The relevance of the epigenetic features of enhancers for germline competence

Inaugural-Dissertation

zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von

Tore Bleckwehl

aus Bremen-Vegesack

Köln
2021

Erster Gutachter: Prof. Dr. Niels H. Gehring

Zweiter Gutachter: Dr. Peter Tessarz

Tag der mündlichen Prüfung: Juni 2021

1. Table of contents

2. List of Abbreviations

4C - Circularized Chromosome Conformation Capture

ATAC - Assay for transposase-accessible chromatin

ChIP - Chromatin immunoprecipitation

CpG - Cytosine-phosphate-Guanine

CRISPR/Cas9 - Clustered regularly interspaced short palindromic repeats

ctrl - control

dCD - catalytic dead Kmt2c/d mutant (KMT2C: Y4792A; KMT2D: Y5477A)

dCT - catalytic dead Kmt2c/d mutant (KMT2C: Y4792N; KMT2D: Y5477N)

Dox - Doxycycline

Dppa3 - Development pluripotency associated protein 3

E14Tg2a - male mouse cell line

EB - Embryoid bodies

ESC - Embryonic stem cells

EpiLC - Epiblast-like cells

EpiSC - Epiblast Stem cells

Esrrb - Estrogen-related receptor β

FACS - Fluorescence-activated cell sorting

gRNA - guide RNA

GEO - Gene Expression Omnibus

H3K4me1/2/3 - Histone H3 lysine4 (mono-/di-/tri-)methylation

H3K9me1/2/3 - Histone H3 lysine9 (/di-/tri-)methylation

H3K27me1/2/3 - Histone H3 lysine27 (/di-/tri-)methylation

H3K27ac - Histone H3 lysine27 acetylation

Klf5 - Krueppel-like factor 5

Kmt2c - Lysine (K)-specific methyltransferase 2C

Kmt2d - Lysine (K)-specific methyltransferase 2D

KO - Knock-out

Lrrc31/34 - Leucine-rich repeat-containing protein 31/34

Nanog - Homeobox protein NANOG

Otx2 - Orthodenticle homeobox 2

p300/CBP - Lysine acetyltransferases

PGC - Primordial germ cells

PGCLC - Primordial germ cell-like cells

Prdm1/14 - PR domain zinc finger protein

RPGC - Reads per genomic content

RPM - Reads per million

RT-qPCR - Quantitative reverse transcription PCR

R1 - male mouse cell line

scRNA-seq - single-cell RNA-seq

TF - Transcription factor

Tfap2c - Transcription factor AP-2 gamma

TPM - Transcripts per million

WT - wild type

3. Abstract

Primordial germ cells (PGC) are the precursor of the gametes, which safeguard the transmission of genetic information to the next generation. In mammals, germline specification occurs through an inductive process whereby competent cells in the post-implantation epiblast differentiate into PGC. Due to the ethical limitations to investigate germline specification in humans, this process is best understood in mice, where PGC specification is initiated in a few epiblast cells at the proximo-posterior end of the mouse embryo (E6.25). Thereby, PGC specification occurs in response to inductive signals emanating from extraembryonic tissues. However, the intrinsic factors that endow mammalian epiblast cells with the competence to respond to germline inductive signals remain unknown. This lack of knowledge is explained, at least partly, due to the scarcity of PGC and peri-implantation epiblast cells *in vivo*. Therefore, here an *in vitro* system that faithfully recapitulates this process by differentiating ESC into primordial germ cell-like cells (PGCLC) has been applied to investigate whether the epigenetic state of *cis*-regulatory elements contributes to germline competence.

First of all, single-cell RNA-sequencing was performed across multiple stages of *in vitro* PGCLC differentiation to further characterize this system. This showed that PGCLC genes are widely expressed in preceding pluripotent embryonic stem cells (ESC), but become homogeneously dismantled in germline competent epiblast like-cells (EpiLC). To understand the regulation of this process, the focus was drawn on distal enhancers that become active in PGCLC (*e.g.* PGCLC enhancer). The analysis of a set of comprehensive epigenetic changes showed that a subset of PGCLC enhancers gets partly decommissioned in EpiLC, retains permissive chromatin features, including H3K4me1, and remains responsive to transcriptional activators. Furthermore, upon further differentiation into epiblast stem cells (EpiSC) in which germline competence is lost, the PGCLC enhancers accumulate higher levels of repressive epigenetic modifications and lose their responsiveness to transcriptional activators, thus getting fully decommissioned.

Subsequently, the relevance of H3K4me1 for germline competence was demonstrated in a H3K4me1/2 deficient cell line, where both H3K4me1/2 methyltransferases (KMT2C/D) were catalytically inactive (dCD cells). The differentiation of the dCD cells revealed a decreased PGCLC differentiation capacity, which was accompanied by reduced expression of PGCLC genes linked to PGCLC enhancers. On the other hand, H3K4me1/2 might not have the same functional impact during the transition from ESC to other pluripotent stages, as gene expression changes during these transitions were relatively normal in the dCD cells. Finally, the analysis of *in vivo* single-cell methylation data suggest an increased epigenetic heterogeneity of PGCLC enhancers in the formative epiblast, before they eventually become activated in PGC.

Overall, the presented work suggests that H3K4me1 can facilitate the (re)activation of enhancers and the establishment of gene expression programs during specific developmental transitions.

4. Introduction

4.1. Lineage competence

Elucidating the molecular basis of lineage competence and specification is important to understand the embryogenesis of multicellular organisms, to gain knowledge about birth defects and to develop potential stem cell therapies, tissue engineering and other applications of regenerative medicine.

Embryogenesis starts with the fertilization of the oocyte. First, the resulting totipotent zygote undergoes several cell divisions, followed by multiple lineage commitments resulting in increasingly specialized stages that will ultimately result in the terminal differentiation of the cells. These lineage commitments during embryogenesis are a series of spatiotemporal inductive events, whereby cellular progenitors undergo an initial phase of specialization during which the cell fate could still be altered. The phase of lineage specialization ends with the cellular determination, when the cell has reached the capability to complete the differentiation autonomously and its cell fate is irreversible. But before lineage specification, the acquisition of lineage competence is mandatory. Lineage competence has been defined as the ability of a cell to differentiate towards a specific cell fate in response to the intrinsic and extrinsic signals (Waddington, 1940). This means that an extrinsic inductive signal will only initiate lineage specification in recipient cells with a compatible intrinsic lineage competence. While the extracellular signals that can induce lineage specification have been extensively studied in the context of multiple cell fates and species (Solini et al., 2017), the intrinsic factors that determine the cellular competence to respond to those signals remain to be elucidated. Importantly, the intrinsic factors that render a cell competent and support the initiation of lineage specification can be considered as the actual origin of lineage specification.

During mouse embryogenesis the first lineage decision divides the blastomeres into the trophoblast and the inner cell mass (ICM). During this process, cells exclusively expressing CDX2 become the trophoblast, while the expression of OCT3/4 specifies the ICM (Dietrich and Hiiragi, 2007). The ICM is later separated into the primitive

endoderm, defined by GATA6 expression, while the NANOG expressing cells will become the epiblast cells (Mohammed et al., 2017; Ohnishi et al., 2014). In both cases, the preceding cellular progenitors simultaneously express the markers for both cell types, like CDX2 and OCT3/4 in the morula. Upon specification, the previous binary decisions towards a particular cell type become reinforced by suppressing the antagonist fates (Niwa et al., 2005). Subsequently, the primordial germ cells (PGCs), which arise from the epiblast cells, become specified and transiently express mesoderm (BRACHYURY) and early PGC specifiers (PRDM1/DPPA3), before getting determined as PGCs (Kurimoto et al., 2008).

However, there are still major open questions regarding germline competence: (i) how do the early murine epiblast cells become competent for the specification of primordial germ cells? (ii) how and why is germline competence lost in all the somatic lineages? Before pursuing these questions, the next chapters are going to summarize how epigenetic modifications in general can affect lineage competence and specification during embryogenesis.

4.2. Epigenetics and Enhancers

Epigenetic research investigates how chemical modifications of the DNA and/or histones, that do not alter the DNA sequence itself, regulate gene expression. Epigenetic modifications might play important roles during lineage specification. Two molecules of each histone of H2A, H2B, H3 and H4 are assembled into one nucleosome that is wrapped by 147 bp of DNA (Pepenella et al., 2014). The N-terminal region of each histone can be post-translational modified by acetylation of lysines, methylation of lysines and arginines, phosphorylation of serines, threonines and tyrosines, ubiquitylation and sumoylation of lysines, etc. (Jenuwein and Allis, 2001; Tan et al., 2011). These modifications regulate major DNA-related cellular processes (*e.g.* transcription, replication, repair, recombination) and change dynamically during development. Thereby histone modifications occur in particular combinations (i.e. histone code) that can contribute to the cellular identity of a cell

and regulate the transcriptome in manifold ways, including the control of *cis*-regulatory elements, like enhancers (David Allis and Jenuwein, 2016).

Enhancers are regions that are located upstream, downstream or within genes and that can positively control gene expression (Arner et al., 2015; Larke et al., 2021). Accordingly, enhancers regulate developmental genes and are highly abundant (Long et al., 2016). One gene can be regulated by multiple enhancers and also one enhancer can control several genes. When several enhancers control the expression of a gene, this can involve additive, competitive or synergistic effects on transcription (Bolt and Duboule, 2020; Hnisz et al., 2013; Osterwalder et al., 2018). In general, enhancers get into physical proximity of their target genes through the formation of chromatin loops (Chen et al., 2018b; Kim et al., 2019; Morgan et al., 2017b). Given that the genome is organized by compartmentalization of the chromatin, enhancer-promoter interactions preferentially occur within the same topological associated domains (Dixon et al., 2012; Nora et al., 2017). However, enhancer states are dynamic and lineage specification often involves major rewiring of enhancer landscapes and changes in gene expression programmes. Interestingly, some enhancer-promoter contacts are already established before activation occurs (Bonev et al., 2017; Schoenfelder and Fraser, 2019). Furthermore, multiple active enhancers and promoters might be organized in large condensates that facilitate transcription in the presence of activated RNA Polymerase II (Beagrie et al., 2017; Ibragimov et al., 2020).

Besides the 3D rewiring of enhancers, their epigenetic states are also dynamically regulated during development. Active enhancers are characterized by displaying low nucleosomal density, high accessibility to transcription factors (TF) and coactivators p300/CBP and are typically identified by their high H3K27ac levels and/or the production of non-coding and short bidirectional transcripts termed enhancer RNAs (eRNAs) (Figure 1) (Whitaker et al., 2015).

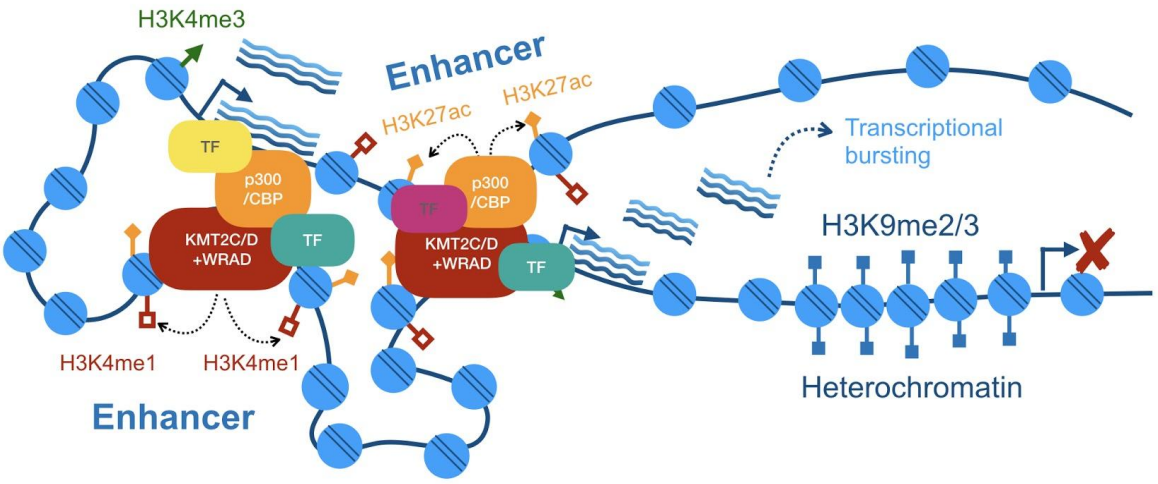


Figure 1: Illustration of enhancer states. The simplified diagram shows the binding of lysine acetyltransferases p300/CBP, the histone methyltransferases KMT2C/D and transcription factors (TF) to two active enhancers. At the active enhancers, H3K4me1 is catalyzed by KMT2C/D and H3K27ac by p300/CBP. The KMT2C/D and p300/CBP and other complexes that mediate the chromatin loop formation might be recruited by lineage-specific TF. The physical proximity of the active enhancers to their target gene results in the increased target gene transcription. In contrast, regions within (facultative) heterochromatin, which consist of H3K9me2/3, are inaccessible and transcriptionally inactive.

Typically, the initial enhancer activation requires the remodeling of the chromatin by histone acetyltransferases (*e.g.* p300/CBP), histone methyltransferases (*e.g.* KMT2C/D WRAD), chromatin remodelers (*e.g.* BAF complex) and communication with the basal transcription machinery located at the promoters (*e.g.* Mediator complex) (Tafessu and Banaszynski, 2020). While all these factors are usually ubiquitously expressed in different tissues, their recruitment to enhancers is typically mediated by lineage specific TF. Moreover, the activation and decommissioning of enhancers involves epigenetic changes, which in turn might also influence the binding of TFs (Long et al., 2016). In the next chapter, the most studied histone H3 modifications and DNA methylation are described in more detail (Figure 2), highlighting how these different epigenetic modifications can regulate enhancers, lineage competence and specification during mammalian embryogenesis.

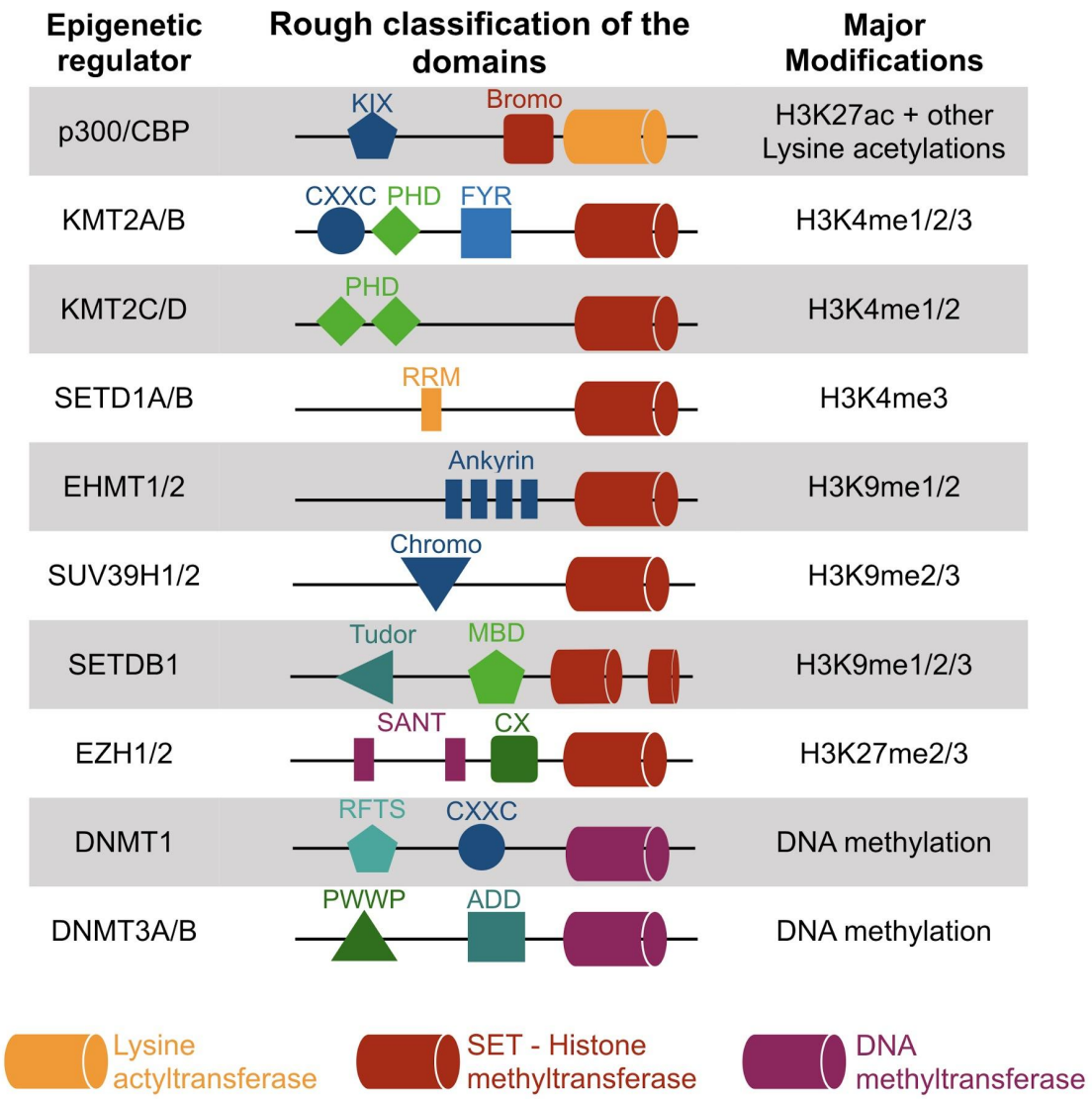


Figure 2: Overview of important epigenetic regulators. Shown are the domains of enzymes responsible for the acetylation or methylation of histones as well as the methylation of DNA. Despite slightly different functions and protein architecture, the illustration summarizes similar epigenetic regulators to outline important differences between non-redundant enzymes. In the last column the most abundant epigenetic modifications that have been reported for each epigenetic regulator are listed.

4.2.1. H3K4 methylation at TSS and enhancers

H3K4 mono-, di- or trimethylation (H3K4me1, H3K4me2, H3K4me3) is associated with active transcription and is catalyzed by seven partially redundant H3K4 methyltransferases of the mixed lineage leukemia (MLL) family. All MLL H3K4

methyltransferases contain a SET1-domain at the C-terminus of the protein (Figure 2) that catalyzes the methylation of H3K4 and are associated with the WRAD (WDR5, RBBP5, ASHL2, DPY30) complex and additional context-specific effectors (Bochyńska et al., 2018). On the other hand, the different H3K4 methyltransferases target distinct genetic elements and catalyze different grades of H3K4 methylation. Here, the focus will be on MLL methyltransferase in the context of enhancers, although other enzymes, like the ones belonging to the SMYD family, also possess H3K4 methyltransferase activities (Tracy et al., 2018).

Among the MLL methyltransferases, SET1A/B (KMT2F/G) catalyze mainly H3K4me3 at promoters. Thereby, the recruitment of SET1A/B to the CpG-rich promoters is mediated by CFP1 (which consists of a CxxC-domain for the recruitment to CpG islands) and pre-existing H3K4me3. Accordingly, SET1A/B are mainly found at actively transcribed gene promoters (Brown et al., 2017; Hughes et al., 2020). Additionally, KMT2E (MLL5) might be associated with promoters but has not been extensively studied yet (Ali et al., 2013). In contrast to SET1A/B, the H3K4 methyltransferases KMT2A/B (MLL1/2) both possess a CxxC-domain and recognize CpG-rich regions by themselves:

KMT2A catalyzes H3K4me2/3 at CpG-rich promoters (Addicks et al., 2019) and enhancers (Rickels et al., 2016). The *Kmt2a* knockout (KO) is embryonic lethal, while the deletion of the KMT2A-SET domain results in a less severe phenotype, including skeletal defects, misexpression of *Hox* genes and decreased H3K4me1 (Terranova et al., 2006). On the other hand, the catalytic function of KMT2A is dispensable for hematopoietic stem cells, as selected target genes do not show severe expression changes or different H3K4me1 or H3K4me3 levels at these genes. In contrast, the deletion of *Kmt2a* in hematopoietic stem cells results in misregulation of the selected genes and decreased MOF-mediated H4K16ac. Therefore it has been suggested that KMT2A recruits MOF to the regulatory elements of these genes, hence pointing out the non-catalytic importance of KMT2A (Mishra et al., 2014).

KMT2B catalyzes mostly H3K4me3 at bivalent promoters and CpG-rich enhancers (Denissov et al., 2014; Hu et al., 2013a, 2017). However, the deletion of KMT2B also

abolishes H3K4me1 at specific sites (Morgan et al., 2017a). KMT2B KO is also embryonic lethal (~E10.5) and chimera experiments revealed its ubiquitous expression and requirement for development (Glaser et al., 2006). Although KMT2B is dispensable for ESC self-renewal, it is actually required for the differentiation of ESC into EBs (Hu et al., 2013a; Lubitz et al., 2007). Differentiation defects of an ESC line harboring an inactive KMT2B SET-domain indicate the importance of H3K4me3 (Hu et al., 2017). Furthermore, the expression of the downregulated genes in KMT2B KO ESC can be rescued by diminishing DNA methylation or H3K27me3, pointing towards an imbalance of epigenetic modifications upon KMT2B loss rather than major non-catalytic functions for this methyltransferase (Douillet et al., 2020).

KMT2C and KMT2D (MLL3/4) are the largest proteins of the MLL family, which both lack a CxxC domain and have been shown to interact with several proteins besides the WRAD complex (Bochyńska et al., 2018; Sze and Shilatifard, 2016). KMT2C and KMT2D catalyze H3K4me1/2 mainly at enhancers and are redundant in their methyltransferase activities (Hu et al., 2013b; Lee et al., 2013; Rickels et al., 2017). Nevertheless, they might have distinct roles during murine embryogenesis as the *Kmt2d* KO is embryonic lethal (~E9.5), while the *Kmt2c* KO is lethal at birth (Ashokkumar et al., 2020; Lee et al., 2013). Besides, the redundant and context-dependent activity of MLL proteins, the relative importance of their non-catalytic (through associated proteins) and catalytic (H3K4 methyltransferase activity) functions are currently under debate. Notably, a recent report have interrogated the importance of the KMT2C/D mediated H3K4me1/2 through the generation of an ESC line with amino acid substitutions in the SET domains of KMT2C and KMT2D that enabled the decoupling of their methyltransferase activity from their non-catalytic functions (Dorighi et al., 2017). The KMT2C/D double KO ESC (Cao et al., 2018; Lin-Shiao et al., 2018; Wang et al., 2016; Yan et al., 2018) results in the loss of H3K4me1 at active enhancers, strong reduction in H3K27ac levels, RNA polymerase II binding and enhancer RNA (eRNA) production. In contrast, the loss of H3K4me1 in *Kmt2c/d* catalytic mutant ESC (*Kmt2c/d* dCD) only partially reduced H3K27ac at enhancers, had no measurable effects on RNA polymerase II binding and displayed minor changes in gene expression and eRNA

levels compared to the *Kmt2c/d* double KO ESC. Overall, the findings suggest that the function of KMT2C/D as long-range co-activators is largely independent of H3K4me1 (Dorighi et al., 2017). In addition, work in *Drosophila melanogaster* showed that the KO of *Trr*, the homolog of *Kmt2c/d* in flies, was embryonic lethal. However, an amino acid substitution in the SET domain of *Trr* that globally reduced H3K4me1 did not impair fly development (Rickels et al., 2017). This suggests that H3K4me1 might be dispensable and supports a major relevance for the non-catalytic functions of KMT2C/D. On the other hand, subsequent work with *Kmt2c/d* dCD ESC showed that the recruitment of the BAF complex, involved in chromatin remodeling (Local et al., 2018), and the establishment of long-range chromatin interactions (Yan et al., 2018) require H3K4me1. Therefore, the functional relevance of H3K4me1 for enhancer function is still under debate.

Furthermore, the KO of *Kmt2c/d* or *Kmt2d* alone has also been studied in more differentiated cellular contexts (*e.g.* embryoid bodies, adipogenesis, neural progenitors, hematopoiesis), where impaired enhancer activation and cellular differentiation were generally observed (Ang et al., 2016; Jang et al., 2019; Lai et al., 2017; Lin-Shiao et al., 2018; Ortega-Molina et al., 2015; Placek et al., 2017; Wang et al., 2016; Yan et al., 2018; Zhang et al., 2015). However, in the light of the recent findings, it remains unclear whether these defects are caused by the catalytic and/or non-catalytic functions of KMT2C/D.

4.3. H3K27ac and active enhancers

Active enhancers are associated with histone acetylation, which is catalysed by lysine acetyltransferases (KAT) (Lee and Workman, 2007). Lysines, which are overrepresented in the histone tails, possess positive charges. While this charge is not altered by methylation, the positive charge of the lysine is lost with acetylation, resulting in a repulsion from the negatively charged DNA backbone and release of nucleosome-nucleosome stacking and finally a more accessible DNA (Barnes et al., 2019; Onufriev and Schiessel, 2019).

Acylation might occur in a chemical reaction of acetyl-CoA with a lysine residue (Wagner and Hirschey, 2014) or through the catalytic activity of KAT. Thereby the target specificity of KAT is often very relaxed and includes many non-histones target sides (Weinert et al., 2018). Nucleosomal KAT belong to the family of GNAT (*e.g. Kat2a/Gcn5* (Salah Ud-Din et al., 2016)), MYST (*e.g. Kat5/Tip60*, *Kat6a/Moz* (Sapountzi and Côté, 2011)) and p300/CBP (Dancy and Cole, 2015). Thereby p300/CBP (KAT3A/B) are quite redundant in their acetyltransferase activity and both contain a KIX domain, which serves as a docking side for CREB (cAMP-response element-binding protein) and other proteins that form a heterodimer with p300/CBP. Furthermore p300/CBP contain a bromodomain that confers binding to acetylated lysines and a lysine acetyltransferase that acetylates H3K27ac and other lysines (Figure 2) (Dancy and Cole, 2015).

Beside other targets, p300/CBP catalyze H3K27ac, which is the most prevalent feature of active enhancers (Rada-Iglesias et al., 2011). However, the acetylation of other histone lysines, like H3K122ac (Pradeepa et al., 2016), H3K9ac (Jin et al., 2011; Karmodiya et al., 2012) and H4K16ac (Taylor et al., 2013) can also occur within active enhancers.

The loss of H3K27ac by a H3.3K27 substitution (Zhang et al., 2020) or the inhibition of the catalytic activity of p300/CBP (Raisner et al., 2018) caused a preferential reduction of H3K27ac at enhancers, while the H3K27ac levels at promoters were more stable. In mouse ESC, the H3.3K27 substitution is accompanied by unaltered chromatin accessibility at enhancers, similar Pol II levels at TSS and minor expression changes. Therefore, under ESC self-renewal conditions, there is not a strong correlation between H3K27ac at enhancers and gene expression (Zhang et al., 2020). Notably, acetylation of other histone lysine residues might compensate for the loss of H3K27ac. Finally, the establishment of KMT2C/D-mediated H3K4me1 seems to precede *de novo* enhancer activation by p300/CBP and H3K27ac (Kang et al., 2021; Wang et al., 2016).

4.3.1. H3K9 methylation and heterochromatin

Transcriptional silencing of enhancers and genes can occur by chromatin compaction and the establishment of facultative heterochromatin, which can act locally or span long distances of the genome. Facultative heterochromatin is often rich in heterochromatin protein (HP1), H3K9me2/3 and DNA methylation. It is usually localized to the periphery of the nucleus and inaccessible for transcription factors and other activators (Mozzetta et al., 2015; Ninova et al., 2019).

H3K9 methylation, and therefore the initiation of heterochromatinization, can be catalyzed by the methyltransferases EHMT1/2 and SETDB1/2 on H3K4me3-devoid nucleosomes (Binda et al., 2010), while SUV39H1/2 is mostly involved in the establishment and maintenance of heterochromatin (Figure 2).

EHMT1/2 (KMT1D/C and also known as GLP/G9a) are two similar histone methyltransferases that are embryonic lethal (~E9.5) and form a heteromeric complex that catalyze H3K9me1/2 at euchromatic regions (Tachibana et al., 2005). EHMT1/2 both possess ankyrin repeat domains that specifically recognize H3K9me1/2 and hence promote the spread of H3K9me2 at already H3K9me pre-marked sites, which then support the silencing of ESC genes upon differentiation (Liu et al., 2015). Furthermore, the non-catalytic functions of EHMT2 promote the establishment of DNA methylation, which might be part of the observed gene silencing (Auclair et al., 2016).

SETDB1/2 (KMT1E/F) consist of a Tudor domain, a methyl-CpG binding domain and a SET domain that is interrupted by a conserved insertion. SETDB1 (also known as ESET) catalyzes all three forms of H3K9 methylation and forms various repressive protein complexes both at euchromatic and heterochromatic regions. A complex of SETDB1, KRAB-ZFP and KAP1, which is also associated with deacetylases, promotes the formation of a microenvironmental heterochromatin at euchromatic sites, while a SETDB1/MBD1/ATF7IP complex maintains the heterochromatin during DNA replication (Zhu et al., 2020). Furthermore SETDB1 can also bind the bivalent chromatin state of H3K9me1/2 and H3K14ac via the Tudor domain (Jurkowska et al.,

2017). Although SETDB2 is the paralog of SETDB1, a H3K9 methyltransferase activity could not be experimentally proven for SETDB2 to date (Roqueta-Rivera et al., 2016).

SUV39H1/2 (KMT1A/B) catalyze H3K9me2/3 and are enriched at heterochromatin, possess a chromatin-organization modifier domain that binds to H3K9me1 and thus result in the establishment of heterochromatic regions (Mozzetta et al., 2015). Furthermore, SUV39H1 sumoylates the heterochromatin associated protein HP1, which facilitates the recruitment of HP1 to heterochromatin. And since SUV39H1 is also recruited by HP1, SUV39H1 is considered broadly for the final establishment and maintenance of heterochromatin and cellular identity (Maison et al., 2016; Muramatsu et al., 2016).

During development the loss of *Setdb1* results in early embryonic lethality (~E4.5) and SETDB1 forms only puntacted patterns in the inner cell mass, but not in the trophectoderm during normal embryogenesis, indicating its absence in the trophectoderm (Cho et al., 2012). Furthermore, ESC with low SETDB1 levels that are aggregated into 2-cell embryos will end up solely in the trophectoderm. And because SETDB1 interacts with POU5F1, which has been shown to bind to and repress trophectoderm genes, the expression of *Setdb1* in the inner cell mass might contribute to the bifurcation of the inner cell mass and the trophectoderm (Yeap et al., 2009; Yuan et al., 2009).

Higher resolution of the heterochromatin formation showed that heterochromatin and H3K9me3 accumulate at protein-coding genes until gastrulation before the heterochromatin regions decrease in a lineage specific manner that leads to the proper specification of hepatic and pancreatic lineages (Nicetto et al., 2019). Interestingly, the majority of genes that lost H3K9me3 during these specification events but did not become induced, gained H3K27me3, indicating a switch in the underlying transcriptional repression mechanisms (Nicetto et al., 2019). Direct evidence for the actual release of gene coding regions from the H3K9me2-related nuclear lamina that result in gene expression was observed during cardiogenesis, where a triggered dissociation of cardiac myocytes genes from the lamina increased

their lineage specification (Poleshko et al., 2017). Furthermore it has been shown that SETDB1 stabilizes the lineage commitment during the differentiation of neural lineages and primordial germ cells (Mochizuki et al., 2018; Tan et al., 2012) or can establish a poised state with H3K4me3 in adipogenesis (Matsumura et al., 2015), while SUV39H1 balances the lineage commitment of T-helper cells (Allan et al., 2012).

Finally, some enhancers might also function as silencers in other cell types. When acting as silencers, these regulatory elements seem to be marked by repressive epigenetic modifications such as H3K9me3 and DNA methylation, although a universal chromatin signature for silencers has not been uncovered yet (Doni Jayavelu et al., 2020; Pang and Snyder, 2020). In addition, HP1 binding and H3K9 methylation of facultative heterochromatin act as a barrier for reprogramming, affecting the activation of pluripotency enhancers (Becker et al., 2016; Cossec et al., 2018), which together with other studies (Leemans et al., 2019), suggests that enhancer activity is repressed by heterochromatin.

4.3.2. H3K27 methylation and poised enhancers

H3K27me1/2/3 is catalyzed by EZH1/2 (KMT6A/B, Figure 2). An *Ezh2* KO is embryonic lethal (~E8.5) and leads to a profound loss of H3K27 methylation. Accordingly, *Ezh2* is considered the main H3K27 methyltransferase (Lee et al., 2018). The catalytic activity of EZH2 requires the assembly of the core polycomb repressive complex 2 (PRC2), consisting of EZH2, EED and SUZ12 (Laugesen et al., 2019). Facultative subunits of the PCR2 complex, like JARID2 (PRC2.2 complex) and MTF2 (PRC2.1 complex), might recruit PRC2 to unmethylated CpG-rich DNA sequences found within promoters and enhancers, establishing H3K27me3 through the catalytic activity of the EZH2 SET domain (van Mierlo et al., 2019; Pachano et al., 2020). Furthermore, the binding of EED to methylated H3K27 facilitates the methylation of H3K27 by EZH2 of the neighboring nucleosome (Poepsel et al., 2018)

and the spreading of H3K27me3. In contrast, H3K27 methylation is counteracted by H3K36 and DNA methylation (Brinkman et al., 2012; Streubel et al., 2018).

While H3K9me2 and H3K27me3 are considered to be part of different modalities of transcriptional silencing (Zylicz et al., 2015), it is interesting to note that EHMT1, besides H3K9me1/2, can also catalyze H3K27me1/2 (Mozzetta et al., 2014), facilitating the recruitment of PRC2 to developmental genes. Moreover, EHMT1 might also be responsible for the establishment of H3K27me2 in the male pronucleus (Meng et al., 2020).

PRC2-mediated gene repression has been mostly studied in the context of bivalent (H3K27me3+PRC2 and H3K4me3+KMT2B) CpG-rich promoters in pluripotent stem cells. This chromatin bivalency might protect the promoters of developmental genes from repressive DNA methylation that otherwise could impair differentiation capacity (Boulard et al., 2015; Douillet et al., 2020; Eckersley-Maslin et al., 2020; Verma et al., 2018). In the context of enhancers, poised enhancers display binding of H3K27me3 and PRC2 as well as H3K4me1 and p300 in mouse and human ESC. Poised enhancers become activated during the differentiation of ESC into anterior-neural progenitors (Cruz-Molina et al., 2017; Rada-Iglesias et al., 2011). Furthermore, poised enhancers are already in physical proximity to their target genes in ESC, which might facilitate the activation of the target genes upon differentiation (Cruz-Molina et al., 2017; Ngan et al., 2020; Russ et al., 2017).

4.3.3. CpG methylation and transcriptional repression

The DNA methyltransferases DNMT1 and DNMT3A/B/C (Figure 2) catalyze cytosine methylation (CpG methylation), that in a hypermethylated state can repress transcriptional activity of genes and enhancers (Amabile et al., 2016; Hon et al., 2014; Kremsky and Corces, 2020; Liu et al., 2016).

The *de novo* DNA methyltransferases *Dnmt3a/b* exist in different isoforms, which participate differently in the methylation of DNA and are both essential for embryogenesis (Baubec et al., 2015). Both DNA methyltransferases contain a

proline-tryptophan-tryptophan-proline (PWWP) domain that recognizes H3K36me2 or H3K36me3 (Weinberg et al., 2019), an ATRX-DNMT3-DNMT3L (ADD) domain, that can interact with H3K9me KMTs or the histone H3 tail, and the DNA methyltransferase domain that catalyzes the methylation of cytosines. Thereby, the methyltransferase activity of DNMT3s are stimulated by unmethylated H3K4, while H3K4 methylation inhibits such activity (Guo et al., 2015; Ooi et al., 2007; Zhang et al., 2010). Interestingly, catalytic inactive DNMT3B isoforms stimulate the *de novo* DNA methyltransferase activity of DNMT3A/B (Duymich et al., 2016; Zeng et al., 2020). Similarly DNMT3L, which is catalytically inactive and essential for gametogenesis, stabilizes a DNMT3A/DNMT3L complex (Veland et al., 2019).

The DNA methyltransferase DNMT3C specifically methylates the promoters of retrotransposons during mouse spermatogenesis, which safeguards the fertility of sperms (Barau et al., 2016). Lastly, DNMT1 differs from DNMT3s and contains a replication foci targeting sequence (RFTS) that promotes the post-replicative DNA methylation. Together with the cofactor UHRF1, DNMT1 forms a complex that is thought to maintain DNA methylation through cell divisions by the methylation of hemimethylated DNA (Bostick et al., 2007).

CpG demethylation can occur through a passive loss of CpG methylation during DNA replication in the absence of the DNA methylation machinery. In ESC and oocytes, DPPA3 binds to UHRF1 and exports UHRF1 to the cytoplasm, resulting in genome-wide DNA hypomethylation (Li et al., 2018b; Mulholland et al., 2020). Alternatively, the DNA demethylases (TET1-3) actively demethylate 5-methylcytosine by conversion to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine and 5-carboxylcytosine, whereby the later two modifications can be converted to an unmethylated cytosine by the thymine DNA glycosylase (TDG) and the base excision repair pathway (Ito et al., 2010; Weber et al., 2016). Both active and passive DNA demethylation mechanisms occur sequentially during the reprogramming of primordial germ cells (Hargan-Calvopina et al., 2016). Furthermore, 5hmC can function as an independent epigenetic mark that can participate in lineage

specification events by inhibiting the establishment of CpG methylation at regulatory elements (Tsiouplis et al., 2020).

On the other hand, the establishment of CpG methylation is important to silence expression programmes, like the one of pluripotent cells (Schmidt et al., 2012). In addition, lineage competence and specification might be regulated by DNA methylation: Ectodermal enhancers for instance are already hypomethylated in naïve pluripotent cells and maintain such epigenetic state in subsequent ectodermal lineages, while becoming methylated in mesodermal or endodermal lineages (Argelaguet et al., 2019). In contrast, meso- and endodermal enhancers progressively gain DNA methylation in the epiblast and undergo lineage-specific enhancer demethylation during differentiation. Furthermore, the differentiation of mature mesodermal lineage is impaired in a *Tet1-3* KO, indicating that proper DNA demethylation is needed for the execution of the mesodermal program (Argelaguet et al., 2019). Another example has been shown in haematopoietic stem cells, which are transcriptionally primed for differentiation towards erythroid progenitors and myelomonocytic progenitors (Izzo et al., 2020). However, in the *Dnmt3a* KO or *Tet2* KO, the specification of erythroid progenitors and myelomonocytic progenitors is favoured, respectively. The transcriptional priming of the WT and the KOs towards different progenitors is also reflected at the DNA methylation levels present at enhancers in haematopoietic stem cells (Izzo et al., 2020). Moreover, in mice and human ESC, DNMT3A/B compete with TET1-3 in the context of somatic enhancers, which undergo a permanent turnover of their DNA methylation status that might be widely restricted to pluripotent stages (Charlton et al., 2020).

Overall, accumulating evidence indicates that the epigenetic state of enhancers can play important regulatory functions in several developmental contexts.

4.4. Primordial germ cell (PGC) formation

From a biological point of view, primordial germ cell (PGC) specification is a crucial developmental event, as it is imperative for reproduction in *metazoans*. Once determined, PGCs are unipotent cells that will only develop into germ cells, oocytes

or sperm. Then, upon fertilization, the fusion of two gametes gives rise to the totipotent zygote and the circle of life is closed. Therefore, in contrast to somatic cell types that support the function of the actual living organism, PGCs are required for the intergenerational transmission of genetic information and the initial steps of embryogenesis.

PGCs are segregated from the soma by two fundamentally different mechanisms during early embryogenesis: (i) Preformation, whereby, soon after fertilization, maternally inherited determinants (*e.g. vasa, nanos*) form a germ plasm or (ii) Epigenesis, whereby PGCs are specified later during embryogenesis, as embryonic cells are exposed to inductive signals (*e.g.* WNT3, BMP4).

4.4.1. PGCs in other species

The two different mechanisms through which PGC can be formed, preformation and epigenesis, are found across *metazoa*. In invertebrates, preformation occurs in the fruit fly *Drosophila melanogaster* (Nakamura et al., 2010), the nematode *Caenorhabditis elegans* (Nakamura and Seydoux, 2008) and the wasp *Nasonia vitripennis* (Quan et al., 2019). In other insects, like the cricket *Gryllus bimaculatus*, PGCs are induced by BMP4 and require the expression of *Prdm1* for their specification (Donoughe et al., 2014; Nakamura and Extavour, 2016). Similarly, the bug *Oncopeltus fasciatus* (Ewen-Campen et al., 2013), the beetle *Tribolium castaneum* (Schröder, 2006) and the honey bee *Apis mellifera* (Dearden, 2006) also lack a germ plasm. Likewise, vertebrates differ in PGC formation. Among other vertebrates, in model organisms such as zebrafish (*Danio rerio*) (Raz, 2003), frog *(Xenopus laevis)* (Venkatarama et al., 2010) and chicken (*Gallus gallus)* (Tsunekawa et al., 2000), PGC are formed by maternally inherited factors, while in the Axolotl *Ambystoma mexicanum* (Chatfield et al., 2014) and all studied mammals (*e.g. Mus musculus*, *Macaca fascicularis*), PGCs are induced and, thus, formed through epigenesis.

This diversity of PGC specification modalities raised the question of the evolution of PGCs. Due to the pleiotropic roles of clade specific germ plasm determinants, like

oskar (*D. melanogaster*), PGL (*C. elegans*) and *bucky ball* (*D. rerio*), epigenesis has been suggested as the ancestor way of PGC formation, while preformation might reflect convergent evolution (Johnson and Alberio, 2015) or spandrel adaption (Whittle and Extavour, 2017). On the other hand, the conservation of *oskar* between the fruit fly and wasp (PGC preformation) would indicate a common ancestor. In contrast, *oskar* orthologues were not identified in beetles and honey bees (presumably PGC epigenesis), which, like the fruit fly and the wasp, belong to the *holometabola* insects (Quan and Lynch, 2016). Therefore, the beetle and honey bee could have either reverted to the PGC induction mode, or the factors for both PGC specification modalities could have been present at the evolutionary origin of *holometabola*. In line with the later idea, the sea urchin *Strongylocentrotus purpuratus* (Juliano et al., 2010) and the sea anemone *Nematostella vectensis* (Chen et al., 2020) display an intermediate step of PGC specification, as their PGC progenitors show early expression of *vasa,* although these cells are not fully committed to PGCs yet, but remain multipotent (Seervai and Wessel, 2013). Therefore, it seems likely that some sort of PGC epigenesis might be the anchestoral mode of PGC specification and that *metazoa* reflect a continuum of PGC specification modalities.

4.4.2. PGCs in mammals

This continuum also includes PGC specification in mammals, where different modes of PGC induction have been documented (Irie et al., 2014; Tang et al., 2016). In mice, PGC specification is initiated in a few epiblast cells at the proximo-posterior end of the mouse embryo (E6.25) right after implantation (Ohinata et al., 2005) and the exit from naïve pluripotency (E4.5 - E5.5). PGC epigenesis occurs in response to inductive signals emanating from extraembryonic tissues: WNT3 from the visceral endoderm (Aramaki et al., 2013) and BMP4 from the extraembryonic ectoderm (Ohinata et al., 2009). Although only the proximo-posterior epiblast cells will eventually be induced for PGCs specification, it has been proposed that from E5.5 to E6.25, most of the epiblast cells are germline competent (Saitou et al., 2002; Tam and Zhou, 1996). But in the natural environment their responsiveness to BMP4 is

antagonised by signalling (*e.g. Cer1*) from the anterior visceral endoderm so that only a few proximo-posterior epiblast cells are normally induced to become PGCs (Ohinata et al., 2009).

Following the initial induction of germline competence in the epiblast, by E7.25 a population of approximately 40 PGCs are formed at the base of the allantois, which then migrate into the developing hindgut endoderm (E7.75), the mesentery (E9.5), finally colonizing the genital ridges at E10.5. The PGC phase of germline development concludes by E13.5, when the PGCs have been expanded, epigenetically reprogrammed and the germ cells enter meiosis and gametogenesis (Saitou and Yamaji, 2012).

The PGC specifications in rats, rabbit and pork arise similarly to the mouse (Kobayashi et al., 2020). Porcine PGCs for instance specify from competent posterior pre-primitive streak epiblast cells (~ E11.5) and, notably, before the formation of the mesoendoderm (Kobayashi et al., 2017). However, in mice, the cup-shaped egg cylinder organizes the topographical proximity of the extraembryonic ectoderm (BMP4 signaling) and the early epiblast that results in the induction of PCGs, while the rabbit or the pig early embryos represent a flat bilaminar disc where PGCs might be induced by cytokines, emanating from the hypoblast (Hopf et al., 2011; Irie et al., 2014).

Interestingly, PGCs of the cynomolgus monkey are specified in the nascent dorsal amnion by autocrine signalling of BMP4 and WNT3 from surrounding amniotic cells (Sasaki et al., 2016). Although the exact time point and modality of human germline competence have not been conclusively determined *in vivo* due to ethical reasons, the observation of a few putative PGCs observed in early human embryos (Chen et al., 2019), comparative single-cell RNA sequencing from *in vitro* models or microfluid modelling of gastrulation (Zheng et al., 2019) suggest that human PGCs could arise from either the posterior epiblast (Popovic et al., 2019), amniotic cells (Ma et al., 2019) or both (Niu et al., 2019).

Conclusively, it can be stated that regardless of the different PGC induction modalities, the intrinsic factors, which render mammalian cells competent for

germline specification, remain broadly elusive. And since the inductive mode of mammalian PGC formation is best studied in mice, this model organism might be the most suitable system to study the intrinsic factors of germline competence.

4.5. Peri-implantation mouse development

This chapter will highlight the current knowledge about germline competence and the induction of PGCs in mice. Firstly, the time-frame of germline competence will be summarized, focusing on the early stages of mouse embryogenesis (i.e. mouse peri-implantation development; E4.5-E6.5) during which mouse epiblast cells undergo a continuum of pluripotency states: naïve, formative and primed (Figure 3) (Kinoshita and Smith, 2018). Then, the enhancer and epigenetic remodelling that occurs before and after germline competence is established will be described. Finally, the formation of the specialized PGC will be covered.

