

1 **Supplementary Information**

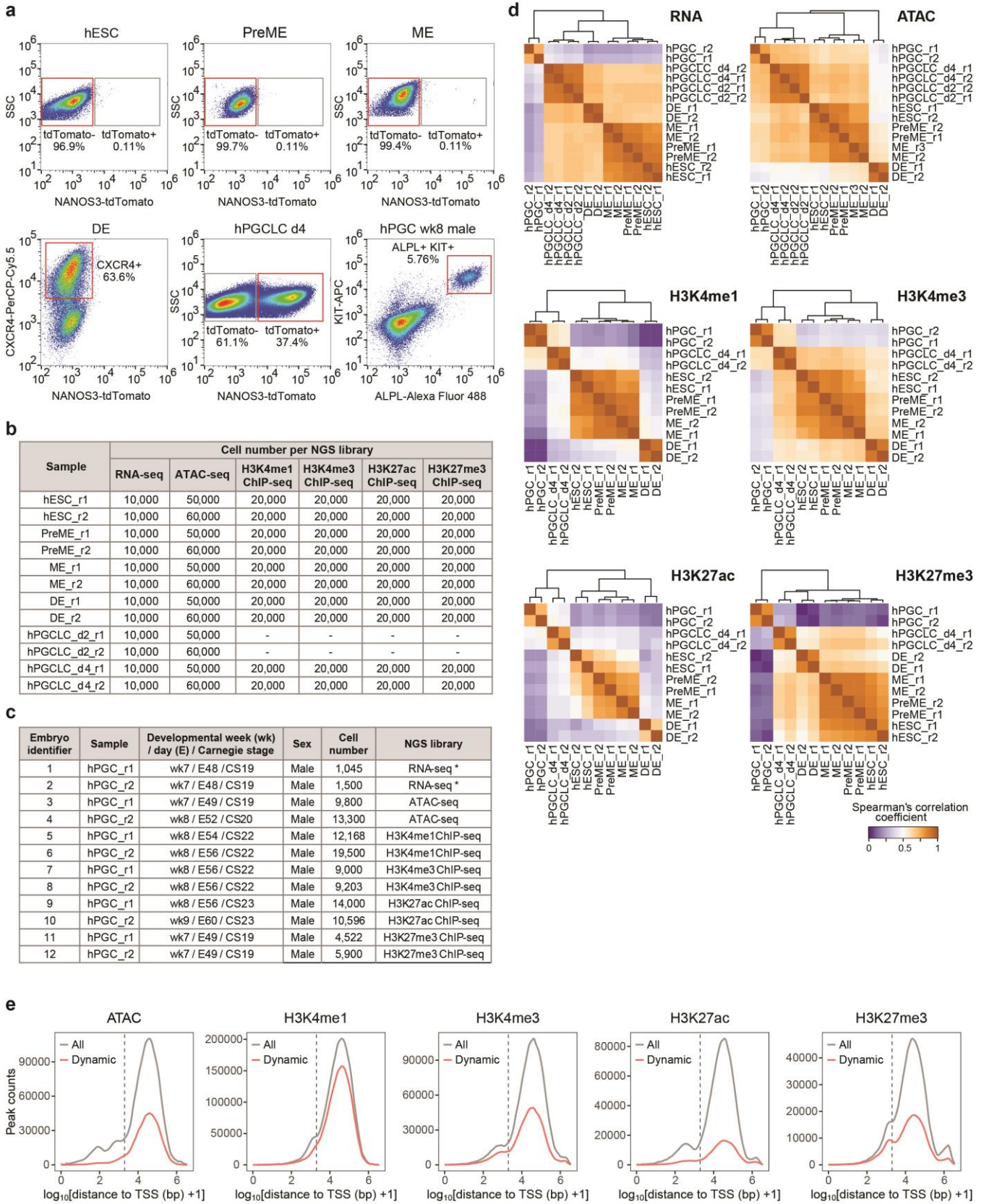
2 **Sequential enhancer state remodelling defines human germline competence and**
3 **specification**

4 **Tang et al**

5

6 **Extended Data Figures**

Extended Data Figure 1



Extended Data Fig. 1: Sample collection and overview of transcriptomic and epigenomic data.

a, Fluorescence-activated cell sorting (FACS) pseudocolor plots showing the cell populations collected for transcriptomic and epigenomic analysis (red gates).

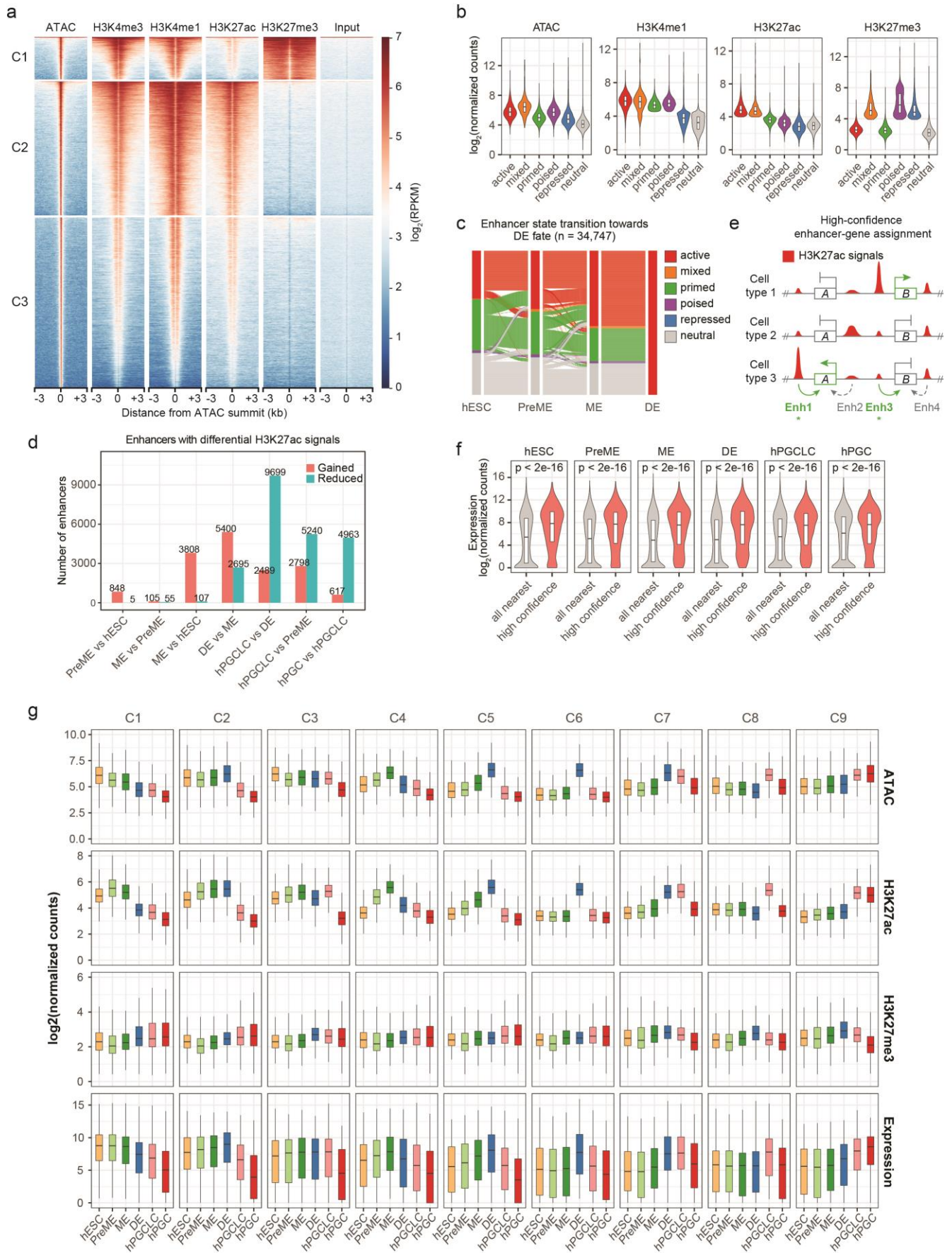
b, Table showing the hESC-derived cell types and the number of cells used to generate RNA-seq, ATAC-seq, and ChIP-seq libraries.

