Title: Enhancer priming by H3K4 methylation safeguards germline competence

Authors: Tore Bleckwehl[1*], Kaitlin Schaaf[1], Giuliano Crispatzu[1,2], Patricia Respuela[1,3], Michaela Bartusel[1,4], Laura Benson[5], Stephen J. Clark[5], Kristel M. Dorighi[6], Antonio Barral[7], Magdalena Laugsch[1,8], Miguel Manzanares[7,9], Joanna Wysocka[6,10,11], Wolf Reik[5,12,13], Álvaro Rada-Iglesias[1,3,14*]

Affiliations:

1 Center for Molecular Medicine Cologne (CMMC), University of Cologne, Germany.

2 Department of Internal Medicine 2, University Hospital Cologne.

3 Institute of Biomedicine and Biotechnology of Cantabria (IBBTEC), CSIC/University of Cantabria, Spain.

4 Department of Biology, Massachusetts Institute of Technology, USA.

5 Epigenetics Programme, Babraham Institute, Cambridge CB22 3AT, UK.

6 Department of Chemical and Systems Biology, Stanford University School of Medicine, USA.

7 Centro Nacional de Investigaciones Cardiovasculares (CNIC), Spain.

8 Institute of Human Genetics, Heidelberg University Hospital, Germany.

9 Centro de Biología Molecular Severo Ochoa (CBMSO), CSIC-UAM, Spain.

10 Department of Developmental Biology, Stanford University School of Medicine, USA.

11 Howard Hughes Medical Institute, Stanford University School of Medicine, USA.

12 Centre for Trophoblast Research, University of Cambridge, CB2 3EG, UK

13 Wellcome Trust Sanger Institute, Cambridge CB10 1SA, UK.

14 Cluster of Excellence Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Germany

***Corresponding author(s) e-mail address(es):** alvaro.rada@unican.es, tbleckwe@uni-koeln.de

SUMMARY:

Germline specification in mammals occurs through an inductive process whereby competent cells in the post-implantation epiblast reactivate a naïve pluripotency expression program and differentiate into primordial germ cells (PGC). The intrinsic factors that endow epiblast cells with the competence to respond to germline inductive signals remain largely unknown.

Here we show that early germline genes that are active in the naïve pluripotent state become homogeneously dismantled in germline competent epiblast cells. In contrast, the enhancers controlling the expression of major PGC genes transiently and heterogeneously acquire a primed state characterized by intermediate DNA methylation, chromatin accessibility, and H3K4me1. This primed enhancer state is lost, together with germline competence, as epiblast cells develop further. Importantly, we demonstrate that priming by H3K4me1/2 enables the robust activation of PGC enhancers and is required for germline competence and specification. Our work suggests that H3K4me1/2 is directly involved in enhancer priming and the acquisition of competence.

INTRODUCTION:

Competence can be defined as the ability of a cell to differentiate towards a specific cell fate in response to intrinsic and extrinsic signals[1]. While the extracellular signals involved in the induction of multiple cell fates have been described[2], the intrinsic factors that determine the cellular competence to respond to those signals remain elusive.

One major example illustrating the dynamic and transient nature of competence occurs early during mammalian embryogenesis, as the primordial germ cells (PGC), the precursors of the gametes, become specified. In mice, following implantation and the exit of naïve pluripotency (E4.5 - E5.5), PGC are induced from the formative epiblast around E6.0 ~ E6.25 at the proximo-posterior end of the mouse embryo[3]. The induction of PGC occurs in response to signals emanating from the extraembryonic tissues surrounding the epiblast: BMP4 from the extraembryonic ectoderm and WNT3 from the visceral endoderm. Furthermore, regardless of their position within the embryo, formative epiblast cells (~E6.0-6.25) are germline competent when exposed to appropriate signals, but this ability gets lost as the epiblast progresses towards a primed pluripotency state (>E6.5)[4]. Nevertheless, only a fraction (typically <20%) of the formative epiblast cells can give rise to PGCs when exposed to the appropriate signals[5], suggesting that the formative epiblast is heterogeneous in terms of its intrinsic germline competence. Importantly, the intrinsic factors that confer germline competence on the formative but not the primed epiblast cells remain obscure[6,7]. This reflects the difficulties with investigating the molecular basis of germline competence and PGC specification in general, due to the limited

cell numbers that can be obtained *in vivo* from mouse peri-implantation embryos (E4.5-E6.5). These limitations were overcome by establishing a robust *in vitro* differentiation system whereby mouse embryonic stem cells (ESC) grown under 2i conditions (naïve pluripotency) can be sequentially differentiated into epiblast-like cells (EpiLC; formative pluripotency) and primordial germ cell-like cells (PGCLC)[5]. This system revealed transcription factors (TFs)[8–10] and epigenomic reprogramming events involved in PGC specification[11–13] and led to a better understanding of the mouse peri-implantation transitions in general[14]. Hence, several TFs, including FOXD3 and OTX2, were found to promote the transition from naïve to formative pluripotency by coordinating the silencing of naïve genes and the activation of early post-implantation epiblast markers[7,15,16]. Subsequently, FOXD3 and OTX2 restrict the differentiation of formative epiblast cells into PGCs and, thus, the silencing of these TFs is required for germline specification. Previous work from our lab showed that the regulatory function of FOXD3 during these developmental transitions involves binding to and silencing of enhancers shared between naïve pluripotent cells and PGCLC[7]. Interestingly, during the transition from naïve to formative pluripotency, FOXD3-bound enhancers lose H3K27ac but retain H3K4me1, suggesting that they do not become fully decommissioned but transiently primed instead. This led us to propose that the priming of PGCLC enhancers in the formative epiblast might confer germline competence[6,7].

Pre-marking of enhancers with H3K4me1 (*i.e.* enhancer priming) can precede their activation and has been implicated in endodermal differentiation, hematopoiesis or brown adipogenesis[17–20]. H3K4me1 at enhancers is catalyzed by the SET domains of the histone methyltransferases MLL3 (KMT2C) and MLL4 (KMT2D), which are part

of the COMPASS Complex[21,22]. In agreement with the relevance of enhancer priming by H3K4me1, the knockout (KO) of *Mll3/4* or *Mll4* alone impairs enhancer activation and results in differentiation defects in various lineages[23–30]. Despite their manifold and relevant findings, one limitation of studies based on KO models is that they can not discriminate between the catalytic and non-catalytic functions[20,31] of MLL3/4 and, thus, can not directly assess the relevance of H3K4me1 for enhancer function. This question was recently addressed by generating ESC lines with amino acid substitutions in the SET domains of MLL3 and MLL4 that enabled the decoupling of their methyltransferase activity from their non-catalytic functions[32]. In agreement with previous reports[23–25,33], in *Mll3/4* double KO ESC the loss of H3K4me1 at active enhancers was accompanied by a strong reduction in H3K27ac levels, RNA polymerase II binding and enhancer RNA (eRNA) production. In contrast, the loss of H3K4me1 in *Mll3/4* catalytic mutant ESC (*Mll3/4* dCD) only partially reduced H3K27ac at enhancers and had no effects on RNA polymerase II binding or eRNA levels. Furthermore, the *Mll3/4* dCD ESC displayed minor gene expression changes compared to the *Mll3/4* double KO ESC, suggesting that the function of MLL3/4 as long-range co-activators is largely independent of H3K4me1[32]. In addition, work in *Drosophila melanogaster* showed that while the KO of *Trr*, the homolog of *Mll3/4* in flies, was embryonic lethal, an amino acid substitution in the SET domain of *Trr* that globally reduced H3K4me1 did not impair development or viability[34]. Altogether, these reports imply that H3K4me1 might be dispensable for the maintenance of enhancer activity as well as for enhancer priming. Nevertheless, subsequent work with *Mll3/4* dCD ESC showed that the recruitment of chromatin remodelers[35] and the establishment of long-range chromatin interactions[25] require H3K4me1. Moreover

MLL3/4 catalyze the deposition of not only H3K4me1 but also H3K4me2 at enhancers[34] and both of these histone marks antagonize *de novo* CpG methylation[36]. As the function of H3K4me1 has been specifically interrogated in naïve ESC and flies, which are both largely devoid of CpG methylation, it is possible that, under strong *de novo* DNA methylation conditions (*e.g.* exit from naïve pluripotency), H3K4me1 might protect enhancers from CpG methylation and render them competent for future activation[37].

Here we performed an extensive comparison of the transcriptional and epigenetic features of formative and primed epiblast cells to gain insights into the molecular basis of germline competence. Importantly, these comparisons revealed that PGCLC enhancers, which tend to be already active in the naïve state, retain H3K4me1 and remain CpG hypomethylated in formative epiblast cells in comparison to primed pluripotency ones. Most importantly, we show that, in the absence of H3K4me1/2, PGCLC enhancers do not get properly re-activated in PGCLC and germline specification is impaired. Overall, our work demonstrates that enhancer priming by H3K4me1/2 is a major determinant of germline competence and highlights the importance of the epigenetic state of enhancers for the robust deployment of developmental gene expression programs.

RESULTS:

Characterization of the PGCLC *in vitro* differentiation system by single-cell RNA-seq

To overcome the scarcity and transient nature of PGCs *in vivo,* we used an *in vitro* system that faithfully recapitulates PGC specification[5]. In this system, mouse embryonic stem cells (ESC, naïve pluripotency) are differentiated into epiblast-like cells (EpiLC, formative pluripotency) from which PGC-like cells (PGCLC) can be obtained within heterogeneous embryoid bodies (EB). In contrast, epiblast stem cells (EpiSC, primed pluripotency), resembling the post-implantation gastrulating epiblast, can not be efficiently differentiated into PGCLC and, thus, display limited germline competence (Fig. 1a).

Previous work indicated that the acquisition of germline competence by the formative epiblast (*i.e.* E5.5-6.5 epiblast; EpiLC) requires the complete dismantling of the naïve gene expression program, which is subsequently re-activated during PGC induction[38]. However, another possibility that has not been thoroughly investigated is that germline competence is associated with a few cells of the formative epiblast in which the naïve expression program is totally or partially retained. To assess these two alternative scenarios, we performed single-cell RNA sequencing (scRNA-seq) across multiple stages of PGCLC differentiation (the scRNA-seq data can be explored using Supplementary Data 1). t-SNE analysis of the resulting single cell transcriptomes (Supplementary Data 2) showed that cells tended to cluster within their corresponding differentiation stage and were characterized by the expression of stage specific markers (Fig. 1b,d). However, d2 and d4 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*) (Fig. 1d, Supplementary Fig. 1a,b). Furthermore, 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[39]. 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* (Fig. 1c, Supplementary Fig. 1a). Overall, these observations demonstrate the quality of our scRNA-seq data and further support the robustness of the PGCLC *in vitro* differentiation system to investigate germline competence and specification.

Germline competence involves a robust dismantling of the naïve gene expression programme

To better characterize the *in vitro*-derived PGCLC, we performed differential expression analysis by comparing the expression of the PGCLC clusters with the remaining cells of the EB and d2 EpiLC (Supplementary Table 1). This led to the identification of a set of genes containing previously described PGC markers (e.g. *Nanog*, *Tfap2c*, *Prdm14, Prdm1, Dppa3*) (Fig. 1d). In agreement with previous reports[5], we found that many of these genes were highly expressed in ESC, rapidly and strongly silenced in individual EpiLC and EpiSC and finally reactivated in PGCLC (Fig. 1e, Supplementary Fig. 1c). The similarity in gene expression

Figure 1

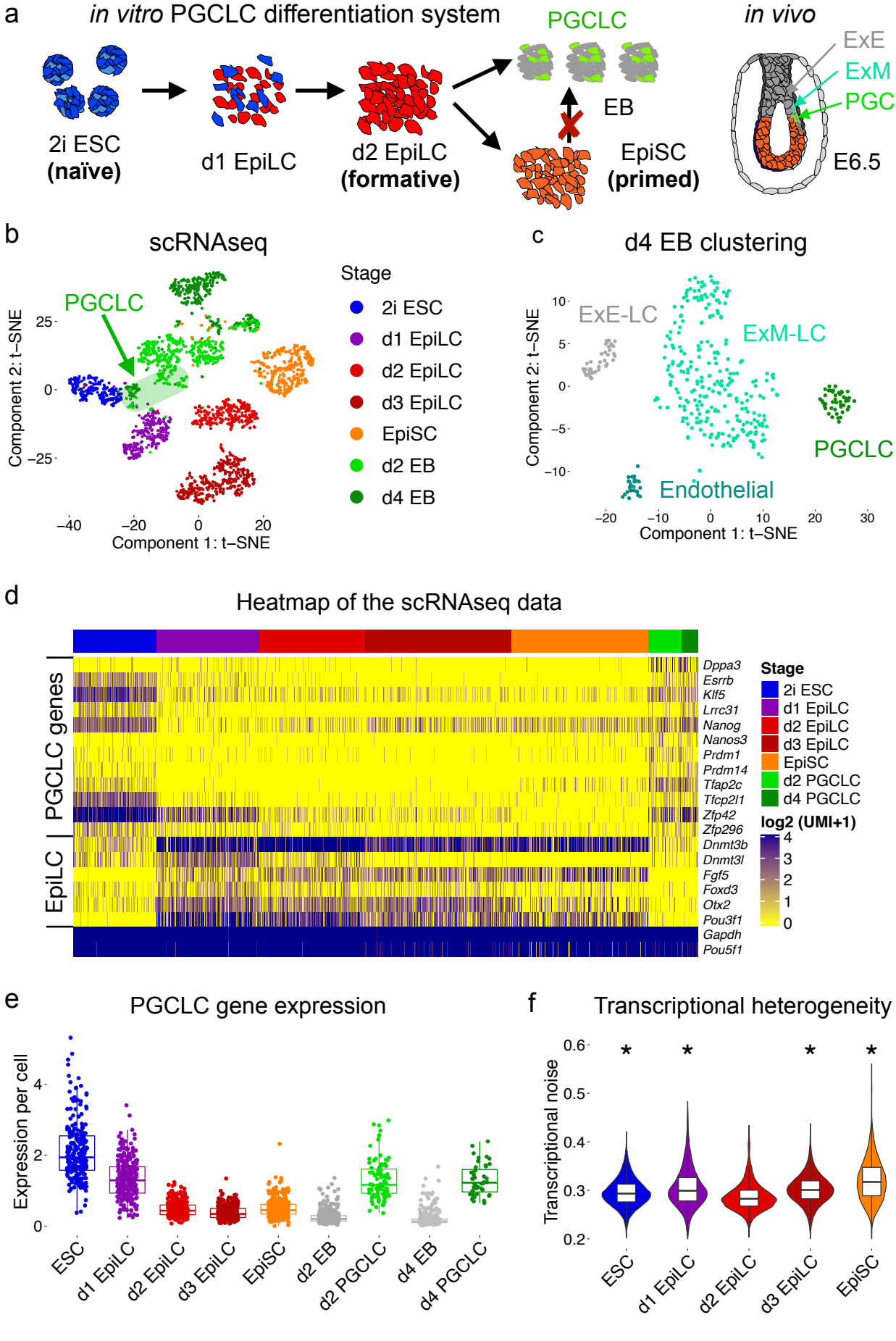

a
in vitro PGCLC differentiation system
2i ESC
(naïve)
d1 EpiLC
d2 EpiLC
(formative)
PGCLC
EB
EpiSC
(primed)
in vivo
ExE
ExM
PGC
E6.5
b
scRNAseq
PGCLC
Component 2: t–SNE
Component 1: t–SNE
Stage
2i ESC
d1 EpiLC
d2 EpiLC
d3 EpiLC
EpiSC
d2 EB
d4 EB
c
d4 EB clustering
ExE-LC
ExM-LC
PGCLC
Endothelial
Component 2: t–SNE
Component 1: t–SNE
d
Heatmap of the scRNAseq data
PGCLC genes
EpiLC
Dppa3
Esrrb
Klf5
Lrrc31
Nanog
Nanos3
Prdm1
Prdm14
Tfap2c
Tfcp2l1
Zfp42
Zfp296
Dnmt3b
Dnmt3l
Fgf5
Foxd3
Otx2
Pou3f1
Gapdh
Pou5f1
Stage
2i ESC
d1 EpiLC
d2 EpiLC
d3 EpiLC
EpiSC
d2 PGCLC
d4 PGCLC
log2 (UMI+1)
e
PGCLC gene expression
Expression per cell
ESC
d1 EpiLC
d2 EpiLC
d3 EpiLC
EpiSC
d2 EB
d2 PGCLC
d4 EB
d4 PGCLC
f
Transcriptional heterogeneity
Transcriptional noise
ESC
d1 EpiLC
d2 EpiLC
d3 EpiLC
EpiSC

Figure 1: Single-cell RNA sequencing (scRNA-seq) profiling of the *in vitro* PGCLC differentiation system.

a.) Schematic representation of the *in vitro* PGCLC differentiation system. 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 diagram to the far right illustrates the *in vivo* location of PGCs at the proximal-posterior end of a mouse embryo.

b.) t-SNE plot of the scRNA-seq data (n = 2,782 cells) generated across the main stages of the *in vitro* PGCLC differentiation protocol. The PGCLC identified within the d2 and d4 EB are shadowed in green and highlighted with an arrow.

c.) t-SNE plot based on the scRNA-seq data generated in d4 EB (n = 368 cells). K-means clustering of d4 EB identified four clusters resembling the transcriptomes of PGCs and their main surrounding tissues *in vivo*: Extraembryonic Ectoderm (ExE), Extraembryonic Mesoderm (ExM), Primordial Germ Cells (PGCs) and Endothelial cells.

d.) Heatmap showing the expression (as UMI (unique molecular identifier) counts) of selected PGCLC genes (*Dppa3-Zpf296*) and epiblast markers (*Dnmt3b-Pou3f1*) within individual cells belonging to the indicated stages. The expression of genes (*Pou5f1, Gapdh*) displaying high and constant expression across all cell types is also shown.

e.) Expression dynamics of the PGCLC genes during PGCLC formation. The PGCLC genes were identified by differential expression analysis of the d2 and d4 PGCLC cluster versus the remaining cells of the EB and d2 EpiLC using a likelihood ratio test. Each dot represents the mean expression of all PGCLC genes in a single cell of the indicated stage. d2 and d4 EB refers to all the cells analyzed within the corresponding EB except those considered as PGCLC.

f.) Violin plots showing 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 stages were compared to d2 EpiLC using wilcoxon tests (*: p-value < $2.2\ 10^{-16}$).

programs between ESC and PGCLC is further supported by the t-SNE plot (Fig. 1b), in which the clusters corresponding to these two cell types are close to each other.

Evaluation of the PGCLC gene set expression dynamics in EpiLC indicated that the silencing of these genes occurs rapidly as the formative pluripotency state is established (Fig. 1d,e), which we also confirmed at the protein level using publicly available proteomic data[40] (Supplementary Fig. 1d). Furthermore, in germline competent d2 EpiLC, all cells clustered together and neither a distinct subpopulation indicative of a retained naïve pluripotency expression program nor signs of precocious germline induction could be identified with different computational methods[41,42] (Supplementary Fig. 1e). Congruently, 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 states (Fig. 1f). This is in full agreement with the E5.5 epiblast bearing the lowest transcriptional noise during mouse peri-implantation development[43]. Notably, using published scRNA-seq data[44], we confirmed that the PGCLC genes also become rapidly and homogeneously silenced *in vivo* following implantation (>E4.5, Supplementary Fig. 1f). Lastly, we analyzed bulk RNA-seq data generated in *Otx2*$^{-/-}$[16] and *Prdm14*$^{-/-}$[11] EpiLC, which despite displaying increased and decreased germline competence, respectively, showed normal gene expression profiles in EpiLC (Fig. 1g).

Although we can not fully exclude the possibility that a rare cell subpopulation exists within the formative stage, the number of single-cell transcriptomes analyzed (EpiLC: 289 cells, E5.5: 126 cells) suggest that the germline competence in the formative

epiblast entails a transcriptionally homogeneous state in which the PGCLC genes are silenced.

PGCLC enhancers acquire a transient primed state in formative pluripotent cells

Most of the PGCLC genes are also inactive in primed pluripotent cells (*i.e.* >E6.5 epiblast: Supplementary Fig. 1f; EpiSC: Fig. 1d,e), which, nevertheless, display limited germline competence. Therefore, the extinction of the naïve program seems to be necessary but not sufficient for the acquisition of germline competence, suggesting that differences, other than transcriptional, should exist between competent (EpiLC, E5.5 epiblast) and non-competent (EpiSC, >E6.5 epiblast) epiblast cells. Based on previous reports[7,45,46], we hypothesized that enhancers involved in PGC specification might display epigenetic differences between EpiLC and EpiSC that could explain the distinct germline competence of these two epiblast stem cell populations.

To test the former hypothesis, we identified active PGCLC enhancers by using H3K27ac ChIP-seq data generated in d2 and d6-sorted PGCLC[12]. Since the H3K27ac peaks were frequently broad, PGCLC enhancers were further refined by the presence of ATAC-seq peaks within the H3K27ac enriched regions, as illustrated in Fig. 2a. Among all distal peaks identified in d2 and d6 PGCLC, only those near the previously defined PGCLC genes (Supplementary Table 1) were considered. Using these criteria, we defined 511 PGCLC enhancers (Supplementary Table 2). On average, one PGCLC gene was linked to 4.5 enhancers, implying that the PGCLC

expression program is controlled by a complex and potentially redundant enhancer network (Supplementary Fig. 2a). Furthermore, and consistent with previous observations[7] and the expression dynamics of the PGCLC genes, the majority of PGCLC enhancers were initially active in ESC (*i.e* high H3K27ac levels), lost H3K27ac in EpiLC and became progressively reactivated in d2 and d6 PGCLC (Fig. 2b). The loss of H3K27ac was specific for PGCLC enhancers and did not affect EpiLC enhancers (Supplementary Fig. 2b).

To explore whether PGCLC enhancers display any differences in their epigenetic status between germline competent (*i.e.* EpiLC) and non-competent (*i.e.* EpiSC) epiblast cells, we performed ChIP-seq experiments for several histone modifications in ESC, EpiLC and EpiSC (Fig. 2c). In agreement with their high H3K27ac levels in ESC, PGCLC enhancers were also enriched in other active histone marks (*i.e.* H3K4me1/2/3) in ESC. All these active chromatin features decreased upon exit from naïve pluripotency, but, H3K4me1 was partly retained in EpiLC in comparison to EpiSC (Fig. 2d, Supplementary Fig. 2c). In contrast, histone modifications associated with polycomb repressive complexes (*i.e.* H3K27me2/3) increased upon exit from the naïve pluripotent stage. However, such an increase only occurred at a small subset of the PGCLC enhancers and no major differences were observed between EpiLC and EpiSC (Supplementary Fig. 2d). Lastly, we also investigated H3K9me2/3, two repressive histone modifications mostly found within heterochromatin, and capable of restricting cellular competence[47,48]. H3K9me2/3 progressively increased as cells exited naïve pluripotency, reaching significantly higher levels in EpiSC than in EpiLC (Supplementary Fig. 2e).

