-205 expression is reached for intermediate E2F1 levels and downregulation for very high E2F1 expression. This system is an incoherent feedforward loop and, according to our hypothesis, nonlinear regulation of the branches in the feedforward loop switches the status of the feedforward loop from one in which the positive branch predominates (*e.g.,* p73 ▲) and promotes target expression to another in which the negative branch predominates (*e.g.,* DNp73 ▲) and the target expression is repressed. The figure is taken from Vera et al., [121].

Appendix G

Time series model simulation of hybrid model

For one of the scenarios considered previously, I simulated the effect of pulse-like drug administration for different time points. The simulation time was discretized ($t = 0, 60, 120, 180, 240, 270$ h) and each target gene and phenotype were updated at those time points.

	t=0	t=60	t=120	t=180	t=240	t=270
GF	1	1	1	1	1	1
FS	1	1	1	1	1	1
TGFB1	1	1	1	1	1	1
GxS	0	0	1	1	1	0
E2F1	10.25	10.25	10.25	14.27	15.37	12.38
p73	6.66	6.66	6.66	8.97	9.98	8.55
DNp73	7.80	7.80	7.80	10.35	11.69	10.19
miR-205	4.56	4.56	4.56	4.61	4.61	4.53
Gab2	1	1	1	2	2	2
CCND1	1	1	1	2	2	2
EGFR	1	1	1	2	2	2
ANCCA	1	1	1	2	2	2
ACTR2	1	1	1	2	2	2
EZH2	1	1	1	2	2	2
BIM	0	0	0	0	0	0
FOXO3	0	0	0	1	1	0
DUSPx	0	0	0	1	1	0
HRK	0	0	0	1	1	0
NOXA	0	0	0	1	1	0
BAX	0	0	0	0	0	0
PUMA	0	0	0	0	0	0
BCL2	0	0	0	1	1	0
ABCA2	0	0	0	1	1	0
ABCA5	0	0	0	1	1	0
Proliferation	1	1	1	0	0	1
Apoptosis	0	0	0	0	0	0
Chemoresistance	0	0	0	1	1	0

Table G.1. Time series model simulation of hybrid model.

Appendix H

Hybrid model parameters

Hybrid model parameters.

Parameters	Description of associated process	Value	Comments/ref. (PubMed ID)
k_1	EGF mediated synthesis of EGFR	0.6932 hr^{-1}	11597398, 23447575
k_2	self activation of EGFR	1 hr^{-1}	assumed for normalization
k_3	E2F1 mediated synthesis of EGFR	0.2012 hr^{-1}	23447575
k_x	Hill's constant	2	estimated
k_4	basal degradation of EGFR	0.06931	23447575, 16473627
k_{xx}	basal degradation of EGFR*	1	23447575, 2335562
k_5	EGFR mediated synthesis of ERK	0.1 hr^{-1}	19935650
k_6	basal degradation of ERK	0.01 hr^{-1}	19935650
k_7	genotoxic-stress mediated synthesis of E2F1 mRNA	1.0 hr^{-1}	23447575
k_8	basal synthesis of E2F1 mRNA	2.085 hr^{-1}	23447575
k_9	basal degradation of E2F1 mRNA	0.139 hr^{-1}	19001483, 23447575
k_{10}	miR-205 mediated repression of E2F1	0.01 hr^{-1}	23447575
k_{11}	mRNA-mediated synthesis of E2F1i	0.231 hr^{-1}	23447575
k_{12}	E2F1 and Rb disassociation by pERK	1.5*0.16075 hr^{-1}	11965250
k_{13}	E2F1 and Rb association	0.235	11965250
k_{14}	basal degradation of E2F1i	0.231 a.u.	10077601, 23447575
k_{15}	E2F1 phophorilation by mitogenic signal	1.5*0.16075 hr^{-1}	11965250
k_{16}	basal degradation of E2F1p	0.235 a.u.	10077601, 23447575
k_{17}	E2F1 mediated synthesis of p73	0.1502	23447575
k_{18}	basal degradation of p73	0.231 hr^{-1}	10469568
k_{19}	E2F1 mediated synthesis of DNp73	0.1317	23447575
k_{20}	basal degradation of DNp73	0.173	23447575, 20185758
k_{21}	p73 mediated synthesis of miR-205	0.101	23447575
k_{22}	threshold effective DNp73 repression of miR-205	9.2672	23447575
k_{23}	basal degradation of miR-205	0.029*1.5	23447575, 20051982
$ERKt$	total ERK concentration	1	assumed

Table H.1. Model definition and parameters: Parameter values used in the core module of this model along with their relevant references. Some of the parameters in this model are not exactly the same as the values given in the citation but were scaled to fit to the performance of our current model.

Appendix I

Hybrid toy model code

Below is the MATLAB code for toy model simulation.

Matlab function to encode the ODEs

```
function  dy = H_model2_13Feb13(t,y)
global p73 Dp73
dy = zeros(2,1);
miR = y(1);      BCLW = y(2);
kx = 1.84;       k11 = 0.3419 / kx^2;
k12 = 0.029;     k13 = 5.58 * kx;
k25 = 0.1;       k26 = 0.1;       k27 = 0.18;
dy(1) = k11 * p73 * (1 + Dp73./k13).^ (-2) - k12 * miR; % miR205
dy(2) = k25./(1 + k27 * miR) - k26 * BCLW; % BCLW
```

Matlab Code for simulations

```
global p73 Dp73
tint = 0:0.1:50; % simulation time
p73log = -1:0.1 * 0.40:2;
Dp73log = -1:0.1 * 0.4:2;
p73vector = 10 .^ p73log; % input vector for model simulation
Dp73vector = 10 .^ Dp73log; % input vector for model simulation
count2 = 0;
numpoints = length(p73vector);
Xl = [0 1]; %initial values
for i = 1 : length(p73vector)
    coun = 0;
    for j = 1:length(Dp73vector)
     Dp73 = Dp73vector(j);
     Dp73matrix(i,j) = Dp73vector(j);
     p73 = p73vector(i);
     p73totalmatrix(i,j) = p73vector(i);
     coun = coun+1;
     Dp73_count=length(Dp73vector) - coun;
```

```
19      [t,x] = ode23(@H_model2_13Feb13,tint,Xl);
20      miRc(i,j) = x(length(tint),1);
21      BCLWc(i,j) = x(length(tint),2);
22      mir_v(i) = x(length(tint),1);
23      % Now this will discritize p73 and DNp73 for hybrid simulation.
24      if  miRc(i,j) < 5
25          miRd(i,j) = 0;
26          BCLWd(i,j) = 2;
27      elseif miRc(i,j) >= 5 && miRc(i,j) <= 20
28          miRd(i,j) = 1;
29          BCLWd(i,j) = 1;
30      else
31          miRd(i,j) = 2;
32          BCLWd(i,j) = 0;
33     end
34     count2 = count2+1;
35             iteration = numpoints^2 - count2;
36         if round(iteration./100)==iteration./100
37             clc
38             iteration
39         end
40     end
41 end
42 % drawing surface plot of ODE model
43 figure('units', 'centimeters', 'position', [2 2 7 5.5], ...
44     'paperunits', 'centimeters', 'paperposition', [2 2 7 5.5], ...
45     'paperpositionmode', 'auto');
46 axes('position', [.20 .20 .80 .70], 'fontsize', 8, 'box', 'on')
47 surf(p73totalmatrix,Dp73matrix,BCLWc,'Edgecolor','none');
48 set(gca, 'FontName', 'Times New Roman')
49 colormap(jet)
50 axis([0.1 100 0.1 100 0 1]);
51 set(gca,'XScale','log','fontsize',9.);
52 set(gca,'YScale','log','fontsize',9.);
53 view([0 0 1])
54 colorbar
55 xlabel('p73','fontsize',10,'fontweight','b');
56 ylabel('DNp73','fontsize',10,'fontweight','b');
57 title('BCL-w ODE model','fontsize',10,'fontweight','b');
58 % drawing surface plot of hybrid model
59 figure('units', 'centimeters', 'position', [2 2 7 5.5], ...
60     'paperunits', 'centimeters', 'paperposition', [2 2 7 5.5], ...
61     'paperpositionmode', 'auto');
62 axes('position', [.20 .20 .80 .70], 'fontsize', 8, 'box', 'on')
63 surf(p73totalmatrix,Dp73matrix,BCLWd,'Edgecolor','none');
64 set(gca, 'FontName', 'Times New Roman')
65 colormap(jet)
66 axis([0.1 100 0.1 100 0 2]);
67 set(gca,'XScale','log','fontsize',8.);
68 set(gca,'YScale','log','fontsize',8.);
69 view([0 0 1])
70 colorbar
71 xlabel('p73','fontsize',10,'fontweight','b');
72 ylabel('DNp73','fontsize',10,'fontweight','b');
73 title('BCL-w hybrid model','fontsize',10,'fontweight','b');
```

Appendix J

Hybrid model code

Below is the MATLAB code for hybrid model simulation.

Matlab function to encode the ODEs

```
function  dy = model_E_miR_12feb13(t,y)
global FS k1 k5 k6 k8 k9 k10 k11 k12 k13 k14 k15...
       DrugInd GF TGFB1 timedd
dy = zeros(9,1);
% Variables
EGFR = y(1); EGFRstar = y(2); pERK = y(3); mE2F1 = y(4); E2F1i = y(5);
E2F1p = y(6); p73 = y(7); DNp73 = y(8); miR = y(9);
if t >= timedd
    DrugInd = k1;
    DS = 1;
else
    DrugInd = 0;
    DS = 0;
end
kegf = 1.5;
EGFRtotal = 10;
k28 = 0.6932; k36 = 2; g2 = 1; k35 = 0.2012; k29 = 0.6931; kxx = 1; k31 = 0.06931;
dy(1) = - k28 * (kegf * GF + kxx * EGFRstar) * EGFR +(k35 * E2F1p .^g2./ (k36 .^g2 +
    E2F1p .^g2 )) - k31 * EGFR; % EGFR synthesis
dy(2) = k28 * (kegf * GF + kxx * EGFRstar) * EGFR - k29 * EGFRstar;% EGFR
    Phosphorylation
% pERK
ERKt = 1; k5 = 0.1; k6 = 0.01;
dy(3) = k5 * EGFRstar * (ERKt - pERK) - k6 * pERK;
% mE2F1
k8 = 2.085; k9 = 0.139; k10 = 0.01;
dy(4) = DrugInd + FS * k8 - k9 * (1 + k10 * miR) * mE2F1;
% E2F1i
k11 = 0.231; k12 = 0.231;
kx = 1.5*0.16075; ky=0.235; % taken from pubmedID 11965250% multiply by 0.16075
dy(5) = k11 * mE2F1 - kx * E2F1i * pERK + ky * E2F1p - k12 * E2F1i;
% E2F1p
dy(6) = kx * E2F1i * pERK - ky * E2F1p;
% p73
```

```
k13 = 0.15020; k14 = 0.2310;
dy(7) = k13 * E2F1p - k14 * p73;
% DNp73
k15 = 0.1317; k16 = 0.173;
dy(8) = k15 * E2F1p - k16 * DNp73;
% mir205
k11a = 0.101; k13a = 9.2672; k12a = 0.0435;
dy(9) = k11a * TGFB1 .^(-1) * p73 * (1 + DNp73 ./ k13a) .^(-2) - k12a * miR;
```

Matlab Code for simulations

```
global timedd FS k1 kdi kapo1 kapo2 kbaxo kbako DrugInd GF TGFB1
ti = 0;       % initial time
tf = 300;    % final time
timeint = ti:1:tf;
timedd = 200;
kdi = 15*0.139;   kapo1 = 0.028;  kapo2 = 1.0;
kbaxo = 0.0;     kbako = 0.0;
% Initial condition
  xo = [0.1 0.1  0.1    0.1  0.1  0.1   0.1  0.1  0.1];
clear aggressive TGFB1matrix FSmatrix chemoresistance
FSvector = 0.1:0.09*5:10; % Simulations for multiple values of FS
TGFB1vector = 0.1:0.09*5:10; % Simulations for multiple values of TGFB1
count2 = 0;
numpoints = length(FSvector);
for i = 1:length(FSvector)
    coun = 0;
    for j = 1:length(TGFB1vector)
     FS = FSvector(i);
     FSmatrix(i,j) = FSvector(i); % Matrix for surface plot
     TGFB1 = TGFB1vector(j);
     TGFB1matrix(i,j) = TGFB1vector(j); % Matrix for surface plot
     GF = 1;
     k1 = 1;
      coun = coun + 1;
           FS_count = length(FSvector) - coun;
    [t,sol] = ode23(@model_E_miR_12feb13,timeint,xo);
% Input matric for H-model
% e2f1p discretization
if max(sol(150:300,6)) <= 1.150
    e2f1p = 0 ;
else    if max(sol(150:300,6)) <= 12
        e2f1p = 1;
else
    e2f1p = 2;
    end
end
% p73 discretization
if max(sol(150:300,7)) > 1 && max(sol(150:250,7)) <= 10
   p73 = 1 ; % very low and high p73 is not effective in non-monotonics
else
   p73 = 0;
end
% dnp73 discretization
if max(sol(150:300,8)) <= 10
```

```
   dnp73 = 0 ;
else
   dnp73 = 1;
end
%mir205 discretization
if DrugInd == 0
 mir205 = 1; % drug administration induced high DNp73 and low mir205
else if max(sol(150:300,9)) < 2.5
   mir205 = 0;
else if p73 <= dnp73
   mir205 = 0;
    else
   mir205 = 1;
    end
    end
end
% GxS discretization
if DrugInd == 0
  gxd = 0 ; % low drug administration
else
   if DrugInd == 0.5
       gxd = 1; % medium drug administration
   else
       gxd = 2; % high drug administration
   end
end
% Commond for compuatation of logical steady state (LSS) of a model "e2f1_h_cna2" for
    initial values of [gxd e2f1p p73 dnp73 mir205]
[spec_lss] = CNAcomputeLSS (e2f1_h_cna2,[gxd e2f1p p73 dnp73 mir205]);
[e2f1_h_cna2.specID num2str(spec_lss')]; %shows LLS value of each species
proliferative(i,j) = spec_lss(23); % output of the model
apoptosis(i,j) = spec_lss(24); % output of the model
chemoresistance(i,j) = spec_lss(25); % output of the model
aggressive(i,j) = spec_lss(26); % output of the model
count2 = count2+1;
            iteration = numpoints ^2 - count2;
        if round(iteration ./ 250) == iteration ./250
            iteration
        end
    end
end
%Surface plot of Chemoresistance for different expressions of E2F1 and TGFB
figure('units', 'centimeters', 'position', [2 2 6 6], ...
'paperunits', 'centimeters', 'paperposition', [2 2 6 6], ...
'paperpositionmode', 'auto');
axes('position', [.20 .20 .70 .70], 'fontsize', 8, 'box', 'on')
surf(FSmatrix, TGFB1matrix, chemoresistance, 'Edgecolor', 'none');
colormap(jet)
set(gca,'fontname','Times New Roman')
set(gca,'XScale','log','fontsize',10.);
set(gca,'YScale','log','fontsize',10.);
view([0 0 1])
colorbar
xlabel('FS (a.u.)');
ylabel('TGFB1 (a.u.)');
title('Chemoresistance','Fontsize',10);
```

Appendix K

Exclusively logical representation of the chemoresistance model

Figure K.1 is a graphical representation of the logical model created in ProMoT [195]. The model was simulated for logical steady states (LSS) analysis using CellNetAnalyzer (CNA). LSSs were calculated for the same combination of inputs as applied to the hybrid model. Simulation results of the logical and the hybrid model are the same for scenarios where growth factor is present. In absence of the growth factor the logical model cannot reproduce the behavior of the system, where for high expression of the E2F1 the tumor cell become independent of growth factors, which is one of the hallmark of cancer [190, 219] (see Table K.2). The logic-based model in CNA representations can be access at[1].

No.	Reactions	Reaction type	References (PubMed IDs)
1	EGF = EGFR	Monotone	11597398, 23447575
2	EGFR = ERK	Monotone	19935650
3a	FS = E2F1i	Monotone	23447575, 22871739
3b	2 FS = 2 E2F1i	Monotone	23447575, 22871739
3c	2 GxS = E2F1i	Monotone	23447575, 22871739
3d	GxS + FS = 2 E2F1i	Monotone	23447575, 22871739
3e	2 GxS + FS = 2 E2F1i	Monotone	23447575, 22871739
3f	GxS + 2FS = 2 E2F1i	Monotone	23447575, 22871739
4a	E2F1i + ERK = E2F1p	Monotone	11965250
4b	2 E2F1i + ERK = 2 E2F1p	Monotone	11965250, 23447575
5	E2F1p = p73	Non-monotone	22871739, 23447575
6a	E2F1p = DNp73	Non-monotone	22871739, 23447575
6b	2 E2F1p = 2 DNp73	Non-monotone	22871739, 23447575
7	P73 + 2 !DNp73 + !TGFB1 = miR-205	Non-monotone	22871739, 23447575

Table K.1. Logic-based reactions for core module of hybrid model. Logical reactions of the core module of the E2F1 transcriptional factor hybrid model. The target gene module and phenotypical modules has the same equations as in Table 4.3 in main text on page 78.

[1] https://sourceforge.net/projects/khan-phd/files/Exclusive_logical_model_CNA.rar/download

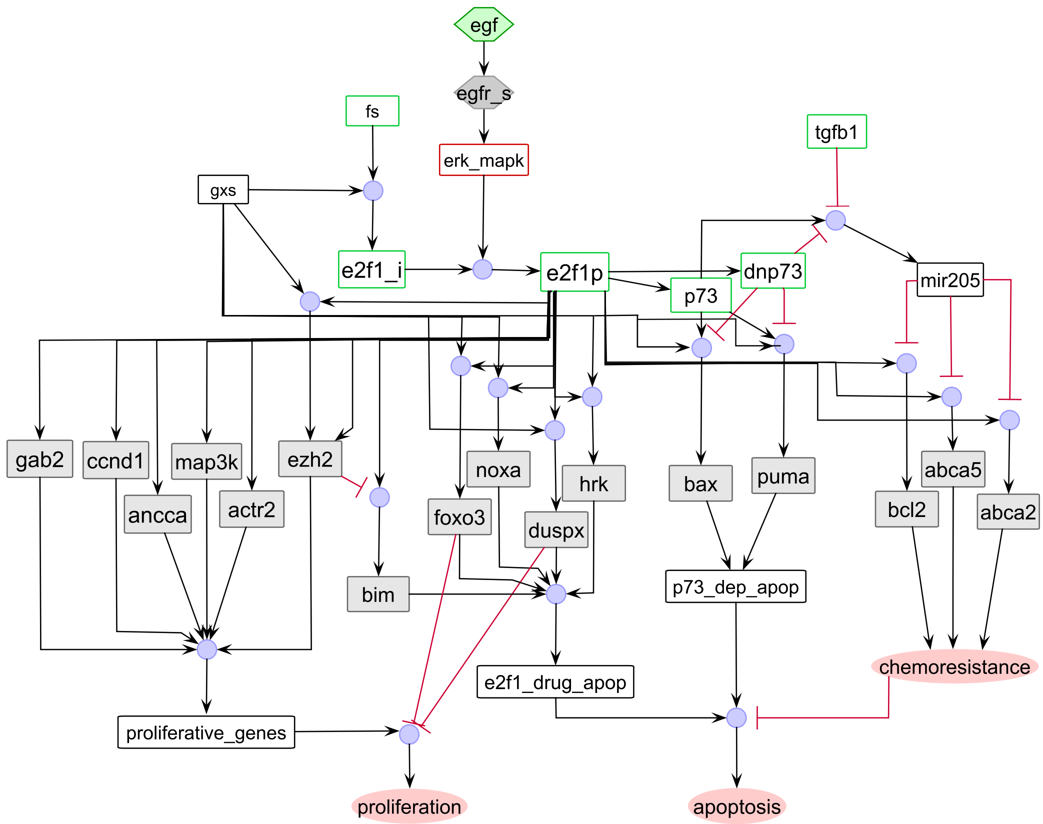


Figure K.1. Logical representation of hybrid model. The Logical representation of E2F1 hybrid model. The model is created using a process modeling tool (ProMoT) [195] and simulations are carried out in CellNetAnalyzer [199].

Inputs	GF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	FS	0	0	0	0	0	0	1	1	1	1	1	1	2	2	2	2	2	2	0	0	0	0	0	0	1	1	1	1	1	1	2	2	2	2	2	2
	TGFB1	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	1	1	1
	GxS	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
Core module	E2F1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	2	2	1	2	2	2	2	2	2	2	2
	p73	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
	DNp73	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
	miR-205	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0
Target genes	Gab2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	2	2	1	2	2	2	2	2	2	2	2
	CCND1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	2	2	1	2	2	2	2	2	2	2	2
	MAP3Ks	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	2	2	1	2	2	2	2	1	2	2	2
	ANCCA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	2	2	1	2	2	2	2	2	2	2	2
	ACTR2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	2	2	1	2	2	2	2	2	2	2	2
	EZH2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	2	2	1	2	2	2	2	2	2	2	2
	BIM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	FOXO3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	1	0	1	1	0	1	1	0	1	1
	DUSPx	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	1	0	1	1	0	1	1	0	1	1
	HRK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	1	0	1	1	0	1	1	0	1	1
	NOXA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	1	0	1	1	0	1	1	0	1	1
	BAX	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	PUMA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	BCL2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	1	1	0	1	1
	ABCA2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	1	1	0	1	1
	ABCA5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	1	1	0	1	1
Phenotype	Proliferation	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0
	Apoptosis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	Chemoresistance	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	1	1	0	1	1

Table K.2. Excusively logical simulation of E2F1 hybrid model. The logical model nicely capture the role of E2F1 in chemoresistance in the presence of growth factor (*i.e.*, $GF = 1$). But it can not produce the system dynamics when external growth factor is removed (*i.e.*, $GF = 0$) due to ignoring feedback loops in steady state analysis of logical model.

Theses

Faiz M. Khan - An integrative workflow for large-scale networks

1. To understand mechanisms underlying drug resistance and tumor invasion, the analysis of the E2F1-focused biochemical network can provide useful insights (Section 2.1).

2. Biochemical networks tend to be large and complex, therefore, a single computational method/approach cannot decode the mechanisms underlying cellular processes in normal and disease conditions (Section 2.2).

3. Using an integrative workflow that combines network structure properties, with high-throughput and biomedical data, one can identify cell type and disease phenotype-specific sub-networks, which are amenable for dynamical systems analysis (Section 2.2).

4. Logic-based modeling formalisms do not require detailed quantitative data which make them suitable for modeling large biochemical networks. Their analysis can identify disease signatures that can successfully stratify patients between early and advanced stages of cancer (Chapter 3).

5. Ordinary differential equations (ODEs) quantitatively describe the dynamics of small molecular networks to quantify expression levels of model variables over time (Section 4.3).

6. A hybrid model can be a good compromise between the convenience of logical models for scalability and the quantitative description of non-linear dynamics by ODE models (Section 4.4).

7. The integrative workflow and hybrid modeling framework can be generalized to any biochemical network underlying diseases (Chapter 5).

Curriculum Vitae

Faiz Muhammad Khan
Ziolkowskistr. 10
18059 Rostock
fm.khan163@gmail.com

Nationality	Pakistani
Date of birth	01 May 1985
Place of birth	Pawaka-Peshawar, Pakistan

Education

since 2012	PhD at the Dept. of Systems Biology & Bioinformatics, Universität Rostock, Rostock, Germany
2009 - 2012	Master of Science in Computational Engineering, at Faculty of Electrical Engineering and Information Technology, University of Rostock , Germany
2003 - 2008	Bachelor of Science in Computer Systems Engineering at University of Engineering and Technology, Peshawar, Pakistan

Contributions to grant proposals

2017	Unraveling mechanisms underlying acute inflammation and inflammation resolution. HEEL GmbH
2015	IMTA-Integrative modeling of Tumour Aggressiveness. DFG

Teaching and Administration Experience

since 2013	Hands on lectures and Labs in Systems Biology I & II Logic-based Model Analysis of Signaling and Transcription Networks
21 Feb 2018	Provided Training on Logic-based Modeling at the EU- and BMBF-funded 3rd OpenMultiMed Training School, Erlangen, Germany
WiSe 2016-17 WiSe 2018-19	Lab: Introduction to High Performance Computing
2018	Supervision of BSc project (Dragana Gjorgevikj): Layouting and Web Representation of a Molecular Interaction Map (Uni. of St. Kliment Ohridski-Bitola)
2016	Supervision of MSc project (Jan Keckeis): Large-scale Mathematical Model of E2F1 Signaling Network.
2013-2017	Research Group Seminar Organizer

Original publications and contributions

Main publications based on which this thesis is written

- **Khan FM**, Marquardt S, Gupta SK, Knoll S, Schmitz U, Spitschak A, Engelmann D, Vera J, Wolkenhauer O, Pützer BM. Unraveling a tumor type-specific regulatory core underlying E2F1-mediated epithelial-mesenchymal transition to predict receptor protein signatures. Nature Communications. 2017 Aug 4; 8(1):198. https://doi.org/10.1038/s41467-017-00268-2.

  I designed and implemented the workflow. I, along with Stephan Marquardt carried out molecular interaction map (MIM) construction and annotations. I performed the network structural analysis and integrated them with high-throughput and biomedical data to identify core-regulatory networks for epithelial-mesenchymal transition (EMT) regulation in bladder and breast cancer. I developed logic-based models of the core networks and calibrate them with expression data. I performed in silico analyses and identified molecular signatures and therapeutic targets for each cancer type. Further, I made a web visualization of the MIM. I wrote the manuscript.

- **Khan FM**, Schmitz U, Nikolov S, Engelmann D, Pützer BM, Wolkenhauer O, Vera J. Hybrid modeling of the crosstalk between signaling and transcriptional networks using ordinary differential equations and multi-valued logic. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics. 2014 Jan 1; 1844(1):289-98. https://doi.org/10.1016/j.bbapap.2013.05.007.

  I and Prof. Julio Vera conceived the original idea of hybrid modeling. The network structure was determined by me along with other co-authors. I developed the mathematical model and performed simulations and interpreted the results. I wrote the manuscript.

- Vera J, Schmitz U, Lai X, Engelmann D, **Khan FM**, Wolkenhauer O, Pützer BM. Kinetic modeling-based detection of genetic signatures that provide chemoresistance via the E2F1-p73/DNp73-miR-205 network. Cancer research. 2013 Feb 27. https://doi.org/10.1158/0008-5472.CAN-12-4095.

  I contributed in developing the methodology, writing, proof reading and revising the manuscript.

Other journal publications

- Dreyer FS., Cantone M, Eberhardt M, Jaitly T, Walter L, Wittmann J, Gupta SK, **Khan FM**, Wolkenhauer O, Pützer BM, Jäck HM, Heinzerling L, Vera J. A web platform for the network analysis of high-throughput data in melanoma and its use to investigate mechanisms of resistance to anti-PD1 immunotherapy. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 2018, 1864(6), 2315-2328. https://doi.org/10.1016/j.bbadis.2018.01.020.

- Sadeghi M, Ranjbar B, Ganjalikhany MR, **Khan FM**, Schmitz U, Wolkenhauer O, Gupta SK. MicroRNA and transcription factor gene regulatory network analysis reveals key regulatory elements associated with prostate cancer progression. PLoS One, 2016, 11(12), e0168760. https://doi.org/10.1371/journal.pone.0168760.

  I contributed in network analysis and interpretation of results. I proof read and edited the manuscript.

Book chapters

- **Khan F.M.**, Sadeghi M., Gupta S.K., Wolkenhauer O. (2018) A Network-Based Integrative Workflow to Unravel Mechanisms Underlying Disease Progression. In: Bizzarri M. (eds) Systems Biology. Methods in Molecular Biology, vol 1702. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-7456-6_12.

- **Khan F.M.**, Gupta S.K., Wolkenhauer O. Integrative workflows for network analysis. Essays in biochemistry. 2018 Oct 26;62(4):549-61. https://doi.org/10.1042/EBC20180005.

Scientific talks

- **Annotation of models and how to benefit from them** at INCOME2018: Conference and Hackathon, Bernried, Lake Starnberg, Germany, 17 Oct. 2018.

- **Integrative workflow for identification of diagnostic and therapeutic markers in tumor invasion** at INCOME2018: Conference and Hackathon, Bernried, Lake Starnberg, Germany, 15 Oct. 2018.

- **Identification of diagnostic and therapeutic markers in tumor invasion using logic-based modeling** in Workshop on Logical Modelling of Cellular Networks at ECCB 2018, Athen, Greece, 08 Sep. 2018.

- **Network-based integrative workflow to identify disease specific biomarkers and therapeutic candidates** at the EU- and BMBF-funded 3rd OpenMultiMed Training School, Erlangen, Germany, 23 Feb. 2018.

- **Detection of molecular signatures using computational methods** in MAJU Islamabad, Pakistan, 07 Oct. 2015.

- **Mathematical (logical) modeling approach to E2F1 signaling in tumor progression and chemoresistance** in Wissens-Forum at Institut für Eperimentalle Gentherapie und Tumorforschung, 18 Apr. 2013.

Poster presentations

- **Integrative workflow for identification of diagnostic and therapeutic markers in tumor invasion** at INCOME2018: Conference and Hackathon, Bernried, Lake Starnberg, Germany, 16 Oct. 2018.
- **An integrative workflow to identify disease specific biomarkers and therapeutic candidates** at Applied Bioinformatics in Life Science (2nd edition), Leuven, Belgium, 8-9 March 2018.
- **An integrative workflow to study large-scale biochemical networks** at ICMSB2017, Garching (near Munich), Germany, 26-28 July 2017.
- **Unraveling the role of E2F1 in tumor aggressiveness using integrative methods.** CASyM Summer School Djurhamn, Sweden, 22-26 June 2015.
- **Systems Biology of Metastasis (SysMet): E2F signaling in tumor progression and metastasis** at BMBF e:Bio Status Seminar, Berlin Germany, 22,23 September 2014.
- **SysMet: Computational multilevel model of E2F1 regulated signaling pathways in Tumor Progression and Metastasis** at BMBF e:Bio Kick off Meeting, Mainz, Germany, 23-25 September 2013.
- **Detection of potential drug targets in cancer signalling by mathematical modelling and optimization** at Forschungscamp Universität Rostock, Germany, 22 June 2012.

Copies of the main publications

Here, I attached original PDF copies of the main publications.

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-017-00268-2 OPEN

Unraveling a tumor type-specific regulatory core underlying E2F1-mediated epithelial-mesenchymal transition to predict receptor protein signatures

Faiz M. Khan[1], Stephan Marquardt[2], Shailendra K. Gupta[1,3], Susanne Knoll[2], Ulf Schmitz[1,4,5], Alf Spitschak[2], David Engelmann[2], Julio Vera[6], Olaf Wolkenhauer[1,7] & Brigitte M. Pützer[2]

Cancer is a disease of subverted regulatory pathways. In this paper, we reconstruct the regulatory network around E2F, a family of transcription factors whose deregulation has been associated to cancer progression, chemoresistance, invasiveness, and metastasis. We integrate gene expression profiles of cancer cell lines from two E2F1-driven highly aggressive bladder and breast tumors, and use network analysis methods to identify the tumor type-specific core of the network. By combining logic-based network modeling, in vitro experimentation, and gene expression profiles from patient cohorts displaying tumor aggressiveness, we identify and experimentally validate distinctive, tumor type-specific signatures of receptor proteins associated to epithelial–mesenchymal transition in bladder and breast cancer. Our integrative network-based methodology, exemplified in the case of E2F1-induced aggressive tumors, has the potential to support the design of cohort- as well as tumor type-specific treatments and ultimately, to fight metastasis and therapy resistance.