4.5.1. The transient nature of germline competence

Naïve pluripotency defines a cellular state during which cells are capable of differentiating into all embryonic lineages. Naïve pluripotency can be tested by the re-injection of cells into a blastocyst and their contribution to the germline and germ layers. *In vitro*, mouse pluripotent embryonic stem cells (ESC) can be derived from the ICM around E4.5 (Boroviak et al., 2014; Nichols and Smith, 2009). However, ESC can not be directly differentiated into PGCs (Hayashi et al., 2011).

Upon implantation, the epiblast cells exit naïve pluripotency and acquire a formative pluripotent state (E5.5 - E6.25). The establishment of formative pluripotency entails an extensive remodelling of enhancers, signaling pathways and epigenetic landscapes, which ultimately enables the emergence of competence for lineage specification, including PGCs (Figure 3) (Mulas et al., 2017; Smith, 2017). Next, as gastrulation starts, the epiblast cells transit towards a primed pluripotent state (Osorno et al., 2012) (E6.5 - E7.5). Primed pluripotent cells display drastically reduced germline competence (Ohinata et al., 2009) and start to express major

specifiers of the three somatic lineages. In fact, the primed pluripotency phase ends with the commitment towards mesoderm, endoderm or ectoderm. Accordingly, epiblast stem cells (EpiSC) derived from the primed epiblast display heterogeneous expression of major somatic specifiers when cultured *in vitro* (Peng et al., 2016; Song et al., 2016; Sugimoto et al., 2015; Tsakiridis et al., 2015) and do not show chimeric contribution when injected into the pre-implantation blastocyst (Masaki et al., 2016). Nevertheless EpiSC are capable of chimeric contribution when injected into post-implantation embryos, being able to contribute to all somatic germ layers, but not to the germline (Huang et al., 2012; Kojima et al., 2014). The marked differences of the previous pluripotency stages with respect to their germline competence is further demonstrated by experiments in which the exogenous expression of the transcription factors *Nanog* (Murakami et al., 2016) or *Prdm14* (Magnúsdóttir et al., 2013) can efficiently induce PGC-like cells (PGCLC) from an *in vitro* formative stage, but not from the naïve (Murakami et al., 2016) or primed pluripotent stages (Gillich et al., 2012).

On the other hand, due to the transient nature of the formative pluripotency and the scarcity of early PGCs, it is challenging to investigate and elucidate the molecular basis of germline competence. To overcome these limitations, an *in vitro* system that faithfully recapitulates PGC specification and enables the establishment of the formative pluripotent stage and the induction of primordial germ cell-like cells (PGCLC) was previously developed by the Saitou lab (Hayashi et al., 2011). In this system, mouse embryonic stem cells (ESC) are cultured in 2i+LIF media (Ying et al., 2008), which renders the ESC most similar to the naïve pluripotent epiblast (E4.5). Subsequently, naïve ESC are differentiated into epiblast-like cells (EpiLC, similar to the formative pluripotent epiblast; E5.5), from which PGCLC can be obtained within heterogeneous embryoid bodies (EB) (Hayashi et al., 2011). In contrast, epiblast stem cells (EpiSC) resembling epiblast cells from the gastrulating embryo (>E6.5) cannot be efficiently differentiated into PGCLC and (Kojima et al., 2014), thus, display limited germline competence (Hayashi and Surani, 2009).

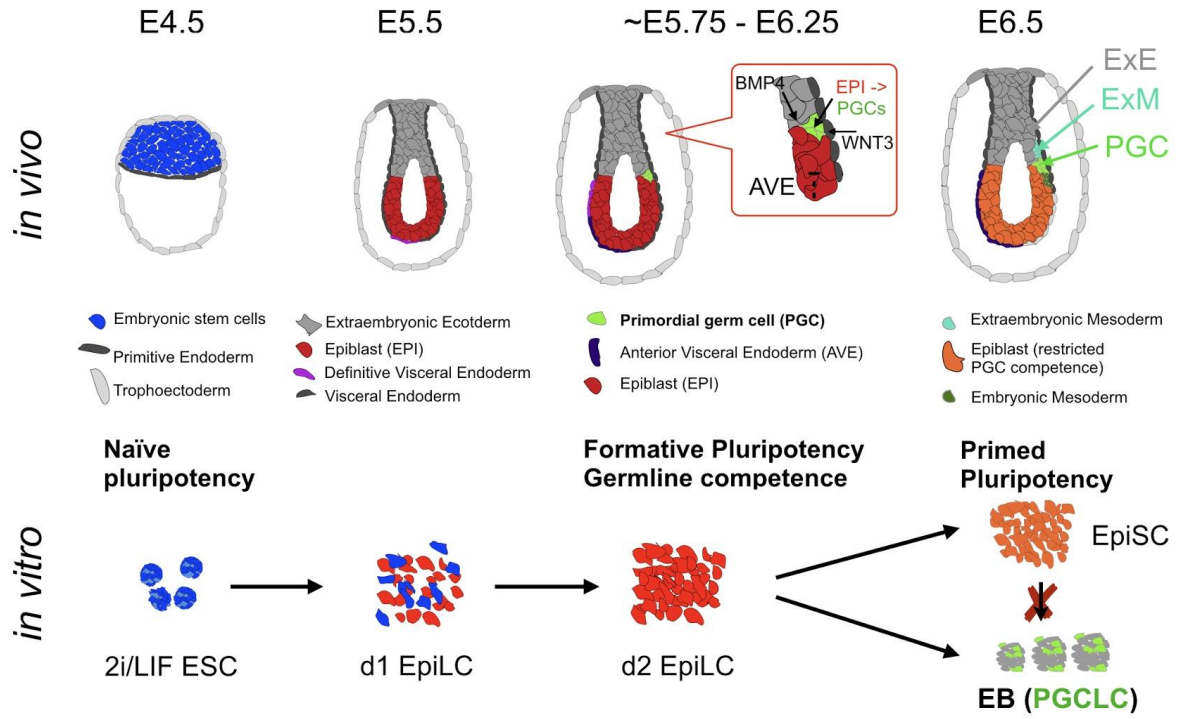


Figure 3: Schematic representation of the germline competent phase. The upper row shows the germline competence in vivo, where epiblast cells at the proximal-posterior end of a mouse embryo are induced around E6.0 by the surrounding signals of BMP4 from the Extraembryonic ectoderm and WNT3 from the visceral endoderm. The lower row shows the in vitro PGCLC differentiation system, whereby embryonic stem cells (ESC) are differentiated into epiblast-like cells (EpiLC). Day 2 (d2) EpiLC are germline competent and can be differentiated further into embryoid bodies (EB) containing primordial germ cell-like cells (PGCLC). In contrast, further progress of EpiLC into epiblast stem cells (EpiSC) restricts germline competence. The figure is adapted from Bleckwehl and Rada-Iglesias 2019.

4.5.2. Naïve, formative and primed pluripotency

The exit of the naïve pluripotent stage (E3.5-E4.5) is an asynchronous process (Kalkan et al., 2017) that occurs *in vivo* upon implantation, and ultimately results in transcriptionally homogenous epiblast cells around E5.5 (Mohammed et al., 2017). During the transition from naïve to formative pluripotency, epiblast cells undergo major epigenetic and transcriptional changes. The naïve expression program is shut down (*e.g. Prdm14, Esrrb, Zfp42*) (Boroviak et al., 2015; Zhang et al., 2018b), while

early epiblast markers become expressed (*e.g. Fgf5, Otx2, Pou3f1, Foxd3*) (Buecker et al., 2014). Meanwhile, the more general pluripotency transcription factors (*e.g. Pou5f1*/*Oct4*, *Sox2)* remain expressed at relatively constant levels throughout all pluripotent stages (Bleckwehl and Rada-Iglesias, 2019).

At the epigenetic level, the hypomethylated state that characterizes the naïve pluripotent state is repealed and global DNA methylation arises in the formative pluripotent stage together with an increase in the levels of the *de novo* DNA methyltransferases (*e.g. Dnmt3a/b/l*) and heterochromatinization (*e.g. Ehmt1/2*, H3K9me2/3) (von Meyenn et al., 2016; Zhang et al., 2018b). The increase in DNA methylation occurs at different velocities in different parts of the genome, with naïve super-enhancers showing the slowest methylation rates. Moreover, CpG-rich genes get strongly marked by H3K27me3, while low CpG-rich regions gain H3K9me2 and CpG methylation (Kalkan et al., 2017; Zylicz et al., 2015).

Furthermore, the transition between the different pluripotency states is controlled by a dedicated network of TFs that has been shown to operate through enhancers. FOXD3 is a TF that acts as a repressor. FOXD3 binds and decomissions enhancers associated with naïve pluripotency and germline genes and hence promotes the exit from naïve pluripotency (Respuela et al., 2016). A subset of the naïve pluripotency enhancers remains bound by POUF51 in the formative pluripotent state, which are then also co-occupied by OTX2 (Yang et al., 2014). However, OTX2 largely redirects POU5F1 to formative enhancers, which then become activated and initiate the expression of a formative expression programme (Buecker et al., 2014). Interestingly, the TFs FOXD3 and OTX2 not only promote formative pluripotency, but also restrict the entry into the germline and need to be silenced for PGC specification to proceed. The inducible expression of either TF reduces the number of PGCLCs, while their loss increases PGCLCs formation (Bleckwehl and Rada-Iglesias, 2019; Respuela et al., 2016; Zhang et al., 2018a). Moreover, ETV5 operates as an activator in both the naïve and formative stage for a different set of genes in each stage (Kalkan et al., 2019). Besides, the expression of general pluripotency factors, which are similar in naïve and formative cells, is controlled by different enhancers in each pluripotent stage. In ESC, the general pluripotency factors are regulated by

enhancers containing KLF4/5 motifs, while upon differentiation towards the formative pluripotent stage, they are regulated by different GRHL2-bound enhancers. This enhancer switching might partition the associated genes into smaller networks to facilitate further and diverse differentiation (Chen et al., 2018a). A divergent usage of enhancer is also evident between the naïve and primed pluripotency stages. In addition to the *de novo* establishment of primed pluripotency enhancers, naïve pluripotent super-enhancers become decommissioned in the primed pluripotent stage by losing H3K27ac, chromatin accessibility and POU5F1 binding. However, some subregions of the super-enhancers remain active throughout the primed pluripotent stage (Bell et al., 2020). Furthermore, long-range interactions between naïve super-enhancers and their target genes are lost in EpiSC (Novo et al., 2018). Notably, primed pluripotency enhancers are more extensively activated in somatic lineages than their naïve counterparts and some of them eventually become part of super-enhancers, active in different somatic lineages (Factor et al., 2014).

In summary, enhancers specific to naïve pluripotent cells, which are often controlled by naïve TF, get decommissioned concomitantly with the silencing of their target genes (Festuccia et al., 2018), while the enhancers of primed pluripotency factors are established *de novo* by formative or primed pluripotency TFs.

4.5.3. Specification of primordial germ cells

Primordial germ cells are induced from the proximal-posterior epiblast around E6.0 and will undergo several steps of reprogramming (Seah and Messerschmidt, 2018; Zeng and Chen, 2019). This reprogramming involves genome-wide reduction of CpG methylation and H3K9me2 as well as increased H3K27me3 levels. Early specifying PGCs, that already express early PGC markers (*e.g.* DPPA3) around E7.25, are not distinguishable in their genome-wide H3K9me2, H3K27me3 and CpG methylation levels from surrounding somatic cells (Seki et al., 2005). However, with the initiation of migration (E7.75), H3K9me2 is progressively lost, followed by transcriptional quiescence (G2 arrest), increased H3K27me3 levels and the reinitiation of transcription, so that post-migrative PGCs at the hindgut (E8.75) show lower levels

of H3K9me2, but higher levels of H3K27me3 compared to the surrounding somatic cells (Seki et al., 2007). Notably, these changes are regulated downstream of PRDM14 (Yamaji et al., 2008) and interestingly a KO of the H3K9 methyltransferase *Ehmt2* does not increase PGCLC specification (Zylicz et al., 2015). During this time-frame, CpG methylation levels decrease due to the absence of DNMTs in PGCs (i.e. passive demethylation), resulting in a hypomethylated genome of PGCs. However, imprinting control regions (ICRs), intracisternal A particles (IAPs) and the promoters of genes involved in meiosis and gamete generation retain DNA methylation until this phase (Seisenberger et al., 2012), which might repress early gametogenesis and meiosis (Hargan-Calvopina et al., 2016). This last barrier is removed actively by TET1/2 in the second wave of active DNA demethylation (Hackett et al., 2013; Zhang et al., 2016) around E9.5. By E10.5 PGCs reach the genital ridges and once the genome is completly hypomethylated, the remethylation occurs during spermatogenesis and oogenesis in a sex-specific manner and finally mature gamets (Larose et al., 2019; Smallwood et al., 2011).

4.6. Aim of the research

The thesis aims to understand how the epigenetic features of enhancers might contribute to germline competence. Due to the scarcity of PGCs and their precursors in the early epiblast, the use of epigenomic methods *in vivo* can be technically challenging and the resulting data can be noisy and difficult to interpret. Hereupon, the well-established *in vitro* PGCLC differentiation system (Figure 3) has been chosen to comprehensively study the dynamics of major epigenetic marks during the establishment of germline competence. Considering the important role of enhancers in the control of cellular identity, we hypothesized that by comparing the epigenetic features of enhancers in EpiLC and EpiSC, which display high and low germline competence, respectively, we could gain major insights into the molecular basis of germline competence. Besides, and from a more technical point of view, we also aimed at performing a detailed characterization of the *in vitro* PGCLC system using single-cell RNA-sequencing, as this could help us to better understand PGCLC specification, while also providing a valuable resource for the scientific community.

5. Material & Methods

5.1. Biological experiments

Most of the procedures and protocols for the biological experiments descripted in this section are also part of a preprint publication (Bleckwehl et al., 2021).

5.1.1. Cell culture and differentiation protocols

For general culture, passage and CRISPR/Cas9 genome editing experiments of mouse ESC, the cells were cultured under serum+LIF conditions. Thereby, ESC were cultured on gelatin-coated plates using knock-out DMEM (*Life Technologies*) supplemented with 15% FBS (*Life Technologies*), 1% Glutamax (*Life Technologies*), 1% Non-essential Amino acids (*Life Technologies*), Antibiotic-Antimycotic solution (*Life Technologies*) and in-house made leukemia inhibitory factor (LIF).

Before each experiment and differentiation, ESC were adapted to 2i media (serum-free N2B27 medium supplemented with MEK inhibitor PD0325901 [0.4 µM, *Miltenyi Biotec*], GSK3β inhibitor CHIR99021 [3 µM, *Amsbio*], and LIF) for at least four days in gelatin-coated tissue culture plates (Hayashi and Saitou, 2013). Consequently, all the experiments performed in ESC and described in the thesis were done under 2i conditions.

EpiLC and PGCLC differentiation were performed according to the protocols from (Hayashi and Saitou, 2013). For the EpiLC differentiation, 6-well plates were coated with 15 µg mL^{-1} Fibronectin (*Merck*) and 2 x 10^5 2i ESC were differentiated in N2B27 media supplemented with 20 ng mL^{-1} Activin A (*PeproTech*), 12 ng mL^{-1} bFGF (*PeproTech*) and 10 µL mL^{-1} KSR (*Invitrogen*) for two days (unless indicated otherwise for particular experiments). For the PGCLC differentiation, d2 EpiLC or EpiSC were plated at a density of 2 x 10^4 cells mL^{-1} and cultured as embryoid bodies (EB) on Ultra-Low attachment multiwell plates (*Corning® Costar®*) in GK15 medium supplemented with growth factors (0.5 µg mL^{-1} BMP4 (*Miltenyi Biotec*), 0.1 µg mL^{-1} SCF (*Miltenyi Biotec*), 1000 U mL^{-1} ESGRO® Recombinant Mouse LIF Protein

(*Merck*) and 50 ng mL^{-1} EGF (*PeproTech*), no BMP8a). The percentage of PGCLC formed within the EBs were typically analyzed after four days by FACS and RT-qPCR measurements of representative PGCLC markers.

EpiSC differentiations were performed according to (Guo et al., 2009). 2i ESC were plated at a density of 2.5 x 10^5 cells mL^{-1} on plates coated with 15 µg mL^{-1} Fibronectin (*Merck*) and differentiated into EpiSC by passaging them in N2B27 media supplemented with 20 ng mL^{-1} Activin A (*PeproTech*) and 12 ng mL^{-1} bFGF (*PeproTech*) for at least eight days or as indicated in the results section for particular experiments.

5.1.2. Enhancer and gene deletions with CRISPR/Cas9

Pairs of gRNAs flanking each of the selected enhancers or genes were designed with *benchling* (https://www.benchling.com). ESC were transfected with pairs of CRISPR/Cas9-gRNA expressing vectors (Cong et al., 2013) specific for each enhancer or gene of interest using Lipofectamine 3000 (*Invitrogen*). After transfection and puromycin selection (2 µg mL^{-1}, *Life Technologies*) for two days, the cells were recovered for another day in Serum+LIF media and then single cell dilutions were seeded into a 96-well plate to screen for clonal cell lines. Subsequently, the clonal lines were screened for the intended deletion by a PCR with the GoTaq® DNA Polymerase (*Promega*) and two primers outside the gRNAs sites. Positive cell lines were further validated by a negative PCR with primers inside the deletion and the PCR product with the outside primers was verified by Sanger-sequencing. The gRNAs used for the *Esrrb, Lrrc31/34, Klf5* and *Prdm14* enhancer deletions in E14Tg2a ESC as well as the *Otx2* gene deletion in R1 ESC are listed in Table 1. The *Otx2*$^{-/-}$ E14Tg2a ESC cell line was kindly provided by Christa Buecker (Max Perutz Labs, Austria).

Table 1: gRNA sequence and genomic coordinates (mm10) of the CRISPR/Cas9 mediated deletions.

Target	gRNA 1	gRNA 2	Coordinates	Size - bp
Esrrb Enhancer	GCTTCTTGAATT GAGAACGG	GTAACCGCCAC AGGACACTC	chr12:86497006-86503356	6350
Lrrc31/34 Enhancer	GTAGGCCCACC ATTGCAACC	GAAGCTGACGA GGTGTTGAG	chr3:30722005-30726105	4100
Klf5 Enhancer	GTTTCAGAAAC AACTTCACA	GTTGGCTAAAC CCAGTACGT	chr14:99338943-99341862	2919
Prdm14 Enhancer 1	GTACAGTCAGT CATCTGAGG	CCTAGCTATATT CCCCTCCA	chr1:13072129-13076981	4852
Prdm14 Enhancer 2	GTTTTGCAAGC TGGTTCTGT	gTAAGATACTAA GTAGAAGCC	chr1:13084831-13085465	634
Prdm14 Enhancer 3	GCCCAGATGTG TGAACTCAG	GAGGTTATCAG TGATAATGT	chr1:13093655-13096290	2635
Otx2 gene	GGTCATGGCCT TCAAAGCAG	GCAAAGTCGGC CCAAATTGG	chr14:48658344-48663679	5335
Prdm14 auxin degradation tag	GAGGCTTCATA AAACCTGCT	N.A.	chr1:13114256 -13114276	N.A.

5.1.3. The generation of transgenic ESC lines

***Kmt2c/d* catalytic mutant cells:** The *Kmt2c/d* (dCD), *Kmt2c/d* (dCT) and *Kmt2d* (CT) catalytic mutant ESC lines were kindly provided by Joanna Wysocka. The dCD was previously described (Dorighi et al., 2017) and all the catalytic mutant cell lines were generated similarly by co-transfecting R1 ESC with a 200 bp single stranded oligonucleotide donor template harboring the desired point mutations and a vector coexpressing Cas9 and gRNAs targeting *Kmt2c* and *Kmt2d*. After the selection of clonal lines, the intended *Kmt2c/d* mutations were confirmed by Sanger sequencing. For the dCD *Otx2*$^{-/-}$ cell line, *Otx2* was deleted in the dCD cell line using the gRNAs from Table 1.

PGCLC reporter cell line: Based on the previous use of DPPA3 expression as a readout for PGCLC quantifications (Hayashi et al., 2011), a DPPA3-GFP reporter ESC line was generated as follows. First, a repair template harbouring homologous arms to the C-terminus of *Dppa3* and sequences encoding the T2A-peptide and

turboGFP was designed. Next, E14 ESC were co-transfected with the repair template and a vector co-expressing Cas9 and a gRNA targeting the C-terminus of *Dppa3*. Then, single cells were sorted into 96-well plates and the resulting clonal ESC lines were genotyped by PCR. A heterozygous ESC line showing high levels of GFP upon PGCLC differentiation in the presence of growth factors was selected for further analyses.

Auxin inducible PRDM14 degron: First, the *Tir1* gene from *Oryza sativa* (*OsTir1*) was integrated at the *Tigre* locus with CRISPR/Cas9 as previously described (Nora et al., 2017) and verified by PCR and Sanger-sequencing. Then, the minimal sequence of the auxin degradable tag and an in-frame FLAG-tag were cloned in a donor plasmid to fuse it in-frame at the C-terminus of *Prdm14.* For this purpose, the CRISPaint technique (Schmid-Burgk et al., 2016) was used and the donor plasmid was co-transfected with the 'Frame selector 2' plasmid and the gRNA listed in Table 1. Finally, the degradation of PRDM14 was induced with 500 µM indole-3-acetic acid (*Sigma Aldrich*) and verified by western blot.

Ectopic and inducible expression of PGCLC transcription factors: Mouse *Prdm14* and *Nanog* cDNAs were amplified and cloned into a Doxycycline (Dox)-inducible piggyBac vector (Calo et al., 2018) using the primers listed in Table 2. The resulting piggyBac vectors enabled the ectopic expression of the selected genes fused with a HA-tag at their C-terminus and were transfected together with a Super PiggyBac Transposase expressing vector (*Systems Bioscience*) and a Tet transactivator in ESC. After transfection, single cells were seeded into a 96 well plate to derive clonal ESC lines. The clonal ESC lines with the lowest expression of the transgenic genes in the absence of Dox were selected. To evaluate the effects of PRDM14 and NANOG ectopic expression on PGCLC specification, 1 µg mL^{-1} Dox was added once PGCLC differentiation had been started. To investigate the effects of the ectopic expression of PRDM14 or NANOG in EpiLC and EpiSC, 1 µg mL^{-1} Dox was added 18 hours before the cells were collected for downstream ChIP-seq experiments (*i.e.* EpiLC differentiation was initiated, after 30 hours Dox was added, and after 48 hours cells were collected; EpiSC were maintained in differentiation media for at least eight passages and then treated with Dox for 18 hours).

Table 2: Primers used for the construction of the piggyBac vectors with the annealing sequence for the cDNA highlighted in blue.

cDNA	Forward primer	Revers primer
Prdm14	AAAAACTCGAGGCCACCATGTC AGAAGTCTAGGGTCCCG	AAAAAACTAGTGCAGGTTTTATG AAGCCT
Nanog	AAAAACTCGAGGCCACCATGAG TGTGGGTCTTCCT	AAAAAACTAGTTATTTCACCTGG TGGAGT

5.1.4. Quantification of PGCLC by flow cytometry

In general, PGCLC were quantified using antibody staining and flow cytometry (Hayashi and Saitou, 2013). Briefly, after four days of PGCLC differentiation, the resulting EB were dissociated and stained for 45 minutes with antibodies against CD61 (*Biolegends*) and CD15 (*eBioscience*) conjugated with PE and Alexa Fluor 647, respectively. The deletions of the enhancers associated with *Esrrb*, *Klf5* and *Lrrc31* were generated in a DPPA3-GFP reporter line. Therefore, PGCLC quantifications in the resulting cell lines with enhancer deletions were performed by measuring the number of cells expressing high levels of GFP (FITC) after four days of PGCLC differentiation.

All PGCLC quantifications were performed using the FACS CantoII Cytometer (*BD Bioscience*) equipped with the BD FACSDiva Software. PGCLC sorting was performed on a FACS AriaIII cell sorter (*BD Bioscience*).

5.1.5. RNA isolation, cDNA synthesis and RT-qPCR

Total RNA from ESC, EpiLC and EpiSC was extracted following the protocol of the innuPREP DNA/RNA mini kit (*Analytik Jena*), while for the RNA extraction from d2 EB, d4 EB and sorted PGCLC the ReliaPrep™ RNA Miniprep Systems (*Promega*) was used. cDNAs were generated using the ProtoScript II First Strand cDNA Synthesis Kit and Oligo(dT) primers (*New England Biolabs*). RT-qPCRs were performed on the LightCycler 480II (*Roche*) using the ORA-qPCR Green ROX L Mix (*HighQu*). For the normalization of the expression *Hrpta* and *Eef1a1* were measured

and all RT-qPCR primers that were used for the measurement of the gene expressions are listed in Table 3 below.

Table 3: Sequence of the RT-qPCR primers for RNA abundance.

Gene	Forward primer	Reverse primer
Hprt	CAAGGGCATATCCAACAACA	GCCCCAAAATGGTTAAGGTT
Eef1a1	TAGACGAGGCAATGTTGCTG	AGCGTAGCCAGCACTGATTT
Gata6	CTACACAAGCGACCACCTCA	TGTAGAGGCCGTCTTGACCT
Dnmt3b	ACAACCGTCCATTCTTCTGG	GTGAGCAGCAGACACCTTGA
Sox17	AGCAGAACCCAGATCTGCAC	GCTTCTCTGCCAAGGTCAAC
Klf4	CAGGCTGTGGCAAAACCTAT	CCTGTGTGTTTGCGGTAGTG
Nanog	AAGTACCTCAGCCTCCAGCA	GCTTGCACTTCATCCTTTGG
Nanos3	CCTGACAAGGCAAAGACACA	CTTCCTGCCACTTTTGGAAC
Prdm14	GCCTGAACAAGCACATGAGA	TGCACTTGAAGGGCTTCTCT
Esrrb	GACCTTTACCGAGCCATCCT	GCCTCCAGGTTCTCAATGTA
Klf5	CGACGTATCCACTTCTGCGAT	CAGGTGCACTTGTAGGGCTT
Lrrc31	AGACGCCTCTGTGTACCTCC	CGTAGCACTCGAAGCAACCT
Lrrc34	ACAATCGCTTAGACTCCGGACA	CCACACCACTGATACATGGCT

5.1.6. Bulk RNA-seq and single-cell RNA-seq

Total RNA from WT and dCD EpiLC (2 replicates), EpiSC (2 replicates) and d4 EB (3 replicates dCD / 4 replicates WT) were purified as described above. Bulk RNA-seq libraries were generated following the protocol of the TruSeq stranded kit (*Illumina*).

For scRNA-seq, EpiSC were dissociated with StemPro® Accutase® (*Life Technologies*) and all other cell types with TypLE™ Express Enzyme (*Life Technologies*). The cells were then centrifuged and resuspended in PBS containing 0.04 % BSA. Next, cells were passed through a cell strainer, the cell concentration was determined, and the scRNA-seq libraries were prepared using the Chromium Single Cell Gene Expression (*10x Genomics*) according to the Single Cell 3’

Reagents Kit (v2) for the single-cell experiment of the PGCLC differentiation and the Single Cell 3' Reagents Kit (v3) protocol for the d4 EB from R1 WT and dCD cells.

5.1.7. Bisulfite sequencing and ChIP-bisulfite sequencing

For the locus-specific bisulfite sequencing experiments, genomic DNA was purified with phenol-chloroform and subjected to bisulfite conversion according to the EZ DNA Methylation-Direct™ Kit (Zymo Research) by incubating 200 ng of purified DNA with the CT-Conversion reagent after a short heat up (98 °C, 8 minutes) for 3.5 hours at 64 °C. Following washing and desulphonation on a column, a fragment of 300 - 500 bp of the eluted DNA was amplified by the EpiTaq HS polymerase (*TaKaRa*). Finally, the resulting PCR products were gel-purified and subjected to blunt-end cloning with the pGEM-T vector system (*Promega*). Individual colonies harboring a bisulfite converted insert were analyzed by Sanger sequencing.

Table 4: Sequences of the bisulfite primers used for locus-specific and ChIP-bisulfite sequencing of PGCLC enhancers.

PGCLC Enhancer locus	Forward bisulfite primer	Reverse bisulfite primer	Genome Coordinates
Esrrb	TGTATTTTAGGTTAGTTTTTGAATTT	ACTAATCATCTCATTTTTAATCTACAC	chr12:86498315-86498631
Lrrc31/34	TTTAGGTATATGAGTATATTGTAGTTGT	TAAATCTTACTTCATTTAAAAAACC	chr3:30724994-30725297
Prdm1	GTATTTTTTGTTTTGTTTTGTTTTG	ACTTACTAATTTCAATCTCCTTC	chr10:44390290-44390668
Prdm14 (E3)	TTGGTTGGGATTAAGTTATTATTGTT	ATCCATTTACCCTCTTAAATCTTTCTATC	chr1:13095658-3095873

ChIP-bisulfite experiments were performed as described in (Thomson et al., 2010) with slight modifications. Firstly, the ChIP protocol described above was followed. After the final DNA purification, all the resulting H3K4me1 and H3K4me2 ChIP DNAs and 200 ng of the corresponding input DNA were subjected to bisulfite conversion according to the EZ DNA Methylation-Direct™ Kit (*Zymo Research*). Then, the

subsequent amplification, purification and sequencing steps were performed as described for the locus-specific bisulfite sequencing experiments. For the *Esrrb*, *Lrrc31/34*, *Prdm1* and *Prdm14* enhancers, the bisulfite primers listed in the following Table 4 were used.

For the genome-wide bisulfite sequencing experiments, libraries were prepared from column-purified DNA of d2 EpiLC using the PBAT method previously described in (Clark et al., 2017). Briefly, for the bisulfite conversion, the instructions of the EZ Methylation Direct MagPrep Kit (*Zymo*) were followed. After purification, bisulfite converted DNAs were eluted from MagBeads directly into 39 µL of first strand synthesis reaction mastermix (1x Blue Buffer (*Enzymatics*), 0.4 mM dNTP mix (*Roche*), 0.4 µM 6NF preamp oligo (*IDT*)), heated to 65 °C for 3 minutes and cooled on ice. 50 U of klenow exo- (*Enzymatics*) was added and the mixture was incubated on a thermocycler at 37 °C for 30 minutes after slowly ramping from 4 °C. Reactions were diluted to 100 µL and 20 U of exonuclease I (*New England Biolabs*) was added and incubated at 37 °C before purification using a 0.8:1 ratio of AMPure XP beads. Purified products were resuspended in 50 µL of second strand master mix (1x Blue Buffer (*Enzymatics*), 0.4 mM dNTP mix (*Roche*), 0.4 µM 6NR adaptor 2 oligo (*IDT*)), then heated to 98 °C for 2 minutes and cooled on ice. 50 U of klenow exo- (*Enzymatics*) was added and the mixture was incubated on a thermocycler at 37 °C for 90 minutes after slowly ramping from 4 °C. Second strand products were purified using a 0.8:1 ratio of AMPure XP beads, resuspended in 50 µL of PCR master mix (1x KAPA HiFi Readymix, 0.2 µM PE1.0 primer, 0.2 µM iTAG index primer), and amplified with 12 cycles. The final libraries were purified using a 0.8:1 volumetric ratio of AMPure XP beads before pooling and sequencing. All libraries were prepared in parallel with the pre-PCR purification steps carried out using a Bravo Workstation pipetting robot (*Agilent Technologies*).

5.1.8. Assay for transposase-accessible chromatin (ATAC-seq)

The ATAC-seq protocol was adapted from (Buenrostro et al., 2013), using 3.8 x 10^4 and 5.0 x 10^4 cells for the two replicates generated for each of the investigated cell

types, respectively. Briefly, the cells were lysed in lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL CA-630) supplemented with freshly added protease inhibitor for 15 minutes. Following centrifugation, the pellet was resuspended in a Tn5 transposase reaction mix (*Illumina*) for 30 minutes at 37°C. Following DNA purification with the MinElute PCR purification kit (*Qiagen*), libraries were prepared with the Nextera DNA library prep kit (*Illumina*) and the fragmentation was checked with the Bioanalyzer (*Agilent*).

5.1.9. Circular Chromatin Conformation Capture sequencing

Circular Chromatin Conformation Capture (4C) followed by sequencing was performed as previously described (Laugsch et al., 2019). Briefly, cells grown to confluence in 10 cm dishes were crosslinked with 1% formaldehyde for 10 minutes, quenched with 0.125 M glycine, washed with PBS and incubated in lysis buffer (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% Triton X-100 and 1x protease inhibitor) on ice for 10 minutes. Nuclei were then re-suspended in 0.5 mL 1.2x CutSMART buffer (*New England Biolbas*) with 0.3% SDS and incubated at 37°C with constant shaking for 1h. Then, Triton X-100 was added to a final concentration of 2%, followed by 1h incubation at 37°C with constant shaking. Next, 600 U of *NlaIII* (*New England Biolabs*) were added to digest chromatin overnight at 37°C with constant shaking (900 rpm). *NlaIII* was inactivated by adding SDS to a final concentration of 1.6% and incubating the mixture for 20 minutes at 65°C. The digested chromatin was transferred to 50 mL tubes, and 1x ligation buffer (50 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 1 mM ATP, 1mM DTT) was added up to a final volume of 7 mL. Triton X-100 was added to a final concentration of 1% and the solution was incubated for 1h at 37°C and 300 rpm. Afterwards, the digested chromatin was ligated using 100 U of T4 DNA ligase (*Life Technologies*) for 4 hours at 16°C, followed by RNase A treatment (300 µg total, *Peqlab*) for 45 minutes at 37°C. Subsequently, chromatin was de-crosslinked by treating with Proteinase K (300 µg total, *Peqlab*) and incubating at 65°C overnight. DNA was purified by phenol/chloroform extraction, followed by ethanol precipitation and re-suspended in

100 µL of water. The purified DNA was digested with 50 U of *DpnII* (*New England Biolabs*) in 500 µL of 1x NEBuffer at 37°C overnight. DNA samples were again purified (phenol/chloroform extraction and ethanol precipitation), re-suspended in 500 µL of water and ligated with 200 U of T4 DNA Ligase in a final volume of 14 mL 1x ligation buffer overnight at 16°C. DNA samples were subjected to another round of phenol/chloroform extraction and ethanol precipitation, re-suspended in 150 µL of water and purified by QIAquick PCR Purification Kit (*QIAGEN*). Finally, the resulting 4C DNA products were amplified by inverse PCR using primers located close to the transcription start site (Table 5). The inverse PCRs were performed with the Expand™ Long Template PCR System (*Roche*) using 32 amplification cycles (94°C 2 minutes, 32x [94°C 10 s, 58°C 1 minute, 68°C 3 min], 68°C 5 min). 4C-seq libraries were sequenced on the HiSeq 2500, generating reads of 74 bp in length.

Table 5: 4C-seq index primers with the annealing sequence highlighted in blue.

Viewpoint	p5 index primer	p7 index
Prdm14 TSS	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT-barcode-CCTTCTGTGTGTAAACGCA	CAAGCAGAAGACGGCATACGATTCCAGGGAACGGTCAAGAG
Esrrb TSS	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT-barcode-TCATACCCGTAGACACGCCC	CAAGCAGAAGACGGCATACGATGGCGGTCCCATCTAAAGTA
Tfap2c TSS	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT-barcode-TGAGGAAAGGCGAATGCAGA	CAAGCAGAAGACGGCATACGAAGAGAGAACCTACTCCCGCC

5.1.10. Chromatin Immunoprecipitation (ChIP-seq) and ChIPmentation

In the standard ChIP protocol, cells were cross-linked with 1 % formaldehyde for 10 minutes, quenched with 0.125 M glycine, harvested with a cell scraper and washed twice with PBS containing cOmplete™ EDTA-free Protease Inhibitor Cocktail (*Roche*). The cells were sequentially lysed in three buffers (lysis buffer 1: 50 mM HEPES, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% TX-100; lysis buffer 2: 10 mM Tris, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA; lysis Buffer 3:

10 mM Tris, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine) with rotation for 10 minutes in between. Then, the chromatin was sonicated with the Epishear™ Probe Sonicator (*Active Motif*) with 20 s ON and 30 s OFF for 8 cycles. After centrifugation, the supernatant was divided into input and ChIP samples. The ChIP samples were incubated overnight with specific antibodies listed in Table 6 below.

Table 6: Antibodies used for ChIP experiments.

Antibody	Source	ID	Dilution
H3K4me1	*Active Motif*	39297	4 µL
H3K4me2	*Active Motif*	39141	4 µL
H3K4me3	*Active Motif*	39159	4 µL
H3K9me2	*Active Motif*	39683	5 µL
H3K9me3	*Active Motif*	39161	5 µL
H3K27ac	*Active Motif*	39133	4 µL
H3K27me2	*Cell Signaling*	9728S	5 µL
H3K27me3	*Active Motif*	39155	5 µL
HA	*Abcam*	ab9110	4 µL

The next day, Dynabeads Protein G beads (*invitrogen*) were washed with Blocking Solution (PBS with 0,5 % BSA) and incubated with the chromatin for another 6-8 hours at 4 °C. Then the beads were washed with RIPA buffer (50 mM Hepes, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.7% Na-Deoxycholate) on a magnet to capture antibody bound chromatin, which was then eluted (50 mM Tris, 10 mM EDTA, 1% SDS) and reverse crosslinked at 65 °C overnight in parallel with the input. Finally, DNAs were purified with the ChIP DNA Clean & Concentrator (*Zymo Research*) and ChIP libraries were generated with the TruSeq kit (*Illumina*) or analyzed by quantitative PCR (qPCR) with the primers listed in Table 7.

Table 7: Sequences of the qPCR primers for ChIP.

Enhancer	Forward primer	Reverse primer
Esrrb	AAAGTGGCCAGGACTCCACT	TGCACGGCAACTCTTGAGAA
Klf5	CCACAGGAACCAGTAACAGAGG	AGCTTTCTGAACCATCTGGACC
Lrrc31/34	ATCTACTCAAGGCCAGCTCCTC	GCTCTGGTATTCATAGCTGTCCG
Espr1	ACTGACAACCCAGCACCAAA	CCACGCTCAGGAACTGTGAT
Prdm1	TCACAATAGCAAGGCTCCGC	CCAACCGGAAGTAGGCTGAC
Prdm14 (E1)	GCAAGCAAGCAAGCTAACAGAG	TGGCAAGTCTCTCCTCCACA
Zfp42	CGGTTCCGTAGTGAAGCTTTGT	AGCCCGTTCCTTACAGGTGA
ctrl region	CCTGAGGCTGGAAGTTTCTG	CTCCTGGGATTAAAGGCACA

For the generation of ChIP-seq profiles from low cell numbers, like d4 EB obtained with the PGCLC differentiation system, the ChIPmentation protocol described by (Schmidl et al., 2015) was adapted. To reduce the sample loss, the crosslinking with 1% Formaldehyde and the later quenching was performed in a 1.5 mL reaction tube and 100 µL reaction volume without pipetting in the solution. After one PBS washing step, the cells were snap-frozen and then resuspended in 160 µL Sonication Buffer (10 mM Tris-HCl [pH8], 0.25% SDS, 2 mM EDTA + freshly added 1,6 µL PIC and 1,6 µL PMSF). Then the lysate was diluted with 80 µL of Equilibration Buffer (30 mM Tris-HCl [pH8], 700 mM NaCl, 5% Triton X-100, 0.5 % Deoxycholic acid, 3mM EDTA) and sonicated with the small probe of the EpiShear (*Active Motif*) sonicator (25% amplitude, pulse for 20 seconds ON and 30 seconds OFF) for 8 cycles. Cellular debris was separated by centrifugation and the supernatant was transferred to a new PCR reaction tube (2 % input was set aside). The chromatin was incubated with 1 µL of antibody overnight and, after washing with PBS + 0.1% BSA, 10 µL of Protein G Dynabeads were incubated separately in 0,1 % BSA on a rotator overnight at 4 °C. The next day, the chromatin and beads were incubated together for 6 hours, washed with RIPA buffers and then subjected to Tagmentation (5 µL Tagmentation Buffer, 1 µL Tagmentation DNA Enzyme (*Illumina*), 19 µL Nuclease free water) in order to incorporate the sequencing adapters. Lastly, DNAs were eluted from the beads and used for library preparation with the Nextera DNA library preparation kit (*Illumina*).

5.1.11. Western Blot

Nuclei were isolated by incubating cells with Lysis Buffer (20 mM Tris pH 7.6, 100 mM NaCl, 300 mM sucrose, 3 mM $MgCl_2$) containing freshly added protease inhibitors for 10 minutes at 4°C and then centrifuged for 10 minutes at 4°C and 3000 rpm. The resulting pellets, containing the cell nuclei, were treated with a high salt buffer (20 mM Tris pH 8.0, 400 mM NaCl, 2 mM EDTA pH 8.0) and disrupted with a glass homogenizer on ice. After incubation on ice for 30 minutes and centrifugation (24000 x g for 20 minutes at 4°C), supernatants were collected and protein concentration was estimated by the Pierce™ BCA Protein Assay Kit (*Thermo Fisher*). Then 20 µg of the protein extracts were heated in Laemmli buffer at 95 °C for 5 minutes, loaded on 4–15% Mini-PROTEAN® TGX™ Precast Protein Gels (*Bio-Rad*) and transferred (190 mM glycine, 25 mM Tris, 20% Methanol, 0.1% SDS) to a PVDF membrane. After blocking with 5 % milk, the membrane was incubated overnight at 4 °C first with a primary antibody (Table 8) and then with a matching secondary antibody (coupled to a horseradish peroxidase (HRP)) for 1h at room temperature. In between the membrane was washed with PBST. Finally, the reaction of the HRP was visualized with the Lumi-LightPlus Western Blotting Substrate (*Roche*).

Table 8: Antibodies used for Western Blots.

Antibody	Source	ID	Dilution
HA-tag	*Abcam*	ab9110	1:5000
FLAG-tag	*Merck*	F1804	1:2500
NANOG	*Bethyl*	A300-397A	1:2000
ESRRB	*Perseus Proteomics*	PP-H6705-00	1:1000
OTX2	*Proteintech*	13497-1-AP	1:1500
beta-TUBULIN	*Sigma-Aldrich*	T0198	1:2000

5.2. Computational data analysis

Most of the computational analyses mentioned in this chapter are part of a preprint publication (Bleckwehl et al., 2021). In-house codes and scripts are available on Github:

https://github.com/ToreBle/Germline_competence

5.2.1. Data availability and public data resources

All public available data are referenced in the results and figures sections and are listed in Table 9. The sequencing datasets generated in the thesis have been deposited in the GEO database (GSE155089) under the following accession numbers:

- GSE155015 - 4C-seq
- GSE155058 - ATAC-seq
- GSE155062 - ChIP-seq (Epigenetic features related to germline competence)
- GSE155069 - ChIP-seq (*Kmt2c/d* dCD and *Otx2*$^{-/-}$ during differentiation)
- GSE155079 - Bulk RNA-seq
- GSE155083 - Genome-wide DNA methylation
- GSE155088 - single-cell RNA-seq (PGCLC differentiation and d4 EB dCD)

Table 9: Public available datasets used in the thesis.

GEO accession	Data set	Reference
E-MTAB-6153	single-cell RNA-seq from whole E8.25 embryos	(Ibarra-Soria et al., 2018)
GSE60204	H3K27ac from PRDM1/DPPA3+ sorted PGCLC	(Kurimoto et al., 2015)
GSE72164	capture Hi-C from ESC (ATAC- and CTCF-site defined interactome)	(Atlasi et al., 2019)
GSE113981	Hi-C data from ESC, EpiLC, EpiSC	(Miura et al., 2019)
DRA003471	Bulk RNA-seq from *Prdm14*$^{-/-}$	(Shirane et al., 2016)
GSE56138	Bulk RNA-seq from *Otx2*$^{-/-}$ EpiLC	(Buecker et al., 2014)

GSE41923 GSE70355	Genome-wide DNA methylation in ESC, EpiLC and EpiSC	(Habibi et al., 2013; Zylicz et al., 2015)
DRA003471	Genome-wide DNA methylation during PGCLC differentiation	(Shirane et al., 2016)
GSE76505	Genome-wide DNA methylation from *in vivo* epiblast (E4.5 - 6.5)	(Zhang et al., 2018b)
GSE109767	DNAseI from OCT4-GFP sorted *in vivo* PGC (E9.5 - E10.5)	(Li et al., 2018a)
GSE121650 GSE121690	single-cell NMT-seq (parallel single-cell NOME-, Methylation and RNA-seq)	(Argelaguet et al., 2019)

5.2.2. Single-cell RNA-seq data analysis

The 10x Genomics scRNA-seq data can be easily explored with the Loupe Cell Browser (https://www.10xgenomics.com) and the .cloupe files deposited at GEO (GSE155088).

For the 10x Genomics data, UMIs were counted using NCBI:GCA_000001635.6 and *cellranger-2.1.0* (Zheng et al., 2017). The resulting UMI values were aggregated into a single matrix with default normalization (“–normalize=mapped”).

M*onocle2* (Trapnell et al., 2014) was used to evaluate the *in vitro* scRNA-seq data generated across the different PGCLC differentiation stages and K-means clustering was performed on the t-SNE plots (with k=3 for d2 EB and k=4 for d4 EB). From the resulting clusters, those containing PGCLC were identified by the enrichment of previously defined core PGC genes from d4/d6 PGCLC and E9.5 PGCs (Nakaki et al., 2013). To determine the cellular identity of the remaining clusters found within the EB, the expression of lineage specific markers identified in E8.25 mouse embryos (Ibarra-Soria et al., 2018) was used. To this end, all markers with a log2FoldChange >2.5 were considered. Each EB cluster was annotated as equivalent to the mouse

embryonic tissue for which the most significant enrichment in the expression of the corresponding marker genes was observed.

