Department for General, Visceral, Vascular and Transplant Surgery,

University Medicine Rostock

Director: Univ. Prof. Dr. med. Clemens Schafmayer

Defining the Role of Human Endogenous Retroviruses in Pluripotency of Colorectal Cancer Cells

Inaugural Dissertation
submitted in partial fulfilment of
the requirements for the academic degree
doctor rerum humanarum (Dr. rer. hum.)
of Rostock University Medical Center

Submitted by
Fatemeh Shokraie, born September 19th 1982, in Iran

Rostock, February 2021

Reviewer:

Prof. Dr. rer. nat. Burkhard Hinz

Universitätsmedizin Rostock

Institut für Experimentelle Chirurgie / Zentrale Versuchstierhaltung

Prof. Dr. med. Friedrich Prall

Universitätsmedizin Rostock

Institut für Pathologie

Prof. Dr. med. Stephan Kersting

Universitätsmedizin Greifswald

Klinik für Allgemeine Chirurgie, Viszeral-, Thorax- und Gefäßchirurgie

Date of the defense: 05.07.2022

Table of Contents

I

1. Introduction

1.1. Colorectal cancer

1.1.1. Definition and basics

Colorectal carcinomas (CRC) are defined as tumors occurring in the colon (colon cancer) and rectum (rectal cancer) (1). These cancers are grouped together based on the similar organ of origin (large bowel), similar anatomical structure, as well as comparable histological features (2). CRC often results from precancerous polyps either localized growth or accumulations of abnormal cells in the inner lining of the colon (1,3). More than 90% of all CRC cases are adenocarcinomas emerging from mucous-secreting cells (4). Neuroendocrine tumors, lymphomas, sarcomas, and squamous cell carcinomas are the rare types found in the colorectum (5).

1.1.2. Epidemiology

CRC is the third most commonly diagnosed tumor entity in both female and male worldwide (6,7). The higher frequency of this type of cancer in developed countries is reflected by a sedentary lifestyle (8). Despite advances in screening and therapeutic approaches that have contributed to improved overall rates, it still ranks second with regard to cancer-related death in Europe (6,9) as well as in the world (8).

Data from the World Health Organization reported over 1.8 million newly diagnosed cases in 2018 with 862,000 deaths (10). In Germany, nearly 60,000 new cases are diagnosed every year with a death rate by almost 30%. The high average age of onset (72 years in men and 75 years in women) characterizes CRC as an age-related illness (11). However, recent epidemiological data describe an increase of about 2% in men and women younger than 50 years. According to recent forecasts, CRC incidence rate will significantly increase until 2030 in patients aged between 20 to 34 years (12).

Survival correlates with tumor stage at time of diagnosis (9). The average of the 5-year survival rate for both sexes is 63% in colon cancer and 67% in rectal cancer (13) which ranges from 90% in patients with stage I (localized tumors) to 15% in patients with stage IV disease (distant tumors) (14).

1.1.3. Risk factors

CRC pathogenesis is influenced by modifiable and nonmodifiable risk factors (3). Some elements such as lack of exercise, obesity, smoking, alcohol, high-fat processed food, and red meat are modifiable factors (1,12). Changes in lifestyle may decrease the risk of developing CRC (3). Other non-modifiable factors include age, gender (53% in men *vs*. 47% in female) (2), race, the presence of chronic inflammatory diseases, such as inflammatory bowel disease, and hereditary conditions (1,12). The latter plays a major role in cancer development (4,12). Up to 35 % of all CRC cases are due to hereditary mutations. Still, driver mutations are only detectable in a small number of patients (around 5 – 10%). Here, germline mutations in tumor suppressors, including adenomatous polyposis coli (*APC*), as well as the DNA mismatch repair (MMR) genes *(MLH1*, *MSH2*, *MSH6*, *PMS2)* are the most common ones (15,16). Germline MMR gene mutations are associated with the Lynch syndrome (formerly known as hereditary nonpolyposis colorectal cancer (HNPCC). This tumor syndrome accounts for 2-3% of all CRC cases and is clinically challenging due to the heterogeneous presentation. Patients suffering from Lynch syndrome develop a complex tumor spectrum, including CRC, gastric cancer, endometrium cancer, and others (17,18). The average age of CRC onset in Lynch syndrome is 45 years and thus decades earlier than the average age seen in the general population. The same accounts for patients harboring *APC* germline mutations. Clinically, these patients suffer from familial adenomatous polyposis (FAP), detectable in about 1% of CRC cases. A hallmark of FAP is the presence of hundreds to thousands of small adenomatous polyps throughout the entire length of the colon with almost complete penetrance by the fourth decade (19).

Still, the large majority of CRC are sporadic with no family history or obvious genetic predisposition (20,21) and contain somatic mutational hotspots in tumor suppressor genes, including tumor protein p53 (*TP53*) and Bcl-2-associated X protein (*BAX*), as well as proto-oncogenes, such as Kirsten rat sarcoma *(KRAS)* and *BRAF* (17,18). Somatic mutations as well as epigenetic alterations such as promotor hypermethylation contribute to CRC formation. Aberrant methylation of tumor suppressor genes prevents gene transcription and causes functional loss of the corresponding gene (3,22).

1.1.4. The adenoma-carcinoma sequence

Typically, CRC arise from adenomatous (precancerous) polyps which gradually accumulate genetic and epigenetic changes as polyps increase in size (23,24). This process from adenoma

to carcinoma and finally metastasis can take decades (23,25). Dividing polyp cells can grow into the colon and rectal wall and may enter the nearby lymph nodes through blood or lymphatic vessels, eventually spreading to distant organs. However, less than 10% of polyps progress to invasive cancer (26). Nonetheless, early detection and resection of precancerous polyps are vital to intervene the adenoma-carcinoma sequence and stop CRC development.

1.1.4.1. Morphological and molecular changes

The morphological and molecular changes during tumorigenesis are outlined in Figure 1 (27). The polyp size enlarges as the cells begin to proliferate, known as hyperplasia. This hyperplasia usually coincides with acquired (somatic) or inherited (germline) DNA mutations and epigenetic changes, which then lead to cytological and histological dysplasia (24). With time, escalation of DNA damage may cause the severe dysplasia, predicting a high risk of progression to invasive cancer (3).
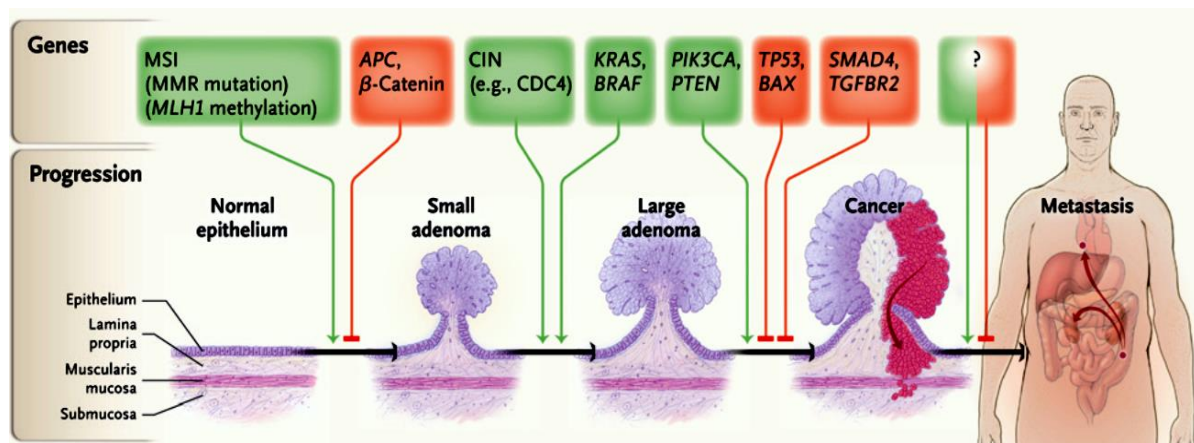
Figure 1. Molecular and histological changes during progression of colorectal cancer.

Genetic or epigenetic alterations increase with polyp growth. These changes cause the specific genes to be on or off. MSI pathway is triggered either by gene mutation or by aberrant methylation. Mutations in APC, TP53 or BAX result in loss-of-function, while KRAS and BRAF are activated. The question mark indicates unknown genetic or epigenetic changes leading to metastatic progression. Green arrows and red blockers denote oncogenic and tumor suppressor factors, respectively (27).

According to the 1990 proposal by Fearon and Vogelstein, the acquisition of ordered genetic alterations is necessary for the development of cancer. Mutations in the *APC* gene are the driving event in adenoma formation that lead to uncontrolled cell proliferation (23,28). APC itself acts as a negative regulator of the *Wnt* signaling pathway, which stimulates the

degradation of β-catenin and inhibits the *Wnt* pathway (Figure 1). APC inactivation leads to sustained intracellular β-catenin levels and prolonged activity of the *Wnt* signaling pathway in CRC cells (22).

Transition from early to intermediate and late adenoma is driven by mutations activating the proto-oncogene *KRAS* and additional mutations in *TGF-β* (transforming growth factor beta), *PIK3CA* (phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha), and *TP53* (23,28). P53 is a key regulator involved in apoptosis, cell cycle regulation, senescence, angiogenesis, immune response, cell differentiation, motility, and migration. Therefore, loss of function of *TP53* affects a variety of cell functions that ultimately results in carcinogenesis (28).

1.1.5. CRC - a heterogeneous disease

Over the past decades, advances in understanding of the genetic and epigenetic mechanisms have identified different molecular subgroups in CRC. To date, at least three major molecular mechanisms have been postulated as promoters of carcinogenesis. These include chromosomal instability (CIN; the most frequent one with about 70%), microsatellite instability (MSI; accounting for 15%) as well as the CpG island methylator phenotype (CIMP, ~15%) (20,29,30). These molecular variations can arise alone or in combination with other alterations. In fact, this phenomenon creates a tumor heterogeneity that contributes to different clinical and pathological characteristics (31).

CIN tumors are microsatellite stable (MSS) (32) and development follows the classical adenoma-carcinoma sequence (3). It is characterized by numerical or structural chromosomal aberrations that arise during cell division because of chromosome mis-segregation (30,33). These tumor frequently harbour mutations in specific oncogenes (*KRAS* and *BRAF*) and tumor suppressor genes (*APC* or *TP53*) as previously described (Figure 1) (22,30). Another molecular mechanism is MSI caused by disruption of DNA MMR genes, as stated earlier in paragraph 1.1.3. (32). Defective MMR system leads to insertion or deletion of microsatellites sequences and the formation of numerous mutations. In sporadic MSI tumors, promoter hypermethylation is the main cause of MMR gene inactivation. While *MLH1* is the main target here, MMR loss in the hereditary counterpart Lynch syndrome is due to germline MMR gene mutations and a somatic second hit in the remaining allele (30). The third pathway, assigned as CIMP, is characterized by loss of function of tumor suppressor genes and/or MMR genes through

widespread epigenetic modifications. Here, CpG island are hypermethylated in promotor regions (4,32). The CIMP pathway is detectable in both MSI and MSS tumors (20).

1.1.6. Prognosis

The prognosis and therapeutic options for cancer patients depend on the stage at diagnosis. By performing extensive surgical resection and/or adjuvant (radio-) chemotherapy, early CRC is usually curable (34). However, when diagnosed at an advanced or even metastatic stage, affected patients have a poor prognosis and conventional treatment is ineffective in the vast majority of patients (35). In stage II and III patients, clinical activity of adjuvant chemotherapy varies due to CRC molecular heterogeneity. Thus, identification of prognostic biomarkers will help to identify patients likely benefitting from adjuvant chemotherapy and/or predicting risk stratification of relapse. This in turn will finally improve prognosis by selecting an optimal therapy regimen (20). Several mutations were already found to be associated with poor patient outcome and/or treatment failure (36). These include, among others, somatic mutations in *TP53*. Still, this does not serve as a routine prognosis marker for CRC (37). Mutant *APC*, which appears in early development of CRC, is considered another helpful biomarker for identifying high-risk individuals (38). Likewise, *KRAS* and *BRAF* as well as MSI status are usually tested in clinical practice (36,39,40). While *KRAS* and particularly *BRAF* mutations are associated with a poor prognosis, MSI is a good prognostic biomarker (40). CIMP without MSI is independently associated with significantly worse outcome (41). *Vice versa*, MSI-High (MSI-H) tumors have the best prognosis (4,41).

To refine diagnosis and predict outcome, comprehensive molecular testing for early and precise diagnosis, such as tumor mutational burden, as well as epigenetic alterations is recommended (16). To this end, next generation sequencing is a highly sensitive state of the art system for precise molecular characterization of both sporadic and hereditary CRC (15).

1.1.7. Cell of origin in CRC

The last decade has seen much effort in molecular understanding and contributed to refinement of treatment. Still, a large number of patients suffer from tumor relapse and metastasis - particularly in previously diagnosed advanced stages (42,43). Conventional treatment based on

application of cytostatic drugs targets rapidly proliferating cells while sparing certain cell types, known as cancer stem cells (CSCs) or cancer initiating cells (44,45).

Stem cells exist in three forms. These include germline, embryonic, and somatic stem cells (ESCs and SSCs). Somatic or adult stem cells are responsible for the normal blood and tissue regeneration in organs such as the skin and colon. This regeneration is controlled through the asymmetric division that creates differentiated progenitor cells for the organ's lifetime (46). There is still an unanswered question in the context of stem cells in cancer biology. The question is whether the primary target in carcinogenesis process is either stem cells, progenitor cells or dedifferentiated cells acquiring stem cell properties (46,47). In fact, a small subpopulation within the bulky tumor mass has the capacity to preserve long-term growth of the tumor. Studies on leukemia and solid cancers confirmed the presence of cells with stem cell capabilities; but no information was given about the origin of CSCs (46). However, stem cells have higher potential to accumulate carcinogenic mutations than differentiated cells. This is thought to be due to their longevity and self-renewal abilities. Hence, they can be the key target for early events in the process of carcinogenesis (46). There is growing evidence for the hierarchical model in carcinogenesis, with CSCs at the apex. According to this model, cancer initiation, development, drug resistance, malignant growth, and metastasis are driven by this rare subpopulation (48,49). Therefore, recognition, successful targeting, and eradication of CSCs may represent a promising therapeutic approach for cancer treatment.

1.2. Cancer stem cells

1.2.1. CSCs history

The idea of CSCs dates back to 1855, when Virchow proposed that cancer growth resulted from remnants of undifferentiated embryonic cells in adult tissues. It was later formulated by Cohnheim in 1875 as an "embryonal rest theory" (48,50). In 1937, leukemia transmission in mice was carried out using a single cell (48). In the 1960s, Pierce and colleagues demonstrated that the single embryonic cancer cell was multipotent and could cause heterogeneous teratocarcinomas in mice (51). Finally, the CSCs concept was explained in 1971 by Park (52). Then, the first indication of the presence of stem cells in cancer was reported in 1997 in human acute myeloid leukemia (53). This identification was the starting point for further investigations in other cancer entities and have contributed to the characterization of stem cells in brain (54),

breast (55), lung (56), colon (57–59), ovary (60), pancreas (61), bladder (62), skin (63), liver (64), and prostate cancers (65,66). Characterization of this rare fraction in various cancers has evoked scientific interest in thinking about the role of these cells in cancer relapse (67).

1.2.2. Cancer stem cell properties

CSCs share specific properties with normal SSCs and ESCs. They have an increased self-renewal ability, achieved through the cooperative interaction of symmetrical and asymmetrical cell divisions (45,68). With this function, a single CSC can maintain the pool of its stem cells and extend its lifespan (68). Another common characteristic among stem cells is their potential of multi-lineage differentiation (45,69). In both ESCs and CSCs, pluripotency is maintained by several transcription factors, namely octamer-binding transcription factor 4 (OCT4), Nanog homeobox (NANOG), and SRY-box 2 (SOX2) (48).

CSCs can, similar to the other stem cells, switch functional features from stem cell to non-stem cell properties (48,70) or in other words, they have a high plasticity (71,72). This includes proliferative or quiescent states, transition between epithelial and mesenchymal state (EMT or MET), and symmetric/asymmetric cell divisions (70). This plasticity can serve as cause for
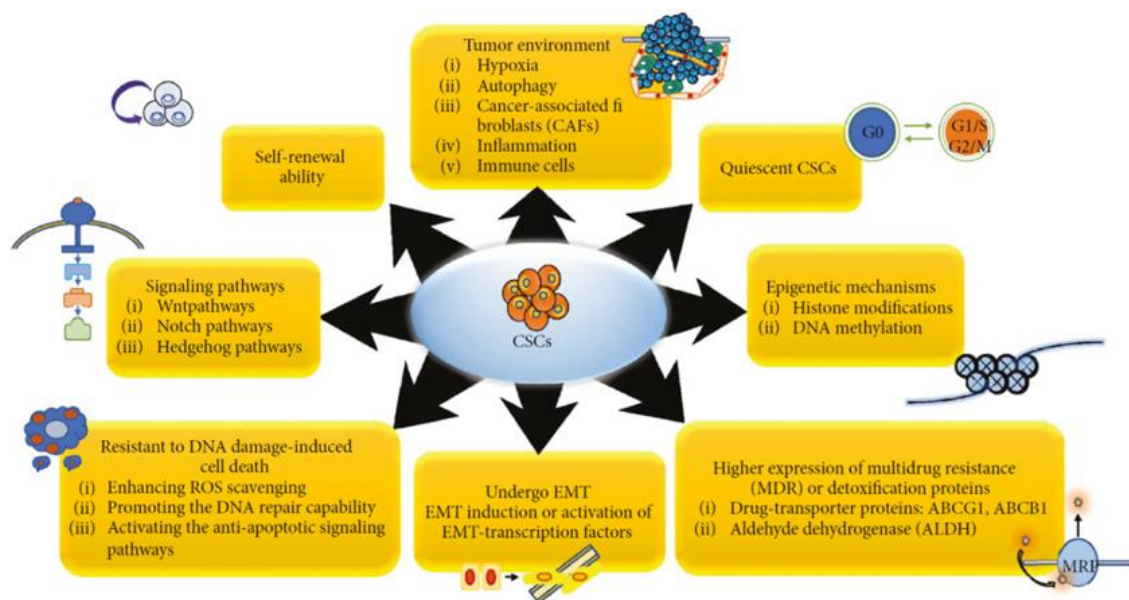
Figure 2. Summary of key strategies involved in therapy resistance of CSCs (73).

incomplete cancer elimination and subsequent cancer recurrence (47,48). To overcome this obstacle, escape mechanisms of CSCs leading to therapy resistance must be understood. Growing evidence suggests that the combination of conventional chemotherapy with agents

that specifically target the CSCs pathways may provide a better efficacy (73). Some of self-protection strategies in CSCs are shown in Figure 2 and include the increased DNA repair capability, EMT and MET, selective quiescent, membrane transporters and drug efflux (48,74), diminished expression of pro-apoptotic p53 proteins, and increased expression of anti-apoptotic proteins such as B-cell lymphoma 2 (Bcl-2) and B-cell lymphoma-extra large (Bcl-xL) (75).

CSCs possess defense mechanisms similar to normal stem cells, such as multidrug resistance (MDR) and pump activity. The existence of a rare fraction, the so-called side population (SP), has been proven in various types of cancer (73).

The SP population was first described by Goodell in 1996 as a rare distinct subset of bone marrow cells. They share their phenotype with hematopoietic stem cells and exhibit low Hoechst staining (76). Zhou *et al.* showed that low Hoechst pattern in a SP fraction is the consequence of highly concentrated ABC transporters and their activities. This method was then introduced as a meaningful strategy for purifying stem cells in normal and cancerous tissues (77). This cell population has highly elaborated efflux systems (including MDR1, ATP-binding cassette transporter G2 (ABCG2), and ATP binding cassette subfamily B member 1 (ABCB1)) to pump out a variety of genotoxins in an ATP-dependent manner (e.g., ATP-binding cassette (ABC) transporter). This leads to the MDR phenomenon (in CSCs) and to protection from chemical damage. The expression of such efflux pumps is strongly associated with tumor relapse and metastasis (78,79). Alisi *et al.* even suggested overexpression of ABC proteins as the most likely important protective mechanism for CSCs in response to conventional chemotherapeutic regimens (75). Later on, expression of ABCG2 and its involvement in drug resistance were enormously investigated in CRC (80–82) and various cancer types (83–85). Principally, increased resistance to commonly used therapy and higher expression of ABC transporters were reported in stem cell populations. The CSCs were isolated based on SP cells, known surface markers, or spheres and finally compared to the matched non-stem cells compartment.

In 2015, an increase of human endogenous retroviruses (HERVs) subfamily H RNA expression was shown in ESCs (86). These HERV-H transcripts possess (potential) binding sites for specific transcription factors recruitment such as LBP9 (lipopolysaccharide binding protein 9) (87), NANOG (86,87) and OCT4 (86–88). All of them are associated with stem cells and

pluripotency. Thus, such endogenous retroviral sequences might play an essential role in induction and maintenance of pluripotency in ESCs (86) and potentially also in CSCs.

1.2.3. Identification/characterization of cancer stem cells

Choosing the most appropriate model to reflect the characteristics of a certain tumor is one of the greatest challenges for cancer research (89). CSCs are commonly isolated via *in vivo* and *in vitro* technologies. To test tumorigenic ability *in vivo*, candidate cancer cells either form human tumors or cancer cell lines are transplanted into immunodeficient mice. Despite its limitations, this method is widely used. For instance, remodeling of a similar tumor niche containing CSCs-specific cytokines has failed in the animal model (90). In addition, humans and mice differ in terms of physiology. This is an obstacle to gain precise insights into the operative mechanisms in human carcinogenesis (91). Another variety is the rarity of the loss of heterozygosity in mice, often occurring in human, such as the *APC* gene in CRC (89).

There are several *in vitro* methods to enrich CSCs population. One of the traditional tactics is to isolate a tumor population that expresses specific surface markers of the corresponding normal stem cells (45,73). However, due to limited reproducibility and the variable expression level, there is an increasing disagreement in marker-based assays (48). CSCs can be identified by colony formation, three-dimensional (3D) cell cultures (73) and isolating SP via sorting (92). Growing tumor spheres in 3D culture reflect behavior of the native cancer tissue at best. Given that, it serves as a reliable model in therapy studies, as well as investigation of proliferation, invasion, cell-cell/matrix interaction, and metabolism (89). Another simple method for detecting CSCs is based on Hoechst dye 33342 staining and flow cytometric analysis. Here, stem or stem-like cells pump out the dye much faster than their normal cellular counterparts – and can be recognized as less-stained "SP " (75,92).

1.2.4. Cancer stem cells in colorectal cancer

Initial studies on CSCs in CRC started in 2007 by two different groups. In both studies, CD133 -a putative stem cell marker- was used to isolate CSCs from patients' specimen (58,59). Isolated CD133$^+$ cells showed outstanding features such as tumor formation in mice, the maintenance of an undifferentiated state in serum-free culture, and the aggressiveness over the time (59). Furthermore, CD133$^-$ cells that comprise the majority of tumor cells, could not

initiate tumor growth in mice (58). Based on these studies, the hierarchical organization is a consistent model in CRC formation. Undifferentiated tumorigenic CD133$^+$ cells were also termed as the stem cell fraction in tumors (58,59). Later in 2007, Todaro *et al.* reported a large population of differentiated cancer cells and a small portion of the undifferentiated cells in CRC patients. It was also shown that CD133$^+$ cells protected themselves against apoptosis by interleukin 4 (IL-4) secretion. Besides, colonospheres were generated in serum free medium supplemented with epidermal growth factor (EGF) and basic fibroblast growth factors (bFGF). Interestingly, expression of CD133 protein was sustained in 90% of all colonospheres that were introduced as a population containing stem-like cells *in vitro* (57). Afterwards, efforts were made to identify possible new markers and to investigate the role of CR-CSCs in carcinogenesis (Table 1).

Table 1. *List of CSCs studies and related findings in CRC.*

Samples	Outcome
CRC lines **(E450, CSC1, and Co100)**	Correlation of Wnt activity with: - CD133, CD166, and CD29 expression - Clonogenicity (93)
Animal models (94) **Patient-derived organoid** (95)	- Prevention of primary tumor growth after exhaustion of LGR5^{+} [1] cells - Filling the LGR5$^+$ CSCs pool by LGR5$^-$ cells (94,95)
CRC lines **(SW480, DLD-1)**	Transduction with *OCT3/4*, *SOX2*, and *KLF4* [2] increased: - Expression of stem cell related genes - Sphere formation - Chemoresistance - Tumorigenicity (96)
15 independent studies **($\sum$ 2300 cases)**	Significant correlation of CD133 overexpression with: - Poorer clinical outcome - T and N stages - Vascular invasion in CRC patients (97)

---

1 . Leucine-rich repeat-containing G-protein coupled receptor 5

2. Kruppel-like factor

CRC and matched cell lines	- CD133[+] in normal and cancer - CD133 = CSCs marker in CRC - Differentiated CSCs: different CD133 glycosylation and localization (98)
CRC lines (Caco-2, LoVo)	- Correlation between CD133 and low survival - No link between CD133 reduction and invasion, proliferation, or colony formation (99)
523 formalin-fixed and paraffin embedded CRC	Significant correlation of: - CD133 with gender and advanced T stage - CD24 with degree of differentiation - CD44 with tumor size (100)
Normal mucosa, normal-appearing mucosa from FAP, adenomas, and carcinomas	ALDH1[3]- expressing cells; - Few at the bottom of normal crypt - Increase in number and migration to upper part of the crypt during transformation - Marker to identify and enrich colonic CSCs (101)
60 CRC and matched normal tissues	CD44[+] *vs.* CD133[+]: - Different localization - Reduced tumorigenicity after being knockdown - Functional role in CR-CSCs and cancer initiation (102)
CRC lines (LS174T, HCT15, SW620), 11 adenocarcinoma tissues, one primary tumor cell line, and five primary xenograft tumors	CD44[hi]: - Enhanced colony forming efficiency - Higher *in vivo* tumorigenicity at low cell numbers - Recapitulation of primary tumors' heterogeneity - Increased tumorigenicity when combined with ALDH - Enriched nuclear localization of β-catenin (103)

Results obtained from these studies will help to move forward in understanding of how CR-CSCs initiate tumor formation, metastasis, and therapy resistance.

---

3. Aldehyde dehydrogenase 1

1.3. Endogenous retroviruses (ERVs)

1.3.1. Definition and basic knowledge

HERVs originate from infectious exogenous retroviruses that were fixed in the germline DNA as a result of endogenization processes millions of years ago (104–106). Approximately 1% of the human genome consists of ERVs (107), classified into several families, ranging between 22 (108) to 31 families (109). This classification is based on the similarities in the *pol* gene sequence (110,111) and the homology of their primer binding site to the complementary host tRNA (112). These include among others tryptophan (W) for HERV-W, lysine (K) for HERV-K, arginine (Arg) for HERV-R (108), phenylalanine (Phe) for HERV-F, and histidine for HERV-H (113). The genomic structure of HERVs is similar to the exogenous retroviruses that consist of four genes including *gag* (group antigens)*, pro* (protease)*, pol* (polymerase), and *env* (envelope). The genes are flanked by identical long terminal repeat (LTR) at the 5' and 3' ends (114,115), functioning as promotor or enhancer. They contain RNA regulatory regions and transcription factor binding sites (114). LTR retroelements comprise roughly 8% of the human genome and consist of at least 200,000 individual loci (104,109).

Random genetic drift, natural selection and post-endogenization mechanisms lead to the loss, fixation or multiple copies of the retrovirus sequences. In the course of evolution, hundreds of various ERVs loci have been accumulated in the human genome, resulting in mutations in the primary viral genome (106). Mutations inhibit the viral particles to assemble normally as well as preventing mobility of these elements. Thus, they are trapped in the hosts' genome and lose their ability to infect other cells (116). Indeed, the vast majority of HERV elements are replication incompetent due to mutations, deletions, and epigenetic mechanisms in open reading frames (ORF) of coding genes during endogenization (104,109). In contrast to the defective coding sequences, there are a few thousands HERV LTR elements with active ORF (111,114). The ORF can regulate transcription of adjacent genes (both viral and host genes) (117) with partially conserved viral promoter, enhancer and polyadenylation signals (111,113). So far, Syncytin-1 (HERV-W) and Syncytin-2 (HERV-FDR) are the only HERV-derived proteins with physiological function expressed during placentation (118). Conversely, aberrant HERV reactivations occurs in several pathological situations such as multiple sclerosis and cancers (110). Studies on the transcription of HERV family members are listed in the Table 2.

Table 2. *List of HERV families associated with various tumor entities.*

Cancer type	HERV family	References
Melanoma	HERV-K (HML2)	(119,120)
Leukemia and Lymphoma	HERV-K (HML2)	(121,122)
Breast cancer	HERV-K (HML2)	(123)
Ovary tumors	HERV-K (HML2)	(124)
Prostate cancer	HERV-E	(125,126)
Seminoma	HERV-W	(126)
Liver and Lung cancers	HERV-R	(127)
Colorectal cancer	HERV-H	(128–131)

1.3.2. Functional roles of HERVs in cancer development

Despite increasing studies on the topic of HERV, functional roles in carcinogenesis are not definitely understood. Some of the possible mechanisms are shown in the Figure 3.

To date, HERVs' transcripts have been documented as non-coding RNA and/or functional peptides that can activate or suppress proto-oncogenes. Besides, activated LTRs' promotor may act as an alternative promotor or enhancer for cancer-related genes, leading to uncontrolled proliferation. LTR elements also provide a binding site for transcription or regulatory factors involved in oncogene activation. Furthermore, non-allelic homologous recombination between HERVs may result in chromosomal rearrangements and genetic mutations (113). The other mechanism is inactivation of tumor suppressor genes due to a *de novo* insertion or translocation of these elements (132). Some HERV-envs may be able to activate signaling pathways that promote the immunosuppressive response by possessing putative immunosuppressive domains or by coding env-derived variants such as *Np9* and *Rec* as potential oncogenes (133).

Interestingly, it was observed that tumor-derived vesicles contained HERVs transcripts, suggesting tumor-derived HERV-RNA may stimulate metastasis by transferring to normal cells (113). The best-known physiological role of HERV is cell fusion, which occurs in placenta development by env proteins. However, env proteins encoded by HERV-K mediated intracellular fusion of cancer cells in melanoma. This phenomenon may contribute to malignancy, tumor progression, and therapy resistance (134).
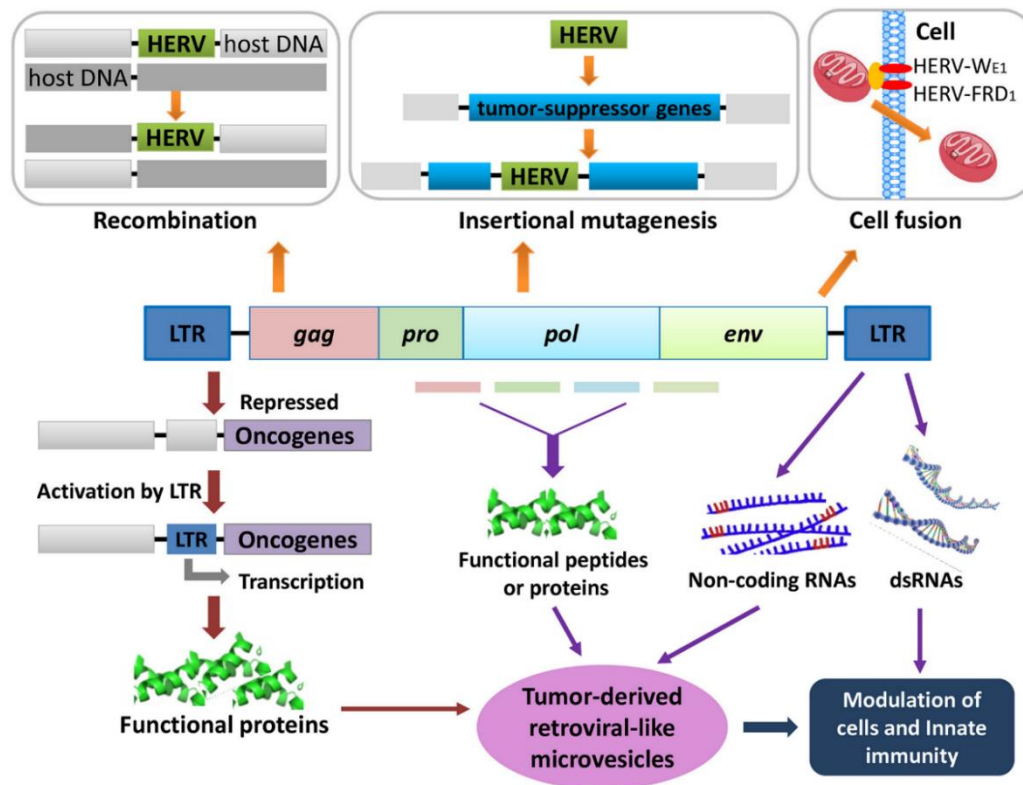
Figure 3. Potential mechanisms of HERV-mediated oncogenesis (113).

1.3.3. HERV-H family

In the human genome, the HERV-H family has approximately 1000 copies in full length, truncated, and solitary, making it the largest HERV family (135,136). The most common integrated ERV-H in the human genome is around 5.8 kb and consists of the truncated *gag* and *pol* genes, while the *env* coding region has been lost (137,138). These fragments are flanked by one of the LTR7 subtypes, namely LTR7, LTR7A, LTR7B, LTR7C, and LTR7Y (139). Conversely, almost 50-100 copies of intact HERV-H sequences with a size of 8.7 kb have been found in human genome. Although some of them have an ORF for the *env* gene, no env protein with physiological function has been reported (138,139). It is believed that the env protein is the most important replicated HERVs in pathological situations with immunosuppressive function. So far, replication of intact env proteins was described in three integrated HERV-H with open ORF such as *env62*, *env60*, and *env59* (136). Given the high similarities within the H family, it is still problematic to analyze the transcription of HERV-H copies (113). Hence, it is largely unknown whether *env* or other HERV-H genes are expressed either as non-coding spliced transcripts or as functional proteins. In the last decade, several studies reported the

accelerated transcription of the HERV-H elements (LTR) in ESCs (88,140). These transcripts were mainly long noncoding RNAs (lncRNAs) needed for maintenance of the stem cell identity (88). Moreover, a study in 2012 discovered binding of NANOG, OCT4, and SOX2 to LTR sequences of transcriptionally active HERV-H (Figure 4). Additionally, HERV-H-RNA reduced with stem cell differentiation similar to that observed for *OCT4* and *NANOG* (140). In conclusion, all these data strongly support the hypothesis of using HERV-H as a reliable indicator of pluripotency in stem cells.
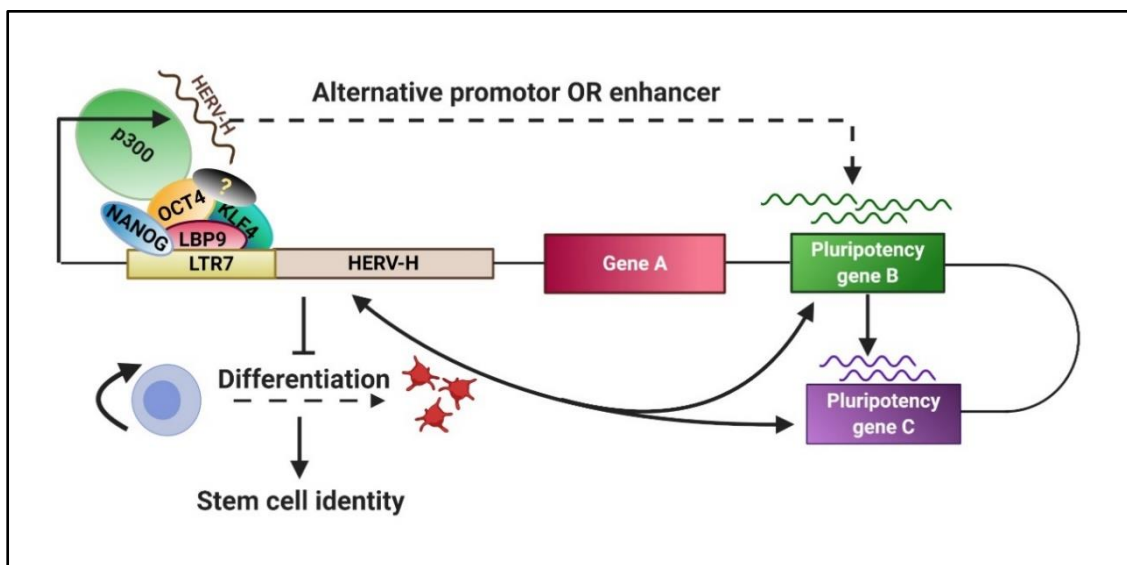
Figure 4. A scheme of the HERV-H functionality in stem cells.

The model shows that LTR7 of HERV-H functions as a scaffold to recruit pluripotency-associated transcription factors and histone modifiers such as p300. This incorporation controls LTR7 enhancer activity and leads to expression of neighboring genes, mainly involved in pluripotency and acquisition of stem cell identity.

created with BioRender.com

1.3.4. The presence and functional role of HERV-H in CRC

HERV-H transcripts have already been identified in normal (placenta, skeletal muscle, spleen, and thymus) and cancer cells (teratocarcinoma, bladder carcinomas, testicular tumors, and lung tumors) (141). Furthermore, Stauffer *et al*. revealed the preferential expression of HERV-H loci in CRC and prostate cancers (107). In 2004, Wentzensen *et al*. identified an upregulated *env* locus in colon adenomas and tumor tissues without notable changes in matched normal tissue (128). Later in 2007, overexpression of a HERV-H element on chromosome X was found in 47% of colorectal, 40% of gastric, and 17% of pancreas cancer samples (129). Likewise,

Alves *et al.* reported the selective transcription of a HERV-H locus on chromosome X in CRC tissues while it was not detected in normal colon (142).

As explained above and outlined in Table 2, the HERV-H family is particularly associated with the colon (128,129). Some of these sequences were exclusively found in CRC, but not in normal mucosa or in other organs (130,131,142). In 2012, it was proposed that CRC-specific HERV-encoded sequences might well represent a new class of tumor-specific antigens (143). Recently, in a pioneer work, CRC-specific HERV-H expression could be linked to the molecular subtype MSI. By integrating expression profiles, molecular patterns, and clinical data of a large CRC patients' cohort, HERV-H expression was strongly correlated with MSI status (131). Based on this initial study, the first monoclonal antibodies against a Gag protein sequence of the HERV-H family were produced and applied on tumor slides. This finally provided a more detailed picture of the subcellular localization of Gag-H proteins (144).