[1] Department of Systems Biology and Bioinformatics, University of Rostock, 18051 Rostock, Germany. [2] Institute of Experimental Gene Therapy and Cancer Research, Rostock University Medical Center, 18057 Rostock, Germany. [3] Department of Bioinformatics, CSIR—Indian Institute of Toxicology Research, Lucknow 206001, India. [4] Gene & Stem Cell Therapy Program, Centenary Institute, Camperdown, NSW 2050, Australia. [5] Sydney Medical School, University of Sydney, Sydney, NSW 2006, Australia. [6] Department of Dermatology, Laboratory of Systems Tumor Immunology, Erlangen University Hospital and FAU University of Erlangen-Nuremberg, 91054 Erlangen, Germany. [7] Stellenbosch Institute for Advanced Study (STIAS), Wallenberg Research Centre at Stellenbosch University, Stellenbosch 7600, South Africa. Faiz M. Khan, Stephan Marquardt, and Shailendra K. Gupta contributed equally to this work. Julio Vera, Olaf Wolkenhauer, and Brigitte M. Pützer jointly supervised this work. Correspondence and requests for materials should be addressed to J.V. (email: Julio.Vera-Gonzalez@uk-erlangen.de)

Recent advances in sequencing and omics technologies provide us with data that can be used to identify and characterize cancer and tumor-specific molecular networks. The analyses of these networks have given insights into various aspects of carcinogenesis, tumor progression, and metastasis[1, 2]. The set of mutated, deregulated, or epigenetically modified cancer genes is highly patient and tumor-type variable, and more important, these genes are integrated in a small set of regulatory pathways[2]. Further, these pathways are not isolated: they crosstalk to shape and fine-tune basic cellular phenotypes that are subverted in cancer. In recent years, several researchers have deployed methodologies based on the reconstruction of cancer-associated networks and used them to analyze high-throughput cancer data[3–5]. Interestingly, it has been found that cancer networks are enriched in regulatory motifs, and beyond, that cancer-related regulatory motifs do crosstalk. Network hubs, feedback, and feedforward loops, the regulatory motifs often encountered in cancer networks, are able to induce a complex regulatory behavior that evades the use of conventional data analysis tools for their understanding[6, 7]. Hence, the utilization of advanced network-based methodologies and mathematical modeling becomes necessary to get a deeper understanding of cancer networks.

An outstanding example of a deregulated cancer network is the one controlled by the E2F family of transcription factors. The most prominent member of this family, E2F1, is involved in a number of essential cancer-related cellular processes such as proliferation, apoptosis, and differentiation[8]. E2F1 is a remarkable example of a network hub as this protein interacts with many genes, proteins, and other transcription factors through a variety of regulatory mechanisms. In the context of solid tumors, unbalanced E2F1 regulation can lead to the emergence of aggressive tumor cells, which drive cancer progression, resistance to anti-cancer drugs, and the rise of metastatic lesions[9–13].

Enforced E2F1 expression in advanced tumors and metastases of different kinds of cancers correlates with pronounced resistance towards therapy and poor patient prognosis[14, 15]. E2F1 drives epithelial–mesenchymal transition (EMT), similar to the classical EMT inducer TGFB1, via signaling pathways that involve non-coding RNAs[16]. As a direct target of E2F1, enforced expression of *miR-224/452* in melanoma cells stimulates a mesenchymal phenotype by repressing the metastasis suppressor *TXNIP* associated with changes in the actin cytoskeleton towards an enhanced invasive cell behavior. TXNIP in turn controls E2F1 activity in a negative regulatory loop. This process is reversible through ablation of endogenous E2F1 in highly aggressive skin cancer cells, leading to increased TXNIP and E-Cadherin and loss of mesenchymal markers SNAI2, ZEB1, and Vimentin[9, 17].

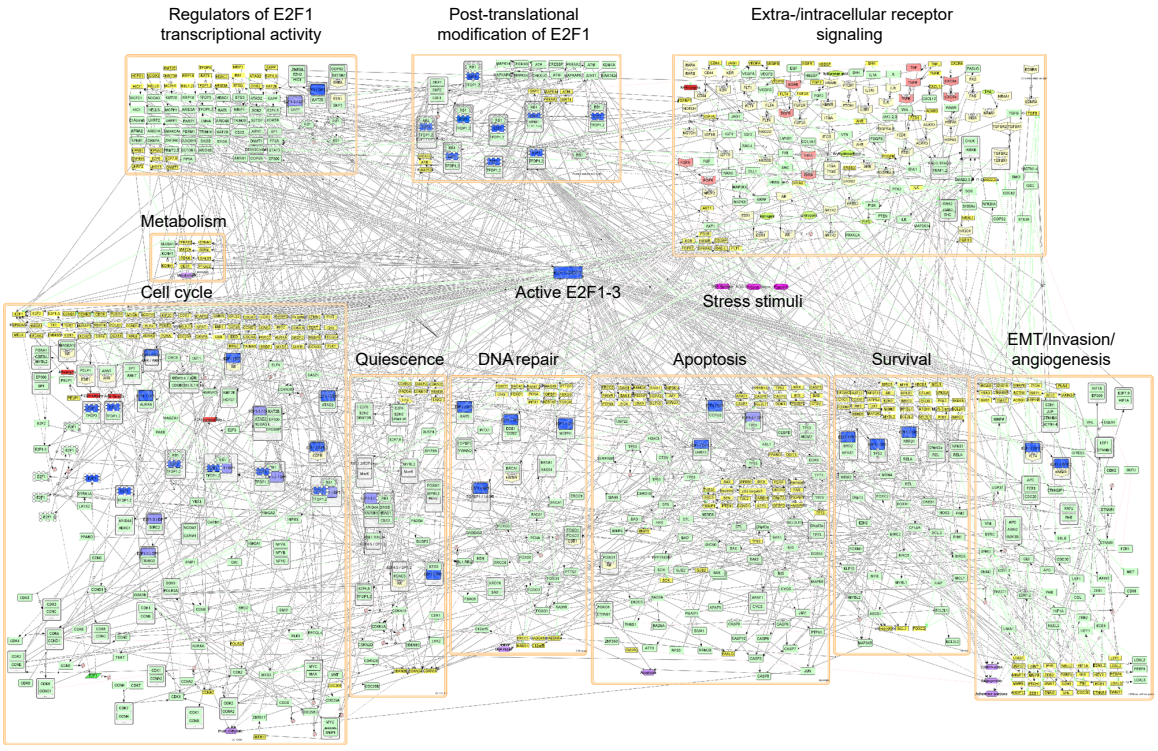


Fig. 1 A modularized map of E2F1 in tumor progression and metastasis. The map contains three E2F1 regulatory compartments: (i) Extra-/intracellular receptor signaling ($n = 113$); (ii) Post-translational modifications of E2F1 ($n = 24$); (iii) Regulators of E2F1 transcriptional activity ($n = 66$). Furthermore, there are seven functional compartments: (i) Cell cycle ($n = 145$); (ii) Quiescence ($n = 29$); (iii) DNA repair ($n = 33$); (iv) Metabolism ($n = 11$); (v) Apoptosis ($n = 89$); (vi) Survival ($n = 52$); (vii) EMT/invasion/angiogenesis ($n = 69$), where n stands for the number of factors in each compartment. Biomolecules are visualized in standard CellDesigner format (gene: *yellow rectangle*; protein: *light green round cornered box*; receptor: *light yellow hexagon*; ligand: *green oval*; phenotype: *violet hexagon*; drug/external stimulus: *pink box*). For better visualization, in the map transcription and translation are condensed to one reaction directly leading from gene to protein. In *red*, we represent place holders for protein families (e.g., FGFR for FGFR1-4, FGF for FGF1-23, ITGA, and ITGB for alpha and beta integrins) and unspecified genes responding to a given transcription factor. The microRNA layer is not included in this CellDesigner diagram, but in the Cytoscape network provided as Supplementary Information. The interactive E2F1 interaction map can be accessed at: https://navicell.curie.fr/pages/maps_e2f1.html

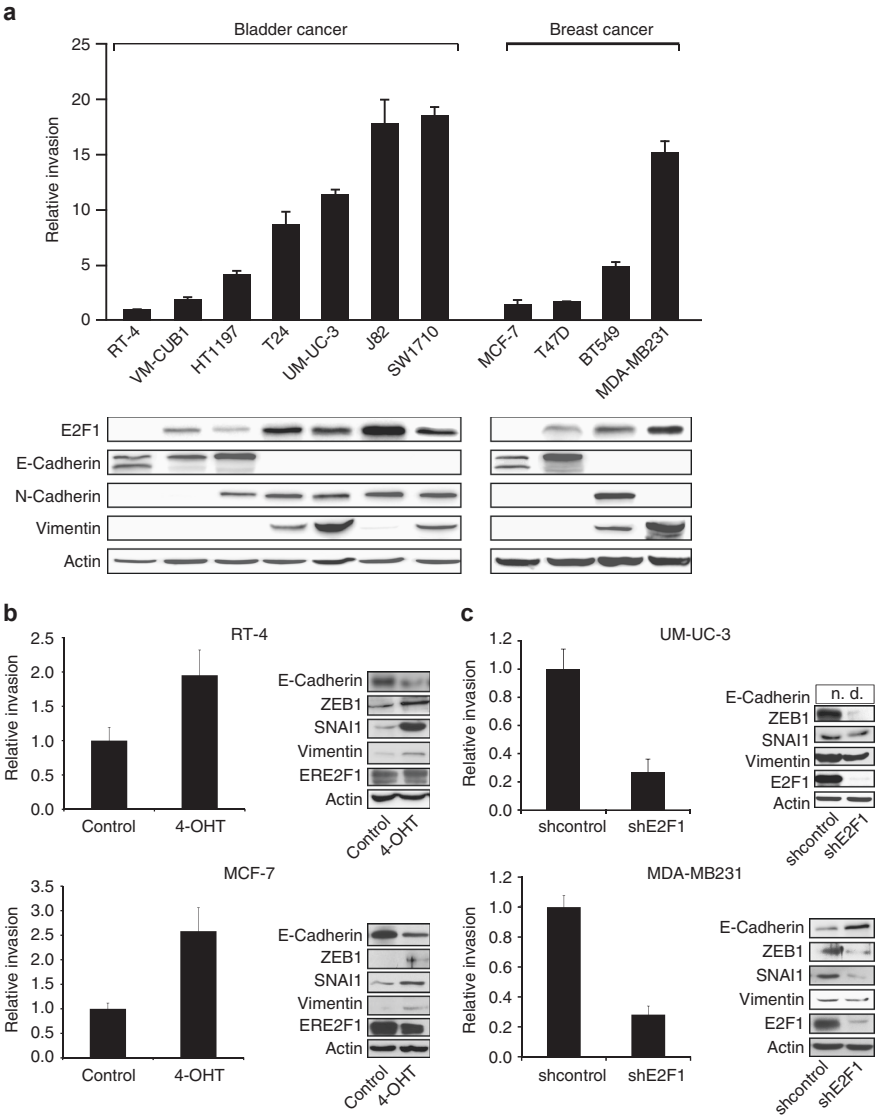


Fig. 2 Invasive potential and EMT marker expression in less-invasive and invasive human bladder and breast cancer cell lines. **a** Boyden chamber assay and western blot showing the invasive potential and the expression of E2F1 and EMT markers of various bladder and breast cancer lines. RT-4 was used as reference. **b** Indicated cells were transduced with adenoviral vector expressing the 4-OHT responsive estrogen receptor (*ER*)-E2F1 fusion protein (Ad.ER-E2F1) to conditionally activate E2F1 nuclear translocation. After 24 h they were applied to Boyden chamber assay and induced with 4-OHT or ethanol as control (*left panels*) or harvested for protein isolation and western blotting (*right panels*). **c** Cells were transduced with adenoviral vector expressing *shE2F1* or *shcontrol*. After 72 h cell invasion was determined (*left*). Protein level and *shE2F1* knockdown is indicated by immunoblots (*right*). All figures are representatives of at least three independent experiments. *Error bars* indicate s.e.m., *n.d.* not detectable

Another mediator of E2F1-induced EMT is *miR-205*[11, 13]. Inhibition of the E-Cadherin repressors *ZEB1* and *ZEB2* by this microRNA results in stabilization of the epithelial cancer cell phenotype[18]. The relevance of this transcription factor to tumor progression was also shown in a genetic model by interbreeding Neu transgenics with *E2F1* knockout mice as well as in *HER2*+ breast cancer patients, in which the E2F activation status predicts relapse and metastatic potential of MMTV-Neu-induced tumors[19]. In fact, E2F-responsive genes define a novel molecular subset of high-grade human tumors of the breast, ovary, and prostate, termed ERGO (E2F-responsive gene overexpressing) cancers[20]. Our studies also revealed that vascular endothelial growth factor-C (*VEGF-C*) and its cognate receptor *VEGFR-3*, both highly upregulated in cancer cells with abundant E2F1 expression, are direct targets of this transcription factor[21]. Co-regulation of *VEGF-C*/*VEGFR-3* by E2F1 stimulates endothelial cells to form tubule-like structures and promotes neovascularization in mice. *E2F1* is activated by VEGFR-3 signaling in a positive feedback loop and both proteins cooperate in the nucleus to co-regulate transactivation of the proangiogenic cytokine *PDGF-B*. In addition, we identified the epidermal growth factor receptor (*EGFR*) as a direct target of E2F1 and demonstrated that inhibition of receptor signaling abrogates E2F1-induced invasiveness[9]. These results provide

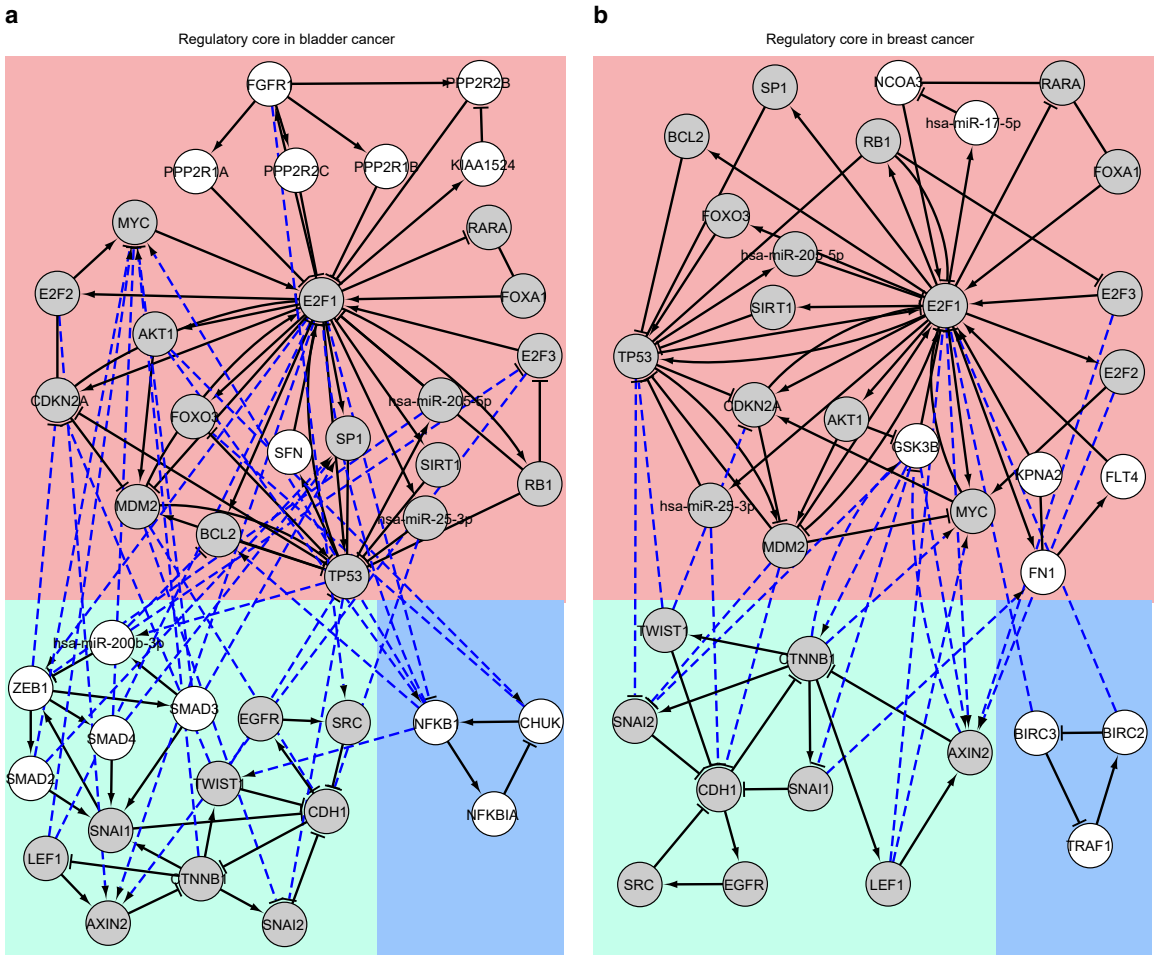


Fig. 3 Core regulatory networks that drive tumor invasiveness in **a** bladder and **b** breast cancer. The regulatory cores contain three disjoint sub-networks (shown in *blue, pink,* and *green background colors*). *Blue dotted lines* represent direct interactions extracted from the comprehensive E2F1 regulatory network to interconnect these sub-networks. The networks were generated using Cytoscape plugin NetDS. *Nodes* in *gray color* are common in both regulatory core networks, whereas those in *white color* are unique

support for the outstanding, cell context-dependent unique role of E2F1 in driving cancer aggressiveness.

Here, we use a network approach to identify the tumor type-specific regulatory core and to predict receptor protein signatures associated with E2F1-mediated EMT transitions in two types of highly aggressive solid tumors, bladder, and breast cancer. To this end, we construct a comprehensive map of the regulatory network around the E2F family. By mapping gene expression profiles from cancer cell lines displaying the features of EMT transition onto the E2F1 interaction map, we identify a tumor type-specific regulatory core. We then analyze the regulatory core to predict tumor-specific receptor protein signatures linked to aggressiveness. By conducting in vitro experiments, we could verify the impact of these molecules on tumor cell invasiveness. We also found a correlation between the molecular signatures and clinical tumor aggressiveness in relevant patient data.

Results

A comprehensive E2F1 interaction map. To understand how E2F1 interacts with different molecules and how it mediates cancer-related processes, we constructed a functionally modularized interaction map based on information retrieved from published literature and databases (Fig. 1). The comprehensive map of E2F1 regulation and activity is based on manual exploration of over 800 publications related to E2F1 and other E2F family proteins, as well as connected pathways having a role in cancer-related cellular processes. The map contains 879 nodes including different types of factors (genes, proteins, microRNAs, or complexes) and 2278 interactions. To improve visualization, we modularized the map into several regulatory and functional compartments (Fig. 1). The map comprises HUGO annotations of all the factors, together with meta-information about isoform expression (e.g., DNp73 or mutant TP53) and corresponding PubMed references.

E2F1 drives EMT in bladder and breast cancer. Recent clinical results indicate that E2F1 is upregulated in high-grade bladder and breast cancers[14, 22, 23]. To further substantiate these findings, we examined the effects of E2F1 activity on the invasive capacity of patient-derived metastatic bladder and breast tumor cell lines using functional invasion assays, western blotting, and PCR analysis. The experiments revealed a clear correlation of E2F1 expression with the invasive behavior and EMT marker expression in both cell models: high levels of E2F1 and mesenchymal markers in invasive bladder

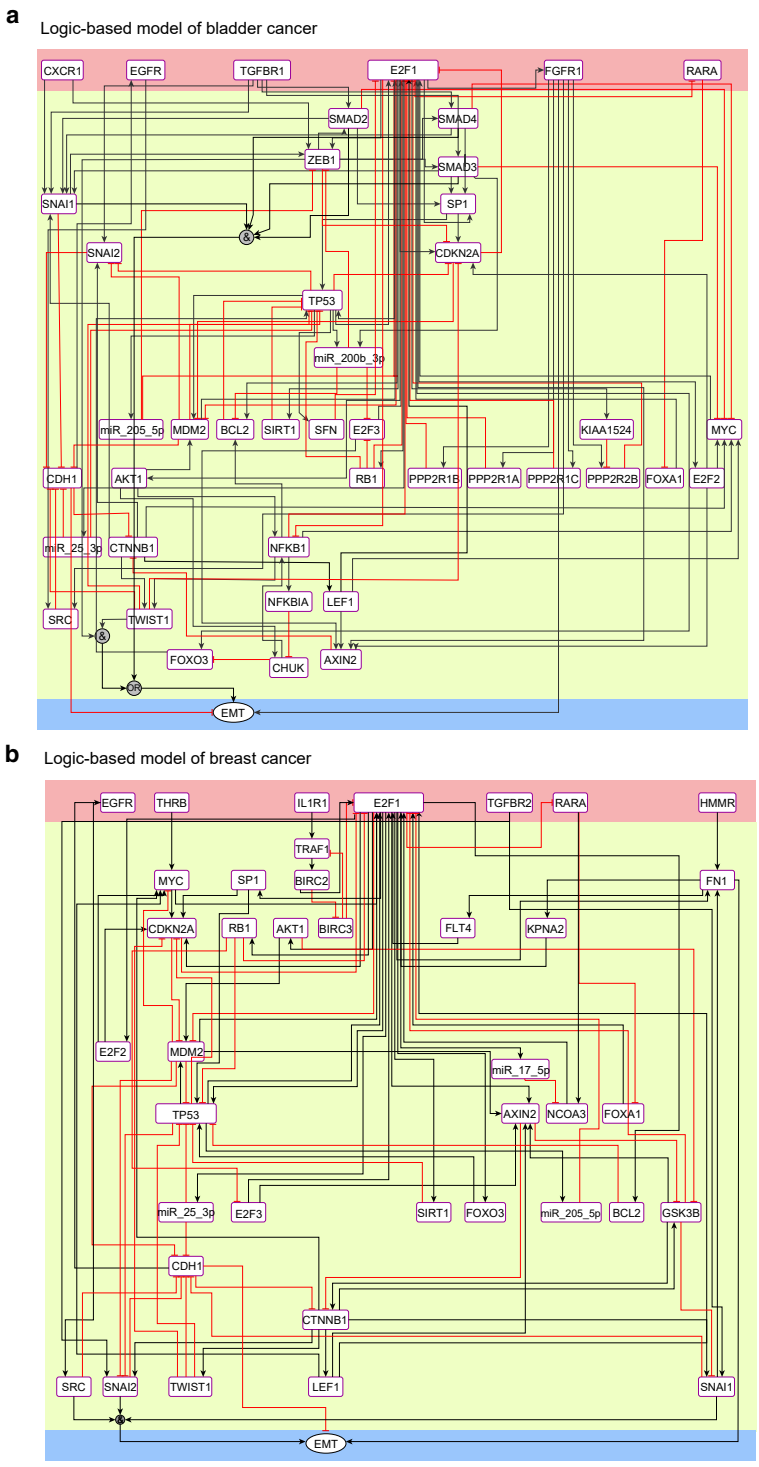


Fig. 4 Logic-based models of the core network regulating invasive phenotypes in **a** bladder and **b** breast cancer. The *black* and *red lines* represent the type of interactions (i.e., activation and inhibition) among the interacting components. The microRNAs, receptors, and proteins are represented by *rectangular nodes*, whereas the EMT phenotype is represented by an *ellipse* in the output layer. The model is divided into three layers, i.e., the input layer (*pink*), the regulatory layer (*yellow*), and the output layer (*blue*)

(UM-UC-3, SW1710, J82, T24) and breast (MDA-MB231, BT549) vs. low expression in non-invasive or less-invasive epithelial bladder (RT-4, VM-CUB1, HT1197) and breast (MCF-7 and T47D) cell lines (Fig. 2a). Our experimental data are supported by gene expression data from the CCLE database (Supplementary Fig. 5). Increased expression of E2F1 commonly observed in EMT-like cell lines is also evident from several other aggressive tumor entities such as pancreatic, lung, and prostate cancer (Supplementary Fig. 6) as well as cutaneous melanoma[9]. Furthermore, overexpression of E2F1 induces an invasive phenotype in RT-4 and MCF-7 cells by upregulation of mesenchymal and downregulation of epithelial markers (Fig. 2b). In contrast, knockdown of the transcription factor in aggressive UM-UC-3 and MDA-MB231 cells results in reduced invasiveness and decreased expression of mesenchymal ZEB1 as well as upregulation of E-Cadherin in MDA-MB231 (Fig. 2c).

Network motif prioritization. From the E2F1 interaction map, we identified a large set of feedback loops ($n = 444$; 213 positive, 228 negative, and 3 neutral) that are responsible for non-intuitive behavior of the system (Supplementary Data 1). From such a large set of feedback loops, our aim was to identify the most important ones involved in the investigated phenotype. Towards this end, we used a weighted multi-objective function containing topological and non-topological network parameters (see "Methods"). Further, we selected different weighting scenarios for motif prioritization to avoid any bias induced by the parameters used in the multi-objective function (for details see Supplementary Methods).

We used KEGG's cancer disease pathway (KEGG: hsa05200) to estimate the number of nodes of a motif associated to a cancer pathway. For calculating the gene prioritization parameter, we first selected all the known EMT markers in our map proposed in Lanouille et al.[24] and calculated the score for all the nodes using a random walk with restart algorithm implemented in the Cytoscape plugin GPEC[25]. A complete list of the selected motifs along with structural and biomedical parameters as well as the motifs ranking scores are provided in Supplementary Data 1. Further, we selected the top ten motifs from each of the weighting scenarios implemented in the multi-objective function. In this way, we obtained 32 non-redundant motifs associated with an invasive phenotype in bladder and 28 with breast cancer, respectively.

Derivation of tumor type-specific core regulatory network. To obtain the core regulatory network, we merged all the unique motifs using the Cytoscape plugin NetDS[26] (v3.0). We retrieved three disjoint sub-networks in both tumor entities, which we connected by reviving interactions among the nodes from the E2F1 regulatory network. Thereby, we obtained tumor-specific (bladder and breast cancer) core regulatory networks (Fig. 3). Both networks include the transcription factors E2F1-3 and the cell cycle regulators RB1, MYC, CDKN2A, TP53/MDM2, SP1, FOXA1, FOXO3, and AKT1. Furthermore, both contain the enzyme SIRT1 that modifies targets like E2F1, TP53, and histones to silence their function. In addition, both core networks contain the CDH1 regulators SNAI1/2 and TWIST1, and interaction partners of CTNNB1 (AXIN2, LEF1).

In bladder cancer, we additionally found fibroblast growth factor receptor 1 (FGFR1) and its downstream regulatory subunits of the protein phosphatase 2 (PP2; inhibitor of cell growth and division), and the pro-proliferative inhibitor of PP2, KIAA1524 (CIP2A). Also, we found the transforming growth factor beta receptor (TGFBR) 1/2 downstream signaling molecules SMAD2-4 and their regulator ZEB1. In the breast cancer core network, we found Fibronectin 1 (FN1), which is related to migration; FLT4 involved in angiogenesis; GSK3B, an anti-proliferative enzyme; KPNA2, a nucleopore transporter and the EP300-associated transcriptional activator NCOA3.

In addition, we observed feedback loops concerning nuclear factor kappa B subunit 1 (NFKB1) activation (CHUK, NFKBIA; related to cell survival) in bladder cancer, anti-apoptotic BIRC2/3, and pro-apoptotic TRAF1 factors in breast cancer, respectively. Interestingly, both loops are related to each other, as NFKB1 and BIRC2/3 are survival molecules and NFKB1 activates BIRC transcription. The regulatory cores, which we consider as the drivers of the invasive phenotypes contain 41 nodes and 107 interactions in bladder cancer and 35 nodes and 86 interactions in breast cancer.

Logic-based models for EMT-driving molecular signatures. To evaluate the input–output relationship of the obtained regulatory core networks, we used logic-based modeling formalism. The input layer of the logic-based models contains E2F1 and all receptors present in the regulatory core (Fig. 4). In bladder and breast cancer, two common receptors are part of the input layer: (i) EGFR and (ii) the retinoic acid receptor alpha (RARA). In addition, we found FGFR1 only in the regulatory core of bladder cancer. In simulations of the model, we considered signal propagation from input to output layer.

To capture all possible input signals to the regulatory network cores, we expanded the input layers by including additional receptors present in the comprehensive interaction network, which are directly connected to nodes constituting the regulatory cores. Thus, we included TGFBR, which is connected to SMADs, and the chemokine (C-X-C motif) receptor 1 (CXCR1) connected to ZEB1 and SNAI1 in the bladder cancer model. Similarly, we expanded the breast cancer input layer with the hyaluronan-mediated motility receptor (HMMR) connected to FN1; the TGFBR connected to SNAI1 and SNAI2; the interLeukin 1 receptor type I (IL1R1) and the thyroid hormone receptor beta (THRB) connected to TRAF1 and MYC, respectively. For bladder cancer, we selected TGFBR1 and for breast TGFBR2 due to their tissue-specific expression profiles. We derived Boolean functions for the input signals and their propagation through the nodes constituting the regulatory layer (Supplementary Data 2).

The output layer of the models comprises a unique node that represents the EMT process as the driver of the invasive phenotype, which is determined by a logic function involving the EMT markers present in the regulatory core. The output, determined through a multi-valued logic function, accepts four ordinal levels, ranging from 0 (no EMT) to 3 (high EMT).

Predictive model simulations. We determined the steady states of each variable in the model for different initial values of the input nodes. We consider two sets of scenarios, characterized by: (i) high expression of E2F1 (i.e., E2F1 = 1); and (ii) low expression (i.e., E2F1 = 0) in all possible Boolean combinations of receptors in the input layer. We obtained 64 input vectors for bladder cancer and 128 for breast cancer. Next, we simulated the network to determine the impact of the input vectors on the level of EMT (Table 1). Our simulation results suggest that when E2F1, TGFBR1, and FGFR1 are simultaneously active, bladder cancer cells become highly invasive (EMT = 3). A similar effect was observed in breast cancer when E2F1, TGFBR2, and EGFR are simultaneously active. Furthermore, we carried out in silico perturbation experiments to identify important nodes that can be exploited for therapeutic interventions. Perturbation experiments were performed for a highly invasive phenotype (EMT = 3) by changing the Boolean state of each node in the regulatory layer to reduce invasiveness. We used single and double perturbation

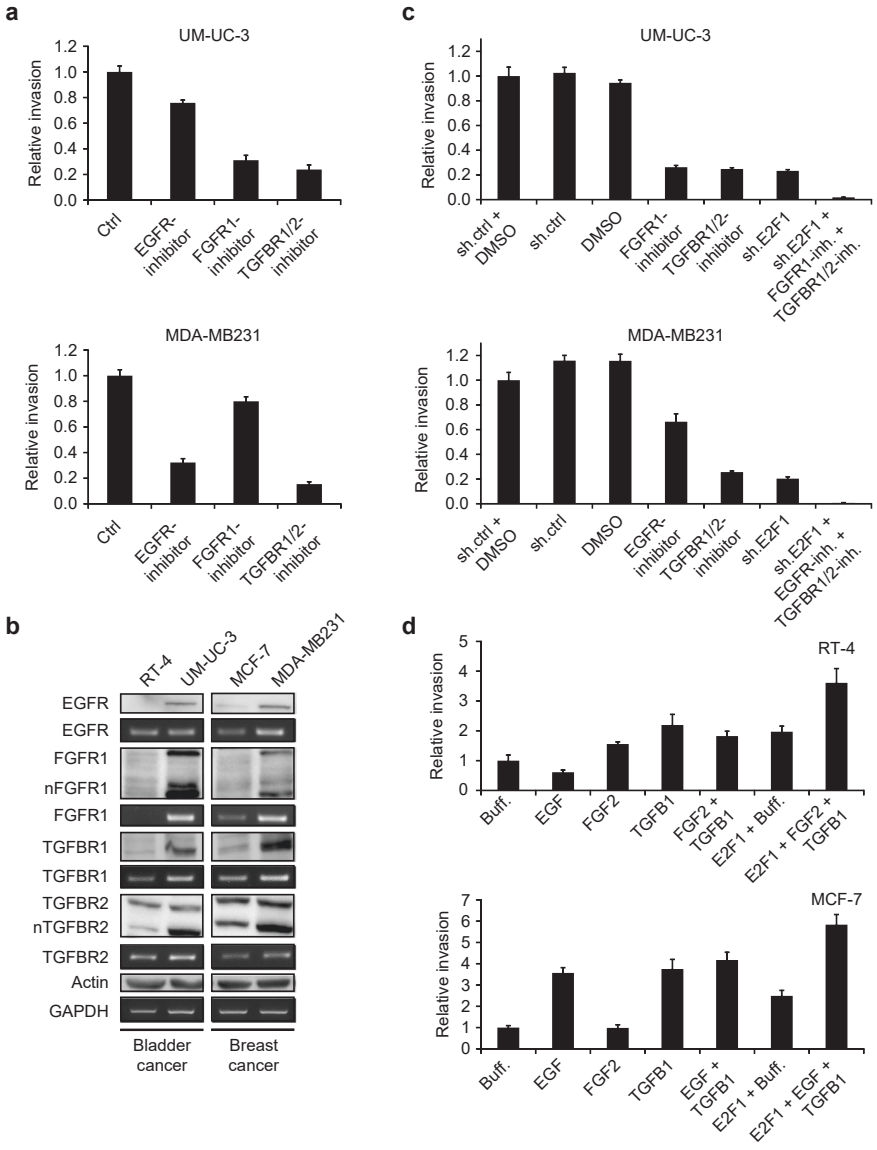


Fig. 5 Effects of the EGFR, FGFR1, and TGFBR pathways on bladder and breast cancer invasion. **a** Different effects of EGFR inhibitor and FGFR1 inhibitor on invasive bladder (UM-UC-3) and breast cancer (MDA-MB231) cell lines. **b** Western blots and PCRs show the expression levels of the indicated receptors in less-invasive (RT-4, MCF-7) vs. invasive (UM-UC-3, MDA-MB231) cells. EGFR, FGFR1, as well as TGFBR1 are highly expressed in both invasive cell lines compared to the less-invasive ones. TGFBR2 expression levels equal within cell lines of each tissue type, whereas the nuclear fraction of TGFBR2 (nTGFBR2 as well as nFGFR1 for FGFR1) shows a clear upregulation in both invasive cancer cell lines. **c** Invasive potential of UM-UC-3 or MDA-MB231 was measured by Boyden chamber assay upon treatment with an *E2F1*-specific shRNA and inhibitors for TGFBR1/2, EGFR, or FGFR1 as indicated. **d** Boyden assay indicating the invasive potential of epithelial cell lines after stimulation with growth factors and/or overexpression of E2F1 transcription factor. Cells were preincubated in growth factor supplemented medium (10 ng/ml of EGF, FGF2, TGFB1, or both) and transduced with adenoviral vector for overexpression of ER-E2F1, which was activated by adding 4-OHT. Fold-changes were calculated relative to control cells (set as 1). All figures are representatives of at least three independent experiments. All *error bars* indicate s.e.m.

iteratively and observed the most prominent reduction of EMT in the latter case (simulation results are provided in Supplementary Data 5). Our perturbation results suggest that in bladder cancer (i) double knockout of *ZEB1* in combination with either *SNAI1*, *TWIST1*, *NFKB1*; (ii) knockout of *ZEB1* and activation of *CDH1*; or (iii) knockout of *SMAD2/3/4* in combination with *TWIST1* or *NFKB1* reduces EMT to 1. In case of breast cancer double perturbation by silencing *SRC*, *FN1*, *SNAI1*, *SNAI2*, or activation of *CDH1* in any of the combinations reduces EMT to 1 (Supplementary Table 4).

