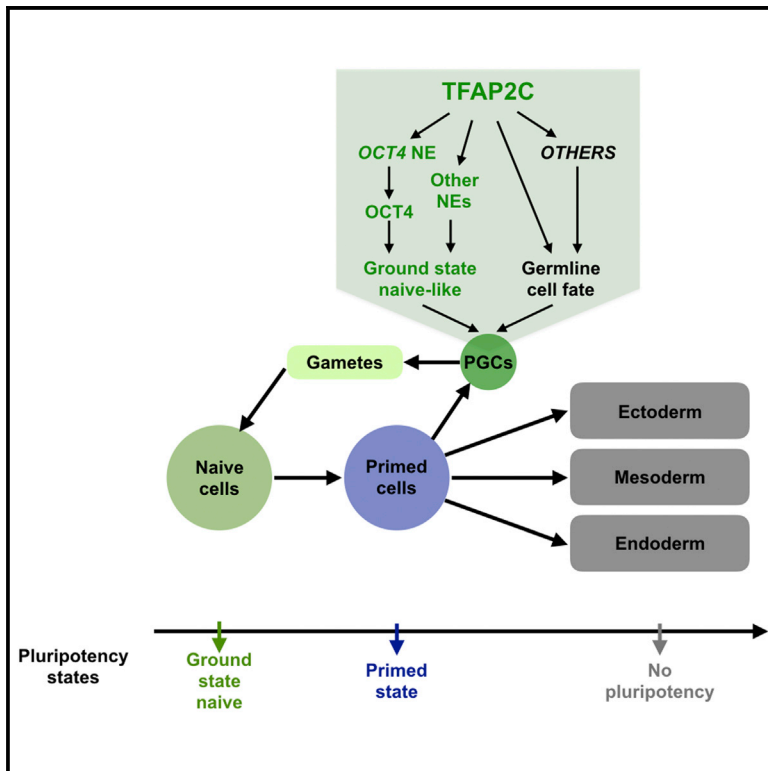


Cell Reports

The TFAP2C-Regulated *OCT4* Naive Enhancer Is Involved in Human Germline Formation

Graphical Abstract



Authors

Di Chen, Wanlu Liu, Jill Zimmerman, ..., Joanna J. Gell, Steven E. Jacobsen, Amander T. Clark

Correspondence

clarka@ucla.edu

In Brief

Combining genomics and functional studies, Chen et al. identify the open chromatin state of human primordial germ cells (hPGCs), leading to the discovery that TFAP2C regulates hPGC development through the opening of naive enhancers.

Highlights

- The chromatin and transcriptome of hPGCs resembles ground-state naive hESCs
- TFAP2C is required for hPGC formation and expression of KLF4
- The TFAP2C-regulated OCT4 naive enhancer is involved in hPGC formation



The TFAP2C-Regulated OCT4 Naive Enhancer Is Involved in Human Germline Formation

Di Chen,^{1,9} Wanlu Liu,^{2,9} Jill Zimmerman,¹ William A. Pastor,^{1,10} Rachel Kim,³ Linzi Hosohama,¹ Jamie Ho,¹ Marianna Aslanyan,¹ Joanna J. Gell,^{4,5} Steven E. Jacobsen,^{1,2,3,6,7} and Amander T. Clark^{1,2,3,8,11,*}

¹Department of Molecular Cell and Developmental Biology, University of California, Los Angeles, Los Angeles, CA, USA

²Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA, USA

³Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, Los Angeles, CA, USA

⁴Department of Pediatrics, Division of Hematology-Oncology, Los Angeles, CA 90095, USA

⁵David Geffen School of Medicine, Los Angeles, CA 90095, USA

⁶Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA

⁷Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, CA, USA

⁸Jonsson Comprehensive Cancer Center, University of California, Los Angeles, Los Angeles, CA, USA

⁹These authors contributed equally

¹⁰Present address: Department of Biochemistry, McGill University, Montreal, QC H3G 1Y6, Canada

¹¹Lead Contact

*Correspondence: clarka@ucla.edu

<https://doi.org/10.1016/j.celrep.2018.12.011>

SUMMARY

Human primordial germ cells (hPGCs) are the first embryonic progenitors in the germ cell lineage, yet the molecular mechanisms required for hPGC formation are not well characterized. To identify regulatory regions in hPGC development, we used the assay for transposase-accessible chromatin using sequencing (ATAC-seq) to systematically characterize regions of open chromatin in hPGCs and hPGC-like cells (hPGCLCs) differentiated from human embryonic stem cells (hESCs). We discovered regions of open chromatin unique to hPGCs and hPGCLCs that significantly overlap with TFAP2C-bound enhancers identified in the naive ground state of pluripotency. Using CRISPR/Cas9, we show that deleting the TFAP2C-bound naive enhancer at the OCT4 locus (also called *POU5F1*) results in impaired OCT4 expression and a negative effect on hPGCLC identity.

