STEM CELL REPORTS
Resource
ISSCR
OPEN ACCESS

High resolution multi-scale profiling of embryonic germ cell-like cell derivation reveals pluripotent state transitions in humans

Sarah Stucchi,[1,2,9] Lessly P. Sepulveda-Rincon,[3,4,9] Camille Dion,[3,4,9] Gaja Matassa,[1,2,9] Alessia Valenti,[1,2,9] Cristina Cheroni,[1] Alessandro Vitriolo,[1,2] Filippo Prazzoli,[1] George Young,[3] Marco Tullio Rigoli,[1,2] Riccardo Nagni,[1] Martina Ciprietti,[5,7] Benedetta Muda,[1,6] Zoe Heckhausen,[3,4] Petra Hajkova,[3,4] Nicolò Caporale,[1,2,*] Giuseppe Testa,[1,2,5,*] and Harry G. Leitch[3,4,8,10,*]

[1]Human Technopole, Milan, Italy
[2]Department of Oncology and Hemato-Oncology, University of Milan, Milan, Italy
[3]MRC Laboratory of Medical Sciences (LMS), London, UK
[4]Institute of Clinical Sciences (ICS), Faculty of Medicine, Imperial College London, London, UK
[5]Department of Experimental Oncology, European Institute of Oncology IRCCS, Milan, Italy
[6]Department of Medical Biotechnology and Translational Medicine, University of Milan, Milan, Italy
[7]Present address: Laboratory of Endometrium, Endometriosis and Reproductive Medicine, Department of Development and Regeneration, KU Leuven, Leuven, Vlaams-Brabant, Belgium and Laboratory of Ion Channel Research, Department of Molecular Medicine, KU Leuven, Leuven, Belgium
[8]Present address: UCL Great Ormond Street Institute of Child Health, Genetics & Genomic Medicine Department, London, UK and North East Thames Regional Genetic Service, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK
[9]These authors contributed equally
[10]Lead contact
*Correspondence: nicolo.caporale@unimi.it (N.C.), giuseppe.testa@fht.org (G.T.), harry.leitch@ucl.ac.uk (H.G.L.)
https://doi.org/10.1016/j.stemcr.2025.102746

SUMMARY

Primordial germ cells (PGCs) are the embryonic precursors of the gametes. In rodents, PGCs readily form self-renewing embryonic germ cell (EGC) lines *in vitro*. Although human PGCs undergo a similar conversion during germ cell tumorigenesis, no comparable *in vitro* system has yet been established in humans. Here we report that hPGC-like cells (hPGCLCs) undergo conversion to human EGC-like cells (hEGCLCs) using the inductive signals previously identified in mice. This feeder-free culture system allows efficient derivation of hEGCLCs that are transcriptionally similar to human induced pluripotent stem cells and can give rise to hPGCLCs once more demonstrating the interconvertibility of pluripotent states. This is also evident at the chromatin level, as the initial DNA demethylation that occurs in hPGCLCs is reversed in hEGCLCs. This new *in vitro* model provides a highly tractable system to study human pluripotent and early developmental transitions, including those driving germ cell tumorigenesis and epigenetic inheritance.

INTRODUCTION

In sexually reproducing organisms, germ cells play a crucial role in transmitting the genetic and epigenetic information necessary for the development of a new organism, serving as an enduring link between generations (Leitch et al., 2013a; Ramakrishna et al., 2021). Primordial germ cells (PGCs), the earliest embryonic precursors of the gametes (Saitou and Hayashi, 2021), are likely specified in humans around embryonic day 12–16 (post-conceptional week 2) (Alberio et al., 2021). At weeks 3–5, human PGCs (hPGCs) migrate from the yolk sac to the developing gonads and commence a wave of global DNA demethylation, including significant loss of DNA methylation at imprinting control regions (Gkountela et al., 2015; Guo et al., 2015; Tang et al., 2015; von Meyenn and Reik, 2015), which heralds the most extensive epigenetic reprogramming process during development.

PGCs are traditionally regarded as unipotent, as physiologically they give rise only to gametes. However, mouse PGC (mPGC) development is dependent on the expression of pluripotency genes (Campolo et al., 2013; Kehler et al., 2004; Zhang et al., 2018), and PGCs exhibit pluripotent characteristics. For instance, mPGCs are the cell of origin of pluripotent germ cell tumors such as teratocarcinomas, from which mouse embryonal carcinoma cells were first derived (Finch and Ephrussi, 1967; Kleinsmith and Pierce, 1964), and, in culture, mPGCs can directly give rise to pluripotent stem cells (PSCs) called embryonic germ cells (EGCs) (Matsui et al., 1992; Resnick et al., 1992). Mouse EGCs can contribute to chimeras (Labosky et al., 1994; Stewart et al., 1994) and their properties are indistinguishable from those of naïve embryonic stem cells (ESCs), despite their different origin (Leitch et al., 2013b). In fact, mouse EGC conversion from mPGCs is highly efficient, calling into question the designation of PGCs as a unipotent lineage (Leitch et al., 2013c). Indeed, we have recently revisited the classic chimaera experiments (Leitch et al., 2014), concluding that mPGCs exhibit a latent form of pluripotency that distinguishes them from somatic precursor cells (Sepulveda-Rincon et al., 2024).

The latent pluripotency of PGCs is not a mouse-specific property. Naïve pluripotent EGC lines have also been derived from rats (Blair et al., 2012; Leitch et al., 2010;

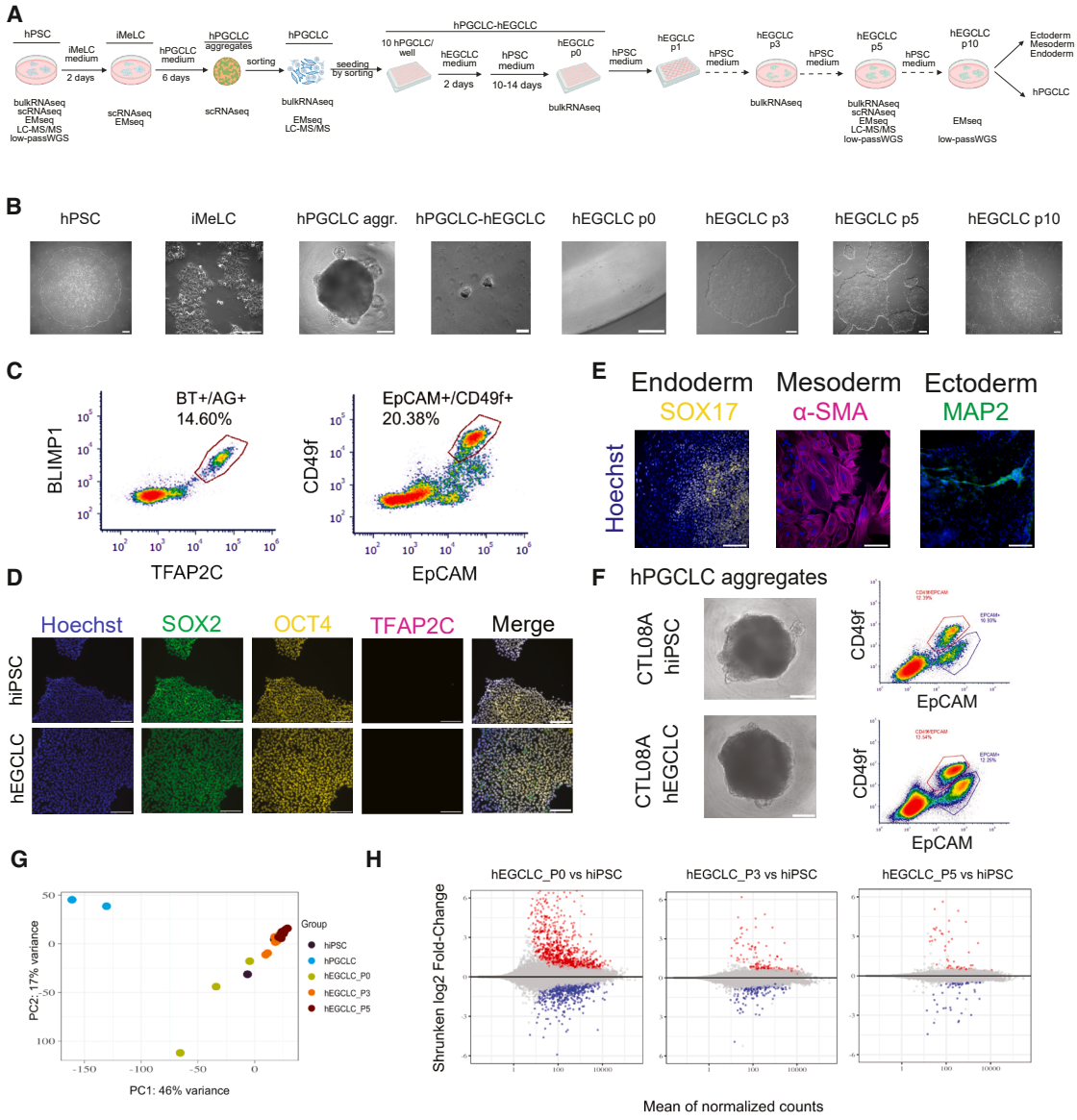

Figure 1. Derivation of pluripotent hEGCLCs from hPGCLCs in defined and feeder-free conditions

(A) Workflow of hPSC differentiation to hPGCLCs, conversion to hEGCLCs, and associated omics assays.

(B) Representative phase contrast images of the key steps and cellular states of the protocol. Scale bars: 250 μm and 25 μm for hPGCLC-hEGCLC image taken 48 h after sorting.

(C) Representative flow cytometry density plot of day 6 hPGCLC aggregates from BTAG (left) and CTL08A (right) lines. Double-positive cell percentages are indicated.

(D) Representative immunofluorescence images of CTL08A hiPSC and hEGCLC colonies stained for SOX2, OCT4, and TFAP2C markers; nuclei stained with Hoechst. Scale bars, 150 μm.

(E) CTL08A hEGCLC tri-germ layer differentiation: representative immunofluorescence staining for lineage specification markers: SOX17 (endoderm), α-SMA (mesoderm), MAP2 (ectoderm); nuclei stained with Hoechst. Scale bars, 150 μm.

(F) Left: representative phase contrast images of day 6 hPGCLC aggregates differentiated in parallel from CTL08A hiPSCs (top) and hEGCLCs (bottom). Scale bars, 250 μm. Right: corresponding flow cytometry density plots. Percentages of $EpCAM^{+}/CD49f^{+}$ cells (i.e., hPGCLCs) are shown.

(G) Principal-component analysis (PCA) of BTAG bulk RNA-seq profiling of the samples indicated in the legend (hiPSC in duplicate; day 4 hPGCLC in duplicate; hEGCLC P0 [14 days after EGC induction] in triplicate; and hEGCLC P3 and hEGCLC P5, nine replicates each).

(legend continued on next page)

Northrup et al., 2011), completing the naïve PSC repertoire in this species, following the derivation of rat ESCs (Buehr et al., 2008; Li et al., 2008) and induced PSCs (iPSCs) (Kobayashi et al., 2010). In humans, hPGCs can give rise to germ cell tumors, many of which have pluripotent properties (Oosterhuis and Looijenga, 2019). However, although the derivation of human ESCs (Thomson et al., 1998) and iPSCs (Takahashi et al., 2007) was reported decades ago, to date, human EGCs that exhibit the same properties have not been derived (Turnpenny et al., 2006). This inability may stem from species-specific differences—such as the lack of SOX2 expression in hPGCs (Perrett et al., 2008)—and the limited access to hPGCs, particularly at the pre-migratory and migratory stages, known to be most amenable to EGC conversion in mice (Labosky et al., 1994; Leitch et al., 2013c). The recent advances in *in vitro* induction of human PGC-like cells (hPGCLCs) have provided a new and exciting tool to study the early human germline (Irie et al., 2015; Sasaki et al., 2015). Of note, several labs have developed culture conditions for hPGCLCs' expansion that maintain their PGC-like properties (Gell et al., 2020; Kobayashi et al., 2022; Murase et al., 2020). Modifications to the media composition can allow the emergence of PSC colonies from proliferating cultures of hPGCLCs (Kobayashi et al., 2022). However, a defined culture system to consistently trigger this conversion is still lacking. Building on our previously established fully defined, feeder-free culture system for the derivation of mouse EGCs (Borkowska and Leitch, 2021; Leitch et al., 2013c), here we report the development of a fully defined, feeder-free culture system that enables the transition of human germline cells (hPGCLCs) to PSCs, called human embryonic germ cell-like cells (hEGCLCs). We demonstrate the utility of our novel culture system by undertaking a high-resolution, multi-omics profiling during the conversion of hPSCs to hPGCLCs and back again, producing a rich dataset that provides insight into these unique and clinically relevant pluripotent transitions and providing a comprehensive resource for the broader research community.

RESULTS

hEGCLC derivation from hPGCLC in defined and feeder-free conditions

Given the similar requirements for PGCLC induction between mouse and human, we hypothesized that the factors that trigger the conversion of PGCs (or PGCLCs) to PSCs might also be conserved.

First, hPGCLCs were differentiated from human PSCs (hPSCs) through incipient mesoderm like-cells (iMeLCs), using a well-established protocol (Sasaki et al., 2015) (Figures 1A and 1B). At day 4 or 6, hPGCLCs were identified by the expression of BLIMP1 and TFAP2C fluorescent reporters (BTAG) (Sasaki et al., 2015) or by using previously published cell surface markers, EpCAM (CD326) and integrinα6 (CD49f) (Sasaki et al., 2015; Yamashiro et al., 2020) (Figure 1C). After sorting, hPGCLCs were plated on Geltrex in an induction medium containing the same factors that trigger efficient EGC conversion in mice (N2B27 basal medium supplemented with human leukemia inhibitory factor [hLIF], the GSK3 inhibitor CHIR99021 [CH], forskolin [FK], basic fibroblast growth factor [bFGF], human stem cell factor [hSCF], and retinoic acid [RA] [see methods]). After 48 h, the medium was transitioned to standard hPSC media by half medium changes. In these defined and feeder-free culture conditions, we observed the emergence of colonies that resembled hPSCs. After 10–14 days, these colonies were picked, expanded, and maintained using standard hPSC culture protocols (Figures 1A and 1B). Using this protocol, we have derived multiple such cell lines from hPGCLCs induced from both hiPSCs and ESCs (male and female, supplemental methods). In keeping with a previous study (Kobayashi et al., 2022), we have designated these cell lines as hEGCLCs to reflect their origin from hPGCLCs.