The ESC, EpiLC, EpiSC and PGCLC gene sets were defined by differential expression using *Seurat* (Stuart et al., 2019) and the “negbinom” option for differential expression testing. ESC, EpiLC and EpiSC genes were determined by differential expression between the stage indicated vs the remaining two (e.g. ESC vs. d2 EpiLC and EpiSC). From this analysis, only the genes upregulated in the particular stage (adjusted p-value < 0.005), a high expression in the stage of interest (more than 20 %, pct.1 >0.2) and a low expression distribution in the others (less than 40%, pct.2<0.4) were considered. Similarly, PGCLC genes were determined by differential expression between the d2+d4 PGCLC clusters and the remaining clusters from d2 and d4 EB. Again, from this analysis only the genes upregulated in PGCLC (adjusted p-value < 0.005) and with high expression in PGCLC (expressed in 20 % of the PGCLC) and a low expression distribution in the other analyzed cells (expressed in less than 40% of non-PGCLC EB cells) were considered. To quantify the expression of PGCLC genes in different analyses, the UMI count matrix (of the *in vitro* stages) or FPKM-normalized data (of the *in vivo* stages) were used to calculate the mean expression of all PGCLC genes within a cell or the mean expression of each PGCLC gene across all cells of a stage.

For the RNA velocity analysis, spliced and unspliced read counts were obtained with *kallisto* (Bray et al., 2016) and *bustools* (Melsted et al., 2019), parsed into R-3.6.1 to create a *Seurat* object and t-SNE plot, which was then overlaid by the RNA velocity calculations from *velocyto.R* (La Manno et al., 2018).

For the scRNA-seq data from E4.5 - E6.5 mouse embryos (Argelaguet et al., 2019) the count matrix (GSE121650) was normalized to FPKM (*edgeR*, Ensembl gene annotation, v87). Then, the previously generated lineage assignments (Argelaguet et al., 2019) were used to solely select epiblast cells for further analysis.

The estimation of transcriptional noise for the epiblast stages *in vivo* and *in vitro* was performed like in (Mohammed et al., 2017). First, the 500 most variable genes for each stage were selected and pairwise comparisons were performed by calculating

Spearman correlations. Then, the Spearman correlation values were transformed into distance measurements ($\sqrt{(1-\rho)/2}$) that were considered to represent transcriptional noise.

For the single-cell RNA-seq data set of the d4 EB from R1 WT and dCD cells, single cells with a high percentage of mitochondrial gene expression were discarded by considering only single cells with at least 2000 expressed genes. The clusters were identified in the combined dataset of R1 WT and dCD cells by shared nearest neighbor (SNN) modularity optimization based clustering algorithm from Seurat with a resolution of 0.3. This resulted in 9 clusters that were classified based on the marker gene expression of E8.25 mouse embryos (Ibarra-Soria et al., 2018). The expression of the markers was verified by differential expression analysis between all clusters, leading also to the identification of those clusters that did not show any specific marker gene expression in the E8.25 mouse embryos (*e.g.* PGCLC, 2-cell-like). The *Prdm1orDppa3+/Klf4-* cells were defined using the following gene expression values; Dppa3 > 0.1 OR Prdm1>1 AND Klf4<1. The single-cell expression correlation values were determined by the spearman correlation of the top 1000 most variable genes in each cluster and the correlation of their expression among the cells belonging to the same cluster.

5.2.3. Differential expression analysis of bulk RNA-seq data

Public available RNA-seq data generated across different time-points of EpiLC differentiation (Yang et al., 2019a), *Otx2*$^{-/-}$ EpiLC (Buecker et al., 2014) and *Prdm14*$^{-/-}$ EpiLC (Shirane et al., 2016) were mapped with *STAR* (Dobin and Gingeras, 2015) to the mouse reference genome (*Ensembl* gene annotation, v99) and reads within genes were counted with *featureCounts* (Liao et al., 2014). The rlog normalization was generated with *DESeq2* (Anders and Huber, 2010).

For the bulk RNA-seq data generated in R1 WT and dCD cells, three replicates for R1 WT and the dCD ESC previously generated by (Dorighi et al., 2017), two replicates for R1 WT and dCD d2 EpiLC and two replicates for R1 WT and dCD EpiSC were analyzed. The transcript abundance was determined with the bootstrap

mode of *kallisto* (-b 100) (Bray et al., 2016) and the differential expression analyses were performed with *sleuth* on the gene-level (*Ensembl* gene annotation, v96) (Bray et al., 2016; Yi et al., 2018).

5.2.4. ATAC-seq data processing

For ATAC-seq data processing, paired end reads were mapped to the mouse genome (*mm10*) using *BWA mem (Li and Durbin, 2009)*. Duplicates and reads within blacklisted regions were discarded. After checking the amount of mononucleosome reads (30-115 bp), the concordance of the ATAC-seq replicates were checked (Spearman correlation of a 2 Kb window for ESC: 0.86; EpiLC: 0.88; EpiSC: 0.86). Then BAM files for each stage were merged and converted into bigWig files by normalization to 1x sequencing depth with *deeptools-3.3.1* (Ramírez et al., 2016).

5.2.5. ChIP-seq visualization, peak calling and motif analysis

For ChIP-seq data processing, single end reads were mapped to the mouse genome (*mm10*) using *BWA mem* (Li and Durbin, 2009). After duplication removal with the *MarkDuplicates* function from the *Picard* tools, reads within blacklisted regions were discarded and the aligned reads were normalized with *deeptools-3.3.1* to 1x sequencing depth (as RPGC: Reads per genomic content, (Ramírez et al., 2016)):

```
bamCoverage --binSize 5 --normalizeUsing RPGC --effectiveGenomeSize
2150570000 --ignoreForNormalization chrX chrY
```

The processed ChIP-seq data of the thesis can be explored in the UCSC browser:

- Comparison of the pluripotent stages:

http://genome-euro.ucsc.edu/s/Tore/Comparison%20of%20the%20pluripotent%20stages

- PRDM14/NANOG overexpression and topological features:

http://genome-euro.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=Tore&hgS_otherUserSessionName=PRDM14%2FNANOG%20overexpression

- H3K4me1/2 deficiency and *Otx2*$^{-/-}$:

http://genome-euro.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=Tore&hgS_otherUserSessionName=H3K4me1%2F2%20deficiency%20and%20Otx2%2D%2F%2D

5.2.6. Definition of stage-specific enhancers

PGCLC enhancers: The H3K27ac ChIP-seq data (SOLiD sequencing) from d2 and d6 sorted PGCLC (Kurimoto et al., 2015) were aligned with the default settings of *novoalignCS* (V1.06.09, *Novocraft Technologies*). For visualization, replicate BAM files for each stage were merged and normalized to 1 x sequencing depth. Briefly, H3K27ac peaks were called from both replicates of d6 PGCLC with MACS2 (Zhang et al., 2008) using broad settings:

```
macs2 callpeak -t ChIP.bam -c input.bam -f BAM
--broad -g mm -m 5 50 --fix-bimodal --extsize 200
```

Next, all peaks with q-values < 1 x 10^{-3} and within a distance of 1 Kb from both replicates of d6 PGCLC H3K27ac peaks were merged with *bedtools*. The resulting regions were subtracted from blacklisted and promoter regions (-/+ 2 Kb of the *Ensembl* gene annotation, v86) with *bedtools*. PGCLC enhancers were defined as the subset of the distal H3K27ac peaks identified in d6 PGCLC that could be physically linked to PGCLC genes according to capture Hi-C data generated in ESC by (Atlasi et al., 2019). Thereby, distal d6 PGCLC H3K27ac peaks and PGCLC genes were linked if the two anchors of a Capture Hi-C interaction occurred within 1 Kb of a d6 PGCLC H3K27ac peak and 3 Kb of the PGCLC gene TSS.

For the statistical comparisons of epigenetic signals within PGCLC enhancers in different cell types, the average signals of the ChIP-seq, ATAC-seq or genome-wide CpG bisulfite sequencing within -/+ 1 Kb of the PGCLC enhancers in d2 EpiLC and EpiSC was determined using *deeptools-3.3.1 (Ramírez et al., 2016)*. Then, the effect size of paired wilcoxon tests was calculated by dividing the z-statistics by the square roots of the sample sizes using *rstatix* and *boots.ci* for the approximation of the confidence intervals (confidence level: 0.95).

EpiLC/EpiSC enhancers: Briefly, H3K27ac peaks were called in both cell types separately with MACS2 (Zhang et al., 2008) using broad settings (see above) and q-values < 1 x 10^{-3}. The resulting regions were subtracted from blacklisted and promoter regions (-/+ 2 Kb of the Ensembl gene annotation, v86) using *bedtools*.

EpiLC and EpiSC enhancers were obtained from these regions by selecting only H3K27ac peaks located within 500 Kb of the gene sets defined by single-cell RNA-seq for each cell type. The enhancer assignment to the proximal genes was done with *GREAT-4.0.4* (McLean et al., 2010), and only enhancers with a minimum distance of 3.5 Kb to the nearest transcription start site (TSS) were considered.

Motif Analysis (MEME-Suite): The enhancer regions and the PGCLC TSS regions were converted into fasta files and MEME was run with classic mode and default settings. The top motif sequences reported by MEME were searched against known transcription factor motifs from the JASPAR database (Tanaka et al., 2011).

5.2.7. 4C-seq data processing and normalization

Reads were assigned to samples based on the first 10 bases of the read containing unique barcodes. Subsequently, the primer sequence was removed from the read and the remaining sequence starting before the restriction site for *NlaIII* (CATG) was trimmed to 41 bases. These 41 bases were aligned to the mouse reference genome (*mm10*) using *HISAT2* (Kim et al., 2015). From these alignments, RPM (reads per million) normalized bedgraph files were generated for downstream visualization and analysis (Thongjuea et al., 2013).

5.2.8. CpG methylation analysis

The locus-specific bisulfite sequencing data were analysed with BISMA (Rohde et al., 2010).

For the genome-wide CpG methylation data, fastq files were mapped with *Bismark* (Krueger and Andrews, 2011). As the data sets were generated by different library preparation protocols, the preprocessing steps were adjusted accordingly: For the whole-genome bisulfite sequencing data from 2i+LIF ESC (Habibi et al., 2013), d2 EpiLC and EpiSC (Zylicz et al., 2015), the adapter trimming was performed with *Trim Galore* using the default settings; for data sets generated by post bisulfite adaptor tagging (pbat), either 9 (data from d2 EpiLC and PGCLC (Shirane et al.,

2016)) or 6 bp (genome-wide methylation data generated in this study) were removed. The DNA methylation data from E4.5 - E6.5 epiblasts were generated by STEM-seq (Zhang et al., 2018b), and, in this case, adapter sequences were removed with *cutadapt* (Martin, 2011) and *Trim Galore,* respectively. For all the previous samples, reads were mapped with *Bismark-v0.16.1* (Krueger and Andrews, 2011) and *bowtie2-2.2.9* (Langmead and Salzberg, 2012), using the "-pbat" setting for the STEM-seq and pbat samples. For paired-end pbat samples the unmapped reads were remapped as single-end reads to increase the coverage. Then, for each cell type the available datasets were combined to estimate the CpG methylation levels with the *Bismark methylation extractor*. Finally, only CpGs with a coverage of 3 - 100 reads were considered.

5.2.9. CpG methylation heterogeneity (scNMT-seq)

The scNMT-seq method records the transcriptome, methylome (CpG methylation) and chromatin accessibility (GpC methylation) from the same single cell (Clark et al., 2018). Among all the single cells analyzed in E4.5, E5.5 and E6.5 mouse embryos by (Argelaguet et al., 2019), only those assigned to the epiblast by scRNAseq (https://github.com/rargelaguet/scnmt_gastrulation) were considered. Then, epiblast cells were additionally filtered according to their methylome (GSE121690) and only cells with a CpG coverage > 10^5 were further analyzed. This resulted in 258 epiblast cells from E4.5, E5.5 and E6.5 with high quality single cell CpG methylation data.

For the scRNA-seq data, the counts were FPKM-normalized with *edgeR* (Robinson et al., 2010) and the mean expression of all PGCLC genes linked to Group I PGCLC enhancers was determined per cell.

The single cell CpG methylation for each cell was stored in a bedGraph file. From each bedGraph file the genome-wide CpG methylation levels of individual cells were determined as the mean methylation of all covered CpGs in each single cell. For the analysis of mCpG levels and CpG coverage within PGCLC enhancer Group I, the genome-wide bedGraph files were subtracted to obtain bedGraph files with the

PGCLC enhancer regions using *bedtools* from which the mean methylation of all covered CpGs was determined in each single cell.

CpG methylation heterogeneity was estimated with the *PDclust* package (Hui et al., 2018). The number of CpGs covered in each pair of cells resulted in approximately 150 CpGs for each pairwise comparison when PGCLC enhancers were considered. Then, the average of the absolute difference in the methylation values for all the CpGs covered for each pairwise comparison were computed as a dissimilarity matrix.

6. Results

6.1. Transcriptional characterization of the PGCLC differentiation system by scRNA-seq

The *in vitro* PGCLC differentiation system described by (Hayashi et al., 2011) was used for the investigation of germline competence, due to the scarcity and transient nature of PGCs *in vivo.* In this protocol, mouse ESC are differentiated into EpiLC from which PGCLC can be obtained within heterogeneous embryoid bodies (EB). In contrast, Epiblast stem cells (EpiSC), resembling the post-implantation gastrulating epiblast, cannot be efficiently differentiated into PGCLC and, thus, display limited germline competence (Figure 3) (Hayashi et al., 2011; Smith, 2017).

6.1.1. Transcriptional dynamics of PGCLC genes

As described in (Bleckwehl et al., 2021), single-cell RNA sequencing (scRNA-seq, n = 2,782 cells) across multiple stages of PGCLC differentiation was performed to (i) gain insights about the formation of PGCLC *in vitro* and (ii) to define a PGCLC gene set in the heterogeneous cell population found within EBs. A first t-SNE analysis of the resulting single cell transcriptomes showed that cells tend to cluster within their corresponding differentiation stage (Figure 4).

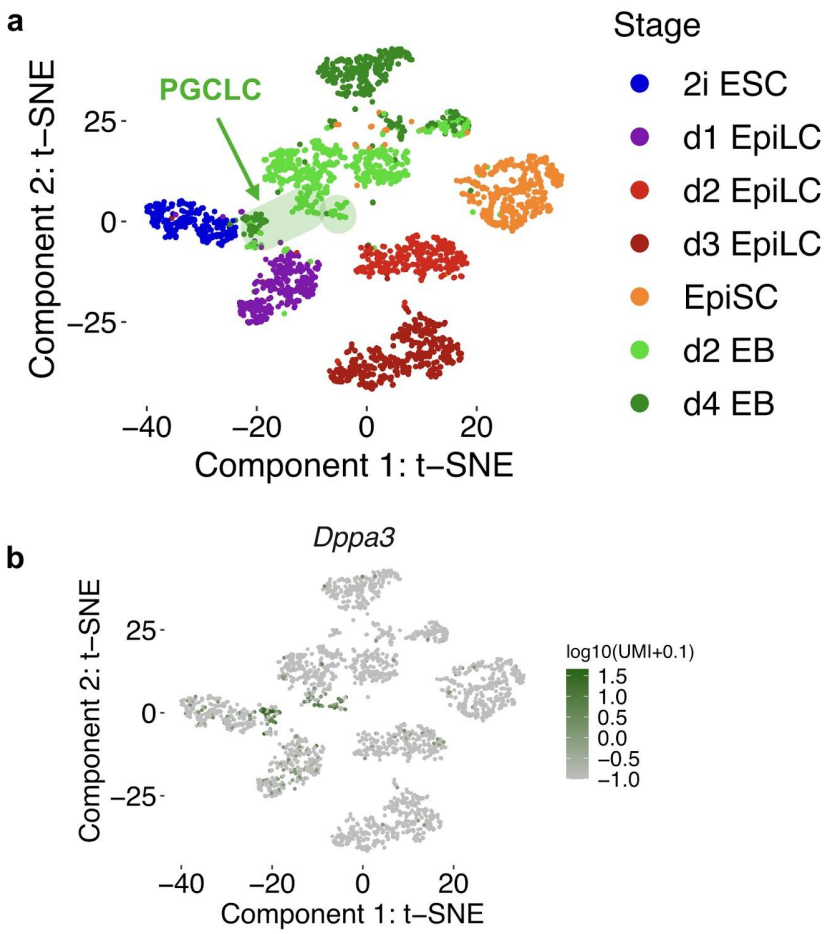


Figure 4: The scRNA-seq data and the identification of PGCLC.

a.) t-SNE plot of the scRNA-seq data generated across the main stages of the in vitro PGCLC differentiation. The PGCLC cluster within the d2 and d4 EBs is highlighted by the arrow. The plot was adapted from Bleckwehl et al. 2021.

b.) Expression of the PGCLC marker Dppa3 in the t-SNE plot of the PGCLC differentiation showing to the left.

However, Day 2 and Day 4 EB showed cellular heterogeneity and formed distinct subclusters. One of these subclusters was identified as PGCLC based on the high expression of major PGC markers (*e.g. Prdm14*, *Prdm1, Tfap2c, Dppa3*), while the additional subpopulations within d4 EB were annotated based on the expression of cell identity markers identified by single-cell transcriptional profiling of E8.25 mouse embryos (Ibarra-Soria et al., 2018). Remarkably, these subclusters were similar to the extraembryonic tissues (*i.e.* extraembryonic ectoderm, extraembryonic

mesoderm and endothelium) that surround PGCs in the proximo-posterior end of the mouse embryo following germline specification *in vivo* (Figure 5).

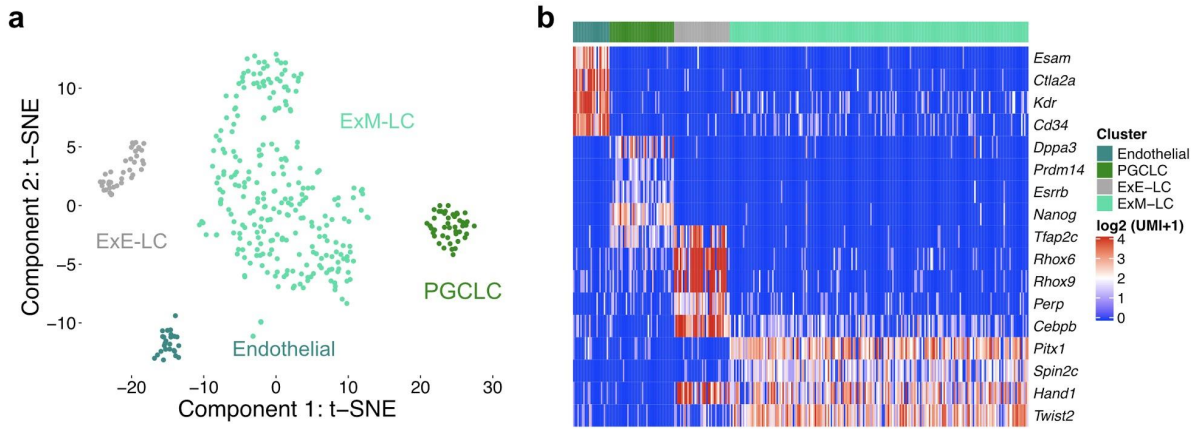


Figure 5: Cluster identity of the d4 EB.
a.) k-means clustering of d4 EB data identified four clusters resembling the transcriptomes of PGCs and their main surrounding tissues in vivo: Extraembryonic Ectoderm like-cells (ExE-LC), Extraembryonic Mesoderm like-cells (ExM-LC) and Endothelial cells.
b.) Expression of selected genes representative of the different cell clusters identified in d4 EBs. The expression values are displayed as UMI (unique molecular identifier) counts in log2 scale. The figures were adapted from Bleckwehl et al. 2021.

Furthermore, differential expression analysis between the PGCLC cluster and the remaining cells of the d2 and d4 EBs (see Methods) led to the identification of a set of 389 PGCLC genes, which included the PGC markers mentioned above as well as major naïve pluripotency regulators (e.g. *Nanog*, *Esrrb*). In agreement with previous reports (Hayashi et al., 2011), several PGCLC genes were highly expressed in ESC, progressively silenced in EpiLC and finally reactivated in PGCLC. Lastly, gene sets specific for the three investigated *in vitro* pluripotent cell types (*i.e*. ESC, d2 EpiLC and EpiSC) were defined by differential gene expression analysis between each cell type and the other two cell types (Figure 6).

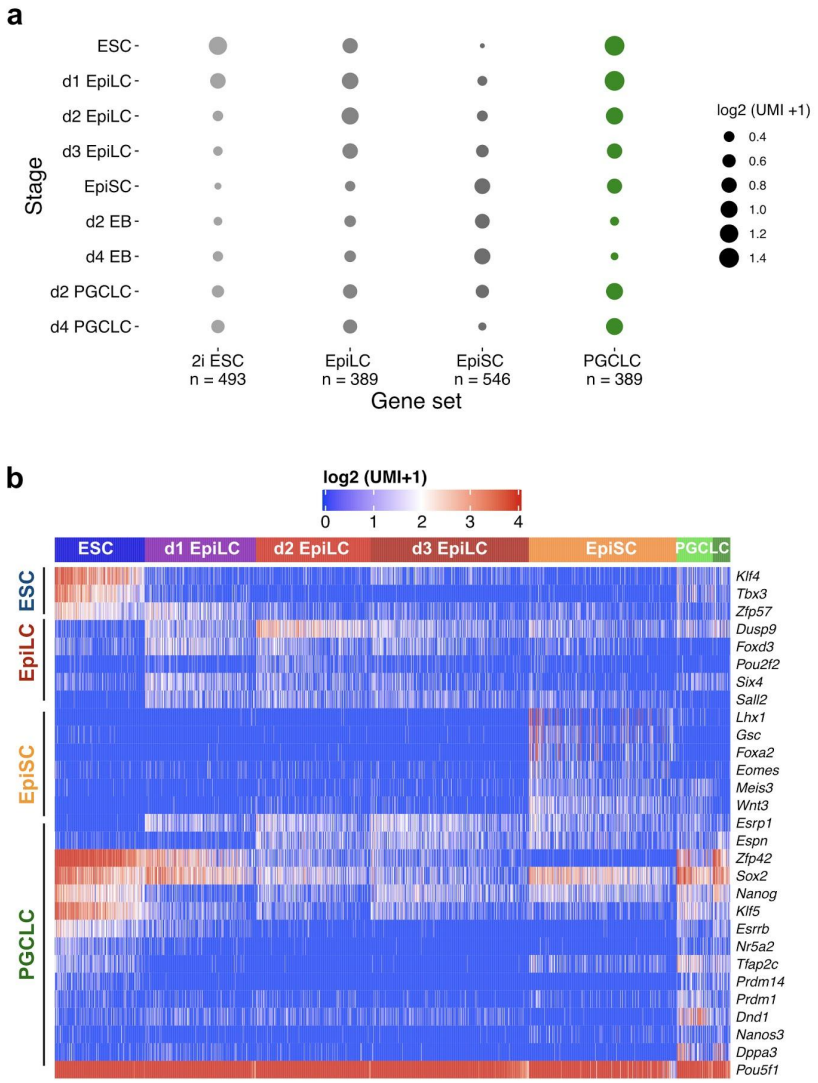


Figure 6: Stage specific expression of the gene sets.

a.) Summary for the different gene sets defined by the single-cell RNA-seq data. The ESC, EpiLC and EpiSC gene sets were determined by considering the genes differentially expressed in each stage compared to the remaining two (i.e. ESC vs. d2 EpiLC and EpiSC). The PGCLC gene set was determined by considering the differentially expressed genes in the d2 and d4 PGCLC cluster compared to the remaining cells of the d2 and d4 EBs. The number of genes per set are indicated below and the size of a dot represents the average expression of all genes within a gene set for the indicated cellular stages.

b.) Exemplary expression of selected genes from each gene set (y-axis labels). Shown are the expression values as UMI counts within individual cells belonging to the indicated cellular stages (colored above the heatmap). Pou5f1, a general pluripotency and germ cell marker, is shown to illustrate the quality of the single-cell RNA-seq data. The figures were adapted from Bleckwehl et al. 2021.

6.1.2. Transcriptional induction of PGCLC

The previously described asynchronous dismantling of the naïve expression programm (Kalkan et al., 2017) as well as the reactivation of most of the naïve genes during PGCLC specification raised the question of whether a small subpopulation in d2 EpiLC might be transcriptionally primed for PGCLC induction. Alternatively, and as indicated in previous work (Kalkan et al., 2017; Mulas et al., 2017), the acquisition of germline competence in d2 EpiLC might entail the complete dismantling of the naïve gene expression program and the subsequent re-activation of some of these naïve genes in PGCLC.

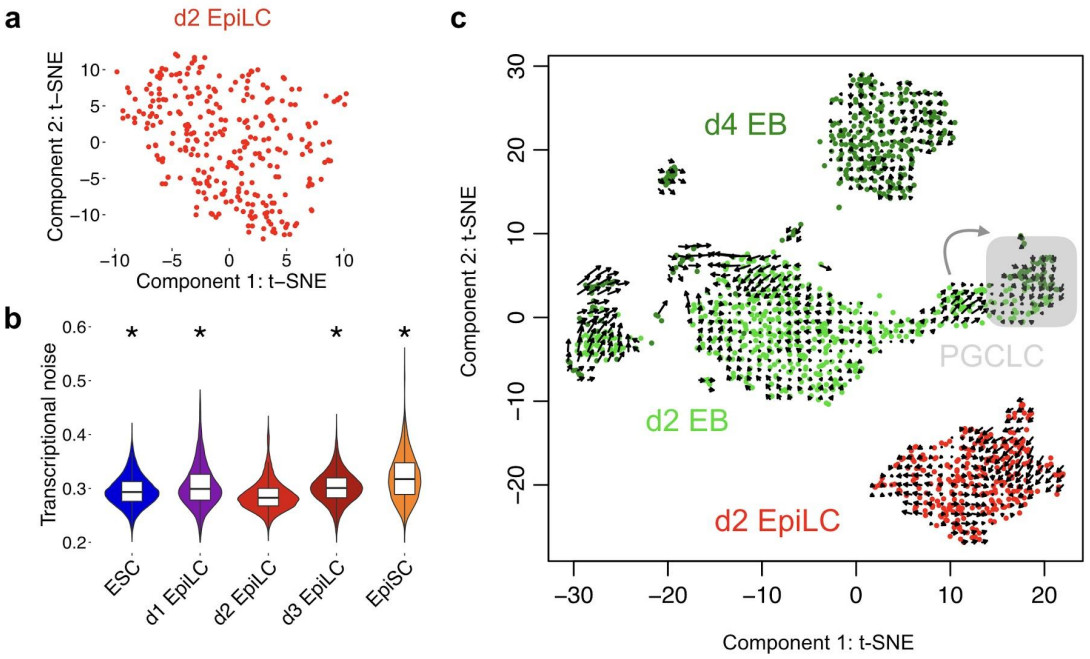


Figure 7: Transcriptional induction of PGCLC.

a.) t-SNE plot of d2 EpiLC alone (n = 289).

b.) Transcriptional noise, defined as cell-to-cell transcript variation for the 500 most variable genes, for ESC, EpiLC and EpiSC. Lower transcriptional noise indicates higher transcriptional similarity between the cells belonging to a particular stage. All cellular stages shown were compared to d2 EpiLC using a two-sided Wilcoxon tests (: p-value < 2.2 10^{-16}).*

c.) RNA velocity analysis of d2 EpiLC and EB, which does show transcriptional priming of PGCLC from cells with a mixed mesodermal/PGCLC programm (gray arrow), but not from germline competent d2 EpiLC. The figures were adapted from Bleckwehl et al. 2021.

Clustering of d2 EpiLC (289 cells) and RNA velocity analysis (La Manno et al., 2018) of the PGCLC differentiation revealed that all d2 EpiLC are transcriptionally similar and neither a distinct subpopulation indicative of a retained naïve pluripotency expression program nor signs of precocious germline induction could be identified. Moreover, the cell-to-cell variability in gene expression levels, defined as transcriptional noise, was significantly lower in d2 EpiLC than in the preceding or subsequent cellular stages (Figure 7). This is in agreement with the transcriptional homogeneity of the E5.5 epiblast *in vivo*, which bears the lowest transcriptional noise during mouse peri-implantation development (Mohammed et al., 2017). Conclusively, the data is in agreement with previous reports (Kalkan et al., 2017; Mulas et al., 2017) and indicates that *in vitro* germline competence entails a transcriptionally homogeneous stage in which the gene expression program shared between naïve pluripotency and PGCLC is silenced.

6.2. Epigenetic features of PGCLC enhancers in EpiLC vs. EpiSC

Given that transcriptional priming in EpiLC was not detectable and that most PGCLC genes are downregulated in EpiLC and EpiSC, but yet only EpiLC display high germline competence, there might be other differences between those two cellular stages that explain why only EpiLC can rapidly induce the PGCLC expression program when exposed to the appropriate signals.

Based on previous observations and considering the importance of enhancers for lineage specification (Respuela et al., 2016; Tischler et al., 2019; Zylicz et al., 2015), one possibility is that the enhancers controlling the expression of PGCLC genes might display epigenetic differences between EpiLC and EpiSC that could explain their distinct germline competence. To test this hypothesis, distal H3K27ac peaks in d6-sorted PGCLC (n=13,105) were identified, using publically available data (Kurimoto et al., 2015). In agreement with previous observations (Respuela et al., 2016), a large fraction of the d6 PGCLC H3K27ac peaks were initially active in ESC, lost H3K27ac in EpiLC and became progressively reactivated in d2 and d6 PGCLC

(Bleckwehl et al., 2021). Since most of the d6 PGCLC H3K27ac peaks were also active in ESC, public available capture Hi-C data generated in ESC (Atlasi et al., 2019) was used to systematically associate these distal peaks to their putative target genes. PGCLC enhancers were defined as those distal d6 PGCLC H3K27ac peaks that could be physically linked to the PGCLC gene set defined by single-cell RNA-seq (Figure 8). This resulted in 415 PGCLC enhancers linked to 216 of the 389 PGCLC genes. Furthermore, to compare epigenetic changes between different enhancer groups, EpiLC and EpiSC enhancers were defined using similar criteria.

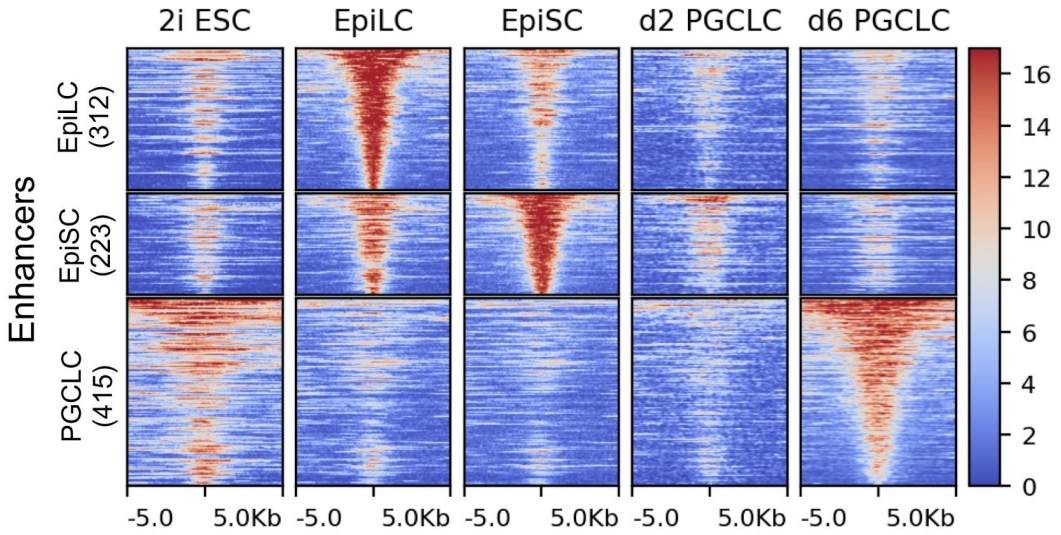
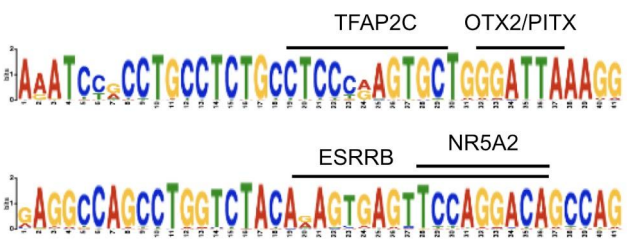


Figure 8: Heatmaps showing the H3K27ac dynamics for EpiLC, EpiSC and PGCLC enhancers during PGCLC differentiation. For EpiLC: 146 out of 389, EpiSC: 164 out of 546, and PGCLC: 216 out of 389 genes from the gene sets defined in Fig. 1 were linked to their enhancers (see Methods for the criteria used to define the different enhancer sets). Enhancers within each set were ordered according to the H3K27ac levels in the corresponding cell type. The heatmap was adapted from Bleckwehl et al. 2021.

6.2.1. Motif Analysis

Motif analysis of the different enhancer groups revealed a strong enrichment of naïve pluripotency TF (ESRRB+NR5A2) motifs in the TSS and enhancers associated with PGCLC genes. In addition, the PGCLC TSS were enriched for the KLF9 motif, while the motifs of TFAP2C and OTX2 were identified next to each other within the PGCLC enhancers (Figure 9). The same TFAP2C/OTX2 motif was also found when

analyzing EpiLC enhancers, consistent with the major role of OTX2 during EpiLC differentiation (Buecker et al., 2014). EpiSC enhancers were enriched for the ZFP281 motif, which together with ZIC2, have been shown to induce the transition to EpiSC (Mayer et al., 2020).



Regulatory element	Motifs
EpiLC enhancer	PITX/OTX2+TFAP2C
EpiSC enhancer	ZFP281, ZFP263
PGCLC enhancer	PITX/OTX2+TFAP2C, NR5A2+ESRRB
PGCLC TSS	KLF9, NR5A2+ESRRB

Figure 9: Motif analysis of the different enhancers and PGCLC TSS. The most abundant motif of PGCLC enhancers (TFAP2C/OTX2 and ESRRB/NR5A2) are shown to the left and other motifs for each enhancer and the PGCLC TSS are listed in the table on the right.

6.2.2. Active chromatin features

Next, an extensive set of ChIP-seq and ATAC-seq in ESC, EpiLC and EpiSC was generated to explore the chromatin differences between EpiLC and EpiSC. ATAC-seq experiments were performed in two replicates to monitor the accessibility of the enhancer sets. PGCLC enhancers were found to be highly accessible in ESC, but lost their accessibility together with the H3K27ac ChIP-seq signal in EpiLC and EpiSC. EpiLC and EpiSC enhancers showed high accessibility in their active stage and to some extent also in the preceding stages. In contrast, the PGCLC TSS displayed high accessibility in all three pluripotent cell types (Figure 10).

H3K4 methylation is widely associated with active transcription, with H3K4me1/2 being more prevalent at enhancers and H3K4me3 at promoters (Heintzman et al., 2007). The ChIP-seq experiments revealed minor changes for H3K4me3 at both enhancers and PGCLC TSS, while H3K4me1/2 showed a much more dynamic behaviour at the different enhancer groups. Notably, among all the investigated active marks, H3K4me1 showed the most obvious differences between EpiLC and EpiSC within the PGCLC enhancers: H3K4me1 was partially retained in EpiLC, but not in EpiSC, indicating an incomplete decommissioning of the PGCLC enhancers in

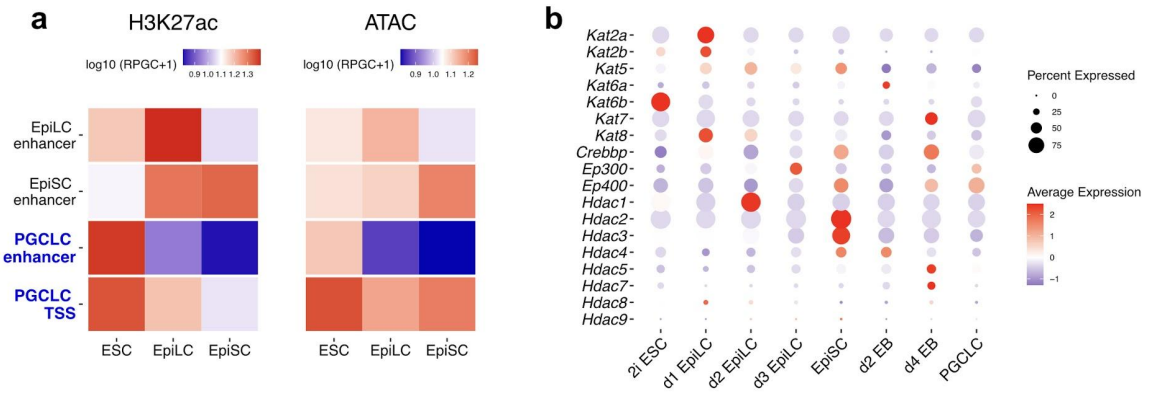


Figure 10: Summary plot of H3K27ac and chromatin accessibility for the enhancer sets.

a.) Average levels of H3K27ac ChIP-seq and ATAC-seq signal in 2i ESC, d2 EpiLC and EpiSC for the EpiLC, EpiSC and PGCLC enhancers as well as the TSS of the PGCLC genes. Quantifications were performed by measuring the average signals of each epigenetic mark within -/+ 1kb of the enhancers or TSS. RPGC: Reads per genomic content. The summary plots were adapted from Bleckwehl et al. 2021.

b.) Expression of the acetyltransferase (Kat2a - Ep400) and classical nucleosomal deacetylases (Hdac1 - Hdac9). The color indicates the scaled expression in each stage and the size represents the number of cells expressing each of the indicated genes.

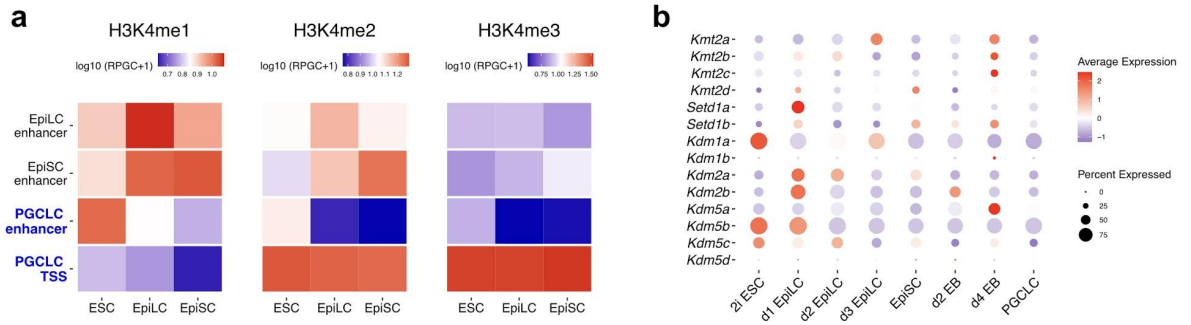


Figure 11: Summary plot for H3K4 methylation and H3K4 (de-) methyltransferases.

a.) Average levels of H3K4me1/2/3 ChIP-seq signals in 2i ESC, d2 EpiLC and EpiSC for the EpiLC, EpiSC and PGCLC enhancers as well as the TSS of the PGCLC genes. The quantifications were performed by measuring the average signals of each epigenetic mark within -/+ 1kb of the enhancers or TSS. The summary plots were adapted from Bleckwehl et al. 2021.

b.) Expression of the H3K4 methyltransferase (Kmt2a - Setd1b) and the demethylases (Kdm1a - Kdm5d). The color indicates the scaled expression in each stage and the size represents the number of cells expressing each of the indicated genes.

EpiLC (Figure 11). On the other hand, the transcription of nearly all H3K4 methyltransferases and demethylases did not change dramatically during the PGCLC differentiation. In summary, the analysis of active epigenetic modifications showed that H3K4me1 persistence was the most pronounced difference at PGCLC enhancers between EpiLC and EpiSC.

6.2.3. Repressive chromatin features

H3K9me2/3 are associated with the silencing of enhancers and with heterochromatinization, while H3K27me2/3, together with H3K4me1, are characteristic of poised enhancers (Cruz-Molina et al., 2017). ChIP-seq experiments of these modifications were performed in ESC, EpiLC and EpiSC. In comparison to ESC, H3K9me2 increased specifically at PGCLC enhancers in EpiLC, while in EpiSC this histone mark increased globally. H3K9me3 levels were generally low in ESC and EpiLC, but interestingly, they were quite high at EpiSC enhancers before their activation in EpiSC. Besides, H3K9me3 increased concurrently in EpiSC at EpiLC enhancers, PGCLC enhancers and PGCLC TSS (Figure 12). Consistent with the low levels of H3K9me2 in PGCLC (Kurimoto et al., 2015; Seki et al., 2005), most of the H3K9 methyltransferases were lowly expressed in PGCLC compared to other cells of the EB, while the expression of H3K9 demethylases (*e.g. Kdm3a/b*) were higher in PGCLC.

H3K27me2 and H3K27me3 showed both similar trends, displaying low levels within PGCLC enhancers in all stages, but increasingly higher levels at the PGCLC TSS upon exit from naïve pluripotency. The highest levels of H3K27me3 for EpiLC and EpiSC enhancers were measured in ESC, indicating, at least partially, a polycomb-associated poising in ESC. Besides, the expression of H3K27 methyltransferases and demethylases remain widely constant in all differentiation stages (Figure 13).

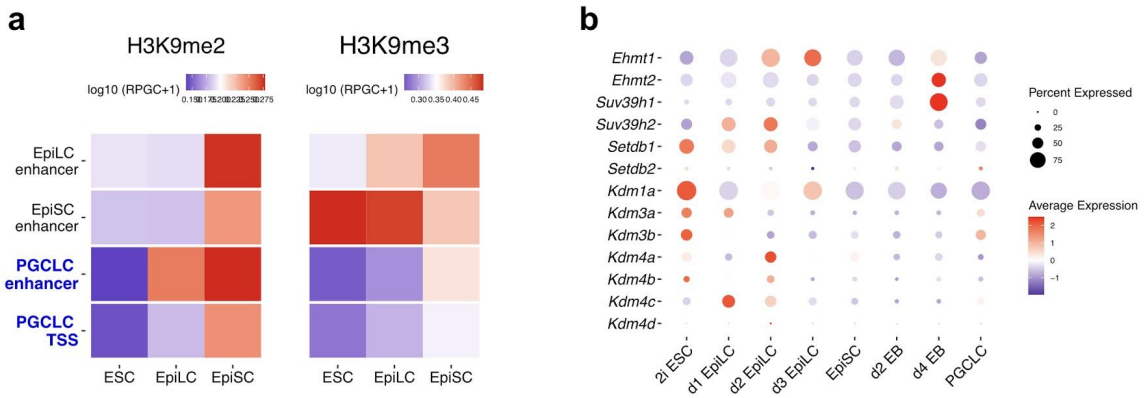


Figure 12: Summary plot for the H3K9 methylation and H3K9 (de-) methyltransferases.

a.) Average levels of H3K9me2/3 ChIP-seq signals in 2i ESC, d2 EpiLC and EpiSC for the EpiLC, EpiSC and PGCLC enhancers as well as the TSS of the PGCLC genes. The quantifications were performed by measuring the average signals of each epigenetic mark within -/+ 1kb of the enhancers or TSS. The summary plots were adapted from Bleckwehl et al. 2021.

b.) Expression of the H3K9 methyltransferase (Ehmt1 - Setdb2) and the demethylases (Kdm1a - Kdm5d). The color indicates the scaled expression in each stage and the size represents the number of cells expressing the gene of the methyltranses or demethylases.

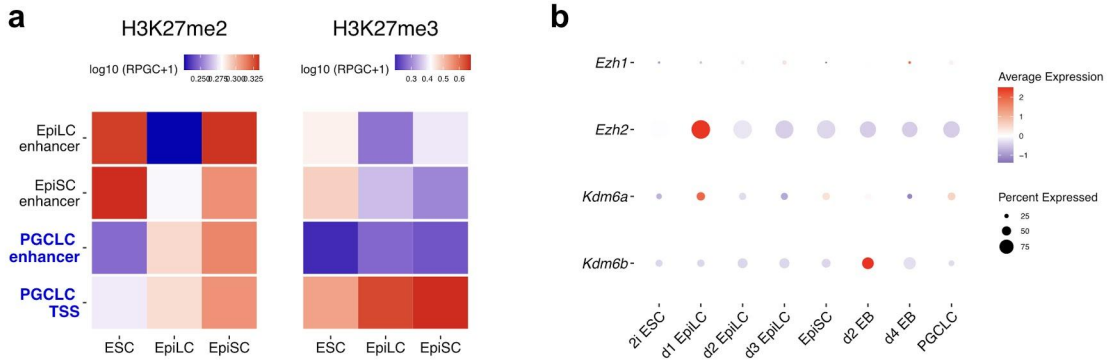


Figure 13: Summary plot for the H3K27 methylation.

a.) Average levels of H3K27me2/3 ChIP-seq signals in 2i ESC, d2 EpiLC and EpiSC for the EpiLC, EpiSC and PGCLC enhancers as well as the TSS of the PGCLC genes. The quantifications were performed by measuring the average signals of each epigenetic mark within -/+ 1kb of the enhancers or TSS. The summary plots were adapted from Bleckwehl et al. 2021.

b.) Expression of the H3K27 methyltransferase (Ezh1/2) and the demethylases (Kdm6a/b). The color indicates the scaled expression in each stage and the size represents the number of cells expressing the gene of the methyltranses or demethylases.

Given that CpG methylation is another repressive epigenetic modification that frequently co-occurs with H3K9me2/3 and is antagonized by H3K4 methylation (Hughes et al., 2020; Ooi et al., 2007), publicly available whole-genome bisulfite sequencing data generated across pluripotent states were analyzed (Zylicz et al., 2015). In agreement with their active state, CpG methylation at PGCLC enhancers was low in ESC. Then, as these enhancers were decommissioned, they gained CpG methylation in EpiLC and became fully methylated in EpiSC (Figure 14). Increased CpG methylation was also observed at the EpiLC enhancers in EpiSC, while the EpiSC enhancers and PGCLC TSS displayed low CpG methylation levels in all stages. CpA methylation increases globally during the establishment of formative pluripotency due to the elevated levels of *de novo* methylation (Zhang et al., 2018b). Accordingly, CpA methylation did not change specifically within any of the considered enhancer groups, but rather showed globally higher levels in EpiLC. Consequently, the expression of the DNA methyltransferase (*Dnmt3a/b/l*) increased sharply in d2 EpiLC.