2. Aim of the study

The aim of this study was to investigate the functional relationship of pluripotent factors/stemness related genes (*LGR5, NANOG)* and the so-called Yamanaka factors (*OCT4, SOX2*, *KLF4,* and *c-MYC*)) with known CRC-HERV-H loci (131) as well as four new ESC-HERV-H sequences. The fact that the HERV-H family is somewhat exclusively expressed in ESCs and CRC lead us to hypothesize that HERV-H loci represent excellent markers for characterization of CRC stem-like cells (CR-CSCs) and might provide targetable structures for cancer therapy. To this end, we performed a comprehensive analysis on patient-derived cell lines from primary as well as metastatic CRC specimen covering all three molecular subtypes.

First, the expression level of pluripotent factors and HERV-H loci (CRC and ESCs relevant loci) was assessed in 56 primary and metastasis CRC cell lines using quantitative real-time PCR.

Then, stem cell enrichment of the candidate cell lines was accomplished by 3D culture and sorting of the SP, based on the hypothesis that CSCs enrichment results in elevated expression levels of stemness factors and of HERV-H loci permitting pluripotency.

Another aim of this study was to compare the mRNA (messenger RNA) expression level of genes in two different populations containing CSCs and non-CSCs. Finally, statistical analysis was performed from all examined samples with the aim to identify HERV-H pluripotency candidate loci.

This thesis handled the following four main tasks:

- Assessment of the expression level of pluripotent factors and HERV-H loci
- Enrichment of the stem cells via spheroid culture and sorting SP
- Comparison of the gene expression pattern in CSCs and non-CSCs populations
- Identification of the HERV-H pluripotency candidate locus/loci

3. Materials and Methods

3.1. Materials

3.1.1. Instruments

Products	Manufacturer
BD FACSAria™ IIIu	Becton Dickinson, Heidelberg, Germany
BD FACS Calibur	Becton Dickinson, Heidelberg, Germany
BD FACS LSRII flow cytometer	Becton Dickinson, Heidelberg, Germany
Biological safety cabinet class II	Thermo Fisher Scientific, Waltham, USA
Centrifuge 5415 D	Eppendorf, Hamburg, Germany
Freezer (-20 °C) and freezer (-80 °C)	Bosch and Kryotech, Germany
Gel documentation system	Vilber Lourmat, Germany
Heater HBT-1-131	Heap Labor Consult, Bovenden, Germany
Incubator	Thermo Fisher Scientific, Waltham, USA
Light microscope CKX 41	Olympus Deutschland, Hamburg, Germany
Microspin	Labnet International, Edison New Jersey, USA
Microwave	Bosch, Germany
Mr. Frosty™ freezing container	Thermo Scientific, Waltham, USA
Multicentrifuge 3 S-R	Thermo Scientific, Waltham, USA
Multistepper	Eppendorf, Hamburg, Germany
Nano Drop 1000	Thermo Fisher Scientific, Waltham, USA
Neubauer-counting chamber	Marienfeld, Lauda Königshofen, Germany
pH-Meter Multi 720	WTW inoLab, Germany
Pipetboy	Integra Bioscience, Fernwald, Germany
Pipettes	Eppendorf and Gilson, Germany
PowerEase 500 power supply	Bio-Rad-Laboratories GmbH, Germany
Refrigerator	Liebherr, Germany
Scale	Kern & Sohn GmbH, Balingen, Germany
Sigma centrifuge 2K15	Sigma, Osterode am Harz, Germany
Thermal cycler system MyCycler™	Bio-Rad, München, Germany
ViiA™ 7 Real-Time PCR System	Applied Biosystems®, Carlsbad, USA

Vortex Genie 2	Scientific Industries, New York, USA

3.1.2. Disposable material

Products	Manufacturer
Cell culture flasks (25 cm^2, 75 cm^2, 175 cm^2)	Greiner Bio-One, Kremsmünster, Austria
Cell culture flask (25 cm^2)-cell repellent	Greiner Bio-One, Kremsmünster, Austria
Cell culture plate (6, 48, 96 well)	Greiner Bio-One, Kremsmünster, Austria
Cell strainer (pore size: 40 μm)	Becton Dickinson, Franklin Lakes, USA
CryoPure tube (1.6 ml)	Sarstedt, Nümbrecht, Germany
FACS tubes	Sarstedt, Nümbrecht, Germany
Falcon tube Cellstar® (15 ml, 50 ml)	Greiner Bio-One, Kremsmünster, Austria
Fast PCR-plate half frame skirt	Sarstedt, Nümbrecht, Germany
Gloves	Hartmann AG, Heidenheim, Germany
Micro tube (0.5 ml, 1.5 ml, 2.0 ml)	Sarstedt, Nümbrecht, Germany
Pasteur pipette (130 and 250 cm length)	Thermo Fisher Scientific, Waltham, USA
Pipette tips (10 μl, 100 μl, 200 μl, 1000 μl)	Eppendorf and Sarstedt, Germany
Polystyrene round-bottom tube with cell-strainer cap (5 ml)	Becton Dickinson, Heidelberg, Germany
Sealing tape	Sarstedt, Nümbrecht, Germany
Serological pipette (5 ml, 10 ml, 25 ml)	Greiner bio-one, Kremsmünster, Austria

3.1.3. Chemicals and Reagents

Products	Manufacturer
Accutase	Sigma-Aldrich, Hamburg, Germany
Acetic acid	Roth, Karlsruhe, Germany
Agarose	Sigma-Aldrich, Hamburg, Germany
Dulbecco's modified Eagle's medium (DMEM/Ham's F12 (1:1))	Gibco, USA
DMEM/Ham's F12 (1:1)	Pan-Biotech, Aidenbach, Germany
DMEM high glucose	Sigma-Aldrich, Hamburg, Germany

Dimethyl sulfoxide (DMSO)	AppliChem, Darmstadt, Germany
Disinfectants Bacillol® AF	Bode Chemie, Hamburg, Germany
Di-sodium hydrogen phosphate dihydrate	MERCK, Darmstadt, Germany
50 bp and 1 kb DNA ladder	Bioron, Ludwigshafen, Germany
100 bp DNA-ladder, equimolar	Roth, Karlsruhe, Germany
DNase I Recombinant	Roche, Germany
Deoxynucleotide triphosphate (dNTP)	Bioron, Ludwigshafen, Germany
Dulbecco's phosphate buffered saline (DPBS)	Pan-Biotech, Aidenbach, Germany
Ethylenediaminetetraacetic acid (EDTA)	AppliChem, Darmstadt, Germany
EGF	Peprotech, USA
Ethanol, absolute	Walter CMP, Kiel, Germany
Ethidiumbromid (EtBr)	Roth, Karlsruhe, Germany
Fetal bovine serum (FBS)	Pan-Biotech, Aidenbach, Germany
Formafix 4%	Formafix GmbH, Düsseldorf, Germany
Gelatin 0.1%	MERCK, Darmstadt, Germany
Hanks' Balanced Salt Solution (HBSS)	Gibco, USA
4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid (HEPES)	Sigma-Aldrich, Hamburg, Germany
Hoechst 33342	Cell signaling, USA
Isopropanol	Walter CMP, Kiel, Germany
Knockout^{TM} SR	Gibco, USA
L-glutamine	Pan-Biotech, Aidenbach, Germany
Loading Dye (6X)	Promega GmbH, Germany
β-mercaptoethanol (β-ME)	Gibco, USA
Mitomycin C	Biomol GmbH, Hamburg, Germany
Non-essential amino acids (NEAAs)	Pan-Biotech, Aidenbach, Germany
Oligo (dt) primer	Metabion, Planegg, Germany
PCR buffer (10X)	Bioron, Ludwigshafen, Germany
Potassium chloride	MERCK, Darmstadt, Germany
Potassium di-hydrogen phosphate	MERCK, Darmstadt, Germany
Primer qPCR	Metabion, Planegg, Germany

Probes qPCR	Metabion, Planegg, Germany
Propidium iodide (PI)	AppliChem, Darmstadt, Germany
Random primer	Metabion, Planegg, Germany
Reverase™ – 200 units	Bioron, Ludwigshafen, Germany
RNase-free water	Roth, Karlsruhe, Germany
ROX	Bioron, Ludwigshafen, Germany
RT buffer complete (5X)	Bioron, Ludwigshafen, Germany
Saponin	AppliChem, Darmstadt, Germany
Sodium chloride	Carl Roth GmbH, Karlsruhe, Germany
SibirRoxHot MasterMix	Bioron, Ludwigshafen, Germany
Tris-Acetate-EDTA (TAE, 50X)	AppliChem, Darmstadt, Germany
TaqMan master mix	Bioron, Ludwigshafen, Germany
Taq DNA Polymerase 5 U/µl	Bioron, Ludwigshafen, Germany
Tris	Roth, Karlsruhe, Germany
TRISure™	Bioline, Luckenwalde, Germany
Trypan blue (0.02%)	Sigma-Aldrich, Hamburg, Germany
Trypsin/EDTA	Pan-Biotech, Aidenbach, Germany
Verapamil 5 mg	Ratiopharm, Germany

3.1.4. Kits

Products	Manufacturer
Direct-zol MiniPrep	Zymo Research Europe GmbH, Germany
EMD Millipore Human STEMCCA lentivirus reprogramming kit	MERCK, Darmstadt, Germany
E.Z.N.A.® MicroElute®	Omega Bio-tek, USA
Fixation/permeabilization solution kit	BD Bioscience, Germany
GeneMATRIX universal RNA purification kit	EURx, Poland
innuPREP RNA Mini kit	Analytik Jena, Germany
peqGOLD MicroSpin total RNA kit	Peqlab, Germany
Quick-RNA™ MiniPrep kit	Zymo Research Europe GmbH, Germany
Wizard genomic DNA purification kit	Promega GmbH, Germany

3.1.5. Probe- and Primer sequences for RTq-PCR

Table 3. *List of the probe and qPCR primers.*

Gene name	Primer sequences	Reference
GAPDH probe	VIC-ACAGCGACACCCACTCCTCCACC-BHQ-1	
GAPDH	F: 5′ -GAAGGTGAAGGTCGGAGTC- 3′ R: 5′ -GAAGATGGTGATGGGATTTC- 3′	(131)
OCT4	F: 5′ -TCCCATGCATTCAAACTGAGG- 3′ R: 5′ -CCAAAAACCCTGGCACAAACT- 3′	(145)
SOX2	F: 5′ -CCATCCACACTCACGCAAAA- 3′ R: 5′ -TATACAAGGTCCATTCCCCCG- 3′	(145)
NANOG	F: 5′ -TGGACACTGGCTGAATCCTTC- 3′ R: 5′ -CGTTGATTAGGCTCCAACCAT- 3′	(145)
KLF4	F: 5′ -CTGCGGCAAAACCTACACAA- 3′ R: 5′ -GGTCGCATTTTTGGCACTG- 3′	(145)
c-MYC	F: 5′ -ACACATCAGCACAACTACG- 3′ R: 5′ -CGCCTCTTGACATTCTCC- 3′	(146)
LGR5	F: 5′ -TGCTCTTCACCAACTGCATC- 3′ R: 5′ -CTCAGGCTCACCAGATCCTC- 3′	(147)
X00041_h_gag	F: 5′ -CAGGCGTTGCTGAGTGTGTCTAATC- 3′ R: 5′ -TGGAGCCTGAGGAAGAATTGGGACC- 3′	(131)
2000045_h_gag	F: 5′ -CCCAAGCGGCGCTGAGTCTT- 3′ R: 5′ -TGGGATGAAGGGAGGGGAGGC- 3′	(131)
2000045_h_pol	F: 5′ -TACAGCATGGGCACCTATAAACTCT- 3′ R: 5′ -TAAGTGAAGGCAAAGAGAGGCTGGG- 3′	(131)
500502_h_L3U3	F: 5′ -CCCAGATGGCCTGAAGTAACTGA- 3′ R: 5′ -AGCCAGGAGAACAATTCACAGGGTT- 3′	(131)
1300360_h_gag	F: 5′ -AGTGCAACTCATTCTGAATCTTCCT- 3′ R: 5′ -CACAGAACGAAACTGTAAGCCAG- 3′	(131)
1300360_h_env	F: 5′ -CTGAACCTCCTTAGGCATTCTCT- 3′ R: 5′ -GTGACATTGAGGGGGTTGTTAGAAG- 3′	(131)

1400035_h_gag	F: 5′ -CCCAAGTGTCGCTGAGTCTTTCT- 3′	(131)
	R: 5′ -GATTACAGGGTGCAGGAGCAGAG- 3′	
hESC_HERV_H _4	F: 5′ -ACGCTTTACAGCCCTAGACC- 3′	
	R: 5′ -GTCGGGAGCAGATTGGGTAA- 3′	
hESC_HERV_H _6	F: 5′ -TGGACCTCTCACAACACAAACT- 3′	
	R: 5′ -AGGGGAATTCCAGTGGGTCT- 3′	
hESC_HERV_H _15	F: 5′ -TGGTGCCAAACCCATATACTC- 3′	
	R: 5′ -AATAGGAAAGAAAGCATGTTTGAGA- 3′	
hESC_HERV_H _19	F: 5′ -TGGTGCCAAACCCATATACTC- 3′	
	R: 5′ -AATAGGAAAGAAAGCATGTTTGAGA- 3′	

Table 4. *List of primers used to examine the presence of the hSTEMCCA vector in gDNA.*

Primer sequences
cMYC F: 5′ -GGAACTCTTGTGCGTAAGTCGATAG- 3′   **WPRE** R: 5′ -GGAGGCGGCCCAAAGGGAGATCCG- 3′
hOCT4 5′ NotI: 5′ -TTTTGCGGCCGCCATGGCGGGACACCTGGCTTCGG- 3′   **hKLF4 3′ BamHI:**   5′ -TGTTGGATCCTTAAAAATGCCTCTTCATGTGTAAGGCG- 3′
hSOX2 5′ NdeI: 5′ -TTTAGTGCATATGATGTACAACATGATGGAGACGGAGCTG- 3′   **hcMYC 3′ AccI:**   5′-TTTAGCAGTGGTACGTCGACTTACGCACAAGAGTTCCGTAGCTGTTC- 3′

3.1.6. Media, Buffer and Solutions

Media/ Buffer/ Solutions	Component
Cancer cell culture medium	445 ml DMEM/Ham's F12 (1:1) 50 ml FBS (final: 10%) 5 ml L-glutamine 200 mM (final concentration: 2 mM)
Mouse embryonic fibroblast (MEF) cell culture medium	44 ml DMEM high glucose 5 ml FBS (final: 10%) 0.5 ml L-glutamine 200 mM (final concentration: 2 mM)

Transduced cell culture medium	195 ml DMEM/Ham's F12 (1:1)
	50 ml Knockout^{TM} Serum Replacement (final: 20%)
	2.5 ml NEAAs (100X)
	2.5 ml β-ME (100X)
Freezing medium	45 ml serum (final 90%)
	5 ml DMSO (final 10%)
DMEM$^+$	48.5 ml DMEM/Ham's F12 without indicator
	1 ml FBS
	0.5 ml 1 M HEPES (final concentration 10 mM)
HBSS$^+$	48.5 ml HBSS
	1 ml FBS
	0.5 ml 1 M HEPES (final concentration 10 mM)
10X PBS pH 7.2-7.4	80 g sodium chloride
	2 g potassium chloride
	18.05 g di-sodium hydrogen phosphate dihydrate
	2 g potassium di-hydrogen phosphate
	Ad 1000 ml dH$_2$O
1M HEPES	23.83 g HEPES
	100 ml 1x PBS
50X TAE buffer pH 8.3	242 g of Tris-base (MW = 121.14 g/mol)
	57.1 ml of acetic acid
	100 ml of 0.5 M EDTA (pH 8.0)
	Ad 900 ml dH$_2$O
1X TAE working solution	20 ml 50X TAE
	Ad 980 ml dH$_2$O
100X Buffer P	0.5 ml FBS (final concentration 1%)
	5 ml 1% Saponin (final concentration 0.1%)
	5 ml 0.1 M HEPES (final concentration 0.01 M)
	39.5 ml 1x PBS
100 ml of MACS buffer	0.5% BSA
	400 µl 0.5 M EDTA (final concentration 2 mM)
	1x PBS

3.1.7. Cell lines

Table 5. List of the colorectal carcinoma cell lines used for this cohort study.

HCT116	HROC103Met	HROC277Met2
HROC18 *	HROC107 cT0 M2	HROC278 T0 M1 *
HROC24 *	HROC111Met1 T0 M2	HROC278Met T2 M2 *
HROC24 T1 M1	HROC112Met T0 M2	HROC284Met
HROC32	HROC113 cT0 M1	HROC285 T0 M2 *
HROC32 T3 M1	HROC126	HROC296
HROC39	HROC131 T0 M3	HROC300 T2 M1 *
HROC39 T0 M2	HROC147Met	HROC309 *
HROC40 *	HROC147 T0 M1 *	HROC313Met1 T0 M2 *
HROC43	HROC173	HROC324 *
HROC46 T0 M1 *	HROC183	HROC334
HROC50 T1 M5 *	HROC183 T0 M2	HROC348Met
HROC57	HROC212	HROC357
HROC59 T1 M1	HROC222 T1 M2	HROC364 *
HROC60	HROC239 T0 M1	HROC370
HROC69	HROC257	HROC374
HROC69 T0 M2	HROC257 T0 M1	HROC383 *
HROC80 T1 M1	HROC277 T0 M1	HHC6548 T1 M1
HROC87 T0 M2 *	HROC277Met1 T0 M2	HROBMC01 *

Cell lines marked with (*) are the selected cell lines and they were applied in further experiments. Abbreviations T, M, and Met stand for transfer, mouse, and metastasis. Complementary cell lines' information is listed in supplementary Table S1.

3.2. Methods

3.2.1. Preliminary work and origin of the cell lines

The Molecular Oncology and Immunotherapy group (headed by PD Dr. Linnebacher) is very experienced in establishing patient-individual tumor models for basic and translational research. Primary CRC resection specimens from in-house performed oncological resections are routinely obtained fresh from surgery, with informed written patient consent and upon pathological evaluation by an experienced pathologist (Prof. Dr. Prall, Institute of Pathology). Here, material was utilized upon cell line establishment either directly from individual patients' tumors (patient-derived cell line) and/or the corresponding xenografts (patient-derived xenograft, PDX). So far, the group has successfully established N>50 patient-derived cell lines and N>110 PDX models. These cell lines and PDX are comprehensively characterized in comparison to the parental tumor by phenotype, morphology, invasiveness, and molecular profile (148). Additionally, response to clinically relevant chemotherapeutics was previously examined *in vitro* and *in vivo* and correlated with clinical data (149–152). These models very closely mirror the clinical behavior and biology of the original tumor – this is reflected by a growing number of high-impact publications resulting from national and international collaborations (153–155).

All cell lines are marked with the prefix HROC (Hansestadt Rostock colorectal cancer) and an ID number, e.g. HROC18. Besides, in the xenograft-derived cell lines, M and T are additional indicators for the mouse number and number of *in vivo* passages. In this pilot study, a large cohort of CRC cell lines (N=56) was used for the preliminary analysis. Finally, 17 cell lines were selected for the further experiments listed in Table 5. Moreover, the SNL76/7 [4] cell line (kindly provided from AG Prof. Robert David) was cultured as feeder cells for the stem cell culture in later experiments.

3.2.2. Cell culture

3.2.2.1. Monolayer or two-dimensional (2D) cultivation

Cell culture was performed under a sterile laminar flow hood. Cells were cultured in 25 cm^2 (T25) culture flasks and kept in a humidified incubator with 5% CO_2 at 37 °C. Tumor cells

---

4. Immortalized cell line derived from a MEF STO cell line and transformed with murine LIF and neomycin resistance genes.

were maintained in the antibiotic-free DMEM/Ham′s F-12 (1:1) medium supplemented with 10% FBS and 2 mM L-glutamine. Cell lines were used in passages ≤50 to assure comparability with the parental tumor in terms of molecular characteristics and growth kinetics. SNL76/7 cells were cultured in DMEM high glucose enriched by 10% FBS, 0.1 mM NEAAs and 2 mM L-glutamine.

Thawing and expansion of cells

One vial of cryo-conserved cells was quickly thawed at room temperature (RT) and resuspended in 5 ml of pre-warmed DMEM culture media following by a centrifugation step for 8 min at 300 x *g*. The cell pellet was suspended in fresh medium and immediately transferred into a T25 culture flask. The medium was exchanged every three or four days. Cells were subsequently expanded in a bigger culture flask when they reached nearly 80% confluence. Therefore, culture medium was completely aspirated and cells were washed with pre-warmed DPBS. Thereafter, enzymatic cell detachment was done by using 0.25% Trypsin/EDTA and incubation at 37 °C for 5-10 min. Cell detachment was checked under a light microscope and the reaction was neutralized by adding an appropriate volume of culture media. Afterwards, cell single suspension was transferred to the tube and cells were counted as described below. At the end, required numbers of viable cells were used for subsequent expansion, experiments and cryopreservation.

Determination of cell number and viability

Total cell number and viability was determined by trypan blue assay and a Neubauer counting chamber. Trypan blue is a cell stain that discriminates between vital and dead cells. It can penetrate the defective cell membrane and enter the cytoplasm. Hence, only nuclei of dead cells absorb trypan blue. For this analysis, 10 µl of cell suspension were diluted with 10 µl of 0.02% trypan blue solution (1:2). Then, the chamber was loaded with 10 µl of the diluted cell suspension and two opposite squares were counted. The number of cells was calculated with the following formulation:

$$\frac{\text{cell number}}{\text{ml}} = \frac{\text{counted viable cells}}{\text{number of the counted squares}} \times 2 \text{ (dilution factor)} \times 10^4 \text{ (chamber factor)}$$

Cryopreservation

Principle

Aim of the cryopreservation is to provide a qualified cell backup at early passages for a long-term study. There are increasing evidences supporting that cell lines' traits are influenced by prolonged cell culture due to the selective pressure and genetic alterations. Consequently, obtained results are not reliable and functional features of the cell line are incompatible with the data of the original source or identified characteristics at earlier passages (156–158).

Procedure

For this purpose, confluent monolayer was dispersed and cell number was counted as described above. After centrifugation (8 min at 300 x *g*), 3×10^6 cells were resuspended in 1.5 ml freezing medium per aliquots. Thereafter, cryo-tubes were kept in freezing containers and stored at -80 °C. This container was filled with isopropanol that provided a consistent drop of the temperature by 1 °C/min for effective cell storage.

3.2.2.2. Spheroid or three-dimensional cultivation (3D)

Principle

Spheroid or so-called 3D culture is one of the approaches to enrich cancer stem-like cells from bulk cancer cells *in vitro*. 3D physical structures stimulate better cell-cell and cell-matrix interactions and develop stronger and also different cellular responses (159). In fact, these interactions can influence cells morphology, proliferation rates, differentiation, as well as molecular alterations such as genes and proteins expression. Therefore, cells environment and their behavior are more reflective of actual *in vivo* state (160). In general, two main points must be considered in this culturing system; including culture surface as well as culture medium. Low adhesion surface prevents cells' adherence to the bottom and promotes cell-cell attachment; resulting in sphere formation. In spite of the classical monolayer culture, serum free medium supplemented with several growth factors is typically applied for culturing tumorspheres[5] in various cancer entities.

---

5. In this study CRC-tumorspheres have been termed "colonospheres or spheroids".

__Procedure__

In this experiment, almost 8×10^5 cells were washed with 1x PBS, pelleted by centrifugation (10 min, at 300 x g). After that, the pellet was suspended in an in-house organoid medium[6] (161) and cells were plated in repellent cell culture flasks (25 cm^2). Cells were checked daily by a light microscopy and cell clumps were broken by pipetting up and down in order to refuse cell aggregations. The medium was replenished every three days and spheres were snap frozen after being cultured for 6 days. They were kept at -80 °C for further analysis.
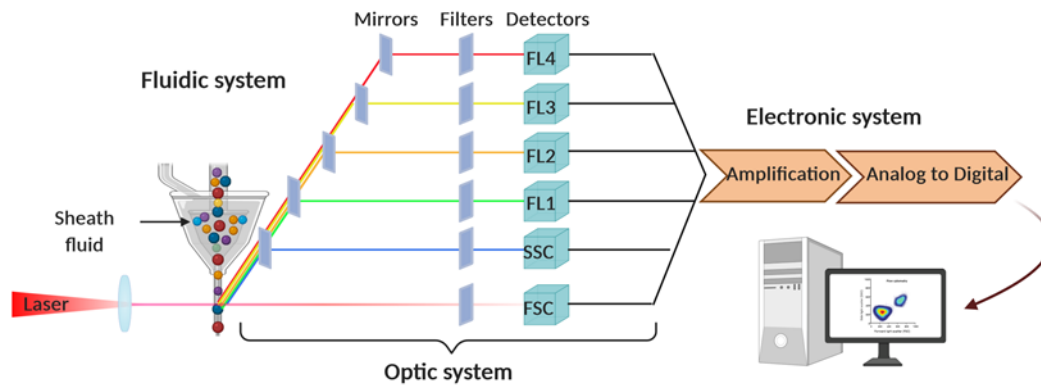
3.2.3. Flow cytometry

__Principle__

Flow cytometry is a technology utilized to individually measure the particle's properties within a cell population. This process provides information by using fluorescence and light scattering. Therefore, it allows cell quantification and discriminates subpopulations based on their size (forward scatter), intracellular components (side scatter, measured at a 90-degree angle), nucleic acid content, surface and intracellular receptors, and cells' functionalities. Principally, labelled cells flow in a single stream of particles produced by hydrodynamic focusing. Thereafter, each particle is presented to a laser beam. Consequently, fluorochromes are excited and emitted signals passes through the various lenses. Filters then block certain wavelengths while passing the others to detectors. At the end, collected lights are converted into electrical or voltage pulses by photomultiplier tubes. These analogue signals are digitized and recorded as an event to desired plots (162) (Figure 5).

In this study, flow cytometry was applied to sort SP fraction and to investigate HERV-H and NANOG expression by intracellular staining in the adherent and spheroid models. Basically, $2\text{-}5 \times 10^5$ cells were intracellularly stained with respective antibodies or isotypes. For acquisition, 10.000 events at low speed were recorded in the gated area. Results of three independent experiments were analyzed for each experiment.

---

6. This medium is pretty much according to the protocol originally described by the Dutch group of Hans Clevers and the patent is in progress.

Figure 5. Flow cytometry components and function.

The three fundamental components are fluidics, optics, and electronics. The fluidic system contains the flowing cells provided by sheath fluid and consequent hydrodynamic focusing effects. The optics system consists of an excitation light source ranging from ultraviolet to far red, lenses for lights collection, various mirrors, filters to detect a particular wavelength, and detectors to measure the emitted fluorescence signal intensity. The electronic device digitizes and converts the light currents to data that will be visualized and interpreted by software (162).

created with BioRender.com

3.2.3.1. Evaluation of HERV-H expression in 2D and 3D cultures

A total of 2×10^6 cells were distributed equally in 5 wells of a 96-well plate and centrifuged 10 min at 300 x *g*. Cell pellets were fixed in 100 µl of 2% Formafix followed by 15 min incubation at RT. After centrifugation (300 x *g*, 10 min) cells were washed with 200 µl 1x PBS and centrifuged again. Next, pellets were incubated 10 min at RT in the diluted buffer P with 1x PBS (1:100). During incubation time, anti-Gag-H antibodies (14H11G1, 1B3H7, and 1D7D11, concentration: 1 µg) and the irrelevant antibody (2G2B3, concentration: 1 µg) were separately mixed with diluted buffer P under sterile conditions. Plate was centrifuged again and cell pellets were resuspended in 200 µl of appropriate antibodies solutions. Incubation was done at RT for 30 min. Then, 100 µl buffer P was added to each well and the plate was spun down at 300 x *g* for 10 min. After removing supernatant, cells were incubated for 30 min in the dark with 50 µl RPE- goat anti-mouse antibody as a secondary antibody (dilution: 1:40 in buffer P). The cells were washed again in 100 µl buffer P and finally resuspended in 2% Formafix for analysis by a FACS Calibur flow cytometer.

3.2.3.2. Evaluation of NANOG expression in 2D and 3D cultures

Samples were washed with 1x PBS and suspended in a final volume of 200 µl MACS buffer. Plate was spun down at 300 x g for 5 min in 4 °C and supernatant was aspirated carefully. Next, maximum of 100 µl Fixation/Permeabilization solution was added to each well to simultaneously fix and permeabilize the cells. Plate was incubated 20 min at RT and 1X Perm/Wash buffer was added up to 250 µl final volume and centrifuged. After discarding supernatant, pellets were mixed with PE-NANOG and isotype antibodies diluted in 50 µl and 160 µl of 1X Perm/Wash, respectively. Following one-hour incubation at 4 °C, 200 µl 1X Perm/Wash buffer was added and centrifuged again. Lastly, cells were resuspended in 250 µl 1x PBS and transferred to the 5 ml FACS tubes for measurement using a BD FACS LSRII flow cytometer (kindly provided from AG Prof. Robert David). Mean fluorescence was analyzed with BD FACSDiva Software 6.1.2 (BD Biosciences, Germany).

3.2.3.3. Sorting of SP by flow cytometry

Principle

The SP has become a hallmark for stem cell characterization (163) and the SP discrimination has emerged as a promising method for primary purification and isolating stem/progenitor cells in different tissues and cancer entities. This assay relies on flow cytometry and is based on the cells' potential to efflux the Hoechst dye via the ABC family of transporter proteins (164).

Procedure

Cells in the exponential growth phase (approximately 80% density) were harvested as described before. After counting, at least 15×10^6 cells were washed with calcium and magnesium free HBSS. Cells were resuspended in 3 ml pre-warmed DMEM$^+$ (DMEM + 2% FBS + 10 mM HEPES) and FACS tubes were prepared for each cell line: unstained (5×10^5 cells), control (5×10^5 cells), and test (remaining cells). Subsequently, 5 µg/ml Hoechst 33342 was added to the test tube. In the control tube, before adding Hoechst 33342, verapamil (50 µM) was added to block the dye efflux by membrane bound pumps. All samples were incubated at 37 °C for 90 min in the dark and cells were shaken occasionally. After 5 min centrifugation at 500 x g, cells were resuspended in ice-cold HBSS supplemented with 2% FBS and 10 mM HEPES. Finally, PI (1 µg/ml) was added shortly before measurement to discriminate living

from dead cells. Moreover, the purity of sorted samples was re-analyzed. Sorted populations were washed with 1x PBS and frozen in liquid nitrogen for subsequent analysis. Importantly, the experiment was repeated for each cell line when re-analysis of sorted populations revealed low purity by not reaching a distinct separation in both gates (Hoechst[low] *vs.* Hoechst[high]). This test was performed in all selected cell lines and samples were stored at -80 °C.

3.2.4. Purification of nucleic acid (DNA and RNA)

3.2.4.1. Genomic DNA (gDNA) extraction

Cell pellets of either fresh or snap frozen cells (up to 3×10^6 cells) were used for gDNA isolation. This was done by using the Wizard Genomic DNA Purification kit according to the manufacturer's instructions. The three main steps in this process were lysis, precipitation, and purification. First, Nuclei Lysis and RNase solutions were added to the samples followed by incubation at 37 °C for 15-30 min. In the next step, protein precipitation solution and isopropanol were added to precipitate DNA. Therefore, DNA was separated from the aqueous phase. In the final step, the DNA was purified by washing the DNA pellet with ethanol and re-dissolved DNA Rehydration solution. Concentration and purity of the DNA was measured with NanoDrop 1000. The value of ~ 1.8 for A260/280 ratio was interpreted as "pure". The samples were kept at -20 °C for longterm storage.

3.2.4.2. RNA extraction

Principle

RNA molecule has a short lifetime and it is particularly unstable due to the pervasive existence of endogenous and exogenous RNases. Some denaturants used in RNA purification inhibit RNases activity. However, handling of samples prior, during, and after RNA extractions can also minimize the exposure of samples to the exogenous RNases. Precautions must be made to obtain high quality and pure RNA, the first and the most critical step in performing molecular techniques.

Procedure of RNA isolation of large number of cells

Total RNA from cell culture pellets (3×10^6 cells) and spheres pellets were isolated using EURx GeneMATRIX universal RNA purification kit according to the manufacturer's

instructions. Briefly, cell pellets were lysed with 400 µl of RL buffer containing ß-ME and were mixed thoroughly by vigorous vortexing. The lysates were transferred to the homogenization spin-column and were centrifuged at maximum speed for 2 min. The flow-through was mixed with 350 µl 70% ethanol by pipetting and mixture, including any precipitate, was transferred to the RNA binding spin-column. After centrifugation for 1 min at 11,000 x g, columns were washed with 400 µl Wash DN1 buffer and centrifuged again. The spin-columns were removed and the fluid was discarded. Next, 50 µl DNase I solution (1 U of DNase I solved in the mixture of 45 µl DEPC water and 5 µl of DNR buffer) was directly added onto the membrane and place on the benchtop at RT for 10 min. The columns were then washed with 400 µl Wash RB1 buffer and subsequently two times with RBW buffer. The spin-columns were placed into the new RNase-free tubes (1.5-2 ml) and maximum 40 µl RNase-free water was applied directly to the membrane to elute the RNA. RNA was immediately placed on ice and quantity and quality were measured. Finally, high-quality samples were aliquoted in order to prevent freeze-thaw damage and stored at -80 °C for further analysis.

Procedure of RNA isolation from small number of the sorted samples

RNA from sorted samples was extracted with the peqGOLD MicroSpin kit. Briefly, frozen sorted cells were lysed with 300 µl of RNA Lysis Buffer T. Lysates were applied to a DNA removing column and centrifuged for 1 min at 10.000 x g. The DNA columns were discarded and 200 µl 70% ethanol was added to filtrates. The whole mixture was placed onto the PerfectBind RNA columns. After centrifugation, flow-throughs were discarded and the RNA columns were washed with 500 µl RNA Wash Buffer I followed by two more washing steps with 500 µl RNA Wash Buffer II. RNA Columns were then centrifuged for 2 min at 10.000 x g to completely dry the PerfectBind matrix. Finally, RNA samples were eluted with 15-30 µl of RNase-free water and placed on ice for quality control.

Assessment of RNA quality

Firstly, the concentration and purity of RNA were photometrically determined with the NanoDrop 1000 via absorbance at 230 nm, 260 nm and 280 nm and corresponding ratios of

A260/230 and A260/280[7] (165). In the second approach, quality tests were completed by using the highly sensitive TaqMan® system described in 3.2.7.1.

3.2.5. Synthesis of complementary DNA (cDNA)

Highly pure RNA (1 µg) samples were utilized for first-strand cDNA synthesis by reverse transcriptase. They were diluted in RNase free water to a volume of 13 µl in PCR tubes. Afterwards, 1 µl of oligo (dt) primer (100 µM) and random primer (50 µM) mixture was added to each tube. The samples were then incubated for 10 min at 70 °C in the Thermal Cycler (BioRad) and placed on ice. Then, a master mix was added as listed in the Table below.

Table 6. Compositions of cDNA master mix.

Component	Volume/Reaction (µl)
5X RT buffer complete	4.0
dNTP mix (10 mM of each dNTP)	1.0
Reverase™ – 200 units	1.0

Cyclin conditions were as follows: 120 min at 45 °C and 10 min at 70 °C. The cDNA samples were immediately put on ice and diluted in the nuclease free water (1:4) to achieve an approximate concentration of 12.5 ng/µl. Synthesized cDNA was aliquoted and kept at -20 °C.

3.2.6. Polymerase chain reaction (PCR)

Principle

PCR is a simple method in molecular biology that enables the amplification of desired DNA segments (amplicons). This method principally rests on three main cyclic reactions including template denaturation, primer annealing, and elongation (Figure 6). Therefore, amplicons are multiplied by using a heat-stable DNA polymerase, primer pairs matched with the 3′ ends of the sense and anti-sense strands of the DNA target sequences, dNTPs (dATP, dTTP, dGTP,

---

7. Protein and phenol contaminations are determined by the A260/A280 ratio, whereas the A260/A230 ratio indicates presence of carbohydrates, lipid, phenol, EDTA or salts contaminants.

dCTP) and buffer system. Theoretically, the amount of the DNA template is doubled in each cycle and then size, sequence, and quantity of the amplified product can be analyzed.
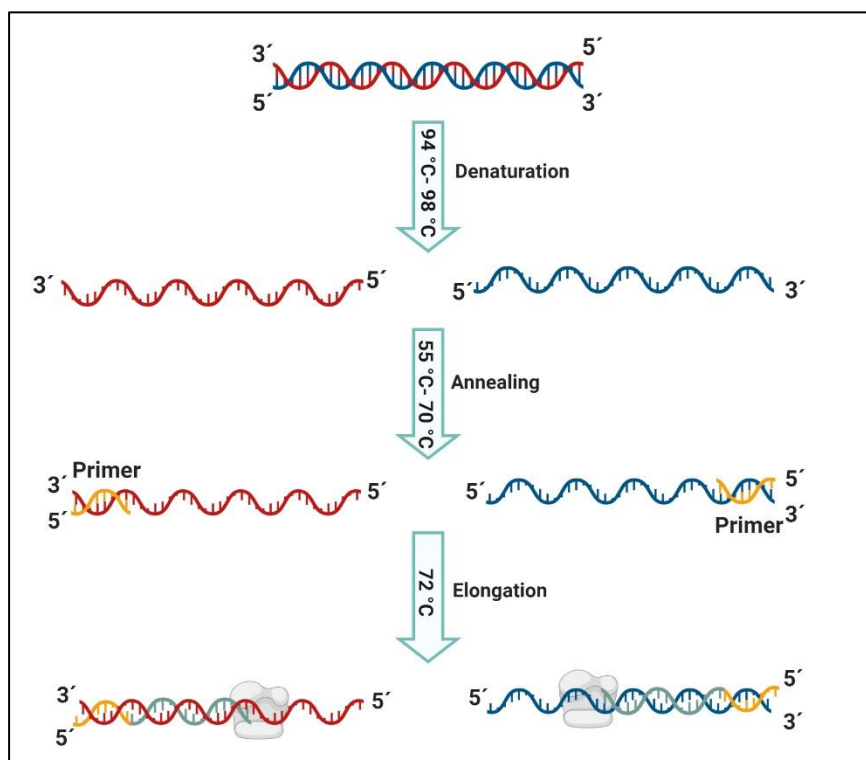
Figure 6. Three key steps in PCR -denaturation, annealing, extension- to amplify DNA target. created with BioRender.com

Procedure

A mixture of PCR components (listed in the Table 7) was prepared and 23 µl was pipet in each tube. As a template, 2 µl of cDNA or gDNA (12.5 ng/µl) was added to each tube. A negative control (non-template control, NTC) for each primer was always included in the PCR experiments to indicate contamination of the reagents as well as primer-dimer formation that could produce false positive results. PCR reaction was done using the program given below.