INTRODUCTION

Germ cells transmit genetic and epigenetic information from one generation to the next and are critical for fertility. Germ cell specification begins in the embryo with the formation of primordial germ cells (PGCs), and this remarkable event has been studied in a range of animal models, including insects, crickets, spiders, worms, fish, frogs, rodents, and non-human primates (Clark et al., 2017; Extavour and Akam, 2003; Leitch et al., 2013b; Magnúsdóttir and Surani, 2014; Nakamura and Extavour, 2016; Raz, 2003; Sasaki et al., 2016; Schwager et al., 2015; Williamson and Lehmann, 1996). The challenge with studying mechanisms involved in early human PGC (hPGC) development is that these cells are specified at the time of em-

bryo implantation, a time that is inaccessible for research. Given this, the mouse model has traditionally been used to extrapolate mechanistic information on PGC formation across all mammals, including humans.

Recent work suggests that humans have evolved a new transcription factor network for inducing hPGC formation centering on the transcription factor SRY Box 17 (SOX17) (Irie et al., 2015). The discovery of SOX17's role in hPGC development began with 4i inhibitor (4i) cultured human pluripotent stem cells (hPSCs) and the differentiation of hPGC-like cells (hPGCLCs) *in vitro*. It is now appreciated that almost all hESC and human induced pluripotent stem cell (hiPSC) lines tested to date are capable of hPGCLC formation. Therefore, hPSCs have emerged as an important model for uncovering new mechanistic insight into hPGC development. Indeed, using CRISPR/Cas9 gene editing approaches, critical roles for PRDM1, EOMESODERMIN (EOMES), and TFAP2C have also been identified in hPGC development (Chen et al., 2017; Irie et al., 2015; Kojima et al., 2017; Sasaki et al., 2015).

hPSCs can be cultured in either the naive or primed states of pluripotency. Ground-state naive pluripotent stem cells are different from 4i and primed pluripotent stem cells in that they are more similar to the pre-implantation epiblast (Gu et al., 2016; Pastor et al., 2016; Theunissen et al., 2014, 2016; Weinberger et al., 2016), whereas primed pluripotent stem cells represent a post-implantation epiblast state (Nakamura et al., 2016). Mammalian PGCLCs can be differentiated from naive ground-state mouse or hPSCs by first re-priming through epiblast-like cells (EpiLCs) followed by aggregate differentiation (Hayashi et al., 2011, 2012; von Meyenn et al., 2016). When starting from primed hESCs or hESCs cultured in 4i, hPGCLCs are consistently generated from a partially differentiated gastrulation-like intermediate called either incipient mesoderm-like cells (iMeLCs) (Sasaki et al., 2015) or mesendoderm precursors (pre-ME) (Kobayashi et al., 2017). In all cases, hPGCLC formation from these early progenitors is induced in three-dimensional aggregates in the presence of



