Sequential enhancer state remodelling defines human germline competence and specification

Walfred W. C. Tang1,2,9, Aracely Castillo-Venzor1,2,3,9, Wolfram H. Gruhn1,2,9, Toshihiro Kobayashi4,5, Christopher A. Penfold1,2,3,6, Michael D. Morgan7,8, Dawei Sun1,2, Naoko Irie1,2 and M. Azim Surani1,2,3

Germline–soma segregation is a fundamental event during mammalian embryonic development. Here we establish the epigenetic principles of human primordial germ cell (hPGC) development using in vivo hPGCs and stem cell models recapitulating gastrulation. We show that morphogen-induced remodelling of mesendoderm enhancers transiently confers the competence for hPGC fate, but further activation favours mesoderm and endoderm fates. Consistently, reducing the expression of the mesendodermal transcription factor OTX2 promotes the PGC fate. In hPGCs, SOX17 and TFAP2C initiate activation of enhancers to establish a core germline programme, including the transcriptional repressor PRDM1 and pluripotency factors POU5F1 and NANOG. We demonstrate that SOX17 enhancers are the critical components in the regulatory circuitry of germline competence. Furthermore, activation of upstream cis-regulatory elements by an optimized CRISPR activation system is sufficient for hPGC specification. We reveal an enhancer-linked germline transcription factor network that provides the basis for the evolutionary divergence of mammalian germlines.

In this Article, we show how morphogens transiently confer human germline competence and direct cell fate choices by sequential epigenetic patterning of enhancer elements. Clustered regularly interspaced short palindromic repeats (CRISPR)-mediated modulation of enhancer activity reveals their importance for regulating critical TFs mediating germline development. Accordingly, an intricate enhancer-regulated TF network underpins hPGC specification and progression.

Results
Epigenetic trajectories upon germline–soma segregation. We investigate the epigenomic dynamics in our in vitro model, which simulates human gastrulation and germline formation. Employing hESCs harbouring the highly specific germline reporter NANOS3–T2A–tdTomato, we examined cell state transitions towards PreME, ME, DE and hPGCLCs (Fig. 1a). We performed RNA sequencing (RNA-seq), transposase-accessible chromatin followed by sequencing (ATAC–seq) and ultralow-input native chromatin immunoprecipitation followed by sequencing (ULI-NChIP–seq) for promoter- and enhancer-associated histone modifications.

1Wellcome Trust/Cancer Research UK Gurdon Institute, Henry Wellcome Building of Cancer and Developmental Biology, Cambridge, UK. 2Physiology, Development and Neuroscience Department, University of Cambridge, Cambridge, UK. 3Wellcome–MRC Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, Cambridge, UK. 4Division of Mammalian Embryology, Center for Stem Cell Biology and Regenerative Medicine, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan. 5Center for Genetic Analysis of Behavior, National Institute for Physiological Sciences, Okazaki, Japan. 6Centre for Trophoblast Research, University of Cambridge, Cambridge, UK. 7Cancer Research UK Cambridge Institute, Li Ka Shing Centre, University of Cambridge, Cambridge, UK. 8European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Cambridge, UK. 9These authors contributed equally: Walfred W. C. Tang, Aracely Castillo-Venzor, Wolfram H. Gruhn.

✉e-mail: walfredtang@gmail.com; a.surani@gurdon.cam.ac.uk
(H3K4me1, H3K4me3, H3K27ac and H3K27me3) (Extended Data Fig. 1a). The hPGCLCs are at the nascent pre-migratory stage, which we compared with the closest available in vivo gonadal hPGCs from individual human male weeks 7–9 embryos (with ethical approval) (Extended Data Fig. 1c).

Unsupervised hierarchical clustering (UHC) of gene expression revealed three main branches; (1) hESCs, PreME and ME; (2) DE and hPGCs; (3) hPGCs, which formed a separate branch, albeit clustering closest to hPGCLCs (Fig. 1b and Extended Data Fig. 1d). Robust expression of most pluripotency factors was observed in all cell types apart from DE, while SOX2 expression diminished in hPGCLCs and hPGCs 9,16,22 (Fig. 1c). There was substantial upregulation of mesendodermal genes TBXT and EOMES in PreME, whereas GSC, GATA4 and GATA6 were induced later in ME and co-expressed with endoderm TFs (for example, FOXA1, FOXA2 and HNF4A) in DE. Strong SOX17 and PRDM1 expression was detected in hPGCLCs, hPGCs and DE. In hPGCLC and hPGC, TFAP2C, NANSOS3 and CD38 were expressed with naive pluripotency factors TFAP2L1 and KLF4,23,24 while DAZL, DDX4 and MAEL were the meiosis-associated RNA-binding proteins, were expressed in the gonadal hPGCs.

Next, Spearman's correlation and UHC of normalized signals at combined peak sets of all cell types showed ATAC, H3K4me1, H3K4me3 and H3K27ac signals exhibited a similar clustering pattern (Extended Data Fig. 1d). Accordingly, hESCs, PreME and ME formed one group separated from DE, whereas hPGCLCs and hPGCs clustered in another branch. Principal component analysis (PCA) of H3K4me1, H3K4me3 and H3K27ac signals linked a germline trajectory to hESCs, hPGCLCs and hPGCs along principal component 1 (PC1), and an endoderm trajectory connects hESCs, PreME, ME and DE along PC2 (Fig. 1d). However, PCA of H3K27me3 signals placed hPGCs away from hPGCLCs and other in vitro-derived cells along PC1, while the endoderm trajectory along PC2 was preserved, reflecting potentially the global reduction of H3K27me3 and DNA demethylation25. Overall, the epigenomic trajectories were consistent with human gastrulation and germline establishment (Fig. 1a).

Activation of enhancers underlies cell fate transitions. Most regions with differential epigenetic signals were 10–100 kb away from the nearest transcription start site (TSS) (Extended Data Fig. 1e), featuring open chromatin (ATAC), H3K4me1 and H3K27ac modifications, the hallmarks of enhancers11 (Extended Data Fig. 2a). To identify enhancer dynamics for the establishment of somatic and germ cell fates, we combined 150,464 distal nucleosome-free regions with differential epigenetic signals were 10–100 kb away from any location (Extended Data Fig. 2b)15,26,28.

Fig. 1 | Genome-wide transcriptome and chromatin profiling revealed the trajectories of gastrulation and hPGC development. a, Generation and collection of in vitro and in vivo samples for RNA-seq, ATAC-seq and histone modification ULI-NChIP-seq. E, embryonic day. b, UHC of gene expression (RNA-seq) using all expressed genes. c, Expression heat maps of lineage-specific genes.
**Fig. 2 | Dynamic activation of enhancers underlies cell fate transitions.**

**a.** Classification of enhancers in hPGCLCs by the intersection of histone modification peaks at combined distal OCRs (ATAC summit ± 500 bp). Note that ‘neutral’ enhancers (distal OCRs that did not overlap with any histone modification peak in the cell type of interest) are not shown. **b.** Alluvial plots showing enhancer state transitions of hPGC-active enhancers. Colour key is shown in **c.** K-means clustering of dynamically active enhancers into nine clusters by H3K27ac signals. Dynamically active enhancers were defined as enhancers that were active in any cell type with differential H3K27ac signals between the contrasting pairs shown in Extended Data Fig. 2d. **d.** Gene Ontology enrichment analysis (DAVID 6.8) on the high-confidence target genes in each dynamically active enhancer cluster. The representative terms and representative genes are shown. The full enrichment list is provided in Supplementary Table 2. **e.** Dot plots showing the enrichment of representative genes in active enhancers.
(for example, SOX17, TFAP2C, UTF1, NANOS3 and PDPN), showing strong H3K27ac enrichment in hPGCLCs and hPGCs. Motif enrichment analysis on hPGCLC-active enhancers suggested that SOX17, TFAP2C and POU5F1 might activate and maintain germ-line enhancers (Fig. 2e).

Next, we defined promoters (TSS ± 1 kb) as active, mixed, poised, repressed or neutral on the basis of their H3K4me3, H3K27ac and H3K27me3 occupancy (Methods and Extended Data Fig. 3a–c). Notably, promoters gaining H3K27me3 during the PreME–hPGCLC transition showed reduced expression in hPGCLCs and enrichment for the PRDM1 motif, which might indicate direct PRDM1-mediated promoter repression (Extended Data Fig. 3d–g).

SOX17 and PRDM1 drive hPGC fate interdependently. To investigate SOX17 and PRDM1 function during hPGCLC specification, we employed a transgenic hESC line allowing doxycycline (Dox)-inducible Myc-tagged PRDM1 and dexamethasone (Dex)-inducible haemagglutinin (HA)-tagged SOX17 expression to conduct ChIP–seq in hPGCLCs (Fig. 3a). Notably, SOX17 and PRDM1 peaks showed minimal overlap, with SOX17 being predominantly found at distal intergenic and intronic regions (>90%), while PRDM1 exhibited pronounced promoter binding (Fig. 3b,c). To identify the direct transcriptional response triggered by SOX17 or PRDM1, we treated PreME aggregates with Dox or Dex (without cytokines) for 12 h and performed RNA-seq (Fig. 3a). Integrated analysis of ChIP–seq peaks and differential gene expression revealed that SOX17 functioned mainly as a transcriptional activator, whereas PRDM1 served primarily as a transcriptional repressor during hPGCLC induction (Extended Data Fig. 4a). SOX17 directly upregulated well-known PGC genes, including PRDM1, CBFA2T2, CBFA2T3, Tet methylcytosine dioxygenase TET2, PDPN and CXCR4 (Fig. 3d, Extended Data Fig. 4b,c and Supplementary Table 3).

Notably, SOX17 was bound to the PRDM1 promoter and an ~6.2 kb upstream putative enhancer, both containing multiple SOX-binding motifs (Fig. 3e). Luciferase reporter assays in hESCs harbouring an inducible SOX17 transgene showed that SOX17 strongly activated the PRDM1 enhancers and promoters, which was abrogated by mutations in their SOX motifs, indicating that SOX17 directly upregulates PRDM1 (Fig. 3f).

Importantly, SOX17 is critical for establishing both hPGC and DE fates,6,39,41 where we found largely different SOX17 binding profiles (Fig. 3g and Extended Data Fig. 4g). Motif enrichment and transcriptional regulator binding site enrichment analyses42 of the SOX17 peaks suggested putative cell-type-specific SOX17 co-factors including POU5F1, NANOG and TFAP2C in hPGCLCs, and EOMES, SMAD2/3/4, FOXA1/A2 and ZIC2/3/5 in DE (Fig. 3h and Extended Data Fig. 4d,e). In hPGCLCs, PRDM1 directly bound promoters of genes involved in development, WNT signalling and neurogenesis, and conferred repression of these genes in PreME aggregates on PRDM1 overexpression (Fig. 3i, Extended Data Fig. 4f and Supplementary Table 4). EOMES and ZIC2/3/5, the putative SOX17 co-factors in DE, were among the direct targets repressed by PRDM1, along with SOX2 (Fig. 3j), a co-factor of POU5F1 in regulating pluripotency genes (Fig. 2e)46. SOX2 repression by PRDM1, and potentially by BMP and WNT signalling44, probably allows POU5F1 to partner with SOX17, resulting in redistribution from SOX2–POU5F1 canonical to compressed SOX17–POU5F1 motifs to activate hPGC genes45. PRDM1 potentially mediates gene repression through co-factors, for example, GATA or TFAP2 TFs (Fig. 3k). In sum, SOX17 directly activates PRDM1, which represses pluripotency- and DE-associated TFs to facilitate SOX17’s function in hPGC specification (Fig. 3l); SOX17 and PRDM1 promote the hPGC transcriptional programme interdependently without cooperative binding.

