Aus dem Institut für Stammzellforschung

des Helmholtz Zentrums München

Direktor: Prof. Dr. Magdalena Götz

A GATA/TFAP2 transcription regulatory network couples human pluripotent stem cell differentiation to trophectoderm with repression of pluripotency

Dissertation

zum Erwerb des Doktorgrades der Naturwissenschaften an der Medizinischen Fakultät

der Ludwig-Maximilians-Universität München

vorgelegt von

Christian Krendl

aus

Waidhofen/Ybbs

2017

Mit Genehmigung der Medizinischen Fakultät der

Universität München

Betreuerin: Prof. Dr. Magdalena Götz

Zweitgutachter: Prof. Dr. Ralph A.W. Rupp

Dekan: Prof. Dr. med. dent. Reinhard Hickel

Tag der mündlichen Prüfung: 25.01.2018

Table of Content

I. Summary

Correct development of the placenta is crucial for the growth of the embryo and the health of both mother and child. Using mainly the mouse as a model system, a cohort of transcription factors (TFs) have been implicated in the specification of trophectoderm lineage progenitors, which gives rise to the placenta. These progenitors are the first differentiated cells that emerge in the embryo, and therefore their specification involves mechanisms that dissolve pluripotency. Importantly, the network configuration of these transcription factors has not been entirely clarified, and so is the degree of conservation in the human. Moreover, it is not well understood how pluripotency is terminated during the commitment of trophectoderm progenitors, and how this is regulated on the epigenetic level.

To address these questions I employed an *in vitro* differentiation system that is based on human embryonic stem cells (ESCs), and treatment with BMP4. This stimulates the differentiation of the cells into a trophoblast fate. To analyze the underlying mechanisms of this specification, I first optimized a purification modality for investigating the intrinsic properties of these progenitor cells. This utilizes an antibody that is specific to APA (CD249, Ly-51, ENPEP), a surface marker of trophoblast progenitors, and cell purification via fluorescence activated cell sorting (FACS). Based on this, I employed global transcriptomic and epigenomic approaches as well as functional approaches to analyze the underlying mechanisms of trophectoderm specification and pluripotency shutdown.

Using this approach I first showed the close homology of the *in vitro* derived APA+ trophoblast progenitors to human trophectoderm progenitors in blastocysts. I then went on to discover a gene regulatory network that governs the differentiation of the human ESC-derived trophoblast progenitors. It consists of the four transcription factors GATA2, GATA3, TFAP2A (AP2-α) and TFAP2C (AP2-γ), which I collectively named the TrophEctoderm four, or in short – the *TEtra*. I found that the TEtra regulate in concert both repression of pluripotency and activation of trophoblast specific genes. I also discovered genome wide distribution of the TEtra network, proven some of its main

components in functional assays, and characterized important epigenetic features that regulate the specification of human trophectoderm progenitors.

The implications of my work are broad and include an important foundation for understanding in a great detail the mechanisms underlying human placental development and numerous pathologies related to placental dysfunction which impact the health of mother and child. Furthermore, I point out important features of regulation of pluripotency dissolvent and TE specification that could be unique to the human. Finally, I discover features of epigenetic turnover that are pertinent for understanding processes undelaying gene regulation in development.

I. Zusammenfassung

Während der Schwangerschaft ist die korrekte Entwicklung der Plazenta essenziell für das Wachstum des Embryos und die Gesundheit von Mutter und Kind. Versuche an Mäusen ermöglichten die Identifikation eines Netzwerks an Transkriptionsfaktoren (TF), das der Entstehung von Trophectoderm- (Vorläuferzellen der Plazenta) und spezialisierten Plazenta-Zellen in Mäusen zugrunde liegt. Da Trophectoderm-Zellen die ersten differenzierten Zellen des Embryos ausmachen geht ihre Entwicklung Hand in Hand mit der Abschaltung von Pluripotenz-Genen. Die Charakterisierung dieses Netzwerks an TF ist jedoch noch nicht komplett ausgereift und es ist nicht klar, ob dieses auch im Menschen für die Entwicklung der Plazenta und das Ende der Pluripotenz wichtig sind und welchen Einfluss Epigenetik auf diese Prozesse hat.

Um diese Fragen zu beantworten habe ich ein *in vitro* Differenzierungsprotokoll, bei dem ich humane Embryonale Stammzellen (ESZ) mit dem Morphogen BMP4 differenziere, benutzt. Um die Homogenität der Population von Trophoblast-Vorläuferzellen sicherzustellen habe ich die Technik fluorescence activated cell sorting (FACS) benutzt um Vorläuferzellen entsprechend der Expression des Oberflächenantigens APA (CD249, Ly-51, ENPEP) anzureichern. Diese Zellen habe ich dann für die Analyse von Transkriptom und Epigenom als auch für funktionelle Assays benutzt um die grundlegenden Mechanismen der humanen Plazentaentwicklung zu entziffern.

Hierdurch konnte ich zeigen, dass diese *in vitro* differenzierten Zellen zum Großteil homolog sind zu *in vivo* Trophectoderm-Zellen. Außerdem konnte ich ein Netzwerk von TF identifizieren, das der Entwicklung von humanen ESZ in frühe Vorläuferzellen der Plazenta zugrunde liegt. Dieses besteht aus den vier TF GATA2, GATA3, TFAP2A (AP2-α) und TFAP2C (AP2-γ), die gleichzeitig Pluripotenz-Gene ab- und Trophoblast–Gene anschalten können. Schließlich habe ich die Verteilung dieser 4 TF über das gesamte Genom während dieser Differenzierung untersucht, durch Manipulation des Genoms gewisse Bestandteile dieses Netzwerks auf deren Bedeutung geprüft und zentrale

epigenetische Merkmale charakterisiert, welche die Entwicklung von humanen Plazenta-Vorläuferzellen regulieren.

Die Ergebnisse dieser Arbeit zeigen die detaillierten Mechanismen, die für die frühe Plazentaentwicklung beim Menschen verantwortlich sind.

II. Introduction

Trophectoderm and placental development

Generation of defined cell types in the preimplantation embryo

The fertilization of the mammalian oocyte leads to the formation of an embryo consisting of totipotent cells called the morula. During the partitioning of the morula and the formation of the blastocyst distinct cell types can be identified for the first time during development. This includes a group of pluripotent cells named the inner cell mass (ICM), and multipotent cells of the trophectoderm (TE). The ICM further gives rise to the primitive endoderm (PE) and the epiblast that forms the soma and the germ line (Figure II-1). At this stage the blastocyst implants in the uterus. The gene networks underlying the cell-commitment processes over time have been studied extensively in the mouse using chiefly classical approaches of gene gain– and loss–of–function. It is not well understood to what degree the molecular processes that were elaborated in the mouse are conserved in our own development. Because it is not anticipated that human embryos will become available for analysis by gene manipulation due to ethical restrictions, this gap can only be filled using *in vitro* cell based models.

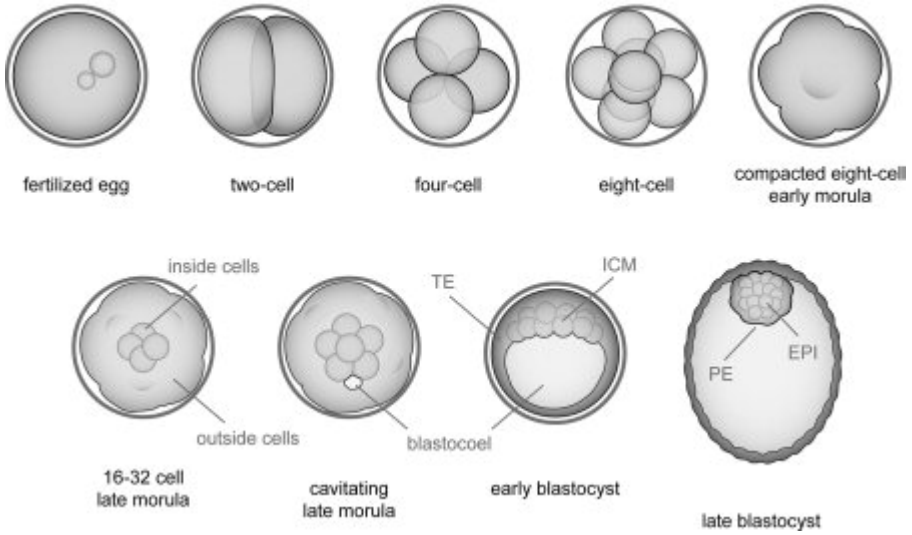

Figure II-1: *Mouse embryo development up to the blastocyst stage*
The fertilized egg develops via cell divisions into 2-, 4-, 8-cell and morula stage embryo. It is not until the early blastocyst stage that defined cell types emerge, namely first, the trophectoderm (TE) and the inner cell mass (ICM), and subsequently the epiblast (EPI) and the primitive endoderm (PE) (Yamanaka et al., 2006).

The first specification

Depletion or rearrangement of totipotent cells in mouse preimplantation embryo have shown that there is plasticity in the cells that can compensate for these manipulations until the 16-32-cell stage (Bedzhov et al., 2014). This led to the hypothesis that it is the position of the cells between the 8- and 32-cell stage that determines the lineage outcome: inner and outer cells into the ICM and the TE, respectively (Figure II-2A). This view has been challenged by depletion of the cell polarity genes *Par3* and *aPKC* that skewed lineage ratio in favor of ICM cells, indicating that it is not only the position of the cells that determine the lineage outcome (Plusa et al., 2005). It has also been proposed that there exists a developmental bias in the blastomeres of the 2-cell stage (Gardner, 2001; Piotrowska et al., 2001; Piotrowska and Zernicka-Goetz, 2001). Although time-lapse microscopy observations initially rejected the view that the two-cell blastomeres are distinct from each other with regard to the ICM and TE outcome (Motosugi et al., 2005), more recently experiments using genetically labeled cells have emphasized the existence of a developmental bias (Tabansky et al., 2013). Another model reconciles these observations by explaining that the differences in cell polarity predispose cells in 2-, 4- and 8-cell stage embryos to give rise to the ICM or TE (Johnson and Ziomek, 1981). This relies on observations that at these stages the cells have apical-basal polarity, and depending on the plane of division, the cells either give rise to one polar TE and one apolar ICM cell or to two polar TE cells. This model is in agreement with the increased production of ICM cells upon the loss of cell polarity (Sutherland et al., 1990) (Figure II-2B).

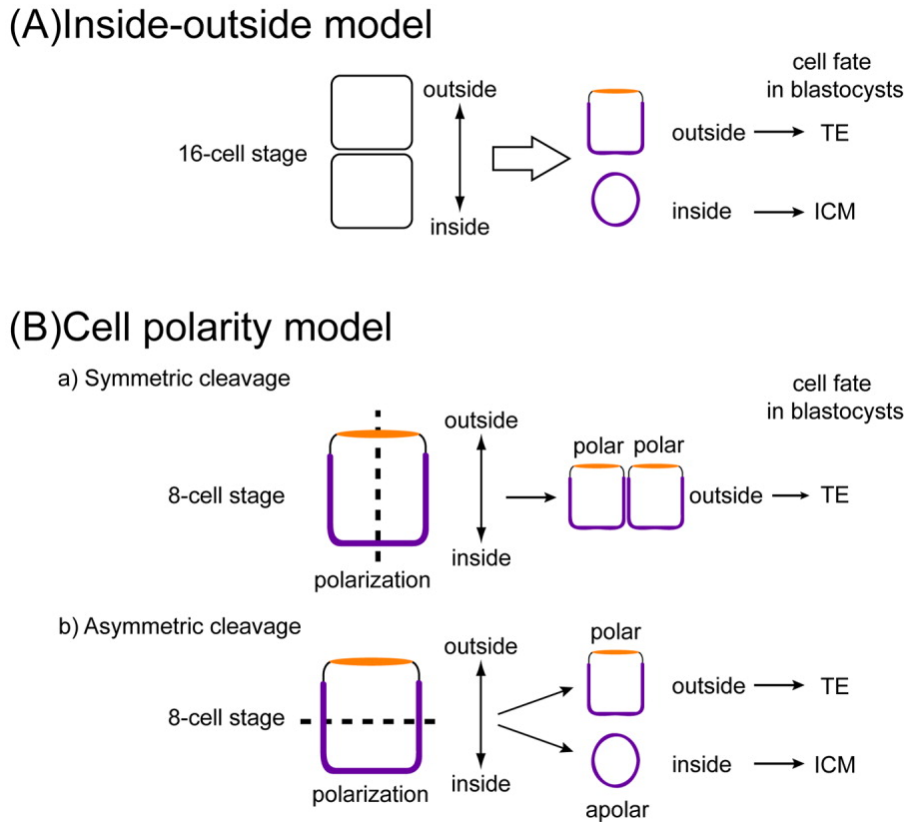

Figure II-2: *Models of TE/ICM specification*
Primary models addressing the regulation of TE and ICM cell location. (A) The inside-outside model proposes that the position of the cells determines the outcome. In contrast, the cell polarity model (B) suggests that the mode of cleavage determines the outcome: if cells divide symmetrically relative to their axis, two outside polar emerge (Ba), and if cells divide asymmetrically this leads to the emergence of one polar and one apolar cell, outside and inside, respectively (Bb) (Yamanaka et al., 2006).

The gene network governing the bifurcation of the ICM and the TE has been extensively studied in the mouse. The most prominent transcriptional regulators implicated in the formation of the ICM are *Oct4* (Nichols et al., 1998), *Nanog* (Mitsui et al., 2003a) and *Sox2* (Avilion et al., 2003), whereas *Cdx2* (Niwa et al., 2005; Strumpf et al., 2005), *Gata3* (Ralston et al., 2010) and *Eomes* (Russ et al., 2000) are important for the TE. The latter are restricted to cells of the 8-16 cell stage embryo that are destined to become TE cells. *Tead4* was shown to regulate *Cdx2*, and its depletion leads to the earliest embryonic lethality due to TE defect (Yagi et al., 2007; Nishioka et al., 2008). Surprisingly, Tead4 is not only expressed in the outer cells that are designated to become TE, but also in inner cells that become the ICM (Nishioka et al., 2008). The phenotypic specificity in TE cells is explained by the nuclear accumulation of Yap and Taz, the two transcriptional co-

activators of Tead4, which are also vital for TE development (Nishioka et al., 2009). Both factors are found in the nucleus of the TE-destined outer cells and in the cytoplasm of the inner cells. What controls Yap and Taz localization in general and specifically in TE cells are kinases of the Hippo pathway: the asymmetry in cell-cell contact in outer cells results in the accumulation of the polarity proteins Par6 and aPKC (Bedzhov et al., 2014), which inhibit the Hippo pathway. As a result unphosphorylated Yap and Taz translocate to the nucleus. In contrast, in ICM cells the Hippo pathway kinases, *Lats1* and *Lats2* are active; they phosphorylate Yap and Taz, leading to cytoplasmic localization and thereby prevention of Tead4-mediated transcription (Nishioka et al., 2009) (Figure II-3). Cellular polarity also governs TE specification by regulation of *Cdx2* mRNA localization to the apical side. This generates inside cells lacking *Cdx2* (Skamagki et al., 2013).

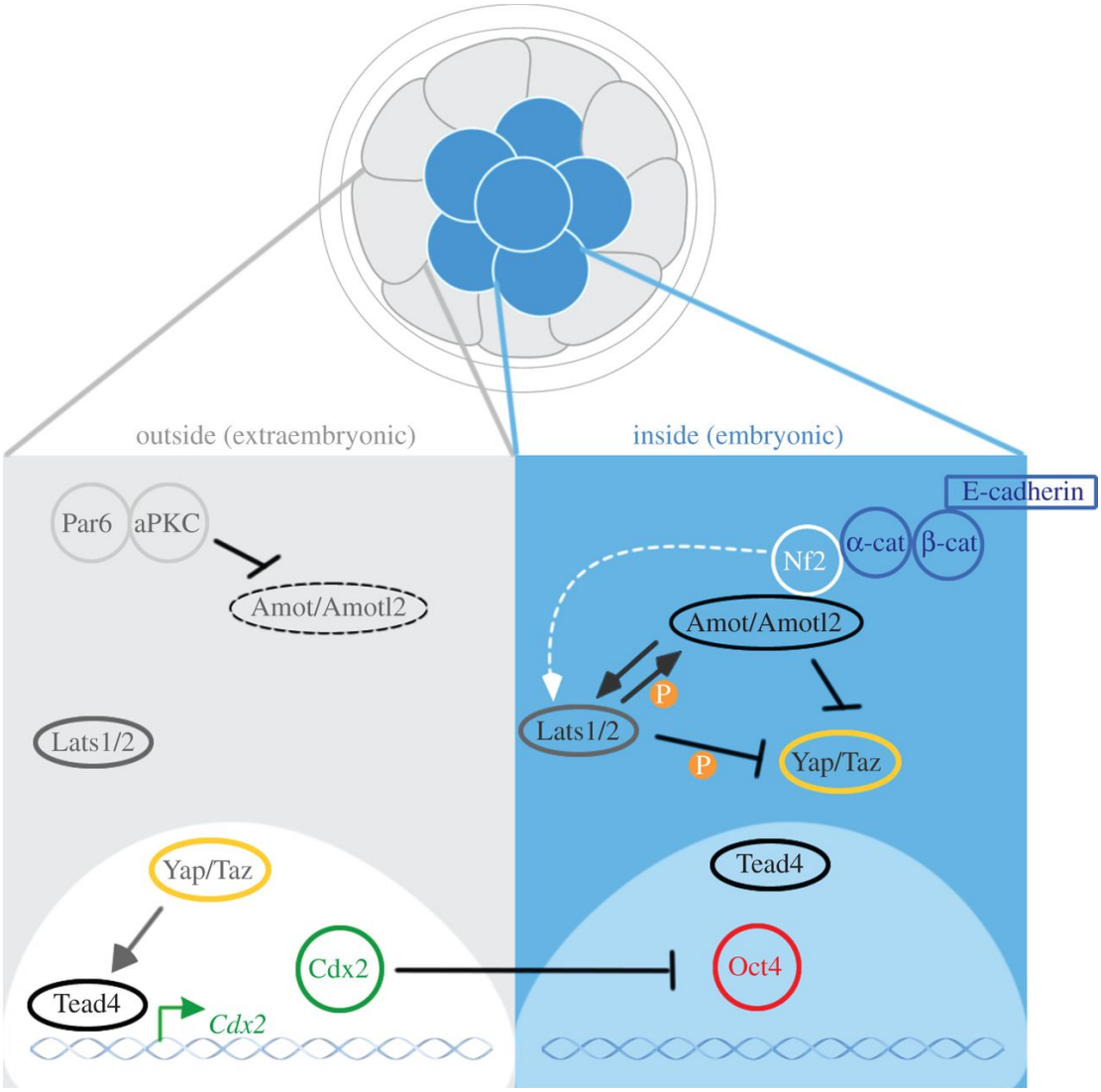

Figure II-3: *Factors involved in the first cell specification in mouse*
Cells of the morula exhibit different features. The outside cells have asymmetric cell-cell contacts and form the TE, whereas the inside cells have symmetrical cell-cell contact and form cells of the ICM. In the mouse it was suggested that in outside cells Amot and Lats1/2, kinases of the Hippo pathway are inactive and Yap/Taz can go to the nucleus of the cell. This leads to activation of Tead4 and Cdx2, which strengthens the TE identity of outside cells. On inside, ICM, cells activity of Amot and Lats1/2 leads to repression of Yap/Taz and Tead4 is not switched on in the nucleus. Therefore, the pluripotency program is maintained by expression of Oct4 (Bedzhov et al., 2014).

Nevertheless, Cdx2 is not vital or alone promoting TE development, because expression of Cdx2 but not Oct4 is variable in the cells of the morula (Dietrich and Hiiragi, 2007), and the initial TE separation is not affected by depletion of maternal and zygotic *Cdx2* (Wu et al., 2010).

Second specification and implantation

The arrangement of the mouse pluripotent epiblast, which is surrounded by a layer of PE cells, is mediated by selective apoptosis and cell migration of epiblast and PE cells. Time-laps microscopy and tracking of ICM cells from mouse early to late blastocyst showed that the precursors of epiblast and PE are intermingled within the ICM (Chazaud et al., 2006; Plusa et al., 2008; Meilhac et al., 2009). With regard to the mechanism, cells of the PE and the epiblast were shown to become established by unique transcriptional programs that can already be detected at the early blastocyst (Chazaud et al., 2006; Kurimoto et al., 2006): *Nanog* (Chambers et al., 2003; Mitsui et al., 2003a) and *Sox2* for the epiblast (Avilion et al., 2003), and endodermal genes such as *Gata6*, *Gata4* (Koutsourakis et al., 1999) or *Sox17* (Niakan et al., 2010) for the PE. As the mechanisms governing the specification of epiblast and PE cells are not the focus of my work, they are not covered here but can be found elsewhere (Bedzhov et al., 2014).

These events take place concomitantly with the translocation of the blastocyst to the uterus where it implants in the endometrium by hatching, a process that involves breakdown on the zona pellucida membrane. The cells of the epiblast further give rise to the fetus, the TE contributes to the placenta and the PE to the yolk sac (Gardner and Johnson, 1973; Papaioannou et al., 1975; Gardner and Rossant, 1979; Gardner, 1985).

Human and mouse preimplantation embryos compared

While mouse and human preimplantation blastocysts are morphologically similar, there are some important differences: mouse and human blastocyst form around E day 3-3.5 and 5, respectively, and implantation takes place at E day 4-4.5 and 7-9, respectively (with another round of cell division in the human) (Hertig et al., 1959; Finn and McLaren, 1967; Norwitz et al., 2001; Cockburn and Rossant, 2010). Known molecular differences include timing of *CDX2* expression, the key TE TF, which is detected in the human only after the formation of the blastocyst (Niakan and Eggan, 2013). Conversely, GATA3 expression is more pronounced in human TE, probably it serves to compensate for the late CDX2 expression (Deglincerti et al., 2016). Differences in expression patterns also exist between human and mouse blastocysts. For example, OCT4 is ubiquitously expressed in all cells of the human blastocyst at E day 5-7, whereas in the mouse it is restricted to the ICM. Moreover, in contrast to the mouse, PE and ICM cells are not sorted at E day 6 in the human (O'Leary et al., 2012; Roode et al., 2012; Niakan and Eggan, 2013; Deglincerti et al., 2016).

Finally, interestingly TE cells that are isolated from the human blastocyst can still give rise to pluripotent cells *in vitro* (De Paepe et al., 2013), which indicates that human blastocyst cells are more plastic than the cells of the murine blastocyst.

Chorion and placenta development

After implantation, the blastocyst committed TE cells develop into an extraembryonic tissue named the chorion, the primary precursor of the placenta. The placenta forms the interface between the mother and the fetus, which is important for the exchange of nutrients, waste products and gases, and for the prevention of immune response towards the fetus. The placenta is also regulating the hormone regimens that are necessary for fetal growth (Rossant and Cross, 2001). Perturbations of chorion development and placental dysfunction are implicated in fetal growth retardation and abnormalities that can also impact the health of the mother.

Mouse placental development

The attachment of the blastocyst to the maternal endometrium is mediated by mural TE, the TE that surrounds the blastocyst cavity opposite of the ICM. The mural TE cells then undergo DNA endoreplication without cell division, a process that leads to the formation of trophoblast giant cells. These cells remodel the vascular system and promote angiogenic processes which are crucial for generating the vascular infrastructure that regulate nutrient, gas and waste product exchange between the fetus and the maternal tissues (Rossant and Cross, 2001; Simmons et al., 2007). Polar TE, which reside adjacent to the ICM, proliferate and generate structures named the extraembryonic ectoderm (ExE) and the ectoplacental cone (Figure II-4).

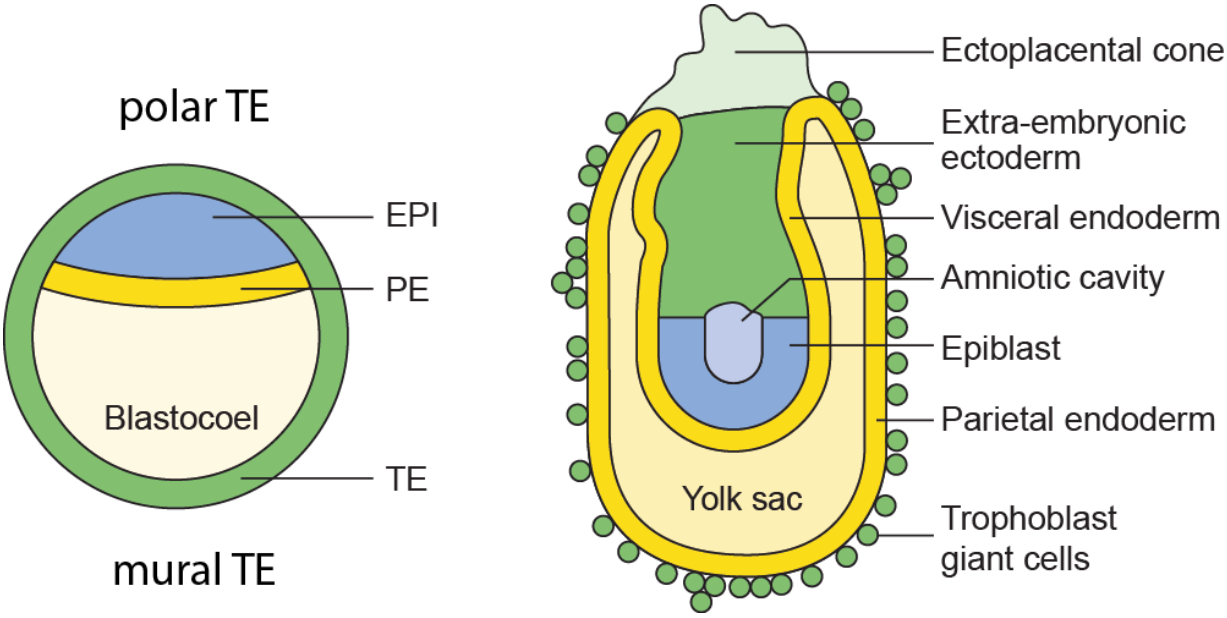

Figure II-4: *Morphology of the late mouse blastocyst*
Mouse TE progenitors form the extraembryonic ectoderm, the ectoplacental cone and trophoblast giant cells. The PE gives rise to the yolk sac. Adapted from (Rossant, 2015).

The cone grows by expansion of a population of trophoblast progenitor cells in the ExE region. This relies on several TFs, like *Elf5*, a TF that regulates the expression of *Cdx2* and *Eomes* (Donnison et al., 2005; Ng et al., 2008) or *Ets2* (Georgiades and Rossant, 2006; Polydorou and Georgiades, 2013) (see section "The transcriptional network of TE development").

Around E day 8.5 the extra embryonic allantois tissue, which is mesoderm derived, makes contact with the chorionic ExE. This perpetuates the folding of the ExE which

creates villous branching structures named the placental labyrinth. This tissue consists of several layers of labyrinthine trophoblasts and it is structurally supported by spongiotrophoblast cells that are derived from the ectoplacental cone. Fused cytotrophoblasts cells named syncytiotrophoblasts form at the tip of the elongating branches of the villous tree (Cross et al., 2003). To support the transport of nutrients and the exchange of gas and waste products, sinuses develop in the spongiotrophoblast tissue, a process leading to invasion of maternal blood to the placenta (Rossant and Cross, 2001) (Figure II-5).

Key TFs that are involved in these processes include glial cell missing 1 (*Gcm1*), which drives the branching and the differentiation of syncytiotrophoblasts (Basyuk et al., 1999; Anson-Cartwright et al., 2000; Schreiber et al., 2000), and TF *AP-2γ* (*Tfap2c*) which promotes the formation of the labyrinth (Werling and Schorle, 2002).

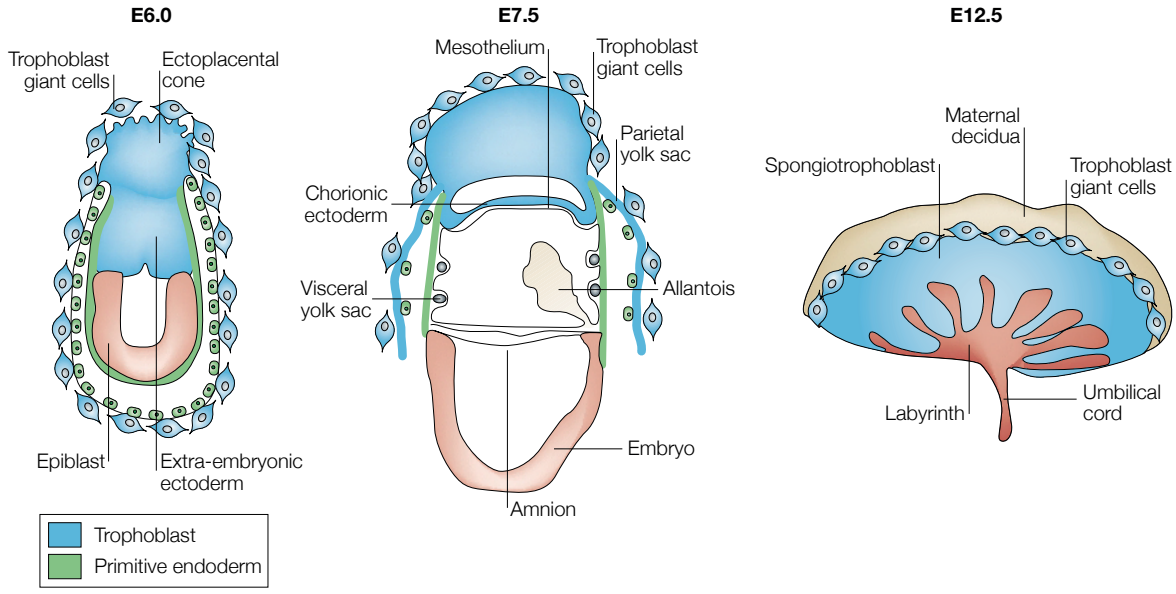

Figure II-5: *Illustrations of mid-stage placenta developmental in the mouse*
Adapted from (Rossant and Cross, 2001).

Human placental development

Our knowledge regarding human placental development is based primarily on hysterectomies. Compared with the mouse, where trophoblast giant cells mediate implantation and invasion of the mouse blastocyst into the uterus, in the human this is established by invasive extravillious trophoblast cells. Common to these cell types is that they are both polyploid (Berezowsky et al., 1995; Zybina and Zybina, 1996; MacAuley et

al., 1998). The chorionic villi of the human placenta are functionally equivalent to the labyrinth of the mouse placenta. In both species these villi/ labyrinth are covered with syncytiotrophoblasts that are in contact with the maternal blood (Rossant and Cross, 2001) and produce hormones like chorionic gonadotropin (CG), which is involved in placentation through activities such as maintaining angiogenesis of the uterine vasculature and promoting differentiation of cytotrophoblasts into syncytiotrophoblasts (Shi et al., 1993; Rao and Alsip, 2001; Zygmunt et al., 2002) (Figure II-6).

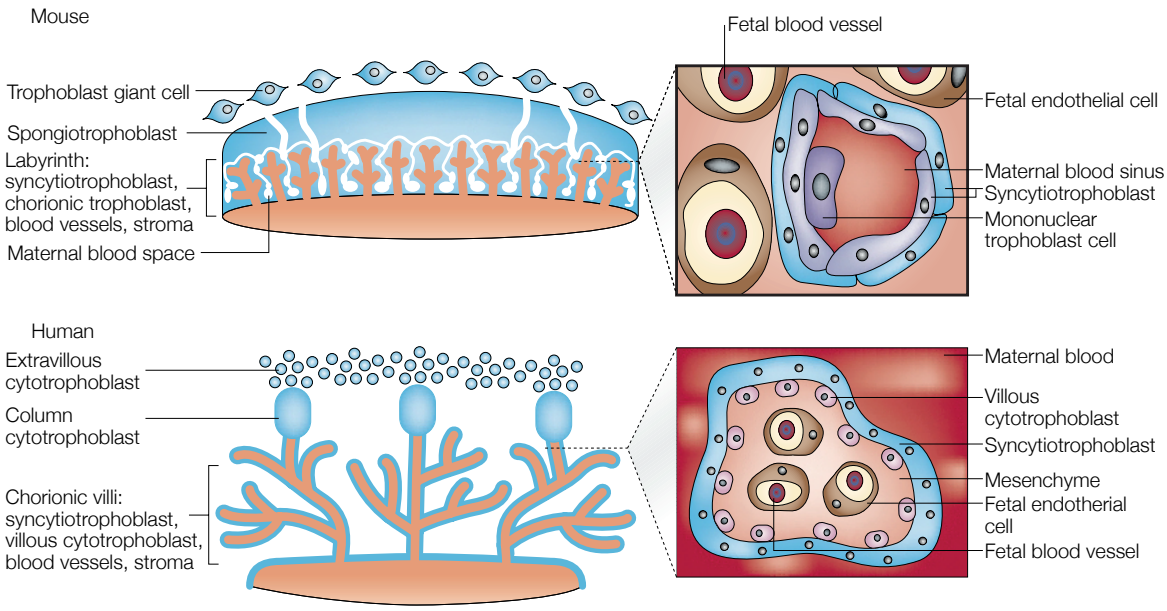

Figure II-6: *Gross anatomy* of *human and mouse placenta*
The inset image shows a cross-section of a chorionic villus, where TE-derived structures are shown in blue and mesoderm-derived tissues are displayed in orange (Rossant and Cross, 2001).

There exist some key differences between the species in relation to TE and blastocyst implantation. This, for example include that while the mouse blastocyst attaches to the uterus at the mural part of the TE, the human blastocyst makes its contact through the polar TE (Herzog, 1909). Another difference is that during early implantation the blastocyst is rapidly engulfed by the expanding decidua (the uterine stroma) in the mouse while during human implantation the TE cells are highly invasive and invade the uterine stoma (Hertig and Rock, 1973; Enders, 1976). Furthermore, it is thought that the mouse yolk sac plays an important role in nourishing the embryo before the placenta is developed, while in the human invasive trophoblasts rather than the yolk sac are important in this respect (Rossant, 2015).

Trophoblast stem cells

In the mouse, all the cell types of the placenta are thought to originate from multipotent precursor cells that exist initially in the polar TE and later in the ExE. It is thought that the renewal of these progenitors relies on FGF signaling because Fgfr2 is specifically expressed in TE cells and the ligand FGF-4 is expressed in the adjacent ICM cells in late mouse blastocysts (Yuan et al., 1995; Arman et al., 1998).

The information about the involvement of FGF4 in TE renewal has facilitated the culturing of mouse trophoblast stem cells (TSCs). Such cultures have been established so far from mouse blastocysts and ExEs, which was dissociated from the early-streak around day 6.5 post coitus (Tanaka et al., 1998; Uy et al., 2002). *In vitro* TSCs are able to renew at the presence of Fgf4 and can produce various cell types of the chorioallantoic placenta including spongiotrophoblast, syncytiotrophoblast and giant cells. When implanted into blastocysts TSCs contribute to the ExE, ectoplacental cone and giant cells, but not to the cells of the epiblast, PE or other extraembryonic tissues that are derived from the ICM (Tanaka et al., 1998).

Several TFs contribute to the establishment or the maintenance of TSCs, including *Cdx2*, *Gata3*, *Eomes*, *Tfap2c*, *Elf5*, *Ets2* and *Esrrb* (Latos and Hemberger, 2014). The functional importance of these factors was recently confirmed by reports showing that ectopic expression of combinations of these factors, *Tfap2c*, *Gata3*, *Eomes* and *Myc* or *Tfap2c*, *Gata3*, *Eomes* and *Ets2*, convert mouse fibroblasts into trophoblast stem-like cells (Benchetrit et al., 2015; Kubaczka et al., 2015). Importantly, derivation of human TSCs or cells that have similar self-renewal / differentiation features has not yet been reported to date.

The transcriptional network of TE development

The putative network of TFs that underlies mouse TE development seems to comprise of three primary layers, involved in the specification, reinforcement and further differentiation, respectively (Senner and Hemberger, 2010) (Figure II-7). *Tead4* and *Cdx2* are atop of this network (Nishioka et al., 2008) and regulate *Eomes* and *Tfap2c*, which subsequently regulate *Gata3*, *Elf5* and *Ets2*. This classification is primarily based on the severity of the respective phenotypes resulting from gene inactivation: loss of

Tead4 results in the earliest known lethality due to a developmental failure of the TE (Yagi et al., 2007; Nishioka et al., 2008), and *Cdx2* depleted blastocysts fail to implant, but not to specify TE precursors (Strumpf et al., 2005). Deletion of *Eomes* leads to a later impairment of TE differentiation and developmental arrest at E4.5 (Russ et al., 2000; Strumpf et al., 2005). Similarly, deletion of *Elf5* produces embryos that lack Cdx2 and Eomes expression, do not form the ExE and no TSCs can be derived from these embryos (Donnison et al., 2005; Ng et al., 2008). *Tfap2c* depletion leads to death at days 7 to 9 of embryonic development (Werling and Schorle, 2002), and knockout of *Ets2* leads to developmental impairments of TE cells (Georgiades and Rossant, 2006; Polydorou and Georgiades, 2013). Finally, *Gata3* seems to play an important role because overexpression of *Gata3* in mouse embryonic stem cells generates trophoblast committed cells, overexpression in TSCs promotes their differentiation (Ralston et al., 2010) and it is part of the cocktail that converts mouse fibroblasts into trophoblast stem-like cells (Benchetrit et al., 2015; Kubaczka et al., 2015).