Table 7. Compositions of PCR master mix.

Component	Volume/Reaction (µl)
10X buffer complete	2.5
dNTP mix (10 mM of each dNTP)	0.5

Forward primer (10 pMol/µl)	0.5
Reverse primer (10 pMol/µl)	0.5
Taq DNA Pol (500 units)	0.2
Nuclease free water	18.8
Final volume	$\sum$ 23

Table 8. *Thermal cycler condition in PCR.*

Step	Temperature	Time	Number of Cycles
Initial denaturation	95 °C	5 min	1
Amplification:			
I. Denaturation	95 °C	15-30 sec	
II. Annealing	The determined temperature	15-60 sec	40
III. Extension	72 °C	30 sec	
Elongation	72 °C	10 min	1

3.2.7. Quantitative reverse transcription PCR (RT-qPCR)

Principle

This method helps to investigate relative gene expression using RNA as the starting material. In a first step, RNA is converted into cDNA via reverse transcriptase as described before. Then, cDNA is applied as a DNA template in qPCR. This technique is principally similar to conventional endpoint PCR. However, in qPCR, fluorescent dyes enable the quantitative measurement of the amplified products. The data collection is done at each cycle before the plateau phase by monitoring reactions during the exponential phase (166). In this process, the fluorescent signal intensity rises proportional to the quantity of the replicated DNA. Hence, information about the expression level of a target is given by a threshold cycle (Ct) value for each sample, meaning the cycle number at which the amount of the collected signals intersects the background fluorescence. It is defined as a real signal of the target gene allowing comparison of Ct values to provide quantitative data of a relative mRNA expression level.

Regardless of the methods used for qPCR in the current project, qPCR reactions were prepared as duplicates in a 96-well plate. The composition of qPCR master mixes can be taken from

Tables 9 and 10. First, 10.5 µl of the master mix was aliquoted in each well and 2 µl of cDNA (25 ng), gDNA (25 ng) or RNA (8.1 ng) was added to the wells. Exposure to the direct light was minimized and all reagents were kept on ice during preparation. The plate was closed with optically clear sealing tape and centrifuged to remove any air bubbles before running the plate. QPCR was accomplished in the ViiA™ 7 Real-Time PCR System according to the set-up program.

Table 9. Compositions of master mix for TaqMan® PCR.

Component	Volume/Reaction (µl)
2X TaqMan MasterMix	6.25
Mix of primer + probe (20X)[8]	0.625
ROX solution, 25 µM (1:10)	0.15
Nuclease free water	3.475
Final volume	$\sum$ 10.5

Table 10. Compositions of master mix for SYBR Green® PCR.

Component	Volume/Reaction (µl)
SibirRoxHot MasterMix	6.25
Forward primer (10 pMol/µl)	0.625
Reverse primer (10 pMol/µl)	0.625
Nuclease free water	3
Final volume	$\sum$ 10.5

3.2.7.1 Probe-based qPCR (TaqMan®)

Principle

In this approach, two single stranded oligonucleotides including primers and a probe complementary to a specific DNA fragment are used. Primers are naturally hybridized with the ssDNA template and initiate the target replication. Conversely, probes are hybridized with

---

8. Concentrated probe-primer system consists of 2 µl probe (final 20 µM), 10 µl of forward and reverse primer (final 100 µM) soluble in 78 µl nuclease free water.

dsDNA for the detection of a particular DNA sequences in a sample. These probes are labeled by a fluorescent reporter dye and a quencher dye at the 5' and 3' ends (Figure 7B).
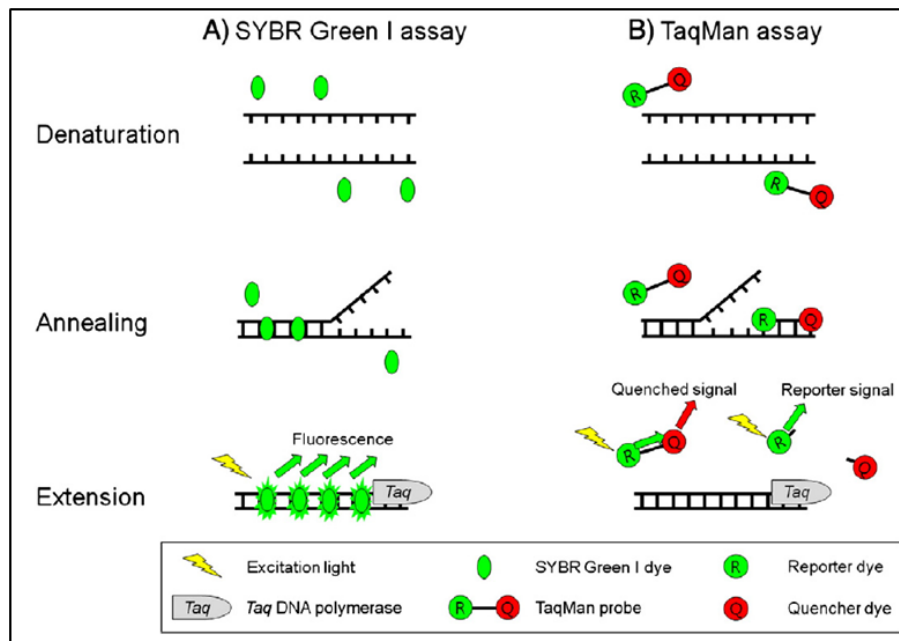
Figure 7. Comparison of the dye-based qPCR and probe-based qPCR (166)*.*

The level of the fluorescent output in an intact probe is reduced due to the energy-absorbing effects of the quencher. After the elongation step, the reporter fluorophore is cleaved by the 5'-3'-exonuclease activity of the Taq-polymerase. This fluorophore liberation and consequent further distance between quencher and fluorophores lead to an increasing fluorescent level in each cycle which is recorded by an instrument.

__Procedure__

Probe-based qPCR was done to assess the quality of the RNA and subsequent cDNA samples via an in-house *GAPDH* primer-probe and using the following program.

- Hold stage (initial denaturation): 95 °C for 10 min
- PCR stage (amplification, 40 cycles):
  I. Denaturation at 95 °C for 15 sec
  II. Annealing at the determined temperature for 1 min

For RNA quality evaluation, 8.1 ng of RNA samples was purely used and *GAPDH* expression was compared with a positive control (gDNA) to ensure absence of the DNA contamination in an isolated RNA sample. This is a mandatory technical issue when analyzing HERVs expression because of the absence of introns in the HERVs sequences (129) and lack of RNA processing. RNA with Ct $\geq$ 32 and cDNA with Ct $\leq$ 25 were considered qualified (Figure 8).
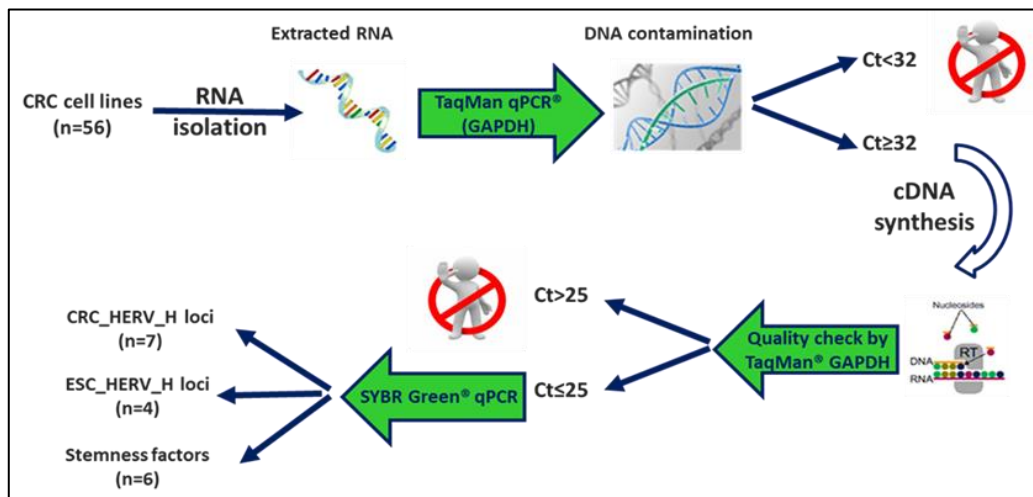
Figure 8. The schematic procedure of quality control.

3.2.7.2 Dye-based qPCR

Principle

This method uses a fluorescent DNA-binding dye, such as SYBR Green. The fluorescence emission of the unbound dye is less than intercalated dye within the dsDNA. Some changes occur in SYBR structure due to intercalation that leads to more fluorescence (Figure 7A).

Procedure

SYBR Green® RT-qPCR was performed using high-quality cDNA samples (Ct $\leq$ 25) and specific primers listed in Table 4. Briefly, expression of the stemness related genes, ESC-HERV-H loci, and CRC-HERV-H loci was examined from 2D and 3D selected groups of all cell lines. PCR was conducted according to the program below with slight changes in annealing temperature, cyclic numbers and duration for each primer.

- Hold stage (Initial denaturation): 95° C for 5-10 min

- PCR stage (Amplification, 40/45 cycles):
    I. Denaturation at 95 °C for 15 sec
    II. Annealing at the determined temperature for 30-60 sec
    III. Elongation at 72 °C for 10 sec (applied in HERV-H primers)
- Melt curve stage: with the default setting of the instrument

GAPDH was used an internal normalizer. Relative expression values were then calculated using $2^{-\Delta Ct}$ and $2^{-\Delta\Delta Ct}$ formulas as follows:

ΔCt = Ct (target gene) – Ct (reference gene)

ΔΔCt = ΔCt (treated sample) – ΔCt (untreated control)

$2^{-\Delta Ct}$ method reflects a gene expression level in relation to a reference gene. However, data given by $2^{-\Delta\Delta Ct}$ method is relative to a normalized calibrator and relies on the comparison between a sample and a reference control.

Validation of amplicon homogeneity

The qPCR primers and consequent synthesized products were assessed with melting curve analysis and gel electrophoresis to ensure reaction specificity, eliminates false positive signals and validate specific product length. Melting temperature alters by two factors including sequence and size of the product. Therefore, nonspecific products and primer dimers generate various peaks with unexpected melting temperature. This assay cannot be combined with TaqMan® technology and it is only applicable in dye-based qPCR.

Agarose gel electrophoresis

Agarose powder was dissolved in boiling 1x Tris-Acetate-EDTA (TAE) buffer to make a 1% agarose (w/v) gel. EtBr was added to the mild solution and cooled down to approximately 55 °C. Upon solidification, a suitable marker (50 bp, 100 bp or 1 kb DNA Ladder) was loaded in the first pocket (Figure 9). Samples were mixed with loading dye (1:5 dilution) and 10-15 µl

of the mixture was pipetted to the other wells. Electrical current was started and lastly the gel was exposed to UV light to record signal and picture DNA fragments.
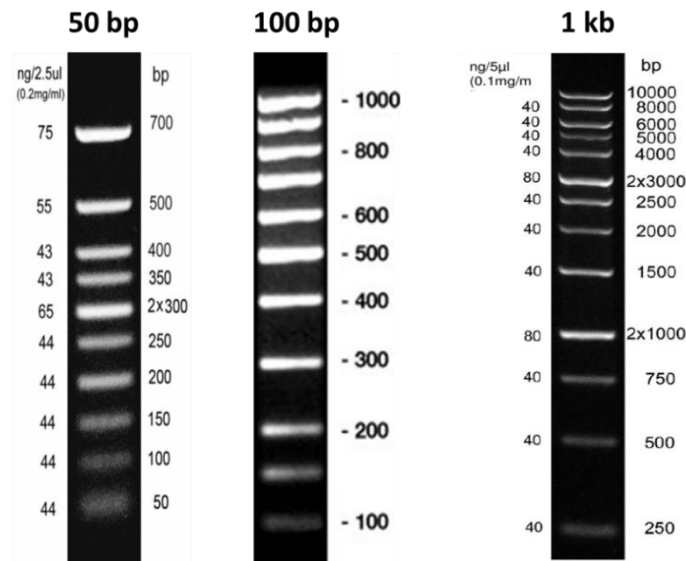
Figure 9. Variety of the DNA ladders used for analysis of the DNA fragments' size.

3.2.8. Lentiviral reprogramming

Principle

This technique is normally used to generate induced pluripotent stem cells (iPSCs) as starting point for tissue engineering. It has, however, also been used to generate CSCs-enriched population (167,168). Paradoxically, human primary cancer cells have proven resistance towards reprogramming. In spite of the significant interest in the iPSCs generation from cancer cells, there have been few studies indicating successful reprogramming of malignant human cells (168).

Preliminary steps for iPSCs establishment

Two steps had to be optimized before starting this line of experiments, including the plating density of target cells and inactivation of the MEF cells (SNL76/7). This cell line is used as a feeder layer to support stem cell growth.

For finding the optimal density, cell numbers ranging from 1×10^4 to 1×10^5 cells were seeded per well of a 6-well plate in 3 ml culture medium. The optimal plating density is the number of cells plated at day 0, reaching 90-95% confluency by day 6. It varies depending on the cell size,

morphology, and proliferation rate. The optimal number was 4×10^4 cells for HROC278Met T2 M2, HROC309, HROC24, and HROBMC01 cells, determined in preliminary analysis.

Feeder cells must be mitotically inactivated to prevent MEF over-growth. Inactivation can be realized by either gamma irradiation or mitomycin-C treatment (169). While the former induces DNA strand breaks, mitomycin-C inhibits nuclear division by DNA cross-linking (170). Applying mitomycin-C seems to be an affordable, convenient, and routine protocol to prepare feeder cells (169). Cells were incubated with 10 µg/ml mitomycin-C for two hours at 37 °C and washed with 1x PBS. After splitting, cells were exposed to gamma irradiation for half an hour (0.68 Gy/min) to ensure complete inactivation.

iPSCs establishment by using EMD Millipore's Human STEMCCA lentivirus reprogramming kit

The kit contains Polybrene® transfection reagent and a polycistronic lentivirus vector (Figure 10) engineered for multicistronic gene expression of the human transcription factors OCT4, KLF4, SOX2, and c-MYC. Theoretically, this single polycistronic cassette can improve the efficiency and produces homogeneous reprogrammed population.
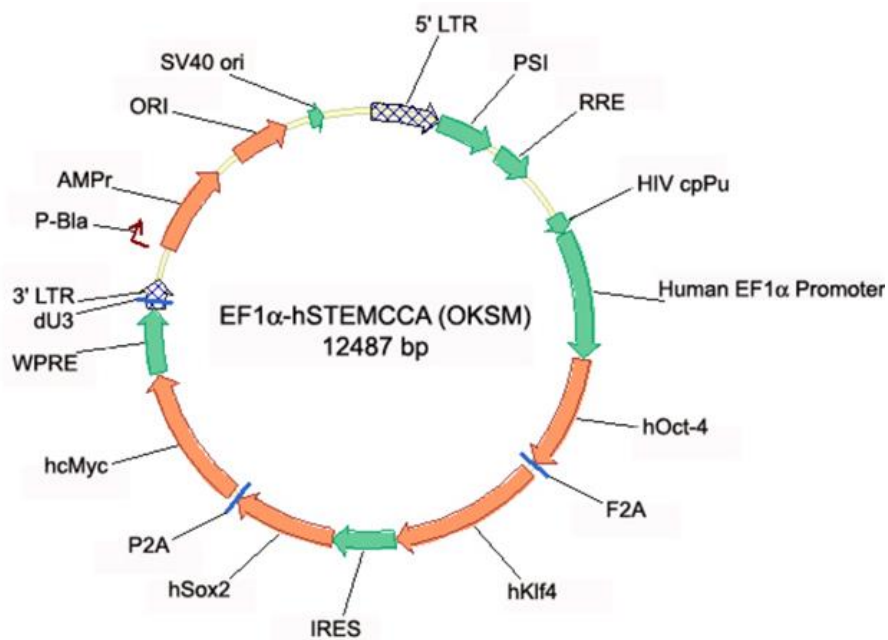
Figure 10. Schematic overview on the EF1α-hSTEMCCA lentiviral vector.

Procedure

Based on the calculated plating density, 4×10^4 HROC24 cells were seeded in 3 wells of a 6 well-plate containing 3 ml of DMEM culture medium. Before transduction, number of cells/well were counted in order to calculate the volume of virus required to achieve a MOI of 20-50 according to the below given formula. Polybrene transfection reagent was diluted 1:10 in dH_2O (final concentration 5 μg/ml) and added to each well, followed by addition of EF1α – hSTEMCCA Lentivirus.

$$\text{Virus Volume (μL)} = \frac{\text{Number of cells seeded (from step 1)}}{\text{Virus titer}} \times \frac{\text{Desired MOI}}{1 \text{ ml}} \times 1000 \text{μL}$$

Infection was repeated on 2nd day following by medium change until the 5th day. Alongside, inactivated MEF cells (SNL76/7) were prepared. For this step, each well of a fresh 6-well plate was coated with 2 ml of 0.1% gelatin coating solution followed by 30 minutes incubation at 37 °C. Gelatin solution was aspirated from each well exactly before seeding the 1.5×10^5 inactivated MEF cells. Normal MEF medium was used to culture the cells and plates were incubated overnight in a 37 °C, 5% CO_2 incubator. The day after, the wells containing virus-infected cells were harvested and a portion of cells was collected for gDNA isolation to check whether integration into the genome was successful. The remaining infected cells were counted and 2×10^4 cells re-cultured on prepared inactivated MEF cells. Infected cells were diluted in ESCs medium containing DMEM/F12, 20% KnockoutTM Serum Replacement, NEAAs, β-ME, and 10 ng/ml bFGF per well and seeded in coated wells with feeder cells. Cell morphology was monitored daily for almost two weeks, medium was also changed every other day and fresh inactivated MEF cells were added every week to replenish older MEF during the reprogramming. Finally, on day 25, desired colonies with sharp boarders and packed cells were picked and each colony was seeded into a well of a 48 well-plate. Colony growth was controlled every day and after 7-10 days, cells were expanded. Cells were regularly harvested and stocks were frozen viable as backup and native for molecular analysis.

To screen the integration of hSTEMCCA vector in the gDNA of the transduced cells and in its subsequent colonies, gDNA was isolated and processed as explained in 3.2.4.1. Next, 25 ng of gDNA was applied in PCR using the following primers and conditions (171):

cMYC F5′-GGAACTCTTGTGCGTAAGTCGATAG-3′

WPRE R5′-GGAGGCGGCCCAAAGGGAGATCCG-3′

Size of the PCR products were controlled by gel electrophoresis and 1 kb ladder used as a DNA marker.

Table 11. Thermal cycler condition.

Step	Temperature	Time	Number of Cycles
Initial denaturation	95 °C	3 min	1
Amplification:			
I. Denaturation	94 °C	30 sec	
II. Annealing	60 °C	30 sec	33
III. Extension	72 °C	1 min	
Elongation	72 °C	5 min	1

3.2.9. Statistical analysis

The statistical significance of differences between groups was assessed by unpaired two-tailed t test using GraphPad Prism 5 software. All values are reported as mean ± standard error of the mean (SEM). A p-value of < 0.05 was considered significant.

Spearman r correlation was done using R package version 0.9-7 (2008). Two following packages were used to draw plots; PreformanceAnalytics for chart correlation (172) and ggcorrplot for correlogram [https://github.com/kassambara/ggcorrplot]. Each significance level is associated with a symbol as follows: p-values (0.001, 0.01, 0.05, 0.1, 1) <=> symbols ("***", "**", "*", ".", " ") [http://www.sthda.com/english/wiki/correlation-matrix-a-quick-start-guide-to-analyze-format-and-visualize-a-correlation-matrix-using-r-software]

4. Results

In this thesis, the association of HERV-H sequences and pluripotency genes was examined. This was conducted by applying multiple cell biology and molecular techniques including monolayer and spheroid culture, RNA and DNA extraction, qPCR analysis, and sorting. The results are outlined below.

4.1. Analysis of stemness factors and HERV-H loci expression in primary CRC cell lines

A large cohort of low-passage CRC cell lines (n=56) was included in the pilot experiment. Here, the expression of stemness related genes (*LGR5*, *NANOG*, *OCT4*, *SOX2*, *KLF4*, and *c-MYC*), seven HERV-H loci (131), and four ESCs/pluripotency-relevant loci was examined by RT-qPCR. This experiment gave an overview on the loci's expression in CRC cell lines.
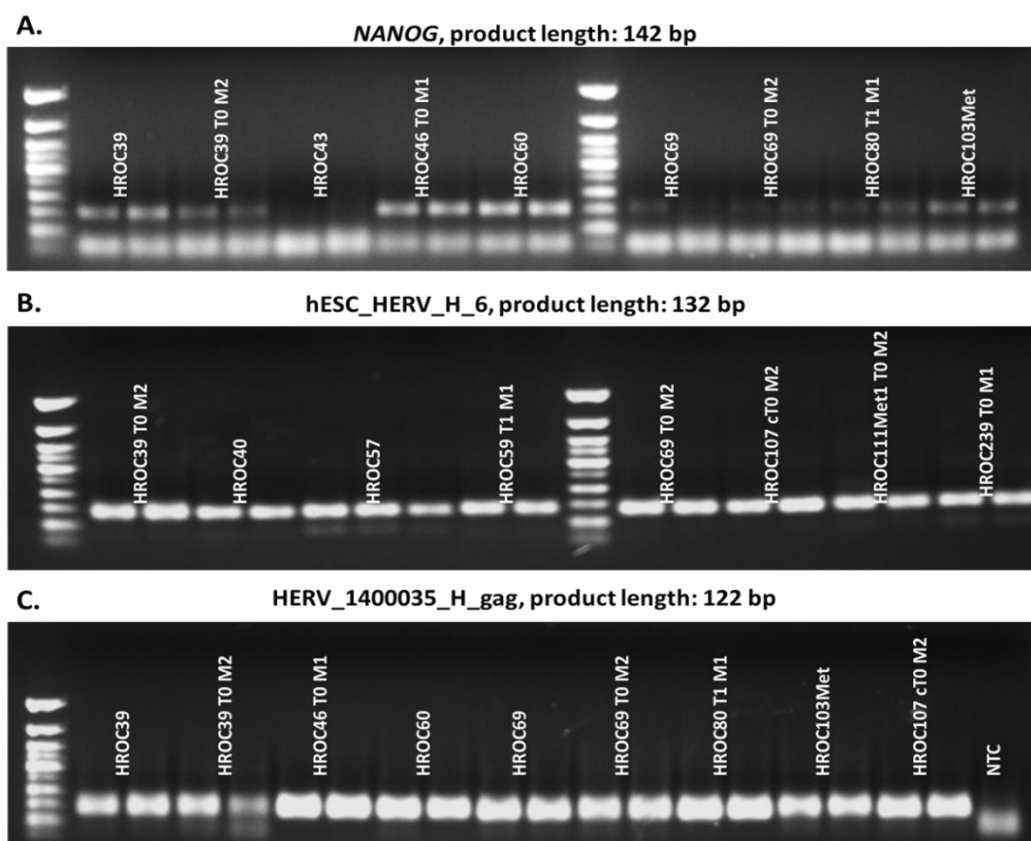
Figure 11. Agarose gel electrophoresis of qPCR products.

These pictures provide information about the size of the PCR products in, A) NANOG (a stemness gene), B) an ESC-HERV-H locus, and C) a CRC-HERV-H locus in exemplary HROC cell lines.
A 50 bp DNA ladder was used to determine product size. All samples were run in duplicates.

The melting curve analysis and the agarose gel electrophoresis showed the primers' specificities and confirmed the size of the products. Representative results are displayed in Figure 11.

According to the data obtained, CRC cell lines expressed stemness related genes at low level (Figure 12). Among them, *c-MYC* had the highest expression level that ranged from 3×10^{-2} to 60.5×10^{-1} and *NANOG* showed the lowest relative expression, which ranged from 5×10^{-5} to 1.9×10^{-2}. With regard to the melt curve analysis and gel electrophoresis of *LGR5* products, a few cell lines produced two melt curves and two products (*data not shown here*). To find out whether a new *LGR5* transcript was particularly expressed in CRC cells, PCR products were extracted from the gel and sequenced. Sequencing data revealed an unspecific product for *LGR5* primer. Consequently, this gene was not addressed further.
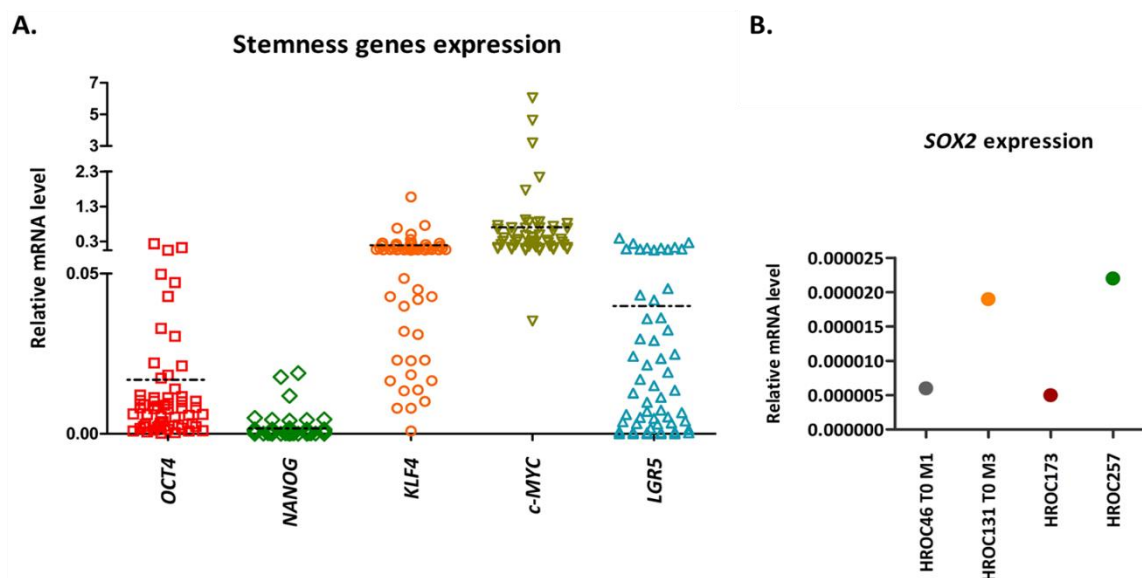
Figure 12. Relative expression level of stemness related genes in 56 cell lines.

The scatter dot plots represent the first screening qPCR results of A) stemness genes including OCT4, NANOG, KLF4, c-MYC and LGR5, and B) SOX2 in CRC cell lines. Expression levels were normalized to GAPDH and the expression rates are presented as $2^{-\Delta Ct}$ values. Each spot indicates the result of one cell line analyzed. Mean value of the data is shown by a line for each gene.

SOX2 expression was confined to four cell lines (HROC46 T0 M1, HROC131 T0 M3, HROC173, and HROC257), while it was not detectable in the remaining 52 cases (Figure 12B). To validate qPCR results, all samples were controlled on the gel to exclude distortion by primer dimer. This analysis confirmed the low or missing *SOX2* expression in most cell lines. Thus, this gene was excluded from further experiments.

46

Analysis of all examined HERV-H loci revealed that CRC-HERV-H loci (Figure 13A) and ESC-HERV-H loci (Figure 13B) were detected in almost all examined samples. However, CRC specific loci had higher mRNA level than ESC loci (Figure 13). Generally, 500502_h_L3U3 and X00041_h_gag were the most highly expressed CRC-HERV-H loci followed by loci on chromosome 13 (1300360_h), chromosome 20 (2000045_h), and 1400035_h_gag. Expression level of two genes (gag/env on chromosome 13 and gag/pol on chromosome 20) from the same locus was nearly similar in each cell line (Figure 13A). In ESC-HERV-H loci, a locus on chromosome 19 indicated highest expression with the maximum level of 28×10^{-2}, while expression of the other ESC loci was in the range of 8×10^{-2} to 1×10^{-5} (Figure 13B).
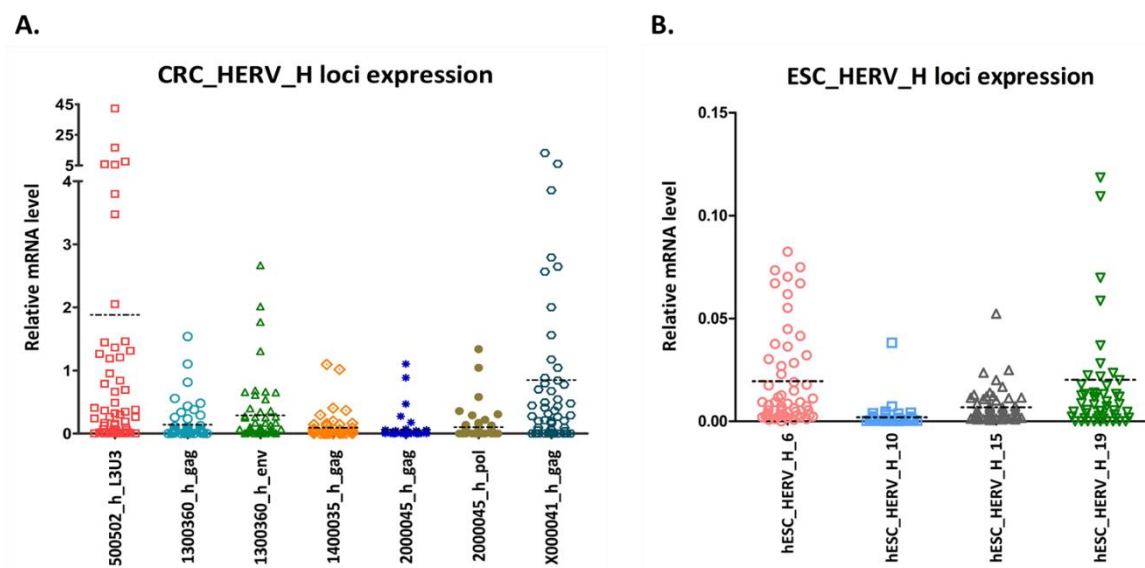

Figure 13. Relative expression level of HERV-H loci in 56 cell lines.

The scatter dot plots represent the first screening qPCR results of A) CRC-HERV-H loci, and B) ESC-HERV-H loci in CRC cell lines. Expression levels were normalized to GAPDH and the expression rates are presented as $2^{-\Delta Ct}$ values. Each spot indicates the result of one cell line analyzed. Mean value of the data is shown by a line for each gene.

4.1.1. Correlation between stemness related genes and HERV-H loci

Preliminary statistical analysis of correlation coefficient was done in R program without assessing normal distribution. Here, correlation was computed by Spearman r to examine any possible associations between HERV-H loci and stemness genes. In the first approach, all samples were applied to the test regardless of molecular subtypes, TNM, and grading

parameters. Moreover, correlation of hESC_HERV_H_19 was determined separately because expression of this locus was lost in 14 cell lines.

The r value indicated either a moderate ($0.4 < r < 0.7$) or a strong ($r \geq 0.7$) positive correlation among CRC-HERV-H loci (Figure 14). Additionally, there was a moderate positive relationship between two loci of ESC-HERV-H located on chromosomes 6 and 10 with almost all of CRC-HERV-H loci. However, stemness indices weakly correlated with some HERV-H loci. Associations were observed between *OCT4* and 1400035_h_gag ($r = 0.36$, $p = 0.003$) and hESC_HERV_H_10 ($r = 0.44$, $p = 0.0003$). The Spearman r values in *NANOG* expression revealed weak positive correlations with all ESC-HERV-H loci. *NANOG* also correlated with 1400035_h_gag ($r = 0.32$, $p = 0.008$). Interestingly, 1400035_h_gag and *c-MYC* had a weak to moderate positive and significant correlations ($p < 0.05$) with all of the examined loci (Figure 14).
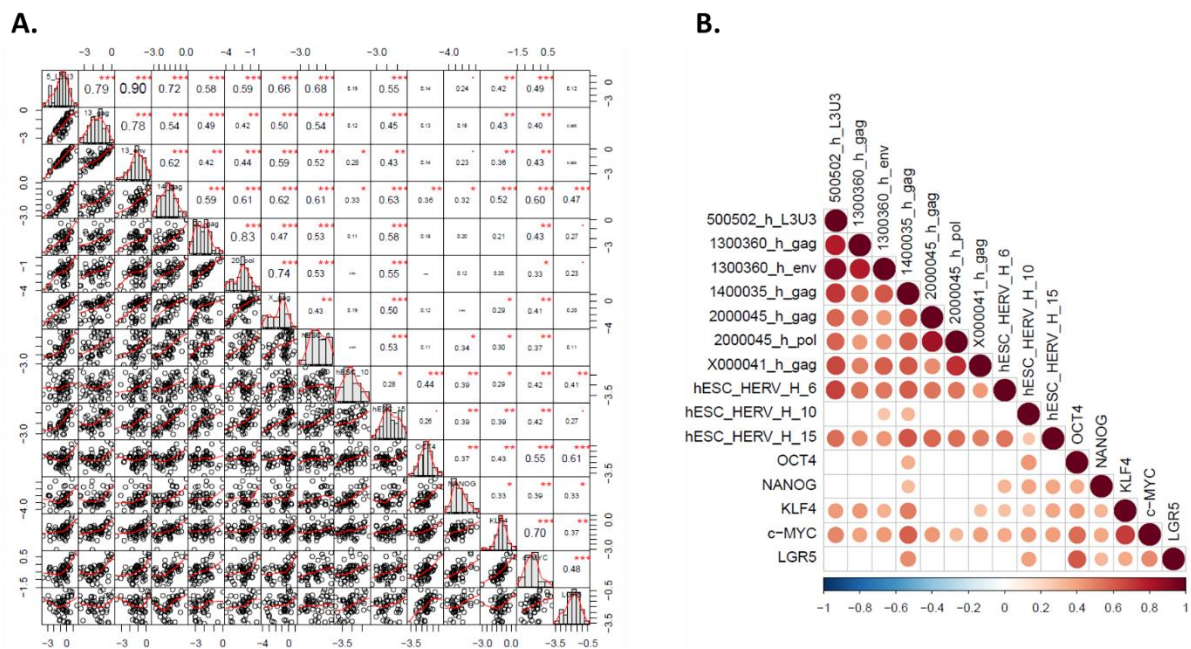
Figure 14. Chart correlation and correlogram of gene expression data in CRC cell lines.

A) The chart indicates scatter plots, distributions, and Spearman r correlation coefficients among variables sorted as follows; 5_L3U3 (500502_h_L3U3), 13_gag (1300360_h_gag), 13_env (1300360_h_env), 14_gag (1400035_h_gag), 20_gag (2000045_h_gag), 20_pol (2000045_h_pol), X_gag (X000041_h_gag), hESC_6 (hESC_HERV_H_6), hESC_10 (hESC_HERV_H_10), hESC_15 (hESC_HERV_H_15), OCT4, NANOG, KLF4, c-MYC and LGR5. Font size of R value and stars are comparative to the correlation' strength. (* p < 0.05; ** p < 0.01; *** p < 0.001). B) This diagonal panel shade visualizes the significant correlation among genes of interest. Insignificant r values are not shown here. The magnitude of the correlation is determined by the depth of the shading and size of the circles.

Furthermore, Spearman analysis in locus 19 revealed a moderate to strong correlation with the other HERV-H loci (*plot not shown here*). Nevertheless, no significant correlations were seen with *OCT4* and *NANOG*.

To find out whether gene expression pattern correlates with specific clinico-pathological characteristics, analysis was refined by including the following parameters: molecular subtypes (MSI, MSS, and CIMP), T (T2/T3 and T4), N (N0 and N1/N2), M (M0, M1), and grading (G2 and G3) (Figure 15).

Generally, the heterogeneity in expression pattern was preserved in the molecular subtypes, without a clear trend. The same was seen for TNM and grading.
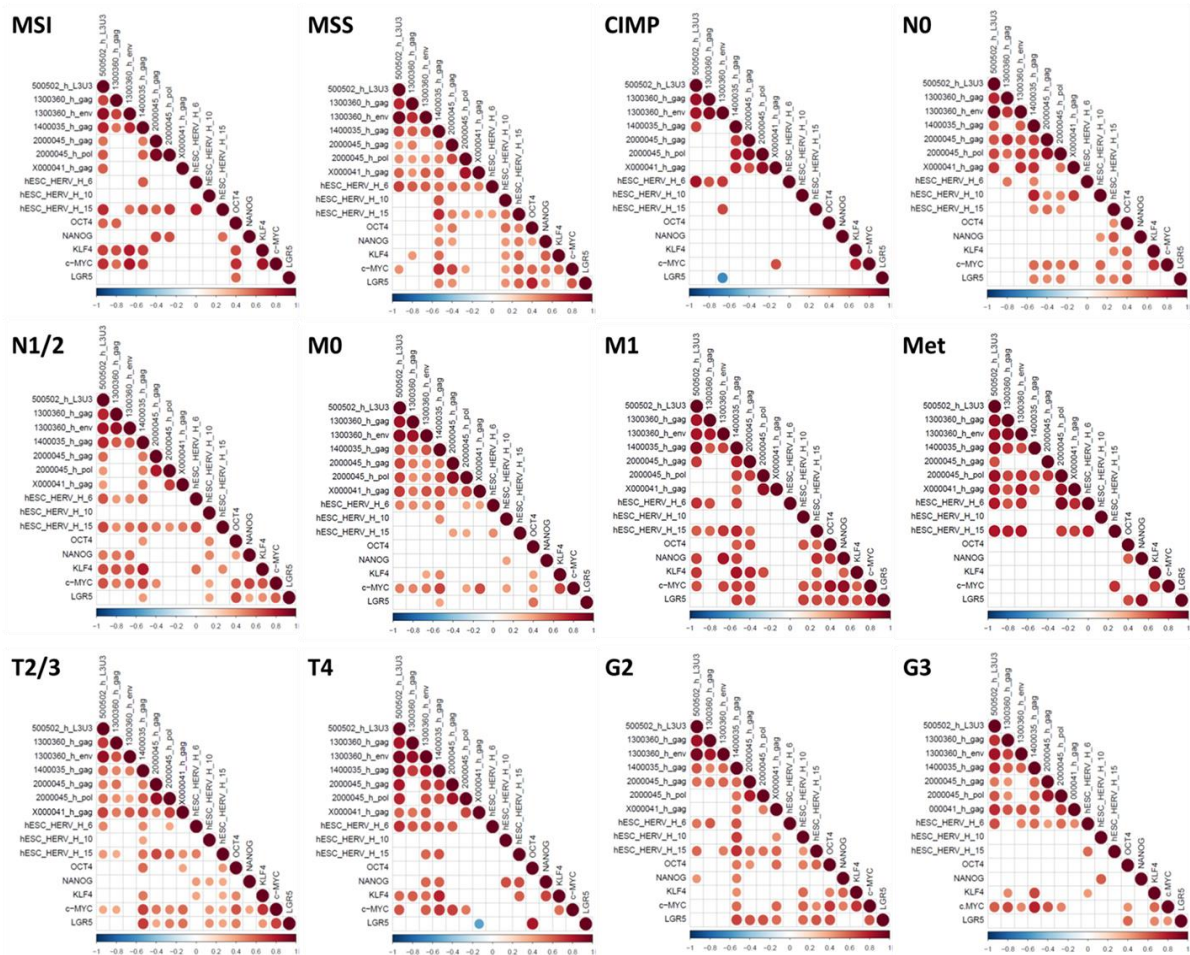
Figure 15. Correlation confidence in classified CRC lines.