Next, we monitored the transition from hPGCLC to hEGCLC using live cell imaging. Cultured hPGCLCs initially exhibited a typical migratory morphology, but over the first week of culture, small colonies began to form and by day 14 colonies with the typical appearance of hPSCs were present (Figure 1B). Live imaging of the BTAG fluorescent reporters (Sasaki et al., 2015) indicated that the hPGC hallmark genes *TFAP2C* and *BLIMP1* are downregulated from around day 7 of the culture and are no longer expressed in hEGCLC lines (Videos S1 and S2). In addition, we performed live imaging using a previously published SOX2 reporter cell line (Balboa et al., 2017), which revealed a reciprocal transcriptional upregulation of the pluripotency gene *SOX2* over a similar time frame. As expected, *SOX2* expression was maintained thereafter in hEGCLCs (Video S3). Thus, live imaging indicates that downregulation of the PGC program occurs in a similar time frame to the upregulation of *SOX2* and also rules out that hEGCLC colonies emerge from contaminating undifferentiated hPSCs. hEGCLCs maintain karyotypic stability when assessed at

(H) Mean average (MA) plot showing the mean normalized counts (x axis) and shrunken log2 fold-change values of expressed genes for the following pairwise comparisons: hEGCLC P0 vs. hiPSC; hEGCLC P3 vs. hiPSC; hEGCLC P5 vs. hiPSC (from left to right, respectively), BTAG line. Differentially expressed genes (DEGs) after p value adjustment (q value < 0.01) are displayed in red (upregulated) or in blue (downregulated).

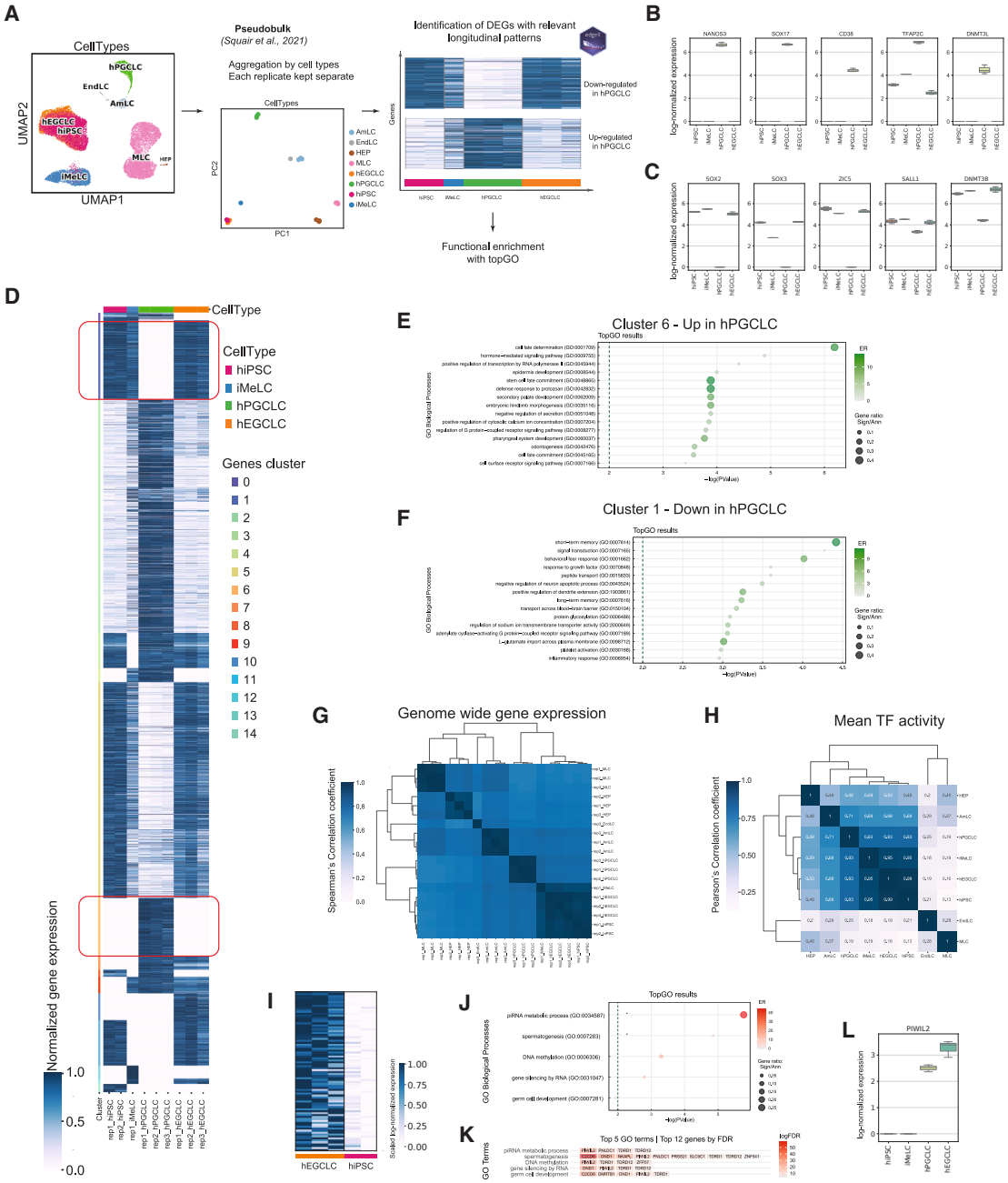

Figure 2. Longitudinal transcriptomic profiling at single-cell resolution

(A) Schematic of the pseudo-bulk differential expression analysis done on the scRNA-seq data (CTL08A line). From left to right: UMAP visualization of all cells in the dataset, after pre-processing and filtering, colored by annotated cell type (hiPSC, human induced pluripotent stem cell; iMeLC, incipient mesoderm-like cell; hPGCLC, human primordial germ cell-like cell; hEGCLC, human embryonic germ cell-like cell; MLC, mesoderm-like cell; AmLC, amnion-like cell; EndLC, endoderm-like cell; HEP, hemato-endothelial progenitor. PCA plot of pseudo-bulk counts, colored by annotated cell types. An unsupervised approach was used to identify the clusters of gene expression patterns, which were then used as input for functional enrichment analysis.

(B and C) Boxplots of pseudo-bulk log-normalized level of expression (*y* axis) of representative genes upregulated or downregulated in hPGCLC. Boxes depicts the lower and upper quartiles of the distribution while whiskers extend to the minimum and maximum of the distribution up to 1.5 of the inter-quartile range of the box.

(D) Heatmap of gene clusters obtained as described in (A). Clusters of genes upregulated (cluster 6) or downregulated (cluster 1) in hPGCLC compared to hiPSC, iMeLC, and hEGCLC are highlighted with a red box.

(legend continued on next page)

passage 5 (P5) and P10 (Figure S1A), can be freeze-thawed, and maintain the same undifferentiated morphology after at least 20 further passages. They express pluripotency markers SOX2, OCT4, NANOG, and SSEA-4 and no longer express PGC markers such as TFAP2C, BLIMP1, and SOX17 (Figures 1D and S1B–S1D; Video S2). hEGCLCs can differentiate into all three germ layers (Figures 1E and S1E) and can form hPGCLCs once again, with an efficiency comparable to that of the parental hiPSCs (Figure 1F). Thus, hEGCLCs show the broad *in vitro* differentiation capacity expected from hPSC lines.

Finally, we investigated the efficiency with which hPGCLCs can give rise to hEGCLCs. In initial experiments, we plated 10 hPGCLCs per well in a 96-well plate and counted the number of hEGCLC colonies that emerged after 14 days. This revealed a derivation efficiency between 3% and 18% (Figure S1F). The variability in derivation efficiency is in keeping with findings in hiPSC derivation and EGC derivation from other species. However, we also observed broad patterns depending on the parental cell line used, which may reflect genetic background or other cell-line-specific factors. The highest efficiencies obtained approach those reported for mouse EGC derivation from wild-type PGCs in optimized conditions (Leitch et al., 2013c). However, as we have previously observed that mPGCs can give rise to more than one EGC colony (Leitch et al., 2013c), we sought to confirm these findings using singly plated hPGCLCs. For the cell line tested, we obtained a broadly similar derivation efficiency (Figure S1G) and observed no examples of wells containing more than one primary colony. This confirms that single hPGCLCs can give rise to hEGCLCs and suggests, unlike for mouse PGCs, that significant overestimation of conversion efficiencies is unlikely to occur due to single hPGCLCs forming multiple hEGCLC colonies. Collectively, these results establish the first fully defined and highly efficient method to derive hEGCLCs.

Transcriptional profiling of the hPGCLC-to-hEGCLC transition

A benefit of an efficient and defined cell culture system is that it facilitates downstream molecular profiling. We first undertook transcriptomic profiling by bulk RNA sequencing (RNA-seq) of the hiPSC parental line, day 4 hPGCLCs, and hEGCLCs at P0 (14 days after hEGCLC induction), P3 (hEGCLCs P3), and P5 (hEGCLCs P5) (Figure 1G). This revealed that by P5, hEGCLCs cluster together with hiPSCs in principle-component analysis (PCA), pointing to the similarity of their transcriptomes (Figure 1G). Further, the number of differentially expressed genes (DEGs) between hEGCLCs and hiPSCs is reduced from 467 at P0 to only 46 at P3 and 50 at P5 (Figure 1H; Table S1). The high number of DEGs at P0 likely reflects a still mixed population of hEGCLCs and non-reprogrammed cell types. Established hEGCLC lines express pluripotency genes at similar levels to hiPSCs and have downregulated PGC genes (Figure S1H). These results confirm that hEGCLCs established from hPGCLCs are transcriptionally similar to hiPSCs, further confirming their identity as PSCs.

To increase the resolution of transcriptome analysis, we profiled hiPSC, iMeLC, day 6 hPGCLC aggregates, and hEGCLC transcriptomes at the single-cell level, obtaining 27,925 high-quality cells (see methods). To reduce the dimensionality of the dataset, we applied PCA (Figure S2A) on highly variable genes (HVG) selected with Triku (Ascensión et al., 2022) and uniform manifold approximation and projection (UMAP) (Figure S2B).

We first confirmed hiPSC, hEGCLC, and iMeLC identity by checking the expression of pluripotency (*SOX2*, *DPPA4*, *PRDM14*, and *ZIC5*) and iMeLC (*EOMES*, *SP5*, and *MIXL1*) markers, respectively (Figures 2A left, S2C, and S2D). This first instance of hEGCLC single-cell RNA sequencing (scRNA-seq) profiling exposed their remarkable, hiPSC-equivalent degree of homogeneity (Messmer et al., 2019) (Figures 2A left and S2B). We then systematically annotated the identity of each cell population captured by analyzing the expression of validated markers from the existing knowledge bases (Chialastri et al., 2022; Choi et al., 2012; Guo et al., 2015; Kojima et al., 2017; Nakamura et al., 2016; Sasaki et al., 2015; Zheng et al., 2021), including markers from a reference single-cell dataset of a gastrulating human embryo (Tyser et al., 2021) (Figure 2A left; Table S2). We annotated cells expressing PGC markers (*NANOS3,*

(E and F) Dot plot of significantly enriched gene ontology (GO) terms (biological processes category, BP) for gene clusters 6 (E) and 1 (F), highlighted in (D). *y* axis: GO term; *x* axis: −log(*p* value). Color gradient indicates the enrichment ratio (ER) > 2, and dot dimension indicates gene ratio. The dotted vertical line indicates the *p* value threshold at 0.01.

(G) Clustered heatmap of Spearman's correlation coefficient between the annotated cell types.

(H) Heatmap of Pearson's correlation coefficient of TF activity between annotated cell types.

(I) Heatmap of pseudo-bulk gene expression of the 82 genes upregulated in hEGCLC compared to hiPSC. Color gradient: scaled log-normalized expression.

(J) Dot plot of significantly enriched GO terms (BP category) of DEGs in (I). Threshold as in (E).

(K) Top 12 genes by false discovery rate (FDR) for each of the top 5 GO terms shown in panel (J). Color gradient indicates logFDR.

(L) Boxplots of pseudo-bulk log-normalized level of expression (*y* axis) of *PIWIL2* gene in hiPSC, iMeLC, hPGCLC, and hEGCLC (*x* axis). Boxes depicts the lower and upper quartiles of the distribution while whiskers extend to the minimum and maximum of the distribution up to 1.5 of the inter-quartile range of the box.

CD38, KLF4, TFAP2C, PRDM1, SOX17, KLF4, and *PIFO*) as hPGCLCs; cells expressing amnion markers (*TFAP2A, GATA3, ISL1, WNT6, GABRP,* and *HAND1*) as amnion-like cells (AmLCs), cells expressing endoderm markers (*FOXA2, APOA1, APOB, TTR, FGB,* and *MTTP*) as endoderm-like cells (EndLCs), cells expressing hemato-endothelial markers (*MEF2C, GMFG, LAPTM5, ICAM2,* and *LMO2*) as hemato-endothelial progenitors, and cells expressing mesoderm markers (*HAND1, GATA6, FOXF1,* and *SNAI2*) as mesoderm-like cells (Figures 2A left, S2C, and S2D). These data outline a heterogeneous composition of hPGCLC aggregates (Figure S2B, green cluster), recapitulating cell types present in the early phases of human embryo development (Rossant and Tam, 2022; Shahbazi and Zernicka-Goetz, 2018), as also highlighted in other studies (Bleckwehl et al., 2021; Castillo-Venzor et al., 2023).