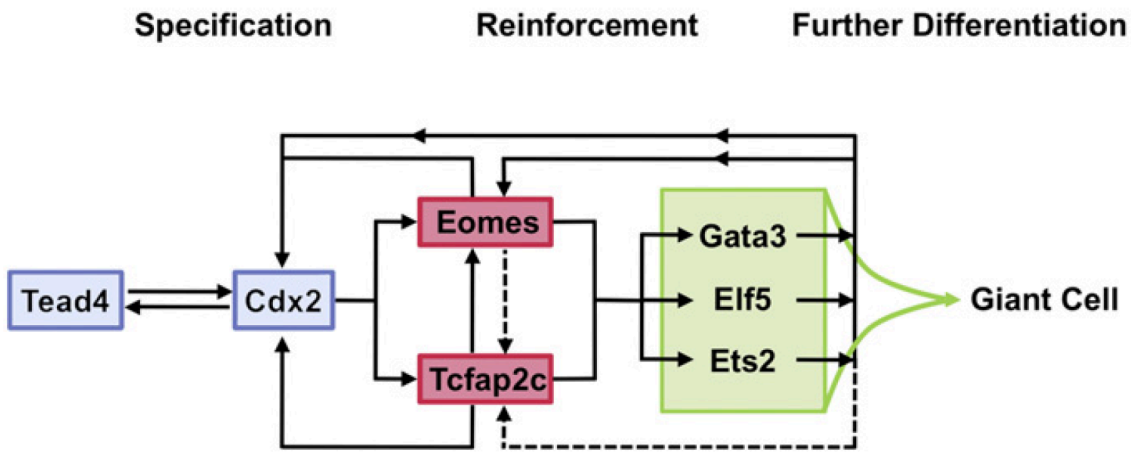

Figure II-7: *A putative transcription circuit of mouse trophoblast development*
The dashed lines indicate putative connections between *Eomes*, *Gata3*, *Elf5* and *Ets2* with *Tcfap2c* (Senner and Hemberger, 2010).

Comparison of human TE, human embryonic stem cells (ESCs, described below) and cells of the adult placenta resulted in a list of 16 TFs induced in the TE and placenta compared to human ESC, including TFs that have been shown to be involved in mouse trophoblast development like GATA3, TFAP2C and GCM1 and some that are still not connected with mouse trophoblast or TE development (Bai et al., 2012).
However, the connectivity of the network is still largely unknown.

Human pluripotent stem cells

Human pluripotent stem cells (PSCs), in the form of ESCs that are derived from blastocyst-stage embryos, or induced pluripotent stem cells (iPSCs) produced by reprograming of somatic cells using a cohort of pluripotency TFs, can be maintained indefinitely *in vitro*. Human ESCs / iPSCs can differentiate into derivatives of all three germ layers, the meso-, endo- and ectoderm. Interestingly, as early as the first report of human ESCs, it was noted that differentiation into progeny resembling cells of the TE lineage takes place, but with a low efficiency (Thomson et al., 1998). This is despite the fact that the gene network regulating human PSCs was found similar to that of the mouse, including *OCT4*, *NANOG* and *SOX2* (Nichols et al., 1998; Boyer et al., 2005; Wang et al., 2012). Moreover, it was shown that in human ESCs OCT4 and SOX2 can form a complex and bind to their own and to the NANOG promoter to activate their gene expression. These three factors also bind genes, as TFAP2C, important for developmental processes and can contribute to their silencing in human ESCs (Boyer et al., 2005). Therefore, OCT4, SOX2 and NANOG are considered the key responsible players for maintenance of the pluripotent state and restriction of differentiation.

OCT4

As Oct4 can antagonize Cdx2 expression, its regulation is relevant for ICM and TE specification, and is elaborated here in detail. During mouse development, Oct4 mRNA and protein is first detected in the oocyte (Scholer et al., 1989; Rosner et al., 1990; Yeom et al., 1991; Palmieri et al., 1994). During the 8-cell stage, Oct4 mRNA and protein expression increases again (Yeom et al., 1991; Palmieri et al., 1994), and becomes restricted to pluripotent cells upon the partitioning of the ICM and the TE. Importantly, although Oct4 is vital for renewal of ICM cells, it is not necessary for specification of the pluripotent cells (Frum et al., 2013; Wu et al., 2013). Later, Oct4 is expressed only in specific tissues of the epiblast, and finally is detected only in the primodial germ cells (PGCs) (Yeom et al., 1996), which give rise to the gametes (Figure II-8).

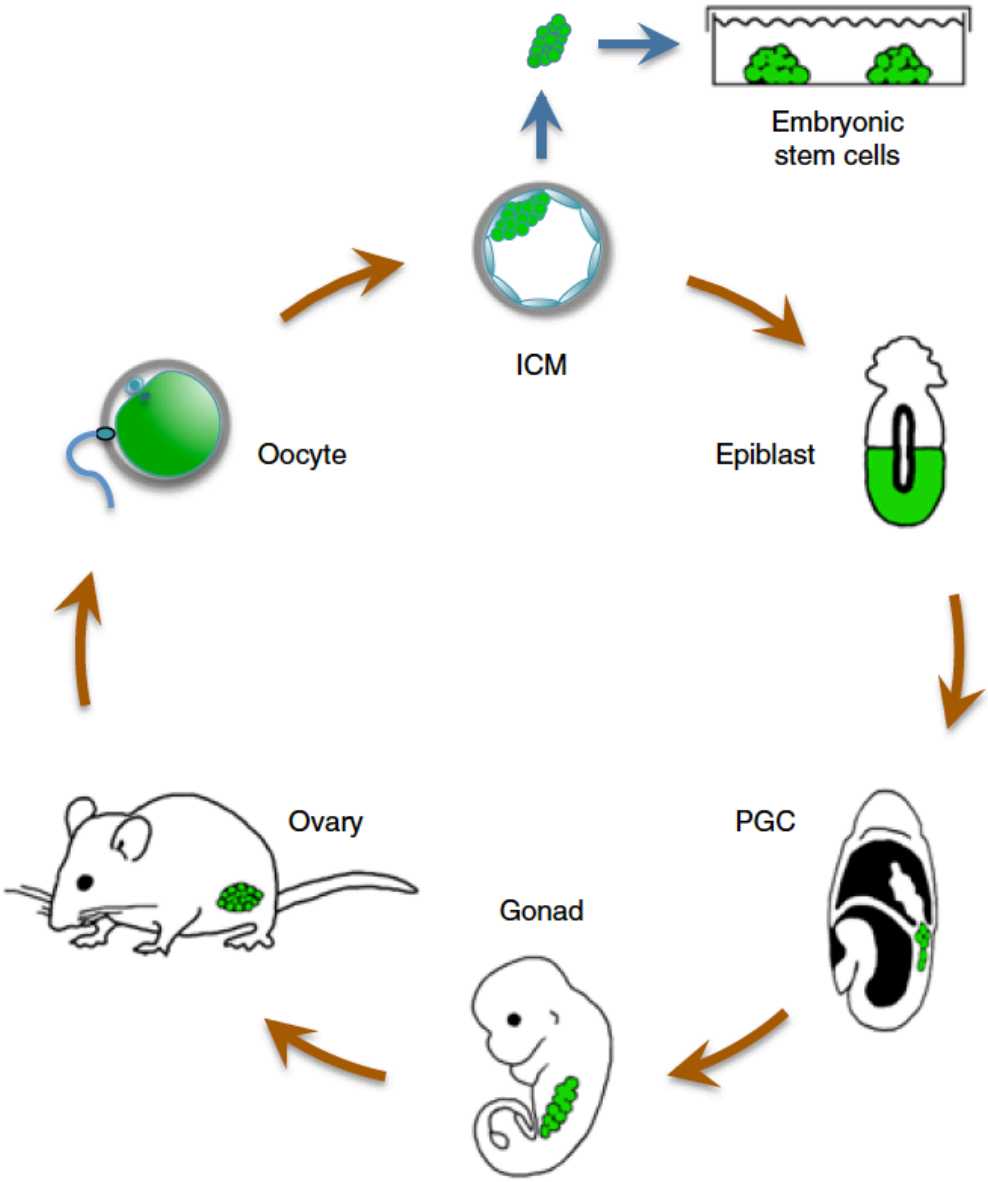

Figure II-8: *Expression of Oct4 during the mouse life cycle*
Oct4 (green) is expressed in the oocyte, the cells of the ICM, in the epiblast and finally in PGCs (Wu and Scholer, 2014).

Several regulatory regions exist in the *OCT4* locus, including the proximal promoter and proximal and distal enhancers. Within the enhancers there are four elements that are conserved between mouse, bovine and human, indicating a common mode of regulation (Nordhoff et al., 2001). The distal enhancer regulates the expression of *Oct4* in pluripotent / embryonic germ cells whereas the proximal enhancer is active in epiblast cells (Yeom et al., 1996; Tesar et al., 2007). This regulation involves proteins that can positively or negatively influence *Oct4* expression by binding to different regulatory elements (Wu and Scholer, 2014). For example, the binding of Cdx2 to the distal

enhancer downregulates *Oct4* in mouse TE cells (Niwa et al., 2005). Conversely, binding of pluripotency factors as Nanog, Sox2 or Oct4 itself to the enhancer regions induces its expression (Wu and Scholer, 2014) (Figure II-9).

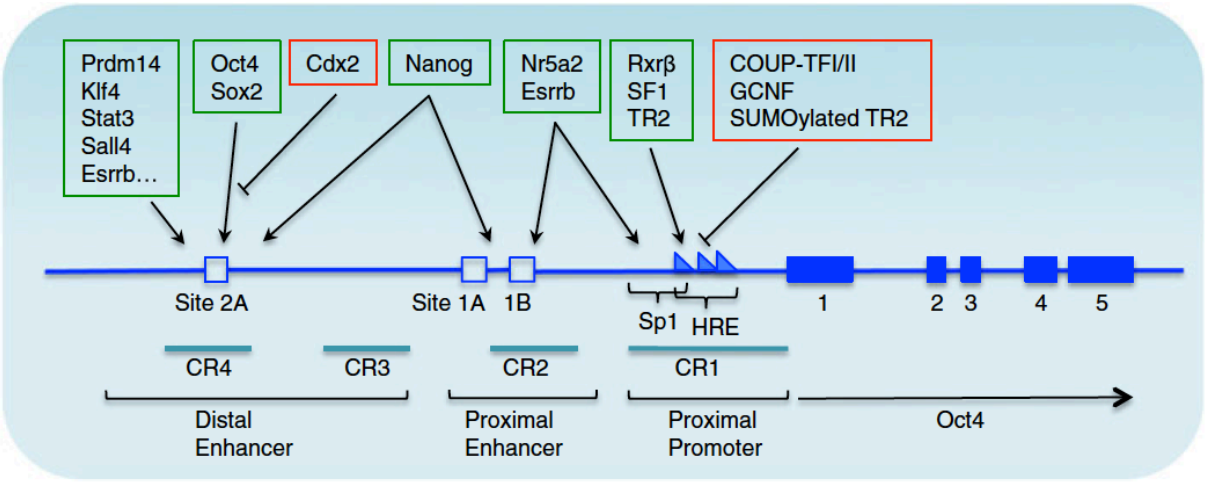

Figure II-9: *Transcriptional regulation of mouse Oct4*
A schematic overview of the *Oct4* gene. TFs and the respective binding sites that regulate *Oct4* are indicated on top. Green and red labels indicate positive and negative influence. CR= conserved region; Sp1= GC-rich site recognized by TFs of the Sp1/Sp3 family; HRE= hormone response element (Wu and Scholer, 2014).

Transcriptional processes in response to Bone Morphogenic Protein (BMP) 4 in human ESCs

BMP4 is a TGF-beta superfamily ligand that binds to type I and II BMP receptors (Allendorph et al., 2006). Binding of the ligand promotes phosphorylation and activation of the receptor regulated SMADs (R-SMAD), SMAD1/5/8, that associate with a common mediator SMAD (Co-SMAD), SMAD4. This heteromeric complex translocates into the nucleus where it regulates gene expression (Mukhopadhyay et al., 2008; Morikawa et al., 2011) (Figure II-10).

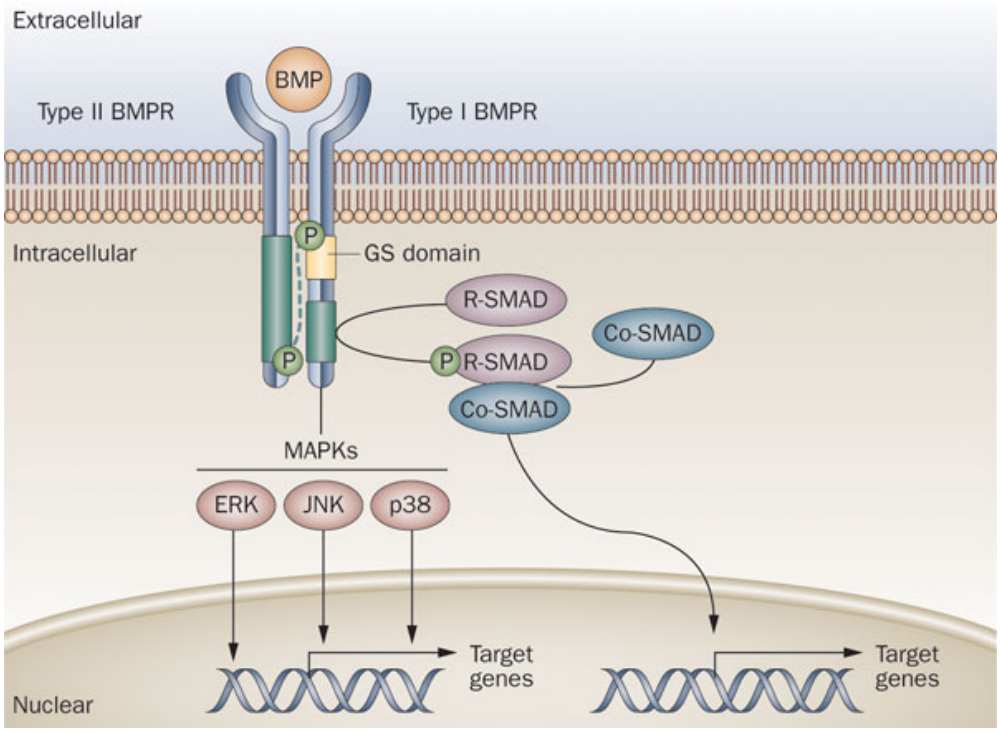

Figure II-10: *The BMP pathway* (Shore and Kaplan, 2010)

In the mouse 16-cell stage blastocysts components of the BMP pathway are already differentially expressed in inside and outside cells, with inside cells expressing mRNA of BMP ligands Bmp4 and Bmp7, whereas outside cells show mRNA expression of the receptor BmprII. Further manipulation of 4 cell stage mouse embryos, using dominant negative forms of Smad4 and BmprII, led to impaired formation of TE and PE, but not epiblast cells at E4.5 (Graham et al., 2014). However, embryos depleted of Smad4 or BmprII develop past the implantation stage, but show abnormalities in the extraembryonic tissues and cannot gastrulate (Beppu et al., 2000; Chu et al., 2004). BMP signaling plays also an important role in the mouse post-implantation embryo during the formation of the node and the primitive streak and it coordinates the left-right asymmetry and patterning in the mouse (Winnier et al., 1995; Fujiwara et al., 2002). Further it plays central parts in formation of PGCs and mesoderm (Lawson et al., 1993; Winnier et al., 1995).

BMP4-mediated trophoblast differentiation of human ESCs

Human tumor cell lines derived from choriocarcinoma, such as JAR or JEG3, primary cytotrophoblast cultures and immortalized extravillous cytotrophoblasts are classical

models for investigating trophoblast physiology (Ringler and Strauss, 1990). Yet, these systems cannot recapitulate the emergence of differentiated trophoblast cell types from progenitor cells (Genbacev et al., 2013). The derivation of human ESCs and later of human iPSCs revolutionized this aspect because it turned out that human PSCs have an intrinsic tendency to differentiate into trophoblast-like progeny by exposure to BMP ligands (Xu et al., 2002). Human ESCs treated with BMP4 (Telugu et al., 2013) or with a combination of BMP4, an Activin A inhibitor and a FGF2 inhibitor (Amita et al., 2013) could mimic the invasive behavior of some trophoblast cells. The phenotypes noted include rapid flatting of the cells, emergence of syncytium-like cells, production of placental hormones, expression of typical early trophoblasts genes and reduction of pluripotency genes (Xu et al., 2002; Das et al., 2007; Marchand et al., 2011; Drukker et al., 2012; Ezashi et al., 2012; Sudheer et al., 2012; Amita et al., 2013).

This trophoblast phenomenon was received initially with a surprise, as human ESCs are thought to represent a developmental stage past the segregation of the ICM and TE. The trophoblast nature of these cells was therefore challenged by a classification system used to classify *in vivo* first trimester trophoblasts. The *in vitro* differentiated cells fulfill some, but not all criteria. They express the tested trophoblast specific proteins (KRT7, TFAP2C and GATA3), their ELF5 promoter gets de-methylated, but the expression of the microRNA cluster C19MC, which is highly expressed in first trimester trophoblast, is only weakly expressed in the BMP4, Activin A inhibitor and FGF2 inhibitor treated cells. Further they do not express the placenta specific HLA-G, which induces immune tolerance during pregnancy (Lee et al., 2016).

Moreover, it was disputed that human ESCs derived trophoblast progeny emerges from TE fates cells: Bernardo and colleagues treated human ESCs with combinations of BMP4, FGF2 and Activin in chemically defined medium. Using BMP4 + FGF2, which does not resemble the culture conditions used before by others, BRACHYURY (BRA), a mesoderm associated TF, as well as CDX2, the trophoblast associated TF, were upregulated. The BRA upregulation was dependent on FGF2. In contrast to previous studies cells treated with BMP4 alone did only result in low amounts of KRT7+ cells (4-8% of all cells at day 7 of differentiation) and no de–methylation of ELF5. RT-PCR of the sorted KRT7+ cells revealed high expression of the trophoblast associated genes GCM1, ELF5 and HCGA, which encodes for hCG, but also of mesoderm associated genes ISL1 and FLK1 in KRT7+ compared to KRT7- cells. Therefore they conclude that this differentiation route represents an extraembryonic mesoderm rather than a trophoblast pathway (Bernardo

et al., 2011). However, a newer study shows that treatment of human ESCs with BMP4 also leads to expression of target genes including WNTs, which promote expression of mesoderm genes. This creates a mixture of trophoblast and mesodermal, WNT induced, cells, rather than extraembryonic mesoderm, as the addition of a WNT inhibitor leads to the formation of trophoblast cells only (Kurek et al., 2015).

Human PSC derived trophoblast progenitors

Understanding of developmental processes using human ESCs is hindered by the emergence of heterogeneous progeny. The differentiation of human ESCs with different factors can lead to the emergence of mixed populations of cells with different nature (Gifford et al., 2013). In our case BMP4 not only produces trophoblast cells, but also cells with other specifications, e.g. mesoderm (Bernardo et al., 2011; Kurek et al., 2015). This complicates the identification of ground laying mechanisms during one specific developmental process, as the unwanted side products can blur the real drivers of the differentiation.

This can be circumvented by using antibodies against surface antigens expressed by unique progenitor populations. Aminopeptidase A (APA, also known as CD249, Ly-51), encoded by the ENPEP gene, was shown to be a suitable candidate to sort purified trophoblast progenitors from BMP4 differentiated cells, as APA+ but not APA- cells develop into placental structures, shown by expression of the placental specific proteins STS, the placental alkaline phosphatase (PLAP) and human placental lactogen (HPL) and others, when implanted into mice (Drukker et al., 2012). Furthermore APA was shown to be present on the surface of human syncytiotrophoblast cells (Ito et al., 2003) and it was also found to be involved in the regulation of maternal blood pressure in the mouse (Mitsui et al., 2003b).

Chromatin associated histone turnover in development and differentiation

The molecular basis of tissue formation is the expression of unique sets of genes from the genome, which is uniform in all cell types. Epigenetics is the discipline that deals with the mechanisms of gene regulation that are heritable over cell and organismal

generations without changes in the DNA sequence. Some of the best-characterized epigenetic mechanisms are biochemical modifications of histone tails and additions of methyl groups to DNA.

Nucleosome architecture and the histone code

The eukaryotic DNA is compressed in the form of a polymer named chromatin (Paweletz, 2001). Two forms of chromatin are distinguished by their density, known as eu- and heterochromatin during interphase that correspond to open and closed configurations of chromatin. The nucleosome is the basic unit of the chromatin, which consist of a histone octamer containing two copies histone H2A, H2B, H3 and H4 with 146bp of DNA wrapped around it, and the histone tails reaching outside of this structure (Luger et al., 1997) (Figure II-11).

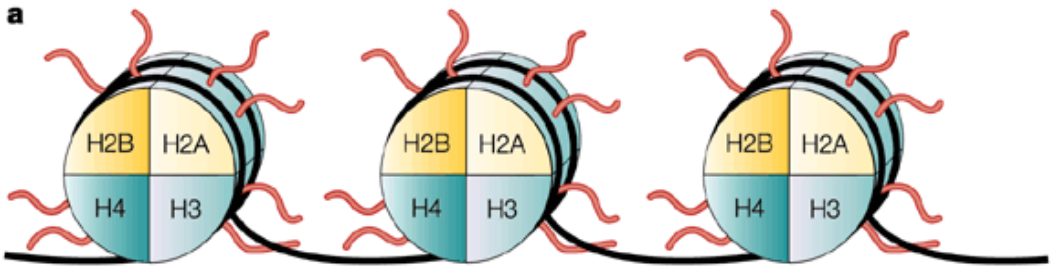

Figure II-11: *The structure of the nucleosome*
DNA that is wrapped around the histone octameres, consisting of each two copies of histones H2A, H2B, H3 and H4. Red lines indicate histone tails projecting from the nucleosome (Marks et al., 2001).

The structure of the chromatin is determined by several key factors, including variants of the core histone proteins encoded by different genes and producing chromatin of different densities. This influences the recruitment of activating or repressing factors. For example, histone H3.3, a variant of H3, is found in transcriptionally active loci, whereas histone H2AZ, a variant of H2A, recruits the heterochromatin specific protein HP1α and thereby promotes the maintenance of heterochromatin (Sarma and Reinberg, 2005). Another process that changes the nucleosome distribution and position is chromatin remodeling, mediated by nucleosome sliding or nucleosome eviction which enhances accessibility of regulatory sequences (Cairns, 2007). Modification of specific

residues on histone tails, called posttranslational modifications (PTM), can further affect the chromatin structure and accessibility of the DNA for TFs. The addition or removal of these reversible PTMs can lead to activation or repression of a specific DNA locus (Jenuwein and Allis, 2001).

Many different histone modifications have been reported and best characterized are those involving histone methylation, acetylation and phosphorylation (Figure II-12). The recognition, establishment and removal of PTMs depend on specific proteins, called readers, writers and erasers.

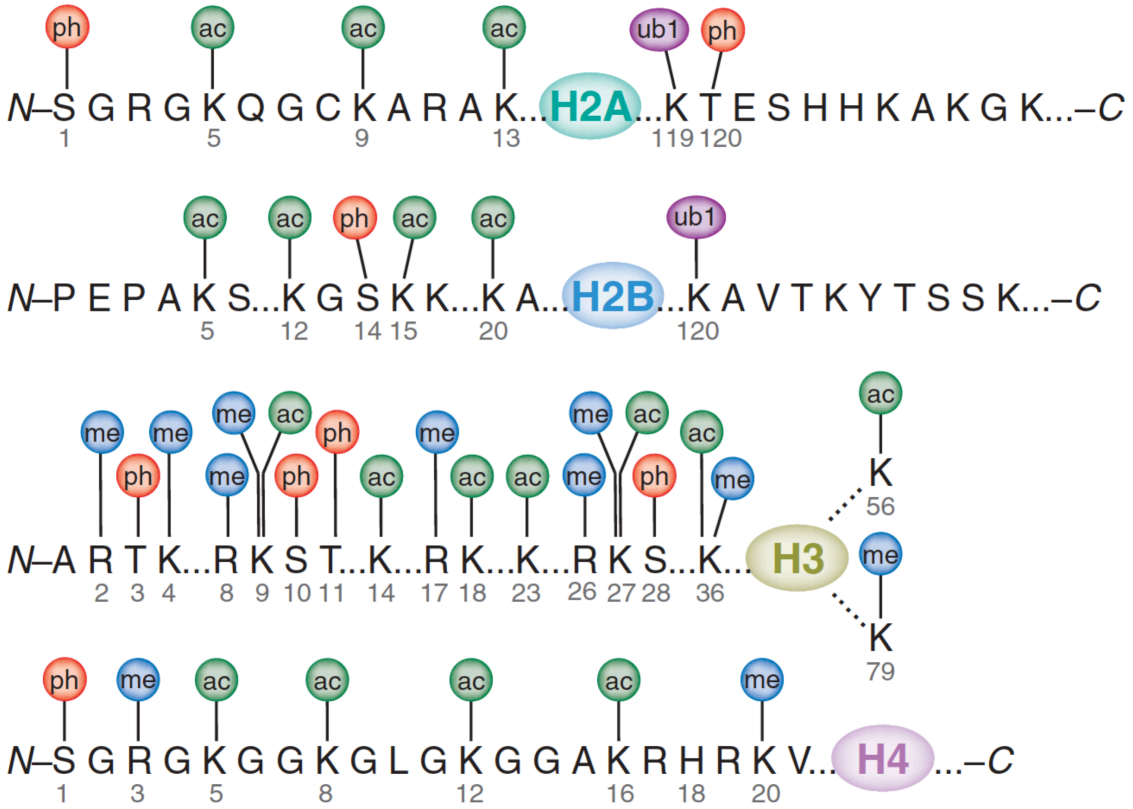

Figure II-12: *Post-translational modifications of histone tails*
A simplified diagram of methylation, acetylation, phosphorylation and ubiquitination marks of histone tails. The amino acid residues of the histone tails H2A, H2B, H3 and H4 are shown. The numbers indicate the position of the respective amino acid in the histone tail. Ph= phosphorylation; ac= acetylation; ub1= ubiquitination; me= methylation (Bhaumik et al., 2007).

Methylated histone tails

Histone methylation takes place on lysine and arginine residues of the histone tails (Shilatifard, 2006; Kouzarides, 2007). While arginines are mono- or dimethylated, lysines can be mono-, di- and trimethylated. The most prominent and best-characterized methylation marks are on lysine 4, 9, 27, 36 and 79 of histone H3 and at lysine 20 of histone H4. These histone marks have been studied extensively with respect to transcriptional regulation.

Acetylated histone tails

Acetylation of specific lysines on histone tails leads to neutralization of their positive charge and alters the DNA-histone interaction leading to a more open chromatin conformation. This makes the DNA more accessible for proteins like TFs (Shahbazian and Grunstein, 2007). The writers of histone acetylation are called histone acetyl transferases (HATs) and the erasers histone deacetylases (HDACs). Many transcriptional coactivators, such as p300 have an intrinsic HAT activity, whereas transcriptional corepressor complexes contain HDAC subunits (Denslow and Wade, 2007; Shahbazian and Grunstein, 2007).

Phosphorylated histone tails

Serine and threonine residues of histone tails can be phosphorylated. Phosphorylation of these amino acids is involved in DNA repair, mitosis and activation of transcription (Rossetto et al., 2012). Transcriptional activation is mediated by a crosstalk of phosphorylation marks with enzymes that can acetylate histone tail residues next to the phosphorylation sites, promote removal of the methylation mark of H3K9me3 or inhibit the removal of methylation at the H3K4 position. All these events lead to an open transcriptionally active chromatin configuration (Lau and Cheung, 2011).

DNA methylation

The methylation of the fifth position of the cytosine nucleoside is conserved throughout the animal kingdom (Feng et al., 2010), and most of the methylation in mammals takes place in C-phosphate-G dinucleotides (CpGs)(Ramsahoye et al., 2000; Ziller et al., 2011). At least three enzymes catalyze DNA methylation in humans: DNMT1 is ubiquitously

expressed and maintain the methylation marks during cell division, and DNMT3A and DNMT3B are developmentally regulated and carry out de-novo methylation (Li et al., 1992; Okano et al., 1999). In the human genome, which is overall depleted of CpGs, there are around 28 million CpG sites of which 60-80% are constitutively methylated. Out of these, less than 10% are located in so called CpG islands, which are regions with a high CpG content. These regions are mostly unmethylated and located at the promoters of developmental and housekeeping genes (Deaton and Bird, 2011). Enzymes, namely TET1, TET2 and TET3 can remove the methylation mark of the CpG by hydroxymethylating the CpG base and therefore reverting the effect of the before methylated CpG (Tahiliani et al., 2009; Guo et al., 2011). CpG hydroxymethylation exists mainly in the brain and PSC, both in humans and mice (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009; Pastor et al., 2011; Szulwach et al., 2011).

Epigenetic profiles of PSCs

Electron microscopy of heterochromatin domains and DNaseI or Micrococcal nuclease (MNase) digestion, which allow the detection of open chromatin, showed that in mouse and human ESCs the chromatin state is more open than in somatic cells, meaning it harbors less heterochromatin (Park et al., 2004; Efroni et al., 2008; Schaniel et al., 2009; Ahmed et al., 2010) (Figure II-13).

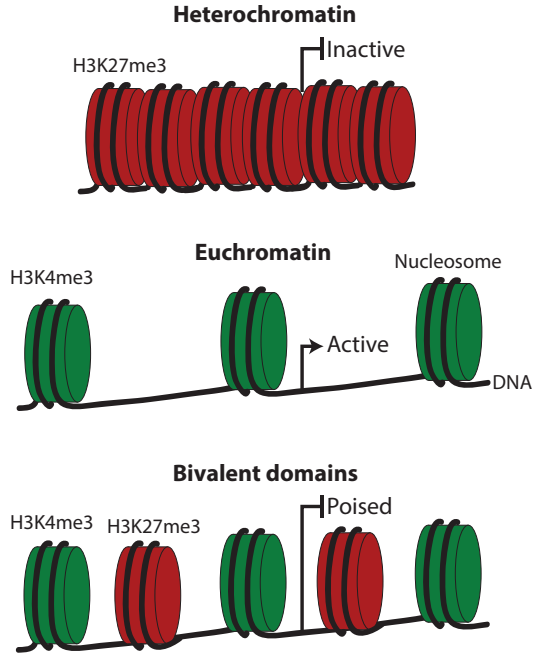

Figure II-13: *Schematic illustration of different chromatin regions.*
The upper pane shows condensed, transcriptional inactive heterochromatin marked with histone H3K27 tri-methylation marks.
The middle pane shows open, transcriptional active euchromatin marked with histone H3K4 tri-methylation marks.
The lower pane shows bivalent domains that are marked by histone H3K4 and K27 tri-methylation marks at the same position, which poises them for transcriptional activation or repression.

Transcriptionally inactive heterochromatin in PSCs is characterized by repressive H3K9 di- and trimethylation marks (H3K9me2, H3K9me3), found at repetitive elements like centromeres (Peters et al., 2001; Lehnertz et al., 2003) or by trimethylation of H3K27 (H3K27me3) at facultative heterochromatin regions (Boyer et al., 2006; Lee et al., 2006; Mikkelsen et al., 2007). In the same cells, genes in transcriptionally active euchromatin are marked by H3K4me3 at their promoter regions (Bernstein et al., 2006; Mikkelsen et al., 2007) and by H3K27 acetylation, H3K4me1 and the presence of the HAT p300 at their respective enhancer region (Rada-Iglesias et al., 2011). Moreover, actively transcribed genes contain H3K36me3 marks in the gene body (Kolasinska-Zwierz et al., 2009). This active histone modification combination, comprising of H3K4me1, H3K4me3, H3K27 acetylation and H3K36me3 distributed in different regions around the gene, is present in pluripotency and housekeeping genes in mouse and human PSCs.

Mouse and human PSCs also feature a unique category of histone modifications named the bivalent domains characterized by co-occupancy of H3K4me3 activating and H3K27me3 repressing marks (Bernstein et al., 2006; Mikkelsen et al., 2007). Bivalent domains are associated with promoters of developmental genes that contain high CpG regions in PSC but bivalent domains are found rarely in differentiated cells (Watanabe et al., 2013). The conceptual framework states that it is the existence of bivalent domains that allows developmental genes to be rapidly silenced or activated during differentiation by loss of the activating H3K4me3 mark or by removal of the repressing H3K27me3 mark, respectively (Pan et al., 2007; Pasini et al., 2010). This concept has been debated initially by researchers who argued that the existence of bivalent domains could be explained by analysis of mixture of cells harboring activating and repressive marks at the same respective loci. This concern has been dismissed by sequential ChIP assays for H3K4me3 and then for H3K27me3 (or vice versa) that confirmed that both marks could be detected at the same region (Pan et al., 2007; De Gobbi et al., 2011). Moreover, analysis of mononucleosomes in conjunction to sequential ChIP in mouse ESCs revealed that H3K4me3 and H3K27me3 can co-exist on the same nucleosome (Voigt et al., 2012). Importantly, there is a correlation between the presence of high and medium CpG promoters and the presence of bivalent domains in mouse and human ESC. Furthermore, practically all promoters that are CpG rich lack DNA methylation marks in human and mouse PSC (Weber et al., 2007; Fouse et al., 2008; Meissner et al., 2008; Mohn et al., 2008).

Promoters with a low percentage of CpGs are less well studied as they are found at a

very small part of the human genes. Only around 6,5% of these low CpG promoters harbor H3K4me3 marks and do not show any H3K27me3 mark in mouse ESCs. Low CpG genes have also been associated with highly tissue specific functions (Mikkelsen et al., 2007) and harbor DNA methylation marks in human ESC. During human ESC differentiation the DNA methylation levels of these low CpG genes anti-correlate with gene expression (Xie et al., 2013).

During differentiation of ESCs it was shown that pluripotency genes acquire repressive marks, both, histone modifications and DNA methylation, whereas genes, that are important during this differentiation state loose the repressive H3K27me3 mark and acquire additional activating H3K4me3 marks. On the contrary, if the gene is not needed during this developmental step it looses the activating H3K4me3 mark and gets silenced (Christophersen and Helin, 2010).

Taken together, all these epigenetic events contribute to the establishment of unique gene expression patterns in specialized cell types resulting from PSC without changes in the DNA sequence.

III. Aims of the study

My interest lies in the first differentiation event of human development – the specification of TE progenitors. I have set the primary objective to utilize human PSCs as a model and novel surface markers defined by the Drukker lab for purifying trophoblast progenitors as a basis for my analysis. Because of the ethical ban on research of human embryos as well as the minute quantity of chromatin in mouse embryos, this human PSC progenitor-progeny system is both essential and uniquely tailored to break grounds understanding the cell intrinsic molecular properties that drive the specification of human TE progenitors. The specific aims of my work include:

1. To resolve the uncertainty concerning the lineage correspondence of human PSC-derived trophoblast progeny

There exists uncertainty regarding the lineage correspondence of BMP4-treated PSC progeny that exhibits trophoblast features, including expression of the trophoblast specific genes *KRT7*, *CDX2*, *GATA3*, *GCM1*, *ELF5* and others, production of the placental hormone human Chorionic Gonadotropin (hCG), presentation of the placental immune regulator HLA-G, and exhibiting demethylation of the *ELF5* promoter (Xu et al., 2002; Amita et al., 2013; Lichtner et al., 2013; Lee et al., 2016). These properties are regarded by many as evidence for a TE decent of the cells (Roberts et al., 2014) however, other investigators claimed that the expression of genes that are characteristic to mesoderm progeny, which includes blood, muscle, and bone tissues, such as *ISL1* and *FLK1* and the absence of KRT7 at the day 7 after differentiation, indicates that the trophoblast progeny emerges from mesoderm precursors (Bernardo et al., 2011). What complicates concluding the lineage classification using the existing data is the fact that BMP4 also induces the expression of WNT3 during the differentiation process leading to heterogeneous cultures consisting of trophoblast and mesoderm cell populations as our lab has shown in collaboration (Kurek et al., 2015). However, this is still not a proof of the TE decent of BMP4-induced trophoblast progeny. To resolve this issue I have set to address the following questions:

a. Does the gene cohort expressed in progenitors during the differentiation of human PSC support mesoderm classification?
b. Does the gene cohort expressed in progenitors during the differentiation of human PSC support trophoblast classification?
c. What is the lineage correspondence of the non-trophoblast progenitors emerging during the differentiation of human PSCs?
d. Does the gene cohort expressed in BMP4 induced purified trophoblast progeny from human PSCs correspond to blastocyst stage human TE progeny?

My experimental approach included global transcriptomic and tissue classification bioinformatics methods coupled to culturing of bulk human PSCs and cell sorting modalities of differentiated derivative populations. Furthermore, I used existing transcriptomic datasets of human blastocyst-stage TE as a basis for comparison to the *in vitro* generated trophoblast progenitors.