Figure 2

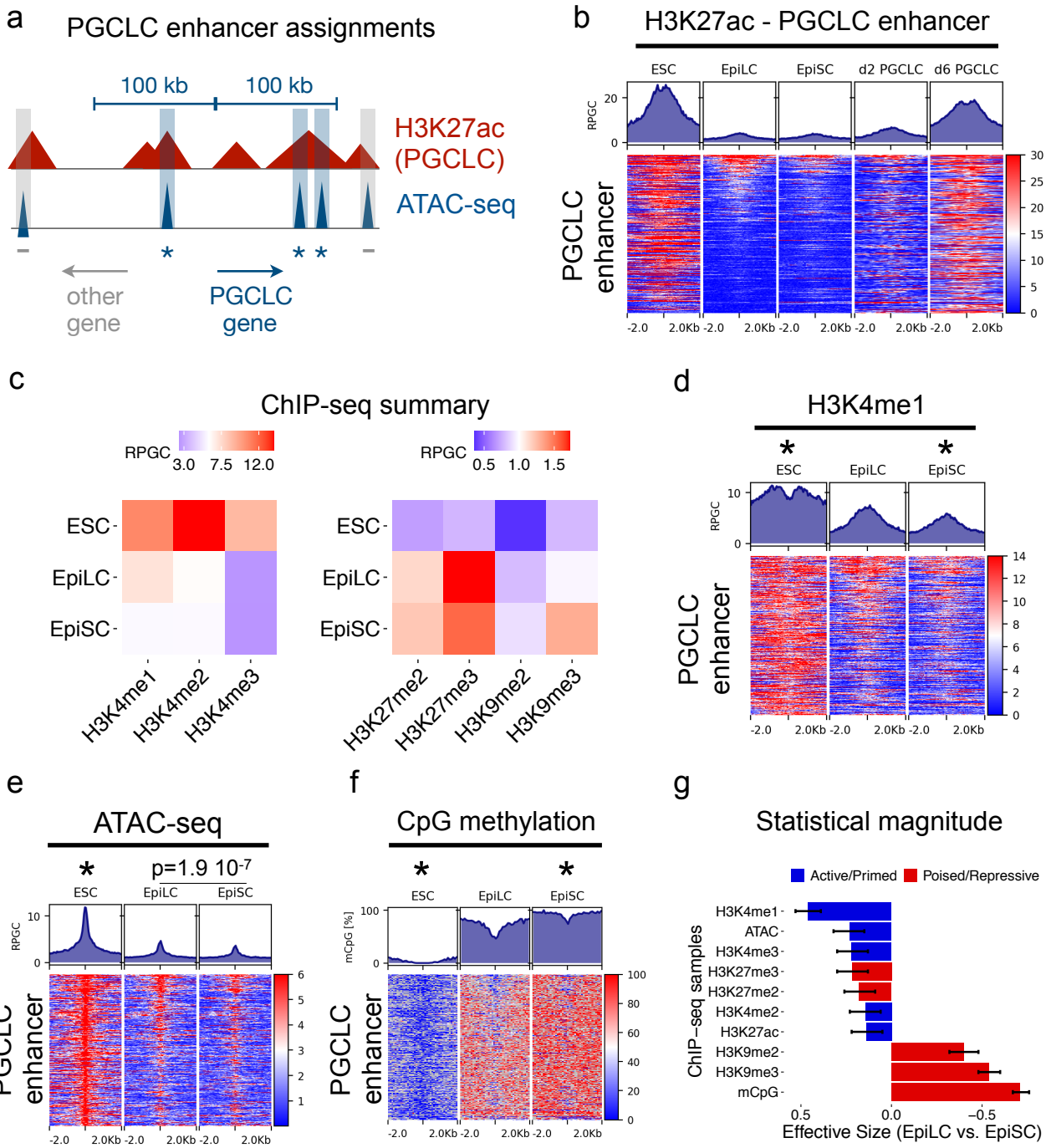

a
PGCLC enhancer assignments
100 kb
100 kb
H3K27ac (PGCLC)
ATAC-seq
other gene
PGCLC gene
b
H3K27ac - PGCLC enhancer
ESC
EpiLC
EpiSC
d2 PGCLC
d6 PGCLC
RPGC
PGCLC enhancer
c
ChIP-seq summary
RPGC 3.0 7.5 12.0
RPGC 0.5 1.0 1.5
ESC
EpiLC
EpiSC
H3K4me1
H3K4me2
H3K4me3
H3K27me2
H3K27me3
H3K9me2
H3K9me3
d
H3K4me1
e
ATAC-seq
p=1.9 10-7
f
CpG methylation
mCpG [%]
g
Statistical magnitude
Active/Primed
Poised/Repressive
ChIP−seq samples
ATAC
mCpG
Effective Size (EpiLC vs. EpiSC)

Figure 2: Epigenomic-based identification and characterization of PGCLC enhancers.

a.) Graphical overview of the strategy used to define PGCLC enhancers (see Methods for additional details). PGCLC enhancers were identified as genomic regions containing overlapping H3K27ac and ATAC-seq peaks in PGCLC and located within 100 kb of at least one of the PGCLC genes defined by scRNA-seq (blue star).

b.) H3K27ac signals within PGCLC enhancers are shown at different stages of PGCLC differentiation as either median profile (top) or heatmap plots (bottom). In the heatmap plots the PGCLC enhancers were ordered by decreasing levels of H3K27ac/H3K27me3 ratios in EpiLC. The PGCLC H3K27ac data was obtained from Kurimoto *et al*. 2015[12].

c.) Summary of the ChIP-seq signals for the PGCLC enhancers in ESC, EpiLC and EpiSC. Mean ChIP-seq signals were calculated for all PGCLC enhancers using a -/+ 1kb window (RPGC: Reads per genomic content).

d-f.) H3K4me1 (d), ATAC-seq (e) and CpG methylation (f) signals within PGCLC enhancers are shown in ESC, EpiLC and EpiSC as median profile (top) and heatmap plots (bottom). In the heatmaps plots the PGCLC enhancers were ordered as in b. For each epigenomic dataset, the mean signals within -/+ 1kb of the PGCLC enhancers were determined and those obtained in EpiLC were compared to the ones measured in ESC or EpiSC using paired wilcoxon tests (n = 511, *: $p < 2.2\ 10^{-16}$). CpG methylation data was obtained from Zylicz *et al.* 2018[89,90].

g.) The magnitude of all epigenetic features between EpiLC and EpiSC at PGCLC enhancers determined as effective size from the paired wilcoxon tests in c-f. The error bar represents the confidence interval.

Since all the previous ChIP-seq experiments were done in E14 ESC, we used another mouse ESC strain (R1) and confirmed that EpiLC displayed higher H3K4me1 levels at PGCLC enhancers in both genetic backgrounds (Supplementary Fig. 2f). As chromatin accessibility has been linked to enhancer priming[44,49], we performed ATAC-seq experiments in ESC, EpiLC and EpiSC and observed that PGCLC enhancers remained more accessible in EpiLC than in EpiSC (Fig. 2e). Furthermore, chromatin accessibility within PGCLC enhancers was orthogonally assessed by analyzing ChIP-seq data for POU5F1 (OCT4)[50,51]. Although *Pou5f1* is similarly expressed in the three cell types (Fig. 1d), its binding within PGCLC enhancers was considerably higher in ESC and EpiLC than in EpiSC (Supplementary Fig. 2g). Given that CpG methylation is a repressive epigenetic modification that frequently co-occurs with H3K9me2/3 and is antagonized by H3K4 methylation[36,52], we also analyzed public whole-genome bisulfite sequencing data generated across pluripotent states[45]. PGCLC enhancers were hypomethylated in ESC, reached intermediate CpG methylation levels in EpiLC and became hypermethylated in EpiSC (Fig. 2f). Overall, among the analyzed epigenetic features of the PGCLC enhancers, the largest differences between EpiLC and EpiSC included higher H3K4me1 and lower CpG methylation in EpiLC (Fig. 2g). Furthermore, such differences were not observed around the transcription start sites (TSS) of the PGCLC genes, which also lost H3K27ac, but retained high levels of H3K4me3 and remained CpG hypomethylated in both EpiLC and EpiSC (Supplementary Fig. 2h-j).

In conclusion, PGCLC enhancers are active in ESC and, as they become decommissioned (loss of H3K27ac), they transiently acquire a primed state in EpiLC

(intermediate H3K4me1 and CpG methylation levels) before becoming fully silenced in EpiSC (CpG hypermethylation, gain of H3K9me2/3 and loss of H3K4me1). Based on these observations, we hypothesize that the priming of PGCLC enhancers in EpiLC might enable their subsequent re-activation in PGCLC and, thus, contribute to germline competence.

The formative epiblast displays epigenetic heterogeneity within PGCLC enhancers

To evaluate whether the priming of PGCLC enhancers also occurs in germline competent cells *in vivo,* we analyzed genome-wide CpG methylation data from mouse epiblasts[53]. In agreement with our *in vitro* observations, PGCLC enhancers displayed intermediate CpG methylation levels in germline competent E5.5 epiblast cells, while they became hypermethylated in the E6.5 epiblast (Fig. 3a), in which germline competence is already reduced[4].

Previous single cell analyses of CpG methylation revealed that, during mouse peri-implantation stages, the formative epiblast is particularly heterogeneous, especially within enhancers with low CpG content (*i.e.* 2.5% CpG)[54]. Interestingly, PGCLC enhancers displayed a low CpG content (2.4%, Supplementary Fig. 3a), which made us hypothesize that the intermediate mCpG levels observed for these enhancers in the formative epiblast could be the result of cell-to-cell variation. To evaluate this idea, we analyzed a comprehensive single cell DNA methylation data set from different epiblast stages (*i.e.* E4.5, E5.5 and E6.5)[44] and measured the CpG

methylation heterogeneity by comparing the methylation status of individual CpGs within PGCLC enhancers[55] (Fig. 3b). This analysis revealed that the formative stage (E5.5) displays the highest variation in CpG methylation, with ~ 30 % of the compared CpG sites being differentially methylated between E5.5 epiblast cells (Fig. 3c,d). Moreover, the CpG methylation heterogeneity of the E5.5 epiblast was more pronounced for PGCLC enhancers than for other enhancers or the whole genome (Fig. 3e, Supplementary Fig. 3b,c). As the CpG coverage was comparable across epiblast stages and the mCpG levels at PGCLC enhancers correlated with the genome-wide levels, the differences in mCpG heterogeneity are unlikely to be caused by technical reasons (Supplementary Fig. 3d).

Next, we reasoned that the CpG methylation heterogeneity observed in the E5.5 epiblast for the PGCLC enhancers could reflect differences in developmental timing and, thus, in germline competence. To explore this possibility, individual cells of each epiblast stage were classified in two groups (high and low mCpG) based on their mean mCpG levels within PGCLC enhancers. Then, cells within the high and low mCpG groups were compared regarding their chromatin accessibility and gene expression levels (Fig. 3f-h). The E4.5 epiblast cells with higher mCpG levels showed decreased chromatin accessibility within PGCLC enhancers and lower expression of PGCLC genes, which could indicate an incipient exit from naïve pluripotency and, hence, a more advanced developmental age. In contrast, although the E5.5 epiblast cells with higher mCpG methylation also displayed lower chromatin accessibility within PGCLC enhancers, they did not show major differences in the expression of the PGCLC genes compared to E5.5 cells with low mCpG levels (Fig. 3f-h). Therefore, in the formative epiblast, the epigenetic status of the PGCLC

Figure 3

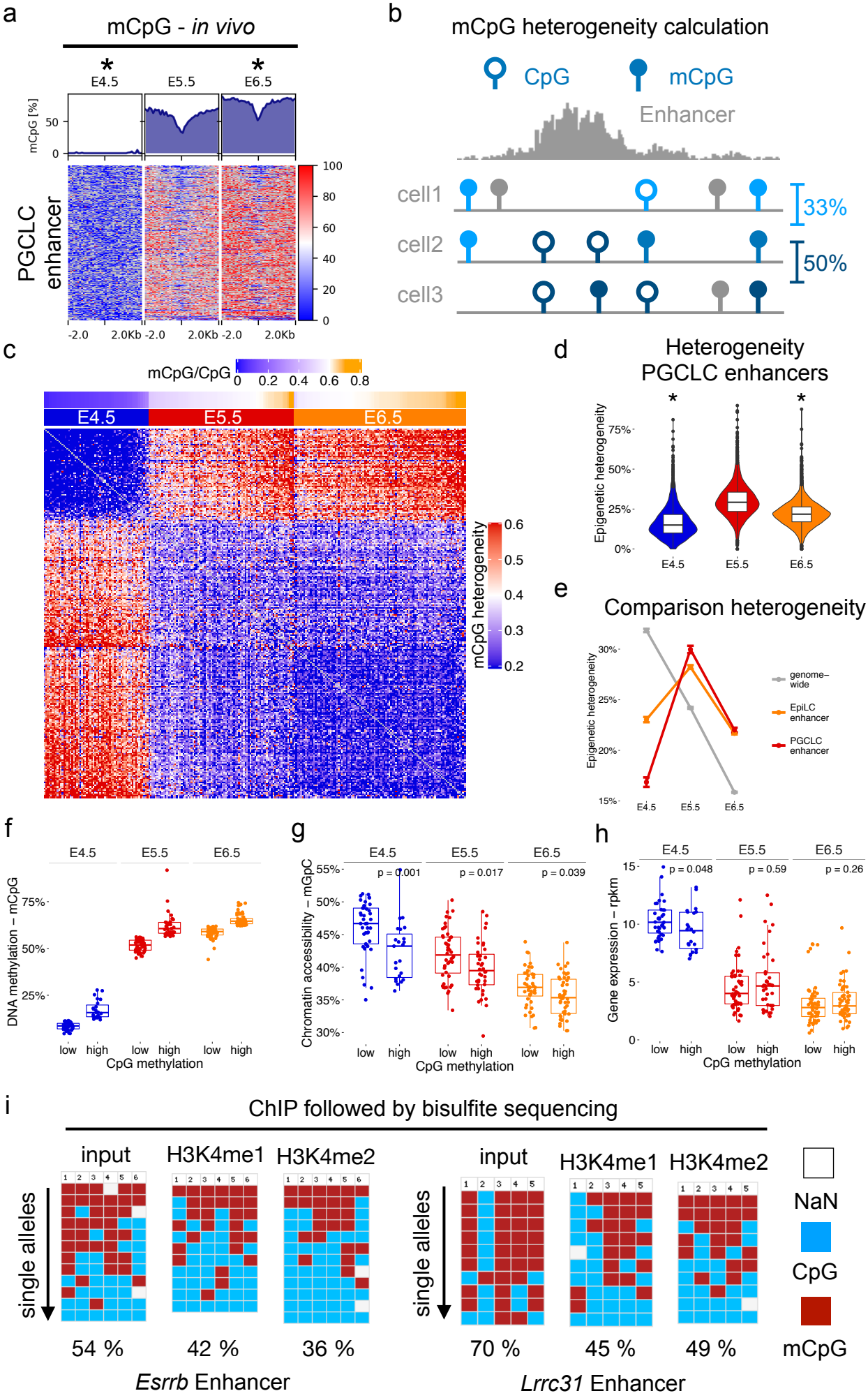

a
mCpG - in vivo
E4.5
E5.5
E6.5
mCpG [%]
PGCLC enhancer
-2.0
2.0Kb
b
mCpG heterogeneity calculation
CpG
mCpG
Enhancer
cell1
cell2
cell3
33%
50%
c
mCpG/CpG
mCpG heterogeneity
d
Heterogeneity PGCLC enhancers
Epigenetic heterogeneity
e
Comparison heterogeneity
genome–wide
EpiLC enhancer
PGCLC enhancer
f
DNA methylation – mCpG
CpG methylation
low
high
g
Chromatin accessibility – mGpC
p = 0.001
p = 0.017
p = 0.039
h
Gene expression – rpkm
p = 0.048
p = 0.59
p = 0.26
i
ChIP followed by bisulfite sequencing
input
H3K4me1
H3K4me2
single alleles
54 %
42 %
36 %
Esrrb Enhancer
70 %
45 %
49 %
Lrrc31 Enhancer
NaN
CpG
mCpG

Figure 3: PGCLC enhancers are epigenetically heterogeneous in the formative epiblast.

a.) The *in vivo* CpG methylation levels of the PGCLC enhancers were analyzed in E4.5, E5.5 and E6.5 epiblasts using data from Zhang *et al.* 2018[53]. The mCpG levels of each PGCLC enhancer (using a -/+ 1kb window) in E5.5 were compared to those measured in E4.5 or E6.5 using a paired wilcoxon test (n = 511, *: $p < 2.2\ 10^{-16}$).

b.) Illustration of the mCpG heterogeneity estimation, which is based on the mCpG dissimilarity concept[91]. 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 100% and if they show opposite methylation patterns a value of 0 %. The mean of all pairwise comparisons reflect the CpG heterogeneity.

c.) 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 epigenetically similar and red that they are different). 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=258 cells). For each developmental stage, the cells were ranked according to average mCpG levels within the PGCLC enhancers.

d.) Violin plots showing the CpG methylation heterogeneity measured for the PGCLC enhancers within E4.5, E5.5 and E6.5 epiblast cells. The CpG methylation heterogeneity in E5.5 was compared to the one measured in E4.5 or E6.5 using wilcoxon tests (*: p-value $< 2.2\ 10^{-16}$).

e.) The CpG methylation heterogeneity in E4.5, E5.5 and E6.5 was measured as described in b for PGCLC enhancers, EpiLC enhancers (Supplementary Fig. 2a) or genome-wide for all CpGs (Supplementary Fig. 3f). The error bars show the confidence interval of each measurement.

f-h.) Single cells from E4.5, E5.5 and E6.5 epiblasts were classified as either low or high based on their average CpG methylation within PGCLC enhancers. Then, for each individual cell of the low and high groups, (f) the average CpG methylation of the PGCLC enhancers (n=258 cells), (g) the average chromatin accessibility (mGpC) of the PGCLC enhancers (n=242 cells) and (h) the average expression of the PGCLC genes (n=254 cells) are shown. In (g) and (h) outliers, likely representing technical differences, were discarded. Within each stage, the differences in chromatin accessibility and PGCLC gene expression for high and low methylated cells were compared by wilcoxon tests.

i.) The CpG methylation status of the PGCLC enhancers associated with *Esrrb* and *Lrrc31* 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 which were not covered are shown in gray. At least 10 alleles were analyzed for each template DNA (rows).

enhancers and the expression of the PGCLC genes seem to become uncoupled. Lastly, since the priming of PGCLC enhancers in the formative epiblast involves intermediate levels of not only mCpG but also H3K4me1, we wondered whether formative epiblast cells with lower mCpG within PGCLC enhancers might show higher H3K4me1 levels. To test this, we performed H3K4me1 ChIP in EpiLC followed by bisulfite treatment and DNA sequencing of two representative PGCLC enhancers (*Esrrb* and *Lrrc31* enhancers). In comparison to the input genomic DNA, the H3K4me1-enriched DNA showed lower mCpG levels at the two analyzed enhancers (Fig. 3i).

Collectively, the previous results show that, in contrast to its transcriptional homogeneity, the formative epiblast displays variable CpG methylation and H3K4me1 within PGCLC enhancers, suggesting that the priming of these enhancers is heterogeneous among formative epiblast cells.

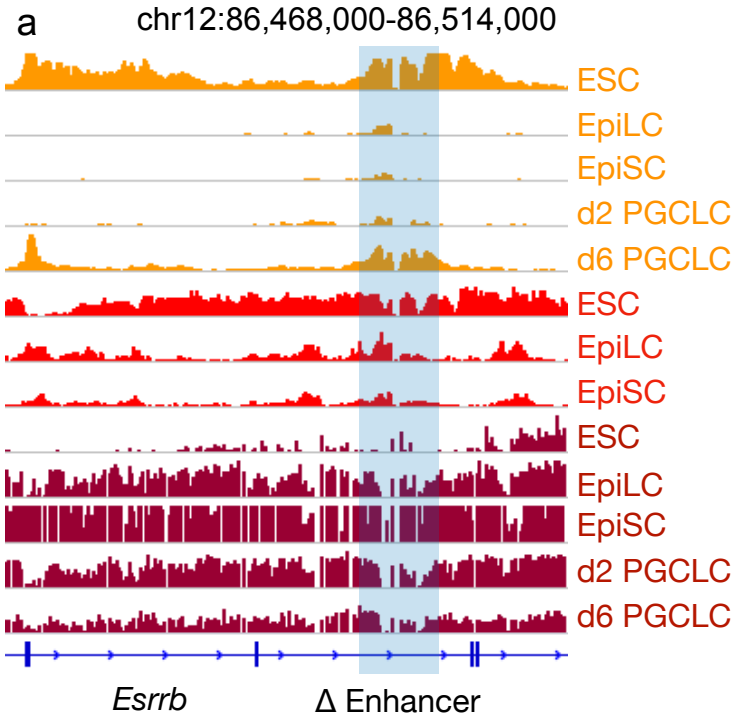
Assessment of PGCLC enhancer function during PGCLC specification

Before investigating PGCLC enhancer priming in more depth, we evaluated the functional relevance of some representative enhancers during PGCLC specification. We selected three PGCLC enhancers associated with *Esrrb*, *Klf5*, and *Lrrc31* that showed the preponderant epigenetic dynamics described in Fig. 2 (Fig. 4a, Supplementary Fig. 4a). We used CRISPR/Cas9 to delete them individually in ESC containing a DPPA3-GFP reporter that facilitates subsequent PGCLC quantification

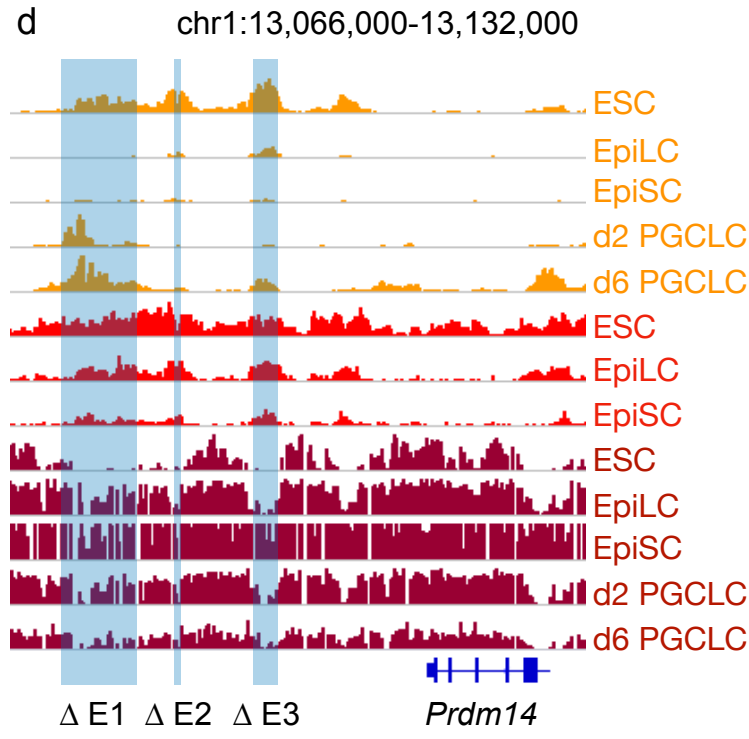
(Supplementary Fig. 4b). Two clonal lines for each individual enhancer deletion were generated and analysed for their PGCLC differentiation capacity. The deletion of the enhancers associated with *Lrrc31* and *Klf5* significantly diminished PGCLC specification (Fig. 4b, Supplementary Fig. 4c). Furthermore, both enhancer deletions strongly reduced the expression of *Klf5* and *Lrrc31* not only in d4 EB but already in ESC (Supplementary Fig. 4d,e). Hence, the reduced PGCLC differentiation capacity of the ESC lines with the *Klf5* and *Lrrc31* enhancer deletions could be indirectly caused by an already compromised naïve pluripotent state that subsequently reduces germline competence and/or by direct effects during PGCLC specification. Regardless, *Klf5* and *Lrrc31* might represent PGCLC regulators shared between naïve and PGCLC, similarly to other naïve/PGCLC TFs with this dual regulatory role (e.g. *Prdm14*, *Nanog*)[6]. On the other hand, the *Esrrb* enhancer deletion moderately reduced the expression of *Esrrb* in ESC and did not significantly affect PGCLC differentiation (Fig. 4b,c). Nevertheless, the expression of *Esrrb* was severely diminished in d4 EB (Fig. 4c), demonstrating that this enhancer is needed for the proper induction of *Esrrb* in PGCLC.