Validation of in silico predictions with cell line models. Our model predictions revealed a common impact of E2F1 and TGFB1 signaling on tumor invasiveness in both cancer types. More specifically, TGFBR1 and FGFR1 in combination with

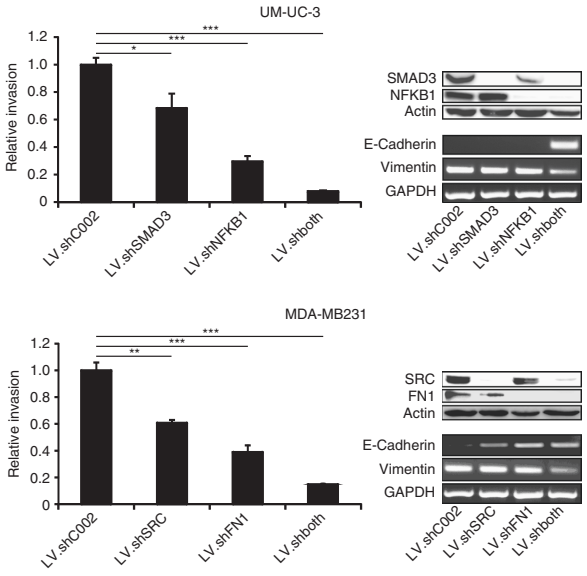


Fig. 6 Validation of in silico knockout simulations by gene knockdown experiments. By in silico simulations, we identified the most effective combination of double knockouts regarding reversal of EMT. The invasive potential of UM-UC-3 and MDA-MB231 was measured by Boyden chamber assay upon treatment with lentiviral vector expressing *SMAD3, NFKB1, SRC,* or *FN1*-specific shRNA, alone or in combination (*left panels*). Changes of EMT markers *E-Cadherin* and *Vimentin* after gene knockdown are shown on transcriptional level (*right panels*). Fold-changes were calculated relative to control cells (treated with control LV.shC002; set as 1). All *error bars* indicate s.e.m. For statistical significance *t*-test was used (**P*-value < 0.05; ***P*-value < 0.01; ****P*-value < 0.001). ShRNA efficiency was confirmed by western blot

highly expressed E2F1 induce the most invasive phenotype in bladder cancer, whereas in breast cancer, it is the combined action of TGFBR2, EGFR, and E2F1 that triggers high levels of invasiveness. To validate the predicted influence of the receptors and E2F1 on the invasive phenotype, we used chemical inhibitors and a shRNA-based approach to target these key players. In line with the in silico simulations, inhibition of EGFR in UM-UC-3 (bladder) or FGFR1 in MDA-MB231 (breast) had a minor influence on cell invasion (Fig. 5a). To confirm that this is not due to a lack of EGFR in UM-UC-3 or FGFR1 in MDA-MB231, respectively, receptor expression was confirmed by PCR and immunoblot in all cell lines. As Fig. 5b shows, all receptors are highly expressed in both invasive cell lines. Furthermore, as predicted by the simulations, inhibition of E2F1, TGFBR1/2 and FGFR1 in UM-UC-3 and E2F1, TGFBR1/2, and EGFR in MDA-MB231 had a tremendous impact on the invasive behavior of the respective cell line with the highest effect upon combined inhibition (Fig. 5c).

Referring to the initial data where we have shown induction of an invasive phenotype in epithelial cell lines (RT-4, MCF-7) by overexpressing E2F1, we now stimulate the signaling pathways by applying their respective ligands (epidermal growth factor (EGF), fibroblast growth factor 2 (FGF2), transforming growth factor beta-1 (TGFB1)) to induce invasion in those cell lines. Figure 5d demonstrates that the in silico predicted selective response to the different stimuli actually occurs: In RT-4 cells, overexpression of E2F1 or stimulation with FGFR1 ligand or TGFBR2 ligand (alone or in combination) raises the invasive potential, whereas stimulation with EGFR ligand has no effect on cell invasion. In contrast, EGFR stimulation of MCF-7 cells promotes cell invasion in a manner comparable with the TGFBR1/2 stimulation or E2F1 overexpression. Here, activation of FGFR1 has no impact on the invasive potential of the MCF-7 cell line. Taken together, by combining the predictions from in silico simulations and the in vitro experimental validation, we were able to find molecular signatures that regulate invasive phenotypes in E2F1-driven bladder and breast cancer.

To validate findings of the in silico perturbation simulations, we decided to knockdown *NFKB1* and *SMAD3* in UM-UC-3, and *SRC* and *FN1* in MDA-MB231, instead of modulating the other well known EMT markers *SNAI1/2*, *TWIST1*, *ZEB1*, or *CDH1*. We performed single and double knockdown of these genes and measured both the transcriptional and EMT/MET response. Although the removal of single genes resulted in a clear reversal of the EMT phenotype (reduced invasion) in both cell lines, the strongest effect was observed after double knockdown, as demonstrated by their lowest invasive capacity, increased E-Cadherin, and decreased Vimentin levels (Fig. 6).

Validation of model predictions with patient data. To further validate the molecular signature that regulates invasiveness in bladder cancer, we used data from a patient cohort ($n = 165$)[14], in which a correlation between *E2F1* expression and superficial to invasive progression was observed (GEO id: GSE13507). We grouped the patients into high and low expression profiles of *E2F1*, *TGFBR1*, and *FGFR1* from their respective median expression values. For each group, we calculated the progression-free survival probability and found that the survival probability was higher in the patient group with low expression of each molecule individually. Furthermore, we identified the subgroups of patients with high vs. low expression of: (i) *E2F1–FGFR1*; (ii) *E2F1–TGFBR1*; and (iii) *E2F1–FGFR1–TGFBR1*. The progression-free survival probability of each subgroup reveals that patients with high expression of *E2F1–FGFR1* have lowest mean survival time (33.79 months), whereas those with low expression of *E2F1–FGFR1–TGFBR1* have the best prognosis (93.35 months) among all the subgroups analyzed (see Kaplan–Meier plots in Fig. 7a–c). Our analyses indicate that patients with low expression survive more than twice as long as the patient subgroup with high expression of the molecular signature.

To validate the molecular signature from the regulatory core in breast cancer, we used data from the TRANSBIG network (GEO id: GSE7390; $n = 198$) generated by Desmedt et al.[22] Similar to bladder cancer patients, we observed that the progression-free survival probability of breast cancer patients was low for high expression of *E2F1*, *EGFR*, and *TGFBR2* in different combinations (Fig. 7d–f). Interestingly, we observed the highest mean survival time (97.29 months) in the patient subgroup with low expression of *E2F1–EGFR–TGFBR2*. Similar to the bladder cancer analyses, the patient subgroup with high expression of all three components had nearly half the mean survival time (50.11 months) compared with the subgroup with low expression.

We further validated molecular signatures in large patient cohorts of TCGA bladder cancer (BLCA; $n = 426$) and TCGA breast cancer (BRCA; $n = 1218$) accessible through UCSC Xena (http://xena.ucsc.edu). We found that signatures predicted using Boolean simulations were able to distribute patients into early vs. advanced stages in bladder cancer and aggressive vs. less-aggressive stages in breast cancer significantly (*P*-value < 0.005) (Fig. 8a, b; Supplementary Fig. 7). To assess the capability of our workflow in predicting significant molecular signatures associated with invasive phenotypes, we generated 30 random signatures of three nodes from each of the regulatory cores and arbitrarily

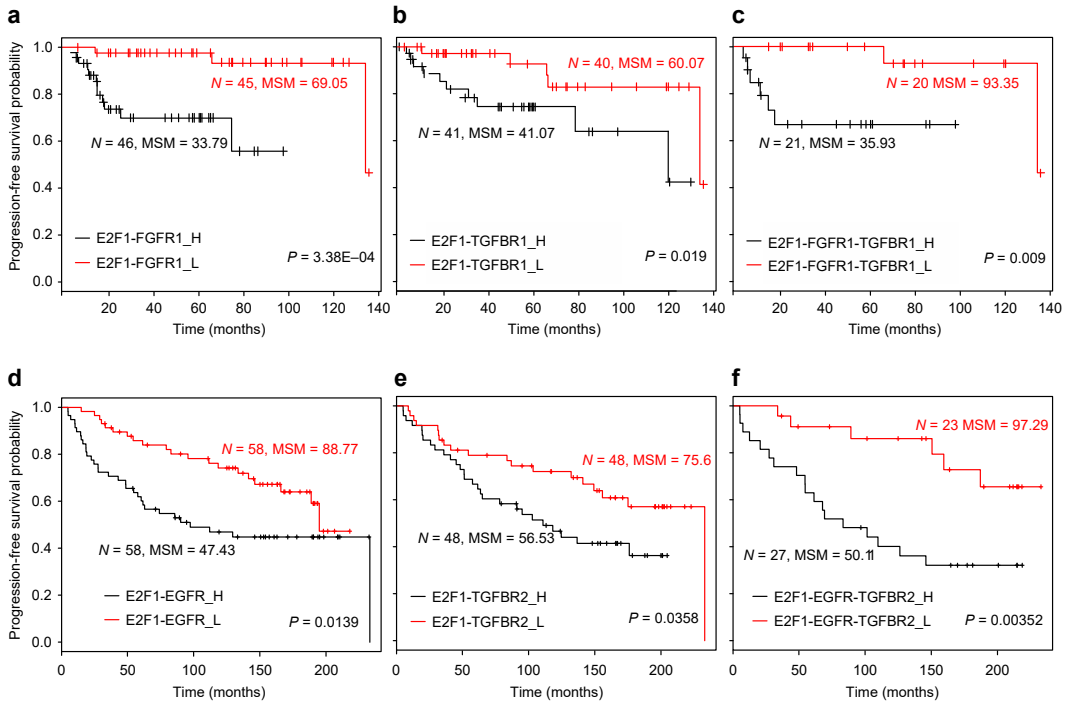


Fig. 7 Kaplan–Meier plots of progression-free survival of patients with bladder and breast cancer. Plots **a**–**c** show the survival curves for patients with high and low expression of combined signatures (*E2F1–FGFR1*; *E2F1–TGFBR1*; and *E2F1–FGFR1–TGFBR1*) in bladder cancer patients, whereas plots **d**–**f** are for the combined signatures of *E2F1–EGFR*; *E2F1–TGFBR2*; and *E2F1–EGFR–TGFBR2* in breast cancer patients. In both cases, patients with low expression of the molecular signatures have high mean survival times and vice versa. High expression of molecular signature(s) is represented as '_H' (*black curve*) and low expression as '_L' (*red curve*). '*N*' is the number of patients observed with high/low expression of molecular signatures and 'MSM' is the mean survival month from the patient group. *P*-values shown in the figures are for log rank test

assigned high or low expression values (Supplementary Data 7 and 8). We observed that the molecular signature predicted for bladder cancer is the only one that nicely distinguishes between the early and advanced stage of disease (Fig. 8c). In case of breast cancer, in addition to the predicted signature, some of the random signatures were also able to distinguish between aggressive and less-aggressive cancer types (Fig. 8d). This might be due to the highly heterogeneous nature of breast cancers. Overall, our analysis reveals that an invasive tumor phenotype in bladder cancer is driven by *E2F1*, *TGFBR*, and *FGFR1*, whereas in case of breast cancer it is driven by *E2F1*, *TGFBR*, and *EGFR*.

Discussion

To improve the treatment outcomes of patients who develop metastases and drug resistance, a mechanistic understanding of the determinants of these processes is indispensable. A large number of clinical studies have recently been published which identified E2F1 as a key transcription factor that switches duties from a tumor suppressor to a driver of metastasis[11, 14, 27]. To understand how E2F1 switches its duties, we derived a comprehensive interaction map (Fig. 1) that includes state of the art knowledge on transcriptional, post-transcriptional, and protein–protein interactions around the E2F family. The map contains 879 nodes and 2278 interactions of gene regulation and signaling processes associated with the E2Fs, thereby providing ways not only for the detailed elucidation of E2F regulation but also for tracing their connections to other cancer-related pathways. Recently, the idea of analyzing regulatory maps by integrating multi-omics data for the detection of disease driving molecules and the identification of therapeutic targets has gained momentum[28–32]. In the context of our work, Calzone et al.[28] reconstructed a comprehensive map of the E2F transcription factor family. Their work focused on the differing roles of E2F family members in the cell cycle, reflecting the complex interplay between the E2Fs, RB1, its homologs RBL1 and 2, the cyclins/cyclin-dependent kinases, and cell cycle arresters. In contrast, our map sets a main focus on the newly discovered role of activating members of the E2F family (E2F1-3) in cancer development and progression, with an emphasis on pro-apoptotic and anti-apoptotic (survival), angiogenic as well as functions relevant for EMT. We included additional key players connected to E2F1 directly or through its neighbors along with a post-transcriptional layer of microRNAs in the context of cancer. Interestingly, the majority of the components in the map by Calzone and coworkers are included in our map (see Supplementary Fig. 8 for further details).

The underlying idea of our approach is that the structural and data-driven analysis of this map allows the identification of key functional modules, here named core regulatory networks, composed of regulatory motifs and critical molecular interactions that drive given cancer phenotypes. To this end, we have integrated coherent workflow tools coming from data analysis, bioinformatics, and mathematical modeling. Precisely, and to the best of our knowledge, we do not find in the literature a precedent of combining network-based high-throughput data analysis, network reduction, and Boolean modeling. Existing work either focuses on network-based analysis[5, 33] or Boolean network construction and simulation[34, 35]. Further, our methodology includes an innovative element in terms of network reduction,

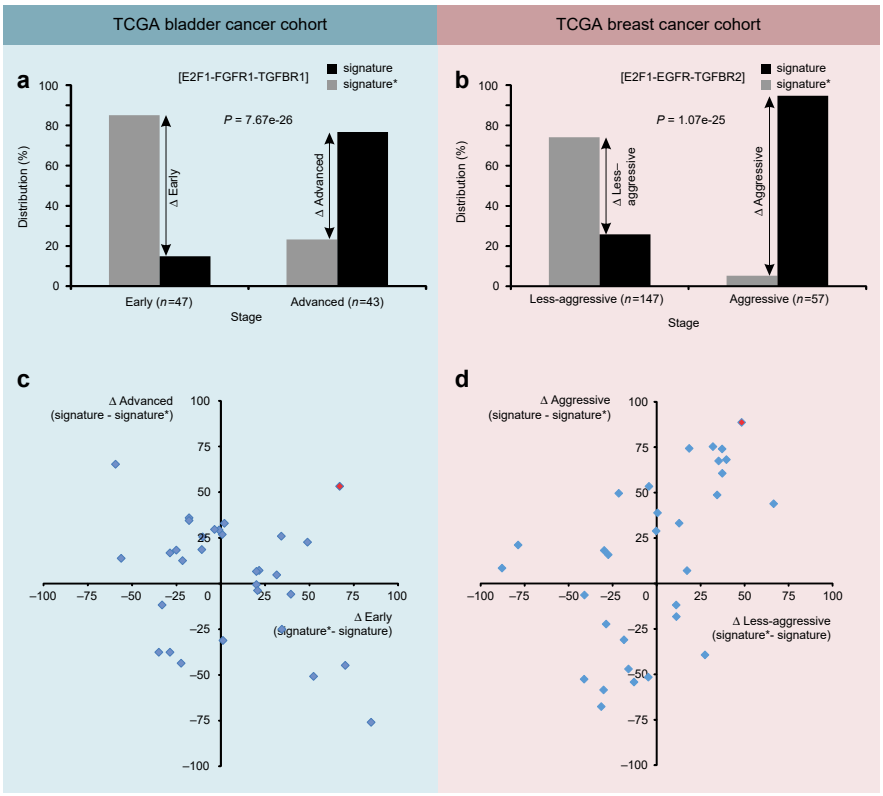


Fig. 8 Validation of molecular signatures in **a** TCGA bladder and **b** TCGA breast cancer cohorts. In the bladder cancer cohort, we first filtered patients diagnosed with early (stage II) and advanced (stage IV) stages. Further in both stages, we identified subsets of patients where expression of genes of the molecular signature (i.e., *E2F1*, *TGFBR1*, and *FGFR1*) was above (signature) or below (signature*) the respective mean expression value. In case of the breast cancer cohort, we classified patients into aggressive (basal & Her2) vs. less-aggressive (luminal A and B) molecular subtypes based on PAM50 stages provided. Afterwards we identified those subsets of patients where expression of genes of the molecular signature (i.e., *E2F1*, *TGFBR2*, and *EGFR*) was above (signature) or below (signature*) the respective mean expression value. We calculated distribution difference (%) of patients within pathological stages in both cancer types. Similarly, distribution differences for 30 random signatures vs. signatures* derived from **c** bladder and **d** breast cancer regulatory cores are shown as *scatter plots*. Signatures in the *upper right* or *lower left corners* nicely distinguish patient phenotypes. The molecular signatures predicted from our workflow are shown in *red*. *P*-value is calculated using Pearson's χ^2 test

namely the use of an algorithm employing multi-objective optimization concepts to rank and select key regulatory motifs, based on network topology features and expression profiles. As far as we know, this has not been explored before in the context of cancer.

The analysis of topological properties can provide important information about the cues that have a significant impact on the dynamics of the network[7, 36]. In our workflow, we analyzed the network properties node degree and betweenness centrality. Nodes with a high degree, often named hubs, are known for their importance in network organization and very often are transcription factors having a central role in orchestrating cell differentiation programs[37, 38]. Nodes with high betweenness centrality serve as gate keepers in the communication between different components of a network[39]. As our network was constructed with all possible regulatory processes around E2F1, one can expect that values of the topological properties for some of the nodes are relatively higher than for others, a potential bias compensated by also considering non-topological properties. Precisely, we assigned different weights to the genes in the network according to their known relatedness to relevant cancer-associated pathways. Furthermore, we used the gene expression fold-change in cell lines reflecting the cancer phenotypes investigated, thereby providing a data-driven approach to make the core network cancer-type and context-specific.

As sets of genes involved in certain phenotypes are highly interconnected and regulate each other through coherent and incoherent regulatory loops (motifs) from different pathways, analysis of these can provide key insights into the structure and dynamics of the network[40] followed by identification of disease biomarkers[41, 42]. In intracellular regulatory networks, feedback loops provide stability and robustness against intrinsic and extrinsic noise, homeostasis or even all-or-nothing patterns of activation[6, 40, 43]. Very often, these network motifs are disrupted or abnormally regulated in cancer and therefore, the analysis of their differential regulation provides important information on the emergence of cancer phenotypes. However, the identification of important feedback loops in a highly connected network is a methodological challenge. From the E2F1 interaction map, we identified a large set of three-nodes feedback loops responsible for non-intuitive behavior of the system (Supplementary Data 1). We considered three-node feedback loops due to the fact that larger sized network loops are typically composed of one or more three-node loops[43].

Motif identification-based methods have been previously used to recognize key network regulators. For example, Zhang et al.[41]

ranked network motifs using gene expression data to detect breast cancer susceptible genes and Koschützki et al.[44] used motifs with various network topological parameters to identify important nodes in a biochemical network. We here introduced a new motif ranking scheme using a weighted multi-objective function that integrates topological (e.g., node degree and betweenness centrality) and non-topological (e.g., gene expression, gene prioritization) properties. Topological properties account for the structural importance of the nodes, and non-topological properties for their cancer-type and context-specific relevance. We used multiple weighting scenarios in the multi-objective function to provide motif ranking as unbiased as possible regarding the properties assessed in the function. Our proposed multi-objective function and ranking scheme can easily be extended to add new information layers in the workflow. Thus, we think that the method proposed can be used for investigating other cancer networks besides those focused on the E2F family discussed here.

Using our proposed multi-objective function for motif prioritization, we have ranked feedback loops identified in the E2F1 interaction map (Supplementary Data 1). To understand the combined effect of top-ranked feedback loops on the regulation of EMT processes in bladder and breast cancer, we interconnected them to generate tumor-specific regulatory core networks (Fig. 3), which we believe are the main drivers of the network dynamics. Furthermore, we analyzed the core regulatory network for the identification of molecular signatures driving the invasive phenotype by using logic-based model simulations[45]. We used Boolean logic for the input and regulatory layers, whereas multi-valued logic representation for the phenotypical output of the network. Multi-valued logic allows us to model several activity levels of the phenotype, which helps assessing the aggregated effect of various network components on the phenotype[29]. However, the use of multi-valued logic increases the complexity of the model, and therefore, we apply it only to the phenotypical output.

The in silico simulation results indicated that high levels of *E2F1–TGFBR1–FGFR1* in bladder cancer and *E2F1–TGFBR2–EGFR* in breast cancer constitute a molecular signature that represent the most aggressive phenotype. Surprisingly, the other receptors that are part of the input layers in the models had no effect on the EMT process in our simulations. Our results are in agreement with previous experimental findings in bladder cancer studies where high levels of *E2F1*[14], *TGFBR1*[46], and *FGFR1*[47] were independently associated with tumor invasion. Similarly, in breast cancer studies, high expression of *E2F1*[48], *TGFBR2*[49], and *EGFR*[50] was separately observed to regulate invasive tumor phenotypes. Furthermore, we confirmed the role of predicted signatures on the regulation of tumor invasion using bladder and breast cancer patient survival data from independent studies in Figs 7 and 8. For all our predicted signatures high expression of the constituent molecules mapped to low patient survival and vice versa. These correlations prove that our approach is successful in identifying tumor-specific molecular signatures regulating EMT processes and driving invasive phenotypes.

By applying an expression signature from highly invasive bladder cancer cells and the respective regulatory core, our model predicted resistance to EGFR inhibition in cells overexpressing *E2F1*. Indeed, treatment of UM-UC-3 with EGFR inhibitors resulted only in a marginal reduction of cell invasion. Likewise, exposure of RT-4 to EGF did not have any observable effect. These results are not intuitive as we showed that UM-UC-3 express *EGFR* (Fig. 5b). On the basis of these results, we propose that EGFR-targeted therapies might be ineffective in muscle invasive bladder cancer exhibiting elevated levels of *E2F1*. In this regard, it is interesting to note that the majority of completed clinical trials using inhibitors of EGFR family RTKs do not show an added benefit over standard of care chemotherapy in an adjuvant or second line setting[51]. These studies show that treatment with chemotherapeutic agents rendered patients with muscle invasive bladder cancer also resistant to EGFR inhibitors for hitherto unknown reasons. Expression of *EGFR* is, like in our model system UM-UC-3, detected in urothelial carcinomas of the bladder. However, the absence of activating *EGFR* mutations at exons 19 to 21 in a number of bladder cancer specimens has been demonstrated and may contribute to EGFR inhibitor resistance[52]. In contrast, aberration of *FGFR1* is a frequent event in bladder cancer contributing to rapid disease progression[53]. In pre-clinical models, the presence of *FGFR1* genetic alterations confers sensitivity to FGFR inhibitors[54]. Respective clinical trials are ongoing. However, molecular pre-selection (e.g., *FGFR1* amplifications present or not) is the primary challenge for the development of FGFR inhibitors. In this context, our model could be a valuable tool to select for patients who potentially benefit from an anti-FGFR therapy through application of model-based signatures for patient classification such as determination of *E2F1* and *TGFBR1* oncogene expression. With regard to our data in bladder cancer cell lines, we observed an opposite situation in breast cancer cells. Although EGFR activation in less-invasive MCF-7 or inhibition in metastatic MDA-MB231 showed a clear regulation of cell invasion, FGF treatment, and exposure to FGFR1 inhibitors, respectively, did not substantially alter cell invasion. However, the FGFR signaling pathway regulates normal mammary gland development and *FGFR1* overexpression has been associated with breast cancer progression[55]. A recent study applied a pre-clinical mammary tumor model to show that FGFR1 inhibitor treatment leads to initial rapid regression which is, nevertheless, finally followed by tumor recurrence[56]. Intriguingly, recurrent tumor tissues revealed elevated levels of activated EGFR compensating for FGFR1 inhibition. It remains to be seen whether or to which extent available FGFR1 inhibitors can improve treatment of metastatic breast cancer[57].

Overall, model-based treatment recommendations of E2F1-driven tumor diseases such as advanced bladder or breast cancer, have the potential to support cohort-specific treatment of patients to avoid therapy resistance and cope with aggressive cancers. Finally, we used the map to investigate E2F1-associated malignant progression in two tumor entities, but the map and the workflow proposed can also be applied to other cancer types in which E2F1 might have a similar role, as well as to uncover other phenotypes related to this transcription factor like chemoresistance or angiogenesis.

Methods

Data retrieval and construction of the E2F1 interaction map. For the construction of the E2F1 interaction map, we derived, curated, and incorporated information from the literature and publically available databases as well as E2F1 cofactors recently identified by our group. More specifically, we retrieved data on protein–protein interactions from STRING[58] (v9.1) and HPRD[59] (release 9). Transcription factors and their target genes were retrieved from databases[60] and relevant literature. Moreover, we included microRNA-target interactions, which were extracted from the miRTarBase[61] database (release 4.5). Transcription factors of microRNAs have been extracted from the TransmiR[62] database (v1.2). In addition, we searched PubMed for publications about validated E2F transcription factors, their post-translational modifications, molecular interactions, and connections to certain diseases, especially to cancer. Furthermore, we manually curated interactions (assigned directions to the interactions, i.e., activation/inhibition and relevant references) that were retrieved from mostly automatically generated databases like STRING, KEGG or the EMBL-EBI search engine PSICQUIC[63] and the text mining tool iHop[64] to search for more interactions described in the literature.

The E2F1 interaction map was built with the process diagram editor CellDesigner[65] (v4.3) and visualized as a SBGN (Systems Biology Graphical Notation) compliant diagram[66]. We created different regulatory and functional

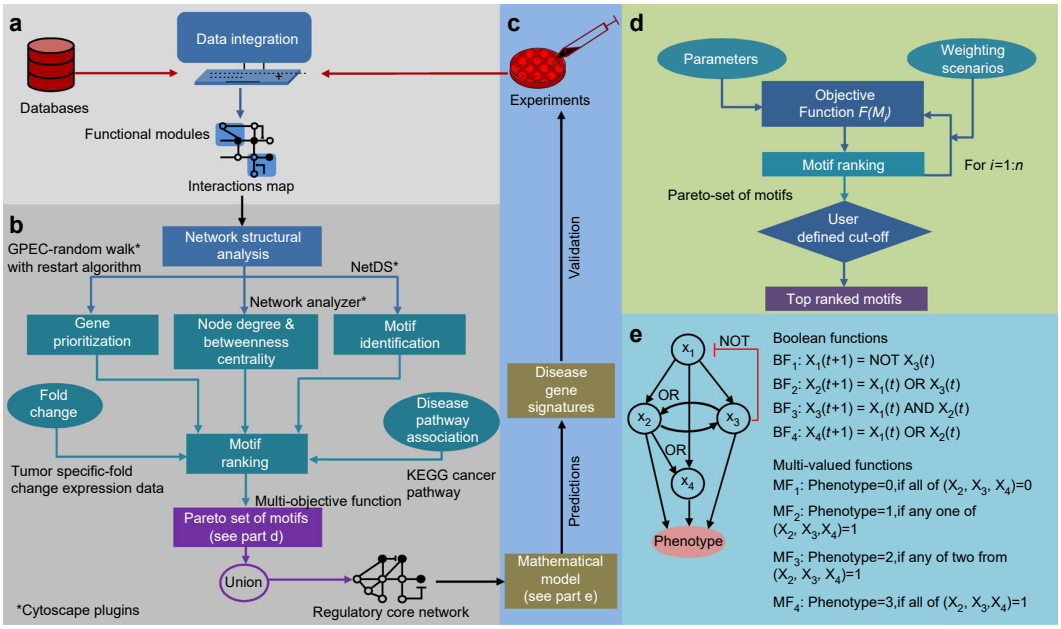


Fig. 9 Methodology for the construction of an interaction map, derivation of the regulatory core, and prediction of molecular signatures using model simulations. **a** Data on molecular interactions are extracted from public databases and the literature to reconstruct the regulatory network underlying the cancer phenotype investigated. **b** Topological and non-topological network properties are used to identify important network motifs. A regulatory core is derived by merging the top-ranked motifs. **c** In silico simulations of the regulatory core help in the detection of molecular signatures that are then subject to experimental validation. **d** An algorithm for motif prioritization. Motifs can be prioritized based on the objective function represented in Eq. (1). This objective function contains parameters accounting for node properties, the expression profile of nodes in relevant cancer cell lines and their relatedness to disease pathways. Weights can be assigned by giving importance to a particular parameter in a user defined manner or iteratively to determine the Pareto sets of motifs. From these sets, top-ranked motifs can be identified based on user defined cutoff. **e** Logic-based representation of biochemical network. *Left*: A toy model of biological network consisting of four nodes ($X_{1\text{-}4}$), and interactions among them regulate certain phenotype. *Right*: Derivation of Boolean functions (*BF*) and multi-valued logic functions (*MF*) for the toy model

compartments based on the role of molecules in influencing E2F1 activities and determining cell fate as follows: (i) Extra-/intracellular receptor signaling: This compartment contains cellular receptors and ligands with known crosstalk to E2F family pathways (e.g., FLT4, PDGFRB) and additionally, cancer-relevant receptors that feed downstream processes like the cell cycle (e.g., growth factor receptors via the MAPK pathway), apoptosis (like TNF/TRAIL receptors) or survival (e.g., IGF1R for Ras and AKT signaling); (ii) E2F1 modifications: To separate the post-translational modifications influencing its transactivation potential of target genes from the effects that cofactors have on the latter, we consider post-translational modifiers of E2F1, containing factors that regulate E2F1 protein stability, for example, upon phosphorylation after DNA damage, including kinases (ATM/R), acetylases (EP300), deacetylases (HDAC1) or methyltransferases (DNMT1), and regulators of E2F1 transcriptional activity, containing protein-binding partners that regulate the affinity or specificity of E2F1 to its DNA targets (like epigenetic modifiers interacting with histones and recruiting E2F1 to target promoters, e.g., ATAD2); (iii) Cell cycle: This compartment includes cyclins, CDKs, and MYC that regulate the cell cycle upon extracellular stimulation by growth factors. It also harbors the regulation of expression and activity (e.g., inhibition by RB1) of E2F1-3 as central cell cycle regulators; (iv) Quiescence: This compartment encloses factors involved in arresting the cell cycle (CDK inhibitors such as p21, p14/ARF), as well as complexes that silence E2F1 targets during G0/G1/G2-phase or quiescence (e.g., SWI/SNF complex or the DREAM complex).