To compare our *in vitro* model with publicly available *in vitro* data, we projected our longitudinal scRNA-seq data on the scRNA-seq dataset from an external *in vitro* reference (Chen et al., 2019), where hPGCLCs were differentiated from human ESCs (hESCs) using the same protocol (Sasaki et al., 2015) (Figure S2E). We confirmed how our hPGCLCs map onto the published *NANOS3*$^{+}$ cells (i.e., hPGCLCs), as well as iMeLCs onto iMeLCs and hiPSCs onto hESCs (Figure S2F). Interestingly, hEGCLCs, present only in our dataset, map onto published hESCs, further confirming their transcriptional overlap with hPSCs (Figure S2F), as also evident from their quite even distribution among the clusters we annotated as hEGCLC and hiPSC in our dataset (Figures 2A left and S2B; Table S2).

Finally, we benchmarked our *in vitro* data with *in vivo* fetal references (human prenatal gonads from 6 to 16 weeks post-fertilization (Chitiashvili et al., 2020; Guo et al., 2021), Figure S2G and Table S2) noting how our hPGCLCs resemble germ cells rather than fetal somatic cells (Figure S2H).

Longitudinal gene expression patterns reveal the emergence and dissolution of germline identity

To identify gene expression patterns throughout differentiation, we analyzed our scRNA-seq data through a pseudobulk approach (Hao et al., 2021; Lun and Marioni, 2017; Zimmerman et al., 2021) to leverage the statistical rigor of generalized linear models for differential expression analysis (Squair et al., 2021). We aggregated the cells according to the annotated cell types (Figure 2A), keeping each replicate separate, and found that different replicates of the same cell type cluster together, as do hiPSCs and hEGCLCs (Figure 2A middle), as also confirmed by Spearman's correlation analysis (Figure 2G).

Next, we defined groups of DEGs showing specific longitudinal expression patterns, in the four key cell types (namely, hiPSCs, iMeLCs, hPGCLCs, and hEGCLCs), using an unsupervised approach. We computed the log fold-change (logFC) of genes in the latter three cell types with respect to their expression levels in hiPSC, filtered by a minimum significance of 0.01, and then used the logFC to identify the groups of interest (Robinson et al., 2010) (supplemental methods; Figure 2A right). Of the 15 clusters identified (Figure 2D; Table S3), we noted that clusters 2 and 6 include genes (such as *NANOS3, SOX17, CD38, TFAP2C,* and *DNMT3L*; Figure 2B) whose expression increases in hPGCLCs and decreases again in hEGCLCs and clusters 5 and 1 that include genes (such as *SOX2, SOX3, ZIC5, SALL1,* and *DNMT3B*; Figure 2C) that show the reciprocal pattern. These expression patterns are aligned with what is already known from hiPSC-to-hPGCLC developmental transition (Chen et al., 2019; Sasaki et al., 2015), demonstrating the robustness of our experimental and analytical approach and adding new insights into hPGCLC-to-hEGCLC transition.

We then performed functional enrichment analysis (Alexa and Rahnenführer, 2022) on these DEGs, focusing on clusters 6 and 1, which show the strongest up/downregulation in hPGCLCs compared to the other three cell types (Table S3). For cluster 6 (genes upregulated in hPGCLCs), we identified terms related to cell fate determination and embryo development and morphogenesis (Figure 2E). Interestingly, top genes (by false discovery rate) associated with these terms are *SOX17*, *CD38, ITGB3,* and *GATA* transcription factors (TFs), known early PGC markers (Irie et al., 2015; Kojima et al., 2021), genes related to WNT signaling pathway (such as *WNT7A* [Castillo-Venzor et al., 2023; Overeem et al., 2021]), already known to be upregulated upon hPGCLC differentiation (Jo et al., 2022; Kojima et al., 2017; Yamashiro et al., 2018), as well as genes involved in PGC migration (such as *EDN1, MSX1,* and *MSX2* Kojima et al., 2017; Kojima et al., 2021; Sun et al., 2016) (Figure S2I). Instead, among DEGs from cluster 1 (genes downregulated in hPGCLCs), we identified terms related to differentiation toward other lineages, such as regulation of dendrite extension, platelet activation, and inflammatory response (Figures 2F and S2J). This is consistent with the need to not only activate germ-cell-specific genes but also to repress genes specific for other cell lineages (Kojima et al., 2017; Sasaki et al., 2015; Surani et al., 2007; Tang et al., 2022).

Next, to investigate the regulatory logic of the developmental transitions exposed by our longitudinal design, we performed TF activity inference using decoupleR package (Badia-I-Mompel et al., 2022) and DoRothEA regulons collection (Garcia-Alonso et al., 2019). First, we conducted a correlation analysis of TF activity that highlighted the highest regulatory similarity between hiPSCs and hEGCLCs (Pearson's correlation coefficient equal to 0.99) (Figure 2H). Among the 292 TFs considered in the analysis

(see supplemental methods), we selected those relevant to pluripotency, as well as germ cell and endoderm specification pathways, and plotted them in a UMAP to visualize the inferred TF activity scores. *NANOG* and *OCT4* (i.e., *POU5F1*) scores are higher in hiPSC and hEGCLC clusters, whereas *TFAP2C* score is higher in AmLC and hPGCLC clusters, and *FOXA2* score in EndLC cluster, consistent with our own and others' previous findings (Castillo-Venzor et al., 2023) (Figure S2K).

To deepen the investigation of finer differences between hiPSCs and hEGCLCs, we performed differential expression analysis focusing specifically on this comparison. We used the pseudo-bulk approach and found 34 genes downregulated (Figure S2L; Table S3) and 82 genes upregulated (Figure 2I; Table S3) in hEGCLCs compared to hiPSCs. By performing a functional enrichment analysis on the hEGCLC upregulated genes, we found terms related to PIWI-interacting RNA (piRNA) metabolic processes and spermatogenesis. *PIWIL2* is one of the top genes associated with these terms (Irie et al., 2023) (Figures 2J and 2K; Table S3), and we noted its expression level is higher in hPGCLCs than in hiPSCs and then does not decrease in hEGCLCs (Figure 2L).

In summary, this longitudinal transcriptomic profiling at single-cell resolution elucidated the gene expression patterns and the transcriptional logic underlying the transitions from the hPSC state to hPGCLCs and back.

GRN analysis highlights key regulators of hEGCLC transcriptional program

To better investigate the regulatory logic governing the reprogramming from PSCs to germ cells and back, we performed gene regulatory network (GRN) analysis using CellOracle (Kamimoto et al., 2023), leveraging our own scRNA-seq dataset alongside publicly available Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) profiles (Chen et al., 2018; Gu et al., 2024). For GRN construction, we associated each regulatory region (identified by chromatin accessibility) with a target gene (among the HVGs and a list of selected genes, see supplemental methods and Table S4) and with one or more TFs, building the TF-target gene graphs. From this first GRN, we harnessed our transcriptomic data to build four different cell-type-specific GRNs (hiPSC, iMeLC, hPGCLC, and hEGCLC GRN, Table S4).

First, we considered two key GRN parameters: the betweenness centrality (i.e., a measure proportional to the number of times each TF connects to other nodes in the network, indicating the importance of each TF in that GRN, Figure 3A) and the out-degree centrality (i.e., a measure of the number of edges that go out from that node, a measure of the number of targets regulated by that TF within the network, Figure S3A). This analysis revealed that the top regulators of hiPSC and hEGCLC GRN are *MYC, HDAC2, ZIC5,* and *SP3*. Instead, for iMeLC, *EOMES* and *SP5* emerged as master regulators, as expected (Sasaki et al., 2015), together with TFs such as *ENO1* and *HDAC2* that are in common with their pluripotent precursors. As for hPGCLC, among the top regulators we identified *TFAP2C, PRDM1* (i.e., *BLIMP1*), and *KLF4*, in line with what is already known in literature (Chen et al., 2018; Kojima et al., 2017, 2021; Saitou and Hayashi, 2021), confirming the robustness of our analysis, but also *REST*, which has been only recently identified as an important TF in PGCs (Cheng et al., 2024) (Figures 3A and S3A). These top regulators also emerged when comparing GRNs of these cell types in pairs (Figure 3B).

We then assessed the cooperativity among the TFs of the GRN, for each cell type, by using a hypergeometric test to evaluate the statistical significance of their overlap in terms of shared targets. In hiPSC and hEGCLC GRNs, among the top cooperating TFs we identified TFs belonging to *ID, TCF, FOXO,* and *SOX* families, known to be associated with pluripotency (Jiang et al., 2022; Santini et al., 2024; Sierra et al., 2018; Wong et al., 2016); in iMeLC again, *ID, TCF, FOXO*, but also *REST, SP3,* and *SP5*; and in hPGCLC, *KLF4, PRDM1, TFAP2C, REST,* and *STAT2* (Figure 3C). We also performed the same analysis, in a supervised way, focusing on known TFs of interest, confirming the results obtained so far and highlighting the high number of common targets between top cooperating TFs for each cell type. For example, the pairs *MYC-HDAC2* and *MYC-SP3* in hiPSC and hEGCLC with 761 and 811, 742 and 806 common targets, respectively, and *TFAP2C* and *PRDM1* with 389 common targets in hPGCLC (Figure S3B).

In summary, GRN analysis highlighted how hiPSC and hEGCLC have a similar regulatory logic in terms of key TFs and target genes, suggesting *MYC, ZIC5, HDAC2,* and *SP3* as the key regulators and how these regulatory networks differ from that of hPGCLC, in which *TFAP2C, PRDM1, KLF4,* and *REST* emerged as master regulators.

Longitudinal DNA methylation profiling to investigate the reversibility of epigenetic states

During their migratory phase, hPGCs initiate a wave of global DNA demethylation, including significant loss of methylation at imprinting control regions (Gkountela et al., 2015; Guo et al., 2015; Tang et al., 2015). To evaluate the extent to which our *in vitro* model recapitulates the genome-wide reprogramming events occurring *in vivo*, we profiled its DNA methylation levels longitudinally by enzymatic methyl sequencing (EM-seq) (Vaisvila et al., 2021). As evident from the PCA plot, samples corresponding to the same differentiation step cluster together, with the first principal component (PC1) explaining 36% of the variance, mainly driven by hPGCLCs (Figure 4A). hiPSCs and

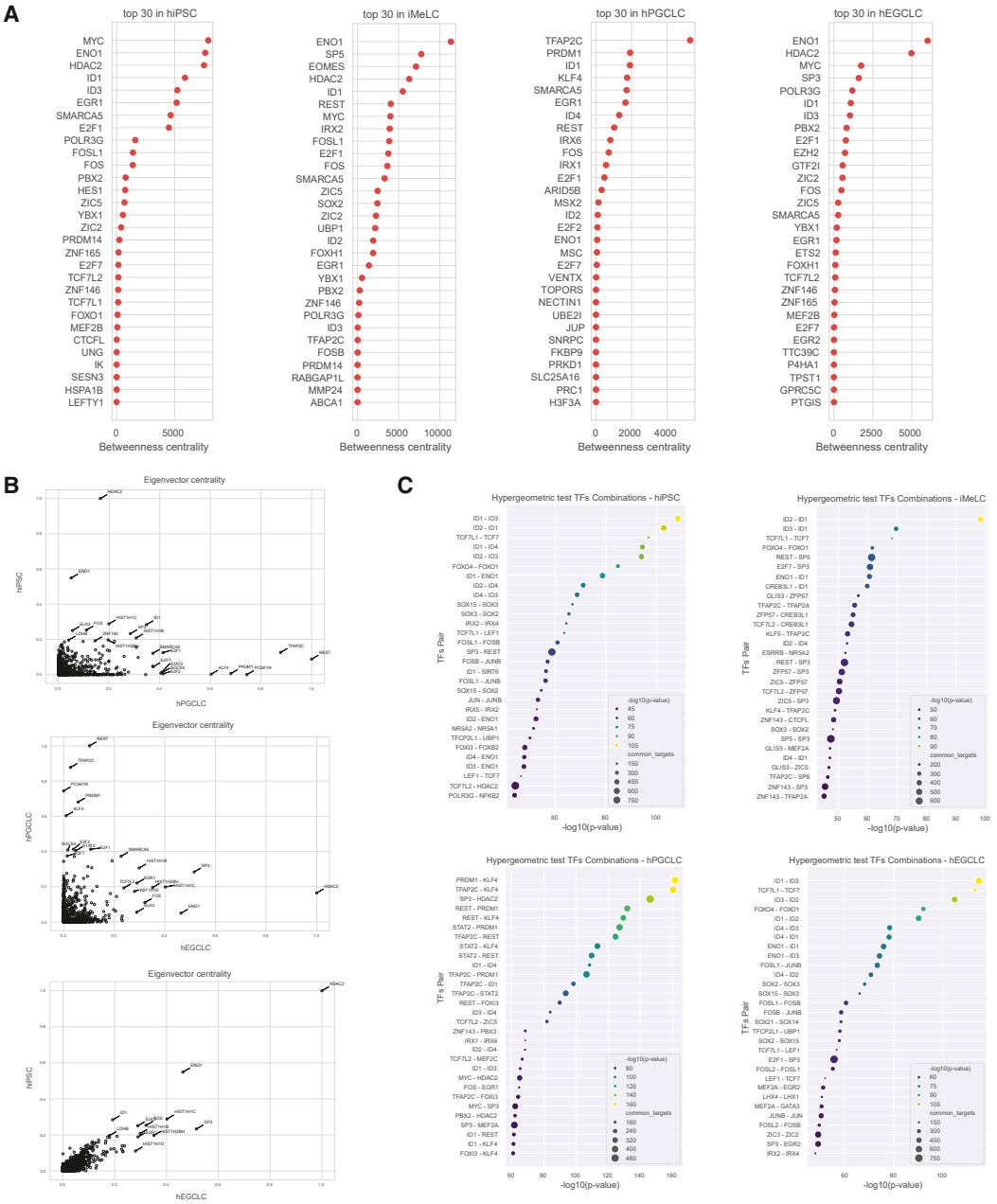

Figure 3. GRN analysis highlights key regulators of hEGCLC transcriptional program

(A) Dot plot of top-ranking transcription factors (TF) based on betweenness centrality (*x* axis) measured for each cell type (reported in each plot title).