Roles of TFAP2C, SOX17 and PRDM1 in hPGCLCs. The TFAP2C–SOX17 interaction in hPGCLCs is essential for specifying hPGCLCs,16,38,42. Analysis of TFAP2C ChIP–seq data of day 4 hPGCLC aggregates46 revealed ~30,000 TFAP2C peaks evenly distributed between promoters and intronic and intergenic regions (Extended Data Fig. 5a). Integrated analysis of wild type (WT) and TFAP2C knockout (KO) hPGCLCs46 revealed that TFAP2C acted as both a transcriptional activator and a repressor (Extended Data Fig. 5b).

We observed substantial overlap between TFAP2C and SOX17 (2,466) and between TFAP2C and PRDM1 peaks (1,843), but little co-binding among the three factors (83) (Fig. 4a). Strikingly, TFAP2C alone bound to 39% of the loci gaining accessibility during hPGCLC induction, while TFAP2C–SOX17 together and SOX17 alone accounted for 13% and 6%, respectively (Extended Data Fig. 5c). Cross-referencing with our chromatin state maps, the boundary sites of TFAP2C alone (21%), SOX17 alone (4%) and TFAP2C–SOX17 (6%) together overlapped more than 30% of enhancers activated during the PreME to hPGCLC transition (Fig. 4b and Extended Data Fig. 5d). Besides being a pioneering TF47,48, TFAP2C might also contribute to promoter activation and promoter repression, both alone and with PRDM1 (Fig. 4b).

To identify individual and cooperative direct target genes of SOX17, TFAP2C and PRDM1, we integrated the DNA profiles of the three TFs with enhancer and promoter epigenetic state maps and loss-of-function RNA-seq data49 (Extended Data Fig. 5e,f, Supplementary Tables 5 and 6, and Methods). Among the
only three cooperative targets of SOX17, TFAP2C and PRDM1 was NANO53, a conserved metazoan germ cell gene (Fig. 4c.d). TFAP2C–SOX17 manifestly cooperated to directly upregulate/sustain the expression of core pluripotency factors POU5F1 and NANOG and the transcriptional repressors PRDM1 and CBF2A2T2. Interestingly, TFAP2C promoted upregulation of H3K9 demethylases KDM4B, KDM4C and ARID5B, which might trigger H3K9me2

erasure and chromatin re-organization in hPGCs[24,25], TFAP2C and PRDM1 directly mediated the expression of the core components of chromatin remodelling BAF (SWI/SNF) complex SMARCA2 and ARID1B, respectively, which maintains lineage-specific enhancers[26]. Furthermore, PRDM1 alone or with TFAP2C repressed somatic genes involved in embryonic development, anterior/posterior patterning and cell differentiation (Fig. 4c and Supplementary Table 6).
TFAP2C alone repressed homoeodomain genes (for example, HOXA1, HOXB6 and HOXB7) and epidermal growth factor-like domain genes (for example, NOTCH1 and LAMA1).

Next, we intersected the cooperative peak sets with DNA-binding profiles of 1,135 transcription regulators in the ReMap2020 database. Strikingly, 28–88% of SOX17/TFAP2C/PRDM1 individual and combinatorial peaks overlapped with the binding sites of the pluripotency factors POU5F1 and NANOG, and of the trophoderm factor TEAD4 (Fig. 4f), which showed robust expression in both hESCs and hPGCs (Fig. 1b). In hESCs, TEAD4, a key
Articles

Fig. 4 | Combinatorial and individual roles of TFAP2C, SOX17 and PRDM1 in epigenetic regulation of target genes in hPGCLCs. a, The intersection of TFAP2C, SOX17 and PRDM1 peaks in hPGCLC aggregates. Statistical significance of overlap was determined by hypergeometric test. b, The enrichment of TFAP2C, SOX17 and PRDM1 peaks in promoters and enhancers that became active or inactive during the PreME to hPGCLC transition (Extended Data Fig. 5d). The TF peaks were categorized into seven cooperativity classes as in a. Dot size represents the fraction of enhancers/promoters that overlapped with the TF peaks. c, The direct up-targeted genes of TFAP2C, SOX17 and PRDM1. The heat maps show the expression of representative target genes during hPGC development (left) and the expression pattern in TFAP2C (day 2), SOX17 (day 2) and PRDM1 (day 4) KO hPGCLCs/aggregates versus CTL (middle). The representative Gene Ontology terms enriched in the direct target genes on the basis of the binding cooperativity of TFAP2C, SOX17 and PRDM1 are shown on the right. d, Genome browser snapshots of representative TFAP2C, SOX17 and PRDM1 direct up-target genes. e, The direct down-target genes of TFAP2C, SOX17 and PRDM1 and the representative Gene Ontology terms. f, Enrichment of NANOG, POU5F1 and TEAD4 binding sites (Remap2020 non-redundant peaks) in TFAP2C, SOX17 and PRDM1 peaks in hPGCLCs. g, The enhancer-linked TF network that establishes the hPGC programme.

Fig. 5 | Enhancer and promoter trigger expression of core hPGC TFs synergistically. a, The epigenetic landscape of the SOX17, TFAP2C and PRDM1 loci in PreME and hPGCLCs. For CRISPRa assay, three to five sgRNAs were used to activate or repress each putative enhancer (Enh or E) (highlighted) and promoter (Pro or P). ‘Neutral’ regions (Neut or N) that do not bear enhancer signature were chosen as negative controls. b, An optimized Dox-inducible dCas9-SunTag-Vp64 CRISPRa system for enhancer and promoter activation in hESCs (Extended Data Fig. 6a). After stable integration of the Dox-inducible CRISPRa transgene and the plasmid encoding enhancer/promoter targeting sgRNAs to the genome, hESCs were treated with Dox for 48 h. GFP-positive cells that express the CRISPRa components were subjected to RT-qPCR and immunofluorescence analysis. c, Induction of SOX17, TFAP2C and PRDM1 mRNA following CRISPRa of enhancers and/or promoters. Stable hESCs harbouring the CRISPRa transgene and the indicated sgRNA combinations were treated with Dox for 2 days. GFP-positive cells (expressing dCas9-SunTag and scFv-sGFP-Vp64) were isolated for RT-qPCR. Average of three biological replicates, with individual replicates shown as data points. d-f, Immunofluorescence showing the induction of SOX17 (d), TFAP2C (e) and PRDM1 (f) protein by CRISPRa in hESC lines after 2 days Dox treatment. The experiment was repeated independently three times with similar results.
activation of SOX17 enhancers and the addition of BMP4 in PreME had synergistic effects with a doubling of the efficiency of hPGCLC induction compared with BMP4 treatment alone. Consequently, the gain of competence in PreME from hESCs might be attributed to a permissive TF combination that can activate SOX17 enhancers (Fig. 7d).
Next, we considered enhancers dynamically activated during ME differentiation, designated as ‘early’ (C4) and ‘late’ (C5) ME enhancers (Fig. 2c,d). Early ME enhancers (C4) lacked H3K27ac and were relatively inaccessible in hESCs but became increasingly opened up and gained H3K27ac in PreME and ME (Fig. 8a). The high-confidence targets of these enhancers were involved in ‘Wnt signalling pathway’ and ‘mesoderm formation’ (Fig. 2d), including EOMES, which is necessary for SOX17 upregulation during hPGCLC specification20,58. Motif enrichment analysis suggested that early ME enhancers were activated by downstream mediators of the FGF (JUN and FOS) and canonical WNT signalling pathway (LEF1, TCF3 and TCF7L2)20 (Fig. 8b). Indeed, EOMES is a known downstream target of the WNT signalling pathway69. On the other hand, late-activated ME enhancers (C5) only became accessible and enriched for H3K27ac in ME, with further chromatin opening and activation in DE (Fig. 8a). These enhancers targeted master mesoderm and endoderm regulators (GSC, GATA4, CER1 and LHX1) and were enriched for GATA motifs, coinciding with GATA4 and GATA6 upregulation in ME and DE (Figs. 2e and 8b,c). Notably, the OTX2 motif was enriched explicitly in late-activated ME enhancers.

Next, we analysed the cellular heterogeneity of hESCs, PreME and ME by single-cell RNA sequencing (scRNA-seq), revealing that these cell types represent distinct transcriptomic states without clear subpopulations (Extended Data Fig. 8a). However, individual genes, including EOMES and OTX2, exhibit heterogeneous expression (Fig. 8d and Extended Data Fig. 8b). In many PreME cells, the OTX2 expression level was reduced compared with hESCs and ME, while EOMES expression increased strongly relative to hESCs. We used our inducible CRISPRi system to test whether a further reduction of OTX2 in PreME could promote hPGCLC specification, and indeed there was a substantial gain of PGCLC specification efficiency (Extended Data Fig. 8c).

Therefore, the temporal reduction of OTX2 expression in PreME cells exhibiting increasing EOMES levels might critically define the gain of germine competence in the absence of later activated ME TFs, for example, GSC and GATA6. High levels of OTX2 and other ME TFs in ME abrogate germline competence and promote somatic fates (Fig. 8f).

Discussion

We demonstrate how an integrated signalling response manifests in altered epigenetic states, and the activation of developmental TFs drives human germline–soma segregation (Fig. 8f). During the hESCs–PreME transition, endogenous FGF and WNT signalling15,21 (Fig. 1a) activates early ME enhancers and genes, including EOMES required for hPGC specification20,21, WNT signalling and elevated NANOG expression in response to NODAL signalling probably contribute to the transient OTX2 reduction in a subset of PreME cells, conferring germline competence62 while delaying the mesendodermal fate. A reciprocal OTX2–NANOG relationship has been reported in human blastocysts and neuronal differentiation62,63. Consistently, CRISPRi-mediated OTX2 knockdown promotes PGCLC competence. Cell-type-specific functions of OTX2 are possible throughout hESCs–to–ME transition44, following redistribution and altered chromatin interactions44 (Fig. 8b); binding to regulatory elements in hESCs might repress hPGCLC specification (Extended Data Fig. 8e). In mice, OTX2 also restricts germine competence by interfering with TFs that drive murine PGC fate while promoting a primed pluripotent state that lacks germine competence63,64.