DNA repair: DNA damage sensing and repair factors (e.g., BRCA1, TOPBP1) are summarized in this compartment; (v) Apoptosis: Factors inducing and executing the cellular apoptotic program like the TP53 family (TP53/63/73), pro-apoptotic BCL-family inhibitors (BID, BAX), or Caspases; (vi) Survival: Factors suppressing pro-apoptotic signaling like the anti-apoptotic BCL-family members BCL2, MDM2, XIAP proteins, MDR-transporters, and the MYB family; (vii) EMT/invasion/angiogenesis: Besides currently known players in EMT-like ZEB1/2, SNAI1/2, VIM, CDH1/2, this compartment contains the CTNNB1 regulation and its influence on CDH1 as well as angiogenesis and extracellular matrix regulating factors such as HIF1A, MMPs, and L1CAM; (viii) Metabolism: Ion channels (KCNH11) and metabolic enzymes (PFKFB2, MAT2A) are summarized in this compartment as there is recent evidence indicating a regulatory role for E2Fs in the cellular metabolome.

Next, we incorporated into our map detailed text-based annotations including HUGO name, HGNC ID, Entrez ID, and UniProt ID. We used complex formation and dissociation information to accommodate activities of protein monomers, dimers, and oligomers (Supplementary Data 3). By assigning web links to the annotations in our map, we turned it into an interactive resource.

The web version of the E2F1 map was constructed using NaviCell[67]. In addition, the map was translated to Cytoscape[68], a format suitable for the use of network analysis tools. For the sake of improving the visualization, the CellDesigner map (Fig. 1) pools proteins of the same family into general terms, whereas the Cytoscape version (Supplementary Fig. 1) contains all respective family members and interactions among them (Supplementary Data 4), thereby allowing the use of existing tools for data integration and network analysis. To assure the accuracy of the network, we randomly selected ~10% of the interactions and asked independent domain experts to cross-validate them. Over 98% of the interactions were derived correctly.

Identification of context-specific regulatory core network. We developed a novel method for the identification of the tumor entity-specific core of large regulatory networks, which we understand as a subnetwork that is responsible for critical systems dynamics and driver of a tumor-specific phenotype. Furthermore, we propose a mathematical modeling-based approach for the prediction of molecular signatures, i.e., sets of network-derived diagnostic/prognostic biomarkers. For an illustration of the workflow for network reconstruction, identification of the regulatory core, and prediction of molecular signatures see Fig. 9.

The method for the identification of the regulatory core and molecular signatures involves the following steps: (i) Network analysis: The purpose is to identify the network structure and node properties, which help in the identification of important nodes and network motifs. Topological and node properties were determined using the Cytoscape plugin NetworkAnalyzer[69]. In particular, we calculated for each node the degree and betweenness centrality, and for the network the clustering coefficient, diameter, radius, characteristic path length, and average number of neighbors to understand the overall organization of the network (Supplementary Table 1; Supplementary Fig. 2). (ii) Network motif identification: Feedback loops from the E2F1 interaction map were identified using Cytoscape plugin NetDS[26] (v3.0). For the identification of feedback loops, we set the loop length to three nodes (see "Discussion" section for detail). (iii) Motif ranking: To identify the most important motifs with respect to the relevance for the disease

Table 1 The effect of E2F1 and receptor molecules in relevant combinations on the EMT phenotype in bladder and breast cancer model

(a) Bladder cancer

E2F1	TGFBR1	FGFR1	EGFR	CXCR1	RARA	EMT
0	0	0	0/1	0/1	0/1	0
0	0	1	0/1	0/1	0/1	1
0	1	0	0/1	0/1	0/1	1
0	1	1	0/1	0/1	0/1	2
1	0	0	0/1	0/1	0/1	1
1	0	1	0/1	0/1	0/1	2
1	1	0	0/1	0/1	0/1	2
1	1	1	0/1	0/1	0/1	3

(b) Breast cancer

E2F1	TGFBR2	EGFR	HMMR	THRB	IL1R1	RARA	EMT
0	0	0	0/1	0/1	0/1	0/1	0
0	0	1	0/1	0/1	0/1	0/1	1
0	1	0	0/1	0/1	0/1	0/1	1
0	1	1	0/1	0/1	0/1	0/1	2
1	0	0	0/1	0/1	0/1	0/1	1
1	0	1	0/1	0/1	0/1	0/1	2
1	1	0	0/1	0/1	0/1	0/1	2
1	1	1	0/1	0/1	0/1	0/1	3

Active state of the molecule is represented by '1' and the inactive state as '0'. The phenotype output (EMT) can take four ordinal levels ranging from '0' (non- invasive) to '3' (highly invasive). Table (a) is the summary of 64 in silico simulations of bladder cancer. Each row represents the result of eight simulations where for the given Boolean state of E2F1, TGFBR1, and FGFR1, all eight combinations of EGFR, CXCR1, and RARA results in the same phenotypical output. Table (b) is the summary of 128 in silico simulations of breast cancer. Each row represents the result of 16 simulations where for the given Boolean state of E2F1, TGFBR2, and EGFR, all 16 combinations of HMMR, THRB, IL1R1, and RARA results in the same phenotypical output

phenotype under investigation, we developed a novel motif ranking scheme (Fig. 9d). The scheme is based on: (a) Topological properties, node degree and betweenness centrality (see "Discussion" section for detail); (b) the involvement of the motif constituents in KEGG's 'Pathways in cancer' pathway (KEGG: hsa05200); (c) the gene prioritization score from the Cytoscape plugin GPEC[25]; and (d) tumor-specific gene expression fold-changes from non-invasive to invasive phenotypes. We used the ArrayExpress database (ArrayExpress accession number: E-MTAB-2706)[70] to find suitable gene expression data in non-invasive and invasive bladder and breast cancer cell lines. In particular, we used RT-4 as non-invasive and UM-UC-3 as invasive cell lines in bladder cancer, whereas MCF-7 as non-invasive and MDA-MB231 as invasive cell lines in breast cancer. Differential expression analysis was performed using the DEseq R-package with *method = 'bind' and fitType = 'local'* (v1.22.1)[71]. Furthermore, we calculated the absolute average fold-change for a motif based on the change in expression values of each node in non-invasive to invasive phenotype. To rank the network motifs considering all these structural and biomedical criteria, we derived the weighted multi-objective function[72] in Eq. (1).

$$S_{ij} = \frac{w_{1j}}{2}\cdot\frac{\langle \mathrm{ND}\rangle_i}{\max(\mathrm{ND})} + \frac{w_{1j}}{2}\cdot\frac{\langle \mathrm{BC}\rangle_i}{\max(\mathrm{BC})} + w_{2j}\cdot\frac{\langle \mathrm{DP}\rangle_i}{\max(\mathrm{DP})} + w_{3j}\cdot\frac{\langle \mathrm{GP}\rangle_i}{\max(\mathrm{GP})} + w_{4j}\cdot\frac{\langle |\mathrm{FC}|\rangle_i}{\max(|\mathrm{FC}|)} \quad (1)$$

Here, S_{ij} is the ranking score of each motif ($i = 1\ldots n$) in different weighting scenarios ($j = 1\ldots13$) as given in Supplementary Table 2. w_{1j} to w_{4j} are weighting factors pounding the importance of the chosen properties, $\langle \mathrm{ND}\rangle_i$: average node degree, $\langle \mathrm{BC}\rangle_i$: average betweenness centrality, $\langle \mathrm{DP}\rangle_i$: number of nodes in a motif involved in disease pathways, $\langle \mathrm{GP}\rangle_i$: average gene prioritization score, and $\langle |\mathrm{FC}|\rangle_i$: average absolute expression fold-change of a motif i.

To give equal importance to each property, the function is normalized to the maximum property value in all the network motifs identified (e.g., max(BC)). In order to not over-emphasize topological properties in motif prioritization, we assigned half of the weighting factor to $\langle \mathrm{ND}\rangle$ and $\langle \mathrm{BC}\rangle$. To generate a ranking of the motifs, we computed the value of the S_{ij} function for every motif i identified. The function proposed is intrinsically multi-objective and may generate a different ranking for same motif depending on the sets of values chosen for the weighting factors (Supplementary Methods). To approximate the Pareto set of all non-dominated motif rankings, we iteratively modified the values of the weighting factors, computed the S_{ij} function for every motif and ranked them according to their S_{ij} values (Supplementary Data 1). Next, we selected the top 10 motifs from each weighting scenario for further analysis (Fig. 9d). (iv) Derivation of the regulatory core: The core regulatory network is obtained by merging the sets of the top-ranked motifs identified using Cytoscape plugin NetDS (v3.0). In case there are disjoint sub-networks in a regulatory core, we connect those using direct interactions between the nodes taken from the complete regulatory network. (v) Prediction of molecular signatures using in silico simulations: We consider a disease gene signature as a group of molecular entities in the regulatory core, which upon perturbation have a significant impact on the disease phenotypes. To identify molecular signatures, the regulatory core is translated into a logic-based model[29, 73] and steady state analysis is performed using the software tool CellNetAnalyzer[74].

We developed logic-based models of the regulatory cores and carried out in silico perturbation experiments. To this end and upon the selection of the relevant network motifs, we established the Boolean rules based on the network structure and the inspection of the available literature about the interactions. The obtained model contains three layers: (i) An input layer, (ii) A regulatory layer, and (iii) An output layer representing the phenotype. In the Boolean-like logic models, nodes $X = (X_{1\ldots n})$ of a network correspond to the Boolean variables that can have values either 1 or 0, and edges define the type of interactions (e.g., activation or inhibition) that can be represented by Boolean gates (ACTIVE, NOT, OR, and AND). In Boolean models, the future state ($t + 1$) of a node is a Boolean function (BF) of the current state (t) of all the nodes regulating it, i.e., $X_i(t + 1) = \mathrm{BF}(X_1(t), X_2(t),\ldots, X_n(t))$ (Fig. 9e). We derived Boolean functions for signals originating from the input layer and their propagation through nodes constituting the regulatory layer based on network structure using Boolean gates (Supplementary Data 2).

In addition, we used qualitative information based on expression data to approximate activation levels[75]. For example, in our bladder cancer model, *CDH1* is inhibited by multiple molecules including SNAI1, SNAI2, TWIST1, SRC, *miR-25*, and MDM2. In case of bladder cancer, the expressions of *SNAI2*, *SRC*, *miR-25*, and *MDM2* were downregulated, whereas *SNAI1* and *TWIST1* were upregulated in the invasive cell line. It is well established that *CDH1* is downregulated in the invasive phenotype, therefore, we consider SNAI1 and TWIST1 more relevant for the regulation of *CDH1* than others (for Boolean rules, see the Supplementary Data 2). Further, we derived multi-valued logic functions that represent the EMT phenotype in four ordinal levels (from 0, accounting for inactive EMT, to 3, accounting for full activation) based on the sum of Boolean states of factor 1: [SMAD2/3/4, SNAI1, ZEB1, TWIST1], factor 2: CDH1, and factor 3: FGFR1, with the following structure:

$$\mathrm{EMT} = [(\mathrm{SMAD2/3/4\ AND\ SNAI1})\ \mathrm{OR}\ (\mathrm{ZEB1\ AND\ TWIST1})] + (\mathrm{NOT\ CDH1}) + \mathrm{FGFR1}$$

The motivation to select these factors as drivers of EMT was due to the fact that CDH1 is a widely accepted hallmark of EMT together with SMAD2/3/4, SNAI1, ZEB1, and TWIST1[24]. We also considered receptor proteins as decisive factors determining EMT phenotype whether they are present in the regulatory core, highly overexpressed in the invasive phenotype and not connected to any of the EMT markers (e.g., FGFR1 in bladder cancer). We validated our methodology for predicting molecular signatures driving a specific phenotype by using an independent TGFB1 signaling network developed by Steinway et al.[34] (Supplementary Figs 3 and 4, Supplementary Table 3, and Supplementary Data 6)

Cell culture and treatment. All cell lines were purchased from ATCC (Rockville, MD, USA) and kindly provided by Dr S. Füssel, Urology Laboratory, University of Dresden (RT-4, UM-UC-3, HT1197, J82, T24, SW1710, and VM-CUB1 bladder cancer cell lines) and by the Department of Obstetrics and Gynecology, University of Rostock (MCF-7, MDA-MB231, BT549, and T47D breast cancer lines). Cells were maintained at 37 °C and 5% CO_2 in Dulbecco's modified Eagle's medium (high glucose, 4.5 g/l) containing 2 mM L-glutamine, 1 mM sodium pyruvate, supplemented with 10% FCS, 0.1 mM non-essential amino acids, 50 U/ml Penicillin and 50 µg/ml Streptomycin. Breast cancer BT549 cells were grown in RPMI medium with the same supplements. RT-4 and MCF-7 cells were incubated with growth factors (EGF, FGF2, TGFB1, from R&D Systems, at 10 ng/ml) for 48 h prior to experiments. UM-UC-3 and MDA-MB231 cells were treated with inhibitors (EGFR inhibitor Tyrphostin AG 1478, FGFR1 inhibitor PD161570, TGFBR inhibitor SB431542, from Santa Cruz, at concentrations ranging from 200 nM to 3 µM) for 24 h prior to Boyden chamber assays. We used non-invasive RT-4 bladder and MCF-7 breast and invasive UM-UC-3 bladder and MDA-MB231 breast cancer cell lines to model the EMT transitions. All cell lines were tested for mycoplasma contamination prior to the experiments according to manufacturer's instructions (Venor GeM Classic, Minerva Biolabs).

Adenoviral transduction. RT-4/MCF-7 and UM-UC-3/MDA-MB231 cells were seeded into cell culture plates and transduced with Ad.ER-E2F1 (MOI 5) and Ad.sh.E2F1/Ad.sh.control (MOI 10) adenoviral vectors, respectively. After 24 h the Ad.ER-E2F1 transduced cells were treated with 4-hydroxy-tamoxifen (4-OHT, 0.02 µM) or 70% ethanol as control.

Lentiviral transduction. For production of lentiviruses expressing shRNA against SMAD3 (shSMAD3, TRCN0000330055), NFKB1 (shNFKB1, TRCN0000006517), SRC (shSRC, TRCN0000038149), FN1 (shFN1, TRCN0000286357) or scrambled shRNA (shscr), Mission shRNA plasmids (Sigma) were used. VSV-G enveloped pseudotyped lentiviral vectors were generated by cotransfection of HEK293T cells with plasmids pMD2.G and psPAX2 (Addgene) using calcium phosphate.

Invasion assays. For Boyden chamber assay (growth factor pretreated) cells were seeded on an 8-µm PET membrane (BD BioCoat™ BD Matrigel™ Invasion Chamber, 6-well) covered with BD Matrigel™ Basement Membrane Matrix (BD Bioscience). Cell invasion was triggered by a concentration gradient of FCS (2% vs. 30%) between insert and well. After 36–48 h cells on the upper membrane surface were removed, whereas those on the lower surface were stained with DAPI and documented by fluorescence microscopy. Migrated cells were counted using ImageJ software (http://imagej.nih.gov/ij/). For UM-UC-3 and MDA-MB231, pretreated cells were seeded into Boyden chambers, supplied with the according amount of inhibitor and incubated for 36 h. For RT-4 and MCF-7, cells were pretreated with 10 ng/ml of each growth factor (alone or in combination) prior to Boyden chamber assay. 24 h after transduction with adenoviral vector (Ad.ER-E2F1), cells were seeded into Boyden chambers and covered with growth factor reduced BD Matrigel™ Basement Membrane Matrix (BD Bioscience) containing growth factors and 4-OHT or EtOH.

Polymerase chain reaction. For semiquantitative PCR, 1 µg of RNA was reverse transcribed using First Strand cDNA Synthesis Kit (Thermo Scientific). cDNA was added to Thermo Scientific PCR Master Mix and amplified with gene-specific primers. Actin was used as loading control. The following primer sequences were used:

EGFR Fwd: AACTGTGAGGTGGTCCTTGG, Rev: GGAATTCGCTCCAC TGTGTT,

FGFR1 Fwd: ACCACCGACAAAGAGATGGA, Rev: GCCCCTGTGCAATA GATGAT,

TGFBR1 Fwd: TTGCTCCAAACCACAGAGTG, Rev: TGAATTCCACCAA TGGAACA,

TGFBR2 Fwd: CTGGTGCTCTGGGAAATGAC, Rev: CAGAAGCTGGGAA TTTCTGG.

CDH1 Fwd: GCTTTGACGCCGAGAGCTACA, Rev: TCCCAGGCGTAGACC AAGAAA

VIM Fwd: CTCCCTGAACCTGAGGGAAAC, Rev: TTGCGCTCCTGAA AAACTGC

GAPDH Fwd: CACCACCCTGTTGCTGTA, Rev: CACAGTCCATGC CATCAC.

Western blots. For western blot analysis cells were lysed using RIPA buffer containing PhosSTOP Phosphatase Inhibitor Cocktail (Roche). Protein concentration was determined by Bradford assay (Bio-Rad). Protein samples were separated by SDS-PAGE and transferred to nitrocellulose membranes (Amersham Biosciences). The following antibodies were used: E2F1 (KH-95; 1:500), EGFR (1:250), FGFR1 (Flg C15; 1:1000), TGFBR1 (V22; 1:1000), TGFBR2 (L21; 1:1000), Vimentin (V9; 1:1500), ZEB1 (H-102; 1:1000), Snail (H-130; 1:500), SMAD2/3 (FL-425; 1:1000), and c-SRC (1:500) from Santa Cruz, E-Cadherin (1:1500), and NFKB1 (C22B4; 1:1500) from Cell Signaling; N-Cadherin (610921; 1:1500) and FN1 (1:1000) from BD Bioscience, Actin (Sigma; 1:4000) and their corresponding HRP-conjugated secondary antibodies (Pierce; 1:2000). Detection of HRP activity was performed with the ChemiDoc TouchTM Imaging System (BioRad) using ECL Plus (Amersham) or Super Signal West Femto (Thermo Scientific) Western Blotting Detection Reagents (GE Healthcare). Uncropped pictures of the immunoblots are shown in Supplementary Figs 9, 10, and 11.

Data availability. The authors declare that all data supporting the findings of this study are available within the paper and its Supplementary Information/Data files. The web version of E2F1 interaction map is available at https://navicell.curie.fr/pages/maps_e2f1.html. The interactions in the Cytoscape version of the map can be found in the Supplementary Data 4. Both the CellDesigner and Cytoscape versions of the maps in xml format, MATLAB code, Boolean models of bladder and breast cancer regulatory core can also be downloaded from https://sourceforge.net/projects/e2f1map/files. Equations used for logic-based model of regulatory cores are given in Supplementary Data 2.

Received: 10 June 2016 Accepted: 15 June 2017
Published online: 04 August 2017

References

1. Kreeger, P. K. & Lauffenburger, D. A. Cancer systems biology: a network modeling perspective. *Carcinogenesis* **31**, 2–8 (2010).
2. Krogan, N. J., Lippman, S., Agard, D. A., Ashworth, A. & Ideker, T. The cancer cell map initiative: defining the hallmark networks of cancer. *Mol. Cell* **58**, 690–698 (2015).
3. Wachi, S., Yoneda, K. & Wu, R. Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics* **21**, 4205–4208 (2005).
4. Jonsson, P. F. & Bates, P. A. Global topological features of cancer proteins in the human interactome. *Bioinformatics* **22**, 2291–2297 (2006).
5. Hofree, M., Shen, J. P., Carter, H., Gross, A. & Ideker, T. Network-based stratification of tumor mutations. *Nat. Methods* **10**, 1108–1115 (2013).
6. Alon, U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461 (2007).
7. Kitano, H. A robustness-based approach to systems-oriented drug design. *Nat. Rev. Drug Discov.* **6**, 202–210 (2007).
8. Polager, S. & Ginsberg, D. E2F - at the crossroads of life and death. *Trends Cell Biol.* **18**, 528–535 (2008).
9. Alla, V. et al. E2F1 in melanoma progression and metastasis. *J. Natl Cancer Inst.* **102**, 127–133 (2010).
10. Pützer, B. M., Steder, M. & Alla, V. Predicting and preventing melanoma invasiveness: advances in clarifying E2F1 function. *Expert Rev. Anticancer Ther.* **10**, 1707–1720 (2010).
11. Alla, V. et al. E2F1 confers anticancer drug resistance by targeting ABC transporter family members and Bcl-2 via the p73/DNp73-miR-205 circuitry. *Cell Cycle* **11**, 3067–3078 (2012).
12. Engelmann, D. & Pützer, B. M. The dark side of E2F1: in transit beyond apoptosis. *Cancer Res.* **72**, 571–575 (2012).
13. Vera, J. et al. Kinetic modeling-based detection of genetic signatures that provide chemoresistance via the E2F1-p73/DNp73-miR-205 network. *Cancer Res.* **73**, 3511–3524 (2013).
14. Lee, J.-S. et al. Expression signature of E2F1 and its associated genes predict superficial to invasive progression of bladder tumors. *J. Clin. Oncol.* **28**, 2660–2667 (2010).
15. Sharma, A. et al. The retinoblastoma tumor suppressor controls androgen signaling and human prostate cancer progression. *J. Clin. Invest.* **120**, 4478–4492 (2010).
16. Knoll, S., Emmrich, S. & Pützer, B. M. The E2F1-miRNA cancer progression network. *Adv. Exp. Med. Biol.* **774**, 135–147 (2013).
17. Knoll, S. et al. E2F1 induces miR-224/452 expression to drive EMT through TXNIP downregulation. *EMBO Rep.* **15**, 1315–1329 (2014).
18. Gregory, P. A. et al. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat. Cell Biol.* **10**, 593–601 (2008).
19. Andrechek, E. R. HER2/Neu tumorigenesis and metastasis is regulated by E2F activator transcription factors. *Oncogene* **34**, 217–225 (2015).
20. Shackney, S. E., Chowdhury, S. A. & Schwartz, R. A novel subset of human tumors that simultaneously overexpress multiple E2F-responsive genes found in breast, ovarian, and prostate cancers. *Cancer Inform.* **13**, 89–100 (2014).
21. Engelmann, D. et al. E2F1 promotes angiogenesis through the VEGF-C/VEGFR-3 axis in a feedback loop for cooperative induction of PDGF-B. *J. Mol. Cell Biol.* **5**, 391–403 (2013).
22. Desmedt, C. et al. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin. Cancer Res.* **13**, 3207–3214 (2007).
23. Rennhack, J. & Andrechek, E. Conserved E2F mediated metastasis in mouse models of breast cancer and HER2 positive patients. *Oncoscience* **2**, 867–871 (2015).
24. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial–mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* **15**, 178–196 (2014).
25. Le, D.-H. & Kwon, Y.-K. GPEC: a Cytoscape plug-in for random walk-based gene prioritization and biomedical evidence collection. *Comput. Biol. Chem.* **37**, 17–23 (2012).
26. Le, D.-H. & Kwon, Y.-K. NetDS: a Cytoscape plugin to analyze the robustness of dynamics and feedforward/feedback loop structures of biological networks. *Bioinformatics* **27**, 2767–2768 (2011).
27. Meier, C. et al. Association of RHAMM with E2F1 promotes tumour cell extravasation by transcriptional up-regulation of fibronectin. *J. Pathol.* **234**, 351–364 (2014).
28. Calzone, L., Gelay, A., Zinovyev, A., Radvanyi, F. & Barillot, E. A comprehensive modular map of molecular interactions in RB/E2F pathway. *Mol. Syst. Biol.* **4**, 173 (2008).
29. Guebel, D. V., Schmitz, U., Wolkenhauer, O. & Vera, J. Analysis of cell adhesion during early stages of colon cancer based on an extended multi-valued logic approach. *Mol. Biosyst.* **8**, 1230–1242 (2012).
30. Matsuoka, Y. et al. A comprehensive map of the influenza A virus replication cycle. *BMC Syst. Biol.* **7**, 97 (2013).
31. Oda, K., Matsuoka, Y., Funahashi, A. & Kitano, H. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol. Syst. Biol.* **1**, 2005.0010 (2005).
32. Wu, G., Zhu, L., Dent, J. E. & Nardini, C. A comprehensive molecular interaction map for rheumatoid arthritis. *PLoS ONE* **5**, e10137 (2010).
33. Alvarez, M. J. et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* **48**, 838–47 (2016).

34. Steinway, S. N. et al. Network modeling of TGFb signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint sonic hedgehog and Wnt pathway activation. *Cancer Res.* **74**, 5963–5977 (2014).
35. Lu, J. et al. Network modelling reveals the mechanism underlying colitis-associated colon cancer and identifies novel combinatorial anti-cancer targets. *Sci. Rep.* **5**, 14739 (2015).
36. Gursoy, A., Keskin, O. & Nussinov, R. Topological properties of protein interaction networks from a structural perspective. *Biochem. Soc. Trans.* **36**, 1398–1403 (2008).
37. He, X. & Zhang, J. Why do hubs tend to be essential in protein networks? *PLoS Genet.* **2**, e88 (2006).
38. Jeong, H., Mason, S. P., Barabási, A.-L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
39. Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
40. Tyson, J. J. & Novák, B. Functional motifs in biochemical reaction networks. *Annu. Rev. Phys. Chem.* **61**, 219–240 (2010).
41. Zhang, Y., Xuan, J., de Los Reyes, B. G., Clarke, R. & Ressom, H. W. Network motif-based identification of breast cancer susceptibility genes. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2008**, 5696–5699 (2008).
42. Wang, P., Lü, J. & Yu, X. Identification of important nodes in directed biological networks: a network motif approach. *PLoS ONE* **9**, e106132 (2014).
43. Yeger-Lotem, E. et al. Network motifs in integrated cellular networks of transcription–regulation and protein–protein interaction. *Proc. Natl Acad. Sci. USA* **101**, 5934–5939 (2004).
44. Koschützki, D., Schwöbbermeyer, H. & Schreiber, F. Ranking of network elements based on functional substructures. *J. Theor. Biol.* **248**, 471–479 (2007).
45. Eduati, F. et al. Drug resistance mechanisms in colorectal cancer dissected with cell type-specific dynamic logic models. *Cancer Res.* **77**, 3364–3375 (2017).
46. Medici, D., Hay, E. D. & Goodenough, D. A. Cooperation between snail and LEF-1 transcription factors is essential for TGF-beta1-induced epithelial–mesenchymal transition. *Mol. Biol. Cell* **17**, 1871–1879 (2006).
47. Tomlinson, D. C., Baxter, E. W., Loadman, P. M., Hull, M. A. & Knowles, M. A. FGFR1-induced epithelial to mesenchymal transition through MAPK/PLCγ/COX-2-mediated mechanisms. *PLoS ONE* **7**, e38972 (2012).
48. Rizwani, W., Schaal, C., Kunigal, S., Coppola, D. & Chellappan, S. Mammalian lysine histone demethylase KDM2A regulates E2F1-mediated gene transcription in breast cancer cells. *PLoS ONE* **9**, e100888 (2014).
49. Dhasarathy, A., Phadke, D., Mav, D., Shah, R. R. & Wade, P. A. The transcription factors Snail and Slug activate the transforming growth factor-beta signaling pathway in breast cancer. *PLoS ONE* **6**, e26514 (2011).
50. Bhargava, R. et al. EGFR gene amplification in breast cancer: correlation with epidermal growth factor receptor mRNA and protein expression and HER-2 status and absence of EGFR-activating mutations. *Mod. Pathol.* **18**, 1027–1033 (2005).
51. Mooso, B. A. et al. The role of EGFR family inhibitors in muscle invasive bladder cancer: a review of clinical data and molecular evidence. *J. Urol.* **193**, 19–29 (2015).
52. Chaux, A. et al. High epidermal growth factor receptor immunohistochemical expression in urothelial carcinoma of the bladder is not associated with EGFR mutations in exons 19 and 21: a study using formalin-fixed, paraffin-embedded archival tissues. *Hum. Pathol.* **43**, 1590–1595 (2012).
53. Helsten, T. et al. The FGFR landscape in cancer: analysis of 4,853 tumors by next-generation sequencing. *Clin. Cancer Res.* **22**, 259–267 (2016).
54. Guagnano, V. et al. FGFR genetic alterations predict for sensitivity to NVP-BGJ398, a selective pan-FGFR inhibitor. *Cancer Discov* **2**, 1118–1133 (2012).
55. Jang, M. et al. FGFR1 is amplified during the progression of in situ to invasive breast carcinoma. *Breast Cancer Res.* **14**, R115 (2012).
56. Holdman, X. B. et al. Upregulation of EGFR signaling is correlated with tumor stroma remodeling and tumor recurrence in FGFR1-driven breast cancer. *Breast Cancer Res.* **17**, 141 (2015).
57. Katoh, M. & Nakagama, H. FGF receptors: cancer biology and therapeutics. *Med. Res. Rev.* **34**, 280–300 (2014).
58. Franceschini, A. et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41**, D808–15 (2013).
59. Keshava Prasad, T. S. et al. Human protein reference database—2009 update. *Nucleic Acids Res.* **37**, D767–772 (2009).
60. Wingender, E. et al. TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res.* **28**, 316–319 (2000).
61. Hsu, S.-D. et al. miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **39**, D163–D169 (2011).
62. Wang, J., Lu, M., Qiu, C. & Cui, Q. TransmiR: a transcription factor-microRNA regulation database. *Nucleic Acids Res.* **38**, D119–D122 (2010).
63. Aranda, B. et al. PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nat. Methods* **8**, 528–529 (2011).
64. Hoffmann, R. & Valencia, A. A gene network for navigating the literature. *Nat. Genet.* **36**, 664–664 (2004).
65. Funahashi, A., Morohashi, M., Kitano, H. & Tanimura, N. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *Biosilico* **1**, 159–162 (2003).
66. Le Novère, N. et al. The systems biology graphical notation. *Nat. Biotechnol.* **27**, 735–741 (2009).
67. Kuperstein, I. et al. NaviCell: a web-based environment for navigation, curation and maintenance of large molecular interaction maps. *BMC Syst. Biol.* **7**, 100 (2013).
68. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
69. Assenov, Y., Ramírez, F., Schelhorn, S. E. S. E., Lengauer, T. & Albrecht, M. Computing topological parameters of biological networks. *Bioinformatics* **24**, 282–284 (2008).
70. Klijn, C. et al. A comprehensive transcriptional portrait of human cancer cell lines. *Nat. Biotechnol.* **33**, 306–312 (2015).
71. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
72. Vera, J., González-Alcón, C., Marín-Sanguino, A. & Torres, N. Optimization of biochemical systems through mathematical programming: methods and applications. *Comput. Oper. Res.* **37**, 1427–1438 (2010).
73. Khan, F. M. et al. Hybrid modeling of the crosstalk between signaling and transcriptional networks using ordinary differential equations and multi-valued logic. *Biochim. Biophys. Acta Proteins Proteomics* **1844**, 289–298 (2014).
74. Klamt, S., Saez-Rodriguez, J., Lindquist, J. A., Simeoni, L. & Gilles, E. D. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics* **7**, 56 (2006).
75. Terfve, C. et al. CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms. *BMC Syst. Biol.* **6**, 133 (2012).

Acknowledgements

This work was supported by the German Federal Ministry of Education and Research (BMBF) as part of the project eBio:SysMet [0316171 to O.W., B.P., and J.V.], eBio: MelEVIR [031L0073A to J.V. and 031L0073B to O.W.], German Cancer Aid, Dr Mildred Scheel Stiftung [109801 to B.M.P. and D.E.], and Rostock University Medical Faculty for the project Systems Medicine of Cancer Invasion and Metastasis to B.P. and O.W. J.V. is funded by the Erlangen University Hospital (ELAN funds, 14-07-22-1-Vera-González and direct Faculty support). We thank Dr. S. Füssel and Prof. T. Reimer for providing bladder and breast cancer cell lines. We are also grateful to Dr. Rahele Amikhah from Nastaran Center for Cancer Prevention, Iran, Prof. Y. Shukla and Nivedita Singh from Tumor Proteomics lab, CSIR-IITR, India for cross validation of random interactions from our map.

Author contributions

The original idea was developed by B.M.P., J.V., and O.W. The computational workflow was designed by S.K.G., J.V., and U.S. The network reconstruction and annotation was carried out by F.M.K. and S.M. under the guidance of D.E. and J.V. The validation experiments were performed by S.M., A.S., and S.K. under the supervision of B.M.P. The mathematical modeling was performed by F.M.K. under the supervision of J.V. and S.K.G. The analysis of the patient data was performed by S.K.G. and S.M. Biological interpretation of the results was done by B.M.P. and D.E. All authors drafted the manuscript.

Additional information

Supplementary Information accompanies this paper at doi:10.1038/s41467-017-00268-2.

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.