(B) Scatterplot of eigenvector centrality measured for each TF in each cell type. Each plot represents the comparison of two cell types reported in the axis title.

(C) Dot plot of enrichments measured in each cell type (indicated in each plot title) for the overlap of targets for each couple of TF considered. The top 30 TF couples are reported based on the significance (−log10[*p* value]) of the hypergeometric test measured on the overlap between targets of the two TF reported on the *y* axis, with respect to the genes expressed in each cell type. Dot size is proportional to the size of the intersection. Color is proportional to significance.

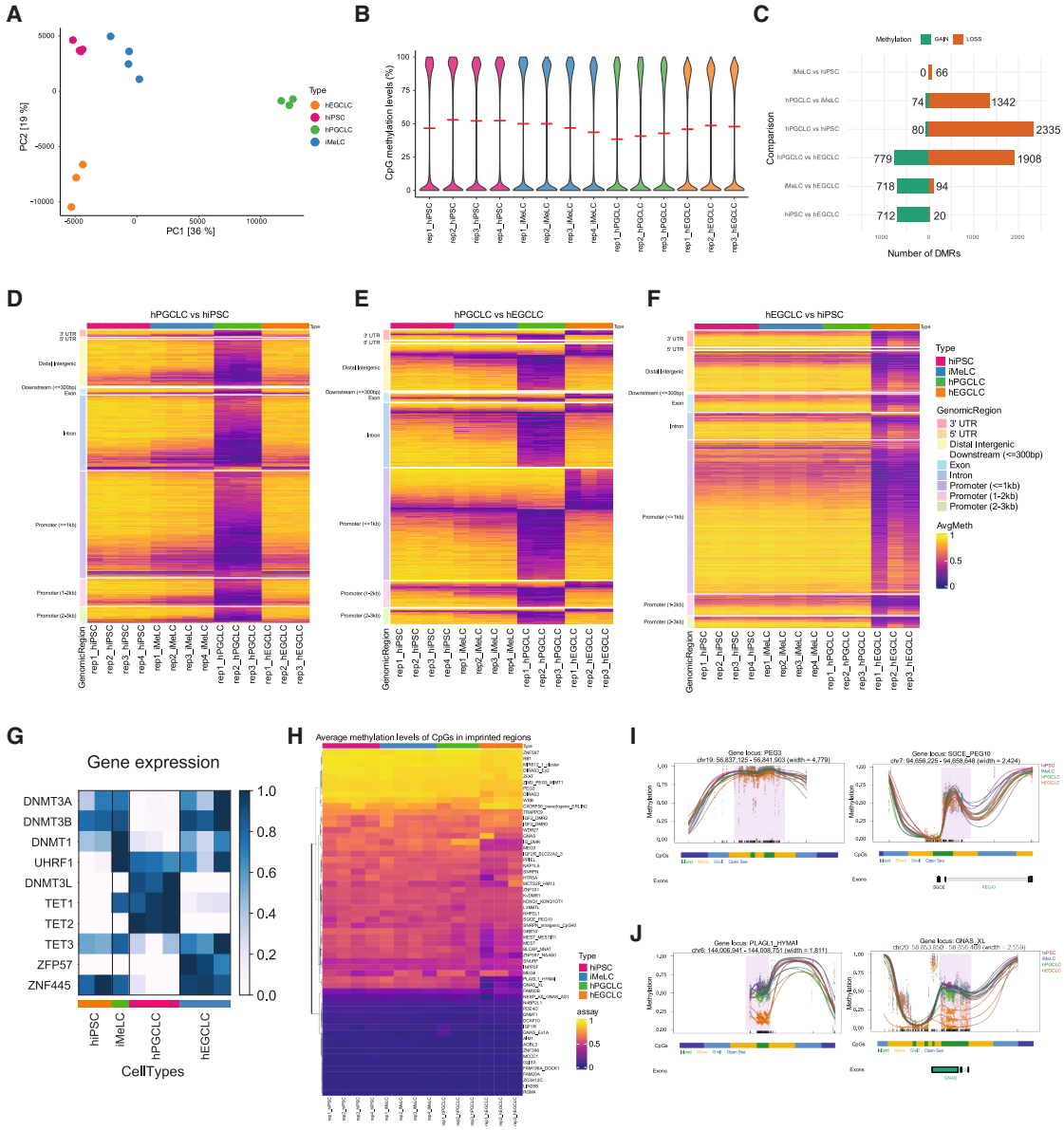

Figure 4. Longitudinal DNA methylation profiling to investigate the reversibility of epigenetic states

(A) PCA plot of CpG methylation levels of hiPSC, iMeLC, day 6 hPGCLC, and hEGCLC (hiPSC and iMeLC in quadruplicate; hPGCLC and hEGCLC in triplicate; CTL08A line). PC1 and PC2 explain 36% and 19% of the variability, respectively.

(B) Violin plots show the distribution of single CpG methylation levels (as percentage, *y* axis) at each measured region in each sample, *x* axis. Red horizontal bars indicate the median methylation level.

(C) Bar plots showing the number of differentially methylated regions (DMRs, *x* axis) for each comparison between the four cell types in pairs (*y* axis). Green/orange bars indicate the number of DMRs that gain/loss methylation in the first term of the comparison, respectively.

(D–F) Heatmaps of CpG average methylation level for each DMR (on the rows), for each of the samples analyzed (columns: hiPSC and iMeLC in quadruplicate, day 6 hPGCLC and hEGCLC in triplicate). (D) hPGCLC vs. hiPSC comparison; (E) hPGCLC vs. hEGCLC comparison; (F) hEGCLC vs. hiPSC comparison. Color gradient indicates the level of CpG methylation. DMRs are clustered on the rows according to the genomic region annotation depicted in the figure legend.

(G) Heatmap of pseudo-bulk expression level in hiPSC, iMeLC, hPGCLC, and hEGCLC of key genes involved in DNA demethylation and imprinting regulation.

(H) Clustered heatmap of average methylation levels of CpGs belonging to imprinted genes (rows) for each of the samples analyzed.

(I and J) DMR plots of four imprinted regions. Each dot represents the methylation level (*y* axis) of an individual CpG locus in a single sample. Size is proportional to the coverage. The lines represent smoothed methylation levels for each sample (hiPSC, iMeLC, hPGCLC, and hEGCLC). Gene exons and CpG annotations are shown below the plot if nearby.

hEGCLCs do not separate along the first principal component (PC1) but do so along the second one (PC2), which explains 19% of the variance (Figure 4A).

Violin plots showing the percentage of global genomic DNA (gDNA) methylation for each sample suggests a gradual decrease in median methylation level along differentiation (hiPSC → iMeLC → day 6 hPGCLC), with a variation in median methylation levels between hiPSCs and hPGCLCs of about 20% (Figure 4B; Table S5), consistent with previous reports (Kobayashi et al., 2022; Sasaki et al., 2015). However, median gDNA methylation recovers in hEGCLCs, reaching levels just slightly below those of hiPSCs, indicating that the global gDNA demethylation in hPGCLCs is largely reversible (Figure 4B; Table S5).

Next, we focused on the regions that in pairwise comparisons between the cell types show significant DNA methylation differences (i.e., greater than 25%), and designated these as differentially methylated regions (DMRs, see supplemental methods). Bar plots showing the total number of DMRs for each comparison confirm that hPGCLCs present lower methylation relative to all other cell types (Figure 4C). This is also evident by plotting CpG mean methylation levels for each DMR in heatmaps (Figures 4D–4F) and violin plots (Figures S4A–S4C; Table S5). Considering the DMRs identified in the hiPSC-hPGCLC comparison (Figure 4D), hEGCLCs have CpG methylation levels similar to hiPSCs. The examples of regions showing a significant loss of methylation in hPGCLCs compared to hiPSCs and hEGCLCs are depicted in Figure S4D. Functional enrichment analysis performed on genes associated with these DMRs (hPGCLC-hiPSC and hPGCLC-hEGCLC) highlights the terms mainly related to the regulation of cell shape, histone deacetylation, and regulation of Hippo signaling (Figure S4E; Table S5).

To quantify the levels of both 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC), we used liquid chromatography-tandem mass spectrometry (LC-MS/MS, see supplemental methods). This analysis confirmed a decrease in 5mC levels in hPGCLCs compared to hiPSCs and hEGCLCs (Figure S4F top) and highlighted an increase in 5hmC levels in hPGCLCs compared to the pluripotent cell types (Figure S4F bottom). This is in line with the initiation of methylome resetting in nascent hPGCLCs via oxidation of 5mC to 5hmC by TET enzymes at certain loci (Iurlaro et al., 2017; Tang et al., 2015; Zeng and Chen, 2019).

We also examined the expression of genes involved in the regulation of DNA methylation in our scRNA-seq dataset. The expression of DNA methyltransferases (*DNMT1* and *DNMT3A/B*) decreases in hPGCLCs compared to hiPSCs but recovers in hEGCLCs. In contrast, the expression of *TET1* and *TET2* demethylases and of *DNMT3L* is low in hiPSCs and hEGCLCs and increases in hPGCLCs (Figure 4G). However, we were surprised to find that the expression of the DNMT1 cofactor *UHRF1* increases as hiPSCs differentiate to hPGCLCs but does not decrease in hEGCLCs (Figure 4G).

Finally, we explored specific DMRs in hEGCLCs compared to hiPSCs, of which we identified 712 hypomethylated and 20 hypermethylated regions (Figures 4C and 4F). Functional enrichment analysis performed on hypomethylated regions highlighted the terms related to cell adhesion and regulation of small GTPase-mediated signal transduction (Figure S4G). Intriguingly, the promoter of *PIWIL2* is among these hypomethylated regions, in keeping with its increased expression in hEGCLCs compared with hiPSCs (Figure S4H). These hypomethylated regions emerge as PSCs passage through a germ cell state and therefore may reflect an "epigenetic memory" of this cell fate transition. To investigate how stable these epigenetic differences are, we analyzed DNA methylation in later passage hEGCLCs (P10) and found almost complete reversion to the pattern observed in hiPSCs (Figures S5A and S5B). Indeed, hEGCLCs_P10 show only 99 hypomethylated and 26 hypermethylated regions compared to hiPSCs (Figure S5C). Some of these DMRs remain demethylated to the same extent as in hEGCLCs_P5 (such as *HTRA4* promoter region, Figure S5D), while other regions show a partial remethylation (such as *ZNF490* promoter region, Figure S5E). It is possible that with further passaging, DNA methylation may completely recover at these regions. Indeed, the majority of DMRs at P5 are fully methylated by P10, including the *PIWIL2* promoter region (Figure S5F).

We next focused on regions subjected to genomic imprinting. Loss of DNA methylation at imprinting control regions has been observed in many mouse EGC lines (Labosky et al., 1994; Shovlin et al., 2008; Tada et al., 1998), although EGC lines derived from early PGCs can emerge with imprints intact (Leitch et al., 2013b). Global methylation levels at the different human imprinted regions are similar in hiPSCs, iMeLCs, hPGCLCs, and hEGCLCs (Figure 4H; Table S6). This pattern is exemplified by *PEG3* (Li et al., 2016; Yeung et al., 2018) and *PEG10-SGCE* (Ono et al., 2001; Peall et al., 2013), the two imprinted regions with the highest coverage in terms of the number of CpGs detected in our dataset (Figures 4I and S5G). As such, hPGCLCs largely recapitulate early hPGC development (around post-conceptional weeks 2–3), when imprinted regions have not yet undergone DNA demethylation (Gell et al., 2020; Hackett et al., 2013b; Hargan-Calvopina et al., 2016; Hill et al., 2018; Vincent et al., 2013; Yamashiro et al., 2020). Interestingly, however, we noted that hPGCLCs downregulate *ZFP57* and *ZNF445* (Figure 4G), two regulators critical for imprinting maintenance (Takahashi et al., 2019), and that selected imprinted regions including *PLAG1-HYMAI* (Arima et al., 2006), *GNAS* (Kalish et al., 2014; Kelsey, 2010; Turan and

Bastepe, 2013) (Figures 4J and S5H), and *MEST* (Kobayashi et al., 1997), exhibited lower levels of methylation in hEGCLCs (both at P5 and P10) when compared with hiPSCs and hPGCLCs. These imprinted regions overlap with those that have recently been reported to be demethylated during the early stages of hPGCLC culture (Murase et al., 2024). It is therefore possible that these loci are more susceptible to DNA demethylation or that factors used to culture hPGCLCs *in vitro* play a role in triggering their demethylation. We also note that imprint instability in culture is a well-described phenomenon during PSC culture (Bar et al., 2017; Humpherys et al., 2001). The extent to which loss of DNA methylation at imprinted regions during *in vitro* culture of PGCs/PGCLCs accurately reflects the *in vivo* process of imprint erasure and whether this can be manipulated by altering the culture environment are important areas for future study.

Overall, these data confirm that the initial demethylation associated with hPGCLC induction is largely reversible, but subtle epigenetic differences remain, including reduced DNA methylation at a subset of imprinted loci.

Multi-omics GRN analysis identifies key regulators of the pluripotency-germline transitions

Leveraging the multi-omics nature of our longitudinal dataset, we integrated DNA methylation and transcriptomic data to further dissect the master regulators governing the pluripotency-germline transitions. Focusing on genes exhibiting an inverse correlation between methylation levels at their regulatory regions and their expression, we specifically examined those that were upregulated or downregulated in hPGCLCs compared to PSCs (both hiPSC and hEGCLC) (Figures 2D and 5A).