During the PreME-to-ME transition, early ME TFs and ACTIVIN–SMAD signalling induce expression of genes such as GSC, OTX2 and GATA457,68, which, in turn, activate somatic enhancers in ME that profoundly change the cellular response to BMP and SOX17 and drive the cells past the ‘point of no return’ for the hPGC fate (Figs. 7d and 8f). Only a fraction of epiblast cells commits to the germ cell lineage in mouse and pig embryos, indicating a high cell-intrinsic barrier for PGC fate69. Similarly, only 10–40% of PreME cells differentiate into hPGCLCs in vitro, suggesting that only cells with the appropriate epigenetic state, ME TF dosage and cell cycle stage69 might commit to the hPGC fate.
Fig. 7 | Induction of hPGCLCs by CRISPRa of key cis-regulatory elements. 

a. Generation of day 4 EBs from hESC lines harbouring the Dox-inducible CRISPRa transgene with the indicated sgRNA combinations. Note that co-activation of (1) SOX17 and PRDM1 or (2) TFAP2C, SOX17 and PRDM1 cis-regulatory elements led to the formation of NANOS3-tdTomato-positive hPGCLCs in the absence of BMP4. The experiment was repeated independently three times with similar results. 

b. Validation of CRISPRa-induced hPGCLCs by RT-qPCR of key hPGC genes. Average of technical replicates, with individual replicates shown as data points and number of replicates indicated in the figure. PCR was replicated three times with similar results. SPT indicates SOX17, PRDM1 and TFAP2C; SP indicates SOX17 and PRDM1. 

c. Induction of hPGCLCs from hESCs, PreME and ME with or without activation of SOX17 enhancers. FACs analysis of day 4 EBs shows that the activation of SOX17 enhancers and the addition of BMP4 synergistically increased the efficiency of hPGCLC induction from hESCs and PreME, but not from ME. Orange boxes indicates gates for PDPN- and NANOS3-positive hPGCLCs.

d. A model elucidating the key role of SOX17 enhancers in human germline competence.

The high hPGCLC specification on CRISPR-mediated SOX17 enhancer activation suggests that SOX17 transcriptional induction represents an essential barrier for hPGC specification. A permissive epigenetic state of the SOX17 cis-regulatory elements is a component of germline competence (Fig. 7c). The oncogenic transformation of hPGCs into pluripotent embryonal carcinoma cells and germ cell tumours entails the loss of SOX17 and the gain of SOX2 function. Therefore, the epigenetic status of the regulatory elements is probably of clinical relevance.8,19,20

EOMES is essential for germline competence, yet additional TFs are probably required for SOX17 induction since only a fraction of EOMES-positive PreME cells acquire the hPGCLC fate19,20,24. BMP4 signalling is unlikely sufficient for SOX17 induction since the expression of BMP-responsive genes ID1, ID2 and MSX2 precedes SOX17 substantially21. Putative TF-binding sites within the SOX17 enhancers, including POUSF1, EOMES, GATA3, TFAP2A/C and SMAD1, suggest a combinatorial and cooperative action of TFs at individual enhancers to drive SOX17 expression beyond a threshold.
for hPGC specification. SOX17 and TFAP2C activate germline enhancers and cooperate with their direct downstream targets to sculpt the epigenome for hPGC fate. Remarkably, CRISPR-mediated activation of the cis-regulators of SOX17, TFAP2C and PRDM1 is sufficient for hPGCLC induction without BMP4.

During hPGCLC specification, PRDM1 is a direct target of SOX17 but not in mice. Despite the mouse–human differences, the human PRDM1 enhancer bears a substantial resemblance to the murine counterpart, which interacts with OTX2 during retina development2. Since the OTX2-binding motif is conserved in the
human PRDM1 enhancer, OTX2 may modulate PRDM1 expression. Since the human and mouse PRDM1 loci show conservation of four out of five SOX motifs in their enhancers and promoters (Extended Data Fig. 8f), SOX17 can probably regulate mouse PRDM1 as exemplified by their co-expression in mouse visceral endoderm13,24,25. Altogether, SOX17 is the critical regulator of hPGC fate, while PRDM1, PRDM14 and potentially SOX2 fulfill this role in mice17,25,26.

Regulatory elements of TFs defining germ cell identity, for example, SOX17 and TFAP2C, are active in nascent hPGCLCs and more advanced gonadal hPGCs (Fig. 2c, C9). During hPGC maturation towards gametogenesis, genes regulating migration, epigenetic resetting, meiotic entry and genome defence become transcriptionally induced with the activation of the associated regulatory elements26,27. While hPGCLCs co-cultured with mouse gonadal tissue can develop an oogonia-like state, the process is highly inefficient (~1%) and requires 4 months of culture28,29. Investigating the regulatory elements in hPGCLCs and hPGCs could help optimize hPGCLC differentiation conditions by determining likely roadblocks that hinder maturation. Our re-designed CRISPRa and CRISPRi systems that allow efficient multiplexed modulation of cis-regulatory elements could be deployed to discover and overcome epigenetic obstacles during the development of hPGCLC towards gametogenesis.

The origin of hPGCs during peri-implantation development remains a challenge, with the posterior epiblast and nascent amnion being possible sites of PGC specification30. In a rare human gastrulating embryo, hPGCs were found in the epiblast31. In some mammalian embryos that develop as bilaminar discs as in humans, PGCs originate in the posterior epiblast32,33. In the future, comparing the epigenetic profiles of PrME or hPGCLCs with amniotic ectoderm-like cells34 might help to determine similarities between these cells.

With the epigenetic principles of human germline competence, specification and development, we establish a framework for in vitro gametogenesis and for decoding the mechanisms promoting the critical epigenetic resetting in the germline for totipotency and its evolutionary divergence among mammals. Understanding germline networks will help to explore the pathogenesis of infertility, germ cell cancer and age-related diseases of somatic tissues that lack the unique epigenetic resetting event present in the ‘immortal’ germline.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41556-022-00878-z.

Received: 8 October 2020; Accepted: 28 February 2022; Published online: 11 April 2022

References
1. Seydoux, G. & Braun, R. E. Pathway to totipotency: lessons from germ cells. Cell 127, 891–904 (2006).
2. Tang, W. W., Kobayashi, T., Irie, N., Dietmann, S. & Surani, M. A. Specification and epigenetic programming of the human germ line. Nat. Rev. Genet. 17, 585–600 (2016).
3. Saitou, M. & Miyauchi, H. Gametogenesis from pluripotent stem cells. Cell Stem Cell 18, 721–735 (2016).
4. Kobayashi, T. & Surani, M. A. On the origin of the human germline. Development 145, dev150433 (2018).
5. Senf, A. D., Bikoff, E. K., Robertson, E. J. & Costello, I. Genetic dissection of Nodal and Bmp signalling requirements during primordial germ cell development in mouse. Nat. Commun. 10, 1089 (2019).
6. Ohnata, Y. et al. A signaling principle for the specification of the germ cell lineage in mice. Cell 137, 571–584 (2009).
7. 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).
8. Magnusdottir, E. et al. A tripartite transcription factor network regulates primordial germ cell specification in mice. Nat. Cell Biol. 15, 905–915 (2013).
9. Kobayashi, T. et al. Principles of early human development and germ cell program from conserved model systems. Nature 546, 416–420 (2017).
10. Sasaki, K. et al. The germ cell fate of cynomolgus monkeys is specified in the nascent amnion. Dev. Cell 39, 169–185 (2016).
11. Lawson, K. A. et al. Bmp4 is required for the generation of primordial germ cells in the mouse embryo. Genes Dev. 13, 424–436 (1999).
12. Ohinata, Y. et al. Blimp1 is a critical determinant of the germ cell lineage in mice. Nature 436, 207–213 (2005).
13. Tyser, R. C. et al. Single-cell transcriptional characterization of a gametogenetic human embryo. Nature 600, 285–289 (2021).
14. D’Amour, K. A. et al. Efficient differentiation of human embryonic stem cells to definitive endoderm. Nat. Biotechnol. 23, 1534–1541 (2005).
15. Loh, K. M. et al. Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations. Cell Stem Cell 14, 237–252 (2014).
16. Irie, N. et al. SOX17 is a critical specifier of human primordial germ cell fate. Cell 160, 253–268 (2015).
17. Tsuchioka, F. et al. Induction of human germ-cell fate by transcription factors in vitro. Nature 501, 222–226 (2013).
18. Hari, K. et al. Evidence for crucial role of hindgut expansion in directing proper migration of primordial germ cells in mouse early embryogenesis. Dev. Biol. 330, 427–439 (2009).
19. Chen, D. et al. The TFAP2C-regulated OCT4 naive enhancer is involved in human germline formation. Cell Rep. 25, 3591–1603 (2018).
20. Kojima, Y. et al. Evolutionarily distinctive transcriptional and signaling programs drive human germ cell lineage specification from pluripotent stem cells. Cell Stem Cell 21, 517–532 (2017).
21. Sasaki, K. et al. Robust in vitro induction of human germ cell fate from pluripotent stem cells. Cell Stem Cell 17, 178–194 (2015).
22. de Jong, I. et al. Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications. J. Pathol. 215, 21–30 (2008).
23. Hoeri-Hansen, C. et al. Transcription factor AP-2γ is a developmentally regulated marker of testicular carcinoma in situ and germ cell tumors. Clin. Cancer Res. 10, 8521–8530 (2004).
24. Eckert, D. et al. Expression of BLIMP1/PRMT5 and concurrent histone H2A/H4 arginine 3 dimethylation in fetal germ cells, CIS/IGCNN and germ cell tumors. BMC Dev. Biol. 8, 106 (2008).
25. Tang, W. W. et al. A unique gene regulatory network resets the human germline genome for development. Cell 161, 1453–1467 (2015).
26. Li, L. et al. Single-cell RNA-seq analysis maps development of human germline cells and gonadal niche interactions. Cell Stem Cell 20, 858–873 (2017).
27. Corces, M. R. et al. An improved ATAC–seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods 14, 959–962 (2017).
28. Brind’Amour, J. et al. An ultra-low-input native ChIP–seq protocol for genome-wide profiling of rare cell populations. Nat. Commun. 6, 6033 (2015).
29. Takashima, Y. et al. Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell 158, 1254–1269 (2014).
30. Pontis, J. et al. Hominoid-specific transposable elements and KLFs facilitate human embryonic genome activation and control transcription in naive human ESCs. Cell Stem Cell 24, 724–735 (2019).
31. Kim, T. K. & Shiekhattar, R. Architectural and functional commonalities between enhancers and promoters. Cell 162, 948–959 (2015).
32. 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).
33. Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470, 279–283 (2011).
34. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? Mol. Cell 49, 825–837 (2013).
35. Zentner, G. E., Tesar, P. J. & Scacheri, P. C. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res. 21, 1237–1243 (2011).
36. Creyghton, M. P. et al. Histone H3K27ac separates active from poised developmental enhancers. Cell 154, 724–735 (2013).
37. Tü, S. et al. Co-repressor CBFA2T2 regulates pluripotency and germline development. Nature 534, 387–390 (2016).
38. Nady, N. et al. ETO family protein Mtrg1 mediates Prdm14 functions in stem cell maintenance and primordial germ cell formation. *eLife* 4, e10150 (2015).