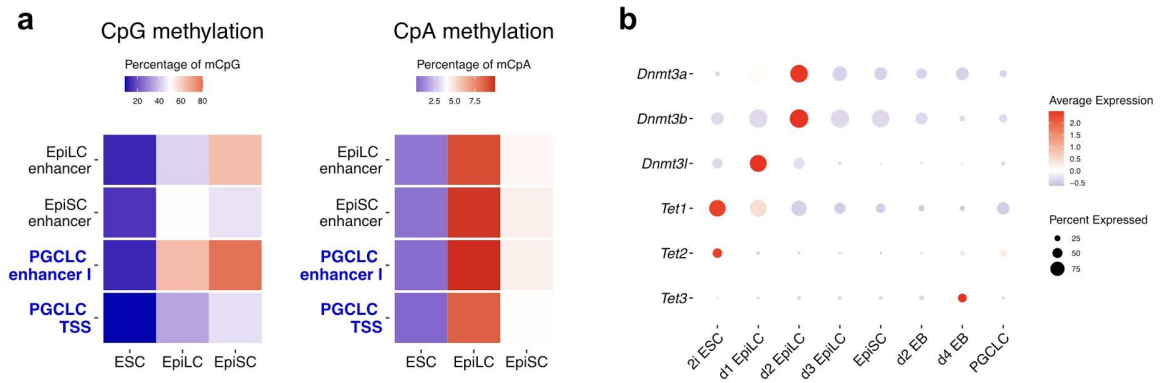


Figure 14: Summary plot for the cytosine modification.

a.) Average percentage of CpG and CpA in 2i ESC, d2 EpiLC and EpiSC for the EpiLC, EpiSC and PGCLC enhancers as well as the TSS of the PGCLC genes. The quantifications were performed by measuring the average percentage of each epigenetic mark within -/+ 1kb of the enhancers or TSS.

b.) Expression of the DNA methyltransferase (Dnmt3a-l) and the demethylases (Tet1-3). The color indicates the scaled expression in each stage and the size represents the number of cells expressing the gene of the methyltranses or demethylases.

6.2.4. H3K4me1 and CpG methylation antagonism within PGCLC enhancers

Overall, the most pronounced epigenetic changes within PGCLC enhancers were observed for H3K4me1, which was higher in EpiLC, and H3K9me3 and mCpG, which were increased in EpiSC (Bleckwehl et al., 2021). The antagonism between these epigenetic features has been reported before (Hughes et al., 2020; Ooi et al., 2007). Here, with focus on the well-known enhancer mark H3K4me1 (Heintzman et al., 2007), its putative function as a regulator of *in vitro* germline competence is explored.

Firstly, two additional H3K4me1 ChIP-seq replicates (which included another mouse ESC strain, *i.e.* R1) were generated in EpiLC and EpiSC to further validate the higher H3K4me1 levels in EpiLC and to identify PGCLC enhancers that show consistent differences between EpiLC and EpiSC (Figure 15). This revealed that 71% of the PGCLC enhancers showed higher H3K4me1 levels in EpiLC than in EpiSC in at least two replicates (71% Group I; 29% Group II). The Group I PGCLC enhancers also retained more H3K27ac in EpiLC, while the differences in H3K4me2 and chromatin accessibility, as measured by ATAC-seq, were rather moderate. In contrast, the Group I enhancers displayed lower mCpG and H3K9me3 levels in EpiLC than in EpiSC, while these differences were considerably less obvious for the Group II enhancers (Figure 16).

Furthermore, H3K4me1 and mCpG were inversely correlated across Group I PGCLC enhancers, particularly in EpiLC (Figure 17). This negative correlation was also investigated by performing H3K4me1 ChIP in EpiLC followed by bisulfite sequencing of two representative Group I PGCLC enhancers (*i.e. Esrrb* and *Lrrc31/Lrrc34* enhancers). In comparison to the input genomic DNA, the H3K4me1-enriched DNA showed lower mCpG levels at the two analyzed enhancers (Figure 17).

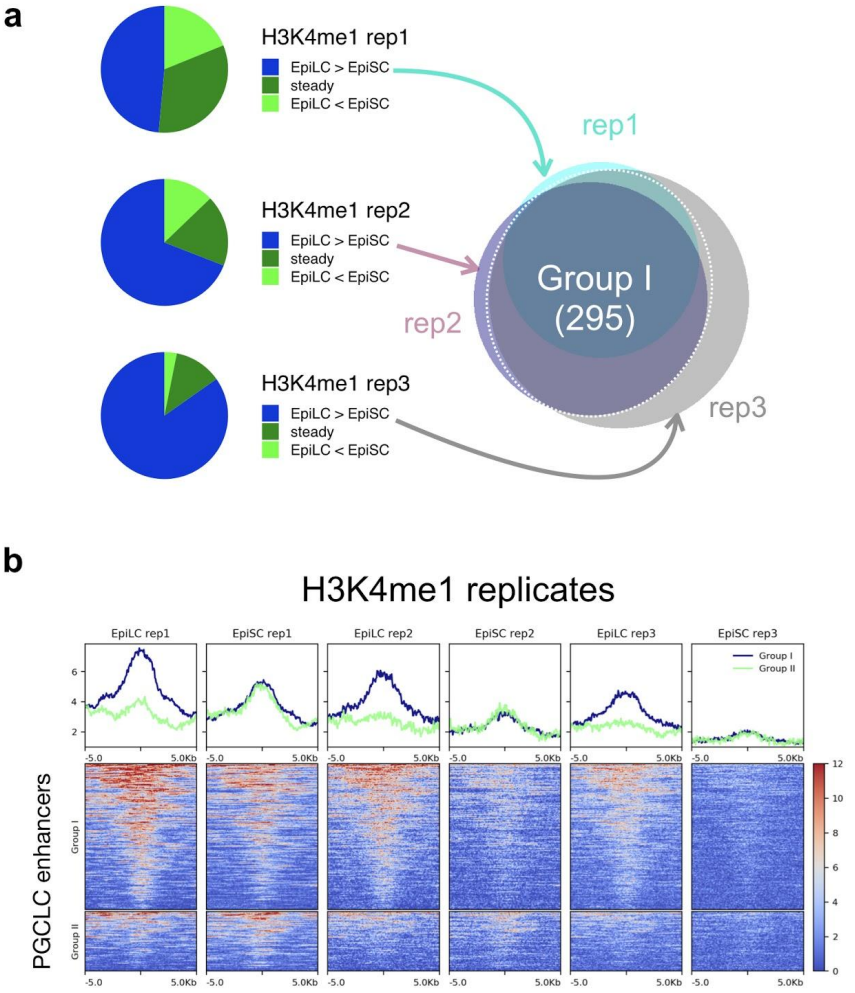


Figure 15: Reproducible loss of H3K4me1 in EpiSC.

a.) Three H3K4me1 ChIP-seq replicates were generated in d2 EpiLC and EpiSC, respectively. For each replicate, the average H3K4me1 signal within -/+ 1kb of each PGCLC enhancer was calculated and compared between EpiLC and EpiSC. Those PGCLC enhancers showing an EpiLC/EpiSC ratio higher than 1.2-fold were assigned to the 'EpiLC > EpiSC' category for each replicate. PGCLC enhancers assigned to the 'EpiLC > EpiSC' category in at least two replicates were defined as Group I (Venn diagram). All other enhancers, including those from the 'steady' (EpiLC/EpiSC H3K4me1 ratio: 0.8>FC>1.2) and 'EpiLC < EpiSC' categories, were assigned to Group II. The figure was adapted from Bleckwehl et al. 2021.

b.) Average profile and heatmap plots showing H3K4me1 levels for the Group I and Group II PGCLC enhancers for all replicates d2 EpiLC and EpiSC.

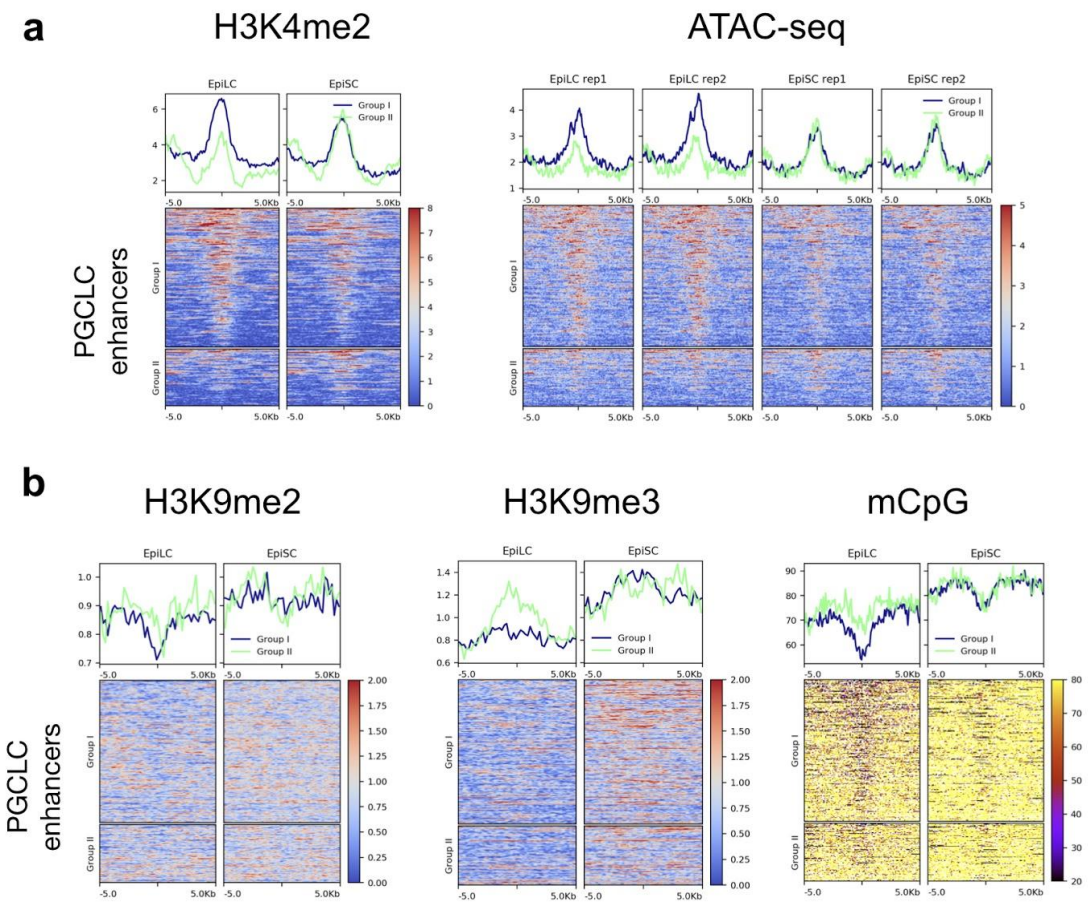


Figure 16: Epigenetic features of Group I and Group II PGCLC enhancers. Average profile and heatmap plots showing ATAC-seq (a), H3K4me2 (b), H3K9me2/3 levels (c) and the percentage of CpG methylation (d) for the Group I and Group II PGCLC enhancers in d2 EpiLC and EpiSC. The figures were adapted from Bleckwehl et al. 2021.

In summary, a subset of the PGCLC enhancers shows partial retention of H3K4me1 and lower levels of heterochromatic marks (i.e. mCpG, H3K9me2/3) in EpiLC compared to EpiSC. These enhancers tend to be initially active in ESC, but lose H3K27ac and TF/coactivator binding (indirectly measured by ATAC-seq) already in EpiLC. Therefore, the data suggest that a significant fraction of PGCLC enhancers are not fully decommissioned in EpiLC.

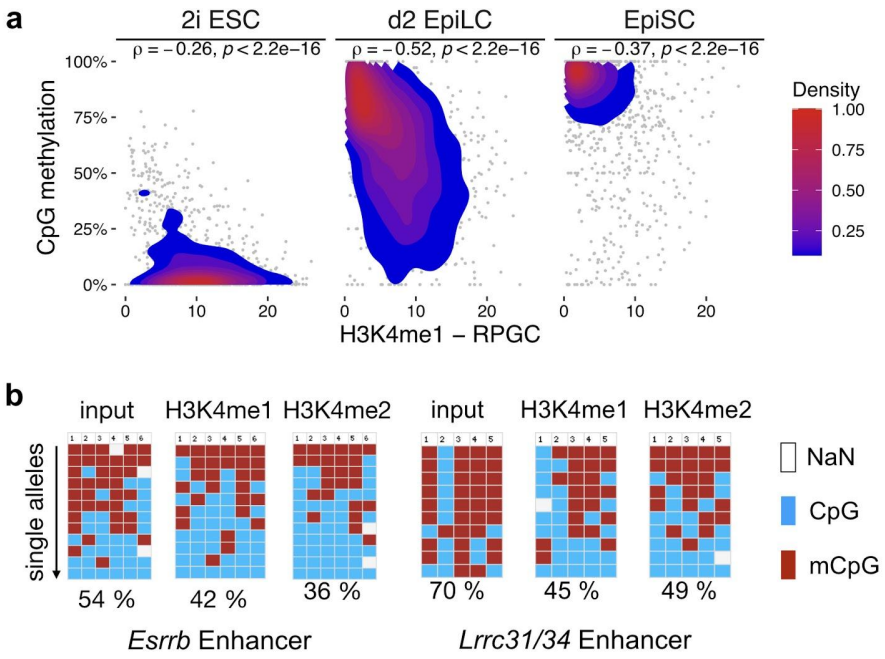


Figure 17: Antagonism between H3K4me1/2 and CpG methylation within PGCLC enhancers.

a.) Correlation plots between CpG methylation and H3K4me1 at Group I PGCLC enhancers in 2i ESC, d2 EpiLC and EpiSC. For each enhancer the average percentage of CpG methylation or level of H3K4me1 was calculated for two 500 bp bins up- and downstream of the enhancer. All four bins of one enhancer are plotted and the spearman correlation was determined. Shown are the measurements of the first H3K4me1 ChIP-seq replicate.

b.) The CpG methylation status of the PGCLC enhancers associated with Esrrb and Lrrc31/34 was investigated by bisulfite sequencing using as templates ChIP input DNA, H3K4me1 ChIP DNA and H3K4me2 ChIP DNA generated in d2 EpiLC. The columns of the plots correspond to individual CpG dinucleotides located within each enhancer. Unmethylated CpGs are shown in blue, methylated CpGs in red and CpGs that were not sequenced are shown in gray. At least 10 alleles were analyzed for each template DNA (rows). The figures were adapted from Bleckwehl et al. 2021.

6.3. PGCLC enhancer deletions and PGCLC specification

The functional relevance of the identified PGCLC enhancers (Figure 8) was investigated by generating CRISPR/Cas9-mediated deletions of selected enhancers and by evaluating the topological dynamics of PGCLC enhancers during pluripotency transitions.

6.3.1. PGCLC enhancers associated with *Esrrb*, *Klf5* and *Lrrc31/34*

Three representative PGCLC enhancers that were linked to *Esrrb*, *Klf5* and *Lrrc31/Lrrc34*, respectively, were chosen. These three enhancers were initially active (H3K27ac) in ESC, got partially decommissioned (lost H3K27ac and H3K4me2, but persistence of H3K4me1) in d2 EpiLC, became reactivated in PGCLC (H3K27ac gain) and were marked by repressive epigenetic features (H3K9me3 and CpG methylation) in EpiSC (Figure 18).

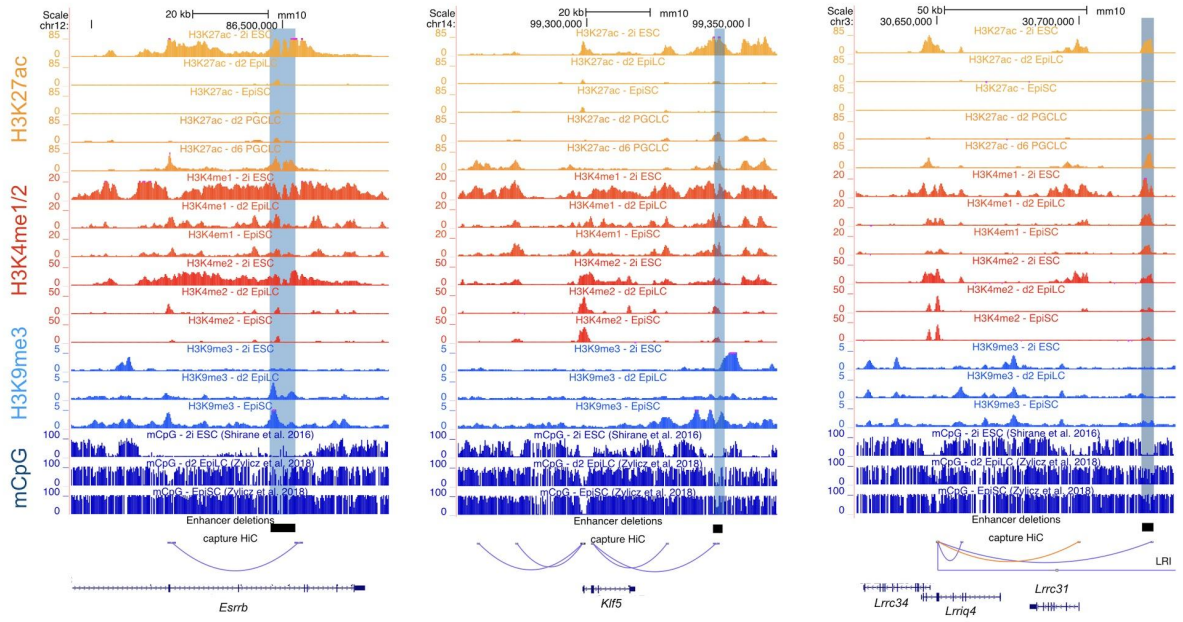


Figure 18: Epigenomic profiles of selected enhancers during PGCLC differentiation. Selected PGCLC enhancers associated with Esrrb (left), Klf5 (middle) and Lrrc31/34 (right) are highlighted in blue. The PGCLC enhancers are physically linked to the transcription start site (TSS) of the indicated genes in ESC according to capture Hi-C data from Atlasi et al. 2019. The Lrrc31/Lrrc34 enhancer is directly linked to the Lrrc34 TSS, which itself interacts with Lrrc31 (orange line). The black rectangles above the interactions denote the enhancer deletions generated in mESC with CRISPR/Cas9 technology. The H3K27ac ChIP-seq data from PGCLC data was obtained from (Kurimoto et al., 2015).

Each enhancer was individually deleted with CRISPR/Cas9 in ESC containing a DPPA3-GFP reporter that facilitates subsequent PGCLC quantification. Two clonal lines for each individual enhancer deletion were generated and analysed for their PGCLC differentiation capacity. The deletion of the enhancer associated with *Esrrb*

had only minor effects on PGCLC specification and showed a similar fraction of DPPA3+ PGCLC within d4 EB compared to the wild type control. On the other hand, the enhancer deletions of *Lrrc31/34* and *Klf5* significantly diminished PGCLC specification (Figure 19).

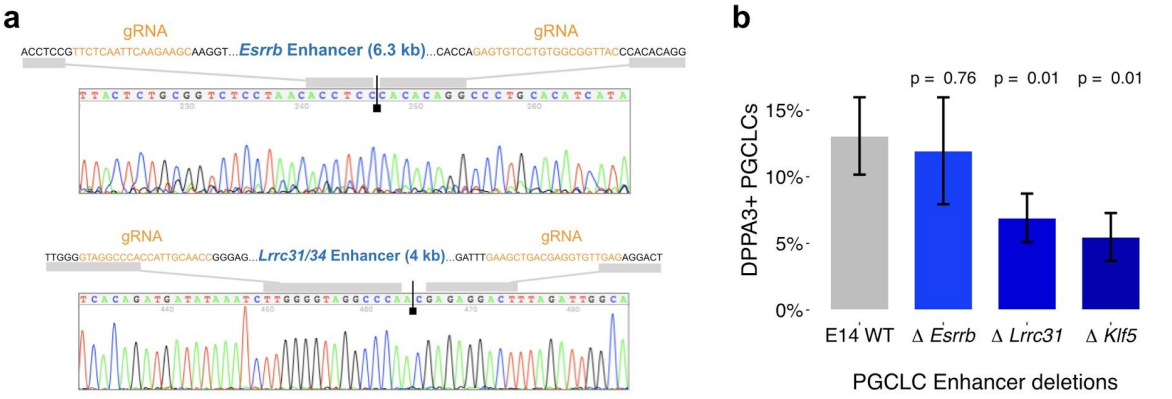


Figure 19: PGCLC enhancer deletions and effects on PGCLC specification.
a.) PGCLC enhancer deletions generated in ESC were confirmed by PCR genotyping followed by Sanger sequencing as shown for one clonal line of the Esrrb and the Lrrc31/34 PGCLC enhancer deletions, respectively. The reference sequence for each enhancer together with the corresponding CRISPR/Cas9 gRNAs (highlighted in orange) are shown above the chromatograms. The chromatograms were adapted from Bleckwehl et al. 2021.
b.) WT ESC and ESC lines with the indicated PGCLC enhancer deletions were differentiated into PGCLC. PGCLC were quantified as the percentage of DPPA3+ cells within d4 EB. Each PGCLC quantification was performed in biological duplicates and two different clonal lines were used for each enhancer deletion (n=2x2). The percentages of PGCLC obtained when differentiating the ESC with the enhancer deletions were compared to those obtained with WT ESC using two-sided wilcoxon tests.

Regardless of the PGCLC quantification, all three enhancer deletions significantly reduced the expression of the corresponding target genes in d4 EB, indicating their regulatory relevance during PGCLC induction (Figure 20). However, the deletions of the enhancers linked to *Klf5* and *Lrrc31/34* reduced the expression of their target genes already in ESC, which could compromise ESC differentiation and indirectly affect the expression of the genes in PGCLC. One the other hand, these results suggest that *Lrrc31/34* and *Klf5* might represent PGCLC regulators shared between naïve and PGCLC, similarly to other naïve/PGCLC TFs (Bleckwehl and Rada-Iglesias, 2019).

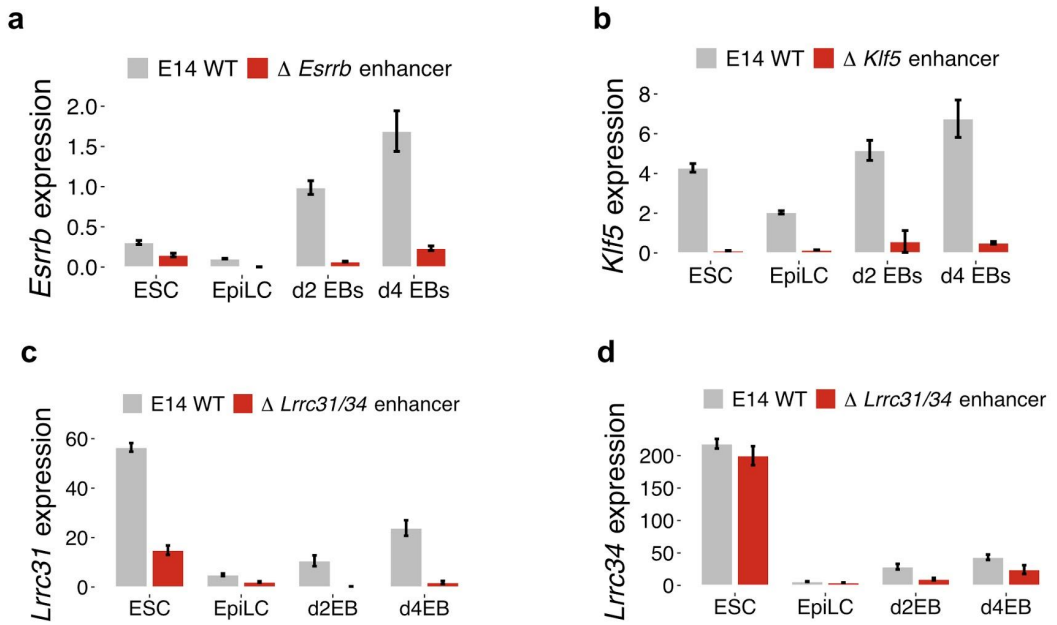


Figure 20: Gene expression measurements of the PGCLC enhancer deletions.
The expression of Esrrb (a), Klf5 (b), Lrrc31(c) and Lrrc34 (d) were measured by RT-qPCR in WT ESC and ESC with the associated enhancer deletion during the differentiation into d4 EB in WT. The expression values were normalized to two housekeeping genes (Eef1a1 and Hprt). Error bars represent standard deviations from 6 measurements (two clonal lines x three technical replicates). The figures were adapted from Bleckwehl et al. 2021.

6.3.2. The complex enhancer regulatory landscape of *Prdm14*

Since many PGCLC genes are associated with multiple and potentially redundant enhancers (52% of PGCLC genes linked to >1 enhancer), three different enhancers (*i.e.* E1, E2 and E3) previously described as components of a *Prdm14* super-enhancer (Hnisz et al., 2015) were individually deleted. In agreement with this previous report, the E2 deletion strongly reduced *Prdm14* expression in ESC, while the deletion of E3 and E1 resulted in considerably smaller or no effects, respectively (Figure 21). Upon PGCLC differentiation, the regulatory importance of each enhancer changed considerably and E1 strongly contributed to *Prdm14* induction, especially during early PGCLC specification (*i.e.* d2 EB). Together with the epigenetic profiles observed within the *Prdm14* locus, these results suggest that, rather than components of an ESC super-enhancer, the E1-E3 elements differentially contribute to *Prdm14* expression in either ESC (*i.e.* E2) or PGCLC (*i.e.* E1).

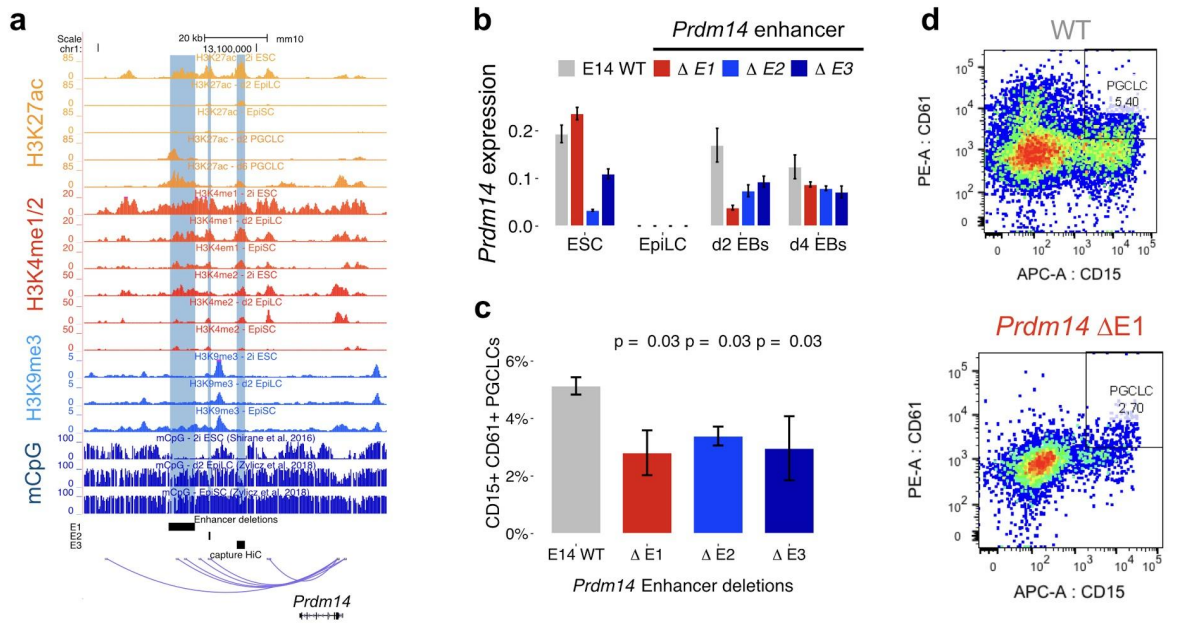


Figure 21: Functional dissection of the Prdm14 super-enhancer during the PGCLC specification.

a.) Genome-browser view of epigenetic features during PGCLC differentiation for three PGCLC enhancers (E1-E3) found within the Prdm14 locus. The PGCLC enhancers that are physically linked to the Prdm14 promoter in ESC according to capture Hi-C data from Atlasi et al. 2019 are shown in blue. The deleted regions are denoted by the black squares shown above the capture Hi-C track.

b.) Prdm14 expression levels were measured by RT-qPCR during the differentiation into d4 EB. The expression values were normalized to two housekeeping genes (Eef1a1 and Hprt). Error bars represent standard deviations from 6 measurements (two clonal lines x three technical replicates).

c.) WT ESC and ESC lines with the indicated Prdm14 enhancer deletions were differentiated into PGCLC. PGCLC were measured as $CD15^{+}CD61^{+}$ cells within d4 EB. Each PGCLC quantification was performed in biological duplicates and two different clonal lines were used for each enhancer deletion (n=2x2). The percentages of PGCLC obtained when differentiating the ESC with the Prdm14 enhancer deletions were compared to those obtained with WT ESC using Wilcoxon tests.

d.) Representative examples of the PGCLC quantifications ($CD15^{+}$ $CD61^{+}$ cells) performed by FACS after four days of PGCLC differentiation using WT ESC (E14 WT) or ESC with the E1 enhancer deletion of Prdm14. The figures were adapted from Bleckwehl et al. 2021.

Furthermore, in agreement with the role of *Prdm14* as a PGC master regulator (Yamaji et al., 2008), the individual E1-E3 deletions significantly impaired PGCLC differentiation (Figure 21). Since the E2 and E3 deletions already reduced the expression of *Prdm14* in ESC, their effects on PGCLC specification could arise secondarily due to compromised naïve pluripotency (Ma et al., 2011; Yamaji et al., 2013). In contrast, the E1 enhancer directly contributes to PGCLC specification, as its deletion affected *Prdm14* expression during PGCLC differentiation, but not in ESC, thus demonstrating its relevance for *Prdm14* induction in PGCLC.

6.3.3. Topological analysis of PGCLC enhancers during differentiation

The results of the PGCLC enhancer deletions showed that some PGCLC enhancers control the expression of their target genes in ESC already, indicating that these PGCLC enhancers might be functionally shared between naïve pluripotency and PGCLC (Respuela et al., 2016). In agreement with this, many PGCLC enhancers physically interact with their target genes in ESC (Figure 8). However, it was unknown how such enhancer-gene contacts might change upon exit from naïve pluripotency. Hence, circular chromatin conformation capture followed by sequencing (4C-seq) experiments were performed and publically available Hi-C data (Miura et al., 2019) were analyzed to elucidate whether there are any topological differences between ESC, EpiLC and EpiSC.

The 4C-seq results showed that especially the *Prmd14* E2 enhancer, to some extent the proximal region of the *Prdm14* E1 enhancer and weakly the *Esrrb* enhancer tended to interact with their target gene in ESC and d2 EpiLC, but not in EpiSC. Besides, the *Prdm14* TSS showed novel interactions in d2 EpiLC within *Ncoa2* introns that have not been identified by capture Hi-C or 4C-seq in ESC. Because these regions gained H3K27ac specifically in PGCLC, they might be additional PGCLC enhancers, which are more relevant in PGCLC than in ESC (Figure 22). Furthermore, the interactions of the *Tfap2c* locus were comparable in all stages. Notably and in contrast to *Esrrb* and *Prdm14*, which are only expressed in ESC, *Tfap2c* is heterogeneously expressed in both ESC and EpiSC (Figure 6).

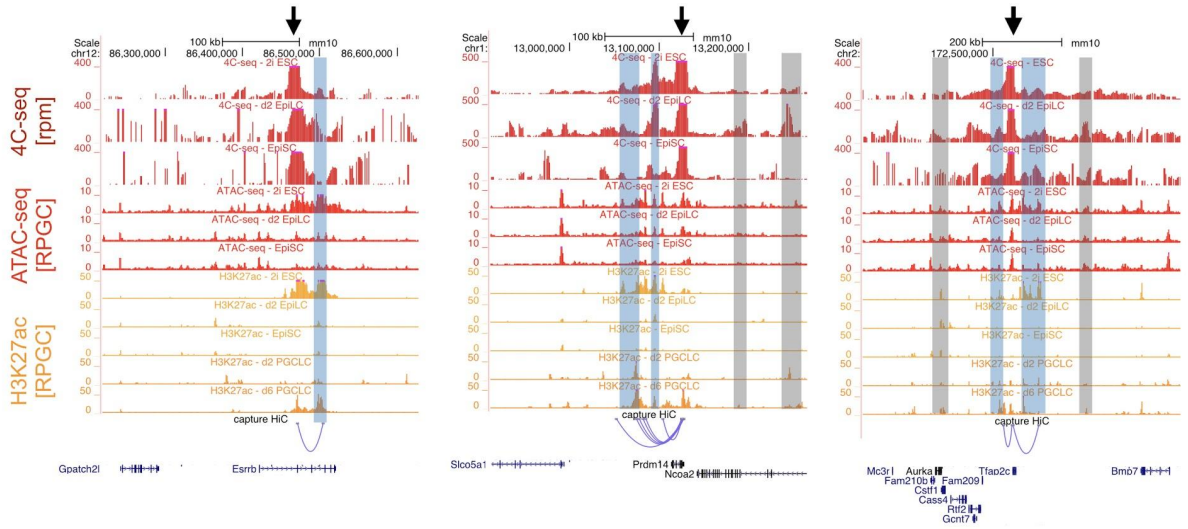


Figure 22: Analysis of the 3D organization of major PGCLC genes by 4C-seq. 4C-seq experiments were performed in ESC and upon differentiation into d2 EpiLC and EpiSC. The TSS of the PGCLC genes Esrrb (left), Prdm14 (middle) and Tfap2c (right) were used as a viewpoint and are highlighted by an arrow above each plot. PGCLC enhancers defined by capture Hi-C are highlighted in blue and other putative PGCLC enhancers are shown in gray. The ATAC-seq and H3K27ac tracks of ESC, EpiLC, EpiSC and PGCLC are shown below the 4C-seq tracks. The 4C-seq experiments for the Prdm14 locus were performed by Michaela Bartusel.

Overall, most of the 4C interactions were rather weak, which might reflect heterogeneity and/or highly transient promoter-enhancer interactions. Next, publically available HiC data were analyzed to extend the previous observations to all PGCLC enhancer-gene pairs (Figure 23). These analyses showed negligible differences in the 3D architecture of the PGCLC loci between ESC, EpiLC and EpiSC. Overall, the previous results indicate that the 3D organization of PGCLC loci might not play a major role in the establishment of germline competence. However, alternative methods (*e.g.* capture Hi-C, HiChIP) that can globally interrogate enhancer-gene contacts with high sensitivity and spatial resolution might be required to fully elucidate this question.

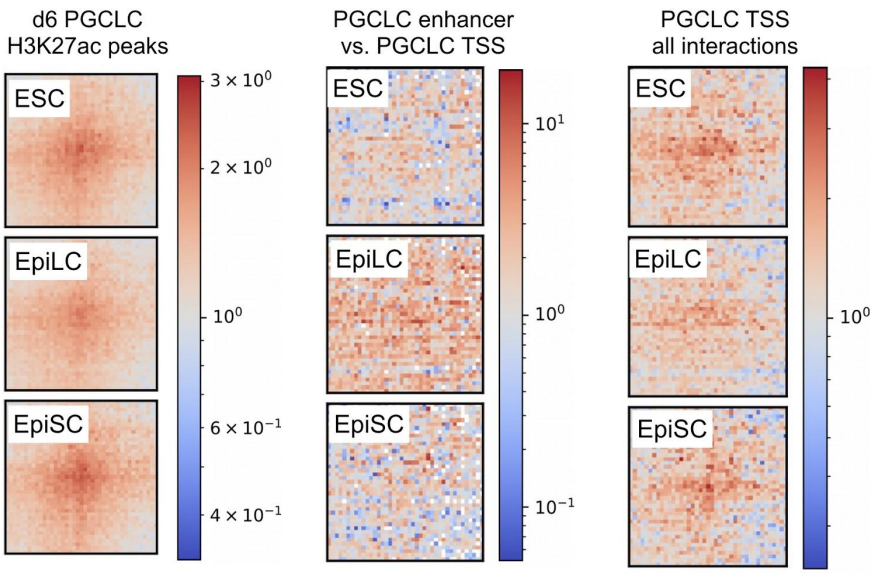


Figure 23: Hi-C analysis for PGCLC gene interactions during differentiation. Shown are aggregate plots for all distal H3K27ac peaks of d6 PGCLC (left, n=39017), all interaction of the PGCLC genes defined by single-cell RNA-seq and the PGCLC enhancer (middle, n=603) and all capture Hi-C interactions within -/+ 3kb of the TSS of all PGCLC genes (right, n=2945) in ESC, EpilC and EpiSC. The data was obtained from Miura et al. 2019 and the Hi-C analysis experiments were performed by Giuliano Crispatzu.

6.4. Cell lines with altered PGCLC competence

Next, cell lines with increased and decreased germline competence were investigated to further evaluate the correlation between germline competence and the epigenetic state of the PGCLC enhancers.

6.4.1. Increased germline competence in *Otx2*$^{-/-}$ cells

The deletion of *Otx2* increases and prolongs germline competence in EpiLC (Zhang et al., 2018a). Before exploring the chromatin features of PGCLC enhancers in *Otx2*$^{-/-}$ cells, their germline competence (Buecker et al., 2014) was evaluated in more detail. Although WT EpiLC were only competent for PGCLC differentiation at the d2 EpiLC stage, increased competence was observed until the d4 EpiSC stage for the *Otx2*$^{-/-}$ cells. Moreover, until d7 EpiSC some *Otx2*$^{-/-}$ cells remained competent for PGCLC differentiation, while in d8 EpiSC germline competence was also lost in *Otx2*$^{-/-}$ cells (Figure 24).

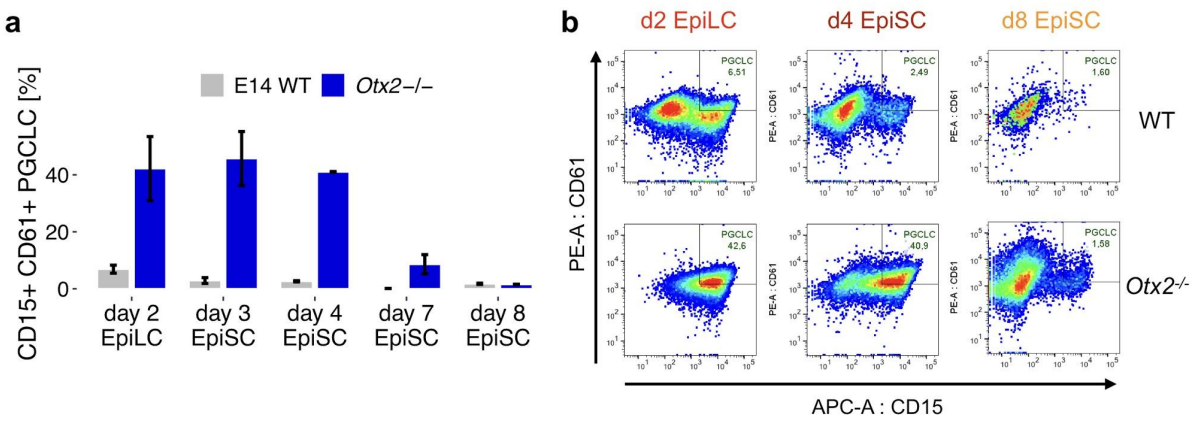


Figure 24: Quantification of Otx2$^{-/-}$ PGCLC induced from different cellular stages.

a.) E14 WT and Otx2$^{-/-}$ ESC were differentiated into d2 EpiLC, d4 EpiSC or d8 EpiSC, which were then further differentiated into PGCLC. Then PGCLC were quantified as the proportion of CD15^{+}CD61^{+} cells found within d4 EB. The error bar represents the standard deviation from two biological replicates.

b.) The panels show representative examples of the PGCLC quantifications from (a). The figures were adapted from Bleckwehl et al. 2021.

6.4.2. Impaired decommissioning of PGCLC enhancers in *Otx2*$^{-/-}$ cells

To test whether the increased PGCLC competence appears with epigenetic changes at PGCLC enhancers, ChIP-seq experiments were performed prior to the PGCLC induction in WT and *Otx2*$^{-/-}$ d2 EpiLC. Strikingly, H3K4me1, H3K4me2 and H3K27ac signals within PGCLC enhancers were higher in *Otx2*$^{-/-}$, particularly within enhancers displaying partial decommissioning in WT EpiLC (*i.e.* Group I PGCLC enhancers). Moreover, the higher H3K4me1/2 levels observed in *Otx2*$^{-/-}$ EpiLC were not due to an increase in these histone modifications already in *Otx2*$^{-/-}$ ESC, nor did they reflect expression changes of the target genes. These results indicate that the increased germline competence of *Otx2*$^{-/-}$ cells could be linked to the delayed decommissioning of a subset of PGCLC enhancers (Figure 25).

To gain further insights into the correlation between impaired PGCLC enhancer decommissioning and increased PGCLC competence, H3K4me1/2 ChIP-seq experiments and bisulfite sequencing were performed at later time points of the EpiSC differentiation. These experiments revealed increased H3K4me1/2 and lower

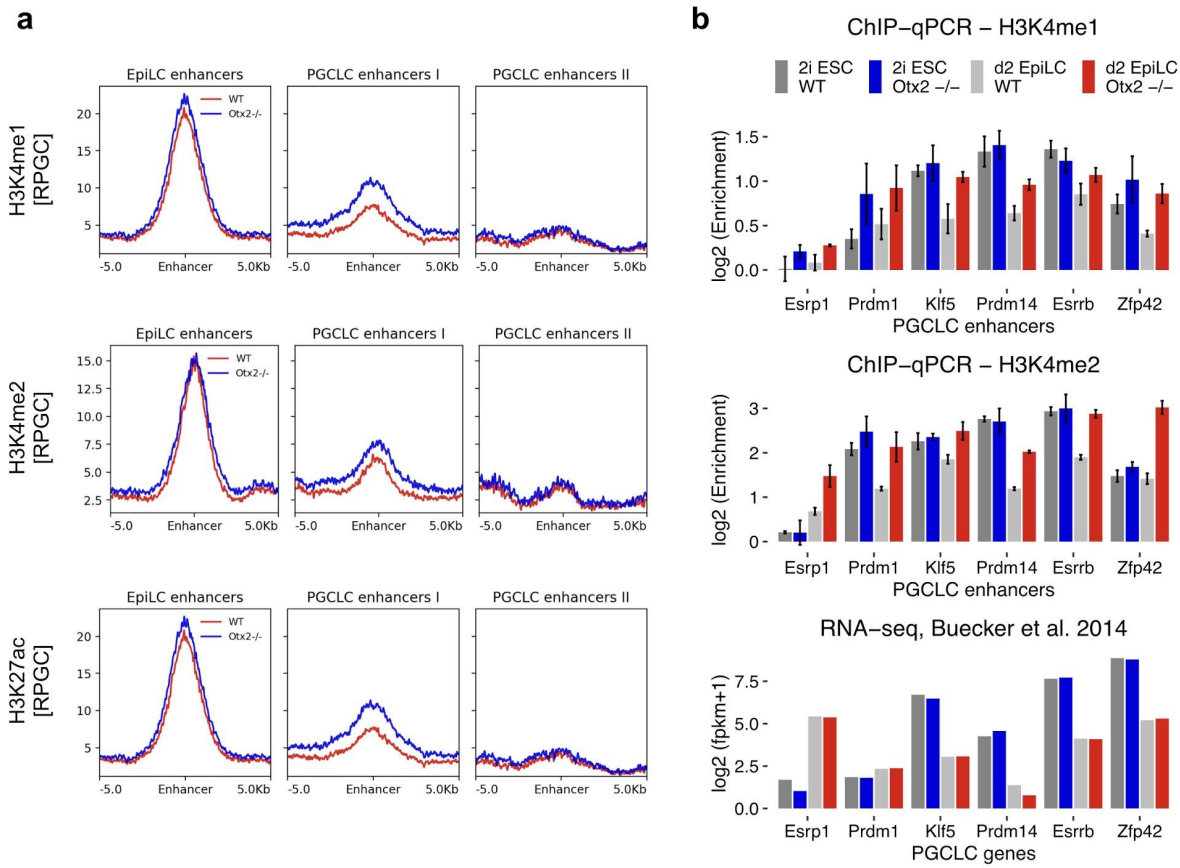


Figure 25: Decreased decommissioning of PGCLC enhancers in Otx2$^{-/-}$ d2 EpiLC.

a.) ChIP-seq profiles of WT (red) and Otx2$^{-/-}$ (blue) EpiLC for the H3K4me1, H3K4me2, H3K27ac levels of EpiLC, Group I and Group II PGCLC enhancers.

b.) H3K4me1 (top) and H3K4me2 (middle) ChIP-qPCR analyses for selected PGCLC enhancers in E14 WT and Otx2$^{-/-}$ cells (2i ESC and d2 EpiLC). H3K4me1 and H3K4me2 enrichments levels were normalized to a negative control region and are shown in log2 scale. The bottom panel shows the expression values for the genes associated with the selected PGCLC enhancers using RNA-seq data from (Buecker et al., 2014).

CpG methylation levels at PGCLC enhancers in *Otx2*$^{-/-}$ d4 EpiSC with extended PGCLC competence. These epigenetic differences between *Otx2*$^{-/-}$ and wild type cells were barely observed in d8 EpiSC, which, importantly, showed limited PGCLC competence in both cases (Figure 26, Figure 27).

