Genetic Heterogeneity of Cancer Stem Cells

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Jörg Otte
Aus Oberhausen

Düsseldorf, Dezember 2018

aus dem Institut für Stammzellforschung und Regenerative Medizin der
Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. James Adjaye

2. Prof. Dr. Constantin Czekelius

Tag der mündlichen Prüfung: 29.01.2019

Die messbare Seite der Welt ist nicht die Welt;
sie ist nur die messbare Seite der Welt.
Martin Seel, 2009

Meinen Eltern

HEINRICH-HEINE-UNIVERSITÄT DÜSSELDORF

Abstract

Doctor rerum naturalium

Genetic Heterogeneity of Cancer Stem Cells

by Jörg Otte

The nature of human stem cells varies enormously. Precious early embryonic stem cells, inheriting the greatest developmental potential, exist only transiently despite their inexhaustible ability to self-renew. Adult stem cells, on the other hand, have a highly restricted developmental potential but maintain our organs life-long. Mechanisms of cellular plasticity emerging during normal tissue regeneration are often hijacked by cancer cells conferring them stem cell-like qualities. These cancer stem cells (CSCs) are an enigmatic cellular specification often described as the root of a tumor's malignancy as they contribute to cancer relapse, progression and the metastatic process. One major aim of this thesis was to better understand the ongoing self-renewal of CSCs. We obtained colon cancer patient-derived organoids as newfangled *ex vivo* cell culture model. By adapting culture conditions, we selected for highly malignant colon cancer cells capable of organoid formation with low cell culture requirements, implying autonomous self-renewal capacities. However, we know from adult and embryonic stem cell culture that the fibroblast growth factor 2 (FGF2) is essential for maintaining self-renewal of stem cells. In view of this, we investigated the role of FGF2 in colorectal cancer biopsy-derived organoids identifying FGF2 as a central player in organoid formation and self-renewal of colon CSCs. Interestingly, when we inhibited FGF2-signaling, we not only observed an abrogated organoid formation and cellular differentiation, but also detected many molecular similarities with mechanisms of resistance against epidermal growth factor (EGF) receptor targeted therapies.

In another study, we were the first to describe the transcriptome of aggressive digital papillary adenocarcinomas (ADPA). In this comprehensive analysis, we detected a deregulated FGF-signaling pathway presumably contributing to the high malignancy of this rare tumor entity.

The heterogeneity of stem cells is not only important in tumor biology. During embryonic development, different cell types originate from one single cell, the zygote.

Within the first cell divisions, a period called pre-implantation development, embryonic genome activation and early cell lineage commitments define the onset of specification programs leading to the tremendous complexity of the human body. Detailed molecular analyses of these events require modern techniques to simultaneously analyze the genome, transcriptome and epigenome of single cells. In a state-of-the art review, we discuss new insights by modern methods and persisting challenges in single-cell analysis in pre-implantation embryos as well as in primordial germ cells. The described methods are just as relevant to understand the heterogeneity of CSCs, their plasticity and mechanisms of tumor progression.

Zusammenfassung

Humane Stammzellen zeichnen sich durch eine große Vielseitigkeit aus. Das größte Entwicklungspotenzial haben embryonale Stammzellen, welche sich theoretisch beliebig oft selbst erneuern können, jedoch nur kurzzeitig während der frühen Embryonalentwicklung existieren. Nach Abschluss der frühen Reifung sind adulte Stammzellen für die lebenslange Aufrechterhaltung und Regeneration von Geweben und Organen unverzichtbar. Sie besitzen lediglich ein begrenztes Potenzial sich in andere Zelltypen zu entwickeln, sind jedoch ebenfalls zu grenzenloser Selbsterneuerung fähig.
Im Falle einer malignen Entartung können diese Eigenschaften einer zellulären Plastizität sowie einer unbegrenzten Selbsterneuerung von Tumorzellen übernommen werden, was ihnen stammzellähnlichen Charakter verleiht. Diese sogenannten Krebsstammzellen sind eine in vielerlei Hinsicht noch unverstandene Ausprägung der Tumorzellen. Es wird jedoch vermutet, dass sie entscheidend zu der Bildung von Rezidiven und Metastasen des Tumors beitragen und somit die Ursache der Malignität von Tumoren darstellen.
Für ein besseres Verständnis der ständigen Selbsterneuerung von Krebsstammzellen untersuchten wir im Hauptteil dieser Arbeit primäre Kolonkarzinomzellen in dreidimensionalen *ex vivo* Organoid-Modellen. Wir modifizierten die Kulturbedingungen des Organoid-Models um ausschließlich hoch-maligne Kolonkarzinomzellen zu kultivieren. Die so selektierten Tumorzellen zeichnen sich dadurch aus, dass sie unter minimalen Wachstumsbedingungen Organoide bilden können, was auf eine hohe intrinsische Fähigkeit zur Selbsterneuerung schließen lässt. Aus Zellkulturexperimenten mit adulten und embryonalen Stammzellen ist bekannt, dass der Fibroblasten-Wachstumsfaktor 2 (FGF2) essentiell ist, um diese Stammzelleigenschaften der ständigen Selbsterneuerung aufrecht zu erhalten. In den beschriebenen Organoid-Kulturen untersuchten wir den Einfluss von FGF2 auf die Selbsterneuerung dieser primären Darmkrebszellen und konnten zeigen, dass, im Rahmen unserer Kulturbedingungen, diese relativ autonomen, hoch-malignen Tumorzellen auf einen aktiven FGF-Signalwegs angewiesen sind. Weiter beobachteten wir nicht nur einen Verlust der Fähigkeit zur Organoidbildung, sondern detektieren viele molekulare Effekte ähnlich der Resistenzbildung gegen die therapeutisch-angewandten Epidermalen-Wachstumsfaktor (EGF) -Rezeptor-Inhibition.
In einer Studie zum aggressiven digitalen papillären Adenokarzinom (ADPA) ist es uns erstmalig gelungen das Transkriptom dieser seltenen Tumorart zu beschreiben. Auch

hier beobachteten wir eine Entartung des FGF-Signalwegs, was vermutlich zu der hohen Malignität dieser Erkrankung beiträgt.

Die allgemeine Heterogenität von Stammzellen ist jedoch nicht nur in der Tumorbiologie von Bedeutung. Während der frühen Embryonalentwicklung entstehen eine Vielzahl von verschiedenen Zelltypen aus nur einer Zelle, der Zygote. Während dieser ersten Zellteilungen, die noch vor der Einnistung der befruchteten Eizelle in die Uteruswand stattfinden, definieren die embryonale Genomaktivierung sowie die ersten zellulären Spezialisierungen den Beginn weiterer Entwicklungen die zu der großen Komplexität aller Zellen des menschlichen Körpers führen. Detaillierte molekulare Analysen dieser frühesten Ereignisse erfordern moderne Techniken um gleichzeitig das Genom, das Transkriptom sowie das Epigenom von Einzelzellen zu untersuchen. In einem aktuellen Review-Artikel diskutieren wir Fortschritte und bestehende Herausforderungen der Einzelzellanalyse während der frühesten Embryonalentwicklung sowie in primordialen Keimzellen. Die beschriebenen Methoden kommen ebenso in der modernen Tumorforschung zum Einsatz, um die Plastizität und die Progression von Tumoren sowie von Krebsstammzellen in ihrer Heterogenität zu verstehen.

Table of Content

List of Figures

Abbreviations

ADPA = aggressive digital papillary adenocarcinoma
ALK = anaplastic lymphoma kinase
ALPI = Alkaline phosphatase intestinal
APC = adenomatous polyposis coli
BMP = bone morphogenic protein
BMPR = BMP receptor
CBC = crypt base columnar
CER = cerberus
CKIα = casein kinase Iα
c-MYC = cancers-Myelocytomatosis
CSC = cancer stem cell
EPI = epiblast
FGF = fibroblast growth factor
FGFR = fibroblast growth factor receptor
FZD = frizzled
GREM = gremlin
GSK3 = glycogen synthase kinase 3
hESC = human embryonic stem cells
ICM = inner cell mass
Il2rg = interleukin 2 receptor gamma
INHBA = Inhibin A
iPSC = induced pluripotent stem cell
KLF4 = Krüppel-like factor 4
LEF = lymphoid-enhancer binding factor
LGR5 = Leucine-rich repeat-containing G-protein coupled receptor 5
LRC = labeling retaining cells
LRP = low-density lipoprotein receptor-related protein
NOD/SCID = non-obese diabetes / severe combined immunodeficiency
NSCLC = non-small cell lung cancer
OCT4 = octamer-binding transcription factor 4
PDX = patient-derived xenograft
PE = primitive endoderm
SOX2 = sex determining region y box 2
TCF = T-cell factor
TGFBR = TGFB receptor
TGF-β = Transforming growth factor β

1. Introduction

1.1. Pluripotent Stem Cells

A general definition of stem cells includes two functional features, the potential of unlimited self-renewal while also being able to produce differentiated progenies. A cell division generating two equal daughter stem cells is called symmetric cell division and increases the number of stem cells. An asymmetric cell division, on the other hand, generates one daughter cell that differentiates to a mature cell type, while the other daughter cell stays in an undifferentiated state and maintains the pool of stem cells within a tissue. A further classification of stem cells is inferred by their developmental potential [1].

The direct product of fertilization, the zygote and its early derivatives, the blastomeres, are defined as totipotent. They hold the potential to develop into the embryo proper and all extraembryonic tissues such as the placenta, the amnion or the yolk sac. By ongoing successive cell divisions, the compacted morula originates, and cells start to polarize generating the blastocyst, the first morphological structure with distinct cell types. Cells residing inside the blastocyst belong to the inner cell mass (ICM) and are considered pluripotent. Pluripotency defines the ability to develop into all embryonic tissues excluding the potential of developing extraembryonic membranes. The potency of an early embryonic stem cell can be assessed by different *in vivo* assays such as the ability to form chimaeras or teratomas as reviewed in [1, 2]. However, it is worth mentioning that no *in vitro* assay exists to define the developmental potential of totipotent cells.

Around the time of implantation of the blastocyst into the uterus wall, pluripotent cells of the ICM further commit to cells of the primitive endoderm (PE) or epiblast cells (EPI) (*for details, see 3.2*). During post-implantation development, within the first weeks of embryogenesis, all totipotent and pluripotent stem cells are gone defining toti- and pluripotency in normal development as transient cell stages.

Given that totipotency is not relevant in *in vitro* assays, it is fair to say that pluripotent cells represent the most crucial cell stage for stem cell research. Even though early embryonic stem cells exist only transiently, their unlimited potential of self-renewal allows the derivation of human embryonic stem cell (hESC) lines from the ICM of the blastocyst, first successfully established by J. Thomson in 1998 (*Figure 1*) [3]. From these cultured and, by that, artificial hESCs, scientists gained substantial insights into mechanisms of maintaining pluripotency and self-renewal. Taking into account that previous knowledge of mouse embryonic stem cells could only partially be transferred

to the human situation, first attempts of culturing hESC demanded intensive effort [4]. Since hESC can only be obtained from a blastocyst, the starting point of a human being, ethical restrictions need to be considered.

A milestone in regenerative medicine was reached when Takahashi and Yamanaka first induced pluripotency in adult mouse fibroblast in 2006 and in human fibroblasts in 2007 [5, 6]. IPSCs are ethically defensible and can nowadays be obtained from nearly every somatic cell type of every individual. They are an appropriate model to understand mechanisms of human pluripotency and stem cell self-renewal and can be used to study developmental and differentiation processes (*Figure 2*). They can further be applied in drug screening assays when, for example, the toxicity of cardio-active drugs needs to be tested on patient-derived cardiomyocytes [7].

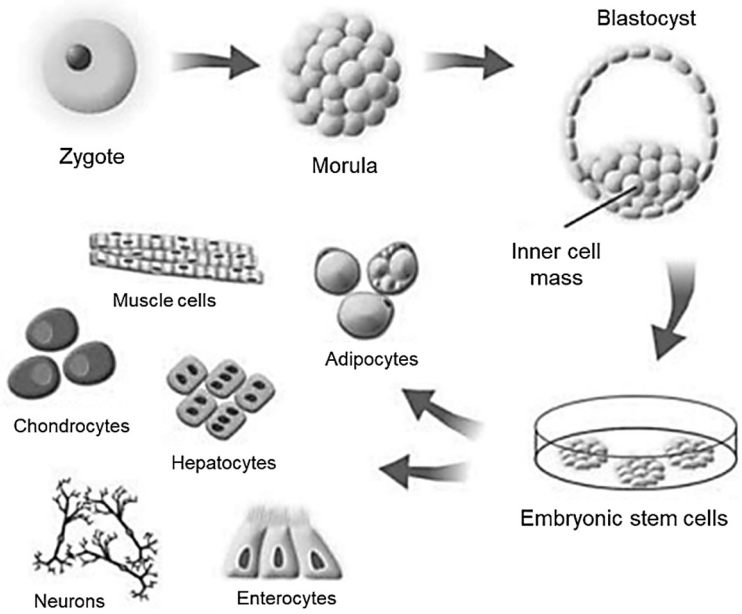


Figure 1. Embryonic Stem Cells

During the pre-implantation stage, the zygote forms the morula by successive cell divisions and later develops into the blastocyst. After implantation, the inner cells of the blastocyst will develop into the embryo proper. In artificial embryonic stem cell culture, these pluripotent cells are capable of infinite self-renewal as well as they can be differentiated into all cell types of the body. With kind permission taken from [4].

1.1.1. Molecular Mechanisms of Pluripotency

It is known from gene expression analyses in preimplantation embryos that the POU-domain containing transcription factor OCT4 (*POU5F1)* is expressed in the unfertilized oocyte and in the blastocyst where it regulates the development of the ICM and maintains its pluripotency [8]. Reprogramming of adult somatic cells is based on the induced expression of core transcription factors facilitating pluripotency. The group of Yamanaka successfully induced pluripotency in human fibroblasts by the transduction of the core transcription factors *OCT4*, *SOX2*, *KLF4* and *c-MYC* when, in the same year, Thomson's group also achieved a stable human iPSC line by transducing the transcription factors *OCT4*, *SOX2*, *NANOG* and *LIN28 [6, 9]*. While OCT4 is the most crucial and defined factor for successful reprogramming, other factors can be replaced or substituted. Despite the fact that NANOG was not party of Yamanaka's initial reprogramming strategy, it is nowadays assumed that *OCT4*, *NANOG* and *SOX2* build a tripartite auto-regulatory loop which sustains pluripotency and self-renewal by collectively binding target gene promoters [1, 10].

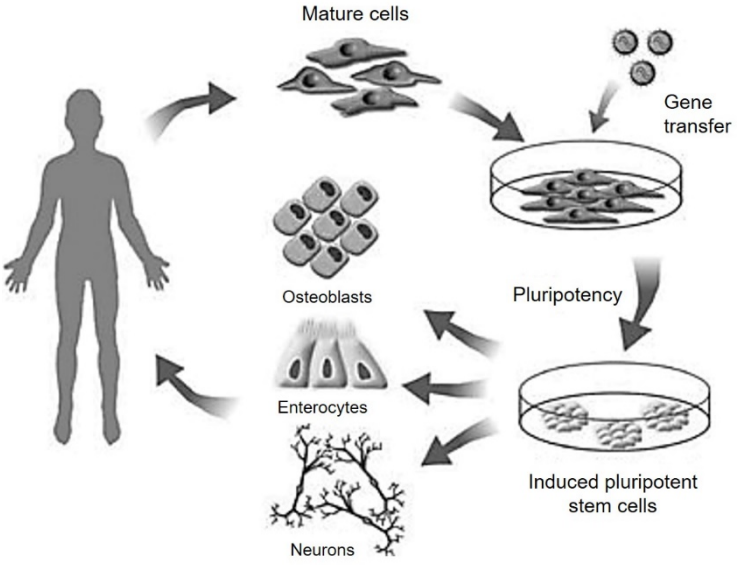


Figure 2. Induced Pluripotent Stem Cells
Mature somatic cells, such as fibroblasts, can be reprogrammed to pluripotent cells, e.g. by the transduction of the Yamanaka factors OCT4, SOX2, KLF4 and c-Myc. Like embryonic stem cells, induced pluripotent stem cells can be differentiated into all cell types of the body. These autologous cells are hoped to become a precious therapeutic tool in regenerative medicine. With kind permission taken from [4].

NANOG and OCT4 represent important molecular markers when testing cells for pluripotency. Especially the isoform OCT4A is highly specific for pluripotent cells and should not be confused with other isoforms or pseudogenes which can lead to misidentification of stem cells and their potential [11, 12]**.**

The core transcription factors OCT4, SOX2 and NANOG not only enable pluripotency but also regulate a wide range of target genes needed for its long-term maintenance. This process is called self-renewal and describes proliferation without differentiation. While the expression of OCT4 is highly restricted to pluripotent stem cells, genes enabling self-renewal are commonly used by many cell types. Some of these signaling cascades are well investigated and highly depend on extrinsic signals from the cellular environment or on the *in vitro* cell culture conditions, respectively. Interestingly, many of these cytokines or transcription factors are induced by OCT4, but also fulfill important functions in higher differentiated progenitor or somatic cells which lack any OCT4 expression. By OCT4 knock-down and chromatin-immunoprecipitation experiments, which analyze OCT4 binding sites within gene promoters, it could be shown that the FGF, TGF-β and Wnt signaling pathways essentially contribute to OCT4-driven self-renewal [10, 13]. Many interconnections between these pathways exist but at least in *in vitro* hESC culture, the cytokine FGF2 is known to extensively modulate TGF-β signaling and also affects Wnt expression. By inducing the expression of TGF-β ligands *TGFB1*, *NODAL,* Activin A (*INHBA*) and simultaneously repressing BMP4 activity by activating its antagonists *GREM1* and *CER1*, FGF2 essentially regulates the self-renewal of pluripotent cells (*Figure 3*). Downstream effects of this complex signaling are integrated by different isoforms of receptor-SMAD proteins. The phosphorylation of SMAD2/3 by the TGF-β receptors TGFBR1 (*ALK5*), ALK4 and ALK7 leads to nuclear translocation of the SMAD-proteins allowing self-renewal [14]. Indeed it was shown that the pharmacological inhibition of the receptors ALK4/5/7 by the small molecule SB431542 prevents the pluripotency promoting effect of FGF2 [15]. The TGF-β receptors ALK1, ALK2 as well as BMPR1A and BMPR1B phosphorylate the proteins SMAD1/5/8 which induce differentiation after nuclear translocation [14, 16]. In this cascade, FGF2 expression is induced by OCT4 and regulates TGF-β signaling which induces the expression of the core transcription factors OCT4, SOX2 and NANOG in terms of a regulatory feedback loop. However, FGF2 is still indispensable as an exogeneous stimulus in cell culture medium to avoid differentiation. This additionally underlines its central role in pluripotency and self-renewal.

As mentioned above, the expression of OCT4 and NANOG can only be detected in pluripotent stem cells, while the reciprocal modulation of self-renewal associated

pathways is found in many cell types and not necessarily stem cell specific. An induction of OCT4 and NANOG by FGF2, however, depends on epigenetic mechanisms, which are also largely regulated by the core transcription factors and other mechanisms as reviewed in [1, 17].

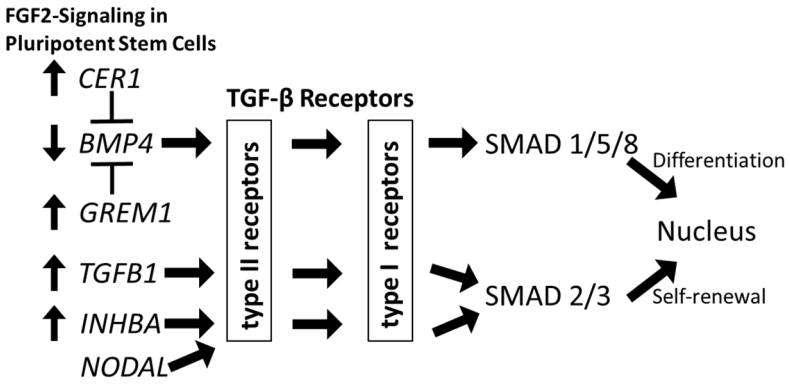


Figure 3. FGF2 Regulates Self-Renewal via TGF-β Members
The self-renewal of pluripotent cells depends on FGF2-signaling regulating the expression of cytokines of the TGF-β family. Heteromeric combinations of different TGF-β receptors lead to SMAD-protein phosphorylation and nuclear translocation. The phosphorylated SMAD2/3 dimer allows self-renewal, while the phosphorylated SMAD1/5/8 heteromer induces differentiation. With kind permission taken from [14].

1.2. Adult stem cells

As every stem cell, adult stem cells are able to self-renew but exhibit a differentiation potential which is much more restricted than that of pluripotent stem cells. Multipotent adult stem cells can produce all cell types within their cell-lineage but *in vivo* never differentiate into cells of other tissues [1]. Unipotent stem cells can generate only one cell type, while progenitor cells are undifferentiated precursors that lack the ability of unlimited self-renewal and therefor do not count as stem cells. In comparison to pluripotent stem cells, which exist only transiently during human development, adult stem cells regenerate and maintain our tissues lifelong. Due to the great complexity of the human body, the pool of adult stem cells is highly heterogeneous with each type having different specifications (*Figure 4*).

Not long ago, adult stem cells were described as rare and quiescent cells with limited self-renewal capacity regenerating the organ by asymmetric cell division [4]. Researches were looking for a well-defined cell population in each tissue, which should be identifiable by a clear marker expression [18]. The prototype of this concept was the hematopoietic system with its stem cells located within the bone marrow. In irradiated patients, e.g. in the case of leukemia therapy, bone marrow transplantation can

regenerate the whole blood system of the recipient. The environment of hematopoietic stem cells was well characterized and putative progenitor cells within the hierarchy of differentiation were identified by surface marker expression.

However, data of the last decade, obtained by complex lineage tracing experiments in genetic modified mouse models and new *ex vivo* cell culture strategies, challenged this view not only in the hematopoietic system but also in other tissues in which adult stem cells were never definitely identified [18, 19]. Especially in organs such as the intestine, the skin or the airway epithelium with high turn-over rates, meaning strained cells need to be replaced frequently, quiescent and asymmetrically dividing cells could not be identified. It was rather revealed that, upon injury and following regeneration, a heterogeneous cell population with equal potency of stem cell-like qualities replaces lost cells [19, 20]. It has further been shown that the fate of a stem cell's daughter cell is not pre-defined by an intrinsic determination to asymmetrically divide but the environment, also referred to as the stem cell niche, extrinsically affects its commitment [19, 21, 22]. We will discuss these characteristics of a heterogeneous stem cell pool with quiescent and proliferative stem cells, which do not necessarily divide asymmetrically, on the intestinal crypt; one well-investigated stem cell compartment.

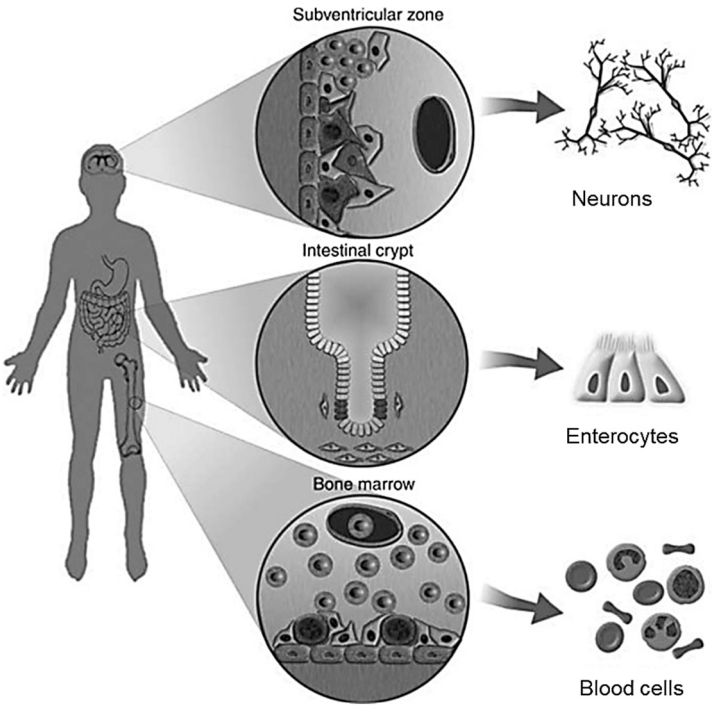


Figure 4. Adult Stem Cells

Adult stem cells are multipotent cells with an unlimited self-renewal capacity. After embryonic development, adult stem cells are the only stem cells within the body maintaining and regenerating all organs and tissues lifelong. Adult stem cells are assumed to reside in a highly specialized environment, called the stem cell niche, which dictates the stem cell's fate. With kind permission taken from [4].

1.2.1. The Intestinal Crypt

The mucosa of the intestinal tract secretes digesting enzymes, absorbs nutrients and is in close contact with microorganisms. These requirements demand an extensive cellular turn-over rate and the epithelial lining of the intestine is completely renewed every five to seven days [23, 24]. Having such a high level of constant self-renewal, the intestine became a most tempting model to study mechanisms of tissue homeostasis and regeneration. Most knowledge about the intestinal stem cell compartment was gained from genetic modified mouse models of which not every detail can be transferred to humans.

The small intestine is built by a structure of protrusions into the intestinal lumen, called villi, and invaginations called crypts. The villi increase the surface area for a better nutrition uptake, while crypts represent intestinal glands producing mucus as well as they harbor the intestinal stem cells. Villi exist only in the small intestine, but crypts can be found across the whole intestine.

It has been known for many decades now that highly mitotic cells reside at the base of the crypt, called crypt base columnar (CBC) cells, and that their progenies continuously repopulate the intestinal epithelium. However, their regenerative potential and useful markers have been a conundrum for many years. Besides the CBC cells, the crypt base is populated by Paneth cells secreting anti-microbial compounds contributing to the immune defense of the intestine. Right above the crypt base at the so called “+4 position” reside labeling retaining cells (LRCs) assuming that these cells are quiescent non-cycling cells. As stated above, it has been assumed that quiescence, which would reduce the risk to acquire DNA mutations during life-time, would be a hallmark of adult stem cells [23].

Early genetic analyses have revealed that a loss-of-function mutation of the negative regulator of Wnt signaling, *APC*, causes familial adenomatous polyposis, a hereditary disease in which colonic epithelial cells massively proliferate, form polyps and nearly always develop colorectal cancer. *APC* as well as other genes of the Wnt signaling pathway such as β-catenin (*CTNNB1*) or *AXIN2,* were further identified as regularly affected in the development of non-hereditary sporadic colorectal cancer [23, 25, 26].

Besides the fact that the Wnt pathway is evolutionary conserved across the entire animal kingdom, its central role in the development of human cancer supported an intensified research in this area, placing Wnt-signaling among the best characterized and also most complex known signaling pathways. The canonical β-catenin-dependent signaling can be delineated as that a Wnt glycoprotein binds its receptor Frizzled (Fzd) together with its co-receptor Lrp5. Receptor activation inhibits the glycogen synthase

kinase 3 (Gsk3) which is part of a destruction complex consisting of Apc, Axin2 and casein kinase Iα (CkIα). The Wnt-induced inhibition of the destruction complex leads to cytoplasmic β-catenin accumulation which translocates to the nucleus where β-catenin binds the transcription factors T-cell factor 4 (Tcf4) and lymphoid-enhancer binding factor (Lef) enabling target gene expression [27, 28].

When Tcf4 was experimentally transduced and gene expression was compared to cells with an artificially disrupted β-catenin activity, the oncogene and stem cell factor c-Myc was identified as one highly regulated target gene of colon cells with aberrant Wnt activity. C-Myc induced a phenotype of strong proliferation, which was abrogated after β-catenin disruption. Additionally, it was shown that the downregulation of c-Myc leads to differentiation of the cells [29]. Searching for stem cells in the colon crypt, the identification of aberrant c-Myc expression as one important oncogene in colon cancer was of significance. However, c-Myc is a ubiquitously expressed transcription factor involved in many complex cellular processes and no specific factor for the differentiation or proliferation of colon cells. Another interesting target gene of high Wnt activity and also associated with the less differentiated phenotype was the Wnt-modulating cell surface protein Leucine-rich repeat-containing G-protein coupled receptor 5 *(Lgr5)*. The Lgr5 protein acts as a receptor for the Wnt agonist R-Spondin *(Rspo)* which stabilizes the expression of the actual Wnt receptor Fzd at the cellular surface [30]. By lineage tracing experiments in genetic modified mice carrying reporter knock-in alleles, an exclusive expression of Lgr5 in the CBC cells was revealed. Further, these experiments have shown that under normal conditions all cells of the intestinal epithelium were derived from these $Lgr5^+$ CBC cells constituting a sufficient stem cell population [31]. This approach was later refined by a stochastic multicolor Cre-reporter, called confetti knock-in, allowing fate mapping of neighboring individual stem cells within the same niche. This study has shown that $Lgr5^+$ stem cells are perfectly intermingled with Paneth cells which were already known to express Wnt ligands and by that induce stemness in the adjacent cells. Even more interesting was the observation that each heterogeneously colored stem cell compartment drifts into monoclonality. Within eight weeks of lineage tracing, crypts and the according villi were composed of single colored daughter cells, indicating that the $Lgr5^+$ adult stem cells do not divide asymmetrically but the fate decision of a daughter cell is stochastically defined by niche space (*Figure 5*) [21, 22, 32].

These observations, described as neutral drift, were confirmed in a following study showing that multipotent $Lgr5^+$ stem cells, which reside at the border of the stem cell compartment, lose their potency when they were passively displaced from the niche by cell divisions of the more central cells. Such cells develop into transit amplifying

progenitors without any cell division [33]. These studies conducted in genetically modified mice assume that Lgr5$^+$-cells play a pivotal role in the homeostasis of the intestinal mucosa. More data underpinning Lgr5$^+$-cells as multipotent stem cells were obtained by the *ex vivo* organoid model showing that single Lgr5$^+$ stem cells can grow "mini-guts" in culture as described in chapter 1.3.2 [34].

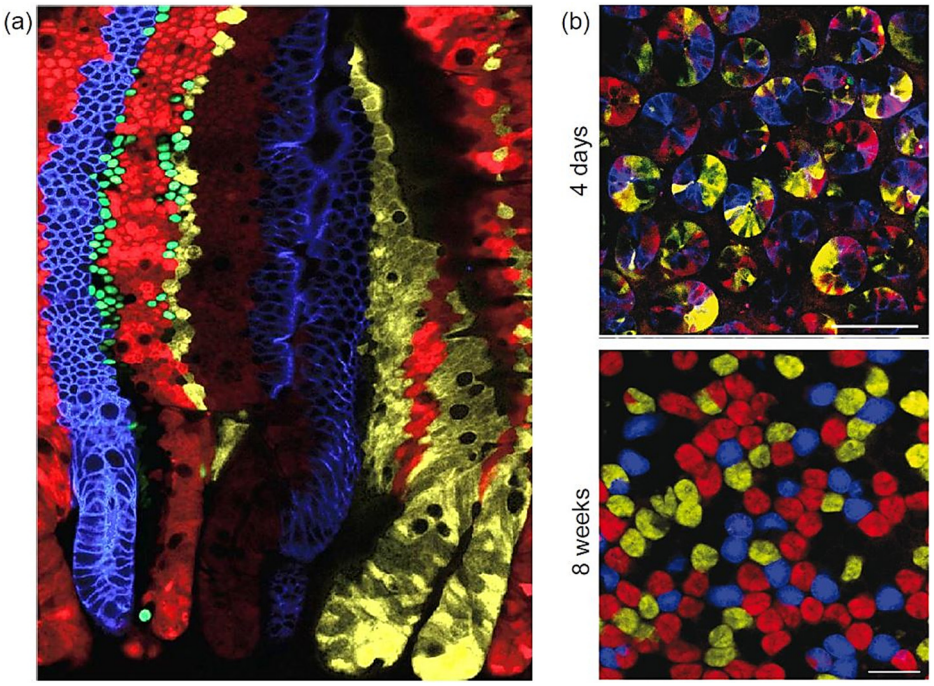


Figure 5. Multicolor Reporter "Confetti"
The confetti reporter labels Lgr5$^+$ intestinal stem cells stochastically with fluorescent proteins. All Lgr5$^+$ daughter cells keep expressing the color during their life time, even after differentiation into Lgr5$^-$ negative cells. (a) Every crypt becomes monoclonal over time while different crypts express different fluorescent proteins. (b, c) Within 4 to 8 weeks former heterochromatic crypts become monochromatic indicating monoclonality by competition of the stem cells for niche space [32]. *https://creativecommons.org/licenses/by/4.0/; no changes were made*

Self-renewal of intestinal crypts can be well explained by highly proliferative Lgr5$^+$ stem cells. However, these cells depend on Wnt modulation and are, due to their strong proliferation, also highly sensitive to radiation. The regenerative potential upon injury of the intestine cannot be explained by Lgr5$^+$-cells alone. Another cell type ascribed with a regenerative potential are the above stated LRC +4 cells which are not affected by Wnt inhibition and proliferate much less than Lgr5$^+$-cells. There is strong evidence that LRCs completely regenerate the intestine when Lgr5$^+$-cells were ablated by radiation or by specific knock-in alleles mediating inducible cell death [35, 36]. In such situations, LRCs reverse to a stem cell state and re-express Lgr5 and also rebuilt the secretory population of Paneth cells reconstructing the niche [37]. From these observations, it has been

concluded that $Lgr5^+$ CBC cells are the intestinal stem cell population sufficiently maintaining homeostasis while LRCs at the +4 position can regain the complete adult stem cell potential after injury constituting a reserve stem cell pool [23]. In a more recent study, it was shown that late precursor cells of the enterocyte lineage, simple columnar epithelial cells which absorb nutrients, can also reverse to a stem cell phenotype. These precursors reside at the upper part of the crypt and were traced by a knock-in reporter allele controlled by the intestinal alkaline phosphatase (Alpi) promotor. After injury, $Alpi^+$ cells migrated down the crypt, dedifferentiated into $Lgr5^+$ or Paneth cells and rebuilt the stem cell niche of the crypt base resulting into a recovered intestinal epithelium of $Alpi^+$-derived progenies [38].

The mouse gut has a tremendous regenerative potential as shown by these comprehensive studies. Many aspects, for example the existence of $LGR5^+$ multipotent stem cells, could be confirmed in humans by organoid cultures [39]. However, the stem cell niche of human colon crypts is differing in some respects such as different Wnt ligands are expressed and Paneth cells do not exist in the terminal colon. It is rather assumed that myofibroblasts surrounding the crypt base express Wnt ligands as well as other stem cell regulating factors [40].

Important regulators of stem cell behavior in rodents and humans in the colonic crypt are BMPs, which were already discussed in pluripotent stem cells as ligands of the TGF-β pathway inhibiting hESC self-renewal (see 0). They fulfill a similar function in adult stem cells, acting as antagonists of Wnt-signaling and inhibiting self-renewal. The local concentration of Wnt and BMP molecules is tightly regulated along the crypt axis. Wnt ligands and their receptors FZD are highly expressed at the bottom of the crypt, while the differentiating BMP ligands and receptors are expressed at the upper crypt parts. This ratio is further controlled by the negative Wnt-regulating factor Dickkopf-related protein (DKK), which is expressed at the crypt base to avoid over-proliferation within the stem cell compartment. The BMP-antagonist GREM1, also involved in hESC self-renewal, attenuates the effect of BMPs at the medium level of the crypt (*Figure 6*) [41].

FGF2, a master regulator of pluripotency in hESCs, is not as well described in the intestine. Different types of FGFRs and several FGF ligands play important roles during embryonic development of the digestive tract and are expressed in the intestine and the surrounding stroma in the adult gut. In mice, it has been reported that FGF2 is mainly expressed by mesenchymal cells of the intestinal stromal cells where it enhances the survival of intestinal stem cells after radiation and is activated during repairing processes [42]. Another study reported increased FGF2 blood levels in patients with chronic

inflammatory bowel disease suggesting a role during inflammation and regeneration [43].

Beyond those cytokines named, an even more complex network including factors of the Ephrin or NOTCH pathway are additionally involved in regulating self-renewal and lineage commitment of the adult stem cells into different cell types of the intestine which was described elsewhere [40, 41].

Intensive research on the intestinal epithelium has shown that the high regenerative potential is achieved by different stem cell populations. Lgr5 expressing CBC cells are the dominant cell type responsible for tissue homeostasis indicating that adult stem cells are not necessarily quiescent or rare and are not predefined to divide asymmetrically [19]. The identification of +4 LRCs revealed that at least one other cell type of the intestine builds a reserve stem cell pool being able to regenerate the full stem cell compartment upon injury indicating that the assumed hierarchy of stem cell differentiation is not unidirectional. This great plasticity further puts the mesenchymal niche of the adult stem cells into focus which is assumed to highly regulate a cells identity. These findings are also of great importance for the understanding of cancer biology, as cancer cells might also acquire stem cell properties, dedifferentiate and show a high plasticity as we will discuss in the following sections.

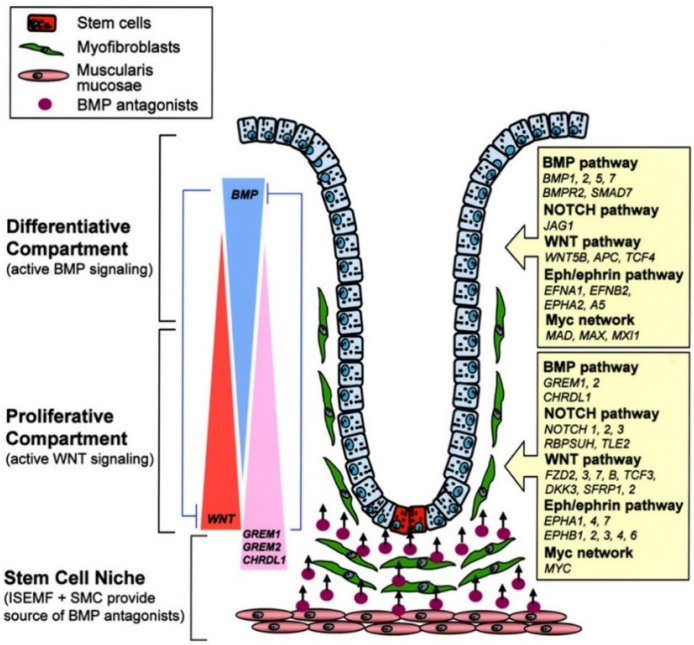


Figure 6. Cell Signaling in the Human Colonic Crypt

A complex signaling network regulates cellular differentiation within the colonic crypt. Most undifferentiated intestinal stem cells ($LGR5^+$) reside at the crypt base determined by a strong Wnt-activity. Towards the upper part of the crypt, increasing BMP activity induces differentiation of colonic enterocytes. ISEMF: Intestinal subepithelial myofibroblasts, SMC: smooth muscle cells. With kind permission taken from [41]. *Copyright (2007) National Academy of Sciences, U.S.A.*

1.3. Cancer

The words *tumor* and *cancer* are often used as synonyms, but they are not interchangeable. Tumor describes a mass of abnormal cell growth, also referred to as neoplasm. Cancer, on the other hand, describes a malignancy of aggressively growing cells invading adjacent lymphatic or blood vessels. Most crucial, however, is the cancer's potential to spread and colonize distant organs, a process called metastatic dissemination. In many cases, metastasis represent the life-threatening incident during cancer progression.

In carcinomas, which are malignant solid tumors of epithelial origin, the progression from a premalignant lesion to an invading and metastasizing cancer is described by a multistep progress of acquired properties. These capabilities were assorted by Douglas Hanahan and Robert Weinberg as the six hallmarks of cancer in 2000 and revisited in 2011 [44, 45]. Besides the already mentioned characteristics of sustaining proliferation, invasion and metastasis, the hallmarks comprise resisting cell death, evading growth suppression, enabling replicative immortality and inducing angiogenesis. Crucial for the stepwise development of these hallmarks is the genetic instability of cancer cells leading to the accumulation of somatic mutations in central regulatory positions of each characteristic. On the one hand, cancer cells are not susceptible for external stimuli such as growth suppressors by neighboring cells or apoptosis induction by immunocytes. On the other hand, cancer cells highly depend on their environment as they stimulate the surrounding stroma to express tumor promoting factors or they induce angiogenesis [45]. It is assumed that the development of cancer starts with one renegade cell acquiring increased proliferation. However, in progressed cancer, different cells are not just clones of the cell of origin but are often highly heterogeneous and build complex tissues with a specialized environment. As cancer often interacts with the whole organism, tumors can be designated as abnormal organs [46]. Equivalent to normal organs, it has been assumed for many years now, that cancer tissues are built by a hierarchy of cells of different potential, with cancer stem cells (CSCs) representing the top of this hierarchy.