39. Gomes Fernandes, M., Biulecka, M., Salvatori, D. C. F. & Chua de Sousa Lopes, S. M. Characterization of migratory primordial germ cells in the aorta-gonad-mesonephros of a 4.5-week-old human embryo: a toolbox to evaluate in vitro early gametogenesis. *Mol. Hum. Reprod.* 24, 233–243 (2018).

40. Sibyra, A. et al. A critical role of PRDM14 in human primordial germ cell fate revealed by inducible degrons. *Nat. Commun.* 11, 1282 (2020).

41. Seguin, C. A., Draper, J. S., Nagy, A. & Rossant, J. Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells. *Cell Stem Cell* 3, 182–195 (2008).

42. Cheneby, J. et al. ReMap 2020: a database of regulatory regions from an integrative analysis of human and *Arabidopsis* DNA-binding sequencing experiments. *Nucleic Acids Res.* 48, D180–D188 (2020).

43. Aksoy, I. et al. Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *EMBO J.* 32, 938–953 (2013).

44. Rao, J. et al. Stepwise clearance of repressive roadblocks drives cardiac induction in human ESCs. *Cell Stem Cell* 18, 341–353 (2016).

45. Jostes, S. V. et al. Unique and redundant roles of SOX2 and SOX17 in regulating the germ cell tumor fate. *Int. J. Cancer* J146, 1592–1605 (2020).

46. Chen, D. et al. Human primordial germ cells are specified by lineage-primed progenitors. *Cell Rep.* 29, 4568–4582 (2019).

47. Rothstein, M. & Simoes-Costa, M. Heterodimerization of TFAP2 pioneer factors drives epigenomic remodeling during neural crest specification. *Genome Res.* 30, 35–48 (2020).

48. Pastor, W. A. et al. TFAP2C regulates transcription in human naive pluripotency by opening enhancers. *Nat. Cell Biol.* 20, 553–564 (2016).

49. Eguizabal, C. et al. Characterization of the epigenetic changes during human gonadal primordial germ cell reprogramming. *Stem Cells* 34, 2418–2428 (2016).

50. Alver, B. H. et al. The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat. Commun.* 8, 14648 (2017).

51. Nishikawa, N. et al. The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophodermot from inner cell mass. *Dev. Cell* 16, 398–410 (2009).

52. Yagi, R. et al. Transcription factor TEAD4 specifies the trophodermot lineage at the beginning of mammalian development. *Development* 134, 3827–3838 (2007).

53. Beyer, T. A. et al. Switch enhancers interpret TGF-β and Hippo signaling to control cell fate in human embryonic stem cells. *Cell Rep.* 5, 1611–1624 (2013).

54. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159, 635–646 (2014).

55. Morita, S. et al. Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Cell* 155, 1479–1491 (2013).

56. Sun, D. et al. A functional genetic toolbox for human tissue-derived cell fate. *Mol. Reprod. Dev.* 97, 850–861 (2017).

57. Bahrami, S. & Drahos, F. Gene regulation in the immediate-early response process. *Adv. Biol. Regul.* 62, 37–49 (2016).

58. Chen, D. et al. Germline competency of human embryonic stem cells depends on eomesodermin. *Birol. Reprod.* 97, 1080–1085 (2016).

59. Chen, B. et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479–1491 (2013).

60. Funá, N. S. et al. β-Catenin regulates primitive streak induction through collaborative interactions with SMAD2/SMAD3 and OCT4. *Cell Stem Cell* 16, 639–652 (2015).
Methods

Ethics statement. Human embryonic tissues were used with permission from the National Health Service Research Ethics Committee, UK (Research Ethics Committee number 96/085). Patients (who had already decided to undergo the termination of pregnancy or conception) fully and freely consented to donate the foetal tissues for medical and academic research. Medical or surgical termination of pregnancy was carried out at Addenbrooke’s Hospital, Cambridge, UK.

Collection of hPCGs from human embryos. Crown–rump length and anatomical features including sex and digit development, were used to determine the developmental stage of human embryos with reference to Carnegie staging. The sex of embryos was determined by sex determination PCR as previously described45. Human embryonic genital ridges from individual male embryos (weeks 7–9) were dissected in PBS and separated from surrounding mesonephric tissues. The embryonic tissues were dissociated with collagenase IV (Sigma, CS138) and DNase I in DMEM/F-12 (Gibco) at 37 °C for 15–30 min (depending on tissue size). Cell suspension was diluted with FACS medium (PBS with 3% foetal bovine serum and 5 mM EDTA) and centrifuged at 500g for 5 min. The cell pellet was suspended with FACS medium and incubated with Alexa Fluor 488-conjugated anti-alkaline phosphatase (BD PharMingen, 561495; 5μl) and APC-conjugated anti-c-KIT (Invitrogen CD170S, 5 μl) antibodies for 20 min at room temperature in the dark. Cells were spun down, resuspended in FACS medium and passed through a 35 μm cell strainer. FACS was performed with SH800Z Cell Sorter (Sony), and FACS plots were generated by FlowJo (10.7.1). The alkaline phosphatase- and cKIT-double-positive populations were sorted ontoPropertyName: a property used to describe the name of a property or feature. For example, the property name of a feature might be 'color' or 'temperature'. The property type describes the type of value that the property can take. A property type might be 'string' or 'integer'.

Biosystems), and RNA was extracted according to manufacturer's protocol with the rapID rapid DR Multiplex System (Nugen). Libraries were quantified by quantitative PCR (qPCR) using KAPA Library Quantification Kit (Kapa Biosystems) according to the manufacturer's protocol. Quantified and validated libraries were subjected to single-end sequencing on the HiSeq 4000 sequencing system (Illumina), resulting in >30 million single-end reads per sample.

RNA-seq libraries of PreME aggregate with SOX17 or PRDM1 overexpression were prepared following the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB, E7760S) and the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, E7490) according to the manufacturer's protocol. Quantified and validated libraries were subjected to single-end sequencing on the HiSeq 4000 sequencing system (Illumina).

Generation of ATAC-seq libraries. Cells were sorted directly into Nuclei EZ Storage Buffer (Sigma, NUC-101) and stored at ~80 °C. ATAC-seq libraries were prepared following the Omni-ATAC protocol described by Corces et al.46 with the following modifications: tagged DNA was amplified using the KAPA HiFi HotStart Real-Time Library Amp Kit (Roche) with modified Nextera dual indexed primers as listed in Supplementary Table 7. Amplified libraries were purified using Ampure XP beads (Beckman Coulter) with double-sided size selection (first bead selection, 0.5%; second bead selection, 1.2x) according to the manufacturer's protocol. Quantified and validated libraries (~150,1000 bp) were subjected to paired-end sequencing on the HiSeq 4000 sequencing system (Illumina), resulting in >30 million single-end reads per sample.

Generation of chromatin ChIP-seq libraries. Histone modification ULI-NChIP-seq was conducted as described by Brind’Amour et al.45. In brief, cells were FACSorted in 3% FCS/PBS, pelleted by centrifugation and stored in 20 μl Nuclei EZ Storage Buffer at ~80 °C. Cells were thawed on ice, incubated with 2μl of 10 μM X-100 and 1% sodium deoxycholate and diluted into single cell suspensions (500 μl). Single cell suspension was incubated in proteinase K digestion buffer (20 mM HEPES pH 8.0, 1 mM EDTA, 0.5% SDS, 1 mg ml−1 RNase and 0.1 mg ml−1 proteinase K) for 15 min at 55 °C and 1 h at 65 °C. The DNA was purified from the eluate through AMPure XP beads and eluted in 20 μl EZ Buffer (MiniElute Reaction Cleanup Kit, Qiagen). ULI-NChIP-seq libraries were generated by the KAPA Hyper Prep Kit (KAPA Biosystems) according to the manufacturer's protocol. To minimize adaptor dimer formation, the NEBNext Adaptor and NEBNext Index PCR Primers from the NEBNext Multiplex Oligos for Illumina (Index Primers Set 1) (NEB, E7335S) were used. After library amplification, libraries were purified by AMPure XP beads with NEBNext 700 bp size selection as described previously47. Quantified and validated libraries were subjected to paired-end sequencing on the HiSeq 4000 sequencing system (Illumina), resulting in 27–96 million paired-end reads per sample. All histone modification antibodies used in this study (Supplementary Table 7) were extensively validated for their sensitivity and specificity by ULI-NChIP qPCR and ULI-NChIP-seq.

Generation of TF ChIP-seq libraries. For HA-SOX17 and Myc-PRDM1 ChIP-seq, PreME cells were induced from NANO3−/−/tdTomato hESCs bearing Dex-inducible SOX17 and Dox-inducible PRDM1 transgenes (WT + iSOX17 + iPRDM1). PreME cells were treated with vehicle (water), 100 μM Dex (Sigma) or 0.5 μg ml−1 Dox (Sigma) in the absence of cytokines. Aggregates were collected for total RNA extraction 12h after transgene induction.

Two biological replicates were collected for each transcription and epigenome analysis.

Generation of RNA-seq libraries. hESCs, PreME, ME, DE, hPCGs and hPCGs were sorted directly into extraction buffer of PicoPure RNA Isolation Kit (Applied Biosystems), and RNA was extracted according to manufacturer’s protocol with on-column DNase I treatment (Qiagen, 79254). RNA-seq libraries were generated from 5 ng total RNA using Ovation RNA-Seq System V2 (Nugen) and Ovation Rapid DR Multiplex System (Nugen)48. Libraries were quantified by quantitative PCR (qPCR) using KAPA Library Quantification Kit (Kapa Biosystems) using QuantStudio Flex Real-Time PCR System (Applied Biosystems) and validated using Agilent TapeStation 2200 with High Sensitivity D1000 ScreenTape. Libraries were subjected to single-end 50bp sequencing on HiSeq 4000 sequencing system (Illumina), resulting in >30 million single-end reads per sample.

RNA-seq libraries of PreME aggregate with SOX17 or PRDM1 overexpression were prepared following the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB, E7760S) and the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, E7490) according to the manufacturer's protocol. Quantified and validated libraries were subjected to single-end sequencing on the HiSeq 4000 sequencing system (Illumina).

To study the transcriptional response after SOX17 or PRDM1 overexpression, PreME were first induced from NANO3−/−/tdTomato hESCs bearing Dex-inducible SOX17 and Dox-inducible PRDM1 transgenes (WT + iSOX17 + iPRDM1). PreME aggregates were treated with vehicle (water), 100 μM Dex (Sigma) or 0.5 μg ml−1 Dox (Sigma) in the absence of cytokines. Aggregates were collected for total RNA extraction 12h after transgene induction.

Two biological replicates were collected for each transcription and epigenome analysis.