2. To explain the genetic mechanism that drives trophoblast specification of human PSCs

Previous studies have identified TFs that are highly expressed following BMP4 treatment of human PSCs (Xu et al., 2002; Marchand et al., 2011; Sudheer et al., 2012). Nevertheless they were not placed in the context of a TF network that explains the mode of trophoblast formation. I reason that this is due to two primary causes:
1. Cellular heterogeneity: studies from our lab and others have shown that there is a high degree of lineage heterogeneity in early differentiating human PSC cultures (Drukker et al., 2012; Kurek et al., 2015). Cell purification therefore has to be applied for analyzing the intrinsic properties of cell lineages, a principle that has been demonstrated in defining the populations of the hematopoietic system, the nervous system and others (Hoppe et al., 2014).
2. Post lineage commitment analysis: previous studies of human PSC differentiation towards trophoblast lineages mainly focused on analyzing cultures that already contain trophoblasts, which are therefore not compatible with attempts to analyze the mechanism leading to trophoblast specification from human PSCs. I have therefore set the following questions for identifying the TF network that underlie human trophoblast specification:

a. What are the key TFs upregulated by BMP4 treatment prior to the emergence of committed trophoblast progenitors from human PSCs, and do they correspond to mouse TE key genes?
b. What are the trajectory categories of the upregulated TFs and do they indicate layers of a putative TF network?
c. Is there a correspondence between the key putative TFs upregulated before trophoblast progenitors and TFs enriched in purified trophoblast progenitors?
d. What is the connectivity of the putative trophoblast TF network?
e. Does reduction of early key trophoblast TF candidates perturb trophoblast progenitor specification and the putative network?
f. How does the trophoblast TF network relate to downregulated genes during the process of trophoblast progenitor specification

To address these questions I have used the above-mentioned approaches, and in addition I determined the bound human genomic loci of key putative TFs for identifying the connectivity of the TE TF network. Moreover, I used functional knock-out assays for proving the network connectivity.

3. To decipher epigenetic mechanisms that underlie trophoblast specification from human PSCs

Analysis of a spectrum of histone modifications of bulk cultures of human ESCs treated by BMP4 (Xie et al., 2013) did not yield a significant insight into epigenetic mechanisms underlying trophoblast commitment. I reason that this is due to the above-mentioned confounding effects of lineage and stage heterogeneity. To identify epigenetic chromatin-related mechanisms that play key roles during this differentiation process I have set to address the following questions:

a. What is the turnover of the activating and repressing histone modifications of key genes in trophoblast progenitors and their precursors?
b. What is the turnover of DNA methylation marks of key genes in trophoblast progenitors and their precursors?

c. Can transcriptional reduction and induction of key pluripotency and TE genes be explained by changes in histone configurations and DNA methylation marks during trophoblast progenitors commitment?

To address the questions related to the histone modification and DNA methylation turnover I analyzed the genome-wide distribution of the H3K4me3 modification, which is correlated with transcribed chromatin, the H3K27me3 modification, which is correlated with non-transcribed chromatin, and of CpG methylation in purified undifferentiated cells and trophoblast progenitors. I have correlated this chromatin data with transcriptomic data for gaining a holistic view of the regulation governing human TE specification, Finally, this combinatorial analysis enabled me to define a human trophoblast chromatin signature similar to the one identified for the cardiac progeny of human PSC (Paige et al., 2012).

IV. Material and Methods

Material

Cell culture	Article number	Vendor
2-Mercaptoethanol	31350-010	Life Technologies
Accutase	A6964-100ML	Sigma-Aldrich
B27 Supplement, minus insulin-10 mL	A1895601	Life Technologies
Collagenase, type IV	17104019	Life Technologies
DMEM	21969035	Life Technologies
DMEM/F12	11320074	Life Technologies
DMSO	D5879-100ml	Sigma-Aldrich
Gelatin powdered, pure Ph. Eur., NF	A1693,0500	AppliChem
GlutaMAX, 100X	35050038	Life Technologies
HyClone™ Fetal Bovine Serum (South America), Research Grade	SV30160.03 (LOT Nr RZB35918)	GE Healthcare
Knockout-Serum Replacement (KSR)	10828028	Thermo Fisher Scientific
Matrigel-Matrix	FALC354230	Schubert&Weiss
MEM Non-Essential Amino Acids Solution (100X) (NEAA)	11140050	Thermo Fisher Scientific
Millex-GP Syringe Filter Unit, 0.22 µm, polyethersulfone, 33 mm, gamma sterilized	SLGP033RS	Merck Millipore
mTESR1	5850	Stemcell Technologies
PBS, 1x	14190094	Life Technologies
Penicillin-Streptomycin	15070063	Life Technologies
Recombinant Human BMP-4 Protein	314-BP	R&D Systems
Recombinant Human FGF-basic (154 a.a.)(bFGF)	100-18B	Peprotech
RPMI-1640 medium	21875034	Life Technologies
Trypsin-EDTA (0.25%), phenol red	25200056	Thermo Fisher Scientific
Y-27632 dihydrochloride (ROCK inhibitor)	1254/10	R&D Systems
µ-Slide 8 Well, ibidi-treat	80826	Ibidi

Human Chorionic Gonadotropin	Article number	Vendor
AccuLite® CLIA	8575-300	Monobind
PREGNANCY TESTS	720-0723	VWR

ChIP-Seq	Article number	Vendor
16% Formaldehyde (w/v), Methanol-free 10 x 10 mL	10321714	Thermo Fisher Scientific
AbSurance Histone H3 Antibody Specificity Array	16-667	Merk Millipore
Agencourt® AMPure® XP, 60 mL	A63881	Beckman Coulter
Albumin from bovine serum (BSA)	A9647-10G	Sigma-Aldrich
ChIP-IT high sensitivity kit	53040	Active Motif
Clarity Western ECL Substrate, 200 ml	170-5060	Bio-Rad Laboratories
Dynabeads(R) Protein A for Immunoprecipitation	10001D	Life Technologies
EDTA Dinatriumsalz Dihydrat >99% 250g	X986.1	Carl Roth
EGTA	3054.1	Carl Roth
Glycine	23391,02	Serva
HEPES, 1M Buffer Solution 20x100ml	15630122	Life Technologies
HiSeq® Rapid SBS Kit v2 (50 Cycle)	FC-402-4022	Illumina
HiSeq® Rapid SR Cluster Kit v2	GD-402-4002	Illumina
NEBNext® ChIP-Seq Library Prep Reagent Set for Illumina®	E6200 S	New England Biolabs
NEBNext® Multiplex Oligos for Illumina® (Index Primer Set 1)	E7335S	New England Biolabs
Nuclease-free water (H2O)	AM9932	Life Technologies
Phenylmethylsulfonylfluorid 5 g (PMSF)	6367,1	Carl Roth
Protease Inhibitor Cocktail Set III, EDTA-Free - Calbiochem	539134	Merk Millipore
Proteinase K Solution 20 mg/ml 1.25 ml	AM2546	Life Technologies
Qiaquick PCR purification kit	28104	Qiagen
Quant-iT PicoGreen dsDNA Assay Kit	P7589	Life Technologies
Qubit dsDNA HS Assay Kit	Q32854	Life Technologies
Rabbit IgG isotype control antibody 100 µg	GTX35035	Biozol Diagnostica
Ribonucleic acid, transfer from bakers yeast (tRNA)	R5636-1ML	Sigma-Aldrich
RNA 6000 Pico kit	5067-1513	Agilent
RNase A (20 mg/ml)	12091021	Life Technologies
RNeasy MinElute cleanup kit	74204	Qiagen
RNeasy mini kit	4104	Qiagen
SDS Solution, 20 %	20768,02	Serva Electrophoresis
Sodium Chloride	P029.2	Carl Roth
Sodium-deoxycholate	D6750-10G	Sigma-Aldrich
TRIS PUFFERAN®	5429,3	Carl Roth
Triton™ X-100	X100-500ML	Sigma-Aldrich
TURBO DNase 2U/µl 1000U	am2238	Life Technologies

Microarray & RNA seq & RT-PCR	Article number	Vendor
Encore™ Biotin Module , 60 pack	4200-60	NuGen
GeneChip® Human Gene 2.0 ST Array, 30 arrays	902113	Affymetrix
GeneChip® Hybridization, Wash, and Stain Kit	900720	Affymetrix
NextSeq 500/550 v2 reagent cartridge, 75x	FC-404-2005	Illumina
Ovation® Pico WTA System V2 , 60 pack	3302-60	NuGen
Power SYBR® Green PCR Master Mix	4367659	Thermo Fisher Scientific
RNA 6000 Pico kit	5067-1513	Agilent
RNeasy MinElute cleanup kit	74204	Qiagen
RNeasy Mini kit	74104	Qiagen
SuperScript® III First-Strand Synthesis System for RT-PCR	18080051	Life Technologies
TaqMan® Gene Expression Master Mix	4369016	Thermo Fisher Scientific
TruSeq Stranded total RNA LT kit (RiboZero Gold)	RS-122-2301	Illumina

DNA methylation	Article number	Vendor
EZ DNA Methylation-Lightning™ Kit	D5030	Zymo Research
Infinium HumanMethylation450 BeadChip Kit	WG-314-1003	Illumina
Wizard DNA isolation kit	A1120	Promega

Immunofluorescence & western blot	Article number	Vendor
2-Mercaptoethanol	M3148	Sigma-Aldrich
Clarity Western ECL Substrate, 200 ml	170-5060	Bio-Rad Laboratories
goat anti-mouse IgM-HRP	sc-2064	Santa Cruz
Ibidi mounting medium	50001	Ibidi
Mini-PROTEAN TGX Stain Free Gels, 4-15%	456-8086	Bio-Rad Laboratories
PhosSTOP™	4906837001	Sigma-Aldrich
Powder Milk, blotting grade	T145.1	Carl Roth
ProLong® Gold Antifade Reagent with DAPI	8961	New England Biolabs
Protease Inhibitor Cocktail Set III, EDTA-Free - Calbiochem	539134	Merk Millipore
TRIS PUFFERAN®	5429.3	Carl Roth
Triton™ X-100	X100-500ML	Sigma-Aldrich
µ-Slide 8 Well, ibidi-treat	80826	Ibidi

Genome editing	Article number	Vendor
Accutase	A6964-100ML	Sigma-Aldrich
Boric acid	B7901	Sigma-Aldrich
Doxycycline hydrochloride	D9891-1G	Sigma-Aldrich
GeneJET Plasmid Miniprep Kit, 50 preps	K0502	Fermentas
LE Agarose, 500 g	840004	Biozym
Lithium acetate dihydrate	L6883	Sigma-Aldrich
NEB® 5-alpha Competent E. coli (High Efficiency)	C2987 I	NEB
P3 Primary Cell 4D-Nucleofector® X Kit L (24 RCT)	V4XP-3024	Lonza
PureLink HiPure Plasmid Filter Maxiprep Kit	K210017	Life Technologies
Quick Ligation™ Kit	M2200 L	NEB
QuickExtract DNA Extraction Solution 1.0	101098	Biozym
SYBR® Safe DNA Gel Stain	5001208	Life Technologies
Taq PCR Master Mix Kit (1000 U)	201445	Qiagen

Tables IV-1-7: *Reagents used in this study*

Media composition

bFGF medium		MEF medium	
400 ml	DMEM/F12	500ml	DMEM
100 ml	KSR	50 ml	FBS
5 ml	NEAA	5.5 ml	NEAA
5 ml	Glutamax	5.5 ml	Glutamax
10 ng/ml	bFGF	5 ml	Penicillin-Streptomycin
1 ml	2-mercaptoethanol		
5 ml	Penicillin-Streptomycin		

Table IV-8: *Media composition of bFGF and mouse embryonic fibroblast (MEF) medium*

Diff medium 1		Diff medium 2	
400 ml	DMEM/F12	49 ml	RPMI medium1640
100 ml	KSR	1 ml	B27 Supplement, minus insulin
5 ml	NEAA	50 ng/ml	BMP4 (added fresh daily)
5 ml	Glutamax		
1 ml	2-mercaptoethanol		
5 ml	Penicillin-Streptomycin		
50 ng/ml	BMP4 (added fresh daily)		

Table IV-9: *Media composition Differentiation (Diff) medium 1 and 2*

mTESR1

mTESR1 was prepared by adding 100 ml mTESR1 5x Supplement to 400 ml mTESR1 Basal Medium and addition of 5 ml Penicillin-Streptomycin.

1x Matrigel (MG)

Matrigel was prepared by thawing on ice and dilution of 1 ml MG in 50 ml cold DMEM/F12. This mix was then stored at 4°C and used for up to 4 weeks.

2x Collagenase

Collagenase was prepared by addition of 100 mg of collagenase to 50 ml of DMEM/F12. The collagenase was dissolved and sterile filtered.

Freezing medium

Freezing medium contained of culture medium (bFGF or mTESR1) plus 10% DMSO.

FACS medium

FACS medium was prepared by addition of 2-4% FBS and 5 mM EDTA to PBS.

Methods

Cell culture

All centrifugation steps were performed at RT for 4 min and 1200 RPM on a Megafuge 40R centrifuge (Thermo Fisher Scientific).

Human ESC lines used:

For this work I used two different human embryonic stem cell lines, namely H9 (WA09) and HUES9 iCRISPR. The H9 cells were used for the main part of the project (Microarray, ChIP-Seq, DNA-methylation, time-course RNA-Seq, immunofluorescent microscopy), whereas the HUES9 iCRISPR cells were used for genetic manipulation experiments. The

reason for this is that the HUES9 iCRISPR cells were genetically manipulated by Transcription activator-like effector nucleases (TALENs) in a way that the DNA endonuclease Cas 9 was inserted in the AAVS1 locus of the genome of HUES9 cells. The expression of Cas 9 can be induced by addition of doxycycline. This system allows editing the genome of the cells just by introducing guide RNAs and addition of doxycycline. Thus it represents a fast and easy way to perform gene editing experiments with high efficiency (Gonzalez et al, 2014).

Culture conditions

Maintenance:

H9 cells

H9 cells were cultured on a layer of irradiated mouse embryonic fibroblasts (MEFs) in bFGF medium. The feeder layer is used because it produces different factors that are important for the maintenance of pluripotency. Passaging of these cells was performed every 3-4 days when cells were 70-80% confluent. Therefore one day prior to passaging, the irradiated feeders were plated on gelatin coated plates and left to attach for one day in MEF medium. For splitting, the H9 cells were incubated for 45 min to 1 hour with 2x collagenase. After the colonies detached the cells were collected with bFGF medium and allowed to settle by gravity. When the majority of cells were settled, the excess medium was sucked off, cells were resuspended in bFGF medium and pipetted up and down for 5 times with a 10 ml pipette to break the colonies into smaller clumps of cells. Then cells were passaged 1:6 on the MEF plates in bFGF medium.

For RNA-Seq experiments, which were performed in collaboration with Dr. Dmitry Shaposhnikov, H9 cells were adapted to mTESR1 medium and cultured on MG coated plates in feeder-free conditions. Passaging was performed as for cells in bFGF medium, except that mTESR was used instead of bFGF and cells were plated on MG-coated plates instead of feeder layers.

HUES9 iCRISPR cells (Gonzalez et al., 2014)

HUES9 iCRISPR cells were maintained in mTESR1 medium and cultured on MG coated plates in feeder-free conditions, as this provided beneficial for further nucleofection

experiments. Passaging was performed as described for H9 cells.

Freezing

Freezing of cells was performed by addition of collagenase as described above, except cells were not plated on plates, but resuspended in freezing medium and transferred to cryotubes. Then, cells were cooled down in Mr. Frosty freezing containers (Thermo scientific) at -80°C. The freezing containers allow the cells to cool down for 1°C per minute, the optimal speed to conserve cells. After 24 hours cells were transferred to liquid nitrogen.

Thawing

Cells were taken from liquid nitrogen, quickly thawed in the water bath at 37°C, resuspended in culture medium, centrifuged and the supernatant was discarded. Cells were taken up in medium and plated in the respective medium, either on MEFs (for bFGF cultured cells) or on MG-coated dishes (for mTESR1 cultured cells).

BMP directed differentiation

H9 in bFGF

Cell culture plates were coated with MG for 30 min. Cells were dissociated to single cells by incubation in 0.25% Trypsin for 5 min at 37°C. Detached cells were collected in MEF medium and centrifuged. Afterwards, cells were resuspended in Diff medium 1 and plated at a density of 105 000 cells/cm^2. BMP4 was added fresh every day at a concentration of 50 ng/ml. Medium was changed daily during differentiation and cells were harvested after 2.5 days using 0.25% Trypsin as described above.

H9 in mTESR1

Cell culture plates were coated with MG for 30 min. H9 cells were dissociated to single cells by incubation with Gentle cell dissociation buffer for 10 min. Cells were plated in mTESR1 plus 10 µM ROCK inhibitor (Y-27632). On the next day cells were washed with PBS and Diff Medium 2 including freshly added BMP4 to a concentration of 50 ng/ml was added. Medium was changed daily. For time-course experiments cells were harvested at time-points 8 hours, 16 hours, 24 hours, 48 hours and 72 hours using PLB

buffer.

HUES9 iCRISPR:
BMP4 directed differentiation of HUES9 cells was performed as described for H9 cells cultured in bFGF medium, except that 10 µM ROCK inhibitor was added for the first 24 hours of differentiation.

Human Chorionic Gonadotropin (hCG) measurement

For hCG measurement cells were differentiated as described above and 1 ml of medium was collected per sample every 24 hours. For initial testing pregnancy test stripes were used following manufacturer's instructions. Samples were then frozen and stored at -80°C for further testing with the AccuLite® CLIA kit according to manufacturer's instructions.

Fluorescence activated cell sorting (FACS)

The following protocol was adjusted to 2.5 x 10^5 cells (¼ of a confluent 6 well plate). If more cells were used, for example for ChIP-Seq, the protocol was up-scaled accordingly.
For FACS cells were incubated with 1 ml 0.25% trypsin per 6 well until they dissociated as single cells. Trypsin was stopped by the addition of 3 ml MEF medium and cells were centrifuged. Cells were then resuspended in 1 ml FACS medium, distributed to 4 different tubes (1x antibody staining, 1x IgG staining and 2x for RNA isolation) and put on ice. Then the appropriate amount of primary antibody was added (Table IV-10) and stainings were performed for 30 min on ice. Afterwards cells were centrifuged at 2000 g for 5 min at 4°C. The supernatant was removed and the pellet was resuspended in 200 µl FACS medium containing secondary antibody (1:250). Cells were incubated in the dark for another 30 min on ice, centrifuged, the supernatant removed and resuspended in 250 µl FACS medium containing Propidium iodide (PI, 1:200). Cells were then filtered through a Falcon 40 µm cell strainer (BD) and analyzed on a FACS ARIA III (BD). For sorting, cells were gated using FSC-A against SSC-A according to morphology of the population and then doublets were excluded according to FSC-A against FSC-W. Next PI

was excited using the 561 nm yellow-green laser and detected using a 610/20 filter. Exclusion of the PI-positive cells diminishes dead cells in the population. Then the respective dye of the secondary antibody was detected using either the 633 nm red laser to excite Alexa Fluor 647 that was detected using a 660/20 filter, or the 488 nm blue laser to excite Alexa Fluor 488, where the fluorescence was detected via a 530/30 filter. For sorting the top and lowest 20% of the APA population were gated and sorted as APA+ and APA-, respectively. The primary SSEA-5 labeled antibody was detected using the 405 nm violet laser for excitation and the 450/40 filter for detection.
FACS data was then analyzed using FlowJo X 10.0.7r2.

Antibodies used for FACS:

Antibody	Vendor	Catalog Number	Lot Number	Concentration used
Purified Mouse Anti-Human CD249, Clone 2D3/APA	BD Biosciences	564532	4197560	1µg/2,5x10^5 cells or 1:200
Mouse Igg	eBioscience	16-4714-85	E034743	1µg/2,5x10^5 cells or 1:400
SSEA5 Pacific Blue	Homemade			5µl/2,5x10^5 cells
Alexa Fluor(R) 488 goat anti-mouse IgG (H+L) 2 mg/ml	Life Technologies	A11001	1726530	1:250
Alexa Fluor(R) 647 goat anti-mouse IgG (H+L) 2 mg/ml	Life Technologies	A21235	1383063	1:250

Table IV-10: *Antibodies used for FACS analysis*

RNA isolation

RNA was isolated using the RNeasy mini kit according to manufacturer's instructions. Isolated RNA was stored at -80°C thereafter.

Quantitative real-time PCR (RT-PCR)

Reverse transcription

RNA concentration was measured using the Nanodrop ND-1000 system. For all samples of an experiment similar RNA concentrations were reverse transcribed using the SuperScript® III First-Strand Synthesis System for RT-PCR according to manufacturer's instructions. cDNA was amplified using oligo dT primers.

RT-PCR

Both, SYBR green and Taq Man reactions were performed using the QuantStudio 12K Flex Real-Time PCR System (Thermo Fisher Scientific) in 384 well plates.

SYBR green

For SYBR green reactions the following mastermix and primers were used:

Volume in µl	Component
5	*Power* SYBR® Green PCR Master Mix
2.25	Primer mix [4 µM]
1	DNA
1.75	H_2O

Table IV-11: *Mastermix used for SYBR green assay*

Primers:

Name	**Forward**	**Reverse**
ELF5	CCTGATGTCGTGGACTGATCTG	GCTTAGTCCAGTATTCAGGGTGG
GAPDH	GAGTCAACGGATTTGGTCGT	ATGACAAGCTTCCCGTTCTC
STS	GCTGGCAAAAGTCAACACGGAG	GTCCGATGTGAAGTAGATGAGGG

Table IV-12: *Primers used for SYBR green assay*

TaqMan

TaqMan reactions were performed using the following mastermix and Taq Man assays:

Volume in µl	Component
5	TaqMan® Gene Expression Master Mix
0.5	TaqMan Probe
1	DNA
3.5	H_2O

Table IV-13: *Mastermix used for TaqMan assay*

TaqMan Assays:

Gene	Assay
ANKRD1	Hs00923599_m1
CD13	Hs00174265_m1
CDX2	Hs01078080_m1
ENPEP	Hs00157366_m1
GAPDH	Hs02758991_g1
GATA2	Hs00231119_m1
GATA3	Hs00231122_m1
GCM1	Hs00961601_m1
MESP1	Hs00251489_m1
POU5f1	Hs01895061_u1
ROR2	Hs00171695_m1
T	Hs00610080_m1
TFAP2A	Hs01029413_m1
TFAP2C	Hs00231476_m1
TP63	Hs00978343_m1
VGLL1	Hs00212387_m1

Table IV-14: *Probes used for TaqMan assay*

Microarray

Microarray was performed using the GeneChip Human Gene 2.0 ST Arrays (Affymetrix). They were hybridized and scanned according to manufacturer's instructions.

Chromatin Immunoprecipitation (ChIP)

Histone modification ChIP:

Antibody testing

Antibodies were tested for specificity using the AbSurance Histone H3 Antibody Specificity Array according to manufacturer's instructions. Briefly, the membrane was rehydrated in methanol and blocked using 5% milk powder in TBS-T for 1 hour at room temperature. Afterwards the antibody was added to the blocking solution and the

membrane was incubated at 4°C over night. The membrane was then washed three times with TBS-T for 5 min each. After incubation with secondary antibody for 1 hour at room temperature three washing steps, as described before, followed and the membrane was incubated for 1 min with Clarity Western ECL Substrate and imaged.

Further antibody testing was performed using ChIP-RT-PCR. Therefore regions enriched and depleted for the respective histone mark were determined using the UCSC genome browser and ENCODE data. Primers were designed in these regions using Primer3 Plus (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi) to span around 80-150bp. (Table IV-15) These primers were then used to test the specificity of the antibodies after the ChIP by RT-PCR using the *SYBR* green mastermix as described above. For analysis of the enrichment or depletion the mean Ct of 2 technical replicates was used per primer set and compared to the adjusted input values (the Ct values of the input need to be adjusted according to the dilution factor) with the formula:

$$\% \text{ input} = 100 \; x \; 2^{(\text{Ct (adjusted Input)} - \text{Ct (IP)})}$$

Primers used:

Name	Forward	Reverse
ACTB	AACGGCAGAAGAGAGAACCA	AAGATGACCCAGGTGAGTGG
B2M	GAGGCTATCCAGCGTGAGTC	GAAGTCACGGAGCGAGAGAG
CCND1	TGAAGAATCCCTGGATGGAG	GCCTGGGGTGAGATACAAGA
ESR1	AGAAAGGCGGGCATTAACTT	GGCCTTGACTTTCATGGTGT
EVX2	CTGAGTCTTCGGGGTTTCAA	GTCAGCGGGAGAAAGAGTTG
GAPDH #1	AGTCCCCAGAAACAGGAGGT	AGAGCGCGAAAGGAAAGAA
GAPDH #2	CTCTCTCCCATCCCTTCTCC	GGGAAGAGGGGAAGCTGTAT
GAPDH #3	AGGCAACTAGGATGGTGTGG	TGGACTCCACGACGTACTCA
HOXD12	GGAAACCCTACACGAAGCAG	TCGCTGAGGTTCAGCCTATT
K4neg	CCAGGCAGATGAATGAGGAT	CCCTTCCAAGGCTCTCTTCT
NEUROG1	CTGCAGGTACCCCTGATCTC	AACTGCCCTTTCCTGAGTGA
SPERT	GCATTAGAAGCTGGGGTGAA	CCTTCTCTCTTGCCCATCTG

Table IV-15: *Primers used for ChIP-RT-PCR*

Sample preparation:

Cells were dissociated from three confluent 10 cm plates using 0.25% trypsin as described before. The pellet was resuspended in 10 ml cold DMEM/F12 and 1 ml of freshly prepared crosslinking solution, consisting of 50 mM Hepes/KOH, 100 mM sodium chloride, 1 mM EDTA, 0.5 mM EGTA, 11% Formaldehyde (diluted from 16%

Formaldehyde) was added. This results in a 1% formaldehyde concentration. After 10 min on the shaker the crosslinking was stopped by the addition of 550 µl 2.5 M glycine. Cells were centrifuged at 2000 g for 5 min at 4°C. The pellet was resuspended in 5 ml ice cold PBS with 50 µl 100 mM PMSF (in ethanol) and centrifuged as before. Then the pellet was resuspended in cold FACS medium and stained as mentioned above. After sorting $1x10^6$ cells per tube cells were washed once more with PBS+PMSF and further flash frozen in liquid nitrogen. Cells were stored at -80°C until ChIP was performed.

ChIP:

All steps were performed on ice unless noted otherwise. For each ChIP reaction 2.5×10^5 cells were used.

Pre-clearing beads preparation:

For 10 pre-clearing reactions 200 µl Dynabeads(R) Protein A for Immunoprecipitation were put on a MagnaRack™ Magnetic Separation Rack (Thermo Fisher Scientific) and excess liquid was discarded. The beads were resuspended in 1 ml TE buffer and for blocking purposes 100 µl tRNA (10 mg/ml) were denaturated for 5 min at 95°C and added with 4 µl of rabbit IgG isotype control (5 µg/µl) to the beads. This mix was then rotated at 4°C over night and then washed three times with 1 ml wash buffer 1 (WB1, 50 mM Tris pH 8.8, 0.1% SDS, 0.1% Na-Deoxycholate, 1% Triton X100, 150 mM sodium chloride, 1 mM EDTA, 0.5 mM EGTA) for 5 min at a rotating wheel at 4°C. Afterwards beads were resuspended in 200 µl TE buffer.

Cell lysis:

Cell pellets (1×10^6 cells) were taken from -80°C and thawed on ice for 45 - 60 min. 500 µl Lysis buffer 1 (50 mM Hepes, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100) were supplemented with protease inhibitor (1X) and the pellet was resuspended by pipetting up and down. Cells were then rotated vertically for 10 min at 4°C. Afterwards cells were centrifuged at 2000 g for 5 min at 4°C. The pellet was resuspended in 500 µl Lysis buffer 2 (LB2, 10 mM Tris, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) supplemented with protease inhibitor (1X) and again rotated and centrifuged as before. Afterwards the pellet was resuspended in 120 µl Lysis buffer 3 (LB3, 10 mM Tris, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate,

0.5% N-lauroylsarcosine, 0.1% SDS) and transferred into microTUBE AFA Fiber Pre-Slit Snap-Cap 6x16mm for sonication.

Sonication:

Sonication was performed in 120 µl LB3 in microTUBE AFA Fiber Pre-Slit Snap-Cap 6x16mm with a Covaris E220 Focused-ultrasonicator with the following parameters: Peak Incident Power = 105, Duty Factor = 10%, Cycles per Burst = 200, Treatment time (s) = 480 at 4°C.

IP:

Chromatin was diluted with 900 µl dilution buffer (50 mM Tris pH 8.0, 167 mM sodium chloride, 1.1% Triton X100 and 0.11% Na-Deoxycholate) and a 50 µl aliquot (equals 5%) was taken as input sample and frozen at -20°C. To the rest of the sample 80 µl beads for pre-clearing were added and the chromatin was rotated for 2 hours at 4°C. Then the supernatant was collected in a new tube and the pre-clearing beads were discarded. The pre-cleared chromatin was then aliquoted into 4 tubes, each aliquot containing chromatin from 2.5 x 10^5 cells. The preferred antibody was then added to each tube and the sample was incubated on the rotating wheel over night at 4°C. In parallel, for each IP reaction 20 µl Dynabeads(R) Protein A for Immunoprecipitation per sample were blocked with 10 µl BSA (10 mg/ml) and 20 µl denaturated tRNA. The beads where filled up to 1 ml with TE buffer and put on a rotating wheel over night at 4°C. Next day the beads were washed three times with WB1 as described for the pre-clearing beads, resuspended in 20 µl TE buffer per IP and then the chromatin was rotated with the beads for 3 hours at 4°C. Next, two washing steps were performed using WB1 followed by one washing step with wash buffer 2 (WB2, 50 mM Tris pH 8.0, 0.1% SDS, 0.1% Na-Deoxycholate, 1%, Triton X100, 500 mM sodium chloride, 1 mM EDTA and 0.5 mM EGTA), one with wash buffer 3 (WB3, 50 mM Tris pH 8.0, 250 mM LiCl, 0.5% Na-Deoxycholate, 0.5% NP40, 1 mM EDTA and 0.5 mM EGTA) and two washing steps with wash buffer 4 (WB4, 50 mM Tris pH 8.0, 10 mM EDTA and 5 mM EGTA). The bound chromatin was then eluted from the beads by incubation of 130 µl elution buffer (EB, 1% SDS, 0.1 M NaHCO3) in a shaker, with full speed shaking for 15 min at 65°C. The liquid was collected and transferred to a new tube. Then the elution step was repeated with 100 µl EB and the sample collected in the same tube as before. At this step the input sample was thawed, filled up with EB to 230 µl and treated as the IP samples. All

samples were filled up with TE buffer to a volume of 300 µl. For digestion of RNA the chromatin was incubated with 3 µl RNAse A resulting in a final concentration of 200 µg/ml for 45 min at 37°C. Then 9 µl 5M sodium chloride and 3 µl Proteinase K were added and the sample decrosslinked in the shaker over night at 65°C. Shaking was on for 15 min every hour. The next day the DNA was purified using the Qiaquick PCR purification kit and eluted in 30 µl elution buffer supplied by the kit. This DNA was tested for enrichment of positive and negative regions as described above and then frozen at -20°C until library preparation.

Transcription Factor ChIP:

For ChIP of GATA2, GATA3, TFAP2A and TFAP2C 1 x 10^7 cells were used and treated according to the ChIP-IT high sensitivity kit.

Antibodies used

Name	Vendor	Catalog number	Lot number	Concentration used
H3K4me3 polyclonal antibody – Premium	Diagenode	C15410003-50	A.5051-001P	1 µg/IP
H3K27me3 polyclonal antibody – Classic	Diagenode	C15410069	A1821D	1 µg/IP
GATA2 antibody (H-116)	Santa Cruz	sc-9008 X	B0514	5 µg/IP
Purified Mouse anti-GATA3, clone L50-823	BD Biosciences	558686	509987	5 µg/IP
AP-2α (TFAP2A) antibody (C-18)	Santa Cruz	sc-184X	J1310	5 µg/IP
AP-2γ (TFAP2C) antibody (6E4/4)	Santa Cruz	sc-12762X	H2012	5 µg/IP

Table IV-16: *Antibodies used for ChIP-Seq experiments*

Next generation sequencing

ChIP-Seq

Library preparation for ChIP-Seq was performed using the NEBNext ChIP-Seq Library

Prep Reagent Set for Illumina and the NEBNext® Multiplex Oligos for Illumina® (Index Primer Set 1) according to manufacturer's instruction with 15 cycles of PCR. The libraries where then multiplexed and sequenced using the HiSeq® Rapid SBS Kit v2 (50 Cycle) and HiSeq® Rapid SR Cluster Kit v2 on a HiSeq 2500 (Illumina).

RNA-Seq:

The following paragraph was submitted to Developmental Cell for publication as Krendl et al. 2017:
3 µg of RNA were treated with TURBO DNase (Life Technologies, am2238) according to manufacturer's instructions followed by Rneasy MinElute RNA cleanup kit (Qiagen, 74204). Microcapillary electrophoresis on Agilent 2100 Bioanalyzer with RNA Pico 6000 kit (Agilent, 5067-1513) was used to analyze RNA quality (RIN values >8). Per RNA-seq library, 1 µg of DNAse-treated RNA was treated with RiboZero Gold (Human/Mouse/Rat) kit (Illumina, RS-122-2301) to remove rRNAs, followed by RNA cleanup using the Rneasy MinElute RNA cleanup kit. Sequencing libraries were prepared using TruSeq Stranded total RNA LT kit (Illumina, RS-122-2301) according to manufacturer's instructions using 11 cycles of PCR followed by purification with Agencourt Ampure XP beads (Beckman-Coulter, A63881). Libraries were evaluated on an Agilent 2100 Bioanalyzer using the DNA 1000 kit (Agilent, 5067-1504). DNA concentration was measured using a Qubit dsDNA HS Assay Kit (Life Technologies, Q32854). Samples were sequenced using a NextSeq 500 instrument to generate 75-nt single-end reads, sequencing depth was 20–40 Mio reads per library. Multiplexing of libraries was performed according to manufacturer's instructions.

Quantification of libraries

The concentration of the libraries was measured using the three methods described below. The molarity was calculated automatically for the Bioanalyzer method, for the two other methods it was calculated using the following formula:

$$\text{nM} = \frac{\text{Concentration}\left(\frac{\text{ng}}{\mu\text{l}}\right) x\ 1000000}{Fragment\ size\ x\ 650}$$

Then the mean value of the three different measurements was determined and libraries

diluted stepwise to 2 nM.

Bioanalyzer

The fragment size and the molarity of the sample was determined using the High Sensitivity DNA Analysis Kit on a 2100 Bioanalyzer instrument (Agilent) according to manufacturer's instructions.

Pico green concentration measurement

The DNA concentration was measured using the Quant-iT PicoGreen dsDNA Assay Kit with a Safire 2 multimode microplate reader (Tecan) according to manufacturer's instructions.

Qubit concentration measurement

DNA concentration measurement was performed using the Qubit dsDNA HS Assay Kit on a Qubit 2.0 Flourometer according to manufacturer's instructions.

DNA methylation analysis

In order to identify differences in DNA methylation levels on a genome wide level, genomic DNA from sorted populations was isolated using the Wizard DNA isolation kit and bisulfite conversion was performed using the EZ DNA Methylation™ Kit. Three replicates of each population were then analyzed with the Infinium HumanMethylation450 BeadChip kit according to manufacturer's instructions. Data analysis was performed using the RnBeads R package (http://rnbeads.mpi-inf.mpg.de).

Immunofluorescent Microscopy

The following paragraph was submitted to Developmental Cell for publication as Krendl et al. 2017:

Cells grown in 8-well chamber slides coated with MG were fixed with 4% formaldehyde/PBS for 15 min at room temperature and permeabilized with 0.3% Triton X-100/5% BSA/PBS for 30 min at room temperature. Primary and secondary antibodies were diluted to working concentrations in 1% BSA/0.1% Triton X-100/PBS. Primary antibody was incubated overnight at 4°C, secondary for 1 hour at room temperature. Specimens were washed with PBS containing DAPI before applying mounting medium. Images were obtained using a Zeiss Axiovert 200M epifluorescent microscope.

Antibodies used:

Name	Vendor	Catalog number	Lot number	Concentration used
GATA-3	Biocare Medical	CM 405 A	031915	1:200
AP-2α (TFAP2A) antibody (C-18)	Santa Cruz	sc-184X	J1310	1:200
AP-2γ (TFAP2C) antibody (6E4/4)	Santa Cruz	sc-12762X	H2012	1:200
Alexa Fluor(R) 488 goat anti-mouse IgG (H+L) 2 mg/ml	Life Technologies	A11001	1726530	1:1000

Table IV-17: *Antibodies used for immunofluorescent microscopy*

Western blot

The following paragraph was submitted to Developmental Cell for publication as Krendl et al. 2017:

Cells were trypsinized and lysed using RIPA buffer, containing phosphatase and protease inhibitors. After addition of 2x SDS loading buffer supplemented with 10% 2-Mercaptoethanol samples were heated at 95°C for 5 min. Samples were run on Mini-PROTEAN TGX Stain Free Gels, 4-15% and blotted using the Mini Trans-Blot Cell (Bio-Rad Laboratories). Following 3 x 5 min washing steps with TBS-T, membranes were blocked with 5% milk powder in TBS-T. Membranes were then incubated over night at 4°C with 5% milk powder in TBS-T containing the primary antibody. After 3 x 5 min

TBS-T washing steps the membrane was incubated with goat anti-mouse IgM-HRP in 5% milk powder in TBS-T. Following 4 washing steps, 15 min each, with TBS-T the membrane was incubated for 1 min with Clarity Western ECL Substrate and imaged with ChemiDoc™ MP System (Bio-Rad Laboratories).