1.3.1. Cancer Stem Cells

The theory of CSCs suggests that a subpopulation of cancer cells possesses stem cell characteristics, similar to adult stem cells, which continuously regenerate the tumor during therapy and colonize metastatic sites. CSCs might also reside in a dormant state making them less sensitive for chemotherapy and causing tumor recurrences even after many years of cancer remission.

In healthy tissue compartments such as the intestinal crypt, cellular heterogeneity is triggered by differentiation programs, epigenetic mechanisms and extrinsic stimuli of the environment. In cancer tissue, an additional important factor is the genetic heterogeneity, emanating from different subclones with diverse mutational patterns. This additional level of complexity bedevils the understanding of a cellular hierarchy within cancer tissue and due to the many known factors influencing tumor growth, an explicit identification of CSCs has not yet been possible.

It has been known for a long time that an undifferentiated cellular phenotype correlates with a worse prognosis, while well-differentiated cancer cells are usually less malignant. This relationship was first described in 1964 in germ cell tumors and teratomas in which progenies of highly proliferating stem-like cells decreased their proliferation upon differentiation while retaining the oncogenic driver mutations [47, 48]. However, only since it was possible to sort living cells from a tissue context and to engraft a defined cell population into immunocompromised mice, the correlation of stem cell marker expression with tumorigenicity could be investigated and the CSC concept gained more attention [47]. A landmark study about CSCs in carcinomas was published in 2003 by Al-Hajj *et al.* showing that breast cancer initiating cells can be defined as $CD44^{high}$ and $CD24^{low}$. In colon cancer, CSCs were identified in 2007, when two independent studies indicated $CD133^{+}$ as an appropriate marker [49, 50]. It was reported that $CD133^{+}$ cells initiate xenografts in mice with high efficiency which resembled the parental tumor and consisted of heterogeneous cell populations implying self-renewal and multi-lineage differentiation. Similar experiments were conducted in other tissues such as in pancreas, brain or ovary cancer as reviewed in [47, 51]. Putative CSCs were identified by different markers always constituting a rare subpopulation of cells with tumorigenic potential. Similar results were initially obtained in melanoma, reporting a rare subpopulation of marker positive cells initiating tumor engraftment [52]. However, the concept of analyzing stem cell qualities by tumor initiation in mice was challenged in the same year by Quintana *et al*. They could show that 27% of all unselected primary melanoma cells can induce tumors when transplanted as single cells [53]. Different to earlier engraftment experiments, Quintana *et al*. used a modified assay based on recipient mice

which are more highly immunocompromised by the loss of the Interleukin-2 gamma receptor ($Il2rg^{-/-}$) additionally to the standard Non-Obese Diabetic/ Severe Combined Immunodeficiency (NOD/SCID) background. Transplantation efficiency could further be improved by mixing the tumor cells with Matrigel presumably protecting the transplanted cells from the foreign environment [53].

It has always been criticized that xenotransplantation assays rather reveal the robustness of a cell against the sorting and engraftment procedure than real stem cell-like qualities. To follow the fate of a potential CSC during tumor formation and cancer progression, lineage tracing in genetic modified mice, as discussed in 1.2.1, has been developed [54].

After $Lgr5^{+}$ cells of the intestinal crypt have been identified as potent adult stem cells in normal tissue, their role during tumor formation has been investigated. Similar to earlier studies in the healthy intestine, lineage tracing experiments have revealed that $Lgr5^{+}$ stem cells can form adenomas, which are benign tumors of the intestine, when the tumor suppressor Apc was genetically deleted. In these mouse models, transformed $Lgr5^{+}$ cells fueled adenoma growth, while maintaining their stem cell qualities of self-renewal and differentiation potential [55]. Further studies with complex lineage tracing strategies, e.g. using the abovementioned confetti-reporter, confirmed a hierarchical organization of adenomas following the CSC concept. Additionally, it was shown that not all transformed $Lgr5^{+}$ cells act as CSCs but are also subject to the mechanisms of neutral drift, meaning that only cells in the proper niche retain their stem cell characteristics and “islands” of stem-like cells of clonal origin develop with tumor progression. The dependency on the niche was also confirmed by the observation that early $Apc^{-/-}$- $Lgr5^{+}$ cells were intermingled with Paneth cells indicating a dependency on Wnt-signaling besides the Apc-negative background [56, 57].

Artificially induced tumors in mice coupled with lineage tracing approaches of different cell types provided tremendous insights in mechanisms of tumor formation and progression. However, these benign adenomas usually do not grow invasively and never metastasize reflecting only some aspects of human colon cancer. Luckily, the establishment of *ex vivo* organoid cultures made human tissue assessible for lineage tracing experiments and other genetic modifications.

1.3.2. Intestinal Organoid Culture

About a decade ago, Toshiro Sato, Hans Clevers and colleagues published the first long-term organoid culture system modeling mouse intestinal crypt-villus units [34]. They could show that a single Lgr5$^+$ stem cell has the potential to initiate these "mini-guts" consisting of many different cell types. Organoids are self-organizing and do not rely on a mesenchymal cellular niche. They have further proven that the self-renewal of intestinal adult stem cells is not limited by intrinsic restrictions in *ex vivo* cultures [34]. In 2011, the same group published a detailed culture protocol for mouse and human intestinal tissue of healthy and tumor origin [39]. Culturing organoids is based on providing an artificial extracellular matrix, e.g. Matrigel, as well as on a cocktail of essential growth factors such as WNT3A, RSPO, EGF, the BMP inhibitor Noggin and, specifically for human culture systems, the TGF-β receptor inhibitor A83-01 and the p38-MAPK inhibitor SB202190. In a systematic comparison, it has been revealed that culturing human tissue requires more complex culture conditions than mouse tissue, but these requirements are highly reduced in colon cancer [39]. By optimized establishment protocols, culture efficiency of human colon tissue reached nearly 100% including most colon cancer subtypes. This allowed the establishment of comprehensive living organoid biobanks also including the according healthy tissue [58, 59]. Genome and transcriptome studies have revealed that organoid cultures of individual patients reflect the intratumor heterogeneity to a high extend which is also stable over many passages [59]. Given the high culture efficiency and great genomic stability, organoid cultures have many advantages compared to earlier cancer models such as cell lines. 2D *in vitro* cultures of primary cancer cells can only be obtained with very low efficiency while the derivation of healthy control cells is not possible in this culture system. In general, only rare clones can be expanded in cell line culture which do not mirror the original tumor's heterogeneity. Another common cancer model uses patient-derived xenografts (PDX) which describes the transplantation of primary tumor material into immune-compromised mice making it an extensive and costly method. While the missing stroma is a drawback of current organoid cultures, PDXs need to adapt to the mouse environment limiting heterogeneity of the growing xenograft [60].

Modern new genome editing technologies now also permit to directly modify patient-derived organoids opening widespread possibilities to analyze mechanisms of cancer development. To study the role of LGR5$^+$ putative CSCs, Cortina *et al.* induced a reporter gene in the *LGR5*-gene locus of patient-derived organoids before xenotransplantation for *in vivo* cell fate mapping. This study showed, that LGR5$^+$ cells are able to self-renew

after transplantation and produce LGR5$^-$ progenies which mostly express the enterocyte differentiation marker KRT20 or the goblet cell marker MUC2. Interestingly, when LGR5$^-$ cells were transplanted, the xenografts also consisted of LGR5$^+$ cells indicating plasticity and dedifferentiation of cancer cells [61].

The process of tumorigenesis is characterized by the acquirement of several genome alterations. Progressed cancers can have highly aberrant genomes but only few mutations act as drivers of tumor progression while most alterations are only passenger mutations. Such driver mutations not only influence the aggressiveness of cancer, but they also define niche requirements for organoid cultures. This aspect was investigated by two independent studies which consecutively induced critical mutations into intestinal stem cell organoid cultures. Organoids with the combined mutations of *APC*, *P53*, *KRAS* and *SMAD4* grew independently of any stem cell niche factors and initiated invasive carcinomas upon xenotransplantation [62]. The work of Mami Matano, Toshiro Sato and colleagues used a similar approach additionally introducing mutations in the *PIK3CA* locus. They also report an invasive growth pattern of transplanted xenografts. Interestingly, organoids carrying all introduced mutations were not able to metastasize. Only when obtained from patients with chromosome-instable adenomas, organoids were able to colonize the liver indicating that additional unknown mutations are needed to grow in the hostile environment [63].

Plasticity and heterogeneity of advanced colon cancers are the biggest challenge when developing new anti-cancer therapies. By using the living organoid bio bank developed by Fujii *et al.* [59], Shimokawa *et al.* gained tremendous insights into mechanisms of stem cell plasticity when targeting colon CSCs. With their strategy of an inducible apoptosis cassette into LGR5$^+$ CSCs, they showed that killing CSCs can stop tumor progression. However, all cancers regained their proliferative potential when apoptosis induction was removed. The combination of complex lineage tracing strategies has shown that free niche space after CSC removal continuously induces stemness in former differentiated cells and that post-mitotic cells regained proliferation [64].

It is nowadays widely accepted that in many cancers stem cell properties are the root of malignancy. By constant self-renewal, CSCs fuel the tumor growth and restore the cancerous tissue after therapy. With their ability to differentiate into several cell types, CSCs also contribute to intratumor heterogeneity. However, the initially postulated unidirectional hierarchy of stem cell differentiation, leading to non-tumorigenic and by that non-malignant bulk cells, is outworn. New insights about healthy intestinal stem cells and their potential to maintain and regenerate the gut epithelium simultaneously

illuminated our understanding of cancer biology. Recent data of therapeutically targeting CSCs indicate that a simple eradication of CSCs, if ever possible, would still not be the final solution in curing advanced cancer. It can rather be assumed that the stem cell niche perpetually recreates CSCs demanding an all-embracing therapeutic strategy targeting CSCs, their niche as well as fast and slow cycling cancer cells [54].

2. Aims of this thesis

A cell's true identity, its potential and fate within the human body are still often puzzling due to its versatility. Powerful new technologies, especially in *ex vivo* culturing methods, lineage tracing strategies and in single-cell analyses have provided fascinating new insights but also raised new questions. The heterogeneity and plasticity of embryonic and adult stem cells cause difficulties when addressing ethical and safety concerns in reproductive and regenerative medicine. Most challenging, however, is the nature of CSCs, which is still mysterious in many aspects.

As the constant self-renewal of CSCs is maybe the most clinically relevant aspect, this thesis aims to understand self-renewal in colon CSCs. In our *ex vivo* organoid model we especially focus on pathways shared with adult and embryonic stem cells. Identifying critical self-renewal-associated pathways we analyze molecular downstream effects of its interruption.

We were the first to analyze the transcriptome of aggressive digital papillary adenocarcinomas (ADPA), a very rare tumor entity of the sweat glands of the skin. This tumor entity is characterized by a high malignancy and metastatic potential, properties which are, in some tumor entities, assumed to be conferred by CSCs. Our aim is to identify according gene expression patterns within the transcriptome and to detect deregulated pathways explaining the aggressiveness of this cancer.

Another aim of this thesis is to describe the heterogeneity of stem cells leading to the enormous complexity of the human body during embryonic development, inducing genetic variety between progenies but also being relevant in tumor biology. In circumstances such as during the pre-implantation development the biological material is very rare. Many tools for the combined analysis of the genome, transcriptome and epigenome of low-input material down to single cells were developed in recent years. In a state-of-the-art review, we elaborate the significance of current methods and future challenges in analyzing stem cell heterogeneity at the single-cell level during embryonic preimplantation development and in primordial germ cells. We further contribute a protocol for the combined analysis of the genome and transcriptome of ultra-low input material.

3. Results

3.1. FGF Signaling in the Self-Renewal of Colon Cancer Organoids

Jörg Otte, Levent Dizdar, Bianca Behrens, Wolfgang Goering, Wolfram T. Knoefel, Wasco Wruck, Nikolas H. Stoecklein* and James Adjaye*

*contributed equally

Abstract

With their ability to self-renew and simultaneously fuel the bulk tumor mass with highly proliferative and differentiated tumor cells, cancer stem cells (CSC) supposedly drive cancer progression. However, the CSC-phenotype in colorectal cancer (CRC) is unstable and dependent on environmental cues. Since FGF2 is essential for adult and embryonic stem cell culture to maintain self-renewal, we investigated its role in advanced CRC using tumor-derived organoids as experimental model. We found that FGF-Receptor (FGFR) inhibition prevents organoid formation in very early expanding cells but induces cyst formation when applied to already established organoids. Comprehensive transcriptome analyses revealed that the induction of the transcription factor activator-protein-1 (AP-1) together with a MAPK stimulation was most prominent after FGFR-inhibition. These effects resemble mechanisms of an acquired resistance against other described tyrosine kinase inhibitors such as targeted therapies against the EGF-Receptor. Furthermore, we detected elevated expression levels of several self-renewal and stemness-associated genes in organoid cultures with active FGF2 signaling. The combined data assumes that CSCs are a heterogeneous subpopulation while self-renewal is a common feature regulated by many different pathways. Finally, we highlight the effects of FGF2 signaling as one of numerous aspects of the complex regulation of stemness in cancer.

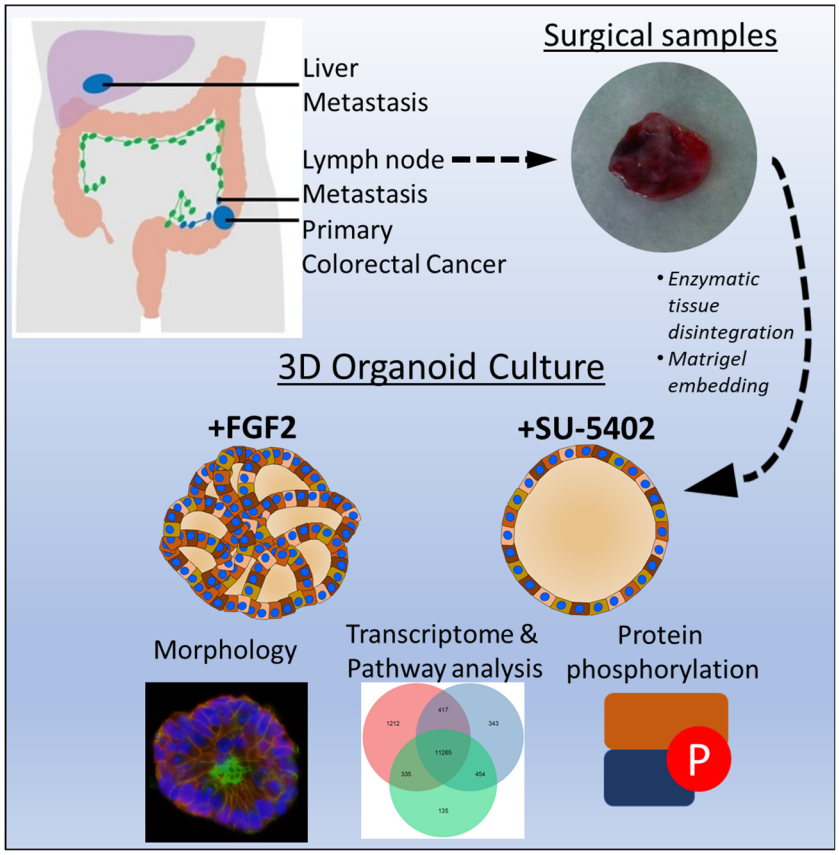


Figure 7. Graphical Abstract

Graphical abstract showing the workflow of the study. Biopsies of primary colorectal cancer, adjacent lymph nodes and a liver metastasis were obtained. An organoid culture was obtained, and effects of FGFR-inhibition were analyzed.

The graphical abstract is not part of the original publication and was designed for the purpose of this thesis.

Share of scientific contribution: 75%

All experiments were designed and executed by Jörg Otte. Jörg Otte established the organoid culture of patient-derived tumor cells. With support from his co-authors, he received and interpreted all data and wrote the manuscript under supervision of Prof. James Adjaye and Prof. Nikolas Stoecklein.

The article was submitted to the *International Journal of Cancer* on November 8th, 2018 and is currently under review.

FGF Signaling in the Self-Renewal of Colon Cancer Organoids

Jörg Otte[1], Levent Dizdar[2], Bianca Behrens[2], Wolfgang Goering[3], Wolfram T. Knoefel[2], Wasco Wruck[1], Nikolas H. Stoecklein[2]* and James Adjaye[1]*

*contributed equally

[1]Institute for Stem Cell Research and Regenerative Medicine, University Hospital and Medical Faculty of the Heinrich-Heine University Düsseldorf, Düsseldorf, Germany.
[2]General, Visceral and Pediatric Surgery, University Hospital and Medical Faculty of the Heinrich-Heine University Düsseldorf, Düsseldorf, Germany
[3]Institute for Pathology, University Hospital and Medical Faculty of the Heinrich-Heine University Düsseldorf, Düsseldorf, Germany.

Conflict of interests
The authors declare no potential conflicts of interest

Correspondence to:
Prof. James Adjaye,
Institute for Stem Cell Research and Regenerative Medicine
Heinrich Heine University
Moorenstrasse 5
Building 14.80
40225, Düsseldorf
Germany
Tel. +49(0)211-81-08191
Fax. +49(0)211-81-17858
James.Adjaye@med.uni-duesseldorf.de

Short title
FGF Signaling in the Self-Renewal of Colon Cancer Organoids

Novelty and Impact
Describing the role of FGF2 in the self-renewal of colon cancer stem cells, we highlight facets of the complexity of stemness in cancer, its heterogeneity and important aspects of primary cancer organoid culture. We found elevated expression levels of self-renewal and stemness-associated genes in organoid cultures with active FGF2 signaling. Pharmacological FGF-receptor inhibition completely abrogated organoid formation in naïve cells but induced cyst formation, MAPK and EGF-receptor signaling when applied to organoids of higher passages.

Article Category
Molecular Cancer Biology

Abbreviations
5-FU: Fluorouracil; **BMP:** bone morphogenic protein; **CNV:** copy number variation; **CRC:** colorectal cancer; **CSC:** cancer stem cells; **EGF:** epidermal growth factor; **EGFR:** epidermal growth factor receptor; **EMT:** epithelial-mesenchymal transition; **ERK:** extracellular signal-regulated kinases; **FGF2:** fibroblast growth factor-2; **FGFR:** fibroblast growth factor receptor; **GO:** gene ontology; **GSEA:** gene set enrichment analysis; **hESC:** human embryonic stem cells; **HISC:** human intestinal stem cell; **MAPK:** mitogen-activated protein kinase; **NSCLC:** non-small-cell lung carcinoma; **RAS:** rat sarcoma; **TCGA:** the cancer genome atlas; **TGF-β:** tumor growth factor beta;

Keywords
Self-Renewal, Cancer Stem Cells, Colon Cancer, FGF2, Organoids

1. Abstract

With their ability to self-renew and simultaneously fuel the bulk tumor mass with highly proliferative and differentiated tumor cells, cancer stem cells (CSC) supposedly drive cancer progression. However, the CSC-phenotype in colorectal cancer (CRC) is unstable and dependent on environmental cues. Since FGF2 is essential for adult and embryonic stem cell culture to maintain self-renewal, we investigated its role in advanced CRC using tumor-derived organoids as experimental model. We found that FGF-Receptor (FGFR) inhibition prevents organoid formation in very early expanding cells but induces cyst formation when applied to already established organoids. Comprehensive transcriptome analyses revealed that the induction of the transcription factor activator-protein-1 (AP-1) together with a MAPK stimulation was most prominent after FGFR-inhibition. These effects resemble mechanisms of an acquired resistance against other described tyrosine kinase inhibitors such as targeted therapies against the EGF-Receptor. Furthermore, we detected elevated expression levels of several self-renewal and stemness-associated genes in organoid cultures with active FGF2 signaling. The combined data assumes that CSC are a heterogeneous subpopulation while self-renewal is a common feature regulated by many different pathways. Finally, we highlight the effects of FGF2 signaling as one of numerous aspects of the complex regulation of stemness in cancer.

2. Introduction

In today's view of tumor biology, cancer stem cells (CSC) are a subpopulation of tumor cells with the ability to self-renew and serve as an ongoing source for differentiated tumor cells. With their stem cell-like properties, they supposedly play a major role in therapeutic resistance and the metastatic process. However, the CSC-phenotype is an unstable feature dictated by environmental cues.[1] Preserving the tumor's hierarchy *in vitro*, organoid cell culture systems have progressed to a versatile model with many applications such as the analysis of driver mutations after gene modifications or complex drug screens.[2]
In colorectal cancer (CRC), stem cell niche requirements seem to reduce with cancer progression and increasing malignancy. With the ongoing acquisition of gene mutations affecting self-renewal associated pathways such as the Wnt/β-catenin, TGF-β, EGF or MAPK-pathways, a more autonomous CSC phenotype develops.[1, 3-5]
For some adult stem cells and especially in human embryonic stem cells (hESC), FGF2 is an important cell culture component essential for the maintenance of self-renewal. In hESC and induced pluripotent stem cells, FGF2 co-ordinates the expression of several TGF-β family members leading to SMAD2/3 phosphorylation and self-renewal while BMPs and the following SMAD1/5/8 phosphorylation are suppressed.[6] These factors are not restricted to pluripotent cells, but also play a major role in intestinal homeostasis with BMPs being one of the major inhibitors of stem cell self-renewal.[7] The influence of the TGF-β pathway is highly context-dependent. In epithelial cells, TGF-β can have tumor suppressive and differentiating effects on healthy or low-malignant cells, whereas it can also promote tumor growth and invasion in mutated cells, which lack the suppressive response.[8] TGF-β can further influence the number of CSCs by stromal expression.[9] In the healthy intestine, FGF2 is important during development, crypt homeostasis and tissue repair after injury. In rat models, FGF2 increased intestinal stem cell survival after radiation.[10]
The aim of our study was to investigate the role of FGF2 in self-renewal of highly malignant CRC. With our organoid culture conditions, we selected for most autonomous CSC with low niche requirements.[3, 11-13] In a first step, we analyzed the transcriptome of obtained organoids, followed by the analysis of effects of FGF-receptor (FGFR) inhibition after SU-5402 treatment.

3. Material and Methods

Tumor dissociation and culture initiation

Surgically resected tumor tissue (metastasis and primary tumors) were obtained from nine patients from the University Hospital Düsseldorf. All patients were diagnosed with colorectal cancer. The study was approved by the institutional ethics committee of the University Hospital Düsseldorf, Germany (#4664). We hereby confirm that all research was performed in accordance with relevant guidelines and regulations and informed consent was obtained from all participants.
Tissue was washed in PBS (Gibco) including antibiotics/antimycotics (200 U/ml Penicillin (Corning), 200 µg/ml Streptomycin (Corning), 500 µg/ml Gentamicin (Gibco), 25 µg/ml Amphotericin B (Sigma)). Tumor fragments were minced using scissors and scalpel until small pieces of maximal size of 0.5 mm^3 were obtained. Smaller pieces were incubated in digestion buffer consisting of Advanced DMEM/F12 (Gibco), 10 mM Hepes (Sigma), 1% GlutaMAX (Gibco), 2% B-27 media supplement without retinoic acid (Gibco), 1% N-2 media supplement (Gibco), 2 µg/ml Heparin (Sigma), 1 mM N-Acetylcystein (Sigma), 50 ng/ml human EGF (Peprotech), 50 ng/ml human FGF2 (Gibco), 10 µM Y-27632 (Tocris), 1 mg/ml Collagenase IV (Biochrom), 20 µg/ml Hyaluronidase IV (Sigma) and antibiotics/antimycotics under constant rotation for 2 h at 37°C with occasionally pipetting up and down.[11, 14]
FBS (Gibco) was added up to 10% to stop the reaction. After centrifugation (300 x g, 5 Min) the supernatant was discarded, and cell suspension was washed in 30 ml PBS and filtered through a 70 µm and a 30 µm nylon mesh (Falcon) and centrifuged (300 x g, 5 Min) afterwards. The cell pellet was resuspended in Cancer Stem Cell Medium consisting of Advanced DMEM/F12, 10 mM Hepes, 1% GlutaMAX, 2% B-27 media supplement without retinoic acid, 1% N-2 media supplement, 2 µg/ml Heparin, 1 mM N-Acetylcystein, 50 ng/ml human EGF, 50 ng/ml human FGF2, 10 µM Y-27632 and antibiotics/antimycotics.
Cells were cultured in ultra-low attachment wells (96 Well (Corning), 5,000 cells per well in 250 µl media) or as organoids in growth-factor-reduced Matrigel (Corning, phenol red-free) droplets (5,000 cells per 50 µl Matrigel droplet). Matrigel droplets were covered with 500 µl medium. Medium was

changed every other day. After 48 h of culturing, Y-27632 was omitted from the media and antibiotics/ antimycotics treatment was reduced to 100 U/ml Penicillin and 100 µg/ml Streptomycin. For FGFR-inhibition experiments, 10 µM SU-5402 (Sigma) was added to the medium. SU-5402 powder was solved in DMSO (30 mM) according to the manufacturer's protocol and was diluted 1:3,000 for a working concentration. The same DMSO dilution was equally added to control experiments.

DNA isolation
Matrigel cultured organoids were treated with digestion buffer for 30-60 Min at 37°C to remove remaining Matrigel. For DNA isolation, the DNeasy Blood & Tissue Kit (Qiagen) was used. According to the manufacturer's protocol, organoids were lysed in Buffer AL with proteinase K and DNA was isolated using silica-based columns.

Analyses of the mutational status of BRAF, KRAS and NRAS using Sanger sequencing
DNA amplification of selected regions from BRAF (Exon 15), KRAS (Exon 2-4) and NRAS (Exon 2-4) was performed using specific primers (Table S1). Amplification was controlled using agarose gel electrophoresis and amplified fragments were subjected to direct Sanger sequencing from either one end or both ends at our core facility (BMFZ, Düsseldorf) using amplification primers as indicated (Table S1). Resulting DNA sequences were analyzed for nucleotide variants using Chromas software (Technelysium), BLAST alignment tool and UCSC genome browser.[15, 16]

RNA isolation
For RNA isolation, Matrigel-cultured organoids were treated with digestion buffer for 30-60 Min at 37°C to remove remaining Matrigel. After washing in PBS, organoids were lysed in 500 µL TRIzol (peqGold Trifast, VWR). RNA isolation was done using the Direct-zol RNA Isolation Kit (Zymo Research) according to the manufacturer's protocol including DNase treatment.

Protein Isolation and Western Blotting
After removing remaining Matrigel using digestion buffer, organoids were lysed in ice-cold RIPA buffer (Sigma) with proteinase-inhibitor (1:7, Roche) and Phosphatase-Inhibitor (1:10, Roche) using a Hamilton syringe. The solution was centrifuged at 4°C, 10,000 x g for 10 Min and the supernatant was taken for BCA (Pierce BCA Protein Assay Kit, Thermo Scientific) protein concentration measurement. 40 µg of protein were analyzed by western blot using antibodies for total-ERK (CST) and phospho-ERK (CST) (Table S2). Equal loading of protein was assessed by Ponceau (Merck) staining after transfer. For signal detection HRP-coupled secondary antibodies were used followed by chemiluminescence detection (Amersham ECL Prime (GE Healthcare), Fusion FX instrument, (PeqLab)).

Organoid embedding and Cryo-sections
Organoids were embedded in OCT (Tissue-Tek, Sakura) by following published protocols with some modifications.[17] Matrigel droplets containing cultured organoids were treated with digestion buffer to remove Matrigel. Free floating organoids were fixed in 4% PFA for 15 Min at RT. Fixed organoids were incubated in 10% sucrose solution, followed by 20% and 30% sucrose solution, each for 30-60 Min for dehydration. To later re-identify the organoids, an alcian-blue staining was done by adding alcian blue (1% in 3% acetic acid, Sigma) solution to the 20% sucrose solution (1:100 dilution). Organoids were placed in an appropriate mold and left to settle down. All remaining sucrose solution was carefully removed. The mold was filled with OCT followed by mild agitation for 15 Min to collect all organoids in the center of the mold. Embedded organoids were frozen at -80°C. 10 µm sections were done using a cryostat (Leica Biosystems) with a cutting temperature of -21°C.

Immunocytochemistry
Slides were washed in PBS to remove all remaining OCT. All steps were carried out in a wet chamber. Unspecific binding was blocked by incubation for 2 h at room temperature (RT) with blocking buffer (PBS, 10% normal donkey serum, 1% BSA). For all intracellular targets, cells were permeabilized by including 0.5% Triton and 0.05% Tween in all blocking, washing and incubation steps.
Primary antibodies for Cytokeratin-20 (abcam) and β-catenin (CST) were diluted in blocking buffer and incubated overnight at 4°C (Table S2). After primary staining, slides were washed three times in PBS. Secondary staining was done at RT for 2 h. For actin filament staining, Alexa Fluor-488 Phalloidin

(Invitrogen) was added to the secondary staining following the manufacturer's protocol. Slides were washed three times and covered with coverslips using DAPI Fluoromount mounting solution (SouthernBiotech). Images were captured using a LSM700 fluorescence microscope (Zeiss)

Microarray transcriptome analysis
Microarray transcriptome profiling was performed at the core facility BMFZ, Düsseldorf. Affymetrix PrimeView Human Gene Chips were used. Biostatistics were conducted as described before.[18] Details are given in the Supplementary Methods.

Comparative Genomic Hybridization with Oligonucleotide Microarrays (aCGH)
ACGH was done using the Kit Oligonucleotide Array-Based CGH for Genomic DNA Analysis, Version 7.5, June 2016 (Agilent Technologies). Genomic DNA (gDNA) isolation was conducted as described above. Sample preparation, labeling and microarray processing was performed following the manufacturer's instructions and as described before.[19] Details are given in the Supplementary Method.

4. Results

Self-renewal in advanced colorectal cancer organoids
In all stem cell cultures, serum replacement is crucial to avoid unwanted differentiation. Our CSC-medium lacks exogenous Wnt, TGF-β, or p38-MAPK modulation, but contains the growth factors EGF and FGF2. We obtained tissue biopsies from nine CRC patients and initiated a spheroid or organoid culture from a single cell suspension of naïve unsorted bulk cells without any foregoing selection. As expected, only biopsies from progressed cancers (with distant metastasis or poorly differentiated) grew in our culture conditions (Table 1). Cells that grew in culture were sequenced for mutational hotspots of *BRAF*, *NRAS* and *KRAS* genes. We found activating mutations in all patients except Patient 2. To further confirm the cancer status of the grown organoids, we analyzed their genomes for copy number variations (CNV) by comparative genomic hybridization. We found highly aberrant genomes in biopsies from Patient 1 and 2, which were derived from metastatic sides and less but significant CNVs in biopsies from patient 3 and 4, which were obtained from primary tumors of the sigmoid colon (Supplementary Fig. S1).
As our medium was intended to maintain cells in an undifferentiated state, we analyzed the transcriptome of the two metastatic biopsies by microarrays and compared their gene expression to hESCs (Fig. 1A, Table S3). We found the majority of expressed genes (11,265) shared between hESC and the metastasis-derived organoids. 1,212 genes were expressed only in hESC, this being the highest number of individually expressed genes, indicating the peculiarity of pluripotent cells. Notably, both cancer samples shared more expressed genes with each other than they express individually.
To identify functional patterns and enriched pathways among these expressed genes, we used the Gene Ontology (GO) tool and specifically focused on biological processes associated with stemness or self-renewal.[20] In the set of genes expressed in all samples, we detected GO-terms of stem cell-associated attributes such as "*telomere organization*" or "*asymmetric cell division*". Self-renewal-associated pathways such as Notch, FGF2, TGF-β or Wnt signaling were also significantly enriched in this gene-set. Factors of these pathways, together with other stem cell genes for example, *SALL4*, *SOX2* or *CDH2* combine in the GOs "*(somatic) stem cell population maintenance*" (Fig. 1B, Table S4). Beyond those listed, we found several GO-terms associated with oxidative stress and hypoxia, proliferation, DNA repair, apoptosis, inflammation or metabolic changes.
Genes expressed by both patient-derived organoids displayed an intestinal differentiation pattern next to stem cell circuits shared with hESC (Fig. 1C, Table S5). The GO-terms "*digestion*", "*maintenance of gastrointestinal epithelium*" or "*bile acid and bile salt transport*" comprise intestinal specific genes of the TFF- or the AKR-family. Interestingly, metastasis-derived organoids exerted an extended Notch and EGF signaling. Beyond those genes shared with hESC, cancer cells expressed Epiregulin (*EREG*) and Amphiregulin (*AREG*) of the EGF-receptor (EGFR) pathway, generally associated with progressed cancer.[21, 22]
Genes expressed by hESC annotated with the GO-term "*cell adhesion*", for example α-catenin (*CTNNA2*) or cadherin-11 (*CDH11*), are associated with adherence junctions but also fulfill functions in stem cell proliferation (Fig. 1D, Table S6).[23, 24] The pluripotency-associated genes *NANOG*, *LIN28*

and *SALL1*, were exclusively expressed by hESC. In accordance with the mechanism of FGF2 regulated TGF-β signaling, the GO-terms "*negative regulation of BMP signaling pathway*" and "*SMAD protein signal transduction"* were detected in our hESC culture.[6] However, most genes associated with this circuit were shared with metastasis-derived organoids indicating that FGF2 also modulates self-renewal associated pathways in our organoid culture. As described above, our CSC-medium contains no further pathway modulators but only serum replacement supplements and the growth factors EGF and FGF2. When we initially established organoid cultures from these chemo-refractory metastases, we observed the withdrawal of EGF had no effect on culture initiation, which is often described in organoids obtained from progressed colon cancer.[3, 12] In our organoid culture model, the independence from exogenous EGF treatment can be explained by activating mutations of genes of the RAS family observed in Patients 1, 3 and 4.[25] Patient 2, however, carried a wild-type RAS gene but developed resistance against the EGFR-inhibitor Cetuximab during pretreatment, indicating a constitutively activated EGF signaling.
Together, these results showed that highly progressed colon cancers can be cultured in conditions modelling minimal niche factors, while intrinsic stem cell-associated pathways remain active.

FGFR-inhibition prevents organoid culture initiation

By comparative transcriptome analyses, we found many self-renewal-associated pathways employed by pluripotent as well as by CSC in our organoid culture. Specifically, many genes linked to FGF2 and TGF-β signaling were shared by these two cell types. Analyzing if this axis of FGF2-TGF-β mediated self-renewal also triggers stemness in our patient-derived organoids, we omitted cytokines from our CSC-medium and added a FGFR-inhibitor, the small molecule SU-5402, which also inhibits auto-phosphorylation of the intracellular receptor domain.[26] To our surprise, we found that no long-term organoid or spheroid culture could be established under FGFR-inhibition (Fig. 2). When we cultured single cell suspensions under low-attachment conditions, we observed spheroid formation within five days. In Patient 1, we observed a luminal organization reminiscent of an adenoma-like architecture upon our standard CSC-medium treatment. Under FGFR-inhibition, however, we found initial spheroid formation, but morphological differences such as a more flattened structure without luminal organization (Fig. 2A). These initial spheroids stopped proliferating after a few weeks and could not be passaged further. We obtained similar results when Patient 4 was analyzed. Spheroids obtained under standard culture conditions appeared more flattened compared to Patient 1. Under FGFR-inhibition, initially formed spheroids stopped growing after 9 d. When naïve single cells were embedded in Matrigel and cultured in our CSC-medium, a more complex self-organization was observed in all patients analyzed (Fig. 2B). As floating spheres, the organoids did not survive under FGFR-inhibition, after initial formation for few days. In general, we observed that spheroid formation was faster than organoid formation in Matrigel, resulting in more and larger spheroids within the first days after culture initiation. This might be due to cell aggregation, which is reduced by the high viscosity of Matrigel. Within 30 d, however, organoids increased massively in size, build more complex structures and proliferated faster than floating spheroids. These results imply that upon our culture conditions without Wnt activation, an active FGF signaling is key to initiate a spheroid or organoid culture.
Interestingly, the effect of FGFR-inhibition was less severe in already established organoid cultures (Fig. 2C). After treatment for 7 d, established organoids were still able to proliferate and could be passaged but changed their morphology. While organoids in CSC-medium usually grew in a densely packed, grape-like or disheveled structure, we observed formation of huge cysts under FGFR-inhibition. These liquid-filled cysts consisted of a thin layer of epithelial cells with dark patches of apoptotic or necrotic cells shed into the lumen. Cysts also appeared in low frequency in our standard culture conditions, but their occurrence increased significantly under FGFR-inhibition. This observation was also confirmed by immunofluorescence staining on cryo-sections of Matrigel-embedded organoids (Fig. 2D, E, Fig. S2). Tissue polarity of a crypt-lumen axis as in healthy mucosa-derived organoids is usually lost in early adenoma and could not be detected in our organoids as expected.[13, 27] In CSC-medium cultured organoids, we observed an increased cell density compared to FGFR-inhibition. By F-actin staining, we assessed cell morphology and found mostly columnar cells with clear epithelial polarity and nuclei located on the basal side of the cells. However, in organoids derived from Patient 3, inhibition of FGF signaling led mainly to smaller cuboidal cells. In Patient 4, a biopsy of a non-metastatic but poorly differentiated tumor, we observed a less pronounced impact on tissue morphology, corroborating previous observations in poorly differentiated carcinomas.[3] In the FGFR-inhibitor treated sample, we observed lower cell density and a clear lumen-like structure. Interestingly, cells kept their polarized,

epithelial structure but were arranged in a stratified order of multiple layers of cells. Wnt signaling is an important driver of stem cell activity and self-renewal. When the canonical pathway is activated, β-catenin (*CTNNB1*) translocates to the nucleus where it induces gene expression.[28] Therefore, we stained β-catenin on sections of organoids cultured with or without FGFR-inhibition and found β-catenin localized on the lateral cell membranes. As a component of adherens junctions, β-catenin interacts with E-cadherin (*CDH1*) and α-catenin contributing to cell polarity in epithelial cells. Besides a weaker cytoplasmic staining, β-catenin signal was always excluded from the nucleus, independent of culture conditions. The intestinal differentiation marker Cytokeratin-20 (*KRT20*) was heterogeneously expressed within individual organoids (Fig. S2). In some cysts after FGFR-inhibition, the staining appeared stronger compared to CSC-medium treated organoids with high cell density. This trend, however, could not significantly be quantified. The cystic morphology after FGFR-inhibition indicated higher differentiation of these organoids controlled by FGF signaling.

FGF signaling inhibits cell differentiation

Aiming for a deeper understanding of the underlying mechanisms that confer cyst formation and induce differentiation, we used comparative microarray transcriptome analysis of organoids cultured in our standard FGF2-containing CSC-medium or after FGFR-inhibition by SU-5402 treatment. Venn diagrams depicting numbers of regulated genes expressed for each condition indicated a heterogeneous response between all samples analyzed (Fig. 3A, B; Table S7,8). While many different genes responded to the modulation by FGF2, only a few were commonly regulated in all three samples. Next to genes associated with glucose or folate metabolism (e.g. *SLC2A12, SLC2A14, FOLR1*), we identified TGB-β-induced (*TGFBI*) among the highest overrepresented genes in our CSC-medium culture (Table S9), indicating FGF2 downstream effects mediated by TGB-β signaling.

In the set of common genes overexpressed after FGFR-inhibition, we detected genes associated with drug metabolism (e.g. *CYP1A1, CYP51A1*) as well as with proliferation and differentiation (e.g. *EGR1, ATF3, FOSB, JUN*) (Table S8). GO-term analysis of mean overexpressed genes yielded only few significant results due to their low numbers (Fig. 3C, Table S10). Most of these terms for the CSC-medium condition were associated with active transcription, cell proliferation and with an increased glucose uptake. After FGFR-inhibition, we found cell-cycle regulators as well as many genes annotated with "*epithelial cell differentiation"*, "*digestion*" or "*bile acid and bile salt transport*" (Fig. 3D, Table S11).

To additionally include genes with minor but still significant changes, we conducted gene-set enrichment analysis (GSEA) including all expressed genes ranked by ratio.