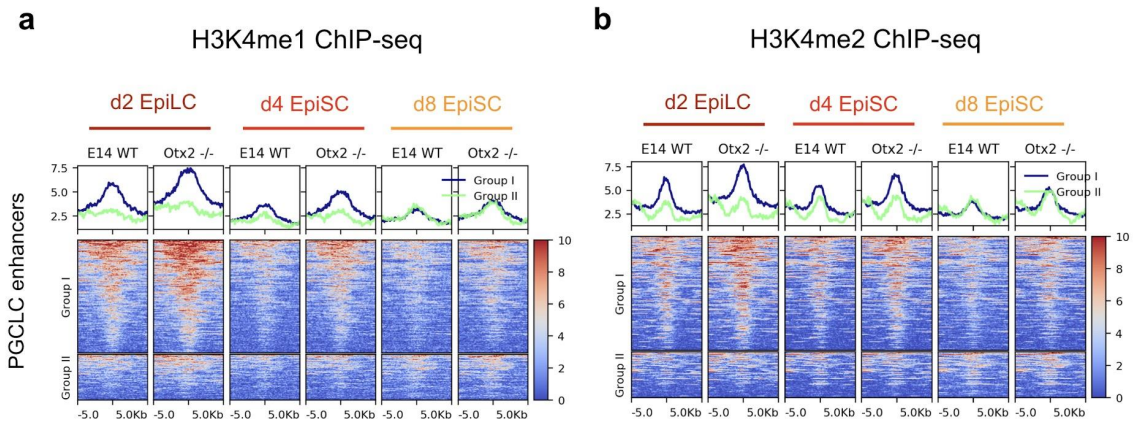


Figure 26: Impaired enhancer decommissioning correlates with PGCLC competence. Average profile (top) and heatmap (bottom) plots showing H3K4me1 (a) and H3K4me2 (b) levels for the Group I and Group II PGCLC enhancers in d2 EpiLC, d4 EpiSC and d8 EpiSC differentiated from either E14 WT or Otx2$^{-/-}$ ESC. The H3K4me1 ChIP-seq data shown for WT d2 EpiLC and WT d8 EpiSC are the same ones used in Figure 15 as second replicates. The figures were adapted from Bleckwehl et al. 2021.

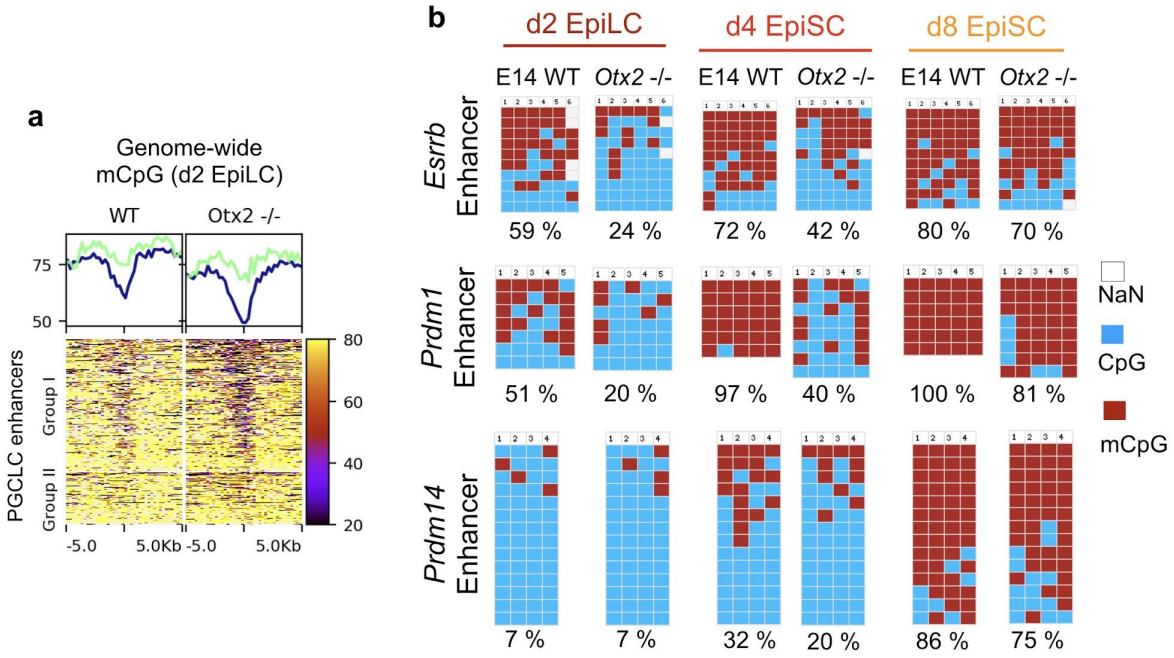


Figure 27: Decreased CpG methylation of Group I PGCLC enhancers in Otx2-/- cells.
a.) Average profile (top) and heatmap (bottom) plots showing the percentage of CpG methylation within Group I and Group II PGCLC enhancers in d2 EpiLC, differentiated from either E14 WT or Otx2$^{-/-}$ ESC.
b.) The CpG methylation levels of selected PGCLC enhancers linked to Esrrb, Prdm1 and Prdm14 genes were measured by bisulfite sequencing in d2 EpiLC, d4 EpiSC and d8 EpiSC

differentiated from E14 WT and Otx2$^{-/-}$ ESC. The columns of the plots correspond to individual CpG dinucleotides located within the indicated enhancer. Unmethylated CpGs are shown in blue, methylated CpGs in red and CpGs that were not sequenced are shown in gray. The rows represent the sequenced alleles in each cell line. The figures were adapted from Bleckwehl et al. 2021.

Nevertheless, the correlation between germline competence and H3K4me1 levels within PGCLC enhancers was not perfect, since *Otx2*$^{-/-}$ d4 EpiSC displayed higher germline competence than WT d2 EpiLC, yet slightly lower H3K4me1 levels within Group I enhancers. Therefore, in addition to H3K4me1, other chromatin features within PGCLC enhancers might contribute to the extended germline competence of *Otx2*$^{-/-}$ cells. In agreement with this possibility, Group I enhancers showed higher H3K4me2 in *Otx2*$^{-/-}$ d4 EpiSC than in WT EpiLC, which might indicate a slower H3K4 demethylation process that results in less CpG methylation of these enhancers in *Otx2*$^{-/-}$ d2 EpiLC and d4 EpiSC (Figure 27).

6.4.3. Decreased germline competence in *Prdm14*$^{-/-}$ cells

Previous reports showed that the loss of *Prdm14* lead to increased CpG methylation in ESC due to the upregulation of *Dnmt3a/b* (Gretarsson and Hackett, 2020; Sim et al., 2017; Yamaji et al., 2013) and reduced numbers of PGCs (Yamaji et al., 2008). While the low CpG methylation of PGCLC enhancers in *Otx2*$^{-/-}$ EpiLC could contribute to their high germline competence, an increase in CpG methylation levels for example in EpiSC or *Prdm14*$^{-/-}$ EpiLC, could restrict PGCLC competence.

To test this hypothesis, genome-wide CpG methylation data generated during the differentiation of *Prdm14*$^{-/-}$ ESC into PGCLC (Shirane et al., 2016) were analyzed. These analyses showed that CpG methylation levels within PGCLC enhancers were increased in *Prdm14*$^{-/-}$ cells throughout the PGCLC differentiation (Figure 28). Furthermore and similarly to *Otx2*$^{-/-}$ EpiLC, *Prdm14*$^{-/-}$ EpiLC did not show major expression changes among PGCLC genes (Bleckwehl et al., 2021; Shirane et al., 2016).

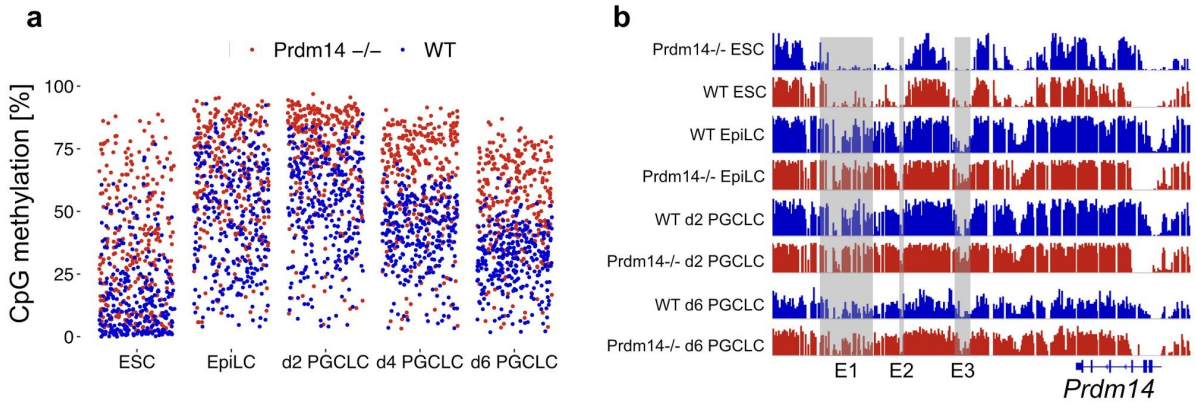


Figure 28: Increased CpG methylation in the Prdm14$^{-/-}$ EpiLC.

a.) Quantification of genome-wide CpG methylation in WT (blue) and Prdm14$^{-/-}$ (red) within -/+ 1kb of PGCLC enhancers during the PGCLC differentiation.

b.) Genome browser view of CpG methylation levels during PGCLC differentiation at the Prdm14 locus in WT and Prdm14$^{-/-}$ cells. The genome-wide CpG methylation data was obtained from (Shirane et al., 2016).

In order to investigate whether the increased CpG methylation in *Prdm14*$^{-/-}$ cells might restrict germline competence, a PRDM14 degron cell line was generated to enable the rapid and reversible degradation of PRDM14 during PGCLC differentiation. Briefly, the *OsTir1* gene was stably integrated into the genome and *Prdm14* was fused to an auxin inducible tag (Nora et al., 2017), which enabled an auxin-dependent degradation of PRDM14 in ESC (Figure 29). The degradation was reversible when auxin was washed off. However, auxin seemed to increase PGCLC formation on its own, as more PGCLC in auxin-treated *OsTir1* cells were observed. Furthermore, the addition of the AID tag to PRDM14 seems to affect its levels and/or function, since PGCLC numbers were reduced in PRDM14-AID cells even in the absence of auxin (Figure 29). These potential effects of the AID on PRDM14 levels could not be properly addressed due to the lack of a suitable antibody to detect endogenous PRDM14. These unexpected effects of Auxin and the AID tag did not allow us to conclusively assess whether the loss of *Prdm14* could compromise germline competence by increasing CpG methylation within PGCLC enhancers. This limitations might be overcome with an updated degradation system (Sathyan et al., 2019; Yesbolatova et al., 2020), or other strategies, like CRISPR/Cas9 mediated

DNA methylation of selected enhancers (Liu et al., 2016), to investigate the impact of CpG methylation for PGCLC competence.

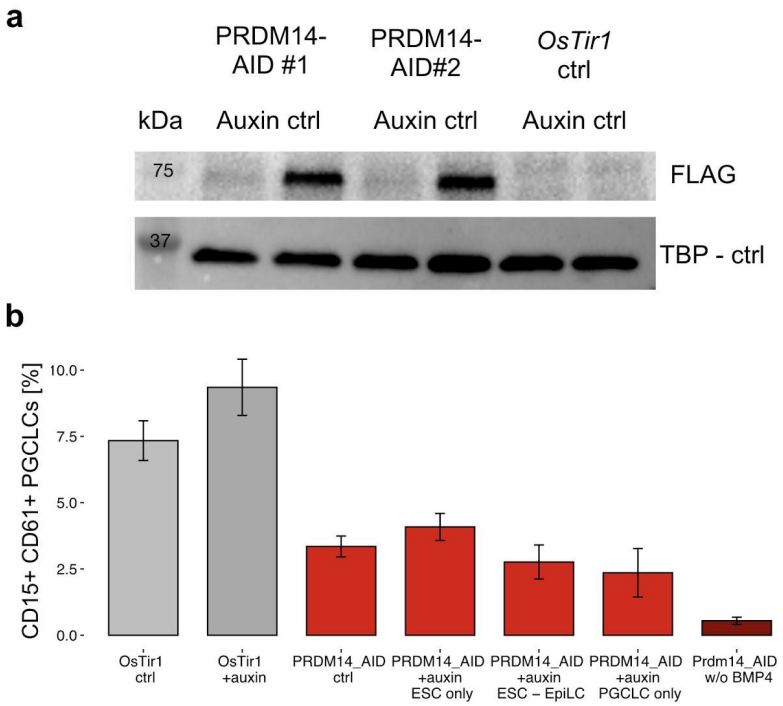


Figure 29: Auxin inducible degradation of PRDM14.

a.) Western blot showing the degradation of PRDM14 in two cell lines, where the auxin degradable tag and a FLAG tag were fused at the C-terminus of Prdm14.

b.) Supplementation of auxin during different time frames of the PGCLC differentiation in the parental line harboring only the OsTir1 insertion (gray) and the OsTir1 cell line with an auxin degradable PRDM14 (red), where no auxin was added (control), auxin was added for 4 days in ESC, but washed off afterwards, auxin was added in ESC and EpiLC but washed off afterwards and where auxin was added with the induction of PGCLC (positive control).

6.5. Accessibility of PGCLC enhancers to germline regulators

To investigate whether the partial decommissioning of PGCLC enhancers in EpiLC compared to EpiSC could have any functional consequences, the accessibility of these enhancers to major germline transcriptional activators (*i.e.* PRDM14 and NANOG) was evaluated. Thereby, clonal ESC lines in which HA-tagged PRDM14 or NANOG could be overexpressed upon addition of Doxycycline (Dox) were generated. In agreement with previous reports, the overexpression of either

PRDM14 or NANOG upon differentiation of EpiLC into PGCLC yielded a high percentage of PGCLC in the absence of growth factors (Magnúsdóttir et al., 2013; Murakami et al., 2016). In contrast, the overexpression of PRDM14 or NANOG upon differentiation from EpiSC resulted in considerably less PGCLC (Figure 30).

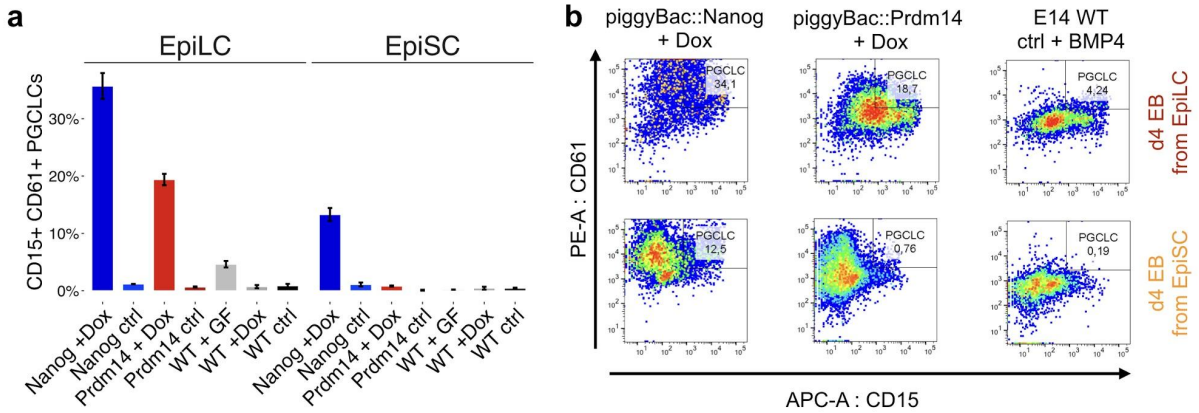


Figure 30: Effects of NANOG and PRDM14 overexpression on PGCLC differentiation.
a.) WT ESC and ESC lines with a Doxycycline-inducible system enabling the overexpression of exogenous NANOG (blue) or PRMD14 (red) were differentiated into d2 EpiLC and EpiSC. Next, d2 EpiLC and EpiSC were differentiated into PGCLC with (+Dox) or without (ctrl) Doxycycline and in the absence of growth factors. As a positive control, WT d2 EpiLC and EpiSC were differentiated with growth factors (WT +GF). The percentage of PGCLC (i.e. $CD15^{+}CD61^{+}$ cells) was measured in d4 EB from two biological replicates.
b.) Representative examples of the PGCLC quantifications shown on the left panel. The figures were adapted from Bleckwehl et al. 2021.

To assess whether PGCLC enhancers were differentially accessible to these TFs in EpiLC and EpiSC, ChIP-seq experiments were performed after a short induction of HA-tagged PRDM14 or NANOG in both cell types. Importantly, PGCLC enhancers, especially those belonging to Group I, were considerably more bound by PRDM14-HA and NANOG-HA in EpiLC than in EpiSC (Figure 31).

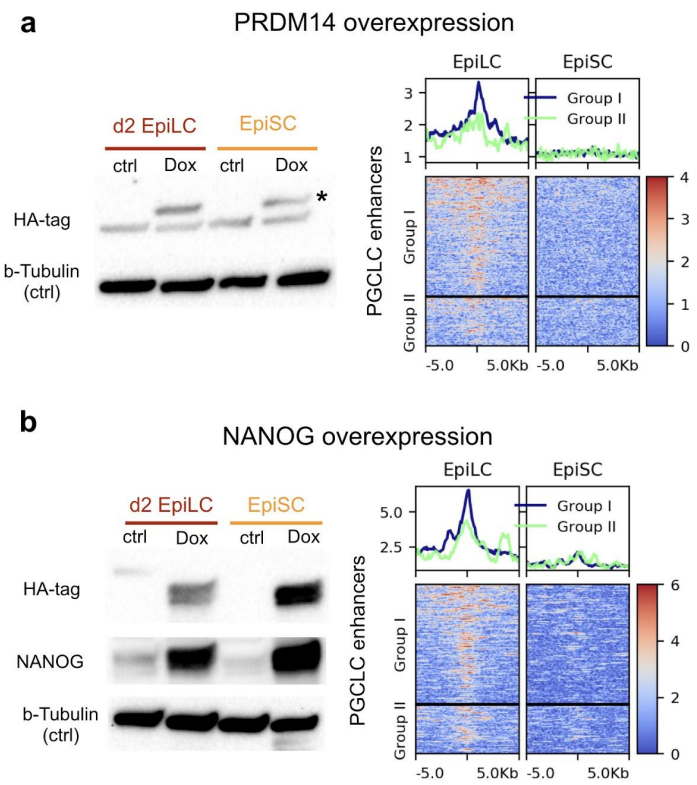


Figure 31: NANOG and PRDM14 binding to PGCLC enhancers.
a.)Western blot analysis of the inducible overexpression of exogenous PRDM14-HA (left). EpiLC or EpiSC were either left untreated (ctrl) or treated with Dox for 18 hours and then the PRDM14-HA levels were measured using an anti-HA antibody. The band corresponding to PRDM14-HA is indicated with an asterisk. B-Tubulin was used as a loading control in all samples. The exogeneous PRDM14 tagged with HA were overexpressed for 18 hours in EpiLC and EpiSC. Then, ChIP-seq experiments were performed using an anti-HA antibody (right). The average profile (top) and heatmap (bottom) plots show the binding of the PRDM14-HA within Group I and Group II PGCLC enhancers.
b.) Experiments as in (a) for the overexpression of exogenous NANOG-HA. For the western blot, endogenous and exogenous NANOG levels were additionally measured with an anti-NANOG antibody. The figures were adapted from Bleckwehl et al. 2021.

Furthermore, NANOG-HA binding to the PGCLC enhancers resulted in increased H3K27ac and H3K4me2 levels in EpiLC, but not in EpiSC, particularly within Group I enhancers (Figure 32). Altogether, these results indicate that PGCLC enhancers, especially those belonging to Group I, are both accessible and responsive to transcriptional activators in EpiLC, but not in EpiSC.

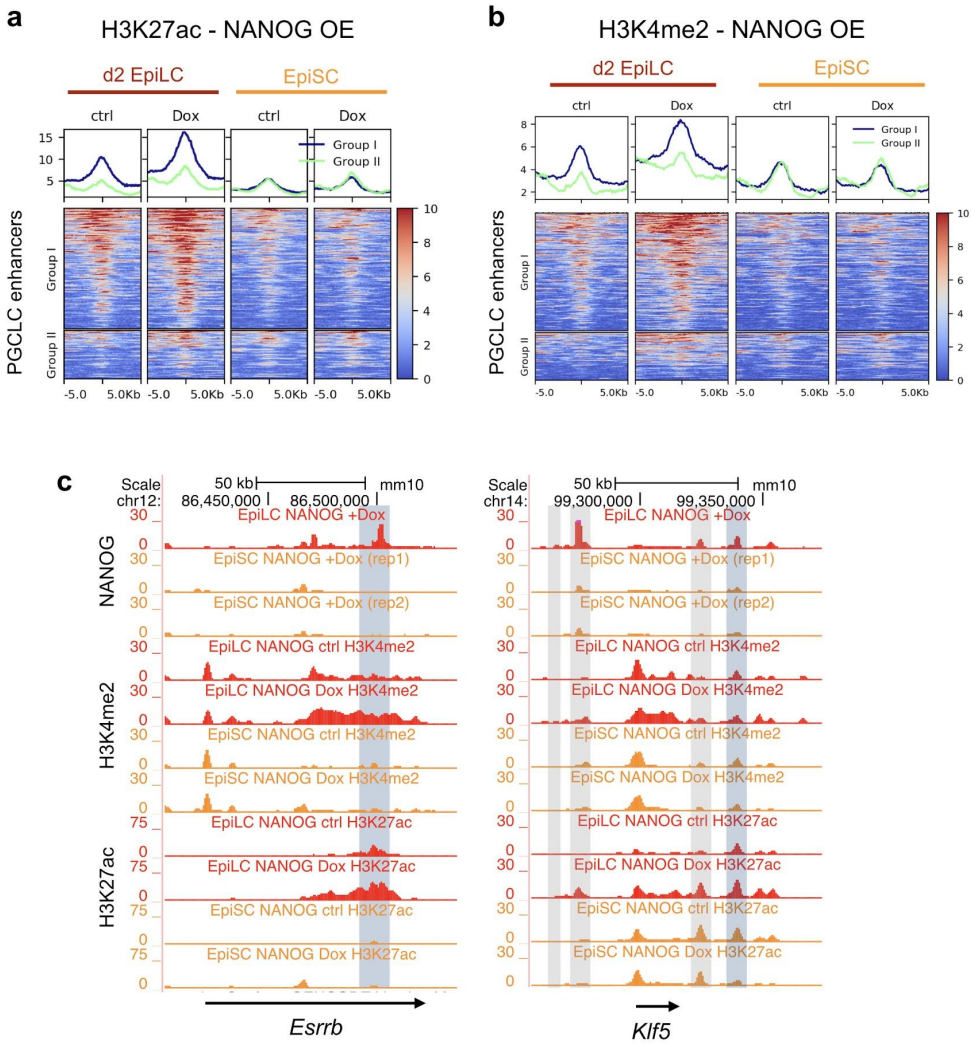


Figure 32: Epigenetic changes of active chromatin features by NANOG-HA overexpression.

a-b.) The ESC line enabling the inducible overexpression of exogenous NANOG-HA was differentiated into d2 EpiLC and EpiSC. Next, the cells were either left untreated (ctrl) or treated with Dox to overexpress NANOG as described in Figure 31 and, subsequently, H3K27ac (a) or H3K4me2 (b) ChIP-seq experiments were performed. The average profile (top) and heatmap (bottom) plots show H3K27ac and H3K4me2 levels within Group I and Group II PGCLC enhancers.

c.) Genome-browser view of the Klf5 and Esrrb loci showing the binding of NANOG-HA in both EpiLC and EpiSC as well as the H3K27ac and H3K4me2 signals in untreated (-Dox) and Dox-treated (i.e. NANOG-HA overexpression) d2 EpiLC and EpiSC. For the NANOG-HA ChIP-seq profiles in EpiSC the results of two independent biological replicates are shown to illustrate that the weak binding of this transcription factor to the PGCLC enhancers was reproducibly observed. The previously deleted PGCLC enhancers (Figure 19) are highlighted in blue and other PGCLC enhancers found within the same loci are indicated in gray. The figures were adapted from Bleckwehl et al. 2021.

6.6. H3K4me1/2 is dispensable for the establishment of the formative and primed *in vitro* pluripotency expression programs

Overall, the presented data suggest that a significant fraction of PGCLC enhancers display permissive chromatin features in EpiLC, including higher H3K4me1 levels, that might contribute to their increased germline competence compared to EpiSC. Therefore, the next aim was to investigate the relevance of H3K4me1 for PGCLC competence.

H3K4me1 is catalyzed by two redundant H3K4 methyltransferases (KMT2C/D). Recent work based on the generation of *Kmt2c* and *Kmt2d* catalytic mutant ESC lines (dCD and dCT ESC lines) reported that the non-catalytic functions of the KMT2C/D complex are more relevant for enhancer activity than H3K4me1; as the expression changes in the *Kmt2c/d* KO ESC were more severe compared to the dCD ESC (Dorighi et al., 2017). Furthermore, in the dCD ESC, active enhancers showed a moderate reduction in H3K27ac, while gene expression, eRNA levels and the binding of KMT2C/D and their associated complexes were not affected. Therefore, H3K4me1 seems to be largely dispensable for the maintenance of enhancer activity (Dorighi et al., 2017; Rickels et al., 2017). However, the differentiation capacity of the dCD ESC and, thus, the importance of H3K4me1 upon mouse ESC differentiation has not been investigated yet. Therefore, the role of H3K4me1 upon exit from the naïve pluripotent state was evaluated first by performing RNA-seq and ChIP-seq experiments for H3K4me1/2/3 and H3K27ac in dCD (KMT2C: Y4792A; KMT2D: Y5477A) and dCT (KMT2C: Y4792N; KMT2D: Y5477N) ESC as well as upon their differentiation into d2 EpiLC and EpiSC.

In order to extend previous observations based on the use of the dCD cell line, the dCT ESC line, which contains a different amino acid substitution at the same position of the SET domain of KMT2C and KMT2D was further characterized. In agreement with the previous results obtained in dCD ESC (Dorighi et al., 2017), the dCT ESC showed an almost complete loss of H3K4me1 and diminished levels of H3K37ac at enhancers (Figure 33).

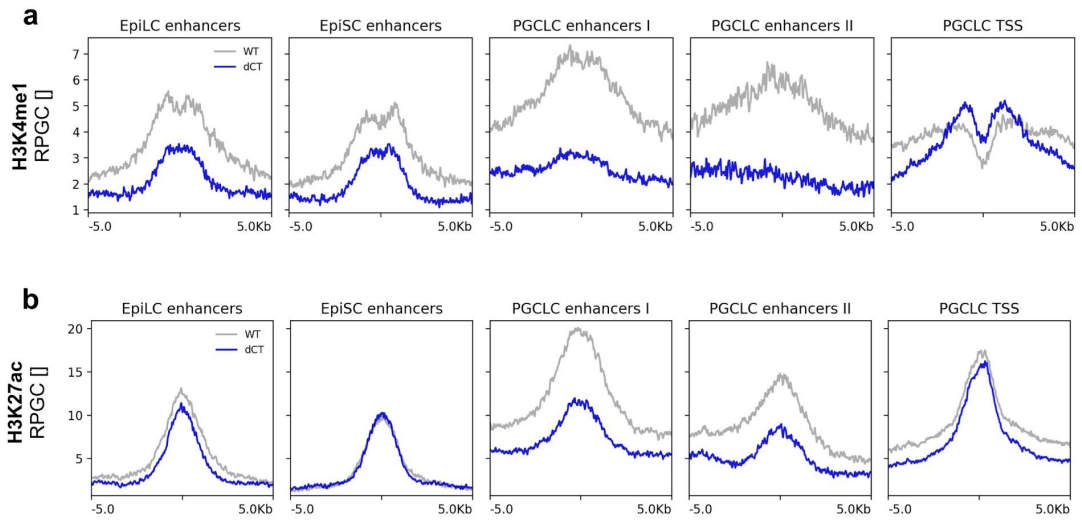


Figure 33: H3K4me1 and H3K27ac signals in WT and dCT ESC.
H3K4me1 (a) and H3K27ac (b) ChIP-seq profiles of WT and dCT ESC for EpiLC, EpiSC, Group I and Group II PGCLC enhancers and PGCLC TSS.

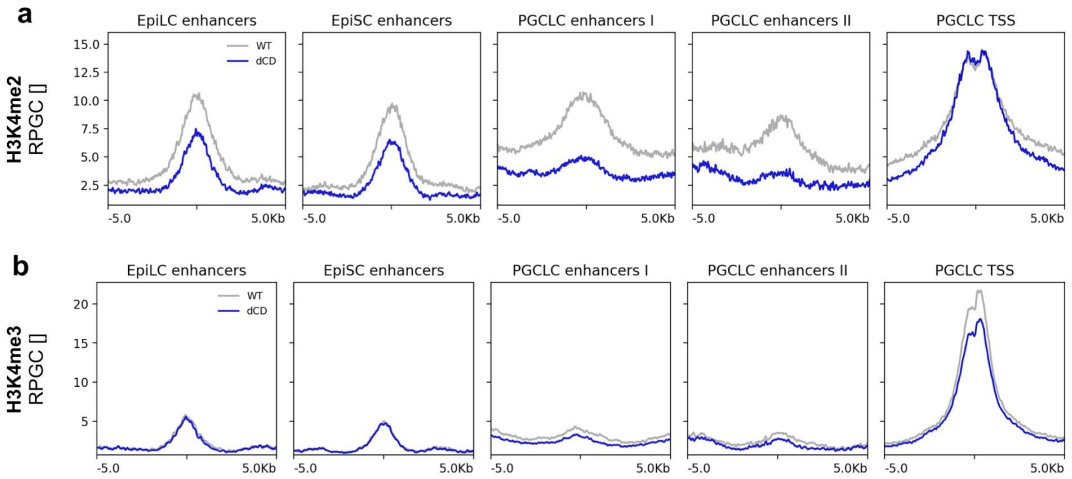


Figure 34: H3K4me2 and H3K4me3 signals in WT and dCD ESC.
H3K4me2 (a) and H3K4me3 (b) ChIP-seq profiles of WT and dCD ESC for EpiLC, EpiSC, Group I and Group II PGCLC enhancers and PGCLC TSS.

Furthermore, to better characterize the epigenetic defects of the dCD cells, H3K4me2 and H3K4me3 ChIP-seq experiments were performed. This revealed that H3K4me2 was also profoundly reduced at active enhancers, while H3K4me2/3 levels at PGCLC TSS were not altered (Figure 34). Conclusively, the dCD cell line is deficient for both H3K4me1 and H3K4me2.

Upon differentiation into EpiLC and EpiSC, H3K4me1 and H3K4me2 were similarly lost at enhancers and also the H3K27ac levels of the EpiLC and EpiSC enhancers were affected (Figure 35). Notably, H3K27ac was gained at EpiSC enhancers in dCD EpiSC compared to dCD EpiLC, indicating at least a partial EpiSC enhancer activation in dCD EpiSC. Additionally, elevated H3K4me1 was observed at the PGCLC TSS in all stages.

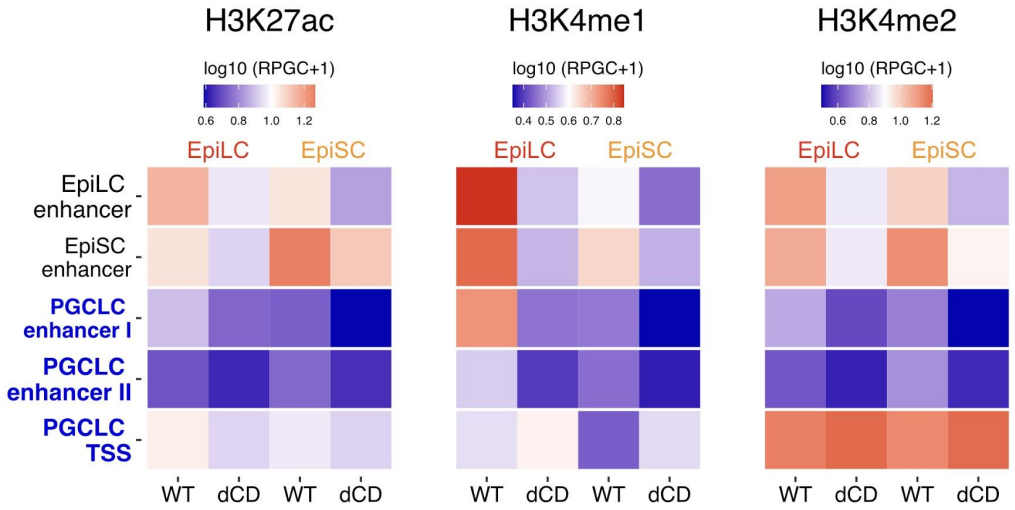


Figure 35: Epigenetic changes in the dCD cell line upon differentiation. Summary plot showing the average H3K27ac, H3K4me1, H3K4me2 and H3K4me3 levels in WT and dCD cells (d2 EpiLC and EpiSC) for the EpiLC, EpiSC and PGCLC enhancers as well as the TSS of the PGCLC genes. Quantifications were performed by measuring the average signals of each epigenetic mark within -/+ 1kb of the enhancers or TSS. The H3K4me1 ChIP-seq data shown for WT d2 EpiLC and EpiSC are the same ones used in Figure 15 as the third replicate. RPGC: Reads per genomic content. The figures were adapted from Bleckwehl et al. 2021.

Next, to investigate how the loss of H3K4me1/2 could impact gene expression, RNA-seq experiments were performed WT and dCD ESC as well as upon their differentiation into EpiLC. Notably, rather minor gene expression changes were observed between WT and dCD cells in both pluripotent states (Figure 36). These results suggest that H3K4me1/2 is dispensable for the activation of the EpiLC expression program.

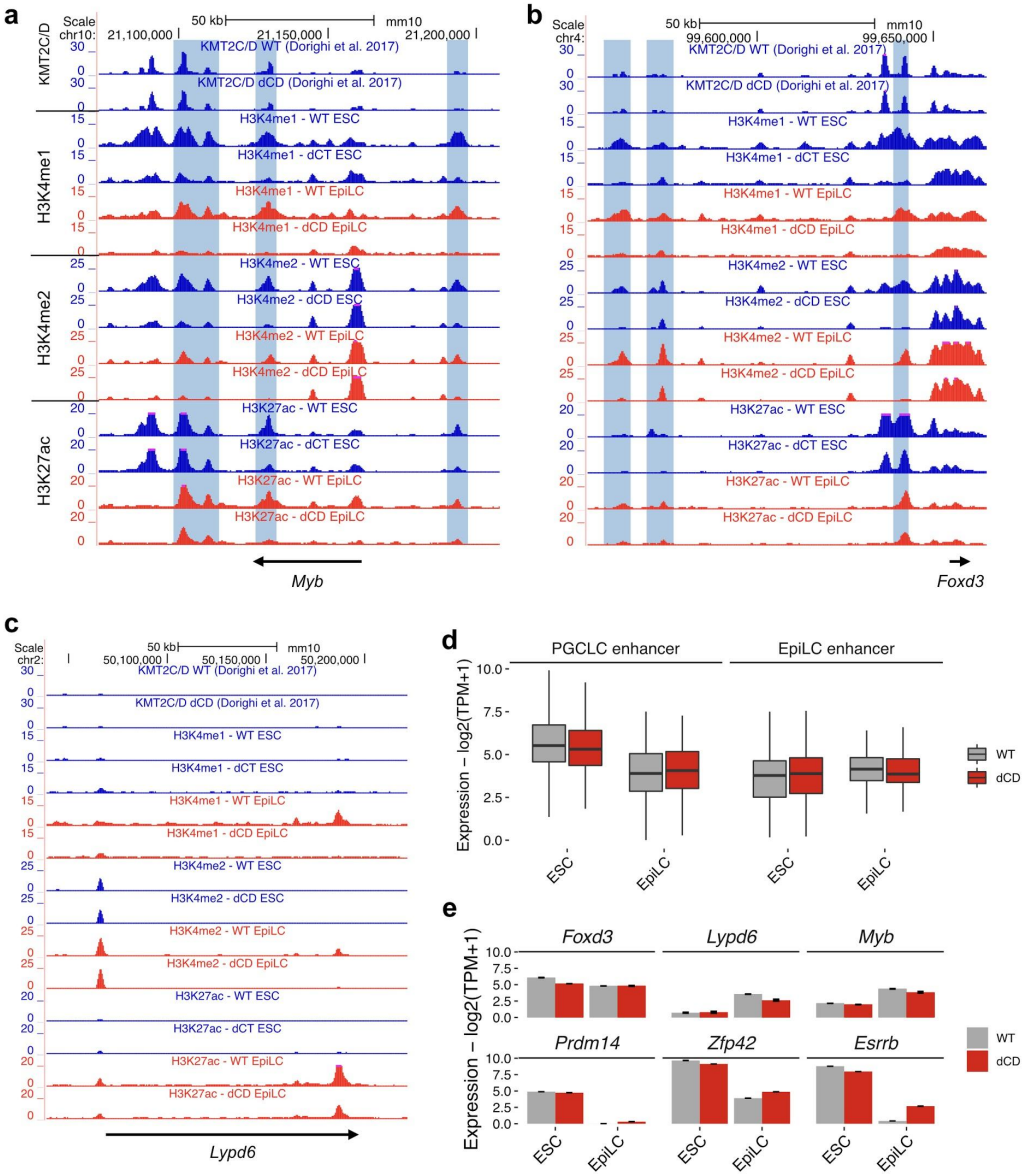


Figure 36: Epigenetic and transcriptional changes of the dCD cell line upon differentiation into EpiLC.

a-c.) Epigenetic profiles of selected EpiLC enhancers (highlighted in dark blue), which were associated with Myb (a), Foxd3 (b) and Lypd6 (c). Shown are the tracks of KMT2C/D, H3K4me1, H3K4me2 and H3K27ac in WT and H3K4me1/2 deficient ESC (blue) and EpiLC (red). The KMT2C/D ChIP-seq data was obtained from Dorighi et al. 2017.

d.) Box plots showing the expression, as measured by RNA-seq, of the genes associated with the PGCLC and EpiLC enhancers in WT (gray) and dCD (red) ESC cells as well as upon their differentiation into EpiLC. The RNA-seq data in ESC was obtained from Dorighi et al. 2017 and the experiments in d2 EpiLC were performed in duplicates.

e.) Expression levels of representative EpiLC (Foxd3, Lypd6, Myb) and PGCLC (Prdm14, Zfp42, Esrrb) genes.

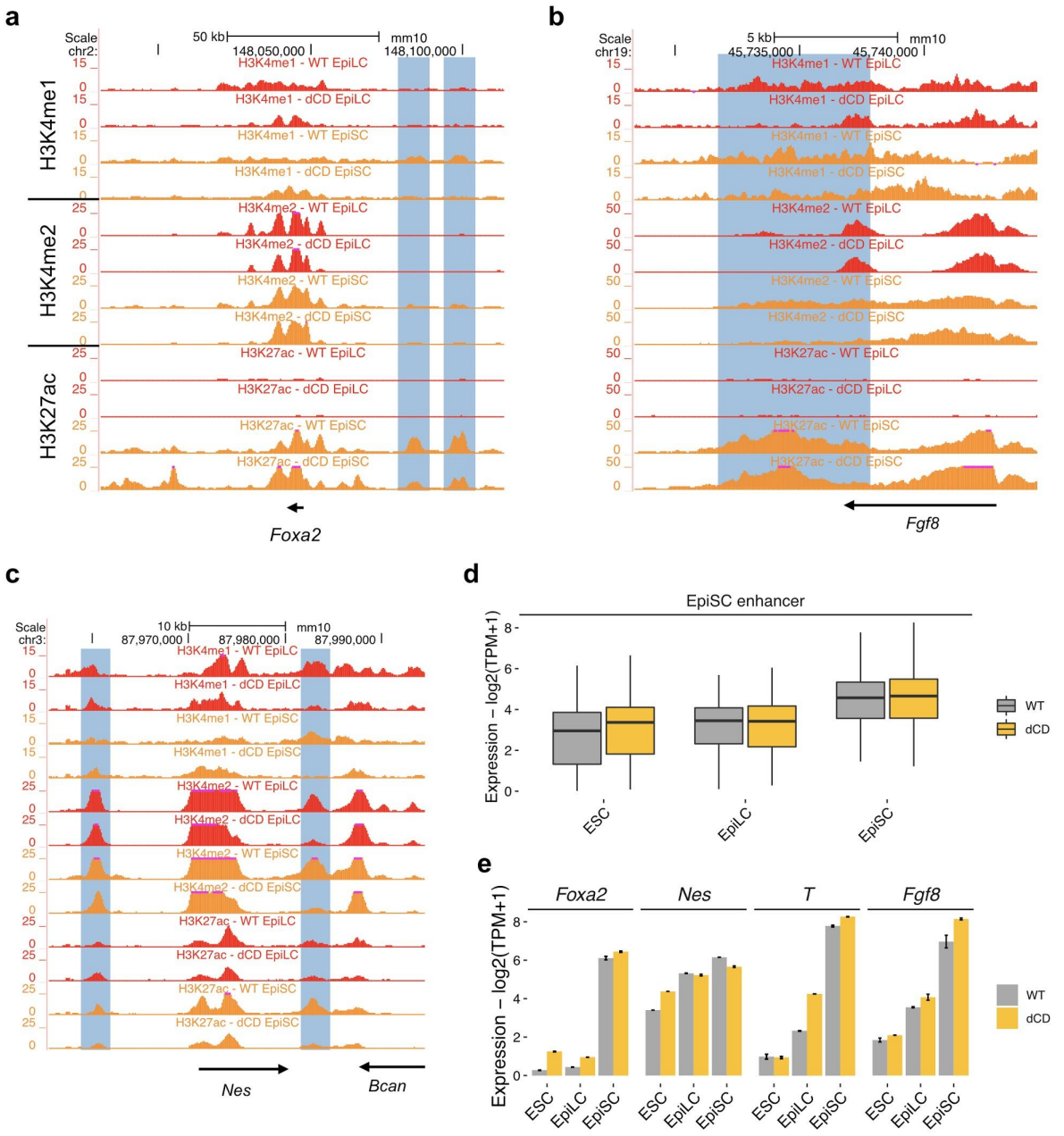


Figure 37: Epigenetic and transcriptional changes of the dCD cell line upon differentiation into EpiSC.

a-c.) Epigenetic profiles of selected EpiSC enhancers (highlighted in dark blue), which were associated with Foxa2 (a), Fgf8 (b) and Nes (c). Shown are the tracks of H3K4me1, H3K4me2 and H3K27ac in WT and dCD EpiLC (red) and EpiSC (orange).

d.) Box plots showing the expression, as measured by RNA-seq, of the genes associated with the PGCLC and EpiLC enhancers in WT (gray) and dCD (red) ESC cells as well as upon their differentiation into EpiLC. The RNA-seq data in ESC was obtained from Dorighi et al. 2017 and the experiments in d2 EpiLC were performed in duplicates.

e.) Expression levels of representative EpiSC genes.

EpiSC show incipient expression of somatic germ layers and can be easily differentiated towards neural, endodermal or mesodermal progenitors (Morgani et al., 2017). Interestingly, evaluation of major somatic specifiers, such as *Nes* (neural), *Foxa2* (endoderm) or *Ffg8* and *T* (mesodermal), revealed that the expression levels of these genes and the H3K27ac levels within their associated enhancers were almost the same in WT and dCD EpiSC. Therefore, EpiSC differentiation potential might not be severely impaired by the H3K4me1/2 deficiency (Figure 37). Overall, the results are in agreement with a recent preprint showing only minor defects upon differentiation of the dCD ESC into either EBs (reflecting the formation of all somatic germ layers) or neural progenitors (Xie et al., 2020).

6.7. H3K4me1 is necessary for in vitro germline competence

The previous results indicate that the persistence of H3K4me1 at PGCLC enhancers might be an important feature of germline competence. Careful evaluation of the PGCLC enhancers in dCD ESC, EpiLC and EpiSC revealed that these enhancers display a similar epigenetic state (*i.e.* low H3K4me1/2 and H3K27ac) in dCD EpiLC and WT EpiSC (Figure 38).

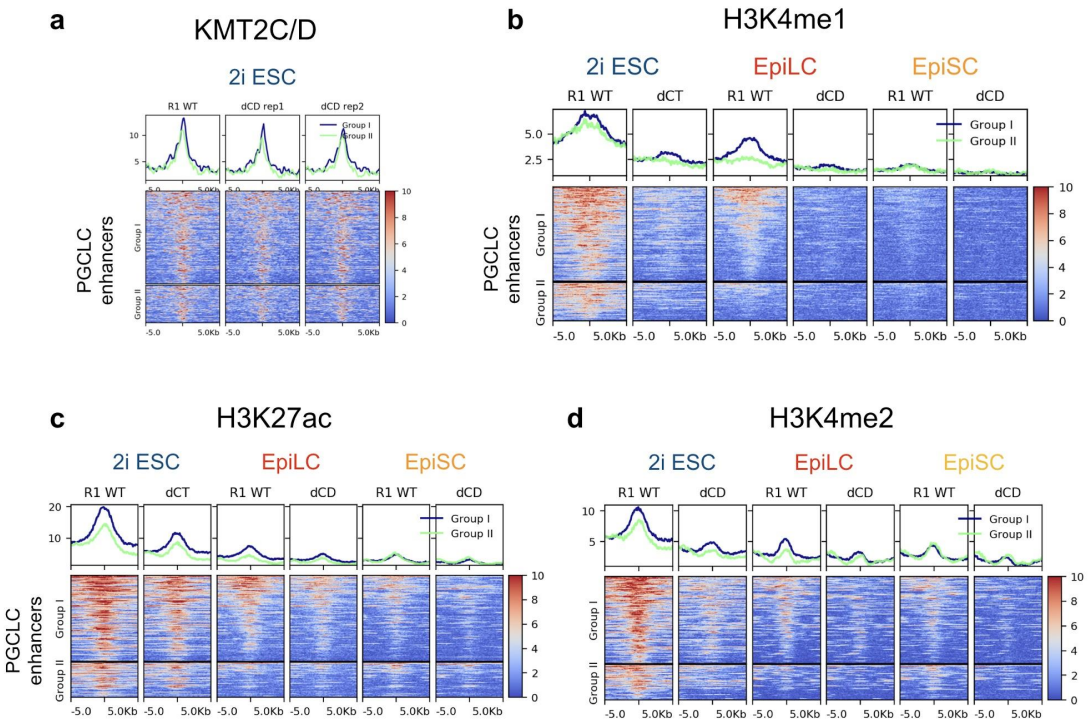


Figure 38: ChIP-seq profiles of PGCLC enhancers in the dCD cell line. As indicated above each panel, the average profile (top) and heatmap (bottom) plots of KMT2C/D in ESC (a) as well as H3K4me1 (b), H3K4me2 (c) and H3K27ac (d) levels for the Group I and Group II PGCLC enhancers in R1 WT and KMT2C/D catalytic mutant ESC lines as well as upon their differentiation into d2 EpiLC and EpiSC. The H3K4me1 ChIP-seq data shown for WT d2 EpiLC and WT EpiSC are the same ones used in Figure 15 as third replicates. The figures were adapted from Bleckwehl et al. 2021.