Master regulator analysis (MRA) on these differentially methylated and expressed genes (dMdE, see supplemental methods) identified 136 key TFs (Figures 5B and S6B). Interestingly, while only 7 of these TFs emerged as key regulators across all 4 cell types, we observed that subsets of the TFs exhibited cell-type specificity, with 52 TFs, including members of the *GATA, SOX* (Kojima et al., 2021), and *ZNF* families (Tang et al., 2015), specific for hPGCLCs, 10 TFs specific for hiPSC (*IRX3, NFATC4, ATOH8, PRDM11, PKNOX2, NR5A2, TBX3, MEF2B, GBX2,* and *HDAC2*), and 4 specific for hEGCLC (*SOX18, ZNF394, RXRG,* and *MYC*). This observation confirms our previous identification—from the transcriptomic-specific analysis—of *MYC* and *HDAC2* as top regulators of pluripotency networks. However, despite the similarities between hiPSC and hEGCLC GRNs, key regulatory differences exist, particularly in genes that display dynamic expression and methylation profiles during the reprogramming process (Figure S6A). Furthermore, our analysis revealed that among the 136 TFs regulating dMdE genes, 11 were themselves dMdE (Figures 5C, 5D, and S6C) and 17 were methylation-sensitive TFs, namely TFs for which the methylation state of their recognition sequence can directly modulate their binding affinity (Grand et al., 2024; Isbel et al., 2022; Kaluscha et al., 2022; Yin et al., 2017) (Figure S6D; Table S7).

These findings highlight the intricate interplay between DNA methylation and transcriptional regulation during germline differentiation and reprogramming, and underscore the importance of considering both epigenetic and transcriptomic dynamics to study these processes.

Escapee regions are hypermethylated compared to non-escapee regions and resistant to reprogramming

Previous studies reported "escapee" regions that exhibit resistance to DNA demethylation in both mouse (Hackett et al., 2013a) and human PGCs (Tang et al., 2015) and proposed these as candidates that might mediate epigenetic inheritance (Tang et al., 2015). Our protocol presents a unique opportunity to study how these epialleles behave during germline induction and subsequent reprogramming to hEGCLCs. Among the 3,680,616 CpG loci analyzed, 9,113 fall within the escapee regions identified by Tang et al. (2015) (Table S6). A PCA plot of CpG methylation levels at the escapee regions shows how hiPSCs and hPGCLCs separate along the first PC (explaining 20% of the variance); however, conversion to hEGCLCs is not accompanied by a reversion in methylation at early passages, as hEGCLCs_P5 cluster together with hPGCLCs along PC1. hPGCLCs and hEGCLCs separate only along PC2 (that explains 16% of the variance) (Figure 6A). This contrasts markedly with the PCA plot of non-escapee regions (Figure S7A). The non-escapee pattern is almost identical to the combined PCA plot (Figure 4A), as expected, since the majority of the CpG loci belong to non-escapee regions.

We then compared the distribution of CpG mean methylation levels in escapees vs. non-escapee regions (Figure 6B), noting how non-escapee regions follow a bimodal distribution (Figures 6B and S7B), as expected, whereas escapee regions follow a unimodal distribution, shifted toward higher methylation levels (Figures 6B and 6C). A recent paper from Saitou's laboratory also identified escapee regions in their hPGCLC expansion culture (Murase et al., 2024). We therefore checked the CpG methylation level of the list of escapee regions from (Murase et al., 2024; 192,309 escapee regions, 207,478 CpGs), confirming a similar unimodal distribution (Figure S7D). We also explored the behavior of escapee regions from Tang et al. (2015) in a whole genome bisulfite sequencing (WGBS) dataset from Hsu et al. (2023), confirming both in hESC and day 4 hPGCLC the unimodal distribution skewed toward high methylation levels

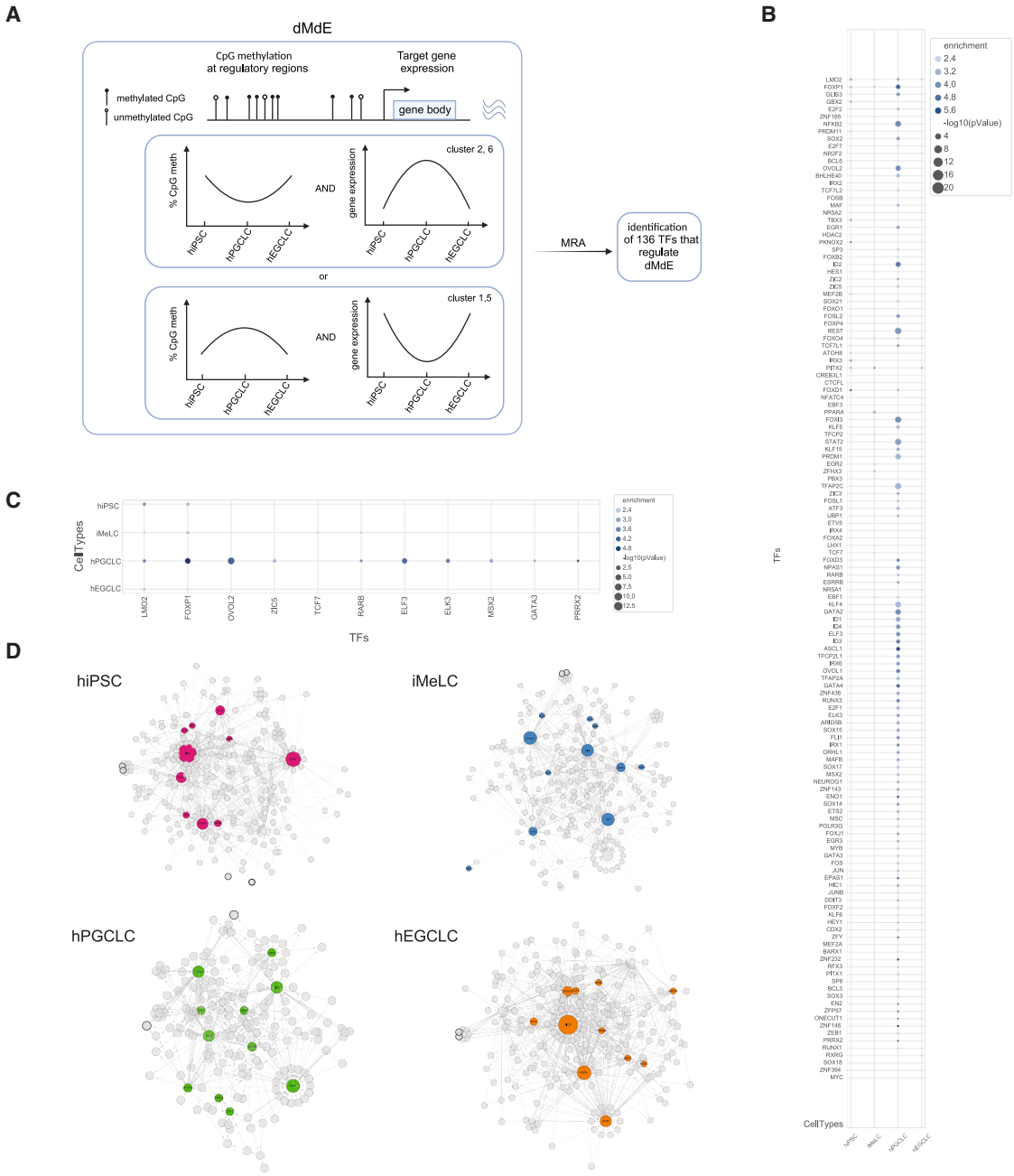

Figure 5. Multi-omics GRN analysis identifies key regulators of pluripotent state transitions

(A) Schematics of the selection strategy for differentially methylated and expressed (dMdE) genes by integration of DEGs and DMRs. MRA was then performed to identify TFs responsible for the regulation of those genes.

(B) Dot plot of TFs (*y* axis) significantly regulating dMdE genes in each cell type (*x* axis). Dot color represents enrichment (>2) and dot size represents −log10(*p* value) > 1.3 (*p* value < 0.05).

(C) Dot plot of dMdE TFs (*x* axis) significantly regulating dMdE genes in each cell type (*y* axis). Dot color represents enrichment (>2) and dot size represents −log10(*p* value), (*p* value < 0.05).

(D) GRNs visualization (one for each of the four cell types), highlighting the 11 TFs (colored) that are differentially expressed and differentially methylated and that regulate dMdE genes. Their targets and regulators are in gray.

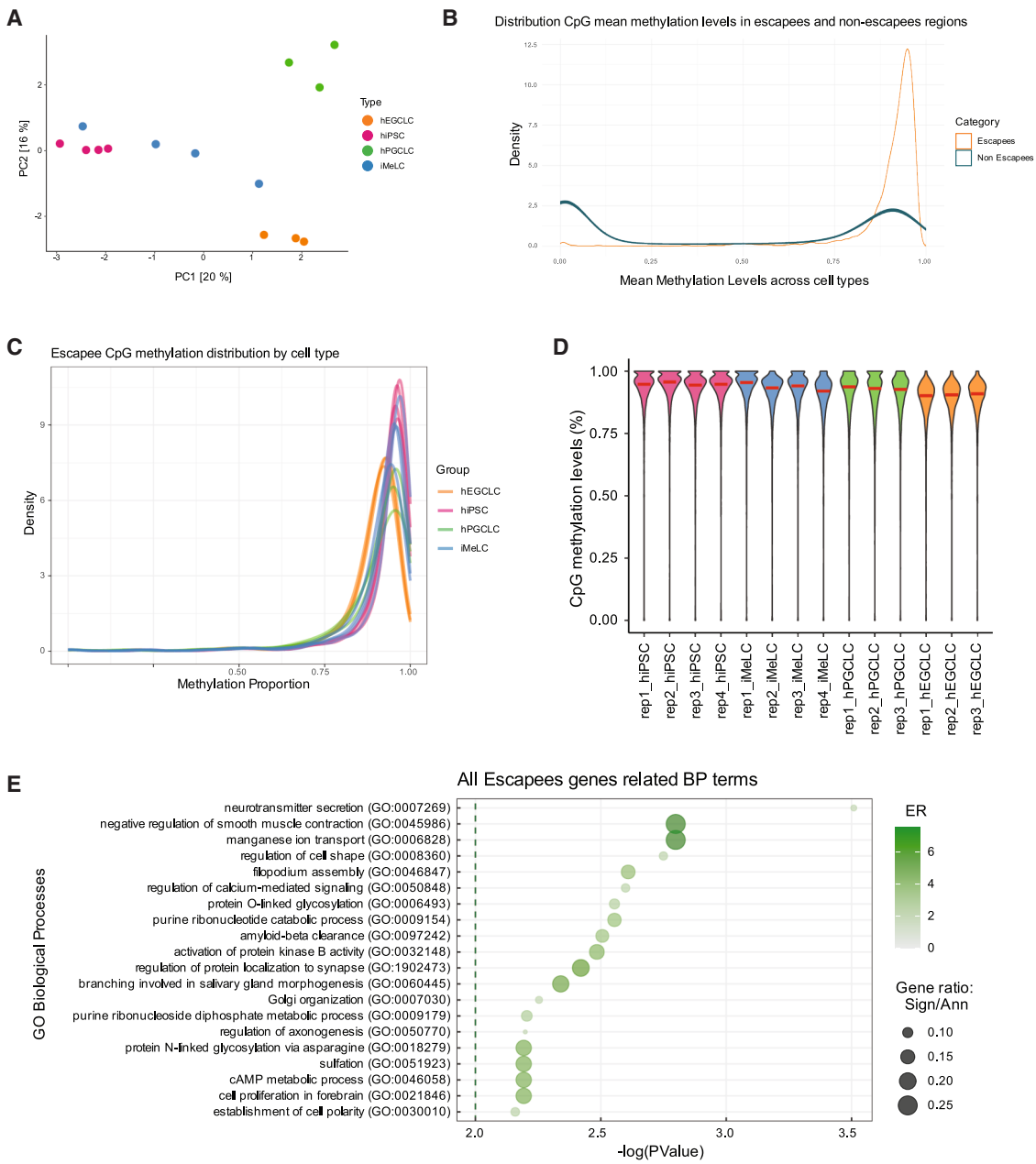

Figure 6. Context-dependent regulation of DNA methylation at escapee regions

(A) PCA plot of CpG methylation levels of escapee regions of hiPSC, iMeLC, day 6 hPGCLC, and hEGCLC, CTL08A line. PC1 and PC2 explain 20% and 16% of the variability, respectively.

(B) Density plot of methylation level distributions of the 9,113 CpGs within escapee regions and of 500 sets of 9,113 randomly selected CpGs within non-escapee regions (cell types collapsed by average).

(C) Density plot of the methylation level distribution of the 9,113 CpGs within escapee regions, colored by cell type.

(D) Violin plots of CpG methylation levels of escapee regions (percentages, *y* axis) for each sample (*x* axis). Red horizontal bars indicate the median methylation level.

(E) Dot plot of enriched GO terms (BP category) for the genes annotated to the escapee regions. *y* axis: GO term; *x* axis: −log(*p* value). Color gradient indicates ER > 2, and dot dimension indicates gene ratio. The dotted vertical line indicates the *p* value threshold at 0.01.

(Figures S7E and S7F). These data are consistent with the tendency of these regions to be hypermethylated and to escape DNA demethylation (Gao et al., 2023; Tang et al., 2015).

However, when looking at the escapee regions CpG methylation distribution by cell type (Figure 6C), we noticed that methylation levels are lower in hPGCLCs compared to hiPSCs and even lower in hEGCLCs_P5 (Figures 6D and S7C) (noticeable despite the unimodal distribution skewed toward increased methylation of these regions in all cell types). This result points to a peculiar behavior of these regions during the conversion of hPGCLCs to hEGCLCs. While the majority of genomic loci re-accumulate DNA methylation to levels similar to hiPSCs by P5, the escapee regions remain relatively demethylated. It is only upon extended culture to P10 that DNA methylation levels fully recover (Figures S7G and S7H). Thus, these regions are not simply characterized by an inherent propensity to maintain high levels of DNA methylation. Rather, they exhibit a context-dependent response to different forms of epigenetic resetting—remaining relatively hypermethylated during epigenetic reprogramming in late PGCs, while remaining relatively hypomethylated during the generalized reacquisition of DNA methylation during hEGCLC derivation.