c, Table showing details of the human embryos used for hPGC isolation and the number of hPGCs used to generate RNA-seq, ATAC-seq, and ChIP-seq libraries. Asterisks indicate RNA-seq samples published in a previous study¹⁶.

d, Heatmaps showing Spearman's correlation coefficient of gene expression, ATAC-seq, H3K4me1, H3K4me3, H3K27ac, and H3K27me3 ChIP-seq signals in biological replicates. For RNA expression, correlation was based on the $\log_2(\text{normalized counts})$ of the top 25% most variable protein coding genes and lincRNA. For ATAC-seq, signals ($\log_2(\text{normalized counts})$) at combined peaks across the 6 cell types were used. For H3K4me1, H3K4me3, H3K27ac, and H3K27me3 ChIP-seq, signals ($\log_2(\text{normalized counts})$) at 1 kb bins of combined peaks were used. The samples were clustered using $(1 - \text{Spearman's correlation coefficient})$ as the distance measure (Ward's method with optimal tree ordering). See Methods.

e, Distance distribution between the summit of ATAC peaks or the centre of histone modification peaks and the closest TSS. Shown are all peaks and dynamic peaks with differential chromatin signals in the sample pairs shown in Extended Data Fig. 2d ($\log_2(\text{signal fold change}) > 1$ and adjusted p-value < 0.05). Note that most of the dynamic ATAC, H3K4me1, H3K4me3, H3K27ac and H3K27me3 peaks were > 2 kb away (dotted line) from the TSS.

Extended Data Figure 2



32

33

34 **Extended Data Fig. 2: Characterisation of dynamically active enhancers.**

35 a, Chromatin profile heatmaps of ATAC, H3K4me3, H3K4me1, H3K27ac, H3K27me3 and input signals
36 in hPGCLCs at ATAC-seq summit $\pm$ 3 kb. Segregation of ATAC-seq summits by K-means clustering
37 using normalised chromatin mark signals.

38 b, Distribution of chromatin mark signals at active, mixed, primed, poised, repressed, and neutral
39 enhancers in hPGCLCs (see Fig. 2a). Enhancers per violin/box plot: 23255 active, 1288 mixed, 36999
40 primed, 5648 poised, 3984 repressed, 79290 neutral. Box plots depict the median, lower and upper
41 hinges correspond to the 25th and 75th percentiles.

42 c, Enhancer state transitions of DE-active enhancers. Distal OCRs not overlapping any histone
43 modification peak in the analysed cell types were referred to as 'neutral' enhancers.

44 d, Putative enhancers with differential H3K27ac levels (absolute(log2(fold change)) >1 and adjusted p-
45 value < 0.05) in the indicated sample pairs.