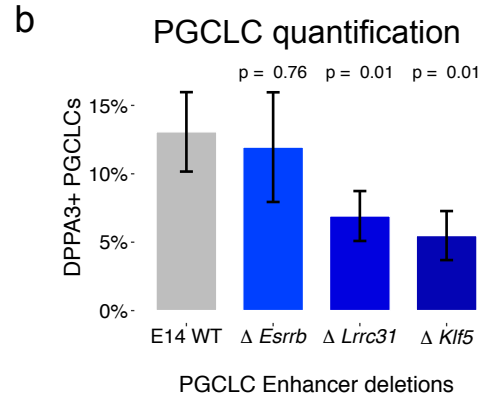
Since most of the PGCLC genes are associated with multiple and potentially redundant enhancers, we used CRISPR/Cas9 in ESC to individually delete three different enhancers (*i.e.* E1, E2 and E3) (Fig. 4d) previously described as components of a *Prdm14* super-enhancer[56]. In agreement with this work, the E2 deletion strongly reduced *Prdm14* expression in ESC, while the deletion of E3 and E1 resulted in considerably smaller or no effects, respectively (Fig. 4e). Interestingly, upon PGCLC differentiation, the regulatory importance of each enhancer changed considerably and E1 strongly contributed to *Prdm14* induction, especially during

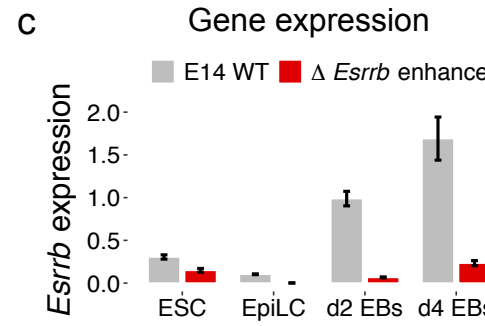
Figure 4

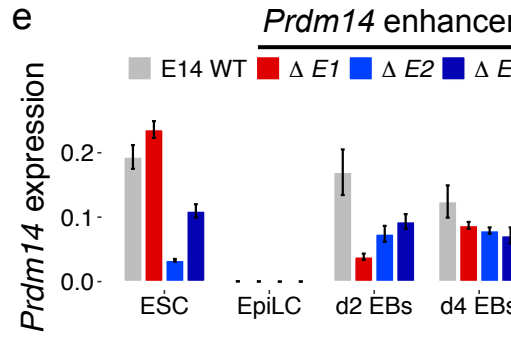

a
chr12:86,468,000-86,514,000
ESC
EpiLC
EpiSC
d2 PGCLC
d6 PGCLC
ESC
EpiLC
EpiSC
ESC
EpiLC
EpiSC
d2 PGCLC
d6 PGCLC
Esrrb
Δ Enhancer


H3K27ac
0 - 100 RPGC
H3K4me1
0 - 30 RPGC
mCpG
0 - 100 %


d
chr1:13,066,000-13,132,000
ESC
EpiLC
EpiSC
d2 PGCLC
d6 PGCLC
ESC
EpiLC
EpiSC
ESC
EpiLC
EpiSC
d2 PGCLC
d6 PGCLC
Δ E1 Δ E2 Δ E3
Prdm14


b
PGCLC quantification
p = 0.76 p = 0.01 p = 0.01
DPPA3+ PGCLCs
15%
10%
5%
0%
E14 WT
Δ Esrrb
Δ Lrrc31
Δ Klf5
PGCLC Enhancer deletions


c
Gene expression
E14 WT
Δ Esrrb enhancer
Esrrb expression
2.0
1.5
1.0
0.5
0.0
ESC
EpiLC
d2 EBs
d4 EBs


e
Prdm14 enhancer
E14 WT
Δ E1
Δ E2
Δ E3
Prdm14 expression
0.2
0.1
0.0
ESC
EpiLC
d2 EBs
d4 EBs

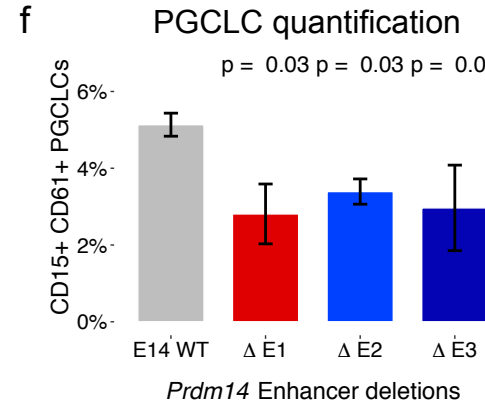

f
PGCLC quantification
p = 0.03 p = 0.03 p = 0.03
CD15+ CD61+ PGCLCs
6%
4%
2%
0%
E14 WT
Δ E1
Δ E2
Δ E3
Prdm14 Enhancer deletions

Figure 4: Functional assessment of representative PGCLC enhancers.
a.) Genome-browser view showing the PGCLC enhancer found within the *Esrrb* locus and its H3K27ac, H3K4me1 and CpG methylation dynamics during PGCLC formation. The H3K27ac PGCLC data was obtained from Kurimoto *et al.* 2015[12] and the CpG methylation data from Shirane *et al.* 2016[11].
b.) WT ESC and ESC lines with the indicated PGCLC enhancer deletions were differentiated into PGCLC. The number of PGCLC present within d4 EB was quantified using a DPPA-GFP reporter. 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 when differentiating WT ESC using wilcoxon tests.
c.) *Esrrb* expression levels were measured by RT-qPCR in d4 EB differentiated from WT ESC and the two different ESC clonal lines with the *Esrrb* enhancer deletion shown in a. 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).
d.) Genome-browser view showing the PGCLC enhancers (E1-E3) found within the *Prdm14* locus and their H3K27ac, H3K4me1 and CpG methylation dynamics during PGCLC formation.
e.) *Prdm14* expression levels were measured by RT-qPCR in d4 EB differentiated from WT ESC and ESC with the indicated *Prdm14* enhancer deletions (two clonal lines for each). 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).
f.) 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 enhancer deletions were compared to those obtained with WT ESC using wilcoxon tests.

early PGCLC specification (*i.e.* d2 EB). Together with the epigenetic profiles observed within the *Prdm14* locus (Fig. 4d), 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). Furthermore, in agreement with the role of *Prdm14* as a PGC master regulator[57], we found that the individual E1-E3 deletions significantly impaired PGCLC differentiation (Fig. 4f, Supplementary Fig. 4f). 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[58,59]. In contrast, the E1 enhancer directly contributes to PGCLC specification, as its deletion affected *Prdm14* expression during PGCLC differentiation but not in ESC.

In summary, these results indicate that PGCLC enhancers frequently control the expression of their target genes already in ESC, further supporting that a significant set of enhancers is functionally shared between naïve pluripotency and PGCLC[7]. Most importantly, the enhancer deletions that reduced the expression of their target genes preferentially *(i.e. Esrrb* enhancer) or solely (*i.e.* E1 *Prdm14* enhancer) in PGCLC support the functional relevance of the identified PGCLC enhancers during germline specification.

The priming of PGCLC enhancers is associated with permissive chromatin and topological features

The priming of enhancers has been suggested to facilitate their future activation and the subsequent induction of their target genes[60]. Mechanistically, the presence of

H3K4me1/2 could protect primed enhancers from mCpG and H3K9me2/3 by blocking the recruitment and/or activity of DNA and H3K9 methyltransferases[36,61–65], while facilitating the recruitment of the BAF chromatin remodelling complex[35]. Therefore, the transient priming of the PGCLC enhancers in formative epiblast cells could endow them with permissive chromatin features. To investigate this possibility, we used an experimental system whereby the inducible overexpression of PRMD14 or NANOG enables PGCLC differentiation in the absence of growth factors[9,66]. Using piggyBac vectors, we generated clonal ESC lines in which HA-tagged PRDM14 or NANOG could be overexpressed upon addition of Doxycycline (Dox) (Supplementary Fig. 5a,b). In line with previous reports, the overexpression of either PRDM14 or NANOG upon differentiation of EpiLC into PGCLC yielded a high percentage of PGCLC (Fig. 5a, Supplementary Fig. 5c). In contrast, the overexpression of PRDM14 or NANOG upon differentiation from EpiSC resulted in considerably less PGCLC. To assess whether PGCLC enhancers were differentially accessible to these TFs in EpiLC and EpiSC, we performed ChIP-seq experiments after a short induction of HA-tagged PRDM14 or NANOG in both EpiLC and EpiSC. Importantly, PGCLC enhancers were more strongly bound by PRDM14-HA and NANOG-HA in EpiLC than in EpiSC (Fig. 5b). Furthermore, the binding of NANOG-HA to the PGCLC enhancers in EpiLC resulted in increased H3K27ac and H3K4me2 levels, indicating that NANOG overexpression can specifically (*i.e.* H3K27ac or H3K4me2 did not increase at EpiLC enhancers) and more effectively activate PGCLC enhancers in EpiLC than in EpiSC (Fig. 5c,d and Supplementary Fig. 5d-f).

Figure 5

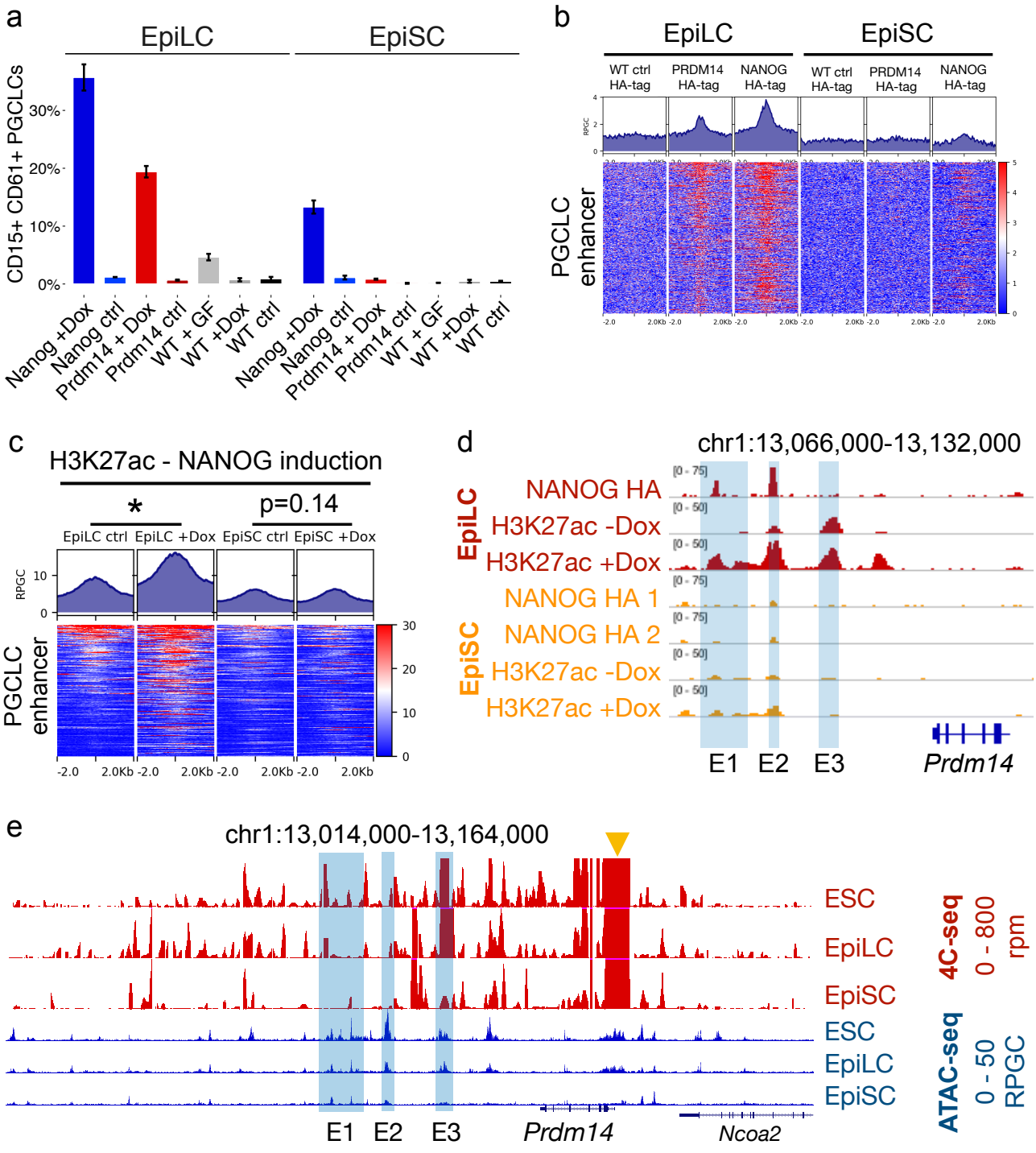

a
EpiLC
EpiSC
CD15+ CD61+ PGCLCs
30%
20%
10%
0%
Nanog +Dox
Nanog ctrl
Prdm14 + Dox
Prdm14 ctrl
WT + GF
WT +Dox
WT ctrl
b
EpiLC
EpiSC
WT ctrl HA-tag
PRDM14 HA-tag
NANOG HA-tag
RPGC
PGCLC enhancer
c
H3K27ac - NANOG induction
*
p=0.14
EpiLC ctrl
EpiLC +Dox
EpiSC ctrl
EpiSC +Dox
RPGC
PGCLC enhancer
-2.0
2.0Kb
d
chr1:13,066,000-13,132,000
EpiLC
NANOG HA
H3K27ac -Dox
H3K27ac +Dox
EpiSC
NANOG HA 1
NANOG HA 2
H3K27ac -Dox
H3K27ac +Dox
E1
E2
E3
Prdm14
e
chr1:13,014,000-13,164,000
ESC
EpiLC
EpiSC
4C-seq
0 - 800
rpm
ATAC-seq
0 - 50
RPGC
E1
E2
E3
Prdm14
Ncoa2

Figure 5: PGCLC enhancers display permissive chromatin and topological features in EpiLC.

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.) Exogeneous PRDM14 or NANOG tagged with HA were overexpressed for 18 hours in EpiLC and EpiSC. Then, ChIP-seq experiments were performed using an anti-HA antibody. As a negative control, similar ChIP-seq experiments were performed in WT EpiLC and WT EpiSC not expressing the HA-tagged TFs. The binding of the exogenous PRDM14 and NANOG within the PGCLC enhancers is shown as median profile (top) and heatmap plots (bottom). In the heatmap plots the PGCLC enhancers were ordered as in Fig. 2b.

c.) The ESC line enabling the inducible overexpression of exogenous HA-tagged NANOG 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 b and H3K27ac ChIP-seq experiments were performed. H3K27ac levels within PGCLC enhancers are shown as median profile (top) and heatmap plots (bottom). In the heatmap plots the PGCLC enhancers were ordered as in Fig. 2b. The mean H3K27ac signals within -/+ 1kb of each PGCLC enhancer for untreated and treated cells were compared by a paired wilcoxon test (n = 511, *: p-value < $2.2\ 10^{-16}$).

d.) Genome-browser view of the *Prdm14* locus showing the binding of the exogenously induced NANOG-HA in both EpiLC and EpiSC as well as the H3K27ac 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 the reproducibility of the weak binding within the PGCLC enhancers. The relevant PGCLC enhancers (*i.e.* E1-E3) found within the *Prdm14* locus are highlighted in blue.

e.) 4C-seq experiments in ESC, EpiLC and EpiSC using the *Prdm14* promoter as a viewpoint (orange triangle). The 4C-seq signals within the *Prdm14* locus and ATAC-seq profiles from the same stages are shown. The relevant *Prdm14* enhancers (*i.e.* E1-E3) are highlighted in blue.

Besides providing an accessible chromatin environment, H3K4me1 can also facilitate the recruitment of the cohesin complex, which in turn can increase chromatin interactions between enhancers and genes[25]. In the context of the PGCLC enhancers, the retention of H3K4me1 in EpiLC could help to maintain the physical proximity between the enhancers and their target genes and, thus, facilitate future gene induction upon PGC specification[67–69]. To investigate the 3D chromatin architecture of a representative locus harboring PGCLC enhancers, we performed circular chromosome conformation capture (4C-seq) experiments in ESC, EpiLC and EpiSC using the *Prdm14* TSS as a viewpoint (Fig. 5e). In agreement with their initial active state, in ESC the three *Prdm14* enhancers (i.e. E1-E3) were in close spatial proximity to the *Prdm14* TSS. Remarkably, the contacts between *Prdm14* and the E1-E3 enhancers were partly preserved in EpiLC but strongly diminished in EpiSC (Fig. 5e).

Overall, these results suggest that the priming of PGCLC enhancers in EpiLC might contribute to germline competence by conferring these enhancers with permissive chromatin and topological features that facilitate their activation and regulatory function.

H3K4me1/2 is necessary for germline competence

Next, we wanted to specifically assess the relevance of H3K4me1 for PGCLC enhancer priming and germline competence. To do so, we used a previously generated mESC line (dCD ESC) that is catalytically deficient for the H3K4 methyltransferases MLL3 (KMT2C) and MLL4 (KMT2D)[32]. In these dCD ESC,

H3K4me1 is completely lost from enhancers, while the binding of MLL3/4 and their associated complexes is largely maintained[32] (Fig. 6a). Gene expression, H3K27ac and eRNA levels within active enhancers were only mildly affected in dCD ESC, indicating that H3K4me1 is dispensable for the maintenance of enhancer activity[32,34]. However, the importance of H3K4me1 for the priming and subsequent activation of enhancers upon ESC differentiation has not been investigated yet. Using previously generated ChIP-seq data[32], we found that PGCLC enhancers were strongly and similarly bound by MLL3/4 in WT and dCD ESC (Fig. 6b). Moreover, H3K4me1 levels within PGCLC enhancers were strongly reduced in dCD ESC as well as upon their differentiation into EpiLC and EpiSC (Fig. 6c, Supplementary Fig. 6a). The loss of H3K4me1 and the milder reduction of H3K27ac within PGCLC enhancers were also observed in an independent MLL3/4 catalytic mutant ESC line (dCT; Fig. 6a, Supplementary Fig. 6a). Additionally, the dCD cells displayed a strong reduction in H3K4me2 within PGCLC enhancers (Supplementary Fig. 6b). In agreement with the histone modification changes observed for active enhancers in MLL3/4 catalytic mutant ESC lines[34], the dCD EpiLC also displayed strong H3K4me1/2 losses and milder H3K27ac reductions within other enhancer categories (e.g. EpiLC enhancers, Supplementary Fig. 6c). Lastly, in agreement with the protective role of H3K4me1/2 against DNA methylation[36,62], mCpG levels within PGCLC enhancers were elevated in dCD ESC and EpiLC (Fig. 6d).

The previous analyses indicated that, in the absence of MLL3/4 catalytic activity, the PGCLC enhancers no longer acquire a primed state in EpiLC. Chiefly, upon differentiation into PGCLC, the MLL3/4 catalytic mutant ESC lines (*i.e.* dCD and dCT) showed a significant reduction in their capability of generating PGCLC (Fig. 6e,

Figure 6

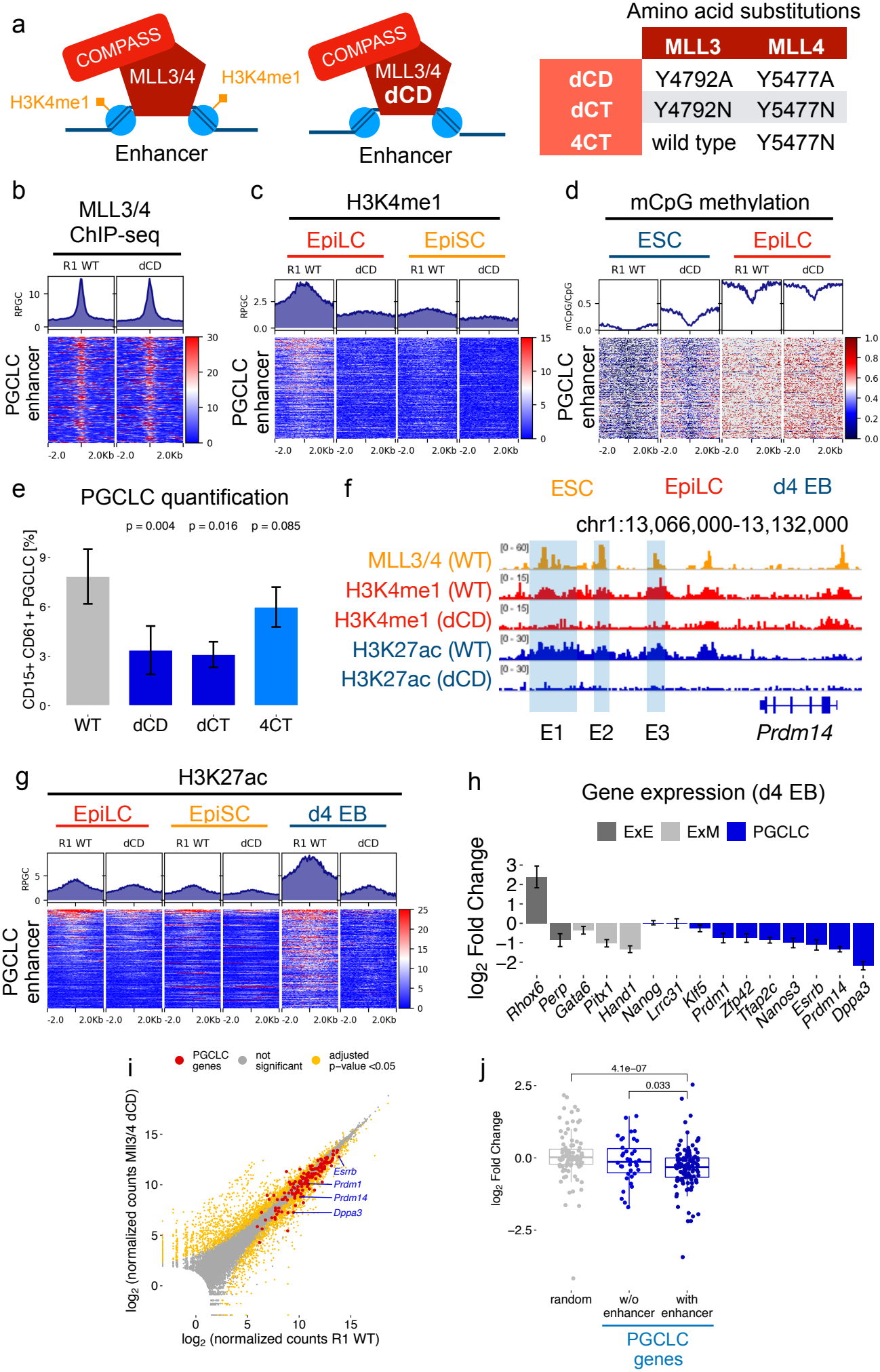

a
COMPASS
MLL3/4
H3K4me1
H3K4me1
Enhancer
COMPASS
MLL3/4
dCD
Enhancer
Amino acid substitutions
MLL3
MLL4
dCD
Y4792A
Y5477A
dCT
Y4792N
Y5477N
4CT
wild type
Y5477N
b
MLL3/4
ChIP-seq
R1 WT
dCD
RPGC
PGCLC
enhancer
-2.0
2.0Kb
c
H3K4me1
EpiLC
EpiSC
d
mCpG methylation
ESC
EpiLC
mCpG/CpG
e
PGCLC quantification
p = 0.004
p = 0.016
p = 0.085
CD15+ CD61+ PGCLC [%]
WT
dCD
dCT
4CT
f
ESC
EpiLC
d4 EB
chr1:13,066,000-13,132,000
MLL3/4 (WT)
H3K4me1 (WT)
H3K4me1 (dCD)
H3K27ac (WT)
H3K27ac (dCD)
E1
E2
E3
Prdm14
g
H3K27ac
EpiLC
EpiSC
d4 EB
h
Gene expression (d4 EB)
ExE
ExM
PGCLC
log2 Fold Change
Rhox6
Perp
Gata6
Pitx1
Hand1
Nanog
Lrrc31
Klf5
Prdm1
Zfp42
Tfap2c
Nanos3
Esrrb
Prdm14
Dppa3
i
PGCLC genes
not significant
adjusted p–value <0.05
log2 (normalized counts Mll3/4 dCD)
log2 (normalized counts R1 WT)
Esrrb
Prdm1
Prdm14
Dppa3
j
4.1e−07
0.033
log2 Fold Change
random
w/o enhancer
with enhancer
PGCLC genes