Biochimica et Biophysica Acta 1844 (2014) 289–298

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Hybrid modeling of the crosstalk between signaling and transcriptional networks using ordinary differential equations and multi-valued logic ☆

Faiz M. Khan [a], Ulf Schmitz [a], Svetoslav Nikolov [c], David Engelmann [b], Brigitte M. Pützer [b], Olaf Wolkenhauer [a,d,1], Julio Vera [a,e,*,1]

[a] Department of Systems Biology and Bioinformatics, Institute of Computer Science, University of Rostock, Rostock, Germany
[b] Institute of Experimental Gene Therapy and Cancer Research, Rostock University Medical Center, Rostock, Germany
[c] Bulgarian Academy of Sciences, Institute of Mechanics, Sofia, Bulgaria
[d] Stellenbosch Institute for Advanced Study (STIAS), Wallenberg Research Centre at Stellenbosch University, Stellenbosch, South Africa
[e] Laboratory of Systems Medicine, Department of Dermatology, Faculty of Medicine, University of Erlangen-Nurnberg, Erlangen, Germany

ARTICLE INFO

Article history:
Received 1 March 2013
Received in revised form 9 May 2013
Accepted 10 May 2013
Available online 18 May 2013

Keywords:
Signaling
E2F1
Non-linear motif
Feedback loop
Chemoresistance
Cancer

ABSTRACT

A decade of successful results indicates that systems biology is the appropriate approach to investigate the regulation of complex biochemical networks involving transcriptional and post-transcriptional regulations. It becomes mandatory when dealing with highly interconnected biochemical networks, composed of hundreds of compounds, or when networks are enriched in non-linear motifs like feedback and feedforward loops. An emerging dilemma is to conciliate models of massive networks and the adequate description of non-linear dynamics in a suitable modeling framework. Boolean networks are an ideal representation of massive networks that are humble in terms of computational complexity and data demand. However, they are inappropriate when dealing with nested feedback/feedforward loops, structural motifs common in biochemical networks. On the other hand, models of ordinary differential equations (ODEs) cope well with these loops, but they require enormous amounts of quantitative data for a full characterization of the model. Here we propose hybrid models, composed of ODE and logical sub-modules, as a strategy to handle large scale, non-linear biochemical networks that include transcriptional and post-transcriptional regulations. We illustrate the construction of this kind of models using as example a regulatory network centered on E2F1, a transcription factor involved in cancer. The hybrid modeling approach proposed is a good compromise between quantitative/qualitative accuracy and scalability when considering large biochemical networks with a small highly interconnected core, and module of transcriptionally regulated genes that are not part of critical regulatory loops. This article is part of a Special Issue entitled: Computational Proteomics, Systems Biology & Clinical Implications. Guest Editor: Yudong Cai.

1. Introduction

In cells, biological processes are driven by complex networks integrating genes, transcripts, like mRNAs and microRNAs (miRNAs), proteins and small molecules. Concentrations and activity of these biomolecules are continuously changing in a concerted manner in response to internal and external cell signals. Those changes are regulated by multiple nested biological circuits that may contain feedback and feedforward loops. These non-linear biological circuits give rise to important cell features like robustness against noise, adaptation despite environmental changes or hysteretic responses with multistability [1]. In the last decade it has been found that some biochemical networks are extremely large and complex: they are commonly integrated by hundreds of compounds and are enriched in non-linear motifs like feedback and feedforward loops. Under these conditions, the analysis of biochemical networks evades human intuition. However, mathematical modeling is an appropriate tool to help in understanding those networks [2–4]. Depending on the features of the biochemical network under consideration and the available experimental data, a variety of frameworks for mathematical modeling are available. These range from simple and abstract approaches to biologically detailed ones, from deterministic to probabilistic or from spatio-temporal continuous to discrete ones [5].

Models in ordinary differential equations (ODEs), accounting for the variation in time of variables representing the concentration or activation state of biochemical molecules, have been used to describe biological networks for decades [6]. These ODE models allow the quantitative simulation of the concentration changes or activity profiles of proteins, genes, RNAs and other molecules over time. These features make it possible to compare the model predictions with most of the

☆ This article is part of a Special Issue entitled: Computational Proteomics, Systems Biology & Clinical Implications. Guest Editor: Yudong Cai.

* Corresponding author at: Department of Systems Biology and Bioinformatics, Institute of Computer Science, University of Rostock, Rostock, Germany. Tel.: +49 16095759623.
E-mail address: julio.vera.g@gmail.com (J. Vera).

[1] Equal contributors.

1570-9639/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bbapap.2013.05.007

standard experimental measurements. When biological compartments are considered, they can also account for a qualitative description of spatial features [7]. Finally, ODE models are the almost "natural" modeling framework when investigating the features of biochemical networks containing regulatory circuits like feedback- and feedforward loops [8]. However, a complete characterization of the model requires to specify the values of several kinds of model parameters (for example, initial protein concentration and rate constants), which in biochemical mid-size networks becomes a computationally intensive task that requires large amounts of quantitative experimental data and sophisticated optimization algorithms. For larger networks parameter estimation becomes cumbersome and difficulties to identify unique values for model parameters emerge [9–11].

An alternative to the ODE models is the *discrete modeling* approach which allows the modeling of interactions among a large number of proteins, genes and other biomolecules to analyze system behavior and make predictions on biologically relevant scenarios [12–14]. It is a qualitative approach that depends only on the network structure (i.e. parameter free) with the simplifying assumption that networks nodes, accounting for the expression level or activation state of biological molecules, exist only in a well-defined set of possible discrete numerical values [10]. Boolean models are the simplest discrete models, in which each element of the network can have one of two possible states at any time (1: ON, expressed or active; 0: OFF, non-expressed or inactive). This assumption is supported by biological evidences, e.g. when genes/proteins exhibit ON/OFF switch like behavior. This can be adopted by assuming a reduced set of biologically meaningful values for the model variables [15,16,10]. A number of recent papers illustrate how discrete logic models nicely capture the behavior of large systems where the interactions are simple [17–19]. However, in some cases discrete logic models do not reproduce some of the time-dependent features associated to non-linear biological circuits, like those containing nested feedback loops.

The recent development of new experimental techniques facilitates high-throughput quantitative proteomics measurements and increases the complexity of the biological networks under investigation. Therefore, it is necessary to find suitable modeling strategies that realize a compromise between the ability to simulate the behavior of large biochemical networks, the complexity associated to the existence of multiple regulatory loops in them and the diversity of sources of experimental data. Hybrid models are mathematical and computational constructs that combine, in a single modeling framework, interdependent variables that distributed over discrete/continuous or deterministic/stochastic domains [20]. Hybrid models account, in an integrative manner, for different spatio-temporal scales of the same biological phenomenon, which are described using different, interconnected modeling frameworks. This approach seems especially valuable to deal with the multifactorial and multi-level nature of biological phenomena like cancer emergence and progression [21].

Here we propose, discuss and analyze the construction of hybrid models, composed of ODE and logic sub-modules, as a strategy to handle large scale, non-linear biochemical networks associated to cancer and other complex diseases. We illustrate the construction of this kind of models using as example a regulatory network centered in E2F1, a key cancer-related transcription factor.

2. Material and methods

For the implementation of our hybrid modeling approach we used built-in ODE solvers from Matlab to simulate the dynamics of the core regulatory module (MathWorks, MA) and CellNetAnalyzer (CNA) to simulate the logical part of the target genes and phenotypical read-outs [14]. To connect both subparts, we develop a Matlab script that discretizes the values of critical variables of the core regulatory model and uses these values as the input vector for the simulations with CNA. The output of the model simulations (continuous and discrete variable values) is stored in a matrix, whose content is visualized using color coded surface plots.

3. Results

3.1. Methodology proposed

Here, we address the problem of modeling large biochemical networks involving dozens to hundreds of receptors, kinase proteins, transcription factors, mRNAs and microRNAs. Their expression and activity are typically regulated by multiple, cross-talking pathways that may even contain overlapping regulatory loops. Our strategy is to organize and divide the network into three parts with distinctive regulatory features (Fig. 1A): 1) the *core regulatory module* of the network, a highly interconnected subnetwork that contains the signaling and transcriptional pathways enriched in feedback and feedforward loops, and therefore is expected to display a high non-linear behavior; 2) the *target genes module*, which accounts for dozens to hundreds of genes whose expression is directly regulated by the proteins and transcription factors included in the core regulatory module and can be experimentally quantified using high-throughput transcriptomics and proteomics techniques; and 3) a set of *phenotypical read-outs*, phenomenological variables accounting for relevant phenotypes triggered upon activation of groups of genes in the target genes module. Input signals external to the networks are assumed to regulate or interact with key compounds of the core regulatory module; although in some cases they could also be direct regulators of the target genes.

In our approach the core regulatory network is modeled using ordinary differential equations, while the target genes module and the phenotypical read-outs are encoded using discrete logic modeling (Boolean or multi-valued logic). Both parts of the model are connected using a discretization interface. Given the set of continuous time-dependent variables of the core regulatory module ($X_{i,C}$,$i = 1,2,\ldots k$), we define a set of auxiliary discrete variables ($X_{i,D}$,$i = 1,2,\ldots k$) for the discretization interface. Furthermore, we discretize the time by considering a discrete and finite set of p time points within the duration of the simulation (t_j,$j = 1,2,\ldots p$). For every time point considered, the interface assigns values to the auxiliary discrete variables, $X_{i,D}(t_j)$, following a set of discretization rules that depend on the current values of the continuous variables, $X_{i,C}(t_j)$, and n_i physiologically relevant thresholds, $TH_{i,m}$,$m = 1,2,\ldots n_i$. The number of thresholds for each continuous variable defines the number of ordinal states for the auxiliary discrete variables. The structure of the discretization rules is displayed in Table 1.

To simulate the whole model, first the core module is initialized and simulated with a set of values defined for the model inputs (Fig. 1). At each of the p time points defined, the values of the auxiliary discrete variables, $X_{i,D}(t_j)$, are calculated following the discretization rules. Those discrete values are passed as inputs to the logical module accounting for the target genes and are used to update the values of its discrete variables following a logical steady-state analysis. Subsequently, the updated values for the model variables in the transcriptional circuitry module are used to update the phenotypical read-outs. In this manner and for every discrete time point (t_j), the model simulation generates: 1) a set of continuous values for the variables of the regulatory core module, $X_{i,C}(t_j)$; 2) a set of discrete values for the auxiliary variables, $X_{i,D}(t_j)$; 3) values for the discrete variables accounting for the expression of the target genes, $Tgt_i(t_j)$; and 4) Boolean values for the phenotypic read-outs (Fig. 1A).

This approach has the advantage that the complex and highly regulated part of the network, which is enriched in regulatory loops is encoded with a modeling approach that provides many computational and analytical tools to investigate its properties [7,22]. On the other hand, the use of logic modeling to describe the associated target genes allows the simulation of massive transcriptional networks with dozens to hundreds of biologically relevant genes with low computational burden [23,24]. In this line, the data produced in high-throughput transcriptomics

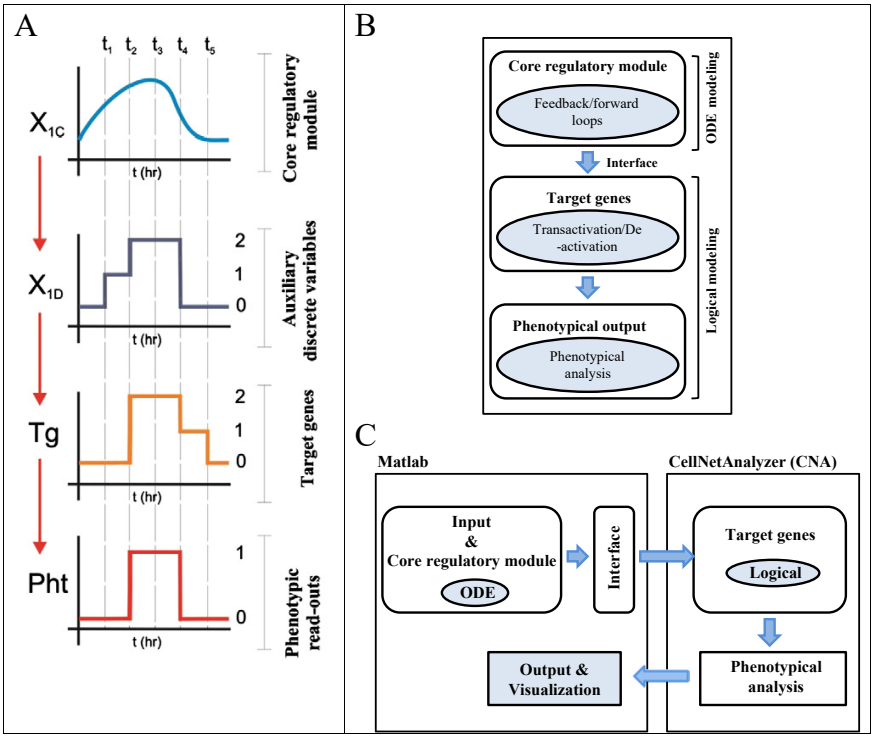


Fig. 1. Hybrid modeling methodology. A: Sketch of a simulation, with continuous and discrete variable sets B: Structure of the model. C: Data workflow in the hybrid model during simulations.

and proteomics experiments can be discretized in the described way to substantiate these large scale transcriptional networks.

To illustrate the functioning of the proposed hybrid modeling approach, we derived a small model describing the regulation of the tumor suppressor miRNA-205 by the transcription factor p73 and its isoform DNp73, and the subsequent regulation of anti-apoptotic targets like BCL-w (Fig. 2A) [25,26]. We have simulated this regulatory module using a model in ordinary differential equations (Fig. 2C) and the hybrid modeling approach here proposed (Fig. 2D). The ODE model was extracted from our previous publication [26]. To construct the hybrid version of the model, we conserved the ODE description of the miR-205 regulation and generated a multi-valued logic model to describe the regulation of BCL-w by miR-205 (Fig. 2B). Furthermore, the continuous ODE values of miR-205 ($miR205_{con}$) were discretized ($miR205_{disc}$) into three possible values using the rule described in Fig. 2B. The two surface plots generated in the model simulations clearly show that the hybrid model preserves the non-linear nature of the BCL-w regulation for different expression values of p73 and DNp73 as displayed by the ODE model [26].

Table 1
Definition of the discretization rules. For the each of the k variables of the core regulatory module ($X_{i,C}, i = 1,2,\ldots k$) at the considered time point, t_j, we assign a value $X_{i,D}(t_j)$ using the following discretization rules:

$$
\begin{array}{l}
\text{for } i = 1,2,\ldots k \\
\text{if } X_{i,C}(t_j) \leq TH_{i,1} \rightarrow X_{i,D}(t_j) = 0 \\
\text{if } TH_{i,1} \leq X_{i,C}(t_j) \leq TH_{i,2} \rightarrow X_{i,D}(t_j) = 1 \\
\text{if } TH_{i,2} \leq X_{i,C}(t_j) \leq TH_{i,3} \rightarrow X_{i,D}(t_j) = 2 \\
\quad \vdots \\
\text{if } TH_{i,n-1} \leq X_{i,C}(t_j) \leq TH_{i,n} \rightarrow X_{i,D}(t_j) = n_i - 1
\end{array}
$$

However, BCL-w expression is now represented by three ordinal states (0: no expression, 1: low expression and 2: high expression).

3.2. Case study: regulation of chemoresistance by a regulatory network centered around the transcription factor E2F1

3.2.1. Background

The protein E2F1 is a transcription factor. Upon stimulation with different signals E2F1 can promote the activation and subsequent expression of different sets of genes involved in DNA replication and repair, cell cycle progression, proliferation and/or apoptosis [27,28]. Activation of E2F1 by growth factor associated signals induces expression of target genes promoting cell cycle progression and proliferation [29,30]. In addition, after genotoxic stress E2F1 regulates the expression and transcriptional activity of the tumor suppressors p53 and p73, promoting the expression of multiple genes related to the regulation of apoptosis and cell cycle arrest [31–33]. Having the ability to regulate cell proliferation and apoptosis, it is not surprising that E2F1 has been found to be deeply involved in tumor progression and the resistance against anti-cancer drugs [34,25,26].

Rather than isolated, E2F1 is part of a vast and complex biochemical network involving regulation events and a transcriptional circuitry. Subparts of this network are enriched in non-linear network motifs, including feedback and feedforward loops. This makes the use of mathematical modeling mandatory to investigate the deregulation of E2F1 in cancer. Several mathematical models have been developed to investigate the regulation of E2F1 in different biological contexts. A number of papers published to date are devoted to the elucidation of the E2F1 role in the regulation of the mammalian G1-S cell cycle transition by means of ODE modeling and bifurcation analysis [35,36]. Others have investigated the mechanisms by which cellular stress (e.g. DNA damage) shifts E2F1 activity to induce apoptosis, and how this mechanism

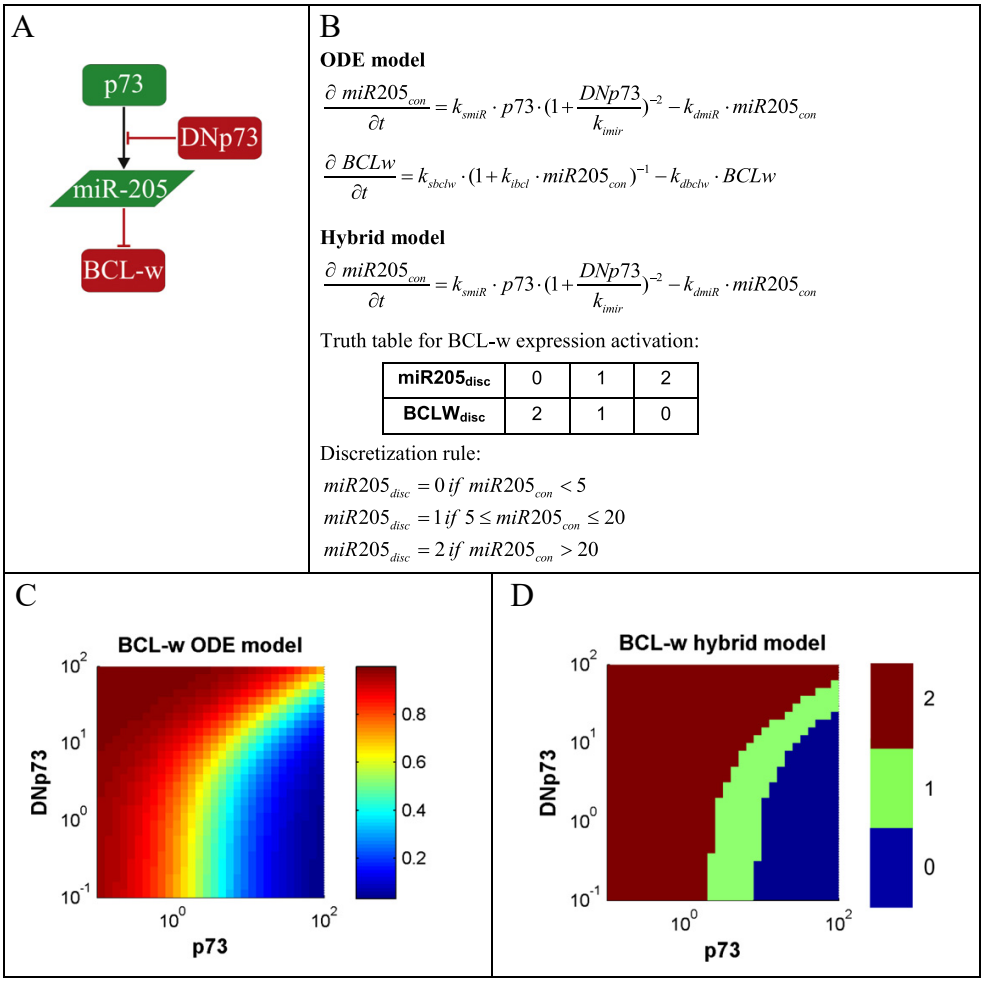


Fig. 2. A: Graphical representation of transcriptional regulation of the oncogene BCL-w by the transcription factor p73 and its isoform DNp73. B: ODE model equations [26] and hybrid model representation. C: ODE model simulations. D: Hybrid model simulations. In the simulations, different expression levels for p73 and DNp73 were considered and the steady-state expression levels of miR-205 and BCL-w were computed using the two versions of the model. ODE model parameters: $k_{dmiR} = 0.029$; $k_{smiR} = 0.1010$; $k_{imir} = 10.2672$; $k_{sbclw} = 0.1$; $k_{dbclw} = 0.1$; $k_{ibc} = 0.18$; p73, DNp73 [10^{-1}, 10^{2}].

is distorted in cancer cells [37–39]. The role of miRNAs in the regulation of E2F1 has also recently been investigated via mathematical modeling [40]. In our own previous work we developed a mathematical model to investigate the E2F1–p73/DNp73–miR-205 regulatory circuit in the emergence of resistance to anti-cancer drugs [26]. Taken together, those papers investigate the regulation of small modules controlling E2F1 in different biological contexts, but none of them has addressed the question of how these events modulating E2F1 expression and activity affect the expression of large transcriptional networks downstream of E2F1. The model here described is the first approach to this question with a focus on the role of E2F1 in chemoresistance.

3.2.2. Model description

The model derived accounts for the critical elements of the signaling and transcriptional network regulating the activity of E2F1 in tumor and normal cells (Fig. 3; for further details see Sup. Mat.). The network can be divided into three parts: 1) the *core regulatory module*, a signaling/transcriptional subnetwork accounting for the regulation of the expression and activity of E2F1 by proliferative and genotoxic stress signals, 2) the *target genes*, a network accounting for the E2F1 modulated expression of critical genes involved in the regulation of cell proliferation, genotoxic drug response and apoptosis; and 3) the *phenotypical read-outs*, a set of phenomenological variables accounting for the activation of relevant phenotypes (e.g., proliferation, apoptosis, chemoresistance) upon the expression of given genes in the target genes (see Sup. Mat.).

The core regulatory module has been extended to account for the regulation of p73, DNp73 and miR-205, which form together with E2F1 a regulatory circuit in the context of the genotoxic stress response and apoptosis initiation [25,26]. Furthermore, we included a sub-module accounting for the activation of E2F1 via EGFR mediated proliferative signals and the E2F1 mediated transcriptional regulation of EGFR [29]. Both sub-modules constitute experimentally confirmed cases of feedback loops, a negative one in the case of E2F1–p73/DNp73–miR-205 [25] and a positive one in the case of E2F1/EGFR [29,26]. In addition, input variables regulating the dynamics of the core module account for: a) the concentration of growth factors activating EGF receptor (GF), b) the cancer-related changes in the expression rate of E2F1 (FS), c) the expression level for the oncogene TGFβ-1 (TGFB1), and d) the dose of genotoxic agent administered as anti-cancer drug (GxS).

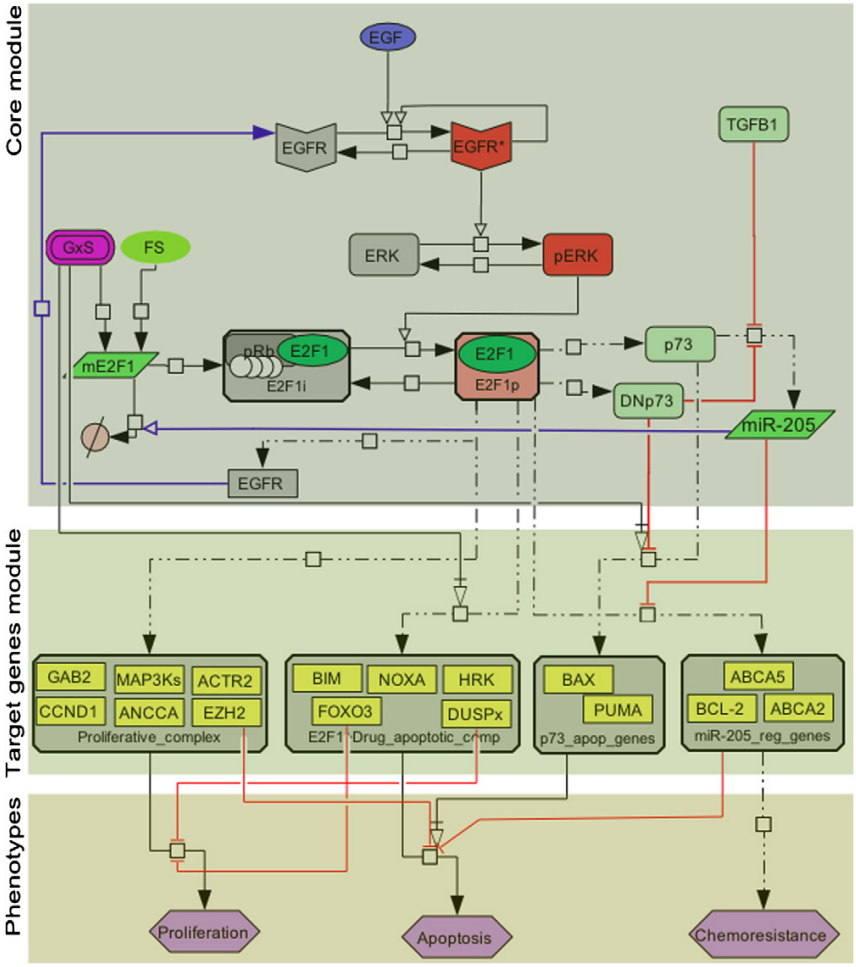


Fig. 3. Regulatory map of the E2F1 centered biochemical network involved in resistance to anti-cancer drugs. The figure displays a CellDesigner map of the system and consists of three parts: 1) the core regulatory module, a transcriptional subnetwork accounting for the regulation of the expression and activity of E2F1 by proliferative and genotoxic stress signals, 2) the target genes module, a network accounting for the E2F1 modulated expression of critical genes involved in the regulation of cell proliferation, genotoxic drug response and apoptosis and 3) the phenotypical read-outs (purple hexagons), a set of phenomenological variables that connect the transcriptional circuitry with phenotypic responses.

We note that TGFβ-1 exemplifies in our model a set of proteins whose dynamics is external to the core module, but regulate the expression and activation of the core regulatory module compounds.

On the other hand, genes considered in the target genes module account for: a) E2F1-regulated genes involved in cancer-related abnormal cell proliferation (e.g., Gab2, CCND1, ANCCA); b) E2F1-modulated genes involved DNA-damage response and apoptosis initiation (e.g., BIM, DUSP1-3, HRK, NOXA); c) p73-regulated genes involved in apoptosis initiation (e.g., BAX, PUMA); and d) miR-205-repressed genes involved in apoptosis repression and chemoresistance (e.g., BCL2, ABCA2). These genes are used here as an example and represent a wider set of genes which undergo similar regulation. With the organization proposed for the network, the core regulatory module contains the molecules and interactions that display non-linear behavior and are involved in regulatory loops.

Following the approach proposed, we derived an ODE model to describe the dynamics of the *core regulatory module*, which was extracted and adapted from our previous publication [26]. For the *target genes* we constructed a discrete multi-valued logic model, which was connected to *phenotypical read-outs* using also multi-valued logic. The *core regulatory module* and the *target genes module* were connected using a computational interface that discretizes the values of the model variables in the regulatory module that control the expression of the genes in the target genes, namely, active E2F1 (E2F1p), p73, DNp73, and miR-205. The discretization rules were established using manual training [26]. Taken together, the model is integrated by nine ordinary differential equations and 35 multi-valued logic functions (Table 2).

3.2.3. Model simulations

We simulated cancer and non-cancer like cells and applied different input vectors by tuning the values of the model inputs GF, FS, TGFB1, and GxS (Fig. 4). We considered two sets of scenarios, characterized by i) the presence of sufficient amount of growth factor (GF = 1) to trigger proliferation and ii) the lack of growth factor stimulation (GF = 0). Next, we modified sequentially and systematically the values of the other input variables, obtaining 36 biological scenarios accounting for the effect of genotoxic drug administration on cancer and non-cancer cells with/without abnormal overexpression of E2F1 and TGFβ-1. The hybrid model was used to compute steady-state values of the model variables. The effect of genotoxic drug administration (GxS) in these different biological scenarios is substantiated in the model by the emerging gene

Table 2
Model description.

Core regulatory module	
$\frac{\partial EGFR}{\partial t} = -k_1 \cdot (GF + k_2 \cdot EGFR^*) \cdot EGFR + \left(k_3 \cdot \frac{E2F1}{k_x + E2F1}\right) - k_4 \cdot EGFR$	[1]
$\frac{\partial EGFR^*}{\partial t} = k_1 \cdot (GF + k_2 \cdot EGFR^*) \cdot EGFR - k_{xx} \cdot EGFR^*$	[2]
$\frac{\partial pERK}{\partial t} = k_5 \cdot EGFR \cdot (ERKt - ERK) - k_6 \cdot pERK$	[3]
$\frac{\partial mE2F1}{\partial t} = GxS \cdot k_7 + FS \cdot k_8 - k_9 \cdot (1 + k_{10} \cdot miR205) \cdot mE2F1$	[4]
$\frac{\partial E2F1i}{\partial t} = k_{11} \cdot mE2F1 - k_{12} \cdot E2F1i \cdot pERK + k_{13} \cdot E2F1p - k_{14} \cdot E2F1i$	[5]
$\frac{\partial E2F1p}{\partial t} = k_{15} \cdot E2F1i \cdot pERK - k_{16} \cdot E2F1p$	[6]
$\frac{\partial p73}{\partial t} = k_{17} \cdot E2F1p - k_{18} \cdot p73$	[7]
$\frac{\partial DNp73}{\partial t} = k_{19} \cdot E2F1p - k_{20} \cdot DNp73$	[8]
$\frac{\partial miR205}{\partial t} = k_{21} \cdot TGFB1^{-1} \cdot p73 \cdot \left(1 + \frac{DNp73}{k_{22}}\right)^{-2} - k_{23} \cdot miR205$	[9]
Discretization rules	
1	$E2F1p_D = 0$ *if* $E2F1p_c < 1.15$
2	$E2F1p_D = 1$ *if* $1.15 \leq E2F1p_c < 12$
3	$E2F1p_D = 2$ *if* $12 \leq E2F1p_c$
4	$miR205_D = 0$ *if* $miR205_c < 5$
5	$miR205_D = 1$ *if* $5 \leq miR205_c$
6	$p73_D = 0$ *if* $p73_c < 1$
7	$p73_D = 1$ *if* $1 \leq p73_c$
8	$DNp73_D = 0$ *if* $DNp73 < 10$
9	$DNp73_D = 1$ *if* $10 \leq DNp73$
Transcriptional circuitry	
Reactions	
1a	E2F1p = Gab2
1b	2 E2F1p = 2 Gab2
2a	E2F1p = CCND1
2b	2 E2F1p = 2 CCND1
3a	E2F1p = ANCCA
3b	2 E2F1p = 2 ANCCA
4a	E2F1p = MAP3Ks
4b	2 E2F1p = 2 MAP3Ks
5a	E2F1p = ACTR2
5b	2 E2F1p = 2 ACTR2
6a	E2F1p = EZH2
6b	2 E2F1p = 2 EZH2
7a	E2F1p + GxS = EZH2
7b	2 E2F1p + GxS = 2 EZH2
8	E2F1p + 2 !EZH2 = BIM
9	E2F1p + GxS = FOXO3
10	E2F1p + GxS = DUSPx
11	E2F1p + GxS = NOXA
12	E2F1p + GxS = HRK
13	P73 + !DNp73 + GxS = BAX
14	P73 + !DNp73 + GxS = PUMA
15	!miR-205 + E2F1p = Bcl-2
16	!miR-205 + E2F1p = ABCA2
17	!miR-205 + E2F1p = ABCA5
Phenotypic read-outs	
Reactions	
1a	Gab2 + ANCCA + CCND1 + MAP3Ks + ACTR + EZH2 = Proliferative_complex
1b	2 Gab2 + 2 ANCCA + 2 CCND1 + 2 MAP3Ks + 2 ACTR + 2 EZH2 = Proliferative_complex
2	BIM + FOXO3 + DUSPx + NOXA + HRK = E2F1_Drug_apoptotic_comp
3	BAX = p73_apop_genes
4	PUMA = p73_apop_genes
5	Bcl-2 = miR-205_reg_genes
6	ABCA2 = miR-205_reg_genes
7	ABCA5 = miR-205_reg_genes
8	Proliferative_complex + !DUSPx + !FOXO3 = Proliferation
9	E2F1_Drug_apoptotic_comp + p73_apop_genes + 2 ! EZH2 + !miR-205_reg_genes = Apoptosis
10	miR-205_reg_genes = Chemoresistance

In CellNetAnalyzer the "+" sign represents logical 'AND', while the "!" sign represents logical 'NOT'.

expression pattern and the triggering of any of the phenotypes considered in the model. The computation of the complete set of simulation took approximately 1 s on a conventional workstation.