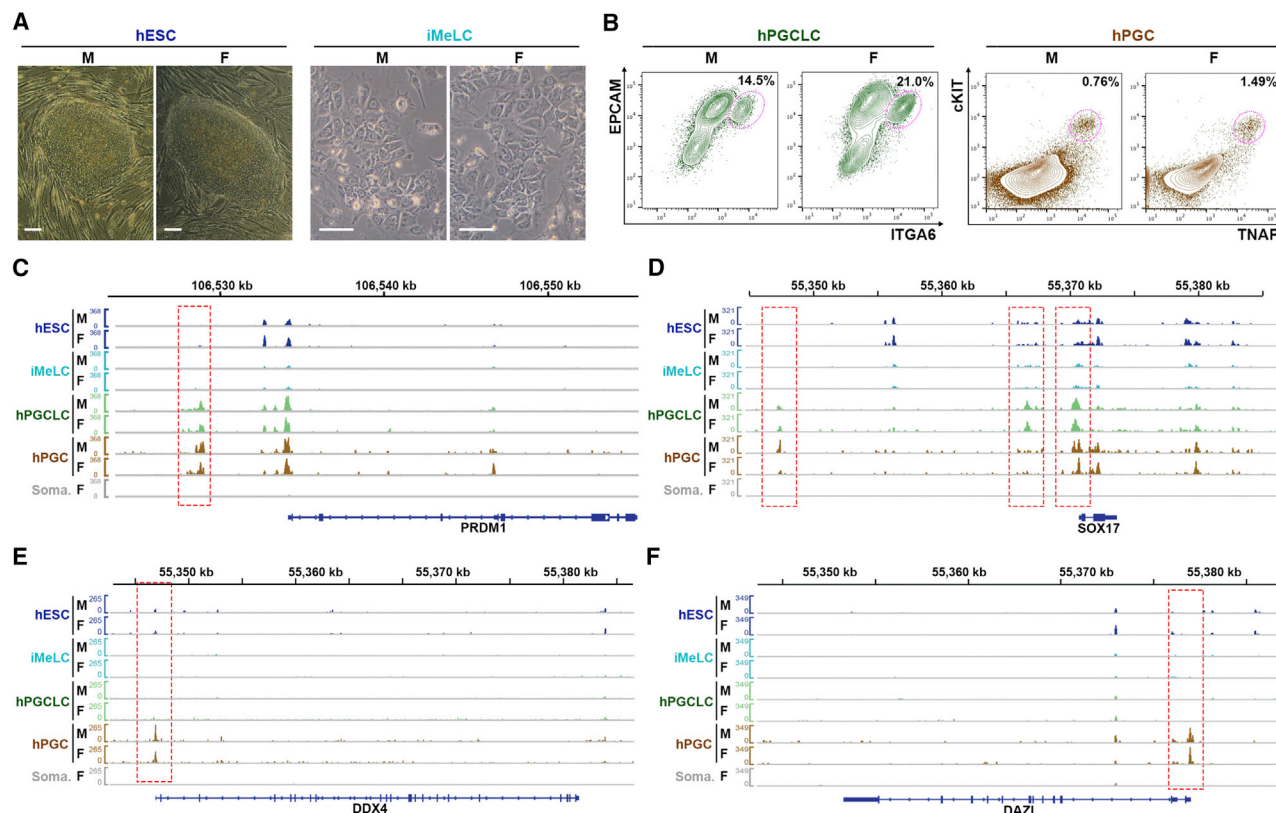


Figure 1. Identifying Unique Regions of Open Chromatin in Human Germline Cells

(A) Morphology of male (UCLA2) and female (UCLA1) primed hESCs, and iMeLCs used for ATAC-seq. Scale bars, 100 μ m.

(B) Male (UCLA2) and female (UCLA1) hPGCLCs were isolated as ITGA6/EPCAM double-positive cells at day 4 of aggregate differentiation. 82d and 89d hPGCs were isolated as TNAP/cKIT double-positive cells from a pair of embryonic testes and ovaries, respectively.

(C–F) Screenshot of the ATAC-seq signal over *PRDM1* (C), *SOX17* (D), *DDX4* (E), and *DAZL* (F) for male and female primed hESCs, iMeLCs, hPGCLCs, hPGCs, and embryonic somatic cells (soma). Red dotted boxes highlight ATAC-seq peaks in hPGCLCs and/or hPGCs, but not in primed hESCs, iMeLCs, or embryonic somatic tissues.

F, female; M, male. See also Figure S1.

growth factors, including bone morphogenetic protein 4 (BMP4). Therefore, exit from naive pluripotency into a primed state followed by short exposure to activin A and WNT are required for inducing hPGCLC differentiation with BMP4. Interestingly, some transcription factors diagnostic of the ground state naive pluripotency are expressed by hPGCs and hPGCLCs (e.g., *KLF4*, *KLF5*, and *DPPA3*), and similar to the naive ground state, hPGCs are also globally demethylated (Gkoutela et al., 2015; Tang et al., 2015). Despite this, any additional requirements for hPGC development are not well understood.

In order to identify molecular mechanisms in hPGC formation, we utilized an unbiased approach by identifying motifs in open chromatin unique to hPGCs and hPGCLCs. From this screen, we used CRISPR/Cas9 approaches to determine that re-opening of a large fraction of TFAP2C-bound ground-state naive enhancers (NEs) combined with a shift in the global transcriptional program toward ground-state naive pluripotency is a major milestone in the regeneration of human germline cells from primed pluripotent stem cells.