This suggests that PGCLC enhancers get fully decommissioned in dCD EpiLC, which, consequently, might reduce their germline competence. This hypothesis was experimentally tested as described in the following sections.

6.7.1. H3K4me1 is necessary for PGCLC formation and the re-activation of PGCLC enhancers

The different KMT2C/D catalytic mutant ESC lines were differentiated into PGCLC. Notably, dCD and dCT ESC both showed a significant reduction in their PGCLC differentiation capacity (Figure 39). Such PGCLC differentiation defect was not observed with cells that were catalytic mutant for KMT2D, but not KMT2C (*i.e.* 4CT cells), consistent with the known redundant function of KMT2C and KMT2D (Hu et al., 2013b). Furthermore, the compromised PGCLC differentiation of the dCD cells was also observed when the differentiation time was extended, indicating that the observed defects are not simply explained by a delay in PGCLC specification (Bleckwehl et al., 2021).

Then, to investigate if the reduced germline competence of the dCD cells could be caused by compromised PGCLC enhancer reactivation upon PGCLC differentiation, a ChIP-seq protocol for reduced cell numbers was used to investigate H3K27ac in d4 EB derived from WT and dCD ESC. The ChIP-seq profiles revealed that PGCLC enhancers, especially those belonging to Group I, displayed higher H3K27ac levels in WT than in dCD d4 EB (Figure 39). Overall, the reduced PGCLC numbers and H3K27ac levels at PGCLC enhancers observed in dCD d4 EB indicate that PGCLC enhancers might not get properly re-activated in the absence of H3K4me1.

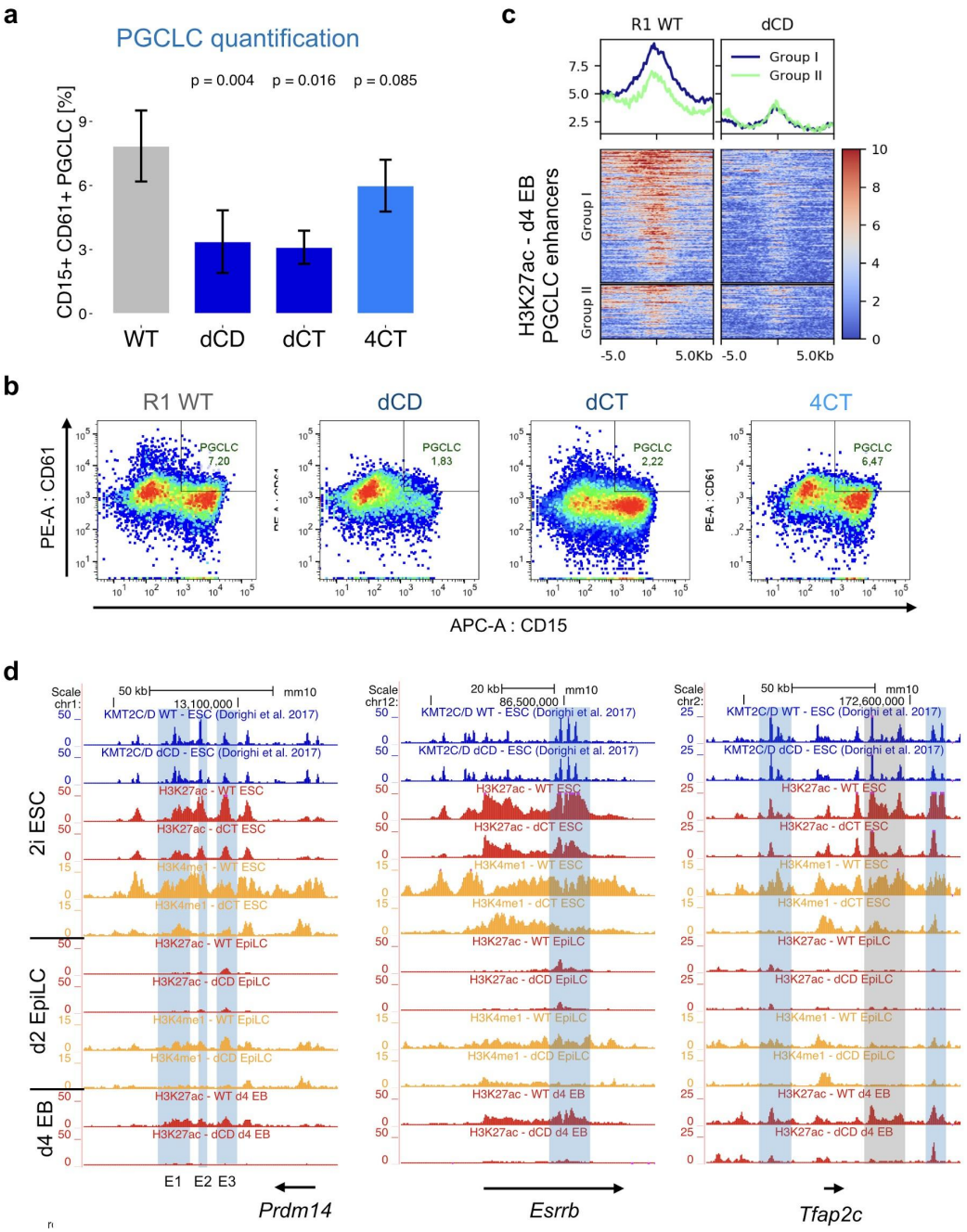


Figure 39: Compromised differentiation of KMT2C/D catalytic mutant ESC lines into PGCLC.

a.) WT ESC and ESC lines with the indicated KMT2C/D amino acid substitutions were differentiated into PGCLC. PGCLC were defined as CD15^{+}CD61^{+} cells within d4 EB. Each PGCLC quantification was performed in at least four biological replicates. The p-values were calculated using Wilcoxon tests.

b.) Representative examples of the PGCLC quantifications shown in the left panel.

c.) *H3K27ac ChIPmentation experiments were performed in d4 EB differentiated from R1 WT and dCD ESC. The average profile (top) and heatmap (bottom) plots show H3K27ac levels for the Group I and Group II PGCLC enhancers.*

d.) Genome browser views showing KMT2C/D ChIP-seq from Dorighi et al. 2017 in ESC (blue), H3K4me1 levels in ESC and EpiLC (orange), and H3K27ac profiles in ESC, EpiLC and d4 EB (red) in R1 WT and dCD. Shown are three representative PGCLC genes (Prdm14, Esrrb and Tfap2c) and their linked enhancers. For Prdm14 and Esrrb the CRISPR-Cas9 deleted enhancer regions are highlighted in blue, while for Tfap2c the enhancers linked by capture-HiC are highlighted in blue and another putative enhancer in gray. The figures were adapted from Bleckwehl et al. 2021.

6.7.2. The induction of the PGCLC expression programme is impaired in the absence of H3K4me1/2

The previous ChIP-seq experiments were performed in EBs and not in sorted PGCLC. Therefore, the low H3K27ac levels in dCD cells could be caused by either a defect in the activation of PGCLC enhancers and their associated genes or by an overall reduction in the number of PGCLC present within d4 EBs. To distinguish between these two possibilities, scRNA-seq analyses of WT (1416 cells) and dCD (1699 cells) d4 EBs were performed. UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) analysis (McInnes et al., 2018) of the resulting single cell transcriptomes confirmed the presence within d4 EB of subclusters resembling the main cell populations (*i.e.* PGCLC, ExEctorderm-like, ExMesoderm-like) that characterize the proximo-posterior end of the mouse embryo following germline specification (Figure 40, Figure 41). Most importantly, the PGCLC cluster consisted mostly of WT cells, thus in agreement with the results obtained using FACS and cell surface markers. An additional ExEndoderm/Gut-like subcluster was identified in both cell types. This ExEndoderm/Gut-like cluster was not observed in previous scRNA-seq analyses, which might be attributed to the lower number of d4 EB cells. By differential expression analysis of the clusters, a 2-cell-like cluster that expressed 2-cell markers (*e.g. Zscan4a-f*) only in dCD cells was identified.

The transcriptomes of some of the extraembryonic-like cell types were similar between WT and dCD cells (e.g. Ex-Endoderm/Gut-like, Endothelial-like, Figure 41). However, many dCD cells were located in clusters with poorly defined identity that did not show differential expression of specific markers. These

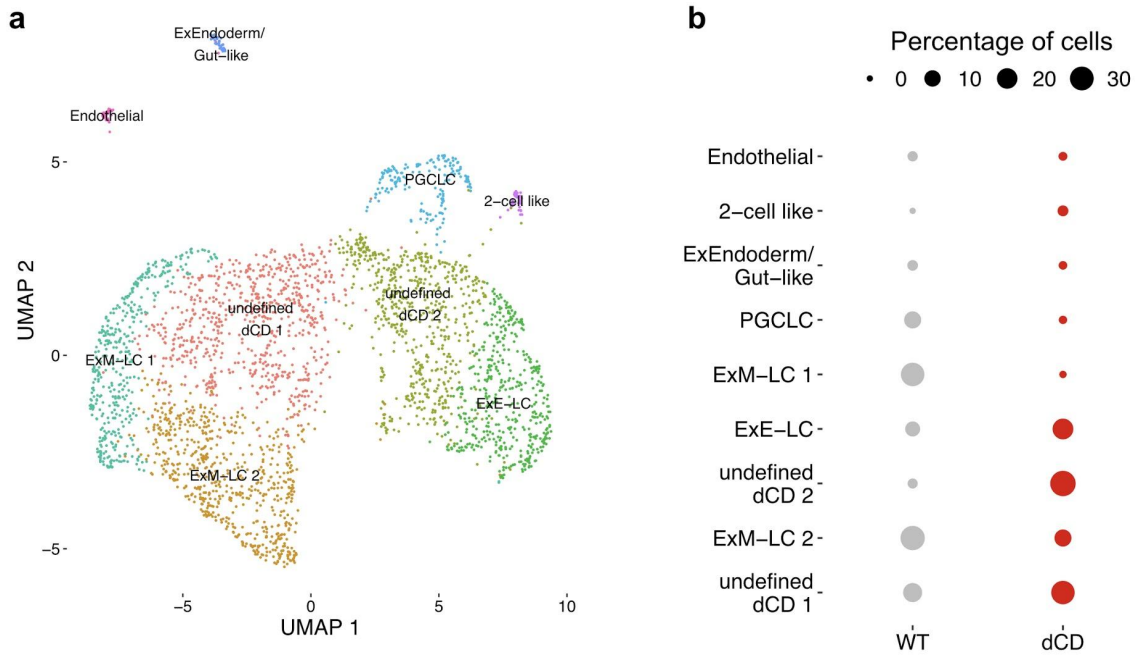


Figure 40: Clusters identified in WT and dCD d4 EBs based on single-cell RNA-seq profiling. a.) UMAP plot showing the different cell clusters identified within WT and dCD d4 EB according to the single-cell RNA-seq data. The single-cell RNA-seq data was subject to clustering analysis (Shared Nearest Neighbor) and the cellular identity of the resulting clusters was determined by differential expression analysis between clusters as well as by using specific markers of the main embryonic and extraembryonic tissues found within E8.25 mouse embryos (Ibarra-Soria et al. 2018).
b.) Percentage of WT and dCD cells present within each of the clusters identified in d4 EB as shown to the left. The figures were adapted from Bleckwehl et al. 2021.

"undefined" clusters were transcriptionally more heterogeneous than the remaining clusters found within d4 EB (Bleckwehl et al., 2021). Furthermore, the undefined clusters of the dCD EBs contained cells that expressed major PGC markers (i.e. *Prdm1* or *Dppa3*) but not naïve pluripotency ones (i.e. *Klf4*), suggesting a cellular identity similar to PGCLC. The proportion of these *Prdm1orDppa3+/Klf4-* cells was similar among WT and dCD EBs (Figure 41). However, while in the WT EBs, these cells were mostly found within the PGCLC subcluster, in the dCD EBs they were part of the subclusters with poorly defined identity, suggesting important transcriptional differences between WT and dCD *Prdm1orDppa3+/Klf4-* cells.

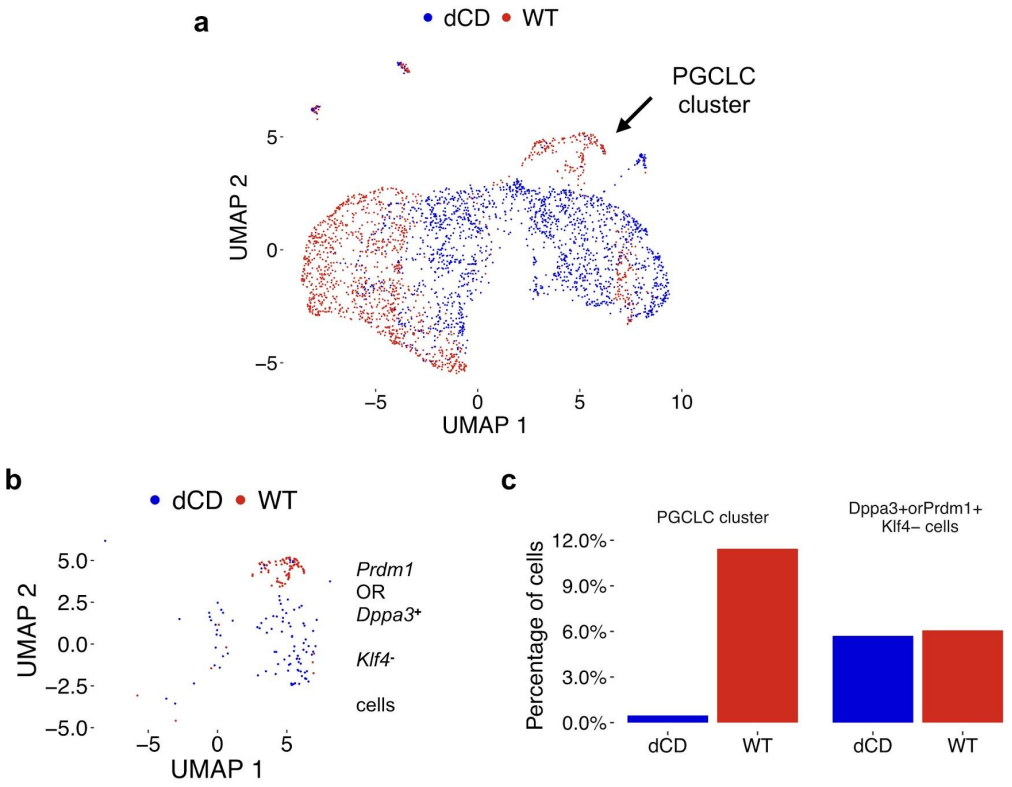


Figure 41: Identification of PGCLC and Prdm1orDppa3+/Klf4- cells within *d4 EBs.*

a.) UMAP plot of the single-cell RNA-sequencing data generated in d4 EB differentiated from R1 WT and dCD ESC. The PGCLC cluster, which was identified by the expression of PGCLC markers, is indicated with an arrow (see also Figure 40). Each red and blue dot in the plot represents a WT and dCD individual cell, respectively.

b.) Shown are those cells that express Prdm1 or Dppa3, but not the naïve pluripotency marker Klf4 (Prdm1orDppa3+/Klf4-) in WT (red) and dCD (blue) d4 EBs within the same UMAP plot as in (a).

c.) Percentage of WT and dCD d4 EB cells found within the PGCLC cluster (a) or expressing Prdm1/Dppa3, but not Klf4 (Prdm1orDppa3+/Klf4-) (b). The figures were adapted from Bleckwehl et al. 2021.

Congruently, the expression of the PGCLC genes associated with PGCLC enhancers were significantly reduced in dCD *Prdm1orDppa3+/Klf4-* cells in comparison to their WT counterparts (Figure 42). This included PGCLC genes that have been individually reported to be major PGC/PGLC regulators (*Nanog*, *Tfap2c*, *Nr5a2*, *Prdm14*, *Cbfa2t2*, *Esrrb*, *Utf1, Erbb3, Kit, Lefty1*) (Farini et al., 2007; Hackett et al., 2018; Kasowitz et al., 2017; Murakami et al., 2016a; Okamura et al., 2019; Schemmer et al., 2013; Toyoda-Ohno et al., 1999; Tu et al., 2016; Wu et al., 2013;

Yamaji et al., 2008). On the other hand, the expression differences for the PGCLC genes without associated enhancers were rather minor. In summary, the scRNA-seq analyses suggest that the induction of the PGCLC expression program, particularly of those genes linked to PGCLC enhancers, is compromised in dCD cells.

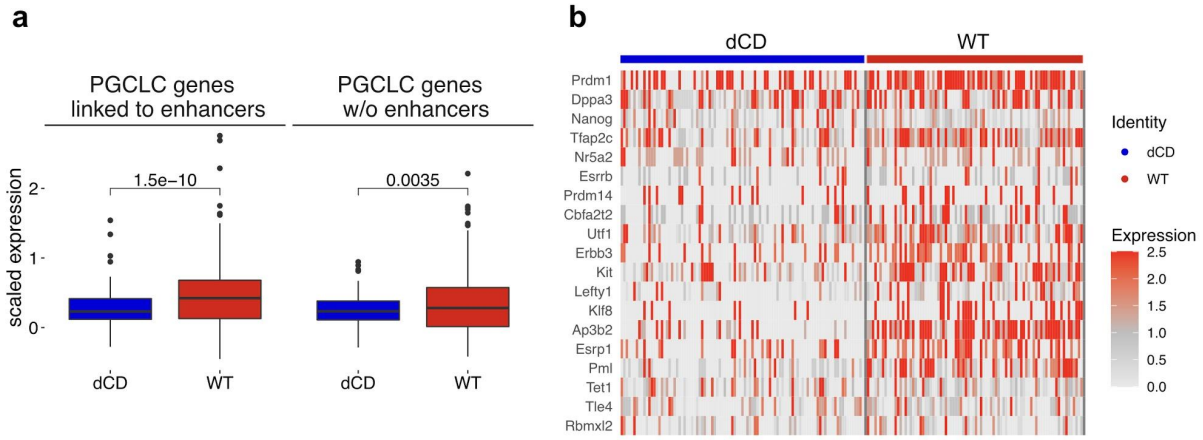


Figure 42: The expression of PGCLC genes linked to PGCLC enhancers is reduced in dCD Prdm1orDppa3+/Klf4- cells.

a.) Box plots showing the expression of the PGCLC genes linked (n=216) or not (n=170) to at least one PGCLC enhancer within WT (red) and dCD (blue) Prdm1orDppa3+/Klf4- cells. WT and dCD were compared using a two-sided paired student t-test.

b.) Heatmap showing the expression of selected PGCLC genes (rows) linked to PGCLC enhancers within individual WT and dCD Prdm1orDppa3+/Klf4- cells (columns). The figures were adapted from Bleckwehl et al. 2021.

6.7.3. The loss of OTX2 can not restore germline competence in the dCD cell line

As shown in previous sections, *Otx2*$^{-/-}$ EpiLC displayed higher H3K4me1/2 levels within PGCLC enhancers and increased *in vitro* germline competence. To further evaluate the importance of H3K4me1 for *in vitro* germline competence, we decided to investigate whether the absence of H3K4me1 could also reduce the PGCLC differentiation capacity of *Otx2*$^{-/-}$ cells. To this purpose, *Otx2* was deleted in the dCD ESC (*i.e.* dCD *Otx2*$^{-/-}$) as well as in their parental WT ESC (*i.e.* R1 *Otx2*$^{-/-}$) and differentiated into PGCLC. As expected, the deletion of the *Otx2* gene in the R1 ESC increased germline competence, although not as pronouncedly as in the E14 *Otx2*$^{-/-}$ ESC line (Figure 24, Figure 43). Most importantly, the dCD *Otx2*$^{-/-}$ cells showed

reduced PGCLC differentiation capacity, comparable to the one observed for the dCD cell line. Furthermore, genome-wide and locus-specific analyses of mCpG levels revealed that Group I PGCLC enhancers were considerably more methylated in dCD and dCD *Otx2*$^{-/-}$ EpiLC than in *Otx2*$^{-/-}$ EpiLC. Thus, these results further support that H3K4me1/2, possibly by keeping PGCLC enhancers accessible and responsive to inductive signals, is required for *in vitro* germline competence.

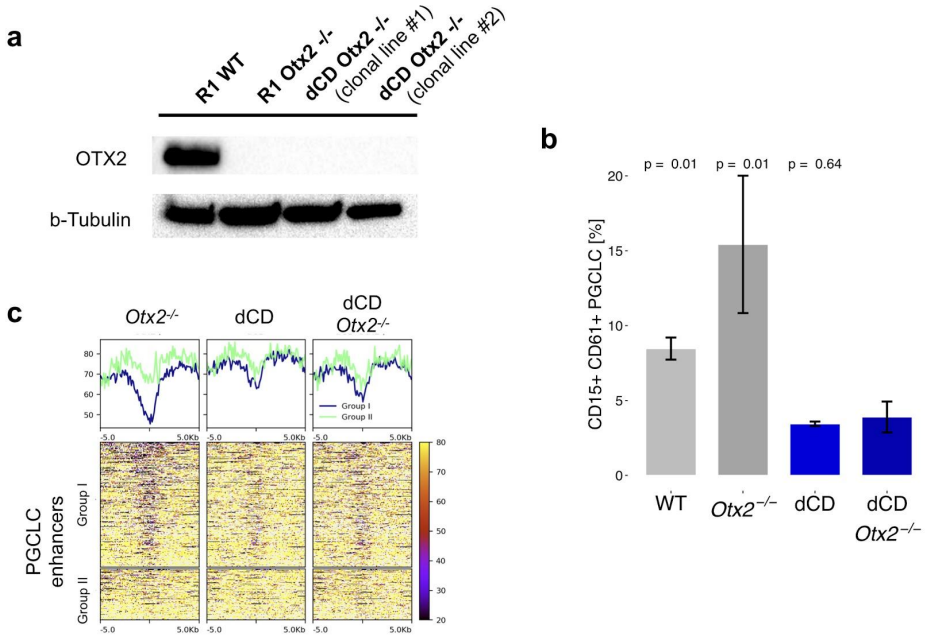


Figure 43: Quantification of the PGCLC differentiation capacity and CpG methylation levels in dCD Otx2-/- cells.

a.) Western blot analysis of the OTX2 levels in R1 Otx2$^{-/-}$ *and dCD Otx2*$^{-/-}$ *d2 EpiLC. B-Tubulin was used as a loading control.*

b.) R1 WT, R1 Otx2$^{-/-}$*, dCD and dCD Otx2*$^{-/-}$ *ESC were differentiated into PGCLC. PGCLC were quantified as the proportion of CD15*$^{+}$*CD61*$^{+}$ *cells found within d4 EB. The PGCLC differentiations for the dCD Otx2*$^{-/-}$ *ESC were performed in three biological replicates and using two different clonal lines (n=3x2). The other PGCLC measurements were performed in biological triplicates.*

c.) Genome-wide bisulfite sequencing experiments were performed in d2 EpiLC differentiated from Otx2$^{-/-}$*, dCD and dCD Otx2*$^{-/-}$ *ESC. The average profile (top) and heatmap (bottom) plots show the percentage of CpG methylation for the Group I and Group II PGCLC enhancers in WT and dCD d2 EpiLC. The figures were adapted from Bleckwehl et al. 2021.*

To get further insights into the molecular mechanisms leading to increased CpG methylation in dCD cells, the expression of the *de novo* DNA methyltransferases (*Dnmt3a/b*), DNA maintenance machinery (*Dnmt1, Uhrf1*) and the major DNA demethylase (*Tet1*) was analyzed using the RNA-seq data generated in WT and dCD EpiLC. These analyses showed that *Dnmt3a* was upregulated and *Dnmt3b* downregulated in dCD EpiLC compared to WT EpiLC (Figure 44). Furthermore, analysis of the d4 EB scRNA-seq data revealed that the *Prdm1orDppa3+/Klf4-* dCD cells failed to downregulate *Dnmt3a/b* and *Uhrf1* in comparison to their WT counterparts (Figure 42). In contrast, the expression of *Dnmt3a/b* in other cell types found within d4 EB was similar in WT and dCD cells. The initial phase of CpG demethylation during PGCLC specification involves the downregulation of DNA methyltransferases (Seki et al., 2005). Therefore, the defective expression of these

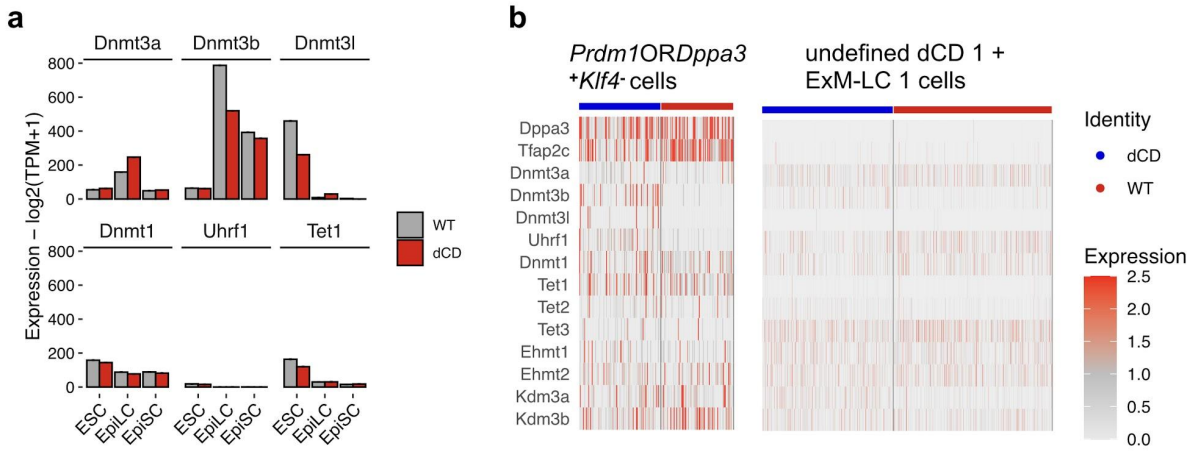


Figure 44: Expression of the DNA methyltransferases in the dCD cell line.

a.) Bar plots showing the expression, as measured by RNA-seq, of the DNA methyltransferases (Dnmt3a/b/l) and DNA methylation maintenance (Dnmt1, Uhrf1) and DNA demethylation (Tet1) in WT (gray) and dCD (red) ESC cells as well as upon their differentiation. The RNA-seq data in ESC was obtained from Dorighi et al. 2017 and the experiments in d2 EpiLC and EpiSC were performed in duplicates.

b.) Heatmap showing the expression of the PGCLC marker (Dppa3, Tfap2c), the genes involved in DNA methylation (Dnmt1, Dnmt3a/b/l, Uhrf1), DNA demethylases (Tet1/2/3) and selected H3K9me methyltransferases (Ehmt1/2) and demethylases (Kdm3a/b) within individual WT (blue column) and dCD (red column) Prdm1orDppa3+/Klf4- cell (left) or cell of the ExM-LC 1 and undefined dCD 1 clusters.

enzymes in dCD cells might contribute to their PGCLC differentiation defects. In addition, since H3K4 methylation can directly interfere with the enzymatic activity of the *de novo* DNA methyltransferases (Guo et al., 2015; Ooi et al., 2007; Zhang et al., 2010), the loss of H3K4me1 in dCD cells could also lead to the increase methylation of PGCLC enhancers through this alternative mechanism.

6.8. Relevance of the findings to understand germline competence *in vivo*

The presented results (chapter 6.1 - 6.7) show that H3K4me1/2 is required for *in vitro* germline. Thereby, the persistence of H3K4me1/2 within the PGCLC enhancers in EpiLC might protect them from rapid CpG methylation, conferring a permissive and responsive chromatin state that facilitates PGCLC enhancer reactivation and PGCLC specification *in vitro*.

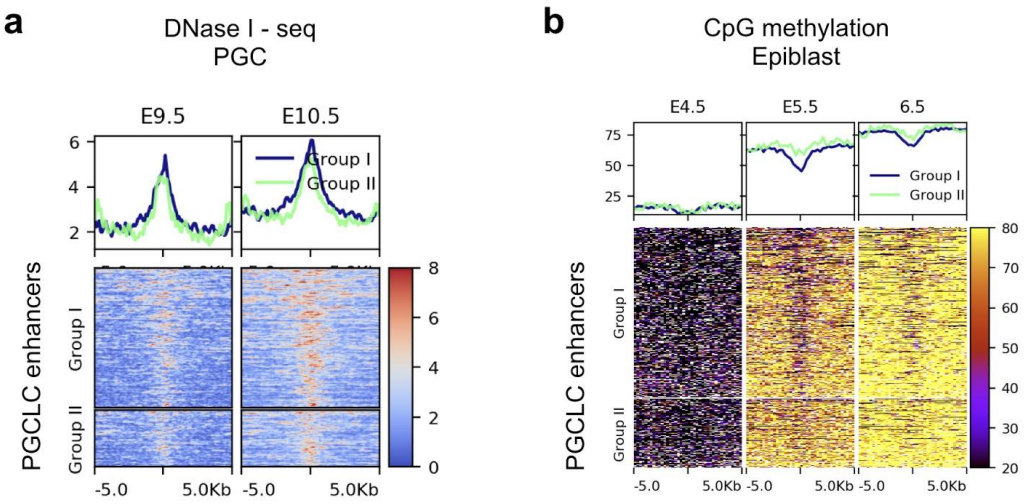


Figure 45: Evaluation of available epigenomic data from in vivo PGCs and epiblasts.

a.) Average profile (top) and heatmap (bottom) plots showing DNase-seq levels for the Group I and Group II PGCLC enhancers in PGCs isolated from E9.5 and E10.5 mouse embryos. The DNase-seq data were obtained from Li et al. 2018.

b.) Average profile (top) and heatmap (bottom) plots showing the percentage of CpG methylation for the Group I and Group II PGCLC enhancers in E4.5, E5.5 and E6.5 mouse epiblasts. The genome-wide mCpG data were obtained from Zhang et al. 2018. The figures were adapted from Bleckwehl et al. 2021.

As mentioned in the introduction, *in vivo* data sets are challenging to generate during mouse peri-implantation development, but due to recent advances on genomic technologies some interesting *in vivo* data sets are already available. Some of these datasets were analyzed to evaluate whether our findings regarding how the epigenetic state of PGCLC enhancers can influence germline competence can be relevant *in vivo*. First of all, analysis of chromatin accessibility based on DNase-seq experiments performed in E9.5 and E10.5 PGC (Li et al., 2018a) showed that our set of PGCLC enhancers is highly accessible *in vivo* (Figure 45) and, thus, probably active. Furthermore, the analysis of genome-wide CpG methylation data from mouse E4.5 - E6.5 epiblast (Zhang et al., 2018b) showed lower CpG methylation levels within Group I PGCLC enhancers in germline competent E5.5 epiblast cells than in the E6.5 epiblast, thus in agreement with our *in vitro* observations (Figure 16+45).

6.9. Epigenetic heterogeneity in the germline competent stage

A previous study of single-cell CpG methylation during mouse peri-implantation stages showed that the formative epiblast is particularly heterogeneous, especially within enhancers with low CpG content (Rulands et al., 2018). Therefore, the lower CpG methylation levels observed for some PGCLC enhancers in the E5.5 epiblast in comparison to the E6.5 epiblast could be the result of increased cell-to-cell variation. To evaluate this idea, we analyzed single-cell genomic data in which DNA methylation and gene expression were measured for the same cells across different epiblast stages (i.e. E4.5, E5.5 and E6.5) (Argelaguet et al., 2019). Due to the limited coverage of single-cell DNA methylation, the comparison of CpG methylation states is challenging. For instance it is possible, that in one cell the CpG methylation information might be covered at the center of the enhancer (more likely to be unmethylated), while in another cell the CpG methylation might be covered at the enhancer periphery (more likely to be methylated). In such cases, the comparison of the average CpG methylation levels would indicate differences between cells, although this could be caused by the differences between different parts of the enhancers rather than by cellular heterogeneity. To circumvent this problem, the CpG methylation heterogeneity was measured by comparing only the methylation states

of individual CpGs within the PGCLC enhancers using a dissimilarity matrix approach illustrated in Figure 46 (Hui et al., 2018). This ensures the direct comparison of the same covered CpGs between individual cells. Notably, this analysis revealed that the formative epiblast (E5.5) displayed the highest variation in mCpG levels within the PGCLC enhancers, while in the primed epiblast (E6.5) these enhancers were more homogeneously methylated (Figure 46).

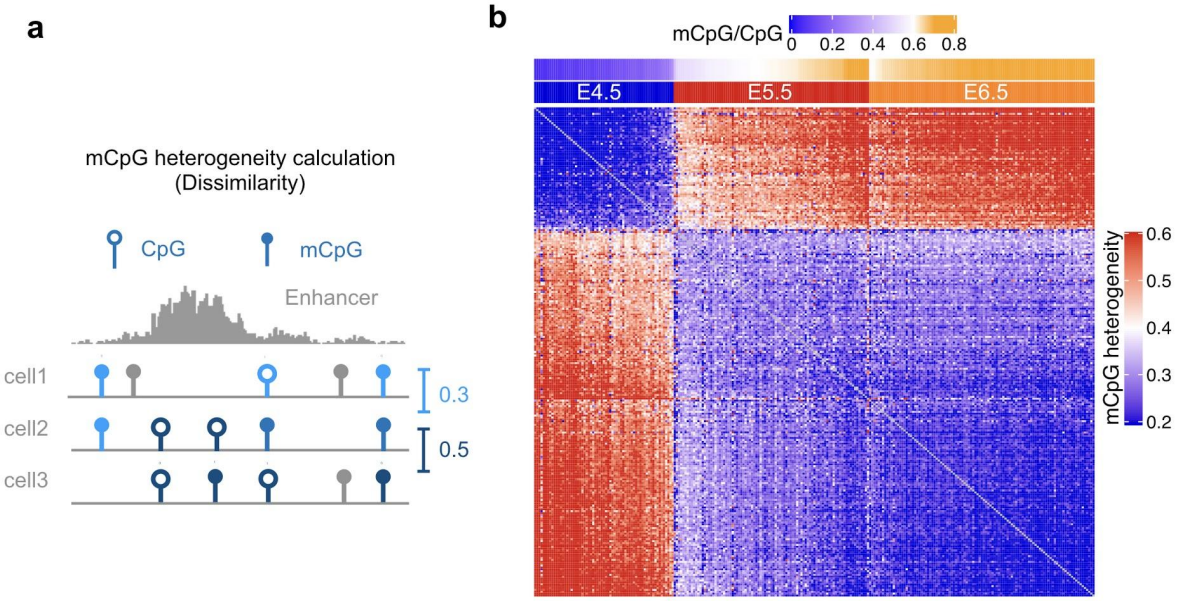


Figure 46: CpG methylation heterogeneity at PGCLC enhancers.

a.) Illustration of the strategy used to measure mCpG, which is based on the mCpG dissimilarity concept. Briefly, for each pairwise comparison, the methylation status of only those CpGs that are covered in the two cells being compared is considered (blue lollipops). Then, if two cells show the same methylation status they receive a value of 0 (similarity) and if they show dissimilar methylation patterns a value of 1. The mean of all pairwise comparisons reflect the dissimilarity of the CpG methylation, or mCpG heterogeneity.

b.) CpG methylation heterogeneity heatmap showing the differences in mCpG within PGCLC enhancers between pairs of individual cells. The CpG methylation heterogeneity values are presented with a blue-red scale (blue means that cells are similar in their methylation status, while red indicates that they are more dissimilar/heterogenous). Above the heatmap, the developmental stages of the investigated cells (E4.5, E5.5 or E6.5) and the average CpG methylation (blue-orange scale) measured for all PGCLC enhancers within each single cell are shown (n=261 cells). For each developmental stage, the cells were ranked according to the average mCpG levels within the PGCLC enhancers. The figures were adapted from Bleckwehl et al. 2021.

Interestingly, when comparing different enhancer sets across epiblast stages, the highest epigenetic heterogeneity (~30 %) was observed for the Group I PGCLC enhancers in the E5.5 epiblast (Figure 47). As the CpG coverage for the Group I PGCLC enhancers was similar across epiblast stages, the differences in mCpG heterogeneity are unlikely to be caused by technical reasons (Bleckwehl et al., 2021). Finally, the correlation of the mCpG levels within Group I PGCLC enhancers and the expression of their associated genes across the profiled epiblast cells showed that the E5.5 and E6.5 epiblast cells displayed comparable gene expression profiles yet different CpG methylation levels (Figure 47). These results suggest that the partial decommissioning of PGCLC enhancers in germline competent cells (EpiLC *in vitro* and E5.5 epiblast cells *in vivo*) might actually be the result of epigenetic heterogeneity. Furthermore, our results also suggest that during the establishment of formative pluripotency, germline competence might be acquired in an asynchronous manner and/or might not be acquired by all the epiblast cells.

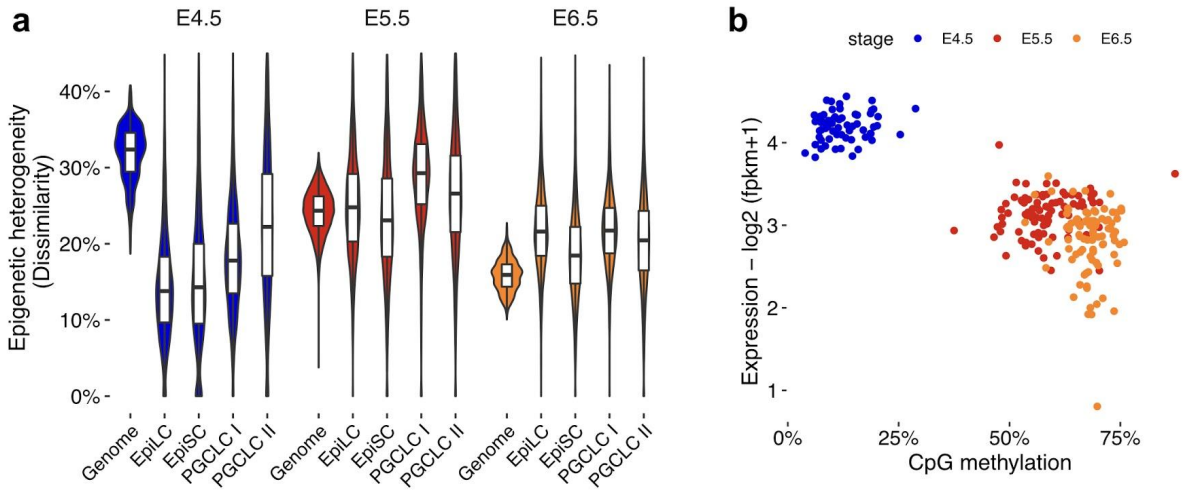


Figure 47: Single-cell analysis of the CpG methylation and expression levels of PGCLC enhancers and their associated genes.

a.) Violin plots showing the CpG methylation heterogeneity measured in E4.5, E5.5 and E6.5 epiblast cells for all covered CpGs between two cells in the mouse genome (genome-wide) as well as within EpiLC, EpiSC and PGCLC enhancers (Group I and II).

b.) Comparison of the single-cell RNA expression and CpG methylation of E4.5, E5.5 and E6.5 epiblasts. Plotted is the average expression of genes linked to Group I PGCLC enhancers and the average CpG methylation of these enhancers (n=256). The figures were adapted from Bleckwehl et al. 2021.

7. Discussion

The presented thesis augmented the knowledge of germline competence and showed that H3K4me1, which persisted at the PGCLC enhancers in EpiLC, is required for the induction of PGCLC. On the other hand, based on gene expression profiling, H3K4me1 was found to be dispensable for the establishment of different *in vitro* pluripotent states (i.e. ESC, EpiLC and EpiSC).

Here, the influence of different epigenetic modifications for germline competence and the necessity of an intermediate competent state (EpiLC *in vitro* or the formative epiblast *in vivo*) for the establishment of PGCLC and its possible mechanistic implementation will be discussed. Afterwards the findings will be discussed in the broader context of lineage competence and development. Finally, I will postulate the hypothesis that the competence of certain lineages might involve different preferences for epigenetic mechanisms that then support the lineage bifurcations towards one lineage during developmental transitions.

7.1. Epigenetic modifications and their importance for murine germline competence

The importance of H3K4me1 at enhancers has been suggested in several studies, which have shown that H3K4me1 can either precede enhancer activation in cells with higher plasticity (Battle et al., 2019; Creyghton et al., 2010; Rada-Iglesias et al., 2011; Wang et al., 2015), or can be established *de novo* during differentiation as enhancers get activated (Choukrallah et al., 2015; Geusz et al., 2020; Lara-Astiaso et al., 2014; Lee et al., 2019). Consequently, *Kmt2d* and *Kmt2c/d* KO showed severe effects in different tissues, which involved imparied activation of lineage specific enhancers (Ang et al., 2016; Jang et al., 2019; Lai et al., 2017; Lin-Shiao et al., 2018; Ortega-Molina et al., 2015; Placek et al., 2017; Wang et al., 2016; Yan et al., 2018; Zhang et al., 2015). Thus, these studies mostly concluded that H3K4me1 was important for enhancer activation. However, recent studies in mouse ESC and

Drosophila melanogaster challenged this view. The catalytic mutant *Kmt2c/d* (dCD cell line) partly reduced H3K27ac, but did not affect transcription from either enhancers or gene promoters under self-renewable conditions in ESC (Dorighi et al., 2017). Furthermore, *Drosophila melanogaster* embryos expressing catalytic deficient *Trr*, the homolog of mammalian *Kmt2c/d*, did not show any developmental effect, which would argue against any relevance of H3K4me1 at enhancers during development (Rickels et al., 2017).

However, the hyperactive version of *Trr*, characterized by hyper-H3K4me2 (Rickels et al., 2017) or the substitution of Histone H3.3K4 with alanine, which can not be methylated at all, resulted in meiosis defects and infertility rather than in severe developmental defects, suggesting that H3.3K4 methylation is largely dispensable for the development of somatic lineages in Drosophila (Hödl and Basler, 2009; Sakai et al., 2009). In contrast, H3.3K4 substitutions in mice have been shown to impair adipogenesis as well as muscle and neural progenitor development (Gehre et al., 2020; Jang et al., 2019), suggesting that this histone residue might be more functionally relevant in mammals. In agreement with this idea, a recent preprint (Xie et al., 2020) showed that *Kmt2c/d* catalytic mutant mice display embryonic lethality (~E8.5) (Xie et al., 2020).

Here, we have shown that H3K4me1/2 is important for germline competence and PGCLC induction, most likely by conferring a relevant set of enhancers with permissive and responsive chromatin features. In contrast and in agreement with the findings in ESC from Dorighi et al. 2017, H3K4me1/2 seems to be dispensable for enhancer activity and gene expression in both EpiLC and EpiSC. These differences in the importance of H3K4me1 for enhancer function might be linked to mechanistic differences during enhancer activation. Namely, while the EpiLC and EpiSC enhancers are widely pre-marked by H3K4me1 but inactive in ESC, the PGCLC enhancers are active in ESC and get partly decommissioned in EpiLC. Therefore, the subset of PGCLC enhancers in which H3K4me1 temporarily persists in EpiLC resemble the so-called latent enhancers previously described in differentiated macrophages (Ostuni et al., 2013). Following an initial round of activation and

silencing, the persistence of H3K4me1 within latent enhancers was proposed to facilitate their subsequent activation upon restimulation (Ostuni et al., 2013).

Nevertheless, it is worth mentioning that, despite the potent reduction of H3K27ac at PGCLC enhancers in dCD PGCLC, the induction of the PGCLC genes linked to these enhancers was reduced, but not completely abolished. Hence, compensatory mechanisms might exist to sustain enhancer activity in the absence of H3K4me1/2 during PGCLC induction. These findings are supported by recent preprints in which it has been shown that the absence of H3K4me1/2 only caused mild differentiation defects upon differentiation from ESC into embryoid bodies or neural progenitor cells (Kubo et al., 2021; Xie et al., 2020). It is possible that the mild expression changes observed in the dCD cells could be explained by the presence of redundant KMT2C/D-independent enhancers, which might partially compensate for the KMT2C/D-dependent enhancers, as recently reported in neural progenitor cells (Kubo et al., 2021). In addition, the KMT2C/D-dependent enhancers might lose some activity in dCD cells, but can still be functional. Hence, H3K4me1/2 might facilitate, rather than being essential for, enhancer (re)activation and the robust induction of developmental gene expression programs (Bleckwehl et al., 2021). The presented data implies that H3K4me1 is required for PGCLC enhancer function, but it cannot be completely ruled out that gene expression and epigenetic changes in ESC and/or extraembryonic-like cell types might also contribute to the PGCLC differentiation defects observed in dCD/dCT cells.