Spearman r correlation was used to compare correlations between level of HERV-H expression and stemness related genes in CRC cell lines based on the molecular and clinical features. Insignificant r values are not shown in the panel. The magnitude of the correlation is determined by the depth of the shading and size of the circles.

In MSS cell lines, ESC_HERV_H_10 and 1400035_h_gag correlated significantly positive with all stemness genes. Remarkably, correlation of ESC_HERV_H_10 was completely vanished in MSI cell lines. Likewise, there was no correlation of 1400035_h_gag with *OCT4* and *NANOG* in MSI samples. Comparison of nodal involvement (N0 and N1/2) revealed correlation between HERV-H loci and stemness gene expression. Accordingly, correlations of 500502_L3U3 and both genes of locus 13 -found in N1 or N2 samples- with *NANOG*, *KLF4*, and *c-MYC* were not detected in N0 tumors. The correlations of stemness genes (in particular *OCT4* and *NANOG*) with HERV-H loci observed in M1, T4, and G3 tumors were missed or reversed in the corresponding lower stages (M0, T2/3, and G2). To support these findings, metastasis-derived cell lines were compared with the matched primary-derived cell line (Figure 15). Here, no correlation between HERV-H and stemness loci was seen. Still, correlations of 1400035 locus against *OCT4* and *NANOG* were more significant in the advanced tumors compared to the lower grade tumors (Figure 15).

Correlation of hESC_HERV_H_19 with the other HERV-H loci and stemness genes was separately performed (Figure 16). Spearman r values in tumors with different molecular types showed no significant correlation of this locus with *OCT4* and *NANOG*. Comparison based on the lymph node involvement (N0 *vs.* N1/N2) revealed no significant relationship of HERV_H_19 expression level with *OCT4* and *NANOG*. Still, significant associations of locus 19 were observed with *KLF4* ($r = 0.66$, $p < 0.01$) and *c-MYC* ($r = 0.57$, $p < 0.05$) in CRC cell lines with higher N stage. Comparable associations were seen for metastasized tumors. The clinic-pathological M1 status revealed a strong positive correlation between locus 19 and *NANOG* ($r = 0.84$, $p < 0.01$). Spearman r values of HERV_H_19 and other stemness genes increased clearly ($0.63 \le r \le 0.67$) in M1 tumors compared to M0 tumors with small r values ($r < 0.3$). Similar to the M0 samples, no correlation of the locus 19 with the pluripotent factors was found in secondary tumors. Contrary to the M status, correlation with *OCT4* was significantly higher in tumors with lower T stage ($r = 0.46$, $p < 0.05$). *Vice versa*, HERV_H_19 correlation with *KLF4* ($r = 0.7$, $p < 0.05$) in T4 tumors was not present in T2/3 tumors. Nonetheless, the correlation with *c-MYC* in T4 ($r = 0.64$, $p < 0.05$) and T2/3 ($r = 0.54$, $p < 0.05$) CRC was almost the same. Finally, stemness expression (*OCT4* and *NANOG*) did not correlate with grading (Figure 16).
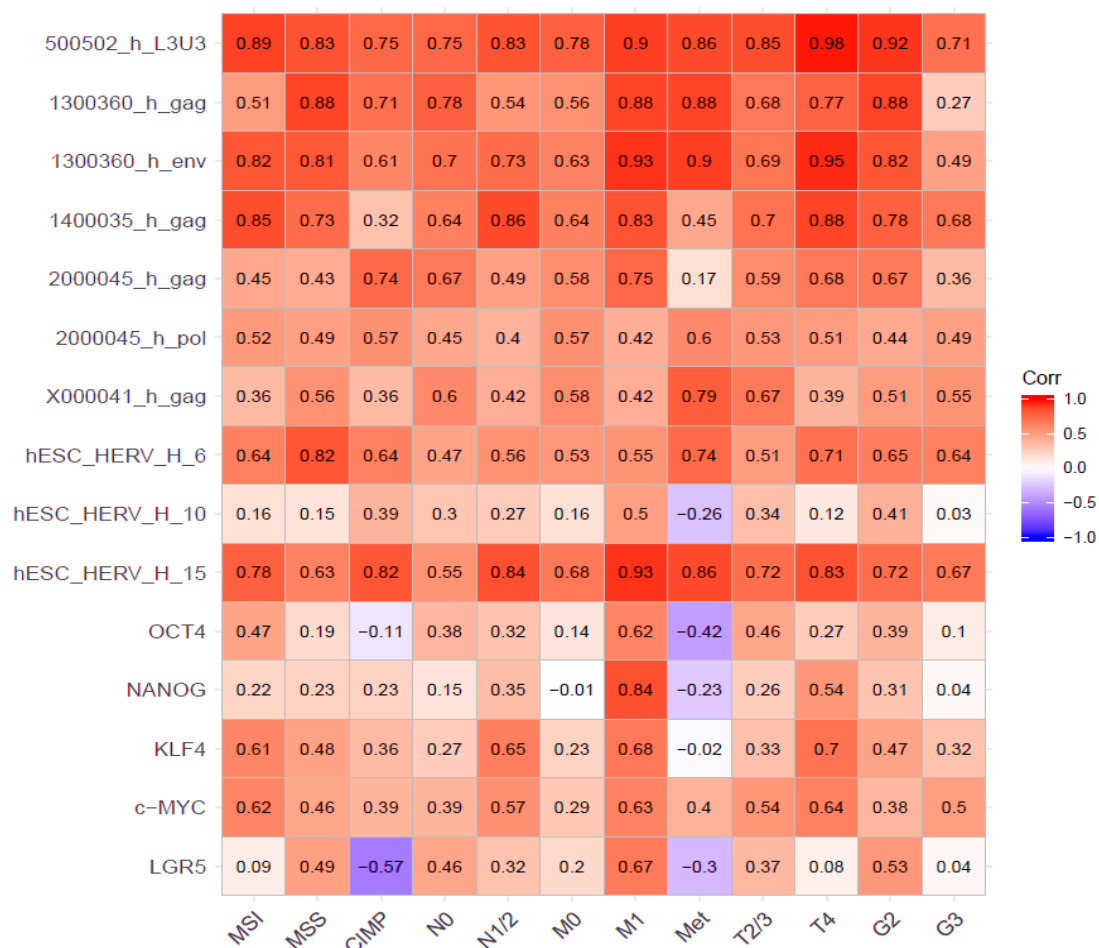
Figure 16. Correlation confidence of HERV_H_19 in classified colorectal cancer lines.

The quadrants and colors represent the Spearman r correlation of the HERV-H loci with pluripotency factors in categorized CRC tumors considering molecular types, TNM stage, and tumor grading (G).

4.1.2. Cell lines selection

Based on the expression levels of CRC-HERV-H loci, selected cell lines were classified as being HERV-H high, intermediate or low. Accordingly, each group consisted of eight cell lines (Figure 17A). Thereafter, qPCR of CRC-HERV-H expression was repeated in these pre-selected cell lines (Figure 17B) in order to confirm the accuracy in samples' selection. Comparison of qPCR data in both datasets indicated deviations in amount of the RNA expression within a same cell line in two different samples. However, the general trend almost remained in each category. Finally, five cell lines per group were chosen for further analysis (Figure 17C).

A.

High	Intermediate	Low
HROC18 (MSS, spSTD)	HROC24 T1 M1 (spMSI-H)	HROC24 (spMSI-H)
HROC46 T0 M1 (MSS, spSTD)	HROC40 (MSS, CIMP-H)	HROC59 T1 M1 (MSS, spSTD)
HROC87 T0 M2 (spMSI-H)	HROC43 (MSS, CIMP-H)	HROC239 T0 M1 (MSS)
HROC131 T0 M3 (spMSI-H)	HROC50 T1 M5 (spMSI-H)	HROC278Met T2 M2 (MSI-L, CIMP-H)
HROC147 T0 M1 (MSI, CIMP-L)	HROC147Met (MSI-L)	HROC296
HROC173 (MSS, spSTD)	HROC278 T0 M1 (MSI-L, CIMP-H)	HROC309
HROC257 T0 M1 (spMSI-H)	HROC285 T0 M2 (MSI-H, HNPCC)	HROC364
HROC313Met1 T0 M2 (MSS)	HROC324 (MSI-H, HNPCC)	HROBMC01 (MSS, spSTD)

B.
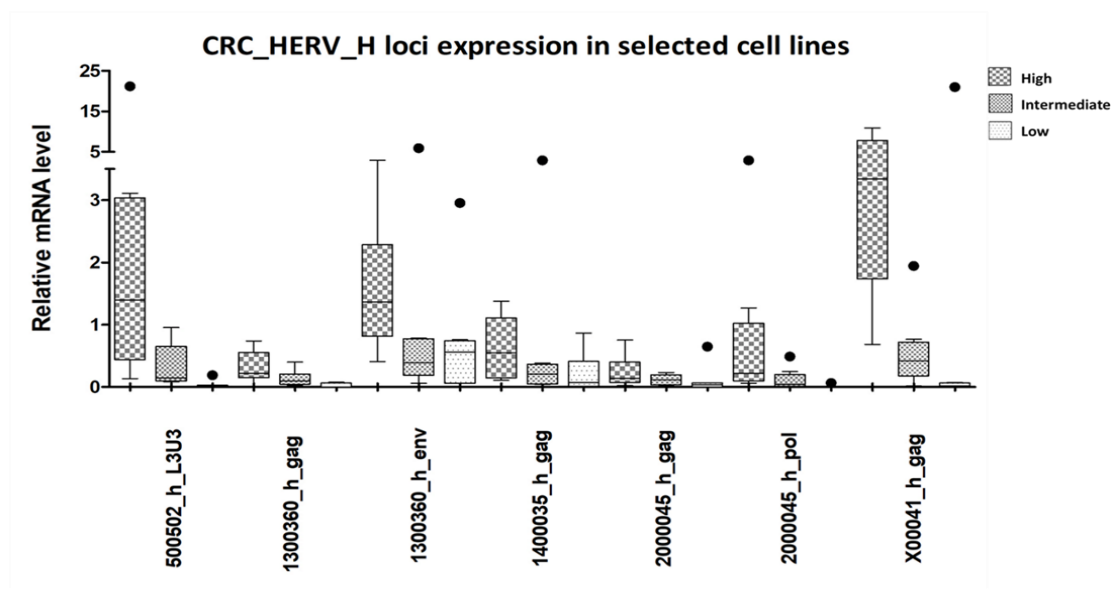
C.

High	Intermediate	Low
HROC18 (MSS, spSTD)	HROC40 (MSS, CIMP-H)	HROC24 (spMSI-H)
HROC46 T0 M1 (MSS, spSTD)	HROC50 T1 M5 (spMSI-H)	HROC278Met T2 M2 (MSI-L, CIMP-H)
HROC87 T0 M2 (spMSI-H)	HROC278 T0 M1 (MSI-L, CIMP-H)	HROC309
HROC147 T0 M1 (MSI, CIMP-L)	HROC285 T0 M2 (MSI-H, HNPCC)	HROC364
HROC313Met1 T0 M2 (MSS)	HROC324 (MSI-H, HNPCC)	HROBMC01 (MSS, spSTD)

Figure 17. Selected CRC cell lines based on the gene expression analysis.

A) The table shows a summary of CRC cell lines and the corresponding molecular subtype in each group. B) Relative expression levels of CRC-HERV-H sequences in pre-selected cell lines depicted in boxplots. Median value is shown by a line in each box ± standard deviation; outliers are identified as dots. C) The table indicates the final selection of five CRC cell lines categorized into one of the three groups: high, intermediate, or low.

Moreover, two new cell lines were added to the panel (HROC300 T2 M1 and HROC383) due to the interesting features. HROC300 T2 M1 expressed high level of most HERV-H loci and stemness genes, whereas HROC383 grew spontaneously in suspension which made it attractive for subsequent sphere culture experiments.

4.2. CSCs enrichment

In order to enrich CSCs, two methods were used: i) the formation of cellular spheres and ii) sorting SP after staining with Hoechst 33342 using the FACS Aria II cell sorter.

4.2.1. Spheroid culture

To establish this method, cell culture flasks (25 cm^2) were coated with 1% agar to provide the non-adherent condition for 3D cell growth. HROC cell lines (HROC24 T1 M1 and HROC50 T1 M5) were then seeded at a density of 5×10^5 cells into the pre-coated flasks using serum-free medium supplemented with 20 ng/ml EGF. EGF was added and replaced every other day to promote cellular dedifferentiation. A few small floating cell clusters or aggregates were observed after three days which did not grow in size over the next days. Shrinkage of the cell and condensed nucleus or pyknosis were visible by light microscopy. Despite that, spheres were collected and RNA isolation failed to produce satisfactory RNA quality.

In the next step, the medium was exchanged with cancer stem cell medium (PromoCell) to provide optimal conditions for growing spheres [http://www.promocell.com/products/cell-culture-media/]. Besides, cell culture flasks were replaced with repellent cell culture flasks. Accordingly, HROC cell lines showed an increased ability to form viable 3D shapes and results of the RNA extraction were promising. Nevertheless, this medium was rather expensive and was not the best choice for all cell lines. For example, the spheres of few cell lines such as HROC24 and HROC383 were mainly cell clusters and not actual spheres. Therefore, RNA extraction was not successful and yielded a very low quantity and quality.

Sphere formation results confirmed the cells' ability to generate 3D structures, but with distinct morphologies (Figure 18). These spheroids were classified as grape-like clusters (Figure 18A), solid spheres (Figure 18B), and hollow or organoids-like spheres (Figure 18C). Grape-like spheres were irregular in shape and similar in cell mass with loose connections. Solid spheres were characterized by a round and a regular colony shape, tightly packed with small cells, and

a smooth rim. The latest group was recognized by combination of solid colonies and hollow regions.

Considering spheroids' morphology, it was hard to discriminate types of the spheres in cell lines including HROC18, HROC50 T1 M5, HROC300 T2 M1, and HROC313Met1 T0 M2. As shown in Figure 18, the sphere of HROC18 looked different compared to other cell lines placed in this group. Moreover, in a few cell lines, a mixture of various spheroids was observed. For example, solid colonies alongside with hollow shapes were detected in HROC46 T0 M1 and HROC313Met1 T0 M2. In HROC46 T0 M1, solid shapes were in majority, but in the other line organoid-like structures were mainly formed. Additionally, solid and grape like clusters were shaped in HROC50 T1 M5. Another exception was recognized in HROC300 T2 M1. This
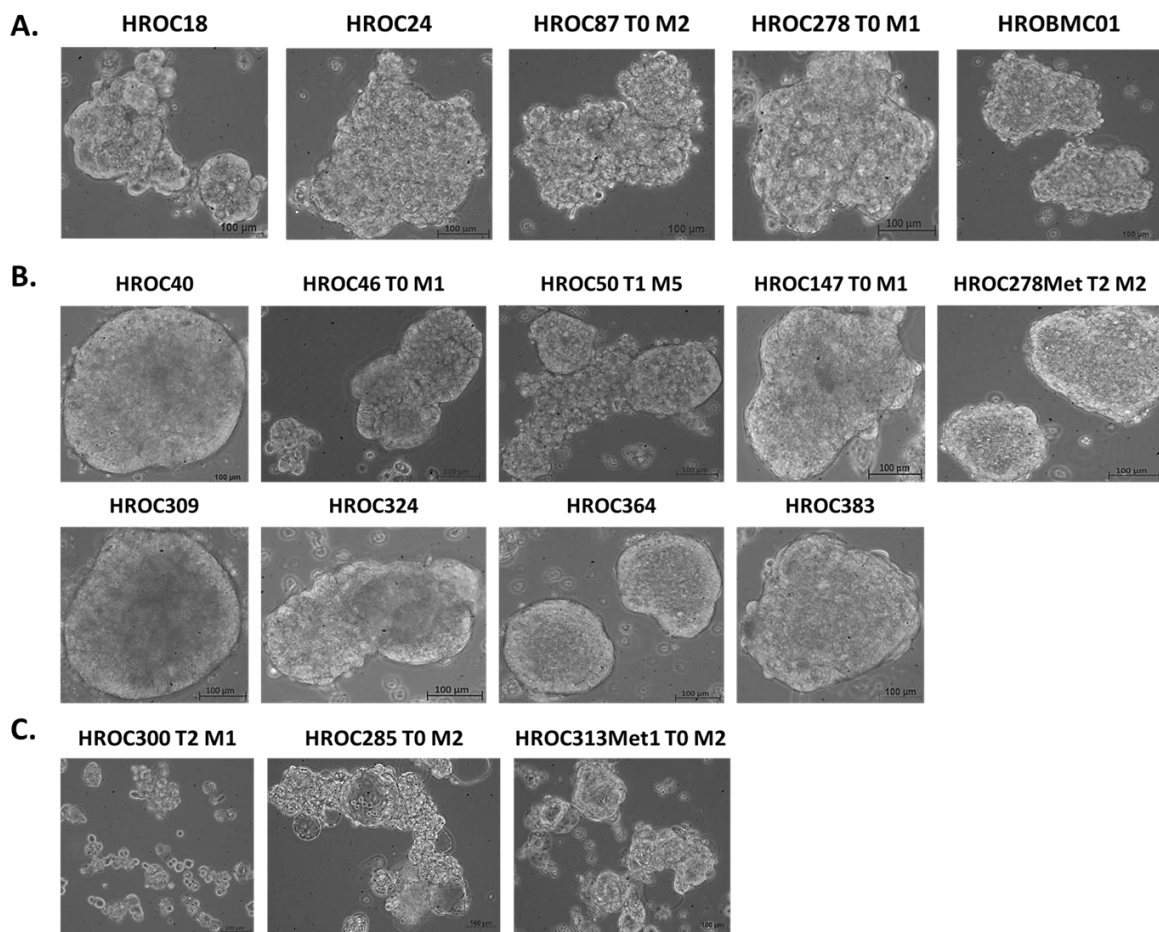
Figure 18. Light microscopy images of distinct spheres' types in 17 CRC lines cultivated in suspension condition.

A) Grape-like colonies appear with irregular shapes and dense central cells surrounded by visible cells. B) Solid spheres were formed in half of the cell lines, were mainly big and difficult to dissociate. C) Organoid-like spheres were formed in a few cell lines and were similar to vacant cavities around other colonies. Original magnification x100; scale bars represent 100µm.

cell line grew slowly and had big cells (Figure 18C). In 3D culture, growth rate was almost similar to 2D and colonospheres growth rate was also slow. As a result, in the day of collection they were definitely smaller than other spheres formed in the examined cell lines.

To investigate the influence of the 3D culture on gene expression, RNA was extracted from all 2D and 3D samples. After controlling RNA quality by the *GAPDH* probe system, RNA was reversely transcribed to cDNA and expression of stemness related genes and CRC_HERV_H loci in spheroids were compared to adherent 2D-cultured maternal cells.

4.2.1.1. Gene expression analysis in colonospheres

<u>Stemness gene</u>

Figures below depict mRNA expression levels of *OCT4*, *NANOG*, *KLF4,* and *c-MYC* in colonospheres that have been normalized to their related 2D groups (Figure 19). According to the data, no consistent pattern was found in *OCT4* (Figure 19A) or *c-MYC* expression (Figure 19E). Some of the alterations observed in spheroids were significant ($p < 0.05$). Furthermore, no noteworthy correlations of *OCT4* and *c-MYC* were found between molecular subtypes and up or down regulation of these genes. In spite of the former genes, *KLF4* expression dropped in all spheroids, although this did not reach statistical significance. Moreover, significations were not correlated to the molecular subtypes. Among analyzed stemness genes, *NANOG* was strongly overexpressed in 15 cell lines grown in 3D condition with minimum enhancement in HROC24 (1.99-fold) and maximum growth (11.42-fold) in HROC87 T0 M2. A similar rising trend was also observed in spheroids of HROC147 T0 M1 ($p = 0.058$) and HROC300 T2 M1 ($p = 0.054$) without meaningful differences compared to the maternal cells.

Figure 19. Comparison of relative stemness genes expression in colonospheres and the corresponding adherent cells.

The graphs show the qPCR results of the stemness genes in 3D culture of the selected cancer cell lines. A) OCT4, B) NANOG, C) KLF4, and D) c-MYC. Expression levels were firstly normalized to GAPDH and the expression rates were calculated by $2^{-\Delta\Delta Ct}$.. The values were transformed to log 2 and 2D culture set as zero. The data are presented as the mean values ± SEM. (*P < 0.05; **P < 0.005; ***P < 0.0001)

CRC-HERV-H loci

At a first glance, expression of HERV-H loci in most of the spheres seemed higher than corresponding adherent cells (Figure 20). However, the upregulations were not consistent. A diverse pattern of the 500502_h_L3U3 expression appeared in all groups (Figure 20A) regardless of any categories such as molecular subtypes, spheroids' morphology, and TNM grading. On the other hand, remarkable overexpression ($p < 0.05$) of 1400035_h_gag was detected in 12 spheroids groups (70% of samples). Among them, expression level of the gag locus on chromosome 14 in HROBMC01 spheres was 4.7-fold more than the matched adherent cells (Figure 20B).

Number of the spheroid groups with the significant upregulation of X00041_h_gag dropped to 8 (47%). Conversely, relative expression of this locus was lower in four spheres groups including HROC87 T0 M2, HROC147 T0 M1, HROC285 T0 M2, and HROC313Met1 T0 M2 (Figure 20C). Moreover, this downregulation was significant in HROC147 T0 M1 and HROC285 T0 M2 ($p < 0.05$).

Interestingly, the expression of two genes (*gag* and *env*) from the same locus on the chromosome 13 presented a strong discrepancy (Figure 20D). Although in spheroids, the increased expression level of *env* gene was the most obvious trend, *gag* expression level was significantly lower ($p < 0.05$) in five and tended to decrease in four spheroid groups. Furthermore, *gag* expression entirely vanished in spheroids of HROC278Met T2 M2 and HROC383. In total, 11 colonospheres had an overall decrease in *gag* expression (13000360_h_gag) compared to their adherent parental cell lines. Besides, coincident overexpression of both genes was significant ($p < 0.05$) in the spheroids of HROC46 T0 M1, HROC285 T0 M2, HROC324, HROC364, and HROBMC01.

In spite of the observed pattern in genes on chromosome 13, the expression level of *gag* and *pol* genes on chromosome 20 increased in all spheres, but reached statistical significance in 11 samples (65%). Additionally, 10 colonospheres showed remarkable expression of both loci (Figure 20E).
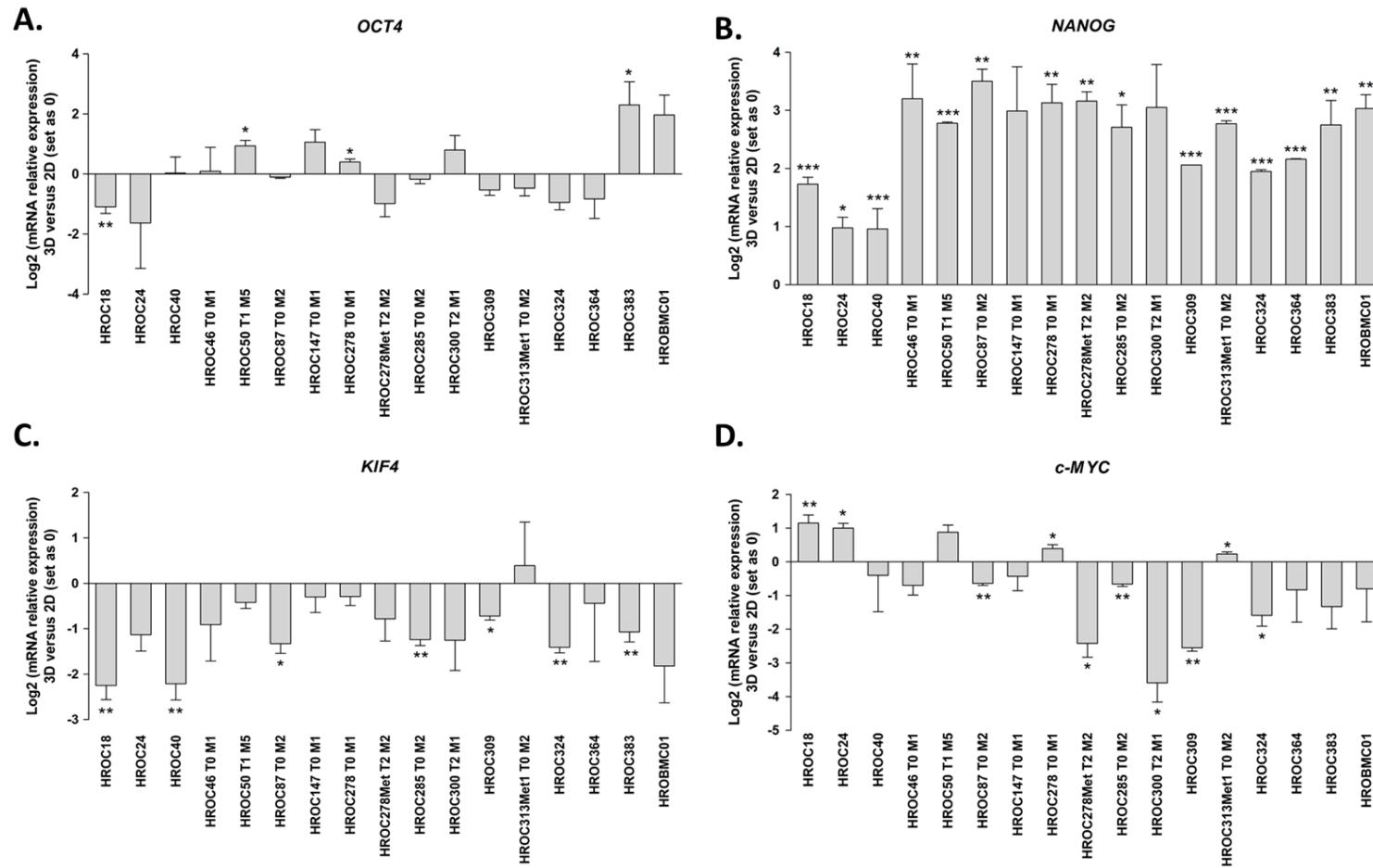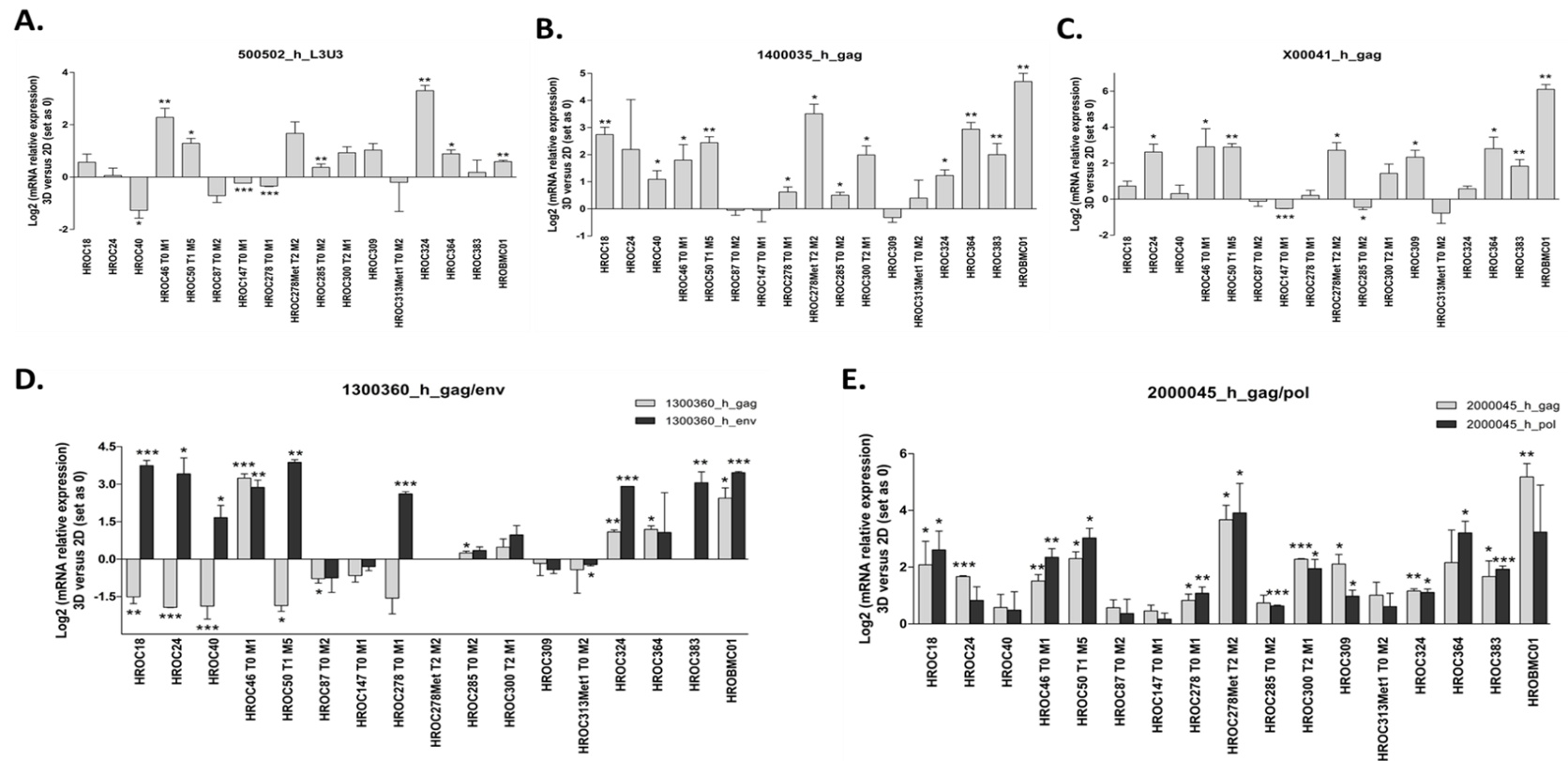
Figure 20. Comparison of relative HERV-H RNA expression in colonospheres and the corresponding adherent cells.

The graphs represent the qPCR results of the CRC-HERV-H loci in 3D culture of selected cancer cell lines. A) 500502_h_L3U3, B) 1400035_h_gag, C) X00041_h_gag, D) gag and env genes form locus 1300360, and E) gag and pol genes form locus 2000045. Expression levels were firstly normalized to GAPDH and the expression rates were calculated by $2^{-\Delta\Delta Ct}$. The values were transformed to log 2 and 2D culture set as zero. The data are presented as the mean values ± SEM. (*P < 0.05; **P < 0.005; ***P < 0.0005)

**4.2.1.2. Investigation of Spearman r correlations in spheroids**

To see correlation between expression level of stemness factors and CRC-HERV-H loci, Spearman r was first computed in all colonospheres followed by the second calculation based on parameters M and grading. This analysis revealed a negative significant link ($p < 0.05$) between *c-MYC* expression level and *gag* gene on chromosome 5 and 13. This negative trend was additionally observed in colonospheres generated from N1/2, M0, and G3 tumors. Besides, *KLF4* positively correlated with 2000045_h_gag (M0) or X00041_h_gag (M0 and G2). However, no relationship was found between *OCT4* and *NANOG* with HERV-H loci (*data not shown*).

Table 12. List of the remarkable correlations observed between pluripotent related genes and CRC-HERV-H in 17 colonospheres.

Correlated loci	Number of XY pairs	Spearman r	P value (two-tailed)	Category
c-MYC and 500502_h_L3U3	17	-0.5221	0.0316	-
c-MYC and 1300360_h_gag	17	-0.7186	0.0012	-
c-MYC and 500502_h_L3U3	7	-0.8929	0.0123	N1/2
c-MYC and 1300360_h_gag	7	-0.8929	0.0123	N1/2
KLF4 and 2000045_h_gag	9	0.6946	0.0433	M0
KLF4 and X00041_h_gag	9	0.8000	0.0138	M0
c-MYC and 1300360_h_gag	9	-0.7333	0.0311	M0
KLF4 and X00041_h_gag	7	0.8571	0.0238	G2
c-MYC and 500502_h_L3U3	7	-0.7857	0.0480	G3
c-MYC and 1300360_h_gag	7	-0.8214	0.0341	G3

4.2.1.3. Spheroid selection for assessment of HERV-H Gag (Gag-H) and NANOG protein expression

According to the mRNA expression level of *NANOG* and all examined HERV-H loci in spheroids, five cell lines with high expression were chosen for evaluating Gag-H and NANOG proteins. Expression of these proteins was measured in the adherent 2D cultures and the matched spheroids of HROC46 T0 M1, HROC50 T1 M5, HROC364, HROC383, and

HROBMC01. In these selected cell lines, at least five HERV-H loci were upregulated in the corresponding spheres.

4.2.1.4. Screening of Gag-H protein in the selected colonospheres

Expression of Gag protein was done by flow cytometry using three monoclonal mouse anti-Gag-H antibodies against the clones 14H11G1, 1B3H7, and 1D7D11 in comparison to the irrelevant antibody 2G2B3 (an anti-His antibody). Controls included cells stained with secondary antibody and cells stained with anti-2G2B3 antibody. The Gag-H antibodies are originally generated with two tags surrounding Gag-H amino acid sequences. These tags are GST at the N-terminus and His at the C-terminus which produces tag-specific signals. According to the study on 2016, His tag signal was above the background and consequently it must be tested while checking Gag-H expression (144).

Fold-change in mean fluorescence intensity (MFI) was calculated in the examined 2D and 3D samples (Figure 21) with the following formula:

Fold-change in MFI = MFI (sample)/MFI (control)
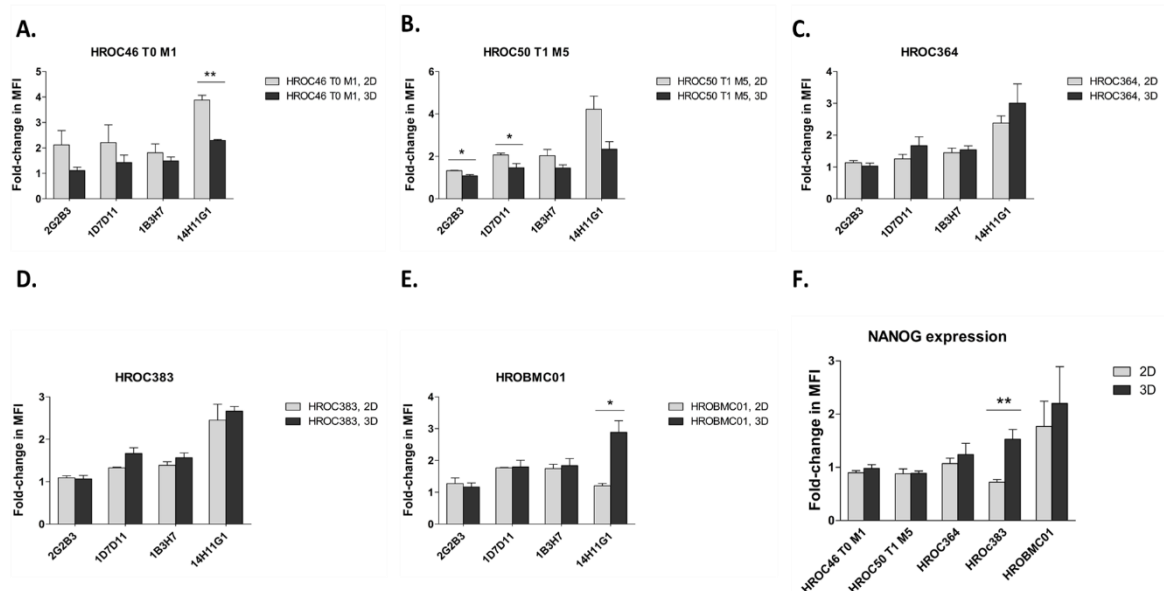
Figure 21. Gag-H and NANOG protein expression in CRC cell lines cultured in 2D and 3D.

*Bar charts exhibit and compare the **fold-change in** MFI of anti-Gag-H in the adherent and the corresponding spheroids in; A) HROC46 T0 M1, B) HROC50 T1 M5, C) HROC364, D) HROC383, and E) HROBMC01, F) NANOG protein expression. Results represent the mean of three-five independent flow cytometry experiments ± SEM. (*P < 0.05; **P < 0.005)*

In general, all five cell lines and the matched 3D spheres showed positive staining with irrelevant and Gag-H antibodies. There was a trend towards higher abundance of anti-Gag-H antigens in 2D-culture samples of HROC46 T0 M1 (Figure 21A) and HROC50 T1 M5 cells (Figure 21B). Among them, overexpression of clone 14H11G1 in HROC46 T0 M1 ($p = 0.001$) and clone 1D7D11 in HROC50 T1 M5 ($p = 0.044$) were significant. This overexpression of anti-Gag clones was detectable in the colonospheres of HROC364 (Figure 21C), HROC383 (Figure 21D), and HROBMC01 (Figure 21E), particularly with a remarkable increase of 14H11G1 in HROBMC01 spheroids ($p = 0.037$).

4.2.1.5. Screening of NANOG protein in the selected colonospheres

To further confirm the overexpression of *NANOG* detected in qPCR, a validation step was done to check whether or not the increased RNA expression of *NANOG* paralleled the respective protein abundance. For this purpose, intracellular staining was done using PE-anti-NANOG and isotype control antibodies on the five colonospheres. Except for HROC383 ($p = 0.005$), no significant difference was seen in the other four colonospheres (Figure 21F). Colonospheres of HROC364 and HROBMC01 tended to upregulate NANOG, which was not seen in HROC46 T0 M1 and HROC50 T1 M5. This finding nicely matches with flow cytometry results showing high abundance of the anti-Gag-H clone 14H11G1 in the same colonospheres.

4.2.2. Cell sorting

In the second approach to enrich cancer stem-like cells, sorting SP and non-SP fractions was done after staining with Hoechst 33342. The SP cells are able to pump the Hoechst dye out of their cytoplasm and are therefore not stained in the control tube. These cells were identified as a Hoechst^{low} (SP) cell fraction (g1) that displayed sensitivity to verapamil. Cells in g2 gate (Hoechst^{high}) were considered as very likely to be non-SP cells (Figure 22).