Buffers used:

RIPA buffer	**2x SDS loading buffer**	**1x SDS running buffer**	**1x wet transfer buffer**	**TBS-T**
50 mM Tris HCL pH 7.5-8	120 mM Tris HCl pH 6.8	2.5 mM Tris base	2.5 mM Tris base	2 mM Tris base
150 mM NaCl	4% SDS	19.2 mM Glycine	19.2 mM Glycine	15 mM NaCl
1% Trition X-100	20% Glycerol	0.01% SDS		0.1% Tween 20
0.5% Na deoxycholate	0.02 % bromophenol blue			Adjust pH to 7.5
0.1% SDS				

Table IV-18: *Buffers used for western blot*

Antibodies used:

Name	**Vendor**	**Catalog number**	**Lot number**	**Concentration used**
Purified Mouse anti-GATA3, clone L50-823	BD Biosciences	558686	509987	1:700
β-Actin (8H10D10) Mouse mAb	Cell signaling	3700		1:5000
goat anti-mouse IgM-HRP	Santa cruz	sc-2064		1:10000

Table IV-19: *Antibodies used for western blot*

CRISPR/Cas9 gene editing

Guide RNA design

Guide RNAs (gRNAs) were designed using the MIT CRISPR design webpage

(http://crispr.mit.edu) according to Ran et al (Ran et al., 2013) instructions. For the GATA3 genome editing experiment two pairs of gRNAs were designed, flanking a region at the first exon/intron boundary, in order to cut out this genomic region. The following gRNAs were used for this experiment:

Name	Target	Sequence
GATA3_gRNA	GATA3	CACCGTACTGCGCCGCGTCCATGT
GATA3_gRNA_rev_comp	GATA3	AAACACATGGACGCGGCGCAGTAC
GATA3_gRNA_2	GATA3	CACCGACACTCTCGCGACGAGCCAG
GATA3_gRNA_rev_comp_2	GATA3	AAACCTGGCTCGTCGCGAGAGTGTC

Table IV-20: *gRNA sequences used for genome editing experiments*

Vector construction

gRNAs were resuspended in TE buffer to a final concentration to 1 µg/µl and 1 µl of each gRNA and reverse compliment fragment were mixed in 100 µl TE buffer. They were denatured in a Mastercycler nexus (Eppendorf) for 5 min at 100°C. Then the temperature was lowered for 5°C/minute until 25°C were reached. This allows annealing of the two complementary gRNA fragments. The annealed gRNAs were then ligated into a pBS U6 Vector with the following reaction:

Reagent	Amount
pBS U6 vector	50 ng
gRNA mix	4 µl
H_2O	5 µl
2x Quick Ligation Reaction Buffer	10 µl
Quick T4 DNA Ligase	1 µl

Table IV-21: *Composition of the ligation reaction of pBS U6 vector and gRNA*

This mix was then incubated for 5 min at room temperature and further transformed into 5-alpha Competent E. coli, which where first thawed on ice for 30 min. For transformation 1 µl of the ligation was mixed with 50 µl bacteria and incubated for 30 min on ice. After this they were incubated for 30 seconds at 42°C, put on ice for 2 min and then supplemented with 550 µl SOC medium. The mix was incubated at a thermomixer at 500 rpm shaking for 45 min at 37°C and 50 µl were then plated on a pre-warmed Agar plate containing 100 µg/ml ampicillin (AMP) and incubated at 37°C over night. The next day colonies were picked and incubated in 3 ml LB medium

containing AMP (100 µg/ml) and again incubated at 37°C over night. The DNA of this mix was then isolated used the GeneJET Plasmid Miniprep Kit according to manufacturer's instructions and sent for Sanger sequencing using the U6 forward primer to GATC Biotech. Sanger sequencing results were analyzed for proper integration of the right gRNA using the Sequencher 5.1 software. When sequences where right the leftovers of the miniprep were incubated in 100 ml LB Medium containing 100 µg/ml AMP and incubated over night at 37°C. Maxi prep was performed the next day using the PureLink HiPure Plasmid FP Maxiprep Kit according to manufacturer's instructions.

Nucleofection

For genetic modification experiments the HUES9 iCRISPR cell line was used and cultured as described above. For nucleofection a 10 cm plate of 70-80% confluent cells was dissociated into single cells using 3 ml Accutase. After dissociation the Accutase was stopped using 3 ml mTESR1 and cells were centrifuged. Cells were resuspended and counted using the Neubauer cell counter. 1 x 10^6 cells were nucleofected with 6 µg plasmid (3 µg per target sequence) on a 4D Nucleofector™ System (Lonza) using the P3 Primary Cell 4D-Nucleofector X Kit. Cells were re-plated on a MG coated 6 well plate in pre-warmed mTESR and doxycycline hydrochloride (Dox) was added to 1 µg/ml. Additionally ROCK inhibitor was added to a concentration of 10 µM for the first 24 hours. After this period the cells were cultured for two days in mTESR1 and Dox. Thereafter they were cultured as described before.

Identification of genetically edited clones

When cells on the 6 well reached confluency they were passaged as single cells using Accutase and plated on MG-coated 10 cm plates in a low density (1:20 - 1:50) in order to obtain single cell colonies. For the first 24 hours ROCK inhibitor was added as described above. The rest of the cells were used for DNA extraction (explained below) and testing of the genome editing efficiency by PCR (explained in the next paragraph). gRNA combinations that showed a cut were used for single clone picking.

After clones reached a size that was sufficient for picking, they were picked using the tip of a 200 µl pipette and put into one well of a 96 well plate. Colonies were then broken up and 1/10 of it was taken for direct DNA isolation.

DNA isolation

The DNA isolation was performed using the QuickExtract DNA Extraction Solution 1.0 according to manufacturer's instructions. Briefly, 50 µl of solution was added to 20 µl media containing the cells, vortexed for 15 seconds and then incubated for 6 min at 65°C. After this cells were vortexed again for 15 seconds and incubated for 2 min at 98°C.

PCR

To detect the mutated genomic DNA, I designed primers flanking the site of excision using Primer 3 plus. Initially primers were tested on genomic DNA using a gradient PCR with an annealing temperature from 60°C to 70°C. The results were then detected on an agarose gel in Lithium Acetate Borate (LAB) buffer (10mM Lithium Acetate and 10mM Boric Acid) including Sybr Safe (1:20000). The percent of the agarose gel used depended on the size of the fragment that should be detected. Fragments smaller than 1500 bp were analyzed on a 1,5% agarose gel, whereas fragments between 1500 and 5000 bp were analyzed on a 1% agarose gel. This testing resulted in the following conditions for GATA3:

PCR Mastermix	Volume(µl) for 1 reaction for bulk	Volume(µl) for 1 reaction for picked colonies
Taq PCR Master Mix	10	10
4µM Primer mix	2	2
DNA	2	6
H_2O	6	2
Total volume	20	20

PCR program	Temperature in °C	Time in min
	94	3
45x	94	0.5
	68.2	1
	72	1
	72	10
	10	hold

Table IV-22: *PCR mastermix and program used for detection of GATA3 mutated clones*

Primers used

Target	Forward primer	Reverse primer
GATA3	CGAGGCCATGGAGGTGACGG	CCCCCACCCCAAAACCTGCA

Table IV-23: *Primers used for detection of GATA3 mutated clones*

Data analysis

Data analysis was performed in collaborations with Dr. Steffen Sass (Institute of Computational Biology, Helmholtz Center Munich, Munich, Germany) and Dr. Tobias Straub (Biomedical Center Munich, ZFP and Bioinformatic Unit, Planegg-Martinsried, Germany)

The following paragraphs were submitted to Developmental Cell for publication as Krendl et al. 2017:

Microarray

CEL files were processed and normalized using the affy package within the R framework for statistical computing. Normalization was performed using robust multiarray average (RMA). Genes with low variation or low signal were removed by applying a filtering approach that is implemented in the genefilter package (Bourgon et al., 2010). Differential gene expression was determined using limma (Smyth, 2004). P-values were corrected for multiple testing using the Benjamini and Hochberg. A gene was considered to be differentially expressed if its corrected p-value was below a threshold of 0.05. The resulting gene sets were subject to literature-based tissue enrichment analysis using the GeneRanker tool within the Genomatix Software Suite.

ChIP-Seq read processing

Sequencing reads were aligned to version hg19 of the human genome using bowtie (Langmead et al., 2009) version 1.1.1 allowing only for single matches to the reference (parameter –m 1). We extended the matched reads to a total of 200 bp and calculated for each sample a per-base genomic coverage vector by cumulating the total spans of all sequenced fragments.

Histone modification enrichment at promoters

Per gene for each 4 kb window the number of matching reads in each replicate sample were calculated. Raw read counts were transformed by $reads.transformed_i = arcsin\left(\sqrt{\frac{reads_i}{total\ reads}}\right)$, where i = promoter window, reads = number of reads covering i. Signal enrichment was analyzed by subtracting the normalized reads of the corresponding input samples from the IP read values. Enrichments across replicates of each histone modification were adjusted by quantile normalization and values were averaged. In case of many promoters-to-one gene relationships we selected the promoter with highest interquartile range of signals across all experimental conditions.

Peak calling and definition of robust peak sets

Transcription factor peaks were called using Homer (Heinz et al., 2010) findPeaks (v 4.7.2) with parameters style=factor, size=200, fragLength=200, inputFragLength=200 and C=0. All peaks were called using the corresponding input samples as control. We defined peaks as robust if the region was called in at least two biologically replicated samples except for TFAP2C, for which only one high quality data set was obtained.

De novo motif discovery and genome-wide motif searches

We search for enriched motifs in peak regions using MEME (Bailey et al., 2009) using the zero or one occurrence per sequence ("zoops") model.

RNA-Seq

RNA-Seq reads were mapped to the hg38 genome using TopHat (Trapnell et al., 2009). The resulting alignments were overlaid with the UCSC "known gene" track (Kent et al., 2002) to obtain exon read coverage for every annotated gene. Genes with a maximum read count of zero across all samples were considered to be non-expressed and thus removed from further analysis. The count data was then pre-processed using voom (Law et al., 2014) including quantile normalization between the individual samples. Differential gene expression between time points was then determined similarly to the

microarray data. For illustration purposes data was normalized using the igv tools option normalize coverage data.

Time series of selected TFs were clustered according to the pairwise Pearson correlation coefficients using k-means clustering analysis with k=6 clusters. Fisher's exact test was employed to test whether a set of genes with a certain histone modification was significantly overrepresented in one of the resulting clusters.

V. Results

Analysis of the lineage correspondence and heterogeneity of progeny generated from human ESCs by BMP4

There is still controversy whether the trophoblast-like progeny that emerges from human ESCs by BMP4 treatment descends from mesoderm or trophoblast precursors (Aims section "To resolve the uncertainty concerning the lineage correspondence of human PSC-derived trophoblast progeny"). Furthermore the degree of heterogeneity in this progeny population is not clear. To address these questions I treated human ESCs plated as single cells with BMP4, a regimen that promotes trophoblast differentiation (Introduction section" Transcriptional processes in response to Bone Morphogenic Protein (BMP) 4 in human ESCs"). To determine the optimal time point for purifying the cells by sorting I first analyzed the expression of APA by flow cytometry (Figure V-1A). This showed only ~10% APA+ cells in the first two days, but at day 3 this population increased to around 60% of the cells, which then plateaued (Figure V-1B).

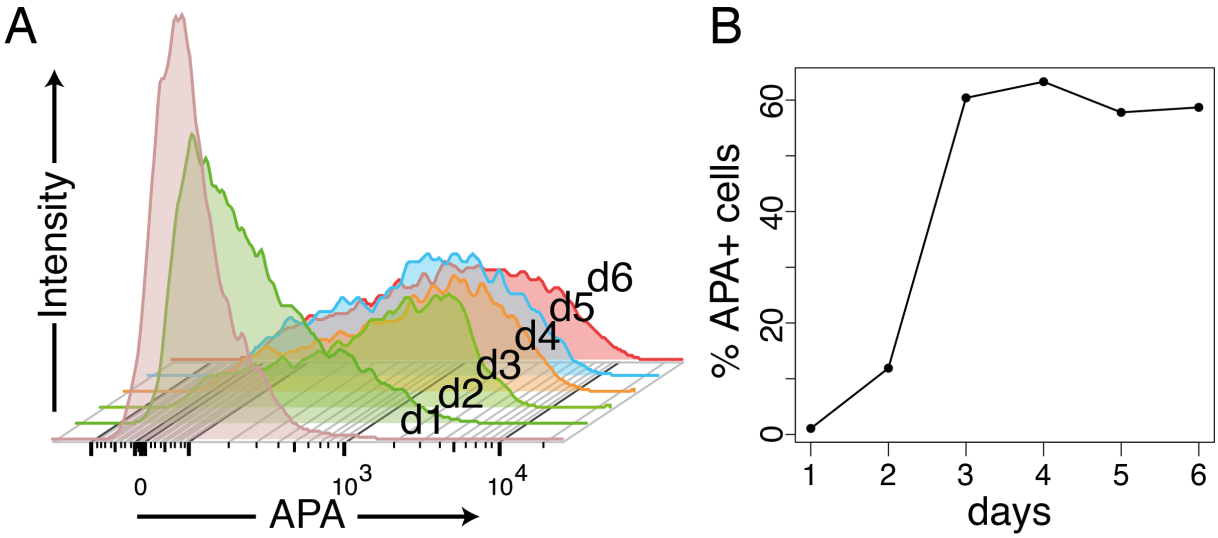

Figure V-1: *Time course analysis of APA+ cells*
A. Representative time-course flow cytometry measurements of the cell surface marker APA in BMP4-treated human ESCs over a period of 6 days.
B. A Diagram showing the percentage of the APA+ population over a period of 6 days.

Based on this analysis, I elected the time-point of 2.5 days for purifying the cells and analyzing their cell intrinsic properties (as earlier time-points do not yield sufficient cell amounts for genome-wide analyses). I sorted the top 20% APA+ and lowest 20% APA- cells for analyzing the corresponding trophoblast and non-trophoblast progenitor

population (Figure V-2A). Residual differentiated cells were eliminated from human ESCs cultures by staining the cells using an antibody specific for the surface marker SSEA-5 and sorting of the positive fraction (Tang et al., 2011).

An important characteristic of trophoblast or placenta cells is the expression of hormones, including the human chorionic gonadotropin (hCG) (Introduction section "Human placental development"). Therefore, to validate the trophoblast enrichment of the BMP-4 treated human ESC cultures, I tested the production of hCG initially using pregnancy "stick" tests and later using a quantitative immunoencymometric assay. My analysis, conducted at days 0, 2, 4, 6 and 8 confirmed trophoblast induction using BMP-4, as only cells of these cultures produced the hormone starting at day 4 and not cells treated with KSR medium only (Figure V-2B).

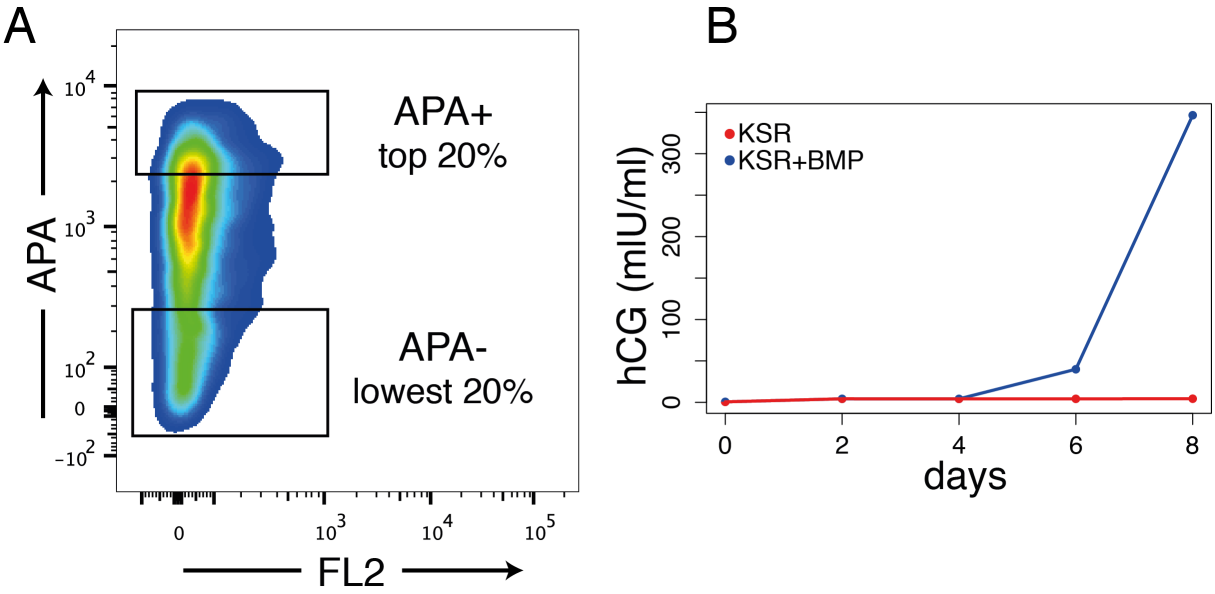

Figure V-2: *Flow cytometry analysis of APA and hormone production of BMP4 treated human ESCs.*
A. A representative flow cytometry analysis of the cell surface marker APA in 2.5 day BMP4-treated human ESCs. Gates used for sorting of top and bottom 20% of the population are shown. FL2 = Fluorescence 2
B. Time course analysis of hCG protein concentration produced by human ESCs treated with KSR based medium alone or KSR+BMP4 (n=2).

To address the lineage identity of the APA+ and APA- populations, I initially analyzed these populations using RT-PCR. Testing the expression of specific assays for pluripotency (*OCT4*), trophoblast (*CDX2, ELF5, ENPEP* and *GCM1*) and meso- or mesendoderm (*MESP1*, *T, CD13, ROR2* and *GSC*) genes showed that both, the APA+ and the APA- population compared to sorted SSEA-5+ cells exhibited a decrease of *OCT4* and an increase of trophoblast genes. Moreover, the expression of *ELF5*, *ENPEP* and *GCM1*

was higher in the APA+ population compared to the APA- population. In addition in the APA- population compared to the SSEA-5+ cells I observed an increase in meso- and mesendoderm gene expression (Figure V-3). These data provided me the confirmation necessary for conducting further analysis pertaining to the identity of the trophoblast progenitors and for mechanistic investigation.

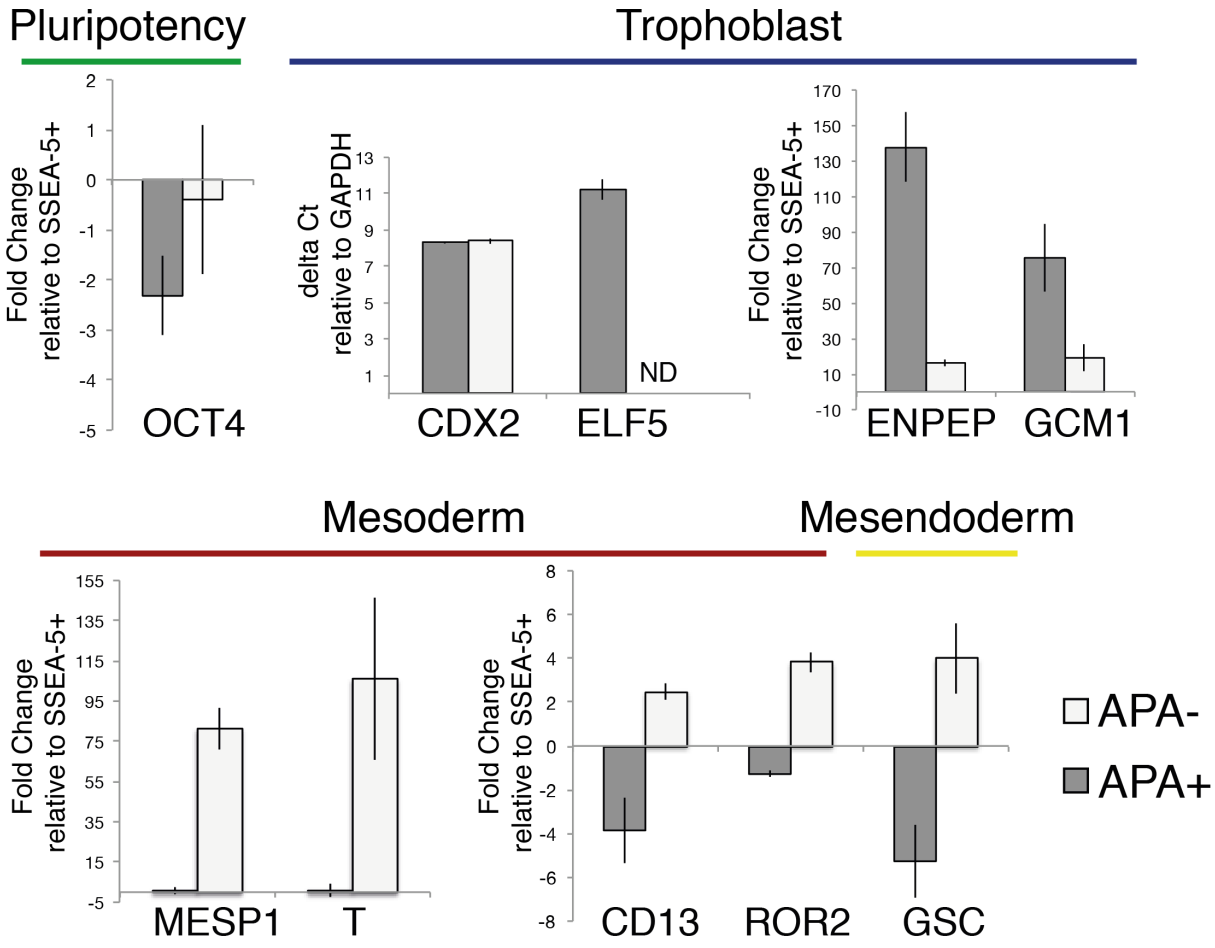

Figure V-3: *Targeted gene expression measurements in APA- and APA+ populations.* Quantitative PCR analysis of canonical developmental TFs and markers, involved in pluripotency, trophoblast, mesoderm and mesendoderm development, assayed in sorted APA-, APA+, SSEA-5+ cell populations. Expression amplitudes were normalized to the SSEA-5+ cell population, or delta Ct relative to GAPDH, if expression was not detected in the SSEA-5+ cell population. Standard error of the mean of two independent experiments (sorted from human ESCs of different passages) is shown. ND refers to not detected.

Next, I analyzed the same samples (three replicates of SSEA-5+, APA- and APA+ sorted cells) using GeneChip Human Gene 2.0 ST arrays (Affymetrix). This platform allows quantification of approximately 40,000 RefSeq transcripts, including nearly 25,000 genes and 11,000 lincRNAs. Analysis of these samples according to gene expression

patterns showed that all replicates of the respective populations clustered together (Figure V-4A). Furthermore, this showed that the APA- population is more similar to the SSEA-5+ population than the APA+ population to the SSEA-5+ population. By applying a two-fold expression change cutoff and a false discovery rate (FDR) cutoff of 5% between biological replicates I found around 720 decreased (Figure V-4B) and around 1000 increased (Figure V-4C) transcripts between all three conditions. I also noted that the proportion of transcripts differentially expressed between the APA+ vs. SSEA-5+ populations is much higher than between the APA- vs. SSEA-5+ populations. I presume the reason for this is because the replicates of the APA- population are heterogeneous, and therefore many genes do not reach the 5% FDR threshold.

This in depth transcriptomic approach corroborated and expanded the RT-PCR data. I found that the pluripotency downregulation also includes *SOX2* and *NANOG* next to *OCT4* (Figure V-4B). Furthermore, among the notable increased transcripts I identified a trophoblast specific signature, consisting of *GCM1, TP63, ENPEP, GATA3* and *TFAP2C* (Introduction section "The transcriptional network of TE development") Among these genes, I found *VGLL1*, whose homolog has been shown to be a co-factor for one of the primary TE formation factors in the mouse, *Tead4* (Nishioka et al., 2008), leading me to postulate that it also may play a role in human TE development.

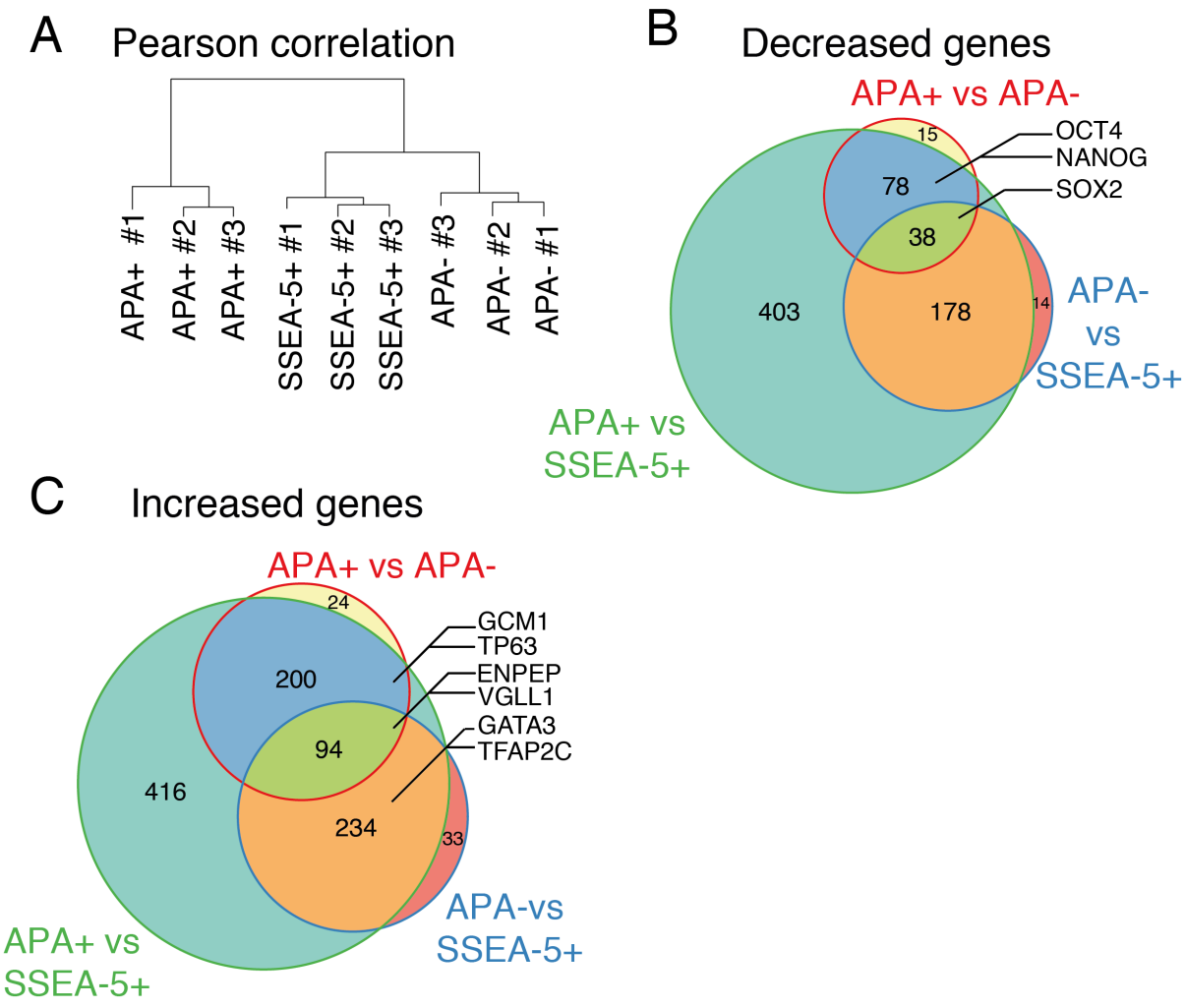

Figure V-4: *Microarray-based transcript expression analysis of APA-, APA+ and SSEA-5+ populations.*
A. An hierarchical clustering of Affymetrix GeneChip Human Gene 2.0 ST Array triplicate APA-, APA+ and SSEA-5+ cell population data calculated using Pearson correlation coefficients as measure of similarity. It shows sample clustering according to cell population identity.
B. - C. Venn diagrams exhibiting the number and overlap of decreased (B) and increased (C) gene transcripts in APA+ versus APA- populations and sorted SSEA-5+ undifferentiated human ESCs. Transcript measurements performed using Affymetrix Human Gene ST 2.0 arrays. Differentially expressed transcripts were analyzed by applying a fold change cut-off of 2 and a false discovery rate (FDR) of 5% between biological repetitions (n=3). Two folds higher number of differentially expressed genes was noted comparing the APA+ versus the SSEA-5+ cell population than the APA- versus the SSEA-5+ and the APA+ versus the APA- cell population, because fewer genes in the APA- population reached the FDR threshold of 5%. The positions of canonical pluripotency and trophoblast genes are indicated respectively in (B) and (C).

Next, I analyzed the genome-wide transcriptome data in the respective populations using Genomatix GO-term analysis of enriched tissue and cell type categories. Starting with the increased genes in the APA+ vs. the SSEA-5+ cell population identified placenta

and trophoblast associated categories among the highest enriched (exhibiting the lowest p-values) (Figure V-5A). In contrast, comparing enriched transcripts of the APA- vs. the SSEA-5+ cell populations, the most enriched tissues were of ectoderm, mesenchymal or embryonic structure categories (Figure V-5B). Finally, comparing the APA+ vs. the APA- increased transcripts we identified only uterus, trophoblast and placenta tissues among the top 7 enriched GO-terms (Figure V-5C).

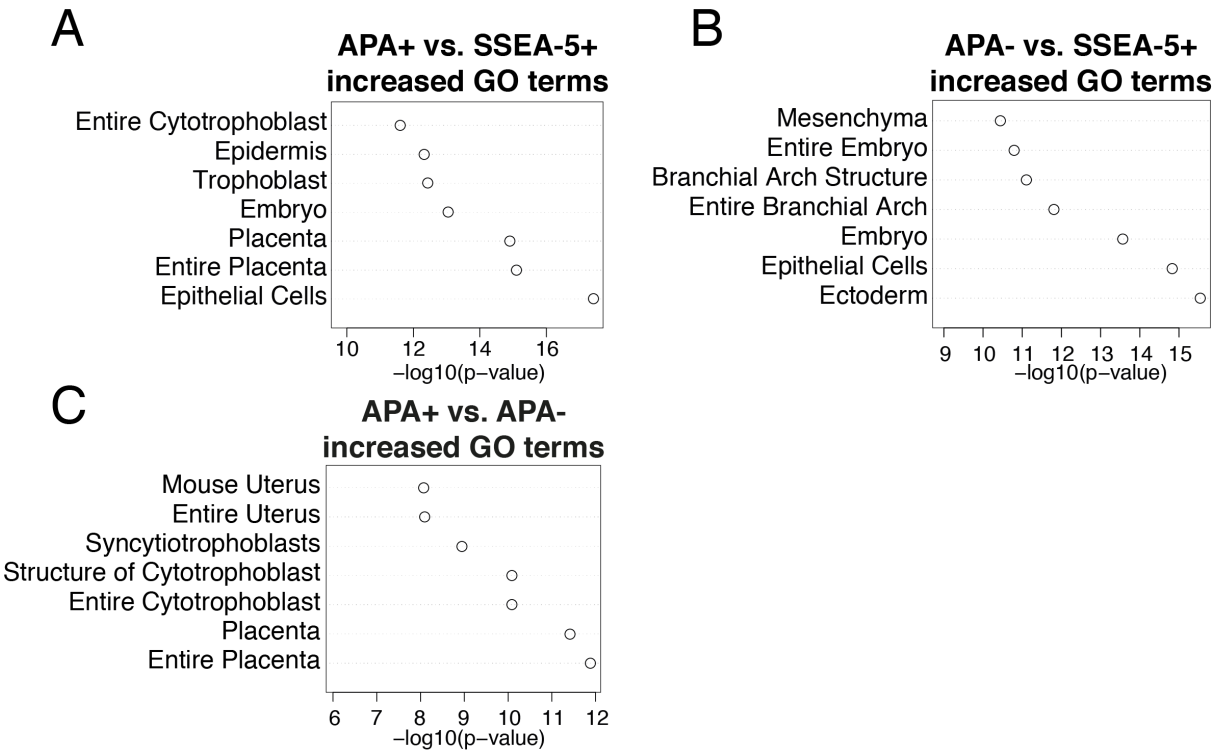

Figure V-5: *GO-term analysis of enriched tissues of APA-, APA+ and SSEA-5+ populations*
Genomatix GO-term analysis of enriched tissue and cell type categories applied to transcripts increased in the APA+ compared to SSEA-5+ (A), the APA- compared to SSEA-5+ (B) and the APA+ compared to the APA- (C) cell populations. (A) and (C) show enrichment for placenta, cytotrophoblast and others, while (B) shows enrichment for different – non-placenta- categories.

After establishing the lineage correspondence of the APA+ population to trophoblasts, I next analyzed the similarity to early embryonic human TE. I relied on a published dataset that compared TE cells isolated from human blastocysts to human ESCs (Bai et al., 2012). I found an overlap of 23 TFs, which underlie the classification of trophoblast/placental tissues, whereas the non-overlapping TFs were mostly enriched for other embryonic structures (Figure V-6). Taken together these data clarify cell/lineage identifies pertinent to my primary study objectives and questions: it strengthens the perspective that the trophoblast progeny of human ESCs is of a TE

origin, it indicates a mixed lineage embryonic character of the non-trophoblast population that emerges during early differentiation of human ESCs using BMP4, and it reveals the correspondence of human trophoblast progenitors to human embryonic TE. Further it corroborates the use of APA+ cells that emerge from human ESCs *in vitro* by treatment using BMP-4 as a differentiation system that enables to study differentiation mechanisms towards human extraembryonic tissues.

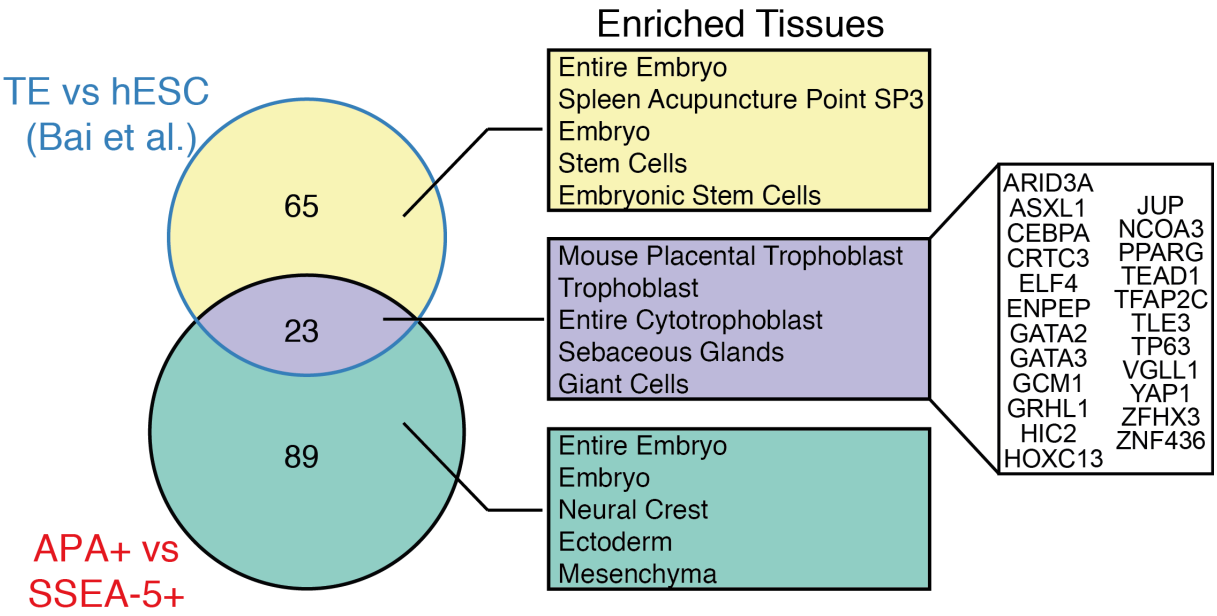

Figure V-6: *Comparison of in vitro generated trophoblasts to TE from human embryos*
A VENN diagram exhibiting the correspondence in upregulation of TFs, co-transcriptional regulators and ENPEP (APA) (extracted using the Genomatix bioinformatics tool) between a published dataset (Bai et al., 2012) (comparing TE versus human ESCs) and my list of genes in the APA+ versus the SSEA-5+ cell population. A 5 and 2 fold-change cut-off were applied to the published and my dataset, respectively. Middle pane shows Genomatix GO-term analysis of enriched tissues of the respective group of genes. Right pane shows the extracted TFs plus *ENPEP* from the group of overlapped genes.