Functional enrichment analysis of the genes associated with escapee regions included in our analysis shows an enrichment for terms related to neurodevelopment (Figure 6E), in keeping with previous findings (Tang et al., 2015). As this original study specifically raised the possibility of intergenerational epigenetic inheritance at these regions, our finding that escapee regions might be specifically protected from initial remethylation during hPGCLC-to-hEGCLC conversion may suggest context-dependent regulation of DNA methylation at these loci and establish our culture system as a new paradigm to investigate the targeted alteration and inheritance of clinically relevant epialleles.

DISCUSSION

Here, we report the first fully defined system to track the pluripotent state transitions between hPSCs and hPGCLCs. We demonstrate that hPGCLCs can efficiently convert back to the PSC state from which they are derived, both at the functional and molecular level. Our multi-omics tracking of these transitions provides a high-resolution map of the transcriptional and epigenomic transitions upon entry to and exit from the human germline.

Since the first reports of EGC derivation in mice, there has been significant progress in identifying the factors that either promote PGC proliferation and/or their conversion to EGCs (Leitch et al., 2013a). These advances led to the development of a fully defined system to culture mouse PGCs and track their conversion to EGCs (Leitch et al., 2013c). Despite these advances in the mouse, *bona fide* human EGCs derived from hPGCs have not yet been reported. While recent studies have highlighted the differences between mouse and human PGC development, there are nevertheless many similarities between the two, both *in vivo* and *in vitro* (Kobayashi and Surani, 2018). Importantly, the same factors used to trigger mPGCLC induction are also effective in hPGCLCs (Saitou and Hayashi, 2021). Our findings establish that the factors that support the derivation of mEGCs also trigger the efficient transition of hPGCLCs to hEGCLCs, suggesting that the mechanisms of conversion might be evolutionarily conserved between mammals, an exciting avenue for future investigation. In particular, future experiments should address whether the essential role of LIF-STAT3 signaling in mice is also conserved in humans.

There has been a previous report of hEGCLC derivation (Kobayashi et al., 2022), in which the authors initiate hPGCLC cultures on a feeder cell layer, transition them to Matrigel in a feeder-cell conditioned medium for long-term expansion, and then, by subsequent removal of conditioned medium and culture with FGF2 and SCF, lead to the formation of PSC colonies that can be picked and expanded as hEGCLC lines. Intriguingly, a recent report also noted that dedifferentiation can occur in long-term hPGCLC cultures maintained on feeder cells, albeit using a different culture medium (Murase et al., 2024). Thus, it seems possible that feeder cells produce a range of factors that can either promote or inhibit the emergence of hEGCLCs in different contexts. Rather than relying on their spontaneous emergence from long-term hPGCLC cultures, we have established a dedicated hEGCLC derivation protocol. Notably, our hEGCLC derivation medium does not support long-term expansion of hPGCLCs, in keeping with the idea that unknown factors produced by feeder cells and present in conditioned medium are key to this long-term expansion. The system reported here may be one route to screen for such factors as, unlike other systems, cultures are initiated and maintained without feeders or feeder cell-conditioned medium. This will also facilitate future studies to identify the essential factors and pathways that enable hPGCLCs to transition to a PSC state. For instance, performing a time course of single-cell assays commencing at early time points will allow further characterization of this transition, including a direct comparison with hPGCLCs that do not convert to hEGCLCs. Future work should also address whether hPGCLCs can also give rise directly to naïve PSCs, helping to deepen our understanding of pluripotent states in humans. In addition, the recent report that hPGCLCs can be differentiated toward

pro-spermatogonia and oogonia (Murase et al., 2024) will enable experiments to address whether there is a specific developmental time window during which hEGCLC derivation is possible.

Our defined culture system facilitated multi-omics longitudinal profiling, which in turn allowed us to capture germline transcriptional program emergence and dissolution. While hiPSCs and hEGCLCs exhibit substantial transcriptomic similarity, we identified a small subset of genes that are upregulated in hPGCLCs but do not return to their previous hiPSC levels in hEGCLCs. In keeping with previous findings (Kobayashi et al., 2022), this included the germ-cell-specific gene *PIWIL2*. Our data demonstrates that this transcriptional memory of transition through a germline state coincides with failure to re-establish DNA methylation at *PIWIL2*'s promoter at early passages, suggesting that the gene expression differences may be driven by an initial incomplete epigenetic resetting. Upon extended passaging, DNA methylation levels recovered at *PIWIL2* and most other loci. Intriguingly, we observed a similar phenomenon at previously reported escapee regions, suggesting that these regions are not simply resistant to DNA demethylation in PGCs but may also exhibit more complex, context-dependent epigenetic regulation than previously appreciated. More broadly, our culture system may enable future studies to investigate the inheritance and targeted alteration of clinically relevant epialleles during transition through a germline state.

Our detailed multi-omics analysis of changes in gene expression and epigenetic state enabled us to describe the gene regulatory logic of the hPSC-to-hPGCLC-to-hEGCLC transitions and identify key transcription factors that control these changes in cell state. This knowledge may have significant applications in both stem cell and cancer fields. For instance, our study represents an important step toward the derivation of hEGC from *in vivo* hPGC. By using our defined protocol, hEGC derivation can now be approached in an iterative and targeted manner, such as by adjusting media components to more specifically regulate TFs and pathways that drive the hPGCLC to hEGCLC transition. Access to early-stage hPGCs is likely to remain a major challenge, although such samples do rarely become available to researchers (Cui et al., 2025; Tyser et al., 2021). The derivation of hEGC would be a major breakthrough in the field, establishing that the ability to give rise to the full repertoire of embryo-derived PSCs is not a rodent-specific phenomenon. This in turn will help us better understand the different pluripotent states in humans, including how well they map to the naïve, formative, primed, and latent states that are increasingly well-defined in mice (Hackett and Surani, 2014; Leitch and Smith, 2013; Nichols and Smith, 2009; Sepulveda-Rincon et al., 2024; Smith, 2017).

Comprehending how hPGCs transition to a PSC state is also of clear relevance to our understanding of the pathogenesis of human germ cell tumors (GCTs), many of which are derived from hPGCs and have pluripotent and stem cell characteristics. In particular, sacrococcygeal teratomas represent the commonest neonatal tumors in humans (Geurten et al., 2020), GCTs are among the most common childhood cancers (Public Health England, 2021). Variants in *SPRY4, GAB2*, and *KITLG* are associated with an increased risk of GCTs (Marcotte et al., 2017), and it is noteworthy that each of the associated signaling pathways are targeted by factors in our hEGCLC derivation medium. A role for environmental factors, including endocrine-disrupting chemicals, has also been proposed (McGlynn and Cook, 2009). Therefore, our culture system represents a new method to screen potential genetic, signaling, and chemical triggers of GCT formation in humans. In addition, the TFs we have identified as drivers of hPGCLC-hEGCLC transition warrant future investigation for their possible role in GCT pathogenesis.

In conclusion, this reference multi-omics longitudinal profiling will be a valuable resource in several domains, from deepening the understanding of pluripotent states in humans to the studying of the genetic and environmental factors involved in GCT pathogenesis and to the investigation of the epigenetic dynamics during the germline cycle.

METHODS

Human PSC lines culture

Healthy control-derived hPSC lines from different genetic backgrounds and sexes were used in this study. Culture conditions and specific line usage are detailed in the supplemental methods.

Induction of iMeLCs and hPGCLCs and fluorescence-activated cell sorting

hPGCLCs were induced from hPSCs via iMeLCs as described in (Sasaki et al., 2015; Yamashiro et al., 2020), with minor modifications detailed in the supplemental methods.

hEGCLCs' derivation from sorted hPGCLCs

To reprogram hPGCLCs into hEGCLCs, sorted day 4 or day 6 hPGCLCs were plated directly into 96 multi-well flat bottom plates (10 cells/well), pre-coated with Geltrex (Thermo Scientific, #A1413201), according to the manufacturer's instructions, and containing 100 μL/well of hEGCLC induction medium (i.e., N2B27 medium [Cold Spring Harbor Laboratory, 2017] containing 10 ng/mL hLIF (PeproTech, #300-05), 3 μM CH (Sigma-Aldrich,

#SML1046-5MG), 10 μM FK (Sigma-Aldrich, #F3917-25MG), 25 ng/mL bFGF (Preprotech, #100-18B), 100 ng/mL hSCF (PeproTech, #300-07), 2 μM RA (Sigma-Aldrich, #R2625), and antibiotics (penicillin 100 U/mL and streptomycin 100 μg/mL, Thermo Fisher, #15140122). They were grown at 37°C, 5% CO_2, and 5% O_2. After 48 h, 100 μL of StemFit (Amsbio, #SFB-504-CT) was added to each well. On day 3, 100 μL of the medium was removed, and 100 μL of StemFit was added. On day 4, 150 μL of the medium was removed, and 100 μL of StemFit was added. On day 5, all medium was removed, and 100 μL of StemFit was added. From day 6 onwards, total media change was performed every other day. In these defined and feeder-free culture conditions, 10%–15% of hPGCLCs undergo conversion to form hEGCLC colonies over a period of two weeks (conversion efficiency was calculated as the total number of hEGCLC colonies obtained divided by the total number of hPGCLCs seeded (i.e., 10 × total number of wells seeded, since we plated 10 hPGCLCs/well). These hEGCLC colonies were then picked, expanded, and maintained in conventional human PSC conditions. hEGCLC pluripotency was validated by immunofluorescence analysis and three germ layers differentiation, as described in the supplemental methods.

RNA extraction, bulk RNA-seq library preparation, and data analysis

Frozen pellets were submitted to GENEWIZ for RNA isolation, library preparation with poly(A) selection, and 150-bp paired-end sequencing on Illumina. The analytical pipeline is described in the supplemental methods.

Single-cell dissociation, scRNA-seq library preparation, and data analysis

hiPSCs, iMeLCs, hEGCLCs, and day 6 hPGCLC aggregates dissociation into single-cell and library preparation are detailed in the supplemental methods. scRNA-seq data were aligned against the probe reference file v1.0.0 from CellRanger. 27,925 high-quality cells were obtained after standard quality checks and filters. The detailed analytical pipeline is described in the supplemental methods.

DNA extraction, EM-seq library preparation, and data analysis

gDNA was extracted using DNeasy Blood & Tissue Kit (Qiagen, #69504). Library preparation and the detailed analytical pipeline are described in the supplemental methods.

RESOURCE AVAILABILITY

Lead contact

Requests for further information should be directed to the lead contact, Harry G. Leitch (harry.leitch@ucl.ac.uk).

Materials availability

hEGCLC clones generated in this study are available from the lead contact with a completed materials transfer agreement.

Data and code availability

- Data generated in this study are available at the following links and accession numbers: low-pass whole-genome sequencing: GEO: GSE288719 (for BTAG line); BioSudies: E-MTAB-15528 for CTL08A line); bulk RNA-seq: GEO: GSE286998; scRNA-seq: BioStudies: E-MTAB-15244; EM-seq: BioStudies: E-MTAB-15142.
- Full code used for the analyses can be retrieved at https://github.com/GiuseppeTestaLab/EGCLC_paper_release.

ACKNOWLEDGMENTS

This project was undertaken in G.T.'s laboratory at Human Technopole, Milan, and in H.G.L's laboratory at MRC Laboratory of Medical Sciences (LMS) and UCL Great Ormond Street Institute of Child Health, London. Work in H.G.L's laboratory was supported by MRC core funding (MC-A652-5QA10) and NIHR Imperial BRC. This work is also supported by the NIHR GOSH BRC. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. G.T. has been funded by the European Union Horizon 2020 research and innovation program grants ENDpoiNTs (825759), NEUROCOV (101057775), RE-MEND (101057604), R2D2-MH (101057385), and PNRR (Centro Nazionale CN3: RNA, "National Center for Gene Therapy and Drugs based on RNA Technology"). We thank the Flow Cytometry (FLARE) and Genomics Facility (GF) of Human Technopole and the LMS/NIHR Imperial Biomedical Research Center Flow Cytometry Facility for support in the experiments and the useful discussions. We thank Matteo Bonfanti (Software Development and Data Curation, SDDC) of Human Technopole for the computational support. We thank the European School of Molecular Medicine (SEMM), at which S.S., G.M., A. Valenti, and M.T.R. are/were enrolled as students for their PhD degree program in Systems Medicine. The HPC Platform of Human Technopole supported computation. We also thank Mitinori Saitou and Timo Otonkoski for sharing their reporter lines. Some components of the schematics and graphical abstract were created in BioRender (https://BioRender.com/).

AUTHOR CONTRIBUTIONS

S.S., L.P.S.-R., C.D., N.C., G.T., and H.G.L. conceived the project, designed the experiments, and critically interpreted the results. G.M. and A. Valenti contributed with critical discussions and interpretation of the results. L.P.S.-R. and C.D. developed the protocol under the supervision of H.G.L. and performed live imaging and bulk RNA-seq experiments. S.S. replicated the protocol in G.T.'s laboratory and performed immunostaining, tri-germ layers, scRNA-seq, and EM-seq experiments, with the initial support of M.T.R. and M.C. in setting the protocol for hPGCLC differentiation and with the support of R.N. for the revision experiments. G.M. performed EM-seq and low-pass whole-genome sequence data analysis, under the supervision of C.C. A. Valenti performed scRNA-seq data analysis under the supervision of C.C. and N.C.

F.P. performed GRN analysis under the supervision of A. Vitriolo. G.Y. performed low-pass whole-genome sequencing and bulk RNA-seq data analysis. S.S. and B.M. curated the list of methylation-sensitive TFs. Z.H. performed the LC-MS/MS experiment under P.H.'s supervision. S.S., N.C., G.T., and H.G.L wrote the manuscript with input from all authors. All authors read and approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.stemcr.2025.102746.