In addition to KMT2C/C-mediated H3K4me1/2, other epigenetic modifications have been studied in the context of germline development. For instance, it has been shown that the KO of another H3K4 methyltransferase (*Kmt2b*) reduced the amount of PGC(LC) (Grosswendt et al., 2020; Hu et al., 2017). Interestingly, the zygotic perturbation of the polycomb subunit *Eed* has been recently reported to drastically increase differentiation towards the PGC lineage (Grosswendt et al., 2020). In contrast, H3K27me3 did not seem to play a major role at PGCLC enhancers at any stage, but instead accumulated at PGCLC TSS in EpiLC and EpiSC (Figure 13). Although speculative, work in terminally differentiated enterocytes showed that a *Eed*

KO led to the reactivation of fetal and embryonic enhancers (Jadhav et al., 2019), which were then accessed by TFs also present at earlier developmental stages. Since many of the PGCLC enhancers also become reactivated and PGCLC genes are silenced in EpiLC, it is conceivable that the loss of H3K27me3 in the *Eed* KO cells results in the increase expression of PGCLC TF, which can then efficiently reactivate the PGCLC enhancers. Indeed, the increased expression of certain PGCLC TF in *Eed* KO EpiLC was measured in the same study by qPCR (Grosswendt et al., 2020) and it was also shown that pluripotency markers are not efficiently downregulated during EB differentiation of *Eed* KO ESC (Obier et al., 2015). Although further research about the role of PRC2 for germline competence is needed, together with the work from (Grosswendt et al., 2020), our findings indicate that H3K27me3 accumulation around the TSS of PGCLC genes might contribute to their silencing in EpiLC.

Beside the persistence of H3K4me1, we also showed that H3K9me2/3 and CpG methylation levels within PGCLC enhancers increase as *in vitro* germline competence is lost. Low levels of H3K9me2/3 are required for ESC maintenance (Loh et al., 2007) and are a hallmark of PGCs (Seki et al., 2007). However, the KO of the H3K9me2 methyltransferase *Ehmt2* did not alter the proportion of PGCs (Grosswendt et al., 2020; Zylicz et al., 2015), while *Setdb1* might be required to silence somatic genes (Mochizuki et al., 2018). The contribution of CpG methylation to germline competence is not fully understood yet. Recent studies, like the zygotic perturbation of *Dnmt1, Dnmt3a* or *Dnmt3b,* suggested that each DNA methyltransferase alone has minor contributions to PGC specification (Grosswendt et al., 2020). However, *Dnmt3a* and *Dnmt3b* might also compensate for the loss of each other (Dahlet et al., 2020). Interestingly, EB differentiation with a CpG methylation inhibitor resulted in increased expression of PGCLC genes and a general differentiation failure from ESC (Kim et al., 2020), indicating that PGCLC formation is favoured from cells with low amounts of CpG methylation. Another work showed that the knock-down of *Dnmt3l* resulted in a decrease of putative DPPA3 positive PGCLC (Neri et al., 2013). However, this study claimed an antagonism of DNMT3L and CpG methylation, which was subsequently disproven in another study

(Veland et al., 2019). Furthermore, the knockout of *Tet1*, which is the major DNA demethylase in ESC and PGC, did not alter PGCLC formation (Vincent et al., 2013), indicating that germline competence might be independent from active DNA demethylation and 5-hydroxymethylcytosine. Taking into consideration these previous reports, and our observations regarding the increase of CpG methylation at PGCLC enhancers in EpiSC and the CpG methylation heterogeneity of the formative epiblast, it is tempting to speculate that the low methylated state of some PGCLC enhancers might contribute to germline competence.

7.2. Mechanisms leading to the partial decommissioning of PGCLC enhancers

An important question that we have not answered in this work is how the H3K4me1, already present within PGCLC enhancers in ESC, partially persisted in EpiLC while other active chromatin features were lost. Notably, it was observed that the subset of PGCLC enhancers with higher levels of H3K4me1 in EpiLC than in EpiSC were more accessible and responsive to major PGC TF (*e.g.* PRDM14, NANOG, Figure 30). It was previously shown that exogenous NANOG binds different sites in ESC and EpiLC and that NANOG can induce PGCLC formation when overexpressed in EpiLC but not in ESC (Murakami et al., 2016). Hence, it seems that the PGCLC enhancers have to undergo a partial decommissioning in EpiLC before their reactivation can induce the PGCLC expression program rather than the reversion to an ESC state (Bleckwehl and Rada-Iglesias, 2019). The reason for this transition might be that it ensures the switch from the naïve expression programm, enabling pluripotency, to the PGC(LC) programm, which confers the unipotency of the germ cell lineage. This probably involves the shut down of the JAK/STAT3 signaling of ESC/ICM (Do et al., 2013). In 2i ESC, JAK/STAT3 represses, through *Klf2* or *Klf4,* the formation of a PGCLC subpopulation and diminishes germline competence upon differentiation (Hackett et al., 2017; Yeo et al., 2014). Therefore it is tempting to speculate that the enhancers controlling mostly pluripotency factors, which do get reactivated in early PGCLC, might get fully decommissioned in EpiLC. In line with

this idea, the *Klf4* enhancer showed the complete loss of active chromatin features in EpiLC compared to the persistence of H3K4me1 at *Prdm14* enhancers (Atlasi et al., 2019). Additionally, this might be also achieved by the specific repression of *Klf4* in PGCLC.

However, the mechanisms involved in the transient persistence of H3K4me1 and other permissive chromatin features in EpiLC are still unknown. In ESC, most PGCLC enhancers are initially active and occupied by a broad set of regulatory proteins that includes core pluripotency (*e.g.* POU5F1, SOX2), naïve pluripotency (*e.g.* ESRRB, PRDM14) and formative pluripotency TFs (*e.g.* FOXD3, OTX2) (Buecker et al., 2014; Hackett and Surani, 2014; Respuela et al., 2016). It is possible that, as the naïve pluripotency factors get silenced, the regulatory function of the formative pluripotency TF (Atlasi et al., 2019; Respuela et al., 2016) becomes more dominant, potentially recruiting corepressors (*e.g.* Histone deacetylases, H3K4 demethylases like KDM1A), which mediate the decommissioning of PGCLC enhancers and further silence their target genes. Yet, PGCLC enhancers remain pervasively bound by POU5F1 and p300 in EpiLC (Buecker et al., 2014; Respuela et al., 2016). Thereby, POU5F1 potentially inhibits the catalytic activity of the histone demethylase KDM1A (AlAbdi et al., 2020). Therefore, it can be envisioned that factors like POU5F1 might transiently protect the PGCLC enhancers in EpiLC from a rapid decommissioning. In addition, H3K4me1 itself can inhibit the activity of DNA methyltransferases (Guo et al., 2015; Ooi et al., 2007; Zhang et al., 2010). Consistent with this model, we showed that the persistence of active epigenetic features within PGCLC enhancers was extended in *Otx2*$^{-/-}$ EpiLC (Figure 26).

On the other hand, POU5F1 and OTX2 are associated with each other in EpiLC, which could indicate a kind of ‘tug-of-war’ mechanism (Buecker and Wysocka, 2012; Buecker et al., 2014), where both TFs might bind to PGCLC enhancers simultaneously. As mentioned above, the binding of POU5F1 to the PGCLC enhancers might counteract the repressive activity of OTX2 and other formative regulators (*e.g*. FOXD3), which might open an opportunity for the PGCLC induction before the enhancers succumb to the repressive activity. A ‘tug-of-war’ might imply

the co-existence of active and repressive features at PGCLC enhancers. The ChIP-seq followed by bisulfite sequencing showed that H3K4me1/2 enriched DNA contains fewer methylated CpG, but also that H3K4me1 and CpG are not mutually exclusive (Figure 17). Thereby, the coexistence of H3K4me1 and CpG methylation seems to occur generally at putative enhancers upon differentiation (Alajem et al., 2020). Besides the bivalency of H3K4me1 and CpG methylation, the PGCLC enhancers of *Esrrb* and *Prdm1* display co-existent H3K27ac and H3K9me2 (Zylicz et al., 2015). Hence, PGCLC enhancers might transiently display both permissive and repressive features during the germline competent state (with respect to germline competence), while the loss of the permissive features eventually closes the germline competence window.

Another interesting observation that we made in this work was that we found that CpG methylation levels, especially at PGCLC enhancers, are highly heterogeneous in the formative stage. This might reflect CpG methylation heterogeneity between single cells as well as between different PGCLC enhancers within a single cell (Figure 46). In the case of the internal PGCLC enhancer heterogeneity, it can be envisioned that a subset of permissive PGCLC enhancers are sufficient to acquire a germline competent state. If this is the case: Is there one or several alternatives of permissive PGCLC enhancers that are needed for germline competence? (Figure 48). The answer to this question remains elusive, while more is known about the cellular CpG heterogeneity. During the early stages of ESC differentiation, increased CpG methylation correlates negatively with the expression of naïve pluripotency marker (*e.g*. *Esrrb, Nr5a2, Zfp42*) (Angermueller et al., 2016; Kalkan et al., 2017). In contrast, the expression of PGCLC genes is quite homogenous in germline competent cells and hence uncoupled from the epigenetic features of the associated PGCLC enhancers. Given the asynchronous exit of the naïve pluripotent stage (Kalkan et al., 2017), the CpG methylation might increase with comparable velocities in all cells, but with a timely off-set which is then reflected by CpG methylation heterogeneity until all cells reach the primed pluripotent stage. This would mean that each cell might have an individual time-frame of competence, but at a given time-window only a fraction can be efficiently induced to PGCLC. Nevertheless, CpG

methylation heterogeneity might arise with genome-wide oscillations in this epigenetic mark (Rulands et al., 2018). Thereby, the oscillation might be also asynchronous, due to the off-set of the exit from naïve pluripotency, resulting in the observed CpG heterogeneity (Figure 48).

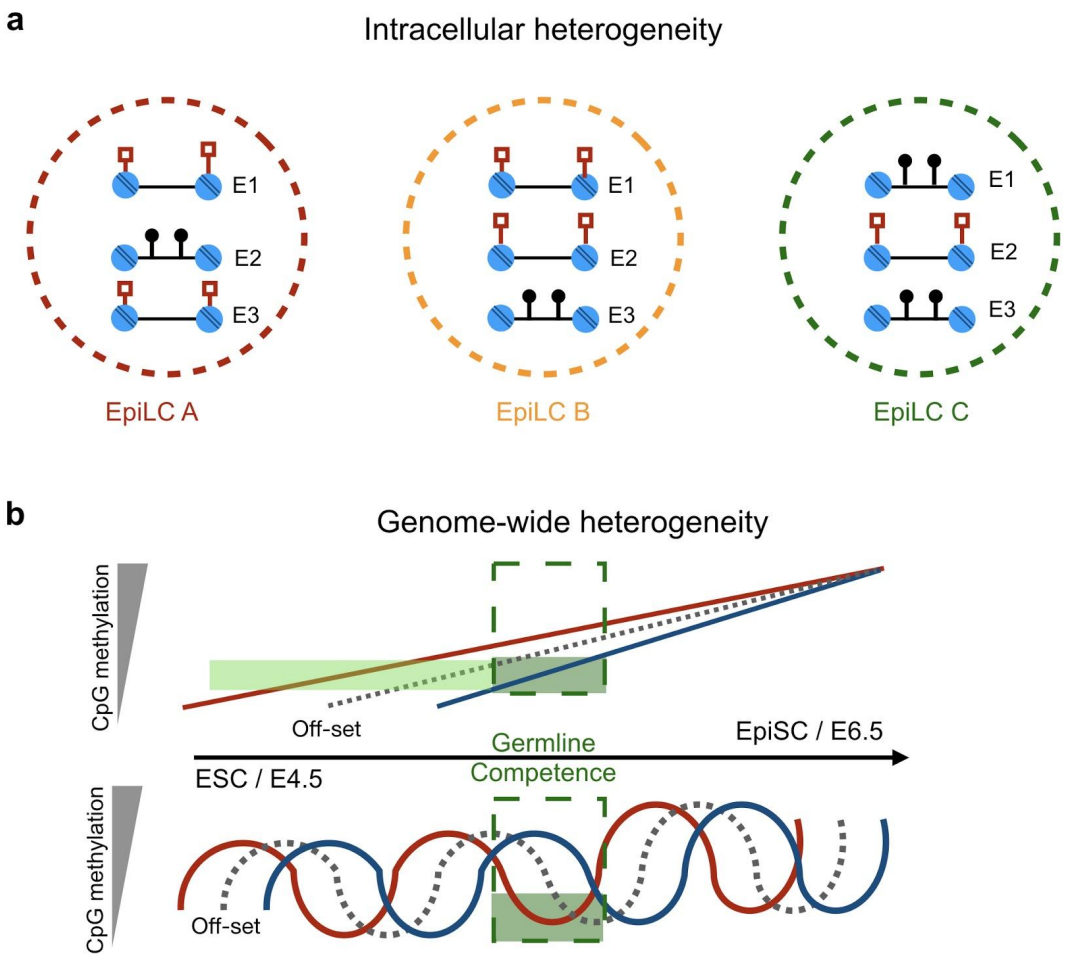


Figure 48: Illustration of epigenetic heterogeneity for germline competence.
a.) Shown are three individual cells of the germline competent phase. Within each cell three representative PGCLC enhancers possess different epigenetic states. Thereby, the representative enhancers might be taken into account as a group of similar PGCLC enhancers. Under the assumption that H3K4me1 at enhancers endows responsiveness to PGCLC induction and is negatively correlated with CpG methylation that restricts the enhancer accessibility, there are different possibilities for germline competence. i) If enhancer E1 and E2 are crucial for the PGCLC fate, only EpiLC B would be germline competent. ii.) If different combinations are possible and E3 can compensate for E1 or E2, then EpiLC A and B would be germline competent.

b.) Genome-wide CpG methylation heterogeneity (of PGCLC enhancers) with a linear or oscillatory increase of CpG methylation above or below the time beam from ESC to EpiSC, respectively. Given the unsynchronized differentiation, the red line denotes a single cell with an early exit of the ESC state and the blue lines, a single-cell with a late exit. The continuum between these two extremes is illustrated by the dotted gray line. With the assumption of a linear increase in CpG methylation over all the PGCLC enhancers, each EpiLC will undergo a germline competent state at a different time point (light green vertical line). At the time point of PGCLC induction only a fraction (which are potentially low methylated, dark green area) of EpiLC are then germline competent. The same applies to the oscillatory establishment of CpG methylation.

In summary, it is tempting to speculate that in addition to the persistence of H3K4me1, CpG methylation heterogeneity within PGCLC enhancers might also act as an important determinant of germline competence. Namely, epigenetic heterogeneity within PGCLC enhancers might increase lineage diversity by ensuring that not all formative epiblast cells are competent to enter the germline fate even when exposed to the appropriate inductive signals, but in fact many will differentiate towards the somatic lineages. This model is in agreement with previous observations showing that typically only a fraction of formative epiblast cells can be differentiated into PGCs when exposed to high inductive signals (Hayashi et al., 2011; Ohinata et al., 2009). The importance of heterogeneous cellular states for subsequent lineage bifurcations and cell fate decision has been previously reported in other lineages (Chang et al., 2008; Olsson et al., 2016; Soldatov et al., 2019).

7.3. General relevance of epigenetic modifications for lineage competence

Our work shows that the loss of H3K4me1/2 impairs PGCLC differentiation, while *in vitro* pluripotent cell types (ESC, EpiLC, EpiSC) are not affected to the same extent. Furthermore, the recent preprint from the Kai Ge lab (Xie et al., 2020) showed minor defects upon differentiation of H3K4me1/2 deficient ESC into EBs, suggesting that *in vitro* differentiation towards somatic lineages can occur in the absence of these

histone marks. Hence, during early mouse development, H3K4me1/2 might play a more important role in the specification of the germline than in the somatic lineages. This might be also true for KMT2B-mediated H3K4me3 (Grosswendt et al., 2020; Hu et al., 2017) and CpG hypomethylation (Kim et al., 2020).

In contrast to the PGCLC enhancers that we studied in this work, mesodermal and endodermal enhancers display high levels of CpG methylation and low chromatin accessibility in epiblast cells before they become active in their corresponding somatic lineage (Argelaguet et al., 2019; Yang et al., 2019b). However, these can not be considered as universal differences between germline and somatic enhancers, since ectodermal enhancers are hypomethylated and already accessible in the epiblast cells before becoming active at later developmental stages (Argelaguet et al., 2019). Furthermore, some of the ectodermal enhancers, particularly those that become active in the neural lineage, display a so-called poised state in ESC, which involves the presence of H3K27me3 and pre-formed contacts with their target genes that facilitate future gene induction (Cruz-Molina et al., 2017; Rada-Iglesias et al., 2011). Conclusively, the enhancers of different early lineages seem to be equipped with different epigenetic features before their activation.

Considering the major roles and cell-type specific activation of enhancers during development (Long et al., 2016), it seems plausible that lineage specification from a common progenitor (*e.g.* epiblast cells) might be organized by privileged epigenetic modifications for different lineage-specifying distal regulatory elements. Namely, the enhancers involved in the specification of a particular lineage could be associated with a specific combination of epigenetic features that could attract or be attracted by the TFs defining that lineage (Xiang et al., 2020). In particular, epiblast cells might possess enhancers with low CpG methylation and high H3K4 methylation that promote differentiation into PGCs, while other enhancers with low CpG methylation and high H3K27me3 favor differentiation towards neural progenitors. These different epigenetic features for different lineages might establish an epigenetic demarcation for lineage-specifying enhancers. Upon lineage induction, the corresponding lineage-specifying TF are more strongly associated with a certain epigenetic state at enhancers, thus safeguarding the induced lineage choice (Figure 49). Several

observations support this model. Firstly, the zygotic knockout of the PRC2 subunit *Eed* mentioned earlier leads to an increase in PGCs and other cell types emerging from the posterior end of the early embryo, while diminishing ectodermal derivatives (Grosswendt et al., 2020). Similar results were obtained with the KO of the PCR1 subunit *Rnf2*. Considering the requirement of PRC2 for neural differentiation (Cruz-Molina et al., 2017), it is possible that the reduced amounts of ectodermal cells in the *Eed* KO embryos cause an increase of posterior lineage identities. In agreement with this idea, a H3K27me3 hypermethylation mutant skews the differentiation towards ectodermal fates (Juan et al., 2017).

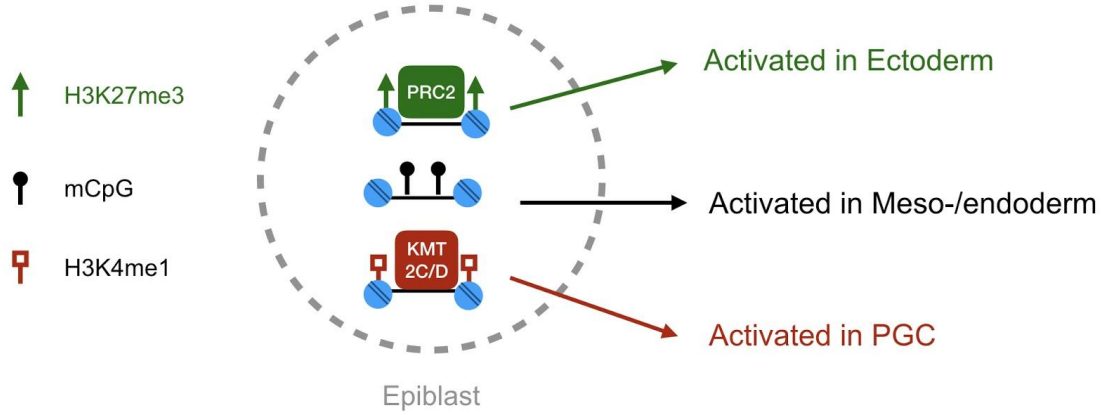


Figure 49: Epigenetic demarcation of lineage-specifying enhancers.
A pluripotent epiblast cell is primed for different lineages in a short period of time. To manage the proper induction of major lineages (Ectoderm, Meso-/endoderm, PGC), the lineage specifying enhancers might be associated with lineage specific epigenetic signatures (e.g. PRC2+H3K27me3 for ectoderm enhancer, KMT2C/D+H3K4me1 for PGC) that enable the rapid activation of the enhancers upon differentiation into the corresponding lineage.

On the other hand, the H3K4me1 levels of H3K27me3-marked poised enhancers were comparable in WT and the dCD cells (Dorighi et al., 2017; Kubo et al., 2021), and the differentiation towards neural progenitors was delayed (Kubo et al., 2021; Xie et al., 2020) but not severely affected as in PGCLC (Bleckwehl et al., 2021). Therefore, in epiblast cells, PRC2/EED bound enhancers might confer the differentiation toward neural progenitors, while KMT2C/D mediated H3K4me1 might

be second-rate for neural progenitor enhancers, but the opposite is true for PGCLC enhancers.

The epigenetic demarcation of enhancers could increase the delimitation of a particular differentiation, endowing a more secure lineage specification due to the interaction of lineage-specifying TF assemblies with a specific epigenetic environment found at the matching lineage specifying enhancers. From an evolutionary perspective, the demarcation might grant more evolutionary latitude within one mode of lineage specification. Also the epigenetic demarcation of lineage-specifying enhancers does not need to be strict, but rather affects highly relevant enhancers controlling the expression of major lineage regulators. Furthermore, this epigenetic demarcation might only exist for a transient time window, when different lineages arise from similar progenitor cells, like in the epiblast and one mode of epigenetic enhancer demarcation is not necessarily limited to the subsequent lineage specification events, but the different modes might be re-utilized, like the occurrence of H3K27me3-marked poised enhancers in early mesodermal cells (Yang et al., 2019b).

Lastly, it is worth mentioning that in addition to the epigenetic state of lineage specifying enhancers, other intrinsic features (*e.g.* metabolic regulation (Tischler et al., 2019), histone variant replacement (Buschbeck and Hake, 2017; Raja et al., 2020)) and extracellular signals (Ohinata et al., 2009; Solini et al., 2017) also contribute to lineage competence.

8. Figure Index

9. Table Index

10. References

Addicks, G.C., Brun, C.E., Sincennes, M.-C., Saber, J., Porter, C.J., Francis Stewart, A., Ernst, P., and Rudnicki, M.A. (2019). MLL1 is required for PAX7 expression and satellite cell self-renewal in mice. Nat. Commun. *10*, 4256.

AlAbdi, L., Saha, D., He, M., Dar, M.S., Utturkar, S.M., Sudyanti, P.A., McCune, S., Spears, B.H., Breedlove, J.A., Lanman, N.A., et al. (2020). Oct4-Mediated Inhibition of Lsd1 Activity Promotes the Active and Primed State of Pluripotency Enhancers. Cell Rep. *30*, 1478–1490.e6.

Alajem, A., Roth, H., Ratgauzer, S., Bavli, D., Motzik, A., Lahav, S., Peled, I., and Ram, O. (2020). DNA Methylation Patterns Expose Variations in Enhancer-Chromatin Modifications during Embryonic Stem Cell Differentiation.

Ali, M., Rincón-Arano, H., Zhao, W., Rothbart, S.B., Tong, Q., Parkhurst, S.M., Strahl, B.D., Deng, L.-W., Groudine, M., and Kutateladze, T.G. (2013). Molecular basis for chromatin binding and regulation of MLL5. Proc. Natl. Acad. Sci. U. S. A. *110*, 11296–11301.

Allan, R.S., Zueva, E., Cammas, F., Schreiber, H.A., Masson, V., Belz, G.T., Roche, D., Maison, C., Quivy, J.-P., Almouzni, G., et al. (2012). An epigenetic silencing pathway controlling T helper 2 cell lineage commitment. Nature *487*, 249–253.

Amabile, A., Migliara, A., Capasso, P., Biffi, M., Cittaro, D., Naldini, L., and Lombardo, A. (2016). Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. Cell *167*, 219–232.e14.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Nature Precedings.

Ang, S.-Y., Uebersohn, A., Spencer, C.I., Huang, Y., Lee, J.-E., Ge, K., and Bruneau, B.G. (2016). KMT2D regulates specific programs in heart development via histone H3 lysine 4 di-methylation. Development *143*, 810–821.

Angermueller, C., Clark, S.J., Lee, H.J., Macaulay, I.C., Teng, M.J., Hu, T.X., Krueger, F., Smallwood, S., Ponting, C.P., Voet, T., et al. (2016). Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nat. Methods *13*, 229–232.

Aramaki, S., Hayashi, K., Kurimoto, K., Ohta, H., Yabuta, Y., Iwanari, H., Mochizuki, Y., Hamakubo, T., Kato, Y., Shirahige, K., et al. (2013). A Mesodermal Factor, T, Specifies Mouse Germ Cell Fate by Directly Activating Germline Determinants. Developmental Cell *27*, 516–529.

Argelaguet, R., Clark, S.J., Mohammed, H., Stapel, L.C., Krueger, C., Kapourani, C.-A., Imaz-Rosshandler, I., Lohoff, T., Xiang, Y., Hanna, C.W., et al. (2019). Multi-omics profiling of mouse gastrulation at single-cell resolution. Nature *576*, 487–491.

Arner, E., Daub, C.O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drabløs, F., Lennartsson, A., Rönnerblad, M., Hrydziuszko, O., Vitezic, M., et al. (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science *347*, 1010–1014.

Ashokkumar, D., Zhang, Q., Much, C., Bledau, A.S., Fu, J., Anastassiadis, K., Francis Stewart, A., and Kranz, A. (2020). MLL4 is required for the first embryonic collective cell migration whereas MLL3 is not required until birth.

Atlasi, Y., Megchelenbrink, W., Peng, T., Habibi, E., Joshi, O., Wang, S.-Y., Wang, C., Logie, C., Poser, I., Marks, H., et al. (2019). Epigenetic modulation of a hardwired 3D chromatin landscape in two naive states of pluripotency. Nat. Cell Biol. *21*, 568–578.

Auclair, G., Borgel, J., Sanz, L.A., Vallet, J., Guibert, S., Dumas, M., Cavelier, P., Girardot, M., Forné, T., Feil, R., et al. (2016). EHMT2 directs DNA methylation for efficient gene silencing in mouse embryos. Genome Res. *26*, 192–202.

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science *354*, 909–912.

Barnes, C.E., English, D.M., and Cowley, S.M. (2019). Acetylation & Co: an expanding repertoire of histone acylations regulates chromatin and transcription. Essays Biochem. *63*, 97–107.

Battle, S.L., Doni Jayavelu, N., Azad, R.N., Hesson, J., Ahmed, F.N., Overbey, E.G., Zoller, J.A., Mathieu, J., Ruohola-Baker, H., Ware, C.B., et al. (2019). Enhancer Chromatin and 3D Genome Architecture Changes from Naive to Primed Human Embryonic Stem Cell States. Stem Cell Reports *12*, 1129–1144.

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schübeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature *520*, 243–247.

Beagrie, R.A., Scialdone, A., Schueler, M., Kraemer, D.C.A., Chotalia, M., Xie, S.Q., Barbieri, M., de Santiago, I., Lavitas, L.-M., Branco, M.R., et al. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. Nature *543*, 519–524.

Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. Trends Genet. *32*, 29–41.

Bell, E., Curry, E.W., Megchelenbrink, W., Jouneau, L., Brochard, V., Tomaz, R.A., Mau, K.H.T., Atlasi, Y., de Souza, R.A., Marks, H., et al. (2020). Dynamic CpG methylation delineates subregions within super-enhancers selectively decommissioned at the exit from naive pluripotency. Nat. Commun. *11*, 1112.

Binda, O., LeRoy, G., Bua, D.J., Garcia, B.A., Gozani, O., and Richard, S. (2010). Trimethylation of histone H3 lysine 4 impairs methylation of histone H3 lysine 9: regulation of lysine methyltransferases by physical interaction with their substrates. Epigenetics *5*, 767–775.

Bleckwehl, T., and Rada-Iglesias, A. (2019). Transcriptional and epigenetic control of germline competence and specification. Curr. Opin. Cell Biol. *61*, 1–8.

Bleckwehl, T., Crispatzu, G., Schaaf, K., Respuela, P., Bartusel, M., Benson, L., Clark, S.J., Dorighi, K.M., Barral, A., Laugsch, M., et al. (2021). Enhancer-associated H3K4 methylation safeguards in vitro germline competence.

Bochyńska, A., Lüscher-Firzlaff, J., and Lüscher, B. (2018). Modes of Interaction of KMT2 Histone H3 Lysine 4 Methyltransferase/COMPASS Complexes with Chromatin. Cells *7*.

Bolt, C.C., and Duboule, D. (2020). The regulatory landscapes of developmental genes. Development *147*.

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.-P., Tanay, A., et al. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell *171*, 557–572.e24.

Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. Nat. Cell Biol. *16*, 516–528.

Boroviak, T., Loos, R., Lombard, P., Okahara, J., Behr, R., Sasaki, E., Nichols, J., Smith, A., and Bertone, P. (2015). Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. Dev. Cell *35*, 366–382.

Bostick, M., Kim, J.K., Estève, P.-O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760–1764.

Boulard, M., Edwards, J.R., and Bestor, T.H. (2015). FBXL10 protects Polycomb-bound genes from hypermethylation. Nat. Genet. *47*, 479–485.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

Brinkman, A.B., Gu, H., Bartels, S.J.J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A., et al. (2012). Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res. *22*, 1128–1138.

Brown, D.A., Di Cerbo, V., Feldmann, A., Ahn, J., Ito, S., Blackledge, N.P., Nakayama, M., McClellan, M., Dimitrova, E., Turberfield, A.H., et al. (2017). The SET1 Complex Selects Actively Transcribed Target Genes via Multivalent Interaction with CpG Island Chromatin. Cell Rep. *20*, 2313–2327.

Buecker, C., and Wysocka, J. (2012). Enhancers as information integration hubs in development: lessons from genomics. Trends Genet. *28*, 276–284.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell *14*, 838–853.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Buschbeck, M., and Hake, S.B. (2017). Variants of core histones and their roles in cell fate decisions, development and cancer. Nat. Rev. Mol. Cell Biol. *18*, 299–314.

Calo, E., Gu, B., Bowen, M.E., Aryan, F., Zalc, A., Liang, J., Flynn, R.A., Swigut, T., Chang, H.Y., Attardi, L.D., et al. (2018). Tissue-selective effects of nucleolar stress and rDNA damage in developmental disorders. Nature *554*, 112–117.

Cao, K., Collings, C.K., Morgan, M.A., Marshall, S.A., Rendleman, E.J., Ozark, P.A., Smith, E.R., and Shilatifard, A. (2018). An Mll4/COMPASS-Lsd1 epigenetic axis governs enhancer function and pluripotency transition in embryonic stem cells. Sci Adv *4*, eaap8747.

Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature *453*, 544–547.

Charlton, J., Jung, E.J., Mattei, A.L., Bailly, N., Liao, J., Martin, E.J., Giesselmann, P., Brändl, B., Stamenova, E.K., Müller, F.-J., et al. (2020). TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. Nat. Genet. *52*, 819–827.

Chatfield, J., O'Reilly, M.-A., Bachvarova, R.F., Ferjentsik, Z., Redwood, C., Walmsley, M., Patient, R., Loose, M., and Johnson, A.D. (2014). Stochastic specification of primordial germ cells from mesoderm precursors in axolotl embryos. Development *141*, 2429–2440.

Chen, A.F., Liu, A.J., Krishnakumar, R., Freimer, J.W., DeVeale, B., and Blelloch, R. (2018a). GRHL2-Dependent Enhancer Switching Maintains a Pluripotent Stem Cell Transcriptional Subnetwork after Exit from Naive Pluripotency. Cell Stem Cell *23*, 226–238.e4.

Chen, C.-Y., McKinney, S.A., Ellington, L.R., and Gibson, M.C. (2020). Hedgehog signaling is required for endomesodermal patterning and germ cell development in the sea anemone. Elife *9*.

Chen, D., Sun, N., Hou, L., Kim, R., Faith, J., Aslanyan, M., Tao, Y., Zheng, Y., Fu, J., Liu, W., et al. (2019). Human Primordial Germ Cells Are Specified from Lineage-Primed Progenitors. Cell Rep. *29*, 4568–4582.e5.

Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J.B., and Gregor, T. (2018b). Dynamic interplay between enhancer–promoter topology and gene activity. Nature Genetics *50*, 1296–1303.

Cho, S., Park, J.S., Kwon, S., and Kang, Y.-K. (2012). Dynamics of Setdb1 expression in early mouse development. Gene Expr. Patterns *12*, 213–218.

Choukrallah, M.-A., Song, S., Rolink, A.G., Burger, L., and Matthias, P. (2015). Enhancer repertoires are reshaped independently of early priming and heterochromatin dynamics during B cell differentiation. Nat. Commun. *6*, 8324.

Clark, S.J., Smallwood, S.A., Lee, H.J., Krueger, F., Reik, W., and Kelsey, G. (2017). Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). Nat. Protoc. *12*, 534–547.

Clark, S.J., Argelaguet, R., Kapourani, C.-A., Stubbs, T.M., Lee, H.J., Alda-Catalinas, C., Krueger, F., Sanguinetti, G., Kelsey, G., Marioni, J.C., et al. (2018). scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nat. Commun. *9*, 781.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Cossec, J.-C., Theurillat, I., Chica, C., Búa Aguín, S., Gaume, X., Andrieux, A., Iturbide, A., Jouvion, G., Li, H., Bossis, G., et al. (2018). SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States. Cell Stem Cell *23*, 742–757.e8.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. *107*, 21931–21936.

Cruz-Molina, S., Respuela, P., Tebartz, C., Kolovos, P., Nikolic, M., Fueyo, R., van Ijcken, W.F.J., Grosveld, F., Frommolt, P., Bazzi, H., et al. (2017). PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. Cell Stem Cell *20*, 689–705.e9.

Dahlet, T., Argüeso Lleida, A., Al Adhami, H., Dumas, M., Bender, A., Ngondo, R.P., Tanguy, M., Vallet, J., Auclair, G., Bardet, A.F., et al. (2020). Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity. Nat. Commun. *11*, 3153.

Dancy, B.M., and Cole, P.A. (2015). Protein lysine acetylation by p300/CBP. Chem. Rev. *115*, 2419–2452.

David Allis, C., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nat. Rev. Genet. *17*, 487–500.

Dearden, P.K. (2006). Germ cell development in the Honeybee (Apis mellifera); vasa and nanos expression. BMC Dev. Biol. *6*, 6.

Denissov, S., Hofemeister, H., Marks, H., Kranz, A., Ciotta, G., Singh, S., Anastassiadis, K., Stunnenberg, H.G., and Stewart, A.F. (2014). Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. Development *141*, 526–537.

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

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Do, D.V., Ueda, J., Messerschmidt, D.M., Lorthongpanich, C., Zhou, Y., Feng, B., Guo, G., Lin, P.J., Hossain, M.Z., Zhang, W., et al. (2013). A genetic and developmental pathway from STAT3 to the OCT4-NANOG circuit is essential for maintenance of ICM lineages in vivo. Genes Dev. *27*, 1378–1390.

Dobin, A., and Gingeras, T.R. (2015). Mapping RNA-seq Reads with STAR. Curr. Protoc. Bioinformatics *51*, 11.14.1–11.14.19.

Doni Jayavelu, N., Jajodia, A., Mishra, A., and Hawkins, R.D. (2020). Candidate silencer elements for the human and mouse genomes. Nat. Commun. *11*, 1061.

Donoughe, S., Nakamura, T., Ewen-Campen, B., Green, D.A., Henderson, L., and Extavour, C.G. (2014). BMP signaling is required for the generation of primordial germ cells in an insect. Proceedings of the National Academy of Sciences *111*, 4133–4138.

Dorighi, K.M., Swigut, T., Henriques, T., Bhanu, N.V., Scruggs, B.S., Nady, N., Still, C.D., 2nd, Garcia, B.A., Adelman, K., and Wysocka, J. (2017). Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. Mol. Cell *66*, 568–576.e4.

Douillet, D., Sze, C.C., Ryan, C., Piunti, A., Shah, A.P., Ugarenko, M., Marshall, S.A., Rendleman, E.J., Zha, D., Helmin, K.A., et al. (2020). Uncoupling histone H3K4 trimethylation from developmental gene expression via an equilibrium of COMPASS, Polycomb and DNA methylation. Nat. Genet. *52*, 615–625.

Duymich, C.E., Charlet, J., Yang, X., Jones, P.A., and Liang, G. (2016). DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. Nat. Commun. *7*, 11453.

Eckersley-Maslin, M.A., Parry, A., Blotenburg, M., Krueger, C., Ito, Y., Franklin, V.N.R., Narita, M., D'Santos, C.S., and Reik, W. (2020). Epigenetic priming by Dppa2 and 4 in pluripotency facilitates multi-lineage commitment. Nat. Struct. Mol. Biol. *27*, 696–705.

Ewen-Campen, B., Jones, T.E.M., and Extavour, C.G. (2013). Evidence against a germ plasm in the milkweed bug Oncopeltus fasciatus, a hemimetabolous insect. Biol. Open *2*, 556–568.

Factor, D.C., Corradin, O., Zentner, G.E., Saiakhova, A., Song, L., Chenoweth, J.G., McKay, R.D., Crawford, G.E., Scacheri, P.C., and Tesar, P.J. (2014). Epigenomic Comparison Reveals Activation of "Seed" Enhancers during Transition from Naive to Primed Pluripotency. Cell Stem Cell *14*, 854–863.

Farini, D., La Sala, G., Tedesco, M., and De Felici, M. (2007). Chemoattractant action and molecular signaling pathways of Kit ligand on mouse primordial germ cells. Dev. Biol. *306*, 572–583.

Festuccia, N., Halbritter, F., Corsinotti, A., Gagliardi, A., Colby, D., Tomlinson, S.R., and Chambers, I. (2018). Esrrb extinction triggers dismantling of naïve pluripotency and marks commitment to differentiation. EMBO J. *37*.

Gehre, M., Bunina, D., Sidoli, S., Lübke, M.J., Diaz, N., Trovato, M., Garcia, B.A., Zaugg, J.B., and Noh, K.-M. (2020). Lysine 4 of histone H3.3 is required for embryonic stem cell differentiation, histone enrichment at regulatory regions and transcription accuracy. Nat. Genet. *52*, 273–282.

Geusz, R.J., Wang, A., Lam, D.K., Vinckier, N.K., Alysandratos, K.-D., Roberts, D.A., Wang, J., Kefalopoulou, S., Qiu, Y., Chiou, J., et al. (2020). A dual mechanism of enhancer activation by FOXA pioneer factors induces endodermal organ fates.

Gillich, A., Bao, S., Grabole, N., Hayashi, K., Trotter, M.W.B., Pasque, V., Magnúsdóttir, E., and Surani, M.A. (2012). Epiblast stem cell-based system reveals reprogramming synergy of germline factors. Cell Stem Cell *10*, 425–439.

Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tufteland, K.R., Aasland, R., Anastassiadis, K., Ang, S.-L., and Stewart, A.F. (2006). Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. Development *133*, 1423–1432.

Gretarsson, K.H., and Hackett, J.A. (2020). Dppa2 and Dppa4 counteract de novo methylation to establish a permissive epigenome for development. Nat. Struct. Mol. Biol. *27*, 706–716.

Grosswendt, S., Kretzmer, H., Smith, Z.D., Kumar, A.S., Hetzel, S., Wittler, L., Klages, S., Timmermann, B., Mukherji, S., and Meissner, A. (2020). Epigenetic regulator function through mouse gastrulation. Nature *584*, 102–108.

Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development *136*, 1063–1069.

Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature *517*, 640–644.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H.D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., et al. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell *13*, 360–369.

Hackett, J.A., and Surani, M.A. (2014). Regulatory principles of pluripotency: from the ground state up. Cell Stem Cell *15*, 416–430.

Hackett, J.A., Sengupta, R., Zylicz, J.J., Murakami, K., Lee, C., Down, T.A., and Surani, M.A. (2013). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science *339*, 448–452.

Hackett, J.A., Kobayashi, T., Dietmann, S., and Surani, M.A. (2017). Activation of Lineage Regulators and Transposable Elements across a Pluripotent Spectrum. Stem Cell Reports *8*, 1645–1658.

Hackett, J.A., Huang, Y., Günesdogan, U., Gretarsson, K.A., Kobayashi, T., and Surani, M.A. (2018). Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. Nat. Commun. *9*, 4292.

Hargan-Calvopina, J., Taylor, S., Cook, H., Hu, Z., Lee, S.A., Yen, M.-R., Chiang, Y.-S., Chen, P.-Y., and Clark, A.T. (2016). Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation. Dev. Cell *39*, 75–86.

Hayashi, K., and Saitou, M. (2013). Generation of eggs from mouse embryonic stem cells and induced pluripotent stem cells. Nat. Protoc. *8*, 1513–1524.

Hayashi, K., and Surani, M.A. (2009). Self-renewing epiblast stem cells exhibit continual delineation of germ cells with epigenetic reprogramming in vitro. Development *136*, 3549–3556.

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells. Cell *146*, 519–532.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Hnisz, D., Schuijers, J., Lin, C.Y., Weintraub, A.S., Abraham, B.J., Lee, T.I., Bradner, J.E., and Young, R.A. (2015). Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. Mol. Cell *58*, 362–370.

Hödl, M., and Basler, K. (2009). Transcription in the Absence of Histone H3.3. Curr. Biol. *19*, 1221–1226.

Hon, G.C., Song, C.-X., Du, T., Jin, F., Selvaraj, S., Lee, A.Y., Yen, C.-A., Ye, Z., Mao, S.-Q., Wang, B.-A., et al. (2014). 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. Mol. Cell *56*, 286–297.

Hopf, C., Viebahn, C., and Püschel, B. (2011). BMP signals and the transcriptional repressor BLIMP1 during germline segregation in the mammalian embryo. Dev. Genes Evol. *221*, 209–223.

Hu, D., Garruss, A.S., Gao, X., Morgan, M.A., Cook, M., Smith, E.R., and Shilatifard, A. (2013a). The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1093–1097.

Hu, D., Gao, X., Morgan, M.A., Herz, H.-M., Smith, E.R., and Shilatifard, A. (2013b). The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Mol. Cell. Biol. *33*, 4745–4754.

Hu, D., Gao, X., Cao, K., Morgan, M.A., Mas, G., Smith, E.R., Volk, A.G., Bartom, E.T., Crispino, J.D., Di Croce, L., et al. (2017). Not All H3K4 Methylations Are Created Equal: Mll2/COMPASS Dependency in Primordial Germ Cell Specification. Molecular Cell *65*, 460–475.e6.

Huang, Y., Osorno, R., Tsakiridis, A., and Wilson, V. (2012). In Vivo differentiation potential of epiblast stem cells revealed by chimeric embryo formation. Cell Rep. *2*, 1571–1578.

Hughes, A.L., Kelley, J.R., and Klose, R.J. (2020). Understanding the interplay between CpG island-associated gene promoters and H3K4 methylation. Biochim. Biophys. Acta Gene Regul. Mech. *1863*, 194567.

Hui, T., Cao, Q., Wegrzyn-Woltosz, J., O'Neill, K., Hammond, C.A., David J H, Laks, E., Moksa, M., Aparicio, S., Eaves, C.J., et al. (2018). High-Resolution Single-Cell DNA Methylation Measurements Reveal Epigenetically Distinct Hematopoietic Stem Cell Subpopulations. Stem Cell Reports *11*, 578–592.

Ibarra-Soria, X., Jawaid, W., Pijuan-Sala, B., Ladopoulos, V., Scialdone, A., Jörg, D.J., Tyser, R.C.V., Calero-Nieto, F.J., Mulas, C., Nichols, J., et al. (2018). Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation. Nat. Cell Biol. *20*, 127–134.

Ibragimov, A.N., Bylino, O.V., and Shidlovskii, Y.V. (2020). Molecular Basis of the Function of Transcriptional Enhancers. Cells *9*.

Irie, N., Tang, W.W.C., and Azim Surani, M. (2014). Germ cell specification and pluripotency in mammals: a perspective from early embryogenesis. Reprod. Med. Biol. *13*, 203–215.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129–1133.

Izzo, F., Lee, S.C., Poran, A., Chaligne, R., Gaiti, F., Gross, B., Murali, R.R., Deochand, S.D., Ang, C., Jones, P.W., et al. (2020). DNA methylation disruption reshapes the hematopoietic differentiation landscape. Nat. Genet. *52*, 378–387.

Jadhav, U., Cavazza, A., Banerjee, K.K., Xie, H., O'Neill, N.K., Saenz-Vash, V., Herbert, Z., Madha, S., Orkin, S.H., Zhai, H., et al. (2019). Extensive Recovery of Embryonic Enhancer and Gene Memory Stored in Hypomethylated Enhancer DNA. Mol. Cell *74*, 542–554.e5.

Jang, Y., Broun, A., Wang, C., Park, Y.-K., Zhuang, L., Lee, J.-E., Froimchuk, E., Liu, C., and Ge, K. (2019). H3.3K4M destabilizes enhancer H3K4 methyltransferases MLL3/MLL4 and impairs adipose tissue development. Nucleic Acids Res. *47*, 607–620.

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

Jin, Q., Yu, L.-R., Wang, L., Zhang, Z., Kasper, L.H., Lee, J.-E., Wang, C., Brindle, P.K., Dent, S.Y.R., and Ge, K. (2011). Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. EMBO J. *30*, 249–262.

Johnson, A.D., and Alberio, R. (2015). Primordial germ cells: the first cell lineage or the last cells standing? Development *142*, 2730–2739.

Juan, A.H., Wang, S., Ko, K.D., Zare, H., Tsai, P.-F., Feng, X., Vivanco, K.O., Ascoli, A.M., Gutierrez-Cruz, G., Krebs, J., et al. (2017). Roles of H3K27me2 and H3K27me3 Examined during Fate Specification of Embryonic Stem Cells. Cell Rep. *18*, 297.

Juliano, C.E., Swartz, S.Z., and Wessel, G.M. (2010). A conserved germline multipotency program. Development *137*, 4113–4126.

Jurkowska, R.Z., Qin, S., Kungulovski, G., Tempel, W., Liu, Y., Bashtrykov, P., Stiefelmaier, J., Jurkowski, T.P., Kudithipudi, S., Weirich, S., et al. (2017). H3K14ac is linked to methylation of H3K9 by the triple Tudor domain of SETDB1. Nat. Commun. *8*, 2057.

Kalkan, T., Olova, N., Roode, M., Mulas, C., Lee, H.J., Nett, I., Marks, H., Walker, R., Stunnenberg, H.G., Lilley, K.S., et al. (2017). Tracking the embryonic stem cell transition from ground state pluripotency. Development *144*, 1221–1234.