Histone modification turnover during BMP4-mediated human ESC differentiation to trophoblast progenitors

Mechanisms of epigenetic regulation produce the phenotypic diversity of our cells and tissues that harbor the same genetic information. The histone composition and specific modifications of histones establish the grounds for differential gene expression among

cell types (Introduction section “Chromatin associated histone turnover in development and differentiation”). To gain understanding of the mechanisms that regulate the establishment of the trophoblast fate I therefore used my human ESC BMP4 differentiation system. I focused on identifying the locations of the activating H3K4me3 and the repressing H3K27me3 in the sorted SSEA-5+, APA- and APA+ cell populations using chromatin immunoprecipitation (ChIP) in conjunction to next generation sequencing (ChIP-Seq). The reason for analyzing these marks is to identify the reshuffling of histone modifications and to establish its correlation with gene expression changes during the differentiation process on a genome wide level.

Optimization of protocols and reagents

I first established a histone H3K4me3 and H3K27me3 ChIP protocol that is suitable for small cell populations that I could readily purify by FACS. I tested different protocols till the ChIP yielded the lowest background and highest enrichment for specific sites, and DNA fragments of around 150 and 500 bps for maximizing the resolution of the sequencing (Methods section “Chromatin Immunoprecipitation”).

In addition, I have tested the specificity of the ChIP antibodies, because binding to other histone modifications could produce false positive hits. I verified the specificity of the antibodies using a membrane blotted with modified histones. This showed that Diagenode antibodies exhibited the highest specificity (for example compare with membranes reacted with H3K27me3 antibodies from Millipore and Active Motif which exhibited unspecific binding to H3K4me3, H3K9me3 and Millipore also to H3K36me3 in Figure V-7).

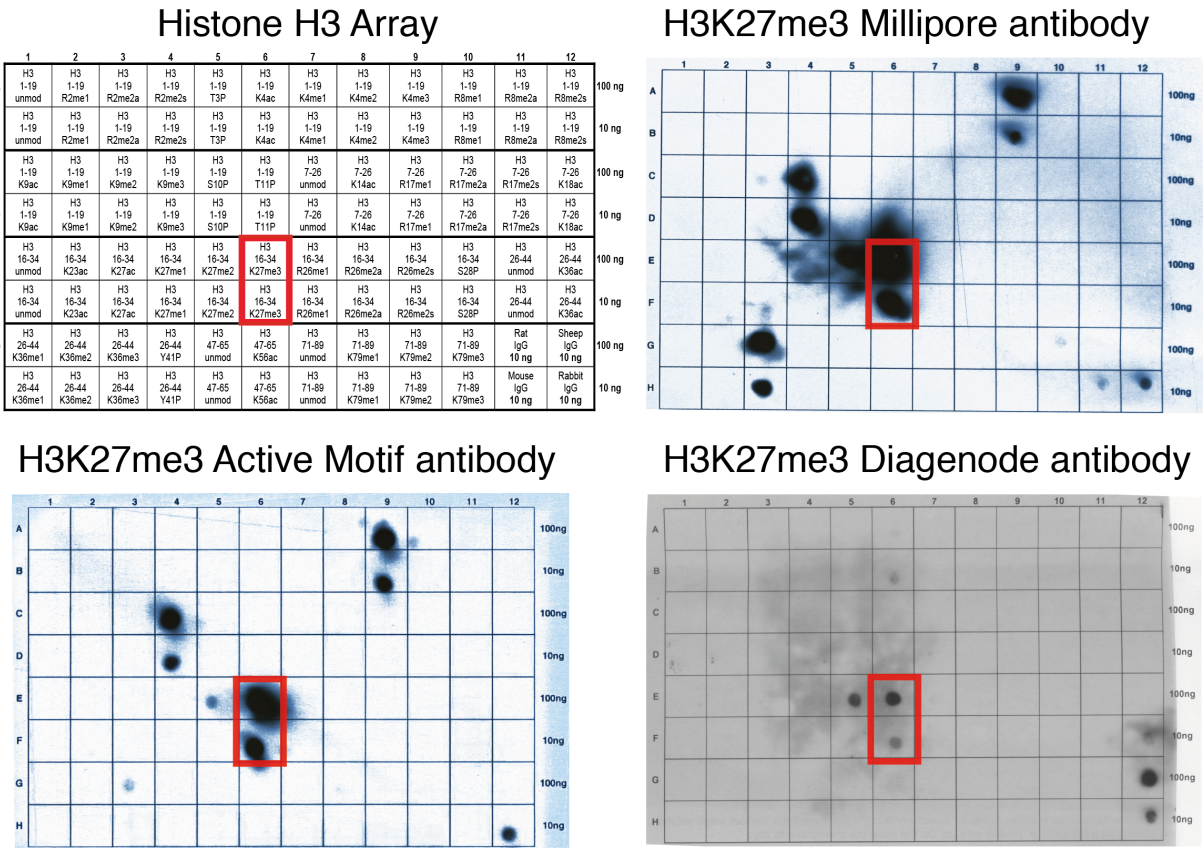

Figure V-7: *Testing of H3K27me3 antibody specificity for ChIP*
AbSurance Histone H3 Antibody Specificity Arrays reacted with Millipore, , Active Motif and Diagenode antibodies. The red rectangle represents the sector spotted with H3K27me3 histones. The signal in the lower right corner corresponds to an internal positive control that detects the secondary antibody.

Furthermore, I have tested the quality of the ChIP by RT-PCR using primers that were designed for genomic regions that harbor the respective H3K27me3 and H3K4me3 modifications in human ESCs and negative control regions where these marks were previously found to be absent. The percent of input DNA was used to calculate the enrichment or depletion of the histone marks. These experiments have validated that the Diagenode antibodies produced the highest enrichment (Figure V-8). Finally, I adjusted the amount of antibody that was optimal for the ChIP of 2.5 x 10^5 sorted cells.

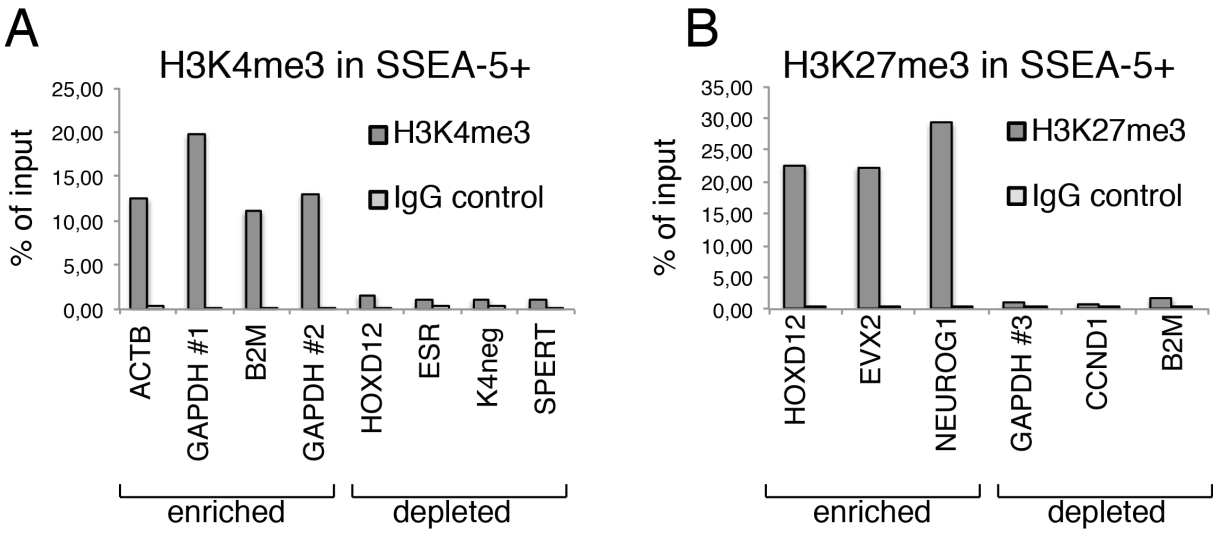

Figure V-8: *Testing of H3K4me3/K27me3 antibody specificity by RT-PCR*
A. RT-PCR analysis of regions within *ACTB*, *GAPDH*, *B2M*, *HOXD12*, *ESR*, and *SPERT* loci using template DNA derived from ChIP of sorted SSEA-5+ cells with H3K4me3 Diagenode premium polyclonal antibody (Cat. Nr. C15410003-50, LotNr: A.5051-001P). Matched IgG control was used, and percentage of enrichment over the input is shown.
B. RT-PCR analysis of regions within *HOXD12*, *EVX2*, *NEUROG1*, *GAPDH*, *CCND1*, and *B2M* loci using template DNA derived from ChIP of sorted SSEA-5+ cells with. H3K27me3 Diagenode standard polyclonal antibody (Cat. Nr. C15410069, LotNr: A1821D). Matched IgG control was used, and percentage of enrichment over the input is shown.

Global analysis of H3K4me3 and H3K27me3 positions during trophoblast specification

To analyze histone modification positions that are relevant to the transcriptome profiles of the sorted APA+, APA- and undifferentiated SSEA-5+ cell populations, I used the same preparations for ChIP-Seq as were used for microarray analysis. I analyzed the DNA fragments by Illumina HiSeq2500 instrument providing at least 20 million 50bp single end reads per sample, and mapped these to the hg19 human genome annotation (Methods section "ChIP-Seq read processing").

Mapping the reads to the human genome showed statistically significant positions of the H3K4me3 mark in narrow peaks around transcription start sides, and the H3K27me3 mark exhibited coverage over broader regions in general and in gene bodies. These results are consistent with the functions of H3K4me3 and H3K27me3 in gene activation and repression, respectively (Introduction section "Epigenetic profiles of PSCs").

Following these validations, I focused on the biological functions of the respective histone modifications to the process of trophoblast specification, and therefore analyzed the differentially expressed genes in the APA+ vs. the SSEA-5+ cell population in this respect.

Analyzing first the histone modification turnover of the increased genes, I discovered that more than half of these genes harbor the activating H3K4me3 mark in their respective promoter regions already in the SSEA-5+ undifferentiated cell population. In addition 22% of genes exhibit bivalency marks (H3K4me3 and H3K27me3 colocalization), and 18% do not harbor both marks (no Mod). Only a very small fraction, around 2%, of the increased genes, is marked by H3K27me3 in SSEA5+ cells (Figure V-9). In the SSEA-5+ to APA+ transition I registered a high degree of turnover in the bivalent genes to H3K4me3, and hardly any gene in the H3K4me3 monovalent group in SSEA5+ cells change their histone mark configuration during differentiation to APA+, which is expected because H3K4me3 marks have been found to be a key feature at active gene promoters. Conversely and unexpectedly, only a small fraction of the genes in the no Mod category change their histone mark configuration during differentiation to APA+ trophoblast progenitors. This is surprising because in many cases active transcription is accompanied by H3K4me3 marks in the promoter region of the respective gene.

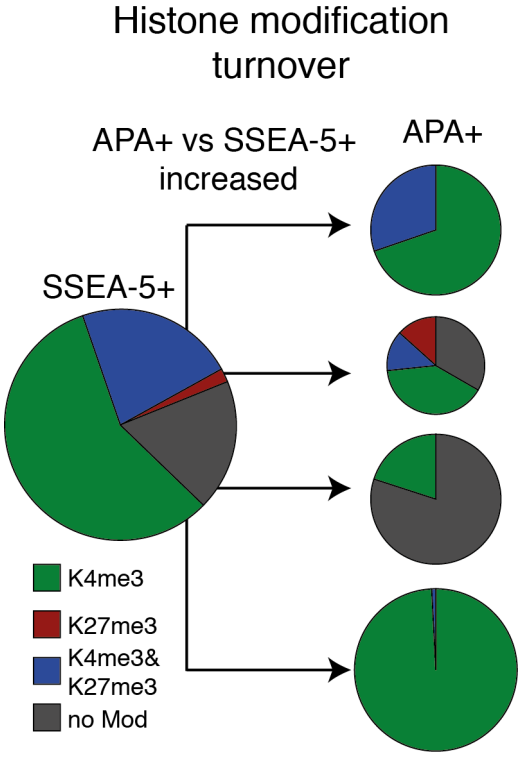

Figure V-9: *Histone modification turnover of increased transcripts in the APA+ vs. SSEA-5+ cell populations*
Transcripts were assigned to four histone modification classes, H3K27me3, H3K4me3, bivalent double positive and no modification (no Mod) based on enrichments of H3K4me3 and H3K27me3 signals in a 4 kb window around their TSS (n=3). Analysis included 870 increased genes from lists of differentially expressed transcripts in the APA+ versus the SSEA-5+ cell population. Presented are the changes in the histone classes between the SSEA-5+ and the APA+ cell population.

Analyzing the histone modification landscape of the decreased transcripts with respect to their configuration in SSEA-5+ undifferentiated cells revealed that the majority of the genes harbor monovalent H3K4me3 marks. Interestingly, although they were transcriptionally downregulated, most monovalent H3K4me3 genes maintain the mark, and this indicates that these genes are still active in APA+ cells, just at a lower level. Interestingly, we find the pluripotency gene *NANOG* in a small group of genes where the H3K4me3 monovalent mark is removed upon differentiation, and this indicates a distinct mode of reducing transcriptional activity (Figure V-10).

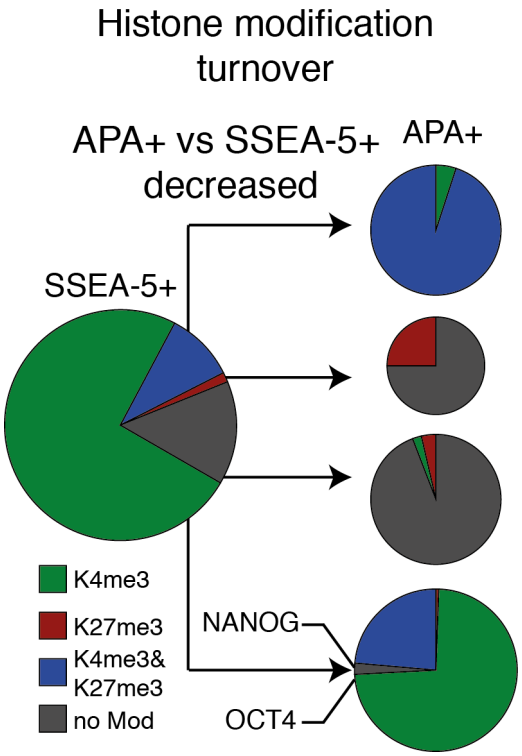

Figure V-10: *Histone modification turnover of decreased transcripts in the APA+ vs. SSEA-5+ cell population*
Transcripts were assigned to four histone modification classes, H3K27me3, H3K4me3, bivalent double positive and no modification (no Mod) based on enrichments of H3K4me3 and H3K27me3 signals in a 4 kb window around their TSS (n=3). Analysis included 592 decreased genes from lists of differentially expressed transcripts in the APA+ versus the SSEA-5+ cell population. Presented are the changes in the histone classes between the SSEA-5+ and the APA+ cell population. The position of pluripotency genes *NANOG* and *OCT4* is indicated.

H3K4me3 and H3K27me3 turnover in TFs during trophoblast specification

Because TFs and transcriptional co-factors are critical drivers of cell fate, I next focused the analysis on them for understanding the mechanisms of trophoblast progenitor specification. Interestingly, these fell into two categories containing classical early and late mouse / human trophoblast genes, e.g. *GATA3, TFAP2C* and *CDX2*, and *ELF5, GCM1, VGLL1* and *TP63*, which belonged to the bivalent and no-mod categories, respectively. Furthermore I noted an enrichment of placenta-associated non-TFs in this no Mod category (Figure V-11). This indicates that bivalent TFs in SSEA5+ cells could be activated faster than the ones carrying none of the H3K4me3 or H3K27me3 marks.

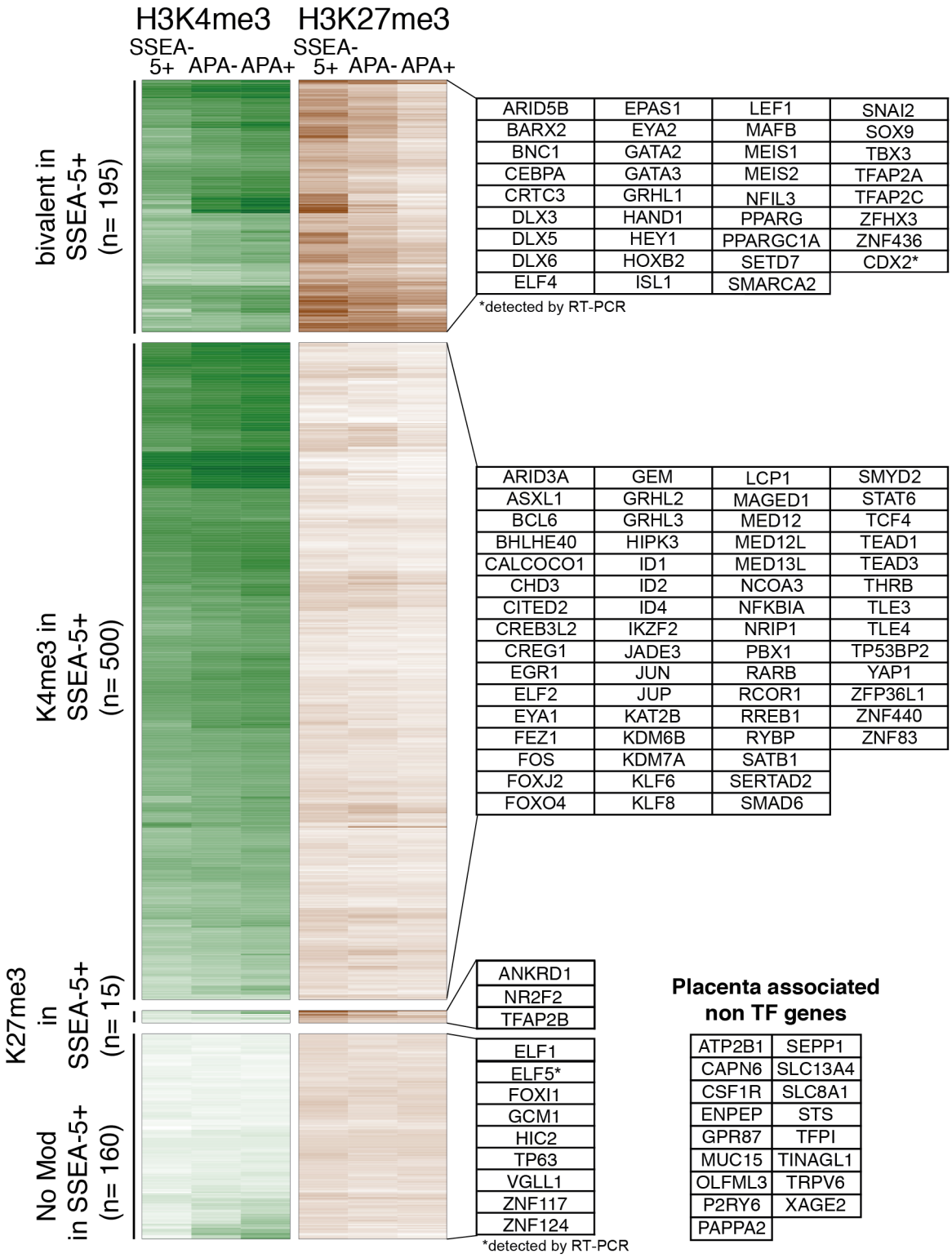

Figure V-11: *Histone modifications of APA+ enriched TFs*
A heatmap display of APA+ versus SSEA-5+ increased genes and their respective histone mark enrichments at gene promoters analyzed in APA-, APA+, and SSEA-5+ cell populations. Genes are grouped according to their histone modification configurations in SSEA-5+ cells. Right pane displays TFs isolated of the respective groups. Additional placenta associated genes (non TFs) of the no-mod category, identified via GO term analysis (Genomatix) are displayed.

Representative examples of bivalent (*GATA3*), H3K4me3 monovalent (*ARID3A*), H3K27me3 monovalent (*TFAP2B*) and no Mod (*GCM1*) genes in SSEA-5+ cells and their histone modification turn-over to the APA- and the APA+ population are shown in Figure V-12.

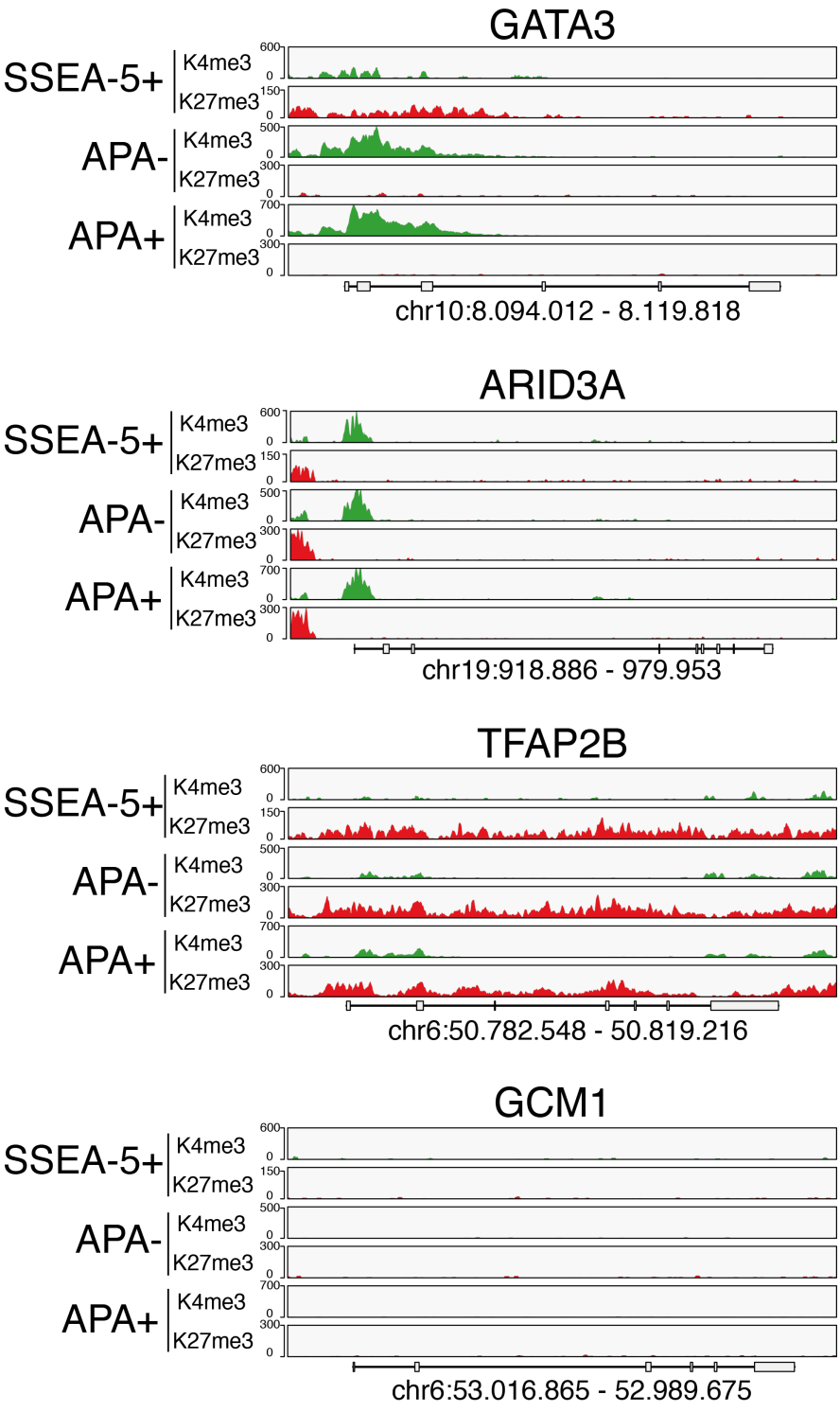

Figure V-12: *Representative histone mark maps of SSEA-5+, APA- and APA+ increased genes*
Representative histone mark maps of genes with characteristics of bivalent (*GATA3*), monovalent H3K4me3 (*ARID3A*) and H3K27me3 (*TFAP2B*) and modification deficient (*GCM1*) loci in SSEA-5+ cells. Histone coverage (y-scale) of one out of three experiments is displayed in SSEA-5+, APA- and APA+ cell populations.

For the downregulated TF *OCT4* a representative example of the histone mark positions in the three purified populations is shown in Figure V-13.

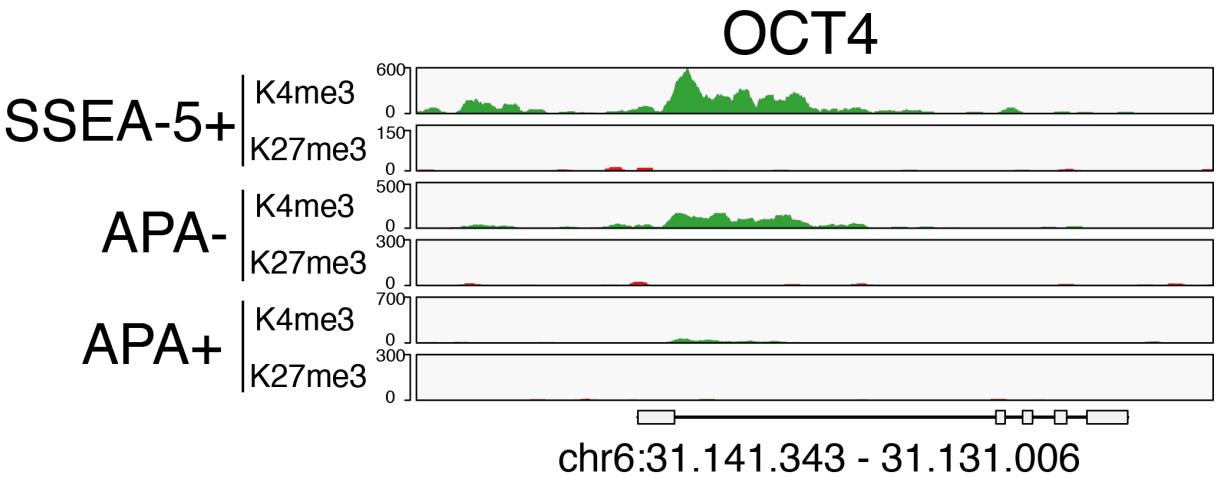

Figure V-13: *Turn over of histone marks in the OCT4 gene*
Histone coverage (y-scale) of one out of three experiments is displayed in SSEA-5+, APA- and APA+ cell populations.

Taken together, the changes in the histone modification that I characterized indicate that the network of the TFs that govern TE differentiation consist of a cascade, which begins with bivalent genes that are transcriptionally poised and therefore loose the H3K27me3 mark when transcribed, continues with genes that harbor neither the modification and have slower activation kinetics and is accompanied by the silencing of pluripotency genes that loose the activating H3K4me3 mark during differentiation.

Identification of a putative human trophoblast gene network

Time-course transcriptomic analysis of BMP4-treated human ESCs

While the analysis of purified APA+ progenitors brought a further step to identify the intrinsic properties of cells committing to the human trophoblast lineage, to gain a

holistic perspective of this process I had to add another layer of information that refers to processes that precede the emergence of the progenitors at day 3 of differentiation. I have therefore performed in collaboration with Dr. Dmitry Shaposhnikov a time-course RNA-Seq analysis of human ESCs treated with BMP until 72 hours, the time point when the APA+ progenitors appear.

To corroborate the results acquired by the different transcriptomic approaches, microarrays for analyzing the progenitors and RNA-sequencing for the time course bulk human ESCs analysis, I analyzed the overlap in the differentially regulated genes. I found a trend of increasing overlap that included over 90% of the genes after 48 and 72 hours of differentiation (Figure V-14). These data therefore indicate that the network of trophoblast genes consist of several layers, and this confirms our lists of involved genes.

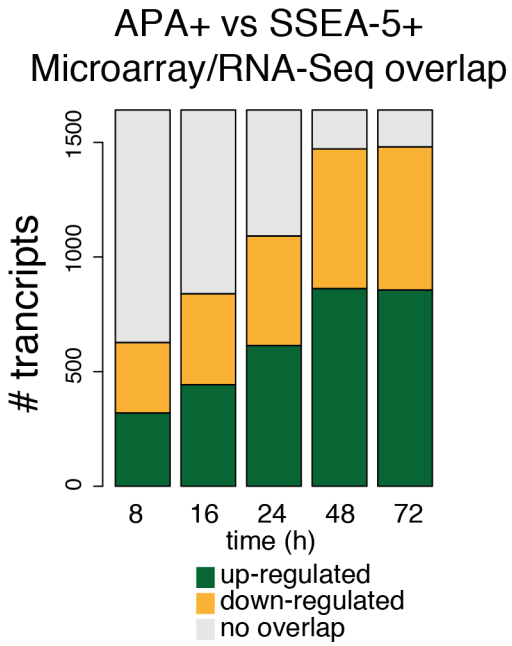

Figure V-14: *Analysis of trophoblast genes by microarray and RNA-seq*
A bar plot exhibiting the number of transcripts overlapping at each of the respective time-point of bulk human ESCs BMP4 treatment (8, 16, 24, 48 and 72 hrs; transcript levels in undifferentiated cells were used as a reference) with genes increased and repressed in the APA+ versus the SSEA-5+ cell population. Green, orange and grey sectors correspond, respectively, to the increased, repressed genes and no overlap. A FDR of less than 5% between replicates (n=2) was used to annotate differentially expressed genes between the respective time points and 0 hrs.

To identify the sequence of activation of the genes in the TE network I analyzed the expression pattern (e.g. trajectory) of the TFs that are upregulated during the time-course bulk RNA-seq measurements (Figure V-15).

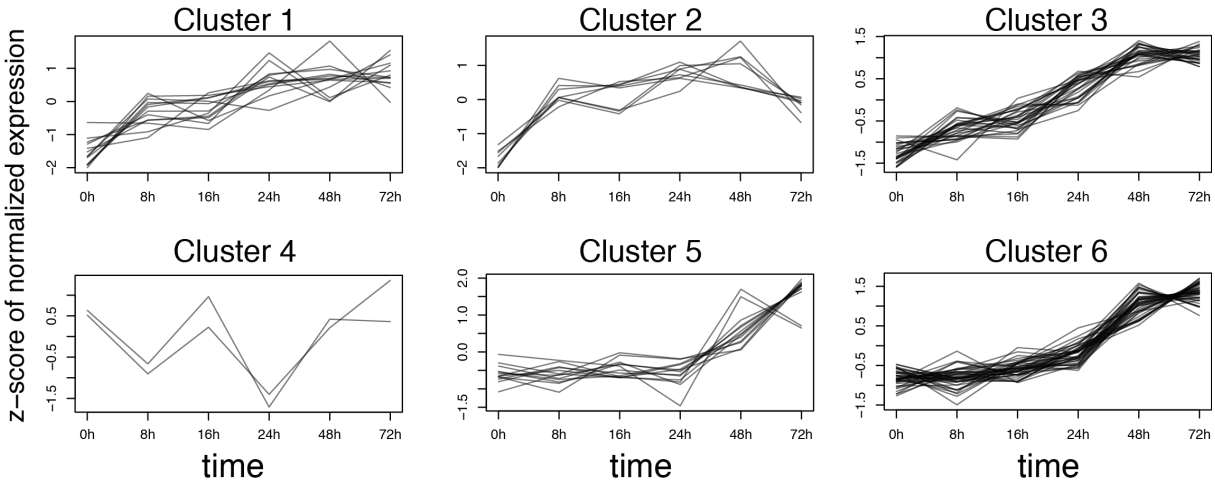

Figure V-15: *Gene expression trajectories of increased TFs*
Clusters of increased TFs (Microarray APA+ versus SSEA-5+) extracted from time-course RNA-seq analysis of human ESCs treatment by BMP4 exhibiting 6 primary cohorts of expression trajectories. Clustering was performed according to pairwise Pearson correlation coefficients using k-means clustering analysis with k=6 clusters.

I identified 6 main cohorts of which 3, termed early, intermediate and late, are the most relevant for my purposes because they could provide clear separation of time wise activation of several group of trophoblast TFs. The genes registered in these clusters, numbered 1, 3, and 6 respectively, are presented in Table V-1. I have validated the expression of some of these early genes by immunohistochemistry (Figure V-16).

early	**intermediate**		**late**		
CRTC3	ANKRD1	MED12L	ARID3A	GRHL1	PPARG
DLX5	ARID5B	MED13L	ASXL1	HIC2	RARB
DLX6	CDX2	NR2F2	BARX2	HIPK3	RCOR1
EGR1	CHD3	NRIP1	BCL6	HOXB2	SETD7
FOS	CITED2	PPARGC1A	BHLHE40	JUP	SMARCA2
GATA2	DLX3	RREB1	BNC1	KDM6B	SMYD2
GATA3	ELF2	RYBP	CALCOCO1	KLF6	SOX9
HEY1	ELF4	SATB1	CEBPA	KLF8	TEAD3
JUN	GRHL2	SERTAD2	CREB3L2	LCP1	TFAP2B
LEF1	HAND1	TBX3	ELF1	MAFB	THRB
MSX2	IKZF2	TEAD1	EPAS1	MED12	TLE4
TFAP2A	ISL1	TFAP2C	EYA2	MEIS1	TP63
TLE3	KAT2B	ZNF117	FOXO4	MEIS2	VGLL1
	MAGED1	ZNF436	GCM1	NFIL3	ZFHX3
		ZNF83			ZNF440

Table V-1: *TFs in early, intermediate and late clusters*
The TFs of clusters 1, 3 and 6 deduced from Figure V-15.

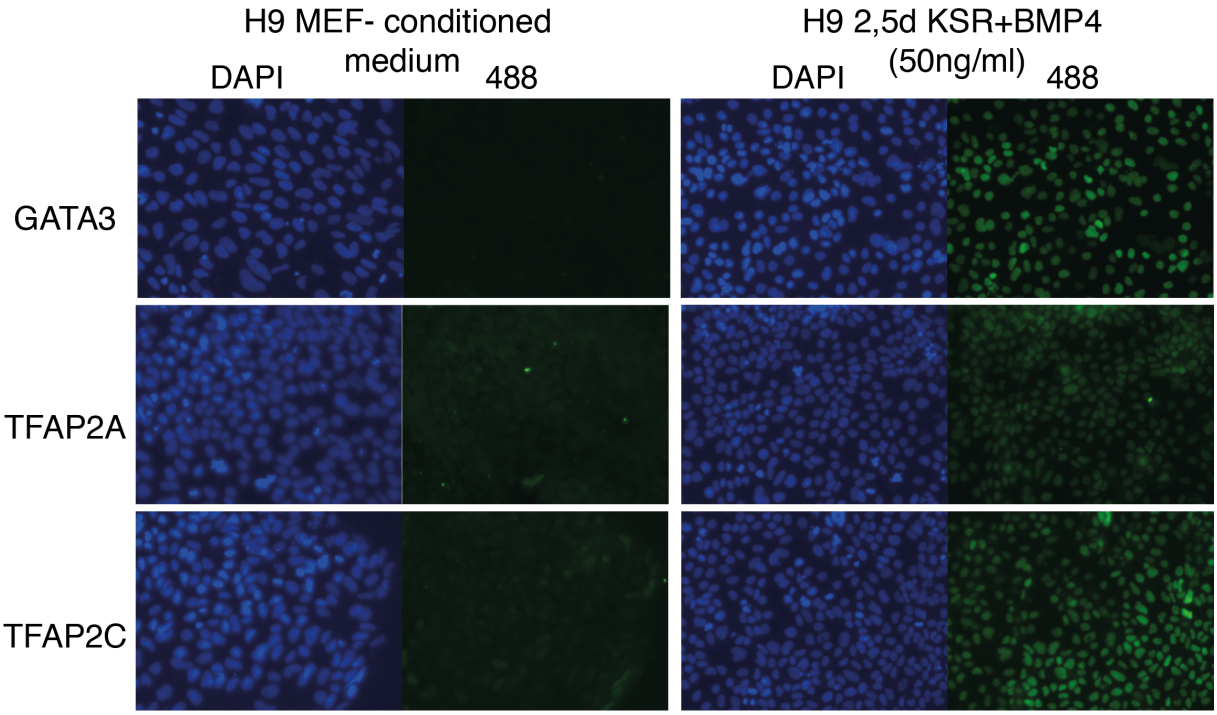

Figure V-16: *Immunohistochemistry of selected TFs of the early trajectory TFs*
Representative immunohistochemistry staining of GATA3, TFAP2A and TFAP2C in undifferentiated and 2.5 days BMP4 treated human ESCs. DAPI is shown in blue, primary antibodies were detected by Alexa 488 labeled secondary antibody.

These indicate that BMP4 exposure activates an early panel of TFs, which subsequently targets additional layers of TFs in the intermediate or late trajectories. Representative RNA-Seq read profiles of an early (GATA3), intermediate (CDX2) and late (GCM1) gene of the respective groups, and one decreased gene (OCT4) are shown in Figure V-17.