Received: February 27, 2025
Revised: November 14, 2025
Accepted: November 18, 2025
Published: December 18, 2025

REFERENCES

Alberio, R., Kobayashi, T., and Surani, M.A. (2021). Conserved features of non-primate bilaminar disc embryos and the germline. Stem Cell Rep. *16*, 1078–1092.

Arima, T., Yamasaki, K., John, R.M., Kato, K., Sakumi, K., Nakabeppu, Y., Wake, N., and Kono, T. (2006). The human HYMAI/PLAGL1 differentially methylated region acts as an imprint control region in mice. Genomics *88*, 650–658.

Badia-I-Mompel, P., Vélez Santiago, J., Braunger, J., Geiss, C., Dimitrov, D., Müller-Dott, S., Taus, P., Dugourd, A., Holland, C.H., Ramirez Flores, R.O., and Saez-Rodriguez, J. (2022). decoupleR: ensemble of computational methods to infer biological activities from omics data. Bioinform. Adv. *2*, vbac016.

Balboa, D., Weltner, J., Novik, Y., Eurola, S., Wartiovaara, K., and Otonkoski, T. (2017). Generation of a SOX2 reporter human induced pluripotent stem cell line using CRISPR/SaCas9. Stem Cell Res. *22*, 16–19.

Bar, S., Schachter, M., Eldar-Geva, T., and Benvenisty, N. (2017). Large-Scale Analysis of Loss of Imprinting in Human Pluripotent Stem Cells. Cell Rep. *19*, 957–968.

Blair, K., Leitch, H.G., Mansfield, W., Dumeau, C.É., Humphreys, P., and Smith, A.G. (2012). Culture parameters for stable expansion, genetic modification and germline transmission of rat pluripotent stem cells. Biol. Open *1*, 58–65.

Bleckwehl, T., Crispatzu, G., Schaaf, K., Respuela, P., Bartusel, M., Benson, L., Clark, S.J., Dorighi, K.M., Barral, A., Laugsch, M., et al. (2021). Enhancer-associated H3K4 methylation safeguards in vitro germline competence. Nat. Commun. *12*, 5771.

Borkowska, M., and Leitch, H.G. (2021). Mouse primordial germ cells: in vitro culture and conversion to pluripotent stem cell lines. Methods Mol. Biol. *2214*, 59–73.

Buehr, M., Meek, S., Blair, K., Yang, J., Ure, J., Silva, J., McLay, R., Hall, J., Ying, Q.-L., and Smith, A. (2008). Capture of authentic embryonic stem cells from rat blastocysts. Cell *135*, 1287–1298.

Campolo, F., Gori, M., Favaro, R., Nicolis, S., Pellegrini, M., Botti, F., Rossi, P., Jannini, E.A., and Dolci, S. (2013). Essential role of Sox2 for the establishment and maintenance of the germ cell line. Stem Cell. *31*, 1408–1421.

Castillo-Venzor, A., Penfold, C.A., Morgan, M.D., Tang, W.W., Kobayashi, T., Wong, F.C., Bergmann, S., Slatery, E., Boroviak, T.E., Marioni, J.C., and Surani, M.A. (2023). Origin and segregation of the human germline. Life Sci. Alliance *6*, e202201706.

Cheng, K., Seita, Y., Whelan, E.C., Yokomizo, R., Hwang, Y.S., Rotolo, A., Krantz, I.D., Kaur, M., Ginsberg, J.P., Lal, P., et al. (2024). Defining the cellular origin of seminoma by transcriptional and epigenetic mapping to the normal human germline. Preprint at bioRxiv. https://doi.org/10.1101/2024.04.20.590406.

Chen, D., Liu, W., Zimmerman, J., Pastor, W.A., Kim, R., Hosohama, L., Ho, J., Aslanyan, M., Gell, J.J., Jacobsen, S.E., and Clark, A.T. (2018). The TFAP2C-Regulated OCT4 Naive Enhancer Is Involved in Human Germline Formation. Cell Rep. *25*, 3591–3602.e5.

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

Chialastri, A., Karzbrun, E., Khankhel, A.H., Radeke, M.J., Streichan, S.J., and Dey, S.S. (2022). Integrated single-cell sequencing reveals principles of epigenetic regulation of human gastrulation and germ cell development in a 3D organoid model. Preprint at bioRxiv. https://doi.org/10.1101/2022.02.10.479957.

Chitiashvili, T., Dror, I., Kim, R., Hsu, F.-M., Chaudhari, R., Pandolfi, E., Chen, D., Liebscher, S., Schenke-Layland, K., Plath, K., and Clark, A. (2020). Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation. Nat. Cell Biol. *22*, 1436–1446.

Choi, K.-D., Vodyanik, M.A., Togarrati, P.P., Suknuntha, K., Kumar, A., Samarjeet, F., Probasco, M.D., Tian, S., Stewart, R., Thomson, J.A., and Slukvin, I.I. (2012). Identification of the hemogenic endothelial progenitor and its direct precursor in human pluripotent stem cell differentiation cultures. Cell Rep. *2*, 553–567.

Cold Spring Harbor Laboratory. (2017). N2B27 Medium. Cold Spring Harb. Protoc. *2017*, pdb.rec096131.

Cui, L., Lin, S., Yang, X., Xie, X., Wang, X., He, N., Yang, J., Zhang, X., Lu, X., Yan, X., et al. (2025). Spatial transcriptomic characterization of a Carnegie stage 7 human embryo. Nat. Cell Biol. *27*, 360–369.

Finch, B.W., and Ephrussi, B. (1967). Retention of multiple developmental potentialities by cells of a mouse testicular teratocarcinoma during prolonged culture in vitro and their extinction upon hybridization with cells of permanent lines. Proc. Natl. Acad. Sci. USA *57*, 615–621.

Gao, R., Zeng, S., Yang, D., Li, X., Liu, W., Gao, Y., Bai, D., Zhang, L., Chen, C., Kang, Y., et al. (2023). Resetting histone modifications during human prenatal germline development. Cell Discov. *9*, 14.

Garcia-Alonso, L., Holland, C.H., Ibrahim, M.M., Turei, D., and Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. *29*, 1363–1375.

Gell, J.J., Liu, W., Sosa, E., Chialastri, A., Hancock, G., Tao, Y., Wamaitha, S.E., Bower, G., Dey, S.S., and Clark, A.T. (2020). An Extended Culture System that Supports Human Primordial Germ Cell-like Cell Survival and Initiation of DNA Methylation Erasure. Stem Cell Rep. *14*, 433–446.

Geurten, C., Geurten, M., Rigo, V., and Dresse, M.-F. (2020). Neonatal cancer epidemiology and outcome: A retrospective study. J. Pediatr. Hematol. Oncol. *42*, e286–e292.

Gkountela, S., Zhang, K.X., Shafiq, T.A., Liao, W.-W., Hargan-Calvopiña, J., Chen, P.-Y., and Clark, A.T. (2015). DNA demethylation dynamics in the human prenatal germline. Cell *161*, 1425–1436.

Grand, R.S., Pregnolato, M., Baumgartner, L., Hoerner, L., Burger, L., and Schübeler, D. (2024). Genome access is transcription factor-specific and defined by nucleosome position. Mol. Cell *84*, 3455–3468.e6.

Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y., et al. (2015). The transcriptome and DNA methylome landscapes of human primordial germ cells. Cell *161*, 1437–1452.

Guo, J., Sosa, E., Chitiashvili, T., Nie, X., Rojas, E.J., Oliver, E., Stukenborg, J.B., DonorConnect, Plath, K., Hotaling, J.M., et al. (2021). Single-cell analysis of the developing human testis reveals somatic niche cell specification and fetal germline stem cell establishment. Cell Stem Cell *28*, 764–778.e4.

Gu, Y., Chen, J., Wang, Z., Shao, Q., Li, Z., Ye, Y., Xiao, X., Xiao, Y., Liu, W., Xie, S., et al. (2024). Integrated analysis and systematic characterization of the regulatory network for human germline development. J. Genet. Genomics. *52*, 204–219.

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

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

Hackett, J.A., Dietmann, S., Murakami, K., Down, T.A., Leitch, H.G., and Surani, M.A. (2013b). Synergistic mechanisms of DNA demethylation during transition to ground-state pluripotency. Stem Cell Rep. *1*, 518–531.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29.

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

Hill, P.W.S., Leitch, H.G., Requena, C.E., Sun, Z., Amouroux, R., Roman-Trufero, M., Borkowska, M., Terragni, J., Vaisvila, R., Linnett, S., et al. (2018). Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. Nature *555*, 392–396.

Hsu, F.-M., Wu, Q.Y., Fabyanic, E.B., Wei, A., Wu, H., and Clark, A.T. (2023). TET1 facilitates specification of early human lineages including germ cells. iScience *26*, 107191.

Humpherys, D., Eggan, K., Akutsu, H., Hochedlinger, K., Rideout, W.M., Biniszkiewicz, D., Yanagimachi, R., and Jaenisch, R. (2001). Epigenetic instability in ES cells and cloned mice. Science *293*, 95–97.

Irie, N., Weinberger, L., Tang, W.W.C., Kobayashi, T., Viukov, S., Manor, Y.S., Dietmann, S., Hanna, J.H., and Surani, M.A. (2015). SOX17 is a critical specifier of human primordial germ cell fate. Cell *160*, 253–268.

Irie, N., Lee, S.-M., Lorenzi, V., Xu, H., Chen, J., Inoue, M., Kobayashi, T., Sancho-Serra, C., Drousioti, E., Dietmann, S., et al. (2023). DMRT1 regulates human germline commitment. Nat. Cell Biol. *25*, 1439–1452.

Isbel, L., Grand, R.S., and Schübeler, D. (2022). Generating specificity in genome regulation through transcription factor sensitivity to chromatin. Nat. Rev. Genet. *23*, 728–740.

Iurlaro, M., von Meyenn, F., and Reik, W. (2017). DNA methylation homeostasis in human and mouse development. Curr. Opin. Genet. Dev. *43*, 101–109.

Jiang, H., Du, M., Li, Y., Zhou, T., Lei, J., Liang, H., Zhong, Z., Al-Lamki, R.S., Jiang, M., and Yang, J. (2022). ID proteins promote the survival and primed-to-naive transition of human embryonic stem cells through TCF3-mediated transcription. Cell Death Dis. *13*, 549.

Jo, K., Teague, S., Chen, B., Khan, H.A., Freeburne, E., Li, H., Li, B., Ran, R., Spence, J.R., and Heemskerk, I. (2022). Efficient differentiation of human primordial germ cells through geometric control reveals a key role for Nodal signaling. eLife *11*, e72811.

Kalish, J.M., Jiang, C., and Bartolomei, M.S. (2014). Epigenetics and imprinting in human disease. Int. J. Dev. Biol. *58*, 291–298.

Kaluscha, S., Domcke, S., Wirbelauer, C., Stadler, M.B., Durdu, S., Burger, L., and Schübeler, D. (2022). Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. Nat. Genet. *54*, 1895–1906.

Kamimoto, K., Stringa, B., Hoffmann, C.M., Jindal, K., Solnica-Krezel, L., and Morris, S.A. (2023). Dissecting cell identity via network inference and in silico gene perturbation. Nature *614*, 742–751.

Kehler, J., Tolkunova, E., Koschorz, B., Pesce, M., Gentile, L., Boiani, M., Lomelí, H., Nagy, A., McLaughlin, K.J., Schöler, H.R., and Tomilin, A. (2004). Oct4 is required for primordial germ cell survival. EMBO Rep. *5*, 1078–1083.

Kelsey, G. (2010). Imprinting on chromosome 20: tissue-specific imprinting and imprinting mutations in the GNAS locus. Am. J. Med. Genet. C Semin. Med. Genet. *154C*, 377–386.

Kleinsmith, L.J., and Pierce, G.B. (1964). Multipotentiality of single embryonal carcinoma cells. Cancer Res. *24*, 1544–1551.

Kobayashi, T., and Surani, M.A. (2018). On the origin of the human germline. Development *145*, dev162594.

Kobayashi, S., Kohda, T., Miyoshi, N., Kuroiwa, Y., Aisaka, K., Tsutsumi, O., Kaneko-Ishino, T., and Ishino, F. (1997). Human PEG1/MEST, an imprinted gene on chromosome 7. Hum. Mol. Genet. *6*, 781–786.

Kobayashi, T., Yamaguchi, T., Hamanaka, S., Kato-Itoh, M., Yamazaki, Y., Ibata, M., Sato, H., Lee, Y.-S., Usui, J.-I., Knisely, A.S.,

et al. (2010). Generation of rat pancreas in mouse by interspecific blastocyst injection of pluripotent stem cells. Cell *142*, 787–799.

Kobayashi, M., Kobayashi, M., Odajima, J., Shioda, K., Hwang, Y.S., Sasaki, K., Chatterjee, P., Kramme, C., Kohman, R.E., Church, G.M., et al. (2022). Expanding homogeneous culture of human primordial germ cell-like cells maintaining germline features without serum or feeder layers. Stem Cell Rep. *17*, 507–521.

Kojima, Y., Sasaki, K., Yokobayashi, S., Sakai, Y., Nakamura, T., Yabuta, Y., Nakaki, F., Nagaoka, S., Woltjen, K., Hotta, A., et al. (2017). Evolutionarily Distinctive Transcriptional and Signaling Programs Drive Human Germ Cell Lineage Specification from Pluripotent Stem Cells. Cell Stem Cell *21*, 517–532.e5.

Kojima, Y., Yamashiro, C., Murase, Y., Yabuta, Y., Okamoto, I., Iwatani, C., Tsuchiya, H., Nakaya, M., Tsukiyama, T., Nakamura, T., et al. (2021). GATA transcription factors, SOX17 and TFAP2C, drive the human germ-cell specification program. Life Sci. Alliance *4*, e202000974.