Generation of RNA-seq libraries. hESCs, PreME, ME, DE, hPCGs and hPCGs were sorted directly into extraction buffer of PicoPure RNA Isolation Kit (Applied Biosystems), and RNA was extracted according to manufacturer’s protocol with on-column DNase I treatment (Qiagen, 79254). RNA-seq libraries were generated from 5 ng total RNA using Ovation RNA-Seq System V2 (Nugen) and Ovation Rapid DR Multiplex System (Nugen)48. Libraries were quantified by quantitative PCR (qPCR) using KAPA Library Quantification Kit (Kapa Biosystems) using QuantStudio Flex Real-Time PCR System (Applied Biosystems) and validated using Agilent TapeStation 2200 with High Sensitivity D1000 ScreenTape. Libraries were subjected to single-end 50bp sequencing on HiSeq 4000 sequencing system (Illumina), resulting in >30 million single-end reads per sample.
RNA-seq data processing. For non-directional RNA-seq libraries listed in Extended Data Fig. 1bc, libraries were checked by FastQC (v0.11.5)93. The low-quality reads and adaptor sequences were removed by Trimm Galore (v0.4.11)94 using the default parameters. The pre-processed reads were mapped to the human reference genome (UCSC GRCh38/hg38) using STAR (2.7.1a)95 (parameters ‘–outFilterMismatchNoverLmax 0.05’–outFilterMultimapNmax 50–outMultimapperOrder Random) guided by the Gencode Human Release 30 comprehensive gene annotation96. Raw read counts per gene were extracted by the featureCounts function of the Subread package (1.6.2) using the default parameters. Normalized read counts and differentially expressed genes (absolute(log(fold change)) > 2) and adjusted P value <0.05) were obtained using DESeq2 (1.26.0) in R (3.6.2)/Bioconductor (3.10.1). For all expression analyses, a log(normalized counts + 1) transformation was applied. Only protein_coding and lincRNA genes were retained in subsequent genome-wide analysis. UHC was performed using the R hclust function with Ward's method using all expressed genes. All UHC dendrograms in this paper were re-ordered using the optimal leaf ordering algorithm in the R cba(0.2–21). Spearman’s correlation analysis was performed using the R cor command, considering the top 25% most variable genes. The accompanying dendrogram was generated using 1 – Spearman’s correlation coefficient as distance measures.

SOX17 or PRDM1 overexpression RNA-seq libraries were processed similarly but with the following modifications at the read counting step. To account for the directional reads, raw read counts per gene were extracted by featureCounts with the parameter ‘-s 2’. To exclude exogenous SOX17 and PRDM1 transcrips originating from the transgene, only reads mapping to the 5′ and 3′ untranslated regions (UTRs) of SOX17 and PRDM1 transcript forms were counted. This allowed the detection of endogenous expression levels of SOX17 and PRDM1 in response to ectopic SOX17 and PRDM1.

The RNA-seq dataset of SOX17, TFA2PC and PRDM1 KO and control hPGLCs/ICs/aggregates was retrieved from NCBI Gene Expression Omnibus (GSE93859). Reads were trimmed to 76 bp and the primers and adapters were trimmed by cutadapt (1.15) with options ‘-e 0.1 -q 20 -n 2 –O 1 -m 30 -a CTCCAGGGGCGCCCGATCC -g CTCCAGGGGCGCCCGATCC 3’ using the default parameters. Trimmed reads were mapped to the human reference genome using STAR, counted by featureCounts and normalized by Deseq2. Differential expression threshold (0.5) was set. For each biological replicate, 50% of reads were randomly chosen at random without replacement. Narrow peaks from each biological replicate were called following the Encode pipeline for ATAC-seq, but without adjustment of read position. The merged ATAC-seq and histone ChIP-seq libraries were called following the Encode protocol using the MACS2 (v4.10.4) function. To study promoters epigenetic dynamics, read counts of ATAC, H3K4me1, H3K27ac and H3K27me3 at promoter regions were obtained by featureCounts using the extended peak calling options and normalized using the log2(normalized counts + 1) across cell types. To evaluate the predictive power of chromatin marks at promoter for gene expression by receiver operating characteristic (ROC), non-neutral promoters were ranked on the basis of RNA expression levels of the associated genes. Promoters with the top 1,000 or the bottom 1,000 expressed genes were used as positives. ROC plots and area under the curve were calculated using the R pROC package. The dynamic peaks were defined as peaks with a log2(signal fold change) >1 and adjusted P value <0.05 and P value <0.0001 for ATAC peaks and <0.001 for histone peaks.

Analysis of individual epigenomic mark. For each histone mark, a combined peak set of all cell types was generated using bedtools (2.26.0) merge99. Raw read counts at genomic 1 kb bins (BEDOPS2 (4.3.45) ) that overlapped any combined peak were extracted using featureCounts (options ’-f -p -O’). Normalized and differential signals at each bin were obtained by DESeq2 in R using relative read depth between libraries as size factors, followed by log2((normalized counts + 1) transformation). Dynamic peaks were defined as absolute(log2(signal fold change)) >1 and adjusted P value <0.05 in the sample pairs shown in Extended Data Fig. 2d. ATAC–seq analysis was performed in a similar manner, except that reads were counted using a combined ATAC peak set (instead of 1 kb genomic bins). Spearman’s correlation analysis of replicates was performed using the R cor command, and the accompanying dendrogram was generated using 1 – Spearman’s correlation coefficient as distance measures (with optimal leaf ordering). PCA was performed using the R procfunction.

For peak distribution analysis (Extended Data Fig. 1e), distance between the summit of ATAC peaks or the centres of histone modification peaks and the nearest protein_coding and lincRNA genes in the Gencode Human Release 30 basic gene annotation96 was extracted using the annotatePeaks.pl script of HOMER (v4.10.4).
enhancers in all six cell types were excluded). Thus, this resulted in 21,652 dynamically active enhancers that were subjected to K-means clustering using the R kmeans function. The histone modification enrichment pattern at distal ATAC peak of each cell type was determined using meta-accessible heatmaps provided by the peakseq package. TF ChIP and input libraries were subsampled using samtools view to eliminate unmapped and low-mapping quality reads (options ‘view -F 4 -q 0.2’). Duplicated reads were removed by samtools rmdup. Reads mapped to non-canonical hg38 contigs and mitochondrial DNA (chrM) were removed by samtools view. Reads mapped to hg38 blacklisted regions were eliminated using bedtools subtract.

To determine SOX17, PRDM1 and TFA2PC cooperativity in hPGCLCs, peaks of the three TFs were merged to generate a combined peak set. Intersection of peaks and generation of Venn diagram were performed using the R Venn package (https://github.com/andrent/ReMapEnrich). The combined peaks were assigned to genes (distance to TSS from peak summit $\leq 100$kb) using BETAMinus. Direct up-target genes were defined as follows: (1) were downregulated in TFA2PC/SOX17/PRDM1 KO hPGCLCs (log, (fold change) versus the WT control $<1$ and adjusted $P$ value $<0.05$) alone or cooperatively as indicated; (2) had the corresponding TFA2PC/SOX17/PRDM1 peaks $<100$kb of the TSS; (3) had associated TF peak(s) that overlapped with ‘active’ or ‘mixed’ enhancer or promoters in hPGCLCs. Similarly, direct down-target genes were defined as follows: (1) were upregulated in TFA2PC/SOX17/PRDM1 KO hPGCLCs (log, (fold change) versus the WT control $<1$ and adjusted $P$ value $<0.05$) alone or cooperatively as indicated; (2) had the corresponding TFA2PC/SOX17/PRDM1 peaks $<100$kb of the TSS; and (3) had associated TF peak(s) that did not overlap with ‘active’ enhancer or promoters in hPGCLCs.

Luciferase reporter assay. Genomic regions containing enhancer (chr6:106,079,826–106,081,103) and promoter (chr6:106,085,395–106,086,553) of PRDM1 were amplified from hESCC genomic DNA. The WT enhancer and promoter were cloned into a piggyBAC-based luciferase (Luc+) reporter plasmid containing a hygromycin-resistant gene driven by a PGK promoter. Subsequently, the IREs motifs (ATGTTG) in the enhancer (3x) and/or promoter (2x) were mutated into AGGCAC by incorporating substitution mutations into PCR primer sequences circularized using the In-Fusion HD Cloning Plus kit (Takara). Using the Lipofectamine Transfection Reagent (Invitrogen), each reporter plasmid was transfected into NANOS3–Tdtomato reporter hESCs, together with a piggyBAC plasmid containing a constitutively expressed Luc+/hygromycin cassette and a neomycin-resistant cassette, a piggyBAC plasmid containing a Dex-inducible SOX17 transgene and a puromycin-resistant cassette, and a plasmid encoding a piggyBAC transposable. Stable cell lines were generated following triple selection by hygromycin, neomycin and puromycin. Following 24h of Dex treatment in Essential 8 medium, cells were collected and subjected to luciferase activity assay using the Dual-Glo Luciferase Assay System (Promega). Normalized luciferase activities were obtained by dividing firefly luciferase activity by Renilla luciferase activity.

CRISPrA. We designed a CRISPrA plasmid and a single guide RNA (sgRNA) targeting a non-coding (Extended Data Fig. 6a) based on the catalytically inactive Cas9 (dCas9)+SuniTag–VP64 system. For the CRISPrA plasmid, we replaced the CMV promoter in the PB–CMV–MCS–EF1–CMV promoter–a PGK promoter. Subsequently, the IREs motifs (ATGTTG) in the enhancer (3x) and/or promoter (2x) were mutated into AGGCAC by incorporating substitution mutations into PCR primer sequences circularized using the In-Fusion HD Cloning Plus kit (Takara). Using the Lipofectamine Transfection Reagent (Invitrogen), each reporter plasmid was transfected into NANOS3–Tdtomato reporter hESCs, together with a piggyBAC plasmid containing a constitutively expressed Luc+/hygromycin cassette and a neomycin-resistant cassette, a piggyBAC plasmid containing a Dex-inducible SOX17 transgene and a puromycin-resistant cassette, and a plasmid encoding a piggyBAC transposable. Stable cell lines were generated following triple selection by hygromycin, neomycin and puromycin. Following 24h of Dex treatment in Essential 8 medium, cells were collected and subjected to luciferase activity assay using the Dual-Glo Luciferase Assay System (Promega). Normalized luciferase activities were obtained by dividing firefly luciferase activity by Renilla luciferase activity.
transgenes were generated after puromycin and neomycin selection for 7–10 days. To activate the enhancers and/or promoters, cells were treated with 0.5 μg ml⁻¹ Dox in Essential 8 medium for 2 days and fixed for immunofluorescence analysis. Alternatively, sfGFP-positive cells were collected by FACS and subjected to qPCR with reverse transcription (RT–qPCR) analysis.

To induce hPGLCs with CRISPRa, hESC lines harboring the indicated sgRNA expression cassettes were differentiated into PreME and ME. Trypsinized hESCs, PreME cells and ME cells were cultured as floating aggregate for 4 days in hPGCLC induction medium supplemented with 0.5 μg ml⁻¹ Dox with or without BMP4. The day 4 EBs were subjected to immunofluorescence or qPCR of NANOS3–tdTomato-positive cells for RT–qPCR analysis. If expression of an analysed transcript was not detectable by RT–qPCR owing to its low expression level (for example, SOX17 expression in control hESCs; Fig. 7b), a Ct value of 40 (maximum cycle number) was assigned.