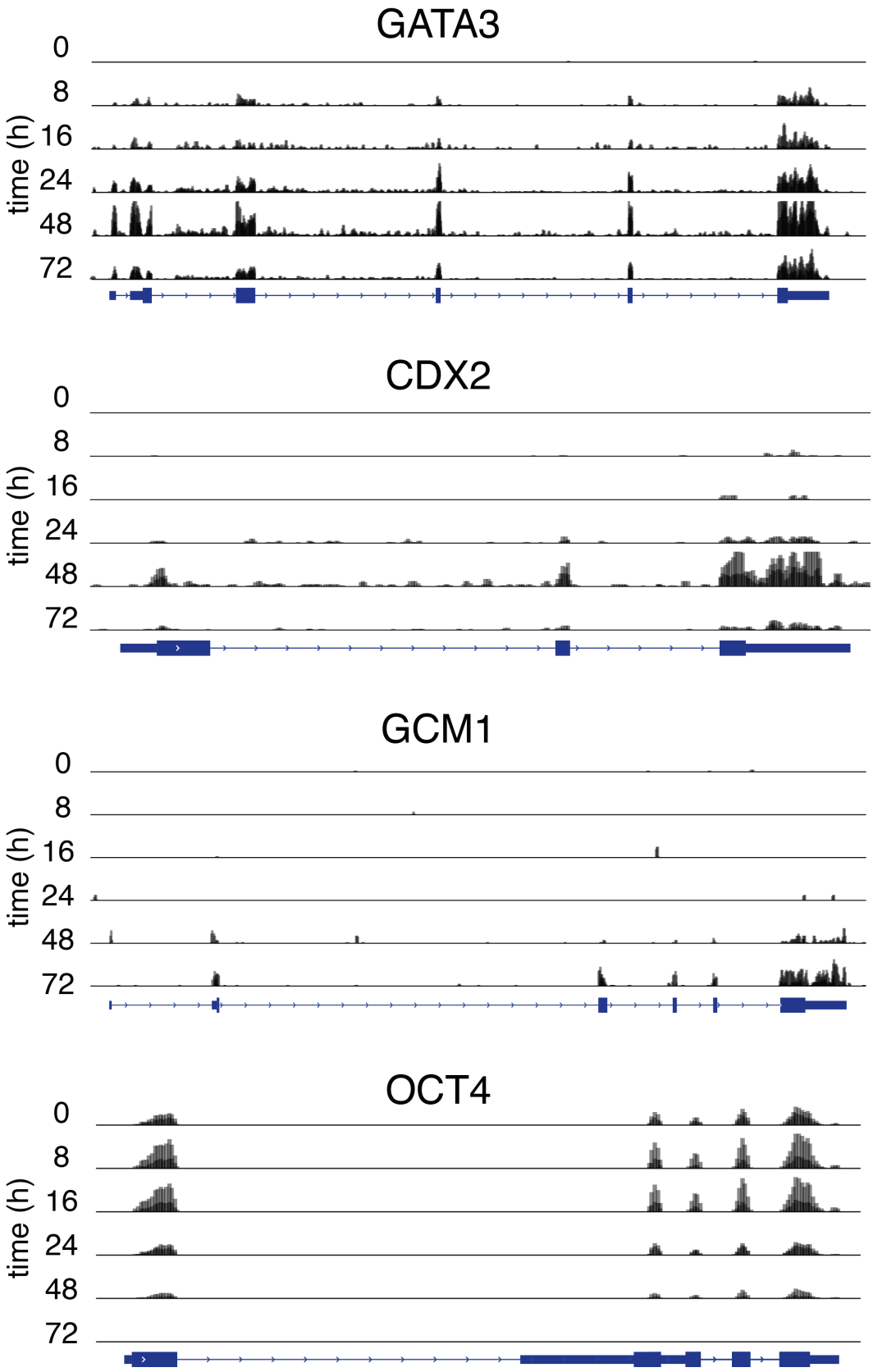

Figure V-17: *Representative RNA-Seq read count profiles*
Representative diagrams exhibiting expression (RNA-Seq) of genes during time-course treatment of human ESCs by BMP4. Genes included are *GATA3*, *CDX2*, *GCM1* (representing the early, intermediate and late clusters) and the downregulated pluripotency gene *OCT4* (n=2). Presented are normalized expression values.

The same clustering approach was also performed for the decreased TFs, but this did not help to separate and produce cohorts that are meaningful for understanding stages of gene repression (Figure V-18).

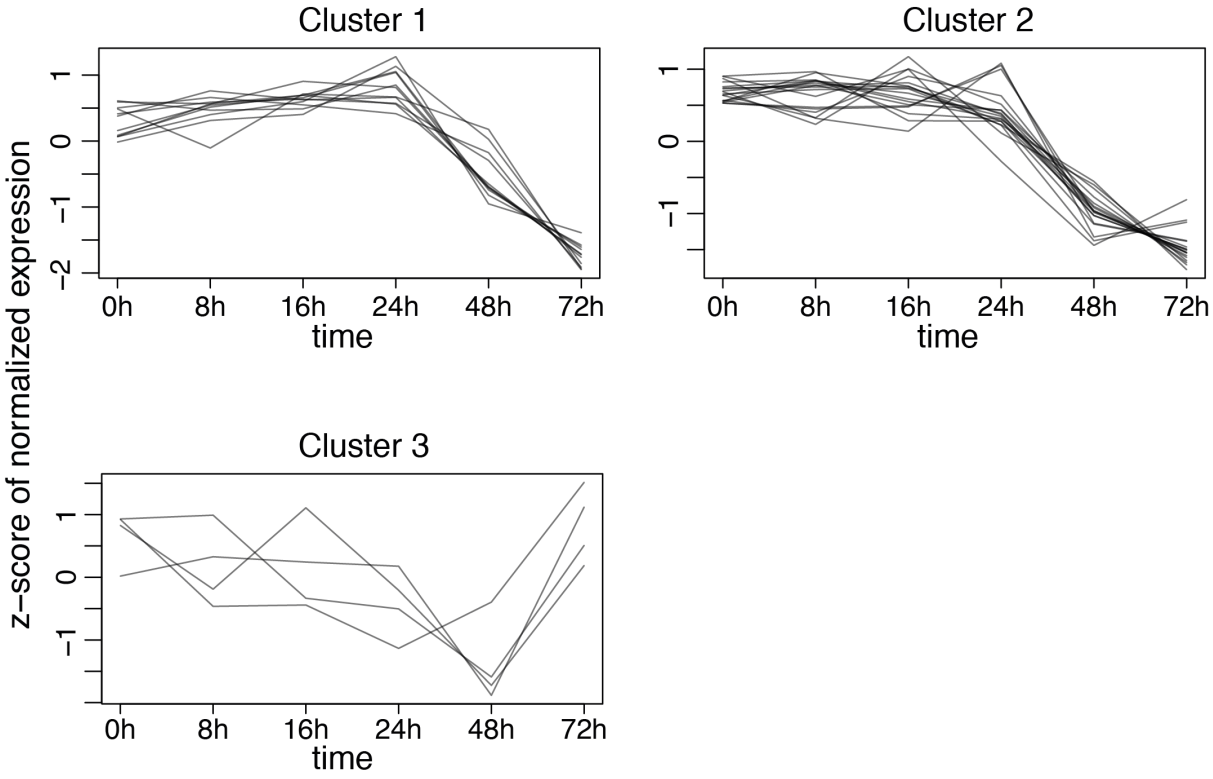

Figure V-18: *Gene expression trajectories of decreased TFs*
Clusters of decreased TFs (Microarray APA+ versus SSEA-5+) extracted from time-course RNA-seq analysis of human ESCs treatment by BMP4 exhibiting 3 primary cohorts of expression trajectories. Clustering was performed according to pairwise Pearson correlation coefficients using k-means clustering analysis with k=3 clusters.

In order to detect the drivers of trophoblast commitment within the early, intermediate and late upregulated gene clusters, I focused on the genes above a log2 fold-change cutoff of 5. The time-points, where this cutoff was applied, were chosen according to the expression patterns of the respective groups. As the early group exhibits a steep increase already after 8 hours the cutoff was applied at this time point, whereas for the intermediate and late groups 24 and 48 hour time-points were chosen, respectively, because then the steepest increase in expression was observed (Figure V-15).
The corresponding TFs are presented in Figure V-19. The TFs exhibiting the highest degree of activation already after 8 hours of BMP-4 treatment in the early trajectory are *GATA2*, *MSX2*, *GATA3* and *TFAP2A*. In the intermediate trajectory I noted trophoblast

regulators including *HAND1*, *CDX2* and *TFAP2C* (Introduction section "The transcriptional network of TE development" and (Riley et al., 1998)). Finally, in the late cluster I noted 15 TFs, including the TFs *VGLL1*, *TP63* and *GCM1*.

early

	log2 FC at 8h
GATA2	9,96
MSX2	7,30
GATA3	7,18
TFAP2A	6,89

intermediate

	log2 FC at 24h		log2 FC at 24h
TBX3	9,45	NR2F2	5,60
HAND1	8,13	TFAP2C	5,36
ANKRD1	7,87	DLX3	5,33
ISL1	7,05	ARID5B	5,26
CDX2	6,35		

late

	log2 FC at 48h		log2 FC at 48h		log2 FC at 48h
EPAS1	11,94	BARX2	8,09	BNC1	5,96
PPARG	9,69	TP63	7,67	GRHL1	5,80
VGLL1	9,40	GCM1	6,24	MEIS2	5,69
MEIS1	9,21	LCP1	6,24	TFAP2B	5,67
HOXB2	8,15	CEBPA	6,21	SMARCA2	5,23

Figure V-19: *Highest induced TFs of early, intermediate and late clusters*
Exhibited are TFs of the respective early, intermediate and late transcript groups increased above a threshold of 5 fold change (log2) at 8, 24 or 48 hrs, respectively. FC = fold change; h = hours

Genome wide mapping of GATA2/3 and TFAP2A/C bound loci during BMP4-mediated human ESC differentiation

As I have hypothesized that early activated TFs propel the trophoblast specification network, I next focused in this regard on the functions of primary candidates in the early gene cluster group. I performed ChIP-Seq experiments in collaboration with Dr. Dmitry Shaposhnikov using antibodies that are specific for human GATA2, GATA3, TFAP2A and TFAP2C. The reason that I chose TFAP2C instead of MSX2, which is in the early cluster while TFAP2C is in the intermediate, is that other work in our lab has indicated that MSX2 is redundantly expressed in other early lineages, and because TFAP2C is involved in mouse TE specification and it is a member of the AP2 family. As the TF ChIP-Seq typically requires more cells than histone modification ChIP-Seq, these experiments were performed using bulk cell preparations. By performing de-novo motif detection of the sequenced reads I discovered that the GATA-motif is overrepresented in the GATA2 and GATA3 ChIP-Seq experiments. Furthermore, same analysis of the TFAP2A and TFAP2C ChIP sequenced reads showed the consensus motif for the TFAP family (Figure V-20). This is reassuring that my TF ChIP readouts are specific for their bound target loci.

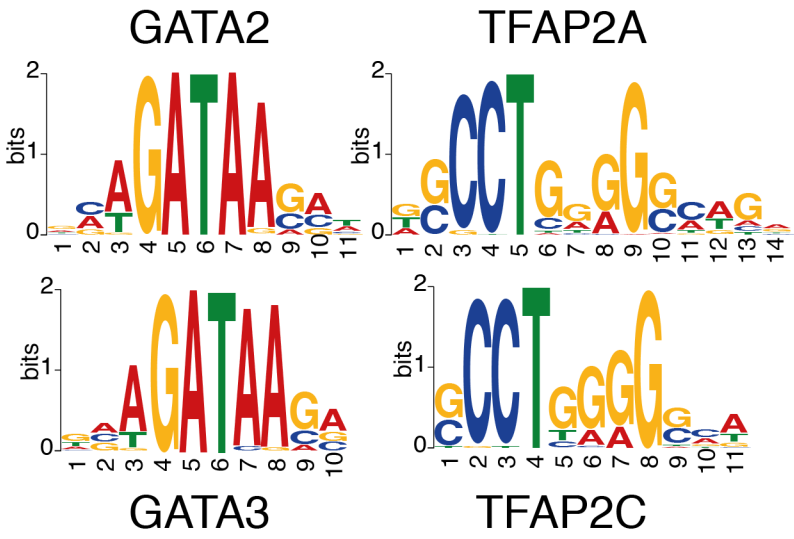

Figure V-20: *De-novo motif analysis of TF ChIP-Seq*
Motifs enriched in the genomic fragments isolated by transcription factor ChIP-Seq of GATA2, GATA3, TFAP2A and TFAP2C.

Next I focused on identifying the genes exhibiting peaks of TF binding within a region of 3.5 kb up- and 5 kb downstream of the TSS. By analyzing the cooperative binding of the TFs, I discovered five categories where 0, 1, 2, 3 or 4 TFs are bound in the same region. To find out whether increased or decreased expression during trophoblast differentiation could be explained by the binding of the genes by the TFs, I analyzed the correlation between the TF bound sites and the induced/repressed genes noted following 72 hours of differentiation compared with undifferentiated cells. As a result I observed, that multiple TF binding correlate with the upregulation of genes rather than the downregulation, with several important exceptions that are discussed below (Figure V-21).

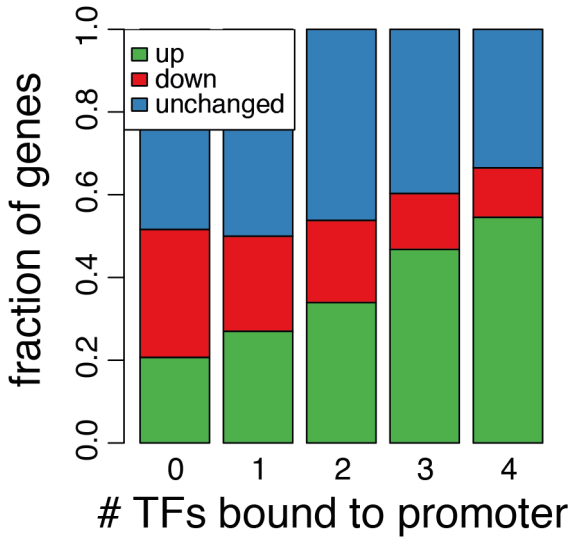

Figure V-21: *Correlation of TF binding and gene expression*
A bar plot exhibiting the correlation between the number of the bound GATA2, GATA3, TFAP2A and TFAP2C TFs (either 0, 1, 2, 3 or all), in 3.5kb up- and 5kb down-stream of gene's TSS, and their direction of regulation: up- (green) and down- (red) regulated, or unchanged (blue), comparing differentially expressed genes at the 72 hour time-point BMP4 treatment versus undifferentiated cells as determined by global RNA-Seq analysis.

In total I found 204 genes that are bound by the all 4 TFs in the region of 3.5 kb up- and 5 kb downstream of the TSS (Table V-2-4).

upregulated								
AADACL3	C8orf31	DOCK9	HRH1	LINC00518	MPZL1	OPN3	S100A16	TRIL
ABCA4	CA12	DOPEY2	INADL	LINC00936	MTUS1	ORAI3	SLC20A2	TRIML2
ACVR1	CDH10	ELMO1	ITGA2	LMOD2	MYLK-AS1	P2RY6	SLC7A2	TTLL7
ADAM18	CDX2	ENPEP	ITGB6	LOC100130476	MYO6	PDE10A	SLCO2A1	USP43
AMOT	CMTR2	ENTPD4	KALRN	LOC100507346	NEDD9	PDZD2	SMTNL2	VTCN1
ANKRD1	COLEC12	FAM89A	KCNC3	LOC729739	NKX2-3	PIK3C2G	SPARC	YPEL2
ANXA3	CRIP1	FOXC1	KCNN4	LPP	NOS2	PPFIBP2	STS	YPEL5
ANXA4	CSF1R	FRS2	KIAA1456	LPP-AS2	NPC2	PRTG	SVOPL	ZAP70
ARHGAP24	CSGALNACT1	GADD45G	KIAA1551	LTBR	NPNT	PTPN14	SYNPO	ZFPM1
BACE1	CTSL3P	GAS7	KMO	MAB21L3	NR2F2	PTPN3	TACC1	ZNF358
BACE1-AS	CYP1B1	GNLY	KRT18	MAP3K8	NR2F2-AS1	PTTG1IP	TBX3	
C10orf10	DGKD	GRHL1	KRT19	MBNL3	NRK	RALBP1	TINCR	
C1QTNF6	DIAPH3	GSTO1	KRT8	MBOAT2	NTAN1	RHOBTB2	TMBIM1	
C21orf2	DLX4	HAPLN1	LGR5	MIR205	NTRK1	RRBP1	TNFAIP3	

unchanged							
ACOT2	FAM65B	LSM4	PPP1R26-AS1	SLC16A9	TLE3	ZFX	ZNF853
ALDH4A1	GLG1	MESDC2	PPP1R9A	SLC44A1	TMEM254	ZMIZ2	ZSWIM6
ATRAID	GMEB1	MRS2	RABGGTB	SLC5A10	TMEM254-AS1	ZNF175	
C14orf80	IQGAP2	MYO1B	RBL2	SLC7A6	TMEM44	ZNF280D	
CHML	ISCA2	NENF	RBM15	SNX13	UBR5	ZNF343	
DCP1A	IVNS1ABP	PDS5A	RPL13AP20	SOCS2	WDR74	ZNF587	
DTNA	LETMD1	PGS1	SETD4	SSBP3	WDR91	ZNF593	
ECH1	LINC00339	PLS3	SLBP	TGIF1	YTHDF1	ZNF826P	

downregulated							
ARTN	CERKL	HSD17B4	LDLR	PDK1	PTPRD	SCNN1A	UNQ6494
ASPRV1	FDFT1	IL17RD	LOC440600	POU5F1	PTPRG	SLC45A4	USP44
CA14	GJA1	JADE1	LRRC9	PRTFDC1	QSOX2	TMEM55A	

Tables V-2-4: *Gene loci bound by all 4 TFs*
Tables showing up-, downregulated, and unchanged transcripts that are bound by all four TFs (GATA2, GATA3, TFAP2A and TFAP2C) 3.5kb up- and 5kb down-stream of their TSS. Lists are derived from the analysis shown in Figure V-21.

By further analyzing the different combinations of three of the TFs, I found that GATA3 exhibits the broadest co-occupancy with the other TFs (Figure V-22) indicating that GATA3 may be the most important driver of this developmental process.

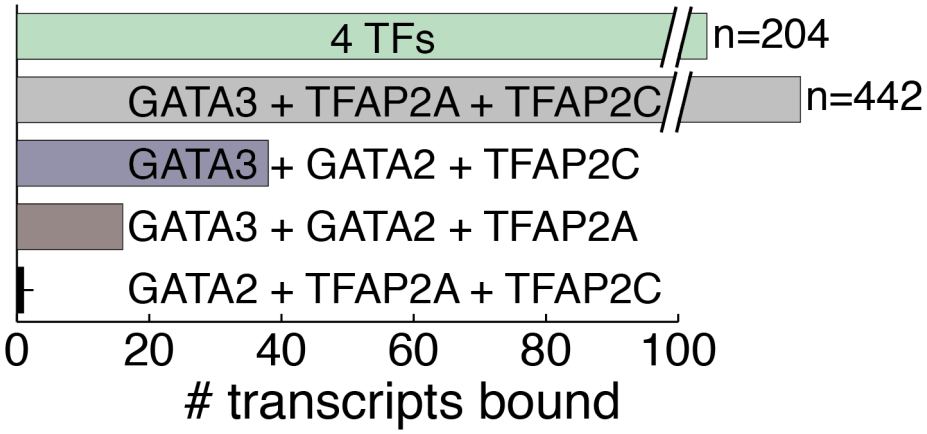

Figure V-22: *The number of transcripts bound by either 3 or 4 TFs*
A graph showing the numbers of genes that are bound either by 4 or different combinations of 3 TFs extracted from GATA2/3 and TFAP2A/C ChIP-Seq data.

Of the 204 bound genes I found that 122 genes were upregulated at the 72 hour time point of BMP4 treatment. (Table V-2). Among these were 11 TFs e.g. *CDX2* and *ANKRD1*, which were identified in the group of genes that exhibited an intermediate upregulation behavior. Other, non TF genes, in this group, include trophoblast associated genes like *STS*, *VTCN1*, *KRT8* or *KRT18*.

The decreased list consisting of the genes that are bound by all four TFs, didn't show enrichment for specific tissues, but importantly included the TF *OCT4* and the transcriptional co-activator *JADE1* (Figure V-23).

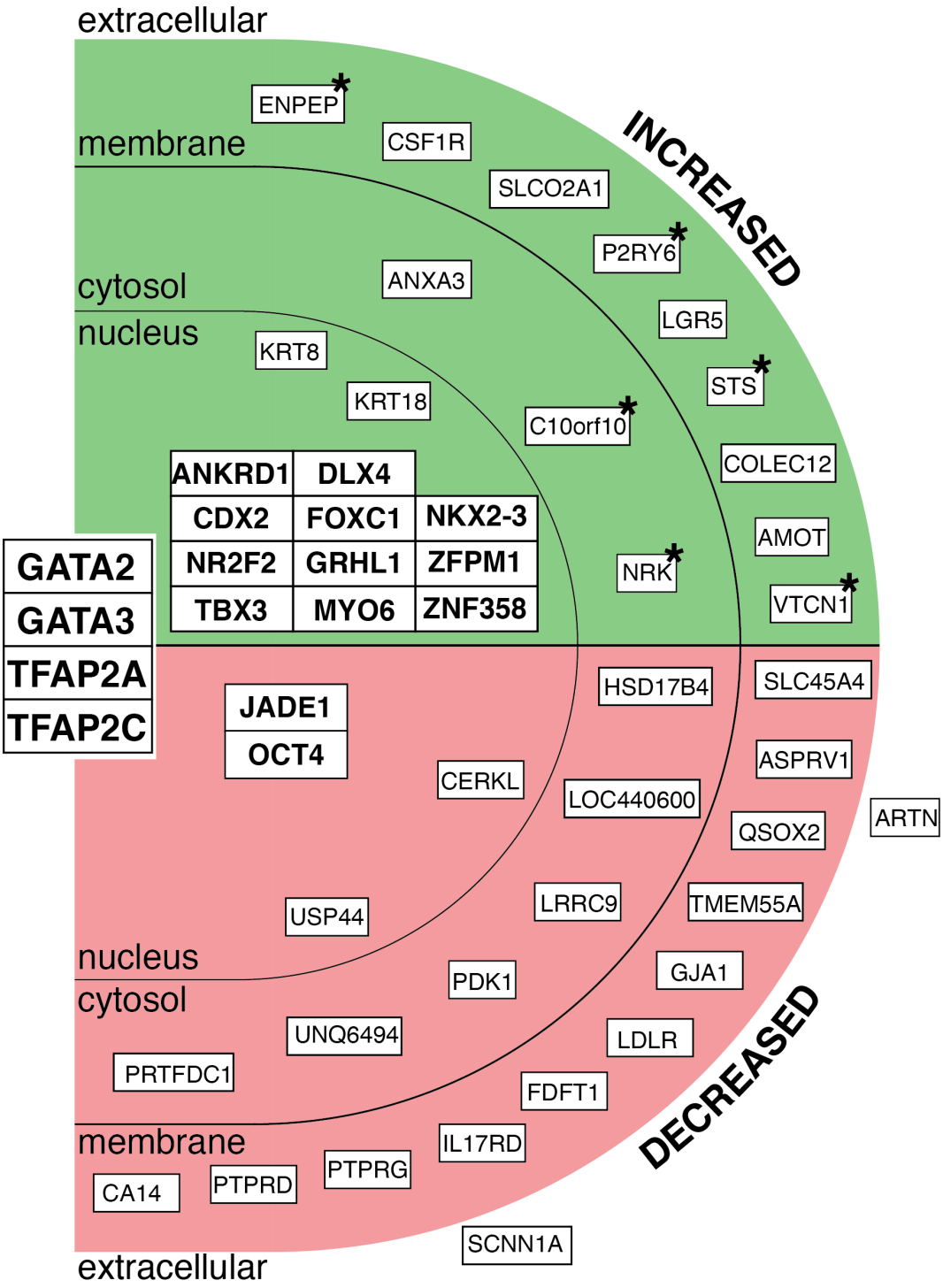

Figure V-23: *Transcripts bound by all 4 TFs*
An overview of genes that are bound by GATA2, GATA3, TFAP2A and TFAP2C TFs (all 4) and the cellular location of the respective proteins as annotated by Genomatix. Increased and decreased genes during the BMP4 time course are exhibited in the upper (green) and lower (red) sections. Transcriptional regulators are marked bold, and representative placenta enriched non-TFs (GO terms analysis) are shown. *OCT4* is the only down-regulated TF bound by all four TFs. The rest of the down-regulated genes were not associated with specific developmental processes.

Interestingly, I found that the TFs bind in the first intron of *CDX2* and *OCT4*, while typically the four exhibit binding upstream or in the near vicinity of TSS, for example as in *ANKRD1*. For *GCM1*, I observed binding of GATA2, GATA3 and to a lesser extend of TFAP2A (Figure V-24).

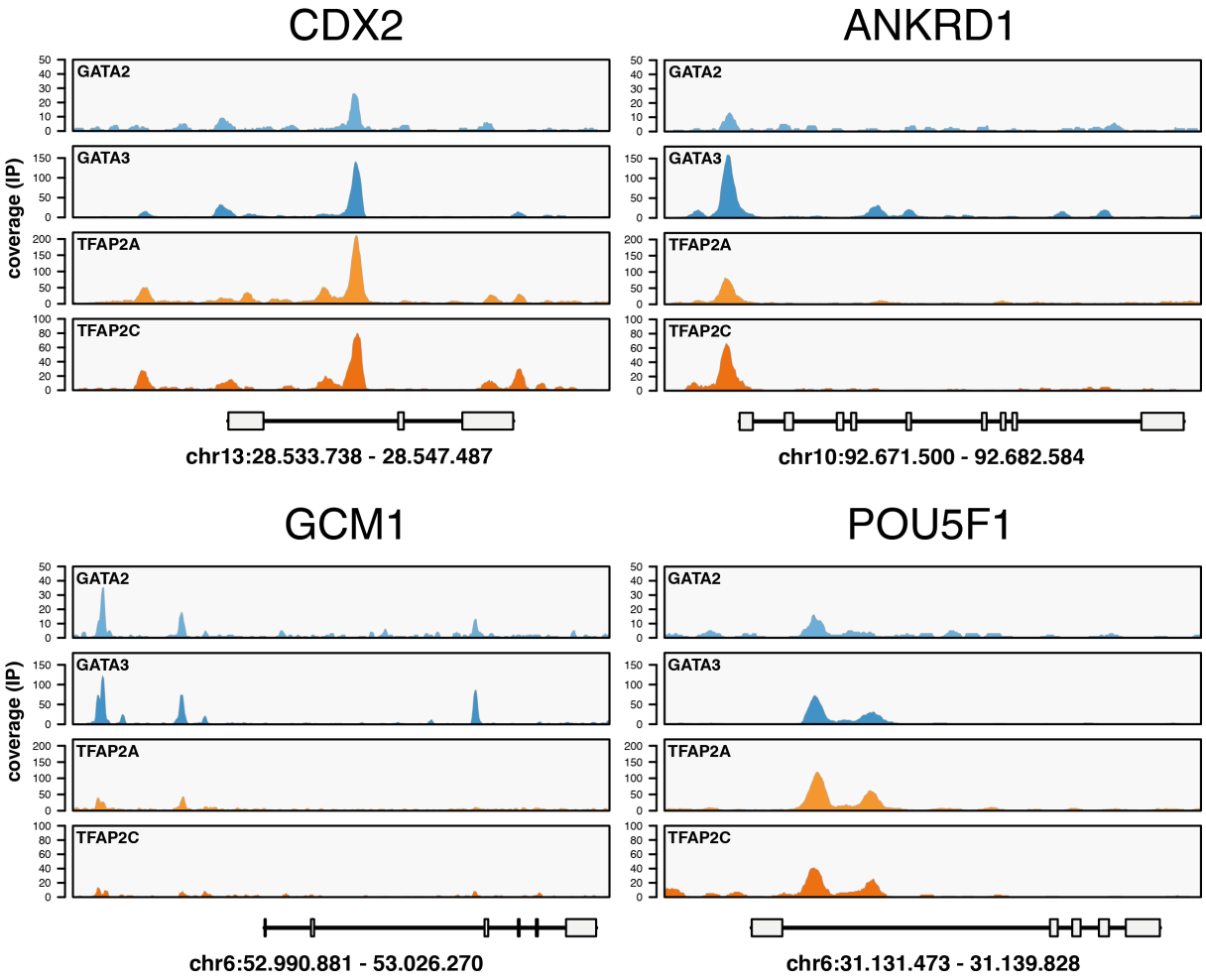

Figure V-24: *Examples of TF binding events by the four TFs*
Genomic maps of representative genes, *CDX2*, *ANKRD1*, *GCM1* and *OCT4*, and the corresponding reads extracted from the ChIP-Seq of GATA2, GATA3, TFAP2A and TFAP2C.

Taken together I took these data as an indication that the four TFs, namely *GATA2*, *GATA3, TFAP2A* and *TFAP2C*, collectively underlie human trophoblast specification. I therefore next set to functionally test the importance of these four TFs, whom I operationally named the TEtra (the TrophEctoderm four), for this differentiation pathway.

Functional analysis of the trophoblast TF network

As mentioned above, I found that GATA3 exhibits the broadest co-occupancy with the other members of the TEtra. Therefore, I decided to focus on this TF for functional validation of the network that I identified. I used a human ESC line that harbors an inducible form of the CRISPR-Cas system (HUES9 iCRISPR, (Gonzalez et al., 2014)) for knocking-out (KO) GATA3. My readout for the function of GATA3 included measurements of the APA+ cell population by FACS, targeted gene expression measurements by RT-PCR and detection of hCG levels by the immunoencymometric assay.

To knockout *GATA3* I used two pairs of guide RNAs (gRNAs) that were designed to cut out a region of 325bp from the second exon and the downstream intron. This resulted in a premature stop codon in this intron of *GATA3*. I have derived two clones harboring such mutations in both alleles, which were verified as knockouts by Western blot analysis after exposure of *GATA3* KO human ESCs to BMP4 (Figure V-25).

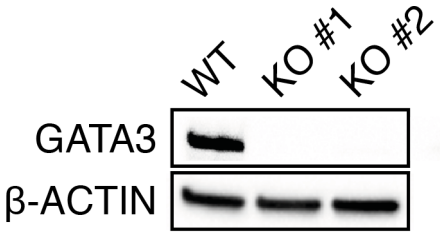

Figure V-25: *Western blot analysis of GATA3 knock-out in the CRISPR-CAS inducible ESC system* A Western blot analysis of wild-type and *GATA3* KO iCRISPR human ESC clones exposed for 3 days to BMP4 treatment.

The impact on differentiation as tested by flow cytometry indicated that the formation of the APA+ cell population is severely impaired, noting a decrease of the cell population from ~60% in the wild-type cells to 5% in the GATA3 KO cell clones (Figure V-26). This validates that GATA3 indeed is one of the primary drivers of the APA+ trophoblast differentiation in human ESCs.

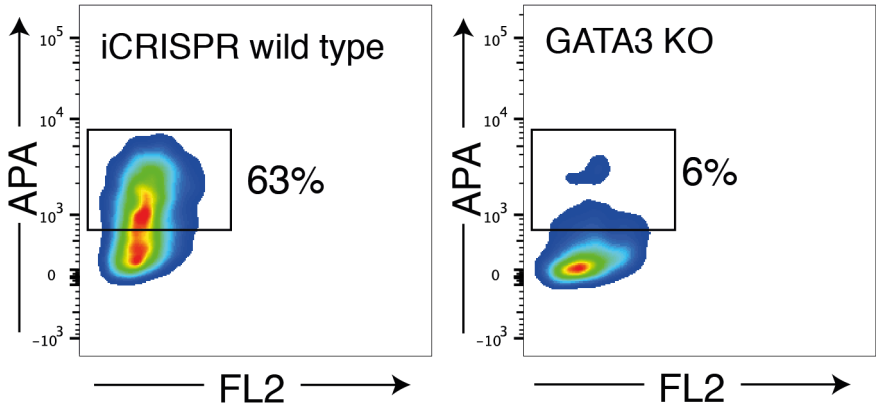

Figure V-26: *Flow cytometry analysis of APA in human ESCs and GATA3 knock-out* Representative APA flow cytometry plots exhibiting analysis of wild-type and GATA3 KO iCRISPR clones following 3 days of BMP4 differentiation. FL2 = Fluorescence 2

To measure the effect of GATA3 on its assumed targets that were identified by TF ChIP-Seq, I measured the expression of candidate genes in the network by RT-PCR and compared to the wild-type cells. I found that the expression of several key genes in the network, including *GATA2, GCM1, VGLL1* and *STS* decreases in comparison to the non-modified cells and as anticipated from my ChIP-Seq data, which have indicated that *OCT4* is inhibited by the TEtra, this gene shows a transcriptional increase. Surprisingly, however several genes exhibited trends that are opposite the anticipated. This includes *TFAP2A, TFAP2C, CDX2* and *TP63* that were increased in the GATA3 KO clones relative to the control (Figure V-27). I hypothesize that *TFAP2A* and *TFAP2C* partially counteract and compensate the GATA3 KO on the transcriptional level by upregulation of *CDX2*. However, it is also likely that *CDX2* is regulated by additional ways that are *GATA3* independent.

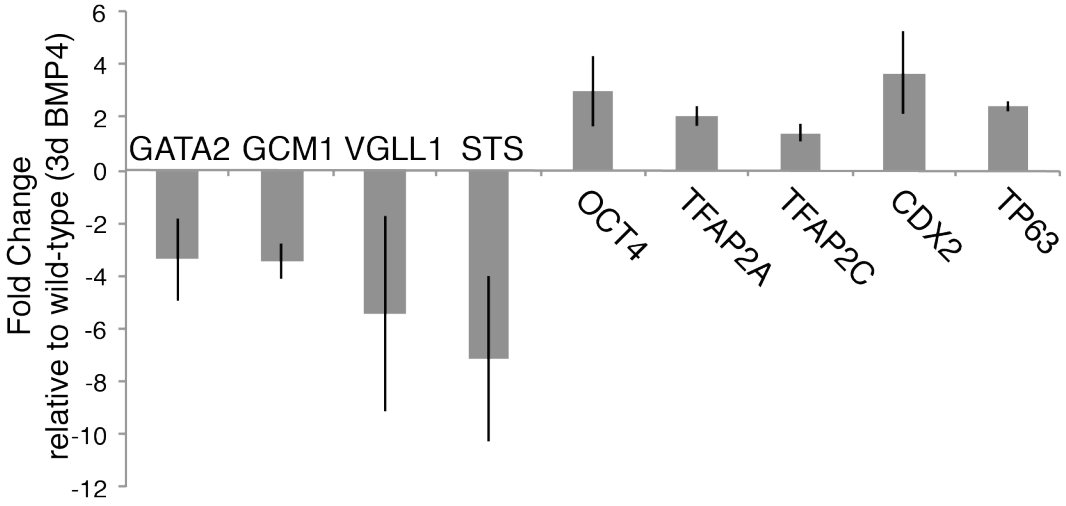

Figure V-27: *Analysis of trophoblast candidate genes after GATA3 KO in human ESCs*
The relative expression fold change of a set of pluripotency and early trophoblast genes was analyzed by RT-PCR comparing *GATA3* KO and wild-type iCRISPR human ESCs following 3-day of BMP4 treatment. Standard error of the mean (SEM) of two independent experiments is shown.

Finally to validate the effect of *GATA3* on placenta hormone production, I compared the hCG production in the GATA3 KO clones to unmodified cells using an immunoencymometric assay. In line with the previous data, I found that also the hCG production is severely impaired during differentiation of GATA3 KO cells to trophoblast by treatment with BMP4 (Figure V-28).

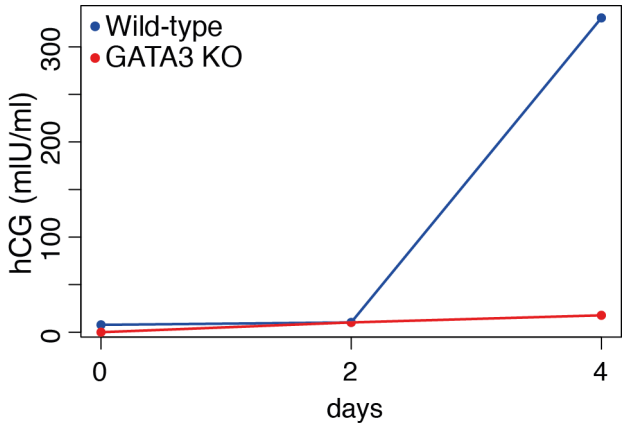

Figure V-28: *Placental hormone production of human ESCs harboring GATA3 KO*
Time course analysis of hCG concentration produced by wild-type or *GATA3* KO human iCRISPR ESCs treated with KSR+BMP4 (n=2).

Identification of developmental epigenetic signatures

Correlation of histone modification turnover and temporal regulation of TFs during BMP4-mediated human ESC differentiation

To identify epigenetic mechanism that underlie human trophoblast differentiation, I analyzed the correlation between transcription and changes of the histone modification patterns in the early, intermediate and late groups of TFs detected by bulk RNA sequencing. For that I tested if any of the histone modification marks and combinations is overrepresented in the populations of SSEA-5+, APA- and APA+ sorted cells. I noted that bivalency is enriched in the SSEA-5+ fraction with respect to the genes with early induction trajectory, that H3K27me3 and to a lesser extent H3K4me3 and bivalent categories characterize the genes exhibiting the intermediate trajectory in SSEA-5+ cells, and that the no Mod and bivalent categories are most prevalent in the group of the genes that exhibit latest induction behavior with respect to all cell populations. Of note, this distribution of histone mark enrichment is similar in the APA- cell population. In the APA+ cell population, however, there is a shift to the H3K4me3 category in the early and intermediate cluster, whereas in the late cluster no Mod and bivalent marks persist (Figure V-29). Taken together, these data indicate that distinct patterns of histone marks characterize TFs that are activated at different stages; bivalent TFs go first, than bivalent/H3K27me3 marked TFs and finally no Mod TFs, such as *GCM1*.