The results obtained are in accordance with those previously derived from an ODE model and with other biological evidences in the literature [26,25,29]. However, our hybrid model delivered some

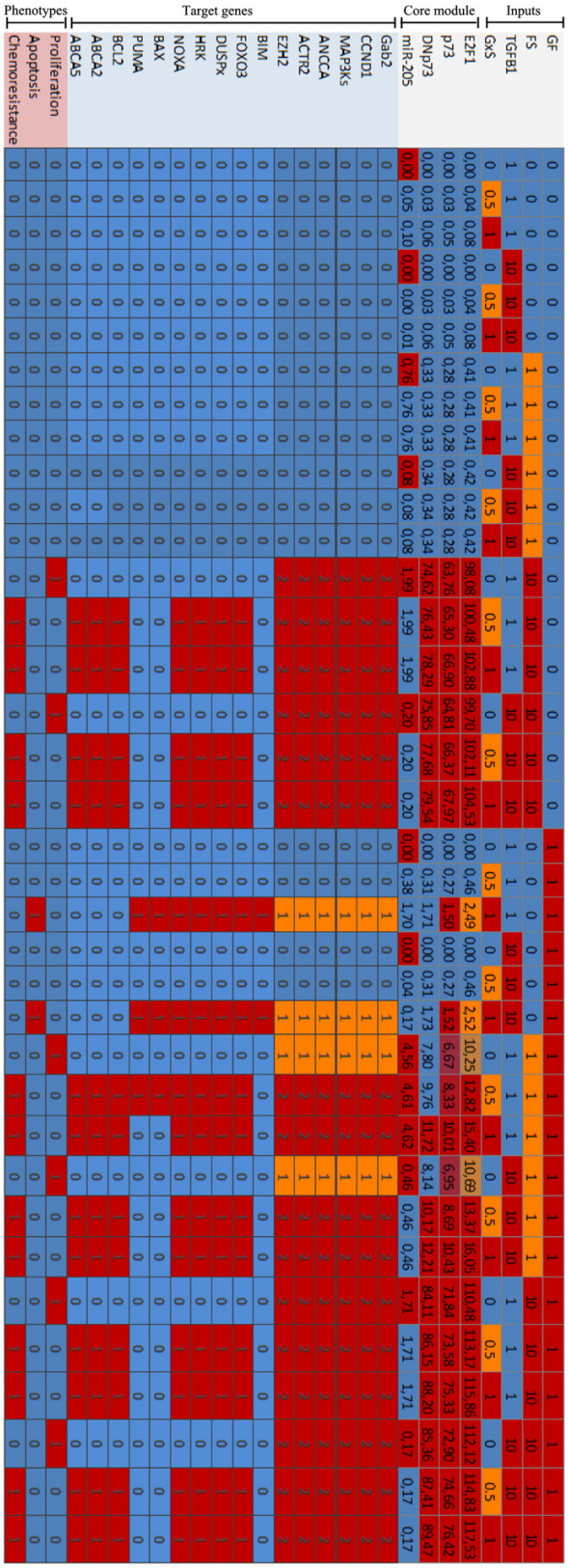


Fig. 4. Model simulations for the effect of genotoxic drug administration on cancer and non-cancer cells with/without overexpression of E2F1 and TGFβ-1, and with/without growth factor stimulation.

interesting results. For example, our model simulations for conditions of non-growth factor stimulation ($GF = 0$) indicate that in the case of sufficient overexpression of E2F1 tumor cells may become self-sufficient in growth signals [41,42]. We link this simulation result to the existence of a positive feedback loop between E2F1 and EGFR [29]. Furthermore, sufficient overexpression of either E2F1 ($E2F1 = 10$) or TGFβ-1 ($TGFB1 = 10$) can promote resistance to anti-cancer drugs by the repression of miR-205. This conclusion is in agreement to the results of our previous publication. On the other hand, a version of the model completely constructed using multi-valued logic fails to reproduce the features associated to the existence of the indicated feedback loops (see Sup. Mat for further details).

Furthermore, we modified the E2F1 synthesis parameter (FS) and TGFβ-1 expression level (TGFB1) to investigate whether they can synergize to provoke the emergence of a resistant phenotype [26]. We investigated the combined effect of deregulation in E2F1 and TGFβ-1 with our model. To this end, the values of the input variables FS and TGFB1 were iteratively modified and we simulated whether resistance to apoptosis initiation (chemoresistance) after stimulation with genotoxic drug emerged (Fig. 5). Our model simulations suggest that a combination of E2F1 and TGFβ-1 overexpressions can trigger chemoresistance. In contrast to Boolean models, the use of a hybrid model is sufficient to determine precise thresholds of E2F1 and TGFβ-1 expressions shifting the phenotype from chemosensitive (Fig. 5, blue area) to chemoresistant (Fig. 5, red area).

4. Discussion

In this work we have proposed and illustrated the construction of hybrid models, composed of ODE and logic sub-modules, as a strategy to handle large scale, non-linear biochemical networks. Our strategy is to organize the network into three parts: 1) the *core regulatory module*, a subnetwork of signaling and transcriptional processes that displays high interconnectivity and is enriched in non-linear regulatory loops, 2) the *target genes*, accounting for genes whose expression is directly regulated by the signaling proteins and transcription factors included in the core module and 3) the *phenotypical read-outs*. The core module is modeled using ODEs, while the target genes and the phenotypical read-outs are described using discrete logic modeling. Furthermore, an interface is designed that links both parts by discretizing the values of the variables in the core module. Although the approach proposed seems simple, a number of questions emerge when trying to generalize the method and make it more accurate.

In our case study, the model construction relied on the manual retrieval of biomedical information and integration into a regulatory map. Real case study networks can be much more complex and extensive, especially in terms of the target genes, which makes it impossible to manually retrieve the information. One approach to predict the structure of these networks could be the systematic use of existing databases of protein–protein interactions and transcriptional regulation [43,44,26].

Similarly, the determination of the core regulatory module is typically not straightforward. It requires, for the biomedical context under investigation, to determine in a systematic manner which parts of the network display structural complexity and are enriched in non-linear regulatory loops. A possible approach could be to employ tools from the graph-based network analysis to detect, classify and rank non-linear network motifs [45].

In addition, in some cases there may be feedback between the core regulatory module and the target genes module. For example, in our case study EGFR is a transcriptional target of E2F1 but also regulates E2F1 activity in a system that has the structure of a positive feedback

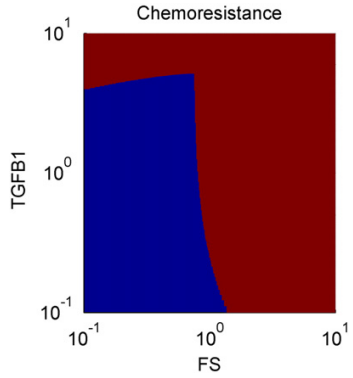


Fig. 5. Emergence of chemoresistance upon deregulation of E2F1 and TGFβ-1 expressions. The values of the input variables accounting for E2F1 synthesis rate (FS) and TGFβ-1 expression (TGB1) were iteratively modified in an interval of physiologically feasible values (values were normalized to the ones in the wild-type chemosensitive cell line described in [25,26]). We simulated the system under genotoxic drug administration and analyzed for each pair of values for FS and TGFB1, respectively, whether apoptosis initiation was triggered (blue area) or the cell displays chemoresistance (red area).

	t=0	t=60	t=120	t=180	t=240	t=270
GF	1	1	1	1	1	1
FS	1	1	1	1	1	1
TGFB1	1	1	1	1	1	1
GxS	0	0	1	1	1	0
E2F1	10,25	10,25	10,25	14,27	15,37	12,38
p73	6,66	6,66	6,66	8,97	9,98	8,55
DNp73	7,80	7,80	7,80	10,35	11,69	10,19
miR-205	4,56	4,56	4,56	4,61	4,61	4,53
Gab2	1	1	1	2	2	2
CCND1	1	1	1	2	2	2
EGFR	1	1	1	2	2	2
ANCCA	1	1	1	2	2	2
ACTR2	1	1	1	2	2	2
EZH2	1	1	1	2	2	2
BIM	0	0	0	0	0	0
FOXO3	0	0	0	1	1	0
DUSPx	0	0	0	1	1	0
HRK	0	0	0	1	1	0
NOXA	0	0	0	1	1	0
BAX	0	0	0	0	0	0
PUMA	0	0	0	0	0	0
BCL2	0	0	0	1	1	0
ABCA2	0	0	0	1	1	0
ABCA5	0	0	0	1	1	0
Proliferation	1	1	1	0	0	1
Apoptosis	0	0	0	0	0	0
Chemoresistance	0	0	0	1	1	0

Fig. 6. Time series model simulation. For one of the scenarios considered previously, we simulated the effect of pulse-like drug administration. Toward this end, the simulation time was discretized (t = 0, 60, 120, 180, 240, 270 h) and the nodes of the target genes and the phenotypes were updated at those time points.

loop [29]. Given the consequences that this motif has on the regulation of the system, the dynamics of the transcriptional target EGFR were included in the non-linear core regulatory module. This idea could be used to account for this kind of motifs in other models. In a more advanced setup of the methodology, network biology methods would be used to detect feedback loops between different parts of the network. These motifs would then be entirely included in the non-linear core regulatory module. In addition, the use of logical modeling would be reserved for the linear transcriptional regulation motifs.

When modeling the core regulatory module with ODEs, two fundamental questions have to be addressed: the derivation of the model equations and the estimation of model parameters. The construction of model equations may benefit from using canonical modeling frameworks like classical kinetic or power-law modeling to obtain more consistent, homogeneous and easy to derive equations [6,46,2]. Secondly, the characterization of the ODE model will require adequate quantitative experimental data (dose–response experiments, quantitative time series, etc; see [2,7] for further details and example), but also adequate methodologies for parameter estimation [11,47].

A critical element of the proposed approach is to determine the discretization rules for the core regulatory variables that control the target genes. These rules are the basis of the interface between the ODE and logic modules. In our case study, the discretization rules were derived by manual training [7], but in a real case with a larger set of genes integrated in the target genes module a more systematic approach will be necessary. This approach will require transcriptomics and proteomics data from experiments, in which expression and activity of the proteins and transcription factors controlling transcriptional programs are modulated in a dose–response fashion. Furthermore, computational methodologies will be required to determine precise discretization rules and gene-dependent activation thresholds for the transcriptional factors [48].

A final but not minor question is that of timing in the model simulations. ODE models are continuous-time models, in which time is, at least theoretically, a continuous variable, while in discrete logic-models also the time is discrete per se. A more advanced version of the computational interface between the ODE and logic submodules will require rules for time discretization. Therefore, a criterion has to be established to re-compute iteratively in defined time-intervals, the discrete values for the regulators of the target genes (Fig. 6). This strategy could even be enhanced by integrating time-dependent high-throughput gene expression and phenotypic data in the analysis.

In this paper we proposed to combine ODE and discrete logic modeling in computational hybrid models, which can be a useful tool for the analysis of large scale biochemical networks. In Table 3 we compare the performance of ODE, discrete logic and hybrid models with respect to a number of biological and computational features, which are important when choosing a strategy to construct a mathematical model for a biochemical system. The table indicates that among them there is not a perfect modeling framework, able to deal better than the others in all the features considered. This is a conclusion that can be extended to a wider pool of modeling frameworks and emerges after inspecting decades of scientific publications on modeling of biochemical models. Some modeling frameworks are clearly better in some features but have a poor performance for other features. When choosing the right modeling framework one has to think which of those features are

Table 3
Comparison of ODE, discrete logic and hybrid models for the analysis of biochemical networks.

	ODE models	Logic models	Hybrid models
Biophysical realism			
Accuracy			
Interpretability			
Encoding of non-linearity			
Calibration techniques			
Integrability of quantitative data			
Scalability			
Computability			

Legend	Good	Acceptable	Poor

more important in the biochemical system investigated. We think that hybrid models like the ones proposed are a good compromise between quantitative/qualitative accuracy and scalability when considering large biochemical networks with a small core of highly interconnected feedback loop regulated processes and a large amount of transcriptionally and post-transcriptionally regulated genes not intervening in those loops.

An interesting option when using a modular representation of regulatory networks connected to phenotypic readouts is the investigation of phenotype probability distributions in heterogeneous cell populations. This kind of approach can be used to connect single-cell and cell-population models. Toward this end, alternative to our method, other approaches based on continuous variables may be considered (i.e., partial differential equations; see for example [49]). In the current study, our approach was used to investigate the behavior of the signaling and transcriptional machinery of single cells belonging to homogenous populations, and their connection to the emergence of given phenotypes. Thus, the discrete representation of the phenotypical readouts remains valid. Furthermore, the hybrid approach proposed could be integrated in more advanced methodologies accounting for the dynamics of cell populations, like for example agent-based models of cell populations [50]. In this case, our hybrid model approach would account for the regulatory machinery of each individual cell, which would be connected to the phenotypic parameters controlling, in the agent-based model, the dynamics of the modeled cell. For models of mid-size cell populations, the use of our hybrid approach for describing the internal regulatory machinery of each cell reduces the computational effort necessary for simulations of the agent-based model. Statistical analysis can be further applied to the output of the model simulations [51].

Acknowledgements

The original idea was developed by JV and FK. E2F1 network structure was determined by JV, US, FK and DE. Mathematical model, numerical simulations and figures were elaborated by FK, SN and JV. The manuscript was drafted by FK, US, JV, DE, BP and OW. This work was supported by the German Federal Ministry of Education and Research (BMBF) as part of the eBio:miRSys [0316175A to JV and OW], and eBio:SysMet [0316171 to JV, OW and BP] and the Rostock University Medical Center for the project Systems Medicine of Cancer Invasion and Metastasis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbapap.2013.05.007.

References

[1] Ronojoy Ghosh, ClaireJ Tomlin, Lateral inhibition through Delta–Notch signaling: a piecewise affine hybrid model, Hybrid Systems: Computation and Control, 2001. 232–246.
[2] J. Vera, Eva Balsa-Canto, Peter Wellstead, Julio R. Banga, O. Wolkenhauer, Power-law models of signal transduction pathways, Cell. Signal. (2007) 1531–1541.
[3] Olaf Wolkenhauer, Mihajlo Mesarovic, Feedback dynamics and cell function: why systems biology is called systems biology, Mol. Biosyst. (2005) 14–16.
[4] Olaf Wolkenhauer, Charles Auffray, Simone Baltrusch, et al., Systems biologists seek fuller integration of systems biology approaches in new cancer research programs, Cancer Res. (2010) 12–13.
[5] Hidde de Jong, Modeling and simulation of genetic regulatory systems: a literature review, J. Comput. Biol. (2002) 67–103.
[6] Michael A. Savageau, Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions, J. Theor. Biol. (1969) 365–369.
[7] J. Vera, O. Wolkenhauer, Mathematical tools in cancer signalling systems biology, in: A. Cesario, F.B. Marcus (Eds.), Cancer Systems Biology, Bioinformatics and Medicine: Research and Clinical Applications, Springer Science, NY, 2011, pp. 185–212.
[8] John J. Tyson, K.C. Chen, B. Novak, Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell, Curr. Opin. Cell Biol. (2003) 221–231.
[9] Melody K. Morris, Julio Saez-Rodriguez, Logic-based models for the analysis of cell signaling networks, Biochemistry (2010) 3216–3224.
[10] Tomas Helikar, Naomi Kochi, Jim A. Rogers, Boolean modeling of biochemical networks, Bioinformatics (2011) 16–25.
[11] Chih-Lung Ko, Eberhard O. Voit, Feng-Sheng Wang, Estimating parameters for generalized mass action models with connectivity information, BMC Bioinforma. 10 (2009) 140.
[12] S.A. Kauffman, Metabolic stability and epigenesis in randomly constructed genetic nets, J. Theor. Biol. (1969) 437–467.
[13] Leon Glass, Stuart A. Kauffman, The logical analysis of continuous, non-linear biochemical control networks, J. Theor. Biol. (1973) 103–129.
[14] Steffen Klamt, Julio Saez-Rodriguez, Ernst D. Gilles, Structural and functional analysis of cellular networks with CellNetAnalyzer, BMC Syst. Biol. 1.1 (2007) 2.
[15] Ritwik Layek, Aniruddha Datta, Cancer therapy design based on pathway logic, Syst. Biol. (2010) 548–555.
[16] Francois Jacob, Jacques Monod, Genetic regulatory mechanisms in the synthesis of proteins, J. Mol. Biol. (1961) 318–365.
[17] Julio, Saez-Rodriguez, et al., A logical model provides insights into T cell receptor signaling, PLoS Comput. Biol. (3.8) (2007) e163.
[18] Rebekka Schlatter, et al., ON/OFF and beyond-a boolean model of apoptosis, PLoS Comput. Biol. 5.12 (2009) e1000595.
[19] Amit Singh, Boolean approach to signalling pathway modelling in HGF-induced keratinocyte migration, Bioinformatics (2012) i495–i501.
[20] Fisher Henzinger, Executable cell biology, Nat. Biotechnol. 25 (2007) 1239–1249.
[21] K.A. Rejniak, A.R. Anderson, Hybrid models of tumor growth, Wiley Interdiscip. Rev. Syst. Biol. Med. (2011) 115–125.
[22] J. Vera, X. Lai, U. Schmitz, O. Wolkenhauer, MicroRNA-regulated networks: the perfect storm for classical molecular biology, the ideal scenario for systems biology, Adv. Exp. Med. Biol. (2013) 978–994.
[23] Julio Saez-Rodriguez, et al., Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction, Mol. Syst. Biol. 5.1 (2009).
[24] Regina Samaga, et al., The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data, PLoS Comput. Biol. 5.8 (2009) e1000438.
[25] V. Alla, Bhavani S. Kowtherapu, David Engelmann, Stephan Emmrich, Ulf Schmitz, Marc Steder, Brigitte M. Pützer, E2F1 confers anticancer drug resistance by targeting ABC transporter family members and Bcl-2 via the p73/DNp73-miR-205 circuitry, Cell Cycle (2012) 2067–3078.
[26] J. Vera, U. Schmitz, X. Lai, D. Engelmann, F.M. Khan, O. Wolkenhauer, B.M. Pützer, Kinetic modeling-based detection of genetic signatures that provide chemoresistance via the E2F1–p73/DNp73-miR-205 network, Cancer Research, 2013. http://dx.doi.org/10.1158/0008-5472.CAN-12-4095 (in press).
[27] James DeGregori, The genetics of the E2F family of transcription factors: shared functions and unique roles, Biochim. Biophys. Acta (2002) 131–150.
[28] Heiko Müller, Adrian P. Bracken, E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis, Genes Dev. (2001) 267–285.
[29] V. Alla, David Engelmann, Brigitte M. Putzer, E2F1 in melanoma progression and metastasis, J. Natl. Cancer Inst. (2010) 127–133.
[30] Shirley Polager, Doron Ginsberg, p53 and E2f: partners in life and death, Nat. Rev. Cancer 9 (2009) 738–748.
[31] David Engelmann, Brigitte M. Putzer, Translating DNA damage into cancer cell death — a roadmap for E2F1 apoptotic signalling and opportunities for new drug combinations to overcome chemoresistance, Drug Resist. Updat. (2010) 119–131.
[32] D. Engelmann, B.M. Pützer, The dark side of E2F1: in transit beyond apoptosis, Cancer Res. (2012) 571–575.
[33] David Engelmann, Susanne Knoll, Daniel Ewerth, Marc Steder, Anja Stoll, Brigitte M. Putzer, Functional interplay between E2F1 and chemotherapeutic drugs defines immediate E2F1 target genes crucial for cancer cell death, Cell. Mol. Life Sci. (2010) 931–948.
[34] B.M. Pützer, D. Engelmann, E2F1 apoptosis counterattacked: evil strikes back, Trends Mol. Med. (2013) 89–98.
[35] M. Swat, A. Kel, H. Herzel, Bifurcation analysis of the regulatory modules of the mammalian G1/S transition, Bioinformatics (2004) 1506–1511.
[36] Guang Yao, Tae Jun Lee, Seiichi Mori, Joseph R. Nevins, Lingchong You, A bistable Rb–E2F switch underlies the restriction point, Nat. Cell Biol. (2008) 476–482.
[37] Gheorghe Craciun, Baltazar Aguda, Avner Friedman, Mathematical analysis of a modular network coordinating the cell cycle and apoptosis, Math. Biosci. Eng. (2005) 473–485.
[38] Baltazar D. Aguda, A structural analysis of the qualitative networks regulating the cell cycle and apoptosis, Cell Cycle (2003) 538–543.
[39] Xiao-Peng Zhang, Feng Liu, Wei Wang, Coordination between cell cycle progression and cell fate decision by the p53 and E2F1 pathways in response to DNA damage, J. Biol. Chem. (2010) 31571–31580.
[40] Baltazar D. Aguda, Yangjin Kim, Clay B. Marsh, MicroRNA regulation of a cancer network: Consequences of the feedback loops involving miR-17-92, E2F, and Myc, PNAS (2008) 19678–19683.
[41] Douglas Hanahan, Robert A. Weinberg, The hallmarks of cancer, Cell (2000) 57–70.
[42] Douglas Hanahan, Robert A. Weinberg, Hallmarks of cancer: the next generation, Cell (2011) 646–674.
[43] J. Vera, S. Nikolov, X. Lai, A. Singh, O. Wolkenhauer, Model-based investigation of the transcriptional activity of p53 and its feedback loop regulation via 14-3-3σ, IET Syst. Biol. (2011) 293–307.

[44] Xin Lai, Ulf Schmitz, Olaf Wolkenhauer, Julio Vera, Computational analysis of target hub gene repression regulated by multiple and cooperative miRNAs, Nucleic Acid Res. (2012) 1–17.
[45] Duc-Hau Le, Yung-Keun Kwon, Yung-Keun, NetDS: a Cytoscape plugin to analyze the robustness of dynamics and feedforward/feedback loop structures of biological networks, Bioinformatics (2011) 2767–2768.
[46] E.O. Voit, Biological Systems Modeling and Analysis, Cambridge University Press, Cambridge, U.K., 2000
[47] I.C. Chou, E.O. Voit, Recent developments in parameter estimation and structure identification of biochemical and genomic systems, Math. Biosci. (2009) 57–83.
[48] F. Eduati, J. De Las Rivas, B. Di Camillo, G. Toffolo, J. Saez-Rodriguez, Integrating literature-constrained and data-driven inference of signalling networks, Bioinformatics 28 (2012) 2311–2317, (s.l.: bioinformatics).
[49] Jean Clairambault, Stéphane Gaubert, Thomas Lepoutre, Math. Comput. Model. (April 2011) 1558–1567.
[50] T. Sütterlin, S. Huber, H. Dickhaus, N. Grabe, Modeling multi-cellular behavior in epidermal tissue homeostasis via finite state machines in multi-agent systems, Bioinformatics 25 (2009) 2057–2063.
[51] I. González-García, R.V. Solé, J. Costa, Metapopulation dynamics and spatial heterogeneity in cancer, PNAS 99 (20, 2002) 13085–13089.

Published OnlineFirst February 27, 2013; DOI: 10.1158/0008-5472.CAN-12-4095

Integrated Systems and Technologies: Mathematical Oncology

Cancer Research

Kinetic Modeling–Based Detection of Genetic Signatures That Provide Chemoresistance via the E2F1-p73/DNp73-miR-205 Network

Julio Vera[1,3], Ulf Schmitz[1], Xin Lai[1,4], David Engelmann[2], Faiz M. Khan[1], Olaf Wolkenhauer[1,5], and Brigitte M. Pützer[2]

Abstract

Drug resistance is a major cause of deaths from cancer. E2F1 is a transcription factor involved in cell proliferation, apoptosis. and metastasis through an intricate regulatory network, which includes other transcription factors like p73 and cancer-related microRNAs like miR-205. To investigate the emergence of drug resistance, we developed a methodology that integrates experimental data with a network biology and kinetic modeling. Using a regulatory map developed to summarize knowledge on E2F1 and its interplay with p73/DNp73 and miR-205 in cancer drug responses, we derived a kinetic model that represents the network response to certain genotoxic and cytostatic anticancer drugs. By perturbing the model parameters, we simulated heterogeneous cell configurations referred to as *in silico* cell lines. These were used to detect genetic signatures characteristic for single or double drug resistance. We identified a signature composed of high E2F1 and low miR-205 expression that promotes resistance to genotoxic drugs. In this signature, downregulation of miR-205, can be mediated by an imbalance in the p73/DNp73 ratio or by dysregulation of other cancer-related regulators of miR-205 expression such as TGFβ-1 or TWIST1. In addition, we found that a genetic signature composed of high E2F1, low miR-205, and high ERBB3 can render tumor cells insensitive to both cytostatic and genotoxic drugs. Our model simulations also suggested that conventional genotoxic drug treatment favors selection of chemoresistant cells in genetically heterogeneous tumors, in a manner requiring dysregulation of incoherent feedforward loops that involve E2F1, p73/DNp73, and miR-205. *Cancer Res; 73(12); 3511–24. ©2013 AACR.*

Major Findings

Our model analysis indicates that tumor cells showing a genetic signature composed of high E2F1, low miR-205, and high ERBB3 expression can be resistant to multiple types of anticancer drugs. Furthermore, we found that the deregulation of E2F1 signaling can promote chemoresistance in concert with some epithelial–mesenchymal transition signals via common miRNA targets. Our analysis shows that the E2F1 signaling network is enriched in feedforward loops, the deregulation of which can play a decisive role in the emergence of chemoresistance. Finally, we found that in *in silico* heterogeneous tumors with cells displaying different signatures for the E2F1 signaling network, genotoxic drugs can favor the selection of subpopulations of chemoresistant tumor cells.

Introduction

Resistance to genotoxic drugs as the major cause of cancer therapy failure is a serious problem for oncologists and their patients that requires the understanding of the pivotal triggering events and its evolution when cancer progresses. Chemoresistance can be innate or acquired and may attributable to a single agent or a class of drugs. Potential mechanisms to counteract the therapeutic effects of DNA-damaging agents include the reduction of effective drug concentrations via enhanced efflux, detoxification enzymes or drug sequestration, modification of drug targets, changes or mutation in mitotic checkpoint signals, and hyperactivation of DNA repair mechanisms (1).

The cellular transcription factor E2F1 is a unique member of the E2F family of proteins as it regulates the tumor suppressor

Authors' Affiliations: [1]Department of Systems Biology and Bioinformatics, Institute of Computer Science, University of Rostock; [2]Institute of Experimental Gene Therapy and Cancer Research, Rostock University Medical Center, Rostock; [3]Laboratory of Systems Tumor Immunology, Department of Dermatology, Faculty of Medicine, Friedrich Alexander Universität, Erlangen-Nürnberg; [4]Department of Inflammation of the Lung / iLung, Philipps Universität, Marburg, Germany; and [5]Stellenbosch Institute for Advanced Study (STIAS), Wallenberg Research Centre at Stellenbosch University, Stellenbosch, South Africa

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

O. Wolkenhauer and B.M. Pützer contributed equally to this work.

Corresponding Author: Julio Vera, Laboratory of Systems Tumor Immunology, Department of Dermatology, Faculty of Medicine, University of Erlangen-Nürnberg, Ulmenweg 18, 91054 Erlangen, Germany. E-mail: julio.vera.g@gmail.com

doi: 10.1158/0008-5472.CAN-12-4095

Published OnlineFirst February 27, 2013; DOI: 10.1158/0008-5472.CAN-12-4095

Quick Guide to Main Model Equations and Assumptions

The kinetic model is organized in four interconnected modules: a core regulatory module, transcriptional modules related to apoptosis and proliferation, and an additional module for the tumor cell population.

The core regulatory module contains five differential equations, accounting for E2F1 mRNA and protein (*mE2F1 and E2F1*), both isoforms of p73 (respectively, *p73* and *DNp73*), and miR-205 (*miR205*). The most relevant equation describes the dynamics of miR-205:

$$\frac{dmiR205}{dt} = k_{11} \cdot TGFB1^{-1} \cdot p73 \cdot \left(1 + \frac{DNp73}{k_{13}}\right)^{-2} - k_{12} \cdot miR205$$

This equation describes the regulation of miR-205 synthesis and degradation (k_{11}, k_{12}). Here, we assume that p73 acts as an activator of miR-205 synthesis, whereas DNp73 plays an inhibitory role. According to the observations reported in several tumor types, we assume that when both isoforms of p73 are overexpressed, the inhibitory effect of DNp73 dominates and cancels out p73 activation of miR-205. To exemplify the set of regulators of miR-205 that are external to the E2F1/p73/miR-205 network, we include a term that describes the repression of miR-205 by the oncogenic signal coming from the TGFβ-1 pathway (*TGFB1*).

The module for apoptosis-related targets includes three differential equations, accounting for the dynamics of E2F1-regulated proapoptotic genes (represented by *Hrk*), p73-regulated proapoptotic genes (represented by *Bax*), and miR-205–repressed antiapoptotic genes (represented by *BCL2*). The most representative equation for this module accounts for *Hrk* dynamics:

$$\frac{dHrk}{dt} = k_{17} \cdot DS \cdot \frac{E2F1^{g}}{k_{16}{}^{g} + E2F1^{g}} - k_{18} \cdot Hrk$$

This equation describes the regulation of *Hrk* synthesis and degradation (k_{17}, k_{18}). In this equation, we include the E2F1-mediated expression of *Hrk* using a Hill function. In addition, we assume that DNA damage signals trigger acetylation and activation of E2F1, which is required to promote *Hrk* expression. Drug-induced DNA damage signals are here encoded as a binary variable (*DS*).

The module for targets associated with proliferation includes ordinary differential equations for the *de novo* inactive EGFR (*EGFR*), growth factor-activated EGFR ($EGFR^*$), cytostatic drug–inhibited EGFR ($EGFR^I$), and active ERBB3 (*ERBB3*). In this module, the most important equation describes the dynamics of $EGFR^I$:

$$\frac{dEGFR^I}{dt} = k_{30} \cdot CyD \cdot \left(k_{36} \cdot \frac{E2F1^{g_2}}{k_{35}{}^{g_2} + E2F1^{g_2}} - EGFR^* - EGFR^I\right) \cdot \left(1 + \frac{ERBB3}{ER^{TH}}\right)^{-1} - k_{31} \cdot EGFR^I$$

The first term accounts for the inhibition of EGFR by cytostatic drugs (*CyD*), which converts EGFR into inactive form ($EGFR^I$). Here, we assume that the expression of EGFR is regulated by E2F1, which is modeled with a Hill equation (k_{36}, g_2). We include a power-law term accounting for the effect of ERBB3 as an inhibitor of the cytostatic drugs regulating EGFR activity (*ER*).

The tumor cell population module is composed of an ordinary differential equation accounting for the size of a population of tumor cells with the considered genetic signature (*TC*). This equation is a phenomenological kinetic equation that connects the components of the regulatory network to the dynamics of the tumor cell population:

$$\frac{dTC}{dt} = k_{24} \cdot EGFR^* \cdot TC - k_{22} \cdot Hrk \cdot Bax \cdot \left(1 + \frac{BCL2}{k_{23}}\right)^{-1} \cdot TC$$

In this equation, active EGFR ($EGFR^*$) connects the proliferation of the tumor cell population to proliferative signals (k_{24}), whereas the representatives for pro- and antiapoptotic proteins (*Hrk*, *Bax*, and *BCL2*) regulate the cell death rate after genotoxic stress (k_{22}, k_{23}).

In some simulations, we consider tumors composed of three subpopulations of tumor cells with different genetic signatures. Accordingly, we construct three instances of the model, which differ in model parameter values describing the different genetic signatures. Further details are available in the main text and Supplementary Material.

p53 and its homologue p73 thereby promoting apoptosis by the activation of a plethora of death pathways (2). One of the first signals recognized to induce E2F1′s apoptotic activity was DNA damage (3). Overexpression of E2F1 was shown to increase the sensitivity of malignant cells to apoptosis upon genotoxic treatment (4), which is critical for tumor growth

Published OnlineFirst February 27, 2013; DOI: 10.1158/0008-5472.CAN-12-4095

reduction (5). In this regard, activation of p73 by E2F1 functions as backup once p53 is defective to ensure that damaged cells can undergo apoptosis (6,7). In such a scenario, high levels of endogenous E2F1 should increase the effectiveness of DNA-damaging agents and discriminate between tumors clinically responsive or resistant to anticancer therapy. This is in line with clinical surveys that associated E2F1 expression with improved survival in patients treated with adjuvant chemotherapy and underscores E2F1′s role as an endogenous chemosensitizer in patients with cancer and beyond, as a predictive factor for therapeutic success and clinical outcome (8).

This beneficial view on E2F1 was thwarted as other studies showed that overexpression or amplification of the E2F1 genomic locus in conjunction with RB1 loss frequently observed during cancer progression is associated with metastatic disease and chemoresistance (9). From this perspective, aberrantly elevated E2F1 levels in high-grade tumors can be considered as a marker for unfavorable patient survival prognosis. In support of these findings, we have recently shown that dysregulated E2F1 causes malignant progression in therapy-resistant metastatic melanoma xenografts in which depletion of endogenously high E2F1 levels abrogates tumor invasion and pulmonary metastasis (10). According to our results, the aggressive behavior of E2F1 in melanoma cells is partially mediated through the induction of the EGF receptor (EGFR) pathway.