Figure 6: H3K4me1/2 priming is required for PGCLC enhancer activation.
a.) Schematic illustration of the dCD cells, which express catalytic mutant MLL3 and MLL4 proteins without histone methyltransferase activity but are capable of binding to its target sites and interacting with other proteins as part of the COMPASS complex. The different *Mll3/4* catalytic mutants used in this study are shown to the right.
b.) ChIP-seq signals for WT and dCD MLL3/4 proteins at the PGCLC enhancers in ESC. The WT MLL3/4 ChIP-seq experiments were merged from two replicates. MLL3/4 levels within PGCLC enhancers are shown as median profile (top) and heatmap plots (bottom). In the heatmap plots the PGCLC enhancers were ordered as in Fig. 2b. The ChIP-seq data were obtained from Dorighi *et al*. 2017[32].
c.) H3K4me1 ChIP-seq experiments in d2 EpiLC and EpiSC differentiated from R1 WT and dCD ESC. H3K4me1 levels within PGCLC enhancers are shown as median profile (top) and heatmap plots (bottom). In the heatmap plots the PGCLC enhancers were ordered as in Fig. 2b.
d.) CpG methylation levels of the PGCLC enhancers in R1 WT and dCD ESC and d2 EpiLC. mCpG levels within PGCLC enhancers are shown as median profile (top) and heatmap plots (bottom). In the heatmap plots the PGCLC enhancers were ordered as in Fig. 2b. The mCpG data from ESC were obtained from Skvortsova *et al.* 2019[92].
e.) WT ESC and ESC lines with the indicated MLL3/4 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.
f.) Genome-browser view of the *Prdm14* locus showing the binding of MLL3/4 in WT ESC and the H3K4me1 and H3K27ac profiles in d2 EpiLC and d4 EB differentiated from WT and dCD ESC, respectively. The relevant *Prdm14* enhancers (*i.e.* E1-E3) are highlighted in blue.
g.) H3K27ac ChIP-seq experiments in d2 EpiLC, EpiSC and d4 EB differentiated from R1 WT and dCD ESC. H3K27ac levels within PGCLC enhancers are shown as median profile (top) and heatmap plots (bottom). In the heatmap plots the PGCLC enhancers were ordered as in Fig. 2b.
h.) Log_2 gene expression fold changes in dCD vs. R1 WT d4 EB for a subset of PGCLC regulators and markers (blue) as well as for genes representative of the Extraembryonic Ectoderm-like cells (ExE) and Extraembryonic Mesoderm-like cells (ExM) (gray). The log_2 fold changes were determined from three biological replicates with DESeq2. The error represents the standard error of the log_2 fold change estimation.
i.) Scatter Plot showing gene expression values (normalized counts; DESeq2) in dCD *vs* R1 WT d4EB. Non-differentially expressed genes are shown in gray, differentially expressed genes (adjusted p-value < 0.05) are indicated in orange and PGCLC genes are marked in red. Some of the PGCLC genes significantly downregulated in dCD d4EB are highlighted (i.e. *Esrrb*, *Prdm1*, *Prdm14* and *Dppa3*).
j.) Gene expression changes for the PGCLC genes in dCD *vs.* R1 WT d4 EB. PGCLC genes were separated into two groups depending on whether they were associated with a nearby PGCLC enhancer (n=113 genes) or not (n=38 genes). As a control group, 100 random genes were sampled. The log_2 fold changes for the PGCLC genes associated with enhancers was compared to the fold changes measured for PGCLC genes without enhancers or the randomly sampled genes, respectively, using one-tailed wilcoxon tests. Log_2 Fold Changes were estimated with DESeq2.

Supplementary Fig. 6d). In agreement with MLL3 and MLL4 being functionally redundant[70], such PGCLC differentiation defects were not observed when using cells that were catalytic deficient for MLL4 but not for MLL3 (*i.e.* 4CT) (Fig. 6e). Furthermore, the compromised PGCLC differentiation of the dCD cells was still observed at a later time point (day 6, Supplementary Fig. 6e), indicating that the observed defects are not simply explained by a delay in PGCLC specification.

As H3K4me1 priming has been suggested to facilitate enhancer activation[19,60,71], we addressed this question in the context of PGCLC differentiation by performing H3K27ac ChIP-seq in EpiLC, EpiSC and d4 EB derived from WT and dCD ESC (Fig. 6f,g and Supplementary Fig. 6f). In EpiLC and EpiSC, H3K27ac levels within PGCLC enhancers were similarly low in both WT and dCD cells. In contrast, most of the PGCLC enhancers, including those associated with major PGCLC regulators, gained H3K27ac in WT but not in dCD d4 EB, indicating that, upon PGCLC differentiation, PGCLC enhancers did not get properly activated in the absence of H3K4me1 (Fig. 6f,g and Supplementary Fig. 6f). Next, to investigate the transcriptional consequences of this defective PGCLC enhancer re-activation, we performed RNA-seq experiments in WT and dCD d4 EB (Supplementary Table 3). The expression of PGCLC genes in dCD cells was particularly reduced among genes associated with at least one PGCLC enhancer, which included relevant PGCLC markers and regulators (e.g. *Prdm1*, *Prdm14*, *Tfap2c, Dppa3*) (Fig. 6h-j). Nonetheless, gene expression changes in dCD cells were in general moderate and some PGCLC genes and enhancers were not affected (Fig. 6h, Supplementary Fig. 6f), which could perhaps be attributed to the compensatory activity of other histone methyltransferases (e.g. MLL1/KMT2A[72], MLL2/KMT2B[73,74]). Additionally, and

consistent with previous work[32], RNA-seq experiments performed in ESC, EpiLC and EpiSC revealed minor gene expression changes in dCD cells, with PGCLC genes being slightly down or upregulated in dCD ESC and EpiLC, respectively (Supplementary Fig. 6g,h and Supplementary Table 3). These minor, but measurable, gene expression changes in dCD ESC and EpiLC could also contribute to the observed PGCLC differentiation defects. In any case, our results show that H3K4me1/2 is required for proper PGCLC specification and supports the importance of PGCLC enhancer priming for germline competence.

H3K4me1/2 is required for the increased competence of OTX2 deficient epiblast cells

Recently, it has been shown that the deletion of *Otx2* markedly increases and prolongs germline competence in epiblast cells[16]. Therefore, we wondered whether the priming of PGCLC enhancers could be involved in the extended germline competence of *Otx2*$^{-/-}$ cells. Firstly, we found that OTX2 binds to PGCLC enhancers in both ESC and EpiLC (Supplementary Fig. 7a), suggesting that this TF might be directly involved in their decommissioning upon exit from naïve pluripotency[16]. Next, we confirmed the increased and prolonged germline competence of *Otx2*$^{-/-}$ cells, which can be robustly differentiated into PGCLC after keeping them for up to eight days in EpiSC culture conditions (Fig. 7a, Supplementary Fig. 7b). H3K4me1 and H3K4me2 ChIP-seq experiments revealed that the increased germline competence of *Otx2*$^{-/-}$ cells was coupled with the retention of H3K4me1/2 at PGCLC enhancers but did not affect EpiLC enhancers (Fig. 7b, Supplementary Fig. 7c,d). Moreover,

bisulfite sequencing of the representative *Esrrb* enhancer showed that the increased competence of *Otx2*$^{-/-}$ cells was also reflected in reduced mCpG levels within PGCLC enhancers (Fig. 7c). Although some PGCLC enhancers displayed elevated H3K27ac levels in *Otx2*$^{-/-}$ EpiLC (Supplementary Fig. 7e), this did not result in major gene expression changes (Supplementary Fig. 1g), suggesting that the increased germline competence of these cells could be preferentially linked to PGCLC enhancer priming rather than activation.

To directly assess whether the extended germline competence of *Otx2*$^{-/-}$ cells requires the priming of PGCLC enhancers by H3K4me1/2, we deleted *Otx2* in dCD ESC (i.e. dCD *Otx2*$^{-/-}$) as well as in their parental WT ESC (i.e. R1 *Otx2*$^{-/-}$) (Supplementary Fig. 7f) and differentiated them into PGCLC. Chiefly, both dCD and dCD *Otx2*$^{-/-}$ EpiLC showed a strong and similar reduction in their PGCLC differentiation capacity (Fig. 7d). Furthermore, genome-wide CpG methylation analysis revealed that PGCLC enhancers were considerably more methylated in dCD *Otx2*$^{-/-}$ EpiLC than in *Otx2*$^{-/-}$ EpiLC (Fig. 7e). Nevertheless, despite their reduced germline competence, dCD *Otx2*$^{-/-}$ EpiLC displayed mCpG levels within PGCLC enhancers comparable to those measured in WT cells (Fig. 7e, Supplementary Fig. 7g), suggesting that, in addition to protecting from CpG methylation, H3K4me1/2 might contribute to germline competence through additional regulatory mechanisms (e.g. permissive 3D chromatin architecture, Fig. 5e). Most importantly, these results further support the importance of H3K4me1/2 and enhancer priming for germline competence (Fig.7f).

Figure 7

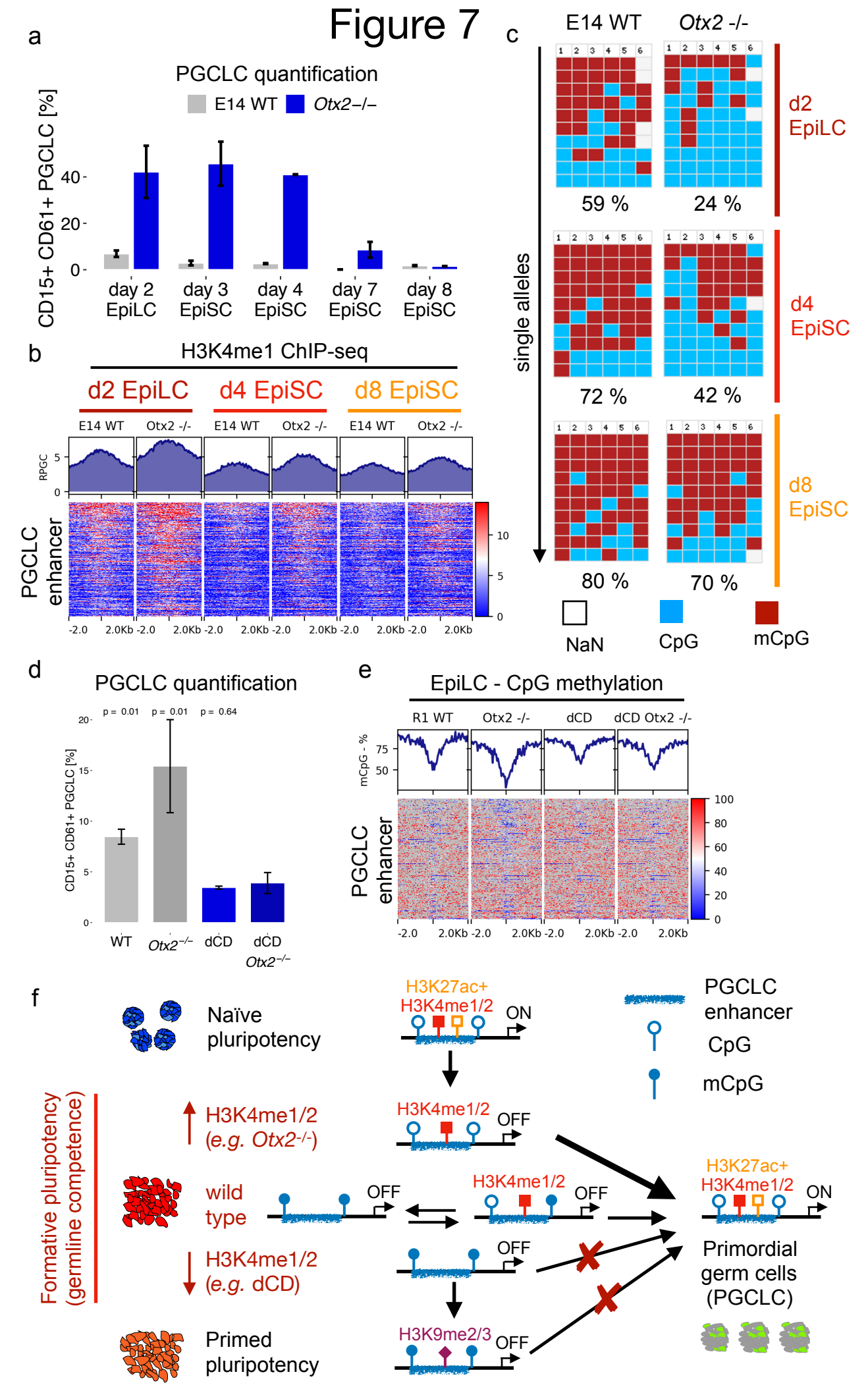

a
PGCLC quantification
E14 WT
Otx2–/–
CD15+ CD61+ PGCLC [%]
40
20
0
day 2 EpiLC
day 3 EpiSC
day 4 EpiSC
day 7 EpiSC
day 8 EpiSC
b
H3K4me1 ChIP-seq
d2 EpiLC
d4 EpiSC
d8 EpiSC
E14 WT
Otx2 -/-
RPGC
PGCLC enhancer
-2.0
2.0Kb
c
E14 WT
Otx2 -/-
single alleles
d2 EpiLC
d4 EpiSC
d8 EpiSC
59 %
24 %
72 %
42 %
80 %
70 %
NaN
CpG
mCpG
d
PGCLC quantification
p = 0.01
p = 0.01
p = 0.64
CD15+ CD61+ PGCLC [%]
WT
Otx2−/−
dCD
dCD Otx2−/−
e
EpiLC - CpG methylation
R1 WT
Otx2 -/-
dCD
dCD Otx2 -/-
mCpG - %
PGCLC enhancer
f
Naïve pluripotency
H3K27ac+ H3K4me1/2
ON
H3K4me1/2
OFF
PGCLC enhancer
CpG
mCpG
H3K4me1/2 (e.g. Otx2-/-)
wild type
H3K4me1/2 (e.g. dCD)
Primed pluripotency
Formative pluripotency (germline competence)
H3K9me2/3
Primordial germ cells (PGCLC)

Figure 7: The increased germline competence of *Otx2*$^{-/-}$ EpiLC requires H3K4me1/2.

a.) E14 WT and *Otx2*$^{-/-}$ ESC were differentiated into d2 EpiLC, d4 EpiSC or d8 EpiSC, which were then further differentiated into PGCLC. 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.) H3K4me1 ChIPseq experiments were performed in d2 EpiLC, d4 EpiSC and d8 EpiSC differentiated from E14 WT and *Otx2*$^{-/-}$ ESC. H3K4me1 levels within PGCLC enhancers are shown as median profile (top) and heatmap plots (bottom). In the heatmap plots the PGCLC enhancers were ordered as in Fig. 2b.

c.) The CpG methylation levels of the *Esrrb* enhancer were determined by bisulfite sequencing in the same cell types and stages described in b. The columns of the plots correspond to individual CpG dinucleotides located within the *Esrrb* enhancer. Unmethylated CpGs are shown in blue, methylated CpGs in red and CpGs which were not covered are shown in gray. At least 10 alleles were analyzed in each cell type (rows).

d.) R1 WT, *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.

e.) Genome-wide bisulfite sequencing experiments were performed in d2 EpiLC differentiated from R1 WT, *Otx2*$^{-/-}$, dCD and dCD *Otx2*$^{-/-}$ ESC. mCpG levels within PGCLC enhancers are shown as median profile (top) and heatmap plots (bottom). In the heatmap plots the PGCLC enhancers were ordered as in Fig. 2b.

f.) Model illustrating the priming of PGCLC enhancers during PGCLC differentiation and its relevance for germline competence. In the naïve pluripotent stage, PGCLC enhancers are active (H3K27ac + H3K4me1 + CpG hypermethylation), while with further development into the formative (EpiLC) and primed (EpiSC) pluripotent states, the PGCLC enhancers become either heterogeneously primed (intermediate H3K4me1 + CpG methylation) or fully decommissioned (H3K9me2/3 + CpG hypermethylation), respectively. Upon PGCLC differentiation, the primed but not the decommissioned PGCLC enhancers can get re-activated, thus indicating that PGCLC enhancer priming is important for germline competence. Congruently, *Otx2*$^{-/-}$ EpiLC display increase PGCLC enhancer priming and germline competence, while the opposite is true for H3K4me1/2 deficient EpiLC in which PGCLC specification is impaired.

DISCUSSION:

Enhancer priming by H3K4me1 has been suggested to precede and facilitate enhancer activation[17,19,60,71]. However, direct evidence supporting the functional relevance of H3K4me1 is scarce, partly due to the difficulties to separate the enzymatic and non-enzymatic functions of histone methyltransferases[75]. Recently, this limitation was overcome by establishing *Mll3/4* catalytic mutant ESC lines (*i.e.* dCD) in which H3K4me1 is lost from enhancers[32]. Under self-renewing conditions, the loss of H3K4me1 in ESC partly reduced H3K27ac but did not affect transcription from either enhancers or gene promoters, suggesting that H3K4me1 is dispensable for enhancer activity[32]. Here we showed that, upon PGCLC differentiation, the lack of H3K4me1/2 impairs PGCLC enhancer activation and the proper induction of major germline regulators. Together with previous studies reporting that the conditional loss of *Mll4* impaired the differentiation but not the maintenance of other progenitor cells[24,27,76], our work indicates that H3K4me1/2 might be important for enhancer priming but dispensable for maintaining enhancer activity.

Seemingly in contradiction with our findings, *Drosophila melanogaster* embryos expressing catalytic deficient *Trr*, the homolog of mammalian *Mll3/4,* develop normally, thus suggesting that H3K4me1 is dispensable for both enhancer activation and priming[34]. However, it is important to consider that, in comparison to mammalian cells, the *D. melanogaster* genome is largely devoid (< 0.03 %) of cytosine methylation throughout its entire life cycle[77]. Previous studies have extensively documented the protective role of H3K4me1/2 against CpG methylation and heterochromatinization[36,61–65]. Congruently, we found that H3K4me1/2 deficient cells

displayed increased mCpG levels within PGCLC enhancers. Therefore, although the importance of mCpG for enhancer activity and gene expression are not fully established[78–80], it is tempting to speculate that the lack of DNA methylation in flies render them less sensitive to H3K4me1/2 loss. Nevertheless, *Trr* catalytic mutant fly embryos displayed aberrant phenotypes under stress conditions, suggesting that H3K4me1/2 might help to fine-tune enhancer activity under suboptimal conditions. Similarly, we found that the induction of PGCLC genes in dCD cells was moderately and partly disrupted. Therefore, we propose that priming by H3K4me1/2 might facilitate, rather than being essential for, enhancer activation and the robust induction of developmental gene expression programs. This H3K4me1/2 facilitator role might involve not only protection from DNA methylation but also increased chromatin accessibility and physical proximity between genes and enhancers[25]. Future work in additional species and cellular transitions will establish the relevance and prevalence of H3K4me1/2 enhancer priming.

In the context of PGCLC differentiation, our scRNA-seq analysis suggests that, despite the remarkable transcriptional similarities between the naïve pluripotency and PGCLC states, germline competence is not attained until the naïve expression programme is robustly and homogeneously dismantled. Remarkably, the transcriptional homogeneity of the formative epiblast is accompanied by a simultaneous increase in epigenetic heterogeneity within enhancers. Hence, in contrast to ESC or the naïve epiblast, where CpG methylation levels within enhancers are negatively correlated with the expression of major naïve pluripotency regulators[54,81], in the formative epiblast the epigenetic heterogeneity within PGCLC enhancers seems to be uncoupled from the transcriptional status of the PGCLC

genes. Therefore, in formative epiblast cells, mCpG levels within enhancers might represent a marker for inferring both developmental timing and germline competence. Moreover, the epigenetic heterogeneity of PGCLC enhancers implies that epiblast cells might acquire germline competence in an unsynchronized and transient manner shortly after exiting the naïve pluripotent state, thus offering a plausible explanation as to why only a fraction of formative epiblast cells can be differentiated into PGCs when exposed to high inductive signals[4,5]. Similarly, the extended germline competence of epiblast cells under certain genetic or metabolic conditions[16,46] might be also explained by an increase in the fraction of cells in which PGCLC enhancers are epigenetically primed, as shown for *Otx2*$^{-/-}$ EpiLC (Fig. 7).

On the other hand, it is still unclear how PGCLC enhancers transiently acquire a primed state in formative epiblast cells. In ESC, most PGCLC enhancers are initially active and occupied by a broad set of regulatory proteins that includes core pluripotency (e.g. OCT4, SOX2), naïve pluripotency (e.g. ESRRB, PRDM14) and primed pluripotency TFs (e.g. FOXD3, OTX2)[7,50,82]. As ESC exit naïve pluripotency, PGCLC enhancers become deactivated (*i.e.* lose H3K27ac) due to the silencing of naïve TFs[81,83,84] and the repressive activity of primed TFs[7,85], which as a net result might lead to the loss of coactivators (e.g. p300) and the recruitment of corepressors (e.g. HDACs, LSD1). Nevertheless, PGCLC enhancers remain pervasively bound by OCT4 in EpiLC but not in EpiSC (Supplementary Fig. 2g). The persistent binding of OCT4 in EpiLC might inhibit the catalytic activity of the histone demethylase LSD1[86], thus potentially explaining how PGCLC enhancers retain H3K4me1/2 and acquire a primed state in EpiLC. Subsequently, a *tug-of-war*[87] might be established between OCT4 and the primed TFs (e.g. OTX2, FOXD3), enabling EpiLC to transiently

maintain a primed state within PGCLC enhancers and to acquire germline competence. OCT4 might eventually succumb to the repressive activity of the primed TFs favoring H3K4 demethylation, which can ultimately lead to the full decommissioning and heterochromatinization (*i.e.* DNA hypermethylation, H3K9me2/3 enrichment, loss of chromatin accessibility, tethering to the nuclear lamina[47]) of PGCLC enhancers and the closure of the germline competence window. Although the mechanistic details and *in vivo* relevance of this *tug-of-war* model need to be carefully dissected, it is supported by the increased priming of PGCLC enhancers in *Otx2*$^{-/-}$ EpiLC and d4 EpiSC as well as by the increased germline competence of *Otx2*$^{-/-}$ and *Foxd3*$^{-/-}$ cells[7,16].

Finally, together with previous reports, our work highlights the diverse mechanisms whereby the epigenetic status of enhancers can contribute to developmental competence. For instance, a subset of early brain enhancers displays a poised state in ESC (*i.e*. H3K4me1+H3K27me3) that might facilitate their activation upon neural differentiation[68], while in endodermal progenitors the binding of pioneer TFs and the consequent priming of relevant enhancers by H3K4me1 might facilitate differentiation[18,19]. In the case of germline competence, we uncovered that the priming of PGCLC enhancers in formative epiblast cells involves the transient and partial retention of H3K4me1 from a preceding active state. The mechanisms involved in this H3K4me1 retention are unknown, but we can envision at least two non-mutually exclusive scenarios: (i) an active maintenance mechanism similar to the one reported in other cell lineages[18,19], whereby the persistent binding of MLL3/4 and of certain TFs in EpiLC enables the retention of H3K4me1 within PGCLC enhancers; (ii) a passive mechanism whereby MLL3/4 binding to PGCLC enhancers

is already lost in EpiLC, but H3K4me1 can still be transiently retained due to the slow dynamics of H3K4 demethylation[86,88]. On the other hand, recent work based on chromatin accessibility or multi-omic profiling indicates that while the commitment towards certain cell lineages might involve the activation of already primed enhancers, in other cases this can occur through *de novo* enhancer activation[44,49]. Future work will elucidate the prevalence and regulatory mechanisms by which the epigenetic state of enhancers can contribute to cellular competence and the robust deployment of developmental gene expression programs.