Labosky, P.A., Barlow, D.P., and Hogan, B.L. (1994). Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of insulin-like growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. Development *120*, 3197–3204.

Leitch, H.G., and Smith, A. (2013). The mammalian germline as a pluripotency cycle. Development *140*, 2495–2501.

Leitch, H.G., Blair, K., Mansfield, W., Ayetey, H., Humphreys, P., Nichols, J., Surani, M.A., and Smith, A. (2010). Embryonic germ cells from mice and rats exhibit properties consistent with a generic pluripotent ground state. Development *137*, 2279–2287.

Leitch, H.G., Tang, W.W.C., and Surani, M.A. (2013a). Primordial germ-cell development and epigenetic reprogramming in mammals. Curr. Top. Dev. Biol. *104*, 149–187.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013b). Naive pluripotency is associated with global DNA hypomethylation. Nat. Struct. Mol. Biol. *20*, 311–316.

Leitch, H.G., Nichols, J., Humphreys, P., Mulas, C., Martello, G., Lee, C., Jones, K., Surani, M.A., and Smith, A. (2013c). Rebuilding pluripotency from primordial germ cells. Stem Cell Rep. *1*, 66–78.

Leitch, H.G., Okamura, D., Durcova-Hills, G., Stewart, C.L., Gardner, R.L., Matsui, Y., and Papaioannou, V.E. (2014). On the fate of primordial germ cells injected into early mouse embryos. Dev. Biol. *385*, 155–159.

Li, P., Tong, C., Mehrian-Shai, R., Jia, L., Wu, N., Yan, Y., Maxson, R.E., Schulze, E.N., Song, H., Hsieh, C.-L., et al. (2008). Germline competent embryonic stem cells derived from rat blastocysts. Cell *135*, 1299–1310.

Li, Y., Xie, C., Murphy, S.K., Skaar, D., Nye, M., Vidal, A.C., Cecil, K.M., Dietrich, K.N., Puga, A., Jirtle, R.L., and Hoyo, C. (2016). Lead Exposure during Early Human Development and DNA Methylation of Imprinted Gene Regulatory Elements in Adulthood. Environ. Health Perspect. *124*, 666–673.

Lun, A.T.L., and Marioni, J.C. (2017). Overcoming confounding plate effects in differential expression analyses of single-cell RNA-seq data. Biostatistics *18*, 451–464.

M Ascensión, A., Ibáñez-Solé, O., Inza, I., Izeta, A., and Araúzo-Bravo, M.J. (2022). Triku: a feature selection method based on nearest neighbors for single-cell data. GigaScience *11*, giac017.

Marcotte, E.L., Pankratz, N., Amatruda, J.F., Frazier, A.L., Krailo, M., Davies, S., Starr, J.R., Lau, C.C., Roesler, M., Langer, E., et al. (2017). Variants in BAK1, SPRY4, and GAB2 are associated with pediatric germ cell tumors: A report from the children's oncology group. Genes Chromosomes Cancer *56*, 548–558.

Matsui, Y., Zsebo, K., and Hogan, B.L. (1992). Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. Cell *70*, 841–847.

McGlynn, K.A., and Cook, M.B. (2009). Etiologic factors in testicular germ-cell tumors. Future Oncol. *5*, 1389–1402.

Messmer, T., von Meyenn, F., Savino, A., Santos, F., Mohammed, H., Lun, A.T.L., Marioni, J.C., and Reik, W. (2019). Transcriptional Heterogeneity in Naive and Primed Human Pluripotent Stem Cells at Single-Cell Resolution. Cell Rep. *26*, 815–824.e4.

Murase, Y., Yabuta, Y., Ohta, H., Yamashiro, C., Nakamura, T., Yamamoto, T., and Saitou, M. (2020). Long-term expansion with germline potential of human primordial germ cell-like cells in vitro. EMBO J. *39*, e104929.

Murase, Y., Yokogawa, R., Yabuta, Y., Nagano, M., Katou, Y., Mizuyama, M., Kitamura, A., Puangsricharoen, P., Yamashiro, C., Hu, B., et al. (2024). In vitro reconstitution of epigenetic reprogramming in the human germ line. Nature *631*, 170–178.

Nakamura, T., Okamoto, I., Sasaki, K., Yabuta, Y., Iwatani, C., Tsuchiya, H., Seita, Y., Nakamura, S., Yamamoto, T., and Saitou, M. (2016). A developmental coordinate of pluripotency among mice, monkeys and humans. Nature *537*, 57–62.

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

Northrup, E., Eisenblätter, R., Glage, S., Rudolph, C., Dorsch, M., Schlegelberger, B., Hedrich, H.-J., and Zschemisch, N.-H. (2011). Loss of Dnd1 facilitates the cultivation of genital ridge-derived rat embryonic germ cells. Exp. Cell Res. *317*, 1885–1894.

Ono, R., Kobayashi, S., Wagatsuma, H., Aisaka, K., Kohda, T., Kaneko-Ishino, T., and Ishino, F. (2001). A retrotransposon-derived gene, PEG10, is a novel imprinted gene located on human chromosome 7q21. Genomics *73*, 232–237.

Oosterhuis, J.W., and Looijenga, L.H.J. (2019). Human germ cell tumours from a developmental perspective. Nat. Rev. Cancer *19*, 522–537.

Overeem, A.W., Chang, Y.W., Spruit, J., Roelse, C.M., and Chuva De Sousa Lopes, S.M. (2021). Ligand-Receptor Interactions Elucidate Sex-Specific Pathways in the Trajectory From Primordial Germ Cells to Gonia During Human Development. Front. Cell Dev. Biol. *9*, 661243.

Peall, K.J., Smith, D.J., Kurian, M.A., Wardle, M., Waite, A.J., Hedderly, T., Lin, J.-P., Smith, M., Whone, A., Pall, H., et al. (2013). SGCE mutations cause psychiatric disorders: clinical and genetic characterization. Brain *136*, 294–303.

Perrett, R.M., Turnpenny, L., Eckert, J.J., O'Shea, M., Sonne, S.B., Cameron, I.T., Wilson, D.I., Rajpert-De Meyts, E., and Hanley, N.A. (2008). The early human germ cell lineage does not express

SOX2 during in vivo development or upon in vitro culture. Biol. Reprod. *78*, 852–858.

Alexa, A., and Rahnenführer, J.. topGO: Enrichment Analysis for Gene Ontology. R package version 2.50.0. https://bioconductor.org/packages/topGO.

Public Health England (2021). Children, teenagers and young adults UK cancer statistics report 2021 .

Ramakrishna, N.B., Murison, K., Miska, E.A., and Leitch, H.G. (2021). Epigenetic Regulation during Primordial Germ Cell Development and Differentiation. Sex. Dev. *15*, 411–431.

Resnick, J.L., Bixler, L.S., Cheng, L., and Donovan, P.J. (1992). Long-term proliferation of mouse primordial germ cells in culture. Nature *359*, 550–551.

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

Rossant, J., and Tam, P.P.L. (2022). Early human embryonic development: Blastocyst formation to gastrulation. Dev. Cell *57*, 152–165.

Saitou, M., and Hayashi, K. (2021). Mammalian in vitro gametogenesis. Science *374*, eaaz6830.

Santini, L., Kowald, S., Sestini, G., Rivron, N.C., and Leeb, M. (2024). FoxO transcription factors actuate the formative pluripotency specific gene expression programme. Preprint at bioRxiv. https://doi.org/10.1101/2024.01.13.575494.

Sasaki, K., Yokobayashi, S., Nakamura, T., Okamoto, I., Yabuta, Y., Kurimoto, K., Ohta, H., Moritoki, Y., Iwatani, C., Tsuchiya, H., et al. (2015). Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. Cell Stem Cell *17*, 178–194.

Sepulveda-Rincon, L.P., Wang, Y.-F., Whilding, C., Moyon, B., Ojarikre, O.A., Maciulyte, V., Hamazaki, N., Hayashi, K., Turner, J.M.A., and Leitch, H.G. (2024). Determining the potency of primordial germ cells by injection into early mouse embryos. Dev. Cell *59*, 695–704.e5.

Shahbazi, M.N., and Zernicka-Goetz, M. (2018). Deconstructing and reconstructing the mouse and human early embryo. Nat. Cell Biol. *20*, 878–887.

Shovlin, T.C., Durcova-Hills, G., Surani, A., and McLaren, A. (2008). Heterogeneity in imprinted methylation patterns of pluripotent embryonic germ cells derived from pre-migratory mouse germ cells. Dev. Biol. *313*, 674–681.

Sierra, R.A., Hoverter, N.P., Ramirez, R.N., Vuong, L.M., Mortazavi, A., Merrill, B.J., Waterman, M.L., and Donovan, P.J. (2018). TCF7L1 suppresses primitive streak gene expression to support human embryonic stem cell pluripotency. Development *145*, dev161075.

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

Squair, J.W., Gautier, M., Kathe, C., Anderson, M.A., James, N.D., Hutson, T.H., Hudelle, R., Qaiser, T., Matson, K.J.E., Barraud, Q., et al. (2021). Confronting false discoveries in single-cell differential expression. Nat. Commun. *12*, 5692.

Stewart, C.L., Gadi, I., and Bhatt, H. (1994). Stem cells from primordial germ cells can reenter the germ line. Dev. Biol. *161*, 626–628.

Sun, J., Ting, M.-C., Ishii, M., and Maxson, R. (2016). Msx1 and Msx2 function together in the regulation of primordial germ cell migration in the mouse. Dev. Biol. *417*, 11–24.

Surani, M.A., Hayashi, K., and Hajkova, P. (2007). Genetic and epigenetic regulators of pluripotency. Cell *128*, 747–762.

Tada, T., Tada, M., Hilton, K., Barton, S.C., Sado, T., Takagi, N., and Surani, M.A. (1998). Epigenotype switching of imprintable loci in embryonic germ cells. Dev. Gene. Evol. *207*, 551–561.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Takahashi, N., Coluccio, A., Thorball, C.W., Planet, E., Shi, H., Offner, S., Turelli, P., Imbeault, M., Ferguson-Smith, A.C., and Trono, D. (2019). ZNF445 is a primary regulator of genomic imprinting. Genes Dev. *33*, 49–54.

Tang, W.W.C., Dietmann, S., Irie, N., Leitch, H.G., Floros, V.I., Bradshaw, C.R., Hackett, J.A., Chinnery, P.F., and Surani, M.A. (2015). A unique gene regulatory network resets the human germline epigenome for development. Cell *161*, 1453–1467.

Tang, W.W.C., Castillo-Venzor, A., Gruhn, W.H., Kobayashi, T., Penfold, C.A., Morgan, M.D., Sun, D., Irie, N., and Surani, M.A. (2022). Sequential enhancer state remodelling defines human germline competence and specification. Nat. Cell Biol. *24*, 448–460.

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

Turan, S., and Bastepe, M. (2013). The GNAS complex locus and human diseases associated with loss-of-function mutations or epimutations within this imprinted gene. Horm. Res. Paediatr. *80*, 229–241.

Turnpenny, L., Spalluto, C.M., Perrett, R.M., O'Shea, M., Hanley, K.P., Cameron, I.T., Wilson, D.I., and Hanley, N.A. (2006). Evaluating human embryonic germ cells: concord and conflict as pluripotent stem cells. Stem Cell. *24*, 212–220.

Tyser, R.C.V., Mahammadov, E., Nakanoh, S., Vallier, L., Scialdone, A., and Srinivas, S. (2021). Single-cell transcriptomic characterization of a gastrulating human embryo. Nature *600*, 285–289.

Vaisvila, R., Ponnaluri, V.K.C., Sun, Z., Langhorst, B.W., Saleh, L., Guan, S., Dai, N., Campbell, M.A., Sexton, B.S., Marks, K., et al. (2021). Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. Genome Res. *31*, 1280–1289.

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

von Meyenn, F., and Reik, W. (2015). Forget the parents: epigenetic reprogramming in human germ cells. Cell *161*, 1248–1251.

Wong, F.C.K., Chambers, I., and Mullin, N.P. (2016). SOX2-Dependent Regulation of Pluripotent Stem Cells, H. Kondoh and R. Lovell-Badge, eds. (Elsevier), pp. 163–185.

Yamashiro, C., Sasaki, K., Yabuta, Y., Kojima, Y., Nakamura, T., Okamoto, I., Yokobayashi, S., Murase, Y., Ishikura, Y., Shirane, K., et al. (2018). Generation of human oogonia from induced pluripotent stem cells in vitro. Science *362*, 356–360.

Yamashiro, C., Sasaki, K., Yokobayashi, S., Kojima, Y., and Saitou, M. (2020). Generation of human oogonia from induced pluripotent stem cells in culture. Nat. Protoc. *15*, 1560–1583.

Yeung, K.S., Ho, M.S.P., Lee, S.L., Kan, A.S.Y., Chan, K.Y.K., Tang, M.H.Y., Mak, C.C.Y., Leung, G.K.C., So, P.L., Pfundt, R., et al. (2018). Paternal uniparental disomy of chromosome 19 in a pair of monochorionic diamniotic twins with dysmorphic features and developmental delay. J. Med. Genet. *55*, 847–852.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science *356*, eaaj2239.

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

Zhang, M., Leitch, H.G., Tang, W.W.C., Festuccia, N., Hall-Ponsele, E., Nichols, J., Surani, M.A., Smith, A., and Chambers, I. (2018). Esrrb Complementation Rescues Development of Nanog-Null Germ Cells. Cell Rep. *22*, 332–339.

Zheng, Y., Yan, R., Kobayashi, M., Xiang, L., Yang, R., Goedel, A., Kang, Y., Xue, X., Esfahani, S., Liu, Y., et al. (2022). Single-Cell Analysis of Embryoids Reveals Lineage Diversification Roadmaps of Early Human Development. Cell Stem Cell *29*, 1402–1419.e8.

Zimmerman, K.D., Espeland, M.A., and Langefeld, C.D. (2021). A practical solution to pseudoreplication bias in single-cell studies. Nat. Commun. *12*, 738.