Cell numbers obtained from the first sorting experiments were very low, resulting in a dramatic drop in purity of the sorted fractions. Approximately, in 65% of sortings, number of events were below 1.5×10^5 (*data not shown*) and the experiment had to be repeated for each cell line. Thus, although the method was already established in the lab, it had to be optimized to isolate higher cell numbers and improve purity.
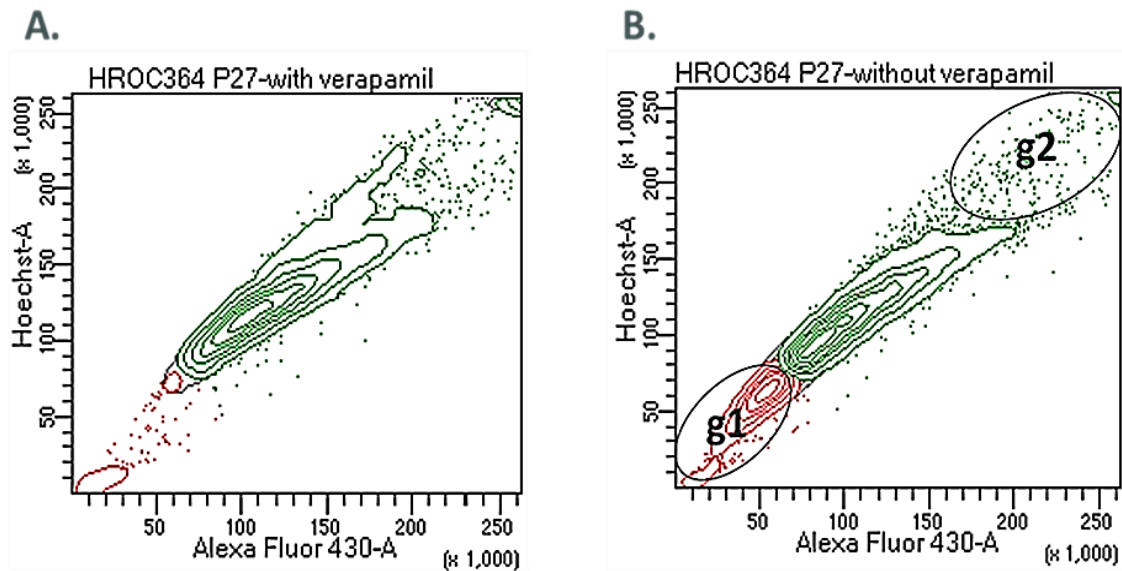
Figure 22. Gating strategy in sorting SP from HROC364 cells.

A) This diagram is an exemplary of a control tube which contains verapamil as a blocker of ABC transporter proteins. Due to the pump blockage, amount of the trapped dye increases that leads to the enhancement of signal intensity and disappearing of the SP in a control measurement. B) The density plot shows the gated area in the stained sample. g1 indicates SP and events in g2 are non-SP.

4.2.2.1. Improvement of cell sorting

To improve sorting purity the following things had to be changed: (I) as many cells as possible were used for sorting, because the SP constitutes a rare portion within a tumor (cell line) and cell loss occurs during each step of preparation as well. (II) Accurate cell sorting requires highly pure single living cells. To achieve this goal, dead cells or debris and cell aggregation need to be minimized. Thus, viability of each cell line was tested before and after staining and the sorting was performed only when viability was at least 60%. (III) Cell suspensions were passed through a 40 μm nylon mesh filter just prior to analysis to remove cell clumps. (IV) Cell concentration ($2 \times 10^6 - 6 \times 10^6$ per ml) had to be set up for some cell lines due to the coincident events[9] that could influence the sorting yield. For these cell lines, the volume used to dilute the cells for the sorting procedure had to be optimized. (V) The gating strategy was refined without actually losing cells of interest. Hence, population gates were chosen which just separated the two distinct populations of SP and non-SP cells without losing more events for sorting (Figure 23).

---

9. Coincident event is a situation when two cells were passing by the intercept so fast that the pulse could not be separated.

(VI) The gating was confirmed by density plot analysis to provide precise localization and distribution of the desired cells (Figure 22). (VII) Working solutions for Hoechst 33342 and verapamil were always fresh prepared before staining.
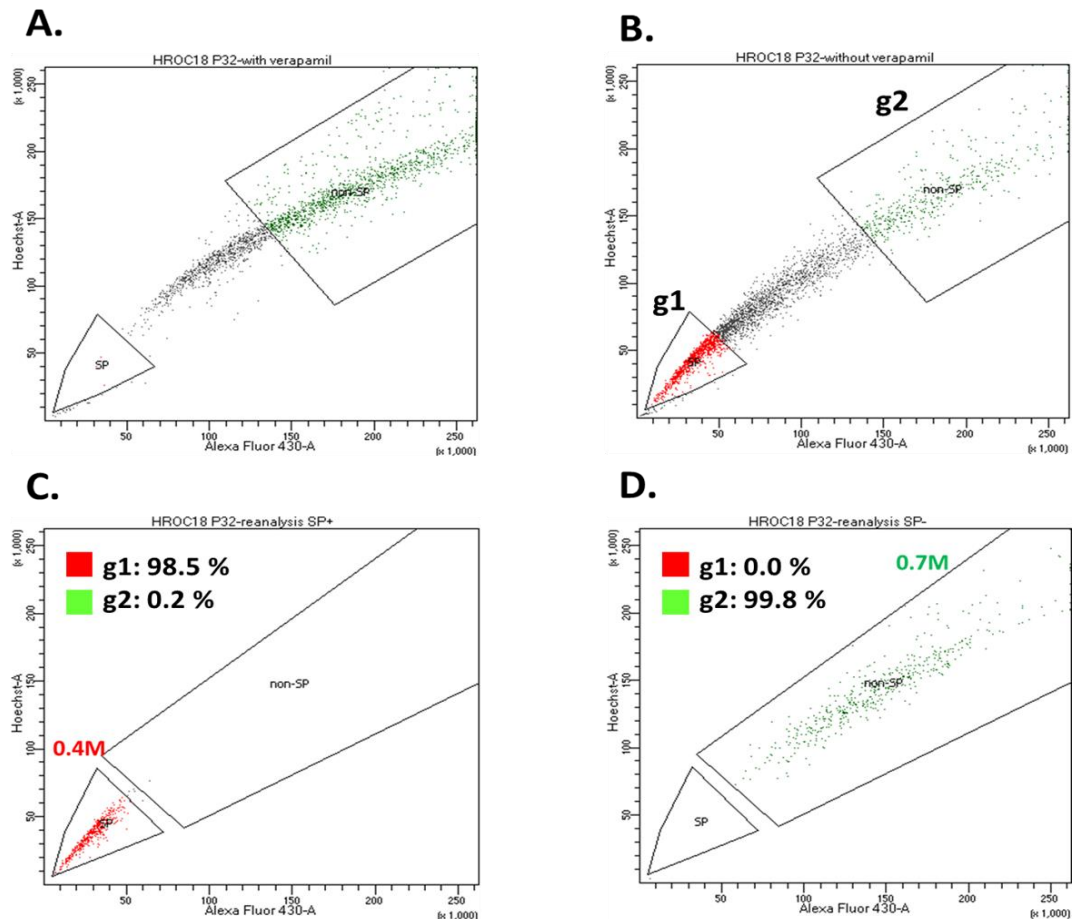
Figure 23. Representative dot plots showing improved sorting data from HROC18.

The dot plots represent; A) a control measurement, B) a stained sample with defined gated areas. G1 indicates SP and events in g2 are non-SP. C) Re-analysis of SP after sorting shows a purity of 98.5% with an approximate 4×10^5 isolated events. D) Re-analysis of 7×10^5 non-SP events after sorting indicates a purity of 99.8%.

Improvement of sorting resulted in higher purity and higher amounts of cells (representative data shown in Figure 23 and supplementary Table S2). Sorting of 17 cell lines with at least three repetitions identified a clear-cut difference among cell lines so that populations shifted in the absence or presence of verapamil just in some cell lines, including HROC18, HROC40, HROC46 T0 M1, HROC300 T2 M1, HROC313Met1 T0 M2, HROC364, and HROBMC01. Interestingly, virtually all of these cell lines were classified as MSS. HROC300 T2 M1 is the

only exception (MSI). In the remaining cell lines, fluctuation was observed and obtaining pure sorted samples was tricky, especially when this shift not occurred.

4.2.2.2. Improving RNA isolation from sorted samples

The overall number of sorted events was low and not enough for RNA purification using the standard procedure explained earlier. Consequently, isolating RNA from small cell numbers became a major problem and initially resulted in poor quality and low concentrated RNA (Figure 24B), which was not sufficient for cDNA synthesis. Hence, this method needed to be optimized in order to have adequate and qualified RNA for subsequent analysis.

A.

Cell numbers	RNA concentration (ng/μl)							
	Quick-RNA™ MiniPrep	innuPREP RNA Mini Kit	Direct-zol MiniPrep	Roboklon Kit	TRISure	TRISure + Glycogen	E.Z.N.A.® MicroElute®	peqGOLD MicroSpin
25×10^3	8.1	8.5	28.2			27.7		
5×10^4	12.8	13.5	24.8	2.7	0.6	24.7	20.6	16.9
1×10^5	25.1		43.3	5.95	9.4	26.5	40.5	**48.2**
2×10^5	40.3		55.2	12.27	61.4	42.6	67.4	
4×10^5				36.50	115.1			

B.   **C.**
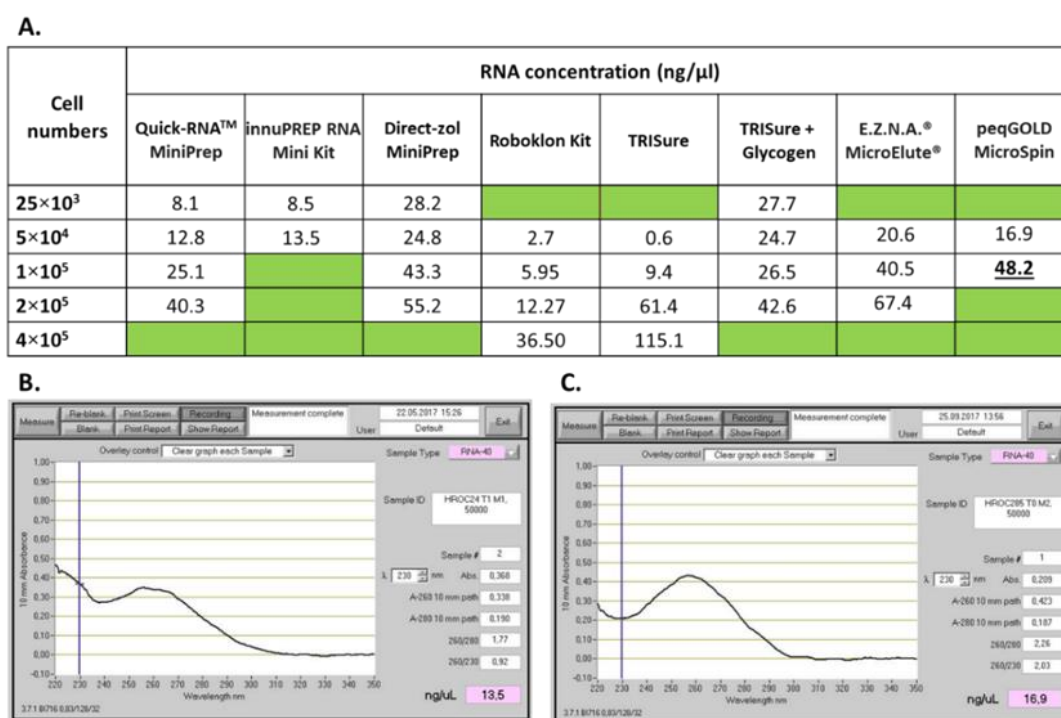
Figure 24. Summary of different kits used for improving RNA isolation from low cell numbers.

A) Several kits were utilized to purify RNA of low cell numbers. Exemplary RNA graphs indicating; B) unqualified RNA isolated with innuPREP RNA Mini kit, and C) qualified RNA extracted using peqGOLD MicroSpin kit.

Several kits (Figure 24A) were utilized to solve this technical issue. The results of the kits' evaluation revealed peqGOLD MicroSpin kit was efficient to provide purified RNA (Figure 24C) that can be used for the downstream applications.

Almost 80 sorting experiments were performed to provide qualified samples of sorted cells. Afterwards, RNA isolation of sorted populations and the matched maternal cells (n = 102 samples) was conducted using the kit. Similar to the procedure mentioned earlier, controlling RNA quality and consequent steps were done. Then, high quality cDNA was stored at -20 ˚C for subsequent qPCR analysis.

4.2.2.3 Gene expression analysis in SP and non-SP populations

In concordance with mRNA expression data obtained in colonospheres, a few loci were chosen for further analysis in sorted populations to investigate whether expression of loci was altered in two purified populations of a tumor samples (so-called SP and non-SP). The candidates consisted of *NANOG* and *gag* locus on chromosome 20 (2000045_h_gag) due to their general upregulation in 3D samples as well as *gag* and *env* genes on chromosome 13 because of the interesting opposite expression pattern.

Although *NANOG* was consistently overexpressed in colon spheroids, a decreasing dominant trend was observed in isolated SP cells. *NANOG* expression in the non-SP was significantly higher ($p < 0.05$) in most tumor cell lines as shown in Figure 25A. In the SP fractions of HROC278Met T2 M2, HROC300 T2 M1, HROC324 ($p = 0.04$), HROC364, and HROC383 ($p = 0.01$) *NANOG* was also upregulated. Expression patterns of *gag* and *env* genes on chromosome 13 as well as *gag* gene on chromosome 20 were nearly similar among sorted populations. Based on that, up or down regulation was observed in all HERV-H loci for SP or non-SP fractions. Some of these changes were remarkable in selected samples (Figure 25B-D). As displayed in Figure 25, expression of these loci was lower in isolated SP fraction. By contrast, isolated SP of HROC324 and HROC383 demonstrated a significant higher value of all examined loci in comparison to the matched non-SP. Likewise, this trend was seen in SP fraction of HROC278Met T2 M2, HROC309, HROC313Met2 T0 M2, and HROC364.

Analysis of Spearman r correlation between *NANOG* and HERV-H loci revealed a different association in SP and non-SP fractions of CRC cell lines. Correlation was seen among all loci and much stronger in the non-SP. In the SP cells, *NANOG* was positively linked with 1300360_h_env ($r = 0.72$, $p = 0.002$) and 2000045_h_gag ($r = 0.63$, $p = 0.009$), but no
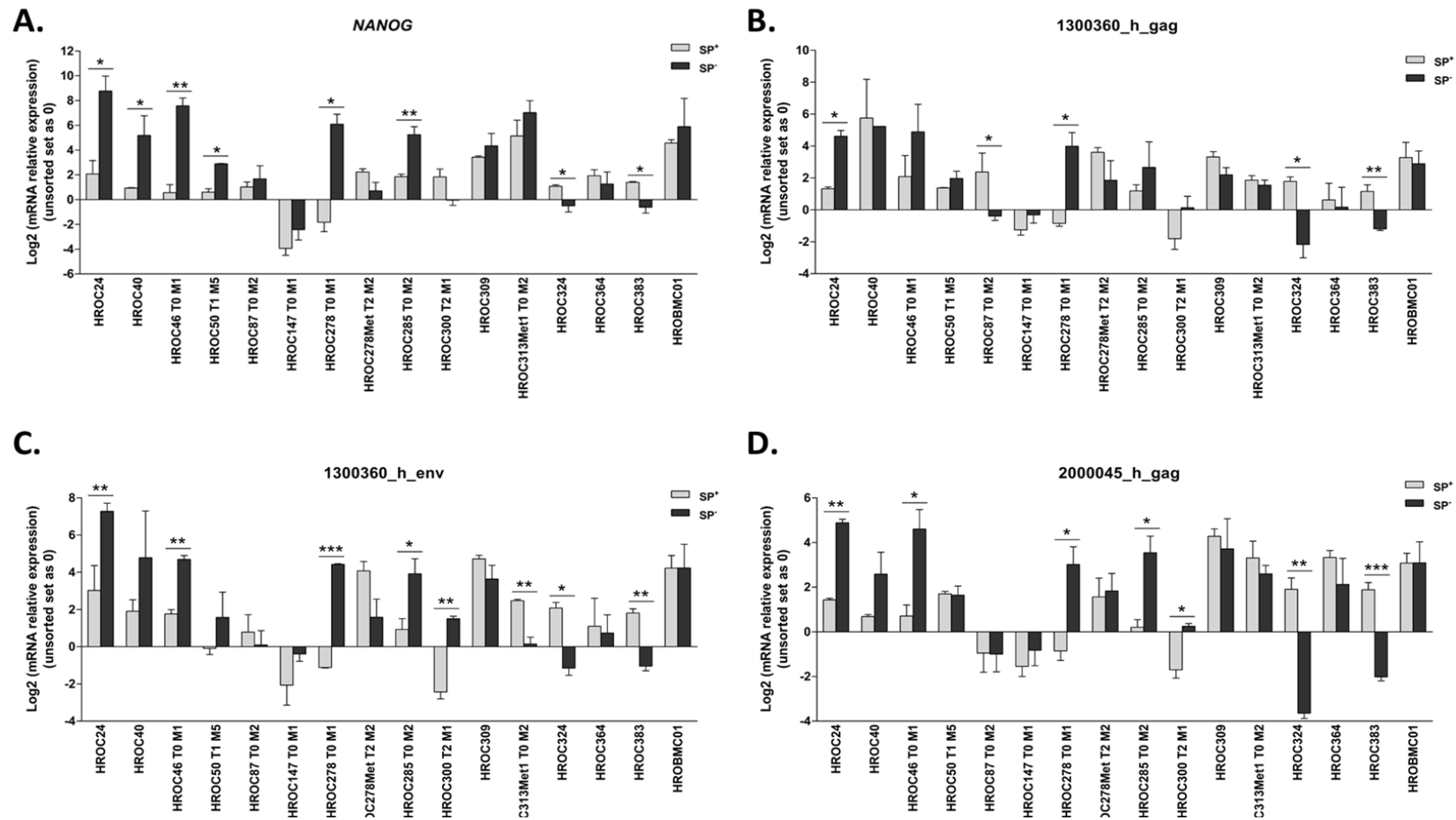
Figure 25. Relative RNA expression of NANOG and selected HERV-H loci in sorted population.

The bar graphs represent the qPCR results of NANOG and three CRC-HERV-H loci in SP (Hoechst^{low}) and non-SP (Hoechst^{high}) of the selected cancer cell lines. A) NANOG, B) gag gene form locus 1300360, C) env gene form locus 1300360, and D) gag gene form locus 2000045. Expression levels were firstly normalized to GAPDH and the expression rates were calculated by $2^{-\Delta\Delta Ct}$. The values were transformed to log 2 and unsorted cells set as zero. The data are presented as the mean values ± SEM. (*P < 0.05; **P < 0.005; ***P < 0.0005)

correlation was found between *NANOG* and 1300360_h_gag. In non-SP, *NANOG* was additionally correlated with 1300360_h_gag (r = 0.81, p = 0.0001). Another dissimilar correlation between SP and non-SP was related to *gag* genes on chromosomes 13 and 20. These loci were not correlated in SP (r =0.37, p = 0.15) but strong positive association (r = 0.85, p < 0.0001) was apparent in non-SP.

4.3. Lentiviral reprogramming

To investigate the influence of pluripotent factors on the identified HERV-H loci, HROC24 with low expression of nominated HERV-H loci and stemness factors was subjected to lentiviral-based reprogramming. The experimental approach aimed at strong upregulation of known stemness factors. This was done by applying EMD Millipore's Human STEMCCA lentivirus reprogramming kit according to the manufacturer's instructions (Figure 26A).
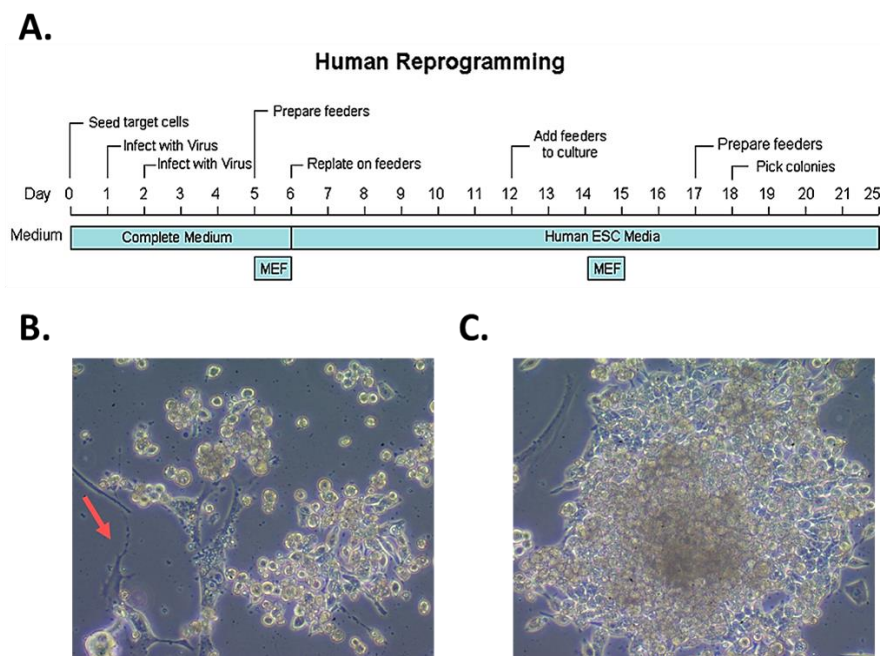
Figure 26. iPSCs generation using a single lentiviral stem cell cassette.

A) The scheme of reprogramming procedure in somatic cells which was applied in transduction of HROC24. Optical microscopy images of B) HROC24 and MEF cells (indicated by a red arrow) in culture, C) a possible iPS colony 25 days after reprogramming.

Briefly, two step transfections were performed and stem cell medium was refreshed every other day. Colony growth was monitored daily by light microscopy. Nearly three weeks after the

first transduction, 40 tightly packed colonies were picked up manually and each one transferred into a well of a 48 well-plate. 29 out of 40 colonies gave rise to grow and were consequently seeded in a bigger culture environment. Then, cells from each colony were frozen for DNA and RNA isolation.

First, gDNA extraction from the frozen transfected HROC24 cells (one day after transfection) was successfully performed. PCR of the gDNA revealed the vector incorporation into the genome (Figure 27A). The size of the PCR product was smaller than 750 bp and harmonized with the expected yield (Figure 27B). Thereby, gDNA extraction of 29 colonies and subsequent PCR with the same primer and program was accomplished. Despite the positive signal of the previous experiment, no band was observed in either colony (*data not shown*). Hence, in colonies, no sign of vector integration was achieved.
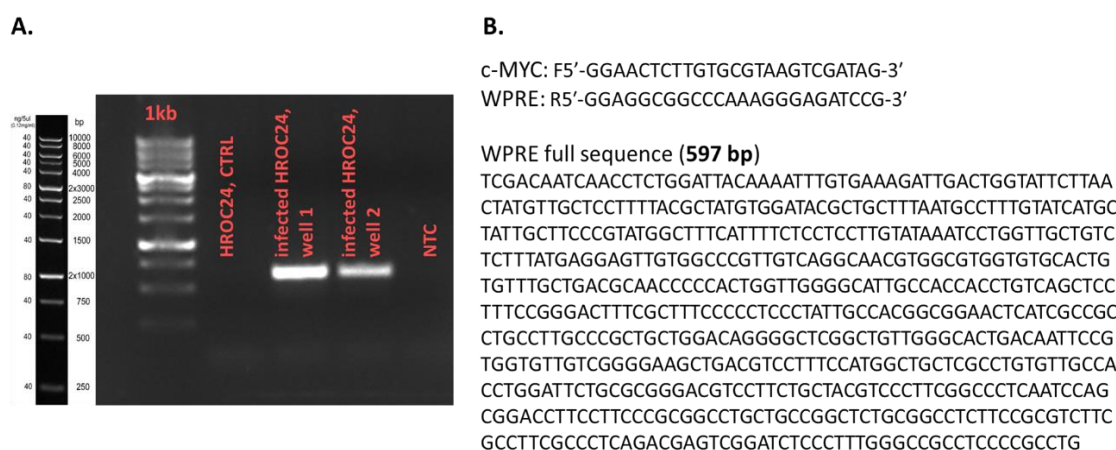
Figure 27. PCR result of infected cells with a single lentiviral stem cell cassette.

A) The PCR of gDNA demonstrates the viral integration into the genomes of HROC24 cells. B) Length of the primer and WPRE fragment (Woodchuck hepatitis virus post-transcriptional regulatory element) is a confirmation for the accuracy of the product.

Several PCR programs were then applied (Table 13) and two additional primer pairs used. These include the following: 1st pair including OCT4 5′ NotI (5′TTT TGC GGC CGC CAT GGC GGG ACA CCT GGC TTC GG-3′), KLF4 3′ BamHI (5′TGT TGG ATC CTT AAA AAT GCC TCT TCA TGT GTA AGG CG-3′); and 2nd pair including SOX2 5′ NdeI (5′-TTT AGT GCA TAT GAT GTA CAA CAT GAT GGA GAC GG AGC TG-3′); cMYC 3′ AccI (5′-TTT AGC AGT GGT ACG TCG ACT TAC GCA CAA GAG TTC CGT AGC TGT TC-3′). The aim of these PCR reactions was to specifically amplify the vector sequences encoding the transcription factors *OCT4*, *KLF4*, *SOX2*, and *c-MYC*. Pfu polymerase was also utilized to

improve PCR efficiency. Unlike Taq polymerase, Pfu has slower polymerization speed, and it possesses 3'-5' exonuclease proofreading activity that recognizes and removes the incorporation errors. Still, none of the experiments delivered valuable data. Although an unspecific band was detected in some PCR reactions, the result was not improved by changing conditions and desired products could not be amplified.

Table 13. Variety of PCR conditions used to amplify vector sequences.

95 °C, 5 min	94 °C, 2 min	94 °C, 2 min	94 °C, 2 min	94 °C, 2 min
40 cycles:	**5 cycles:**	**5 cycles:**	**5 cycles:**	**5 cycles:**
95 °C, 20 sec	94 °C, 45 sec	94 °C, 10 sec	94 °C, 45 sec	94 °C, 45 sec
58 °C, 30 sec	58 °C, 45 sec	58 °C, 20 sec	58 °C, 45 sec	58 °C, 45 sec
72 °C, 2.5 min	72 °C, 2 min	72 °C, 5 min	72 °C, 5 min	72 °C, 5 min
	30 cycles:	**30 cycles:**	**20 cycles:**	**20 cycles:**
72 °C, 10 min	94 °C, 45 sec	94 °C, 10 sec	94 °C, 45 sec	94 °C, 45 sec
	62 °C, 45 sec	60 °C, 20 sec	60 °C, 45 sec	62 °C, 45 sec
	72 °C, 2 min	72 °C, 5 min	72 °C, 5 min	72 °C, 5 min
Taq	*Taq*	*Pfu*	*Pfu*	*Pfu*

5. Discussion and conclusion

CSCs constitute a rare subpopulation within tumors that act as a putative therapeutic target. The importance of CSCs in tumor growth, development, metastasis, and drug resistance has been documented enormously in several cancer entities. Nevertheless, the proper identification and isolation of this cell fraction from tumor tissues remains controversial. Previously established methods mainly rely on cell surface markers. However, growing understanding showed that most flow cytometry approaches were not ideal to purify CSCs. Some of the markers are expressed in both normal and cancer precursor cells, and none of them are completely specific for CSCs. Seemingly, new biomarkers are still needed to specifically differentiate and purify stem cells in cancers (173).

The aim of this study was to find a HERV-H locus involved in pluripotency as an innovative biomarker for stem cell recognition in CRC. For this purpose, expression of desired stemness genes and HERV-H loci was analyzed in a large number of patient-derived CRC models, corresponding spheroids, and sorted populations.

5.1. HERV-H loci correlation with pluripotency in primary CRC cell lines

5.1.1. HERV-H involvement in CRC development

The unique expression of the HERV-H family in CRC specimens and cell lines was repeatedly determined in several studies (128,129,142,174). In 2012, Perot *et al*. analyzed the microarray data obtained from a collection of 10,035 different HERV elements containing 6 HERV families (i.e., family H) in cancerous and corresponding normal RNA samples of breast, colon, lung, ovary, prostate, testis, and uterus. This analysis revealed that patterns of HERVs expression were extremely influenced by tissue origin and the status of the differentiation. The result repeatedly confirmed previous findings and suggested the H family to be the dominant expressed HERVs element in CRC. A total of 166 HERV-H loci were detected in colon tissue, while 21 HERV-H elements were specifically expressed in CRC (130). A further analysis of the nominated loci in the previous study led to the characterization of the five individual loci (seven regions) that were only reactivated in CRC tissues (131). In metastatic samples, HERV-H mRNA level was lower than primary tumors. Moreover, expression profiles of tumors and matched metastases from the same patient revealed no differences (131).

In the current study, the expression of these seven HERV-H sequences was tested in 56 CRC cell lines. The HERV-H expression patterns were correlated with pluripotency-related genes, clinico-pathological parameters (TNM classification, grading), and the molecular subtype. Virtually, all cell lines had different expression patterns of these sequences. Sequences on chromosomes 5 (500502_h) and X (X00041_h) had the highest expression level in CRC cell lines, while 1400035_h had a lowest level. Co-expression of two genes from the same locus on chromosome 13 (*gag* and *env*, r = 0.78) and 20 (*gag* and *pol*, r = 0.83) indicated a significant positive correlation (p < 0.001) analyzed in all samples. With respect to metastasis-derived samples, only one cell line, namely HROC278Met T2 M2 had no expression of the *gag* and *env* genes of chromosome 13. By contrast, in the matched primary cell line (HROC278 T0 M1) both genes were expressed at low level. In 2015, Perot *et al*. reported exclusive expression of *gag* gene on chromosome 13 in 69% of metastasizing CRC specimen, while *env* was hardly detectable and only in one metastasis case (131). Additional comparisons are warranted. However, this could not be done here because of lack of more matched primary and metastatic cell lines.

Perot *et al.* (2015) also identified a correlation of HERV-H expression with the molecular subtype and lymph node infiltration in primary and matched metastasis tumor material (131). For the former, HERV-H expression was much higher in MSI tumors compared to the other molecular subtypes. With regard to the latter, lymph node infiltration was significantly associated with HERV-H reactivation. By contrast, no correlation was seen with tumor grading, localization, age, and mutations in classical CRC drivers, including *APC, TP53, KRAS*, and *BRAF* (131). The data obtained in our study were in contrast to Perot *et al.* findings. With regard to the molecular subtype, correlations were mainly seen among HERV-H loci with the MSS, but not the MSI subtype (Figure 15). Specifically, observed correlations of *gag* and *env* genes of locus 13 with *gag* and *pol* genes of locus 20 and *gag* on chromosome X was not detectable in MSI cell lines. These data suggest that HERV-H loci on chromosomes 13, 20, and X may contribute to progression of MSS in CRC.

Furthermore, Roschke *et al.* found enhanced expression of genes involved in migration, invasion, metastasis, and adhesion in CIN cancer (175). Besides, the identification of HERV-RNAs in tumor-derived vesicles suggests a role in stimulating metastasis (113). Thus, based on the previous findings and overexpression of HERV-H loci in CIN cell lines in this work, we conclude that HERV-H transcripts may be implicated as regulatory RNA in EMT.

Most of the chromosomal alterations are not clonal (nonclonal chromosomal alterations, NCCAs), are ignored in the analysis, and therefore not reported (176). By causing deletions, translocations, gene amplifications, inversions, and duplications, chromosomal aberrations play a causal role in cancer development, progression, and contribute to a poor prognosis in several cancer types (176,177). Given the possible role of HERVs in chromosomal rearrangement, the idea arises of assigning these sequences in NCCAs. In chromosomal abnormalities in CIN tumors, copy number of a gene of interest can be higher in cancerous cells due to variable number and/or structure of chromosomes. In addition, HERVs' elements are retrotransposon elements that allow them to integrate into new genomic locations, providing a better and more opportunity for HERVs gene expression. Consequently, they can be highly expressed in cancer cells and in this case especially in CIN tumors. These cumulative data let us to hypothesize that HERVs expression is highly correlated with CIN tumors. By going one step further, one may even conclude that HERV-H transcripts can be an indicator of this subtype in CRC. However, this issue must be examined in detail by comparison of normal and tumor cells – or ideally fresh tumor and normal tissue.

5.1.2. HERV-H implications in pluripotency

The recent decade has seen a surge of interest in investigating whether transposable elements such as ERVs can serve as a novel binding site for transcription factors (178–180) or regulatory sequences (181). In this regard, Kunarso and colleagues found that binding sites of pluripotent transcription factors including OCT4 and NANOG with transposable elements are species-specific. It was also illustrated that ERV1 (including HERV-H) was the major key in human stem cells (182). RNA seq-data from Santoni *et al*. confirmed that HERV-H was the most expressed repetitive element in ESCs/iPSCs. Its transcripts correlated with OCT4 and NANOG and decreased with differentiation. *Vice versa*, SOX2 was stable and exhibited no link to HERV-H. Notably, the degree of active histone mark H3K4me3 was substantially associated with the status of differentiation. Another point of this study was evaluation of possible implications between HERV-H genomic location and binding site of the pluripotency transcription factors. In here, association was strong with NANOG, weak with OCT4, and no relationship was evident for SOX2 and KLF4 (140). Contrary to this study, Ohnuki *et al.* found KLF4 to be particularly important for HERV-H reactivation and re-suppression, while NANOG had a weaker role. During reprogramming, simultaneous overexpression of *KLF4* and

transient active LTR7 was detected (183). A previous study also described a link between HERV-H and lncRNAs (140), emphasizing HERV-H as an enhancer for lncRNAs and nearby genes (88). Besides, the LTR7 subfamily had the highest representatives at NANOG and OCT4 binding sites (140). This indicates that expression of HERV-H LTR was triggered by OCT4, so that *OCT4* depletion led to a reduced activity of these LTR enhancers (88).

These observations imply that HERV-H transcripts play a remarkable role in human pluripotency by acting as lncRNAs to supply a platform for pluripotency transcription factors or histone modifiers (184). Thus, in this thesis, HERV-H expression was correlated with stemness factors to find a new biomarker for CR-CSCs identification. Obtained data from all cell lines (n = 56) revealed significant correlations of 1400035_h_gag with all stemness genes. This trend was likewise between *c-MYC* and all HERV-H sequences. Notably, complementary analysis indicated subtype-specific differences (summarized in Figure 28), with the strongest connection of stemness markers *KLF4* and *OCT4* in the MSI subtype, but virtually no correlation in the CIMP subtype. Moreover, we identified stronger correlations between HERV-H loci and pluripotency genes (mainly *NANOG*, *KLF4*, and *c-MYC*) in advanced nodal involvement (N1/2), metastasis (M1) at time of resection, and tumor stages (T4). Interestingly, observed correlations were completely lost in the metastases cell lines.
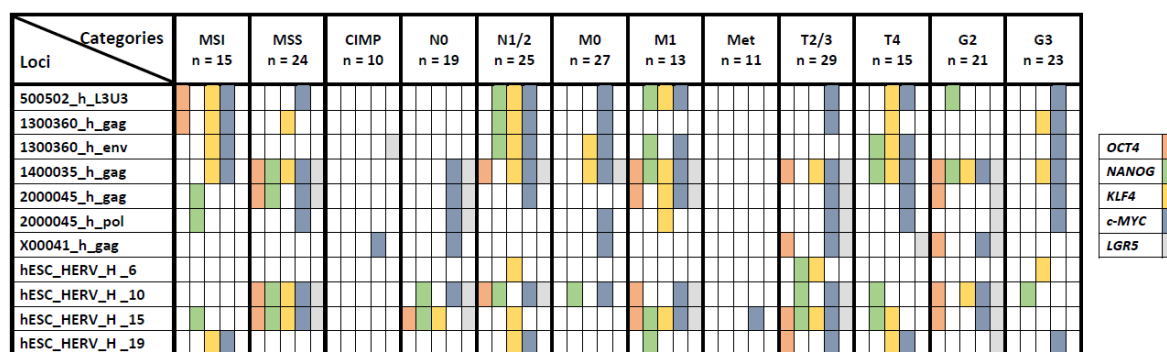
Figure 28. Summary of the significant correlations of HERV-H loci with stemness related genes in CRC cell lines based on the molecular subtype, TNM, and grading.

Few studies found that expression of HERV-H was enormously monopolized in human stem cells (140,178,185). Lu *et al.* showed that HERV-H activation was controlled by OCT3/4 and p300 (88). HERV-H was indispensable for generation and maintenance of self-renewal ability in iPSCs (87,88,183), as loss of pluripotency was a direct result of HERV-H knockdown in SCs (87,183). Given a role for ERVs family H in pluripotency and the exclusive expression of the same family in CRC, some new loci were selected based on a study in 2014 (87). Wang *et*

al. revealed that HERV-H had the highest expression in stem cell samples using RNA seq-data of human fibroblasts, ESCs, and iPSCs. Uniquely aligned reads of the 1,225 full-length HERV-H genomic region in 43 normal somatic tissues, 8 cancer cell lines, 55 ESCs, and 26 iPSC samples indicated 117 highly activated LTR7 in human stem cells. Comparison of HERV-H transcription level between iPSCs at early passages and ESCs showed a higher value in iPSCs, suggesting that HERVH expression is induced after reprogramming. ChIP-seq (chromatin immunoprecipitation-sequencing) data also confirmed that active HERV-H were marked with active histone modifications. Additionally, active LTR7 was hypomethylated in stem cells. These findings highlight the role of HERV-H-LTR as promotor or enhancer. Further investigation showed that HERV-H transcription in stem cells is driven by transcription factors such as OCT4, NANOG, KLF4, and LBP9 (87). Accordingly, four new loci on chromosomes 6, 10, 15, and 19 were selected from 117 full-length and highly activated HERV-H LTR. The RNA-seq analysis data in a single pluripotent cell showed that the candidate sites showed less activity after 10 passages (87).

The expression level of these selected ESC-HERV-H sequences was almost similar to CRC-HERV-H, showing low transcription in most cell lines. However, the locus 19 transcript was undetectable in 25% of samples (14 out of 56 cell lines). Spearman r correlation indicated a strong link between hESC_HERV_H_10 and the pluripotency factors at the transcription level. Interestingly, there was barely a correlation with the CRC-HERV-H loci detectable. Unlike, loci 6 and 15 were strongly associated with all the CRC-HERV-H, but weakly with *NANOG*, *KLF4*, and *c-MYC*. This suggests implication of locus 10 in pluripotency, while loci 6 and 15 contribute to CRC tumorigenesis. Besides, the correlation in the classified groups resulted in a subtype-specific pattern. Overall, the loci 10 and 15 had the most frequent connections with pluripotency factors such as *OCT4* and *NANOG*.