It is well established that E2F1 stimulates the expression of the tumor suppressor p73 and its N-terminally truncated isoform (named DNp73) via direct transactivation of the *TP73* gene (11). While full-length p73 inhibits cancer development by inducing cell-cycle arrest and apoptosis through its ability to bind p53 DNA target sites, DNp73 acts as antagonist of wild-type p53 family members by either directly interfering with DNA binding or forming inactive heteromeric complexes with transcriptionally active p73 (12). Strikingly, in melanoma metastases DNp73 isoforms are strongly upregulated in conjunction with full-length p73 compared with primary tumors (13). Because of the previous results, DNp73 exhibits antiapoptotic activity in human melanoma cells and specific suppression of individual isoforms enhances the sensitivity towards cytotoxic drugs (14). Investigating mechanisms responsible for dysregulated E2F1 losing its apoptotic function in human skin cancer, we identified miR-205 as a specific target of p73 and found that upon genotoxic stress its expression is sufficiently abrogated by endogenous DNp73 (15). Significantly, we showed that metastatic cells can be rescued from drug resistance by selective knockdown of DNp73 or overexpression of miR-205 in p73 depleted cells, leading to increased apoptosis and the reduction of tumor growth *in vivo*. These results suggest the E2F1-p73/DNp73-miR-205 network as a crucial mechanism for chemoresistance and thus as a target for prevention of metastasis.

The structure of the E2F1-centered biochemical network that mediates the potential resistance against anticancer therapies is a complex system involving signaling and transcriptional regulation. This network is enriched with network motifs, including feedback and feedforward loops. Furthermore, these network motifs are well interconnected, overlap, and cross-talk with other cancer-related signaling pathways (4, 16). Networks containing several of these motifs often show nonintuitive regulatory patterns, which require the use of mathematical modeling to understand their function and regulation (17,18).

In this work, we constructed a regulatory map that summarizes the current literature on E2F1 signaling dysregulation in cancer and complemented it with information about E2F1 and miR-205 targets, which have known or putative roles in drug response. On the basis of this map, we derived a mathematical model and used it to investigate under which conditions dysregulation of this network provides tumor cells with resistance mechanisms against several families of anticancer drugs. In connection to this, we further explored the consequences of tumor heterogeneity in the effectiveness of these therapies. Our results indicate that kinetic modeling under the systems biology paradigm can be an effective method to detect genetic signatures providing tumor cells with chemoresistance mechanisms and thereby to design therapeutic strategies to counteract them.

Materials and Methods

Cell culture, PCR, Western blot, and apoptosis assays

The human melanoma cell line SK-Mel-147 characterized by morphology and cytogenetics was obtained from M. Soengas (Department of Dermatology, University of Michigan, Comprehensive Cancer Center, Ann Arbor, MI). Tested and authenticated cells were cultured as described (15). Plasmid miRZIP-205 encoding antagomir-205 was purchased from BioCat. Transfections were conducted with Lipofectamine 2000 (Invitrogen). Adenoviral vector with shE2F1 was described previously (10). Expression of hsa-miR-205 and hsa-RNU6B (for normalization) was detected using specific TaqMan assays (Applied Biosystems). PCR, Western blot, and apoptosis assays were conducted as described. The wet lab data were obtained previously and are published in (15).

Regulatory map construction

Information about the network components and their interactions was extracted from relevant published reports and databases (HPRD and STRING for protein–protein interactions; TRED for E2F1 transcriptional targets; TarBase 6.0, miRTarBase, miRecords and miR2Disease for validated targets of miR-205; and TransmiR for transcription factors regulating the expression of miR-205). Putative miR-205 targets and miR-205 transcription regulators were extracted from additional databases (see Supplementary Materials and Methods for a detailed description of the data retrieval process). As we were interested in the role of E2F1 and miR-205 in pro- and anti-apoptotic processes as well as in drug responses, we filtered the obtained set of molecular interactions based on the corresponding Gene Ontology (GO) terms (GO:0043065, GO:0043066 and GO:0042493). Finally, all the information collected was integrated into a regulatory map in Cytoscape (available in Supplementary Materials and Methods).

Kinetic modeling

Representative interactions in the constructed regulatory map were translated into a kinetic model based on ordinary

Published OnlineFirst February 27, 2013; DOI: 10.1158/0008-5472.CAN-12-4095

differential equations. The model constructed accounts for the dynamics of miRNA, mRNA and protein expression levels, as well as cell populations in different cancer-relevant scenarios, including tumor proliferation and genotoxic or cytostatic drug administration. The model has the following general structure:

$$\frac{dmiRNA_i}{dt} = \sum_{k1} D_{k1}\left(GS, Prot_j, mRNA_i, miRNA_i\right);$$
$$\frac{dmRNA_i}{dt} = \sum_{k1} F_{k1}\left(GS, Prot_j, mRNA_i, miRNA_i\right);$$

$$\frac{dProt_i}{dt} = \sum_{k2} G_{k2}\left(GS, Prot_i, Prot_j, mRNA_i\right);$$
$$\frac{dC_i}{dt} = \sum_{k3} H_{k3}\left(C_i, Prot_j\right),$$

where $Prot_i$ accounts for the expression level of the proteins involved in the network, $mRNA_i$ for the messenger RNAs mediating their synthesis, $miRNA_i$ for miRNAs repressing expression of given proteins, and C_i for the size of proliferative, arrested, or apoptotic cell populations. *GS* is an input signal that accounts for the effect of genotoxic or cytostatic drugs in the synthesis and regulation of proteins and mRNAs. D_{k1}, F_{k1} and G_{k2} account for the description of biochemical processes as rate equations that depend on the expression levels of proteins and mRNAs, or the time-dependent genotoxic stress signal *GS*. H_{k3} are rate equations describing the dependence of cell population dynamics on the expression level of critical proteins in the network. Values of the model parameters characterizing the reaction rates were assigned following a hybrid strategy, composed of: (i) extracting parameter values from published information (e.g., protein, mRNA, and miRNA half-lives); (ii) estimating a subset of parameter values by tuning them to fit published quantitative and qualitative data; (iii) reducing some parameters by normalizing variables around the basal, nonstressed levels of mRNA and protein. Model equations and chosen parameter values are provided in Supplementary Materials and Methods.

In silico cell lines. To search for genetic signatures providing tumor cells with a chemoresistant phenotype, we produced *in silico* (kinetic model-based) cell lines by randomly perturbing values of selected model parameters using Latin hypercube sampling. Thereby, we generated 10^5 *in silico* cell lines, each one with a distinctive set of parameter values accounting for an individual genetic background.

Model simulations

For each *in silico* cell line, we conducted simulations in 3 cancer relevant scenarios: (i) nonstressed growth conditions; (ii) apoptosis under; and (iii) proliferation under cytostatic drug administration. We investigated genetic signatures providing the tumor cells with a chemoresistant phenotype, by randomly perturbing parameter values and simulating the response of the model to different cancer-relevant scenarios. We analyzed the effects of perturbations in the expression levels of critical network components by iteratively modifying their synthesis rate constants accounting for their induction and simulating the behavior of the network. We assessed the efficacy of conventional anticancer treatment in *in silico* cells with different E2F1-related genetic signatures by simulating periodic cycles of genotoxic or cytostatic drug injection. Model calibration, computational simulations, and data analysis were conducted using MATLAB running on a high-end workstation Fujitsu Celsius V840, 2x CPU AMD Opteron 2214, 2.2 GHz. ECC. Complete information about the construction of the regulatory map, as well as about the model structure, calibration, and analysis is included in Supplementary Materials and Methods.

Results

A network with multiple feedforward loops determines the E2F1 drug response

Our recent results suggest that the E2F1-p73/DNp73-miR-205 pathway is involved in the regulation of pro- and antiapoptotic genes, which confers chemoresistance (15). We investigated whether this system is part of a wider and more complex network regulating the response of tumor cells to anticancer drugs. Towards this end, we constructed a regulatory map that contains the E2F1-p73/DNp73-miR-205 pathway as core module and includes genes targeted by E2F1 or miR-205, with known or putative relation to the drug response and drug-induced apoptosis (Fig. 1A). Overall, the map unravels a tightly interlocked regulatory system. Computational analysis of the network (81 nodes, 194 edges) shows a robust structure and a topology that is composed of many regulatory motifs, especially enriched in feedforward loops (for details, see Supplementary Materials and Methods). Interestingly, we detected several instances of network motifs in which a direct E2F1 target gene interacts with a miR-205–repressed target. Because E2F1 regulates miR-205 via p73, processes downstream of the interaction between both targets are regulated by E2F1 following 2 independent routes and can be considered as feedforward loop (Fig 1A). Several of those loops are associated with the efficient apoptosis initiation and involve the positive regulation of proapoptotic genes by E2F1 and the repression of antiapoptotic genes by miR-205. For example, the E2F1 target Hrk regulates apoptosis through interaction with the miR-205 targets BCL-2 and BCL-XL (Fig 1A; ref. 19). In addition, we detected at least one additional instance of this motif, in which E2F1 promotes EGFR expression and miR-205 represses ERBB3 (both belonging to the HER receptor family). In several tumors, both receptors are associated with abnormal proliferation and upon heterodimerization they influence sensitivity to cytostatic drugs like erlotinib (20). Furthermore, DNp73 can inhibit the expression of p73 transcriptional targets, including miR-205. The expression of these targets may depend on a tight trade-off between p73 and DNp73 expression and activity, which are both transcriptionally regulated by E2F1. These E2F1-p73/DNp73-target motifs can be considered as incoherent feedforward loops. Overall, the activation of miR-205 and therefore of its targets may depend on the interplay among E2F1, p73, DNp73, and other regulators of miR-205 like TWIST1

Published OnlineFirst February 27, 2013; DOI: 10.1158/0008-5472.CAN-12-4095

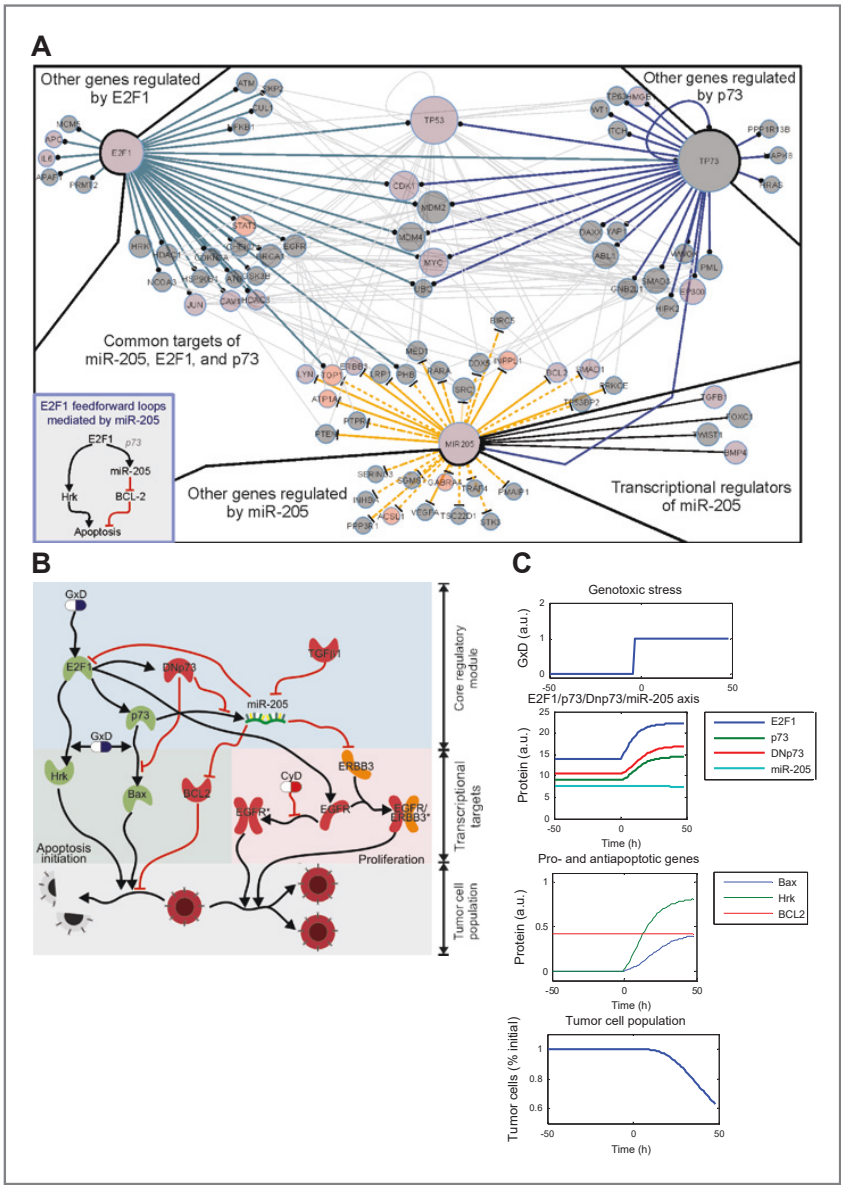


Figure 1. A, E2F1-p73/DNp73-miR-205 regulatory map. This Cytoscape figure shows the targets of E2F1, p73, and miR-205 as well as transcription factors of miR-205 that are involved in apoptosis (gray nodes), drug response (pink nodes), or both (purple nodes) according to GO. Edges are color coded with respect to the following processes: transcriptional regulation (dark and light blue), posttranscriptional regulation (orange), and protein–protein interaction (gray), whereas solid lines represent validated interactions and dashed lines denote predicted interactions. B, kinetic model for the E2F1-p73/DNp73-miR-205 network. The model is composed of four modules: The core regulatory module (blue) describes the dynamics of E2F1, p73, DNp73, and miR-205. The transcriptional targets module accounts for representative apoptosis and proliferation-associated targets of E2F1, p73, and miR-205, which may play a role in drug response. The tumor cell population module connects the biochemical network to the fate of a population of tumor cells. GxD accounts for genotoxic and CyD for cytostatic drug; red-colored symbols for oncogenes and green for tumor suppressors. Twelve ordinary differential equations accounting for the evolution in time of: E2F1 mRNA (in the model represented with the variable mE2F1) and protein (E2F1), both isoforms of p73 (respectively, p73 and DNp73), miR-205 (miR205), E2F1-regulated proapoptotic genes (Hrk), p73-regulated proapoptotic genes (Bax), miR-205–repressed antiapoptotic genes (BCL2), E2F1-regulated and active EGFR (EGFR*), cytostatic drug–inhibited EGFR (EGFRI), and miR-205–repressed ERBB3 (ERBB3). In addition, we consider another differential equation accounting for the population size of tumor cells with the genetic background defined by our model (TC). The node size is in accordance to its degree (i.e., number of edges connected). C, prototypical simulation made with the model. Upon administration of a drug at $t = 0$, genotoxic stress is triggered (top). This activates the E2F1-p73/DNp73-miR-205 network (middle), which in turn promotes the transcription of apoptosis-related genes (middle). When proapoptotic genes are sufficiently expressed, they trigger apoptosis and the population of tumor cell decreases (bottom).

Published OnlineFirst February 27, 2013; DOI: 10.1158/0008-5472.CAN-12-4095

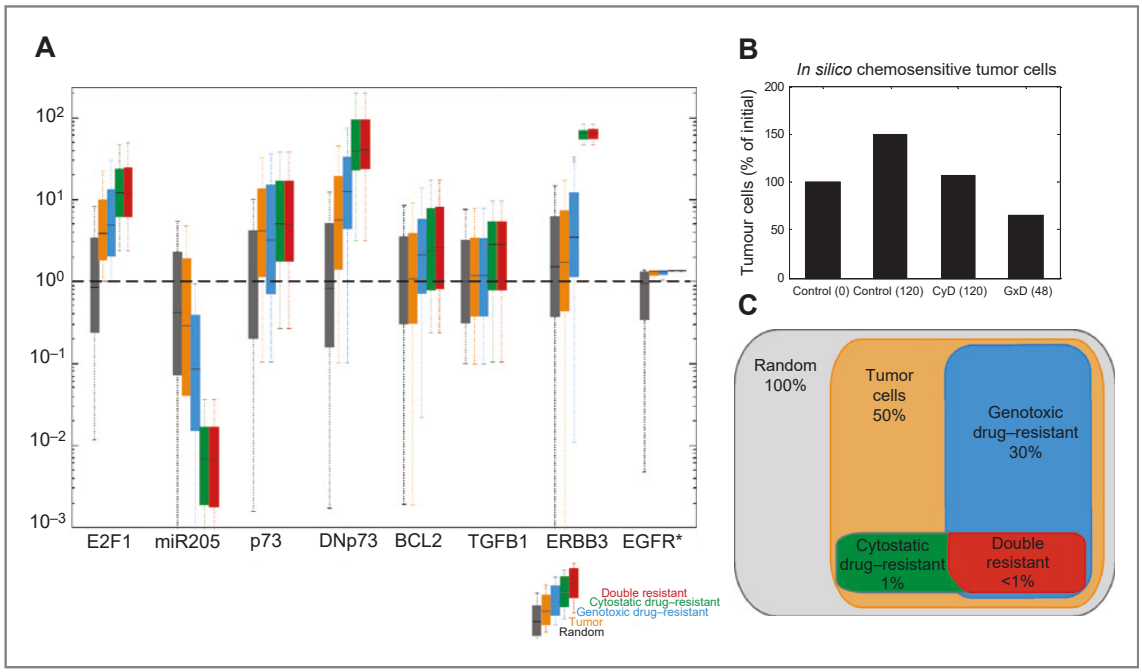


Figure 2. A, genetic signatures of the E2F1-p73/DNp73-miR-205 network that confer chemoresistance. Box-and-whisker plots of simulated nonstress levels for E2F1, p73, DNp73, miR-205, BCL2, TGFβ-1, ERBB3, and EGFR for the different subsets of *in silico* cells generated: gray, entire set of randomly generated solutions; orange, tumor cells; blue, genotoxic drug–resistant; green, cytostatic drug–resistant; red, double drug–resistant. In each bar, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points. The dashed horizontal line in the plot accounts for the values in the nominal chemosensitive genetic signature. B, phenotype of the *in silico* chemosensitive tumor cell line. Percentage of cells with respect to initial population. Control(0): population size at 0 hours, no drug injection; control(120): population at 120 hours, no drug injection; CyD(120): 120 hours after cytostatic drug; GxD (48): 48 hours after genotoxic drug. C, distribution of the initial set of randomly generated 10^4 *in silico* cell lines in the different subpopulations. Percentage over the total is indicated.

and TGFβ-1 signaling, as suggested by our regulatory map. Because of the network complexity, mathematical modeling is necessary to investigate nonintuitive regulation of this system in cancer-relevant scenarios.

A modularized kinetic model to investigate E2F1-p73/DNp73-miR-205 network chemosensitivity

We set up and analyzed a kinetic model using ordinary differential equations. To construct the model, we selected parts of the regulatory map that are representative for the network response to different anticancer drugs (Fig. 1B). The model is organized into four interconnected modules: the core regulatory module, two modules of transcriptional targets, and one for tumor cell population. The core regulatory module of the kinetic model is composed of E2F1, both isoforms of p73 (respectively, p73 and DNp73) and miR-205 (miR-205) and accounts for their temporal evolution under drug administration (15). We include here the transcriptional regulation of miR-205 expression by the known oncogene TGFβ-1 (TGFB1), which exemplifies the set of regulators of miR-205 expression external to the core module. The inclusion of this variable allows for the investigation of cross-talk, via miR-205, between E2F1 signaling and other oncogenic signals involved in the epithelial–mesenchymal transition (EMT). EMT has been associated with the emergence of chemoresistance (21).

In addition, the model includes a module for pro- and antiapoptotic transcriptional targets, which are regulated by the network components in response to genotoxic stress. Precisely: (i) for representing the dynamics of proapoptotic proteins whose expression is promoted by E2F1, we included a variable accounting for Harakiri expression (22); (ii) similarly, for proapoptotic proteins whose expression is positively regulated by p73 and negatively regulated by DNp73, a variable accounting for Bax expression is defined (23); and finally (iii) for antiapoptotic proteins whose expression is repressed by miR-205, we chose BCL2 as a representative and defined a corresponding variable (15). These pro- and antiapoptotic proteins are used here as an example and represent a wider set of proteins which undergo similar regulation. This kind of simplification has been successfully applied to reduce the complexity of models of biochemical networks, for example, in the study conducted by Aguda and colleagues (24). Furthermore, the model includes a module for receptors whose activity is regulated in response to cytostatic drugs. This module includes variables accounting for the dynamics of EGFR, whose expression is promoted by E2F1 (10), and ERBB3, which is

Published OnlineFirst February 27, 2013; DOI: 10.1158/0008-5472.CAN-12-4095

repressed by miR-205 (25,26). In several tumors the heterodimerization of these receptors has been linked to abnormal cell proliferation and changes in the sensitivity of tumor cells to cytostatic drugs (20). Additional input variables account for the effect of genotoxic (GxD) and cytostatic (CyD) drugs triggering DNA damage-E2F1–mediated apoptosis and abolishment of tumor cell proliferation, respectively (4).

Finally, the last module includes a phenomenological kinetic equation, which connects the components of the network to the dynamics of the tumor cell population. In this equation, the representatives for pro- and antiapoptotic proteins regulate the cell death rate after genotoxic stress, whereas EGFR and ERBB3 control the tumor cell proliferation rate. Taken together, our model is able to describe the cascade of regulatory events after drug administration including stress signals, downstream intracellular responses, the regulation of pro- and antiapoptotic signals (for genotoxic drugs) or proliferative signals (for cytostatic drugs), and finally the global effect on tumor cell population (Fig. 1C; see Supplementary Materials and Methods for a complete description of the model).

Genetic signatures for the E2F1-p73/DNp73-miR-205 network that confer chemoresistance to anticancer drugs

In our analysis, we considered two families of drugs: genotoxic drugs, like doxorubicin and cisplatin, inducing apoptosis; and cytostatic drugs, like erlotinib and lapatinib, repressing tumor cell proliferation. In some scenarios, it has been observed that these drugs lose efficacy when some of the network components are dysregulated (15,27). With the help of the kinetic model, we analyzed whether specific genetic signatures of the core module can be linked to a phenotype response of resistance to these anticancer drugs. We define a genetic signature as a group of genes in a tumor cell whose combined expression pattern is linked to a specific phenotype (i.e., chemoresistant or chemosensitive). In line with this, we designed a nominal *in silico* chemosensitive genetic signature, inspired by that of SK-Mel-147 (15). This signature accounts for the expression of the network components for which an *in silico* cell line holds the phenotype of a chemosensitive tumor cell; that is, abnormal proliferation but responsiveness to genotoxic drugs (via triggering of apoptosis) and cytostatic drugs (via inhibition of proliferation, see Fig. 2B).

To detect genetic signatures providing chemoresistance, we generated a population of 104 model configurations obtained by random perturbation of the model parameter values of the nominal chemosensitive genetic signature. In the following, we define these distinct configurations as in *silico cell* lines. For each one of them, we simulate the model and obtained the expression levels of the network components and the tumor cell population in three cancer-relevant scenarios: (i) under nonstress conditions, at time zero, and after 120 hours; (ii) after genotoxic drug administration, at 48 hours; and (iii) after cytostatic drug administration, at 120 hours. The *in silico* cell lines were classified into the following groups: (i) cell lines that show abnormal proliferation in nonstress conditions, named as tumor cells; (ii) those resistant to genotoxic drugs; (iii) those resistant to cytostatic drugs; and (iv) those resistant to both genotoxic and cytostatic drugs. For each group, the data describing the expression levels of the network components at time zero were normalized with respect to the nominal chemosensitive values (Fig. 2C). For the computational definition of the simulation scenarios and classification of the resulting cell phenotypes, see Supplementary Materials and Methods.

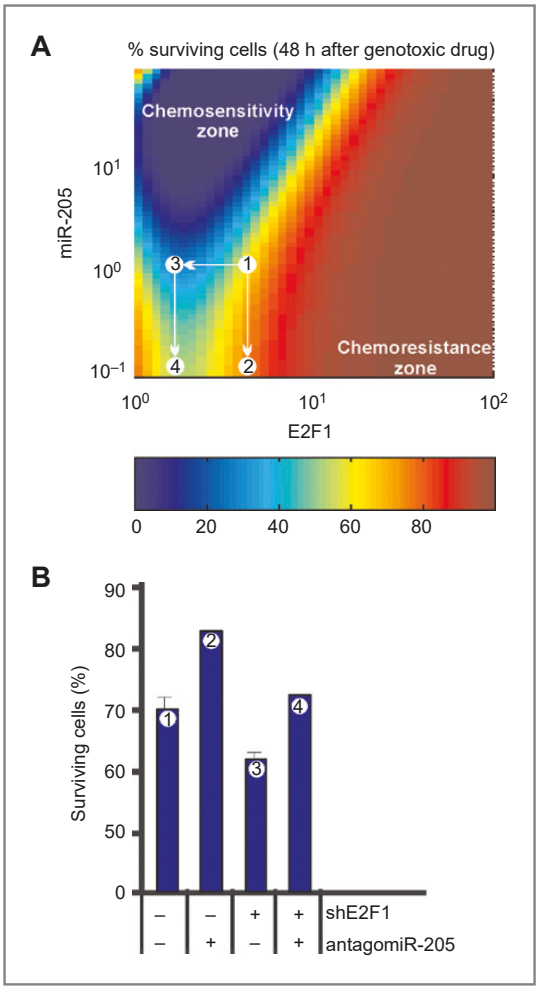


Figure 3. The interplay between E2F1 and miR-205 promotes resistance to genotoxic drugs. A, model simulations. For different values of the synthesis rates for E2F1 and miR-205 (k_{E2F1}, $k_{mir-205} \in [10^{-1}, 10^{1}]$), we determined the percentage of apoptotic cells 48 hours after genotoxic stress: Apoptotic cells $= 100 * [1- TC(t = 48)]$, where *TC* is the number of surviving tumor cells at 48 hours. B, validation experiments with SK-Mel-147 cells. The percentage of apoptotic cells was measured 48 hours after cisplatin treatment. 1, control; 2, after miR-205 knockdown; 3, after E2F1 knockdown; and 4, after miR-205 and E2F1 knockdown. Data were extracted and processed from Alla and colleagues (15).

Our results indicate that 30% of the *in silico* cell lines are resistant to genotoxic drugs (Fig. 2C). These cells display high basal levels (at time zero, nonstress conditions) of E2F1 and

Published OnlineFirst February 27, 2013; DOI: 10.1158/0008-5472.CAN-12-4095

DNp73, whereas miR-205 is downregulated compared with the nominal chemosensitive tumor cell and the group of *in silico* tumor cells (Fig. 2A). In addition, approximately 1% of the cell lines display resistance to cytostatic drugs. These cells have a genetic signature composed of much higher basal E2F1, DNp73, and ERBB3 levels, whereas miR-205 appears strongly downregulated and TGFβ-1 is moderately upregulated. In both cases, EGFR is expressed at its maximum level, but our analysis indicates that this is a property already acquired by the group of tumor cells as a consequence of high levels of E2F1 (see Supplementary Materials and Methods). Finally, a small fraction of the cell lines display a phenotype of double resistance and their genetic signature is very similar to that of cytostatic drug–resistant cells.

Taken together, our results suggest a relevant role for E2F1, DNp73, and miR-205 in the regulation of resistance to several anticancer drugs. To validate this model-based prediction, we compared them with our recently published data on the interplay of E2F1 and miR-205 under genotoxic stress administration (15). To this end, we conducted simulations in which we iteratively perturbed the values of the parameter accounting for E2F1 and miR-205 induction and computed the percentage of apoptotic cells 48 hours after genotoxic drug administration. As shown in Fig. 3A, *in silico* cell lines with high levels of E2F1 display lower percentage of apoptotic cells in response to chemotherapy. Our previous experiments with melanoma SK-Mel-147 cells confirm the simulation results (Fig. 3B). SK-Mel-147 cells were treated with cisplatin and high percentage of cells became apoptotic (Scenario 1 in Fig. 3B). Inhibition of miR-205 with antagomir-205 reduces the percentage of apoptotic cells after treatment (Scenario 2), whereas selective knockdown of endogenously high E2F1 leads to a pronounced increase in apoptosis (Scenario 3). Finally, repression of both E2F1 and miR-205 reduces the apoptotic effect of cisplatin (Scenario 4). These experimental results confirm the model predictions displayed in Fig. 3A for the scenarios named with the same numbers.

To verify whether dysregulation of couples of network components affects chemosensitivity to genotoxic and cytostatic drugs, we conducted additional simulations in which we iteratively modified the values of the protein synthesis rates (Fig 4). Precisely: (i) because the p73/DNp73 ratio is known to play a major role in the aggressiveness of cancer cells (11, 28, 29), we iteratively modified the parameters accounting

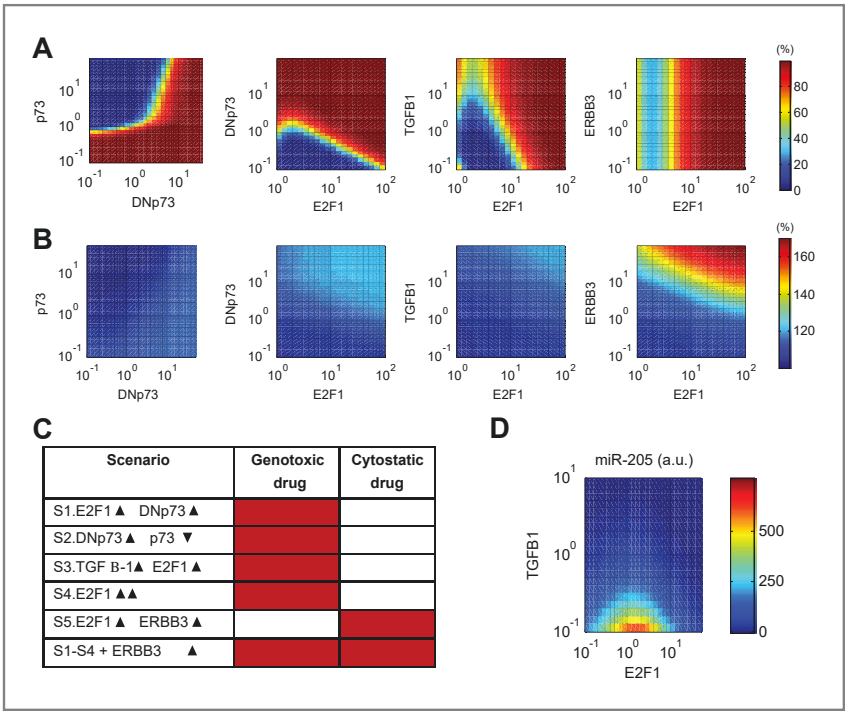


Figure 4. A, genotoxic drug resistance for different protein synthesis rates. In all cases, we iteratively modified the synthesis rates for pairs of network components and computed the percentage of tumor cells surviving 48 hours after genotoxic drug administration. B, cytostatic drug resistance for different protein synthesis rates. The percentage of tumor cells surviving 120 hours after cytostatic drug administration was computed. *x*- and *y*-axes account for normalized changes in the indicated protein synthesis rates (10^0 represents the protein levels in an *in silico* cell line that has the nominal chemosensitive genetic signature). We note that E2F1 is often upregulated in tumors to achieve enhanced proliferation [Alla and colleagues, (10)]. Thus, in our simulations, E2F1 synthesis rate must be equal or greater than 10^0. C, protein expression signatures providing chemoresistance. Compared with the values in the nominal chemosensitive genetic signature: ▲ accounts for overexpression (10^1); ▲▲ for extreme overexpression (10^2); and ▼ for repression (10^{-1}). D, multiple transcriptional signals regulate miR-205.

Published OnlineFirst February 27, 2013; DOI: 10.1158/0008-5472.CAN-12-4095

for p73 and DNp73 synthesis; ii) given that overexpression of E2F1 is associated with aggressiveness and chemoresistance (4,15), we also considered iterative modulation of E2F1 and DNp73 synthesis parameters to test whether their dysregulation suffices to induce resistance; (iii) considering that dysregulation of TGFβ-1 signaling promotes invasion and metastasis (30), we modified the E2F1 synthesis parameter and TGFβ-1 expression level to investigate whether they can synergize in the emergence of a resistance phenotype; and (iv) given that the expression of ERBB3 associates to the resistance to certain cytostatic drugs (20,31,32), we modified E2F1 and ERBB3 synthesis parameters to test whether their combined dysregulation can cause enhanced chemoresistance.