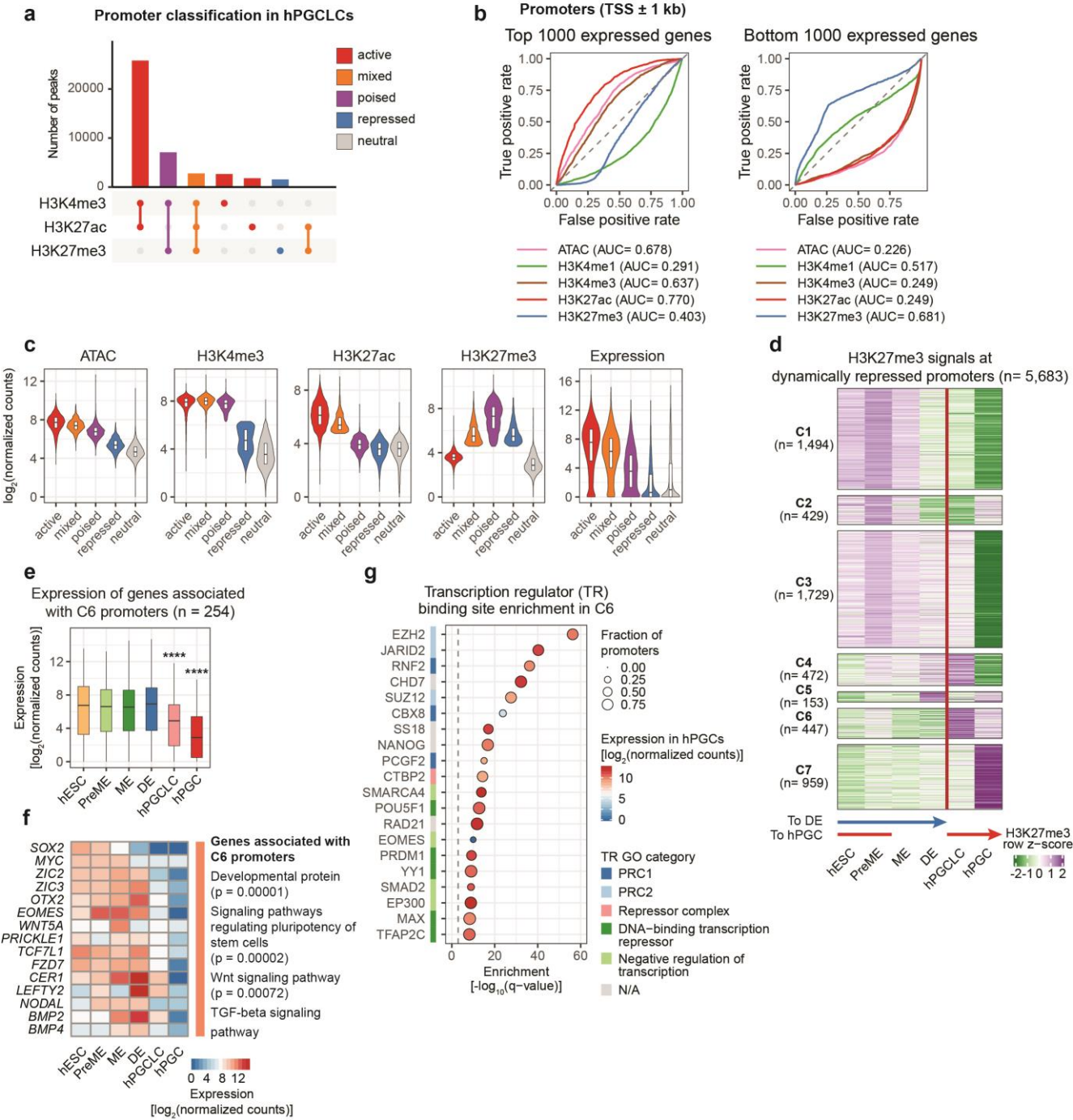
46 e, High confidence enhancer-gene associations. Putative enhancers were assigned to the nearest TSS.
47 The relevance of the enhancer-gene pair was assessed by the Kendall's rank correlation analysis
48 between the enhancer H3K27ac signals and the expression levels of the associated genes across the 6
49 cell types and 2 replicates (see Methods). In the simplified model shown, the Enh1-geneA and Enh3-
50 geneB pairs (green text and arrows) were identified as high confidence associations based on positive
51 correlation between H3K27ac and gene expression levels.

52 f, Expression levels of genes associated with active enhancers in different cell types. Compared to simply
53 associating genes to the nearest active enhancer, high confidence active enhancer associated genes
54 (Kendall rank correlation coefficient >0.3; empirical p-value < 0.05) exhibited significantly higher
55 expression in all cell types. Two-sided Wilcoxon rank sum test with FDR correction. Gene number per
56 violin/box plot (all nearest, high confidence): hESC (8686, 1279), PreME (9782, 1413), ME (11494, 1853),
57 DE (12279, 2224), hPGCLC (9954, 1726), hPGC (6612, 1024). Box plot organisation as in Extended
58 Data Fig. 2b.

59 g, Distribution of ATAC, H3K27ac and H3K27me3 signals in dynamically active enhancers and high
60 confidence target gene expression. Enhancers were segregated into nine clusters (Fig. 2c). Enhancer
61 per clusters as in Fig. 2c. Box plot organisation as in Extended Data Fig. 2b.

62

Extended Data Figure 3



Extended Data Fig. 3: Characterisation of dynamically active and repressed promoters during hPGC development.

a, Promoter classification in hPGCLCs based on the intersection of histone modification peaks at promoter regions (TSS $\pm$ 1 kb).

b, Receiver operating characteristic (ROC) curves of ATAC, H3K4me3, H3K4me1, H3K27ac, and H3K27me3 signals at promoter (TSS $\pm$ 1 kb) as predictors of gene activity in hPGCLCs. The top 1000 (top panel) or the bottom 1000 (bottom panel) expressed genes were used as positives. Promoters not overlapping any chromatin peak were excluded. Note that H3K27ac (area under the curved (AUC) of top

73 genes = 0.770) and H3K27me3 (AUC of bottom genes = 0.681) were the best predictors of expressed
74 and repressed genes, respectively.

75 c, Distribution of chromatin mark signals at active, mixed, poised, repressed, and neutral promoters and
76 the expression of their associated genes in hPGCLCs. Promoter number: 30261 active, 2833 mixed,
77 7089 poised, 1579 repressed, 19832 neutral. Associated gene number: 12629 active, 1526 mixed, 3662
78 poised, 1144 repressed, 13038 neutral. Box plot organisation as in Extended Data Fig. 2b