CRISPRi. For CRISPRi, we used the CRISPRa plasmid as the backbone and inserted a KRAB–dCas9–ecDHFR and a IRES–EGFP fragment downstream of the TRE3G promoter using the In-Fusion HD Cloning plus kit (Takara). The resulting plasmid encodes a KRAB–ecDHFR under the tight transcriptional control of a Dox-inducible promoter and a protein destabilization degron DHFR. The addition of Dox and trimethoprim (TMP) allows robust mRNA and protein expression of KRAB–ecDHFR and sgRNA transgenes, cells were selected for 7–10 days of combined puromycin and neomycin treatment after nucleofection.

To functionally test the role of the specific enhancers and neutral regions on PGLC specification, CRISPRi lines were first induced into PreME and then cultured as floating aggregate for 4 days in hPGCLC induction medium with or without 0.5 μg ml⁻¹ Dox and 10 μM TMP to induce CRISPRi. The day 4 EBs were analysed as above. Cells were first gated by EGFP status followed by quantification of hPGCLC induction efficiency in each population (EGFP⁺ or EGFP⁻) using the NANOS3–tdTomato reporter and antibody staining for PDPN–PECy7 (BioLegend, 337014; 5 μl M⁻¹) or PDPN–BV421 (BD Biosciences 566456; 5 μl M⁻¹). Induction efficiency in EGFP⁺ (CRISPRi⁺) cells was first normalized by that in EGFP⁻ cells in the same lane and relative normalized induction efficiency between CRISPRi lines was calculated in reference to the non-targeting control line.

To functionally test the role of OTX2 on hPGCLC competence, OTX2 promoter-targeting and non-targeting CRISPRi lines were pre-treated for 24 h in E8 medium followed by PreME induction with or without 0.5 μg ml⁻¹ Dox and 10 μM TMP to induce CRISPRi. PreME cells were trypsinized and cultured as floating aggregates for 4 days in hPGCLC induction medium without TMP and Dox. At day 4, EBs were analysed by FACs as described above.

Generation of scRNA-seq libraries. hESCs, PreME and ME cells were FACs sorted into PBS with 0.04 % w/v BSA (400 g ml⁻¹). Sorted populations were loaded into the 10x Genomics Chromium using the single-cell 3 reagents kit v2. Libraries were prepared as per the manufacturer’s instructions and pooled for sequencing. Libraries were sequenced on an Illumina HiSeq 4000 (paired-end; read 1: 26 cycles; i7 index: 8 cycles, i5 index: 0 cycles; read 2: 98 cycles) aiming at a minimum coverage of 50,000 raw reads per cell.

Single-cell data processing and analysis. Multiplexed single-cell libraries were processed using the 10x Genomics cell ranger pipeline. Reads were aligned to a reference genome (Homo sapiens Grc38) using STAR, and quantification of genes against an annotation reference (based on Ensembl Grc38 v90). Initial analysis of our data was done using Seurat (v3.1.4). Count data were normalized and scaled using NormalizeData based on log counts per 10,000 (logCP10K) and scaled using ScaleData. UMAP plots were calculated using the first 20 PCs. Diffusion maps of our data was done using Seurat (v3.1.4). Count data were normalized and scaled using Normalization Toolkit (GSE60138), TF KO RNA-seq (GSE99300), TFAP2C ChIP–seq (GSE140021) and OTX2 ChIP–seq (GSE61475). Genome databases used were: UCSC GRCh38/hg38, Ensembl GRCh38 v90 and Gencode Human Release 30. All other data supporting the findings of this study are available from the corresponding authors on reasonable request. Source data are provided with this paper.

RT–qPCR. Total RNA was extracted using PicoPure RNA Isolation Kit (Thermo Fisher Scientific) and complementary DNA was synthesized using QuantiFast Reverse Transcription Kit (QIAGEN) according to manufacturer’s protocols. qPCR was performed on a QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems) using SYBR Green JumpStart Taq ReadyMix (Sigma) and specific primers (Supplementary Table 7). The ΔΔCt method was used for quantification of gene expression.

Western blot analysis. Western blot analysis was performed as described previously100. In brief, proteins were separated on a 10% polyacrylamide gel using the Mini-PROTEAN system (Bio-Rad) and transferred to an Immobilon-P transfer membrane (Millipore). After blocking in 5% skimmed milk, the membrane was cut according to the molecular weight marker and decorated with rabbit anti-H3 (Abcam, ab1791; 1:10,000) and mouse anti-OTX2 (R&D Systems, AF1979; 1:1,000). Histone antibody binding was visualized using IRDye 680RD (LI-COR; 1:2,000) and the LI-COR Odyssey CLX system. OTX2 antibody binding was detected by horseradish-peroxidase-conjugated anti-goat IgG (Dako; 1:2,000 in 5% skimmed milk, 0.01% TBST) in conjunction with the Western Detection System (GE Healthcare).

Statistics and reproducibility. For ChiP–seq, ATAC–seq and RNA-seq, two independent biological replicates were included according to the guidelines of the Encode Consortium101. No statistical method was used to pre-determine sample size in other experiments. Low-quality replicates of ATAC–seq and ChiP–seq libraries were excluded from the analysis, as determined by percentage of reads in peaks, number of peaks, and genome browser visualization. As all results involved equipment-based quantitative measure and no subjective rating of data was involved, blinding is not relevant. Wilcoxon rank sum test was performed using R (v4.2.0). Hypergeometric test was performed using the R phyper command.

Availability of materials. Any queries on reagents and cell lines can be directed to a.suranigurdon.car@inc.ac.uk. CRISPRa and CRISPRi plasmids have been deposited to Addgene (183409, 183410 and 183411). Other plasmids generated in this study will be made freely available upon request. Modified hESC lines generated in this study will be made available on request upon completion of a Material Transfer Agreement.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
ChiP–seq and RNA-seq datasets are available on NCBI GEO (GSE196654). Single-cell sequencing datasets are available on ArrayExpress (E-MTAB-11135). Previously published data that were re-analysed here are: hPGC RNA-seq (GSE60138), TF KO RNA-seq (GSE99300), TFAP2C ChiP–seq (GSE140021) and OTX2 ChiP–seq (GSE61475). Genome databases used are: UCSC GRCh38/hg38, Ensembl GRCh38 v90 and Gencode Human Release 30. All other data supporting the findings of this study are available from the corresponding authors on reasonable request. Source data are provided with this paper.

References
82. Brijya, J. & Konceny, A. Fast sex identification in wild mammals using PCR amplification of the Sry gene. Folia Zool. 52, 269–274 (2003).
83. FastQC (Babraham Bioinformatics, 2010).
84. Trim Galore (Babraham Bioinformatics, 2012).
85. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
86. Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. 47, D766–D773 (2019).
87. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 29, 2078–2079 (2013).
88. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Bioinformatics 29, 57–59 (2012).
89. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
90. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44, W160–W165 (2016).
91. Gorkin, D. U. et al. An atlas of dynamic chromatin landscapes in mouse fetal development. Nucleic Acids Res. 38, R137 (2010).
92. Neph, S. et al. BEDOPS: high-performance genomic feature operations. Bioinformatics 27, 1919–1920 (2012).
93. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic regions. Bioinformatics 25, 2078–2079 (2009).
94. Neph, S. et al. BEDOPS: high-performance genomic feature operations. Bioinformatics 27, 1919–1920 (2012).
95. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589 (2010).

NATURE CELL BIOLOGY | www.nature.com/naturecellbiology
96. Wang, S. et al. Target analysis by integration of transcriptome and ChIP–seq data with BETA. Nat. Protoc. 8, 2502–2515 (2013).
97. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760 (2009).
98. Zerbino, D. R., Johnson, N., Juettemann, T., Wilder, S. P. & Flicek, P. WiggleTools: parallel processing of large collections of genome-wide datasets for visualization and statistical analysis. Bioinformatics 30, 1008–1009 (2014).
99. Horlbeck, M. A. et al. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. eLife 5, e19760 (2016).
100. Tischler, J. et al. Metabolic regulation of pluripotency and germ cell fate through α-ketoglutarate. EMBO J. 38, e99518 (2019).
101. Landt, S. G. et al. ChIP–seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 22, 1813–1831 (2012).

Acknowledgements
M.A.S. was supported by Wellcome Investigator Awards in Science (209475/Z/17/Z and 096738/Z/11/Z), an MRC research grant (RG85305) and a BBSRC research grant (G103986). W.W.C.T. received a Croucher Postdoctoral Research Fellowship and was supported by the Isaac Newton Trust. A.C.-V. was supported by the Wellcome 4-Year PhD Programme in Stem Cell Biology and Medicine and the Cambridge Commonwealth European and International Trust (203831/Z/16/Z). W.H.G. was supported by a BBSRC research grant (G103986). T.K. and M.A.S. were supported by the Astellas Foundation for Research on Metabolic Disorders. D.S. was supported by a Wellcome Trust PhD studentship (109146/Z/15/Z) and the Department of Pathology, University of Cambridge. N.I. was supported by an MRC research grant (RG85305). We thank R. Barker and X. He for providing human embryonic tissues, and C. Bradshaw for bioinformatic support. We also thank The Weizmann Institute of Science for the W1S2 hESC line and the Genomics Core Facility of CRUK Cambridge Institute for sequencing services, and R. Alberio and members of the Surani lab for insightful comments and critical reading of the manuscript.

Author contributions
M.A.S. and W.W.C.T. conceived the study. W.W.C.T. designed experiments, collected human embryonic tissues and performed bioinformatic analysis. A.C.-V. designed experiments, performed cell culture, cloning and luciferase assay and collected in vitro samples. W.W.C.T. and W.H.G. optimized and generated ATAC–seq and ULI-NChIP–seq libraries. T.K. and N.I. generated TF ChIP–seq libraries. T.K. and A.C.-V. generated RNA-seq libraries. A.C.-V. generated the scRNA-seq libraries. A.C.-V., M.D.M. and C.A.P. analysed the scRNA-seq data. W.W.C.T., A.C.-V., T.K. and D.S. designed and performed CRISPRa and interference assay. W.W.C.T., M.A.S., A.C.-V. and W.H.G. wrote the manuscript with inputs from all authors.

Competing interests
W.W.C.T. is currently employed by Adrestia Therapeutics Ltd. The other authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41556-022-00878-z.
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41556-022-00878-z.
Correspondence and requests for materials should be addressed to Walfred W. C. Tang or M. Azim Surani.