The induction of several Wnt and TGF-β modulators in CSC-media cultured organoids correlates with gene-sets of colorectal adenomas compared to healthy mucosa or with an adult stem cell signature (Fig. 3E, F, Table S12).[29, 30] An active TGF-β signaling can induce epithelial-mesenchymal-transition (EMT), a critical process for cancer's invasiveness and dissemination.[8] Genes directly associated with EMT together with a mesenchymal and self-renewing phenotype (*TGFB3, ID1, ID2* and *ID4*) were found to be overexpressed in organoids of the CSC-medium condition, while their antagonist, *BMP4* and other epithelial genes such as E-cadherin (*CDH1*), *FOS, ATF3, EGR1 or DUSP1* were induced by FGFR-inhibition (Fig. 3G).[8, 31, 32] Interestingly, these genes belong to the group of immediate-early genes (IEGs) which are usually transiently induced by growth factors for example, EGF or FGF2.[6, 33] We, however, detected these genes induced after 7 d of FGFR-inhibition, when no growth factors were added to the medium. In addition to the constitutive expression of the IEGs, we found several EGFR target genes up-regulated and, consequently, many EGF signaling-related gene-sets enriched in our gene list of FGFR-inhibitor treated organoids (Fig. 3H).[34] Most prominent was the persistent up-regulation of the subunits of the Activator Protein-1 (AP-1) which is a homo- or heterodimer of proteins of the FOS family (*FOS, FOSB, FOSL1, FOSL2*) and the JUN family (*JUN, JUNB, JUND*) as well as the additional interaction partner *ATF3* (Fig. 3I).[35] These genes are associated with a strong MAPK signaling. Since many MAPK inhibiting dual specificity phosphatases (DUSP) were also upregulated (e.g. *DUSP1, DUSP4, DUSP5, DUSP8)*, we analyzed the phosphorylation of ERK, the most downstream kinase of the MAPK cascade. An increased ERK-phosphorylation is well described in RAS-mutant as well as in chemorefractory cancers which applies for most our patients. However, we found an altered ERK-phosphorylation induced by FGFR-inhibition in all patients. This might imply that FGF signaling can modulate MAPK-phosphorylation in RAS-mutant (Patient 1,3,4) as well as in RAS-wild-type patients (Patient 2).

Besides the described effects detected in all patients, we also found individual differences in each patient (Table S13 and Figure S3)**.** The strongest correlation of differentiation and increased proliferation under FGFR-inhibition but induced self-renewal under CSC-medium treatment was observed in Patient 1. We found several gene-sets and pathways overrepresented in the FGFR-inhibitor condition not only connected to RAS-induced MAPK activity, but also enrichment of genes regulated by STAT3, JNK and AKT as well as an increased expression of the proliferation marker *MKI67*. In CSC-medium cultured organoids, on the other hand, we detected an induction of TGF-β and of all four members of the inhibitor-of-differentiation (ID) gene family, which are directly involved in self-renewal of colon cancer stem cells.[31] Most significant induction of genes associated with stem cells was found in Patient 3 in the CSC-medium condition. Several gene-sets of breast and glioma CSC and of resistance against alkylating against, where induced. On the single-gene-level, an up-regulation of *TGFB2* and *INHBA* together with the Wnt-target genes *LGR5*, *TCF4* and *LEF1* was most prominent in organoids derived from this biopsy. In Patient 4, the gene expression pattern was also dominated by a strong RAS-ERK signaling under FGFR-inhibition. Distinct to other patients, however, was a slight downregulation of *TGFB1* under CSC-medium treatment as well as a downregulation of *ID1*, *ID2* and *ID3*, but at the same time a reduced expression of the differentiation markers *BMP2* and *KRT20* and a simultaneous up-regulation of the stem cell factor *SOX2*.
These results imply that FGF2 signaling has heterogeneous molecular effects among different patients, but it has also common effects in the inhibition of cell differentiation, a hallmark of self-renewal.

5. Discussion

Today's living organoid biobanks comprise nearly all kinds of colon cancer subtypes with 60- 100% efficiency of culture establishment.[2, 3, 12, 36] While most projects focus on primary cancer tissue, metastatic biopsies are rarely analyzed.[2, 37, 38] From many organoid-based studies, it became clear that niche requirements and, by that, cell culture specifications reduce with tumor progression while the number of putative CSCs increases.[1] However, it is still challenging to culture primary or metastatic colorectal cancer tissue in a way that self-renewal and self-organization are maintained. In some protocols of adult and CSC culture, FGF2 is an important ingredient for maintaining stemness, while the human intestinal stem cell (HISC) medium described by Sato and Clevers uses a complex modulation of the Wnt, ALK/BMP, TGF-β and p38-MAPK pathways by Rspo1, Noggin, A83-01 and SB202190.[11, 12, 36] Our medium contained only serum replacement supplements as well as EGF and FGF2. Pathways shared by hESC and metastatic cancers corroborated with our hypothesis that FGF2 regulates self-renewal via TGF-β modulation. While the HISC-medium uses A83-01, an inhibitor of TGFβR1 (*ALK5*), to avoid TGF-β-induced cytostasis, we observed an induced TGF-β expression after FGF2 stimulation without any growth inhibitory effects. This tumor promoting effect of TGF-β has been reported to be restricted to highly mutated and progressed cancers and might further be context dependent.[8]
The fact that other protocols omit FGF2 in their culture media but use, depending on tumor's stage, other activators of different pathways instead, indicates that self-renewal can be activated via different pathways which have many interconnections.[4] We indeed found differential expression of genes of other well-described pathways such as the NFκB- or the Notch-pathway.
Increased expression of FGFRs as in prostate or in non-small cell lung cancer (NSCLC) is not described in CRC.[39, 40] However, in a genome-scale analysis by The Cancer Genome Atlas (TCGA) 26% (51/195) of all patients had genomic or transcriptomic alterations, mostly mRNA up-regulation, in at least one of the FGFR genes or its adaptor proteins FRS2/3 indicating that cancer cells might benefit from an increased FGF signaling.[5, 41]
To our best knowledge, this is the first study analyzing the effect of FGF2 signaling on CSC-organoid culture. The small molecule-SU-5402, used for FGFR-inhibition, also inhibits auto-phosphorylation of the FGFR1.[26] This active suppression of FGFR downstream effects might have broader consequences than only omitting FGF stimulation. When we inhibited FGFR signaling, no EGF was present in the culture medium. Interestingly, we found many enriched gene-sets of an induced EGF-pathway after FGFR-inhibition. Similar to observations in patients treated with anti-EGFR therapy, we detected increased RAS signaling and ERK-phosphorylation. For some cancers, for example NSCLC, it has been described that anti-EGFR therapy induced FGFR signaling.[40] In many aspects, we found the reversed effect of an increased endogenous EGF signaling after FGFR-inhibition.

It has been shown for MAPK signaling that regulation is preserved in RAS-mutant tumors.[42] While FGFR-inhibition led to a stronger MAPK signaling indicated by AP-1 induction and increased ERK-phosphorylation, we also observed induced differentiation, a loss of most stem cell markers and an epithelial phenotype. This contrasts with results of Blaj et al., who reported EMT-induction together with increased expression of stem cell antigens CD44, ASCL2 and EPHB2 upon MAPK stimulation in KRAS wild-type and mutant CRC cells.[42]

As for MAPK signaling in RAS-mutant tumors, a modulation of β-catenin and the Wnt signaling pathway has been described in APC-mutant CRC. Even though we did not add any direct Wnt regulators to our medium, many Wnt-modulating proteins, such as *DKK1/2/4, LGR5, NOTUM or APCDD1* were regulated by differential FGF signaling. Nuclear translocation of β-catenin, however, was not observed in our organoids. In a study of patients with APC-mutant CRC, nuclear translocation was restricted to the proliferative front of highly invasive cancers.[43] In intestinal organoids of mice with floxed APC alleles ($APC^{fl/fl}$), β-catenin was located to the lateral cell membranes.[27] Control wild-type organoids in this study grew in a highly branched, crypt-like morphology, while $APC^{fl/fl}$-organoids grew as cysts. In this context, cyst-formation indicates a loss of tissue-polarity representing a model of an early and well differentiated adenoma. Only in more advanced cancers with acquired mutations grape-like organoids with high cell density can be observed.[3, 27] In our experiments, cyst-formation increased massively under FGFR-inhibition and accordingly, transcriptional changes indicated differentiation. It can be assumed that the heterogeneity of cells within cysts after FGFR-inhibition was decreased while the RAS-mutated phenotype became dominant, also inducing increased proliferation reflected by increased *MKI67* expression.

With ongoing culturing, cells adapt to culture conditions, which might be accompanied with a loss of heterogeneity and with an increased robustness against unfavorable culture conditions. This might explain why FGFR-inhibition led to a complete suppression of organoid formation on naïve cells, while established organoids changed their morphology but kept, or even increased, their ability to proliferate. Additionally, the presence of cell-cell-contact in already established organoids allows juxtacrine and paracrine signaling, which might increase the viability of the cells.

Selection of highly proliferative clones, and by that neglecting quiet or dormant stem cells, is a drawback of many cell culture models and might apply for our organoid cultures as well. By demonstrating that a subpopulation of CSCs is sensitive to FGF-modulation, we tried to highlight one of many aspects of CSCs great heterogeneity, maybe the biggest challenge in cancer research and its treatment.

Acknowledgements

This work was supported by the Düsseldorf School of Oncology (funded by the Comprehensive Cancer Centre Düsseldorf/Deutsche Krebshilfe and the Medical Faculty, Heinrich Heine University Düsseldorf).

J.A. and N.H.S. acknowledge support from the Medical faculty of the Heinrich Heine University Düsseldorf.

Authors contribution

Study concept and design: J.O., N.H.S., J.A.
Performed the experiments: J.O., B.B., W.G.
Analyzed and interpreted the data: J.O., L.D., B.B., W.G., W.W., N.H.S., J.A.
Contributed reagents, materials, provided facilities: L.D., W.G., W.T.K., N.H.S., J.A.,
Wrote the paper: J.O.
Critical revision of the manuscript for important intellectual content: N.H.S., J.A.
Study supervision: N.H.S., J.A.
All Authors read and approved the final manuscript

6. References

1. Batlle E, Clevers H. Cancer stem cells revisited. *Nat Med* 2017;**23**: 1124-34.

2. Drost J, Clevers H. Organoids in cancer research. *Nat Rev Cancer* 2018.

3. Fujii M, Shimokawa M, Date S, Takano A, Matano M, Nanki K, Ohta Y, Toshimitsu K, Nakazato Y, Kawasaki K, Uraoka T, Watanabe T, et al. A Colorectal Tumor Organoid Library Demonstrates Progressive Loss of Niche Factor Requirements during Tumorigenesis. *Cell Stem Cell* 2016;**18**: 827-38.

4. Pan T, Xu J, Zhu Y. Self-renewal molecular mechanisms of colorectal cancer stem cells. *Int J Mol Med* 2017;**39**: 9-20.

5. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012;**487**: 330-7.

6. Greber B, Lehrach H, Adjaye J. Fibroblast growth factor 2 modulates transforming growth factor beta signaling in mouse embryonic fibroblasts and human ESCs (hESCs) to support hESC self-renewal. *Stem Cells* 2007;**25**: 455-64.

7. He XC, Zhang J, Tong WG, Tawfik O, Ross J, Scoville DH, Tian Q, Zeng X, He X, Wiedemann LM, Mishina Y, Li L. BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-beta-catenin signaling. *Nat Genet* 2004;**36**: 1117-21.

8. Massague J. TGFbeta in Cancer. *Cell* 2008;**134**: 215-30.

9. Calon A, Lonardo E, Berenguer-Llergo A, Espinet E, Hernando-Momblona X, Iglesias M, Sevillano M, Palomo-Ponce S, Tauriello DV, Byrom D, Cortina C, Morral C, et al. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat Genet* 2015;**47**: 320-9.

10. Danopoulos S, Schlieve CR, Grikscheit TC, Al Alam D. Fibroblast Growth Factors in the Gastrointestinal Tract: Twists and Turns. *Developmental Dynamics* 2017;**246**: 344-52.

11. Prasetyanti PR, Zimberlin C, De Sousa EMF, Medema JP. Isolation and propagation of colon cancer stem cells. *Methods Mol Biol* 2013;**1035**: 247-59.

12. Sato T, Stange DE, Ferrante M, Vries RG, Van Es JH, Van den Brink S, Van Houdt WJ, Pronk A, Van Gorp J, Siersema PD, Clevers H. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* 2011;**141**: 1762-72.

13. Sato T, Vries RG, Snippert HJ, van de Wetering M, Barker N, Stange DE, van Es JH, Abo A, Kujala P, Peters PJ, Clevers H. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* 2009;**459**: 262-5.

14. Jung P, Sato T, Merlos-Suarez A, Barriga FM, Iglesias M, Rossell D, Auer H, Gallardo M, Blasco MA, Sancho E, Clevers H, Batlle E. Isolation and in vitro expansion of human colonic stem cells. *Nat Med* 2011;**17**: 1225-7.

15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of Molecular Biology* 1990;**215**: 403-10.

16. Kent WJ, Sugnet Cw Fau - Furey TS, Furey Ts Fau - Roskin KM, Roskin Km Fau - Pringle TH, Pringle Th Fau - Zahler AM, Zahler Am Fau - Haussler D, Haussler D. The human genome browser at UCSC. *Genome Research* 2002;**12**: 1088-9051 (Print).

17. Gomes IC, Acquarone M, Maciel Rde M, Erlich RB, Rehen SK. Analysis of pluripotent stem cells by using cryosections of embryoid bodies. *J Vis Exp* 2010.

18. Graffmann N, Ring S, Kawala MA, Wruck W, Ncube A, Trompeter HI, Adjaye J. Modeling Nonalcoholic Fatty Liver Disease with Human Pluripotent Stem Cell-Derived Immature Hepatocyte-Like Cells Reveals Activation of PLIN2 and Confirms Regulatory Functions of Peroxisome Proliferator-Activated Receptor Alpha. *Stem Cells Dev* 2016;**25**: 1119-33.

19. Mohlendick B, Bartenhagen C, Behrens B, Honisch E, Raba K, Knoefel WT, Stoecklein NH. A robust method to analyze copy number alterations of less than 100 kb in single cells using oligonucleotide array CGH. *PLoS One* 2013;**8**: e67031.

20. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;**4**: 44-57.

21. Wang B, Yong H, Zhu H, Ni D, Tang S, Zhang S, Wang W, Zhou Y, Zhao W, Ding G, Zhu J, Li X, et al. Abnormal amphiregulin expression correlates with gastric cancer prognosis. *Oncotarget* 2016;**7**: 76684.

22. Riese DJ, 2nd, Cullum RL. Epiregulin: roles in normal physiology and cancer. *Semin Cell Dev Biol* 2014;**28**: 49-56.

23. Alimperti S, Andreadis ST. CDH2 and CDH11 act as regulators of stem cell fate decisions. *Stem Cell Res* 2015;**14**: 270-82.

24. Flores ER, Halder G. Stem cell proliferation in the skin: alpha-catenin takes over the hippo pathway. *Sci Signal* 2011;**4**: pe34.

25. Misale S, Di Nicolantonio F, Sartore-Bianchi A, Siena S, Bardelli A. Resistance to anti-EGFR therapy in colorectal cancer: from heterogeneity to convergent evolution. *Cancer Discov* 2014;**4**: 1269-80.

26. Cheng W, Wang M, Tian X, Zhang X. An overview of the binding models of FGFR tyrosine kinases in complex with small molecule inhibitors. *Eur J Med Chem* 2017;**126**: 476-90.

27. Fatehullah A, Appleton PL, Nathke IS. Cell and tissue polarity in the intestinal tract during tumourigenesis: cells still know the right way up, but tissue organization is lost. *Philos Trans R Soc Lond B Biol Sci* 2013;**368**: 20130014.

28. Krausova M, Korinek V. Wnt signaling in adult intestinal stem cells and cancer. *Cell Signal* 2014;**26**: 570-9.

29. Sabates-Bellver J, Van der Flier LG, de Palo M, Cattaneo E, Maake C, Rehrauer H, Laczko E, Kurowski MA, Bujnicki JM, Menigatti M, Luz J, Ranalli TV, et al. Transcriptome profile of human colorectal adenomas. *Mol Cancer Res* 2007;**5**: 1263-75.

30. Boquest Ac, Shahdadfar S, Frønsdal K, Sigurjonsson O, Tunheim SH, Collas P, Brinchmann JE. Isolation and Transcription Profiling of Purified Uncultured Human Stromal Stem Cells: Alteration of Gene Expression after In Vitro Cell Culture. *Mol Biol Cell* 2005;**16**: 1131-41.

31. O'Brien CA, Kreso A, Ryan P, Hermans KG, Gibson L, Wang Y, Tsatsanis A, Gallinger S, Dick JE. ID1 and ID3 regulate the self-renewal capacity of human colon cancer-initiating cells through p21. *Cancer Cell* 2012;**21**: 777-92.

32. Jechlinger M, Grunert S, Tamir IH, Janda E, Ludemann S, Waerner T, Seither P, Weith A, Beug H, Kraut N. Expression profiling of epithelial plasticity in tumor progression. *Oncogene* 2003;**22**: 7155-69.

33. Bahrami S, Drablos F. Gene regulation in the immediate-early response process. *Adv Biol Regul* 2016;**62**: 37-49.

34. Nagashima T, Shimodaira H, Ide K, Nakakuki T, Tani Y, Takahashi K, Yumoto N, Hatakeyama M. Quantitative transcriptional control of ErbB receptor signaling undergoes graded to biphasic response for cell differentiation. *J Biol Chem* 2007;**282**: 4045-56.

35. Ashida R, Tominaga K, Sasaki E, Watanabe T, Fujiwara Y, Oshitani N, Higuchi K, Mitsuyama S, Iwao H, Arakawa T. AP-1 and colorectal cancer. *Inflammopharmacology* 2005;**13**: 113-25.

36. van de Wetering M, Francies HE, Francis JM, Bounova G, Iorio F, Pronk A, van Houdt W, van Gorp J, Taylor-Weiner A, Kester L, McLaren-Douglas A, Blokker J, et al. Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell* 2015;**161**: 933-45.

37. Weeber F, Van de Wetering M, Hoogstraat M, Dijkstra KK, Krijgsman O, Kuilman T, Gadellaa-van Hooijdonk CGM, van der Velden D, Peeper DS, Cuppen E, Vries R, Clevers H, et al. Preserved genetic diversity in organoids cultured from biopsies of human colorectal cancer metastases. *PNAS* 2015;**112**: 13308-11.

38. Vlachogiannis G, Hedayat S, Vatsiou A, Jamin Y, Fernández-Mateos Y, Khan K, Lampis A, Robinson SP, Cunningham D, Valeri N. Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science* 2018;**359**: 920 6.

39. Ahmad I, Iwata T, Leung HY. Mechanisms of FGFR-mediated carcinogenesis. *Biochim Biophys Acta* 2012;**1823**: 850-60.

40. Luo M, Fu L. Redundant kinase activation and resistance of EGFR-tyrosine kinase inhibitors. *Am J Cancer Res* 2014;**4**: 608-28.

41. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012;**2**: 401-4.

42. Blaj C, Schmidt EM, Lamprecht S, Hermeking H, Jung A, Kirchner T, Horst D. Oncogenic Effects of High MAPK Activity in Colorectal Cancer Mark Progenitor Cells and Persist Irrespective of RAS Mutations. *Cancer Res* 2017;**77**: 1763-74.

43. Brabletz T, Jung A, Reu S, Porzner M, Hlubek F, Kunz-Schughart LA, Knuechel R, Kirchner T. Variable beta-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment. *Proc Natl Acad Sci U S A* 2001;**98**: 10356-61.

Supplementary Methods

Microarray transcriptome analysis

As described before, gene expression was determined using a detection p-value <0.05 (1). Transcriptome data of the hESC lines H1 and H9 were obtained and described before (2). Mean expression of H1 and H9 was calculated and used for comparative analyses. Expressed genes in different cell types or under different conditions were compared by Venn diagrams using the R package VennDiagram (3). The limma Bioconductor package and the p-value adjustment method from the q-value package were used to assess differential expression (4,5). Genes were graded as differentially expressed for limma p-value <0.05. Genes were considered as upregulated with ratio >1.33, or as downregulated with ratio < 0.75.
Gene-sets of commonly or individually expressed genes by different cell types, or significantly up- or downregulated after SU-5402 treatment were further analyzed for Gene Ontology terms of biological processes (GO-BP) using the DAVID webtool (https://david.ncifcrf.gov/) (6).
Gene-set enrichment analysis (GSEA) was performed with the GSEA java application downloaded from Broad institute for the comparison of all treated patients together versus all controls together and on the level of individual treated versus untreated patients. For the joint comparison n=1000 permutations and default parameters were used, for the individual patient comparisons also n=1000 permutations and default parameters were used except for the metric "Diff of classes" and the permutation over the "gene_set" (7).

Comparative Genomic Hybridization with Oligonucleotide Microarrays (aCGH) and data analysis

As describe before, 1 µg of DNA was used for restriction digestion of gDNA using restriction enzymes RsaI and AluI to obtain fragments of 200-500 bp (8). Digested gDNA was random-primed labeled with Cyanine-5 or Cyanine-3-dUTP and purified afterwards. After determining sufficient labeling efficiency, samples were hybridized to the 4x180k platform. For evaluation, the Microarray Scanner G2565CA (Agilent Technologies) was used (3mm resolution, 16-bit color depth). A gender-mismatched reference DNA was used as an internal technical control for each sample. Aberration calling was determined by the HaarSeq-based algorithm using the following filter thresholds: ±3 oligos and ± 0.2 log2ratio (9).

Table S1.
Oligonucleotides for BRAF, KRAS and NRAS amplification or sequencing, respectively.

Gene	Exon	Name	Seq (5'à3')	Used for sequencing
BRAF	15	BRAF_15_for	TGCTTGCTCTGATAGGAAAATG	Yes
BRAF	15	BRAF_15_rev	AGCCTCAATTCTTACCATCCA	No
KRAS	2	KRAS_2_for	AGGCCTGCTGAAAATGACTGAA	No
KRAS	2	KRAS_2_rev	AAAGAATGGTCCTGCACCAG	Yes
KRAS	3	KRAS_3_for	GGATTCCTACAGGAAGCAAGT	No
KRAS	3	KRAS_3_rev	TGGCAAATACACAAAGAAAGC	Yes
KRAS	4	KRAS_4_for	GGACTCTGAAGATGTACCTATGG	Yes
KRAS	4	KRAS_4_rev	TCAGTGTTACTTACCTGTCTTGT	Yes
NRAS	2	NRAS_2_for	ACAGGTTCTTGCTGGTGTGA	No
NRAS	2	NRAS_2_rev	CACTGGGCCTCACCTCTATG	Yes
NRAS	3	NRAS_3_for	GTGGTTATAGATGGTGAAACCTGT	No
NRAS	3	NRAS_3_rev	TGGCAAATACACAGAGGAAGC	Yes
NRAS	4	NRAS_4_for	TTCCCGTTTTTAGGGAGCAGA	Yes
NRAS	4	NRAS_4_rev	TGCAAACTCTTGCACAAATGC	Yes

Table S2.
Antibodies used for Western Blotting and Immunocytochemistry (mAB: monoclonal antibody, pAB: polyclonal AB)

Target	Host	Catalog number	Dilution
Total ERK, p44/42 MAP Kinase	mouse mAB	CST #4696	1:1,000
Phospho-ERK, p44/42 (Thr202/Tyr204)	rabbit mAB	CST #4370	1:1,000
anti-rabbit IgG, HRP-linked	goat	CST #7074	1:2,000
anti-mouse IgG, HRP-linked	goat	abcam, ab6789	1:10,000
Cytokeratin-20	mouse mAB	abcam, ab854	1:200
β-catenin	rabbit pAB	CST #9562	1:200

Supplementary References

1. Graffmann N, Ring S, Kawala MA, Wruck W, Ncube A, Trompeter HI, et al. Modeling Nonalcoholic Fatty Liver Disease with Human Pluripotent Stem Cell-Derived Immature Hepatocyte-Like Cells Reveals Activation of PLIN2 and Confirms Regulatory Functions of Peroxisome Proliferator-Activated Receptor Alpha. Stem Cells Dev 2016;25(15):1119-33.
2. Bohndorf M, Ncube A, Spitzhorn LS, Enczmann J, Wruck W, Adjaye J. Derivation and characterization of integration-free iPSC line ISRM-UM51 derived from SIX2-positive renal cells isolated from urine of an African male expressing the CYP2D6 *4/*17 variant which confers intermediate drug metabolizing activity. Stem Cell Res 2017;25:18-21.
3. Chen H, Boutros PM. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 2011;12(35).
4. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004;3(1544-6115 (Electronic)).
5. Storey JD. The positive false discovery rate: a bayesian interpretation and the q-value. The Annals of Statistics 2003;31(6):2013-35.
6. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4(1):44-57.
7. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102(43):15545-50.
8. Mohlendick B, Bartenhagen C, Behrens B, Honisch E, Raba K, Knoefel WT, et al. A robust method to analyze copy number alterations of less than 100 kb in single cells using oligonucleotide array CGH. PLoS One 2013;8(6):e67031.
9. Ben-Yaacov E, Eldar YC. A fast and flexible method for the segmentation of aCGH data. Bioinformatics 2008;24(16):i139-45.

ID	Gender	Age	Location sampled tissue	UICC	Primary Tumor (TNM)	Differentiation	Pretreatment	BRAF/NRAS/ KRAS
Patient 1	M	62	Liver	IV	pT2 pN1 M1hep	Moderate	5-FU, Oxaliplatin, Bevacizumab	wt/wt/G12D
Patient 2	M	64	para rectal LN	IV	pT2 pN2 M1hep	Moderate	5-FU, Oxaliplatin, Cetuximab	wt/wt/wt
Patient 3	F	81	Sigmoid	IV	pT3 pN1 M1hep	Moderate	/	wt/G12V/wt
Patient 4	M	51	Sigmoid	IB	pT2 pN0 M0	Poor	/	wt/G12S/wt
Patient 5	F	59	Rectum	IB	pT2 pN0 M0	Well	Capecitabine, Radiotherapy (50,4 Gy)	
Patient 6	F	58	Rectum		HGIEN		/	
Patient 7	F	85	Rectum	IIA	pT3 pN0 M0	Well	/	
Patient 8	M	74	Liver	IV	pT2 pN0 M1hep	Moderate	/	
Patient 9	M	75	Sigmoid	IIA	pT3 pN0 M0	Well	/	

Table 1. Clinical data
Clinical data of sampled patients. Grey shadow: organoid cultures could be established. Only after successful culture establishment mutational hotspots of the BRAF, NRAS, KRAS loci were sequenced. HGIEN= high-grade intra-epithelial neoplasia; M= male, F= female, LN= lymph node, 5-FU= 5-Fuoruracil, wt= wild-type.

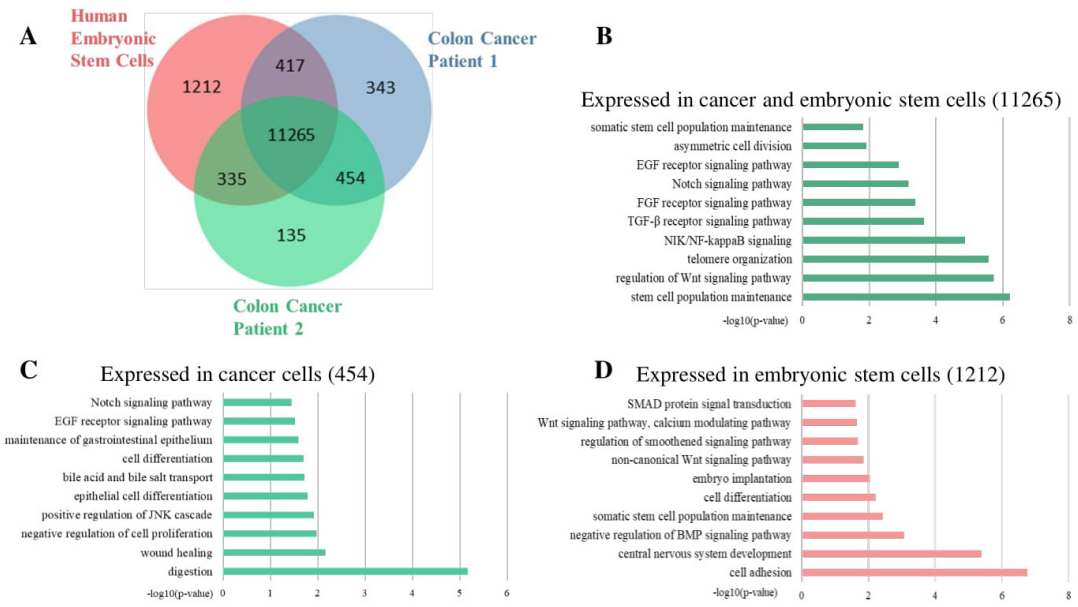


Figure 1.
Comparative microarray transcriptome analysis of genes expressed in human embryonic stem cells and organoids derived from Patient 1 and Patient 2. **A**, Venn diagram showing the number of expressed genes commonly or specifically expressed by each sample. **B-D,** Gene Ontology (GO) terms of significantly enriched gene patterns. Only GOs of biological process with a p-value < 0.05 were considered.

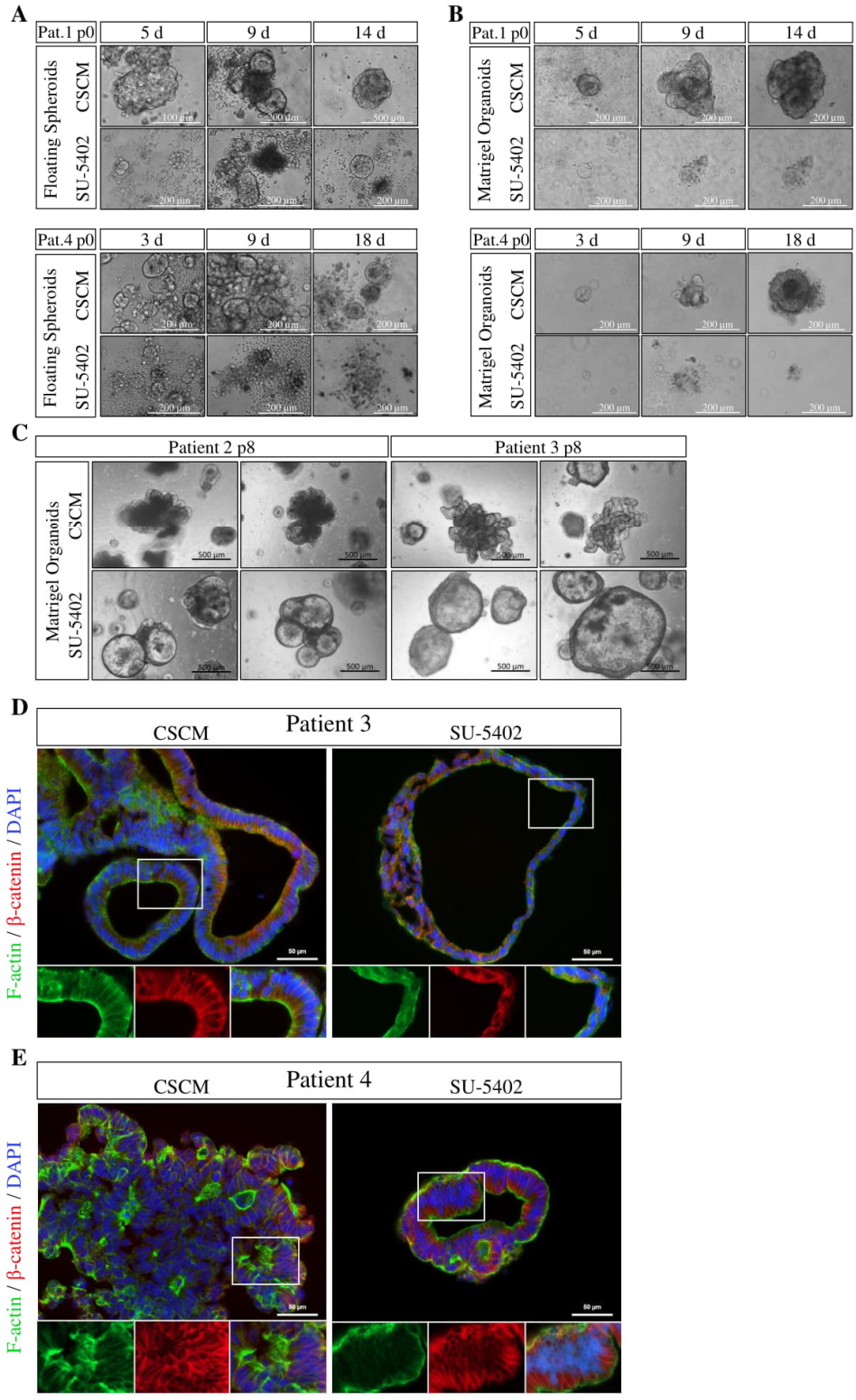


Figure 2.
Morphological Effects of FGFR-inhibition on primary CRC cells. **A,** Cell culture initiation of naïve unsorted bulk cells (p0) after tissue disintegration, cultured in CSC-media (CSCM) or with FGFR-inhibitor (SU-5402) under low-attachment conditions as floating spheres or **B,** cultured as Matrigel-embedded organoids. **C,** Cyst induction by FGFR-inhibition in Matrigel-embedded organoids of p8 after 7 d of SU-5402 treatment. **D, E,** Cryo-sections of organoids cultured with CSCM or SU-5402, stained for F-actin and β-catenin.

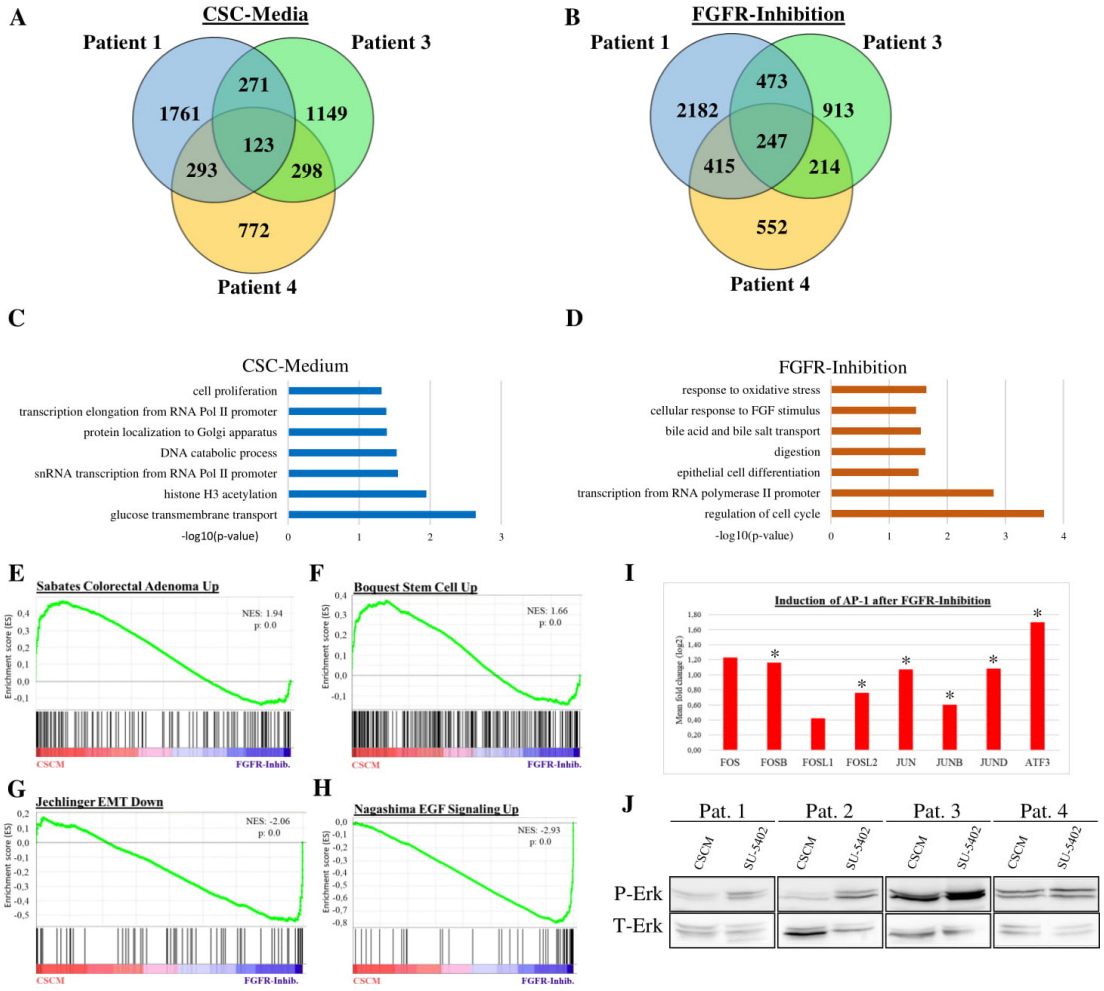


Figure 3.
FGFR-inhibition induces cell differentiation and upregulates MAPK-signaling. **A,** Venn Diagram of genes ($p<0.05$) higher expressed in organoids cultured in CSC-Media or **B,** after FGFR-inhibition. **C,** Significantly enriched GO-terms of biological processes of the genes higher expressed after CSC-Media culturing or **D,** higher expressed after FGFR-inhibition. Only genes with $p<0.05$ and limma-$p<0.05$ were considered for GO-analysis **E-H,** gene sets significantly over represented in the CSC-media condition or the FGFR-inhibition condition. NES= normalized enrichment score **I,** mean fold change (log2) values of single genes belonging to the AP-1 transcription factor complex of Patient 1, 3 and 4 (n=3). Differential expression values were obtained by microarray analysis, * indicates significance as limma p-value <0.05. See Table S9 for log2 ratio of single probes detected. **J,** Western blot of protein lysates of organoids cultured in CSC-Media or with SU-5402. Phosphorylated ERK (P-ERK) is compared to total amount of ERK-protein (T-ERK).

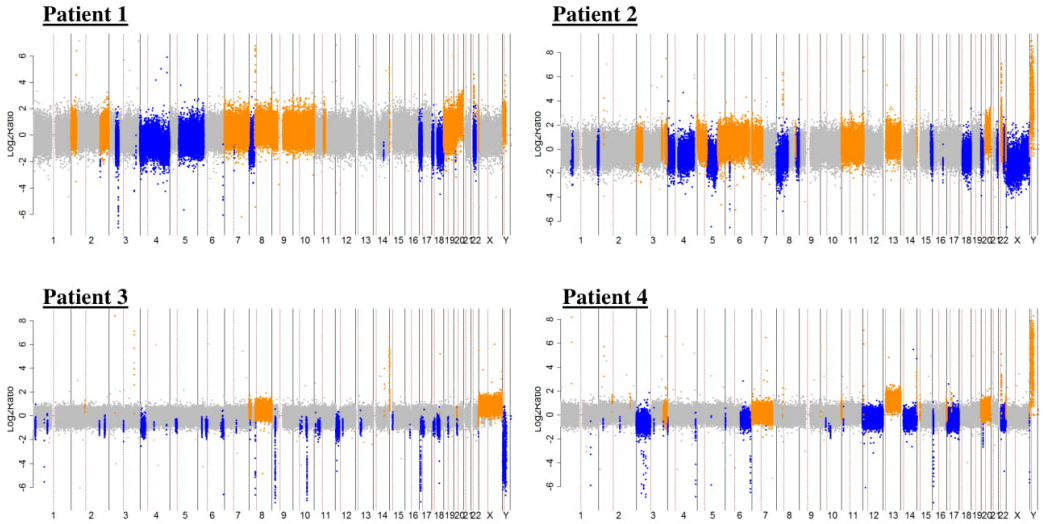


Figure S1.
Call plot of copy number variations (CNV) after comparative genome hybridization (aCGH). The HaarSeg-based algorithm (±3 oligos and ± 0.2 log2ratio) was used. Orange represents a gain, blue a loss of genomic DNA. As an internal control, a gender-mismatched reference DNA was used.