Summarizing these findings, we here confirmed that overexpression of putative HERV-H transcript controls expression of pluripotency factors. This includes, among others, that the *gag* gene of chromosome 14 and LTR of chromosome 10 are implicated in stemness state (Figure 28).

5.2. HERV-H expression in enriched populations of cancer stem-like cells

To examine this further, the qPCR and correlation analysis were accomplished on stem cell enriched populations. CSCs-enrichment was based on functional properties. The lack of

generally accepted surface markers for CR-CSCs led us to focus on their ability to efflux Hoechst dye (referred as SP) and to form spheroids.

5.2.1. Cancer stem cells enrichment in CRC

By applying flow cytometry, our cell lines did not contain significant amounts of SP cells. Therefore, SP low and high were sorted to enrich these rare populations. However, getting a high purity of the sorted fraction was a challenge since there were no two distinct cell subsets. Still, SP cells and non-SP sorting were successful from all 17 cell lines.

Another point worth mentioning is the fact that SP percentage was not stable in the cell lines and changed during culture. Moreover, it was observed that the percentage of SP cells was cell line/tumor dependent. These cells could not be discriminated in all cell lines (186) and some cancers did not even contain SP cells (164). The SP cells may be functionally characterized as CSCs, although expression of ABC transporters is not an exclusive phenotype of all stem cells (164). Despite this limitation, sorting SP has become a valuable and established strategy to enrich stem cells from different solid tissues (187) and numerous cancers (188,189). Several studies also showed that only SP cells have self-renewal capacity, the ability to develop colonies from single cells, and express stemness genes. Finally, the high tumorigenic potential *in vivo* following serial transplantation of low cell numbers or even single cells functionally confirms their cancer-initiating potential (83,85,190). However, investigations on ABCG2 expression in CRC patients were discordant. Several studies reported that *ABCG2* mRNA level was lower in tumor tissues than in normal tissues (191–193), which did not match with the hypothesis of increased expression of ABC transporters in cancers. Moreover, abundancy of ABCG2 protein was positively linked with lymph node metastasis (194), shortened patient survival (195), and worse prognosis in CRC patients (196). On the contrary, in a recent study, high expression of this protein significantly correlated with the better survival rate (197). Given the large discrepancies in clinical studies and those seen in CRC cell lines, sorting SP or evaluation of ABCG2 expression may not be a valuable approach to isolate CR-CSCs.

Parallel to the sorting, candidate cell lines were cultivated in 3D using a non-adherent environment in serum free medium. The rationale for this is based on the stem cells' ability to adapt to a new niche, survive, and proliferate, whereas differentiated cells are forced to die by apoptosis (198). Since cell surface markers have a dynamic expression that may be present in multiple cell types, they are inappropriate for isolating the population of interest (199). By

contrast, spheroid culture might be an optimal model with even better reflecting *in vivo* tumor growth, progression, and drug delivery (200). In fact, morphology and physiology of cells differ in the 3D culture environment resulting from new arrangement of cell surface receptors and physical restraints. These aspects modify signaling pathways, gene expression, and cellular activities (160). Nowadays, there are increasing evidences to acknowledge that spheroid forming culture is the ideal method for screening stem-like features in normal or cancer cells (173). Spheroid culture has been used extensively to enrich CSCs or tumor-initiating cells in various cancers (201–205). This technique has already been utilized in CRC to address whether tumorigenicity, therapy resistance, DNA methylation pattern, expression of pluripotency markers, and stem characteristics resembled the original tumor in the spheroid culture (206–210). Altogether, there was an agreement that 3D cellular structures closely match with the patients' tumors *in vivo*. Hence, this technique is a promising model in stem cell research.

In this work and in agreement with former studies, spheres were successfully generated from 17 cell lines, confirming the presence of a rare population within our patient-derived cell lines that maintained growth in an environment that mimicked the stem cell state. Of note, morphological heterogeneity was observed among cell lines during sphere formation which is a new finding in CRC-spheroids. These 3D structures were so-called named as grape-like clusters (Figure 18A), solid spheres (Figure 18B), and hollow spheres (Figure 18C). These variations in 3D structure bring up the question of whether different progenitors or stem cells are responsible for tumor maintenance in CRC, or whether tumorigenicity, expression of stemness markers, and the capacity to grow in long-term culture are influenced by morphological differences. Similar to our observations, spheroids of sensory epithelium cells produced solid and hollow spheres (211). It was noticed that hollow spheres resulted from solid spheres transition as they enlarged. This was followed by sphere cell maturation and overexpression of E-cadherin. Interestingly, hollow spheres did not retain the sustained growth (211). In this study, spheres' reproducibility was not investigated since this was not an ultimate goal. Nonetheless, we indeed observed differences in growth kinetics, with hollow spheres growing continuously, accompanied by enlargement in number and size. On day 7, they were structurally similar to organoids and alive.

5.2.2. Stemness gene expression in enriched population

In line with pluripotency of colonospheres, many studies revealed that certain self-renewal genes such as *OCT4*, *NANOG*, *SOX2*, and *KLF4* were upregulated in spheroid models compared to 2D-cultured monolayers (212–215). Herein, *OCT4* showed no considerable changes in expression, although it was upregulated in a few colonospheres. Additionally, no pattern or correlation could be detected regarding the molecular subtype, TNM, and grading. Another surprising finding was downregulation of *KLF4* in all spheroids. This transcription factor is recognized as one the foremost regulators involved in normal and cancerous SCs compartment in intestine (216).

KLF4 in CRC

KLF4 is well-known as one of the Yamanaka factors for reprogramming of somatic cells (217). It regulates transcription and functions as a tumor suppressor or as an oncogene in various types of cancer (218). Recent studies have repeatedly recognized the regulatory role of KLF4 in CSCs (219–221), although conflicting results have been obtained in CRC and its stem cells compartment (222,223) as well as other tumor entities (194,195). In CRC, KLF4 is frequently acknowledged as a tumor suppressor (226,227) and inhibits cellular proliferation, colony formation, migration, invasion, and tumor formation in animal models (223,228,229). Additionally, downregulation of *KLF4* led to the increased DNA synthesis, hyperproliferation, and less differentiated phenotype (230). Moreover, *KLF4* expression in CRC tissues or cell lines is induced or lost (231,232). Xu *et al*. reported that KLF4-positive patients had better survival rates compared to the KLF4-negative counterpart (231). By contrast, Lee *et al*. found that elevated KLF4 to be associated with worse prognosis and shorter survival (232). In another study, lower KLF4 was correlated with poor prognosis (227). These divergent findings hamper any conclusion on the role of KLF4 on patients' outcome.

In another study by Leng *et al.,* colon spheroids had higher KLF4 level in comparison to the matched parental cells and *KLF4* knockdown significantly reduced *in vitro* and *in vivo* stemness phenotype of spheroids, confirming that KLF4 acts as an oncogene in CRC development (219,222). Recently, Stankevicius *et al.* observed that KLF4 expression in colonospheres was downregulated on day two and elevated on day 6 (233). Another investigation on nasopharyngeal carcinoma also indicated that KLF4 overexpression

negatively regulated cell proliferation and stemness features such as sphere efficiency, migration, invasion, and expression of related stem cells markers (224).

The conflicting observations in CRC may be due to the complex involvement of KLF4 in numerous signalling pathways. These can be further affected by sample types, different methods applied for analysis, and therefore finally bias results. Another point worth mentioning is that most studies focus on, long-term cultured commercial cell lines. Though such cell lines are widely used in cancer research, they may carry additional mutations caused by serial passaging and the rather artificial *in vitro* culture. Hence, obtained results must always be interpreted with caution and – in worst case – do not adequately resemble the cell of the origin (157). However, in this work, low passage patient-derived and PDX models were used. These cancer sources provide better and valuable results with higher relevance for the clinic.

As explained earlier, cell cultivation in non-adherent and serum-free environment exerts apoptosis to differentiated cells, while rare cancer stem-like cells adjust to the condition and keep their activities. Based on the literature and in line with KLF4 enrolment in cell growth arrest in CRC, KLF4 may be prohibited in 3D structures for a limited time so that more cells proliferate and become spheres. Collectively, downregulation of *KLF4* is suggested to be involved in hyperproliferation and malignant transformation of CRC. In this line, Hu *et al.* discovered a differentiation dependent expression pattern for *KLF4* in CRC. Data revealed that transcriptional and translational level of KLF4 was positively associated with differentiation grade of tumor cells (234). Therefore, a possible explanation for the reduced *KLF4* mRNA levels observed in the current study may be dedifferentiation of the mature cancer cells or more possibly activation of the dormant CSCs due to the non-adherent culture. However, this issue needs to be investigated in detail prospectively.

c-MYC in CRC

Another pluripotency gene measured in 3D structures was *c-MYC*. The obtained qPCR data indicated that 70% of colonospheres had lower *c-MYC* transcripts compared to the matched monolayers. It was assumed that *c-MYC* would be overexpressed in the non-adherent niche since it plays a key role in maintenance of self-renewal and survival of SCs in CRC (235). Nevertheless, expression of *c-MYC* has not been studied well in tumorspheres, while its implication has been demonstrated in a wide range of cellular activities in most cancer types, including tumor initiation and proliferation (236), angiogenesis (237,238), migration and

metastasis (239,240), immune evasion (241), resistance to therapy (235,242), cell reprogramming, and stemness (235,243,244). The latter proposes a link between malignancy and stemness (235).

Activity of c-MYC in CRC is regulated via Wnt/β-catenin signaling pathway. Nuclear accumulation of β-catenin trigger *c-MYC* and *Cyclin D1* transcription that prompt cell proliferation and tumor development (242). Furthermore, overexpression of c-MYC in CRC mediated organoid formation and SC/progenitor expansion through reactivating self-renewing cells and relevant signaling (245). Recently, c-MYC expression and its contribution in sphere formation, mobility, *in vivo* tumorigenicity, and chemosensitivity in the stem cell-enriched population (CD133 positive) from HT29 have been explained. As compared to CD133$^-$ cells, *c-MYC* transcripts were higher in CD133$^+$ cells. Moreover, stemness features were attenuated in the CD133$^+$ fraction when *c-MYC* was downregulated (235). Contrarily, Stankevicius *et al.* detected no remarkable changes in *c-MYC* expression in DLD1 and HT29 cells grown in 3D models (233).

There are also studies demonstrating *c-MYC* activation in a dose-dependent manner (246,247). Low level c-MYC expression enhanced proliferation without increasing cell death, size of the SP population, expression of pluripotency transcription factors, and sphere forming efficiency in hepatic cancer cells (244). Besides, c-MYC protein was degraded in oxygen, glucose, and glutamine deprivation. This downregulation was positively correlated with increased cell survival in CRC (248). ATP depletion in deprived condition induced senescence and cell death through necrosis, apoptosis, or autophagy by c-MYC activation (248–250). Similarly, Wong *et al.* examined mechanisms of c-MYC inhibition under hypoxia in human CRC. It was demonstrated that c-MYC protein level, activity, stability, and transcription of its target genes were reduced in HCT116 cells (251).

Taken all together, spheroids create a microenvironment in which external cells receive sufficient oxygen and nutrients from the medium, while distant cells are in shortage. Energy consumption in the hypoxic regions and consequently DNA replication, metabolism, and cell proliferation decreases by preventing *c-MYC* (251). Given the role of c-MYC in promoting cell death and its involvement in maintenance of CSCs, c-MYC expression pattern in CR-CSCs is possibly dependent on the availability of supplies in the tumor niche or/and c-MYC abundance in the target cell. These characteristics lead to the several fluctuations to keep the basal c-MYC level stable. Thus, transient *c-MYC* downregulation can serve as a survival adaptive strategy

for CSCs or tumor initiating cells in nutrients/ATP deprivation. As a result, c-MYC protein must be broken down and most likely *c-MYC* transcription also diminishes in colonospheres.

NANOG in CRC

In this study, *NANOG* was the only pluripotency gene that was upregulated in all examined spheroids. This is in agreement with data emphasizing that *NANOG* is a key modulator in CR-CSCs. *NANOG* affects tumorigenicity, spheroid formation efficiency of individual cells, and size of the SP (213). It was also reported that CRC patients with elevated NANOG protein had poorer prognosis and more advanced lymph node status. Furthermore, NANOG was introduced as an EMT- inducer promoting proliferation, invasion, and mobility of cancer cells during CRC progression (252). Consistent with this report, independent loss of function studies reinforced the idea that NANOG is the most important factor in controlling CSCs population in CRC. Knockdown or genetic ablation of *NANOG* in CRC cell lines reduced colonospheres (213), impeded cell proliferation, and inhibited *in vivo* tumorigenicity (213,253). Likewise, it has recently been shown that the CSCs phenotype in CRC diminished in knockdown *NANOG* cells. These cells had lower levels of NANOG, OCT4, and c-MYC protein. The size of the tumor xenografts in nude mice and the size and frequency of the spheroids were also reduced (254). Strikingly, Tamura *et al.* demonstrated for the first time the presence of two functionally distinct subpopulations in clinical CRC samples based on the E-cadherin expression. Of note, *NANOG* transcripts and protein were remarkably upregulated in positive sorted cells, which coincided with higher proliferation rate. Actually, NANOG was introduced as a main regulator of proliferation in potential CR-CSCs (255). In another study, it was found that therapy with doxorubicin, irinotecan, or 5-fluorouracil led to the increased SP fraction and elevated NANOG and E-cadherin in CRC lines (256). On the other hand, compulsory overexpression of *NANOG* enhanced clonogenicity, EMT-related properties, and tumorigenic potential *in vivo* (257). Finally, NANOG expressing CRC cells indicated more capability to evade the immune system, supporting tumor growth (258,259). All these data confirm that NANOG plays a crucial role in stemness traits and maintenance of pluripotency in CRC.

To further examine *NANOG* expression behavior in a stem cell population, measurement was performed in SP and non-SP sorted fractions. Unexpectedly, *NANOG* transcription was higher in non-SP cells (Hoechst [high]), apart from a significant upregulation in SP cells isolated form

HROC324 and HROC383. Similarly, *NANOG* was overexpressed in HROC278Met T2 M2 and HROC300 T2 M1 but the difference was not statistically significant.

Oligo microarray analysis indicated different genes expressed in SP and non-SP of a liver cancer cell line (Huh7). Mainly, genes related to stemness signaling and drug resistance were upregulated in SP (260). In CRC cell lines, inhibition of *NANOG* reduced the size of the SP cells compared to the parental population (213). Immunofluorescent analysis of sorted SP cells from fresh primary ovary specimen specified the coexpression of *NANOG* and *ABCG2* (261). In addition, induction of *NANOG* in prostate (LNCaP) and breast (MCF7) cancer cells caused upregulation in *ABCG2* expression (262). Reciprocally, it was observed that MCF7 SP cells were enriched with *NANOG* mRNA (263). Similarly, pancreatic cancer SP cells had a higher *NANOG* transcript and protein level compared to the non-SP (264). Very recently, Kawai *et al.* showed that the rate of SP decreased with the inhibition of *ABCG2*, while the ability to form spheres did not change in pancreatic cancer cell lines. It was concluded that *ABCG2* inhibition has no effect on cancer stemness, although high ABCG2 cells are enriched with CSCs (265). Considering these findings, synchronous spheroid culture and sorting SP may incredibly improve enrichment of the CR-CSCs. Perhaps there is a subfraction of NANOG cells that express high level of ABCG2. Thus, if sphere formation can enrich NANOG positive cells, it may additionally contribute to increased SP size. However, this must be studied in detailed.

Next, expression of NANOG and Gag-H proteins were tested in the spheroids from 5 cell lines including HROC46 T0 M1, HROC50 T1 M5, HROC364, HROC383, and HROBMC01. Using flow cytometry analysis, marginal differences were found between 2D and matching 3D samples. Thus, protein expression was not consistent with increased expression of *NANOG* and *gag* transcripts. One possible explanation is that these transcripts may function as a regulatory RNA as proposed by Salmena *et al.* (266). The other explanation is the nucleus and cytoplasmic localization of NANOG, suggesting a migratory behavior similar to that of OCT4, KLF4, c-MYC, and SOX2 in various cancers (267). Concordant with these results, Meng *et al.* found that only a small fraction of CRC cells accumulated NANOG in their nucleus (252). Therefore, NANOG protein can be washed away through the staining procedure when it is located in the cytoplasm. Another and the most important argument could be the size of the spheroids. This can lead to the incomplete fixation, permeabilization, and/or antibody penetration especially at depth, which then distorts the data. Quality of staining in deep regions of 3D structures will decrease if any of these steps are performed imperfectly (268). Therefore, it is recommended

to re-establish the staining procedures for 3D models, irrespective of widespread protocols used in 2D samples.

5.2.3. Evaluation of HERV-H loci in enriched population

HERV-H expression has never been studied in 3D spheroids or SP cells of CRC. One recent analysis by Giovinazzo *et al.* reported a significant positive correlation of HERV-K and HERV-H with stemness related genes. It was seen that enhanced HERV-K transcription was accompanied by overexpression of *OCT4* and *NANOG* in melanoma spheres. Similarly, concurrent upregulations of HERV-K and HERV-H with *OCT4* and *NANOG* was observed in spheroids of liver cancer. Additionally, HERV-H and *SOX2* had a significantly increased expression rate in the spheroids of lung cancer (269). Our gene expression data are almost in accordance with the previous findings showing higher expression of HERV-H elements in colonospheres examined. However, this was not confirmed on protein level in examined spheroid system. The levels of CRC-HERV-H transcripts in spheroids were higher compared to their corresponding monolayers in most cases. Furthermore, all spheroids overexpressed both genes of *gag* and *pol* on chromosome 20 and this pattern was followed by upregulated *gag* on loci 14 and X in 82% of colonospheres (14 out of 17). In contrast, transcription of the *gag* gene on chromosome 13 diminished in 9 colonospheres and was undetectable in HROC383, while *env* gene of the same locus was overexpressed in 64% of the spheroids. The disparities in expression pattern between two genes of the same locus (1300360_h_gag) have already been observed in few CRC samples (131). However, in the current work, more than half of the colonospheres had the lower 1300360_gag transcript, which was the dominant trend. Perot *et al.* declared that alternative splicing or genomic loss of the *env* gene could be the possible mechanisms during the metastatic progression, leading to the solitary expression of only one gene (131). Probably, the *env* gene is also involved in pluripotency of CR-CSCs. Analysis of *gag* and *env* genes of locus 13 in SP and non-SP did not revealed any variation and both loci were expressed in the isolated fractions. Besides, there was no consistent up or down regulation of 1300360_h_gag in sorted SP and non-SP. However, the *env* gene was mainly higher in non-SP than SP. This was also seen in *gag* gene of locus 20. Lower *NANOG* and HERV-H transcription levels in SP cells disagree with the notion that SP cells are enriched with CSCs. There are numerous variations in SP size in a single cell line, in the presence or absence of SP in CRC lines, and in the gene expression pattern obtained from 17 patient cell

lines. As a result, sorting SP cells does not appear to be a method of choice for CSCs isolation in CRC. Contrariwise, the literature indicates that spheroid culture is a practical way to propagate and expand cancer-initiating cells which can be used to study cancer development, metastases, and drug delivery (201–205).

Spearman r analysis identified new correlations in colonospheres. First, *gag* genes of loci 20 and X were positively linked to *KLF4,* suggesting the importance of these loci in stem-like cell growth. Secondly, there were significant negative correlations of *c-MYC* with 1300360_gag (r = -0.72, p = 0.001) and with 500502_L3U3 (r = -0.52, p = 0.03). However, these loci were positively linked in the 2D parental cells. These contradictory correlations in two different cultivating systems may be just a coincidence. It can also be proposed that *c-MYC* plays dual conflicting roles in two different environments (2D *vs.* 3D) by applying the same molecular network. Therefore, in a deprived niche that promotes CSCs activation, its temporary downregulation does not lead to the same pattern in the HERV-H loci 13 and 5, if they are essential to retain the expression of stemness genes in spheroids. Since association between *c-MYC* and HERV-H has never been reported, further studies will be required to define the reciprocal expression pattern of *c-MYC* and HERV-H in stem cells.

Finally, Spearman correlation was performed in SP and non-SP to find out whether level of *NANOG* mRNA was associated with 1300360_gag/env and 2000045_gag. Surprisingly, *NANOG* was positively correlated with 1300360_env and 2000045_gag in SP, whereas in non-SP it was additionally linked to 1300360_gag, as well. This data provided more clues that locus 13 and most probably the *env* gene plays an important role in CR-CSCs.

5.3. Reprogramming of HROC24

The reprogramming of human cancer cells to pluripotency is becoming an important tool in cancer research. Resected samples and matching derived cell lines mainly represent late-stage markers and phenotype. However, the iPSCs approach potentiates recapitulation of the earliest tumor stage that possibly provides new insights into the cancer progression and identification of early diagnostic biomarkers (270). The rarity of success described in literature underlines the difficulties in de-differentiating cancer cells into pluripotent precursors. It is thought that particular cancer related mutations, aberrant epigenetic modifications (270,271), accumulated DNA damage, and induced cellular senescence are the biological barriers (270). So far, most

of the iPSCs generated were limited to leukemia (272–275) and pancreatic cancers (276). Reprogramed cells were usually derived using vectors expressing Yamanaka factors. Carette *et al.* demonstrated that all Yamanaka factors are necessary for successful reprogramming in CML (KBM7) (272). In a few studies, *NANOG* and *LIN28* genes were additionally used in various combinations with Yamanaka factors (274,277). In a study in 2010, the transfection of pancreatic, liver, and CRC cells was mediated by using Yamanaka factors combined with *NANOG*, *LIN28*, *BCL2*, *KRAS*, and shRNA for tumor suppressors genes (278). Regardless of the methods used for reprogramming, obtained induced cancer cells enhanced expression of ESCs markers such as *OCT3/4* (279,280), stage specific embryonic antigen-3 (*SSEA-3*), *SSEA-4*, *SOX2* (279), *NANOG* (278–280), *SSEA-1*, *TRA-1–60*, and *TRA-1–81* (278). The induced cancer cells could differentiate into different cell lineages (272,273,275,278). Besides, they became insensitive to the known chemo-drug in an opposite behavior to the parental cells (272,273,281). Furthermore, reduced or loss of malignancy such as invasion (278), migration, division rate (279), and tumorigenicity were observed in some iPSCs (277,278). Nonetheless, findings of different independent studies indicate that reprogramming of cancer cells is not adequately efficient. Most of the cancer cells are resistant to reprogramming (282–284), in some cancers full pluripotency was not gained (285,286), and in some cell lines reprogramming had to be optimized individually (278). Our PCR data revealed differences between infected samples at day one after transfection and after three passages. This brings a theory that integration of the vector to the cancer cell genome was temporary or transient, because no bands could be detected in PCR of stem-like colonies. Most likely, the vector was excised from the genomic region and integration was not constitutive although we aimed at permanent integration into the genome to enable stable CRC-iPSCs.

To sum up these findings, reprogramming cannot be performed by commonly used factors in all types of cancers and using new factors may improve iPSCs generation in tumor models.

6. Summary

The mortality rate in colorectal cancer (CRC) patients is still high. One of the main consequences for treatment failure is tumor relapse, which occurs at advanced stages. It is presumed that cancer stem cells (CSCs) population is the cause of tumor relapse and metastasis due to its cellular plasticity. This rare population is also chemo-resistant, arising from highly concentrated drug efflux pumps that weaken drug response and induce survival pathways. The combination of these properties with the capability to evade the immune system and endure nutritional shortage results in CSCs survival after treatment and restoring tumor growth after cancer cell elimination. Thus, successful targeting and eradication of CSCs can improve cancer treatment. Still, before developing novel approaches, precise characterization of the CSCs is mandatory.

Several studies already reported the upregulation of human endogenous retroviruses family H (HERV-H) in embryonic and induced pluripotent stem cells as well as CRC tumor samples and cell lines, which was not found in normal tissues and other cancer types. Considering these findings, this study aimed to investigate the correlation between previously identified CRC-HERV-H loci and stemness genes in CRC cell lines and to find the potential HERV-H pluripotency locus as a biomarker in colorectal CSCs. Initial experiments in 56 CRC cell lines revealed positive significant correlations between *OCT4*, *NANOG*, and *c-MYC* with the specific HERV locus 1400035_h_gag. Correlation analysis with clinic-pathological characteristics including molecular subtype, TNM, and grading parameters indicated different correlation patterns in each subtype. The strongest correlation was between *OCT4* and *KLF4* in microsatellite instable (MSI) cell lines, while no correlation was seen in the other subtypes. Furthermore, correlations between HERV-H loci and *NANOG*, *KLF4*, and *c-MYC* were stronger in cell lines with advanced tumor stage. However, in the metastases-derived cell lines, no correlations were detectable indicating that HERV-H loci expression drives malignancy but plays a minor role in metastasis.

A comparison of the gene expression data between 3D-cultured colonospheres and the matching 2D-monolayers revealed a higher expression of CRC-HERV-H loci in the aforementioned. Nevertheless, the expression of the examined CRC-HERV-H loci showed an inverse pattern and was lower in side population (SP) than in non-SP, classified as non-CSCs. The same was seen for *NANOG*. Here, colonospheres overexpressed *NANOG*, but *KLF4* and

c-MYC were downregulated, which is in sharp contrast to the literature. Still, spheroid culture provides enough clues to be a practical method to enrich colorectal CSCs.

Summarizing these findings, we found that HERV-H expression in CR-CSCs was mainly influenced by *NANOG*. Although we could not specify a locus endowing pluripotency, the *gag* gene on chromosome 14 and the *env* gene on chromosome 13 were the most important candidate loci that warrant further investigations.

7. Zusammenfassung

Die Sterblichkeitsrate bei Patienten mit kolorektalem Karzinom (engl. *colorectal cancer*, CRC) ist noch immer hoch. Die Gründe für das Versagen entsprechender Therapien sind vielfältig und führen häufig zum Rezidiv. Es wird vermutet, dass eine spezifische Subpopulation, die sogenannten Krebsstammzellen (engl. *cancer stem cells*, CSCs) eine der Hautursachen für ein Rezidiv und eine Metastasierung darstellt. Diese seltene Population weist spezifische Charakteristika, wie zelluläre Plastizität und Chemoresistenz, auf. Letztere resultiert aus der Anwesenheit hochkonzentrierter Arzneimittel-Efflux-Pumpen, die die Wirksamkeit von Therapeutika abschwächen und zelluläre Überlebenswege induzieren. Die Kombination dieser Eigenschaften, sowie die Fähigkeit, auch Nährstoffmangel zu tolerieren und der immunologischen Kontrolle zu entgehen, führt zum Überleben von CSCs nach der Behandlung und letztlich zum Rezidiv. Ein erfolgreiches Targeting und die spezifische Elimination von CSCs stellt einen vielversprechenden therapeutischen Ansatz dar, die Prognose von CRC Patienten langfristig zu verbessern. Vor der Entwicklung neuer Ansätze ist jedoch eine genaue Charakterisierung der CSCs entscheidend. Dies war Gegenstand der vorliegenden Dissertation mit dem Schwerpunkt, eine Assoziation zwischen humanen endogenen Retroviren (HERV) und der Pluripotenz von CSCs zu erfassen.

Mehrere Studien berichteten bereits über eine Hochregulation der HERV Familie H (HERV-H) in embryonalen und induzierten pluripotenten Stammzellen sowie in CRC-Tumorproben und abgeleiteten Zelllinien. Interessanterweise wurde HERV-H nicht in normalen Geweben und anderen Krebsarten gefunden, was auf ein CRC-spezifisches Expressionsprofil hindeutet. Vor diesem Hintergrund war das primäre Ziel der Arbeit, eine Korrelation zwischen zuvor identifizierten CRC-HERV-H-Loci und Stammzellgenen in CRC-Zelllinien zu untersuchen, um in einem zweiten Schritt den potenziellen HERV-H-Pluripotenz-Locus als Biomarker in CR-CSCs zu identifizieren. Erste Experimente in 56 CRC-Zelllinien ergaben positive Korrelationen zwischen *OCT4*, *NANOG* und *c-MYC* mit dem spezifischen HERV-Locus 1400035_h_gag. Die Korrelationsanalyse mit klinisch-pathologischen Merkmalen, einschließlich dem molekularem Subtyp, TNM, und Tumor Grading zeigte unterschiedliche Korrelationsmuster in jedem Subtyp. Die stärkste Korrelation bestand zwischen *OCT4* und *KLF4* in Mikrosatelliten-instabilen (MSI) Zelllinien, bei den anderen Subtypen wurde keine Korrelation beobachtet. Darüber hinaus waren die Korrelationen zwischen HERV-H-Loci und *NANOG, KLF4* sowie *c-MYC* in Zelllinien, die aus fortgeschrittenen Tumoren etabliert

wurden, stärker. In den von Metastasen abgeleiteten Zelllinien waren jedoch keine Korrelationen nachweisbar, was darauf hinweist, dass die HERV-H-Loci-Expression die Malignität fördert, jedoch eine untergeordnete Rolle bei der eigentlichen Metastasierung spielt.

Ein Vergleich der Genexpressionsdaten zwischen 3D-kultivierten Kolonosphären und den dazugehörigen 2D-Monolayern ergab eine höhere Expression von CRC-HERV-H-Loci in den Ersteren. Hinsichtlich der Expressionsstärke der untersuchten CRC-HERV-H-Loci zeigte sich hingegen ein inverses Bild mit geringer Expression in der Seitenpopulation (SP), welche typischerweise als CSCs bezeichnet werden. Interessanterweise war die *NANOG*-Expression in SP-Zellen ebenfalls niedrig, während Kolonosphären *NANOG* überexprimierten. Im Gegensatz dazu waren *KLF4* und *c-MYC* in den meisten Kolonosphären herunterreguliert, was im Gegensatz zur Literatur steht. Dementsprechend ist die 3D-Sphäroidkultur eine praktische Methode zur Anreicherung von CSCs.

Zusammenfassend konnte in dieser Arbeit gezeigt werden, dass die HERV-H-Expression in CR-CSCs hauptsächlich durch *NANOG* beeinflusst wurde. Obgleich kein exakter Locus für die Pluripotenz gefunden wurde, konnten das *gag*-Gen auf Chromosom 14 und das *env*-Gen auf Chromosom 13 als wichtigste Kandidaten identifiziert werden, die in weiteren Untersuchungen näher charakterisiert werden sollten.

8. References

1. Jiang S, Good D, Wei MQ. Vaccinations for colorectal cancer: Progress, strategies, and novel adjuvants. Int J Mol Sci. 2019;20(14).

2. Paschke S, Jafarov S, Staib L, Kreuser ED, Maulbecker-Armstrong C, Roitman M, et al. Are colon and rectal cancer two different tumor entities? A proposal to abandon the term colorectal cancer. Int J Mol Sci. 2018;19(9):1–24.

3. Simon K. Colorectal cancer development and advances in screening. Clin Interv Aging. 2016;11:967–76.

4. Fleming M, Ravula S, Tatishchev SF, Wang HL. Colorectal carcinoma: Pathologic aspects. J Gastrointest Oncol. 2012;3(3):153–73.

5. Cuffy M, Abir F, Longo WE. Management of less common tumors of the colon, rectum, and anus. Vol. 5, Clinical Colorectal Cancer. 2006. p. 327–37.

6. McGregor LM, von Wagner C, Vart G, Yuen WC, Raine R, Wardle J, et al. The impact of supplementary narrative-based information on colorectal cancer screening beliefs and intention. BMC Cancer. 2015;15(1):1–9.

7. Zeuner A, Todaro M, Stassi G, De Maria R. Colorectal cancer stem cells: From the crypt to the clinic. Cell Stem Cell. 2014.

8. Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: Incidence, mortality, survival, and risk factors. Prz Gastroenterol. 2019;14(2):89–103.

9. Fanali C, Lucchetti D, Farina M, Corbi M, Cufino V, Cittadini A, et al. Cancer stem cells in colorectal cancer from pathogenesis to therapy: Controversies and perspectives. World J Gastroenterol. 2014;

10. WHO [Internet]. 2018. Available from: https://www.who.int/news-room/fact-sheets/detail/cancer

11. Ralf-Dieter Hofheinz , Dirk Arnold , Markus Borner , Gunnar Folprecht , Ullrich Graeven , Holger Hebart , Susanna Hegewisch-Becker , Volker Heinemann , Thomas Meybier , Ron Pritzkuleit , Werner Scheithauer , Josef Thaler , Jürgen Weitz BW. Kolonkarzinom [Internet]. 2018. Available from: https://www.onkopedia.com/de/onkopedia/guidelines/kolonkarzinom/@@guideline/html/index.html

12. Thanikachalam K, Khan G. Colorectal cancer and nutrition. Nutrients. 2019;11(1).

13. Cancer.org. Survival Rates for Colorectal Cancer [Internet]. Cancer. 1996. p. 1254–62. Available from: https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/survival-rates.html

14. Brenner H, Kloor M, Pox CP. Colorectal cancer. In: The Lancet. 2014.

15. De Rosa M, Pace U, Rega D, Costabile V, Duraturo F, Izzo P, et al. Genetics, diagnosis

and management of colorectal cancer (Review). Oncol Rep. 2015;34(3):1087–96.

16. Lorans M, Dow E, Macrae FA, Winship IM, Buchanan DD. Update on Hereditary Colorectal Cancer: Improving the Clinical Utility of Multigene Panel Testing. Clin Colorectal Cancer. 2018;17(2):e293–305.

17. Carethers JM, Stoffel EM. Lynch syndrome and Lynch syndrome mimics: The growing complex landscape of hereditary colon cancer. World J Gastroenterol. 2015;21(31):9253–61.

18. Biller LH, Syngal S, Yurgelun MB. Recent advances in Lynch syndrome. Fam Cancer. 2019;18(2):211–9.

19. Ibrahim A, Barnes DR, Dunlop J, Barrowdale D, Antoniou AC, Berg JN. Attenuated familial adenomatous polyposis manifests as autosomal dominant late-onset colorectal cancer. Eur J Hum Genet [Internet]. 2014;22(11):1330–3. Available from: http://dx.doi.org/10.1038/ejhg.2014.20

20. Nguyen HT, Duong HQ. The molecular characteristics of colorectal cancer: Implications for diagnosis and therapy (review). Oncol Lett. 2018;16(1):9–18.

21. Rao C V., Yamada HY. Genomic instability and colon carcinogenesis: From the perspective of genes. Front Oncol. 2013;3 MAY(May):1–13.

22. Armaghany T, Wilson JD, Chu Q, Mills G. Genetic alterations in colorectal cancer. Gastrointestinal Cancer Research. 2012.

23. Fearon EF, Vogelstein B. for Colorectal Tumorigenesis. 1989;61:759–67.

24. Lochhead P, Chan AT, Giovannucci E, Fuchs CS, Wu K, Nishihara R, et al. Progress and opportunities in molecular pathological epidemiology of colorectal premalignant lesions. Am J Gastroenterol. 2014;

25. Stracci F, Zorzi M, Grazzini G. Colorectal Cancer Screening: Tests, Strategies, and Perspectives. Front Public Heal. 2014;2(October):1–9.

26. American Cancer Society. Colorectal Cancer Facts & Figures 2020-2022. Atlanta: Am Cancer Soc Inc. 2020;(page 32).

27. Markowitz SD, Bertagnolli MM. Molecular basis of colorectal cancer. N Engl J Med. 2009;

28. Pino MS, Chung DC. NIH Public Access THE CHROMOSOMAL INSTABILITY PATHWAY IN COLON. Gastroenterol 138(6) 2059–2072. 2010;138(6):2059–72.

29. Ogino S, Goel A. Molecular classification and correlates in colorectal cancer. J Mol Diagnostics [Internet]. 2008;10(1):13–27. Available from: http://dx.doi.org/10.2353/jmoldx.2008.070082

30. Yamagishi H, Kuroda H, Imai Y, Hiraishi H. Molecular pathogenesis of sporadic colorectal cancers. Chin J Cancer. 2016;35(1):1–8.

31. Blanco-Calvo M, Concha Á, Figueroa A, Garrido F, Valladares-Ayerbes M. Colorectal

cancer classification and cell heterogeneity: A systems oncology approach. Int J Mol Sci. 2015;16(6):13610–32.

32. Birgisson H, Edlund K, Wallin U, Påhlman L, Kultima HG, Mayrhofer M, et al. Microsatellite instability and mutations in BRAF and KRAS are significant predictors of disseminated disease in colon cancer. BMC Cancer. 2015;15(1):1–11.

33. Sinicrope FA, Sargent DJ. Molecular pathways: Microsatellite instability in colorectal cancer: Prognostic, predictive, and therapeutic implications. Clin Cancer Res. 2012;

34. Schmoll HJ, Van cutsem E, Stein A, Valentini V, Glimelius B, Haustermans K, et al. Esmo consensus guidelines for management of patients with colon and rectal cancer. A personalized approach to clinical decision making. Ann Oncol [Internet]. 2012;23(10):2479–516. Available from: http://dx.doi.org/10.1093/annonc/mds236

35. Lan YT, Chang SC, Yang SH, Lin CC, Wang HS, Jiang JK, et al. Comparison of clinicopathological characteristics and prognosis between early and late recurrence after curative surgery for colorectal cancer. Am J Surg. 2014;

36. Reimers MS, Zeestraten ECM, Kuppen PJK, Liefers GJ, van de Velde CJH. Biomarkers in precision therapy in colorectal cancer. Gastroenterol Rep. 2013;1(3):166–83.

37. Naccarati A, Polakova V, Pardini B, Vodickova L, Hemminki K, Kumar R, et al. Mutations and polymorphisms in TP53 gene - An overview on the role in colorectal cancer. Mutagenesis. 2012;27(2):211–8.

38. Tsang AHF, Cheng KH, Wong ASP, Ng SSM, Ma BBY, Chan CML, et al. Current and future molecular diagnostics in colorectal cancer and colorectal adenoma. World J Gastroenterol. 2014;20(14):3847–57.

39. Phipps AI, Limburg PJ, Baron JA, Burnett-hartman AN, Weisenberger DJ, Laird PW, et al. Association Between Molecular Subtypes of Colorectal Cancer and Patient Survival. Gastroenterology. 2015;148(1):77–87.

40. Stintzing S. Recent advances in understanding colorectal cancer [version 1; peer review: 2 approved]. F1000Research. 2018;7(0):1–8.