Peer review information Nature Cell Biology thanks Hiromitsu Nakauchi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | See next page for caption.
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 study\(^1\). 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 log2(normalized counts) of the top 25% most variable protein coding genes and lincRNA. For ATAC-seq, signals (log2(normalized counts)) at combined peaks across the 6 cell types were used. For H3K4me1, H3K4me3, H3K27ac, and H3K27me3 ChIP-seq, signals (log2(normalized counts)) at 1 kb bins of combined peaks were used. The samples were clustered using (1 - 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 (log2(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 Fig. 2 | See next page for caption.
Extended Data Fig. 2 | Characterisation of dynamically active enhancers. a, Chromatin profile heatmaps of ATAC, H3K4me3, H3K4me1, H3K27ac, H3K27me3 and input signals in hPGCLCs at ATAC-seq summit ± 3 kb. Segregation of ATAC-seq summits by K-means clustering using normalised chromatin mark signals. b, Distribution of chromatin mark signals at active, mixed, primed, poised, repressed, and neutral enhancers in hPGCLCs (see Fig. 2a). Enhancers per violin/box plot: 23255 active, 1288 mixed, 36999 primed, 5648 poised, 3984 repressed, 79290 neutral. Box plots depict the median, lower and upper hinges correspond to the 25th and 75th percentiles. c, Enhancer state transitions of DE-active enhancers. Distal OCRs not overlapping any histone modification peak in the analysed cell types were referred to as ‘neutral’ enhancers. d, Putative enhancers with differential H3K27ac levels (absolute(log2(fold change)) >1 and adjusted p-value < 0.05) in the indicated sample pairs. e, High confidence enhancer-gene associations. Putative enhancers were assigned to the nearest TSS. The relevance of the enhancer-gene pair was assessed by the Kendall’s rank correlation analysis between the enhancer H3K27ac signals and the expression levels of the associated genes across the 6 cell types and 2 replicates (see Methods). In the simplified model shown, the Enh1-geneA and Enh3-geneB pairs (green text and arrows) were identified as high confidence associations based on positive correlation between H3K27ac and gene expression levels. f, Expression levels of genes associated with active enhancers in different cell types. Compared to simply associating genes to the nearest active enhancer, high confidence active enhancer associated genes (Kendall rank correlation coefficient >0.3; empirical p-value < 0.05) exhibited significantly higher expression in all cell types. Two-sided Wilcoxon rank sum test with FDR correction. Gene number per violin/box plot (all nearest, high confidence): hESC (8686, 1279), PreME (9782, 1413), ME (11494, 1853), DE (12279, 2224), hPGCLC (9954, 1726), hPGC (6612, 1024). Box plot organisation as in Extended Data Fig. 2b. g, Distribution of ATAC, H3K27ac and H3K27me3 signals in dynamically active enhancers and high confidence target gene expression. Enhancers were segregated into nine clusters (Fig. 2c). Enhancer per clusters as in Fig. 2c. Box plot organisation as in Extended Data Fig. 2b.
Extended Data Fig. 3 | See next page for caption.
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 ± 1 kb). b, Receiver operating characteristic (ROC) curves of ATAC, H3K4me3, H3K4me1, H3K27ac, and H3K27me3 signals at promoter (TSS ± 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 genes = 0.770) and H3K27me3 (AUC of bottom genes = 0.681) were the best predictors of expressed and repressed genes, respectively. c, Distribution of chromatin mark signals at active, mixed, poised, repressed, and neutral promoters and the expression of their associated genes in hPGCLCs. Promoter number: 30261 active, 2833 mixed, 7089 poised, 1579 repressed, 19832 neutral. Associated gene number: 12629 active, 1526 mixed, 3662 poised, 1144 repressed, 13038 neutral. Box plot organisation as in Extended Data Fig. 2b. d, K-means clustering of dynamically repressed promoters into 7 clusters (C) by H3K27me3 signals. Dynamically repressed promoters were promoters that exhibited ‘mixed’, ‘poised’ or ‘repressed’ state (see Methods) in any cell type with differential H3K27me3 signals. e, Box plots showing expression levels of genes associated with the C6 dynamically repressed promoters in c. **** p-value < 0.00001 (Wilcoxon rank sum test adjusted by the Holm method) in marked against each unmarked cell type. Box plot organisation as in Extended Data Fig. 2b. f, Heatmaps showing the expression levels of representative genes associated with C6 in c. The right panel shows the representative enriched gene ontology terms. g, Dot plots showing the enrichment of transcription regulator (TR) binding site in the dynamic repressed promoters in C6 in c. The top 20 enriched TRs (out of 1,135 in the ReMap2020 database42) are shown. TRs were annotated against 5 gene ontology terms associated with repressive functions. The dot size represents promoter fraction overlapping with the TR binding sites. The dot colour indicates the expression levels of the enriched transcription regulators in hPGCs.
Extended Data Fig. 4 | See next page for caption.
**Extended Data Fig. 4 | Direct targets of SOX17 in hPGCLCs and DE.** a, Cumulative distribution function plot showing the functional prediction of SOX17 and PRDM1. The ChIP peaks of SOX17/PRDM1 in hPGCLCs 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 TSS. The genes were then divided into upregulated, downregulated and unchanged according to their expression patterns upon SOX17 or PRDM1 overexpression. 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. Note that upregulated genes (but not downregulated genes) upon SOX17 induction have a significantly higher tendency to be bound by SOX17. In contrast, genes downregulated upon PRDM1 overexpression tend to be bound by PRDM1. b, GO biological process terms that were enriched in SOX17 direct up targets (red dots in Fig. 3d). c, Expression heatmap of SOX17 direct up target genes. Shown are the genes which were (1) upregulated both by SOX17 alone (log2(fold change) >1 and adjusted p-value <0.05 between Dex-treated and non-treated 12h PreME aggregates); (2) upregulated by cytokines (log2(fold change) >2 and adjusted p-value <0.05 between day 2 hPGCLCs and PreME); (3) downregulated in SOX17 KO (log2(fold change) >1 and adjusted p-value <0.05 between SOX17 KO day 2 aggregate and WT day 2 aggregate)20. d, Top eight TFs with binding sites (ReMap2020) enriched in hPGCLC-specific and DE-specific SOX17 peaks. e, Genome browser snapshots showing the epigenetic landscape of DE-specific (CER1 and LEFTY2) and hPGCLC-specific (NANOS3 and PDPN) SOX17-bound gene targets. f, Heatmap showing expression of genes associated with the top enriched motifs (Fig. 3h) and the top enriched TF binding sites (in d) in hPGCLC-specific and DE-specific SOX17 peaks. g, UniProtKB Keywords that were enriched in PRDM1 direct down targets (blue dots in Fig. 3i).
Extended Data Fig. 5 | Direct target genes of TFAP2C, SOX17 and PRDM1. a, Genomic distribution of the TFAP2C peaks in hPGCLC aggregates. 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 TSS. 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 control. 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/aggregates. Upregulated and downregulated genes were defined as (log2(fold change versus wild-type control) >1 and adjusted p-value < 0.05) and (log2(fold change versus wild-type control) < -1 and adjusted p-value < 0.05), respectively. e, The number of direct up and down target genes of TFAP2C, SOX17 and PRDM1 based on their cooperative binding. f, 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 Fig. 6 | See next page for caption.
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 Fig. 7 | See next page for caption.
Extended Data Fig. 7 | Characterisation of CRISPRa-induced hPGCLCs. 

a and b, FACS analysis of day 4 EBs generated from PreME of hESC lines harbouring the Dox-inducible CRISPRa transgene with the indicated sgRNA combinations. 

c, Immunofluorescence showing the co-expression of hPGCLC markers NANOS3-tdTomato, POU5F1 and SOX17 in hPGCLCs (yellow arrowheads) induced by CRISPRa in the absence of BMP4. White arrowheads indicate SOX17 single-positive cells (presumably DE). 

Representative results of 3 biological replicates. 

d, Induction of hPGCLCs from hESCs, PreME and ME with or without CRISPR-mediated activation of SOX17 and PRDM1 enhancers and promoters. FACS analysis of day 4 EBs shows that the co-activation of SOX17 and PRDM1 act synergistically with BMP4 to increase the efficiency of hPGCLC induction from hESCs and PreME, but not from ME. 

The appearance of the EBs under brightfield and tdTomato fluorescence filter are shown next to the corresponding FACS plots. Representative results of 3 independent experiments. 

e, Alluvial plots showing enhancer state transitions of hPGCLC-active enhancers in hESCs, PreME and ME.
Extended Data Fig. 8 | See next page for caption.
Extended Data Fig. 8 | Sequence conservation of the human PRDM1 regulatory element. a, PCA analysis scRNA-seq 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 promoters 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.
Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a  Confirmed

☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

☐ The statistical test(s) used AND whether they are one- or two-sided

Only common tests should be described solely by name; describe more complex techniques in the Methods section.

☐ A description of all covariates tested

☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

☐ For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted

Give P values as exact values whenever suitable.

☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

☐ Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection
- qPCR was performed on a QuantStudio 6 Flex Real-Time PCR Systems (Applied Biosystems)
- FACS was performed using a SH800Z Cell Sorter (Sony)
- Immunofluorescences were imaged under a Leica SP8 upright or inverted scanning confocal microscope
- RNAseq, ATACseq, ChIPseq libraries were sequenced using HiSeq 4000 sequencing system (Illumina)

Data analysis
- Flow cytometry data was analysed using Flowjo (10.7.1)
- Immunofluorescence images were analysed using Velocity (6.3)
- The DACT method was used for quantification of gene expression in qPCR analysis.
- Bioinformatic software: FastQC v0.11.5, Trim Galore v0.4.1, STAR (2.7.1a), Subread (1.6.2), R (3.6.2), Bioconductor (3.10.1), DSeq2 (1.26.0), cba(0.2-21), Trimmmomatic (0.39), cutadapt (1.15), Bowtie 2 (2.3.4.1), samtools (1.7), Picard Tools (2.9.4-SNAPSHOT), deepTools(3.0.2), MACS2 (2.1.2), BEDOPS (2.4.35), HOMER (v4.10.4), plotROC(2.2.1), bedtools(2.26.0), ggalluvial (0.12.3), BETA (1.0.7), Burrows–Wheeler Aligner (v0.7.17-r1188), WiggleTools (v1.2), Vennerable [3.1.0.9000], DAVID (v6.8), RDAVIDWebService[1.24.0], ReMapEnrich (0.99.0), Seurat [v3.1.4], destiny [2.12.0], ggpubr(0.4.0), IGV (2.4.10)

For manuscripts utilizing custom algorithms or software that are not central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.
Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

ChIPseq and RNAseq datasets are available on NCBI GEO (GSE159654). Single cell sequencing datasets are available on ArrayExpress (E-MTAB-11135). Previously published data that were re-analysed here are: TF knockout RNA-seq (GSE99350), TFAP2C ChIP-seq (GSE140021) and OTX2 ChIP-seq (GSE61475). Genome databases used are: UCSC GRCh38/hg38, Ensembl GRCh38 v90 and Gencode Human Release 30. Source data are provided with this study. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioral & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nrr-reporting-summary-list.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

For NGS libraries, two biological replicates were generated according to the practice of the Encode Consortium. For other experiments, no sample size calculation was performed. As per standard practice in molecular and cell biology, at least 3 replicates were generated.