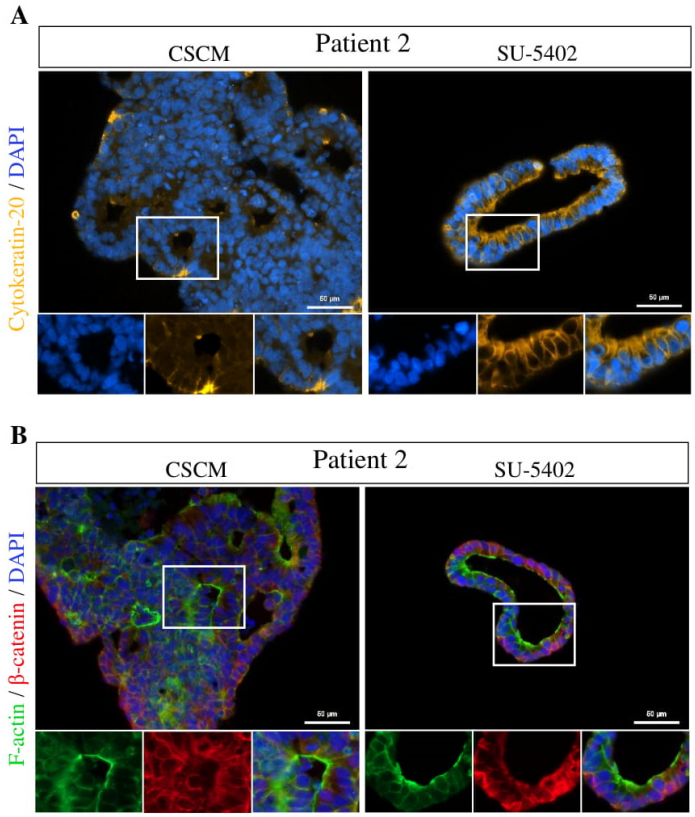


Figure S2.
Cryo-sections of organoids cultured with CSCM or SU-5402 indicating cyst formation after FGFR-inhibition SU-5402, **A,** stained for Cytokeratin 20 or **B,** for F-actin and β-catenin.

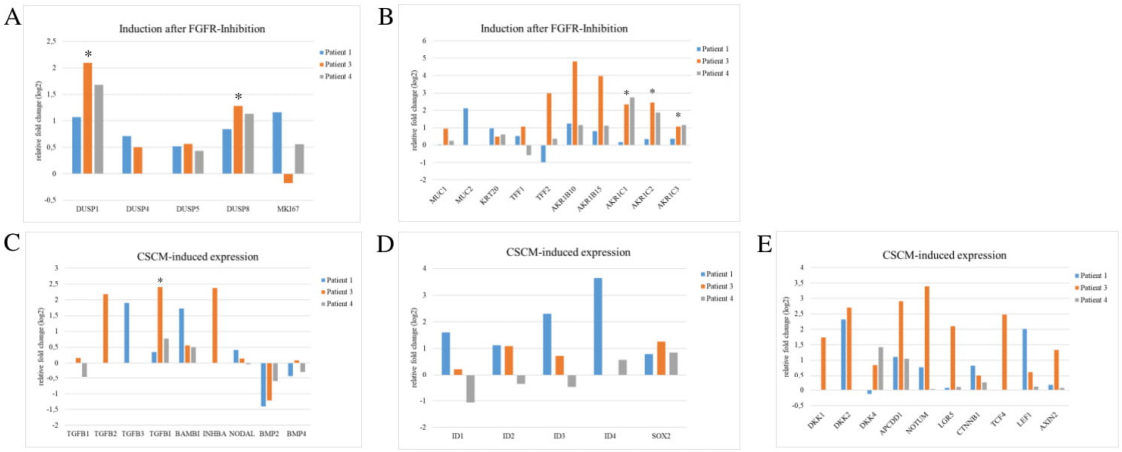


Figure S3
Differential expression of single genes in Patient 1,3 and 4. Expression values were obtained by microarray transcriptome analysis. For each gene in a single patient, the detection p-value is <0.05. * indicates significant regulation among all samples analyzed (limma p-value <0.05, see Table S9). Relative expression compared to FGFR-inhibition **(A, B)** or to CSCM treatment **(C, D, E)** is shown as indicated. **A,** Several genes of dual specificity protein phosphatases (DUSP) as well as the proliferation marker MKI67 were induced by FGFR-inhibition. **B,** Colon-specific differentiation markers were induced by FGFR-inhibition. **C,** In comparison to FGFR-inhibition, the TGF-β pathways was heterogeneously induced by CSCM-treatment. **D,** Self-renewal associated genes were stronger expressed in CSCM-treated organoids of Patient 1 and Patient 3 compared to FGFR-inhibition. **E,** Several Wnt modulating factors were upregulated by CSCM-treatment. The intestinal stem cell factors *LGR5* and *AXIN2* were induced in Patient 3

3.2. Gene expression profiling in aggressive digital papillary adenocarcinoma sheds light on the architecture of a rare sweat gland carcinoma

H. M. Surowy[*], A. K. Giesen[*], **J. Otte**[*], R. Büttner, D. Falkenstein, H. Friedl, F. Meier, P. Petzsch, T. Wachtmeister, D. Westphal, D. Wieczorek, W. Wruck, J. Adjaye, A. Rütten[+], S. Redler[+]

[*], [+] These authors contributed equally to the work.

Abstract

Background: Sweat gland carcinomas are rare cutaneous adnexal malignancies. Aggressive digital papillary adenocarcinoma (ADPA) represent a very rare sub-entity, thought to arise almost exclusively from sweat glands of the fingers and toes. The aetiology of sweat gland carcinomas and ADPA is largely unknown. ADPAs are most likely driven by somatic mutations. However, somatic mutation patterns are largely unexplored, creating barriers to the development of effective therapeutic approaches to the treatment of ADPA.

Objective: To investigate the transcriptome profile of ADPA using a sample of seven formalin-fixed, paraffin embedded (FFPE) tissue samples of ADPA and healthy control tissue.

Methods: Transcriptome profiling was performed using the Affymetrix PrimeView Human Gene Expression Microarray and findings were validated via reverse-transcription of RNA and real-time qPCR.

Results: Transcriptome analyses showed increased tumour expression of 2,083 genes, with significant involvement of cell cycle, ribosomal and crucial cancer pathways. Our results furthermore point to tumour-overexpression of *FGFR2* ($p = 0.003$).

Conclusions: Our results indicate the involvement of crucial oncogenic driver pathways, highlighting cell cycle and ribosomal pathways in the aetiology of ADPA. Suggested tumour-overexpression of *FGFR2* raises the hope that targeting the FGF/FGFR axis might be a promising treatment for ADPA and probably for the overall group of sweat gland carcinomas.

Share of scientific contribution: 25%

Jörg Otte supported the comprehensive gene expression analyses. He was involved in identifying potential marker genes after microarray transcriptome analysis and provided advice in the design of following confirmation steps and the design of specific primers. He conducted all qRT-PCR analyses and did the quality control in tight collaboration with other authors. He wrote the according section in the manuscript.

Published: *British Journal of Dermatology*

Link to the publication: https://onlinelibrary.wiley.com/doi/10.1111/bjd.17446

DR SILKE REDLER (Orcid ID : 0000-0002-0991-252X)

Article type : Original Article

Gene expression profiling in aggressive digital papillary adenocarcinoma sheds light on the architecture of a rare sweat gland carcinoma

H.M. Surowy[1*], A.K. Giesen[1*], J. Otte[2*], R. Büttner[3], D. Falkenstein[1], H. Friedl[4], F. Meier[5,6], P. Petzsch[7], T. Wachtmeister[7], D. Westphal[5,6], D. Wieczorek[1], W. Wruck[2], J. Adjaye[2], A. Rütten[8*], S. Redler[1*]

[1]Heinrich-Heine-University, Medical Faculty, Institute of Human Genetics, Düsseldorf, Germany

[2]Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[3]Pathology, Institute for Pathology, University Hospital Cologne

[4]Department of Dermatology, Darmstadt, Germany

[5]Department of Dermatology, Carl Gustav Carus Medical Center, TU Dresden, Dresden, Germany

[6]National Center for Tumour Diseases (NCT), Partner Site Dresden, Germany

[7]Biological and Medical Research Center (BMFZ), Heinrich-Heine-University, Düsseldorf, Germany

[8]Dermatopathology, Bodensee, Siemensstrasse 6/1, 88048 Friedrichshafen, Germany

* These authors contributed equally to the work.

Correspondence to:
PD Dr. Silke Redler, MD
Institute of Human Genetics
University of Düsseldorf
Universitätsstr. 1
D-40225 Düsseldorf
Germany
Phone: +49 (0)211 81 10423
Fax: +49 (0)211 81 12358
Email: silke.redler@med.uni-duesseldorf.de

Running title: Gene expression profiling in aggressive digital papillary adenocarcinoma

Conflicts of interest: none declared.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/bjd.17446

What is already known about this topic?

Aggressive digital papillary adenocarcinoma (ADPA) is a rare sporadic tumour, arising predominantly from sweat glands of fingers or toes. ADPA is characterized by a local aggressive behaviour, a high recurrence rate, and an emerging metastatic potential. ADPA are most probably driven by somatic mutations. However, somatic mutation patterns are largely unexplored. Knowledge of the biology of ADPA is almost completely lacking, creating barriers to the development of effective therapy strategies.

What does this study add?

We performed gene expression profiling in a well-defined sample of formalin-fixed, paraffin-embedded (FFPE) tissue samples of ADPA. Array-based transcriptome analyses revealed increased tumour expression of 2,266 genes with significant involvement of cell cycle, ribosomal and cancer pathways and suggests tumour over-expression of *FGFR2*. This broadens our understanding of the molecular mechanisms driving these tumours and raises the possibility of potential targeted therapies comprising FGFR kinase inhibitors.

Abstract

Background: Sweat gland carcinomas are rare cutaneous adnexal malignancies. Aggressive digital papillary adenocarcinoma (ADPA) represents a very rare sub-entity, thought to arise almost exclusively from sweat glands of the fingers and toes. The aetiology of sweat gland carcinomas and ADPA is largely unknown. ADPAs are most likely driven by somatic mutations. However, somatic mutation patterns are largely unexplored, creating barriers to the development of effective therapeutic approaches to the treatment of ADPA.

Objective: To investigate the transcriptome profile of ADPA using a sample of eight formalin-fixed, paraffin embedded (FFPE) tissue samples of ADPA and healthy control tissue.

Methods: Transcriptome profiling was performed using the Affymetrix PrimeView Human Gene Expression Microarray and findings were validated via reverse-transcription of RNA and real-time qPCR.

Results: Transcriptome analyses showed increased tumour expression of 2,266 genes, with significant involvement of cell cycle, ribosomal and crucial cancer pathways. Our results furthermore point to tumour-overexpression of *FGFR2* (p = 0.001).

Conclusions: Our results indicate the involvement of crucial oncogenic driver pathways, highlighting cell cycle and ribosomal pathways in the aetiology of ADPA. Suggested tumour-overexpression of *FGFR2* raises the hope that targeting the FGF/FGFR axis might be a promising treatment for ADPA and probably for the overall group of sweat gland carcinomas.

Introduction

Sweat gland carcinomas are rare cutaneous adnexal malignancies. They display as a heterogeneous group of well-defined tumour entities with clear differences regarding clinical course and prognosis.[1,2] Aggressive digital papillary adenocarcinoma (ADPA) represents a very rare, clinically distinctive, homogeneous sub-entity, thought to arise almost exclusively from sweat glands of the fingers and toes.[1–4]. ADPA occurs in both sexes with a supposed peak incidence between age 50 and 70 years.[5–7] Typically, ADPA presents as a slow-growing, painless, solid to cystic mass on the digits or toes, often inconspicuous of a malignant process.[1,2] Originally described as a tumour spectrum from benign digital papillary adenoma to an aggressive digital papillary adenocarcinoma, it turns out by clinical follow-up data, that all tumours from the spectrum behave as sweat gland carcinomas with an infiltrative growth pattern, a tendency to recur and to give rise to metastases. Histologically the tumours are characterized by a solid or cystic growth pattern consisting of basaloid tumour cells (Fig. 1a, Fig1b). The neoplasm often shows small papillae, which attributed to the name digital papillary adenocarcinoma (Fig. 1c, Fig. 2a, Fig. 2b). Cytology covers a wide spectrum from totally benign looking tumours to a clear carcinomatous growth pattern (Fig. 2b). Immunohistochemistry shows a reproducible feature: all tubular structures stain with cytokeratin 7 (Fig. 1d, Fig. 2d) and they are surrounded by an actin positive layer of myoepithelial cells at the periphery (Fig. 1e, Fig. 2c). These findings are helpful to separate these tumours from poromas, hidradenomas and myoepitheliomas in acral skin (Fig. 1 and 2).

The course of disease and the biological behaviour of ADPA is variable. However, ADPA are characterized by aggressive local invasion, a high recurrence rate and an emerging metastatic potential.[1,2,8–10] To date, surgical excision of the tumour lesion is the mainstay of treatment, comprising radical, wide local tumour excision, often including amputation of the affected digit.[5,7,11–14] There are no approved treatment approaches for metastatic disease. Its aetiologic basis has remained largely undefined, raising the question how to proceed in case of this rare tumour.

This stresses the demand for insights into the genomic landscape and the genetic alterations that drive this malignancy. To date, the genetic architecture of ADPA is largely unexplored. In a cohort of 14 distinct metastasizing adnexal carcinomas, comprising one ADPA, Dias-Santagata et al. identified a somatic mutation in the established tumour suppressor gene *TP53*.[15] In 2015, Bell et al. reported a

Sequenom genotyping based approach of 50 genes of nine ADPA cases and solely identified the somatic *BRAF*-V600E mutation in one single ADPA case.[16] To the best of our knowledge, no other mutations have been reported. No exome or genome-wide sequencing approach and no transcriptome analysis has been performed so far.

Taken together, understanding of the biology of ADPA is largely missing, creating barriers to the development of effective, targeted therapeutic approaches.

We aimed to establish a more in-depth understanding of the pathogenic causes of ADPA by performing genome-wide transcriptome profiling for the first time. It is well established that gene expression profiling have led to more individually tailored recommendations in skin cancer patients.[17–20]

Material and methods

Patients/samples

A total of eight formalin-fixed, paraffin-embedded (FFPE) tissue samples of ADPA were included in the present study. The tumours were obtained from eight unrelated individuals of Central European origin (5 women, 3 men). The samples were primary tumours in all cases and no recurrences or metastases. Inclusion criterion was the histological diagnosis of ADPA by the experienced and designated histopathologists R. B. and A. R.. Age at diagnosis ranged from 36 to 96 years (mean age at onset 60.25 years). The tumours arose on the hand/ fingers (n = 7) and the foot (n = 1) (Table 1). Familial occurrence was reported in none of the patients. Matched surrounding healthy tissue, respectively blood samples, were used as controls. Ethics approval was obtained from the respective ethics committees, and all participants provided written informed consent (Ethics vote 5360/13, Heinrich Heine University Düsseldorf). The study was conducted in accordance with the Declaration of Helsinki Principles.

RNA extraction:

Distinction of tumour and surrounding healthy tissue by microdissection was performed according to standard procedures. 10µm-thick sections were cut from the FFPE samples. Deparaffination and RNA isolation was done using the FormaPure DNA kit (BECKMANN COULTER Life Science®) according to the protocol for isolation of total nucleic acids (DNA/RNA) from FFPE tissues using FormaPure DNA with minor modifications. RNA from blood samples was isolated according to standard procedures using the PAXgene Blood RNA Kit (Qiagen®).

Transcriptome Array:

Transcriptome analyses were performed in an explorative set of four tumour samples and matched surrounding healthy tissue. RNA was quantified by fluorometric Qubit assay (RNA HS, thermofisher scientific) measurement. Synthesis of cDNA and subsequent biotin labeling of cRNA was performed according to the manufacturer's protocol (GeneChip 3' IVT Pico Kit 703308 Rev 1; Thermofisher Scientific). Briefly, 50 ng of total RNA was converted to cDNA and preamplified (8 cycles), followed by in vitro transcription and biotin labeling of aRNA. After fragmentation, labeled aRNA was hybridized to Affymetrix PrimeView Human Gene Expression Microarrays for 16 h at 45°C, stained by streptavidin/phycoerythrin conjugate, and scanned as described in the manufacturers' protocol.

qPCR Validation:

The criteria to select genes for validation were as follows: > 1 transcript significantly detected in the tumour (signal quality significantly above the background), with at least one transcript exhibiting a fold change (FC) > 6 and a significant limma p-value, as well as absence of other transcripts with reverse FC ratios. Genes found to be significantly differentially expressed in the initial KEGG cancer pathway analysis but displayed a FC < 6 were also included if they corresponded to the other criteria, resulting in four further genes (Figure 3, Table 2). Pseudogenes were excluded.

Validation of the selected candidate genes was done via reverse-transcription of RNA and real-time qPCR, using *ACTB* (beta Actin) as internal reference gene. Reverse transcription was performed with the AccuScript High Fidelity First Strand cDNA Synthesis Kit (Agilent Technologies®) kit according to the manufacturer's instructions. For each reaction 5 ng of RNA was used as starting material and eluted in 20 µl. Quantitative real-time PCR (qPCR) was performed using the Power SYBR® Green Master Mix (Life technologies, ThermoFisher Scientific®) according to the manufacturer's protocol in a 384-well format and a total volume of 10 µl. Primer sequences were designed to span exon borders and are available upon request. All amplifications were performed in technical triplicates on the ViiA7 (Life technologies, ThermoFisher Scientific®) system using the following PCR conditions: 95°C, 10 Min; followed by 45 cycles of 95°C, 60°C and 72°C (30 s each step). Cycle threshold (Ct) mean values per triplicate were generated. If a gene was not expressed in a control tissue sample, but firmly in the corresponding tumour sample, the maximum Ct was assumed for downstream calculations. The mean Ct values were then normalized to *ACTB* (ΔCt method) and fold changes between tumour and control tissue pairs were calculated as $2^{(-\Delta\Delta Ct)}$.

Statistical analysis:

Tumour and control tissue samples hybridized on the Affymetrix PrimeView Human Gene Expression microarray platform were analysed with R/Bioconductor.[21] The Bioconductor package 'affy' was used for pre-processing and normalization of the microarray data via the Robust Multi-Array average (RMA) normalization method.[22] Gene expression was determined by a detection p-value as described in Graffmann et al.[23] Differential expression was assessed employing the 'limma' package.[24] The

clustering dendrogram was generated using the R function 'hclust' parametrized with the complete linkage method and Pearson correlation as similarity measure. The heatmap was produced via the 'heatmap2()' function from the 'gplots' package.[25] Gene expression in normal RNA samples was compared with gene expression in matched tumour samples using the package 'VennDiagram'.[26] From the Venn diagram the subsets exclusively expressed in tumours, exclusively expressed in common in tumour and control tissue were further investigated in follow-up processing. Differentially expressed genes were determined based on a limma p-value for differential expression, a threshold for the ratio between tumour and control, and a detection p-value indicating if the gene is expressed in at least one condition. Assessment of multiple testing was performed by the Bioconductor package 'qvalue'.[27] The identified sets from the Venn diagram and the sets of differentially up- and down-regulated genes were further analysed for over-representation of gene ontologies (GOs) and KEGG pathways with the packages 'GOstats' and 'KEGG.db'.[28,29]

Normalized qPCR gene expression values (ΔCt) were log2-transformed and differences between paired tumour and normal samples were tested with a paired t-test using the 't.test()' R function. Column plots were produced using the R/Bioconductor packages 'ggplot2' and 'reshape'.[30,31]

Results

Differential gene expression in ADPA tumours

Hierarchical clustering revealed clusters that correspond to the studied patients (Fig. 3a), indicating more similarity within the tumour and control tissue pairs than between tumour or control tissues across patients. Comparison of mean gene expression in tumour and control tissue samples (Fig. 3b) revealed a higher abundance of genes expressed exclusively in the tumour (n=2,266) than in control tissue (n=204), while most genes were commonly expressed in tumour and control tissue (n=8,827). This is reflected in the analysis of differentially expressed genes where most cancer-related genes are up regulated. We first focused on genes from the KEGG "Pathways in cancer" which were differentially up- (ratio > 1.5, $p < 0.05$) or down-regulated (ratio < 0.67, $p < 0.05$) in the tumour samples. Figure 3c displays a heatmap of these genes that cluster in four groups. These comprise clusters with consistently high expression in both tumours that include, amongst others, the central oncogene *KRAS* (KRAS proto-oncogene, GTPase), which is generally overexpressed in tumours. Also consistently down-regulated clusters are present that contain for example *FZD5* (frizzled class receptor 5) and CXCL12 (C-X-C motif chemokine ligand 12). Figures 3d and 3e show the KEGG pathways significantly over- or under-represented among all up-regulated genes, suggesting the highly significant involvement of cell cycle and ribosomal genes as well as of genes associated with many cancer types.

Since these initial findings included a number of genes involved in angiogenesis-related pathways, we performed a more detailed analysis comprising the gene ontology "angiogenesis" (Fig 4). This revealed two clearly separated tumour and normal tissue clusters. Among the angiogenesis-related

genes that are overexpressed in tumours are *FGFR2* (Fibroblast growth factor receptor 2), *VEGFA* (vascular endothelial growth factor A) and *CXCL8* (C-X-C motif chemokine ligand 8, formerly *IL8*, interleukin 8) while *FDZ5* showed a clear downregulation (Fig 4).

FGFR2 is consistently overexpressed in ADPA tumours

We sought to validate our findings from the transcriptome array analyses in an expanded set of in total eight available ADPA tumours and paired control tissues. To account for the expected degree of variation due to the heterogeneous nature of the samples and limited quantities of RNA that could be extracted from the small FFPE tumours available to us, we opted to validate those genes with the most substantial evidence of an over-expression in tumour tissue. These were selected based on the transcriptome array results, regardless of KEGG pathway membership (see Methods, Table 2).

For two of the tumor-normal pairs the available FFPE RNA material could not successfully be utilized in the qPCR validation and were excluded. Of the 15 genes, six were not reliably detected in more than one of the remaining tumours and were excluded from further analysis (Table 2). Consistent tumour-overexpression in the tumour-normal pairs was confirmed for *FGFR2* ($p = 0.001$), while the other tested genes did not show significant overexpression (Fig. 5, Table 2). Interestingly, a tendency towards overexpression in ADPA tumours was present for *GPR89A* (G protein-coupled receptor 89A, $p = 0.05$), *SMS* (spermine synthase, $p = 0.08$) and *VEGFA* ($p = 0.08$). The two genes with previously reported occasional somatic mutations in ADPA, *TP53* and *BRAF*, did not show evidence for differential expression (Fig. 6a, Fig. 6b).

Discussion

Advances in cancer transcriptome profiling have turned out to be one of the most utilized approaches to understand malignancies at the molecular level. High-throughput techniques to study RNA have paved the way to the development of therapeutic targets, and have shown to be a powerful tool to predict drug sensitivity, likewise in skin as in non-skin malignancies.[17–20,32,33]

However, the architecture of gene expression is largely unknown for many skin tumours including sweat gland carcinomas. This is most probably due to i) the rarity of many of these tumour entities with widely dispersed, but small patient bases, ii) missed diagnosis of these easily over-looked tumour entities, iii) the often poor quality and quantity of available tissue, and iv) the lack of fresh tissue and the necessity to take up the challenge of molecular analyses of FFPE tissue.

Diagnosis of sweat gland carcinomas implicitly requires diagnostic histology, leading to tissue fixation and FFPE samples.[1,2] FFPE tissue provides a number of technical challenges and problems and has for a long time not been considered a reliable source of RNA. To the best of our knowledge, there are only few reports of well-defined entities of sweat gland carcinomas, not including ADPA, in which RNA analyses have been performed.[34,35] These studies on FFPE samples were limited to single PCR reactions to screen for defined fusion transcript variants.[34,35] No systematic expression profiles have been reported so far. Even in more common skin malignancies, studies of systematic transcriptome analyses on FFPE samples are nearly completely missing.[36]

Here, we report for the first time a comprehensive transcriptome profile of a well-defined sub-entity of sweat gland carcinomas, namely ADPA. The explorative transcriptome analyses by microarrays generated surprisingly highly significant results, pointing to strong genetic effects, sufficient to detect even in small cohorts.

We found 2,266 genes over-expressed in the ADPA tumours, comprising the crucial oncogenic driver genes *KRAS*, *MDM2* (MDM2 proto-oncogene)[37], *PDGFA* (platelet derived growth factor subunit A)[38], *BCL2L11* (BCL2 like 11)[39], *IL6ST* (interleukin 6 signal transducer)[40], *SMAD2* (SMAD family member 2)[41], *CCND1* (cyclin D1)[42], *LEF1* (lymphoid enhancer binding factor 1)[43], *VEGFA*[44], *RAF1* (Raf-1 proto-oncogene, serine/threonine kinase)[45], *CDK6* (cyclin dependent kinase 6)[46] or *CDK4* (cyclin dependent kinase 4)[46]. These genes play essential roles in tumourigenesis and the development of various types of cancer. Moreover, KEGG pathway analysis pointed to significant involvement of some of these genes in cell cycle and ribosomal processes, increasing our mechanistic understanding of this rare unexplored tumour entity.

A cluster consistently down-regulated in tumour contains the genes *FZD5*[47] and surprisingly *CXCL12* which is reported to be associated with tumour growth and metastasis.[48] Interestingly, there have recently been reports on *CXCL12* down-regulation promoting breast cancer metastasis.[49] The rare ADPA investigated here is to be annotated in a KEGG pathway, as such only projections to other cancer pathways could be identified in our study. Among all down-regulated genes in tumours, those from the KEGG Glutathione metabolism and "Wnt signalling pathways are most notable (Fig. 3e)[50].

Given the rarity of ADPA, our study is limited by the quite small sample size. Nevertheless, the number of analysed tumour samples was reasonable for a very rare disorder and our RNA samples were of extraordinary good quality. Of note, despite the small sample size, the overexpression of several distinct KEGG pathways remained significant even after adjustment for false discovery rate (Fig 3d). Furthermore, our initial findings based on the transcriptome array could be validated with qPCR in a further expanded set of tumours.

Our results point towards a consistent tumour-overexpression of *FGFR2*. *FGFR2* encodes a protein that belongs to the group of fibroblast growth factor receptors (FGFRs), which interacts with fibroblast growth factors (FGFs) within a tyrosine-signalling pathway. The FGF/FGFR signalling pathway is crucially involved in fundamental biological processes. These comprise e.g. cellular proliferation,

differentiation, survival, migration, angiogenesis and embryonic development.[51–53] Research has established that somatic *FGFR2* alterations play a crucial role in the origin of numerous solid tumours and that alterations in the activity (overexpression) are associated with a poor outcome in certain cancer types. Interestingly, many of these malignancies display a papillary structure and/ or an adenocarcinoma morphology like ADPA, pointing to common underlying molecular mechanisms. [54–59], [60,61] Oncogenic activation of *FGFR2* is supposed to have a key role in dysregulation of cell division (proliferation), cell movement and in the development of new blood vessels that nourish a growing tumour.[51–53] Targeting the FGF/FGFR axis is supposed to exert direct effects on cancer cells as well as indirect effects through changes of tumour microenvironment, in particular angiogenesis and immunity.[51,53,62] Our findings are surprising as *FGFR2* overexpression is indeed known to be crucial in the origin of various solid tumour entities[56,57,63] but to the best of our knowledge, overexpression of this gene was previously not in the focus of interest in skin tumours. Only in a small patient base of Kaposi`s sarcomas, up-regulation of FGFR2 expression could have been shown by performing immunohistochemistry and qRT-PCR.[64] Ramsey et al. proposed that p63-regulated FGFR2 signalling is activated in squamous cell carcinoma.[65] Gartside et al. identified structural *FGFR2* variants, mainly loss of function variants in a subset of melanomas.[66] Our unanticipated findings put further emphasis on the possible role of *FGFR2* in skin tumours and in particular sweat gland carcinomas. This is of high interest as currently numerous FGFR inhibitors are in active development and under study, operating through either selective or multi-targeted approaches. These anti-FGFR drugs are at different clinical phases of development and dependent on the specific FGFR inhibitor, they exert the effect on distinct tumours, comprising e.g. lung cancer, gastric cancer, breast cancer, cholangiosarcomas, melanoma and a broad range of solid tumours. [51,52,67] This raises the hope that FGFR2-targeting or a multi-targeted approach might also represent a potential treatment approach for patients with ADPA. However, further studies, including a more detailed characterization of this locus in an enlarged sample, are implicitly required to prove this hypothesis.

Apart from *FGFR2*, we could demonstrate a tendency towards significance for three additional genes, namely *GPR89A SMS* and *VEGFA* (Table 2, Fig. 5). *GPR89A* is involved in pH regulation of the Golgi apparatus. Knockdown of GPR89A has been associated with dysfunction of Golgi apparatus and higher apoptosis. It is a susceptibility cancer gene in breast cancer.[68,69] *SMS* is involved in polyamine metabolism. Dysregulation of polyamine metabolism has been reported to be associated with carcinogenic risk and targeting of this metabolism has raised hope to constitute a cancer therapy.[70,71] Vascular endothelial growth factor (VEGF) signalling is well-established to have a critical function in tumour angiogenesis. Targeting of *VEGFA* are ascertained innovative and effective therapeutic approaches in a broad range of malignancies.[44,72,73] It seems feasible that our gene expression–based approach has detected additional oncogenic pathway signatures. An independent replication is necessary before a potential association of these genes with ADPA can be regarded as proven.

Previous studies suggested an involvement of *TP53 and BRAF*-V600E mutations in the origin of ADPA. We found variability between patients and a not significantly differentially expression of *TP53* and *BRAF* in the tumor samples compared to the normal tissue samples (Fig. 6a, 6b). Our data set therefore provides no indication for expression changes of either gene in ADPA tumours, thus not supporting a crucial role of *TP53 or BRAF*.

We could not validate further initial findings, which is most probably due to the limited power of our sample set. Further research and independent replication of our findings is needed in larger sets of ADPA samples, preferentially comprising fresh frozen tissue to avoid the considerable difficulties in RNA extraction and analyses from FFPE tissue.

Nonetheless, our findings place rare skin tumours in the spotlight and encourage employing genetically technologies, even in challenging FFPE samples with poor quantity and quality of available tissue. This will be a step forward to make serious progress in the large group of unexplored and life-threatening rare skin malignancies with the aim to provide best care for patients and to form a basis for further research and drug development. It will be of high importance to consider rare tumour entities in the differential diagnosis of skin tumours and to take up the challenge to carry out genetic analyses. These insights will allow us to better understand the molecular taxonomy of skin tumours and to unveil disease mechanisms. To achieve this, it will be necessary to combine data of independent sample sets to i) expand sample sizes to identify new loci that exceed threshold for significance, to ii) further characterize candidate genes and to iii) apply high-throughput sequencing technologies to generate multi-omics data. Due to the sparse material of the mostly very small tumours (median diameter of 13mm (range 4 – 50 mm)[74], it is difficult to obtain enough material to perform comprehensive genomic profiling from the same tumour sample, the reason why we focused on transcriptome analyses in the present study. In order to investigate a homogenous sample, we included only solid, primary tumours. A molecular characterization of a subset of recurrent and metastatic ADPA in a next step will be important to gain further insights into the biogenesis of ADPA. It is plausible that somatic mutational and transcriptional profiles consistent with spread of the tumour to either lymph nodes or other tissue may be uncovered and that these are discordant to solid tumour formation. Molecular characterization and knowledge of the affected corresponding pathways might be the key towards a better understanding of this rare tumour entity. These approaches will advance the field towards precision medicine in ADPA and rare skin malignancies in general.

In conclusion, the results of this study provide evidence for the involvement of crucial oncogenic pathways in the aetiology of ADPA, pointing out cell cycle and ribosomal pathways. Suggested *FGFR2* overexpression raises for the first time the possibility of targeted therapeutic approaches comprising FGFR inhibitors. However, future studies are warranted to elucidate the genetic architecture and functional effects of risk loci more detailed and to learn about the effectiveness of existing therapy options and FGFR-targeted therapy in the treatment of patients with ADPA.

Acknowledgements

We are grateful to the patients for participating in this study.

JA acknowledges financial support from the Medical Faculty- Heinrich Heine University and the Duesseldorf School of Oncology (funded by the Comprehensive Cancer Centre Duesseldorf/Deutsche Krebshilfe).

We are grateful to Karl Köhrer, head of the Biological and Medical Research Center (BMFZ), Heinrich-Heine-University, Düsseldorf, Germany.

References

1 Rütten A, Requena L. Schweißdrüsenkarzinome der Haut. *Der Hautarzt* 2008; **59**:151–60.
2 Rütten A. [Eccrine sweat gland carcinoma of the skin]. *Pathologe* 2002; **23**:79–88.
3 Kao GF, Helwig EB, Graham JH. Aggressive digital papillary adenoma and adenocarcinoma. A clinicopathological study of 57 patients, with histochemical, immunopathological, and ultrastructural observations. *J Cutan Pathol* 1987; **14**:129–46.
4 Duke WH, Sherrod TT, Lupton GP. Aggressive Digital Papillary Adenocarcinoma. *Am J Surg Pathol* 2000; **24**:775–84.
5 Kraus A, Altmann S, Damert H-G, *et al.* Aggressive digital papillary adenocarcinoma – a rare malignant tumor of the sweat glands: two case reports and a review of the literature. *Clin Cosmet Investig Dermatol* 2015; **Volume 8**:143.
6 Frey J, Shimek C, Woodmansee C, *et al.* Aggressive digital papillary adenocarcinoma: A report of two diseases and review of the literature. , 2009 doi:10.1016/j.jaad.2008.07.038.
7 Bazil MK, Henshaw RM, Werner A, Lowe EJ. Aggressive Digital Papillary Adenocarcinoma in a 15-year-old Female. *J Pediatr Hematol Oncol* 2006; **28**:529–30.
8 Heiwig EB. Eccrine acrospiroma. *J Cutan Pathol* 1984; **11**:415–20.
9 Geethamani V, Shetty A, Chiwdappa V. Aggressive Digital Papillary Adenocarcinoma-A rare Entity Posing a Diagnostic Challenge. *Malays J Med Sci* 2014; **21**:54–6.
10 Gole GN, Tati SY, Deshpande AK, Gole SG. Aggressive Digital Papillary Adenocarcinoma in a Young Female-a Rare Presentation. *J Hand Microsurg* 2011; **3**:31–3.
11 Bakotic B, Antonescu CR. Aggressive digital papillary adenocarcinoma of the foot: the clinicopathologic features of two cases. *J Foot Ankle Surg*; **39**:402–5.
12 Brodbeck R, Schmitz M, Horch R. Daumenrekonstruktion nach onkologischer Resektion eines seltenen aggressiven digitalen papillären Adenokarzinoms: Ein Fallbericht. *Handchirurgie · Mikrochirurgie · Plast Chir* 2015; **47**:322–35.
13 Chi CC, Kuo TT WS. Aggressive digital papillary adenocarcinoma: a silent malignancy masquerading as acquired digital fibrokeratoma. *Am J Clin Dermatol* 2007; **8**:243–5.
14 Hayashi A, Matsumura T, Horiguchi M, *et al.* The Medial Plantar Flap Vascularized by the Reverse Flow Lateral Plantar Artery: A Novel Variation through the Case of Aggressive Digital Papillary Adenocarcinoma of the Sole. *J Reconstr Microsurg* 2012; **28**:427–30.
15 Dias-Santagata D, Lam Q, Bergethon K, *et al.* A potential role for targeted therapy in a subset of metastasizing adnexal carcinomas. *Mod Pathol* 2011; **24**:974–82.
16 Bell D, Aung PP, Prieto VG, Ivan D. Next-generation sequencing reveals rare genomic alterations in aggressive digital papillary adenocarcinoma. *Ann Diagn Pathol* 2015; **19**:381–4.
17 Lauss M, Donia M, Harbst K, *et al.* Mutational and putative neoantigen load predict clinical benefit of adoptive T cell therapy in melanoma. *Nat Commun* 2017; **8**:1738.
18 Hugo W, Zaretsky JM, Sun L, *et al.* Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* 2016; **165**:35–44.

19 Schramm S-J, Campain AE, Scolyer RA, *et al.* Review and Cross-Validation of Gene Expression Signatures and Melanoma Prognosis. *J Invest Dermatol* 2012; **132**:274–83.

20 Cirenajwis H, Ekedahl H, Lauss M, *et al.* Molecular stratification of metastatic melanoma using gene expression profiling : Prediction of survival outcome and benefit from molecular targeted therapy. *Oncotarget* 2015; **6**:12297–309.

21 Gentleman RC, Carey VJ, Bates DM, *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004; **5**:R80.

22 Gautier L, Cope L, Bolstad BM, Irizarry RA. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 2004; **20**:307–15.

23 Graffmann N, Ring S, Kawala M-A, *et al.* Modeling Nonalcoholic Fatty Liver Disease with Human Pluripotent Stem Cell-Derived Immature Hepatocyte-Like Cells Reveals Activation of PLIN2 and Confirms Regulatory Functions of Peroxisome Proliferator-Activated Receptor Alpha. *Stem Cells Dev* 2016; **25**:1119–33.

24 Smyth GK. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Stat Appl Genet Mol Biol* 2004; **3**. doi:10.2202/1544-6115.1027.

25 Warnes GR, Bolker B, Bonebakker L, *et al.* gplots: Various R Programming Tools for Plotting Data. , 2015.

26 Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* 2011; **12**:35.

27 Storey JD. A direct approach to false discovery rates. *J R Stat Soc Ser B (Statistical Methodol* 2002; **64**:479–98.

28 Kanehisa M, Sato Y, Kawashima M, *et al.* KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 2016; **44**:D457-462.

29 Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association. *Bioinformatics* 2007; **23**:257–8.

30 Wickham H. Ggplot2 : elegant graphics for data analysis. , Springer, 2009.

31 Wickham H. Reshaping Data with the **reshape** Package. *J Stat Softw* 2007; **21**:1–20.

32 Cieślik M, Chinnaiyan AM. Cancer transcriptome profiling at the juncture of clinical translation. *Nat Rev Genet* 2017; **19**:93–109.

33 Kumar-Sinha C, Chinnaiyan AM. Precision oncology in the age of integrative genomics. *Nat Biotechnol* 2018; **36**:46–60.

34 Kazakov D V, Ivan D, Kutzner H, *et al.* Cutaneous Hidradenocarcinoma: A Clinicopathological, Immunohistochemical, and Molecular Biologic Study of 14 Cases, Including < *Am J Dermatopathol* 2009; **31**:236–47.

35 North JP, McCalmont TH, van Zante A, *et al.* Detection of MYB Alterations and Other Immunohistochemical Markers in Primary Cutaneous Adenoid Cystic Carcinoma. *Am J Surg Pathol* 2015; **39**:1347–56.

36 Xie H, Lee L, Caramuta S, *et al.* MicroRNA Expression Patterns Related to Merkel Cell Polyomavirus Infection in Human Merkel Cell Carcinoma. *J Invest Dermatol* 2014; **134**:507–17.

37 Zheng T, Wang J, Zhao Y, *et al.* Spliced MDM2 isoforms promote mutant p53 accumulation and gain-of-function in tumorigenesis. *Nat Commun* 2013; **4**:2996.

38 MacDonald TJ, Brown KM, LaFleur B, *et al.* Expression profiling of medulloblastoma: PDGFRA and the RAS/MAPK pathway as therapeutic targets for metastatic disease. *Nat Genet* 2001; **29**:143–52.

39 Carné Trécesson S de, Souazé F, Basseville A, *et al.* BCL-XL directly modulates RAS signalling to favour cancer cell stemness. *Nat Commun* 2017; **8**:1123.

40 Sanz-Moreno V, Gaggioli C, Yeo M, *et al.* ROCK and JAK1 Signaling Cooperate to Control Actomyosin Contractility in Tumor Cells and Stroma. *Cancer Cell* 2011; **20**:229–45.

41 Yeh H-W, Hsu E-C, Lee S-S, *et al.* PSPC1 mediates TGF-β1 autocrine signalling and Smad2/3 target switching to promote EMT, stemness and metastasis. *Nat Cell Biol* 2018; **20**:479–91.

42 Fusté NP, Fernández-Hernández R, Cemeli T, *et al.* Cytoplasmic cyclin D1 regulates cell invasion and metastasis through the phosphorylation of paxillin. *Nat Commun* 2016; **7**:11581.