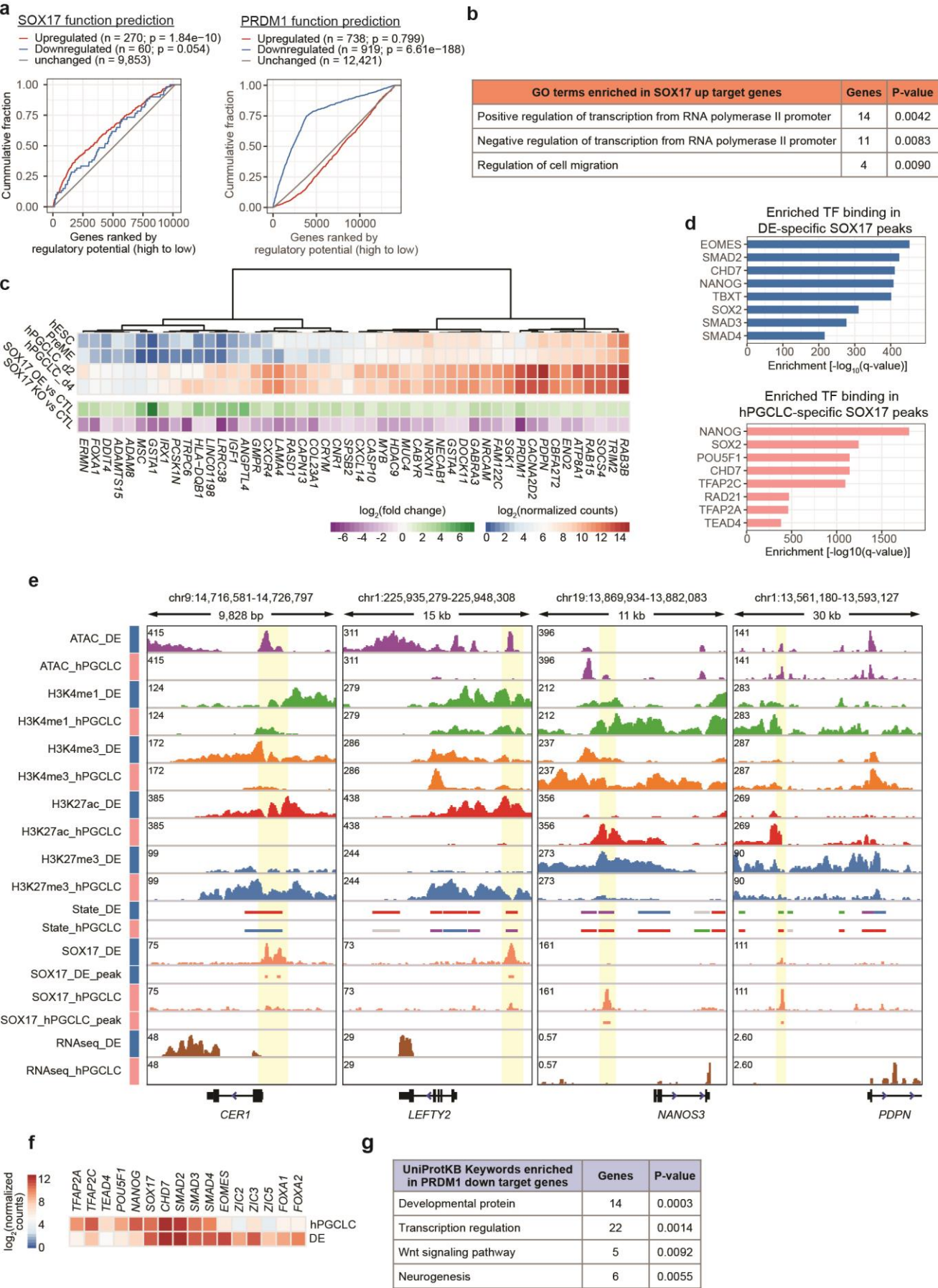
79 d, K-means clustering of dynamically repressed promoters into 7 clusters (C) by H3K27me3 signals.
80 Dynamically repressed promoters were promoters that exhibited 'mixed', 'poised' or 'repressed' state
81 (see Methods) in any cell type with differential H3K27me3 signals.

82 e, Box plots showing expression levels of genes associated with the C6 dynamically repressed promoters
83 in c. **** p-value < 0.00001 (Wilcoxon rank sum test adjusted by the Holm method) in marked against
84 each unmarked cell type. Box plot organisation as in Extended Data Fig. 2b

85 f, Heatmaps showing the expression levels of representative genes associated with C6 in c. The right
86 panel shows the representative enriched gene ontology terms.

87 g, Dot plots showing the enrichment of transcription regulator (TR) binding site in the dynamic repressed
88 promoters in C6 in c. The top 20 enriched TRs (out of 1,135 in the ReMap2020 database⁴²) are shown.
89 TRs were annotated against 5 gene ontology terms associated with repressive functions. The dot size
90 represents promoter fraction overlapping with the TR binding sites. The dot colour indicates the
91 expression levels of the enriched transcription regulators in hPGCs.

Extended Data Figure 4



92

93

94 **Extended Data Fig. 4: Direct targets of SOX17 in hPGCLCs and DE.**

95 a, Cumulative distribution function plot showing the functional prediction of SOX17 and PRDM1.
96 The ChIP peaks of SOX17/PRDM1 in hPGCLCs were assigned to genes with TSS within 100 kb of
97 the peak summits. A regulatory potential score was calculated for each gene based on the distance
98 between the peak summit and the TSS⁹⁶. The genes were then divided into upregulated,
99 downregulated and unchanged according to their expression patterns upon SOX17 or PRDM1
100 overexpression. Cumulative distribution function plot was generated for each group with genes
101 ranked by decreasing regulatory potential. A one-tailed Kolmogorov-Smirnov test was used to
102 determine the statistical significance between the differentially expressed groups and the
103 unchanged group. Note that upregulated genes (but not downregulated genes) upon SOX17
104 induction have a significantly higher tendency to be bound by SOX17. In contrast, genes
105 downregulated upon PRDM1 overexpression tend to be bound by PRDM1.

106 b, GO biological process terms that were enriched in SOX17 direct up targets (red dots in Fig. 3d)

107 c, Expression heatmap of SOX17 direct up target genes. Shown are the genes which were (1)
108 upregulated both by SOX17 alone ($\log_2(\text{fold change}) > 1$ and adjusted p-value < 0.05 between Dex-
109 treated and non-treated 12h PreME aggregates); (2) upregulated by cytokines ($\log_2(\text{fold change})$
110 > 2 and adjusted p-value < 0.05 between day 2 hPGCLCs and PreME); and (3) downregulated in
111 SOX17 KO ($\log_2(\text{fold change}) > 1$ and adjusted p-value < 0.05 between SOX17 KO day 2 aggregate
112 and WT day 2 aggregate)²⁰.