41. Maletzki C, Huehns M, Knapp P, Waukosin N, Klar E, Prall F, et al. Functional characterization and drug response of freshly established patient-derived tumor models with cpg island methylator phenotype. PLoS One. 2015;10(11):1–14.

42. Boman BM, Huang E. Human colon cancer stem cells: A new paradigm in gastrointestinal oncology. Vol. 26, Journal of Clinical Oncology. 2008. p. 2828–38.

43. Susman S, Tomuleasa C, Soritau O, Mihu C, Rus-Ciuca D, Sabourin JC, et al. The colorectal cancer stem-like cell hypothesis: A pathologist's point of view. Vol. 17, Journal of B.U.ON. 2012. p. 230–6.

44. Yeung TM, Mortensen NJ. Colorectal cancer stem cells. Dis Colon Rectum. 2009;

45. Rocco A. Cancer stem cell hypothesis and gastric carcinogenesis: Experimental evidence and unsolved questions. World J Gastrointest Oncol. 2012;4(3):54.

46. Gudjonsson T, Magnusson MK. Stem cell biology and the cellular pathways of carcinogenesis. Apmis. 2005;113(11–12):922–9.

47. Alison MR, Lim SML, Nicholson LJ. Cancer stem cells: Problems for therapy? J Pathol. 2011;223(2):148–62.

48. Moltzahn F, Thalmann GN. Cancer stem cells in prostate cancer. Transl Androl Urol. 2013;2(3):242–53.

49. Song Y, Wang Y, Tong C, Xi H, Zhao X, Wang Y, et al. A unified model of the hierarchical and stochastic theories of gastric cancer. Br J Cancer [Internet]. 2017;116(8):973–89. Available from: http://dx.doi.org/10.1038/bjc.2017.54

50. Sell S. Stem cell origin of cancer and differentiation therapy. Crit Rev Oncol Hematol. 2004;51(1):1–28.

51. Kleinsmith LJ, Pierce GB. Multipotentiality of Single Embryonal Carcinoma Cells'. October. 1964;

52. Park CH, Bergsagel DE, McCulloch EA. Mouse myeloma tumor stem cells: A primary cell culture assay. J Natl Cancer Inst. 1971;46(2):411–22.

53. Dick JE, Bonnet D. Human Acute Myeloid Leukaemia is organised as a heirarchy that originates from a primitive haematopoetic cell. Nat Med [Internet]. 1997;3(7):730–7. Available from: http://www.nature.com/naturemedicine

54. Singh SK, Clarke ID, Terasaki M, Bonn VE, Hawkins C, Squire J, et al. Identification of a cancer stem cell in human brain tumors. Cancer Res. 2003;63(18):5821–8.

55. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. 2003;100(7):3983–8.

56. Bender Kim CF, Jackson EL, Woolfenden AE, Lawrence S, Babar I, Vogel S, et al. Identification of bronchioalveolar stem cells in normal lung and lung cancer. Cell. 2005;121(6):823–35.

57. Todaro M, Alea MP, Di Stefano AB, Cammareri P, Vermeulen L, Iovino F, et al. Colon Cancer Stem Cells Dictate Tumor Growth and Resist Cell Death by Production of Interleukin-4. Cell Stem Cell. 2007;1(4):389–402.

58. O'Brien CA, Pollett A, Gallinger S, Dick JE. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature. 2007;445(7123):106–10.

59. Ricci-Vitiani L, Lombardi DG, Pilozzi E, Biffoni M, Todaro M, Peschle C, et al. Identification and expansion of human colon-cancer-initiating cells. Nature. 2007;445(7123):111–5.

60. Bapat SA, Mali AM, Koppikar CB, Kurrey NK. Stem and progenitor-like cells contribute to the aggressive behavior of human epithelial ovarian cancer. Cancer Res. 2005;65(8):3025–9.

61.	Li C, Heidt DG, Dalerba P, Burant CF, Zhang L, Adsay V, et al. Identification of pancreatic cancer stem cells. Cancer Res. 2007;67(3):1030–7.

62.	Volkmer JP, Sahoo D, Chin RK, Ho PL, Tang C, Kurtova A V., et al. Three differentiation states risk-stratify bladder cancer into distinct subtypes. Proc Natl Acad Sci U S A. 2012;109(6):2078–83.

63.	Fang D, Nguyen TK, Leishear K, Finko R, Kulp AN, Hotz S, et al. A tumorigenic subpopulation with stem cell properties in melanomas. Cancer Res. 2005;65(20):9328–37.

64.	Suetsugu A, Nagaki M, Aoki H, Motohashi T, Kunisada T, Moriwaki H. Characterization of CD133+ hepatocellular carcinoma cells as cancer stem/progenitor cells. Biochem Biophys Res Commun. 2006;351(4):820–4.

65.	Richardson GD, Robson CN, Lang SH, Neal DE, Maitland NJ, Collins AT. CD133, a novel marker for human prostatic epithelial stem cells. J Cell Sci. 2004;117(16):3539–45.

66.	Collins AT, Berry PA, Hyde C, Stower MJ, Maitland NJ. Prospective identification of tumorigenic prostate cancer stem cells. Cancer Res. 2005;65(23):10946–51.

67.	Abetov D, Mustapova Z, Saliev T, Bulanin D. Biomarkers and signaling pathways of colorectal cancer stem cells. Tumor Biology. 2015.

68.	Neethan A. Lobo YS, Dalong Qian and MFC. The biology of cancer stem cells. Annu Rev Cell Dev Biol. 2007;675–99.

69.	Hermann PC, Huber SL, Herrler T, Aicher A, Ellwart JW, Guba M, et al. Distinct Populations of Cancer Stem Cells Determine Tumor Growth and Metastatic Activity in Human Pancreatic Cancer. Cell Stem Cell. 2007;1(3):313–23.

70.	Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. Nature. 2013;501(7467):328–37.

71.	Vermeulen L, Snippert HJ. Stem cell dynamics in homeostasis and cancer of the intestine. Nat Rev Cancer. 2014;14(7):468–80.

72.	Macarthur BD, Lemischka IR. XStatistical mechanics of pluripotency. Cell [Internet]. 2013;154(3):484–9. Available from: http://dx.doi.org/10.1016/j.cell.2013.07.024

73.	Phi LTH, Sari IN, Yang YG, Lee SH, Jun N, Kim KS, et al. Cancer stem cells (CSCs) in drug resistance and their therapeutic implications in cancer treatment. Stem Cells Int. 2018;2018.

74.	Fábián Á, Barok M, Vereb G, Szöllosi J. Die hard: Are cancer stem cells the bruce willises of tumor biology? Cytom Part A. 2009;75(1):67–74.

75.	Alisi A, Cho WC, Locatelli F, Fruci D. Multidrug resistance and cancer stem cells in neuroblastoma and hepatoblastoma. Int J Mol Sci. 2013;14(12):24706–25.

76.	Goodell MA, Brose K, Paradis G, Conner AS, Mulligan RC. Isolation and functional

properties of murine hematopoietic stem cells that are replicating in vivo. J Exp Med. 1996;183(4):1797–806.

77.   Zhou S, Schuetz JD, Bunting KD, Colapietro AM, Sampath J, Morris JJ, et al. The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype. Nat Med. 2001;7(9):1028–34.

78.   Wang T, Shigdar S, Gantier MP, Hou Y, Wang L, Shamaileh H Al, et al. Cancer stem cell targeted therapy : progress amid controversies A BRIEF VIEW OF ANTICANCER. Oncotarget. 2015;6(42):44191–206.

79.   Fletcher JI, Haber M, Henderson MJ, Norris MD. ABC transporters in cancer: More than just drug efflux pumps. Nat Rev Cancer. 2010;10(2):147–56.

80.   Guo P, Wang J, Gao W, Liu X, Wu S, Wan B, et al. Salvianolic acid B reverses multidrug resistance in nude mice bearing human colon cancer stem cells. Mol Med Rep. 2018;18(2):1323–34.

81.   Yuan Z, Liang X, Zhan Y, Wang Z, Xu J, Qiu Y, et al. Targeting CD133 reverses drug-resistance via the AKT/NF-κB/MDR1 pathway in colorectal cancer. Vol. 122, British Journal of Cancer. 2020. p. 1342–53.

82.   Ma L, Liu T, Jin Y, Wei J, Yang Y, Zhang H. ABCG2 is required for self-renewal and chemoresistance of CD133-positive human colorectal cancer cells. Tumor Biol [Internet]. 2016;37(9):12889–96. Available from: http://dx.doi.org/10.1007/s13277-016-5209-5

83.   Jiang Y, Gao H, Liu M, Mao Q. Sorting and biological characteristics analysis for side population cells in human primary hepatocellular carcinoma. Am J Cancer Res. 2016;6(9):1890–905.

84.   Leon G, MacDonagh L, Finn SP, Cuffe S, Barr MP. Cancer stem cells in drug resistant lung cancer: Targeting cell surface markers and signaling pathways. Pharmacol Ther [Internet]. 2016;158:71–90. Available from: http://dx.doi.org/10.1016/j.pharmthera.2015.12.001

85.   Liu BJ, Xu QY, Yu WD, Li N, Yao T, Zhao LJ, et al. Study of the Characterization of Side Population Cells in Endometrial Cancer Cell Lines: Chemoresistance, Progestin Resistance, and Radioresistance. Front Med. 2020;7(March):1–9.

86.   Glinsky G V. Viruses, stemness, embryogenesis, and cancer: A miracle leap toward molecular definition of novel oncotargets for therapy-resistant malignant tumors? Oncoscience. 2015;2(9):751–4.

87.   Wang J, Xie G, Singh M, Ghanbarian AT, Raskó T, Szvetnik A, et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature. 2014;516(7531):405–9.

88.   Lu X, Sachs F, Ramsay LA, Jacques PÉ, Göke J, Bourque G, et al. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nat

Struct Mol Biol. 2014;21(4):423–5.

89.   Breitenbach M, Hoffmann J. Editorial: Cancer models. Front Oncol. 2018;8(OCT):1–4.

90.   Franco SS, Szczesna K, Iliou MS, Al-Qahtani M, Mobasheri A, Kobolák J, et al. In vitro models of cancer stem cells and clinical applications. BMC Cancer [Internet]. 2016;16(Suppl 2). Available from: http://dx.doi.org/10.1186/s12885-016-2774-3

91.   Balani S, Nguyen L V., Eaves CJ. Modeling the process of human tumorigenesis. Nat Commun [Internet]. 2017;8(May):1–10. Available from: http://dx.doi.org/10.1038/ncomms15422

92.   Kondo T, Setoguchi T, Taga T. Persistence of a small subpopulation of cancer stem-like cells in the C6 glioma cell line. Proc Natl Acad Sci U S A. 2004;101(3):781–6.

93.   Essex A, Pineda J, Acharya G, Xin H, Evans J, Iorns E, et al. Replication study: Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. Elife. 2019;8(2010):1–18.

94.   De Sousa E Melo F, Kurtova A V., Harnoss JM, Kljavin N, Hoeck JD, Hung J, et al. A distinct role for Lgr5 + stem cells in primary and metastatic colon cancer. Nature [Internet]. 2017;543(7647):676–80. Available from: http://dx.doi.org/10.1038/nature21713

95.   Shimokawa M, Ohta Y, Nishikori S, Matano M, Takano A, Fujii M, et al. Visualization and targeting of LGR5 + human colon cancer stem cells. Nature [Internet]. 2017;545(7653):187–92. Available from: http://dx.doi.org/10.1038/nature22081

96.   Oshima N, Yamada Y, Nagayama S, Kawada K, Hasegawa S, Okabe H, et al. Induction of cancer stem cell properties in colon cancer cells by defined factors. PLoS One. 2014;9(7).

97.   Chen S, Song X, Chen Z, Li X, Li M, Liu H, et al. CD133 Expression and the Prognosis of Colorectal Cancer: A Systematic Review and Meta-Analysis. PLoS One. 2013;8(2):1–9.

98.   Kemper K, Sprick MR, De Bree M, Scopelliti A, Vermeulen L, Hoek M, et al. The AC133 epitope, but not the CD133 protein, is lost upon cancer stem cell differentiation. Cancer Res. 2010;70(2):719–29.

99.   Horst D, Scheel SK, Liebmann S, Neumann J, Maatz S, Kirchner T, et al. The cancer stem cell marker CD133 has high prognostic impact but unknown functional relevance for the metastasis of human colon cancer. J Pathol. 2009;219(4):427–34.

100.   Choi D, Lee HW, Hur KY, Kim JJ, Park GS, Jang SH, et al. Cancer stem cell markers CD133 and CD24 correlate with invasiveness and differentiation in colorectal adenocarinoma. World J Gastroenterol. 2009;15(18):2258–64.

101.   Huang EH, Hynes MJ, Zhang T, Ginestier C, Dontu G, Appelman H, et al. Aldehyde Dehydrogenase 1 Is a Marker for Normal and Malignant Human Colonic Stem Cells (SC) and Tracks SC Overpopulation during Colon Tumorigenesis. Cancer Res

[Internet]. 2009 Apr 7;69(8):3382–9. Available from: http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-08-4418

102. Du L, Wang H, He L, Zhang J, Ni B, Wang X, et al. CD44 is of functional importance for colorectal cancer stem cells. Clin Cancer Res. 2008;14(21):6751–60.

103. Chu P, Clanton DJ, Snipas TS, Julia L, Mitchell E, Nguyen ML, et al. Characterization of a subpopulation of colon cancer cells with stem cell-like properties. Int J Cancer. 2009;124(6):1312–21.

104. Griffiths DJ. Endogenous retroviruses in the human genome sequence. Genome Biol. 2001;2(6):1–5.

105. Flockerzi A, Burkhardt S, Schempp W, Meese E, Mayer J. Human Endogenous Retrovirus HERV-K14 Families: Status, Variants, Evolution, and Mobilization of Other Cellular Sequences. J Virol. 2005;79(5):2941–9.

106. Johnson WE. Origins and evolutionary consequences of ancient endogenous retroviruses. Nat Rev Microbiol. 2019;17(6):355–70.

107. Yves Stauffer, Gregory Theiler, Peter Sperisen, Yuri Lebedev and CVJ. Digital expression profiles of human endogenous retroviral families in normal and cancerous tissues. Cancer Immun. 2004;4:2.

108. Tristem M. Identification and Characterization of Novel Human Endogenous Retrovirus Families by Phylogenetic Screening of the Human Genome Mapping Project Database. J Virol. 2000;74(8):3715–30.

109. Belshaw R, Katzourakis A, Pačes J, Burt A, Tristem M. High copy number in human endogenous retrovirus families is associated with copying mechanisms in addition to reinfection. Mol Biol Evol. 2005;22(4):814–7.

110. Patzke S, Lindeskog M, Munthe E, Aasheim HC. Characterization of a novel human endogenous retrovirus, HERV-H/F, expressed in human leukemia cell lines. Virology. 2002;303(1):164–73.

111. Liang Q, Ding J, Xu R, Xu Z, Zheng S. Identification of a novel human endogenous retrovirus and promoter activity of its 5′ U3. Biochem Biophys Res Commun [Internet]. 2009 May;382(2):468–72. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0006291X09005245

112. Kim TH, Jeon YJ, Yi JM, Kim DS, Huh JW, Hur CG, et al. The distribution and expression of HERV families in the human genome. Mol Cells. 2004;18(1):87–93.

113. Zhang M, Liang JQ, Zheng S. Expressional activation and functional roles of human endogenous retroviruses in cancers. Rev Med Virol. 2019;29(2):1–11.

114. Alcazer V, Bonaventura P, Depil S. Human endogenous retroviruses (HERVs): Shaping the innate immune response in cancers. Cancers (Basel). 2020;12(3):1–12.

115. Fu B, Ma H, Liu D. Endogenous Retroviruses Function as Gene Expression Regulatory Elements During Mammalian Pre-implantation Embryo Development. Int J Mol Sci

[Internet]. 2019 Feb 12;20(3):790. Available from: http://www.mdpi.com/1422-0067/20/3/790

116. Meyer TJ, Rosenkrantz JL, Carbone L, Chavez SL. Endogenous retroviruses: With us and against us. Front Chem. 2017;5(APR):1–8.

117. Dolci M, Favero C, Tarantini L, Villani S, Bregni M, Signorini L, et al. Human endogenous retroviruses env gene expression and long terminal repeat methylation in colorectal cancer patients. Med Microbiol Immunol [Internet]. 2020;209(2):189–99. Available from: https://doi.org/10.1007/s00430-020-00662-6

118. Mallet F, Bouton O, Prudhomme S, Cheynet V, Oriol G, Bonnaud B, et al. The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology. Proc Natl Acad Sci U S A. 2004;101(6):1731–6.

119. Schiavetti F, Thonnard J, Colau D, Boon T, Coulie PG. A human endogenous retroviral sequence encoding an antigen recognized on melanoma by cytolytic T lymphocytes. Cancer Res. 2002;62(19):5510–6.

120. Büscher K, Hahn S, Hofmann M, Trefzer U, Özel M, Sterry W, et al. Expression of the human endogenous retrovirus-K transmembrane envelope, Rec and Np9 proteins in melanomas and melanoma cell lines. Vol. 16, Melanoma Research. 2006. p. 223–34.

121. Contreras-Galindo R, Kaplan MH, Leissner P, Verjat T, Ferlenghi I, Bagnoli F, et al. Human Endogenous Retrovirus K (HML-2) Elements in the Plasma of People with Lymphoma and Breast Cancer. J Virol. 2008;82(19):9329–36.

122. Lamprecht B, Bonifer C, Mathas S. Repeat-element driven activation of proto-oncogenes in human malignancies. Cell Cycle. 2010;9(21):4276–81.

123. Wang-Johanning F, Frost AR, Jian B, Epp L, Lu DW, Johanning GL. Quantitation of HERV-K env gene expression and splicing in human breast cancer. Oncogene. 2003;22(10):1528–35.

124. Wang-Johanning F, Liu J, Rycaj K, Huang M, Tsai K, Rosen DG, et al. Expression of multiple human endogenous retrovirus surface envelope proteins in ovarian cancer. Int J Cancer. 2007;120(1):81–90.

125. Wang-Johanning F, Frost AR, Jian B, Azuero R, Lu DW, Chen DT, et al. Detecting the expression of human endogenous retrovirus E envelope transcripts in human prostate adenocarcinoma. Cancer. 2003;98(1):187–97.

126. Gimenez J, Montgiraud C, Pichon JP, Bonnaud B, Arsac M, Ruel K, et al. Custom human endogenous retroviruses dedicated microarray identifies self-induced HERV-W family elements reactivated in testicular cancer upon methylation control. Nucleic Acids Res. 2010;38(7):2229–46.

127. Ahn K, Kim HS. Structural and quantitative expression analyses of HERV gene family in human tissues. Vol. 28, Molecules and Cells. 2009. p. 99–103.

128. Wentzensen N, Wilz B, Findeisen P, Wagner R, Dippold W, von Knebel Doeberitz M,

et al. Identification of differentially expressed genes in colorectal adenoma compared to normal tissue by suppression subtractive hybridization. Int J Oncol. 2004;24(4):987–94.

129. Wentzensen N, Coy JF, Knaebel HP, Linnebacher M, Wilz B, Gebert J, et al. Expression of an endogenous retroviral sequence from the HERV-H group in gastrointestinal cancers. Int J Cancer. 2007;121(7):1417–23.

130. Pérot P, Mugnier N, Montgiraud C, Gimenez J, Jaillard M, Bonnaud B, et al. Microarray-based sketches of the HERV transcriptome landscape. PLoS One. 2012;7(6).

131. Pérot P, Mullins CS, Naville M, Bressan C, Hühns M, Gock M, et al. Expression of young HERV-H loci in the course of colorectal carcinoma and correlation with molecular subtypes. Oncotarget. 2015;6(37):40095–111.

132. Mullins CS, Linnebacher M. Human endogenous retroviruses and cancer: Causality and therapeutic possibilities. World J Gastroenterol. 2012;18(42):6027–35.

133. Gonzalez-Cao M, Iduma P, Karachaliou N, Santarpia M, Blanco J, Rosell R. Human endogenous retroviruses and cancer. Cancer Biol Med. 2016;13(4):483–8.

134. Huang G, Li Z, Wan X, Wang Y, Dong J. Human endogenous retroviral K element encodes fusogenic activity in melanoma cells. J Carcinog. 2013;12:1–15.

135. Ramsay LA, Marchetto MC, Caron M, Chen SH, Busche S, Kwan T, et al. Conserved expression of transposon-derived non-coding transcripts in primate stem cells. BMC Genomics. 2017;18(1):1–13.

136. De Parseval N, Casella JF, Gressin L, Heidmann T. Characterization of the three HERV-H proviruses with an open envelope reading frame encompassing the immunosuppressive domain and evolutionary history in primates. Vol. 279, Virology. 2001. p. 558–69.

137. Mager DL, Freeman JD. HERV-H Endogenous Retroviruses: Presence in the New World Branch but Amplification in the Old World Primate Lineage. Virology. 1995;213(2):395–404.

138. Izsvák Z, Wang J, Singh M, Mager DL, Hurst LD. Pluripotency and the endogenous retrovirus HERVH: Conflict or serendipity? BioEssays. 2016;38(1):109–17.

139. Römer C, Singh M, Hurst LD, Izsvák Z. How to tame an endogenous retrovirus: HERVH and the evolution of human pluripotency. Curr Opin Virol. 2017;25:49–58.

140. Santoni FA, Guerra J, Luban J. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. Retrovirology. 2012;9:1–15.

141. Yi JM, Kim HM, Kim HS. Human endogenous retrovirus HERV-H family in human tissues and cancer cells: Expression, identification, and phylogeny. Cancer Lett. 2006;231(2):228–39.

142. Alves PMS, Lévy N, Stevenson BJ, Bouzourene H, Theiler G, Bricard G, et al. Identification of tumor-associated antigens by large-scale analysis of genes expressed in

human colorectal cancer. Cancer Immun. 2008;8(May):1–11.

143. Mullins CS, Linnebacher M. Endogenous retrovirus sequences as a novel class of tumor-specific antigens: An example of HERV-H env encoding strong CTL epitopes. Cancer Immunol Immunother. 2012;61(7):1093–100.

144. Mullins CS, Hühns M, Krohn M, Peters S, Cheynet V, Oriol G, et al. Generation, characterization and application of antibodies directed against HERV-H Gag protein in colorectal samples. PLoS One. 2016;11(4):1–16.

145. Finicelli M, Benedetti G, Squillaro T, Pistilli B, Marcellusi A, Mariani P, et al. Expression of stemness genes in primary breast cancer tissues: The role of SOX2 as a prognostic marker for detection of early recurrence. Oncotarget. 2014;

146. Ortega-Paino E, Fransson J, Ek S, Borrebaeck CAK. Functionally associated targets in mantle cell lymphoma as defined by DNA microarrays and RNA interference. Blood. 2008;111(3):1617–24.

147. Sipos F, Constantinovits M, Valcz G, Tulassay Z, Muzes G. Association of hepatocyte-derived growth factor receptor/caudal type homeobox 2 co-expression with mucosal regeneration in active ulcerative colitis. World J Gastroenterol. 2015;

148. Mullins CS, Micheel B, Matschos S, Leuchter M, Bürtin F, Krohn M, et al. Integrated biobanking and tumor model establishment of human colorectal carcinoma provides excellent tools for preclinical research. Cancers (Basel). 2019;11(10):1–18.

149. Linnebacher M, Maletzki C, Ostwald C, Klier U, Krohn M, Klar E, et al. Cryopreservation of human colorectal carcinomas prior to xenografting. BMC Cancer. 2010;10.

150. Maletzki C, Stier S, Gruenert U, Gock M, Ostwald C, Prall F, et al. Establishment, Characterization and Chemosensitivity of Three Mismatch Repair Deficient Cell Lines from Sporadic and Inherited Colorectal Carcinomas. PLoS One. 2012;7(12).

151. Maletzki C, Gock M, Randow M, Klar E, Huehns M, Prall F, et al. Establishment and characterization of cell lines from chromosomal instable colorectal cancer. World J Gastroenterol. 2015;21(1):164–76.

152. Kuehn F, Mullins CS, Krohn M, Harnack C, Ramer R, Krämer OH, et al. Establishment and characterization of HROC69 - A Crohn's related colonic carcinoma cell line and its matched patient-derived xenograft. Sci Rep [Internet]. 2016;6(March):1–10. Available from: http://dx.doi.org/10.1038/srep24671

153. Oddo D, Sennott EM, Barault L, Valtorta E, Arena S, Cassingena A, et al. Molecular landscape of acquired resistance to targeted therapy combinations in BRAF-mutant colorectal cancer. Cancer Res. 2016;

154. Medico E, Russo M, Picco G, Cancelliere C, Valtorta E, Corti G, et al. The molecular landscape of colorectal cancer cell lines unveils clinically actionable kinase targets. Nat Commun. 2015;6:1–10.

155. Lazzari L, Corti G, Picco G, Isella C, Montone M, Arcela P, et al. Patient-derived xenografts and matched cell lines identify pharmacogenomic vulnerabilities in colorectal cancer. Clin Cancer Res. 2019;

156. Hughes P, Marshall D, Reid Y, Parkes H, Gelber C. The costs of using unauthenticated, over-passaged cell lines: How much more data do we need? Biotechniques. 2007;43(5):575–86.

157. Kaur G, Dufour JM. Cell lines: Valuable tools or useless artifacts. Spermatogenesis [Internet]. 2012;2(1):1–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22553484%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3341241

158. Mouriaux F, Zaniolo K, Bergeron MA, Weidmann C, De La Fouchardière A, Fournier F, et al. Effects of long-term serial passaging on the characteristics and properties of cell lines derived from uveal melanoma primary tumors. Investig Ophthalmol Vis Sci. 2016;57(13):5288–301.

159. Saji Joseph J, Tebogo Malindisa S, Ntwasa M. Two-Dimensional (2D) and Three-Dimensional (3D) Cell Culturing in Drug Discovery. In: Cell Culture [Internet]. IntechOpen; 2019. p. 13. Available from: https://www.intechopen.com/books/cell-culture/two-dimensional-2d-and-three-dimensional-3d-cell-culturing-in-drug-discovery

160. Edmondson R, Broglie JJ, Adcock AF, Yang L. Three-dimensional cell culture systems and their applications in drug discovery and cell-based biosensors. Assay Drug Dev Technol. 2014;12(4):207–18.

161. Van De Wetering M, Francies HE, Francis JM, Bounova G, Iorio F, Pronk A, et al. Prospective derivation of a living organoid biobank of colorectal cancer patients. Cell [Internet]. 2015;161(4):933–45. Available from: http://dx.doi.org/10.1016/j.cell.2015.03.053

162. Flow Cytometry Fundamental Principle [Internet]. Available from: https://www.bosterbio.com/protocol-and-troubleshooting/flow-cytometry-principle

163. Sales-Pardo I, Avendaño A, Martinez-Muñoz V, García-Escarp M, Celis R, Whittle P, et al. Flow cytometry of the Side Population: Tips & tricks. Cell Oncol. 2006;28(1–2):37–53.

164. Golebiewska A, Brons NHC, Bjerkvig R, Niclou SP. Critical appraisal of the side population assay in stem cell and cancer stem cell research. Cell Stem Cell [Internet]. 2011;8(2):136–47. Available from: http://dx.doi.org/10.1016/j.stem.2011.01.007

165. Lucena-Aguilar G, Sánchez-López AM, Barberán-Aceituno C, Carrillo-Ávila JA, López-Guerrero JA, Aguilar-Quesada R. DNA Source Selection for Downstream Applications Based on DNA Quality Indicators Analysis. Biopreserv Biobank. 2016;14(4):264–70.

166. Kim J, Lim J, Lee C. Quantitative real-time PCR approaches for microbial community

studies in wastewater treatment systems: Applications and considerations. Biotechnol Adv [Internet]. 2013;31(8):1358–73. Available from: http://dx.doi.org/10.1016/j.biotechadv.2013.05.010

167. Stricker S, Pollard S. Reprogramming cancer cells to pluripotency. Epigenetics. 2014;9(6):798–802.

168. Kim J, Zaret KS. Reprogramming of human cancer cells to pluripotency for models of cancer progression. EMBO J. 2015;34(6):739–47.

169. Li P, Wang S, Zhan L, He X, Chi G, Lv S, et al. Efficient feeder cells preparation system for large-scale preparation and application of induced pluripotent stem cells. Sci Rep [Internet]. 2017;7(1):1–15. Available from: http://dx.doi.org/10.1038/s41598-017-10428-5

170. Roy A, Krzykwa E, Lemieux R, Néron S. Increased efficiency of γ-irradiated versus mitomycin C-treated feeder cells for the expansion of normal human cells in long-term cultures. J Hematotherapy Stem Cell Res. 2001;10(6):873–80.

171. Somers A, Jean JC, Sommer CA, Omari A, Ford CC, Mills JA, et al. Generation of transgene-free lung disease-specific human induced pluripotent stem cells using a single excisable lentiviral stem cell cassette. Stem Cells. 2010;28(10):1728–40.

172. Peterson BG, Carl P, Kris Boudt , Ross Bennett, Joshua Ulrich, Eric Zivot, Dries Cornilly, Eric Hung, Matthieu Lestel, Kyle Balkissoon, Diethelm Wuertz, Anthony Alexander Christidis, R. Douglas Martin, Zeheng "Zenith" Zhou JMS. Package 'PerformanceAnalytics.' 2020.

173. Wu W, Cao J, Ji Z, Wang J, Jiang T, Ding H. Co-expression of Lgr5 and CXCR4 characterizes cancer stem-like cells of colorectal cancer. Oncotarget. 2016;7(49):81144–55.

174. Liang Q, Xu Z, Xu R, Wu L, Zheng S. Expression patterns of non-coding spliced transcripts from human endogenous retrovirus HERV-H elements in colon cancer. PLoS One. 2012;7(1).

175. Roschke A V., Glebov OK, Lababidi S, Gehlhaus KS, Weinstein JN, Kirsch IR. Chromosomal instability is associated with higher expression of genes implicated in epithelial-mesenchymal transition, cancer invasiveness, and metastasis and with lower expression of genes involved in cell cycle checkpoints, DNA repair, and chromatin mai. Neoplasia. 2008;10(11):1222–30.

176. Vargas-Rondón N, Villegas VE, Rondón-Lagos M. The role of chromosomal instability in cancer and therapeutic responses. Cancers (Basel). 2018;10(1):1–21.

177. Rangel N, Forero-Castro M, Rondón-Lagos M. New insights in the cytogenetic practice: Karyotypic chaos, non-clonal chromosomal alterations and chromosomal instability in human cancer and therapy response. Genes (Basel). 2017;8(6):2–29.

178. Jacques PÉ, Jeyakani J, Bourque G. The Majority of Primate-Specific Regulatory

Sequences Are Derived from Transposable Elements. PLoS Genet. 2013;9(5).

179. Bourque G, Leong B, Vega VB, Chen X, Yen LL, Srinivasan KG, et al. Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res. 2008;18(11):1752–62.

180. Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, et al. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. Proc Natl Acad Sci U S A. 2007;104(47):18613–8.

181. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science (80-). 2016;351(6277):1083–7.

182. Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, et al. Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet. 2010;42(7):631–4.

183. Ohnuki M, Tanabe K, Sutou K, Teramoto I, Sawamura Y, Narita M, et al. Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. Proc Natl Acad Sci U S A. 2014;111(34):12426–31.

184. Tsai M-C, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, et al. Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. Science (80-) [Internet]. 2010 Aug 6;329(5992):689–93. Available from: https://www.sciencemag.org/lookup/doi/10.1126/science.1192002

185. Kelley D, Rinn J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol. 2012;13(11):R107.

186. Zhao L, Zhao Y, Schwarz B, Mysliwietz J, Hartig R, Camaj P, et al. Verapamil inhibits tumor progression of chemotherapyresistant pancreatic cancer side population cells. Int J Oncol. 2016;49(1):99–110.

187. von Furstenberg RJ, Buczacki SJA, Smith BJ, Seiler KM, Winton DJ, Henning SJ. Side population sorting separates subfractions of cycling and non-cycling intestinal stem cells. Stem Cell Res. 2014;12(2):364–75.

188. Patrawala L, Calhoun T, Schneider-Broussard R, Zhou J, Claypool K, Tang DG. Side population is enriched in tumorigenic, stem-like cancer cells, whereas ABCG2+ and ABCG2- cancer cells are similarly tumorigenic. Cancer Res. 2005;65(14):6207–19.

189. Hirschmann-Jax C, Foster AE, Wulf GG, Nuchtern JG, Jax TW, Gobel U, et al. A distinct "side population" of cells with high drug efflux capacity in human tumor cells. Proc Natl Acad Sci U S A. 2004;101(39):14228–33.

190. Wu C, Alman BA. Side population cells in human cancers. Cancer Lett. 2008;268(1):1–9.

191. Candeil L, Gourdier I, Peyron D, Vezzio N, Copois V, Bibeau F, et al. ABCG2 overexpression in colon cancer cells resistant to SN38 and in irinotecan-treated metastases. Int J Cancer. 2004;

192. Gupta N, Martin PM, Miyauchi S, Ananth S, Herdman A V., Martindale RG, et al. Down-regulation of BCRP/ABCG2 in colorectal and cervical cancer. Biochem Biophys Res Commun. 2006;343(2):571–7.

193. Andersen V, Vogel LK, Kopp TI, Sæbø M, Nonboe AW, Hamfjord J, et al. High ABCC2 and low ABCG2 gene expression are early events in the colorectal adenoma-carcinoma sequence. PLoS One. 2015;

194. Liu HG, Pan YF, You J, Wang OC, Huang K Te, Zhang XH. Expression of ABCG2 and its significance in colorectal cancer. Asian Pacific J Cancer Prev. 2010;

195. Wang X, Xia B, Liang Y, Peng L, Wang Z, Zhuo J, et al. Membranous ABCG2 expression in colorectal cancer independently correlates with shortened patient survival. Cancer Biomarkers. 2013;

196. Giampieri R, Scartozzi M, Loretelli C, Piva F, Mandolesi A, Lezoche G, et al. Cancer Stem Cell Gene Profile as Predictor of Relapse in High Risk Stage II and Stage III, Radically Resected Colon Cancer Patients. PLoS One. 2013;8(9):1–6.

197. Kang D, Park JM, Jung CK, Lee BI, Oh ST, Choi MG. Prognostic impact of membranous ATP-binding cassette Sub-family G member 2 expression in patients with colorectal carcinoma after surgical resection. Cancer Biol Ther. 2015;

198. Ma XL, Sun YF, Wang BL, Shen MN, Zhou Y, Chen JW, et al. Sphere-forming culture enriches liver cancer stem cells and reveals Stearoyl-CoA desaturase 1 as a potential therapeutic target. BMC Cancer. 2019;19(1):1–12.

199. Pastrana E, Silva-Vargas V, Doetsch F. Eyes wide open: A critical review of sphere-formation as an assay for stem cells. Cell Stem Cell [Internet]. 2011;8(5):486–98. Available from: http://dx.doi.org/10.1016/j.stem.2011.04.007

200. Thurber AE, Douglas G, Sturm EC, Zabierowski SE, Smit DJ, Ramakrishnan SN, et al. Inverse expression states of the BRN2 and MITF transcription factors in melanoma spheres and tumour xenografts regulate the NOTCH pathway. Oncogene. 2011;30(27):3036–48.

201. Zhang Y, Xu W, Guo H, Zhang Y, He Y, Lee SH, et al. NOTCH1 signaling regulates self-renewal and platinum chemoresistance of cancer stem-like cells in human non-small cell lung cancer. Cancer Res. 2017;

202. Liu C, Liu R, Zhang D, Deng Q, Liu B, Chao HP, et al. MicroRNA-141 suppresses prostate cancer stem cells and metastasis by targeting a cohort of pro-metastasis genes. Nat Commun. 2017;

203. Wang J, Zhang B, Wu H, Cai J, Sui X, Wang Y, et al. CD51 correlates with the TGF-beta pathway and is a functional marker for colorectal cancer stem cells. Oncogene. 2017;

204. Keysar SB, Le PN, Miller B, Jackson BC, Eagles JR, Nieto C, et al. Regulation of head and neck squamous cancer stem cells by PI3K and SOX2. J Natl Cancer Inst. 2017;

205. Bahmad HF, Cheaito K, Chalhoub RM, Hadadeh O, Monzer A, Ballout F, et al. Sphere-Formation Assay: Three-dimensional in vitro culturing of prostate cancer stem/Progenitor sphere-forming cells. Front Oncol. 2018;8(AUG):1–14.

206. Pereira LP, Silva P, Duarte M, Rodrigues L, Duarte CMM, Albuquerque C, et al. Targeting colorectal cancer proliferation, stemness and metastatic potential using Brassicaceae extracts enriched in isothiocyanates: A 3D cell model-based study. Nutrients. 2017;

207. Jeppesen M, Hagel G, Glenthoj A, Vainer B, Ibsen P, Harling H, et al. Short-term spheroid culture of primary colorectal cancer cells as an in vitro model for personalizing cancer medicine. Wu M-H, editor. PLoS One [Internet]. 2017 Sep 6;12(9):e0183074. Available from: https://dx.plos.org/10.1371/journal.pone.0183074

208. De Angelis ML, Bruselles A, Francescangeli F, Pucilli F, Vitale S, Zeuner A, et al. Colorectal cancer spheroid biobanks: Multi-level approaches to drug sensitivity studies. Cell Biol Toxicol. 2018;34(6):459–69.

209. Shaheen S, Ahmed M, Lorenzi F, Nateri AS. Spheroid-Formation (Colonosphere) Assay for in Vitro Assessment and Expansion of Stem Cells in Colon Cancer. Stem Cell Rev Reports [Internet]. 2016;12(4):492–9. Available from: http://dx.doi.org/10.1007/s12015-016-9664-6

210. Gao W, Wu D, Wang Y, Wang Z, Zou C, Dai Y, et al. Development of a novel and economical agar-based non-adherent three-dimensional culture method for enrichment of cancer stem-like cells. Stem Cell Res Ther. 2018;9(1):243.

211. Diensthuber M, Oshima K, Heller S. Stem/progenitor cells derived from the cochlear sensory epithelium give rise to spheres with distinct morphologies and features. JARO - J Assoc Res Otolaryngol. 2009;10(2):173–90.

212. Wang S, Kanojia D, Lo P, Chandrashekaran, Varun Duan X, Berger FG, Wang Q, et al. Enrichment and Selective Targeting of Cancer Stem Cells in Colorectal Cancer Cell Lines. Hum Genet Embryol [Internet]. 2013;01(S2). Available from: https://www.omicsonline.org/enrichment-and-selective-targeting-of-cancer-stem-cells-in-colorectal-cancer-cell-lines-2161-0436.S2-006.php?aid=5623

213. Zhang J, Espinoza LA, Kinders RJ, Lawrence SM, Pfister TD, Zhou M, et al. NANOG modulates stemness in human colorectal cancer. Oncogene [Internet]. 2013 Sep 22;32(37):4397–405. Available from: http://www.nature.com/articles/onc2012461

214. Ying J, Tsujii M, Kondo J, Hayashi Y, Kato M, Akasaka T, et al. The effectiveness of an anti-human IL-6 receptor monoclonal antibody combined with chemotherapy to target colon cancer stem-like cells. Int J Oncol. 2015;46(4):1551–9.