43 Kim D, You E, Jeong J, *et al.* DDR2 controls the epithelial-mesenchymal-transition-related gene expression via c-Myb acetylation upon matrix stiffening. *Sci Rep* 2017; **7**:6847.

44 Jayson GC, Kerbel R, Ellis LM, Harris AL. Antiangiogenic therapy in oncology: current status and future directions. *Lancet* 2016; **388**:518–29.

45 Palanisamy N, Ateeq B, Kalyana-Sundaram S, *et al.* Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. *Nat Med* 2010; **16**:793–8.

46 Otto T, Sicinski P. Cell cycle proteins as promising targets in cancer therapy. *Nat Rev Cancer* 2017; **17**:93–115.

47 Koval A, Katanaev VL. Dramatic dysbalancing of the Wnt pathway in breast cancers. *Sci Rep* 2018; **8**:7329.

48 Domanska UM, Kruizinga RC, Nagengast WB, *et al.* A review on CXCR4/CXCL12 axis in oncology: No place to hide. *Eur J Cancer* 2013; **49**:219–30.

49 Yu PF, Huang Y, Xu CL, *et al.* Downregulation of CXCL12 in mesenchymal stromal cells by TGFβ promotes breast cancer metastasis. *Oncogene* 2017; **36**:840–9.

50 Sherwood V, Leigh IM. WNT Signaling in Cutaneous Squamous Cell Carcinoma: A Future Treatment Strategy? *J Invest Dermatol* 2016; **136**:1760–7.

51 Babina IS, Turner NC. Advances and challenges in targeting FGFR signalling in cancer. *Nat Rev Cancer* 2017; **17**:318–32.

52 Katoh M. Therapeutics Targeting FGF Signaling Network in Human Diseases. 2016. doi:10.1016/j.tips.2016.10.003.

53 Katoh M. FGFR inhibitors: Effects on cancer cells, tumor microenvironment and whole-body homeostasis (Review). *Int J Mol Med* 2016; **38**:3–15.

54 Basturk O, Berger MF, Yamaguchi H, *et al.* Pancreatic intraductal tubulopapillary neoplasm is genetically distinct from intraductal papillary mucinous neoplasm and ductal adenocarcinoma. *Mod Pathol* 2017; **30**:1760–72.

55 Chakrabarty S, Varghese VK, Sahu P, *et al.* Targeted sequencing-based analyses of candidate gene variants in ulcerative colitis-associated colorectal neoplasia. *Br J Cancer* 2017; **117**:136–43.

56 Ahn S, Lee J, Hong M, *et al.* FGFR2 in gastric cancer: protein overexpression predicts gene amplification and high H-index predicts poor survival. *Mod Pathol* 2016; **29**:1095–103.

57 Nakamura H, Arai Y, Totoki Y, *et al.* Genomic spectra of biliary tract cancer. *Nat Genet* 2015; **47**:1003–10.

58 Medico E, Russo M, Picco G, *et al.* The molecular landscape of colorectal cancer cell lines unveils clinically actionable kinase targets. *Nat Commun* 2015; **6**:7002.

59 Stelloo E, Bosse T, Nout RA, *et al.* Refining prognosis and identifying targetable pathways for high-risk endometrial cancer; a TransPORTEC initiative. *Mod Pathol* 2015; **28**:836–44.

60 Tsimafeyeu I, Khasanova A, Stepanova E, *et al.* FGFR2 overexpression predicts survival outcome in patients with metastatic papillary renal cell carcinoma. *Clin Transl Oncol* 2017; **19**:265–8.

61 Fu YT, Zheng HB, Zhang DQ, Zhou L SH. MicroRNA-1266 suppresses papillary thyroid carcinoma cell metastasis and growth via targeting FGFR2. *Eur Rev Med Pharmacol Sci* 2018; **22**:3430–8.

62 Chae YK, Ranganath K, Hammerman PS, *et al.* Inhibition of the fibroblast growth factor receptor (FGFR) pathway: the current landscape and barriers to clinical application. *Oncotarget* 2017; **8**:16052–74.

63 Timsah Z, Ahmed Z, Ivan C, *et al.* Grb2 depletion under non-stimulated conditions inhibits PTEN, promotes Akt-induced tumor formation and contributes to poor prognosis in ovarian cancer. *Oncogene* 2016; **35**:2186–96.

64 Cottoni F, Ceccarelli S, Masala MV, *et al.* Overexpression of the fibroblast growth factor receptor 2-IIIc in Kaposi's sarcoma. *J Dermatol Sci* 2009; **53**:65–8.

65 Ramsey MR, Wilson C, Ory B, *et al.* FGFR2 signaling underlies p63 oncogenic function in squamous cell carcinoma. *J Clin Invest* 2013; **123**:3525–38.

66 Gartside MG, Chen H, Ibrahimi OA, *et al.* Loss-of-function fibroblast growth factor receptor-2 mutations in melanoma. *Mol Cancer Res* 2009; **7**:41–54.

67 Chae YK, Ranganath K, Hammerman PS, *et al.* Inhibition of the fibroblast growth factor receptor (FGFR) pathway: the current landscape and barriers to clinical application. *Oncotarget* 2017; **8**:16052–74.

68 Maeda Y, Ide T, Koike M, *et al.* GPHR is a novel anion channel critical for acidification and functions of the Golgi apparatus. *Nat Cell Biol* 2008; **10**:1135–45.

69 Patel N, Weekes D, Drosopoulos K, *et al.* Integrated genomics and functional validation identifies malignant cell specific dependencies in triple negative breast cancer. *Nat Commun* 2018; **9**:1044.

70 Casero RA, Murray Stewart T, Pegg AE. Polyamine metabolism and cancer: treatments, challenges and opportunities. *Nat Rev Cancer* 2018; :1.

71 Gerner EW, Meyskens FL. Polyamines and cancer: old molecules, new understanding. *Nat Rev Cancer* 2004; **4**:781–92.

72 Ferrara N, Adamis AP. Ten years of anti-vascular endothelial growth factor therapy. *Nat Rev Drug Discov* 2016; **15**:385–403.

73 De Palma M, Biziato D, Petrova T V. Microenvironmental regulation of tumour angiogenesis. *Nat Rev Cancer* 2017; **17**:457–74.

74 Rismiller K, Knackstedt TJ. Aggressive Digital Papillary Adenocarcinoma: Population-Based Analysis of Incidence, Demographics, Treatment, and Outcomes. doi:10.1097/DSS.0000000000001483.

Figure legends

Figure 1: Histopathologic and immunohistochemically features of aggressive digital papillary adenocarcinoma (ADPA), exemplary shown for case 1.

(a) Scanning magnification of an aggressive digital papillary adenocarcinoma shows a solid, in parts cystic tumour, in acral skin without connection to the overlying epidermis. Basaloid cell type. HE x 2.

(b) Solid hyper-cellular areas of the tumour with densely packed tumour cells, showing atypical nuclei and many mitotic figures. Densely packed spindle cells with mitoses are a clue for a malignant sweat gland tumor. HE x 40.

(c) Areas with cystic configuration show characteristic papillary projections into the lumina. These projections contributed to the name„papillary" carcinoma . HE x 20.

(d) Staining with cytokeratin 7 shows typical staining of the more centrally located tubular cells, while the more peripheral myoepitheliall cells are negative. A typical finding for these carcinomas in acral skin. CK 7 Staining X 40.

(e) An actin stain shows a positive staining of the myoepithelial cells at the periphery. All aggregates are surrounded by an actin positive myoepithelial layer. Actin staining x 40.

Figure 2: Scanning picture of a second example of an aggressive digital papillary adenocarcinoma.

(a) The tumor is not connected to the epidermis. There are cystic spaces lined by papillary projections. HE x 10.

(b) Bluish tumor cells. Papillary projections into the cystic lumina. HE X 2.

(c) The actin stains is positive in all the myopepithelial cells at the periphery. These finding separates the tumor from hidradenomas, poromas and myoepitheliomas in acral skin. Actin X 20.

(d) The inner part of the tumor aggregates stain for cytokeratin 7. Cytokeratin 7 X 20.

Figure 3: Cancer-pathway related motifs are up-regulated in the transcriptome of ADPA tumour cells**.**

(a) Global transcriptomes clustered by patients. Hierarchical cluster analysis yielded clusters of tumour and control tissues corresponding to the patients.

(b) The Venn diagram of expressed genes ($p < 0.05$) shows that more genes are expressed in the mean of both tumour samples (n=2266) than in the mean of both control tissue samples (n=204). However, most genes are expressed commonly in tumour and control tissue.

(c) Heatmap of genes KEGG "pathways in cancer" differentially expressed between tumour and control tissue demonstrates up-regulation of the central oncogene *KRAS* in tumours, together with the cell cycle genes.

(d) The pathways over-represented in the up-regulated genes confirms this central cancer motif by the overlap of multiple different cancers, pathways in cancer and cell cycle. The rare cancer investigated here was not yet modelled in a KEGG pathway and thus only projections to other cancer pathways can be identified.

(e) Down-regulated genes are annotated with pathways including glutathione metabolism and Wnt signalling pathways.

Figure 4: Differentially expressed genes ($p<0.05$, detection-$p<0.05$, ratio<0.67 or >1.5) associated with the gene ontology „angiogenesis" (GO:0001525) were subjected to a cluster analysis and the resulting heatmaps and dendrograms showed a clear separation of a tumor (T) and a normal tissue (N) cluster. Note: the official gene symbol for interleukin 8 is now *CXCL8*.

Figure 5: Validation real-time qPCR results in the additional ADPA tumour and normal sample pairs (T3-T8). For the indicated genes and tumour-control sample pairs, log2-transformed fold changes of expression between tumour and control tissue are depicted for the indicated genes, respectively.

Figure 6: Bar chart of gene expression array probesets annotated with *TP53* and *BRAF*. The red dashed line indicates gene expression at a level corresponding to a detection p-value of 0.05. Normal tissue samples (N) are in blue colors and tumor samples (T) in red colors. There is variability between patients and a non-significant tendency for higher expression of *TP53* and *BRAF* in the tumor samples than in the normal tissue samples. However, no significant differential expression was detected.

Table 1: Clinical features

						Control tissue	
Case No.	**Tissue**	**Sex**	**Year of Birth**	**Age of onset (years)**	**Site**	**Matched surrounding healthy tissue**	**Blood**
1	Tumour (ADPA)	F	1910	96	Distal phalanx of the thumb, right hand	√	-
2	Tumour (ADPA)	F	1934	71	Right side of the hand	√	-
3	Tumour (ADPA)	M	1977	40	Fingertip of the third finger, left hand	√	√
4	Tumour (ADPA)	M	1938	77	5th finger, left hand	-	√
5	Tumour (ADPA)	M	1958	51	Forefinger, right hand	-	√
6	Tumour (ADPA)	F	1952	59	Middle finger, right hand	√	-
7	Tumour (ADPA)	F	1968	36	Distal phalanx of the fourth finger, right hand	√	-
8	Tumour (ADPA)	F	1957	52	Left heel	-	√

ADPA indicates Aggressive digital papillary adenocarcinoma; Transcriptome arrays were performed using the GeneChip PrimeView Human Gene Expression Array (Affymetrix©)

Table 2: qPCR validation for ADPA tumour-normal pairs

Gene Symbol	Gene Name	ADPA Over-Expression P-value[a]
ELF3	E74 like ETS transcription factor 3	-[b]
FGFR2	fibroblast growth factor receptor 2	**0.001**
GPR89A	G protein-coupled receptor 89A	0.054
HIST1H2BH	histone cluster 1 H2B family member h	0.181
JUP	junction plakoglobin	0.123
KIAA1324	KIAA1324	-[b]
LEF1	lymphoid enhancer binding factor 1	-[b]
MDM2	MDM2 Proto-Oncogene	-[b]
RPL17	ribosomal protein L17	-[b]
RPL30	ribosomal protein L30	0.504
SLC12A4	solute carrier family 12 member 4	0.307
SMS	spermine synthase	0.079
CXCL8	C-X-C Motif Chemokine Ligand 8	0.636
VEGFA	vascular endothelial growth factor A	0.081
ZFC3H1	zinc finger C3H1-type containing	-[b]

a) P-value of paired t-test (ΔCt of tumour and control tissue)
b) Not reliably detected in >1 ADPA tumours

Figure 1

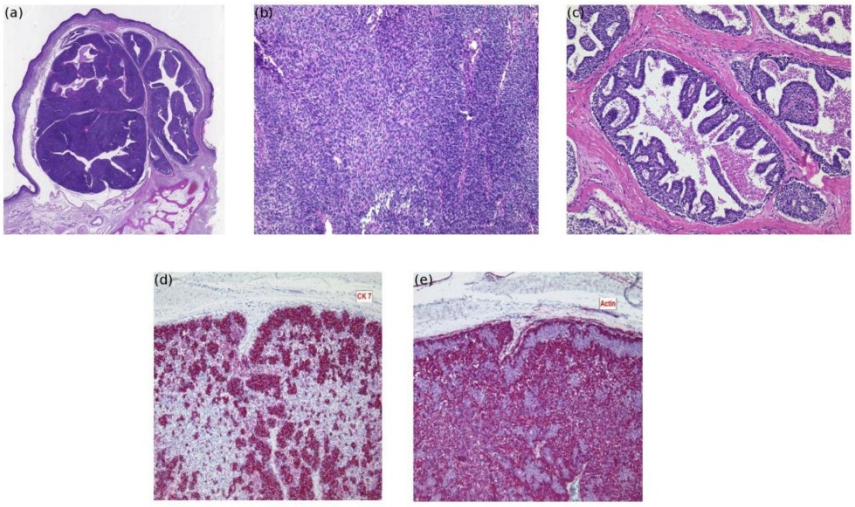


Figure 2

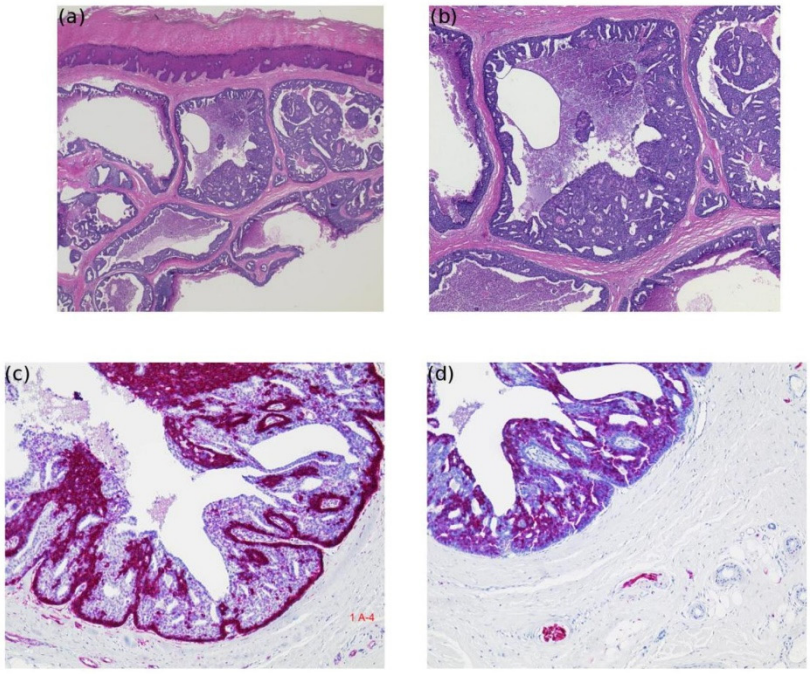
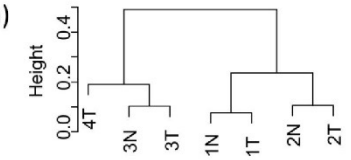
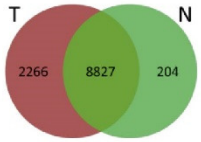


Figure 3

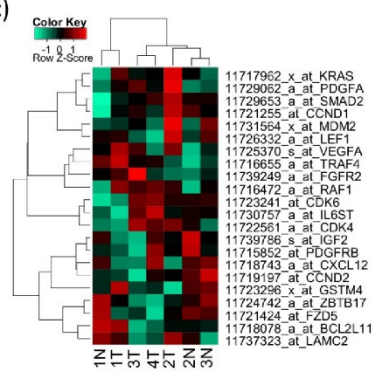
a)



b)



c)



d)

kegg_name	p_hyper	q_hyper
Bladder cancer	1,39939E-05	0,001810348
Glioma	1,7613E-05	0,001810348
Pancreatic cancer	2,88003E-05	0,001810348
Melanoma	3,16218E-05	0,001810348
Prostate cancer	0,000136033	0,006230306
Pathways in cancer	0,000246558	0,009410288
Chronic myeloid leukemia	0,000330562	0,010814096
Non-small cell lung cancer	0,000614793	0,017598442
Ribosome	0,001140352	0,025982217
Colorectal cancer	0,001160763	0,025982217
Cell cycle	0,001248054	0,025982217
VEGF signaling pathway	0,002879826	0,054956676
Endometrial cancer	0,004393494	0,076431817
Small cell lung cancer	0,004672687	0,076431817
Thyroid cancer	0,005586768	0,091397587
Acute myeloid leukemia	0,006495293	0,092953876
Cysteine and methionine metabolism	0,010965877	0,147716808
p53 signaling pathway	0,011896739	0,149651786
Renal cell carcinoma	0,012496328	0,149651786
Wnt signaling pathway	0,013070025	0,149651786
Focal adhesion	0,014392482	0,156946589
Tight junction	0,028410716	0,285973038
Gap junction	0,028722183	0,285973038
Insulin signaling pathway	0,032612618	0,31117873
Endocytosis	0,046898548	0,429590698

e)

kegg_name	p_hyper	q_hyper
Glutathione metabolism	0,00494008	1
Wnt signaling pathway	0,020380915	1
ECM-receptor interaction	0,02104046	1
Cytokine-cytokine receptor interaction	0,035882803	1
Cyanoamino acid metabolism	0,047911992	1
Focal adhesion	0,049650182	1

Figure 4

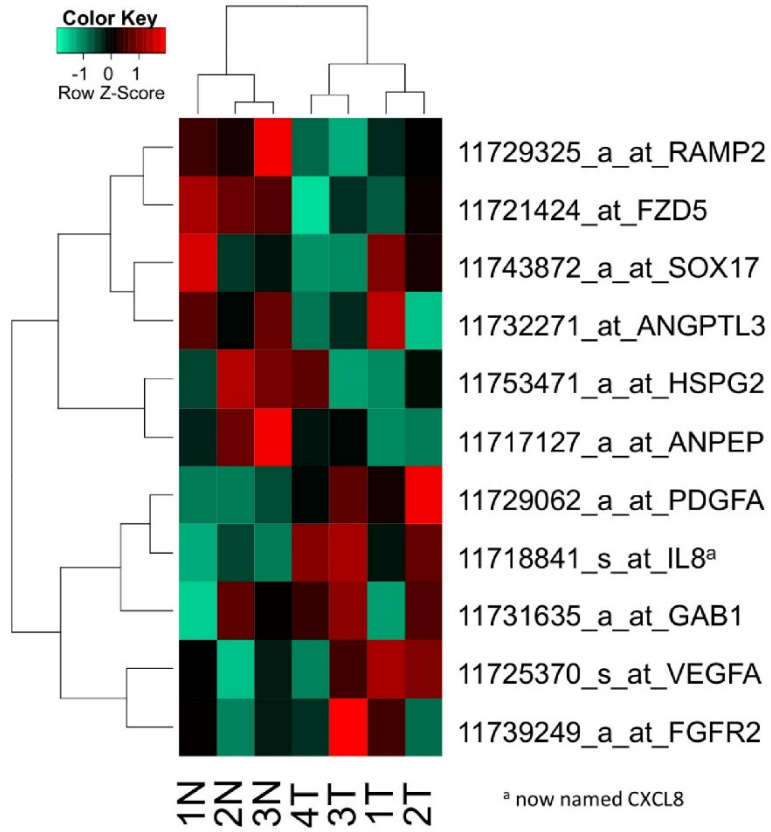
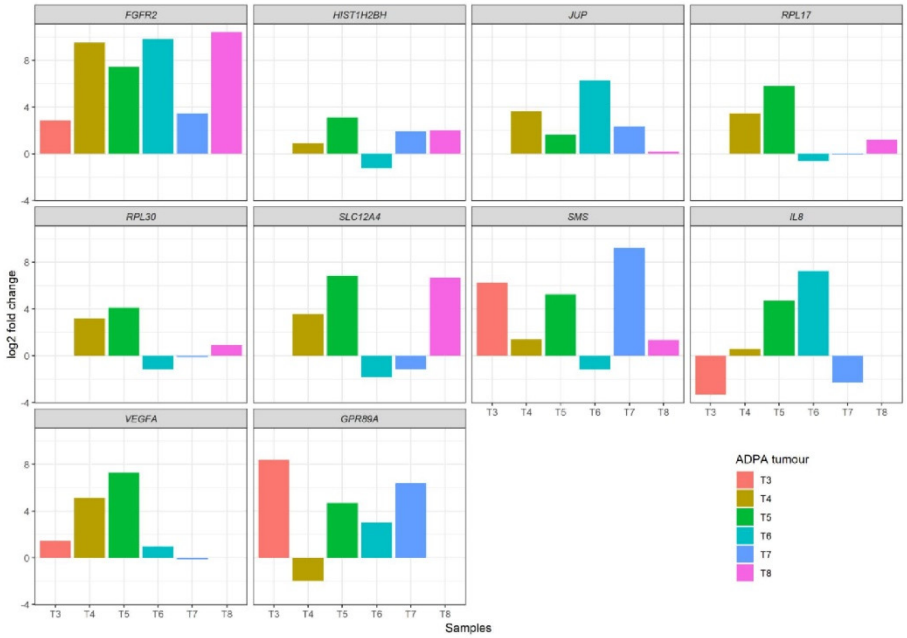


Figure 5



3.3. New insights into human primordial germ cells and early embryonic development from single-cell analysis

Jörg Otte, Wasco Wruck and James Adjaye

Abstract

Human preimplantation developmental studies are difficult to accomplish due to associated ethical and moral issues. Preimplantation cells are rare and exist only in transient cell states. From a single cell, it is very challenging to analyze the origination of the heterogeneity and complexity inherent to the human body. However, recent advances in single-cell technology and data analysis have provided new insights into the process of early human development and germ cell specification. In this Review, we examine the latest single-cell datasets of human preimplantation embryos and germ cell development, compare them to bulk cell analyses, and interpret their biological implications.

Share of scientific contribution: 80%

Jörg Otte did the literature research and the conception under supervision of Prof. Dr. James Adjaye. Jörg Otte wrote the paper.

Published: *FEBS Letters,* June 18, 2017, Vol.591, Issue 15, Pages 2226-2240,
DOI: 10.1002/1873-3468.12716

Link to the publication:
https://febs.onlinelibrary.wiley.com/doi/full/10.1002/1873-3468.12716

FEBS Letters

FEBS PRESS science publishing by scientists

REVIEW ARTICLE

New insights into human primordial germ cells and early embryonic development from single-cell analysis

Jörg Otte, Wasco Wruck and James Adjaye

Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, Duesseldorf, Germany

Correspondence
J. Adjaye, Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, 40225 Duesseldorf, Germany
Fax: +49 211 81 19147
Tel: +49 211 81 08191
E-mail: James.Adjaye@med.uni-duesseldorf.de

(Received 10 May 2017, revised 7 June 2017, accepted 9 June 2017)

doi:10.1002/1873-3468.12716

Edited by Wilhelm Just

Human preimplantation developmental studies are difficult to accomplish due to associated ethical and moral issues. Preimplantation cells are rare and exist only in transient cell states. From a single cell, it is very challenging to analyse the origination of the heterogeneity and complexity inherent to the human body. However, recent advances in single-cell technology and data analysis have provided new insights into the process of early human development and germ cell specification. In this Review, we examine the latest single-cell datasets of human preimplantation embryos and germ cell development, compare them to bulk cell analyses, and interpret their biological implications.

Keywords: embryonic development; primordial germ cells; single-cell analysis

The definition of life includes the capabilities of reproduction, inheritance and development [1]. At the start of life during human reproduction, the potential to develop a highly complex organism is inherent to one single cell, the zygote. The transfer of genetic information from one generation to another is facilitated by germ cells, which are designed to induce biological variety between different progenies. After fertilization, heterogeneity in terms of an asymmetry of daughter cells induces complexity.

Detailed molecular understanding of these life-defining qualities requires high-resolution analytic tools that cope with minimal amounts of sample material. The field of single-cell analyses of mammalian cells opened in 2009, when Tang *et al.* [2] introduced their protocol for a whole transcriptome analysis of a single mouse blastomere [2]. Not only being able to analyse low input materials but also the understanding of growing heterogeneity within developing tissues are the strengths of single-cell analyses. In this review, we describe new insights made by single-cell analyses in germ cell development and preimplantation embryo.

Bioinformatic tools

When analysing heterogeneity, the main challenge of all single-cell approaches is distinguishing between real biological differences and technical variabilities. In developmental biology, most cells exist in transient states that complicate the definition of the present cell type usually defined by permanent properties. Influenced by the spatial context, extrinsic stimuli or intrinsic factors like the cell cycle, a cells' identity can be highly diffuse and difficult to analyse by temporal snap-shot methods [3,4].

Besides these biological confounders, single-cell genome and transcriptome studies also have to manage similar technical challenges: a reliable identification and isolation of the correct cells, capture and amplification of DNA or RNA molecules, sequencing and

Abbreviations
DNMT, *de novo* methyltransferases; EPI, epiblast; ICM, inner cell mass; PCA, principal component analysis; PE, primitive endoderm; PGC, primordial germ cell; RRBS, reduced representative bisulphite sequencing; SNP, single-nucleotide polymorphism; TE, trophectoderm.

analysis with respect to cell number and sequencing depth as well as the interpretation of the data with bioinformatic tools considering technical limitations [5].

The most error-prone part is probably the amplification of the investigated samples, achieved predominantly via PCR but likely to produce unspecific by-products [6]. Thus, improvement of amplification methods as demonstrated by Yan *et al.* [7] is a major issue. Next generation sequencing-based techniques have the advantage of covering the whole sequence of the genome whereas microarrays only cover representative parts of it. Single-cell sequencing can be analysed in a similar manner to other sequencing experiments. Guo *et al.* [8] used the established pipeline of tools TopHat, CuffLinks [9] and DeSeq [10] to align the reads to a reference genome assemble of transcripts, quantify expression and detect differentially expressed genes. Though the tools TopHat and CuffLinks might be updated to their improved successors HISAT and StringTie [11], the alignment and transcript assembly software STAR [12] would be a better alternative. To account for single-cell characteristics, normalization was done via spike-ins following a method developed by Treutlein *et al.* [13]. A linear regression was performed between the log2-transformed FPKM (fragments per kilobase exon per million reads) values and the vendor-provided number of transcript for each single cell; the results were used for normalization in subsequent processing. Several publications have highlighted the high false-negative rates due to dropout genes as an inherent problem in the single-cell approach and recommended various models to account for this [4]. Zero-inflated models combine the true distribution of successful detection of a transcript with the distribution of failed detection due to technical effects. Other strategies work via logistic regression of false-negative curves comparing expected and observed expressions, determine a cell's technical detection efficiency and calculate probabilities for technical dropouts [14].

Due to its propagating use mainly in cancer research and developmental biology, the potential of single-cell technologies is continuously increasing [15]. Further strategies to optimize the output of single-cell DNA and RNA analyses have been reviewed in detail elsewhere [3,5,16–19].

Germ cell genome sequencing

The genome of a healthy organism is usually very stable throughout life. Single-cell DNA analysis of germ cells, however, pose a challenge due to their haploidy. Advancements in reliable amplification and sequencing methods have made reaching new milestones in single germ cell analyses possible. In 2012, Wang *et al.* [20] used the multiple strand displacement amplification method, characterized by high sensitivity and specificity, to analyse the whole genome of human sperm cells. Hou *et al.* [21], who sequenced the whole genome of single oocytes in 2013, used the multiple annealing and looping-based amplification cycles (MALBAC) method. Its high uniformity across the whole genome allowed the detection of meiotic crossover events with low sequencing depths [22]. Both studies depended on the phased genome information of the paternal DNA meaning the assignment of heterozygous single-nucleotide polymorphisms (SNPs) to each of the alleles per chromosome [23,24]. A high accuracy of the amplification of the haploid genome, the subsequent sequencing as well as the solid statistical testing to identify significant events was essential.

Since certain genes follow a mono-allelic expression due to epigenetic imprinting, the detection of heterozygous SNPs in protein coding regions in the embryonic genome allows for the reidentification of maternal or paternal transcripts. This requires a high single-nucleotide resolution of the genome and the transcriptome of an individual organism. It represents a good opportunity to study epigenetic mechanisms like the inactivation of one of the X-chromosomes in female embryos as well as its reactivation during germ cell development [8]. Furthermore, a random mono-allelic expression of autosomal genes can result in cellular heterogeneity [25].

Challenges in transcriptomics in early embryonic development

In contrast to the fairly stable genome, the transcriptome of a developing organism is highly dynamic. Though this may allow for the best insight into a cell's identity it also implies greater challenges on data quality. The sensitivity of single-cell RNA sequencing approaches is lower compared to bulk analyses, meaning that low-expressing genes might not be detected [26]. RNA molecules are very fragile and the degree of degradation of RNA during the cell isolation and the RNA capturing process might account for further technical variability. An ultimate goal is to preserve not only the relative amount of each RNA molecule during the amplification step, but the exact count of each initial molecule in single-cell RNA sequencing experiments [17,18].

Unfortunately, not only technical reasons but also biological principles impede the precise analysis of the

transcriptome. The transcriptional noise within a single cell is influenced by a different half-life of each RNA molecule as well as the fact that RNA is transcribed from the DNA in a burst-like manner [27]. Furthermore, the cell cycle state influences the overall transcriptional activity and many genes show a cell cycle-dependent regulation [28].

Early embryonic cells have very low transcriptional activity. All mRNAs detected within the first cell stages are maternal transcripts from the oocyte that are continually dividing in the daughter cells during cell cleavage [29]. However, thousands of expressed genes can be detected by RNA sequencing within an embryonic cell [30]. Their interpretation in terms of signalling pathway activities or distinction of different cell types by hierarchical clustering maybe one of the most important benefits of single-cell RNA sequencing [3]. Principal component analysis (PCA) is often used to investigate heterogeneity, similarity and lineage-associated developmental processes of single cell populations. It visualizes the main subgroups within a dataset by reducing a high-dimensional dataset into a lower dimensional projection. The entirety of gene expression levels within distinct cell types lead to the description of an individual fingerprint of a cell's identity as well as question the use of established marker genes and already led to identification of new marker genes [3,13,31,32].

In bulk cell analyses, cellular heterogeneity and subpopulations of cells cannot be identified and slight changes in the transcriptome can be masked by an average expression level. Very low-expressed genes may not be detected by single-cell approaches due to its low sensitivity. Detected differences in single-cell analysis on the other hand might not correlate with the actual phenotype of the cell, due to the above-mentioned technical and biological variety in gene expression. These limitations of single-cell methods require enough cells to be analysed to obtain significant results.

Ethical restrictions

The use of embryonic tissue underlies ethical and legal restrictions, which vary in different countries. Early embryonic blastomeres can be obtained from fertilized oocytes through assisted reproductive technology for infertile couples. Nonviable embryos with genomic aberrations that cannot be transferred to the uterus are the most common source for embryonic research. Other embryos with a nonaberrant genome but a monogenic disease and thus, excluded for further reproduction cycles usually have an unaffected potency. In countries such as China, USA, United Kingdom or Switzerland, surplus embryos after a successful assisted reproduction can be donated for research purposes under strict guidance from the respective national ethical boards, for example in the UK-The Human Fertilisation and Embryology Authority. These cells are viable with full potency of development and represent materials of the highest quality. For analyses of older embryonic tissues like primordial germ cells, aborted embryos from a medically or surgically terminated pregnancy is required. The scarcity of embryonic material as well as differences in quality of the isolated cells results in a limited number of studies analysing embryonic tissue with comparable data quality.

Primordial germ cells

The early precursor cells of the germ cell lineage called primordial germ cells (PGCs), originate within the first days of embryonic development. In their later evolution, they develop into haploid oocytes or sperm giving rise to a new generation of totipotent stem cells after fertilization. Until then, these cells fulfil a unique process of maturation.

The very first onset of PGC induction in humans is still unknown. Two weeks after fertilization around the time of gastrulation, PGCs can be found in the primary ectoderm in the posterior epiblast [33]. From this intraembryonic location, the PGCs migrate from extraembryonic tissue into the yolk sac wall. While re-entering the embryo proper, PGCs start proliferating and finally assemble in the genital ridge at about 6 to 9 weeks after fertilization, where they continue to manifold by mitosis. After 9 weeks, female PGCs undergo meiotic divisions before entering meiosis, and later arrest at the diplotene stage of prophase I. Male PGCs, on the other hand, undergo a mitotic arrest later in development to enter meiosis after puberty [34].

Identification of early germ cells

Not only ethical and legal restrictions impede the analysis of human PGCs, also technical limitations complicate the study of such early embryonic cell stages. The most recent transcriptome analyses provide insights into the development of early germ cells from 4 to 19 week old embryos by analysing foetal ovaries, testes or the aorta-gonad-mesonephros region, respectively. These studies, do not describe the early migratory cells though but only focus on gonadal PGCs. The work of Guo *et al.* [8] is the only one that used a single-cell

RNA sequencing approach. Other prominent studies by Tang *et al.* [34] and Gkountela *et al.* [35] analysed bulk cells at different stages [8].

For RNA and epigenetic studies, a precise identification of the correct cell type is critical to obtaining a pure dataset. Initial attempts to analyse PGCs within the gonads showed cells identified by eye as of a larger size and morphology as well as staining by alkaline phosphatase, which showed heterogeneity upon intensity (Fig. 1) [36,37].

In cell sorting experiments, the cell surface marker *CD117* (c-KIT), was described to result in a 100% pure population of primordial germ cells [34,35,38]. However, single-cell analyses of early PGCs, that include foetal cells of the mesonephros as control, have shown that PGC-neighbouring somatic cells also express *c-KIT* in early stages [8,34]. Immunofluorescence analyses revealed that *c-KIT* expression no longer correlate with the highly conserved germ cell marker *DDX4* (*VASA*) in 9.5-week-old female or in 11-week-old male embryos [38]. This indicates that the reliability of *c-KIT* as a marker for developing germ cells is limited to the early migratory and first postmigratory gonadal stages but other subpopulations occurring later in germ cell development could be missed.

Once the RNA of a single putative primordial germ cell has been obtained, the cells' identity can easily be confirmed by its remarkable transcriptional network. Gonadal PGCs share several factors with pluripotent cells like high alkaline phosphatase activity or the expression of core pluripotency genes such as *OCT4*, *NANOG*, *LIN28*, whereas human PGCs lack the expression of *SOX2* [37,39]. Depending on the developmental stage, they express germline related genes such as *BLIMP1*, *PIWIL2*, *DDX4*, as well as specifiers of other germ layers like Brachyury (*T*) of the mesoderm lineage, *SOX17* and *GATA4* of the endodermal lineage or even trophectodermal markers like *TEAD4* or *TFAP2C*. The germ cell identity is maintained by an active suppression of other somatic programs facilitated by *BLIMP1*. It was shown in *in vitro* experiments that the loss of BLIMP1 leads to an expression of mesodermal as well as endodermal markers [34,39].

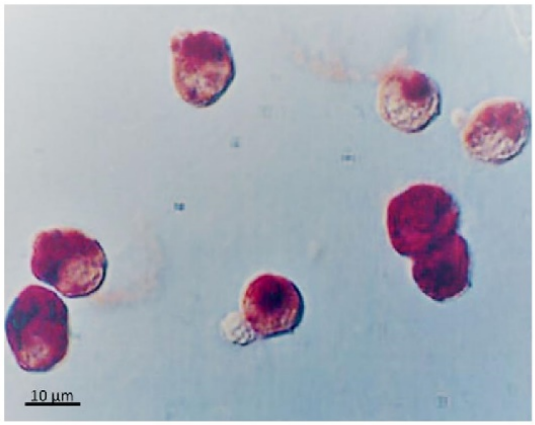


Fig. 1. Photograph showing positive alkaline phosphatase (AP) staining of human foetal germ cells at 10 weeks of gestation. Differences in staining intensity are detectable. A smaller, AP-negative cell is also seen. Figure adapted from [36].

Heterogeneity within the transcriptome

Single-cell research is a powerful tool to analyse heterogeneity of a cell population, especially during development where snapshots of different time points are essential. By analysing 233 single PGCs from 15 individual embryos compared to cells of the inner cell mass (ICM) and somatic cells, Guo *et al.* [8] could show that the transcriptomes of early PGCs between 4- and 11-weeks of development were quite stable. Surprisingly, gender did not discriminate distinct cells in the unsupervised hierarchical clustering and PCA, indicative of a relatively low heterogeneity between early PGCs [8].

A significant change in the transcriptome including upregulation of the meiosis marker in female PGCs was described in 17-week-old embryos. Male PGCs which undergo mitotic arrest and enter meiosis after puberty, also express the meiosis marker in 19-week-old embryos. Gene expression in these cells are not only significantly different compared to earlier PGCs but the heterogeneity within this population has increased enormously [8]. The exact timepoint of the onset of meiosis induction cannot be concluded from this dataset, due to missing data from PGCs between 11 and 17 weeks of development.

Different results, however, were obtained by Tang *et al.* [34] who analysed pooled PGCs of individual embryos of 5–9 weeks of age. In this dataset, the PGCs clustered by gender and age, with 9-week-old PGCs differing from younger cells were analysed. They specifically found an upregulation of meiosis and sexual differentiation markers compared to PGCs of 5.5-week-old embryos [34]. This induction of maturation has not been found in 11-week-old embryos in the Guo dataset.

All PGCs at the age of 17-weeks after fertilization were obtained from an individual embryo. To further analyse their heterogeneity, Guo *et al.* [8] chose two strategies to test for biases due to cell cycle effects or

technical errors. First, they applied the single-cell latent variability model. In this computational approach, genes that are related to the cell cycle or with an expression correlated with at least one cell cycle gene were removed from the analysis [28]. Guo *et al.* [8] also observed that PGCs before 11 weeks clustered together whereas PGCs of 19-week-old male and 17-week-old female embryos show great heterogeneity and cluster distinctively from the earlier cells [8].

In a second analysis, the "pool-and-split" tool developed by Marinov *et al.* [40], was used. It allows for comparison of a single cell to an 'average' single cell. After pooling nine single cells and splitting the isolated RNA into nine distinct analyses, the transcriptome of a real single PGC can be compared to an 'average' single PGC. For each developmental stage, only cells from the same embryo were analysed revealing that the number of heterogeneously expressed genes increased during development. The highest number of heterogeneous genes was found in 17-week-old female PGCs, including meiosis marker genes e.g. *SYCP1* [8]. Interestingly, PGC marker genes like *NANOS3*, *c-KIT* and *POU5F1* were heterogeneously expressed in 19-week-old male embryos. These bioinformatic tools show that the heterogeneity of PGCs increases during development with a stronger accession after 11 weeks. In the case of female PGCs, the asynchronous onset of meiosis induces the strongest gain of heterogeneity.

In the studies by Guo *et al.* [8] and Tang *et al.* [34], the timelines of PGC development differ in terms of the onset of meiosis in female PGCs and the initiation of a broader heterogeneity that might reveal the origination of different subpopulations of germ cell progenitors. The differences described here might be due to different technical approaches. Tang *et al.* [34] used a bulk RNA approach of isolated PGCs whereas Guo *et al.* [8] analysed single cells. Despite careful cell capturing, bulk datasets are prone to contamination by other cell types. The identity of a single cell can be confirmed after sequencing by the expression or absence of additional markers and data of low quality can eventually be omitted from the analysis.

Single-cell RNA sequencing approaches often miss low-expressed genes which might explain the detection of an earlier onset of meiosis in 9-week-old-embryos detected by Tang *et al.* [34], whereas Guo *et al.* [8] did not describe an expression of meiosis markers in 10- and 11-week-old PGCs. Focusing on the onset of meiosis, Tang *et al.* [34] analysed two 9-week-old female embryos, whereas Guo *et al.* [8] found meiosis marker in 31 single cells from one 17-week-old female embryo. When comparing single-cell RNA data to bulk data, Marinov *et al.* [40] described that at least 30–100 single cells were necessary to obtain a dataset that approached the quality of bulk datasets.