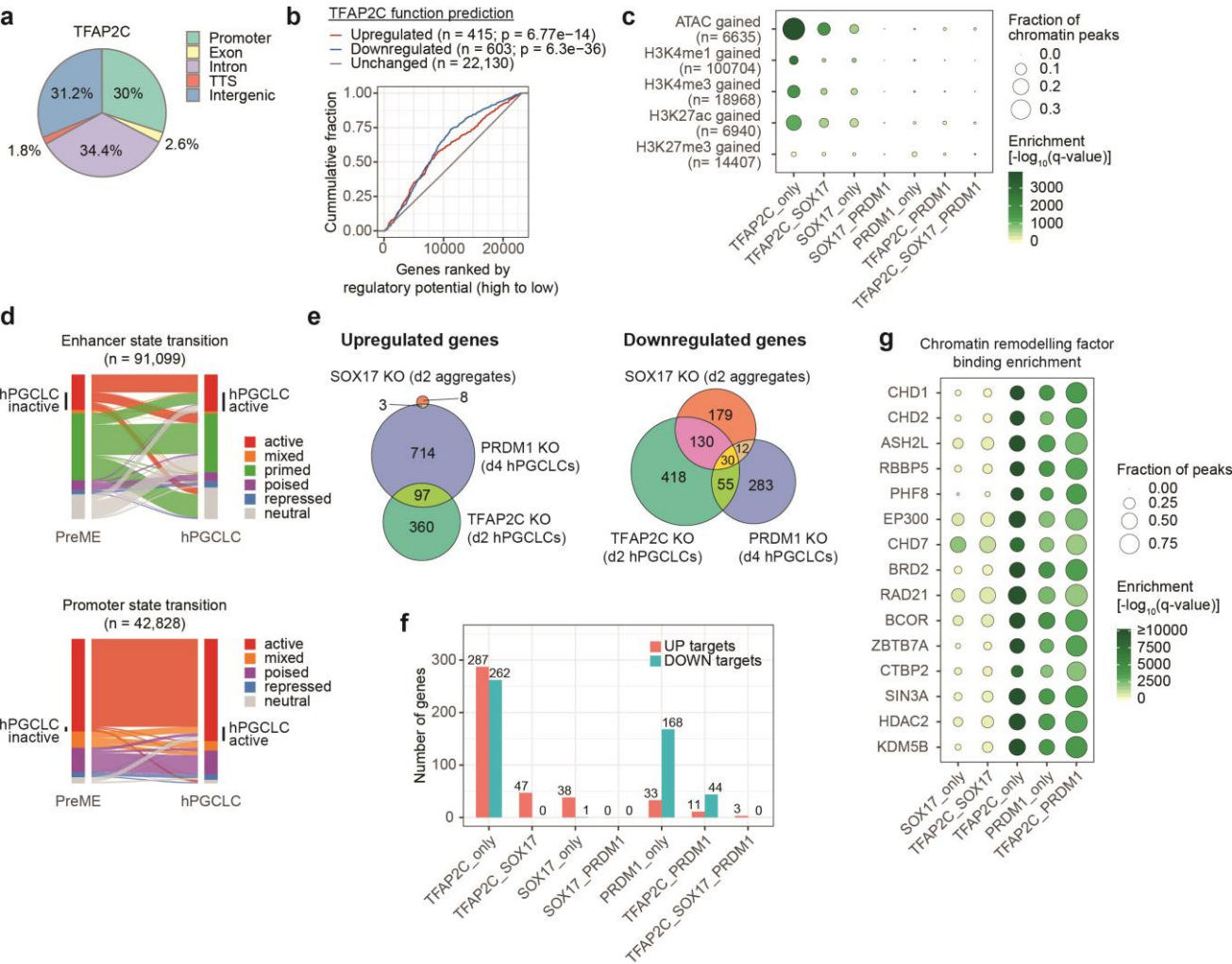
113 d, Top eight TFs with binding sites (ReMap2020) enriched in hPGCLC-specific and DE-specific
114 SOX17 peaks.

115 e, Genome browser snapshots showing the epigenetic landscape of DE-specific (CER1 and
116 LEFTY2) and hPGCLC-specific (NANOS3 and PDPN) SOX17-bound gene targets.

117 f, Heatmap showing expression of genes associated with the top enriched motifs (Fig. 3h) and the
118 top enriched TF binding sites (in d) in hPGCLC-specific and DE-specific SOX17 peaks.

119 g, UniProtKB Keywords that were enriched in PRDM1 direct down targets (blue dots in Fig. 3i).

Extended Data Figure 5



120

121

Extended Data Fig. 5: Direct target genes of TFAP2C, SOX17 and PRDM1.

a, Genomic distribution of the TFAP2C peaks in hPGCLC aggregates⁴⁶.

b, Cumulative distribution function plot showing the functional prediction of TFAP2C. The TFAP2C peaks were assigned to genes with TSS within 100 kb of the peak summits. A regulatory potential score was calculated for each gene based on the distance between the peak summit and the TSS⁹⁶. The genes were then divided into three groups (upregulated, downregulated and unchanged) according to their expression patterns in TFAP2C KO day2 hPGCLCs versus the wild-type control²⁰. Cumulative distribution function plot was generated for each group with genes ranked by decreasing regulatory potential. A one-tailed Kolmogorov-Smirnov test was used to determine the statistical significance between the differentially expressed groups and the unchanged group.

c, The enrichment of TFAP2C, SOX17 and PRDM1 peaks in genomic loci that gained ATAC, H3K4me1, H3K4me3, H3K27ac or H3K27me3 signals during the PreME to hPGCLC transition. The TF peaks were categorized into seven cooperativity classes as in Fig. 4a. The dot size represents the fraction of chromatin peaks that overlapped with the TF peaks. Dot color indicates enrichment significance.