As shown in Fig. 4A, *in silico* cell lines with strong DNp73 and/or weak p73 synthesis display a very high percentage of cells resisting apoptosis after genotoxic drug administration, whereas strong p73 synthesis and reduced DNp73 induction provide efficient cell death. Furthermore, our kinetic simulations suggest that very strong E2F1 induction alone is sufficient to provoke chemoresistance to genotoxic drugs and can therefore compensate low DNp73 levels (Fig. 4B). In addition, high induction of TGFβ-1 reduces the level of E2F1 required to induce resistance. Taken together, strong E2F1 synthesis can synergize with strong induction of DNp73 and TGFβ-1 to provide genotoxic drug resistance (Fig. 4A). When analyzing the emergence of resistance to cytostatic drugs through the considered network, our analysis indicates that strong ERBB3 and E2F1 induction can synergize to achieve chemoresistance (Fig. 4B). Taken together, the results suggest that resistance to both genotoxic and cytostatic drugs can emerge with strong ERBB3 and E2F1 induction. We summarized our results in a qualitative manner in Fig. 4C, which lists possible genetic signatures conferring chemoresistance. We note that in our model, the regulation of antiapoptotic genes like BCL-2 and ERBB3 by E2F1, p73/DNp73, and TGFβ-1 is mediated via miR-205. Therefore, we hypothesize that the coregulation of miR-205 by its transcriptional regulation can foster chemoresistance (Fig. 4D).

Conventional genotoxic drug treatment favors selection of chemoresistant cells in genetically heterogeneous tumors

It is a well-established fact that cell populations with different genetic signatures influencing cancer hallmarks coexist in malignant tumors (33). In line with this, we used our kinetic model to investigate how tumor heterogeneity (i.e., coexistence of several populations with different genetic signatures) affects the efficacy of conventional anticancer therapy (Fig 5). In our analysis, we focused on genotoxic drugs. Towards this end, we reformulated our model to consider two kinds of tumors: (i) tumors with homogeneous expression of E2F1 and (ii) tumors with heterogeneous expression of E2F1. In our simulations, heterogeneity in E2F1 expression is modeled by making 90% of cells with a defined E2F1 level, 5% cells with 0.5-fold down-regulation, and 5% cells with 2.5-fold upregulation of E2F1. In accordance with our previous simulations, we considered two biologic scenarios: E2F1 overexpression (10^1, represented by E2F1▲) and extreme overexpression (10^2, represented by E2F1▲▲).

We used the model to simulate the effect of an anticancer treatment composed of periodic injections of a conventional genotoxic agent. In a tumor with homogenous E2F1 overexpression, our results indicate that the tumor is abolished after several injection cycles (Fig. 5 top left). However, in a

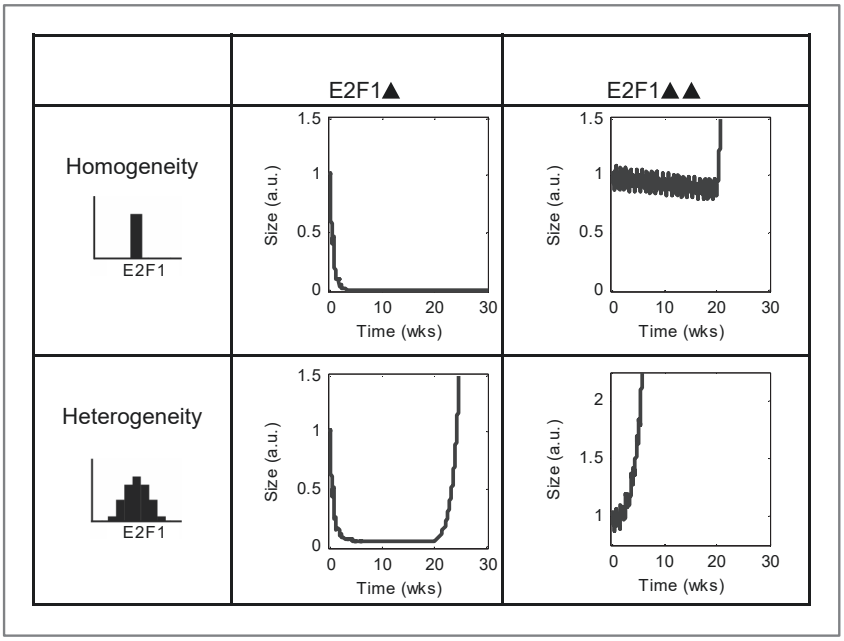


Figure 5. Effectiveness of genotoxic drug treatment in tumors with homo- and heterogeneous genetic signatures for E2F1. We simulated the effect of an anticancer treatment composed of periodic injections of a conventional genotoxic agent, twice a week during 20 weeks (see Supplementary Materials and Methods for details). The size of the tumor was assumed 1 (n.u.) at time zero. Plots in the first row illustrate the treatment-dependent tumor growth for tumors with homogeneous E2F1 levels, whereas the second row accounts for tumors with heterogeneous E2F1 levels. First column accounts for scenarios with E2F1 overexpression (10^1, represented by E2F1▲) and the second column shows extreme overexpression (10^2, represented by E2F1▼▼).

Published OnlineFirst February 27, 2013; DOI: 10.1158/0008-5472.CAN-12-4095

tumor with homogenous E2F1 extreme overexpression, our simulation suggests that the tumor is resistant to the therapy and recovers fast growth afterwards (Fig. 5 top right). Interestingly, in a tumor with heterogeneous E2F1 overexpression, the tumor becomes virtually abolished after several cycles of drug injection, but the tumor reemerges after the drug injections are discontinued (Fig. 5 bottom left). Finally, for a tumor with heterogeneous E2F1 extreme overexpression, model simulations indicate that the tumor is fully drug resistant with continuous proliferation (Fig. 5 bottom right).

We further analyzed the scenario of heterogeneous E2F1 overexpression and found that in our simulations during the treatment, the composition of the tumor shifts from one of predominantly chemosensitive cells at time point zero (Fig. 6A, dark blue curve) towards a composition of entirely genotoxic-drug resistant cells at twenty weeks (Fig. 6A, red curve). This suggests that the genotoxic drug treatment can induce selective pressure on tumor cells. This is because chemosensitive cells undergo apoptosis during treatment, whereas cells with the resistant genotype withstand and continue proliferating. Therefore, the tumor reemerges after discontinuation of the treatment after 20 weeks (Fig. 6B). Our simulations indicate that during treatment the genetic signature of E2F1 and miR-205 in the tumor shift from that of chemosensitive cells to the one in chemoresistant cells (upregulation of E2F1 and downregulation of miR-205, Fig. 6C). Strikingly, we were able to confirm our prediction with experiments, in which we stimulated SK-Mel-147 cells with iteratively increasing levels of cisplatin (1–50 μmol/L) and measured E2F1 and miR-205 expression before and after the treatment with this genotoxic drug (respectively ① and ② in Fig. 6C; ref. 15).

Discussion

E2F1-p73/DNp73-miR-205 network mediates chemoresistance

We developed a kinetic model to investigate the emergence of chemoresistance in tumor cells, mediated by a network with a core module composed of E2F1, p73, DNp73, and miR-205. In previous efforts, mathematical modeling was applied to address the role of E2F1 and other network components in cancer and cell proliferation (24, 34–36). However, to our knowledge, this work is the first modeling-based analysis of E2F1-mediated chemoresistance.

Our results suggest that dysregulation of this network can provoke chemoresistance to multiple anticancer drugs. In line

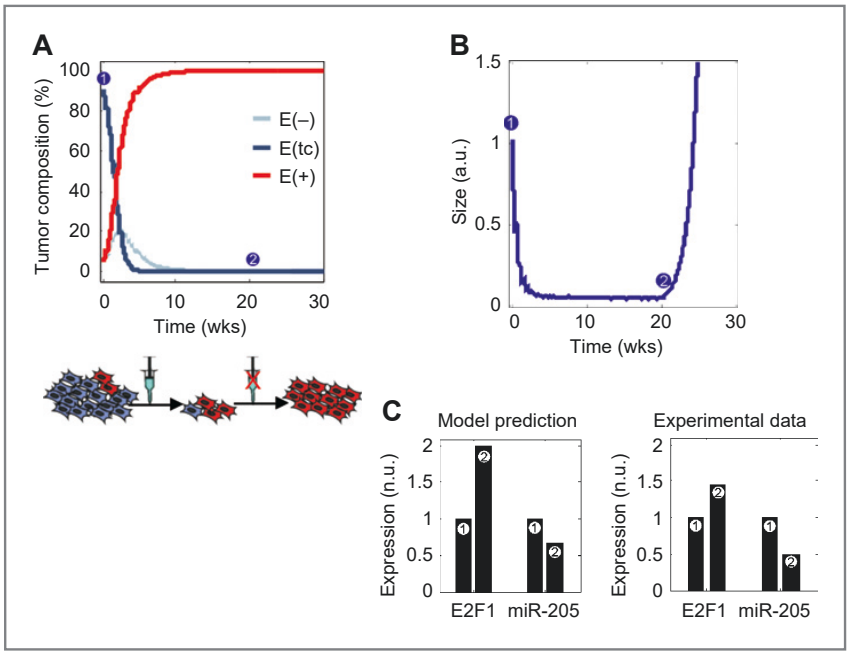


Figure 6. Selection of chemoresistant clones after conventional genotoxic drug treatment in tumors with heterogeneous genetic signature for E2F1. We simulated the effect of periodic injections of a conventional genotoxic agent over an *in silico* tumor, with a heterogeneous genetic signature for E2F1. The average E2F1 expression level, $E2F1^{av}$, is similar to that in the *in silico* chemosensitive tumor cell line; 90% cells display $E2F1 \approx E2F1^{av}$; 5% cells with $E2F1 \approx 0.5 \cdot E2F1^{av}$; 5% cells with $E2F1 \approx 2.5 \cdot E2F1^{av}$. A, time evolution of tumor cell composition during drug treatment. Dark blue curve: percentage of cells with $E2F1 = E2F1^{av}$ (E(tc)). Light blue curve: percentage of cells with $E2F1 \approx 0.5 \cdot E2F1^{av}$ (E(-)). Red curve: percentage of cells with $E2F1 \approx 2.5 \cdot E2F1^{av}$ (E(+)). B, time evolution of the tumor size during drug treatment; the numbered circles indicate the time, when E2F1 expression was measured (1 = before; 2 = after treatment). C, normalized E2F1 and miR-205 levels before ① and after ② injection of repeated cycles of genotoxic drug as predicted by our model. Bottom right, SK-Mel-147 cells stimulated with increasing levels of cisplatin (1–50 μmol/L). E2F1 and miR-205 levels were measured before ① and after ② treatment. Relative miR-205 expression was measured with TaqMan PCR, whereas E2F1 levels were analyzed by Western blot analysis, both data sets were normalized with respect to initial prestress values (more details in ref. 15).

Published OnlineFirst February 27, 2013; DOI: 10.1158/0008-5472.CAN-12-4095

with this, our model unraveled several genetic signatures of tumor cells that can promote chemoresistance. Regarding genotoxic drugs, we found a signature of high E2F1 and low miR-205 expression that promotes resistance. Downregulation of miR-205 in this signature can be mediated by an imbalance in the p73/DNp73 ratio or by the dysregulation of other cancer-related pathways like TGFβ-1 signaling, whose components positively or negatively regulate miR-205 expression (37,38).

With respect to the imbalance in the p73/DNp73 ratio, our simulations predict that tumors with high E2F1 and normal/high DNp73 expression show a higher percentage of surviving cells after genotoxic stress than tumors expressing high E2F1 and low DNp73. This supports the idea that DNp73-mediated repression of miR-205 is a mechanism by which drug resistance is achieved. Because DNp73 is promoted by E2F1, high levels of E2F1 can induce chemoresistance via the E2F1-DNp73-miR-205 axis due to the negative regulation of miR-205 by DNp73 (Fig. 7A). However, DNp73 is upregulated in melanoma metastases with a concomitant rise of the full-length p73, a positive regulator of miR-205 (13). Taken together, this module has the structure of an incoherent feedforward loop. This suggests that the specific ratio between both p73 isoforms expression and activity determines the functional outcome of E2F1 and cell fate upon genotoxic treatment. This ratio is regulated in malignant melanoma by as yet unknown factors involved in alternative splicing of p73 pre-mRNA, possibly independent of E2F1 (39). In tumor cells with high E2F1 induction, our analysis indicates that downregulation of DNp73 and/or upregulation of p73 can restore chemosensitivity, which emphasizes the importance of the p73/DNp73 ratio.

When constructing the network, we found that some proteins involved in the epithelial-to-mesenchymal transition (EMT), like TGFβ-1, BMP, or TWIST1 can regulate miR-205 expression in some tumors (38,40). Furthermore, our simulations indicate that when E2F1 is upregulated, high levels of TGFβ-1 or other EMT-related signals regulating miR-205, external to the E2F1 network, can provoke genotoxic drug resistance in cancer cells. This suggests an intriguing hypothesis, in which E2F1 signaling cross talks and synergizes with some EMT-related signals, and their regulation of common miRNAs can mediate this synergy in the context of chemoresistance (Fig. 7B).

Moreover, our analysis gives additional insights into the resistance to cytostatic drugs that inhibit the activation of HER receptor family members. We found a genetic signature composed of high E2F1, low miR-205, and high ERBB3 that can render tumor cells insensitive to cytostatic drugs. As cited before, miR-205 posttranscriptionally represses ERBB3. Interestingly, recent results suggest that tumor cell lines sensitive to erlotinib show high levels of miR-205 and other miRNAs (41), which supports our predictions. According to our model analysis, tumor cells with E2F1 overexpression and miR-205 downregulation do not show cytostatic drug resistance, but synergy with high *ERBB3* gene activation is required to acquire this feature. In addition, the simulations suggest that in most of the cases, this genetic signature can lead to double resistance, this means the tumor cells become insensitive to some genotoxic and cytostatic drugs at the same time, a feature that may have important consequences in the design of personalized cancer treatments. The underlying molecular mechanism providing this double resistance is to be further elucidated.

Taken together, our results indicate that differences in the expression of some of the proteins in this network are critical for tumor cells to become chemoresistant, and that this resistance is substantiated through regulatory loops including

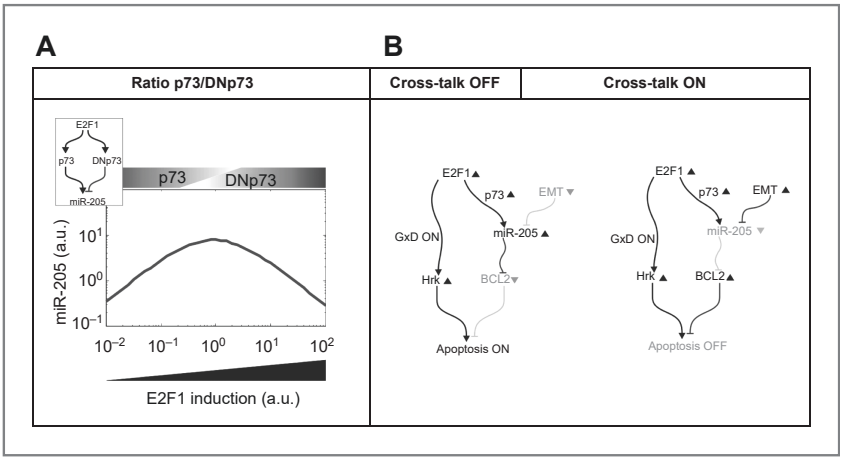


Figure 7. A, E2F1-regulated shift of the p73/DNp73 ratio can induce nonmonotonic, nonlinear regulation of miR-205 expression. The peak of miR-205 expression is reached for intermediate E2F1 levels and downregulation for very high E2F1 expression. This system is an incoherent feedforward loop and, according to our hypothesis, nonlinear regulation of the branches in the feedforward loop switches the status of the feedforward loop from one in which the positive branch predominates (e.g., p73 ▲) and promotes target expression to another in which the negative branch predominates (e.g., DNp73▲) and the target expression is repressed. B, feedforward loops involved in the E2F1-mediated chemosensitivity may be active regulatory structures, whose activation may depend on the cross-talk between E2F1 and other cancer-related signals.

Published OnlineFirst February 27, 2013; DOI: 10.1158/0008-5472.CAN-12-4095

miR-205. We note that the predictions we made are based on qualitative modeling, similar to those made in the studies conducted by Aguda and colleagues and Yao and colleagues (28,39). To make the predictions of our model quantitative, the model should be further refined in iterative cycles of experimentation and model calibration (17,18).

In the model parameterization analyzed, the feedback loop between E2F1 and miR-205 has minor effects in the dynamics of the network. Our previous results of modeling miRNA regulation of cancer-related genes suggest that the strength of miRNA regulation is context-dependent and may be affected by the coregulation of other miRNAs or proteins (42). Thus, we hypothesize that this feedback loop could be negligible in the scenarios investigated in this article, but still crucial in other biologic contexts which have not been investigated in this paper. In this way, miR-205–mediated repression of some antiapoptotic targets in our network may be affected by similar coregulation. Our modeling results complement those of Aguda and colleagues (24) and others (43) who also used kinetic modeling to investigate the role of miRNA regulation on E2F1 in cancer. Aguda and colleagues (24) found that the coupling between the E2F/Myc positive feedback loop and the E2F/Myc/miR-17-92 negative feedback loop is crucial to regulate a bistable on–off switch in E2F/Myc protein levels, a feature that is critical in the dysregulation of E2F1 activity during cancer progression. We here claim that E2F1-mediated regulation of miRNAs may as well be critical to explain the emergence of chemoresistance.

Finally, our simulations indicate that tumor heterogeneity in terms of the genetic signature of the E2F1-p73/DNp73-miR-205 network can affect the efficacy of anticancer therapy. In addition, our simulations suggest that the therapy can induce selective pressure on tumor cells favoring those showing chemoresistance. Other authors have suggested that under conditions of tumor genetic instability, heterogeneous tumors contain populations of one or more resistant clones (44, 45). These resistant subpopulations can be residual compared with other clones in a pretreated tumor due to the phenotypic cost associated with chemoresistance, but could get favored when the anticancer therapy is applied (46). This hypothesis is supported by our results. In our analysis, we only considered the role of the E2F1-p73/DNp73-miR205 network in developing chemoresistance. However, these molecules interact with additional genes involved in the regulation of other phenotypic responses and therefore display pleiotropy (4,16). Thus, it is possible that some features of the resistant phenotype here described may have negative effect on other cancer-associated traits, which are out of the scope of this article. In a more realistic case, one could enrich the analysis by expanding the network with interaction partners representative of these other phenotypes and conducting a more extensive analysis.

Design principles underlying chemoresistance

Our analysis indicates that the network under investigation is enriched in feedforward loops. Most of these motifs are central to the efficient apoptosis initiation after genotoxic stress and therefore their dysregulation can cause chemoresistance (47,48). Previous publications have emphasized the important role of feedforward loops regulating the transient dynamics of biochemical pathways (49). We here support the hypothesis that these loops may further control important features of the long-term response in signaling and transcriptional systems. The paradigmatic case is the regulation of miR-205 by p73 (positive) and DNp73 (negative), two transcriptional targets of E2F1. This system is a canonical case of an incoherent feedforward loop. Several publications suggest that the regulation and activity of p73 and DNp73 is nonlinear and varies for different levels of E2F1 expression (28,50). Under these conditions, our simulations indicate that the incoherent feedforward loop allows a single protein (e.g., E2F1) to nonmonotonically regulate components downstream of the loop (e.g., miR-205; see Fig. 7A). Thus, for an interval of E2F1 expression levels the incoherent feedforward loop is dominated by the p73-mediated positive branch and promotes expression of miR-205. However, for higher expression of E2F1, the negative, DNp73-mediated branch predominates and the downstream target is repressed (Fig. 7A). This suggests that incoherent feedforward loops can be active structures of regulation when the different branches of the loops get distinctively regulated.

Furthermore, we identified several network motifs with the structure of incoherent feedforward loops, in which a component downstream the network is regulated by a direct E2F1 target gene and a miR-205–regulated target. In our study, we focus on apoptosis-related targets (e.g., BCL-2 and HRK, see Fig. 7B). As we mentioned before, miR-205 expression is regulated by other cancer-related signals, external to the E2F1 network (e.g., EMT-related signals). Therefore, changes in the activity of those external signals may alter the activation status of the antiapoptotic branch and as a consequence the balance between pro- and antiapoptotic signals is also altered (Fig. 7B). In addition, miR-205 regulates proteins downstream the EMT signaling, like E-cadherin, and some members of the BCL2 family are regulated by E2F1. This suggests the existence of additional feedforward loops which were not investigated in this work (25). Taken together, feedforward loops involved in chemosensitivity can be considered active regulatory structures, whose activation status may depend on the cross-talk with other cancer-related signals.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: J. Vera, D. Engelmann, B.M. Pützer
Development of methodology: J. Vera, X. Lai, F.M. Khan
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): U. Schmitz
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): J. Vera, U. Schmitz, X. Lai, D. Engelmann
Writing, review, and/or revision of the manuscript: J. Vera, U. Schmitz, X. Lai, D. Engelmann, F.M. Khan, O. Wolkenhauer, B.M. Pützer
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): U. Schmitz
Study supervision: B.M. Pützer

Grant Support

This work was supported by the German Federal Ministry of Education and Research (BMBF) as part of the projects eBio:miRSys (0316175A to J. Vera

Published OnlineFirst February 27, 2013; DOI: 10.1158/0008-5472.CAN-12-4095

and O. Wolkenhauer), eBio:SysMet (0316171 to J. Vera, O. Wolkenhauer, and B.M. Pützer), GERONTOSYS-ROSAge (0315892A to J. Vera, O. Wolkenhauer), and Rostock University Medical Center for the project Systems Medicine of Cancer Invasion and Metastasis.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 6, 2012; revised January 17, 2013; accepted February 7, 2013; published OnlineFirst February 27, 2013.

References

1. Chang A. Chemotherapy, chemoresistance and the changing treatment landscape for NSCLC. Lung Cancer 2011;71:3–10.
2. Stanelle J, Pützer BM. E2F1-induced apoptosis: turning killers into therapeutics. Trends Mol Med 2006;12:177–85.
3. Blattner C, Sparks A, Lane D. Transcription factor E2F-1 is upregulated in response to DNA damage in a manner analogous to that of p53. Mol Cell Biol 1999;19:3704–13.
4. Engelmann D, Pützer BM. Translating DNA damage into cancer cell death-A roadmap for E2F1 apoptotic signalling and opportunities for new drug combinations to overcome chemoresistance. Drug Resist Updat 2010;13:119–31.
5. Rödicker F, Stiewe T, Zimmermann S, Pützer BM. Therapeutic efficacy of E2F1 in pancreatic cancer correlates with TP73 induction. Cancer Res 2001;61:7052–5.
6. Stiewe T, Pützer BM. Role of the p53-homologue p73 in E2F1-induced apoptosis. Nat Genet 2000;26:464–9.
7. Chung J, Irwin MS. Targeting the p53-family in cancer and chemosensitivity: triple threat. Curr Drug Targets 2010;11:667–81.
8. Lee J, Park CK, Park JO, Lim T, Park YS, Lim HY, et al. Impact of E2F-1 expression on clinical outcome of gastric adenocarcinoma patients with adjuvant chemoradiation therapy. Clin. Cancer Res 2008;14:82–8.
9. Han S, Park K, Bae B-N, Kim KH, Kim H-J, Kim Y-D, et al. E2F1 expression is related with the poor survival of lymph node-positive breast cancer patients treated with fluorouracil, doxorubicin and cyclophosphamide. Breast Cancer Res Treat 2003;82:11–6.
10. Alla V, Engelmann D, Niemetz A, Pahnke J, Schmidt A, Kunz M, et al. E2F1 in melanoma progression and metastasis. J Natl Cancer Inst 2010;102:127–33.
11. Pützer BM, Tuve S, Tannapfel A, Stiewe T. Increased DeltaN-p73 expression in tumors by upregulation of the E2F1-regulated, TA-promoter-derived DeltaN'-p73 transcript. Cell Death Differ 2003;10:612–4.
12. Stiewe T, Zimmermann S, Frilling A, Esche H, Pützer BM. Transactivation-deficient DeltaTA-p73 acts as an oncogene. Cancer Res 2002;62:3598–602.
13. Tuve S, Wagner SN, Schittek B, Pützer BM. Alterations of DeltaTA-p 73 splice transcripts during melanoma development and progression. Int J Cancer 2004;108:162–6.
14. Emmrich S, Wang W, John K, Li W, Pützer BM. Antisense gapmers selectively suppress individual oncogenic p73 splice isoforms and inhibit tumor growth *in vivo*. Mol Cancer 2009;8:61.
15. Alla V, Kowtharapu BS, Engelmann D, Emmrich S, Schmitz U, Steder M, et al. E2F1 confers anticancer drug resistance by targeting ABC transporter family members and Bcl-2 via the p73/DNp73-miR-205 circuitry. Cell Cycle 2012;11:3067–78.
16. Engelmann D, Pützer BM. The dark side of E2F1: in transit beyond apoptosis. Cancer Res 2012;72:571–5.
17. Vera J, Wolkenhauer O. A system biology approach to understand functional activity of cell communication systems. Methods Cell Biol 2008;90:399–415.
18. Vera J, Rath O, Balsa-Canto E, Banga JR, Kolch W, Wolkenhauer O. Investigating dynamics of inhibitory and feedback loops in ERK signalling using power-law models. Mol Biosyst 2010;6:2174–91.
19. Inohara N, Ding L, Chen S, Núñez G. harakiri, a novel regulator of cell death, encodes a protein that activates apoptosis and interacts selectively with survival-promoting proteins Bcl-2 and Bcl-X(L). EMBO J 1997;16:1686–94.
20. Zhu S, Belkhiri A, El-Rifai W. DARPP-32 increases interactions between epidermal growth factor receptor and ERBB3 to promote tumor resistance to gefitinib. Gastroenterology 2011;141:1738–48. e1–2.
21. Rosanò L, Cianfrocca R, Spinella F, Di Castro V, Nicotra MR, Lucidi A, et al. Acquisition of chemoresistance and EMT phenotype is linked with activation of the endothelin A receptor pathway in ovarian carcinoma cells. Clin Cancer Res 2011;17:2350–60.
22. Engelmann D, Knoll S, Ewerth D, Steder M, Stoll A, Pützer BM. Functional interplay between E2F1 and chemotherapeutic drugs defines immediate E2F1 target genes crucial for cancer cell death. Cell Mol Life Sci 2010;67:931–48.
23. Melino G, Bernassola F, Ranalli M, Yee K, Zong WX, Corazzari M, et al. p73 Induces apoptosis via PUMA transactivation and Bax mitochondrial translocation. J Biol Chem 2004;279:8076–83.
24. Aguda BD, Kim Y, Piper-Hunter MG, Friedman A, Marsh CB. MicroRNA regulation of a cancer network: consequences of the feedback loops involving miR-17-92, E2F, and Myc. Proc Natl Acad Sci USA 2008;105:19678–83.
25. Gandellini P, Folini M, Longoni N, Pennati M, Binda M, Colecchia M, et al. miR-205 Exerts tumor-suppressive functions in human prostate through down-regulation of protein kinase Cepsilon. Cancer Res 2009;69:2287–95.
26. Wu H, Zhu S, Mo Y-Y. Suppression of cell growth and invasion by miR-205 in breast cancer. Cell Res 2009;19:439–48.
27. Frolov A, Schuller K, Tzeng C-WD, Cannon EE, Ku BC, Howard JH, et al. ErbB3 expression and dimerization with EGFR influence pancreatic cancer cell sensitivity to erlotinib. Cancer Biol Ther 2007;6:548–54.
28. Buhlmann S, Pützer BM. DNp73 a matter of cancer: mechanisms and clinical implications. Biochim Biophys Acta 2008;1785:207–16.
29. Guan M, Chen Y. Aberrant expression of DeltaNp73 in benign and malignant tumours of the prostate: correlation with Gleason score. J Clin Pathol 2005;58:1175–9.
30. Massagué J. TGFbeta in Cancer. Cell 2008;134:215–30.
31. Frogne T, Benjaminsen RV, Sonne-Hansen K, Sorensen BS, Nexo E, Laenkholm A-V, et al. Activation of ErbB3, EGFR and Erk is essential for growth of human breast cancer cell lines with acquired resistance to fulvestrant. Breast Cancer Res Treat 2009;114:263–75.
32. Garrett JT, Olivares MG, Rinehart C, Granja-Ingram ND, Sánchez V, Chakrabarty A, et al. Transcriptional and posttranslational up-regulation of HER3 (ErbB3) compensates for inhibition of the HER2 tyrosine kinase. Proc Natl Acad Sci USA 2011;108:5021–6.
33. González-García I, Solé RV, Costa J. Metapopulation dynamics and spatial heterogeneity in cancer. Proc Natl Acad Sci USA 2002;99:13085–9.
34. Yao G, Lee TJ, Mori S, Nevins JR, You L. A bistable Rb-E2F switch underlies the restriction point. Nat. Cell Biol 2008;10:476–82.
35. Zhang X-P, Liu F, Wang W. Coordination between cell cycle progression and cell fate decision by the p53 and E2F1 pathways in response to DNA damage. J Biol Chem 2010;285:31571–80.
36. Wong JV, Yao G, Nevins JR, You L. Viral-mediated noisy gene expression reveals biphasic E2f1 response to MYC. Mol Cell 2011;41:275–85.
37. Katoh Y, Katoh M. Hedgehog signaling, epithelial-to-mesenchymal transition and miRNA (review). Int J Mol Med 2008;22:271–5.
38. Wiklund ED, Bramsen JB, Hulf T, Dyrskjøt L, Ramanathan R, Hansen TB, et al. Coordinated epigenetic repression of the miR-200 family and miR-205 in invasive bladder cancer. Int J Cancer 2011;128:1327–34.
39. Castillo J, Goñi S, Latasa MU, Perugorría MJ, Calvo A, Muntané J, et al. Amphiregulin induces the alternative splicing of p73 into its oncogenic isoform DeltaEx2p73 in human hepatocellular tumors. Gastroenterology 2009;137:1805–15.e1–4.

Published OnlineFirst February 27, 2013; DOI: 10.1158/0008-5472.CAN-12-4095

40. Samavarchi-Tehrani P, Golipour A, David L, Sung H-K, Beyer TA, Datti A, et al. Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. Cell Stem Cell 2010;7:64–77.
41. Bryant JL, Britson J, Balko JM, Willian M, Timmons R, Frolov A, et al. A microRNA gene expression signature predicts response to erlotinib in epithelial cancer cell lines and targets EMT. Br J Cancer 2012;106: 148–56.
42. Lai X, Schmitz U, Gupta SK, Bhattacharya A, Kunz M, Wolkenhauer O, et al. Computational analysis of target hub gene repression regulated by multiple and cooperative miRNAs. Nucleic Acids Res 2012;40: 8818–34.
43. Li Y, Li Y, Zhang H, Chen Y. MicroRNA-mediated positive feedback loop and optimized bistable switch in a cancer network Involving miR-17-92. PLoS ONE 2011;6:e26302.
44. Merlo LMF, Pepper JW, Reid BJ, Maley CC. Cancer as an evolutionary and ecological process. Nat Rev Cancer 2006;6:924–35.
45. Iwasa Y, Nowak MA, Michor F. Evolution of resistance during clonal expansion. Genetics 2006;172:2557–66.
46. Gatenby RA, Brown J, Vincent T. Lessons from applied ecology: cancer control using an evolutionary double bind. Cancer Res 2009; 69:7499–502.
47. Polager S, Ginsberg D. p53 and E2f: partners in life and death. Nat Rev Cancer 2009;9:738–48.
48. Marin JJG, Briz O, Monte MJ, Blazquez AG, Macias RIR. Genetic variants in genes involved in mechanisms of chemoresistance to anticancer drugs. Curr. Cancer Drug Targets 2012;12:402–38.
49. Mangan S, Alon U. Structure and function of the feed-forward loop network motif. Proc Natl Acad Sci U S A 2003;100:11980–5.
50. Nakagawa T, Takahashi M, Ozaki T, Watanabe K-I, Todo S, Mizuguchi H, et al. Autoinhibitory regulation of P73 by ΔNp73 to modulate cell survival and death through a P73-specific target element within the ΔNp73 promoter. Mol Cell Biol 2002;22: 2575–85.

Published OnlineFirst February 27, 2013; DOI: 10.1158/0008-5472.CAN-12-4095

Cancer Research
The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

AACR American Association for Cancer Research

Kinetic Modeling• Based Detection of Genetic Signatures That Provide Chemoresistance via the E2F1-p73/DNp73-miR-205 Network

Julio Vera, Ulf Schmitz, Xin Lai, et al.

Cancer Res 2013;73:3511-3524. Published OnlineFirst February 27, 2013.

Selbstständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbstständig angefertigt habe. Ich versichere, dass ich nur die angegebenen Quellen und Hilfen in Anspruch genommen habe, und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Rostock, 6. November 2019 Faiz M. Khan