ACKNOWLEDGMENTS:

We thank the Rada-Iglesias lab members for insightful comments and critical reading of the manuscript, Antonio Simeone and Christa Buecker for generously providing the *Otx2*$^{-/-}$ cell line.

Library preparations and next-generation sequencing experiments were performed in the NGS Core Facility of the Cologne Center for Genomics (CCG). Computational analyses were performed on the Cologne High Efficient Operating Platform for Science (CHEOPS).

Cell sorting and Flow cytometry experiments were performed in the FACS Facility of the Centre for Molecular Medicine (CMMC) and the FACS & Imaging Core Facility at the Max Planck Institute for Biology of Ageing.

FUNDING:

Tore Bleckwehl was supported by a doctoral fellowship from the *Studienstiftung des deutschen Volkes* (Germany). Kaitlin Schaaf was supported by a Research Internships in Science and Engineering (RISE) Scholarship of the Deutscher Akademischer Austauschdienst (DAAD). Work in the Rada-Iglesias laboratory was supported by CMMC intramural funding (Germany), the German Research Foundation (DFG) (Research Grant RA 2547/2-1), *"Programa STAR-Santander Universidades, Campus Cantabria Internacional de la convocatoria CEI 2015 de Campus de Excelencia Internacional"* (Spain) and the Spanish Ministry of Science, Innovation and Universities (Research Grant PGC2018-095301-B-I00). AB and MM were supported by the Spanish Ministry of Science, Innovation and Universities (BFU2017-84914-P), and the CNIC is supported by the Instituto de Salud Carlos III (ISCIII), the Spanish Ministry of Science, Innovation and Universities and the Pro CNIC Foundation, and is a Severo Ochoa Center of Excellence (SEV-2015-0505).

Author Contributions

Conceptualization, T.B., A.R-I.; Experimental investigation, T.B., K.S., G.C., P.R., M.B., S.J.C, L.B., A.B., M.L, K.D; Data analysis: T.B., G.C.; Writing, Review & Editing, T.B., A.R-I; Resources, M.M., J.W., W.R., A.R.-I.; Supervision and Funding Acquisition, A.R.-I.;

Declaration of Interests

The authors declare no competing interests.

Methods

Data availability

The sequencing datasets generated during this study will be made available through GEO upon publication.

Codes used in the computational analyses of the sequencing data are available upon request and will be made publicly available through https://github.com/ToreBle/PGC_enhancer_priming.

UCSC sessions to visualize all the processed sequencing data are available using the links provided in the Methods section.

WT and transgenic ESC lines

Two different male mouse ESC lines (*i.e.* E14Tg2a and R1) were used as indicated for each particular experiment. The *Mll3/4* dCD, *Mll3/4* dCT and *Mll4* CT ESC lines were derived from WT R1 ESC as described in Dorighi *et al.* 2017[32]. The E14 *Otx2*$^{-/-}$ ESC line [50,93] and its parental E14Tg2a ESC were kindly provided by Christa Buecker. Other transgenic ESC lines generated in this study are more extensively described in the following method details section.

Cell culture and differentiation protocols

All ESC lines were cultured under serum+LIF conditions both for regular maintenance and CRISPR/Cas9 genome editing experiments. More specifically,

ESC were cultured on gelatin-coated plates using Knock-out DMEM (*Life Technologies*) supplemented with 15% FBS (*Life Technologies*) and LIF. Before each PGCLC differentiation, ESC were adapted to 2i+LIF 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[94]. All measurements described in the manuscript for ESC were performed under 2i conditions.

For EpiLC and PGCLC differentiation, the protocols from Hayashi and Saitou [94] were followed. For the EpiLC differentiation, 6-well plates were coated with 15 μg ml^{-1} Fibronectin and 2 x 10^5 2i ESC were differentiated in N2B27 media supplemented with 20 ng ml^{-1} Activin A, 12 ng ml^{-1} bFGF and 10 μl ml^{-1} KSR for two days (unless indicated otherwise for particular experiments). For the PGCLC differentiation, 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, 0.1 μg ml^{-1} SCF, 1000 U ml^{-1} LIF and 50 ng ml^{-1} EGF, no BMP8a). Unless otherwise indicated, PGCLC were typically analyzed after four days.

EpiSC differentiation was performed according to Guo *et al.* 2009 [95]. 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 and differentiated into EpiSC by passaging them in N2B27 media supplemented with 20 ng ml^{-1} Activin A and 12 ng ml^{-1} bFGF for at least eight days or as indicated in the results section for particular experiments.

Enhancer and gene deletions with CRISPR/Cas9

Pairs of gRNAs flanking each of the selected enhancers/genes were designed with *benchling* (https://www.benchling.com). ESC were transfected with pairs of gRNA-Cas9 expressing vectors[96] specific for each enhancer/gene using Lipofectamine 3000 (*Invitrogen*). After transfection and puromycin selection, single cells were seeded into a 96 well plate to derive clonal cell lines. Then, clonal lines were genotyped by PCR and the presence of the intended deletions in each clonal line was verified by Sanger-sequencing. gRNA sequences and genomic coordinates of the different deletions are listed in Supplementary Table 4.

Generation of the DPPA3-GFP reporter ESC line

Based on the previous use of DPPA3 expression as a readout for PGCLC quantifications[5], we generated a DPPA3-GFP reporter ESC line as follows. First, we designed a repair template harbouring homologous arms to the C-terminus of *Dppa3* and sequences encoding the T2A-peptide and turboGFP (Supplementary Table 4). 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.

Generation of the Mll3/4 dCD, Mll3/4 dCT and Mll4 CT ESC lines

The *Mll3/4* dCD ESC line was previously generated and described[32]. The *Mll3/4* dCT and *Mll4* CT ESC lines were also generated as described in Dorighi *et al*. 2017[32]. Briefly, R1 ESC were co-transfected with a 200 bp single stranded oligonucleotide donor template harboring the desired point mutations and a vector coexpressing Cas9 and gRNAs targeting *Mll3* and *Mll4*. After 48 hr, single GFP+ cells were sorted into 96 well plates coated with fibronectin (5 mg ml^{-1}). The resulting clonal ESC lines were genotyped by PCR amplification of a region spanning the cleavage site, followed by digestion with a restriction enzyme predicted to only cut the wild-type PCR products. Once candidate ESC lines were selected, the presence of the intended *Mll3/4* mutations was confirmed by Sanger sequencing.

Ectopic and inducible expression of candidate transcription factors

Mouse *Prdm14* and *Nanog* cDNAs were amplified and cloned into a Doxycycline (Dox)-inducible piggyBac vector[97] enabling the ectopic expression of selected genes fused with a HA-tag at their C-terminus. The resulting piggyBac vectors were transfected together with a Super PiggyBac Transposase expressing vector (*Systems Bioscience*) and a Tet transactivator. 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 was 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 cells were collected after 48 hours; EpiSC were maintained in differentiation media for at least eight passages and then treated with Dox for 18 hours).

Quantification of PGCLC by flow cytometry

In general, PGCLC were quantified using antibody staining and flow cytometry[94]. 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*).

ATAC-seq

The ATAC-seq protocol was adapted from Buenrostro *et al.* 2013[98], 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 min. Following centrifugation the pellet was resuspended in a Tn5 transposase reaction mix (*Illumina*) for 30 min at 37°C. Following DNA purification with the MinElute PCR purification kit (*Qiagen*), libraries were prepared with the Nextera DNA library prep kit (*Illumina*).

ChIP-seq and ChIPmentation

Cells were cross-linked with 1 % formaldehyde for 10 min, followed by quenching with 0.125 M glycine, harvesting and washing in PBS containing protease inhibitors. 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 min 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 with specific antibodies (Supplementary Table 4) overnight, followed by immunoprecipitation with Dynabeads Protein G beads (*invitrogen*). Next, 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, 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*).

When ChIP-seq profiles were generated from low cell numbers (e.g. d4 EB obtained with the PGCLC differentiation protocol), the ChIPmentation protocol described by Schmidl *et al.* 2015[99] was adapted. Following immunoprecipitation with Dynabeads Protein G beads (*invitrogen*) in PCR tubes, samples were subject to Tagmentation (5 µl Tagmentation Buffer, 1 µl Tagmentation DNA Enzyme (*Illumina*), 19 µl Nuclease free water) in order to incorporate sequencing adapters. Lastly, DNAs were eluted from the beads and used for library preparation with the Nextera DNA library preparation kit (*Illumina*).

Locus-specific bisulfite sequencing

Genomic DNA was purified with phenol-chloroform and subjected to bisulfite conversion according to the EZ DNA Methylation-Direct™ Kit (Zymo Research). Then, primer pairs specific for each investigated enhancer (*Esrrb Enhancer*: TGTATTTTAGGTTAGTTTTTGAATTT and ACTAATCATCTCATTTTTAATCTACAC, 316 bp; *Lrrc31 Enhancer:* TTTAGGTATATGAGTATATTGTAGTTGT and TAAATCTTACTTCATTTAAAAAACC, 303 bp) were used for PCR amplification with the EpiTaq HS polymerase (*TaKaRa*). Finally, the resulting amplicons were gel-purified, subjected to blunt end cloning and analyzed by Sanger sequencing (at least 10 alleles were sequenced for each amplicon).

ChIP-bisulfite sequencing

ChIP-bisulfite experiments were performed as described in Thomson *et al.* 2010[100] 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.

Genome-wide bisulfite sequencing

Bisulfite sequencing libraries were prepared from column-purified DNA of d2 EpiLC using the PBAT method previously described in Clark *et al.* 2017[101]. 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 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 (*NEB*) 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 mastermix (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 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 and resuspended in 50 µl of PCR mastermix (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*).

4C-seq

Circular Chromatin Conformation Capture (4C) followed by sequencing for the *Prdm14* locus were performed as previously described[102]. Briefly, cells grown to confluence in 10 cm dishes were crosslinked with 1% formaldehyde for 10 min, 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 min. Nuclei were then re-suspended in 0.5 mL 1.2x CutSMART buffer (*NEB*) 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* (*NEB*) 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 min 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 min 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* (*NEB*) 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 *Prdm14* TSS (Supplementary Table 4). The inverse PCRs were performed with the expand long template PCR system (*Roche*) using 32 amplification cycles (94°C 2 min, 32x [94°C 10 s, 58°C 1 min, 68°C 3 min], 68°C 5 min). 4C-seq libraries were sequenced on the HiSeq2500, generating reads of 74 bp in length.

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-qPCR were performed on the Light Cycler 480II (*Roche*) with the primers listed in Supplementary Table 4 using *Eef1a1* and *Hprt* as housekeeping controls.

RNA-seq and scRNA-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 were generated following the protocol of the TruSeq stranded kit (*Illumina*).

For scRNA-seq, EpiSC were dissociated with Accutase and all other cell types with TripleExpress. Cells were then centrifuged and resuspended in PBS containing 0.04 % BSA. Next, cells were passed through a 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) protocol.

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 min at 4°C and then centrifuged for 10 min 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 min and centrifugation (24000 x g for 20 min at 4°C), supernatants were collected and protein concentration was estimated by a BCA-Assay. 20 µg of the resulting protein extracts were heated in Laemmli buffer at 95 °C for 5 min, 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

primary antibody (Supplementary Table 4) was incubated overnight at 4°C and the secondary antibody (coupled to a horseradish peroxidase (HRP)) for 1h at RT with washes in between. Finally, proteins were visualized using the lumi-light plus western blotting substrate (*Roche*).

scRNAseq data processing

The 10x Genomics scRNA-seq data generated in this study across different stages of PGCLC differentiation can be easily explored by opening the Supplementary Data 1 file with the *Loupe Cell Browser* (https://www.10xgenomics.com). UMIs were counted using NCBI:GCA_000001635.6 and *cellranger-2.1.0*[103]. The resulting UMI values were aggregated into a single matrix with default normalization (''–normalize=mapped'') (Supplementary Data 2).

For the scRNAseq data from E4.5 - E6.5 mouse embryos[44] the count matrix (GSE121650) was normalized to FPKM (*edgeR*, Ensembl gene annotation, v87). Then, the previously generated lineage assignments[44] were used to solely select epiblast cells for further analysis.

scRNAseq data analysis

The code used to define PGCLC genes will be fully available through Github. Briefly, *monocle2*[42] was used to evaluate the *in vitro* scRNAseq data generated across the different PGCLC differentiation stages. Thereby, 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[104]. 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[39] was used. Each EB cluster was annotated as equivalent to the mouse embryonic tissue for which we observed the most significant enrichment in the expression of the corresponding marker genes.

The EpiLC and PGCLC gene sets were defined by differential expression using *Seurat*[105] and the "negbinom" option for differential expression testing. EpiLC genes were determined by differential expression between 2i ESC and d2 EpiLC. From this analysis only the genes upregulated in d2 EpiLC (adjusted p-value < 0.001) and a low expression distribution in ESC (in less than 40% of ESC) were considered. Similar PGCLC genes were determined by differential expression between the d2+d4 PGCLC clusters and the remaining clusters from d2 and d4 EB as well as d2 EpiLC. Again, from this analysis only the genes upregulated in PGCLC (adjusted p-value < 0.001) and with a low expression distribution in the other analyzed cells (expressed in less than 40% of d2 EpiLC and non-PGCLC EB cells) were considered. In the case of PGCLC genes this resulted in 151 PGCLC genes (Supplementary Table 1). 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 (Supplementary Fig. 1e), spliced and unspliced read counts were obtained with *kallisto*[106] and *bustools*[107], 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*[41].

The estimation of transcriptional noise for the epiblast stages *in vivo* and *in vitro* was performed like in Mohammed *et al.* 2017[43]. First, the 500 most variable genes for each stage were selected and pairwise compared by Spearman correlations. Then the Spearman correlation values were transformed into distance measurements ($\sqrt{(1-\rho)/2}$) that were considered to represent transcriptional noise.

ChIP-seq data processing

For ChIP-seq data processing, single end reads were mapped to the mouse genome (*mm10*) using *BWA*[108]. After duplication removal with the *MarkDuplicates* function from the *Picard* tools (http://broadinstitute.github.io/picard/), reads within blacklisted regions (https://www.encodeproject.org/annotations/ENCSR636HFF/) were discarded and the aligned reads were normalized with *deeptools-3.3.1*[109] to 1x sequencing depth (as RPGC: Reads per genomic context). An overview of the ChIP-seq data generated is listed in Supplementary table 4 and the processed ChIP-seq data can be explored using the following UCSC browser sessions:

- Comparison of the pluripotent stages (related to Fig. 2):

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

- PRDM14/NANOG overexpression and topological features (related to Fig. 5):

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

- H3K4me1/2 deficiency and *Otx2*$^{-/-}$ (related to Fig. 6,7):

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

ATAC-seq data processing

For ATAC-seq data processing, paired end reads were mapped to the mouse genome (*mm10*) using *BWA*[108]. Read duplicates and reads within blacklisted regions were discarded.

ATAC-seq peaks from ESC, EpiLC and EpiSC were called with *MACS2*[110] (--nomodel --shift -75 --extsize 150) and filtered for peaks with a q-value < 1 x 10^{-5}. Given the concordance of the ATAC-seq replicas (Spearman correlation of a 2 kb window for ESC: 0.86; EpiLC: 0.88; EpiSC: 0.86), BAM files for each stage were merged and converted into bigWig files by normalization to 1x sequencing depth with *deeptools-3.3.1*[109]. The processed ATAC-seq data can be also found in the UCSC browser session mentioned above.

PGCLC Enhancer definition

The H3K27ac ChIP-seq data (SOLiD sequencing) from day2 and day6 sorted PGCLC[12] 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. The code for the definition of the PGCLC enhancers will be fully available through Github. Briefly, H3K27ac peaks were called

from both replicates of d2 and d6 PGCLC with MACS2[110] using broad settings (--broad -m 5 50 --fix-bimodal --extsize 200) and q-values < 1 x 10^{-5}. Next, all peaks from both replicates of d2 and d6 PGCLC were merged with *bedtools* (https://github.com/arq5x/bedtools2). The resulting regions were intersected with the merged ATAC-seq peaks from ESC, EpiLC and EpiSC which were generated as described above. Then the regions were further processed by a subtraction from blacklisted and promoter regions (-/+ 2 kb of the Ensembl gene annotation, v86). The intersection and subtraction of the bed files was done with *bedtools*. The resulting regions can be considered as general PGCLC enhancers that might be active in other cell types and control the expression of non-PGCLC specific genes. To obtain PGCLC specific enhancers, among the previous regions we selected those located within 100 Kb of one of the previously defined PGCLC genes (Supplementary Table 1) using *GREAT-4.0.4*[111]. Finally, only PGCLC enhancers with a minimum distance of 3.5 kb to the nearest transcription start site (TSS) were considered (Supplementary Table 2).

For the statistical analysis, we first determined the average signals of the ChIP-seq, ATAC-seq or genome-wide CpG bisulfite sequencing within -/+ 1kb of the PGCLC enhancers in ESC, EpiLC and EpiSC using *deeptools-3.3.1*[109].. Then, for the signal comparison relative to d2 EpiLC the p-values were estimated with a paired and two-sided wilcoxon test (confidence level: 0.95). The effective size of paired wilcoxon tests were 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.

CpG methylation analysis

The analysis of the local bisulfite sequencing experiments performed for selected PGCLC enhancers was performed with BISMA[112].

Genome-wide DNA methylation data was analysed with *Bismark*[113]. However, as the considered data sets were prepared with slightly different protocols, the preprocessing steps were adjusted accordingly: For the whole-genome bisulfite sequencing data from 2i+LIF ESC (GSE41923), d2 EpiLC and EpiSC (GSE70355), the adapter trimming was performed with *Trim Galore* (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) using the default settings; for data sets generated by post bisulfite adaptor tagging (pbat), either 9 (data from d2 EpiLC and PGCLC (DRA003471)) 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 (GSE76505), and, in this case, adapter sequences were removed with *cutadapt*[114] and *Trim Galore,* respectively. For all the previous samples, reads were mapped with *Bismark-v0.16.1*[113] and *bowtie2-2.2.9*[115], 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. Then, for each cell type all 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.

CpG methylation heterogeneity and scNMT-data analysis

With the scNMT-seq method the transcriptome, methylome (CpG methylation) and chromatin accessibility (GpC methylation) are recorded from the same single cell[116]. Among all the single cells analyzed in E4.5, E5.5 and E6.5 mouse embryos by Argelaguet *et al.* 2019[44], we only considered those assigned to the epiblast by scRNAseq (https://github.com/rargelaguet/scnmt_gastrulation). Then, epiblast cells were additionally filtered according to their methylome (GSE121690) and only cells with a CpG coverage > 10^5 were considered. This resulted in 258 epiblast cells from E4.5, E5.5 and E6.5 with high quality single cell CpG methylation data.

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 enhancers, the genome-wide bedGraph files were subtracted to obtain bedGraph files with the PGCLC enhancer regions (Supplementary Table 2) using *bedtools* (https://github.com/arq5x/bedtools2) from which the mean methylation of all covered CpGs was determined in each single cell. Next, the average PGCLC enhancer methylation for all the epiblast cells within a developmental stage was determined and the cells in each stage were classified as either *low* or *high* if their mean PGCLC enhancer methylation was below or above the stage average, respectively (Fig. 3f). From the parsed chromatin accessibility data, the mean mGpC level for each PGCLC enhancer within a -/+ 1kb window was determined and the values of all PGCLC enhancers were summarized for each cell. For the scRNAseq data, the mean expression of all PGCLC genes (FPKM-normalized) was determined per cell.

CpG methylation heterogeneity was estimated with the *PDclust* package[55]. The number of CpGs covered in each pair of cells being compared was determined, which when PGCLC enhancers were considered resulted in approximately 150 CpGs for each pairwise comparison. 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.

4C-seq

Reads were assigned to samples based on the first 10 bases of the read. 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*[117]. From these alignments, RPM (reads per million) normalized bedgraph files were generated for downstream visualization and analysis[118].

Bulk RNA-sequencing

Paired-end RNA-seq data was mapped with *STAR*[119] to the mouse reference genome (Ensembl gene annotation, v99) and reads within genes were counted with *featureCounts*[120]. The $\log_2$ fold changes (Supplementary Table 3), rlog and read count normalization were estimated with *DESeq2* from all replicates generated[121].

Supplemental items

Supplementary Data 1 - Loupe Cell Browser scRNA-seq file (opens with the *Loupe Cell Browser*: https://support.10xgenomics.com/single-cell-gene-expression/software/visualization/latest/installation)

Supplementary Data 2 - UMI count matrix of all cells analyzed by scRNA-seq

Supplementary Table 1 - PGCLC genes and mean expression

Supplementary Table 2 - PGCLC enhancers and linked genes

Supplementary Table 3 - RNA-seq results in WT vs Mll3/4 dCD cells

Supplementary Table 4 - Resources

References

1. Waddington, C. H. *Organisers and Genes by C. H. Waddington*. (1940).
2. Solini, G. E., Dong, C. & Saha, M. Embryonic transplantation experiments: Past, present, and future. *Trends Dev. Biol.* **10**, 13–30 (2017).
3. Ohinata, Y. *et al.* Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* **436**, 207–213 (2005).
4. Ohinata, Y. *et al.* A signaling principle for the specification of the germ cell lineage in mice. *Cell* **137**, 571–584 (2009).
5. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).
6. Bleckwehl, T. & Rada-Iglesias, A. Transcriptional and epigenetic control of germline

competence and specification. *Curr. Opin. Cell Biol.* **61**, 1–8 (2019).

7. Respuela, P. *et al.* Foxd3 Promotes Exit from Naive Pluripotency through Enhancer Decommissioning and Inhibits Germline Specification. *Cell Stem Cell* **18**, 118–133 (2016).
8. Hackett, J. A. *et al.* Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. *Nat. Commun.* **9**, 4292 (2018).
9. Murakami, K. *et al.* NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers. *Nature* **529**, 403–407 (2016).
10. Mitani, T. *et al.* Principles for the regulation of multiple developmental pathways by a versatile transcriptional factor, BLIMP1. *Nucleic Acids Res.* **45**, 12152–12169 (2017).
11. Shirane, K. *et al.* Global Landscape and Regulatory Principles of DNA Methylation Reprogramming for Germ Cell Specification by Mouse Pluripotent Stem Cells. *Dev. Cell* **39**, 87–103 (2016).
12. Kurimoto, K. *et al.* Quantitative Dynamics of Chromatin Remodeling during Germ Cell Specification from Mouse Embryonic Stem Cells. *Cell Stem Cell* **16**, 517–532 (2015).
13. von Meyenn, F. *et al.* Comparative Principles of DNA Methylation Reprogramming during Human and Mouse In Vitro Primordial Germ Cell Specification. *Dev. Cell* **39**, 104–115 (2016).
14. Morgani, S., Nichols, J. & Hadjantonakis, A.-K. The many faces of Pluripotency: in vitro adaptations of a continuum of in vivo states. *BMC Dev. Biol.* **17**, 7 (2017).
15. Kalkan, T. *et al.* Complementary Activity of ETV5, RBPJ, and TCF3 Drives Formative Transition from Naive Pluripotency. *Cell Stem Cell* **24**, 785–801.e7 (2019).
16. Zhang, J. *et al.* OTX2 restricts entry to the mouse germline. *Nature* **562**, 595–599 (2018).
17. Lara-Astiaso, D. *et al.* Immunogenetics. Chromatin state dynamics during blood formation. *Science* **345**, 943–949 (2014).