Kalkan, T., Bornelöv, S., Mulas, C., Diamanti, E., Lohoff, T., Ralser, M., Middelkamp, S., Lombard, P., Nichols, J., and Smith, A. (2019). Complementary Activity of ETV5, RBPJ, and TCF3 Drives Formative Transition from Naive Pluripotency. Cell Stem Cell *24*, 785–801.e7.

Kang, Y., Kim, Y.W., Kang, J., and Kim, A. (2021). Histone H3K4me1 and H3K27ac play roles in nucleosome depletion and eRNA transcription, respectively, at enhancers.

Karmodiya, K., Krebs, A.R., Oulad-Abdelghani, M., Kimura, H., and Tora, L. (2012). H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. BMC Genomics *13*, 424.

Kasowitz, S.D., Luo, M., Ma, J., Leu, N.A., and Wang, P.J. (2017). Embryonic lethality and defective male germ cell development in mice lacking UTF1. Sci. Rep. *7*, 17259.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Kim, I.S., Wu, J., Rahme, G.J., Battaglia, S., Dixit, A., Gaskell, E., Chen, H., Pinello, L., and Bernstein, B.E. (2020). Parallel Single-Cell RNA-Seq and Genetic Recording Reveals Lineage Decisions in Developing Embryoid Bodies. Cell Rep. *33*, 108222.

Kim, J.H., Rege, M., Valeri, J., Dunagin, M.C., Metzger, A., Titus, K.R., Gilgenast, T.G., Gong, W., Beagan, J.A., Raj, A., et al. (2019). LADL: light-activated dynamic looping for endogenous gene expression control. Nat. Methods *16*, 633–639.

Kinoshita, M., and Smith, A. (2018). Pluripotency Deconstructed. Dev. Growth Differ. *60*, 44–52.

Kobayashi, T., Zhang, H., Tang, W.W.C., Irie, N., Withey, S., Klisch, D., Sybirna, A., Dietmann, S., Contreras, D.A., Webb, R., et al. (2017). Principles of early human development and germ cell program from conserved model systems. Nature *546*, 416–420.

Kobayashi, T., Kobayashi, H., Goto, T., Takashima, T., Oikawa, M., Ikeda, H., Terada, R., Yoshida, F., Sanbo, M., Nakauchi, H., et al. (2020). Germline development in rat revealed by visualization and deletion of. Development *147*.

Kojima, Y., Kaufman-Francis, K., Studdert, J.B., Steiner, K.A., Power, M.D., Loebel, D.A.F., Jones, V., Hor, A., de Alencastro, G., Logan, G.J., et al. (2014). The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. Cell Stem Cell *14*, 107–120.

Kremsky, I., and Corces, V.G. (2020). Protection from DNA re-methylation by transcription factors in primordial germ cells and pre-implantation embryos can explain trans-generational epigenetic inheritance. Genome Biol. *21*, 118.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.

Kubo, N., Hu, R., Ye, Z., and Ren, B. (2021). MLL3/MLL4 Histone Methyltranferase Activity Dependent Chromatin Organization at Enhancers during Embryonic Stem Cell Differentiation.

Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K., and Saitou, M. (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev. *22*, 1617–1635.

Kurimoto, K., Yabuta, Y., Hayashi, K., Ohta, H., Kiyonari, H., Mitani, T., Moritoki, Y., Kohri, K., Kimura, H., Yamamoto, T., et al. (2015). Quantitative Dynamics of Chromatin Remodeling during Germ Cell Specification from Mouse Embryonic Stem Cells. Cell Stem Cell *16*, 517–532.

Lai, B., Lee, J.-E., Jang, Y., Wang, L., Peng, W., and Ge, K. (2017). MLL3/MLL4 are required for CBP/p300 binding on enhancers and super-enhancer formation in brown adipogenesis. Nucleic Acids Res. *45*, 6388–6403.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., et al. (2014). Immunogenetics. Chromatin state dynamics during blood formation. Science *345*, 943–949.

Larke, M.S.C., Schwessinger, R., Nojima, T., Telenius, J., Beagrie, R.A., Downes, D.J., Oudelaar, A.M., Truch, J., Graham, B., Bender, M.A., et al. (2021). Enhancers predominantly regulate gene expression during differentiation via transcription initiation. Mol. Cell.

Larose, H., Shami, A.N., Abbott, H., Manske, G., Lei, L., and Hammoud, S.S. (2019). Gametogenesis: A journey from inception to conception. Curr. Top. Dev. Biol. *132*, 257–310.

Laugesen, A., Højfeldt, J.W., and Helin, K. (2019). Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. Mol. Cell *74*, 8–18.

Laugsch, M., Bartusel, M., Rehimi, R., Alirzayeva, H., Karaolidou, A., Crispatzu, G., Zentis, P., Nikolic, M., Bleckwehl, T., Kolovos, P., et al. (2019). Modeling the Pathological Long-Range Regulatory Effects of Human Structural Variation with Patient-Specific hiPSCs. Cell Stem Cell *24*, 736–752.e12.

Lee, K.K., and Workman, J.L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. Nat. Rev. Mol. Cell Biol. *8*, 284–295.

Lee, C.-H., Holder, M., Grau, D., Saldaña-Meyer, R., Yu, J.-R., Ganai, R.A., Zhang, J., Wang, M., LeRoy, G., Dobenecker, M.-W., et al. (2018). Distinct Stimulatory Mechanisms Regulate the Catalytic Activity of Polycomb Repressive Complex 2. Mol. Cell *70*, 435–448.e5.

Lee, J.-E., Wang, C., Xu, S., Cho, Y.-W., Wang, L., Feng, X., Baldridge, A., Sartorelli, V., Zhuang, L., Peng, W., et al. (2013). H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. Elife *2*, e01503.

Lee, K., Cho, H., Rickert, R.W., Li, Q.V., Pulecio, J., Leslie, C.S., and Huangfu, D. (2019). FOXA2 Is Required for Enhancer Priming during Pancreatic Differentiation. Cell Rep. *28*, 382–393.e7.

Leemans, C., van der Zwalm, M.C.H., Brueckner, L., Comoglio, F., van Schaik, T., Pagie, L., van Arensbergen, J., and van Steensel, B. (2019). Promoter-Intrinsic and Local Chromatin Features Determine Gene Repression in LADs. Cell *177*, 852–864.e14.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, J., Shen, S., Chen, J., Liu, W., Li, X., Zhu, Q., Wang, B., Chen, X., Wu, L., Wang, M., et al. (2018a). Accurate annotation of accessible chromatin in mouse and human primordial germ cells. Cell Res. *28*, 1077–1089.

Li, Y., Zhang, Z., Chen, J., Liu, W., Lai, W., Liu, B., Li, X., Liu, L., Xu, S., Dong, Q., et al. (2018b). Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. Nature *564*, 136–140.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Lin-Shiao, E., Lan, Y., Coradin, M., Anderson, A., Donahue, G., Simpson, C.L., Sen, P., Saffie, R., Busino, L., Garcia, B.A., et al. (2018). KMT2D regulates p63 target enhancers to coordinate epithelial homeostasis. Genes Dev. *32*, 181–193.

Liu, N., Zhang, Z., Wu, H., Jiang, Y., Meng, L., Xiong, J., Zhao, Z., Zhou, X., Li, J., Li, H., et al. (2015). Recognition of H3K9 methylation by GLP is required for efficient establishment of H3K9 methylation, rapid target gene repression, and mouse viability. Genes Dev. *29*, 379–393.

Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. Cell *167*, 233–247.e17.

Local, A., Huang, H., Albuquerque, C.P., Singh, N., Lee, A.Y., Wang, W., Wang, C., Hsia, J.E., Shiau, A.K., Ge, K., et al. (2018). Identification of H3K4me1-associated proteins at mammalian enhancers. Nat. Genet. *50*, 73–82.

Loh, Y.-H., Zhang, W., Chen, X., George, J., and Ng, H.-H. (2007). Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. Genes Dev. *21*, 2545–2557.

Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell *167*, 1170–1187.

Lubitz, S., Glaser, S., Schaft, J., Stewart, A.F., and Anastassiadis, K. (2007). Increased apoptosis and skewed differentiation in mouse embryonic stem cells lacking the histone methyltransferase Mll2. Mol. Biol. Cell *18*, 2356–2366.

Ma, H., Zhai, J., Wan, H., Jiang, X., Wang, X., Wang, L., Xiang, Y., He, X., Zhao, Z.-A., Zhao, B., et al. (2019). In vitro culture of cynomolgus monkey embryos beyond early gastrulation. Science *366*.

Ma, Z., Swigut, T., Valouev, A., Rada-Iglesias, A., and Wysocka, J. (2011). Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. Nat. Struct. Mol. Biol. *18*, 120–127.

Magnúsdóttir, E., Dietmann, S., Murakami, K., Günesdogan, U., Tang, F., Bao, S., Diamanti, E., Lao, K., Gottgens, B., and Azim Surani, M. (2013). A tripartite transcription factor network regulates primordial germ cell specification in mice. Nat. Cell Biol. *15*, 905–915.

Maison, C., Bailly, D., Quivy, J.-P., and Almouzni, G. (2016). The methyltransferase Suv39h1 links the SUMO pathway to HP1α marking at pericentric heterochromatin. Nat. Commun. *7*, 12224.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10.

Masaki, H., Kato-Itoh, M., Takahashi, Y., Umino, A., Sato, H., Ito, K., Yanagida, A., Nishimura, T., Yamaguchi, T., Hirabayashi, M., et al. (2016). Inhibition of Apoptosis Overcomes Stage-Related Compatibility Barriers to Chimera Formation in Mouse Embryos. Cell Stem Cell *19*, 587–592.

Matsumura, Y., Nakaki, R., Inagaki, T., Yoshida, A., Kano, Y., Kimura, H., Tanaka, T., Tsutsumi, S., Nakao, M., Doi, T., et al. (2015). H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation. Mol. Cell *60*, 584–596.

Mayer, D., Stadler, M.B., Rittirsch, M., Hess, D., Lukonin, I., Winzi, M., Smith, A., Buchholz, F., and Betschinger, J. (2020). Zfp281 orchestrates interconversion of pluripotent states by engaging Ehmt1 and Zic2. EMBO J. *39*, e102591.

McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. J. Open Source Softw. *3*, 861.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nature Biotechnology *28*, 495–501.

Melsted, P., Ntranos, V., and Pachter, L. (2019). The barcode, UMI, set format and BUStools. Bioinformatics *35*, 4472–4473.

Meng, T.-G., Zhou, Q., Ma, X.-S., Liu, X.-Y., Meng, Q.-R., Huang, X.-J., Liu, H.-L., Lei, W.-L., Zhao, Z.-H., Ouyang, Y.-C., et al. (2020). PRC2 and EHMT1 regulate H3K27me2 and H3K27me3 establishment across the zygote genome. Nat. Commun. *11*, 6354.

von Meyenn, F., Berrens, R.V., Andrews, S., Santos, F., Collier, A.J., Krueger, F., Osorno, R., Dean, W., Rugg-Gunn, P.J., and Reik, W. (2016). Comparative Principles of DNA Methylation Reprogramming during Human and Mouse In Vitro Primordial Germ Cell Specification. Dev. Cell *39*, 104–115.

van Mierlo, G., Veenstra, G.J.C., Vermeulen, M., and Marks, H. (2019). The Complexity of PRC2 Subcomplexes. Trends Cell Biol. *29*, 660–671.

Mishra, B.P., Zaffuto, K.M., Artinger, E.L., Org, T., Mikkola, H.K.A., Cheng, C., Djabali, M., and Ernst, P. (2014). The histone methyltransferase activity of MLL1 is dispensable for hematopoiesis and leukemogenesis. Cell Rep. *7*, 1239–1247.

Miura, H., Takahashi, S., Poonperm, R., Tanigawa, A., Takebayashi, S.-I., and Hiratani, I. (2019). Single-cell DNA replication profiling identifies spatiotemporal developmental dynamics of chromosome organization. Nat. Genet. *51*, 1356–1368.

Mochizuki, K., Tando, Y., Sekinaka, T., Otsuka, K., Hayashi, Y., Kobayashi, H., Kamio, A., Ito-Matsuoka, Y., Takehara, A., Kono, T., et al. (2018). SETDB1 is essential for mouse primordial germ cell fate determination by ensuring BMP signaling. Development *145*.

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., et al. (2017). Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. Cell Rep. *20*, 1215–1228.

Morgan, M.A.J., Rickels, R.A., Collings, C.K., He, X., Cao, K., Herz, H.-M., Cozzolino, K.A., Abshiru, N.A., Marshall, S.A., Rendleman, E.J., et al. (2017a). A cryptic Tudor domain links BRWD2/PHIP to COMPASS-mediated histone H3K4 methylation. Genes Dev. *31*, 2003–2014.

Morgan, S.L., Mariano, N.C., Bermudez, A., Arruda, N.L., Wu, F., Luo, Y., Shankar, G., Jia, L., Chen, H., Hu, J.-F., et al. (2017b). Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. Nat. Commun. *8*, 15993.

Morgani, S., Nichols, J., and Hadjantonakis, A.-K. (2017). The many faces of Pluripotency: in vitro adaptations of a continuum of in vivo states. BMC Dev. Biol. *17*, 7.

Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., Margueron, R., and Ait-Si-Ali, S. (2014). The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. Mol. Cell *53*, 277–289.

Mozzetta, C., Boyarchuk, E., Pontis, J., and Ait-Si-Ali, S. (2015). Sound of silence: the properties and functions of repressive Lys methyltransferases. Nat. Rev. Mol. Cell Biol. *16*, 499–513.

Mulas, C., Kalkan, T., and Smith, A. (2017). NODAL Secures Pluripotency upon Embryonic Stem Cell Progression from the Ground State. Stem Cell Reports *9*, 77–91.

Mulholland, C.B., Nishiyama, A., Ryan, J., Nakamura, R., Yiğit, M., Glück, I.M., Trummer, C., Qin, W., Bartoschek, M.D., Traube, F.R., et al. (2020). Author Correction: Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive DNA demethylation in mammals. Nat. Commun. *11*, 6443.

Murakami, K., Günesdogan, U., Zylicz, J.J., Tang, W.W.C., Sengupta, R., Kobayashi, T., Kim, S., Butler, R., Dietmann, S., and Azim Surani, M. (2016a). NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers. Nature *529*, 403–407.

Muramatsu, D., Kimura, H., Kotoshiba, K., Tachibana, M., and Shinkai, Y. (2016). Pericentric H3K9me3 Formation by HP1 Interaction-defective Histone Methyltransferase Suv39h1. Cell Struct. Funct. *41*, 145–152.

Nakaki, F., Hayashi, K., Ohta, H., Kurimoto, K., Yabuta, Y., and Saitou, M. (2013). Induction of mouse germ-cell fate by transcription factors in vitro. Nature *501*, 222–226.

Nakamura, A., and Seydoux, G. (2008). Less is more: specification of the germline by transcriptional repression. Development *135*, 3817–3827.

Nakamura, T., and Extavour, C.G. (2016). The transcriptional repressor Blimp-1 acts downstream of BMP signaling to generate primordial germ cells in the cricket Gryllus bimaculatus. Development *143*, 255–263.

Nakamura, A., Shirae-Kurabayashi, M., and Hanyu-Nakamura, K. (2010). Repression of early zygotic transcription in the germline. Curr. Opin. Cell Biol. *22*, 709–714.

Neri, F., Krepelova, A., Incarnato, D., Maldotti, M., Parlato, C., Galvagni, F., Matarese, F., Stunnenberg, H.G., and Oliviero, S. (2013). Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. Cell *155*, 121–134.

Ngan, C.Y., Wong, C.H., Tjong, H., Wang, W., Goldfeder, R.L., Choi, C., He, H., Gong, L., Lin, J., Urban, B., et al. (2020). Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development. Nat. Genet. *52*, 264–272.

Nicetto, D., Donahue, G., Jain, T., Peng, T., Sidoli, S., Sheng, L., Montavon, T., Becker, J.S., Grindheim, J.M., Blahnik, K., et al. (2019). H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. Science *363*, 294–297.

Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell Stem Cell *4*, 487–492.

Ninova, M., Fejes Tóth, K., and Aravin, A.A. (2019). The control of gene expression and cell identity by H3K9 trimethylation. Development *146*.

Niu, Y., Sun, N., Li, C., Lei, Y., Huang, Z., Wu, J., Si, C., Dai, X., Liu, C., Wei, J., et al. (2019). Dissecting primate early post-implantation development using long-term in vitro embryo culture. Science *366*.

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.

Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930–944.e22.

Novo, C.L., Javierre, B.-M., Cairns, J., Segonds-Pichon, A., Wingett, S.W., Freire-Pritchett, P., Furlan-Magaril, M., Schoenfelder, S., Fraser, P., and Rugg-Gunn, P.J. (2018). Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition. Cell Rep. *22*, 2615–2627.

Obier, N., Lin, Q., Cauchy, P., Hornich, V., Zenke, M., Becker, M., and Müller, A.M. (2015). Polycomb protein EED is required for silencing of pluripotency genes upon ESC differentiation. Stem Cell Rev Rep *11*, 50–61.

Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S.C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A., et al. (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. Nature *436*, 207–213.

Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., and Saitou, M. (2009). A signaling principle for the specification of the germ cell lineage in mice. Cell *137*, 571–584.

Ohnishi, Y., Huber, W., Tsumura, A., Kang, M., Xenopoulos, P., Kurimoto, K., Oleś, A.K., Araúzo-Bravo, M.J., Saitou, M., Hadjantonakis, A.-K., et al. (2014). Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. Nat. Cell Biol. *16*, 27–37.

Okamura, E., Tam, O.H., Posfai, E., Li, L., Cockburn, K., Lee, C.Q.E., Garner, J., and Rossant, J. (2019). Esrrb function is required for proper primordial germ cell development in presomite stage mouse embryos. Dev. Biol. *455*, 382–392.

Olsson, A., Venkatasubramanian, M., Chaudhri, V.K., Aronow, B.J., Salomonis, N., Singh, H., and Grimes, H.L. (2016). Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. Nature *537*, 698–702.

Onufriev, A.V., and Schiessel, H. (2019). The nucleosome: from structure to function through physics. Curr. Opin. Struct. Biol. *56*, 119–130.

Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714–717.

Ortega-Molina, A., Boss, I.W., Canela, A., Pan, H., Jiang, Y., Zhao, C., Jiang, M., Hu, D., Agirre, X., Niesvizky, I., et al. (2015). The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. Nat. Med. *21*, 1199–1208.

Osorno, R., Tsakiridis, A., Wong, F., Cambray, N., Economou, C., Wilkie, R., Blin, G., Scotting, P.J., Chambers, I., and Wilson, V. (2012). The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. Development *139*, 2288–2298.

Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B.J., Afzal, S.Y., Lee, E.A., Zhu, Y., Plajzer-Frick, I., Pickle, C.S., et al. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. Nature *554*, 239–243.

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). Latent enhancers activated by stimulation in differentiated cells. Cell *152*, 157–171.

Pachano, T., Sánchez-Gaya, V., Mariner-Faulí, M., Ealo, T., Asenjo, H.G., Respuela, P., Cruz-Molina, S., van Ijcken, W.F.J., Landeira, D., and Rada-Iglesias, Á. (2020). Orphan CpG islands boost the regulatory activity of poised enhancers and dictate the responsiveness of their target genes.

Pang, B., and Snyder, M.P. (2020). Systematic identification of silencers in human cells. Nat. Genet. *52*, 254–263.

Peng, G., Suo, S., Chen, J., Chen, W., Liu, C., Yu, F., Wang, R., Chen, S., Sun, N., Cui, G., et al. (2016). Spatial Transcriptome for the Molecular Annotation of Lineage Fates and Cell Identity in Mid-gastrula Mouse Embryo. Dev. Cell *36*, 681–697.

Pepenella, S., Murphy, K.J., and Hayes, J.J. (2014). Intra- and inter-nucleosome interactions of the core histone tail domains in higher-order chromatin structure. Chromosoma *123*, 3–13.

Placek, K., Hu, G., Cui, K., Zhang, D., Ding, Y., Lee, J.-E., Jang, Y., Wang, C., Konkel, J.E., Song, J., et al. (2017). MLL4 prepares the enhancer landscape for Foxp3 induction via chromatin looping. Nat. Immunol. *18*, 1035–1045.

Poepsel, S., Kasinath, V., and Nogales, E. (2018). Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. Nat. Struct. Mol. Biol. *25*, 154–162.

Poleshko, A., Shah, P.P., Gupta, M., Babu, A., Morley, M.P., Manderfield, L.J., Ifkovits, J.L., Calderon, D., Aghajanian, H., Sierra-Pagán, J.E., et al. (2017). Genome-Nuclear Lamina Interactions Regulate Cardiac Stem Cell Lineage Restriction. Cell *171*, 573–587.e14.

Popovic, M., Bialecka, M., Gomes Fernandes, M., Taelman, J., Van Der Jeught, M., De Sutter, P., Heindryckx, B., and Chuva De Sousa Lopes, S.M. (2019). Human blastocyst outgrowths recapitulate primordial germ cell specification events. Mol. Hum. Reprod. *25*, 519–526.

Pradeepa, M.M., Grimes, G.R., Kumar, Y., Olley, G., Taylor, G.C.A., Schneider, R., and Bickmore, W.A. (2016). Histone H3 globular domain acetylation identifies a new class of enhancers. Nat. Genet. *48*, 681–686.

Quan, H., and Lynch, J.A. (2016). The evolution of insect germline specification strategies. Curr Opin Insect Sci *13*, 99–105.

Quan, H., Arsala, D., and Lynch, J.A. (2019). Transcriptomic and functional analysis of the oosome, a unique form of germ plasm in the wasp Nasonia vitripennis. BMC Biol. *17*, 78.

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.

Raisner, R., Kharbanda, S., Jin, L., Jeng, E., Chan, E., Merchant, M., Haverty, P.M., Bainer, R., Cheung, T., Arnott, D., et al. (2018). Enhancer Activity Requires CBP/P300 Bromodomain-Dependent Histone H3K27 Acetylation. Cell Rep. *24*, 1722–1729.

Raja, D.A., Subramaniam, Y., Aggarwal, A., Gotherwal, V., Babu, A., Tanwar, J., Motiani, R.K., Sivasubbu, S., Gokhale, R.S., and Natarajan, V.T. (2020). Histone variant dictates fate biasing of neural crest cells to melanocyte lineage. Development *147*.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.

Raz, E. (2003). Primordial germ-cell development: the zebrafish perspective. Nat. Rev. Genet. *4*, 690–700.

Respuela, P., Nikolić, M., Tan, M., Frommolt, P., Zhao, Y., Wysocka, J., and Rada-Iglesias, A. (2016). Foxd3 Promotes Exit from Naive Pluripotency through Enhancer Decommissioning and Inhibits Germline Specification. Cell Stem Cell *18*, 118–133.

Rickels, R., Hu, D., Collings, C.K., Woodfin, A.R., Piunti, A., Mohan, M., Herz, H.-M., Kvon, E., and Shilatifard, A. (2016). An Evolutionary Conserved Epigenetic Mark of Polycomb Response Elements Implemented by Trx/MLL/COMPASS. Mol. Cell *63*, 318–328.

Rickels, R., Herz, H.-M., Sze, C.C., Cao, K., Morgan, M.A., Collings, C.K., Gause, M., Takahashi, Y.-H., Wang, L., Rendleman, E.J., et al. (2017). Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. Nat. Genet. *49*, 1647–1653.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Rohde, C., Zhang, Y., Reinhardt, R., and Jeltsch, A. (2010). BISMA - Fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. BMC Bioinformatics *11*, 230.

Roqueta-Rivera, M., Esquejo, R.M., Phelan, P.E., Sandor, K., Daniel, B., Foufelle, F., Ding, J., Li, X., Khorasanizadeh, S., and Osborne, T.F. (2016). SETDB2 Links Glucocorticoid to Lipid Metabolism through Insig2a Regulation. Cell Metab. *24*, 474–484.

Rulands, S., Lee, H.J., Clark, S.J., Angermueller, C., Smallwood, S.A., Krueger, F., Mohammed, H., Dean, W., Nichols, J., Rugg-Gunn, P., et al. (2018). Genome-Scale Oscillations in DNA Methylation during Exit from Pluripotency. Cell Syst *7*, 63–76.e12.

Russ, B., Olshansky, M., Li, J., Nguyen, M.L.T., Gearing, L.J., Nguyen, T.H.O., Olson, M.R., McQuilton, H.A., Nüssing, S., Khoury, G., et al. (2017). Regulation of H3K4me3 at Transcriptional Enhancers Characterizes Acquisition of Virus-Specific CD8+ T Cell-Lineage-Specific Function. Cell Rep. *21*, 3624–3636.

Saitou, M., and Yamaji, M. (2012). Primordial Germ Cells in Mice. Cold Spring Harbor Perspectives in Biology *4*, a008375–a008375.

Saitou, M., Barton, S.C., and Surani, M.A. (2002). A molecular programme for the specification of germ cell fate in mice. Nature *418*, 293–300.

Sakai, A., Schwartz, B.E., Goldstein, S., and Ahmad, K. (2009). Transcriptional and developmental functions of the H3.3 histone variant in Drosophila. Curr. Biol. *19*, 1816–1820.

Salah Ud-Din, A.I.M., Tikhomirova, A., and Roujeinikova, A. (2016). Structure and Functional Diversity of GCN5-Related N-Acetyltransferases (GNAT). Int. J. Mol. Sci. *17*.

Sapountzi, V., and Côté, J. (2011). MYST-family histone acetyltransferases: beyond chromatin. Cell. Mol. Life Sci. *68*, 1147–1156.

Sasaki, K., Nakamura, T., Okamoto, I., Yabuta, Y., Iwatani, C., Tsuchiya, H., Seita, Y.,

Nakamura, S., Shiraki, N., Takakuwa, T., et al. (2016). The Germ Cell Fate of Cynomolgus Monkeys Is Specified in the Nascent Amnion. Dev. Cell *39*, 169–185.

Sathyan, K.M., McKenna, B.D., Anderson, W.D., Duarte, F.M., Core, L., and Guertin, M.J. (2019). An improved auxin-inducible degron system preserves native protein levels and enables rapid and specific protein depletion. Genes Dev. *33*, 1441–1455.

Schemmer, J., Araúzo-Bravo, M.J., Haas, N., Schäfer, S., Weber, S.N., Becker, A., Eckert, D., Zimmer, A., Nettersheim, D., and Schorle, H. (2013). Transcription factor TFAP2C regulates major programs required for murine fetal germ cell maintenance and haploinsufficiency predisposes to teratomas in male mice. PLoS One *8*, e71113.

Schmid-Burgk, J.L., Höning, K., Ebert, T.S., and Hornung, V. (2016). CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. Nat. Commun. *7*, 12338.

Schmidl, C., Rendeiro, A.F., Sheffield, N.C., and Bock, C. (2015). ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat. Methods *12*, 963–965.

Schmidt, C.S., Bultmann, S., Meilinger, D., Zacher, B., Tresch, A., Maier, K.C., Peter, C., Martin, D.E., Leonhardt, H., and Spada, F. (2012). Global DNA hypomethylation prevents consolidation of differentiation programs and allows reversion to the embryonic stem cell state. PLoS One *7*, e52629.

Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer–promoter contacts in gene expression control. Nature Reviews Genetics *20*, 437–455.

Schröder, R. (2006). vasa mRNA accumulates at the posterior pole during blastoderm formation in the flour beetle Tribolium castaneum. Dev. Genes Evol. *216*, 277–283.

Seah, M.K.Y., and Messerschmidt, D.M. (2018). From Germline to Soma: Epigenetic Dynamics in the Mouse Preimplantation Embryo. Curr. Top. Dev. Biol. *128*, 203–235.

Seervai, R.N.H., and Wessel, G.M. (2013). Lessons for inductive germline determination. Mol. Reprod. Dev. *80*, 590–609.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol. Cell *48*, 849–862.

Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M., and Matsui, Y. (2005). Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. Dev. Biol. *278*, 440–458.

Seki, Y., Yamaji, M., Yabuta, Y., Sano, M., Shigeta, M., Matsui, Y., Saga, Y., Tachibana, M., Shinkai, Y., and Saitou, M. (2007). Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. Development *134*, 2627–2638.

Shirane, K., Kurimoto, K., Yabuta, Y., Yamaji, M., Satoh, J., Ito, S., Watanabe, A., Hayashi, K., Saitou, M., and Sasaki, H. (2016). Global Landscape and Regulatory Principles of DNA Methylation Reprogramming for Germ Cell Specification by Mouse Pluripotent Stem Cells. Dev. Cell *39*, 87–103.

Sim, Y.-J., Kim, M.-S., Nayfeh, A., Yun, Y.-J., Kim, S.-J., Park, K.-T., Kim, C.-H., and Kim, K.-S. (2017). 2i Maintains a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms. Stem Cell Reports *8*, 1312–1328.

Smallwood, S.A., Tomizawa, S.-I., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S.R., and Kelsey, G. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat. Genet. *43*, 811–814.

Smith, A. (2017). Formative pluripotency: the executive phase in a developmental continuum. Development *144*, 365–373.

Soldatov, R., Kaucka, M., Kastriti, M.E., Petersen, J., Chontorotzea, T., Englmaier, L., Akkuratova, N., Yang, Y., Häring, M., Dyachuk, V., et al. (2019). Spatiotemporal structure of cell fate decisions in murine neural crest. Science *364*.

Solini, G.E., Dong, C., and Saha, M. (2017). Embryonic transplantation experiments: Past, present, and future. Trends Dev. Biol. *10*, 13–30.

Song, L., Chen, J., Peng, G., Tang, K., and Jing, N. (2016). Dynamic Heterogeneity of Brachyury in Mouse Epiblast Stem Cells Mediates Distinct Response to Extrinsic Bone Morphogenetic Protein (BMP) Signaling. J. Biol. Chem. *291*, 15212–15225.

Streubel, G., Watson, A., Jammula, S.G., Scelfo, A., Fitzpatrick, D.J., Oliviero, G., McCole, R., Conway, E., Glancy, E., Negri, G.L., et al. (2018). The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. Mol. Cell *70*, 371–379.e5.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21.

Sugimoto, M., Kondo, M., Koga, Y., Shiura, H., Ikeda, R., Hirose, M., Ogura, A., Murakami, A., Yoshiki, A., Chuva de Sousa Lopes, S.M., et al. (2015). A simple and robust method for establishing homogeneous mouse epiblast stem cell lines by wnt inhibition. Stem Cell Reports *4*, 744–757.

Sze, C.C., and Shilatifard, A. (2016). MLL3/MLL4/COMPASS Family on Epigenetic Regulation of Enhancer Function and Cancer. Cold Spring Harb. Perspect. Med. *6*.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., and Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes Dev. *19*, 815–826.

Tafessu, A., and Banaszynski, L.A. (2020). Establishment and function of chromatin modification at enhancers. Open Biol. *10*, 200255.

Tam, P.P.L., and Zhou, S.X. (1996). The Allocation of Epiblast Cells to Ectodermal and Germ-Line Lineages Is Influenced by the Position of the Cells in the Gastrulating Mouse

Embryo. Developmental Biology *178*, 124–132.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., et al. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell *146*, 1016–1028.

Tan, S.-L., Nishi, M., Ohtsuka, T., Matsui, T., Takemoto, K., Kamio-Miura, A., Aburatani, H., Shinkai, Y., and Kageyama, R. (2012). Essential roles of the histone methyltransferase ESET in the epigenetic control of neural progenitor cells during development. Development *139*, 3806–3816.

Tanaka, E., Bailey, T., Grant, C.E., Noble, W.S., and Keich, U. (2011). Improved similarity scores for comparing motifs. Bioinformatics *27*, 1603–1609.

Tang, W.W.C., Kobayashi, T., Irie, N., Dietmann, S., and Surani, M.A. (2016). Specification and epigenetic programming of the human germ line. Nat. Rev. Genet. *17*, 585–600.

Taylor, G.C.A., Eskeland, R., Hekimoglu-Balkan, B., Pradeepa, M.M., and Bickmore, W.A. (2013). H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. Genome Res. *23*, 2053–2065.

Terranova, R., Agherbi, H., Boned, A., Meresse, S., and Djabali, M. (2006). Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. Proc. Natl. Acad. Sci. U. S. A. *103*, 6629–6634.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W., Deaton, A., Andrews, R., James, K.D., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082–1086.

Thongjuea, S., Stadhouders, R., Grosveld, F.G., Soler, E., and Lenhard, B. (2013). r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. Nucleic Acids Res. *41*, e132.

Tischler, J., Gruhn, W.H., Reid, J., Allgeyer, E., Buettner, F., Marr, C., Theis, F., Simons, B.D., Wernisch, L., and Azim Surani, M. (2019). Metabolic regulation of pluripotency and germ cell fate through α-ketoglutarate. The EMBO Journal *38*.

Toyoda-Ohno, H., Obinata, M., and Matsui, Y. (1999). Members of the ErbB receptor tyrosine kinases are involved in germ cell development in fetal mouse gonads. Dev. Biol. *215*, 399–406.

Tracy, C., Warren, J.S., Szulik, M., Wang, L., Garcia, J., Makaju, A., Russell, K., Miller, M., and Franklin, S. (2018). The Smyd Family of Methyltransferases: Role in Cardiac and Skeletal Muscle Physiology and Pathology. Curr Opin Physiol *1*, 140–152.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Tsakiridis, A., Huang, Y., Blin, G., Skylaki, S., Wymeersch, F., Osorno, R., Economou, C.,

Karagianni, E., Zhao, S., Lowell, S., et al. (2015). Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. Development *142*, 809.

Tsiouplis, N.J., Bailey, D.W., Chiou, L.F., Wissink, F.J., and Tsagaratou, A. (2020). TET-Mediated Epigenetic Regulation in Immune Cell Development and Disease. Front Cell Dev Biol *8*, 623948.

Tsunekawa, N., Naito, M., Sakai, Y., Nishida, T., and Noce, T. (2000). Isolation of chicken vasa homolog gene and tracing the origin of primordial germ cells. Development *127*, 2741–2750.

Tu, S., Narendra, V., Yamaji, M., Vidal, S.E., Rojas, L.A., Wang, X., Kim, S.Y., Garcia, B.A., Tuschl, T., Stadtfeld, M., et al. (2016). Co-repressor CBFA2T2 regulates pluripotency and germline development. Nature *534*, 387–390.

Veland, N., Lu, Y., Hardikar, S., Gaddis, S., Zeng, Y., Liu, B., Estecio, M.R., Takata, Y., Lin, K., Tomida, M.W., et al. (2019). DNMT3L facilitates DNA methylation partly by maintaining DNMT3A stability in mouse embryonic stem cells. Nucleic Acids Res. *47*, 152–167.

Venkatarama, T., Lai, F., Luo, X., Zhou, Y., Newman, K., and King, M.L. (2010). Repression of zygotic gene expression in the Xenopus germline. Development *137*, 651–660.

Verma, N., Pan, H., Doré, L.C., Shukla, A., Li, Q.V., Pelham-Webb, B., Teijeiro, V., González, F., Krivtsov, A., Chang, C.-J., et al. (2018). TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. Nat. Genet. *50*, 83–95.

Vincent, J.J., Huang, Y., Chen, P.-Y., Feng, S., Calvopiña, J.H., Nee, K., Lee, S.A., Le, T., Yoon, A.J., Faull, K., et al. (2013). Stage-specific roles for tet1 and tet2 in DNA demethylation in primordial germ cells. Cell Stem Cell *12*, 470–478.

Waddington, C.H. (1940). Organisers and Genes by C. H. Waddington.

Wagner, G.R., and Hirschey, M.D. (2014). Nonenzymatic protein acylation as a carbon stress regulated by sirtuin deacylases. Mol. Cell *54*, 5–16.

Wang, A., Yue, F., Li, Y., Xie, R., Harper, T., Patel, N.A., Muth, K., Palmer, J., Qiu, Y., Wang, J., et al. (2015). Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. Cell Stem Cell *16*, 386–399.

Wang, C., Lee, J.-E., Lai, B., Macfarlan, T.S., Xu, S., Zhuang, L., Liu, C., Peng, W., and Ge, K. (2016). Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. Proc. Natl. Acad. Sci. U. S. A. *113*, 11871–11876.

Weber, A.R., Krawczyk, C., Robertson, A.B., Kuśnierczyk, A., Vågbø, C.B., Schuermann, D., Klungland, A., and Schär, P. (2016). Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. Nat. Commun. *7*, 10806.

Weinberg, D.N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K.N., Horth, C., McGuire, J.T., Xu, X., Nikbakht, H., et al. (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. Nature *573*,

281–286.

Weinert, B.T., Narita, T., Satpathy, S., Srinivasan, B., Hansen, B.K., Schölz, C., Hamilton, W.B., Zucconi, B.E., Wang, W.W., Liu, W.R., et al. (2018). Time-Resolved Analysis Reveals Rapid Dynamics and Broad Scope of the CBP/p300 Acetylome. Cell *174*, 231–244.e12.

Whitaker, J.W., Nguyen, T.T., Zhu, Y., Wildberg, A., and Wang, W. (2015). Computational schemes for the prediction and annotation of enhancers from epigenomic assays. Methods *72*, 86–94.

Whittle, C.A., and Extavour, C.G. (2017). Causes and evolutionary consequences of primordial germ-cell specification mode in metazoans. Proc. Natl. Acad. Sci. U. S. A. *114*, 5784–5791.

Wu, Q., Kanata, K., Saba, R., Deng, C.-X., Hamada, H., and Saga, Y. (2013). Nodal/activin signaling promotes male germ cell fate and suppresses female programming in somatic cells. Development *140*, 291–300.

Xiang, Y., Zhang, Y., Xu, Q., Zhou, C., Liu, B., Du, Z., Zhang, K., Zhang, B., Wang, X., Gayen, S., et al. (2020). Epigenomic analysis of gastrulation identifies a unique chromatin state for primed pluripotency. Nat. Genet. *52*, 95–105.

Xie, G., Lee, J.-E., McKernan, K., Park, Y.-K., Jang, Y., Liu, C., Peng, W., and Ge, K. (2020). MLL3/MLL4 methyltransferase activities regulate embryonic stem cell differentiation independent of enhancer H3K4me1.

Yamaji, M., Seki, Y., Kurimoto, K., Yabuta, Y., Yuasa, M., Shigeta, M., Yamanaka, K., Ohinata, Y., and Saitou, M. (2008). Critical function of Prdm14 for the establishment of the germ cell lineage in mice. Nat. Genet. *40*, 1016–1022.

Yamaji, M., Ueda, J., Hayashi, K., Ohta, H., Yabuta, Y., Kurimoto, K., Nakato, R., Yamada, Y., Shirahige, K., and Saitou, M. (2013). PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. Cell Stem Cell *12*, 368–382.

Yan, J., Chen, S.-A.A., Local, A., Liu, T., Qiu, Y., Dorighi, K.M., Preissl, S., Rivera, C.M., Wang, C., Ye, Z., et al. (2018). Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers. Cell Res. *28*, 387.

Yang, P., Humphrey, S.J., Cinghu, S., Pathania, R., Oldfield, A.J., Kumar, D., Perera, D., Yang, J.Y.H., James, D.E., Mann, M., et al. (2019a). Multi-omic Profiling Reveals Dynamics of the Phased Progression of Pluripotency. Cell Syst *8*, 427–445.e10.

Yang, S.-H., Kalkan, T., Morissroe, C., Marks, H., Stunnenberg, H., Smith, A., and Sharrocks, A.D. (2014). Otx2 and Oct4 drive early enhancer activation during embryonic stem cell transition from naive pluripotency. Cell Rep. *7*, 1968–1981.

Yang, X., Hu, B., Liao, J., Qiao, Y., Chen, Y., Qian, Y., Feng, S., Yu, F., Dong, J., Hou, Y., et al. (2019b). Distinct enhancer signatures in the mouse gastrula delineate progressive cell fate continuum during embryo development. Cell Res. *29*, 911–926.

Yeap, L.-S., Hayashi, K., and Surani, M.A. (2009). ERG-associated protein with SET domain (ESET)-Oct4 interaction regulates pluripotency and represses the trophectoderm lineage.

Epigenetics Chromatin *2*, 12.

Yeo, J.-C., Jiang, J., Tan, Z.-Y., Yim, G.-R., Ng, J.-H., Göke, J., Kraus, P., Liang, H., Gonzales, K.A.U., Chong, H.-C., et al. (2014). Klf2 is an essential factor that sustains ground state pluripotency. Cell Stem Cell *14*, 864–872.

Yesbolatova, A., Saito, Y., Kitamoto, N., Makino-Itou, H., Ajima, R., Nakano, R., Nakaoka, H., Fukui, K., Gamo, K., Tominari, Y., et al. (2020). The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. Nat. Commun. *11*, 5701.

Yi, L., Pimentel, H., Bray, N.L., and Pachter, L. (2018). Gene-level differential analysis at transcript-level resolution. Genome Biol. *19*, 53.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

Yuan, P., Han, J., Guo, G., Orlov, Y.L., Huss, M., Loh, Y.-H., Yaw, L.-P., Robson, P., Lim, B., and Ng, H.-H. (2009). Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. Genes Dev. *23*, 2507–2520.

Zeng, Y., and Chen, T. (2019). DNA Methylation Reprogramming during Mammalian Development. Genes *10*, 257.

Zeng, Y., Ren, R., Kaur, G., Hardikar, S., Ying, Z., Babcock, L., Gupta, E., Zhang, X., Chen, T., and Cheng, X. (2020). The inactive Dnmt3b3 isoform preferentially enhances Dnmt3b-mediated DNA methylation. Genes Dev.

Zhang, J., Dominguez-Sola, D., Hussein, S., Lee, J.-E., Holmes, A.B., Bansal, M., Vlasevska, S., Mo, T., Tang, H., Basso, K., et al. (2015). Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. Nat. Med. *21*, 1190–1198.

Zhang, J., Zhang, M., Acampora, D., Vojtek, M., Yuan, D., Simeone, A., and Chambers, I. (2018a). OTX2 restricts entry to the mouse germline. Nature *562*, 595–599.

Zhang, T., Zhang, Z., Dong, Q., Xiong, J., and Zhu, B. (2020). Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. Genome Biol. *21*, 45.

Zhang, W., Xia, W., Wang, Q., Towers, A.J., Chen, J., Gao, R., Zhang, Y., Yen, C.-A., Lee, A.Y., Li, Y., et al. (2016). Isoform Switch of TET1 Regulates DNA Demethylation and Mouse Development. Mol. Cell *64*, 1062–1073.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., et al. (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res. *38*, 4246–4253.

Zhang, Y., Xiang, Y., Yin, Q., Du, Z., Peng, X., Wang, Q., Fidalgo, M., Xia, W., Li, Y., Zhao,

Z.-A., et al. (2018b). Dynamic epigenomic landscapes during early lineage specification in mouse embryos. Nat. Genet. *50*, 96–105.

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049.

Zheng, Y., Xue, X., Shao, Y., Wang, S., Esfahani, S.N., Li, Z., Muncie, J.M., Lakins, J.N., Weaver, V.M., Gumucio, D.L., et al. (2019). Controlled modelling of human epiblast and amnion development using stem cells. Nature *573*, 421–425.

Zhu, Y., Sun, D., Jakovcevski, M., and Jiang, Y. (2020). Epigenetic mechanism of SETDB1 in brain: implications for neuropsychiatric disorders. Transl. Psychiatry *10*, 115.

Zylicz, J.J., Dietmann, S., Günesdogan, U., Hackett, J.A., Cougot, D., Lee, C., and Surani, M.A. (2015). Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. Elife *4*.

11. Acknowledgments

Most importantly, I would like to start by expressing my deepest appreciation to Álvaro Rada-Iglesias, who has always been there to discuss new experiments, to question and support data analysis and who has always been curious about every single result. His permanent positive attitude is very motivational and supports a pleasant environment. I would especially like to thank Álvaro for the opportunity to learn new techniques such as bioinformatics analysis.

Also, I am extremely grateful to Patricia Respuela for sharing her experience and knowledge that made it easy to learn new techniques and which supported and structured me in the best way possible for this PhD.

I would like to thank Sara de la Cruz Molina, Magdalena Laugsch and Michaela Bartusel for teaching new methods and for our interesting scientific discussions. A special thanks goes to my overall supportive bench neighbours Tomás Pachano and Rizwan Rehimi. And big thanks to Milos Nikolic and Giuliano Crispatzu for the establishment and support with bioinformatic analysis.

Also, I am grateful for the support of Michal-Ruth Schweiger and Achim Tresch of the Thesis Advisory Committee, and Niels Gehring, Peter Tesarz and Siegfried Roth for agreeing to be part of the thesis defense committee.

During my PhD I also had the chance to work with talented students, namely Nils Grotehans, Moritz Reize and Kaitlin Schaaf, who were very pleasant to work with and also great fun after dropping the pipettes.

For the unlimited support during the PhD, I would like to thank Sina Ossenbrink as well as my family and friends.

A special experience of my PhD has been my visit to the Wolf Reik lab at the Babraham Institute where, thanks to the support of Stephen Clark and Laura Benson, I learned a lot and had a great time. I would like to thank Wolf Reik for the support and the hosting of this collaboration.

Finally, I would like to thank Simon Pöpsel and his group for granting me space and for having a good time finishing my PhD project in their lab at the CMMC.

12. Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Köln im März 2021, Tore Bleckwehl

Folgende Teilpublikationen liegen vor:

Tore Bleckwehl & Álvaro Rada-Iglesias, Transcriptional and epigenetic control of germline competence and specification. Current Opinion in Cell Biology 61, 1–8 (2019), doi: 10.1016/j.ceb.2019.05.006

Tore Bleckwehl, Giuliano Crispatzu, Kaitlin Schaaf, Patricia Respuela, Michaela Bartusel, Laura Benson, Stephen J. Clark, Kristel M. Dorighi, Antonio Barral, Magdalena Laugsch, Wilfred F. J. van IJcken, Miguel Manzanares, Joanna Wysocka, Wolf Reik, Álvaro Rada-Iglesias, Enhancer-associated H3K4 methylation safeguards in vitro germline competence. bioRxiv (2021), doi:https://doi.org/10.1101/2020.07.07.192427