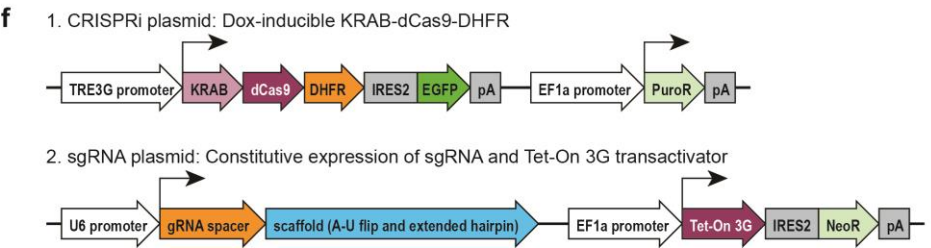
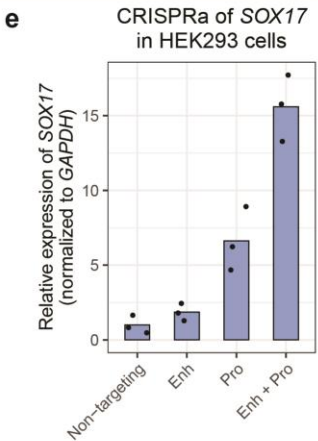
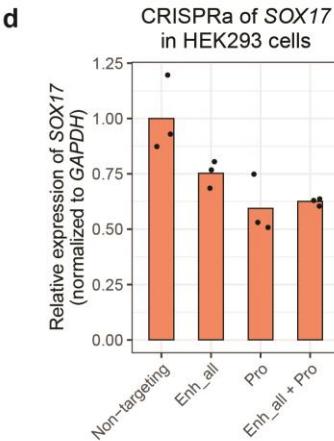
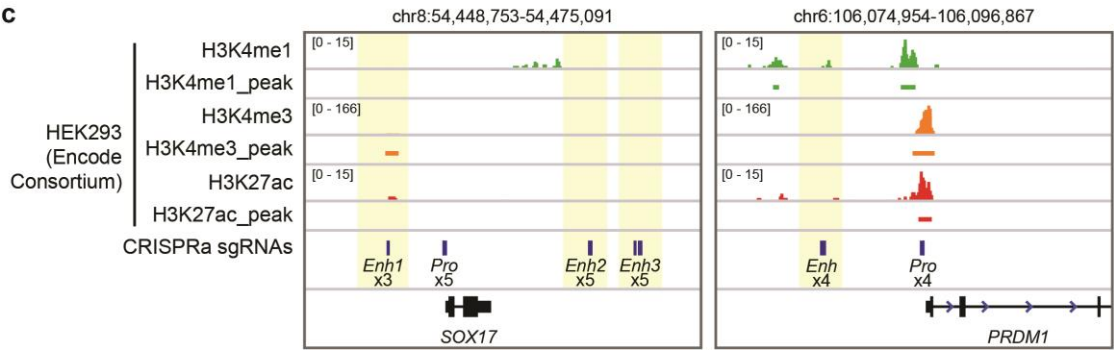
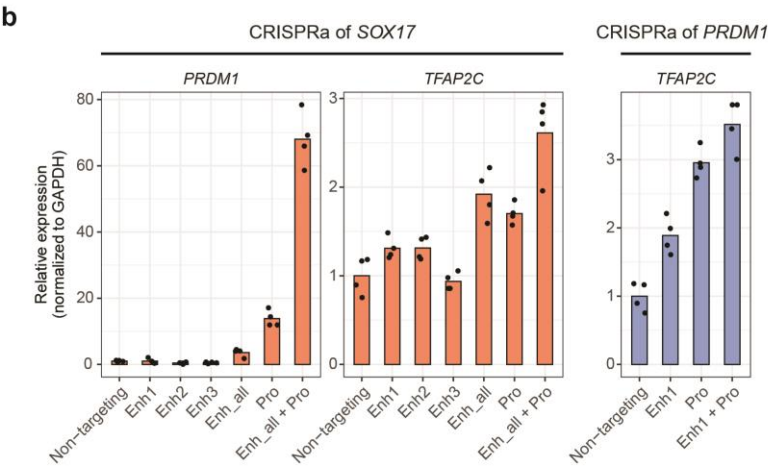
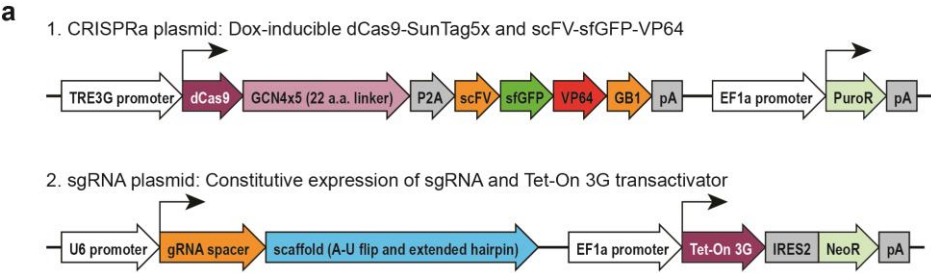
d, Alluvial plots showing the enhancer (upper panel) and promoter (lower panel) state transition from PreME to hPGCLC. The enhancers/promoters that became active/inactive in hPGCLCs were used for TF binding enrichment analysis in Fig. 4b.

e, Venn diagram showing the intersection of upregulated and downregulated genes in SOX17, TFAP2C and PRDM1 KO hPGCLCs/aggregate²⁰. Upregulated and downregulated genes were defined as ($\log_2(\text{fold change versus wild-type control}) > 1$ and adjusted p-value < 0.05) and ($\log_2(\text{fold change versus wild-type control}) < (-1)$ and adjusted p-value < 0.05), respectively.

f, The number of direct up and down target genes of TFAP2C, SOX17 and PRDM1 based on their cooperative binding.

g, Enrichment of chromatin remodelling factor binding sites in TFAP2C, SOX17 and PRDM1 peaks in hPGCLCs. The y-axis shows the chromatin remodelling factors that were amongst the top 10 enriched transcriptional regulators (ReMap2020) in any of the five peak sets (x-axis).

Extended Data Figure 6



149

150

Extended Data Fig. 6: Inducible CRISPR activation and interference systems for activation and repression of hPGC TFs.

a, The piggyBAC plasmids encoding an optimized doxycycline-inducible dCas9-SunTag-VP64 CRISPR activation system. Upon integration of both the CRISPRa plasmid and the sgRNA plasmid into the genome and Dox treatment, the Tet-On 3G doxycycline-binding transactivator protein encoded in the sgRNA plasmid will drive the transcription of dCas9-GCN4x5-P2A-scFV-sfGFP-VP64 through the TRE3G promoter. After translation of the mRNA, the recombinant protein will be split into dCas9-GCN4x5 and scFV-sfGFP-VP64 through the P2A self-cleaving peptide. Subsequently, the dCas9-GCN4x5 will be guided to enhancer/promoter by the constitutively expressed sgRNA and recruit up to 5 copies of the scFV-sfGFP-VP64 recombinant transactivator. To improve epigenome editing efficiency, the GCN4 epitopes were separated by optimized 22-amino-acid linkers⁵⁵. To increase sgRNA expression and to enhance sgRNA-dCas9 affinity, a sgRNA scaffold with an A-U flip and extended hairpin was used⁵⁶.

b, RT-qPCR showing CRISPR activation of SOX17 (2 days Dox treatment in hESCs) induced PRDM1 and TFAP2C mRNA expression. Activation of PRDM1 also upregulated TFAP2C. Average of 4 biological replicates, with individual replicates shown as data points.