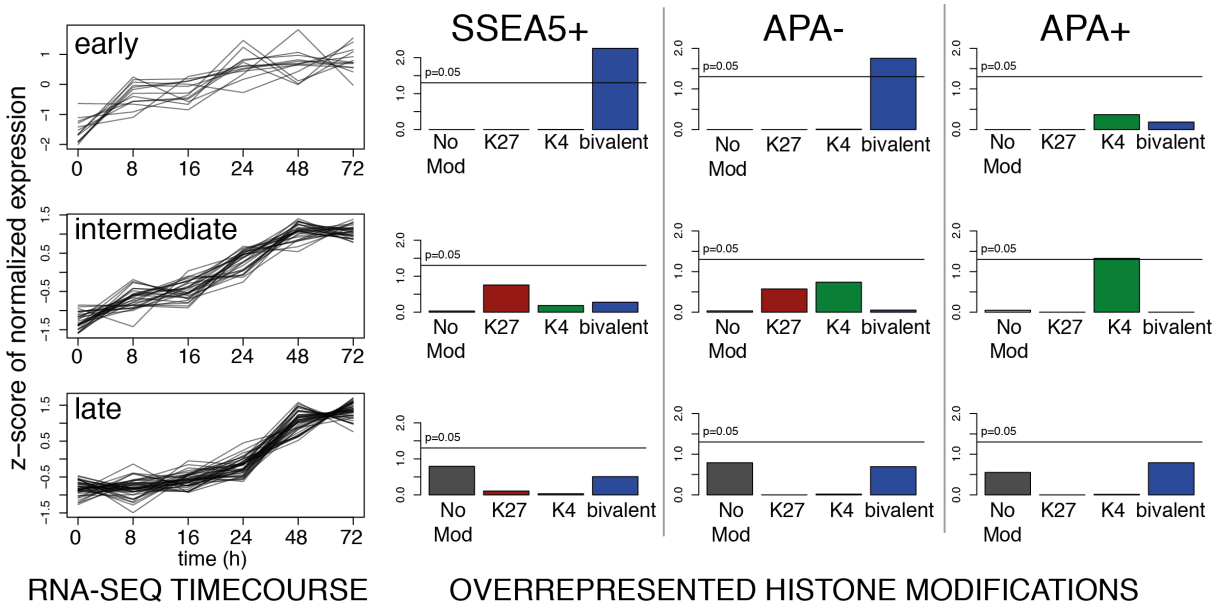

Figure V-29: *Trophoblast gene expression kinetics and histone mark turnover*
The left panel exhibits the trajectories of early, intermediate and late increased TF cohorts over a time-course of 72 hrs where human ESCs were treated by BMP4. Only APA+ versus SSEA-5+ cell population induced TFs are plotted. The right panes exhibit the overrepresented histone modification categories of these TFs in SSEA-5+, APA- and APA+ cell populations, classified as bivalent, monovalent and TFs where these histone modifications were not detected. P-values were calculated by applying Fisher's exact test for the observed number of overlaps between the group of genes with the respective histone modifications and the genes in the three cohorts.

Key genes of the early, intermediate and late cohorts and their corresponding histone modification status in SSEA-5+ cells is shown in Figure V-30.

early

	log2 FC at 8h
GATA2	9,96
MSX2	7,30
GATA3	7,18
TFAP2A	6,89

intermediate

log2 FC at 24h		log2 FC at 24h	
TBX3	9,45	NR2F2	5,60
HAND1	8,13	TFAP2C	5,36
ANKRD1	7,87	DLX3	5,33
ISL1	7,05	ARID5B	5,26
CDX2	6,35		

late

log2 FC at 48h		log2 FC at 48h		log2 FC at 48h	
EPAS1	11,94	BARX2	8,09	BNC1	5,96
PPARG	9,69	TP63	7,67	GRHL1	5,80
VGLL1	9,40	GCM1	6,24	MEIS2	5,69
MEIS1	9,21	LCP1	6,24	TFAP2B	5,67
HOXB2	8,15	CEBPA	6,21	SMARCA2	5,23

Figure V-30: *Histone marks of TFs from early, intermediate and late cohorts*
TFs of the early, intermediate and late cohorts exhibiting > log2 5 fold change at 8, 24 or 48 hrs, respectively. Font color labels correspond to the histone modification categories in SSEA-5+ undifferentiated cells (extracted from Figure IV-19): bivalent (blue), H3K4me3 monovalent (green), H3K27me3 monovalent (red) and H3K27me3, H3K4me3 modification negative (grey).

DNA methylation changes during human trophoblast differentiation

As DNA methylation changes play an important role in developmental gene regulation, I investigated the 5-Methylcytosine status of hundreds of thousands of CpG loci in the SSEA-5+, APA- and APA+ cell population. For consistency, this analysis was conducted using the same batch of samples used for microarray and histone modification analyses. As a background for explaining the results below, it should be mentioned that in human ESCs bivalent genes are typically CpGs rich, often harboring CpG islands at their promoters, and do not exhibit DNA methylation marks in this region.

In line with previous observations, I did not observe CpG methylation in CpG islands at promoters of early bivalent TFs neither in undifferentiated cells not in progenitor populations, where they are H3K4me3 monovalent e.g. *GATA2* (Figure V-31). TFs of the no Mod category, e.g. *GCM1* exhibited CpG methylation in SSEA-5+ cells, but surprisingly although they were induced in the APA+ progenitor population, they did not show a significant removal of methylated CpGs (Figure V-31).

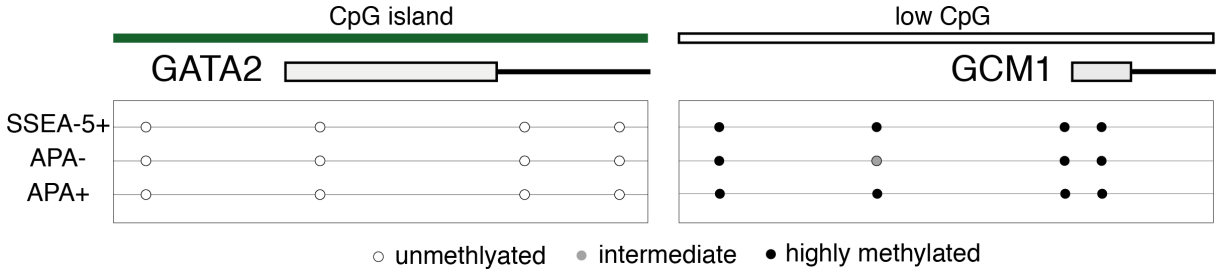

Figure V-31: *Representative CpG sites in promoter regions of early and late TFs*
Displayed are the TSS of the *GATA2* and *GCM1* genes, and positions of the respective CpG island and the CpG poor region. Bottom, the DNA-methylation state of analyzed CpGs at the corresponding genomic positions in the SSEA-5+, APA- and APA+ cell populations. DNA-methylation analysis was conducted in triplicates and classification was performed as follows: unmethylated: 0-20%, intermediate: 21-60% and highly methylated: 61-100%.

In fact I identified only few instances where significant changes in CpGs methylation could be observed at induced trophoblast genes, always in the direction of de-methylation comparing the APA+ to SSEA-5+ cell populations. I found that this de-methylation in fact takes place in the sites of TEtra binding, the most notable of which I observed in *VTCN1* (Figure V-32). I noted a similar trend of CpG de-methylation and

TEtra co-localization in several other structural genes that are induced in APA+ trophoblast progenitors (highlighted with an asterisk in Figure V-23)

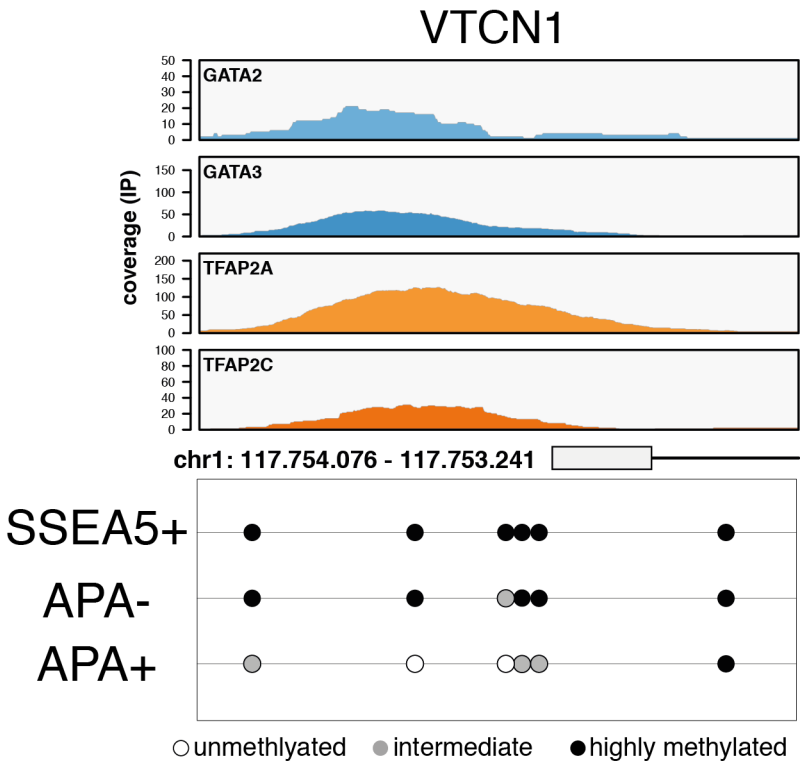

Figure V-32: *TF binding and developmental changes of DNA methylation in VTCN1*
A genomic map of the low CpG gene *VTCN1*, which is induced in the APA+ versus the SSEA-5+ cell population, exhibiting binding of GATA2, GATA3, TFAP2A and TFAP2C. Bottom, the DNA-methylation status of analyzed CpGs at the corresponding genomic positions in the SSEA-5+, APA- and APA+ cell populations. Analysis was conducted as indicated in Figure V-31.

VI. Discussion

My results break grounds in general understanding of trophoblast progenitor differentiation and in particular in the human. This is because I was able to prove lineage correspondence and analyze cell intrinsic properties of human trophoblast progenitors using purified populations of progenitors during their process of differentiation from human pluripotent ESCs. I discovered the TEtra TFs that are at the basis of this differentiation pathway, and proved functionally that it is a critical part of the network that drives human trophoblast specification. This circumvents ethical limitations on studying human embryogenesis trophoblast development, and provides critical grounds to explore mechanisms of placental disease that have long-term effects on child's and mother's health. Moreover, I discovered key modes of epigenetic regulation that govern human trophoblast specification, and these modes may have implications to understanding human development to various cell types.

Resolving the uncertainty concerning the lineage correspondence of human PSC-derived trophoblast progeny

As a foundation for my doctoral thesis, I began my investigation of the progenitor population that emerges from human ESCs, termed here the APA+ progenitors, by analyzing their lineage correspondence. I have obtained several lines of evidence, which in my opinion, remove the doubts that the progeny of human PSCs treated with BMP4 is of a trophoblast origin (Introduction section "BMP4-mediated trophoblast differentiation of human ESC"). This includes:

1. I have shown that the progeny of human ESCs treated by BMP-4 includes other lineages and possibly residual undifferentiated cells in the APA- population.

I have demonstrated this by:

a. RT-PCR analysis of the APA- population indicates heterogeneity of this population, as in parallel to trophoblast genes also genes of other lineages (e.g. meso- and

mesendoderm represented by MESP1 and T) are induced. Further I saw lower *OCT4* expression in the APA+ compared with the APA- population indicating that there are still cells with pluripotency characteristics in the APA- population.
b. The increase of trophoblast genes was less pronounced in the APA- population, indicating lower numbers of trophoblast cells in this population. *ELF5* for example was not detected in APA- population and *GCM1* expression, which is important for syncytiotrophoblast development, was very low in the APA- population.
c. Global gene expression analysis and GO-term bioinformatics showed that the APA- population is enriched for lineages including ectoderm and mesenchyme and not trophoblast or placenta.

2. I have shown that the purified APA+ population has a trophoblast correspondence.

I have demonstrated this by:

a. RT-PCR analysis of purified APA+ cells demonstrated that trophoblast associated genes (*CDX2, ELF5, GCM1* and *ENPEP*) and not genes of meso- and mesendoderm (*MESP1, T, CD13, ROR2* and *GSC*) are expressed in this population. Further I showed that the pluripotency gene *OCT4* is downregulated more in this population compared to APA- cells.
b. Global gene expression analysis and GO-term bioinformatics comparing the APA+ population to both, the APA- and the pluripotent population showing that the APA+ population is enriched for tissues of trophoblast and placental nature.
c. Using existing transcriptomic datasets of human blastocyst-stage TE as a basis for comparison to the *in vitro* generated trophoblast progenitors. Comparison of the datasets led me to the conclusion that the core transcriptional TFs in my system correlate with the *in vivo* dataset.

A recent study pointed out that BMP4 differentiated human ESCs fulfill some, but not all criteria of *in vivo* first trimester trophoblasts (Introduction section "BMP4-mediated trophoblast differentiation of human ESC"). This study shows for example that only a portion of the in vitro differentiated cells expresses *GATA3* and *TFAP2C, in vivo* markers for trophoblasts. However, they have utilized clumps of cells as a starting material for differentiation, which in my opinion represent the physiology of ICM cells rather than

polar TE cells and could then lead to a less efficient BMP4 mediated differentiation process. Our protocol, which is based on seeding human ESCs as single cells before differentiation, can overcome at least partially this heterogeneity as immunohistochemistry identified GATA3 and TFAP2C expression in a majority of the cells.
Despite of this, Lee and colleagues noted that CpGs at the ELF5 promoter become de-methylated, resembling the state of the *in vivo* first trimester trophoblast. My analysis, however, did not show de-methylation of this trophoblast specific promoter, but as mentioned before we look at a very early time-point and generally identify very few changes in DNA methylation over the entire genome.

Heterogeneity during differentiation is in my opinion the source of debate on the lineage correspondence of the trophoblast progenitors that emerge from human PSCs.
For example, Bernardo and colleagues showed that human ESCs treated with BMP4 in chemically defined medium express both, trophoblast and mesoderm genes and conclude that these are of extraembryonic mesoderm therefore challenging the trophoblast nature of these cells (Introduction section "BMP4-mediated trophoblast differentiation of human ESC"). However, the experimental system used by Bernardo and colleagues harbors important limitations:

a. They apply a different cultivation and differentiation protocol than all studies before by using chemically defined medium on gelatin membranes soaked in FBS and do not state whether they started the differentiation using single cells or cell clumps.
b. The differentiation performed with BMP4 only results in 4-8% KRT7+ cells. Compared with studies performed later by several groups this does not reflect their findings, where at least 40% (Amita et al., 2013) and in other cases most of the BMP4 treated cells were KRT7+ (Telugu et al., 2013; Lee et al., 2016). Still it is important to mention that Lee and colleagues utilized an Activin A inhibitor and ALK and FGF receptor inhibitors, which suppress mesoderm differentiation in combination with BMP4 for differentiation.
c. They analyze these 4-8% of KRT7+ cells and although they state that the KRT7+ cells express both trophoblast genes (GCM1, ELF5 and HCGA) and mesoderm genes (ISL1 and FLK1), the amount of genes characterized limit the assignment of these cells to a certain lineage. Genome wide analysis of this KRT7+ population could have shed light into the nature of these KRT7+ cells.

d. There is still the possibility of heterogeneity in this KRT7+ population that can lead to expression of these mesoderm and trophoblast genes in different subpopulations of cells similar to what we saw analyzing the APA+ and APA- population.

It has been repeatedly shown that cell purification is an imperative paradigm for determining the intrinsic properties of stem cells and progenitors (Hoppe et al., 2014). Because previous studies were not based on purified progenitors their conclusions may have derived on the basis of co-existing populations, including a mes/mesendoderm population. Further, one mechanism of the emerging heterogeneity during BMP4 treatment of human ESCs was discovered as BMP4 can also activate components of the WNT pathway leading to induction of mesoderm genes and further a mesoderm subpopulation in the culture (Introduction section "BMP4-mediated trophoblast differentiation of human ESC").

In conclusion our data show that this differentiation approach represents a promising experimental method to follow cells from a precursor state to defined trophoblast cell types during human development with the ability to eliminate unwanted side products of the BMP4 mediated differentiation process.
Taken together the data obtained from the gene expression analysis allowed me to successfully address the first aim of my study.

The TF circuit of human APA+ trophoblast progenitors

I took a combined approach for identifying the network of genes that govern human trophoblast specification. I relied on purified populations of the BMP4 driven differentiation, as I have proved that these progenitors recapitulate most of the human TE and trophoblast progenitor characteristics (section above), in conjunction with global transcriptional, epigenetic mark, and time-course RNA-seq analyses. In addition, I used TF ChIP-seq and functional assays to prove my finding of the network (Figure VI-1).

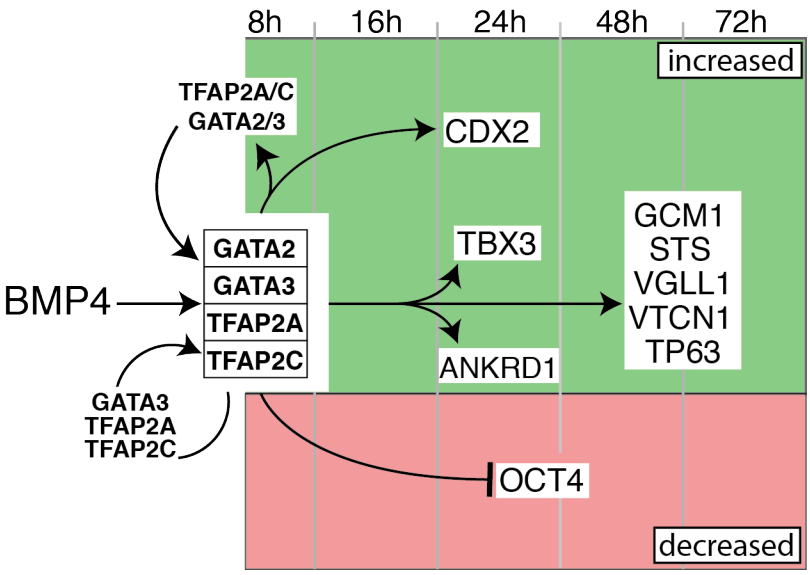

Figure VI-1: *Regulatory network of human trophoblast development*
A summary of the observed time-course increased (green) and decreased (red) trophoblast and pluripotency genes following BMP4 treatment of human ESCs. Reciprocal interactions (binding in the promoter regions) within the cohort of the early four TFs, the TEtra factors, are outlined according to ChIP-seq analysis.

The following evidence support the outline of this network including:

a. We showed that the four TFs, i.e. the TEtra - *GATA2*, *GATA3*, *TFAP2A* and *TFAP2C* are highly upregulated at the earliest time point analyzed (8 hrs BMP-4 treatment), and bind to the promoters of intermediate and late differentiation genes (Figure VI-1). Additional evidence supporting the classification of the TEtra TF as drivers of TE differentiation include that *Gata3* and *Tfap2c* are known key players of early mouse TE development (Introduction section "The transcriptional network of TE development").
b. The targets of the TEtra include a very well known mouse TE marker - *CDX2*, of the intermediate TF cohort, and *GCM1*, a key syncytiotrophoblast TF of the late cohort (Introduction section "Mouse placental development").
c. The TEtra TFs bind to the first intron of *OCT4* leading to repression of this key pluripotency gene. I have shown this by genome wide TF binding analysis of the Tetra using ChIP-Seq and subsequent knock-out of *GATA3*. This links the progression to a trophoblast fate with the inhibition of the pluripotency network.

Importantly, the logic of the TEtra TF network is built in away that the binding sites of these two families are in may cases overlapping. This, in my opinion, confers robustness to this differentiation pathway by redundancy. The human network in this regard seems to have features that are dissimilar to the mouse including the rewiring of the network;

while in the human *GATA2, GATA3, TFAP2A* and *TFAP2C* seem to regulate *CDX2*, in the mouse it is *Cdx2* that regulates *Tfap2c* and *Gata3* (Introduction section "The transcriptional network of TE development"). In line with our observations are recent single cell analysis studies of human blastocyst derived TE cells, which have indicated that in the human *GATA2* and *GATA3* are expressed earlier than *CDX2* (Petropoulos et al., 2016). This was further confirmed by another group using an *in vitro* 3D assay that mimics early human development (Deglincerti et al., 2016). Nevertheless, the full picture seems to be more complex as recent studies have placed mouse *Gata3* in parallel to *Cdx2* (Ralston et al., 2010) and *Tfap2c* upstream of *Cdx2* (Cao et al., 2015). Also my discoveries indicate that the TEtra network is auto-regulatory to some extend as they are found to be bound at their respective promoters during differentiation and *GATA3* depletion leads to upregulation of TFAP2A and TFAP2C, which could indicate a possible compensation mechanism (Figure V-27). One additional difference between mouse and human development is that *EOMES* was not detected upregulated in our data, while in the mouse it plays a critical role (Russ et al., 2000; Strumpf et al., 2005). Again, our observations are in line with *in vivo* measurements (Petropoulos et al., 2016).
Furthermore, our identified TF network could partly explain why mouse fibroblasts, when treated with a combination of 4 TFs, including *Gata3* and *Tfap2c* but not *Cdx2*, reprogram to a trophoblast stem-like state (Benchetrit et al., 2015; Kubaczka et al., 2015).

What is perplexing is that KO of *GATA3* alone seems to severely impair human trophoblast commitment (Figure V-26). I would not expect this severe phenotype because of the putative TF redundancies, but still I have not analyzed time points earlier than 8 hrs and the direct targets of the Smads transducers and therefore additional studies are necessary to understand in detail the function of *GATA3*.

The fact that there is redundancy in the key players of the early human trophoblast development by *GATA2* - *GATA3* and *TFAP2A* - *TFAP2C* could add stability to this developmental step. This suggests that this early cell fate decision is controlled in a similar way as the second wave of development, where the epiblast separates from the primitive endoderm, as the factors involved in this process have been shown to be *GATA4* – *GATA6* and *SOX7* – *SOX17* (Niakan et al., 2010; Blakeley et al., 2015). Although this hypothesis needs further validation this could mean that very critical early cell fate

decisions in the human embryo are controlled by a mechanism that relies on the redundancy of TFs. Taken together, these data, obtained by time-course gene expression analysis and TF ChIP-Seq analysis of TEtra, allowed me to address the second aim of my study.

Regulation of *CDX2* and *OCT4* during human trophoblast differentiation

My discoveries include novel features of regulation of trophoblast-pluripotency bifurcation. This began with my finding that *CDX2* is up- and *OCT4* is downregulated during differentiation from human ESCs to APA+ trophoblast progenitors. Then I observed that the TEtra TFs bind novel sites in the first introns of these genes. Furthermore, I found that in these sites the four TFs overlap. This indicates the presence of a regulatory element at this site and, according to the expression direction, I hypothesized that CDX2 and OCT4 are positively and negatively regulated, respectively.
In the case of CDX2, there has been a previous indication that this intronic site plays a role in mouse trophoblast development. It has been shown that binding of Gata3 and Tfap2c to this site can regulate *Cdx2* expression in mouse TSCs (Home et al., 2009; Cao et al., 2015). Another study showed that the *Cdx2* promoter region, as well as the first intron, are targets of Nanog and Oct4 in mouse ESCs and that during the transition of pluripotency to TSCs this region is bound by Cdx2 itself, which could lead to an autoregulation feedback loop (Chen et al., 2009). In mouse blastocysts *Cdx2* expression is controlled by a TE-specific enhancer (TEE) that lies upstream of the *Cdx2* promoter (Rayon et al., 2014), but this regulatory element is not essential for *Cdx2* expression in mouse TSCs. Taken together this indicates that two sites regulate Cdx2 expression sequentially: the TEE in early development and the intronic site later during development. However, the intronic region was ruled out as an extraembryonic specific enhancer as it is also active in mouse ESCs, shown by reporter assays. Yet, the explanation for the activity of the reporter assay in mouse ESCs could be that Nanog and Oct4 bind this fragment in mouse ESCs, therefore leading to a positive result in the reporter assay. The mechanism behind this could be that Nanog and Oct4 silence *Cdx2* in mouse ESCs.

In human ESCs *CDX2* is also bound by *NANOG* and *OCT4* in the promoter region and together they have been proposed to silence the gene in ESCs (Boyer et al., 2005). Furthermore, by reviewing the published ENCODE datasets of NANOG and OCT4 ChIP-Seq I found that the same intronic region of *CDX2*, which is bound by the Tetra TFs, is also bound and possibly regulated by NANOG and OCT4 in human ESCs.

Taken together, this has led me to propose a model for human *CDX2* regulation (Figure VI-2) that takes into account previous findings and my data. It proposes mutual exclusive regulation of *CDX2*, either in the direction of repression in pluripotent cells or in the direction of induction when OCT4/NANOG are replaced by the TEtra TFs.

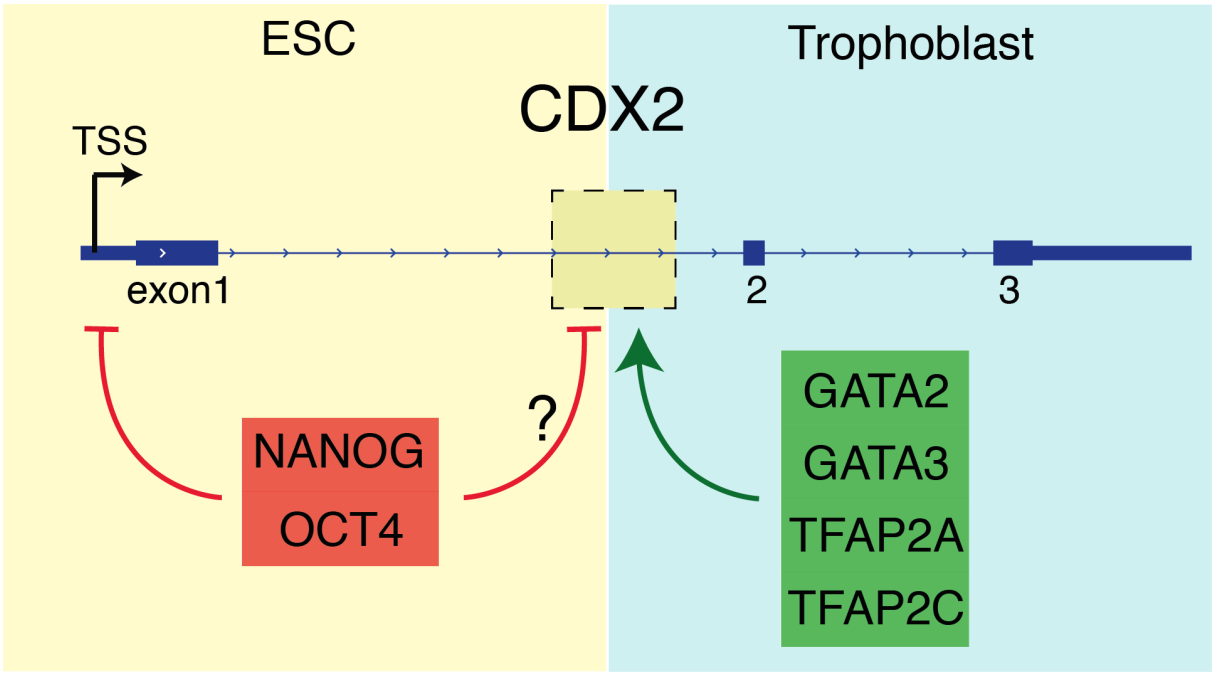

Figure VI-2: *Proposed regulation mode of CDX2*
Shown is the genetic structure of the *CDX2* locus with its three exons. Left (yellow) indicates the regulation of *CDX2* in human ESCs. Right (blue) shows the regulation in trophoblast progenitors. Green indicates activation and red indicates repression.

An unexpected finding is that following differentiation of the GATA3 KO cells with BMP4 I have not noted reduced levels of *CDX2* compared to control. Surprisingly I noted an opposite effect - *CDX2* is upregulated when GATA3 is absent during BMP4 treatment of human ESCs. One explanation for this phenomenon could be the before mentioned compensation by TFAP2A/C or another compensatory route that effects the level of *CDX2* in this case.

Although the regulation of *Oct4* / *OCT4* was rather extensively studied and several regulatory regions were identified, including promoter upstream proximal and distal enhancers, I am not aware of a prior discovery of regulatory elements in the first intron as I have made (Figure VI-3).

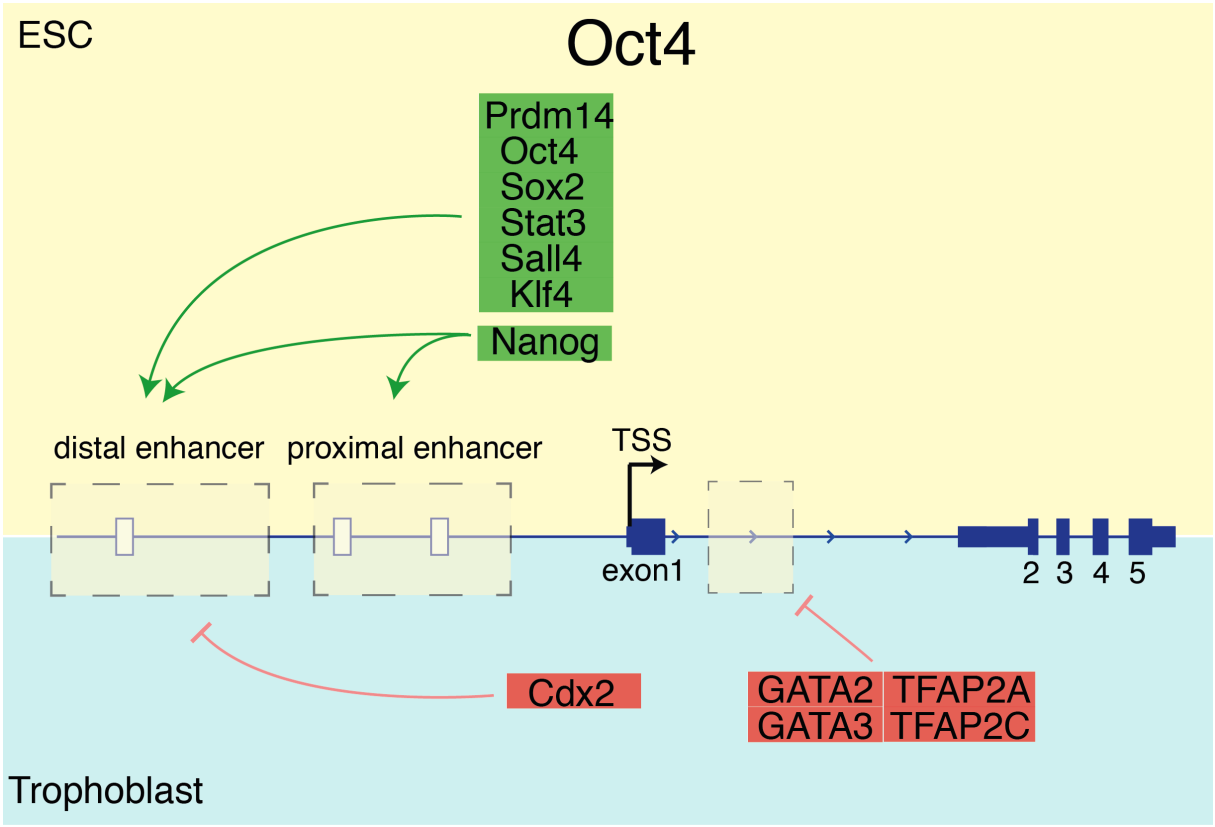

Figure VI-3: Proposed *regulation mode of OCT4*
Shown is the genomic assembly of the *OCT4* locus with its five exons (blue), the distal and the proximal enhancer. Upper and bottom panes outlines the modes of positive and negative regulation in human ESCs and trophoblast progenitors, respectively. TSS = transcription start site; green indicates activating and red indicates repressing factors.

The functional importance of the TEtra binding to the first intron of *OCT4* requires functional validation. I have begun this by using the *GATA3* KO cell line, which verified the direction of negative regulation of *OCT4* by the TEtra TFs, because the levels of *OCT4* increased in this cell line compared to control (Figure V-27). Removal of the GATA and TFAP2 binding sites should be performed in the endogenous locus or using a reporter assay. My hypothesis is that without GATA and TFAP2 binding sites *OCT4* levels will be maintained. One reason why I did not perform such experiments is that I noted several

tandem GATA and TFAP2 motifs and several bound regions within the intron. Therefore, a thorough mutational screen is necessary, and this was beyond the scope of my thesis.

Epigenetic regulation of trophoblast differentiation

My study shows that the early and many intermediate key regulators of the trophoblast differentiation have CpG rich promoters and are marked with both, the activating H3K4me3 and the repressing H3K27me3 marks concomitantly in undifferentiated human ESCs. I further showed that these bivalent TFs are upregulated very rapidly during differentiation and loose the repressing H3K27me3 mark within 2.5 days of differentiation. In contrast to this, I observed that some of the key trophoblast regulatory TFs, which are upregulated later during differentiation, are CpG poor and harbor no activating or repressing marks in undifferentiated cells. In this group I found *GCM1, TP63* and *VGLL1*. In the progenitors only *TP63* acquires low levels of H3K4me3, while *GCM1* and *VGLL1* remain unmarked. This suggests that in genes harboring CpG poor promoters, transcriptional activation precedes the acquisition of activating H3K4me3 marks. Furthermore, very little changes, if any, take place in methylated CpG islands during the transition from pluripotent to APA+ trophoblast cells. The changes that I did observe were mainly de-methylation of CpGs in CpG poor genes, which become activated during differentiation. Interestingly, the site of de-methylation often co-localizes with the TEtra peak sites (Figure V-32), indicating that binding of TFs either leads to active de-methylation or interferes with the maintenance of CpG methylation. The epigenetic characteristics of the genes activated in the trophoblast progenitors fits with their order of induction. It has been shown that bivalent CpG rich promoters are typically developmental regulators, and that CpG poor genes, which are not marked with histone modifications, typically encode for structural proteins (Introduction section "Epigenetic profiles of PSCs"). However, I found a novel feature in this regard - also late trophoblast specific TFs included in this group, indicating that I discovered a novel form of developmental regulation. Additionally, it was shown that gene expression and DNA methylation negatively correlate during differentiation (Xie et al., 2013), but in our relatively early time-point of analysis we did not observe this phenomenon as I still detect DNA-methylation marks at many CpG poor promoters of expressed genes in the APA+ population. Therefore I conclude that the transcriptional activation of these genes

happens before the erasure of DNA methylation marks at their promoters. All these observations allow me to successfully address the third aim of my study.

Biomedical relevance of the discovery of the TEtra TFs

My identification of the TF circuit that underlies human trophoblast differentiation opens a path to understand developmental pathologies of the placenta. This is because it provides a platform to discover their molecular etiology that may in many cases originate from misregulation of early transcriptional networks. Breaking grounds in understanding placental impairments and disorders causing miscarriages or preeclampsia is important for offspring and mother's health.

Preeclampsia for example is a hypertensive disorder, which is usually diagnosed by proteinurea and hypertension after the 20th week of gestation (Bulletins--Obstetrics, 2002). Importantly the link between preeclampsia and members of the TEtra has been made already, although it was almost unnoticed. Published evidences include datasets showing that *GATA2* and *TFAP2A* are both downregulated in preeclampsia compared to placentas from normal pregnancies. Furthermore, a strong enrichment of the *AP2* (*TFAP2*) binding motif was noted at many downregulated genes in preeclampsia placentas (Sober et al., 2015).

Finally, our findings could also be pertinent for reproductive therapies by artificial reproductive technologies (ART) such as in vitro fertilization (IVF). A study in mice found that ART increases the chances for maldevelopment and dysfunction of the placenta, which is linked to reduced weight of the fetal mice (Chen et al., 2015). Examining this data they found misregulation of *GCM1* following IVF compared to normal pregnancies, which we found to be a target of GATA2 and GATA3 during trophoblast differentiation.

Taken together, this provides an initial indication that by discovering the TEtra, I have made a first crucial step in the direction of understanding early placental and pregnancy defects, that this could translate to improve ARTs and treatment of conditions with impaired placental development.

VII. Acknowledgements

I would like to thank Dr. Micha Drukker for giving me the chance to work on this interesting project. I would like to thank him for scientific input and discussions but also for general advice and mentoring.

I would also like to thank Prof. Dr. Magdalena Götz for being my supervisor at the university, for being part of my thesis advisory committee and for valuable input and discussions.

Further, I want to express my gratitude to Prof. Dr. Gunnar Schotta for being a member of my thesis advisory board and for his advice and discussions.

I also want to thank Dr. Tobias Straub and Dr. Steffen Sass for help with the data analysis and scientific input.