18. Lee, K. *et al.* FOXA2 Is Required for Enhancer Priming during Pancreatic Differentiation. *Cell Rep.* **28**, 382–393.e7 (2019).
19. Wang, A. *et al.* Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell* **16**, 386–399 (2015).
20. Lai, B. *et al.* MLL3/MLL4 are required for CBP/p300 binding on enhancers and super-enhancer formation in brown adipogenesis. *Nucleic Acids Res.* **45**, 6388–6403 (2017).
21. Sze, C. C. & Shilatifard, A. MLL3/MLL4/COMPASS Family on Epigenetic Regulation of Enhancer Function and Cancer. *Cold Spring Harb. Perspect. Med.* **6**, (2016).
22. Bochyńska, A., Lüscher-Firzlaff, J. & Lüscher, B. Modes of Interaction of KMT2 Histone H3 Lysine 4 Methyltransferase/COMPASS Complexes with Chromatin. *Cells* **7**, (2018).
23. Lin-Shiao, E. *et al.* KMT2D regulates p63 target enhancers to coordinate epithelial homeostasis. *Genes Dev.* **32**, 181–193 (2018).
24. Wang, C. *et al.* Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11871–11876 (2016).
25. Yan, J. *et al.* Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers. *Cell Res.* **28**, 387 (2018).
26. Jang, Y. *et al.* H3.3K4M destabilizes enhancer H3K4 methyltransferases MLL3/MLL4 and impairs adipose tissue development. *Nucleic Acids Res.* **47**, 607–620 (2019).
27. Placek, K. *et al.* MLL4 prepares the enhancer landscape for Foxp3 induction via chromatin looping. *Nat. Immunol.* **18**, 1035–1045 (2017).
28. Ang, S.-Y. *et al.* KMT2D regulates specific programs in heart development via histone H3 lysine 4 di-methylation. *Development* **143**, 810–821 (2016).
29. Ortega-Molina, A. *et al.* The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nat. Med.* **21**, 1199–1208 (2015).

30. Zhang, J. *et al.* Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. *Nat. Med.* **21**, 1190–1198 (2015).
31. Wang, S.-P. *et al.* A UTX-MLL4-p300 Transcriptional Regulatory Network Coordinately Shapes Active Enhancer Landscapes for Eliciting Transcription. *Mol. Cell* **67**, 308–321.e6 (2017).
32. Dorighi, K. M. *et al.* Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Mol. Cell* **66**, 568–576.e4 (2017).
33. Cao, K. *et al.* An Mll4/COMPASS-Lsd1 epigenetic axis governs enhancer function and pluripotency transition in embryonic stem cells. *Sci Adv* **4**, eaap8747 (2018).
34. Rickels, R. *et al.* Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nat. Genet.* **49**, 1647–1653 (2017).
35. Local, A. *et al.* Identification of H3K4me1-associated proteins at mammalian enhancers. *Nat. Genet.* **50**, 73–82 (2018).
36. Ooi, S. K. T. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* vol. 448 714–717 (2007).
37. Rada-Iglesias, A. Is H3K4me1 at enhancers correlative or causative? *Nature genetics* vol. 50 4–5 (2018).
38. Mulas, C., Kalkan, T. & Smith, A. NODAL Secures Pluripotency upon Embryonic Stem Cell Progression from the Ground State. *Stem Cell Reports* **9**, 77–91 (2017).
39. Ibarra-Soria, X. *et al.* 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 (2018).
40. Yang, P. *et al.* Multi-omic Profiling Reveals Dynamics of the Phased Progression of Pluripotency. *Cell Syst* **8**, 427–445.e10 (2019).

41. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
42. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
43. Mohammed, H. *et al.* Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. *Cell Rep.* **20**, 1215–1228 (2017).
44. Argelaguet, R. *et al.* Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* (2019) doi:10.1038/s41586-019-1825-8.
45. Zylicz, J. J. *et al.* Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. *Elife* **4**, (2015).
46. Tischler, J. *et al.* Metabolic regulation of pluripotency and germ cell fate through α-ketoglutarate. *The EMBO Journal* vol. 38 (2019).
47. Poleshko, A. *et al.* Genome-Nuclear Lamina Interactions Regulate Cardiac Stem Cell Lineage Restriction. *Cell* **171**, 573–587.e14 (2017).
48. Lochs, S. J. A., Kefalopoulou, S. & Kind, J. Lamina Associated Domains and Gene Regulation in Development and Cancer. *Cells* **8**, (2019).
49. Rauch, A. *et al.* Author Correction: Osteogenesis depends on commissioning of a network of stem cell transcription factors that act as repressors of adipogenesis. *Nat. Genet.* **51**, 766 (2019).
50. Buecker, C. *et al.* Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838–853 (2014).
51. Matsuda, K. *et al.* ChIP-seq analysis of genomic binding regions of five major transcription factors highlights a central role for ZIC2 in the mouse epiblast stem cell gene regulatory network. *Development* **144**, 1948–1958 (2017).
52. Rose, N. R. & Klose, R. J. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochim. Biophys. Acta* **1839**, 1362–1372 (2014).
53. Zhang, Y. *et al.* Dynamic epigenomic landscapes during early lineage specification in

mouse embryos. *Nat. Genet.* **50**, 96–105 (2018).

54. Rulands, S. *et al.* Genome-Scale Oscillations in DNA Methylation during Exit from Pluripotency. *Cell Syst* **7**, 63–76.e12 (2018).
55. Hui, T. *et al.* High-Resolution Single-Cell DNA Methylation Measurements Reveal Epigenetically Distinct Hematopoietic Stem Cell Subpopulations. *Stem Cell Reports* vol. 11 578–592 (2018).
56. Hnisz, D. *et al.* Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol. Cell* **58**, 362–370 (2015).
57. Yamaji, M. *et al.* Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat. Genet.* **40**, 1016–1022 (2008).
58. Yamaji, M. *et al.* PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. *Cell Stem Cell* **12**, 368–382 (2013).
59. Ma, Z., Swigut, T., Valouev, A., Rada-Iglesias, A. & Wysocka, J. Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. *Nat. Struct. Mol. Biol.* **18**, 120–127 (2011).
60. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21931–21936 (2010).
61. Hughes, A. L., Kelley, J. R. & Klose, R. J. Understanding the interplay between CpG island-associated gene promoters and H3K4 methylation. *Biochim. Biophys. Acta Gene Regul. Mech.* **1863**, 194567 (2020).
62. Guo, X. *et al.* Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* **517**, 640–644 (2015).
63. Zhang, Y. *et al.* 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 (2010).

64. Li, B.-Z. *et al.* Histone tails regulate DNA methylation by allosterically activating de novo methyltransferase. *Cell Res.* **21**, 1172–1181 (2011).

65. Nady, N. *et al.* Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. *J. Biol. Chem.* **286**, 24300–24311 (2011).

66. Magnúsdóttir, E. *et al.* A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nat. Cell Biol.* **15**, 905–915 (2013).

67. Melo, C. A. *et al.* eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol. Cell* **49**, 524–535 (2013).

68. Cruz-Molina, S. *et al.* PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell* **20**, 689–705.e9 (2017).

69. Schoenfelder, S. *et al.* Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat. Genet.* **47**, 1179–1186 (2015).

70. Hu, D. *et al.* The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol. Cell. Biol.* **33**, 4745–4754 (2013).

71. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).

72. Terranova, R., Agherbi, H., Boned, A., Meresse, S. & Djabali, M. 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 (2006).

73. Hu, D. *et al.* Not All H3K4 Methylations Are Created Equal: Mll2/COMPASS Dependency in Primordial Germ Cell Specification. *Mol. Cell* **65**, 460–475.e6 (2017).

74. Morgan, M. A. J. *et al.* A cryptic Tudor domain links BRWD2/PHIP to COMPASS-mediated histone H3K4 methylation. *Genes Dev.* **31**, 2003–2014 (2017).

75. Aubert, Y., Egolf, S. & Capell, B. C. The Unexpected Noncatalytic Roles of Histone Modifiers in Development and Disease. *Trends Genet.* **35**, 645–657 (2019).

76. Lee, J.-E. *et al.* H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *Elife* **2**, e01503 (2013).

77. Deshmukh, S., Ponnaluri, V. C., Dai, N., Pradhan, S. & Deobagkar, D. Levels of DNA cytosine methylation in the genome. *PeerJ* **6**, e5119 (2018).

78. Barnett, K. R. *et al.* ATAC-Me Captures Prolonged DNA Methylation of Dynamic Chromatin Accessibility Loci during Cell Fate Transitions. *Mol. Cell* **77**, 1350–1364.e6 (2020).

79. Sapozhnikov, D. M. & Szyf, M. Unraveling the functional role of DNA methylation using targeted DNA demethylation by steric blockage of DNA methyltransferase with CRISPR/dCas9. *bioRxiv 2020*, doi:10.1101/2020.03.28.012518.

80. Ford, E. *et al.* Frequent lack of repressive capacity of promoter DNA methylation identified through genome-wide epigenomic manipulation. *bioRxiv 2020*, doi:10.1101/170506.

81. Bell, E. *et al.* Dynamic CpG methylation delineates subregions within super-enhancers selectively decommissioned at the exit from naive pluripotency. *Nature Communications* vol. 11 (2020).

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

83. Atlasi, Y. *et al.* Epigenetic modulation of a hardwired 3D chromatin landscape in two naive states of pluripotency. *Nat. Cell Biol.* **21**, 568–578 (2019).

84. Adachi, K. *et al.* Esrrb Unlocks Silenced Enhancers for Reprogramming to Naive Pluripotency. *Cell Stem Cell* **23**, 266–275.e6 (2018).

85. Krishnakumar, R. *et al.* FOXD3 Regulates Pluripotent Stem Cell Potential by Simultaneously Initiating and Repressing Enhancer Activity. *Cell Stem Cell* **23**, 306–307 (2018).

86. AlAbdi, L. *et al.* Oct4-Mediated Inhibition of Lsd1 Activity Promotes the Active and

Primed State of Pluripotency Enhancers. *Cell Rep.* **30**, 1478–1490.e6 (2020).

87. Buecker, C. & Wysocka, J. Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet.* **28**, 276–284 (2012).
88. Maltby, V. E. *et al.* Histone H3K4 demethylation is negatively regulated by histone H3 acetylation in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 18505–18510 (2012).
89. Zylicz, J. J. *et al.* G9a regulates temporal preimplantation developmental program and lineage segregation in blastocyst. *Elife* **7**, (2018).
90. Habibi, E. *et al.* Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360–369 (2013).
91. Hui, T. *et al.* High-Resolution Single-Cell DNA Methylation Measurements Reveal Epigenetically Distinct Hematopoietic Stem Cell Subpopulations. *Stem Cell Reports* **11**, 578–592 (2018).
92. Skvortsova, K. *et al.* DNA Hypermethylation Encroachment at CpG Island Borders in Cancer Is Predisposed by H3K4 Monomethylation Patterns. *Cancer Cell* **35**, 297–314.e8 (2019).
93. Acampora, D., Di Giovannantonio, L. G. & Simeone, A. Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. *Development* **140**, 43–55 (2013).
94. Hayashi, K. & Saitou, M. Generation of eggs from mouse embryonic stem cells and induced pluripotent stem cells. *Nat. Protoc.* **8**, 1513–1524 (2013).
95. Guo, G. *et al.* Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* **136**, 1063–1069 (2009).
96. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
97. Calo, E. *et al.* Tissue-selective effects of nucleolar stress and rDNA damage in

developmental disorders. *Nature* **554**, 112–117 (2018).

98. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
99. Schmidl, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat. Methods* **12**, 963–965 (2015).
100. Thomson, J. P. *et al.* CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**, 1082–1086 (2010).
101. Clark, S. J. *et al.* Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nat. Protoc.* **12**, 534–547 (2017).
102. Laugsch, M. *et al.* Modeling the Pathological Long-Range Regulatory Effects of Human Structural Variation with Patient-Specific hiPSCs. *Cell Stem Cell* **24**, 736–752.e12 (2019).
103. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
104. Nakaki, F. *et al.* Induction of mouse germ-cell fate by transcription factors in vitro. *Nature* **501**, 222–226 (2013).
105. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).
106. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
107. Melsted, P. *et al.* Modular and efficient pre-processing of single-cell RNA-seq. *bioRxiv 2020*, doi:10.1101/673285.
108. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler

transform. *Bioinformatics* **25**, 1754–1760 (2009).

109. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–5 (2016).

110. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

111. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology* vol. 28 495–501 (2010).

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

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

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

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

116. Clark, S. J. *et al.* scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* **9**, 781 (2018).

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

118. Thongjuea, S., Stadhouders, R., Grosveld, F. G., Soler, E. & Lenhard, B. 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 (2013).

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

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

121. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Nature Precedings* (2010) doi:10.1038/npre.2010.4282.1.

Supplementary Figures

Enhancer priming by H3K4 methylation safeguards germline competence

Tore Bleckwehl, Kaitlin Schaaf, Giuliano Crispatzu, Patricia Respuela, Michaela Bartusel, Laura Benson, Stephen J. Clark, Kristel M. Dorighi, Antonio Barral, Magdalena Laugsch, Miguel Manzanares, Joanna Wysocka, Wolf Reik, Álvaro Rada-Iglesias

Figure S1

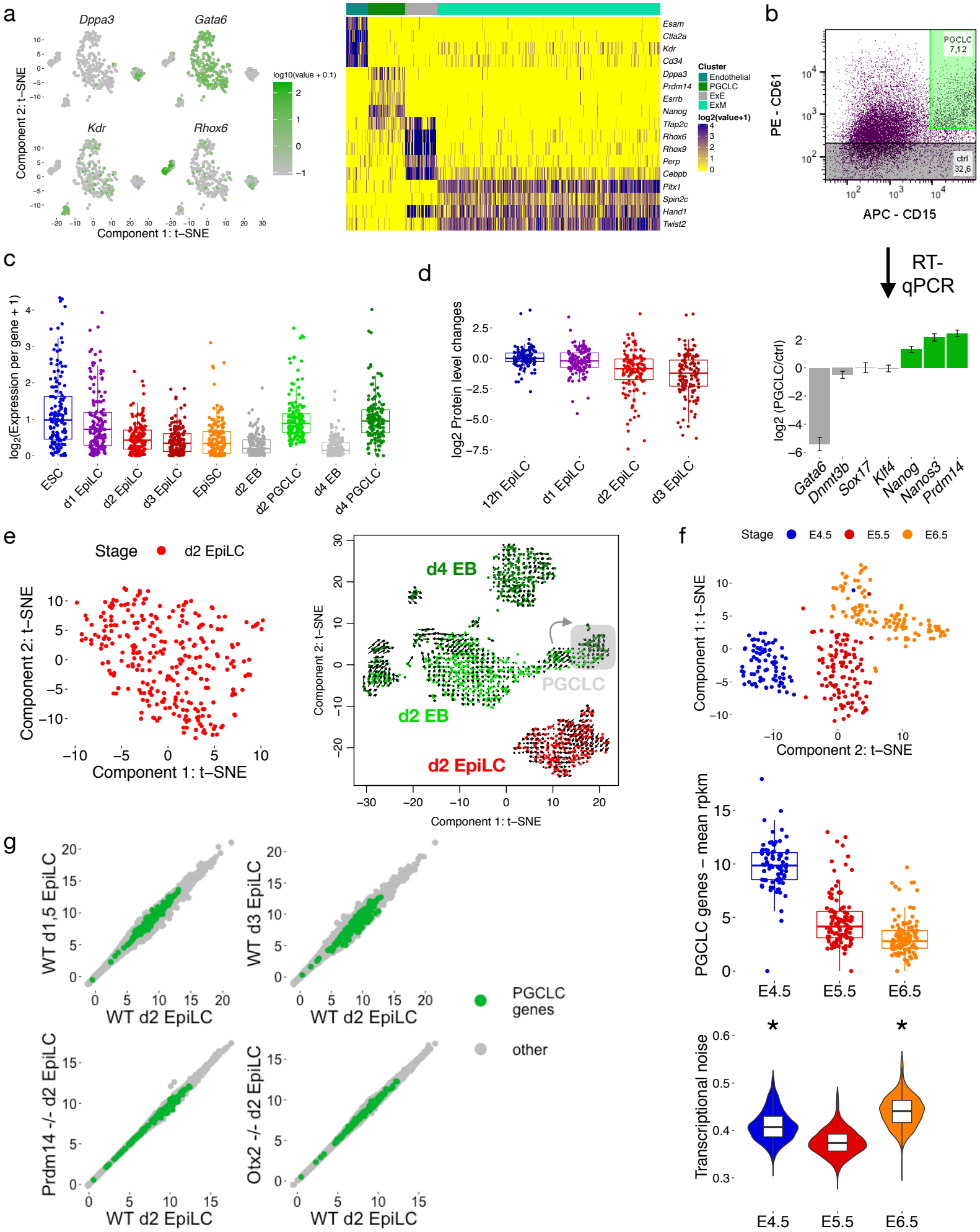

a
Dppa3
Gata6
Kdr
Rhox6
Component 2: t–SNE
Component 1: t–SNE
log10(value + 0.1)
Esam
Ctla2a
Kdr
Cd34
Dppa3
Prdm14
Esrrb
Nanog
Tfap2c
Rhox6
Rhox9
Perp
Cebpb
Pitx1
Spin2c
Hand1
Twist2
Cluster
Endothelial
PGCLC
ExE
ExM
log2(value+1)
b
PE - CD61
APC - CD15
PGCLC
7,12
ctrl
32,6
c
log2(Expression per gene + 1)
ESC
d1 EpiLC
d2 EpiLC
d3 EpiLC
EpiSC
d2 EB
d2 PGCLC
d4 EB
d4 PGCLC
d
log2 Protein level changes
12h EpiLC
d1 EpiLC
d2 EpiLC
d3 EpiLC
RT-qPCR
log2 (PGCLC/ctrl)
Gata6
Dnmt3b
Sox17
Klf4
Nanog
Nanos3
Prdm14
e
Stage
d2 EpiLC
Component 2: t–SNE
Component 1: t–SNE
d4 EB
d2 EB
PGCLC
d2 EpiLC
f
Stage
E4.5
E5.5
E6.5
Component 1: t–SNE
Component 2: t–SNE
PGCLC genes – mean rpkm
Transcriptional noise
g
WT d1,5 EpiLC
WT d3 EpiLC
Prdm14 -/- d2 EpiLC
Otx2 -/- d2 EpiLC
WT d2 EpiLC
PGCLC genes
other

Figure S1: Characterization of the PGCLC differentiation system by scRNA-seq (related to Figure 1).

a.) t-SNE plot based on the scRNA-seq data generated in d4 EB (n = 368 cells) (left) showing the expression of a representative marker for each main cluster. For each of the main four clusters described in Fig. 1c, the expression of representative gene markers are shown for individual cells as a UMI count heatmap (right) .

b.) D4 EB cells were sorted by FACS into $CD15^{+}CD61^{+}$ (*i.e.* PGCLC) or $CD61^{-}$ (*i.e.* non-PGCLC control cells). Then, the expression of the indicated genes, including the PGC markers *Prdm14*, *Nanos3* and *Nanog,* was determined by RT-qPCR in those two cell populations. The error bars represent standard deviations from three technical replicates.

c.) Expression dynamics of the PGCLC genes during PGCLC formation. Each dot represents the mean expression of one PGCLC gene as measured in all individual cells belonging to the indicated stages. EB refers to any cell of the EB except those identified as PGCLC.

d.) Box plots displaying the protein levels measured for the PGCLC genes at different time points during the differentiation of ESC into EpiLC. Dots represent the mean protein levels expressed by each PGCLC gene. The data for each EpiLC time point was normalized to the protein levels measured in ESC and presented as $\log_2$ values. The proteomic data was obtained from Yang *et al.* 2019[40].

e.) Computational approaches used to investigate potential transcriptional priming of the PGCLC expression program in EpiLC. Left: The t-SNE plot of d2 EpiLC scRNA-seq data did not reveal any obvious clusters suggestive of differential cell identity based on gene expression. Right: The RNA velocity analysis, which takes into account unspliced transcript reads for lineage tracing[41], did not show any evidences of transcriptional priming of d2 EpiLC (red) towards the PGCLC cluster (gray). In contrast, this same analysis revealed that a fraction of the cells profiled for the d2 EB (with a mixed mesodermal/PGCLC identity; gray arrow) seem to acquire a transcriptional program similar to the one found for PGCLC (shadowed in gray), indicating the feasibility of the RNA velocity method.

f.) The dismantling of the PGCLC genes upon exit from the naïve pluripotent state (E4.5) *in vivo* was evaluated using scRNA-seq data obtained from Argelaguet *et al.* 2019[44]. Top: tSNE plot based on scRNA-seq data from E4.5, E5.5 and E6.5 epiblasts. Middle: The dots represent the mean expression for all PGCLC genes in individual cells belonging to the indicated stages. Bottom: The violin plots show the transcriptional noise, defined as cell-to-cell transcript variation for the 500 most variable genes, for E4.5, E5.5 and E6.5 epiblast cells. The * indicates that the transcriptional noise for the indicated stage was significantly different in comparison to E5.5 (wilcoxon test, *: $p < 2.2\ 10^{-16}$).

g.) Scatter plots comparing the transcriptomes of WT d2 EpiLC (x-axis) with those of d1.5 EpiLC, d3 EpiLC, *Prdm14*$^{-/-}$ d2 EpiLC or *Otx2*$^{-/-}$ d2 EpiLC (y-axis). All genes considered as expressed are shown as dots, with the PGCLC genes highlighted in green. The gene expression values are r-log normalized (DESeq2). The RNA-seq data was obtained from[11,40,50].