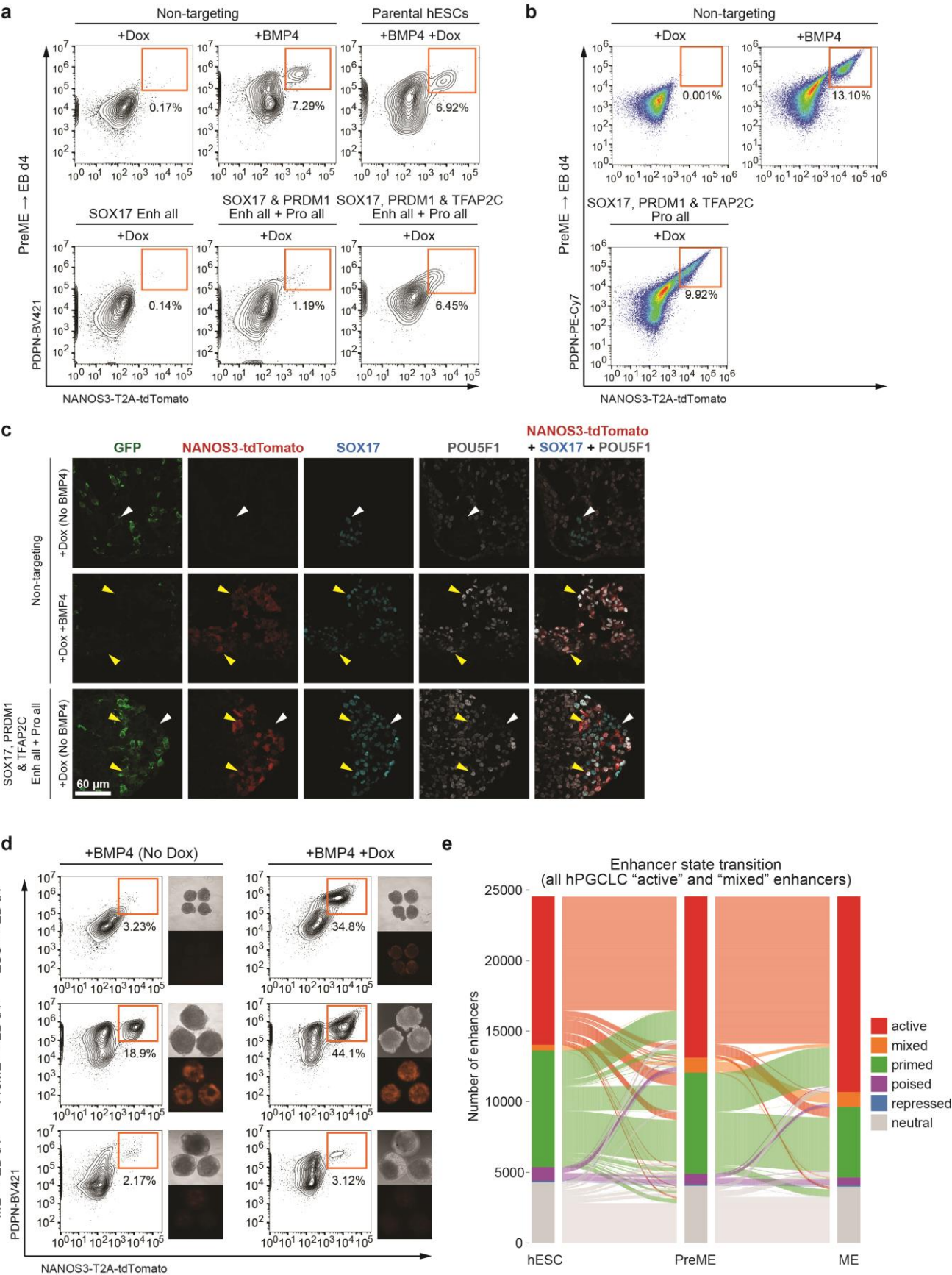
c, The epigenetic landscape of the SOX17 and PRDM1 loci in hESC, PreME, hPGCLCs and HEK293 cells. Note that the SOX17 locus in HEK293 cells does not bear H3K4me1, H3K4me3 and H3K27ac marks. For CRISPR activation (CRISPRa) assay, 3-5 sgRNAs were used to activate each putative enhancer (highlighted) and promoter.

d-e, RT-qPCR of SOX17 and PRDM1 following CRISPR activation of enhancers and promoters in HEK293 cells. HEK293 cells were transiently transfected with CRISPRa (dCas9-SunTag-VP64) and sgRNA plasmids

and treated with Dox for 2 days. GFP-positive cells (expressing dCas9-SunTag and scFV-sfGFP-VP64) were isolated for RT-qPCR. Average of 3 technical replicates, with individual replicates shown as data points. Assay has been performed two times independently with similar results.

f, The piggyBAC plasmids encoding a re-engineered doxycycline-inducible KRAB-dCas9-DHFR CRISPR interference system (also see Fig. 6a).

Extended Data Figure 7



179

180

181 **Extended Data Fig. 7: Characterisation of CRISPRa-induced hPGCLCs.**

182 a and b, FACS analysis of day 4 EBs generated from PreME of hESC lines harbouring the Dox-inducible

183 CRISPRa transgene with the indicated sgRNA combinations.

184 c, Immunofluorescence showing the co-expression of hPGCLC markers NANOS3-tdTomato, POU5F1

185 and SOX17 in hPGCLCs (yellow arrowheads) induced by CRISPRa in the absence of BMP4. White

186 arrowheads indicate SOX17 single-positive cells (presumably DE). Representative results of 3 biological

187 replicates.