We compared expression levels of PGCs isolated and analysed by different approaches. We used the median expression level of single-cell RNA sequencing data from Guo *et al.* [8] in comparison to a dataset of bulk PGCs isolated by Goto *et al.* in 1999 [36]. The mRNA was reamplified in 2012 by Diedrichs *et al.* [37] and expression levels were determined employing microarrays (Fig. 2). The male and female PGCs were 10 weeks old. All single cells expressed the pluripotency markers *NANOG*, *POU5F1* and *KLF4*, but were negative for *SOX2* as expected for PGCs. In the bulk dataset, *NANOG* and *POU5F1* were expressed at low levels. The cells showed a high expression for a number PGC markers such as *PRDM1*, *UTF1* and *TDRD9*. Interestingly, in the single-cell data, other PGC marker genes were detected as expressed like *NANOS3*, *DAZL* and *DDX4*. In early embryonic development, *T* is a marker for the meso-endoderm lineage that was later restricted to the mesoderm [41]. Only 4-week-old PGCs expressed *T* together with other endodermal markers. Guo *et al.* [8] used a sorting approach to isolate *c-KIT* positive PGCs, whereas Goto *et al.* [36] isolated cells by eye and alkaline phosphataseactivity (Fig. 1). This might explain why different subpopulations of PGCs were isolated as they expressed different PGC marker genes. Furthermore, somatic markers were detected in the bulk dataset of Guo *et al.* [8] that might show a contamination by somatic cells. A discrimination of the gender by the male-specific *SRY* gene could not be detected at the analysed time points.

Technical limitations as well as biological variations, for example in the expression of marker genes, demand more extensive sampling of PGCs at different developmental stages and by different sorting approaches. The limited amount of human embryonic tissue, which might be of bad quality after abortion, complicates the comprehensive study of PGC development.

Epigenetic reprogramming

PGCs are not only unique for their simultaneous expression of pluripotency genes combined with somatic and germ line specifiers, but their epigenetic quality is also remarkable and highly relevant for the cells' identity. During embryonic development, ongoing differentiation together with the associated loss of potency is facilitated by modifications in the genome's accessibility, namely by DNA methylation, histone modifications or chromatin accessibility. The methylation of cytosine bases in the context of cytosine-

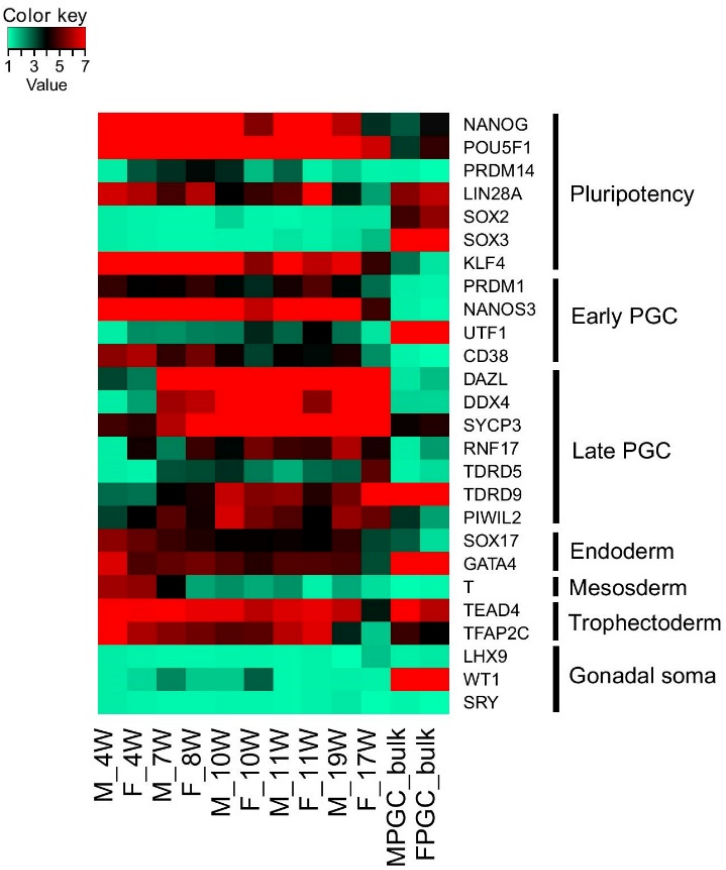


Fig. 2. Meta-analysis of bulk RNA [37] and single-cell data [8] related to distinct developmental stages of PGCs for selected primordial germ cell stage-specific marker genes; heatmap: high expression = red, low expression = green. Bulk samples were10 weeks old and labelled '_bulk', single cells were labelled by their age, e.g. '_4W'.

guanine-dinucleotides (CpG islands) within the DNA controls RNA transcription. Hyper-methylation of promoter or enhancer sequences induce gene silencing, hypo-methylation on the other hand, allows gene expression. In adult somatic tissue, about 80% of all CpG sites are methylated. Genomic imprinting describes allele-specific gene silencing, depending on maternal or paternal inheritance. The DNA methylation status can be analysed by genome sequencing after bisulphite-conversion [42]. The first application for a haploid genome of mouse germ cells still suffering from a low resolution was published by Guo *et al.* in 2013 [43]. DNA methylation analyses on the single-cell and whole genome levels are technically very challenging and a comprehensive analysis of human embryonic cells is still missing.

In early preimplantation development, embryonic cells undergo a first wave of global DNA demethylation, until cells of the blastocyst reach a minimal level of about 30% CpG methylation. Two weeks after fertilization, the DNA of the postimplantation embryo is completely remethylated [34]. Early germ cells, however, must erase epigenetic imprinting signatures to facilitate totipotency after fertilization. This process, called whole-genome epigenetic reprogramming is associated with a second wave of a more comprehensive CpG demethylation as well as global histone modifications [33,38].

Most of the studies on the PGC epigenome were done in mice or other nonprimate animals [33]. Data on human PGC reprogramming are limited, some aspects covering these studies were published by Guo

et al. [8], Tang *et al.* [34], Gkountela *et al.* [35] and reviewed by von Meyenn and Reik [44]. To the best of our knowledge, there is no single-cell analysis of the human PGC methylome but transcriptome data allow some conclusions on the dynamics of imprint erasure. Even though the above-mentioned studies differ in timeline and exact level of CpG methylation, a significant global demethylation of PGCs has already been found in 4-week-old embryos compared to somatic cells [44]. This rapid reprogramming indicates an active process rather than a passive demethylation by inhibition of *de novo* methylation [33]. Indeed, the demethylating enzyme methylcytosine dioxygenase 1 (*TET1*) was expressed in 4- to 19-week-old PGCs [8]. Interestingly, transcripts of *de novo* methyltransferases (*DNMT*) *DNMT1*, *DNMT3A* and *DNMT3B* were detected as expressed in the analysed PGCs, but protein expression could not be confirmed using immunofluorescence-based detection [35]. Pathway analyses of genes exclusively expressed by PGCs compared to ICM or somatic cells revealed activity of the base excision repair mechanism. This indicates that the global restructuring of the PGC epigenome requires active DNA repair machinery [8,33].

In female PGCs, another epigenetic phenomenon is the reactivation of the formerly inactivated X chromosome. In the earliest female PGCs analysed at 4 or 5.5 weeks after fertilization, both X chromosomes were active [8,34]. Interestingly, in single-cell RNA sequencing analyses, this correlates with a 1.6-fold increase in X chromosomal genes in female compared to male PGCs. This indicates that there are further gene-dosage regulating mechanisms in terms of an upregulation in male cells or a partial downregulation in female X chromosome encoded genes. Luckily, some of the female PGCs analysed carried heterozygous SNPs on their X chromosomes. As a prove of concept, these cells showed biallelic expression of the maternal and paternal inherited alleles, whereas somatic cells showed a monoallelic expression of either the one or the other SNP, depending on random X-inactivation in somatic cells [8].

Especially in their early development, PGCs are deeply embedded in the embryo and low in number. Technological advancements allow the analysis of such low cell numbers down to single cells. Bulk cell analyses might still have higher sensitivity and detect low expressed genes, whereas single cell analyses might allow insight into the heterogeneity of PGC subpopulations that increase during development. On this issue, the capturing of the right cells is of prime importance. In the studies presented, cells were isolated based on c-*KIT* expression. Immunofluorescent studies have shown that c-*KIT* negative germ cell progenitors identified by VASA expression exist in the foetal gonad. The same study has also showed that *OCT4* expression, which was exclusively identified in the nucleus in early PGCs (6–10.5 weeks), was found in the cytoplasm in older PGCs (9.5–16.5 weeks) [38]. This critical aspect is another source of heterogeneity that cannot be covered by transcriptome analysis alone. It would be interesting to analyse if the increasing heterogeneity of older PGCs was due to different temporal stages in their development or if it showed the origin of the different subpopulations. Normally in development, ongoing differentiation coincides with promoter silencing by methylation. Since majority of the PGC genome is free of any CpG methylation, a correlation of promoter methylation and RNA transcription on the single-cell level could elucidate the developmental program.

All recent studies analysed total RNA instead of mRNA alone, and found long noncoding RNAs (lncRNAs) expressed in PGCs like *XIST* which mediates X chromosome inactivation. The regulation of gene expression by noncoding RNAs in PGCs like micro RNA (miRNA) or lncRNA has not yet been analysed in depth but it shows significance in other developmental processes and cancer studies.

Heterogeneity within mature sperm

With the onset of male puberty, undifferentiated germ cells called spermatogonia differentiate into primary spermatocytes which are diploid for their chromosome number (2N), but tetraploid for their DNA content (4C). Following two meiotic divisions, each primary spermatocytes first divides into two secondary spermatocytes (MI) and then, the latter, into four aploid round spermatids (MII) which mature into spermatozoa. This process persists throughout life [45].

Meiotic cell division includes the process of recombination of homologous chromosomes to increase genomic diversity. Heterozygous SNPs which appear in every genome can be linked to one of the two alleles of a somatic genome. After sequencing the haploid genome of a single sperm cell, the identified SNPs can be mapped to one of the paternal haplotypes. Depending on the density of heterozygous SNPs detected, the number as well as average size of crossover events per sperm cell could be estimated [20,23].

Two individual studies analysed the genome of single human sperm cells, each from a different donor, and described their heterogeneity with similar results. By using heterozygous SNPs, Wang *et al.* [20] identified 22.8 crossover events per sperm cell with a median

distance of 87 Mb. Lu *et al.* [23] found 26 recombination events per cell, also with a median distance of about 80 Mb. The average distance between two crossover events was higher than expected if it appeared randomly (49 Mb) [20]. This indicates a positive recombination interference to ensure that two crossover events do not happen close to each other. The high resolution of genomic sequencing has shown that recombination happens less often close to genes, especially close to the transcription start site. 45% of all crossovers were located adjacent to a PRDM9 binding motif, a histone methyltransferase that is assumed to be connected to the distribution of recombination events [23]. From earlier population studies, a recombination map with hotspots for crossover events has been derived. Data from 91 single sperm showed an overlap of 58% with historical hotspots. However, a range from 0 to 100% overlap was reported within this pool of single cells [20].

A false segregation of chromosomes during meiosis can result in aneuploid sperm. In single cell analyses, 7% of the investigated sperm were found to be aneuploid, which is in accord with earlier studies that conducted FISH analyses on sperm [20]. Considering a sperm production of 150–275 million sperm per day, this relatively high amount of aneuploid sperm may contribute to the relatively low efficiency of human reproduction [45]. When comparing crossover numbers between single sperm cells, aneuploid sperm were found to have less recombination events. Furthermore, a decreased number of crossover events is associated with infertility [23].

Besides meiotic recombination *de novo* mutations as they occur during DNA replication contribute to the diversity within the population of single sperm cells. After excluding different biases and amplification errors, Wang *et al.* [20] confirmed 16 point mutations on average per sperm cell, of which three were identified as potential missense mutations in a protein coding region. Whether a point mutation already exists in the spermatogonia cells in the testis, or if it had newly formed during the differentiation process into spermatids, cannot be concluded from such a small dataset. However, the mutation rate in sperm is four to five times higher than in oocytes, which might be due to the higher number of germ line cell divisions in males [20].

The DNA of mature sperm is highly methylated. Their progenitor cells, however, undergo extensive epigenetic reprogramming during early embryonic development that necessitates broad remethylation. As was reviewed by Laurentino *et al.* [46], an aberrant DNA methylation of imprinted genes is linked to infertility. If epigenetic defects affect all of a man's sperm or subpopulations exist that might have originated from a small number of epimutated progenitor cells, further investigation of adequate number by single sperm analyses needs to be done. Such datasets are missing due to technical challenges of whole genome bisulphite sequencing of single sperm cells.

Heterogeneity of sperm is, at first, achieved by meiotic recombination. It has been shown to be a highly controlled, active process that drives genetic diversity within one generation and might also increase the probability of a successful fertilization. Furthermore, *de novo* mutations are also a source of diversity that, by their accumulation over many generations, rather have an impact on evolutionary time scales. However, epigenetic heterogeneity of sperm has not yet been comprehensively investigated in humans. From studies conducted till date, it can be suspected that epigenetic alterations that might have already appeared in early PGCs and by that affected the whole germ cell lineage until mature sperm decrease the fitness of sperm cells. New data on the sperm epigenome might benefit latest strategies of assisted reproductive technologies.

Genomics of single oocytes

After proliferation of female PGCs during embryonic development, PGCs start the first of two cycles of meiosis, whereas only those oocytes that will be fertilized later will complete both cycles. Still in the foetal period, right after completing the prophase of meiosis I, the diploid primary oocytes containing two of each chromosome, enter a meiotic arrest that will last until the specific egg is chosen to continue its development in the menstrual cycle after puberty.

Homologous recombination that shuffles the chromosomes to produce only unique oocytes, ends before entry into the meiotic I arrest. Continuing cell division of meiosis I, the primary oocyte divides into a secondary oocyte and a primary polar body, both equipped with the diploid and full set of each of the two chromatids. Now in meiosis II, the secondary oocyte halts in the metaphase II stage. Only after fertilization, can meiosis II be completed resulting in a female pronucleus of the fertilized zygote (as well as the male pronucleus) and another polar body, both haploids carrying one chromatid of each chromosome.

Each male spermatogonia produces four sperm cells, whereas the female primary oocyte develops into one oocyte and two polar bodies that are dispensable for embryonic development.

In an imposing study, Hou *et al.* [21] analysed the genome of single oocytes and the corresponding polar bodies to investigate recombination events. In most

cases (91%), they could predict the haplotype and the ploidy status of the female pronucleus by analysing polar bodies only, as they confirmed their results by sequencing the pronucleus [21].

In a similar approach as was described for the analysis of sperm [20,23], heterozygous SNPs were analysed to identify crossover events. On average, 43 crossover events were observed in oocytes of eight different donors. The higher amount compared to sperm was explained by the different longitudinal organization in female chromosomes as was also observed in the mouse and previously described [21,47]. The distribution of crossovers within the genome was comparable to that observed in sperm. The frequency was decreased close to transcription start sites of genomic regions, 44% of all events correlated with a PRDM9 binding motif [21].

The study by Hou *et al.* [21] showed that, if a potential genetic disease is carried only by the mother, analysis of the polar bodies is sufficient to select an unaffected zygote during *in vitro* fertilization. If the father carries a potential genetic defect, cells of the blastocyst stage must be analysed [21]. Possible *de novo* mutations of the female pronucleus, as were found in sperm cannot be deduced by sequencing the polar bodies.

When sequencing the DNA of a single sperm cell or an oocyte, only the haploid genome is available. In the present studies, the identification of crossover events depended on phased information for detected SNPs. Wang *et al.* [20] used lymphocytes of the donor's blood to analyse the diploid genome. Lu *et al.* [23] and Hou *et al.* [21] did not analyse the genome of the sperm or oocyte donors. Lu *et al.* [23] used the combined SNP linkage information across all sperm cells, whereas Hou *et al.* [21] exploited the diploid genome of the primary polar bodies of the oocytes. However, all approaches require accurate DNA amplification and sequencing methods as well as good algorithms to reliably predict recombination and *de novo* mutations in the gametes genome.

The transcriptome of an oocyte is also of great interest since it drives the first steps of embryonic development. For distinguishing between maternal transcripts and embryonic transcription, the haplotype and the phased genome of both parents are important as will be discussed later.

In vitro modelling of germ cell development

Recent attempts to model germ cell development by human pluripotent cells showed promising results. Azim Surani's laboratory published a protocol for PGC differentiation from naïve human embryonic stem cells including a *NANOS3* reporter, which is a well-established PGC marker gene [39]. Using this model system of human PGC-like cells (hPGCLC), they could dissect early regulatory networks that maintain PGC identity e.g. *SOX17* and *BLIMP1* expression. The role of *BLIMP1* in early PGC development has been tested by a *BLIMP1*-knock out, showing its function in repressing mesoendoderm specification and the initiation of methylome resetting [34,39]. A further development beyond the early PGC stage could not be simulated in this artificial system. A next step in analysing the full potential of these hPGCLCs would be transplantation into foetal gonads, which is contradicted by safety and ethical reasons.

The full process of oogenesis in mice has been reproduced in a dish by Hikabe *et al.* [48] who could grow mature oocytes from embryonic and induced pluripotent stem cells. These oocytes were used to produce healthy and fertile pups.

Growing heterogeneity in the early preimplantation embryo

After fertilization, male and female pronuclei fuse to form the zygote. This one-cell-embryo fulfils the first developmental step while it moves through the ovary until it implants in the uteral wall. This period of preimplantation development takes about seven days. The zygote divides into totipotent blastomeres by cell cleavage, denoting cell division without growth in total size. After reaching the multicellular stage of a compact morula, the cells start to specialize and lose their totipotency. Only within this short period, which is accompanied by a comprehensive demethylation of the embryonic genome, totipotent stem cells exist.

First wave of DNA demethylation

The zygote and early blastomeres undergo the first wave of demethylation after fertilization, followed by a second wave of demethylation only in germ cell development. The mature sperm and oocytes are highly methylated. This dynamic process is technically very challenging to analyse on the single-cell level with single nucleotide resolution. By reduced representative bisulphite sequencing (RRBS), only whole CpG islands which are CpG-rich hotspots for DNA methylation can be analysed. Other genomic regions with a lower content of CpG dinucleotides are missed. Whole genome bisulphite sequencing (WGBS) allows a more detailed insight into the methylation landscape of the

genome, but this has not yet been applied to human embryonic stem cells.

A comparison of single sperm cells, oocytes and the accompanying pronuclei has shown that the DNA methylation level decreases within the first day after fertilization with an increasing heterogeneity between individual cells over time. Interestingly, the demethylation in sperm and male pronuclei which initially have a higher methylation level than oocytes is more rapid and comprehensive than in the female counterpart [49]. It has been shown in the mouse that the maternal genome is passively demethylated by replication, whereas the paternal genome is actively demethylated by TET3 enzymes [43]. If this also holds true for human embryos, we currently do not know, nonetheless, single-cell transcriptome data show significant expression of the TET3 enzyme in the zygote [49].

During preimplantation development, DNA methylation stays at a low level with its minimum at the early blastocyst stage before implantation. The DNA of the postimplantation embryo, however, is highly methylated and this level only slightly increases when the cells differentiate into fully mature somatic cells [34,49].

The quality of RRBS-obtained single-cell data is limited by its low resolution, as RRBS covers only 10% of all CpG-dinucleotides in the genome. Whole-genome bisulphite sequencing that allows single nucleotide resolution has not yet been done on the single-cell level. When data obtained by whole-genome bisulphite sequencing were compared to RRBS obtained data, the methylation levels were comparable [49]. This indicates that the majority of epigenetic regulation by DNA methylation happens at the CpG-rich promoters and enhancers and the low coverage of RRBS data yields satisfying results when global methylation levels are compared.

Many developmental defects in early embryogenesis, as well as infertility in men can be caused by false imprinting. For assessing imprinted regions, an allele-specific DNA methylation analysis by utilizing heterozygous SNPs within methylated regions is necessary [49]. A sufficient coverage of such differentially methylated heterozygous SNPs on the single cell level could yield further insights into cell heterogeneity and tissue mosaicism caused by yet unknown imprinted regions.

Embryonic genome activation

The first cDNA library of single human preimplantation embryos for transcriptome analysis was published in 1997, 1999 [50,51]. A targeted profiling for known transcripts was then accessible using quantitative PCR or microarray approaches. It has long been known that the first spark of development is driven by maternal mRNAs provided in the cytoplasm of the oocyte and that the embryonic genome takes over at later stages. This was already analysed in 1988, when Braude *et al.* [52] showed that cell cleavage is not affected by transcriptional inhibitors until after the four cell stage of development. Besides the tremendous gain of insights by highly sensitive single-cell RNA sequencing, the exact timing of embryonic genome activation is still under debate. A clear evidence struggles with different detection rates and thresholds set for significant expression by different studies. As claimed by Blakeley *et al.* [53], the independent validation of computational approaches as well as an agreement on a common threshold for expression is missing in many studies.

The first cell cleavage into two blastomeres happens without any transcription. The cytoplasm of the zygote becomes partitioned between two daughter cells. One result is that the fewer transcripts per cell may be causing higher technical variability between the two blastomeres. Since the embryonic material per cell decreases with each cleavage, technical variations increase, rather affecting low expressed transcripts than highly expressed genes. Another biological effect is that cell cleavage results in a partitioning error leading to the first onset of heterogeneity in embryonic development [29,54].

With the onset of embryonic transcription and the degradation of maternal transcripts, the transcription noise increases [29]. The above-mentioned missing uniformity in the computational analysis of sequencing data leads to various conclusions about the dynamics of embryonic transcripts. All recent studies agree that 8-cell stage embryos differ mostly from earlier stages, indicating an almost complete take-over of the embryonic genome as well as a degradation of maternal transcripts [8,30,53,55,56]. Some studies using single-cell RNA sequencing also report a minor wave of embryonic genome activation at the two to four cell stage [30,55,56]. In contrast to this, Blakeley *et al.* [53] described only one wave of embryonic genome activation at the 4 to 8-cell stage. In this study, a much higher threshold of RPKM > 5 was used compared to the study by Yan *et al.* [30] who counted genes of RPKM > 0.1 as expressed. Interestingly, Blakely *et al.* [53] could identify *POU5F1 (OCT4)* only in the mouse zygote but not in human embryos until the 8-cell stage, concluding that maternal transcripts are different in mouse and humans. The absence of *POU5F1* in very early embryos in the dataset of Blakeley *et al.*

[53] might be due to technical or computational effects rather than biological reasons. Even though the dynamics of early transcription is different between species, *POU5F1* is well described as a maternal transcript regulating embryonic genome activation in mammals and vertebrates [57,58]. The absence of OCT4 leads to a developmental arrest at the 2-cell stage [59].

The initial onset of transcription from the embryonic genome has been investigated earlier also without the help of RNA-sequencing methods. Vassena *et al.* [60] applied the transcription blocker α-amanitin on early cell-stages and identified by targeted PCR which genes were maternally expressed or were sensitive to transcription inhibition implying embryonic transcription. They concluded that embryonic transcription starts as early as the 2-cell stage with the initiation of the expression of pluripotency related genes, started by *POU5F1* transcription which is already expressed in mature oocytes. Its expression level increases from the 2-cell stage onward due to embryonic transcription. The full set of pluripotency genes including *SOX2*, *NANOG* and *KLF4* was then detected at the 8-cell stage due to active transcription by the embryonic genome [60].

In some applications, a strategy using PCR methods might be a beneficial procedure to identify gene expression of known targets such as the detection of POU5F1 in the zygote. A quantitative analysis by single-cell RNA sequencing on the other hand allows insights into potentially unsuspected or unknown biological processes. Yan *et al.* [30] described that among 2500 genes upregulated during embryonic genome activation, many are related to RNA processing, translation and epigenetic regulation. This shows that the control of the embryonic genome becomes more comprehensive [30].

Embryos analysed by Xue *et al.* [55] were obtained by intracytoplasmic injection of sperm cells of one donor into oocytes from different egg donors. By following paternal SNPs after fertilization, 15–20% of all detected gene transcripts per stage could be assigned to the parent-of-origin. Interestingly, stage-specific paternal expression of a number of genes has been reported. The cell-cycle related gene *CDCA2* showed a maternal-exclusive expression in the 2- and 4-cell stage. After embryonic genome activation, the paternal transcript was detected in the 8-cell stage. At the morula stage, however, the paternal allele could not be identified [55]. Besides technical or cell-cycle reasons, heterogeneity of sperm cells used in different fertilization approaches might be an explanation for this observation. Furthermore, a mosaic-like or random monoallelic expression of some imprinted genes has also been described before [25].

Early lineage specification in the blastocyst

After compaction of the morula, approximately 5 days after fertilization, the first lineage specification of blastomeres takes place. The cells of the outer layer form the trophectoderm (TE) build the placenta and extraembryonic tissues. Cells inside in the blastocyst build the inner cell mass developing into embryonic tissue. By dissecting these two tissues by immunosurgery, a comparative transcriptome analysis was conducted and this revealed that pluripotency factors such as *POU5F1*, *SOX2* and *NANOG* become restricted to the ICM [61]. Within the ICM, the cells further commit to distinct lineages. Presumably caused by different cellular localization, cells closer to the inner fluid of the blastocyst develop into the primitive endoderm (PE), later also taking part in developing the extraembryonic membranes. The remaining cells which develop into the embryo proper are defined as the primitive ectoderm, also called epiblast (EPI) cells. Due to the very low number of cells in this developmental step, the originating heterogeneity can only be assessed by single-cell analysis as the transcriptome was described by several studies [30,53,56].

Principle component analyses showed a clear distinction of cells at day 5 expressing known TE or ICM markers. Most genes were expressed mutually exclusive but subpopulations with an unclear pattern also exist. These cells were described as immature and could commit to one of the lineages later in development. A further differentiation was described 6–7 days after fertilization, when gene expression of EPI and PE identifiers were clearly restricted, whereas in day 5 embryos some of those genes were commonly expressed [56]. After defining a clear gene set for the three lineages TE, EPI and PE, Petropoulos *et al.* [56] analysed the expression of the top 100 differentially expressed genes in each lineage in earlier blastomeres of day 3 and 4, respectively. Early EPI genes and a smaller subset of PE and TE genes were detected in day 4 embryos. They concluded that cells at this early stage were not yet committed to a lineage, but the segregation starts much earlier than previously expected. Furthermore, these data suggest that the differentiation of cells into the TE, EPI or PE lineage happens simultaneously and EPI vs PE restriction is not initiated after segregation to the ICM lineage [56].

A bias by the gender of the embryo could not be detected, as only 173 genes were differentially expressed in male and female preimplantation embryos

with a decreasing level during ongoing development from the 8-cell embryo to the blastocyst stage [56]. Among all upregulated genes in EPI cells, Yan *et al.* [30] found a significant enrichment of gene ontologies connected to reproduction and germ cell development, showing that these cells might later indicate germ cell development [30].

By single-cell analyses, many new insights were made into the growing heterogeneity of the very first blastomeres until the cell fate decision during implantation. The switch from maternal transcripts to embryonic control could not be completely elucidated. A combination of single-cell transcriptome sequencing methods with those detecting nascent RNA, for example by their introns can help to clarify this issue [62]. The period of preimplantation development is highly dynamic and very early expressed lineage specifiers can influence the cell fate without making any strict commitment in cell lineage decision. During *in vitro* fertilization, one or two blastomeres are usually isolated for preimplantation diagnostics without any negative effect to the developing embryo. This indicates that every blastomere keeps its full potential and the observed heterogeneity is only of minor functional relevance.

Outlook

Within the last 8 years, since Tang *et al.* [2] first analysed mouse blastomeres on the single-cell level, many more studies also on human embryos have been conducted and yielded new insights into human development. Many improvements in throughput, amplification biases, sequencing errors and computational strategies have been made. The combined information of the DNA and RNA sequence, its dynamics (regulated by epigenetic mechanisms) and the resulting protein interactions can one day lead to the goal of a comprehensive human cell atlas [4].

Many new methods for a combined analysis of the genome and the transcriptome as well as the transcriptome and the methylome from the same samples were published in the last years [63–66]. The analysis of proteins on the single cell level has been done by single-cell western blot [67] and single-cell mass-cytometry [68]. Even though the amount of information that can be obtained from a single cell has increased enormously, the relevance of the detected heterogeneity on DNA or RNA levels needs to be considered within a biological context. It can be assumed that cells react to changes in the amount of mRNA or protein expression in a threshold-dependent manner [18,69,70]. These combined approaches, together with information on the spatial context of a cell [71], will increase the understanding of the functional phenotype of a cell and we might obtain a better picture where changes finally induce a phenotypic switch during development.

Acknowledgement

This work was supported by the Düsseldorf School of Oncology (funded by the Comprehensive Cancer Centre Düsseldorf/Deutsche Krebshilfe and the Medical Faculty, Heinrich Heine University Düsseldorf).

Author contributions

JO, WW, JA did the conception. JO wrote the paper. WW did the bioinformatics. JA edited and approved the final version.

References

1 Koshland DE Jr (2002) The seven pillars of life. *Science* **295**, 2215–2216.
2 Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A *et al.* (2009) mRNA-seq whole-transcriptome analysis of a single cell. *Nat Methods* **6**, 377–382.
3 Grun D and van Oudenaarden A (2015) Design and analysis of single-cell sequencing experiments. *Cell* **163**, 799–810.
4 Wagner A, Regev A and Yosef N (2016) Revealing the vectors of cellular identity with single-cell genomics. *Nat Biotechnol* **34**, 1145–1160.
5 Gawad C, Koh W and Quake SR (2016) Single-cell genome sequencing: current state of the science. *Nat Rev Genet* **17**, 175–188.
6 Hoppe PS, Coutu DL and Schroeder T (2014) Single-cell technologies sharpen up mammalian stem cell research. *Nat Cell Biol* **16**, 919–927.
7 Yan L, Huang L, Xu L, Huang J, Ma F, Zhu X, Tang Y, Liu M, Lian Y, Liu P *et al.* (2015) Live births after simultaneous avoidance of monogenic diseases and chromosome abnormality by nextgeneration sequencing with linkage analyses. *Proc Natl Acad Sci U S A* **112**, 15964–15969.
8 Guo F, Yan L, Guo H, Li L, Hu B, Zhao Y, Yong J, Hu Y, Wang X, Wei Y *et al.* (2015) The transcriptome and DNA methylome landscapes of human primordial germ cells. *Cell* **161**, 1437–1452.
9 Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL and Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protocols* **7**, 562–578.
10 Anders S and Huber W (2010) Differential expression analysis for sequence count data, *Genome Biol* **11**, R106.

11 Kim D, Langmead B and Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357–360.
12 Dobin A, Davis C, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M and Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.
13 Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, Desai TJ, Krasnow MA and Quake SR (2014) Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375.
14 Gaublomme JT, Yosef N, Lee Y, Gertner RS, Yang LV, Wu C, Pandolfi PP, Mak T, Satija R, Shalek AK *et al.* (2015) Single-cell genomics unveils critical regulators of Th17 cell pathogenicity. *Cell* **163**, 1400–1412.
15 Wang Y and Navin NE (2015) Advances and applications of single-cell sequencing technologies. *Mol Cell* **58**, 598–609.
16 Van Loo P and Voet T (2014) Single cell analysis of cancer genomes. *Curr Opin Genet Dev* **24**, 82–91.
17 Shapiro E, Biezuner T and Linnarsson S (2013) Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat Rev Genet* **14**, 618–630.
18 Eberwine J, Sul J-Y, Bartfai T and Kim J (2014) The promise of single-cell sequencing. *Nat Methods* **11**, 25–27.
19 Stegle O, Teichmann SA and Marioni JC (2015) Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet* **16**, 133–145.
20 Wang J, Fan HC, Behr B and Quake SR (2012) Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* **150**, 402–412.
21 Hou Y, Fan W, Yan L, Li R, Lian Y, Huang J, Li J, Xu L, Tang F, Xie XS *et al.* (2013) Genome analyses of single human oocytes. *Cell* **155**, 1492–1506.
22 Zong C, Lu S, Chapman AR and Xie SX (2012) Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622–1626.
23 Lu S, Zong C, Fan W, Yang M, Li J, Chapman AR, Zhu P, Hu X, Xu L, Yan L *et al.* (2012) Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. *Science* **338**, 1627–1630.
24 Fan HC, Wang J, Potanina A and Quake SR (2011) Whole-genome molecular haplotyping of single cells. *Nat Biotechnol* **29**, 51–57.
25 Reinius B and Sandberg R (2015) Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. *Nat Rev Genet* **16**, 653–664.
26 Streets AM, Xiannian Z, Caoa C, Panga Y, Wua X, Xionga L, Yanga L, Fua Y, Zhaoa L, Tanga F *et al.* (2014) Microfluidic single-cell whole-transcriptome sequencing. *Proc Natl Acad Sci U S A* **111**, 7048–7053.
27 Eldar A and Elowitz MB (2010) Functional roles for noise in genetic circuits. *Nature* **467**, 167–173.
28 Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, Teichmann SA, Marioni JC and Stegle O (2015) Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol* **33**, 155–160.
29 Shi J, Chen Q, Li X, Zheng X, Zhang Y, Qiao J, Tang F, Tao Y, Zhou Q and Duan E (2015) Dynamic transcriptional symmetry-breaking in pre-implantation mammalian embryo development revealed by single-cell RNA-seq. *Development* **142**, 3468–3477.
30 Yan L, Yang M, Guo H, Yang L, Wu J, Li R, Liu P, Lian Y, Zheng X, Yan J *et al.* (2013) Single-cell RNA-seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol* **20**, 1131–1139.
31 Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lönnerberg P, Manno GL, Juréus A, Marques S, Munguba H, He L, Betsholtz C *et al.* (2015) Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142.
32 Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM *et al.* (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214.
33 Leitch HG, Tang WW and Surani MA (2013) Primordial germ-cell development and epigenetic reprogramming in mammals. *Curr Top Dev Biol* **104**, 149–187.
34 Tang WW, Dietmann S, Irie N, Leitch HG, Floros VI, Bradshaw CR, Hackett JA, Chinnery PF and Surani MA (2015) A unique gene regulatory network resets the human germline epigenome for development. *Cell* **161**, 1453–1467.
35 Gkountela S, Zhang KX, Shafiq TA, Liao WW, Hargan-Calvopina J, Chen PY and Clark AT (2015) DNA demethylation dynamics in the human prenatal germline. *Cell* **161**, 1425–1436.
36 Goto T, Adjaye J, Rodeck CH and Monk M (1999) Identification of genes expressed in human primordial germ cells at the time of entry of the female germ line into meiosis. *Mol Hum Reprod* **5**, 851–862.
37 Diedrichs F, Mlody B, Matz P, Fuchs H, Chavez L, Drews K and Adjaye J (2012) Comparative molecular portraits of human unfertilized oocytes and primordial germ cells at 10 weeks of gestation. *Int J Dev Biol* **56**, 789–797.
38 Gkountela S, Li Z, Vincent JJ, Zhang KX, Chen A, Pellegrini M and Clark AT (2013) The ontogeny of cKIT+ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint

erasure and in vitro differentiation. *Nat Cell Biol* **15**, 113–122.

39 Irie N, Weinberger L, Tang WW, Kobayashi T, Viukov S, Manor YS, Dietmann S, Hanna JH and Surani MA (2015) SOX17 is a critical specifier of human primordial germ cell fate. *Cell* **160**, 253–268.

40 Marinov GK, Williams BA, McCue K, Schroth GP, Gertz J, Myers RM and Wold BJ (2014) From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Res* **24**, 496–510.

41 Chu LF, Leng N, Zhang J, Hou Z, Mamott D, Vereide DT, Choi J, Kendziorski C, Stewart R and Thomson JA (2016) Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. *Genome Biol* **17**, 173.

42 Pixberg CF, Schulz WA, Stoecklein NH and Neves RP (2015) Characterization of DNA methylation in circulating tumor cells. *Genes (Basel)* **6**, 1053–1075.

43 Guo H, Zhu P, Wu X, Li X, Wen L and Tang F (2013) Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res* **23**, 2126–2135.

44 von Meyenn F and Reik W (2015) Forget the parents: epigenetic reprogramming in human germ cells. *Cell* **161**, 1248–1251.

45 Neto FT, Bach PV, Najari BB, Li PS and Goldstein M (2016) Spermatogenesis in humans and its affecting factors. *Semin Cell Dev Biol* **59**, 10–26.

46 Laurentino S, Borgmann J and Gromoll J (2016) On the origin of sperm epigenetic heterogeneity. *Reproduction* **151**, R71–R78.

47 Kleckner N, Storlazzi A and Zickler D (2003) Coordinate variation in meiotic pachytene SC length and total crossover/chiasma frequency under conditions of constant DNA length. *Trends Genet* **19**, 623–628.

48 Hikabe O, Hamazaki N, Nagamatsu G, Obata Y, Hirao Y, Hamada N, Shimamoto S, Imamura T, Nakashima K, Saitou M *et al.* (2016) Reconstitution in vitro of the entire cycle of the mouse female germ line. *Nature* **539**, 299–303.

49 Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, Yan J, Ren X, Lin S, Li J *et al.* (2014) The DNA methylation landscape of human early embryos. *Nature* **511**, 606–610.

50 Adjaye J, Daniels R, Bolton V and Monk M (1997) cDNA libraries from single human preimplantation embryos. *Genomics* **46**, 337–344.

51 Adjaye J, Bolton V and Monk M (1999) Developmental expression of specific genes detected in high-quality cDNA libraries from single human preimplantation embryos. *Gene* **237**, 373–383.

52 Braude P, Bolton V and Moore S (1988) Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature* **332**, 459–461.

53 Blakeley P, Fogarty NM, Del Valle I, Wamaitha SE, Hu TX, Elder K, Snell P, Christie L, Robson P and Niakan KK (2015) Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* **142**, 3613.

54 Tohonen V, Katayama S, Vesterlund L, Jouhilahti EM, Sheikhi M, Madissoon E, Filippini-Cattaneo G, Jaconi M, Johnsson A, Burglin TR *et al.* (2015) Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nat Commun* **6**, 8207.

55 Xue Z, Huang K, Cai C, Cai L, Jiang CY, Feng Y, Liu Z, Zeng Q, Cheng L, Sun YE *et al.* (2013) Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593–597.

56 Petropoulos S, Edsgard D, Reinius B, Deng Q, Panula SP, Codeluppi S, Plaza Reyes A, Linnarsson S, Sandberg R and Lanner F (2016) Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* **165**, 1012–1026.

57 Leichsenring M, Maes J, Mössner R, Driever W and Onichtchouk D (2013) Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science* **341**, 1005.

58 Lee MT, Bonneau AR, Takacs CM, Bazzini AA, DiVito KR, Fleming ES and Giraldez AJ (2013) Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **503**, 360–364.

59 Zuccotti M, Merico V, Sacchi L, Bellone M, Brink TC, Bellazzi R, Stefanelli M, Redi CA, Garagna S and Adjaye J (2008) Maternal Oct-4 is a potential key regulator of the developmental competence of mouse oocytes. *BMC Dev Biol* **8**, 97.

60 Vassena R, Boue S, Gonzalez-Roca E, Aran B, Auer H, Veiga A and Izpisua Belmonte JC (2011) Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. *Development* **138**, 3699–3709.

61 Adjaye J, Huntriss J, Herwig R, BenKahla A, Brink TC, Wierling C, Hultschig C, Groth D, Yaspo ML, Picton HM *et al.* (2005) Primary differentiation in the human blastocyst: comparative molecular portraits of inner cell mass and trophectoderm cells. *Stem Cells* **23**, 1514–1525.

62 Ameur A, Zaghlool A, Halvardson J, Wetterbom A, Gyllensten U, Cavelier L and Feuk L (2011) Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat Struct Mol Biol* **18**, 1435–1440.

63 Mertes F, Lichtner B, Kuhl H, Blattner M, Otte J, Wruck W, Timmermann B, Lehrach H and Adjaye J (2015) Combined ultra-low input mRNA and whole-genome sequencing of human embryonic stem cells. *BMC Genom* **16**, 925.

64 Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, Teng MJ, Goolam M, Saurat N, Coupland P, Shirley LM *et al.* (2015) G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat Methods* **12**, 519–522.