Moreover, I want to thank Dr. Dmitry Shaposhnikov for fruitful collaboration, scientific advice and exhausting skiing days.

Additionally I want to thank all former and present members of the Drukker lab for creating such a nice working atmosphere and for numerous social events.

I am likewise grateful for many nice scientific and non-scientific interactions with my fellow doctoral students from all over the Helmholtz Center Munich.

Last but not least, I want to thank my family for supporting me in whatever I was doing and I would like to welcome Paula to the world of science.

VIII. References

Ahmed, K., Dehghani, H., Rugg-Gunn, P., Fussner, E., Rossant, J., and Bazett-Jones, D.P. (2010). Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. PloS one *5*, e10531.

Allendorph, G.P., Vale, W.W., and Choe, S. (2006). Structure of the ternary signaling complex of a TGF-beta superfamily member. Proceedings of the National Academy of Sciences of the United States of America *103*, 7643-7648.

Amita, M., Adachi, K., Alexenko, A.P., Sinha, S., Schust, D.J., Schulz, L.C., Roberts, R.M., and Ezashi, T. (2013). Complete and unidirectional conversion of human embryonic stem cells to trophoblast by BMP4. Proceedings of the National Academy of Sciences of the United States of America *110*, E1212-1221.

Anson-Cartwright, L., Dawson, K., Holmyard, D., Fisher, S.J., Lazzarini, R.A., and Cross, J.C. (2000). The glial cells missing-1 protein is essential for branching morphogenesis in the chorioallantoic placenta. Nature genetics *25*, 311-314.

Arman, E., Haffner-Krausz, R., Chen, Y., Heath, J.K., and Lonai, P. (1998). Targeted disruption of fibroblast growth factor (FGF) receptor 2 suggests a role for FGF signaling in pregastrulation mammalian development. Proc Natl Acad Sci USA *95*, 5082-5087.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes & development *17*, 126-140.

Bai, Q., Assou, S., Haouzi, D., Ramirez, J.M., Monzo, C., Becker, F., Gerbal-Chaloin, S., Hamamah, S., and De Vos, J. (2012). Dissecting the first transcriptional divergence during human embryonic development. Stem cell reviews *8*, 150-162.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic acids research *37*, W202-208.

Basyuk, E., Cross, J.C., Corbin, J., Nakayama, H., Hunter, P., Nait-Oumesmar, B., and Lazzarini, R.A. (1999). Murine Gcm1 gene is expressed in a subset of placental trophoblast cells. Developmental dynamics : an official publication of the American Association of Anatomists *214*, 303-311.

Bedzhov, I., Graham, S.J., Leung, C.Y., and Zernicka-Goetz, M. (2014). Developmental plasticity, cell fate specification and morphogenesis in the early mouse embryo. Philosophical transactions of the Royal Society of London Series B, Biological sciences *369*.

Benchetrit, H., Herman, S., van Wietmarschen, N., Wu, T., Makedonski, K., Maoz, N., Yom Tov, N., Stave, D., Lasry, R., Zayat, V.*, et al.* (2015). Extensive Nuclear Reprogramming Underlies Lineage Conversion into Functional Trophoblast Stem-like Cells. Cell stem cell *17*, 543-556.

Beppu, H., Kawabata, M., Hamamoto, T., Chytil, A., Minowa, O., Noda, T., and Miyazono, K. (2000). BMP type II receptor is required for gastrulation and early development of mouse embryos. Developmental biology *221*, 249-258.

Berezowsky, J., Zbieranowski, I., Demers, J., and Murray, D. (1995). DNA ploidy of hydatidiform moles and nonmolar conceptuses: a study using flow and tissue section image cytometry. Mod Pathol *8*, 775-781.

Bernardo, A.S., Faial, T., Gardner, L., Niakan, K.K., Ortmann, D., Senner, C.E., Callery, E.M., Trotter, M.W., Hemberger, M., Smith, J.C.*, et al.* (2011). BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages. Cell stem cell *9*, 144-155.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bhaumik, S.R., Smith, E., and Shilatifard, A. (2007). Covalent modifications of histones during development and disease pathogenesis. Nat Struct Mol Biol *14*, 1008-1016.

Blakeley, P., Fogarty, N.M., Del Valle, I., Wamaitha, S.E., Hu, T.X., Elder, K., Snell, P., Christie, L., Robson, P., and Niakan, K.K. (2015). Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. Development.

Bourgon, R., Gentleman, R., and Huber, W. (2010). Independent filtering increases detection power for high-throughput experiments. Proceedings of the National Academy of Sciences of the United States of America *107*, 9546-9551.

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.*, et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K.*, et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Bulletins--Obstetrics, A.C.o.P. (2002). ACOG practice bulletin. Diagnosis and management of preeclampsia and eclampsia. Number 33, January 2002. Obstetrics and gynecology *99*, 159-167.

Cairns, B.R. (2007). Chromatin remodeling: insights and intrigue from single-molecule studies. Nat Struct Mol Biol *14*, 989-996.

Cao, Z., Carey, T.S., Ganguly, A., Wilson, C.A., Paul, S., and Knott, J.G. (2015). Transcription factor AP-2gamma induces early Cdx2 expression and represses HIPPO signaling to specify the trophectoderm lineage. Development *142*, 1606-1615.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643-655.

Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. Developmental cell *10*, 615-624.

Chen, L., Yabuuchi, A., Eminli, S., Takeuchi, A., Lu, C.W., Hochedlinger, K., and Daley, G.Q. (2009). Cross-regulation of the Nanog and Cdx2 promoters. Cell research *19*, 1052-1061.

Chen, S., Sun, F.Z., Huang, X., Wang, X., Tang, N., Zhu, B., and Li, B. (2015). Assisted reproduction causes placental maldevelopment and dysfunction linked to reduced fetal weight in mice. Scientific reports *5*, 10596.

Christophersen, N.S., and Helin, K. (2010). Epigenetic control of embryonic stem cell fate. The Journal of experimental medicine *207*, 2287-2295.

Chu, G.C., Dunn, N.R., Anderson, D.C., Oxburgh, L., and Robertson, E.J. (2004). Differential requirements for Smad4 in TGFbeta-dependent patterning of the early mouse embryo. Development *131*, 3501-3512.

Cockburn, K., and Rossant, J. (2010). Making the blastocyst: lessons from the mouse. The Journal of clinical investigation *120*, 995-1003.

Cross, J.C., Simmons, D.G., and Watson, E.D. (2003). Chorioallantoic morphogenesis and formation of the placental villous tree. Annals of the New York Academy of Sciences *995*, 84-93.

Das, P., Ezashi, T., Schulz, L.C., Westfall, S.D., Livingston, K.A., and Roberts, R.M. (2007). Effects of fgf2 and oxygen in the bmp4-driven differentiation of trophoblast from human embryonic stem cells. Stem cell research *1*, 61-74.

De Gobbi, M., Garrick, D., Lynch, M., Vernimmen, D., Hughes, J.R., Goardon, N., Luc, S., Lower, K.M., Sloane-Stanley, J.A., Pina, C.*, et al.* (2011). Generation of bivalent chromatin domains during cell fate decisions. Epigenetics & chromatin *4*, 9.

De Paepe, C., Cauffman, G., Verloes, A., Sterckx, J., Devroey, P., Tournaye, H., Liebaers, I., and Van de Velde, H. (2013). Human trophectoderm cells are not yet committed. Human reproduction *28*, 740-749.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes & development *25*, 1010-1022.

Deglincerti, A., Croft, G.F., Pietila, L.N., Zernicka-Goetz, M., Siggia, E.D., and Brivanlou, A.H. (2016). Self-organization of the in vitro attached human embryo. Nature *533*, 251-254.

Denslow, S.A., and Wade, P.A. (2007). The human Mi-2/NuRD complex and gene regulation. Oncogene *26*, 5433-5438.

Dietrich, J.E., and Hiiragi, T. (2007). Stochastic patterning in the mouse pre-implantation embryo. Development *134*, 4219-4231.

Donnison, M., Beaton, A., Davey, H.W., Broadhurst, R., L'Huillier, P., and Pfeffer, P.L. (2005). Loss of the extraembryonic ectoderm in Elf5 mutants leads to defects in embryonic patterning. Development *132*, 2299-2308.

Drukker, M., Tang, C., Ardehali, R., Rinkevich, Y., Seita, J., Lee, A.S., Mosley, A.R., Weissman, I.L., and Soen, Y. (2012). Isolation of primitive endoderm, mesoderm, vascular endothelial and trophoblast progenitors from human pluripotent stem cells. Nature biotechnology *30*, 531-542.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D., Buetow, K.H.*, et al.* (2008). Global transcription in pluripotent embryonic stem cells. Cell stem cell *2*, 437-447.

Enders, A.C. (1976). Cytology of human early implantation. Research in reproduction *8*, 1-2.

Ezashi, T., Telugu, B.P., and Roberts, R.M. (2012). Model systems for studying trophoblast differentiation from human pluripotent stem cells. Cell and tissue research *349*, 809-824.

Feng, S., Jacobsen, S.E., and Reik, W. (2010). Epigenetic reprogramming in plant and animal development. Science *330*, 622-627.

Finn, C.A., and McLaren, A. (1967). A study of the early stages of implantation in mice. Journal of reproduction and fertility *13*, 259-267.

Fouse, S.D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G. (2008). Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. Cell stem cell *2*, 160-169.

Frum, T., Halbisen, M.A., Wang, C., Amiri, H., Robson, P., and Ralston, A. (2013). Oct4 Cell-Autonomously Promotes Primitive Endoderm Development in the Mouse Blastocyst. Developmental cell *25*.

Fujiwara, T., Dehart, D.B., Sulik, K.K., and Hogan, B.L. (2002). Distinct requirements for extra-embryonic and embryonic bone morphogenetic protein 4 in the formation of the node and primitive streak and coordination of left-right asymmetry in the mouse. Development *129*, 4685-4696.

Gardner, R.L. (1985). Regeneration of endoderm from primitive ectoderm in the mouse embryo: fact or artifact? Journal of embryology and experimental morphology *88*, 303-326.

Gardner, R.L. (2001). Specification of embryonic axes begins before cleavage in normal mouse development. Development *128*, 839-847.

Gardner, R.L., and Johnson, M.H. (1973). Investigation of early mammalian development using interspecific chimaeras between rat and mouse. Nature: New biology *246*, 86-89.

Gardner, R.L., and Rossant, J. (1979). Investigation of the fate of 4-5 day post-coitum mouse inner cell mass cells by blastocyst injection. Journal of embryology and experimental morphology *52*, 141-152.

Genbacev, O., Lamb, J.D., Prakobphol, A., Donne, M., McMaster, M.T., and Fisher, S.J. (2013). Human trophoblast progenitors: where do they reside? Seminars in reproductive medicine *31*, 56-61.

Georgiades, P., and Rossant, J. (2006). Ets2 is necessary in trophoblast for normal embryonic anteroposterior axis development. Development *133*, 1059-1068.

Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R.*, et al.* (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell *153*, 1149-1163.

Gonzalez, F., Zhu, Z., Shi, Z.D., Lelli, K., Verma, N., Li, Q.V., and Huangfu, D. (2014). An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. Cell stem cell *15*, 215-226.

Graham, S.J., Wicher, K.B., Jedrusik, A., Guo, G., Herath, W., Robson, P., and Zernicka-Goetz, M. (2014). BMP signalling regulates the pre-implantation development of extra-embryonic cell lineages in the mouse embryo. Nature communications *5*, 5667.

Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell *145*, 423-434.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell *38*, 576-589.

Hertig, A.T., and Rock, J. (1973). Searching for early fertilized human ova. Gynecologic investigation *4*, 121-139.

Hertig, A.T., Rock, J., Adams, E.C., and Menkin, M.C. (1959). Thirty-four fertilized human ova, good, bad and indifferent, recovered from 210 women of known fertility; a study of biologic wastage in early human pregnancy. Pediatrics *23*, 202-211.

Herzog, M. (1909). A contribution to our knowledge of the earliest known stages of placentation and embryonic development in man. American Journal of Anatomy *9*, 361-400.

Home, P., Ray, S., Dutta, D., Bronshteyn, I., Larson, M., and Paul, S. (2009). GATA3 is selectively expressed in the trophectoderm of peri-implantation embryo and directly regulates Cdx2 gene expression. The Journal of biological chemistry *284*, 28729-28737.

Hoppe, P.S., Coutu, D.L., and Schroeder, T. (2014). Single-cell technologies sharpen up mammalian stem cell research. Nature cell biology *16*, 919-927.

Ito, N., Nomura, S., Iwase, A., Ito, T., Ino, K., Nagasaka, T., Tsujimoto, M., Kobayashi, M., and Mizutani, S. (2003). Ultrastructural localization of aminopeptidase A/angiotensinase and placental leucine aminopeptidase/oxytocinase in chorionic villi of human placenta. Early human development *71*, 29-37.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Johnson, M.H., and Ziomek, C.A. (1981). The foundation of two distinct cell lineages within the mouse morula. Cell *24*, 71-80.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome research *12*, 996-1006.

Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S., and Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. Nature genetics *41*, 376-381.

Koutsourakis, M., Langeveld, A., Patient, R., Beddington, R., and Grosveld, F. (1999). The transcription factor GATA6 is essential for early extraembryonic development. Development *126*, 723-732.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science *324*, 929-930.

Kubaczka, C., Senner, C.E., Cierlitza, M., Arauzo-Bravo, M.J., Kuckenberg, P., Peitz, M., Hemberger, M., and Schorle, H. (2015). Direct Induction of Trophoblast Stem Cells from Murine Fibroblasts. Cell stem cell *17*, 557-568.

Kurek, D., Neagu, A., Tastemel, M., Tuysuz, N., Lehmann, J., van de Werken, H.J., Philipsen, S., van der Linden, R., Maas, A., van, I.W.F*., et al.* (2015). Endogenous WNT signals mediate BMP-induced and spontaneous differentiation of epiblast stem cells and human embryonic stem cells. Stem cell reports *4*, 114-128.

Kurimoto, K., Yabuta, Y., Ohinata, Y., Ono, Y., Uno, K.D., Yamada, R.G., Ueda, H.R., and Saitou, M. (2006). An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. Nucleic acids research *34*, e42.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Latos, P.A., and Hemberger, M. (2014). Review: the transcriptional and signalling networks of mouse trophoblast stem cells. Placenta *35 Suppl*, S81-85.

Lau, P.N., and Cheung, P. (2011). Histone code pathway involving H3 S28 phosphorylation and K27 acetylation activates transcription and antagonizes polycomb

silencing. Proceedings of the National Academy of Sciences of the United States of America *108*, 2801-2806.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome biology *15*, R29.

Lawson, P.T., Lovaglio, J., Liu, C.C., and Lipkin, E.W. (1993). A histomorphometric comparison of bone in young growing rats fed an elemental diet versus a chemically defined polymeric diet. Journal of the American College of Nutrition *12*, 53-60.

Lee, C.Q., Gardner, L., Turco, M., Zhao, N., Murray, M.J., Coleman, N., Rossant, J., Hemberger, M., and Moffett, A. (2016). What Is Trophoblast? A Combination of Criteria Define Human First-Trimester Trophoblast. Stem cell reports *6*, 257-272.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.*, et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Current biology : CB *13*, 1192-1200.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915-926.

Lichtner, B., Knaus, P., Lehrach, H., and Adjaye, J. (2013). BMP10 as a potent inducer of trophoblast differentiation in human embryonic and induced pluripotent stem cells. Biomaterials *34*, 9789-9802.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

MacAuley, A., Cross, J.C., and Werb, Z. (1998). Reprogramming the cell cycle for endoreduplication in rodent trophoblast cells. Mol Biol Cell *9*, 795-807.

Marchand, M., Horcajadas, J.A., Esteban, F.J., McElroy, S.L., Fisher, S.J., and Giudice, L.C. (2011). Transcriptomic signature of trophoblast differentiation in a human embryonic stem cell model. Biology of reproduction *84*, 1258-1271.

Marks, P., Rifkind, R.A., Richon, V.M., Breslow, R., Miller, T., and Kelly, W.K. (2001). Histone deacetylases and cancer: causes and therapies. Nature reviews Cancer *1*, 194-202.

Meilhac, S.M., Adams, R.J., Morris, S.A., Danckaert, A., Le Garrec, J.F., and Zernicka-Goetz, M. (2009). Active cell movements coupled to positional induction are involved in lineage segregation in the mouse blastocyst. Developmental biology *331*, 210-221.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B.*, et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003a). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631-642.

Mitsui, T., Nomura, S., Okada, M., Ohno, Y., Kobayashi, H., Nakashima, Y., Murata, Y., Takeuchi, M., Kuno, N., Nagasaka, T.*, et al.* (2003b). Hypertension and angiotensin II hypersensitivity in aminopeptidase A-deficient mice. Molecular medicine *9*, 57-62.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schubeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Molecular cell *30*, 755-766.

Morikawa, M., Koinuma, D., Tsutsumi, S., Vasilaki, E., Kanki, Y., Heldin, C.H., Aburatani, H., and Miyazono, K. (2011). ChIP-seq reveals cell type-specific binding patterns of BMP-specific Smads and a novel binding motif. Nucleic acids research *39*, 8712-8727.

Motosugi, N., Bauer, T., Polanski, Z., Solter, D., and Hiiragi, T. (2005). Polarity of the mouse embryo is established at blastocyst and is not prepatterned. Genes & development *19*, 1081-1092.

Mukhopadhyay, P., Webb, C.L., Warner, D.R., Greene, R.M., and Pisano, M.M. (2008). BMP signaling dynamics in embryonic orofacial tissue. Journal of cellular physiology *216*, 771-779.

Ng, R.K., Dean, W., Dawson, C., Lucifero, D., Madeja, Z., Reik, W., and Hemberger, M. (2008). Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. Nature cell biology *10*, 1280-1290.

Niakan, K.K., and Eggan, K. (2013). Analysis of human embryos from zygote to blastocyst reveals distinct gene expression patterns relative to the mouse. Developmental biology *375*, 54-64.

Niakan, K.K., Ji, H., Maehr, R., Vokes, S.A., Rodolfa, K.T., Sherwood, R.I., Yamaki, M., Dimos, J.T., Chen, A.E., Melton, D.A.*, et al.* (2010). Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. Genes & development *24*, 312-326.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379-391.

Nishioka, N., Inoue, K., Adachi, K., Kiyonari, H., Ota, M., Ralston, A., Yabuta, N., Hirahara, S., Stephenson, R.O., Ogonuki, N., *et al.* (2009). The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. Developmental cell *16*, 398-410.

Nishioka, N., Yamamoto, S., Kiyonari, H., Sato, H., Sawada, A., Ota, M., Nakao, K., and Sasaki, H. (2008). Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. Mechanisms of development *125*, 270-283.

Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell *123*, 917-929.

Nordhoff, V., Hubner, K., Bauer, A., Orlova, I., Malapetsa, A., and Scholer, H.R. (2001). Comparative analysis of human, bovine, and murine Oct-4 upstream promoter sequences. Mammalian genome : official journal of the International Mammalian Genome Society *12*, 309-317.

Norwitz, E.R., Schust, D.J., and Fisher, S.J. (2001). Implantation and the survival of early pregnancy. The New England journal of medicine *345*, 1400-1408.

O'Leary, T., Heindryckx, B., Lierman, S., van Bruggen, D., Goeman, J.J., Vandewoestyne, M., Deforce, D., de Sousa Lopes, S.M., and De Sutter, P. (2012). Tracking the progression of the human inner cell mass during embryonic stem cell derivation. Nature biotechnology *30*, 278-282.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Paige, S.L., Thomas, S., Stoick-Cooper, C.L., Wang, H., Maves, L., Sandstrom, R., Pabon, L., Reinecke, H., Pratt, G., Keller, G., *et al.* (2012). A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. Cell *151*, 221-232.

Palmieri, S.L., Peter, W., Hess, H., and Scholer, H.R. (1994). Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. Developmental biology *166*.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell stem cell *1*, 299-312.

Papaioannou, V.E., McBurney, M.W., Gardner, R.L., and Evans, M.J. (1975). Fate of teratocarcinoma cells injected into early mouse embryos. Nature *258*, 70-73.

Park, S.H., Park, S.H., Kook, M.C., Kim, E.Y., Park, S., and Lim, J.H. (2004). Ultrastructure of human embryonic stem cells and spontaneous and retinoic acid-induced differentiating cells. Ultrastructural pathology *28*, 229-238.

Pasini, D., Cloos, P.A., Walfridsson, J., Olsson, L., Bukowski, J.P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature *464*, 306-310.

Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P.*, et al.* (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature *473*, 394-397.

Paweletz, N. (2001). Walther Flemming: pioneer of mitosis research. Nature reviews Molecular cell biology *2*, 72-75.

Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A.*, et al.* (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell *107*, 323-337.

Petropoulos, S., Edsgard, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Reyes, A.P., Linnarsson, S., Sandberg, R., and Lanner, F. (2016). Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. Cell *167*, 285.

Piotrowska, K., Wianny, F., Pedersen, R.A., and Zernicka-Goetz, M. (2001). Blastomeres arising from the first cleavage division have distinguishable fates in normal mouse development. Development *128*, 3739-3748.

Piotrowska, K., and Zernicka-Goetz, M. (2001). Role for sperm in spatial patterning of the early mouse embryo. Nature *409*, 517-521.

Plusa, B., Frankenberg, S., Chalmers, A., Hadjantonakis, A.K., Moore, C.A., Papalopulu, N., Papaioannou, V.E., Glover, D.M., and Zernicka-Goetz, M. (2005). Downregulation of Par3 and aPKC function directs cells towards the ICM in the preimplantation mouse embryo. Journal of cell science *118*, 505-515.

Plusa, B., Piliszek, A., Frankenberg, S., Artus, J., and Hadjantonakis, A.K. (2008). Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. Development *135*, 3081-3091.

Polydorou, C., and Georgiades, P. (2013). Ets2-dependent trophoblast signalling is required for gastrulation progression after primitive streak initiation. Nature communications *4*, 1658.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Ralston, A., Cox, B.J., Nishioka, N., Sasaki, H., Chea, E., Rugg-Gunn, P., Guo, G., Robson, P., Draper, J.S., and Rossant, J. (2010). Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. Development *137*, 395-403.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA

methyltransferase 3a. Proceedings of the National Academy of Sciences of the United States of America *97*, 5237-5242.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature protocols *8*, 2281-2308.

Rao, C.V., and Alsip, N.L. (2001). Use of the rat model to study hCG/LH effects on uterine blood flow. Seminars in reproductive medicine *19*, 75-85.

Rayon, T., Menchero, S., Nieto, A., Xenopoulos, P., Crespo, M., Cockburn, K., Canon, S., Sasaki, H., Hadjantonakis, A.K., de la Pompa, J.L.*, et al.* (2014). Notch and hippo converge on Cdx2 to specify the trophectoderm lineage in the mouse blastocyst. Developmental cell *30*, 410-422.

Riley, P., Anson-Cartwright, L., and Cross, J.C. (1998). The Hand1 bHLH transcription factor is essential for placentation and cardiac morphogenesis. Nature genetics *18*, 271-275.

Ringler, G.E., and Strauss, J.F., 3rd (1990). In vitro systems for the study of human placental endocrine function. Endocrine reviews *11*, 105-123.

Roberts, R.M., Loh, K.M., Amita, M., Bernardo, A.S., Adachi, K., Alexenko, A.P., Schust, D.J., Schulz, L.C., Telugu, B.P., Ezashi, T.*, et al.* (2014). Differentiation of trophoblast cells from human embryonic stem cells: to be or not to be? Reproduction *147*, D1-12.

Roode, M., Blair, K., Snell, P., Elder, K., Marchant, S., Smith, A., and Nichols, J. (2012). Human hypoblast formation is not dependent on FGF signalling. Developmental biology *361*, 358-363.

Rosner, M.H., Vigano, M.A., Ozato, K., Timmons, P.M., Poirier, F., Rigby, P.W., and Staudt, L.M. (1990). A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. Nature *345*.

Rossant, J. (2015). Mouse and human blastocyst-derived stem cells: vive les differences. Development *142*, 9-12.

Rossant, J., and Cross, J.C. (2001). Placental development: lessons from mouse mutants. Nature reviews Genetics *2*, 538-548.

Rossetto, D., Avvakumov, N., and Cote, J. (2012). Histone phosphorylation: a chromatin modification involved in diverse nuclear events. Epigenetics : official journal of the DNA Methylation Society *7*, 1098-1108.

Russ, A.P., Wattler, S., Colledge, W.H., Aparicio, S.A., Carlton, M.B., Pearce, J.J., Barton, S.C., Surani, M.A., Ryan, K., Nehls, M.C.*, et al.* (2000). Eomesodermin is required for mouse trophoblast development and mesoderm formation. Nature *404*, 95-99.

Sarma, K., and Reinberg, D. (2005). Histone variants meet their match. Nature reviews Molecular cell biology *6*, 139-149.

Schaniel, C., Ang, Y.S., Ratnakumar, K., Cormier, C., James, T., Bernstein, E., Lemischka, I.R., and Paddison, P.J. (2009). Smarcc1/Baf155 couples self-renewal gene repression with changes in chromatin structure in mouse embryonic stem cells. Stem cells *27*, 2979-2991.

Scholer, H.R., Balling, R., Hatzopoulos, A.K., Suzuki, N., and Gruss, P. (1989). Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. The EMBO journal *8*.

Schreiber, J., Riethmacher-Sonnenberg, E., Riethmacher, D., Tuerk, E.E., Enderich, J., Bosl, M.R., and Wegner, M. (2000). Placental failure in mice lacking the mammalian homolog of glial cells missing, GCMa. Molecular and cellular biology *20*, 2466-2474.

Senner, C.E., and Hemberger, M. (2010). Regulation of early trophoblast differentiation - lessons from the mouse. Placenta *31*, 944-950.

Shahbazian, M.D., and Grunstein, M. (2007). Functions of site-specific histone acetylation and deacetylation. Annual review of biochemistry *76*, 75-100.

Shi, Q.J., Lei, Z.M., Rao, C.V., and Lin, J. (1993). Novel role of human chorionic gonadotropin in differentiation of human cytotrophoblasts. Endocrinology *132*, 1387-1395.

Shilatifard, A. (2006). Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. Annual review of biochemistry *75*, 243-269.

Shore, E.M., and Kaplan, F.S. (2010). Inherited human diseases of heterotopic bone formation. Nature reviews Rheumatology *6*, 518-527.

Simmons, D.G., Fortier, A.L., and Cross, J.C. (2007). Diverse subtypes and developmental origins of trophoblast giant cells in the mouse placenta. Developmental biology *304*, 567-578.

Skamagki, M., Wicher, K.B., Jedrusik, A., Ganguly, S., and Zernicka-Goetz, M. (2013). Asymmetric localization of Cdx2 mRNA during the first cell-fate decision in early mouse development. Cell reports *3*, 442-457.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology *3*, Article3.

Sober, S., Reiman, M., Kikas, T., Rull, K., Inno, R., Vaas, P., Teesalu, P., Marti, J.M., Mattila, P., and Laan, M. (2015). Extensive shift in placental transcriptome profile in preeclampsia and placental origin of adverse pregnancy outcomes. Scientific reports *5*, 13336.

Strumpf, D., Mao, C.A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F., and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development *132*, 2093-2102.

Sudheer, S., Bhushan, R., Fauler, B., Lehrach, H., and Adjaye, J. (2012). FGF inhibition directs BMP4-mediated differentiation of human embryonic stem cells to syncytiotrophoblast. Stem cells and development *21*, 2987-3000.

Sutherland, A.E., Speed, T.P., and Calarco, P.G. (1990). Inner cell allocation in the mouse morula: the role of oriented division during fourth cleavage. Developmental biology *137*, 13-25.

Szulwach, K.E., Li, X., Li, Y., Song, C.X., Han, J.W., Kim, S., Namburi, S., Hermetz, K., Kim, J.J., Rudd, M.K.*, et al.* (2011). Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. PLoS genetics *7*, e1002154.

Tabansky, I., Lenarcic, A., Draft, R.W., Loulier, K., Keskin, D.B., Rosains, J., Rivera-Feliciano, J., Lichtman, J.W., Livet, J., Stern, J.N.*, et al.* (2013). Developmental bias in cleavage-stage mouse blastomeres. Current biology : CB *23*, 21-31.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L.*, et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.

Tanaka, S., Kunath, T., Hadjantonakis, A.K., Nagy, A., and Rossant, J. (1998). Promotion of trophoblast stem cell proliferation by FGF4. Science *282*, 2072-2075.

Tang, C., Lee, A.S., Volkmer, J.P., Sahoo, D., Nag, D., Mosley, A.R., Inlay, M.A., Ardehali, R., Chavez, S.L., Pera, R.R.*, et al.* (2011). An antibody against SSEA-5 glycan on human pluripotent stem cells enables removal of teratoma-forming cells. Nature biotechnology *29*, 829-834.

Telugu, B.P., Adachi, K., Schlitt, J.M., Ezashi, T., Schust, D.J., Roberts, R.M., and Schulz, L.C. (2013). Comparison of extravillous trophoblast cells derived from human embryonic stem cells and from first trimester human placentas. Placenta *34*, 536-543.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196-199.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145-1147.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Uy, G.D., Downs, K.M., and Gardner, R.L. (2002). Inhibition of trophoblast stem cell potential in chorionic ectoderm coincides with occlusion of the ectoplacental cavity in the mouse. Development *129*, 3913-3924.

Voigt, P., LeRoy, G., Drury, W.J., 3rd, Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell *151*, 181-193.

Wang, Z., Oron, E., Nelson, B., Razis, S., and Ivanova, N. (2012). Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. Cell stem cell *10*, 440-454.

Watanabe, A., Yamada, Y., and Yamanaka, S. (2013). Epigenetic regulation in pluripotent stem cells: a key to breaking the epigenetic barrier. Philosophical transactions of the Royal Society of London Series B, Biological sciences *368*, 20120292.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nature genetics *39*, 457-466.

Werling, U., and Schorle, H. (2002). Transcription factor gene AP-2 gamma essential for early murine development. Molecular and cellular biology *22*, 3149-3156.

Winnier, G., Blessing, M., Labosky, P.A., and Hogan, B.L. (1995). Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. Genes & development *9*, 2105-2116.

Wu, G., Gentile, L., Fuchikami, T., Sutter, J., Psathaki, K., Esteves, T.C., Arauzo-Bravo, M.J., Ortmeier, C., Verberk, G., Abe, K.*, et al.* (2010). Initiation of trophectoderm lineage specification in mouse embryos is independent of Cdx2. Development *137*.

Wu, G., Han, D., Gong, Y., Sebastiano, V., Gentile, L., Singhal, N., Adachi, K., Fischedick, G., Ortmeier, C., Sinn, M.*, et al.* (2013). Establishment of totipotency does not depend on Oct4A. Nature cell biology *15*.

Wu, G., and Scholer, H.R. (2014). Role of Oct4 in the early embryo development. Cell regeneration *3*, 7.

Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D.*, et al.* (2013). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell *153*, 1134-1148.

Xu, R.H., Chen, X., Li, D.S., Li, R., Addicks, G.C., Glennon, C., Zwaka, T.P., and Thomson, J.A. (2002). BMP4 initiates human embryonic stem cell differentiation to trophoblast. Nature biotechnology *20*, 1261-1264.

Yagi, R., Kohn, M.J., Karavanova, I., Kaneko, K.J., Vullhorst, D., DePamphilis, M.L., and Buonanno, A. (2007). Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. Development *134*, 3827-3836.

Yamanaka, Y., Ralston, A., Stephenson, R.O., and Rossant, J. (2006). Cell and molecular regulation of the mouse blastocyst. Developmental dynamics : an official publication of the American Association of Anatomists *235*, 2301-2314.

Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development *122*.

Yeom, Y.I., Ha, H.S., Balling, R., Scholer, H.R., and Artzt, K. (1991). Structure, expression and chromosomal location of the Oct-4 gene. Mechanisms of development *35*.

Yuan, H., Corbi, N., Basilico, C., and Dailey, L. (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes & development *9*, 2635-2645.

Ziller, M.J., Muller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C.B., Bernstein, B.E., Lengauer, T.*, et al.* (2011). Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. PLoS genetics *7*, e1002389.

Zybina, E.V., and Zybina, T.G. (1996). Polytene chromosomes in mammalian cells. Int Rev Cytol *165*, 53-119.

Zygmunt, M., Herr, F., Keller-Schoenwetter, S., Kunzi-Rapp, K., Munstedt, K., Rao, C.V., Lang, U., and Preissner, K.T. (2002). Characterization of human chorionic gonadotropin as a novel angiogenic factor. The Journal of clinical endocrinology and metabolism *87*, 5290-5296.

IX. Abbreviations

%	Percent
°C	Degree Celsius
3D	Three Dimensional
APA	Aminopeptidase A
bFGF	Basic Fibroblast Growth Factor
BMP	Bone Morphogenic Protein
BMPRI	Bone Morphogenic Protein Receptor 1
BMPRII	Bone Morphogenic Protein Receptor 2
bp	Base Pair
Cat.Nr.	Catalog Number
cDNA	Complementary DNA
ChIP	Chromatin Immunoprecipitation
ChIP-Seq	Chromatin Immunocprecipitation paired with Next Generation Sequencing
cm2	Square Centimeter
CpG	Cytosine-phosphate-Guanine
CRISPR/Cas	Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein 9
Ct	Cycle threshold
DAPI	4', 6-Diamidino-2-Phenylindole
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl Sulfoxide
DNMT1	DNA Methyltransferase 1
DNMT3A	DNA Methyltransferase 3A
DNMT3B	DNA Methyltransferase 3B
Dox	Doxycycline
E	Embryonic
EB	Elution Buffer
EDTA	Ethylenediaminetetraacetic Acid
EGTA	Ethylene Glycol-bis(β-aminoethyl ether)-N,N,N',N'-Tetraacetic Acid
ENCODE	Encyclopedia of DNA Elements
ESC	Embryonic Stem Cell
ExE	Extraembryonic Ectoderm
FACS	Fluorescence-Activated Cell Sorting
FBS	Fetal Bovine Serum
FDR	False Discovery Rate
FGF2	Fibroblast Growth Factor 2
FSC-A	Forward Scatter-Area
FSC-W	Forward Scatter-Width
g	Gramm
GC	Guanine-Cytosine
GO-term	Gene Ontology term
gRNA	Guide RNA

H3K27	Histone 3 Lysine 27
H3K4	Histone 3 Lysine 4
H3K9	Histone 3 Lysine 9
H3S28	Histone 3 Serine 28
H9	WA09 Embryonic Stem Cells
HAT	Histone Acetyl Transferase
hCG	Human Chorionic Gonadotropin
HDAC	Histone Deacethylase
ICM	Inner Cell Mass
IgG	Immunoglobulin G
IP	Immunoprecipitation
IVF	In Vitro Fertilization
kb	Kilo Bases
KO	Knock-Out
KSR	Knockout Serum Replacement
LB	Lysis Buffer
lincRNA	Long Intergenic Non-Coding RNA
log2	Log Base 2
Lot Nr.	Lot number
M	Molar
me1	Mono-Methylation
me2	Di-Methylation
me3	Tri-Methylation
MEF	Mouse Embryonic Fibroblasts
MG	Matrigel
mg	Milligram
min	Minutes
Mio	Million
ml	Milliliter
MM	Mastermix
MNase	Micrococcal Nuclease
mRNA	Messenger RNA
NaCl	Sodium Chloride
ND	Not Detected
NEAA	Non Essential Amino Acids
ng	Nano Gramm
nm	Nano Meter
nM	Nano Molar
no Mod	No Modification
nt	Nucleotide
PBS	Phosphate-buffered saline
PE	Primitive Endoderm
PGC	Primodial Germ Cell
pH	Potential of Hydrogen
PI	Propidium Iodide
PMSF	Phenylmethane Sulfonyl Fluoride

PSC	Pluripotent Stem Cell
PTM	Photomultiplier Tube
RefSeq	Reference Sequence
RIN	RNA Integrity Number
RNA-Seq	RNA Sequencing
ROCKi	Rho- associated, Coiled-Coil Containing Protein Kinase Inhibitor
RPM	Revolutions Per Minute
rRNA	Ribosomal RNA
RT	Room Temperature
RT-PCR	Real-Time PCR
SDS	Sodium Dodecyl Sulfate
SSC-A	Side Scatter-Area
TALEN	Transcription Activator-Like Effector Nuclease
TBS-T	Tris-Buffered Saline with Tween20
TE	Trophectoderm
TE buffer	Tris-EDTA Buffer
TEE	TE-specific Enhancer
TET	Ten-Eleven Translocation Methylcytosine Dioxygenase
TF	Transcription Factor
tRNA	Transfer RNA
TSC	Trophoblast Stem Cell
TSS	Transcription Start Side
UCSC	University of California, Santa Cruz
WB1	Wash Buffer 1
WT	Wild Type
µg	Micro Gram
µl	Micro Liter
µM	Micro Molar
µm	Micro Meter

Eidesstattliche Versicherung

Krendl, Christian
Name, Vorname

Ich erkläre hiermit an Eides statt,
dass ich die vorliegende Dissertation mit dem Thema

A GATA/TFAP2 transcription regulatory network couples human pluripotent stem cell differentiation to trophectoderm with repression of pluripotency

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 12.2.18

Ort, Datum

Unterschrift Doktorand