215. Lin H, Wang B, Yu J, Wang J, Li Q, Cao B. Protein arginine methyltransferase 8 gene enhances the colon cancer stem cell (CSC) function by upregulating the pluripotency transcription factor. J Cancer. 2018;9(8):1394–402.

216. Gleizes A, Cavaillès V, Lapierre M. Transcriptional Regulation of the Intestinal Cancer

Stem Cell Phenotype We are IntechOpen , the world ' s leading publisher of Open Access books Built by scientists , for scientists. 2019;

217. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell. 2007;

218. Ghaleb AM, Yang VW. Krüppel-like factor 4 (KLF4): What we currently know. Gene. 2017;611:27–37.

219. Leng Z, Li Y, Zhou G, Lv X, Ai W, Li J, et al. Krüppel-like factor 4 regulates stemness and mesenchymal properties of colorectal cancer stem cells through the TGF-β1/Smad/snail pathway. J Cell Mol Med. 2020;

220. Murgai M, Ju W, Eason M, Kline J, Beury DW, Kaczanowska S, et al. KLF4-dependent perivascular cell plasticity mediates pre-metastatic niche formation and metastasis. Nat Med. 2017;

221. Yu F, Li J, Chen H, Fu J, Ray S, Huang S, et al. Kruppel-like factor 4 (KLF4) is required for maintenance of breast cancer stem cells and for cell migration and invasion. Oncogene. 2011;

222. Leng Z, Tao K, Xia Q, Tan J, Yue Z, Chen J, et al. Krüppel-Like Factor 4 Acts as an Oncogene in Colon Cancer Stem Cell-Enriched Spheroid Cells. PLoS One. 2013;8(2).

223. Wei D, Kanai M, Huang S, Xie K. Emerging role of KLF4 in human gastrointestinal cancer. Carcinogenesis. 2006;27(1):23–31.

224. Li X, Zhao Z, Zhang X, Yang S, Lin X, Yang X, et al. Klf 4 reduces stemness phenotype triggers mesenchymal-epithelial transition (MET)-like molecular changes, and prevents tumor progression in nasopharygeal carcinoma. Oncotarget. 2017;8(55):93924–41.

225. Karagonlar ZF, Akbari S, Karabicici M, Sahin E, Avci ST, Ersoy N, et al. A Novel Function for KLF4 in Modulating the De-differentiation of EpCAM-/CD133- nonStem Cells into EpCAM+/CD133+ Liver Cancer Stem Cells in HCC Cell Line HuH7. Cells. 2020;9(5).

226. Zhao W, Hisamuddin IM, Nandan MO, Babbin BA, Lamb NE, Yang VW. Identification of Krüppel-like factor 4 as a potential tumor suppressor gene in colorectal cancer. Oncogene. 2004;

227. Patel N V., Ghaleb AM, Nandan MO, Yang VW. Expression of the tumor suppressor Krüppel-like factor 4 as a prognostic predictor for colon cancer. Cancer Epidemiol Biomarkers Prev. 2010;

228. Wei D, Gong W, Kanai M, Schlunk C, Wang L, Yao JC, et al. Drastic down-regulation of krüppel-like factor 4 expression is critical in human gastric cancer development and progression. Cancer Res. 2005;65(7):2746–54.

229. Dang DT, Chen X, Feng J, Torbenson M, Dang LH, Yang VW. Overexpression of Krüppel-like factor 4 in the human colon cancer cell line RKO leads to reduced tumorigenecity. Oncogene [Internet]. 2003 May 29;22(22):3424–30. Available from:

http://www.nature.com/articles/1206413

230. Shie JL, Chen ZY, O'Brien MJ, Pestell RG, Lee ME, Tseng CC. Role of gut-enriched Kruppel-like factor in colonic cell growth and differentiation. Am J Physiol - Gastrointest Liver Physiol. 2000;279(4 42-4):806–14.

231. Xu J, Lü B, Xu F, Gu H, Fang Y, Huang Q, et al. Dynamic down-regulation of Krüppel-like factor 4 in colorectal adenoma-carcinoma sequence. J Cancer Res Clin Oncol. 2008;134(8):891–8.

232. Lee H young, Ahn JB, Rha SY, Chung HC, Park KH, Kim TS, et al. High KLF4 level in normal tissue predicts poor survival in colorectal cancer patients. World J Surg Oncol. 2014;12(1):1–7.

233. Stankevicius V, Kunigenas L, Stankunas E, Kuodyte K, Strainiene E, Cicenas J, et al. The expression of cancer stem cell markers in human colorectal carcinoma cells in a microenvironment dependent manner. Biochem Biophys Res Commun [Internet]. 2017;484(4):726–33. Available from: http://dx.doi.org/10.1016/j.bbrc.2017.01.111

234. Hu R, Zuo Y, Zuo L, Liu C, Zhang S, Wu Q, et al. KLF4 expression correlates with the degree of differentiation in colorectal cancer. Gut Liver. 2011;5(2):154–9.

235. Zhang H Le, Wang P, Lu MZ, Zhang SD, Zheng L. c-Myc maintains the self-renewal and chemoresistance properties of colon cancer stem cells. Oncol Lett. 2019;17(5):4487–93.

236. Gabay M, Li Y, Felsher DW. MYC Activation Is a Hallmark of Cancer Initiation and Maintenance. Cold Spring Harb Perspect Med [Internet]. 2014 Jun 1;4(6):a014241–a014241. Available from: http://perspectivesinmedicine.cshlp.org/lookup/doi/10.1101/cshperspect.a014241

237. Baudino TA, McKay C, Pendeville-Samain H, Nilsson JA, Maclean KH, White EL, et al. c-Myc is essential for vasculogenesis and angiogenesis during development and tumor progression. Genes Dev. 2002;

238. Brandvold KA, Neiman P, Ruddell A. Angiogenesis is an early event in the generation of myc-induced lymphomas. Oncogene. 2000;

239. Wang T, Cai B, Ding M, Su Z, Liu Y, Shen L. c-Myc Overexpression Promotes Oral Cancer Cell Proliferation and Migration by Enhancing Glutaminase and Glutamine Synthetase Activity. Am J Med Sci. 2019;

240. Wolfer A, Ramaswamy S. MYC and metastasis. Cancer Research. 2011.

241. Topper MJ, Vaz M, Chiappinelli KB, DeStefano Shields CE, Niknafs N, Yen RWC, et al. Epigenetic Therapy Ties MYC Depletion to Reversing Immune Evasion and Treating Lung Cancer. Cell. 2017;

242. Elbadawy M, Usui T, Yamawaki H, Sasaki K. Emerging roles of C-myc in cancer stem cell-related signaling and resistance to cancer chemotherapy: A potential therapeutic target against colorectal cancer. Int J Mol Sci. 2019;20(9).

243. Poli V, Fagnocchi L, Fasciani A, Cherubini A, Mazzoleni S, Ferrillo S, et al. MYC-driven epigenetic reprogramming favors the onset of tumorigenesis by inducing a stem cell-like state. Nat Commun [Internet]. 2018;9(1). Available from: http://dx.doi.org/10.1038/s41467-018-03264-2

244. Akita H, Marquardt JU, Durkin ME, Kitade M, Seo D, Conner EA, et al. MYC activates stem-like cell potential in hepatocarcinoma by a p53-dependent mechanism. Cancer Res. 2014;74(20):5903–13.

245. Kim MJ, Xia B, Suh HN, Lee SH, Jun S, Lien EM, et al. PAF-Myc-Controlled Cell Stemness Is Required for Intestinal Regeneration and Tumorigenesis. Dev Cell [Internet]. 2018 Mar;44(5):582-596.e4. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1534580718301059

246. Murphy DJ, Junttila MR, Pouyet L, Karnezis A, Shchors K, Bui DA, et al. Distinct Thresholds Govern Myc's Biological Output In Vivo. Cancer Cell. 2008;

247. Schuhmacher M, Eick D. Dose-dependent regulation of target gene expression and cell proliferation by c-Myc levels. Transcription. 2013;4(4):192–7.

248. Okuyama H, Endo H, Akashika T, Kato K, Inoue M. Downregulation of c-MYC protein levels contributes to cancer cell survival under dual deficiency of oxygen and glucose. Cancer Res. 2010;70(24):10213–23.

249. Yuneva M, Zamboni N, Oefner P, Sachidanandam R, Lazebnik Y. Deficiency in glutamine but not glucose induces MYC-dependent apoptosis in human cells. J Cell Biol. 2007;

250. Askew DS, Ashmun RA, Simmons BC, Cleveland JL. Constitutive c-myc expression in an IL-3-dependent myeloid cell line suppresses cell cycle arrest and accelerates apoptosis. Oncogene. 1991;

251. Wong WJ, Qiu B, Nakazawa MS, Qing G, Simon MC. MYC Degradation under Low O 2 Tension Promotes Survival by Evading Hypoxia-Induced Cell Death . Mol Cell Biol. 2013;33(17):3494–504.

252. Meng HM, Zheng P, Wang XY, Liu C, Sui HM, Wu SJ, et al. Overexpression of nanog predicts tumor progression and poor prognosis in colorectal cancer. Cancer Biol Ther. 2010;9(4):295–302.

253. Ishiguro T, Sato A, Ohata H, Sakai H, Nakagama H, Okamoto K. Differential expression of nanog1 and nanogp8 in colon cancer cells. Biochem Biophys Res Commun [Internet]. 2012;418(2):199–204. Available from: http://dx.doi.org/10.1016/j.bbrc.2011.10.123

254. Wang H, Liu B, Wang J, Li J, Gong Y, Li S, et al. Reduction of NANOG Mediates the Inhibitory Effect of Aspirin on Tumor Growth and Stemness in Colorectal Cancer. Cell Physiol Biochem. 2017;44(3):1051–63.

255. Tamura S, Isobe T, Ariyama H, Nakano M, Kikushige Y, Takaishi S, et al. E-cadherin regulates proliferation of colorectal cancer stem cells through NANOG. Oncol Rep.

2018;40(2):693–703.

256. Was H, Czarnecka J, Kominek A, Barszcz K, Bernas T, Piwocka K, et al. Some chemotherapeutics-treated colon cancer cells display a specific phenotype being a combination of stem-like and senescent cell features. Cancer Biol Ther [Internet]. 2018;19(1):63–75. Available from: https://doi.org/10.1080/15384047.2017.1385675

257. Ibrahim EE, Babaei-Jadidi R, Saadeddin A, Spencer-Dene B, Hossaini S, Abuzinadah M, et al. Embryonic NANOG activity defines colorectal cancer stem cells and modulates through AP1- and TCF-dependent mechanisms. Stem Cells. 2012;30(10):2076–87.

258. Lin CW, Liao MY, Lin WW, Wang YP, Lu TY, Wu HC. Epithelial cell adhesion molecule regulates tumor initiation and tumorigenesis via activating reprogramming factors and epithelial-mesenchymal transition gene expression in colon cancer. J Biol Chem. 2012;

259. Noh KH, Kim BW, Song KH, Cho H, Lee YH, Kim JH, et al. Nanog signaling in cancer promotes stem-like phenotype and immune evasion. J Clin Invest. 2012;

260. Haraguchi N, Utsunomiya T, Inoue H, Tanaka F, Mimori K, Barnard GF, et al. Characterization of a Side Population of Cancer Cells from Human Gastrointestinal System. Stem Cells. 2006;24(3):506–13.

261. Hu L, McArthur C, Jaffe RB. Ovarian cancer stem-like side-population cells are tumourigenic and chemoresistant. Br J Cancer. 2010;102(8):1276–83.

262. Jeter CR, Liu B, Liu X, Chen X, Liu C, Calhoun-Davis T, et al. NANOG promotes cancer stem cell characteristics and prostate cancer resistance to androgen deprivation. Oncogene [Internet]. 2011 Sep 18;30(36):3833–45. Available from: http://www.nature.com/articles/onc2011114

263. Jeter CR, Badeaux M, Choy G, Chandra D, Patrawala L, Liu C, et al. Functional Evidence that the Self-Renewal Gene NANOG Regulates Human Tumor Development. Stem Cells [Internet]. 2009 May;27(5):993–1005. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf

264. Wang X, Liu Q, Hou B, Zhang W, Yan M, Jia H, et al. Concomitant targeting of multiple key transcription factors effectively disrupts cancer stem cells enriched in side population of human pancreatic cancer cells. PLoS One. 2013;8(9):1–17.

265. Kawai N, Id YH, Ebihara Y, Saito T, Murai A, Saito T, et al. ABCG2 expression is related to low 5-ALA photodynamic diagnosis (PDD) efficacy and cancer stem cell phenotype , and suppression of ABCG2 improves the efficacy of PDD. 2019;1–15.

266. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: The rosetta stone of a hidden RNA language? Cell [Internet]. 2011;146(3):353–8. Available from: http://dx.doi.org/10.1016/j.cell.2011.07.014

267. Van Schaijik B, Davis PF, Wickremesekera AC, Tan ST, Itinteang T. Subcellular localisation of the stem cell markers OCT4, SOX2, NANOG, KLF4 and c-MYC in

cancer: A review. J Clin Pathol. 2018;71(1):88–91.

268. Smyrek I, Stelzer EHK. Quantitative three-dimensional evaluation of immunofluorescence staining for large whole mount spheroids with light sheet microscopy. Biomed Opt Express. 2017;8(2):484.

269. Giovinazzo A, Balestrieri E, Petrone V, Argaw-Denboba A, Cipriani C, Miele MT, et al. The Concomitant Expression of Human Endogenous Retroviruses and Embryonic Genes in Cancer Cells under Microenvironmental Changes is a Potential Target for Antiretroviral Drugs. Cancer Microenviron [Internet]. 2019 Dec 5;12(2–3):105–18. Available from: http://link.springer.com/10.1007/s12307-019-00231-3

270. Kim JJ. Applications of iPSCs in cancer research. Biomark Insights. 2015;2015:125–31.

271. Gong L, Yan Q, Zhang Y, Fang X, Liu B, Guan X. Cancer cell reprogramming: A promising therapy converting malignancy to benignity. Cancer Commun [Internet]. 2019;39(1):1–13. Available from: https://doi.org/10.1186/s40880-019-0393-5

272. Carette JE, Pruszak J, Varadarajan M, Blomen VA, Gokhale S, Camargo FD, et al. Generation of iPSCs from cultured human malignant cells. Blood. 2010;115(20):4039–42.

273. Kumano K, Arai S, Hosoi M, Taoka K, Takayama N, Otsu M, et al. Generation of induced pluripotent stem cells from primary chronic myelogenous leukemia patient samples. Blood [Internet]. 2012 Jun 28;119(26):6234–42. Available from: https://ashpublications.org/blood/article/119/26/6234/105596/Generation-of-induced-pluripotent-stem-cells-from

274. Hu K, Yu J, Suknuntha K, Tian S, Montgomery K, Choi KD, et al. Efficient generation of transgene-free induced pluripotent stem cells from normal and neoplastic bone marrow and cord blood mononuclear cells. Blood. 2011;117(14):109–20.

275. Gandre-Babbe S, Paluru P, Aribeana C, Chou ST, Bresolin S, Lu L, et al. Patient-derived induced pluripotent stem cells recapitulate hematopoietic abnormalities of juvenile myelomonocytic leukemia. Blood. 2013;121(24):4925–9.

276. Kim J, Hoffman JP, Alpaugh RK, Rhim AD, Reichert M, Stanger BZ, et al. An iPSC Line from Human Pancreatic Ductal Adenocarcinoma Undergoes Early to Invasive Stages of Pancreatic Cancer Progression. Cell Rep [Internet]. 2013 Jun;3(6):2088–99. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2211124713002696

277. Zhang X, Cruz FD, Terry M, Remotti F, Matushansky I. Terminal differentiation and loss of tumorigenicity of human cancers via pluripotency-based reprogramming. Oncogene [Internet]. 2013 May 9;32(18):2249–60. Available from: http://www.nature.com/articles/onc2012237

278. Miyoshi N, Ishii H, Nagai KI, Hoshino H, Mimori K, Tanaka F, et al. Defined factors induce reprogramming of gastrointestinal cancer cells. Proc Natl Acad Sci U S A. 2010;107(1):40–5.

279. Lin SL, Chang DC, Chang-Lin S, Lin CH, Wu DTS, Chen DT, et al. Mir-302 reprograms human skin cancer cells into a pluripotent ES-cell-like state. Rna. 2008;14(10):2115–24.

280. Utikal J, Maherali N, Kulalert W, Hochedlinger K. Sox2 is dispensable for the reprogramming of melanocytes and melanoma cells into induced pluripotent stem cells. J Cell Sci. 2009;122(19):3502–10.

281. Moore JB, Loeb DM, Hong KU, Sorensen PH, Triche TJ, Lee DW, et al. Epigenetic reprogramming and re-differentiation of a Ewing sarcoma cell line. Front Cell Dev Biol. 2015;3(MAR):1–13.

282. Bang JS, Choi NY, Lee M, Ko K, Park YS, Ko K. Reprogramming of cancer cells into induced pluripotent stem cells questioned. Int J Stem Cells. 2019;12(3):430–9.

283. Iskender B, Izgi K, Canatan H. Reprogramming bladder cancer cells for studying cancer initiation and progression. Tumor Biol [Internet]. 2016;37(10):13237–45. Available from: http://dx.doi.org/10.1007/s13277-016-5226-4

284. Vatanmakanian M, Yousefi H, Mashouri L, Aref AR, Khamisipour G, Bitaraf A, et al. Generation of Induced Pluripotent Cancer Cells from Glioblastoma Multiform Cell Lines. Cell Reprogram. 2019;

285. Corominas-Faja B, Cufí S, Oliveras-Ferraros C, Cuyàs E, López-Bonet E, Lupu R, et al. Nuclear reprogramming of luminal-like breast cancer cells generates Sox2-overexpressing cancer stem-like cellular states harboring transcriptional activation of the mTOR pathway. Cell Cycle. 2013;12(18):3109–24.

286. Stricker SH, Feber A, Engström PG, Carén H, Kurian KM, Takashima Y, et al. Widespread resetting of DNA methylation in glioblastoma-initiating cells suppresses malignant cellular behavior in a lineage-dependent manner. Genes Dev. 2013;27(6):654–69.

9. Appendix

9.1. Abbreviations

Abbreviation	Meaning
ABC	ATP-binding cassette
ABCB1	ATP binding cassette subfamily B member 1
ABCG2	ATP-binding cassette transporter G2
ALDH1	Aldehyde dehydrogenase 1
APC	Adenomatous polyposis coli
BAX	Bcl-2-associated X protein
Bcl-2	B-cell lymphoma 2
Bcl-xL	B-cell lymphoma-extra large
bFGF	Basic fibroblast growth factors
bp	Base pair
cDNA	Complementary DNA
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CO$_2$	Carbon dioxide
CRC	Colorectal carcinomas
CSCs	Cancer stem cells
Ct	Threshold cycle
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DPBS	Dulbecco's phosphate buffered saline
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EGF	Epidermal growth factor
EMT	Epithelial-mesenchymal transition
env	Envelope
ERVs	Endogenous retroviruses
ESCs	Embryonic stem cells

EtBr	Ethidiumbromid
FAP	Familial adenomatous polyposis
FBS	Fetal bovine serum
g	Gravity
gag	Group antigens
GAPDH	Glyceraldehyde 3 phosphate dehydrogenase
gDNA	Genomic DNA
kb	Kilo-base pair
HBSS	Hanks' Balanced Salt Solution
HEPES	4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid
HERVs	Human endogenous retroviruses
HNPCC	Hereditary nonpolyposis colorectal cancer
HROC	Hansestadt Rostock colorectal cancer
IL-4	Interleukin 4
iPSCs	Induced pluripotent stem cells
KLF4	Kruppel-like factor
KRAS	Kirsten rat sarcoma
LBP9	Lipopolysaccharide binding protein 9
LGR5	Leucine-rich repeat-containing G-protein coupled receptor 5
lncRNAs	Long noncoding RNAs
LTR	Long terminal repeat
MDR	Multidrug resistance
MEF	Mouse embryonic fibroblast
MET	Mesenchymal-epithelial transition
MFI	Mean fluorescence intensity
min	Minute
mM	millimolar
MMR	Mismatch repair
mRNA	Messenger RNA
MSI	Microsatellite instability
MSI-H	MSI-High
MSS	Microsatellite stable
NANOG	Nanog homeobox

NCCAs	Nonclonal chromosomal alterations
NEAAs	Non-essential amino acids
ng	nanogram
nm	nanometre
NTC	Non-template control
OCT4	Octamer-binding transcription factor 4
ORF	Open reading frame
PCR	polymerase chain reaction
PDX	Patient-derived xenograft
PI	Propidium iodide
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
pMol	picomole
Pol	Polymerase
Pro	Protease
RNA	Ribonucleic acid
RT	Room temperature
RT-qPCR	Reverse transcription quantitative PCR
sec	Second
SEM	Standard error of the mean
SOX2	SRY-box 2
SP	Side population
SSCs	Somatic stem cells
SSEA-3	Stage specific embryonic antigen-3
TAE	Tris-Acetate-EDTA
TGF-β	Transforming growth factor beta
TP53	Tumor protein 53
WPRE	Woodchuck hepatitis virus post-transcriptional regulatory element
°C	Centigrade
2D	Two-dimensional
3D	Three-dimensional
β-ME	β-mercaptoethanol
μg	Microgram
μM	Micromolar

9.2. List of figures

9.3. List of tables

9.4. Supplementary information

Supplementary Table S1- List of CRC cell lines with information about molecular subtype, tumor site, primary or metastasis tumor type, TNM, and grading parameters.

Cell line name	Molecular type	Localization	T	N	M	G
HCT116	MSI	----	----	----	----	----
HROC18	MSS-spSTD	Cecum	T2	N0	M0	G2
HROC24	spMSI-H	Ascending colon	T2	N0	M0	G2
HROC24 T1 M1						

HROC32	MSS, spSTD	Ascending colon	T4	N2	M1	G2
HROC32 T3 M1						
HROC39	MSS, spSTD	Ascending colon	T4	N0	M0	G3
HROC39 T0 M2						
HROC40	MSS, CIMP-H	Descending colon	T3	N1	M0	G3
HROC43	MSS, CIMP-H	Ascending colon	T3	N2	M0	G3
HROC46 T0 M1	MSS, spSTD	Ascending colon	T3	N0	M1	G3
HROC50 T1 M5	spMSI-H	Ascending colon	T4	N0	M0	G2
HROC57	MSS	Ascending colon	T3	N2	M1	G3
HROC59 T1 M1	MSS, spSTD	Ascending colon	T3	N1	M1	G2
HROC60	MSS, CIMP-H	Ascending colon	T2	N0	M0	G2
HROC69	MSS, spSTD	Ascending colon	T3	N0	----	G3
HROC69 T0 M2						
HROC80 T1 M1	MSS, spSTD	Cecum	T3	N2	M0	G2
HROC87 T0 M2	spMSI-H	Ascending colon	T3	N0	M0	G3
HROC103Met	MSS, spSTD	Liver metastasis				
HROC107 cT0 M2	MSS, spSTD	Sigmoid	T3	N2	M1	G2
HROC111Met1 T0 M2	MSS, spSTD	Brain metastasis (Prim. Recto-sigmoid)	----	----	----	----
HROC112Met T0 M2	MSS, CIMP-H	Liver metastasis (Prim. Cecum)	----	----	----	----
HROC113 cT0 M1	MSI (Lynch)	Ascending colon	T4	N2	----	G3
HROC126	MSS, spSTD	Rectum	T3	N1	M0	G2
HROC131 T0 M3	spMSI-H	Ascending colon	T3	N1	M0	G3
HROC147Met	MSI-L, CIMP-L	Liver metastasis	----	----	----	----
HROC147 T0 M1		Recto-colon sigmoid	T3	N2	M1	G3
HROC173	MSS, spSTD	Ascending colon	T4	N2	M1	G3
HROC183	MSS, CIMP-H	Ascending colon	T3	N2	M0	G3
HROC183 T0 M2						
HROC212	spMSI-H	Cecum	T4	N2	M1	G3
HROC222 T1 M2	MSS, spSTD	Transverse colon	T3	N0	M0	G2
HROC239 T0 M1	MSS	Rectum	T4b	N2b	M0	G2
HROC257	spMSI-H	Ascending colon	T4	N2	M0	G3
HROC257 T0 M1						

HROC277 T0 M1	MSS, spSTD	Cecum	T4	N0	M1	G2
HROC277Met1 T0 M2		Liver metastasis	----	----	----	----
HROC277Met2		Liver metastasis	----	----	----	----
HROC278 T0 M1	MSI-L, CIMP-H	Ascending colon	T4	N2	M1	G3
HROC278Met T2 M2		Peritoneal metastasis	----	----	----	----
HROC284Met	MSS	Liver metastasis (Prim. Rectum)	----	----	----	----
HROC285 T0 M2	MSI-H (Lynch)	Descending colon	T4	N2	M1	G2
HROC296	----	Ascending colon	T3	N0	M0	G2
HROC300 T2 M1	spMSI-H	Rectum	T4	N1	M1	G2
HROC309	----	Descending colon	T3	N0	M0	G2
HROC313Met1 T0 M2	MSS	----	----	----	----	----
HROC324	MSI-H (Lynch)	Cecum	T3	N2	M0	G3
HROC334	MSS	Cecum	T3	N0	M0	G2
HROC348Met	----	Liver metastasis	----	----	----	----
HROC357	MSS	Transverse colon	T2	N0	M0	G3
HROC364	----	Hepatic flexure	T3	N0	M0	G2
HROC370	spMSI-H	----	T2	N0	M0	G2
HROC374	----	----	T3	N2b	M0	G3
HROC383	----	Transverse colon	T3	N0	M0	G3
HHC6548 T1 M1	MSI-H (Lynch)	Ascending colon	T3	N2		G3
HROBMC01	MSS, spSTD	Brain metastasis (Prim. Colon)	----	----	----	----

Supplementary Table S2- Summary of successful sorting results including number and purity of the SP and non-SP fractions in 17 HROC cell lines.

Cell lines	SP		non-SP	
	Events number	Purity %	Events number	Purity %
HROC18				
	2×10^5	99	3.6×10^5	99
	5.5×10^5	97.1	3.2×10^5	98.5

	9.6×10^5	98.5	5.4×10^5	99.8
	4.3×10^5	97.8	7.7×10^5	99.5
HROC24				
	3.4×10^5	80	1.2×10^6	92
	2.6×10^5	91.1	2.6×10^6	98,5
	2.2×10^5	86	5.4×10^5	99.5
	4.1×10^5	96	9.1×10^5	95.1
HROC40				
	2.4×10^5	90	3.4×10^5	93.5
	3.4×10^5	88.4	8.6×10^5	94.7
	6.9×10^5	97.5	4.8×10^5	98
	9.1×10^5	93	1.1×10^6	99.8
HROC46 T0 M1				
	3.9×10^5	85	3.9×10^5	99
	3.6×10^5	96.5	3.4×10^5	95.5
	6×10^5	96.7	2.7×10^5	92.1
	4.9×10^5	99.6	2.1×10^5	95.2
HROC50 T1 M5				
	1.1×10^5	88	2.5×10^5	99
	1.6×10^5	97.3	2.2×10^5	95.8
	1.7×10^5	88.3	3.2×10^5	96
	1.4×10^5	90.5	4.1×10^5	98.4
HROC87 T0 M2				
	1.6×10^5	98	3.9×10^5	92.1
	1×10^5	92.5	3.8×10^5	95.5
	2.8×10^5	98.7	1.4×10^6	93.9
	3×10^5	93.2	8.1×10^5	92.7
	3.1×10^5	88.8	7.3×10^5	95.8
HROC147 T0 M1				
	1.2×10^5	97	4.2×10^5	84.3
	2.1×10^5	88.5	5.8×10^5	95.2
	5.6×10^5	93	1.1×10^6	93.2
HROC278 T0 M1				
	5.5×10^5	91.2	7.5×10^5	98.4
	3.6×10^5	88.1	4.8×10^5	99.8
	3.8×10^5	95.5	7.1×10^5	97.7
HROC278Met T2 M2				
	4.7×10^5	91	1.1×10^6	93.3

	1.9×10^5	92.1	9.3×10^5	92.5
	8.5×10^5	93.7	1.5×10^6	95.6
HROC285 T0 M2				
	1.4×10^5	93.5	2.7×10^5	92.3
	3.1×10^5	94.9	4.1×10^5	85.9
	2.4×10^5	93.4	2.6×10^5	84.2
	2.4×10^5	91.4	7.2×10^5	89.8
HROC300 T2 M1				
	1.9×10^5	97.3	2.2×10^5	97.9
	4.7×10^5	94.5	3.5×10^5	99.3
	6.6×10^5	95.0	2.2×10^5	98.3
HROC309				
	3.8×10^5	96.5	5.3×10^5	97.8
	5.1×10^5	92.4	1.9×10^6	99.5
	5.1×10^5	94.6	7.2×10^5	98.8
HROC313Met1 T0 M2				
	2.5×10^5	88	7.7×10^5	90
	5.2×10^5	95.3	6.2×10^5	98.4
	1.5×10^6	97.2	1.3×10^6	95.1
	3×10^5	96.6	3.1×10^5	98.7
HROC324				
	2.4×10^5	85	1.1×10^6	97.9
	2.7×10^5	88	9.6×10^5	99.5
	3.7×10^5	89.8	1.4×10^6	94.4
HROC364				
	3.7×10^5	92.8	1.9×10^5	90.4
	6.4×10^5	98.2	5.5×10^5	93.5
	7.6×10^5	97.5	8.7×10^5	97.9
HROC383				
	3.3×10^5	92.1	4.6×10^5	100
	5.9×10^5	92.9	8.6×10^5	96.9
	2.5×10^5	95.7	3.5×10^5	94.3
HROBMC01				
	3.7×10^5	87.7	7×10^5	90
	1×10^5	90	2.9×10^5	94.4
	7×10^5	84.3	1.1×10^6	94.6
	4.4×10^5	84.8	1.5×10^6	98

10. Statutory declaration

I hereby officially declare that I have written this dissertation independently. Any help and assistance in creating this work are clearly indicated in the acknowledgments. In addition, I affirm that I have cited all publications and other sources used in the preparation of this academic work in the appropriate place. I further confirm that my work has been accomplished in accordance with the "Rules to ensure good scientific practice and to avoid scientific misconduct" of Rostock University Medical Center.

Ich versichere eidesstattlich durch eigenhändige Unterschrift, dass ich die Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen sind, habe ich als solche kenntlich gemacht.

Die Arbeit ist noch nicht veröffentlicht und ist in gleicher oder ähnlicher Weise noch nicht als Studienleistung zur Anerkennung oder Bewertung vorgelegt worden. Ich weiß, dass bei Abgabe einer falschen Versicherung die Prüfung als nicht bestanden zu gelten hat.

Rostock   10.2020   Fatemeh Shokraie

11. Acknowledgement

I would like to express my heartfelt gratitude and appreciation to Dr. Dietmar Zechner for his dedication and friendly supervision as well as providing me with the guidance and valuable advice I need to succeed in the PhD program.

I would like to deeply thank Dr. Claudia Maletzki, not only for supporting me with useful advices and showing me the new methods, but also for the valuable time she spent editing my dissertation. I cannot help but mention that I have learnt so much from her in both professional and personal spheres.

I would like to extend my sincere thanks to Prof. Carl Friedrich Classen for giving me the opportunity to work on his laboratory and for his continuous support in many occasions.

My big thanks go to Shaima Almansor for the great amount of assistance. She was always ready to help me and performed a large part of the qPCR experiments. She became my best friend in the lab and we had a wonderful time together.

I would like to gratefully acknowledge Sarah Fischer for her kind and amazing assistance in bioinformatic and statistical analysis of the Spearman correlation.

I also acknowledge the assistance of Mario Udinotti. He helped me with flow cytometry experiments during my pregnancy when I was helpless and not allowed to use chemical reagents. I enjoyed every minute of working with him and had so much fun and laughter.

Special thanks to Dr. Tomas Fiedler for encouraging and patiently supporting me during a very tough time.

Furthermore, I thank my friends and my officemates for standing by my side and brightening my days.

I very much appreciate the state scholarship program (Landesgraduiertenförderung) and Lieselotte-Beutel-Stiftung for their financial support.

Last but not least, I want to sincerely appreciate my parents, my dear husband, and my loving daughters who supported me with their love, endured this long procedure, listened to my worries when I was disappointed, and encouraged me to move on.

Thank you all for giving me many memorable moments.

12. Curriculum Vitae

Personal Information

Surname: Shokraie

First name: Fatemeh

Date of birth: 19/09/1982

Nationality: Iranian

Email: f.shokraie@gmail.com

Education/ Professional Qualifications

12/2020 – 05/2021	Medical Reporter in Covid team, CENTOGENE GmbH, Rostock, Germany.
10/2015 – 10/2020	PhD student in the Molecular Oncology and Immunotherapy Laboratory (AG MOI), Faculty of Medicine, University of Rostock, Rostock, Germany.
10/2017 – 09/2019	Research Assistant in the Children's and Youth Clinic (AG Hirntumorvakzine), Rostock University Medical Center, Rostock, Germany
06/2010 – 10/2012	Royan Institute for Stem Cell Biology and Technology, Tehran, Iran.
09/2009 – 10/2012	Master of Science in Cellular-Developmental Biology, Joint program between **Royan Institute** and University of Science and Culture, Tehran, Iran. Thesis title**: "**Histone Modification Analysis of *CDH1* Regulatory Region in Human Prostate Cancer Stem Cells, Prostate Cancer Cells and Embryonal Carcinoma Cells"
09/2004 – 01/2009	Bachelor of Science in Zoology **University of Shahid Beheshti**, Tehran, Iran. Thesis title: "Methods for the Ferrous and Sulfur-Oxidizing Acidophilic Bacteria Isolation".

Award

10/2016 – 01/2020   State Scholarship Program (Landesgraduiertenförderung), University of
Rostock, Rostock, Germany.

04/2018 – 01/2020   Grant by Lieselotte-Beutel-Stiftung

List of Publications

Articles

Riess C, Schneider B, Kehnscherper H, Gesche J, Irmscher N, **Shokraie F**, Classen C.F, Wirthgen E, Domańska G, Zimpfer A, Strüder D, Junghanss C, & Maletzki C. Activation of the Kynurenine Pathway in Human Malignancies Can Be Suppressed by the Cyclin-Dependent Kinase Inhibitor Dinaciclib. *Frontiers in Immunology 11* (**2020**).

Shokraie F, Moharrami M, Motamed N, Shahhoseini M, Totonchi M, Ezzatizadeh V, Firouzi J, Khosravani P, Ebrahimi M. Histone Modification Marks Strongly Regulate *CDH1* Promoter in Prostospheres as A Model of Prostate Cancer Stem Like Cells. *Cell J; 21(2).* (**2019**).

Riess C, **Shokraie F**, Classen C.F, Kreikemeyer B, Fiedler T, Junghanss C, & Maletzki C. Arginine-Depleting Enzymes – An Increasingly Recognized Treatment Strategy for Therapy-Refractory Malignancies. *Cellular Physiology and Biochemistry 51* (**2018**), 854 - 870.

Abstracts and Poster/Oral Presentation

Torres-Duran C, **Shokraie F**, Ruiz-Cabello F, Aptsiauri N, Linnebacher M, & Garrido F. Comparative Analysis of Cell Surface Expression of HLA-I in Colorectal Cancer (CRC) Cell Lines Derived from Primary Tumors, form Autologous Metastases and from Patient Derived Xenografts (PDX). *TIMO* **2020** (postponed to February 2021)

Torres-Duran C, Jimenez-Gamiz MP, **Shokraie F**, Munoz-Herrera CM, Ruiz-Cabello F, Aptsiauri N, Linnebacher M, & Garrido F. Heterozygosity of HLA-C Group Prevents Colorectal Cancer with LOH from Evading NK Cytotoxicity. 41 Congreso de la Sociedad Española de Inmunología, Seville, Spain, May **2019**

Torres-Duran C, Jimenez-Gamiz MP, **Shokraie F**, Munoz-Herrera CM, Ruiz-Cabello F, Aptsiauri N, Linnebacher M, & Garrido F. Comparative Analysis of Cell Surface Expression of HLA-I in Colorectal Cancer Cell Lines and corresponding Patient Derived Xenografts (PDX). 41 Congreso de la Sociedad Española de Inmunología, Seville, Spain, May **2019**

Torres-Duran C, Jimenez-Gamiz MP, **Shokraie F**, Munoz-Herrera CM, Ruiz-Cabello F, Aptsiauri N, Linnebacher M, & Garrido F. Analysis of HLA Class I Expression in Tumor Cells Lines Derived from Primary Colorectal Cancer and Autologous Metastases. 41 Congreso de la Sociedad Española de Inmunología, Seville, Spain, May **2019**

Shokraie F, Riess C, Fiedler T, Schneider B, Classen C.F, & Maletzki C. ARG-Auxotrophic Patient-Derived Glioblastoma Multiforme Cell Lines Show Increased Sensitivity Towards Doxycycline-Induced Growth Inhibition. *Neuro-oncology 20*, **2018**

Shokraie F, Moharrami M, Shahhoseini M, & Ebrahimi M. CDH1, CD44 and CD49b as Markers for Identification and Isolation of Prostate Cancer Stem-Like Cells. The 9th congress on Stem Cell Biology and Technology, Tehran, Iran, September **2013**

Moharrami M, Ebrahimi M, Motamed N, Totonchi M, Janzamin E, & **Shokraie F**. The Assessment of CDH1 Expression and Methylation in Human Prostate Cancer Stem- Like Cells. Cell Symposia: Hallmarks of Cancer, San Francisco, CA, USA, October **2012**

Shokraie F, Shahhoseini M, Moharrami M, Khosravani P, Firozi J, & Ebrahimi M. Is CD133 an Applicable Marker for the Isolation of Prostate Cancer Stem-Like Cells? The 8th Congress on Stem Cell Biology and Technology, Tehran, Iran, September **2012**

Moharrami M, Motamed N, **Shokraie F**, Janzamin E, Firozi J, & Ebrahimi M. Isolation and Characterization of Prostate Cancer Stem-Like Cells from PC3 Cell Line. The 8th Congress on Stem Cell Biology and Technology, Tehran, Iran, September **2012**