65 Dey SS, Kester L, Spanjaard B, Bienko M and van Oudenaarden A (2015) Integrated genome and transcriptome sequencing of the same cell. *Nat Biotechnol* **33**, 285–289.

66 Angermueller C, Clark SJ, Lee HJ, Macaulay IC, Teng MJ, Hu TX, Krueger F, Smallwood SA, Ponting CP, Voet T *et al.* (2016) Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods* **13**, 229–232.

67 Hughes AJ, Spelke DP, Xu Z, Kang CC, Schaffer DV and Herr AE (2014) Single-cell western blotting. *Nat Methods* **11**, 749–755.

68 Bendall SC, Simonds EF, Qiu P, Amir el AD, Krutzik PO, Finck R, Bruggner RV, Melamed R, Trejo A, Ornatsky OI *et al.* (2011) Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum, *Science* **332**, 687–696.

69 Tanay A and Regev A (2017) Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338.

70 Li GW and Xie XS (2011) Central dogma at the single-molecule level in living cells. *Nature* **475**, 308–315.

71 Nichterwitz S, Chen G, Aguila Benitez J, Yilmaz M, Storvall H, Cao M, Sandberg R, Deng Q and Hedlund E (2016) Laser capture microscopy coupled with smart-seq2 for precise spatial transcriptomic profiling. *Nat Commun* **7**, 12139.

3.4. Combined ultra-low input mRNA and whole genome sequencing of human embryonic stem cells

Florian Mertes, Björn Lichtner, Heiner Kuhl, Mirjam Blattner, **Jörg Otte**, Wasco Wruck, Bernd Timmermann, Hans Lehrach and James Adjaye,

Abstract

Background: Next Generation Sequencing has proven to be an exceptionally powerful tool in the field of genomics and transcriptomics. With recent development it is nowadays possible to analyze ultra-low input sample material down to single cells. Nevertheless, investigating such sample material often limits the analysis to either the genome or transcriptome. We describe here a combined analysis of both types of nucleic acids from the same sample material.

Methods: The method described enables the combined preparation of amplified cDNA as well as amplified whole-genome DNA from an ultra-low input sample material derived from a sub-colony of in-vitro cultivated human embryonic stem cells. cDNA is prepared by the application of oligo-dT coupled magnetic beads for mRNA capture, first strand synthesis and 3'-tailing followed by PCR. Whole-genome amplified DNA is prepared by Phi29 mediated amplification. Illumina sequencing is applied to short fragment libraries prepared from the amplified samples.

Results: We developed a protocol which enables the combined analysis of the genome as well as the transcriptome by Next Generation Sequencing from ultra-low input samples. The protocol was evaluated by sequencing sub-colony structures from human embryonic stem cells containing 150 to 200 cells. The method can be adapted to any available sequencing system.

Conclusions: To our knowledge, this is the first report where sub-colonies of human embryonic stem cells have been analyzed both at the genomic as well as transcriptome level. The method of this proof of concept study may find useful practical applications for cases where only a limited number of cells are available, e.g. for tissues samples from biopsies, tumor spheres, circulating tumor cells and cells from early embryonic development. The results we present demonstrate that a combined analysis of genomic DNA and messenger RNA from ultra-low input samples is feasible and can readily be applied to other cellular systems with limited material available.

Share of scientific contribution: 10%

Jörg Otte contributed to the writing of the manuscript during the revision process addressing reviewers' comments.

Published: BMC Genomics, Nov 12, 2015, Vol.925, Issue 16,
DOI: 10.1186/s12864-015-2025-z

Link to the publication:
https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-015-2025-z

Mertes et al. BMC Genomics (2015) 16:925
DOI 10.1186/s12864-015-2025-z

BMC Genomics

RESEARCH ARTICLE Open Access

Combined ultra-low input mRNA and whole-genome sequencing of human embryonic stem cells

Florian Mertes[1,4*], Björn Lichtner[1], Heiner Kuhl[2], Mirjam Blattner[1], Jörg Otte[3], Wasco Wruck[3], Bernd Timmermann[2], Hans Lehrach[1] and James Adjaye[1,3*]

Abstract

Background: Next Generation Sequencing has proven to be an exceptionally powerful tool in the field of genomics and transcriptomics. With recent development it is nowadays possible to analyze ultra-low input sample material down to single cells. Nevertheless, investigating such sample material often limits the analysis to either the genome or transcriptome. We describe here a combined analysis of both types of nucleic acids from the same sample material.

Methods: The method described enables the combined preparation of amplified cDNA as well as amplified whole-genome DNA from an ultra-low input sample material derived from a sub-colony of in-vitro cultivated human embryonic stem cells. cDNA is prepared by the application of oligo-dT coupled magnetic beads for mRNA capture, first strand synthesis and 3'-tailing followed by PCR. Whole-genome amplified DNA is prepared by Phi29 mediated amplification. Illumina sequencing is applied to short fragment libraries prepared from the amplified samples.

Results: We developed a protocol which enables the combined analysis of the genome as well as the transcriptome by Next Generation Sequencing from ultra-low input samples. The protocol was evaluated by sequencing sub-colony structures from human embryonic stem cells containing 150 to 200 cells. The method can be adapted to any available sequencing system.

Conclusions: To our knowledge, this is the first report where sub-colonies of human embryonic stem cells have been analyzed both at the genomic as well as transcriptome level. The method of this proof of concept study may find useful practical applications for cases where only a limited number of cells are available, e.g. for tissues samples from biopsies, tumor spheres, circulating tumor cells and cells from early embryonic development. The results we present demonstrate that a combined analysis of genomic DNA and messenger RNA from ultra-low input samples is feasible and can readily be applied to other cellular systems with limited material available.

Keywords: Next generation sequencing, RNA-seq, Whole-genome sequencing, Ultra-low input sequencing, Single cell, Pluripotency, Embryonic stem cells

Background

Within recent years an overwhelming number of specific methods and protocols emerged for next-generation sequencing [1]. Amongst them, transcriptome as well as whole-genome sequencing were of prime interest. Both sequencing methods have tremendously accelerated our understanding of both the more dynamic function of RNAs and the more static composition of the genome within a functional cell. Transcriptome sequencing focuses on deciphering the complex expression pattern of RNAs [2, 3], therefore identifying novel expressed RNAs and transcript variants as well as isoforms which in turn lead to a better understanding of cell regulation, function and networks. Whole-genome sequencing has for example highlighted insights into the subtle differences amongst the human population [4, 5] or major genomic re-arrangements found in cancer cells [6] with both having a significant impact on cell fate and the living organism.

* Correspondence: mertes@molgen.mpg.de; james.adjaye@med.uni-duesseldorf.de
[1]Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany
Full list of author information is available at the end of the article

Major improvements for the preparation of sequencing libraries for RNA-seq as well as DNA-seq have emerged [7, 8]; continually reducing the input amount needed which is generally in the microgram range. Within recent years the field of single-cell sequencing for transcriptome and genome sequencing has advanced significantly [9, 10]. There are already many examples available were either RNA [11–15], or DNA [8, 16, 17] have been analyzed down to the single-cell resolution. Recently, studies with the combined analysis of the genome and transcriptome of the same cell have also been published [18, 19]. This is especially advantageous for applications where only a small fraction of the sample should be analyzed, or more importantly, where the sample is composed of a few cells only. Such scenarios include for example early embryonic development which starts from a single cell expanding to a few dozen cells within the first developmental stages [20]. The elucidation of intra-tumor heterogeneity in biopsies [21, 22] as well as in *in-vitro* grown primary tumor spheres [23], or the characterization of circulating tumor cells [24] rely on the analyses of limited cell material. In addition, *in-vitro* cultured stem cells from both mouse and human are limited in the number of cells if sub-population and sub-colony differences in terms of gene expression are under investigation. For all settings, already subtle changes in genome integrity can have a major impact on the expression and regulation of RNAs, and proteins within cells.

Despite the advancements for both areas of sequencing minute amounts of either RNA or DNA, an assay enabling the combined sequencing of RNA and DNA from the very same sample still in the ultra-low input range would add to our understanding of the regulation and developmental processes affected by both, the function of genome integrity as well as RNA expression and gene function.

Here we describe a method which enables the preparation of whole transcriptome amplified cDNA as well as the generation of whole-genome amplified DNA from the same ultra-low input material derived from a sub-colony of *in-vitro* cultivated human embryonic stem cells. Firstly, whole transcriptome amplified cDNA was prepared from mRNA only by using oligo-dT coupled magnetic beads, following cDNA synthesis, 3'-tailing and PCR amplification. Secondly, after magnetic coupling of the mRNA/oligo-dT beads, whole-genome amplified DNA was prepared from the retained DNA by Phi29 mediated amplification. Both, the amplified cDNA as well as DNA were subjected to standard procedures for multiplex short fragment library preparation enabling Illumina sequencing. Using this approach, both the transcriptome as well as the genome of the same sample could be analyzed on both levels of nucleic acids present in cells, the RNA and DNA.

Results

Ultra-low input RNA sequencing

In brief, cells for RNA-seq were collected from human embryonic stem cells (hESCs) serving as biological samples. Colonies of hESCs were mechanically dissociated into 200 μm × 200 μm square fragments consisting of 150–200 cells (Fig. 1b). The undifferentiated and pluripotent state of the cells was verified by microscopic assessment of morphology (small, densely-packed cells with high nuclei:cytoplasm-ratio growing in a homogeneous monolayer) and positive immunocytochemical co-staining for the well-established transcription factors and hESC-markers OCT3/4 and NANOG [25] (Fig. 1b).

The picked sub-colony fragment was directly transferred into lysis buffer. After cell lysis, the solution was supplemented with oligo-dT coupled magnetic micro-beads and transferred to columns placed in a magnetic field for further processing. To selectively enrich mRNA out of the total RNA, cDNA synthesis was performed with oligo-dT coupled magnetic beads. After on column cDNA synthesis, beads with cDNA were retained by centrifugation followed by 5′-tailing and PCR amplification. The size distribution of amplified cDNA ranged from 200–3000 base pairs. PCR products were fragmented by sonication to 150 to 300 base pairs and multiplex fragment library preparation was performed for paired-end Illumina sequencing. Figure 1 gives an overview of the developed methodology.

In this study we report data from sequencing of two hESC samples in the low sub-colony range (150–200 cells, Fig. 1b) which were analyzed by 100 base pair paired-end sequencing on a single flow cell on an Illumina HiSeq instrument. We obtained 314.2 million raw reads on a single lane, after barcode mapping for sample allocation we obtained 65.2 million reads for RNA-seq sample 1 and 58.8 million reads for RNA-seq sample 2 respectively. Furthermore 190.5 million reads belonged to whole genome sequencing performed for sample 1. The RNA-seq reads were mapped with Tophat resulting in 58.9 and 54.7 million mapped reads (90.2 and 93.5 %) correspondingly. The number of duplicate reads was found to be 1.9 % for the WGA-DNA sample and 54.6 and 52.8 % for the RNA-seq sample 1 and 2 respectively. Duplicate read counts were based on mapped reads with the same start and end point. In total 11,755 Refseq genes with a read coverage of ≥ 5 reads were identified (sample 1: 8523; sample 2: 10,908; overlap: 7676) in both hESC samples.

Next imperative parameters for RNA-seq were determined based on the Refseq dataset, specifically the total length of mRNA showing sequence coverage and coverage distribution along the 5'- to 3'-orientation. Since we performed cDNA synthesis by oligo-dT priming it is common sense to observe a bias towards the 3′-end of genes, especially for long transcripts (Fig. 2a). The median length of Refseq cDNAs observed were around

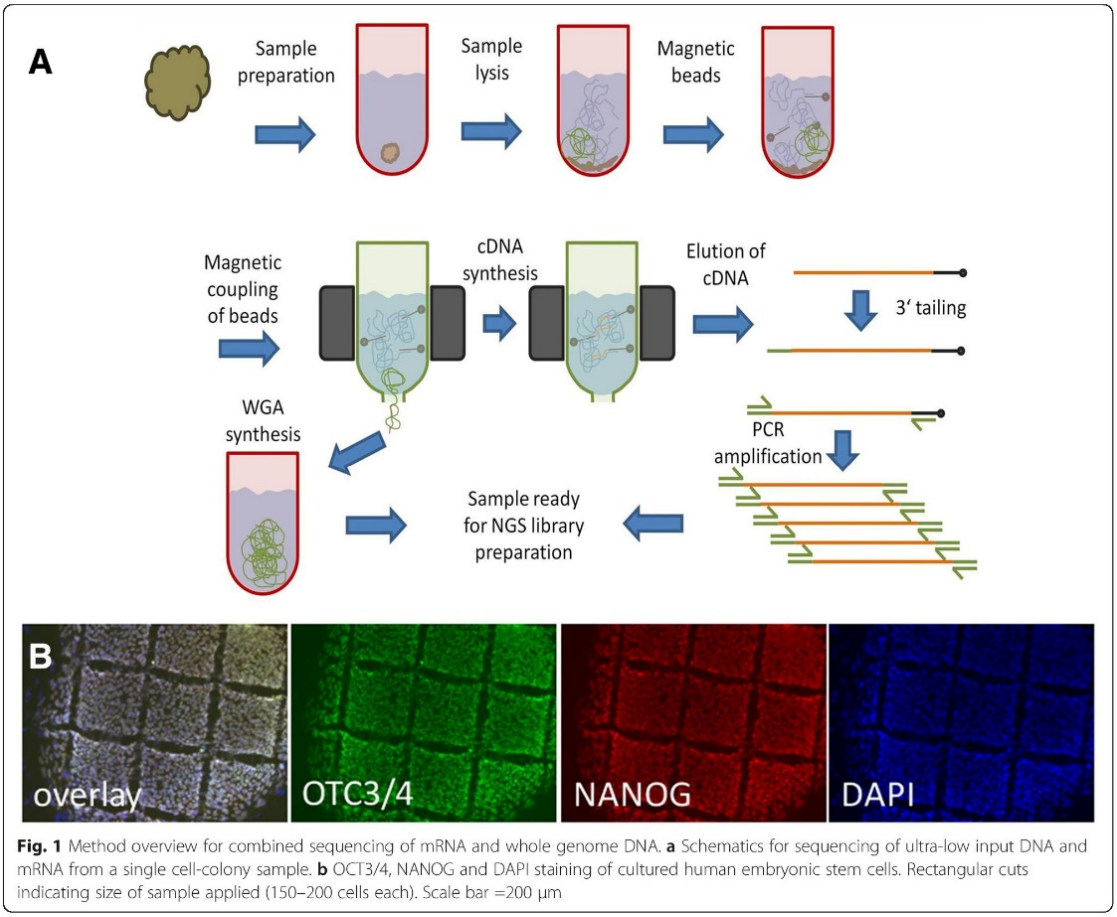


Fig. 1 Method overview for combined sequencing of mRNA and whole genome DNA. **a** Schematics for sequencing of ultra-low input DNA and mRNA from a single cell-colony sample. **b** OCT3/4, NANOG and DAPI staining of cultured human embryonic stem cells. Rectangular cuts indicating size of sample applied (150–200 cells each). Scale bar =200 μm

900–1100 base pairs ranging from approximately 450 to 2000 base pairs (lower and upper quartile) with single cDNAs longer than 10 kb (Fig. 2b). The average coverage distribution along the 5'- to 3'-orientation of Refseq genes was calculated for transcript size intervals of 0–1 kb, 1–2 kb, 2–3 kb, 3–4 kb, 4–5 kb and 5–15 kb. For transcripts in the range of 1–2 kb normalized coverage was almost 80 % over the full transcript length with decreasing coverage towards the last 15 % of bases at the 5′-end. For transcripts ranging from 2–5 kb normalized coverage was at least 50 % (Fig. 2a). Subsequently we evaluated the correlation between both RNA-seq samples. This was done by comparing FPKM values obtained for expressed Refseq genes resulting in a Pearson's correlation factor of 0.85 for RNA-seq sample 1 and sample 2 (Fig. 2c).

To further evaluate the RNA-seq data we used expression analysis performed with an Illumina BeadArray with the same hESC line. The BeadArray experiments were performed with the appropriate amount of mRNA in comparison to the low input RNA-seq experiments. A comparison of expressed Refseq genes for both RNA-seq samples and the Illumina BeadArray showed a high degree of concordance for both methods. For analysis only genes were included which gave rise to FPKM >0.5 for NGS data and p-value <0.05 in BeadArray and were considered significant. In total 13,630 genes were identified in both RNA-seq samples whereas the BeadArray identified 10,834 genes. The total overlap between both sequencing experiments and the BeadArray was found to be 3486 Refseq genes. Moreover the overlap for RNA-seq sample 1 and BeadArray was found to be 4081 and for RNA-seq sample 2 to be 5172, respectively (Fig. 3a). Next, a Consensus Pathway Data Base (CPDB) overrepresentation analysis was performed to identify congruence of BeadArray and NGS experiments in terms of overlapping genes and categories (Fig. 3a). Significant genes from all experiments were analyzed in CPDB for categories using pathways from KEGG, Reactome, BioCarta and Wikipathways and compared for categories with a p-value threshold of 0.05. In total 506 categories

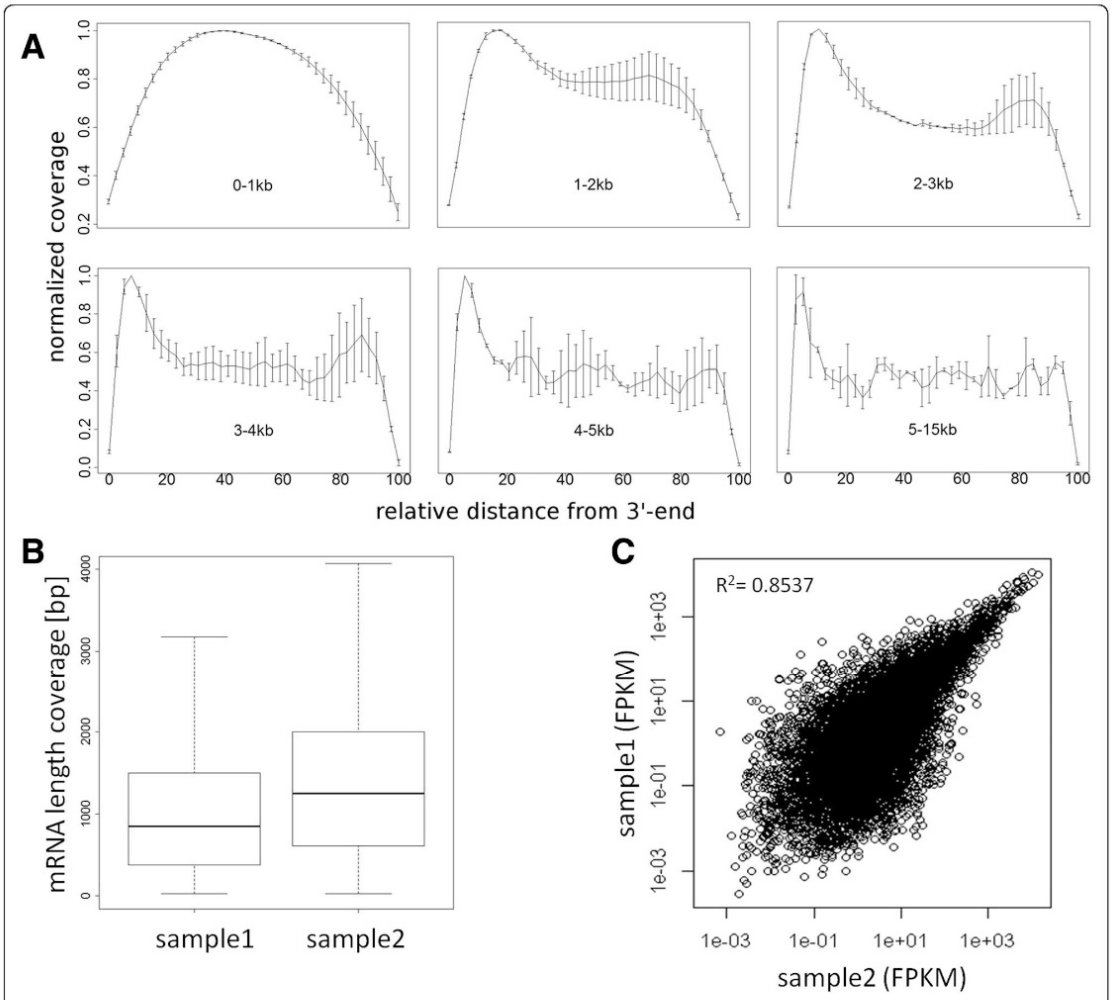


Fig. 2 Comparison of RNAseq data in terms of transcript identification. **a** Read coverage across transcripts with coverage for transcripts of sizes 0 kb-1 kb, 1 kb-2 kb, 2 kb-3 kb, 3 kb-4 kb, 4 kb-5 kb and 5 kb-15 kb. Transcripts were divided into 40 equally sized bins and coverage was averaged over all transcripts of the dedicated size interval. Mean values of mRNA sample1 and sample2 are displayed, error bars denoting standard deviation of samples. **b** Boxplot of total length of sequenced Refseq genes for RNAseq sample 1 and sample 2. **c** Correlation plot for RNAseq sample 1 and sample 2 in terms of FPKM values for expressed Refseq genes

were identified for the BeadArray experiment and 375 and 415 categories for RNA sequencing sample 1 and sample 2 respectively. Overall overlap for all experiments was found to be 238 and for BeadArray and combined RNA samples 322 categories. For RNA sequencing sample 1 and RNA sample 2 only the overlap was found to be 320 categories.

To evaluate pluripotency of hESC samples in BeadArray and RNA sequencing experiments common pluripotency marker genes were compared. Comparison was performed after normalization of gene expression by graphical representation analysis (Fig. 3b). The differential analysis of gene expression showed highest similarity between both RNA samples followed by BeadArray. Analysis of expressed genes among all samples showed clustering of genes in groups of two for *DNMT3B* and *CD9*, *SOX2* and *NANOG* and *POU5F1* (*OCT3/4*) and *IFITM1* with very similar gene expression in BeadArray and both RNA sequencing samples. Examples of sequencing coverage for single pluripotency marker genes (*NANOG, POU5F1* and *SOX2)* as well as housekeeping gene (*ACTB*) are shown in Fig. 4b.

One inherent advantage of RNA-seq over microarray-based analysis is the identification of splice variants and isoforms. Hence both RNA-seq samples were analyzed for expressed isoforms focusing on genes which are

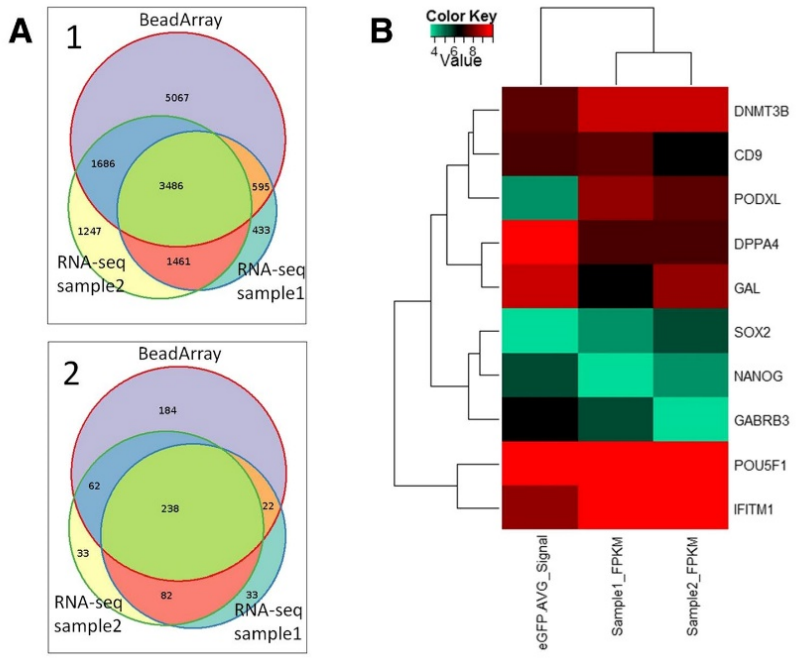


Fig. 3 Overlap between microarray and sequencing measurements. **a.1** Venn diagram displaying overlap of significant genes found in the Illumina microarray and next-generation sequencing (NGS) experiments. Genes expressed with Illumina detection *p*-value < 0.05 in the microarray experiment or FPKM > 0.5 in the NGS experiments were considered significant. Genes were compared via HGNC gene symbol annotation. **a.2** Venn diagram displaying overlap of significant CPDB categories found in the Illumina microarray and deep sequencing experiments. Significant genes from microarray and NGS experiments were analysed with the CPDB functional annotation tool and categories with a *q*-value <0.05 were considered significant. **b** Pluripotency associated genes from microarray experiment and NGS experiments of sample1 and sample2 were compared. Logarithmic values (base 2) of Illumina average signals and FPKM values from expressed genes (detection *p*-value < 0.05, FPKM > 0.5) were quantile normalized and subjected to cluster analysis via R heatmap2 function using Euclidean distance as distance measure

known to be important for the maintenance of the undifferentiated and pluripotent state of both hESCs and induced pluripotent stem cells (Table 1).

Ultra-low input whole genome sequencing

The sequencing of DNA was performed from the same sample as the sequencing of mRNA. The DNA contained in the human embryonic stem cells was collected during magnetic coupling of the mRNA/oligo-dT complexes and before cDNA synthesis was performed (Fig. 1a). The retained DNA was subjected to Phi29 mediated whole genome amplification (WGA) producing high molecular mass DNA. The WGA DNA was subjected to sonication (range 200–250 base pairs), followed by paired-end fragment library preparation and multiplex sequencing on the same flow cell as the mRNA samples on an Illumina HiSeq instrument. In total 190.5 million reads corresponding to the DNA of sample 1 were obtained. From these 153.5 million reads (80.6 %) could be mapped to hg19, giving rise to an average 6-fold genome coverage. The coverage for individual chromosomes ranged from 3-fold coverage for the X-chromosome to 13-fold coverage for chromosome 19 respectively. Furthermore chromosome coverage on a single base pair resolution was found to be ≥90.0 % for twelve chromosomes (chromosomes 2–5, 7, 10–12, 17–20). Lowest coverage was observed for the Y-chromosome with 34.2 % of sequenced bases. The average coverage for the full set of chromosomes was 82.8 % sequenced base pairs. A sequencing coverage overview over all chromosomes is presented in Fig. 4a.

Discussion

To date the vast majorities of analyses of minute amounts of cell material down to single cells for next generation sequencing are still limited to either transcriptome or genome sequencing [9, 10]. Aiming the great advantages of a combined analysis, latest studies include different approaches for genome and transcriptome sequencing of the same cell [18, 19, 26]. Within the pre-NGS era the analysis of single cells was performed by a combination of microarray-based techniques for gene expression and

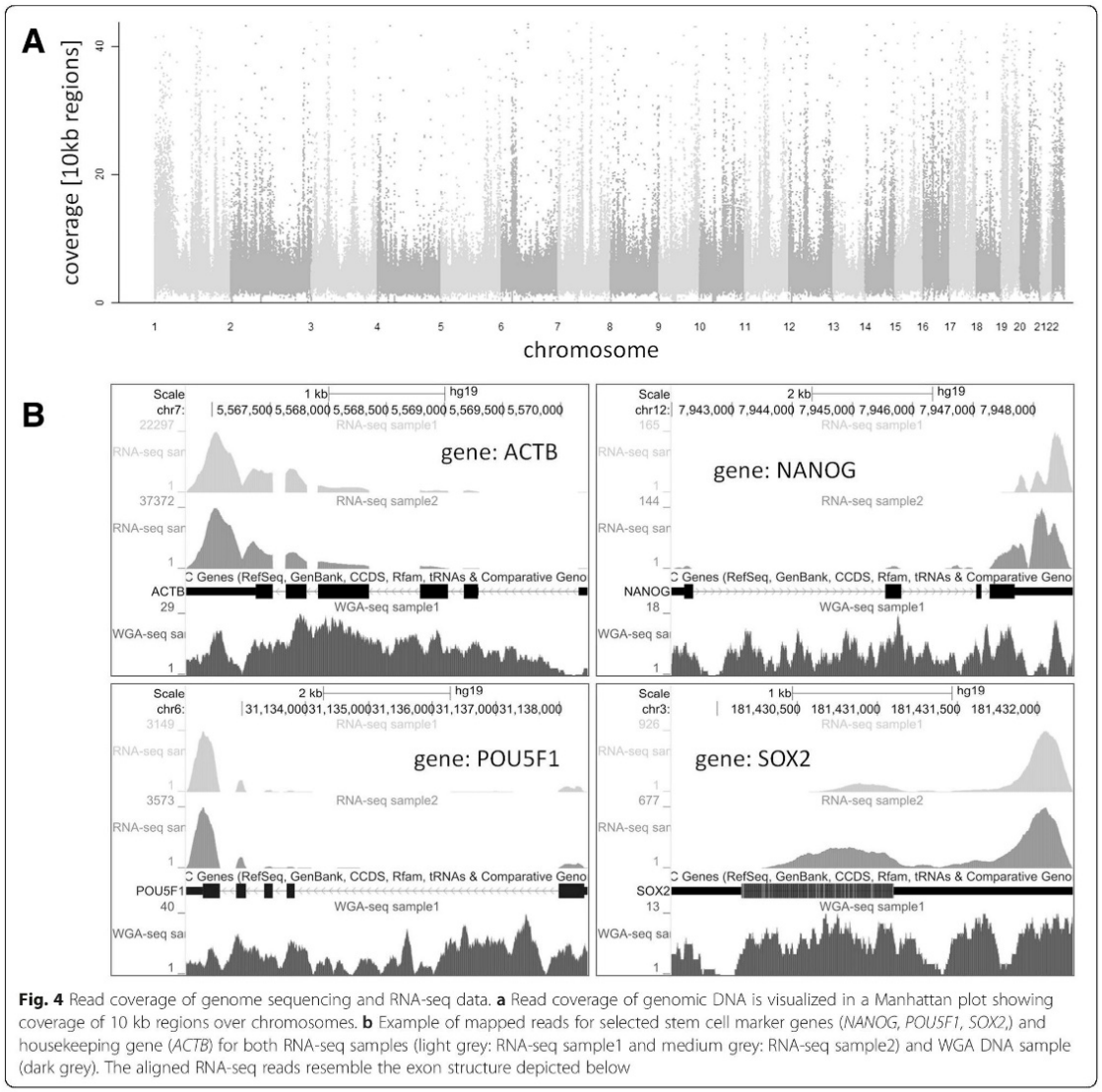


Fig. 4 Read coverage of genome sequencing and RNA-seq data. **a** Read coverage of genomic DNA is visualized in a Manhattan plot showing coverage of 10 kb regions over chromosomes. **b** Example of mapped reads for selected stem cell marker genes (*NANOG*, *POU5F1*, *SOX2*,) and housekeeping gene (*ACTB*) for both RNA-seq samples (light grey: RNA-seq sample1 and medium grey: RNA-seq sample2) and WGA DNA sample (dark grey). The aligned RNA-seq reads resemble the exon structure depicted below

comparative genomic hybridization to study genomic aberrations from the same cell [27]. Enabling a profound transcriptome as well as genome analysis by next generation sequencing from minute amounts of sample material will have many applications and will enable the study of rare cells. These rare cells may originate from cell-type specific differentiation of stem cells, but also from cancer tissue [22], circulating tumor cells [28] and early embryonic development [11] amongst others. Furthermore, it is known that micro-heterogeneity plays a fundamental role not only for stem cell function but for many biological processes [29]. Having tools to dissect these micro-heterogeneities on a transcriptomic as well as genomic level can help to further understand the underlying function of biological processes and might become of clinical relevance not only in cancer therapies but also in personalized medicine like pharmacogenomics or prenatal diagnosis.

Our results demonstrate that a combined transcriptome and genome analysis is feasible from minute amounts of sample material. Sequencing was performed from 150–200 hESC which equals about 200 pg total RNA according to the findings of Islam et al. [30] and about 1 ng of DNA for a diploid genome. In our approach the developed method for combined transcriptome and genome sequencing generated robust and

Table 1 Detected isoforms of selected human embryonic stem cell marker genes

Gene	Nearest ref ID	Locus	Length	Sample1 FPKM	Sample2 FPKM	Number of isoforms
DNMT3B				212.59	132.05	4/7 (13)
	NM_006892	chr20:31350190-31397167	789	9.71	4.87	
	NM_006892	chr20:31350190-31397167	721	0.00	3.00	
	NM_006892	chr20:31350190-31397167	3463	18.81	0.00	
	NM_006892	chr20:31350190-31397167	4131	0.00	21.60	
	NM_006892	chr20:31350190-31397167	4203	175.74	0.54	
	NM_006892	chr20:31350190-31397167	4336	8.33	65.00	
	NM_175848	chr20:31350190-31397167	4276	0.00	29.39	
	NM_175849	chr20:31350190-31397167	4087	0.00	7.65	
NANOG				3.91	6.62	1/1 (1)
	NM_024865	chr12:7941994-7948655	2089	3.91	6.62	
POU5F1				507.35	484.58	3/2 (21)
	NM_001173531	chr6:31132113-31138451	1247	156.01	0.00	
	NM_002701	chr6:31132113-31138451	1401	348.96	483.12	
	NM_203289.6	chr6:31132113-31138451	1733	2.38	1.46	

Gene expression is shown in FPKM values including detected isoforms. Number of detected isoforms for RNA-seq sample1 and sample2 are separated by dash; total number of known isoforms in brackets

quantitative data for studying gene expression, isoforms etc. on one hand and the genome on the other hand. This appears an especially advantageous situation in the case of studying cancer where remarkable chromosome instability causes genome heterogeneity and is directly linked to alterations in transcriptome dynamics [31]. The effect of genomic copy number variations and SNPs on the transcriptome has been investigated with an integrative approach [32]. A further elucidation of the impact of genomic alterations on the phenotype of a cell can be analyzed preferentially by the presented method, where DNA and RNA are sequenced from the very same cell material.

Obtaining full length coverage of expressed genes is vital for the identification of isoforms and splice variants. Therefore we specifically evaluated the average coverage along Refseq genes from the 5'- to the 3'-end. We found noteworthy variation in terms of average coverage from the 5'- to the 3'-end with significant more reads observed at the 3′-end of the transcript. Hence our RNA-seq data reflect a bias towards the 3'-end mainly attributed to the application of oligo-dT primed cDNA preparation. This finding is in agreement with findings of other groups who analyzed minute amounts of sample material down to single cells [1, 11, 28, 33]. Other protocols offer highly multiplexed single-cell sequencing, nevertheless only a limited number of bases either from the 5'-end [30] or the 3'-end [34] are sequenced to enable mainly gene expression studies. The duplication rate for mapped reads found in both RNA-seq samples is in line with comparable approaches eg by Adiconis et al. [35] where also an oligo-dT RNA preparation was performed and a duplication rate for low input samples of ~20 % (SMARTseq) and ~90 % (TRUseq) was reported. In their publication Dey et al. [18] do not present any information about duplicate reads observed in DNAseq and RNAseq, presumably due to their unique combined RNA and DNA amplification approach. The number of duplicate read counts observed by Macaulay et al. [19] ranged from 10 to <50 % for genome sequencing depending on cell line and cell number analyzed, unfortunately no read count for duplicate reads are given for RNA-seq, only the number of total and mapped reads are presented. Therefore the duplicate reads we found for RNA-seq seem currently more or less normal, nevertheless it is desirable to reduce the number in future, to make sequencing more efficient.

Purification of mRNA from the much more abundant ribosomal RNA using magnetic micro beads coupled to oligo-dT primers has become a widely established method in the last years. The recovery of complementary poly-T sequences is relatively high, whereas non-target RNAs represent an insignificant part of the enriched molecules [36]. Beyond the high target selectivity, the practicability and short assay duration of the method, further advantages are its compatibility to modifications. It has been shown that chemical conditions can also be changed to preserve proteins in their native state for further proteomic analyses [37]. However, the abovementioned 3′ skew bias has been faced in different approaches, like the cDNA transcription using SMARTer [28] or the additional mRNA enrichment by its 5′ cap [38, 39]. The latter procedure also deals with the intricate capturing of mRNAs with short poly-A tails since

its length might be influenced in the course of translational control [40]. Some new protocols avoid a physical separation of RNA from DNA because it might be detrimental when automatic liquid handling is conducted in small volumes. In such approaches, the oligo-dt primer contains an additional barcode sequence to identify the amplified cDNA within the pool of genomic DNA [41]. The limitation of all methods using oligo-dT primer to target polyadenylated mRNA is the inability to detect microRNAs or lncRNAs, which are also of importance for the transcriptome and the cell's phenotype.

All current NGS platforms require prior amplification of DNA if minute amounts of DNA are analyzed. Novel approaches specifically utilized for single-cell whole genome amplification promise an improved genome coverage [17]. However, compared to established methods locus dropout is still observed and comparable to the results we obtained for whole genome sequencing. Alternatively a transposon-mediated library preparation strategy omitting any amplification reaction prior to NGS library preparation [8] may offer an alternative approach for the method presented here.

Several aspects of the presented method offer room for improvement. Beyond the separation of the mRNA the preparation of amplified cDNA as well as WGA DNA involves many discrete steps which are prone for technical variation, such as the magnetic coupling and binding of mRNA to the column, the elution of DNA from the column, the PCR amplification of the double-stranded cDNA and library preparation which adds another step of PCR amplification. Secondly, retaining full length coverage of expressed RNAs in ultra-low input preparations is especially challenging. Furthermore locus drop out on the genomic DNA is more likely to occur with decreasing cell numbers within the WGA reaction. All these technical challenges are exacerbating if the cell number is decreased down to single cells and all currently available methods need to cope with [10].

Conclusion

In conclusion the presented approach for combined ultra-low mRNA and whole-genome sequencing from minute amounts of starting material offers new possibilities for many applications where limited material is available. Furthermore it enables one to directly study both the transcriptome and genome in one analytical approach from the same sample material which might be of interest for both basic as well as clinical research.

Methods

Cell culture and cell picking

Human embryonic stem cells (line H1) were obtained from WiCell Research Institute. Cells were cultured in 6-well-plates (TPP) coated with Matrigel (Becton Dickinson) on Mitomycin C-inactivated mouse embryonic fibroblasts (MEFs) as described before [42]. After 1 week, undifferentiated colonies were mechanically fragmented using the StemProEZPassage Disposable Stem Cell Passaging Tool (Invitrogen, cat# 23181–010) according to the recommendations of the manufacturer, leading to squares of relatively uniform size (ca. 200 μm × 200 μm, see Fig. 1b). Fragments from the middle of undifferentiated colonies were detached using a non-rotatable cell spatula (TPP, cat#99010) under microscopic control (Stereo microscope Leica MZ9.5 with cold light source KL 1500 LCD; Leica Microsystems) and sterile conditions inside a HERAguard® HPH 9 Laminar flow clean bench (Heraeus). Detached single squares were individually isolated by very gentle aspiration using a sterile 20 μl filter pipette tip (Biozym Scientific) and used for further downstream processing.

Immunocytochemistry

Cells were fixed with 4 % paraformaldehyde (Electron Microscopy Sciences) in PBS (Gibco/Invitrogen) for 15 min, washed two times with PBS and then stained as described before [42]. Primary antibodies: anti-OCT3/4 (C-10) Mouse monoclonal antibody (Santa Cruz Biotechnology, cat#sc-5279) and anti-NANOG Goat polyclonal antibody (R&D Systems, cat#AF1997). Secondary antibodies: anti-Mouse IgG (H + L) (from chicken) labelled with red-fluorescent Alexa Fluor594 (Invitrogen, cat#A-21201) and anti-Goat IgG (H + L) (from donkey) labelled with green-fluorescent Alexa Fluor488 (Invitrogen, cat#A-11055). Nuclei were counterstained with DAPI. Fluorescence microscopy and photographing was performed using Axiovert 200 M (Zeiss) and Software AxioVision Rel. 4.8 (Zeiss).