188 d, Induction of hPGCLCs from hESCs, PreME and ME with or without CRISPR-mediated activation of

189 SOX17 and PRDM1 enhancers and promoters. FACS analysis of day 4 EBs shows that the co-activation

190 of SOX17 and PRDM1 act synergistically with BMP4 to increase the efficiency of hPGCLC induction

191 from hESCs and PreME, but not from ME. The appearance of the EBs under brightfield and tdTomato

192 fluorescence filter are shown next to the corresponding FACS plots. Representative results of 3

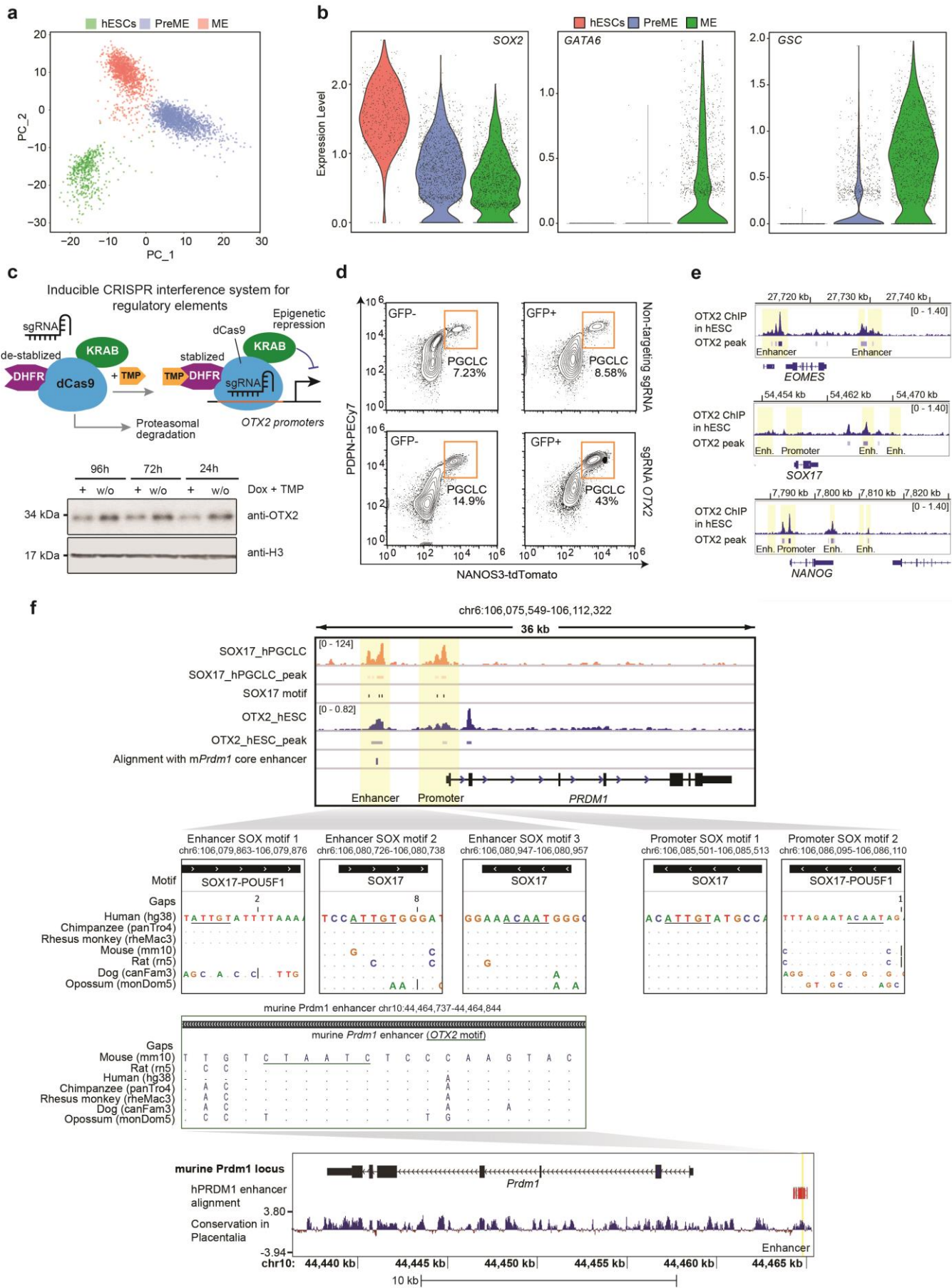
193 independent experiments.

194 e, Alluvial plots showing enhancer state transitions of hPGCLC-active enhancers in hESCs, PreME and

195 ME.

196

Extended Data Figure 8



197

198

Extended Data Fig. 8: Sequence conservation of the human PRDM1 regulatory element.

a, PCA analysis scRNAseq profiles of cells in the hESC, PreME and ME state

b, Violin plots summarizing expression levels of the indicated genes in individual cells in the hESC, PreME and ME state analysed by scRNA-seq.

c, Schematics of the inducible CRISPR interference (CRISPRi) system used to repress the two OTX2 promoters (upper panel). Western Blots depicting OTX2 and H3 levels in transgenic hESCs treated with vehicle or Doxycycline and TMP for the indicated time periods (lower panel). Molecular weights of marker proteins are depicted in kilodaltons (kDa). Representative experiment, knockdown efficiency tested two times independently at the shown timepoints.

d, FACS analysis of PGCLCs expressing non-targeting or sgRNA targeting the OTX2 promoter in the presence or absence of KRAB-dCas9-ecDHFR (GFP). Representative experiment out of 3 independent technical replicates shown in Fig. 8e.

e, Genome browser snapshots showing OTX2 ChIPseq signals and peaks in hESCs (GSE61475)⁶⁴. Enhancer identified in this work are indicated in yellow.

f, Upper panels: BLAT alignment of the core murine Prdm1 enhancer (B108)⁷² to the human genome. Conservation of the SOX motifs in the putative enhancer and promoter of human PRDM1 across seven mammalian species. MULTIZ whole-genome alignment showed that 4 out of 5 core SOX motifs ('ATTGT', underlined) in the human PRDM1 enhancer and promoter are conserved in mice. Grey dot indicates exact match. Blank space represents absence of the corresponding sequence in the indicated species. Lower panels: BLAT alignment of the human PRDM1 enhancer to the murine genome showing the conservation of the OTX2 motif in the murine PRDM1 enhancer⁷².

Supplementary Tables

Supplementary Table 1. Sequencing and mapping summary of RNA-seq and ChIP-seq libraries.

Supplementary Table 2. Gene ontology enrichment analysis of dynamically active enhancers.

Supplementary Table 3. Direct target genes of SOX17 by overexpression.

Supplementary Table 4. Direct target genes of PRDM1 by overexpression.

Supplementary Table 5. Direct up target genes of TFAP2C, SOX17 and PRDM1.

Supplementary Table 6. Direct down target genes of TFAP2C, SOX17 and PRDM1.

Supplementary Table 7. Antibodies, oligos and sgRNAs.

229

230 **Source Data**

231 Source data Figure 3

232 Source data Figure 5

233 Source data Figure 6

234 Source data Figure 7

235 Source data Figure 8

236 Source data Extended data figure 6