Figure S2

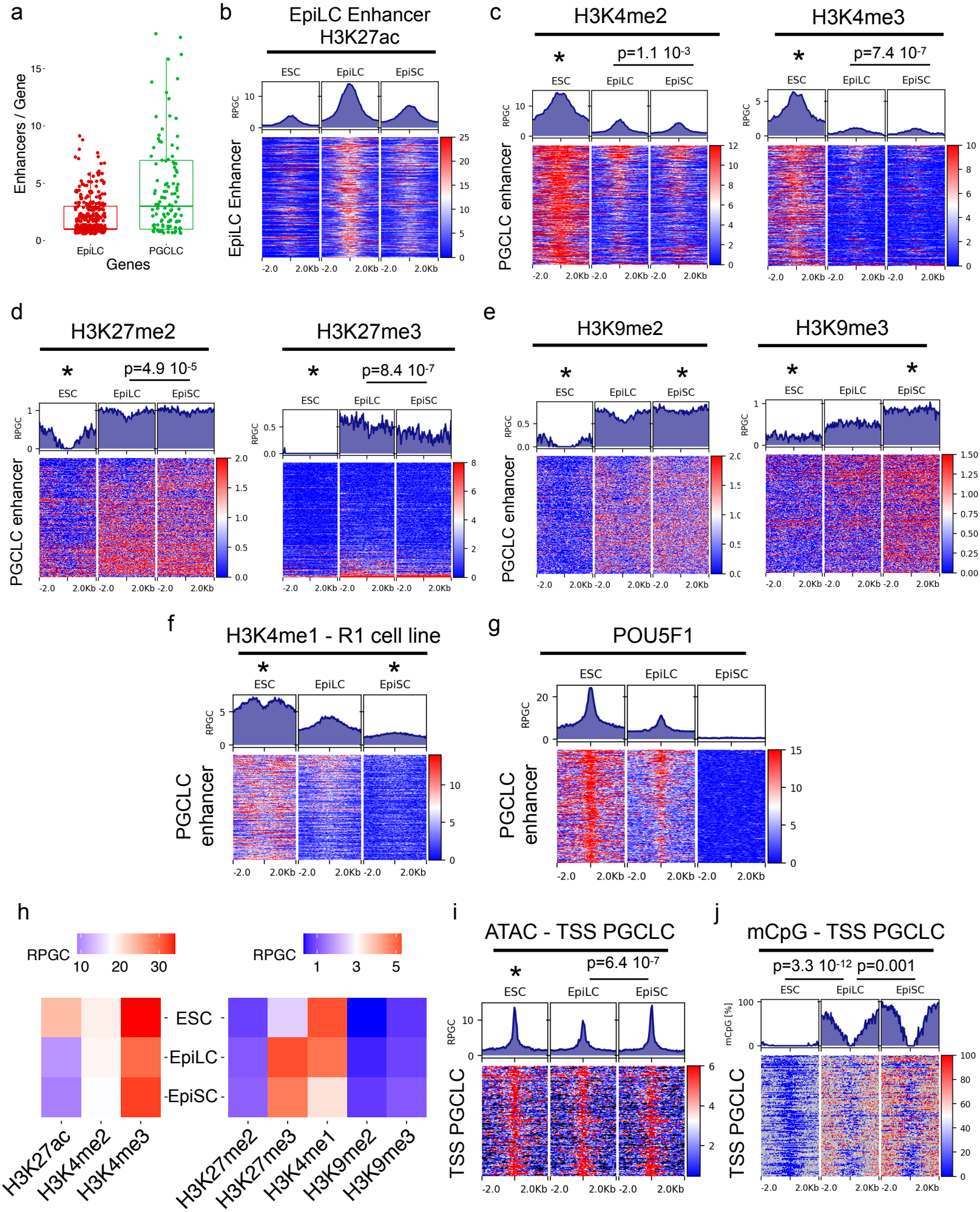

a
Enhancers / Gene
EpiLC
PGCLC
Genes
b
EpiLC Enhancer
H3K27ac
ESC
EpiLC
EpiSC
RPGC
EpiLC Enhancer
c
H3K4me2
p=1.1 10-3
H3K4me3
p=7.4 10-7
PGCLC enhancer
d
H3K27me2
p=4.9 10-5
H3K27me3
p=8.4 10-7
e
H3K9me2
H3K9me3
f
H3K4me1 - R1 cell line
g
POU5F1
h
RPGC
ESC
EpiLC
EpiSC
H3K27ac
H3K4me2
H3K4me3
H3K27me2
H3K27me3
H3K4me1
H3K9me2
H3K9me3
i
ATAC - TSS PGCLC
p=6.4 10-7
TSS PGCLC
j
mCpG - TSS PGCLC
p=3.3 10-12
p=0.001
mCpG [%]

Figure S2: Epigenomic profiles of PGCLC enhancers and promoters in different pluripotent states (related to Figure 2).

a.) The box plots show the number of active enhancers associated per PGCLC or d2 EpiLC gene, respectively. The PGCLC and EpiLC genes were defined based on their differential expression in comparison to EB and d2 EpiLC or ESC, respectively, using the scRNA-seq data described in Fig. 1. Enhancers were linked to genes by proximity as illustrated in Fig. 2a. EpiLC and PGCLC genes with at least one associated enhancer were considered.

b.) H3K27ac signals within EpiLC enhancers are shown in ESC, EpiLC and EpiSC as either median profile (top) or heatmap (bottom) plots (n = 689 enhancers).

c-e.) ChIP-seq signals for (c) active histone modification (*i.e.* H3K4me2, H3K4me3), (d) polycomb associated histone modifications (*i.e.* H3K27me2 and H3K27me3) and (e) heterochromatic histone modifications (*i.e.* H3K9me2 and H3K9me3) within PGCLC enhancers are shown in ESC, EpiLC and EpiSC as median profile (top) and heatmap plots (bottom). In the heatmaps plots the PGCLC enhancers were ordered as in Fig. 2b. The ChIP-seq signals of each PGCLC enhancer (using a -/+ 1kb window) in EpiLC were compared to those measured in ESC or EpiSC using paired wilcoxon tests (n = 511, *: $p < 2.2\ 10^{-16}$).

f.) H3K4me1 ChIP-seq experiments were performed in R1 ESC as well as upon their differentiation into EpiLC and EpiSC. The H3K4me1 ChIP-seq signals within PGCLC enhancers are shown as median profile (top) and heatmap plots (bottom). The ChIP-seq signals of each PGCLC enhancer (using a -/+ 1kb window) in EpiLC were compared to those measured in ESC or EpiSC using paired wilcoxon tests (n = 511, *: $p < 2.2\ 10^{-16}$).

g.) ChIP-seq signals for POU5F1 (OCT4) in ESC, EpiLC and EpiSC. POU5F1 levels within PGCLC enhancers are shown as median profile (top) and heatmap plots (bottom). In the heatmap plots the PGCLC enhancers were ordered as in Fig. 2b. The POU5F1 ChIP-seq data were obtained from Buecker *et al.* 2014[50] and Matsuda *et al*. 2017[51].

h.) Summary of the ChIP-seq signals measured around the TSS of the PGCLC genes in ESC, EpiLC and EpiSC. The average ChIP-seq signals were calculated around the TSS of all the PGCLC genes using a -/+ 1kb window (RPGC: Reads per genomic content).

i.) Chromatin accessibility around the TSS of the PGCLC genes is shown in ESC, EpiLC and EpiSC as median profile (top) and heatmap plots (bottom). The ATAC-seq signals around the TSS of each PGCLC gene (using a -/+ 1kb window) in EpiLC were compared to those measured in ESC or EpiSC using paired wilcoxon tests (n = 511).

j.) CpG methylation levels around the TSS of PGCLC genes are shown in ESC, EpiLC and EpiSC as median profile (top) and heatmap plots (bottom). The mCpG levels around the TSS of each PGCLC gene (using a -/+ 1kb window) in EpiLC were compared to those measured in ESC or EpiSC using paired wilcoxon tests (n = 511). The CpG methylation data was obtained from Zylicz *et al.* 2018[89,90].

Figure S3

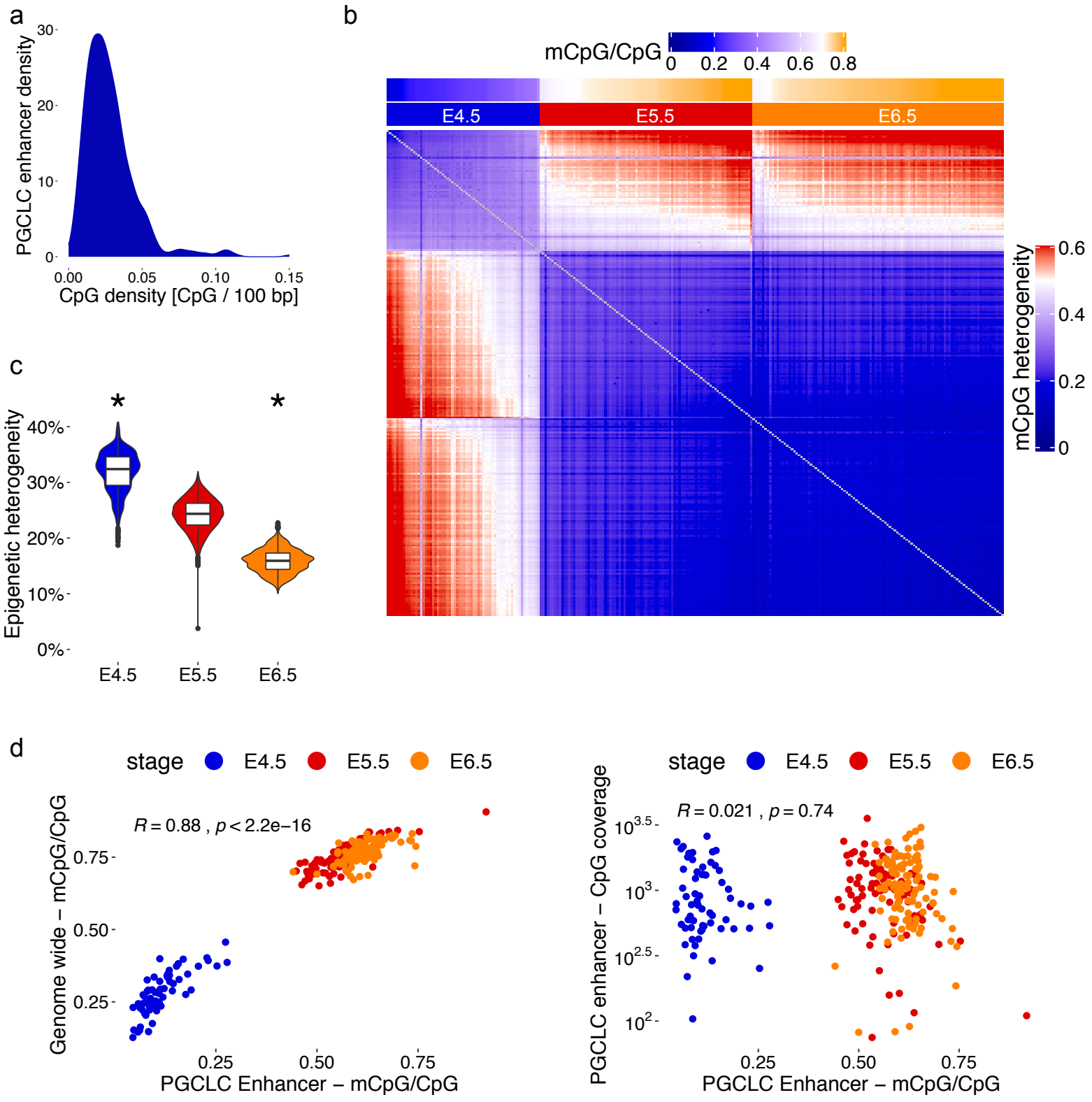

a
PGCLC enhancer density
30
20
10
0
0.00
0.05
0.10
0.15
CpG density [CpG / 100 bp]
b
mCpG/CpG
0
0.2
0.4
0.6
0.8
E4.5
E5.5
E6.5
mCpG heterogeneity
0.6
0.4
0.2
0
c
Epigenetic heterogeneity
40%
30%
20%
10%
0%
*
*
E4.5
E5.5
E6.5
d
stage
E4.5
E5.5
E6.5
R = 0.88 , p < 2.2e−16
Genome wide – mCpG/CpG
0.75
0.50
0.25
0.25
0.50
0.75
PGCLC Enhancer – mCpG/CpG
stage
E4.5
E5.5
E6.5
R = 0.021 , p = 0.74
PGCLC enhancer – CpG coverage
10^3.5
10^3
10^2.5
10^2
0.25
0.50
0.75
PGCLC Enhancer – mCpG/CpG

Figure S3: Analysis of PGCLC enhancers using *in vivo* single-cell CpG methylation data (related to Figure 3).

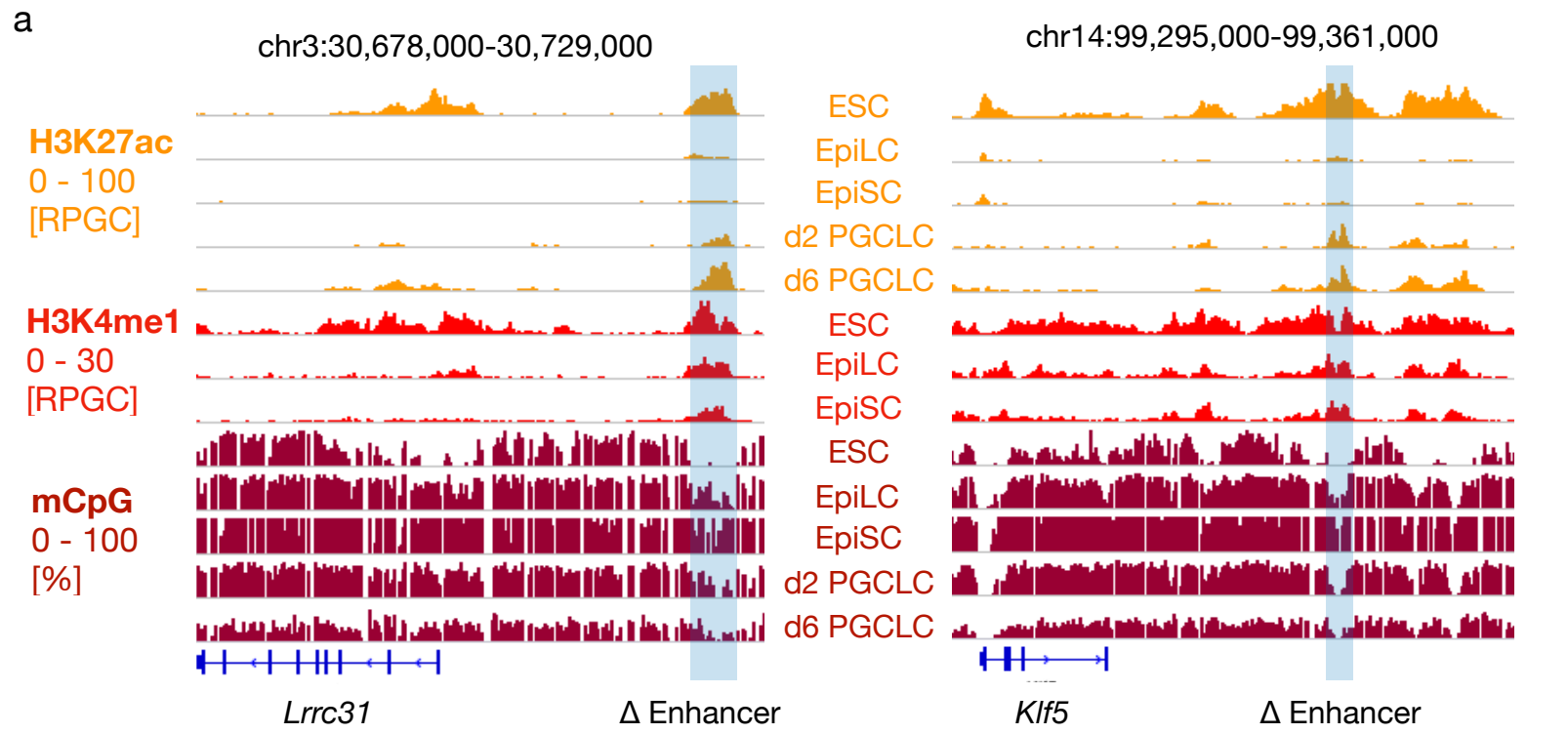
a.) Density plot showing the distribution of PGCLC enhancers according to their CpG densities. CpG density is defined as the fraction of CpGs found within 100 bp of each PGCLC enhancer sequence. For each PGCLC enhancer, the average CpG density within a -/+ 500 kb window was considered. PGCLC enhancers with a CpG density > 0.15, which are about 4.5 %, were omitted in the graph.

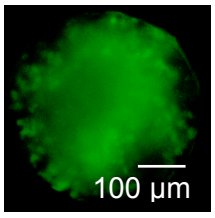
b) CpG methylation heterogeneity heatmap showing the pairwise comparisons between individual cells based on their genome-wide mCpG levels. The CpG methylation heterogeneity values are presented with a blue-red scale (blue means that cells are epigenetically similar and red that they are different). Above the heatmap, the developmental stages of the analyzed cells (E4.5, E5.5 or E6.5) and the average mCpG/CpG ratio (blue-orange scale) measured for all PGCLC enhancers within each single cell are shown (n = 258 cells). Cells were ranked according to genome-wide mCpG levels within each stage.

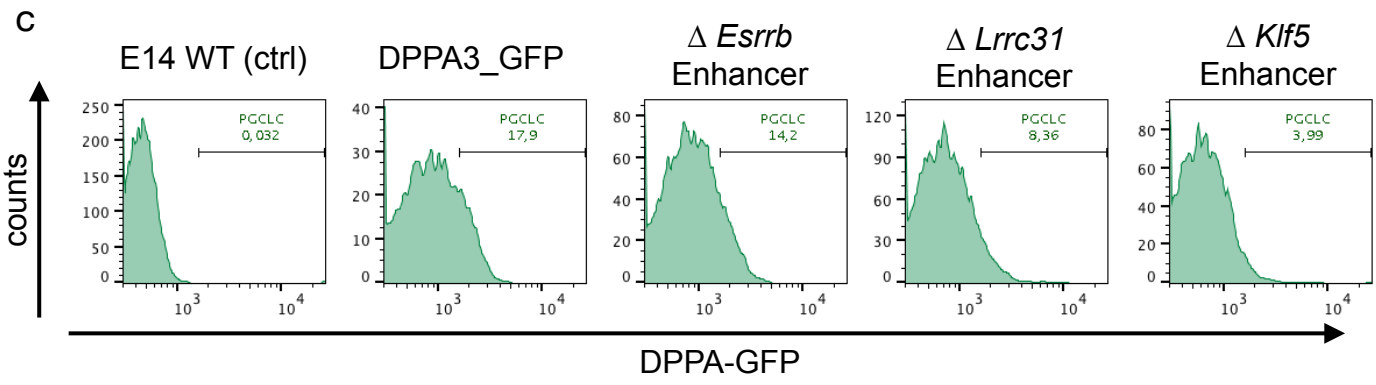
c.) Violin plots showing genome-wide CpG methylation heterogeneity measured in E4.5, E5.5 and E6.5 epiblast cells. The CpG methylation heterogeneity in the E5.5 epiblast cells was compared to the one measured in the E4.5 or E6.5 cells using wilcoxon tests (*: $p < 2.2\ 10^{-16}$).

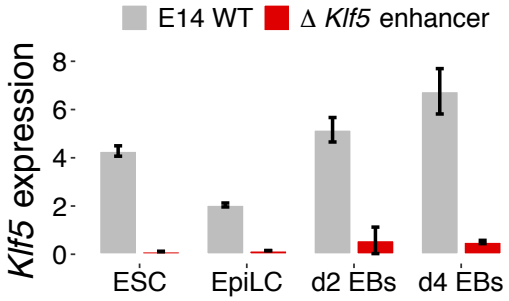
d.) Scatterplot showing the correlation between (left) genome-wide mCpG or (right) CpG coverage and the mean mCpG/CpG ratio of PGCLC enhancers for individual cells belonging to E4.5, E5.5 and E6.5 epiblasts (n=258 cells), respectively. The correlation coefficient (R) and p-value were calculated using Spearman correlation.

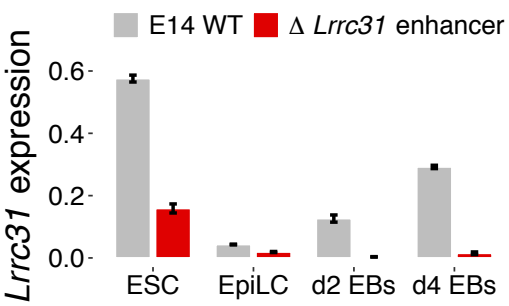
Figure S4


a
chr3:30,678,000-30,729,000
chr14:99,295,000-99,361,000
H3K27ac
0 - 100
[RPGC]
H3K4me1
0 - 30
[RPGC]
mCpG
0 - 100
[%]
ESC
EpiLC
EpiSC
d2 PGCLC
d6 PGCLC
Lrrc31
Δ Enhancer
Klf5
Δ Enhancer


b
d4 EBs:
DPPA3-GFP
100 µm


c
E14 WT (ctrl)
DPPA3_GFP
Δ Esrrb Enhancer
Δ Lrrc31 Enhancer
Δ Klf5 Enhancer
PGCLC 0,032
PGCLC 17,9
PGCLC 14,2
PGCLC 8,36
PGCLC 3,99
counts
DPPA-GFP


d
E14 WT
Δ Klf5 enhancer
Klf5 expression
ESC
EpiLC
d2 EBs
d4 EBs


e
E14 WT
Δ Lrrc31 enhancer
Lrrc31 expression
ESC
EpiLC
d2 EBs
d4 EBs

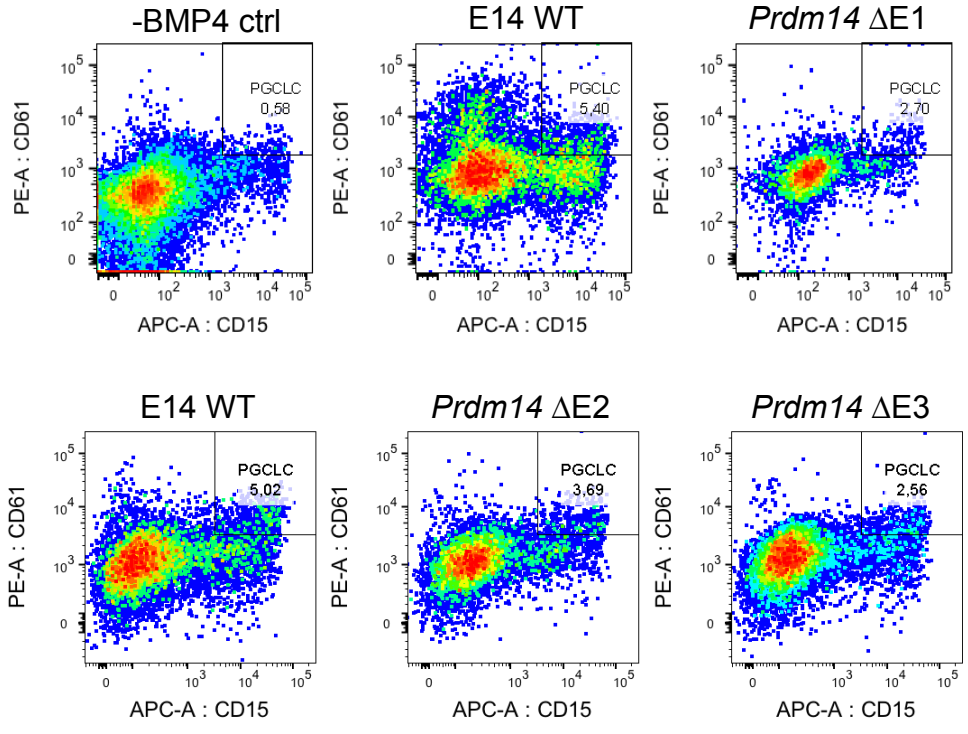

f
-BMP4 ctrl
E14 WT
Prdm14 ΔE1
Prdm14 ΔE2
Prdm14 ΔE3
PGCLC 0,58
PGCLC 5,40
PGCLC 2,70
PGCLC 5,02
PGCLC 3,69
PGCLC 2,56
PE-A : CD61
APC-A : CD15

Figure S4: Deletion and functional evaluation of representative PGCLC enhancers (related to Figure 4).

a.) Genome-browser view showing the PGCLC enhancers found within the *Lrrc31* (left) and *Klf5* (right) loci and their H3K27ac, H3K4me1 and CpG methylation dynamics during PGCLC formation. The H3K27ac PGCLC data was obtained from Kurimoto *et al.* 2015[12] and the CpG methylation data from Shirane *et al.* 2016[11].

b.) Evaluation of the DPPA3-GFP ESC reporter line. As shown in the image, upon differentiation of ESC, the induction of *Dppa3* leads to high GFP expression levels in the PGCLC found within d4 EB.

c.) Representative examples of PGCLC quantifications performed after four days of PGCLC differentiation using WT ESC, DPPA3-GFP ESC or DPPA3-GFP ESC in which the enhancers associated with *Esrrb*, *Lrrc31* or *Klf5* were deleted.

d-e.) *Lrrc31* (d) *and Klf5* (e) expression levels were measured by RT-qPCR in d4 EB differentiated from WT ESC and two different ESC clonal lines with the *Lrrc31* (d) or *Klf5* (e) enhancer deletions. 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).

f.) Representative examples of the PGCLC quantifications performed by FACS after four days of PGCLC differentiation using WT ESC (E14 WT) or ESC with deletions of each of the three *Prdm14* enhancers (∆E1-E3). As a negative control, the WT ESC were also differentiated in the absence of BMP4 (-BMP4 ctrl). PGCLC were defined as the $CD15^{+}CD61^{+}$ cells found within d4 EB.