Data exclusions

Low quality replicate of ATAC-seq and ChIP-seq libraries were excluded from the analysis, as determined by percentage of reads in peaks, number of peaks, and genome browser visualisation. New biological replicate libraries were generated to ensure there are two biological replicate per library type and cell type.

Replication

The experimental findings were reliably reproduced with at least three replications.

Randomization

In each experiment, cells started from the same conditions and treatments were randomly allocated to experimental groups.

Blinding

All results involved equipment-based quantitative measure and no subjective rating of data was involved, hence blinding is not relevant.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a Involved in the study

n/a Involved in the study

Antibodies

The primary antibodies used for immunofluorescence were:
1. anti-GFP, Chicken, Polyclonal, Abcam, Cat. number ab13970, Lot. number GR361051-1 (1:1000)
2. anti-PRDM1, Rabbit, Monoclonal, Cell Signaling Technology, Cat. number 9115, Lot. number 6, Clone number C144A (1:1:200)
3. anti-SOX17, Goat, Polyclonal, R&D, Cat. number AF1924, Lot. number KGA1019061 (1:500)
4. anti-TFAP2C, Rabbit, Polyclonal, Santa Cruz Biotechnology, Cat. number sc-8977, Lot. number H0715 (1:200)
5. anti-OCT4, Mouse, Monoclonal, BD Biosciences, Cat. number 611203, Lot. number 808969, Clone number 40/Oct-3 (1:500)
The antibodies used for FACS were:
1. Alexa Fluor 488-conjugated anti-alkaline phosphatase, Mouse, Monoclonal, BD Pharmingen, Cat. number 561495, Lot. number 7132712, Clone number B4-78 (5 uL/sample)
2. APC-conjugated anti-c-KIT, Mouse, Monoclonal, Invitrogen, Cat. number CD11705, Lot. number 20289675A, Clone number 104D2 (5 uL/sample)
3. BV421-conjugated anti-PDPN, Mouse, Monoclonal, BD Biosciences, Cat. number 566456, Lot. number 9003907, Clone number LpMab-1 (50 uL/million(M) cells)
4. PECy7-conjugated anti-PDPN, Rat, Monoclonal BioLegend, Cat. number 337014, Lot. number b274229, Clone number NC-08, (5 uL/million(M) cells)
5. PerCP-Cy5.5 conjugated anti-CXCR4 antibody, Mouse, Monoclonal, BioLegend, Cat. number 306516, Lot. number B247219, Clone number 12G5 (5 uL/M cells)

The antibodies used for WB were:
1. Anti-H3, rabbit, polyclonal, Abcam, Cat. number ab1791 (1:10000)
2. Anti-OTX2, goat, polyclonal, R&D Systems, Cat. number AF1979 (1:1000)

The antibodies used for ChIP-seq (validated by ChIP-qPCR in house against known target DNA) were:
1. anti-HA, rabbit, monoclonal, Cell Signaling Technology, Cat. number 3724, Clone number C29F4 (2 ug per Chip)
2. anti-Myc, mouse, monoclonal, Cell Signaling Technology, Cat. number 2276, Clone number 9811 (2 ug per Chip)
3. anti-H3K4me1, rabbit, polyclonal/Cell, Active Motif, Cat. number 392928, Lot. number 19417002 (0.25 µg per Chip)
4. anti-H3K4me3, rabbit, polyclonal, Abcam, Cat. number ab8880, Lot. number GR73043-4 (0.25 µg per Chip)
5. anti-H3K27ac, rabbit polyclonal, Active Motif, Cat. number 39134, Lot. number 19017009 (0.25 µg per Chip)
6. anti-H3K27me3, rabbit, monoclonal, Cell Signaling, Cat. number 9733, Lot. number 8 (0.5 µg per Chip)

No home made antibodies were used in this study. All IF, WB and FACS antibodies were commercially validated as below:
- anti-GFP, Abcam, Cat. number ab13970, Lot. number GR3361051-1
  Validation: https://www.abcam.com/gfp-antibody-ab13970.html
- anti-PRDM13, Cell Signaling Technology, Cat. number 9115, Lot. number 6, Clone number C14A4
  Validation: https://www.cellsignal.co.uk/products/primary-antibodies/blimp-1-prdi-bf1-c14a4-rabbit-mab/9115
- anti-SOX17, R&D systems, Cat. number AF1924, Lot. number KGA1019061
  Validation: https://www.rndsystems.com/products/human-sox17-antibody_af1924
- anti-TFAP2C, Santa Cruz Biotechnology, Cat. number sc-8977, Lot. number H0715
  Validation (33 citations): https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5561267/
- anti-OCT4, BD Biosciences, Cat. number 611203, Lot. number 8087969, Lot number 40/Oct-3
  Validation: https://www.bdbsciences.com/en-us/products/reagents/microscopy-imaging-reagents/immunofluorescence-reagents/purified-mouse-anti-oct3-4.611202

Alexa Fluor 488-conjugated anti-alkaline phosphatase, BD Pharmingen, Cat. number 561495, Lot. number 7132712, Clone number B4-78
Validation: https://www.bdbsciences.com/en-no/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/alexa-fluor-488-mouse-anti-human-alkaline-phosphatase-561495

APC-conjugated anti-c-KIT, Invitrogen, Cat. number CD11705, Lot. number 20289675A, Clone number 104D2
Validation (7 references): https://www.thermofisher.com/antibody/product/c-kit-Antibody-clone-104D2-Monoclonal/CD11705

BV421-conjugated anti-PDPN, BD Biosciences, Cat. number 566456, Lot. number 9003907, Clone number LpMab-1
Validation: https://www.bdbsciences.com/en-gb/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/bv421-mouse-anti-human-podoplanin-566456

PECy7-conjugated anti-PDPN, BioLegend, Cat. number 337014, Lot. number b274229, Clone number NC-08
Validation: https://www.biolegend.com/en-us/products/pe-cyanine7-anti-human-podoplanin-antibody-96157clone=NC-08

PerCP-Cy5.5 conjugated anti-CXCR4 antibody, BioLegend, Cat. number 306516, Lot. number B247219, Clone number 12G5
https://www.biolegend.com/en-us/products/percp-cyanine5-5-anti-human-cd184-cxcr4-antibody-67777clone=12G5

Anti-H3, Abcam, Cat. number ab1791
Validation: https://www.abcam.com/histone-h3-antibody-nuclear-marker-and-chip-grade-ab1791.html

Anti-OTX2, R&D Systems, Cat. number AF1979
Validation: https://www.rndsystems.com/products/human-otx2-antibody_af1979

All ChIP-seq antibodies were commercially validated and additionally validated in house by ChIP-qPCR or ChIP-seq against known target DNA in hESC and/or hPSCLC.

anti-HA (Cell Signaling Technology, 3724)
Validation: https://www.cellsignal.co.uk/products/primary-antibodies/ha-tag-c29f4-rabbit-mab/3724

anti-Myc (Cell Signaling Technology, 2276)
Validation: https://www.cellsignal.co.uk/products/primary-antibodies/myc-tag-bb11-mouse-mab/2276
Eukaryotic cell lines

Policy information about cell lines

Cell line source(s) NANOS3-tdTomato reporter hESCs [WT], NANOS3-tdTomato hESCs bearing Dex-inducible SOX17 and Dex-inducible PRDM1 transgenes [WT + SOX17 + IPROM1], SOX17 knockout hESC bearing Dex inducible SOX17 and Dex inducible PRDM1 transgenes [SOX17 KO + SOX17 + IPROM1] were established previously. Luciferase reporter cells and CRISPRa and CRISPRi lines were generated in this study according to the materials and method sections.

Authentication Authentication of hESCs and PGCCLs by qPCR, RNAseq and immunofluorescence.

Mycoplasma contamination All lines tested negative for mycoplasma contamination.

Commonly misidentified lines (See ICTAC register) No commonly misidentified lines were used for this study

Human research participants

Policy information about studies involving human research participants

Population characteristics Human embryonic genital ridges were collected from 12 individual male embryos (wk7-9) for RNA-seq and Hi-C-seq as shown in Extended Data Fig. 1c. Patient’s identity is anonymised and authors have no access to donors’ metadata.

Recruitment Patients (who had already decided to undergo the termination of pregnancy operation) fully and freely consented to donate the foetal tissues for medical and academic research. Medical or surgical termination of pregnancy was carried out at Addenbrooke’s Hospital, Cambridge, UK.

Ethics oversight Human embryonic tissues were used under permission from NHS Research Ethical Committee, UK (REC Number: 96/085).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

☑ Confirm that both raw and final processed data have been deposited in a public database such as GEO.
☑ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE159654

Files in database submission

Fastq files of 54 UCI-NCChIP-seq libraries, 14 ATAC-seq libraries, 8 transcription factor ChIP-seq libraries and 18 RNA-seq libraries.

All processed peaks files (bed format) and two gene expression table from RNAseq experiments.

Genome browser session

No longer applicable

Methodology

Replicates Two biological replicates. Replicates are highly concordant as illustrated by tight clusters in genome-wide PCA (Fig. 1d). Spearman’s correlation analysis and unsupervised hierarchical clustering (Extended Data Fig. 1d)

Sequencing depth 27-96 million reads were generated per NGS library. Full sequencing statistics are provided in Supplementary Table 1.

Antibodies

anti-H3K4me1, Rabbit, Polyclonal/ Serum, Active Motif, Cat. number 39298, Lot. Number 19417002
anti-H3K4me3, Rabbit Polyclonal, Abcam, Cat. number ab8580, Lot. Number GR273043-4
anti-H3K27ac, Rabbit, Polyclonal, Active Motif, Cat. number 39134, Lot. Number 20017009
anti-H3K27me3, Rabbit, Monoclonal, Cell Signaling, Cat. number 9733, Lot. Number 8, Clone Number C36811
Flow Cytometry

Plots

- Confirm that:
  - The axis labels state the marker and fluorochrome used (e.g., CD4-FITC).
  - The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
  - All plots are contour plots with outliers or pseudocolor plots.
  - A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation
To collect hESCs, PreME, ME, DE, hPGCLCs, cells were trypsinised with 0.25% trypsin/EDTA at 37 °C for 5-15 min. Cells were stained with conjugated antibodies for 15 min at room temperature. Cell suspension was subjected to FACS.

Instrument
FACS was performed using a Sony SH800/Cell Sorter

Software
Cells were collected using the Sony SH800S software and analysis was performed using FlowJo 10.7.1

Cell population abundance
At least 10,000 cells were analysed in each FACS experiment. The percentage of tdTomato-ve hESC, PreME, ME, CXC4+ve DE, tdTomato+ve/PDPN+ve hPGCLC and and ALPL+ve/KIT single or double positive hPGC are shown in Fig. 7c, Extended Data Fig. 1a, 7a, 7b, 7d and 8d. Number of cells sorted for RNA-seq and ChIP-seq experiments are shown in Extended Data Fig. 1b and 1c.

Gating strategy
Cells were gated first based on their SSC-A FSC-A profile to avoid cell debris and dead cells, followed by gating with the respective antibody combination or knock-in-fluorescent reporter.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.