Ultra low input cDNA and WGA-DNA preparation

Preparation and amplification of nucleic acids were performed with a customized version of the μMACS SuperAmp Kit (Miltenyi Biotec); if not mentioned explicitly, procedures were according to the manufacturer protocol. The protocol is based on magnetic coupling of mRNA and retaining the nucleic acid in low volume flow-through columns for greatly simplified handling. For selective mRNA isolation, magnetic micro beads coupled to oligo-dT primers were applied. To retain the genomic DNA the eluates from the first two washing steps after loading the cell lysate onto the column were collected into a 1.5 mL reaction tube for later DNA precipitation and whole genome amplification. On column cDNA synthesis was performed at 42 °C for 60 min according to the following protocol by applying the total reaction master mix onto the column: 20 μL contained 2 μL 10× Reverse Transcriptase Buffer (Ambion), 0.5 mM dNTPs, 1 μg T4 Gene 32 Protein (NEB), 400 U

M-MLV Reverse Transcriptase (Enzymatics), 20 U RNase Inhibitor (Ambion). After collection of magnetic beads containing synthesized cDNA by centrifugation and 3'-tailing according to the manufacturer, PCR amplification was performed. To the 3'-tailing reaction, in total 30 μL, the following PCR master mix was added: 76.5 μL PCR master mix contained 14 μL 5× Phusion HF buffer (Finnzymes), 0.5 mM dNTPs, 60 μL resuspended μMACS SuperAmp PCR mix, 2 U PhusionTaq (Finnzymes); the following cycling conditions were applied on a PTC-200 (MJ Research) thermal cycler: 78 °C for 30 s, 95 °C for 1 min, [98 °C for 3 s, 64 °C for 30 s, 72 °C for 2 min]× 40 cycles, 72 °C for 5 min.

Amplification of genomic DNA was performed with the REPLI-g Midi Kit (Qiagen) for 16 h at 30 ° C according to the manufacturer's recommendations. Before whole genome amplification, DNA was ethanol precipitated by adding 0.1 volumes of 3 M sodium acetate solution and 5 μg glycogen (Ambion) to 1 volume of DNA sample. After precipitation the pellet was resuspended in 10 μL of Elution Buffer (Qiagen).

Library preparation and NGS

Library preparation for next generation sequencing was performed according to the Illumina TruSeq DNA Sample Preparation Guide with the Low-Throughput protocol. The indexed paired-end libraries had an insert size in the range of 150–300 base pairs. Subsequent pooling of the samples with a ratio of 3:1:1 (wgaDNA:mRNA1:mRNA2) cluster generation and DNA sequencing was performed on a single lane of an IlluminaHiSeq instrument with a 100 base pair paired-end sequencing chemistry.

Mapping and data analysis

RNA-seq mapping

RNA-seq data was mapped to the human genome by Tophat v1.3.3. Prebuild bowtie index files and annotations in GTF-format were downloaded from Illumina's iGenomes ftp-server (ussd-ftp.illumina.com/Homo_sapiens/UCSC/hg19/). Duplicate read counts were estimated on mapped reads using Picard (http://broadinstitute.github.io/picard/). To compare Tophat mappings to Illumina BeadArray data, we considered only reads that mapped to annotated exons (UCSC genes) and reached peak coverage of 5 or higher. Furthermore we restricted the analysis to Refseq genes.

Whole genome mapping

Genomic DNA sequencing reads were mapped using bowtie with the same parameters that Tophat uses for its first mapping round (–v 2). Duplicate read counts were estimated on mapped reads using Picard (http://broadinstitute.github.io/picard/). Genomic coverage was visualized as Manhattan plot via the *mhtplot* function from the R package *gap*.

Transcriptome read coverage analysis

Coverage was calculated via the IGVtools command count [43] from the exon aligned BAM files for RNA sample1 and sample2 using default settings. For transcript coverage window size 25 was used, for genomic coverage window size 10,000 was used. Transcript calculations were based on exon unions of human genes from ENSEMBL V74 for plus and minus strand separately. Each transcript was divided into 40 equally sized bins according to the method of [28]. To compensate for missing values data points corresponding to the 40 bins were determined by interpolation of the IGVtools results via cubic spline curve fitting (function *spline*) from the statistical software package R. Resulting values were normalized via division by the maximum. These transcript coverages were averaged for transcript size intervals 0 kb –1 kb, 1 kb –2 kb, 2 kb –3 kb, 3 kb –4 kb, 4 kb –5 kb and 5 kb –15 kb. Plus and minus strand were summarized by calculating mean values for all transcript size intervals. Finally, mean values and standard deviations were determined between the two RNA samples shown in the coverage plots. Boxplots of sequenced mRNA length were plotted by the R package.

Overlap between microarray and sequencing measurements

Congruence of Illumina microarray and next generation sequencing experiments was determined in terms of overlapping genes and overlapping categories found via Consensus Pathway Data Base (CPDB) overrepresentation analysis [44]. Genes were considered significantly expressed when the FPKM (Fragments per kilobase of exon per million fragments mapped) values were greater than 0.5 in the sequencing experiments or Illumina detection p-value was less than 0.05. Additionally, these genes were subjected to a CPDB overrepresentation analysis using pathways from KEGG, Reactome, BioCarta and Wikipathways and the resulting categories were compared using a threshold of 0.05 for p-values adjusted via the Benjamini-Hochberg method. The results were displayed in Venn diagrams from R package Vennerable.

Cluster analysis of pluripotency associated genes

Illumina average signals from microarray experiment and FPKM values from sequencing experiments of sample1 and sample2 were compared with respect to pluripotency associated genes [29] which were expressed in the microarray experiment (detection p-value < 0.05) and in the sequencing experiments (FPKM > 0.5). Logarithmic values (base 2) of these measurements were quantile normalized and subjected to cluster analysis via R heatmap2 function using Euclidean distance as distance measure.

Availability of supporting data

The data sets supporting the results of this article are available in the GEO repository at GEO accession number GSE69471 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE69471).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

FM designed the study, performed research and wrote the manuscript. MB and BL performed research and contributed to the manuscript. JO contributed to the writing of the manuscript. HK and WW performed bioinformatic data analysis and contributed to the writing of the manuscript. BT and HL reviewed the manuscript. JA contributed to the manuscript and reviewed the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We are grateful to Stefan Wild from Miltenyi Biotec for the support with the μMACS SuperAmp Kit. Furthermore we would like to thank Ilona Hauenschild, Daniela Roth and Sonia Paturrej from the sequencing core facility for their excellent technical assistance. The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement n° 115234, resources of which are composed of financial contribution from the European Union's Seventh Framework Program (FP7/2007–2013) and EFPIA companies' in kind contribution. JA acknowledges support from the EU FP7 project AgedBrainSYSBIO (Grant Agreement N ° 305299) (http://www.agedbrainsysbio.eu), the Duesseldorf School of Oncology (funded by the Comprehensive Cancer Center Düsseldorf/Deutsche Krebshilfe and the Medical Faculty HHU Düsseldorf).

Author details

[1]Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany. [2]Next Generation Sequencing Group, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany. [3]Institute for stem cell research and regenerative medicine, Medical Faculty, Heinrich Heine University, Moorenstr. 5, 40225 Düsseldorf, Germany. [4]Molecular Exposomics, Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany.

Received: 7 July 2015 Accepted: 7 October 2015
Published online: 12 November 2015

References

1. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010;11(1):31–46.
2. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet. 2011;12(2):87–98.
3. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57–63.
4. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061–73.
5. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56–65.
6. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell. 2011;144(1):27–40.
7. Adiconis X, Borges-Rivera D, Satija R, DeLuca DS, Busby MA, Berlin AM, et al. Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nat Methods. 2013;10(7):623–9.
8. Parkinson NJ, Maslau S, Ferneyhough B, Zhang G, Gregory L, Buck D, et al. Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA. Genome Res. 2012;22(1):125–33.
9. Kalisky T, Blainey P, Quake SR. Genomic analysis at the single-cell level. Annu Rev Genet. 2011;45:431–45.
10. Shapiro E, Biezuner T, Linnarsson S. Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat Rev Genet. 2013;14(9):618–30.
11. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods. 2009;6(5):377–82.
12. Dalerba P, Kalisky T, Sahoo D, Rajendran PS, Rothenberg ME, Leyrat AA, et al. Single-cell dissection of transcriptional heterogeneity in human colon tumors. Nat Biotechnol. 2011;29(12):1120–7.
13. Goetz JJ, Trimarchi JM. Transcriptome sequencing of single cells with Smart-Seq. Nat Biotechnol. 2012;30(8):763–5.
14. Adjaye J, Bolton V, Monk M. Developmental expression of specific genes detected in high-quality cDNA libraries from single human preimplantation embryos. Gene. 1999;237:373–83.
15. Patel P, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 2014;344(6190):1396–401.
16. Peters BA, Kermani BG, Sparks AB, Alferov O, Hong P, Alexeev A, et al. Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells. Nature. 2012;487(7406):190–5.
17. Zong C, Lu S, Chapman AR, Xie XS. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science. 2012;338(6114):1622–6.
18. Dey SS, Kester L, Spanjaard B, Bienko M, van Oudenaarden A. Integrated genome and transcriptome sequencing of the same cell. Nat Biotechnol. 2015;33(3):285–9.
19. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, Teng MJ, et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods. 2015.
20. Adjaye J, Huntriss J, Herwig R, BenKahla A, Brink TC, Wierling C, et al. Primary differentiation in the human blastocyst: comparative molecular portraits of inner cell mass and trophectoderm cells. Stem Cells. 2005;23(10):1514–25.
21. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366(10):883–92.
22. Sottoriva A, Spiteri I, Piccirillo SG, Touloumis A, Collins VP, Marioni JC, et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proc Natl Acad Sci U S A. 2013;110(10):4009–14.
23. Sachs N, Clevers H. Organoid cultures for the analysis of cancer phenotypes. Curr Opin Genet Dev. 2014;24:68–73.
24. Yu M, Ting DT, Stott SL, Wittner BS, Ozsolak F, Paul S, et al. RNA sequencing of pancreatic circulating tumour cells implicates WNT signalling in metastasis. Nature. 2012;487(7408):510–3.
25. Prigione A, Fauler B, Lurz R, Lehrach H, Adjaye J. The senescence-related mitochondrial/oxidative stress pathway is repressed in human induced pluripotent stem cells. Stem Cells. 2010;28(4):721–33.
26. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178–92.
27. Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol. 2012;30(8):777–82.
28. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 update. Nucleic Acids Res. 2013;41 (Database issue):D793–800.
29. Wolfrum K, Wang Y, Prigione A, Sperling K, Lehrach H, Adjaye J. The LARGE principle of cellular reprogramming: lost, acquired and retained gene expression in foreskin and amniotic fluid-derived human iPS cells. PLoS One. 2010;5(10), e13703.
30. Lichtner B, Knaus P, Lehrach H, Adjaye J. BMP10 as a potent inducer of trophoblast differentiation in human embryonic and induced pluripotent stem cells. Biomaterials. 2013;34(38):9789–802.
31. Guzvic M, Braun B, Ganzer R, Burger M, Nerlich M, Winkler S, et al. Combined genome and transcriptome analysis of single disseminated cancer cells from bone marrow of prostate cancer patients reveals unexpected transcriptomes. Cancer Res. 2014;74(24):7383–94.
32. Klein CA, Seidl S, Petat-Dutter K, Offner S, Geigl JB, Schmidt-Kittler O, et al. Combined transcriptome and genome analysis of single micrometastatic cells. Nat Biotechnol. 2002;20(4):387–92.
33. Ramos CA, Bowman TA, Boles NC, Merchant AA, Zheng Y, Parra I, et al. Evidence for diversity in transcriptional profiles of single hematopoietic stem cells. PLoS Genet. 2006;2(9), e159.
34. Islam S, Kjällquist U, Moliner A, Zajac P, Fan J-B, Lönnerberg P, et al. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. 2011;21(7):1160–7.

35. Stevens JB, Horne SD, Abdallah BY, Ye CJ, Heng HH. Chromosomal instability and transcriptome dynamics in cancer. Cancer Metastasis Rev. 2013.
36. Laurila K, Autio R, Kong L, Narva E, Hussein S, Otonkoski T, et al. Integrative genomics and transcriptomics analysis of human embryonic and induced pluripotent stem cells. BioData Min. 2014;7(1):32.
37. Van Loo P, Voet T. Single cell analysis of cancer genomes. Curr Opin Genet Dev. 2014;24:82–91.
38. Hashimshony T, Wagner F, Sher N, Yanai I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep. 2012;2(3):666–73.
39. Adams NM, Bordelon H, Wang KK, Albert LE, Wright DW, Haselton FR. Comparison of three magnetic bead surface functionalities for RNA extraction and detection. ACS Appl Mater Interfaces. 2015;7(11):6062–9.
40. Petersen TS, Andersen CY. Simultaneous isolation of mRNA and native protein from minute samples of cells. Biotechniques. 2014;56(5):229–37.
41. Blower MD, Jambhekar A, Schwarz DS, Toombs JA. Combining different mRNA capture methods to analyze the transcriptome: analysis of the Xenopus laevis transcriptome. PLoS One. 2013;8(10), e77700.
42. Weiss B, Curran JA. CAP selection: A combined chemical-enzymatic strategy for efficient eukaryotic mRNA enrichment via the 5' cap. Anal Biochem. 2015.
43. Laird-Offringa IA, De Wit CL, Elfferich P, Van Der Eb AJ. Poly(A) tail shortening is the translation-dependent step in c-myc mRNA degradation. Mol Cell Biol. 1990;10(12):6132–40.
44. Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods. 2013;10(11):1096–8.

4. Conclusion

In this thesis, we have discussed and analyzed stem cells of different qualities. We were able to culture primary colon CSCs as 3D organoids by conditions relying on minimal niche requirements independent of any marker expression. Despite the high malignancy of the sampled tissues, we could show that a forced suppression of FGFR signaling abrogates organoid formation indicating a dependency on this crucial pathway. Cyst formation after FGFR inhibition on established organoids indicates differentiation suggesting that colon CSCs need an active FGF signaling to maintain self-renewal as it is also described for hESCs. These results further suggest that within highly malignant cancer cells, stemness is still regulated by extrinsic factors, putting the stem cell niche into focus. The first studies describing CSCs in solid tumors depended on isolating cells according to cell surface marker expression. As discussed in 1.3.1, these markers might rather select for fitter cells sustaining the stressful sorting and engraftment procedure than identifying real stemness qualities. As recent data pointed out the high fluidity of stem cell-like properties of cells in healthy as well as in cancerous tissue, the search for stem cell functions instead of marker expression gains growing significance [19, 54]. The organoid cell culture model, which can easily be combined with modern lineage tracing strategies as well as with extensive animal models, meets these modern functional requirements [60].

The level of cellular differentiation within cancerous tissue often inversely correlates with its malignancy. This aspect is underlined by our finding that the *FGFR2* gene was consistently and significantly overexpressed in ADPA tumors indicating an active role in the tumorigenesis of this highly malignant and metastasizing cancer. Interestingly, besides the *FGFR2* gene we could also identify *SMAD2* as overexpressed in ADPA tumors suggesting an interaction of FGF and TGF-β signaling as it has also been observed in our colon cancer organoid model. Whether the modern CSC concept also holds true for ADPA and whether FGF2 plays a pivotal and functional role in stem cell-like cells of this tumor entity needs to be subject of future investigations. While the hair follicle bulge of the skin is also a well-described adult stem cell compartment, the cellular composition of sweat glands, especially in regards of its differentiation hierarchy, has not yet been investigated [65].

In other tumor entities such as non-small cell lung cancer (NSCLC), prostate or bladder cancer, which commonly overexpress FGFRs, therapies targeting the FGF-signaling pathway were developed and approved substances are in clinical use [66, 67]. Beyond its role in oncogene addiction, however, FGF-signaling might also be involved in escape

mechanisms against EGFR-targeted therapies [66, 68]. This redundant kinase activation, which we also observed after FGFR-inhibition in our organoid model, enforces a comprehensive and detailed understanding of molecular downstream effects when targeting tyrosine kinase receptors. Pan- and multi target tyrosine kinase inhibitors are promising achievements of the recent years [67].

A cancer's cellular heterogeneity fosters its malicious potential to withstand all modern treatment options, including surgery, chemo- and radiation therapy, anti-angiogenic or modern immune- and targeted therapies. The unique identity of each cell is not only based on stochastic (epi-)genetic alterations but also on mutually dependent factors like environmental stimuli, the spatial context or the cell cycle state [69]. Modern, single-cell based research in tumor biology, regenerative medicine and embryonic development needs to cope with this complexity. Even though live will always be greater than the sum of its parts.

5. Literature

1. De Los Angeles, A., Ferrari, F., Xi, R., Fujiwara, Y., Benvenisty, N., Deng, H., Hochedlinger, K., Jaenisch, R., Lee, S., Leitch, H. G., Lensch, M. W., Lujan, E., Pei, D., Rossant, J., Wernig, M., Park, P. J. & Daley, G. Q. (2015) Hallmarks of pluripotency, *Nature.* **525**, 469-78.
2. Singh, V. K., Saini, A., Kalsan, M., Kumar, N. & Chandra, R. (2016) Describing the Stem Cell Potency: The Various Methods of Functional Assessment and In silico Diagnostics, *Front Cell Dev Biol.* **4**, 134.
3. Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S. & Jones, J. M. (1998) Embryonic Stem Cell Lines Derived from Human Blastocysts, *SCIENCE.* **282**, 1145-1147.
4. Chagastelles, P. C. & Nardi, N. B. (2011) Biology of stem cells: an overview, *Kidney Int Suppl (2011).* **1**, 63-67.
5. Takahashi, K. & Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors, *Cell.* **126**, 663-76.
6. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. & Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors, *Cell.* **131**, 861-72.
7. Matz, P., Spitzhorn, L. S., Otte, J., Kawala, M. A., Woestmann, J., Yigit, H., Wruck, W. & Adjaye, J. (2017) 4.10 - Use of Stem Cells in Toxicology A2 - Chackalamannil, Samuel in *Comprehensive Medicinal Chemistry III* (Rotella, D. & Ward, S. E., eds) pp. 177-194, Elsevier, Oxford.
8. Adjaye, J., Bolton, V. & Monk, M. (1999) Developmental expression of specific genes detected in high-quality cDNA libraries from single human preimplantation embryos, *Gene.* **237**, 373-383.
9. Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, II & Thomson, J. A. (2007) Induced pluripotent stem cell lines derived from human somatic cells, *Science.* **318**, 1917-20.
10. Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R. & Young, R. A. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells, *Cell.* **122**, 947-56.
11. Zhao, S., Yuan, Q., Hao, H., Guo, Y., Liu, S., Zhang, Y., Wang, J., Liu, H., Wang, F., Liu, K., Ling, E. A. & Hao, A. (2011) Expression of OCT4 pseudogenes in human tumours: lessons from glioma and breast carcinoma, *J Pathol.* **223**, 672-82.
12. Wang, X. & Dai, J. (2010) Concise review: isoforms of OCT4 contribute to the confusing diversity in stem cell biology, *Stem Cells.* **28**, 885-93.
13. Babaie, Y., Herwig, R., Greber, B., Brink, T. C., Wruck, W., Groth, D., Lehrach, H., Burdon, T. & Adjaye, J. (2007) Analysis of Oct4-Dependent Transcriptional Networks Regulating Self-Renewal and Pluripotency in Human Embryonic Stem Cells, *Stem Cells.* **25**, 500-510.
14. Greber, B., Lehrach, H. & Adjaye, J. (2007) Fibroblast growth factor 2 modulates transforming growth factor beta signaling in mouse embryonic fibroblasts and human ESCs (hESCs) to support hESC self-renewal, *Stem Cells.* **25**, 455-64.
15. Vallier, L., Morgan, A. & Pedersen, R. A. (2005) Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells, *Journal of Cell Science*, 4495-4509.
16. Massague, J. (2008) TGFbeta in Cancer, *Cell.* **134**, 215-30.
17. Singh, A. M., Chappell, J., Trost, R., Lin, L., Wang, T., Tang, J., Matlock, B. K., Weller, K. P., Wu, H., Zhao, S., Jin, P. & Dalton, S. (2013) Cell-cycle control of developmentally regulated transcription factors accounts for heterogeneity in human pluripotent cells, *Stem Cell Reports.* **1**, 532-44.

18. Goodell, M. A., Nguyen, H. & Shroyer, N. (2015) Somatic stem cell heterogeneity: diversity in the blood, skin and intestinal stem cell compartments, *Nat Rev Mol Cell Biol.* **16**, 299-309.
19. Clevers, H. (2015) What is an adult stem cell?, *Science.* **350**, 1319-1320.
20. Krieger, T. & Simons, B. D. (2015) Dynamic stem cell heterogeneity, *Development.* **142**, 1396-406.
21. Snippert, H. J., van der Flier, L. G., Sato, T., van Es, J. H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A. M., van Rheenen, J., Simons, B. D. & Clevers, H. (2010) Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells, *Cell.* **143**, 134-44.
22. Lopez-Garcia, C., Klein, A. M., Simons, B. & Douglas, J. W. (2010) Intestinal Stem Cell Replacement, *Science.* **330**, 822-825.
23. Clevers, H. (2013) The intestinal crypt, a prototype stem cell compartment, *Cell.* **154**, 274-84.
24. Vries, R. G., Huch, M. & Clevers, H. (2010) Stem cells and cancer of the stomach and intestine, *Mol Oncol.* **4**, 373-84.
25. Cancer Genome Atlas, N. (2012) Comprehensive molecular characterization of human colon and rectal cancer, *Nature.* **487**, 330-7.
26. Wood, L. D., Parsons, D. W., Jones, S., Lin, J., Sjoblom, T., Leary, R. J., Shen, D., Boca, S. M., Barber, T., Ptak, J., Silliman, N., Szabo, S., Dezso, Z., Ustyanksky, V., Nikolskaya, T., Nikolsky, Y., Karchin, R., Wilson, P. A., Kaminker, J. S., Zhang, Z., Croshaw, R., Willis, J., Dawson, D., Shipitsin, M., Willson, J. K., Sukumar, S., Polyak, K., Park, B. H., Pethiyagoda, C. L., Pant, P. V., Ballinger, D. G., Sparks, A. B., Hartigan, J., Smith, D. R., Suh, E., Papadopoulos, N., Buckhaults, P., Markowitz, S. D., Parmigiani, G., Kinzler, K. W., Velculescu, V. E. & Vogelstein, B. (2007) The genomic landscapes of human breast and colorectal cancers, *Science.* **318**, 1108-13.
27. Niehrs, C. (2012) The complex world of WNT receptor signalling, *Nat Rev Mol Cell Biol.* **13**, 767-79.
28. Polakis, P. (2012) Wnt signaling in cancer, *Cold Spring Harb Perspect Biol.* **4**.
29. van de Wetering, M., Sancho, E., Verweij, C., de Lau, W., Oving, I., Hurlstone, A., van der Horn, K., Batlle, E., Coudreuse, D., Haramis, A.-P., Tjon-Pon-Fong, M., Moerer, P., Van den Born, M., Soete, G., Pals, S., Eilers, M., Medema, R. & Clevers, H. (2002) The B-Catenin/TCF-4 Complex Imposes a Crypt Progenitor Phenotype on Colorectal Cancer Cells, *Cell.* **111**, 241-250.
30. Yu, J. & Virshup, D. M. (2014) Updating the Wnt pathways, *Biosci Rep.* **34**.
31. Barker, N., van Es, J. H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P. J. & Clevers, H. (2007) Identification of stem cells in small intestine and colon by marker gene Lgr5, *Nature.* **449**, 1003-7.
32. Lander, A. D., Kimble, J., Clevers, H., Fuchs, E., Montarras, D., Buckingham, M., Calof, A. L., Trumpp, A. & Oskarsson, T. (2012) What does the concept of the stem cell niche really mean today?, *BMC Biology.* **10**.
33. Ritsma, L., Ellenbroek, S. I. J., Zomer, A., Snippert, H. J., de Sauvage, F. J., Simons, B. D., Clevers, H. & van Rheenen, J. (2014) Intestinal crypt homeostasis revealed at single-stem-cell level by in vivo live imaging, *Nature.* **507**, 362-365.
34. Sato, T., Vries, R. G., Snippert, H. J., van de Wetering, M., Barker, N., Stange, D. E., van Es, J. H., Abo, A., Kujala, P., Peters, P. J. & Clevers, H. (2009) Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche, *Nature.* **459**, 262-5.
35. Yan, K. S., Chia, L. A., Li, X., Ootani, A., Su, J., Lee, J. Y., Su, N., Luo, Y., Heilshorn, S. C., Amieva, M. R., Sangiorgi, E., Capecchi, M. R. & Kuo, C. J. (2012) The intestinal stem cell markers Bmi1 and Lgr5 identify two functionally distinct populations, *PNAS.* **109**, 466-471.
36. Tian, H., Biehs, B., Warming, S., Leong, K. G., Rangell, L., Klein, O. D. & de Sauvage, F. J. (2011) A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable, *Nature.* **478**, 255-9.

37. Buczacki, S. J., Zecchini, H. I., Nicholson, A. M., Russell, R., Vermeulen, L., Kemp, R. & Winton, D. J. (2013) Intestinal label-retaining cells are secretory precursors expressing Lgr5, *Nature.* **495**, 65-9.
38. Tetteh, P. W., Basak, O., Farin, H. F., Wiebrands, K., Kretzschmar, K., Begthel, H., van den Born, M., Korving, J., de Sauvage, F., van Es, J. H., van Oudenaarden, A. & Clevers, H. (2016) Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters, *Cell Stem Cell.* **18**, 203-13.
39. Sato, T., Stange, D. E., Ferrante, M., Vries, R. G., Van Es, J. H., Van den Brink, S., Van Houdt, W. J., Pronk, A., Van Gorp, J., Siersema, P. D. & Clevers, H. (2011) Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium, *Gastroenterology.* **141**, 1762-72.
40. Humphries, A. & Wright, N. A. (2008) Colonic crypt organization and tumorigenesis, *Nat Rev Cancer.* **8**, 415-24.
41. Kosinski, C., Li, V. S., Chan, A. S., Zhang, J., Ho, C., Tsui, W. Y., Chan, T. L., Mifflin, R. C., Powell, D. W., Yuen, S. T., Leung, S. Y. & Chen, X. (2007) Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors, *Proc Natl Acad Sci U S A.* **104**, 15418-23.
42. Danopoulos, S., Schlieve, C. R., Grikscheit, T. C. & Al Alam, D. (2017) Fibroblast Growth Factors in the Gastrointestinal Tract: Twists and Turns, *Developmental Dynamics.* **246**, 344-352.
43. Kanazawa, S., Tsunoda, T., Onuma, E., Majima, T., Kagiyama, M. & Kikuchi, K. (2001) VEGF, Basic-FGF, and TGF-b in Crohn's Disease and Ulcerative Colitis: A Novel Mechanism of Chronic Intestinal Inflammation, *Am Coll of Gastroenterology.* **96**, 822-828.
44. Hanahan, D. & Weinberg, R. A. (2000) The Hallmarks of Cancer, *Cell.* **100**, 57-70.
45. Hanahan, D. & Weinberg, R. A. (2011) Hallmarks of cancer: the next generation, *Cell.* **144**, 646-74.
46. Egeblad, M., Nakasone, E. S. & Werb, Z. (2010) Tumors as organs: complex tissues that interface with the entire organism, *Dev Cell.* **18**, 884-901.
47. Magee, J. A., Piskounova, E. & Morrison, S. J. (2012) Cancer stem cells: impact, heterogeneity, and uncertainty, *Cancer Cell.* **21**, 283-96.
48. Kleinsmith, L. J. & Pierce, G. B. (1964) Multipotentiality of Single Embryonal Carcinoma Cells, *Cancer Research.* **24**, 1544.
49. O'Brien, C. A., Pollett, A., Gallinger, S. & Dick, J. E. (2007) A human colon cancer cell capable of initiating tumour growth in immunodeficient mice, *Nature.* **445**, 106-10.
50. Ricci-Vitiani, L., Lombardi, D. G., Pilozzi, E., Biffoni, M., Todaro, M., Peschle, C. & De Maria, R. (2007) Identification and expansion of human colon-cancer-initiating cells, *Nature.* **445**, 111-5.
51. Clevers, H. (2011) The cancer stem cell: premises, promises and challenges, *Nat Med.* **17**, 313-9.
52. Schatton, T., Murphy, G. F., Frank, N. Y., Yamaura, K., Waaga-Gasser, A. M., Gasser, M., Zhan, Q., Jordan, S., Duncan, L. M., Weishaupt, C., Fuhlbrigge, R. C., Kupper, T. S., Sayegh, M. H. & Frank, M. H. (2008) Identification of cells initiating human melanomas, *Nature.* **451**, 345.
53. Quintana, E., Shackleton, M., Sabel, M. S., Fullen, D. R., Johnson, T. M. & Morrison, S. J. (2008) Efficient tumour formation by single human melanoma cells, *Nature.* **456**, 593-8.
54. Batlle, E. & Clevers, H. (2017) Cancer stem cells revisited, *Nat Med.* **23**, 1124-1134.
55. Barker, N., Ridgway, R. A., van Es, J. H., van de Wetering, M., Begthel, H., van den Born, M., Danenberg, E., Clarke, A. R., Sansom, O. J. & Clevers, H. (2009) Crypt stem cells as the cells-of-origin of intestinal cancer, *Nature.* **457**, 608-11.
56. Schepers, A. G., Snippert, H. J., Stange, D. E., van den Born, M., van Es, J. H., van de Wetering, M. & Clevers, H. (2012) Lineage Tracing Reveals Lgr5⁺ Stem Cell Activity in Mouse Intestinal Adenomas, *Science.* **337**, 730.

57. Kozar, S., Morrissey, E., Nicholson, A. M., van der Heijden, M., Zecchini, H. I., Kemp, R., Tavare, S., Vermeulen, L. & Winton, D. J. (2013) Continuous clonal labeling reveals small numbers of functional stem cells in intestinal crypts and adenomas, *Cell Stem Cell.* **13**, 626-33.
58. van de Wetering, M., Francies, H. E., Francis, J. M., Bounova, G., Iorio, F., Pronk, A., van Houdt, W., van Gorp, J., Taylor-Weiner, A., Kester, L., McLaren-Douglas, A., Blokker, J., Jaksani, S., Bartfeld, S., Volckman, R., van Sluis, P., Li, V. S., Seepo, S., Sekhar Pedamallu, C., Cibulskis, K., Carter, S. L., McKenna, A., Lawrence, M. S., Lichtenstein, L., Stewart, C., Koster, J., Versteeg, R., van Oudenaarden, A., Saez-Rodriguez, J., Vries, R. G., Getz, G., Wessels, L., Stratton, M. R., McDermott, U., Meyerson, M., Garnett, M. J. & Clevers, H. (2015) Prospective derivation of a living organoid biobank of colorectal cancer patients, *Cell.* **161**, 933-45.
59. Fujii, M., Shimokawa, M., Date, S., Takano, A., Matano, M., Nanki, K., Ohta, Y., Toshimitsu, K., Nakazato, Y., Kawasaki, K., Uraoka, T., Watanabe, T., Kanai, T. & Sato, T. (2016) A Colorectal Tumor Organoid Library Demonstrates Progressive Loss of Niche Factor Requirements during Tumorigenesis, *Cell Stem Cell.* **18**, 827-38.
60. Drost, J. & Clevers, H. (2018) Organoids in cancer research, *Nat Rev Cancer*.
61. Cortina, C., Turon, G., Stork, D., Hernando-Momblona, X., Sevillano, M., Aguilera, M., Tosi, S., Merlos-Suarez, A., Stephan-Otto Attolini, C., Sancho, E. & Batlle, E. (2017) A genome editing approach to study cancer stem cells in human tumors, *EMBO Mol Med.* **9**, 869-879.
62. Drost, J., van Jaarsveld, R. H., Ponsioen, B., Zimberlin, C., van Boxtel, R., Buijs, A., Sachs, N., Overmeer, R. M., Offerhaus, G. J., Begthel, H., Korving, J., van de Wetering, M., Schwank, G., Logtenberg, M., Cuppen, E., Snippert, H. J., Medema, J. P., Kops, G. J. & Clevers, H. (2015) Sequential cancer mutations in cultured human intestinal stem cells, *Nature.* **521**, 43-7.
63. Matano, M., Date, S., Shimokawa, M., Takano, A., Fujii, M., Ohta, Y., Watanabe, T., Kanai, T. & Sato, T. (2015) Modeling colorectal cancer using CRISPR-Cas9-mediated engineering of human intestinal organoids, *Nat Med.* **21**, 256-62.
64. Shimokawa, M., Ohta, Y., Nishikori, S., Matano, M., Takano, A., Fujii, M., Date, S., Sugimoto, S., Kanai, T. & Sato, T. (2017) Visualization and targeting of LGR5(+) human colon cancer stem cells, *Nature.* **545**, 187-192.
65. Jaks, V., Kasper, M. & Toftgard, R. (2010) The hair follicle-a stem cell zoo, *Exp Cell Res.* **316**, 1422-8.
66. Luo, M. & Fu, L. (2014) Redundant kinase activation and resistance of EGFR-tyrosine kinase inhibitors, *Am J Cancer Res.* **4**, 608-628.
67. Hallinan, N., Finn, S., Cuffe, S., Rafee, S., O'Byrne, K. & Gately, K. (2016) Targeting the fibroblast growth factor receptor family in cancer, *Cancer Treat Rev.* **46**, 51-62.
68. Ware, K. E., Marshall, M. E., Heasley, L. R., Marek, L., Hinz, T. K., Hercule, P., Helfrich, B. A., Doebele, R. C. & Heasley, L. E. (2010) Rapidly acquired resistance to EGFR tyrosine kinase inhibitors in NSCLC cell lines through de-repression of FGFR2 and FGFR3 expression, *PLoS One.* **5**, e14117.
69. Wagner, A., Regev, A. & Yosef, N. (2016) Revealing the vectors of cellular identity with single-cell genomics, *Nat Biotechnol.* **34**, 1145-1160.

Curriculum Vitae

Education

10/14 - 12/18	**PhD-Thesis at the Institute for Stem Cell Research and Regenerative Medicine, Heinrich-Heine-University Düsseldorf** Prof. Dr. James Adjaye Titel: Genetic Heterogeneity of Cancer Stem Cells
10/12 - 09/14	**Master of Science in Molecular Medicine** University of Erlangen-Nuremberg, Germany Thesis: The role of iPS-related transcription factors in glioma PD Dr. med N.E. Savaskan
10/09 - 07/12	**Bachelor of Science in Molecular Medicine** University of Erlangen-Nuremberg, Germany Thesis: The expression pattern of Sox13 in embryonic mouse development of the central and peripheral nervous system PD. Dr. C. C. Stolt, Prof. Dr. M. Wegner
11/06 - 10/09	**Vocational School of Physiotherapy** Heinrich-Heine-University Düsseldorf, Germany

Publications

Otte, J., Dizdar, L., Behrens, B., Wruck, W., Stoecklein, N.H. and Adjaye, J. (2018). FGF-driven self-renewal in primary colon cancer organoids. Manuscript submitted.

Otte, J., Wruck, W. and Adjaye, J. (2017). New insights into human primordial germ cells and early embryonic development from single-cell analysis. FEBS Lett 591, 2226-2240.

Surowy, H.M., Giesen, A.K., **Otte, J**., […] and Redler, S. (2018). Gene expression profiling in aggressive digital papillary adenocarcinoma sheds light on the architecture of a rare sweat gland carcinoma. Manuscript under revision.

Matz, P., Spitzhorn, L.S., **Otte, J**., […] and Adjaye, J. (2017). 4.10 - Use of Stem Cells in Toxicology A2 - Chackalamannil, Samuel. In Comprehensive Medicinal Chemistry III (Rotella, D. and Ward, S.E.), pp. 177-194. Elsevier, Oxford.

Mertes, F., Lichtner, B., Kuhl, H., Blattner, M., **Otte, J**., Wruck, W., […] and Adjaye, J. (2015). Combined ultra-low input mRNA and whole-genome sequencing of human embryonic stem cells. BMC Genomics*, 16*, 925. doi:10.1186/s12864-015-2025-z

Hafner, S., Zolk, K., Radaelli, F., **Otte, J**., Rabenstein, T. and Zolk, O. (2015). Water infusion versus air insufflation for colonoscopy. Cochrane Database Syst Rev, CD009863.

Grants & Awards

10/14 - 12/17	**Competitive PhD scholarship** of the Düsseldorf School of Oncology (DSO)
04/17	**Best Talk Award** at the DSO Retreat
03/16	**Best Poster Award** at the DSO Retreat

Conference Presentations

05/18	Oral presentation at the internal retreat of the Stem Cell Network North Rhine-Westphalia
11/17	Oral presentation and poster presentation at the retreat of the Disseminating Cancer Cell (DCC)-Net Düsseldorf
05/17	Poster presentation at the International Meeting of the Stem Cell Network North Rhine-Westphalia
04/17	Oral presentation at the 5th DSO-Retreat
09/16	Poster presentation at the 4th International Conference of the German Stem Cell Network (GSCN)
03/16	Poster presentation at the 4th DSO-Retreat
09/15	Poster presentation at the 3rd International GSCN Conference
04/15	Oral presentation and Poster presentation at the 3rd DSO-Retreat

Professional Skills

Cell Culture and Molecular Biology	Tumor tissue dissociation, Primary cell culture, 3D organoid models, RNA expression analysis (qRT-PCR, Microarray transcriptome, SMARTer RNA amplification), Protein analysis (Western Blot, fluorescent microscopy), Gene knock-down (siRNA/ microRNA), Cloning, Stem cell culture
Animal experiments	**B-FELASA certified,** tumor implantation in rat, mouse embryo preparation, hippocampus and dura mater preparation
Transferable Skills	Member of the *Interdisciplinary Graduate and Research Academy Düsseldorf* (iGRAD); participation in workshops • **Presenting Science** – How to own the stage on international conferences • **Discussing in Science** – How to succeed in discussions, debates and meetings • **Career Planning in Business** – How to shape your future • **Time Management** – Get more done with less effort • **Good Scientific Practice for Doctoral Researchers**
Clinical experience	**Working experience as Physiotherapist**
04/07 - 10/09	at the University Hospital Düsseldorf, Germany In particular: Neurosurgery, Neurology, Paediatrics
08/06 - 10/06	**Practical nursing internship**, City Hospital Salzkotten, Germany
Languages	German (native) English (fluent) Italian (basic)

Danksagung

An dieser Stelle möchte ich mich bei allen Personen bedanken, die zum Gelingen dieser Arbeit beigetragen haben.

Zu allererst möchte ich mich bei meinem Betreuer Herrn Prof. Dr. James Adjaye bedanken. Er hat stets meine Stärken und Schwächen erkannt und mein Weiterkommen gefördert. Mit großem Freiraum konnte ich immer meine eigenen Ideen entwickeln.

Des Weiteren bedanke ich mich bei meinem Mentor Herrn Prof. Dr. Nikolas Stoecklein. In vielen spannenden wissenschaftlichen Diskussionen konnte er stets meine Begeisterung neu wecken und häufig mit neuen Ansätzen und Ideen meine Arbeit nach vorne bringen.

Ich bedanke mich auch bei Herrn Prof. Dr. Constantin Czekelius für die Übernahme des Zweitgutachtens. Mit seiner freundlichen und unkomplizierten Art zeigte er als fachfremder Betreuer stets Interesse an meiner Arbeit.

Der Düsseldorf School of Oncology danke ich für die Finanzierung meines Stipendiums. Ebenso möchte ich mich für das herausragende Förderprogramm bedanken. Von den Vorlesungen, Seminaren und Retreats sowie von den iGRAD Kursen konnte ich sehr profitieren und mich nicht nur in wissenschaftlichen Bereichen erheblich weiterentwickeln.

Ein großer Dank für die tolle Zeit, die ich im Institut für Stammzellforschung hatte, geht an das gesamte ISRM-Team. Für die große Hilfsbereitschaft in jeder Phase meiner Arbeit bedanke ich mich sehr. Dank Euch war der Laboralltag stets eine Freude.

Auch bei dem Team der chirurgischen Forschung möchte ich mich herzlich bedanken. Meinen besonderen Dank möchte ich Bianca ausdrücken, die mich ab dem ersten Tag an die Hand genommen hat und bei vielen Herausforderungen als erfahrener Coach zur Stelle war.

Meiner Familie sowie meinen Freunden danke ich für die Unterstützung auf diesem langen Weg. Ihr konntet immer die schweren Zeiten leichter und die guten Zeiten intensiver machen. Vielen Dank!

Eidesstattliche Versicherung

Ich, Jörg Otte, versichere an Eides statt, dass die vorliegende Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Düsseldorf, der 10. Dezember 2018

_____

(Jörg Otte)