Figure S5

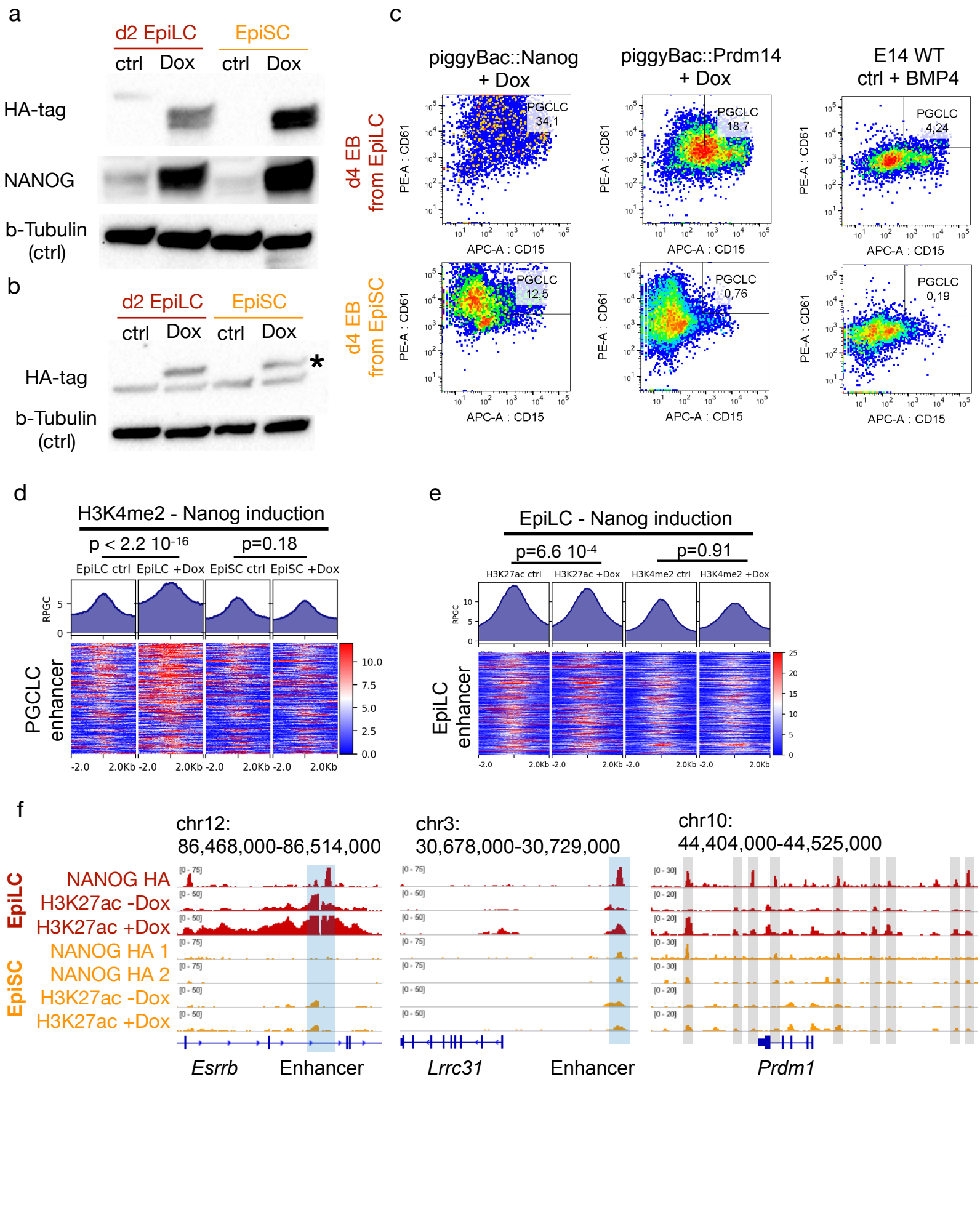

a
d2 EpiLC
EpiSC
ctrl
Dox
ctrl
Dox
HA-tag
NANOG
b-Tubulin (ctrl)
b
d2 EpiLC
EpiSC
ctrl
Dox
ctrl
Dox
HA-tag
b-Tubulin (ctrl)
c
piggyBac::Nanog + Dox
piggyBac::Prdm14 + Dox
E14 WT ctrl + BMP4
d4 EB from EpiLC
d4 EB from EpiSC
PE-A : CD61
APC-A : CD15
PGCLC 34,1
PGCLC 18,7
PGCLC 4,24
PGCLC 12,5
PGCLC 0,76
PGCLC 0,19
d
H3K4me2 - Nanog induction
p < 2.2 10-16
p=0.18
EpiLC ctrl
EpiLC +Dox
EpiSC ctrl
EpiSC +Dox
RPGC
PGCLC enhancer
-2.0
2.0Kb
e
EpiLC - Nanog induction
p=6.6 10-4
p=0.91
H3K27ac ctrl
H3K27ac +Dox
H3K4me2 ctrl
H3K4me2 +Dox
RPGC
EpiLC enhancer
f
chr12: 86,468,000-86,514,000
chr3: 30,678,000-30,729,000
chr10: 44,404,000-44,525,000
EpiLC
EpiSC
NANOG HA
H3K27ac -Dox
H3K27ac +Dox
NANOG HA 1
NANOG HA 2
H3K27ac -Dox
H3K27ac +Dox
Esrrb
Enhancer
Lrrc31
Enhancer
Prdm1

Figure S5: Inducible overexpression of NANOG and PRDM14 reveals that PGCLC enhancers display permissive chromatin features in EpiLC (related to Figure 5).
a-b.) Western blot analysis of the inducible overexpression of exogenous NANOG-HA (a) and PRDM14-HA (b). EpiLC or EpiSC were either left untreated (ctrl) or treated with Dox for 18 hours. (a) NANOG levels were measured using anti-HA or anti-NANOG antibodies. (b) PRDM14-HA levels were measured using an anti-HA antibody and the band corresponding to PRDM14-HA is indicated with an asterisk. B-Tubulin was used as a loading control.
c.) ESC lines enabling the inducible overexpression of exogenous HA-tagged NANOG (piggyBac::Nanog) or PRMD14 (piggyBac::Prdm14) were differentiated into d2 EpiLC and EpiSC. Next, the resulting d2 EpiLC or EpiSC were differentiated into PGCLC upon overexpression (+Dox) of NANOG-HA or PRDM14-HA and in the absence of growth factors. Similarly, d2 EpiLC and EpiSC E14 WT were also differentiated into PGCLC in the presence of growth factors (ctrl + BMP4). Representative examples of the PGCLC quantifications performed by FACS after four days of PGCLC differentiation are shown. PGCLC were quantified as the percentage of $CD15^{+}$ $CD61^{+}$ cells found within d4 EB.
d.) The ESC line enabling the inducible overexpression of exogenous HA-tagged NANOG was differentiated into d2 EpiLC and EpiSC. Next, the cells were either left untreated (ctrl) or treated with Dox in order to overexpress NANOG as described in Fig. 5b. H3K4me2 ChIP-seq experiments were performed in the untreated and Dox-treated EpiLC and EpiSC. H3K4me2 levels within PGCLC enhancers are shown as median profile (top) and heatmap plots (bottom). In the heatmap plots the PGCLC enhancers were ordered as in Fig. 2b. To compare untreated and treated cells, the average H3K4me2 signals within each PGCLC enhancer (using a -/+ 1kb window) were compared by paired wilcoxon tests (n = 511, *: $p < 2.2\ 10^{-16}$).
e.) The ESC line enabling the inducible overexpression of exogenous HA-tagged NANOG was differentiated into d2 EpiLC, which were then left untreated (ctrl) or treated with Dox as described in Fig. 5b. H3K27ac and H3K4me2 ChIP-seq experiments were performed in the untreated and Dox-treated d2 EpiLC. H3K27ac and H3K4me2 levels within EpiLC enhancers are shown as median profile (top) and heatmap plots (bottom). To compare untreated and treated cells, the average ChIP-seq signals within each EpiLC enhancer (using a -/+ 1kb window) were compared by paired wilcoxon tests (n = 689).
f.) Genome-browser view of the *Esrrb*, *Lrrc31* and *Prdm1* loci showing the binding of the exogenous NANOG-HA in both EpiLC and EpiSC as well as the H3K27ac signals in untreated (-Dox) and Dox-treated (*i.e.* NANOG-HA overexpression) EpiLC and EpiSC. For the NANOG-HA ChIP-seq profiles in EpiSC the results of two independent biological replicates are shown to illustrate the reproducibility of the weak binding within PGCLC enhancers. The PGCLC enhancers highlighted in blue for the *Esrrb* and *Lrrc31* loci are the same ones that were deleted in Fig. 4. For *Prdm1* several putative PGCLC enhancers are highlighted in gray.

Figure S6

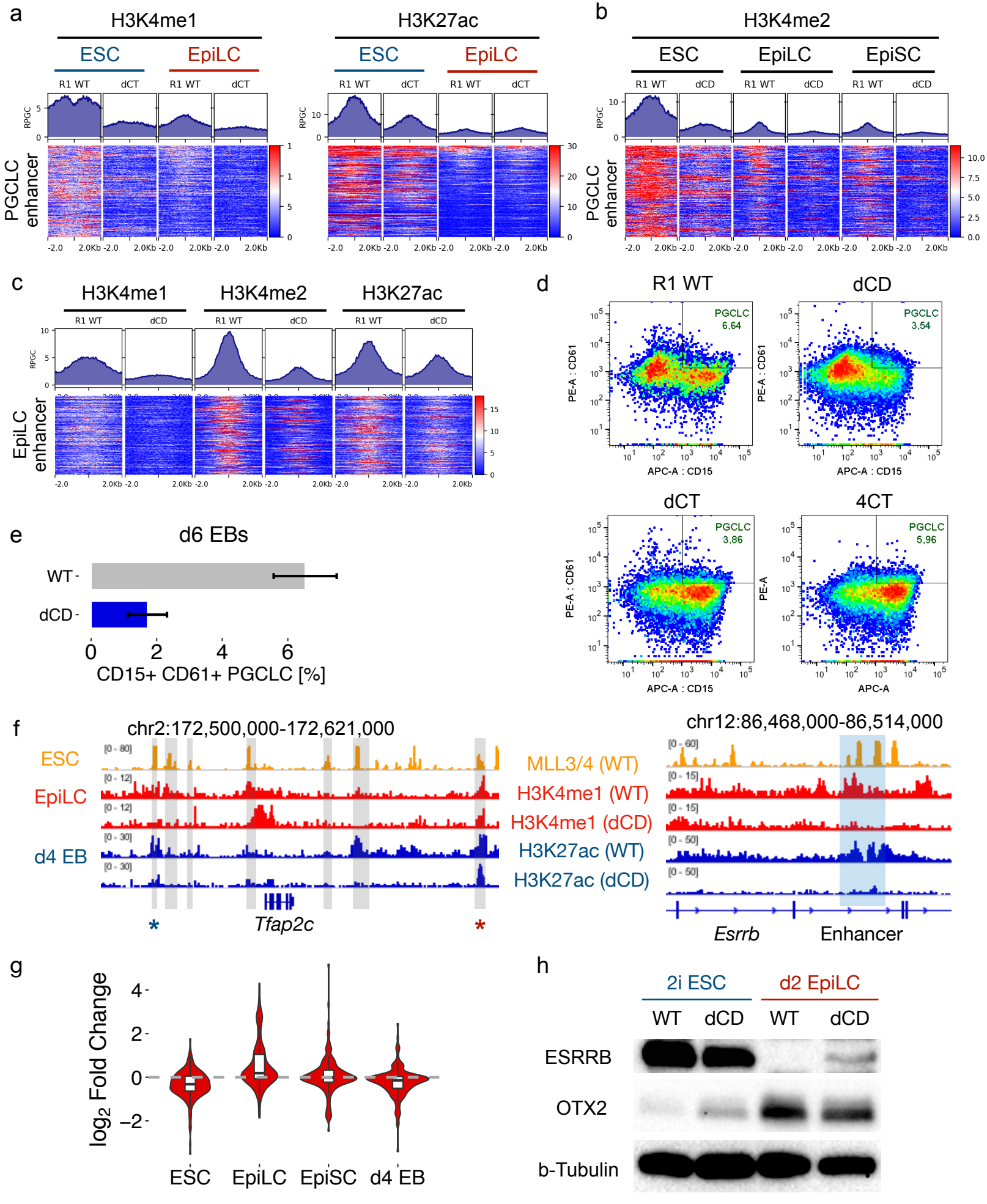

a
H3K4me1
H3K27ac
ESC
EpiLC
R1 WT
dCT
RPGC
PGCLC enhancer
b
H3K4me2
ESC
EpiLC
EpiSC
R1 WT
dCD
RPGC
PGCLC enhancer
c
H3K4me1
H3K4me2
H3K27ac
R1 WT
dCD
RPGC
EpiLC enhancer
d
R1 WT
dCD
dCT
4CT
PGCLC
6,64
3,54
3,86
5,96
PE-A : CD61
APC-A : CD15
PE-A
APC-A
e
d6 EBs
WT
dCD
CD15+ CD61+ PGCLC [%]
f
chr2:172,500,000-172,621,000
ESC
EpiLC
d4 EB
MLL3/4 (WT)
H3K4me1 (WT)
H3K4me1 (dCD)
H3K27ac (WT)
H3K27ac (dCD)
Tfap2c
chr12:86,468,000-86,514,000
Esrrb
Enhancer
g
log2 Fold Change
ESC
EpiLC
EpiSC
d4 EB
h
2i ESC
d2 EpiLC
WT
dCD
ESRRB
OTX2
b-Tubulin

Figure S6: Functional and epigenomic characterization of MLL3/4 catalytic deficient ESC lines upon EpiLC and PGCLC differentiation (related to Figure 6).

a.) H3K4me1 (left) and H3K27ac (right) ChIP-seq experiments in R1 WT and *Mll3/4* dCT ESC as well as upon their differentiation into d2 EpiLC. H3K4me1 and H3K27ac levels within PGCLC enhancers are shown as median profile (top) and heatmap plots (bottom). In the heatmap plots the PGCLC enhancers were ordered as in Fig. 2b.

b.) H3K4me2 ChIP-seq experiments in R1 WT and *Mll3/4* dCD ESC as well as upon their differentiation into d2 EpiLC and EpiSC. H3K4me2 levels within PGCLC enhancers are shown as median profile (top) and heatmap plots (bottom). In the heatmap plots the PGCLC enhancers were ordered as in Fig. 2b.

c.) H3K4me1, H3K4me2 and H3K27ac ChIP-seq experiments in R1 WT and *Mll3/4* dCD d2 EpiLC. H3K4me1, H3K4me2 and H3K27ac levels within EpiLC enhancers are shown as median profile (top) and heatmap plots (bottom). In the heatmap plots the PGCLC enhancers were ordered as in Fig. 2b.

d.) Representative examples of the PGCLC quantifications performed by FACS after four days of PGCLC differentiation using WT ESC (R1 WT) or with the indicated MLL3/4 catalytic mutant ESC lines. PGCLC were quantified as the $CD15^{+}$ $CD61^{+}$ cells found within d4 EB.

e.) WT ESC and *Mll3/4* dCD ESC lines were differentiated into PGCLC and PGCLC were quantified in d6 instead of d4 EB. PGCLC were defined as $CD15^{+}CD61^{+}$ cells within d6 EB from two biological replicates.

f.) Genome-browser view of the *Tfap2c* and *Esrrb* loci showing the binding of MLL3/4 in WT ESC, H3K4me1 ChIP-seq profiles in WT and dCD d2 EpiLC as well as H3K27ac ChIP-seq profiles in WT and dCD d4 EB. The PGCLC enhancers highlighted in blue for the *Esrrb* locus is the same that was deleted in Fig. 4. The putative *Tfap2c* enhancers are highlighted in gray and among them, a PGCLC enhancer that did not show a major loss of H3K4me1 in dCD EpiLC (blue star) and another PGCLC enhancer that became activated in dCD d4 EB (gained of H3K27ac) despite losing H3K4me1 in the dCD EpiLC (red star) are indicated.

g.) Violin plots showing the $\log_2$ gene expression fold changes for all PGCLC genes in *Mll3/4* dCD vs. R1 WT ESC as well as upon their differentiation into d2 EpiLC, EpiSC and d4 EB. The $\log_2$ fold changes were determined from at least two biological replicates with DESeq2. The RNA-seq data from ESC were obtained from Dorighi *et al.* 2017[32].

h.) Western Blot analysis of the ESRRB and OTX2 protein levels present in WT and dCD ESC as well as upon their differentiation into d2 EpiLC. B-Tubulin was used as a loading control.

Figure S7

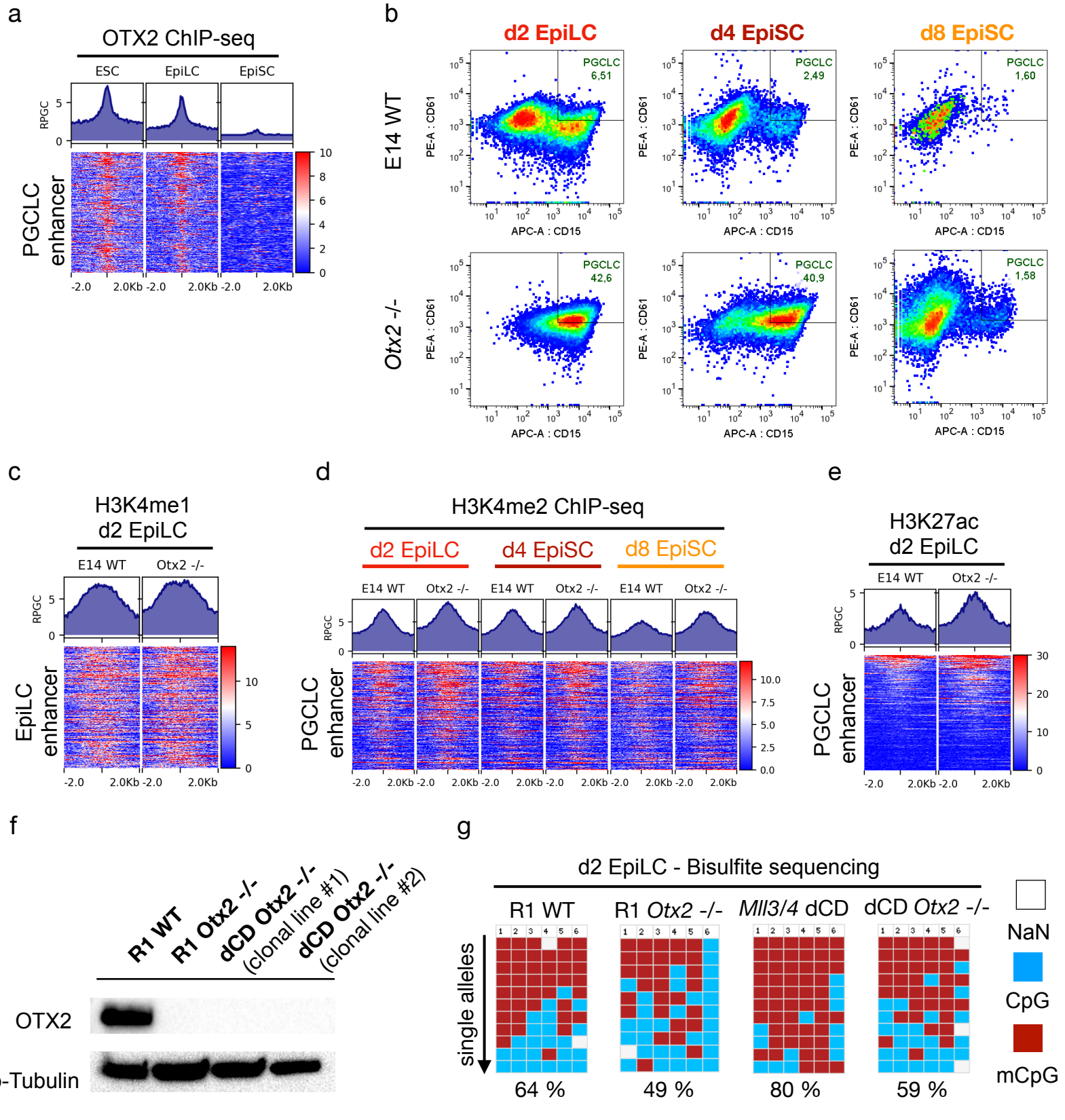

a
OTX2 ChIP-seq
ESC
EpiLC
EpiSC
RPGC
PGCLC enhancer
-2.0
2.0Kb
b
d2 EpiLC
d4 EpiSC
d8 EpiSC
E14 WT
Otx2 -/-
PGCLC 6,51
PGCLC 2,49
PGCLC 1,60
PGCLC 42,6
PGCLC 40,9
PGCLC 1,58
PE-A : CD61
APC-A : CD15
c
H3K4me1
d2 EpiLC
E14 WT
Otx2 -/-
EpiLC enhancer
d
H3K4me2 ChIP-seq
d2 EpiLC
d4 EpiSC
d8 EpiSC
PGCLC enhancer
e
H3K27ac
d2 EpiLC
PGCLC enhancer
f
R1 WT
R1 Otx2 -/-
dCD Otx2 -/- (clonal line #1)
dCD Otx2 -/- (clonal line #2)
OTX2
b-Tubulin
g
d2 EpiLC - Bisulfite sequencing
R1 WT
R1 Otx2 -/-
Mll3/4 dCD
dCD Otx2 -/-
single alleles
64 %
49 %
80 %
59 %
NaN
CpG
mCpG

Figure S7: Characterization of the *Otx2*$^{-/-}$ and dCD Otx2$^{-/-}$ ESC lines upon EpiLC and PGCLC differentiation (related to Figure 7).

a.) ChIP-seq signals for OTX2 in ESC, EpiLC and EpiSC. OTX2 binding levels within PGCLC enhancers are shown as median profile (top) and heatmap plots (bottom). In the heatmap plots the PGCLC enhancers were ordered as in Fig. 2b. The ESC and EpiLC ChIP-seq data was obtained from Buecker *et al.* 2014[50] and the EpiSC data from Matsuda *et al.* 2017[51].

b.) Representative examples of the PGCLC quantifications performed by FACS after four days of PGCLC differentiation starting from d2 EpiLC, d4 EpiSC and d8 EpiSC for either E14 WT (top) or E14 *Otx2*$^{-/-}$ (bottom) cells. PGCLC were quantified as the CD15^{+}CD61^{+} cells found within d4 EB.

c.) H3K4me1 ChIP-seq signals within EpiLC enhancers are shown in E14 WT and *Otx2*$^{-/-}$ d2 EpiLC as either median profile (top) or heatmap plots (bottom).

d.) H3K4me2 ChIPseq experiments were performed in d2 EpiLC, d4 EpiSC and d8 EpiSC differentiated from E14 WT and *Otx2*$^{-/-}$ ESC. H3K4me2 levels within PGCLC enhancers are shown as median profile (top) and heatmap plots (bottom). In the heatmap plots the PGCLC enhancers were ordered as in Fig. 2b.

e.) H3K27ac ChIP-seq signals within PGCLC enhancers are shown in E14 WT and *Otx2*$^{-/-}$ d2 EpiLC as either median profile (top) or heatmap plots (bottom).

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

g.) CpG methylation at the *Esrrb* enhancer was investigated by bisulfite sequencing of d2 EpiLC obtained from the indicated ESC lines. The columns of the plots correspond to individual CpG dinucleotides located within the *Esrrb* enhancer. Unmethylated CpGs are shown in blue, methylated CpGs in red and CpGs which were not covered are shown in gray. At least 10 alleles were analyzed in each cell line (rows).