RESULTS

Application of ATAC-Seq to the Induction of Human Germline Cells

In order to identify unique transcription factors and regulatory elements that may function in hPGC development, we performed the assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Buenrostro et al., 2013) on chromatin isolated from hPGCLCs (ITGA6/EPCAM) and hPGCs (TNAP/cKIT) collected by fluorescence-activated cell sorting (FACS), as well as primed hESCs and iMeLCs (Figures 1A and 1B). Controls for ITGA6/EPCAM staining involved analysis of undifferentiated hESCs that were positive (Figure S1A) as shown previously (Sasaki et al., 2015). In order to confirm hPGCLC identity of the FACS-isolated cells from female (UCLA1) and male (UCLA2) hESC lines (Chen et al., 2017), we compared RNA sequencing (RNA-seq) of these cells with hPGCLCs differentiated from hESCs cultured in 4i (Irie et al., 2015) and hPGCLCs differentiated from hiPSCs cultured in StemFit followed by 48 hr in iMeLC media (Sasaki et al., 2015) (Figure S1B). This comparison shows

that the hPGCLCs differentiated from UCLA1 and UCLA2 using a 1-day iMeLC differentiation step prior to aggregate formation in BMP4 are transcriptionally comparable to hPGCLCs differentiated from 4i and StemFit pluripotent cells. Similarly, the hPGCs isolated using TNAP/cKIT in the current study also cluster together with hPGCs isolated by Irie et al. (2015).

Given that the number of hPGCs isolated from a pair of embryonic gonads is limited (1,000–10,000 TNAP/cKIT hPGCs per embryo), we first tested ATAC-seq on different numbers of hESCs ranging from 1,000 to 50,000 cells (Figure S1C). We found concordance of ATAC-seq peaks even down to as few as 1,000 cells (Figure S1C), indicating that our ATAC-seq approach could be used on sorted hPGCs/hPGCLCs where cell number is more limiting.

Next, we collected hESCs, iMeLCs, and ITGA6/EPCAM-sorted hPGCLCs using UCLA1 and UCLA2 hESC lines. We also collected TNAP/cKIT hPGCs isolated by FACS from a pair of 82 days post-fertilization (82d) fetal testes and a pair of 89d fetal ovaries (Figures 1A and 1B). We constructed ATAC-seq libraries from all samples to characterize chromatin accessibility in the different cell types. In order to identify regions of open chromatin unique to germline cells, but not somatic cells, we also made ATAC-seq libraries from embryonic somatic tissues (76d female embryo), including embryonic heart, liver, lung, and skin. ATAC-seq reads from the different somatic libraries were merged together to create a composite “somatic” sample (called soma.). Analysis of ATAC-seq peaks across different cell types at the promoter region of the housekeeping genes, for example *TUBB* and *RHOB* (Figures S1D and S1E), indicated that the quality of the libraries were the same between samples, and this was further confirmed by equivalent expected size distributions across all samples (Figure S1F) (Buenrostro et al., 2013).

Clustering of all samples revealed overlaps between the ATAC-seq peaks of different biological replicates rather than sample sex (Figure S1G). Given the high concordance between replicates independent of sex, we combined reads from male and female hPGCs and male and female hPGCLCs to create composite “hPGC” and “hPGCLC” data sets respectively for further analysis. Similarly, reads from male and female hESCs and male and female iMeLCs were merged to create the “hESC” and “iMeLC” sets. Analysis of ATAC-seq signal occupancy at the early hPGC genes *PRDM1* and *SOX17* loci revealed regions of open chromatin distal to the transcription start site (TSS) in hPGCLCs and hPGCs, but not other samples (Figures 1C and 1D). Similarly, at the *NANOG* gene locus, a differentially open germline cell-specific region was identified in hPGCLCs and hPGCs, but not primed pluripotent stem cells (Figure S1H). Moreover, differentially open ATAC-seq peaks for late PGC genes *DDX4* and *DAZL* are detected in hPGCs, but not hPGCLCs or other samples (Figures 1E and 1F). These dynamic observations at known germ cell-expressed genes indicate that the ATAC-seq libraries generated in this study could be used to systematically uncover insights into human germline cell-specific open chromatin.

Characterization of Candidate Transcription Factors for Human Germline Cell Formation

In order to identify the regions of open chromatin unique to hPGCs and hPGCLCs, we first identified open chromatin re-

gions that were specific to primed hESCs, iMeLCs, hPGCLCs, and hPGCs relative to embryonic somatic cells (Figures 2A and S2A). Next, we identified transcription factor motifs enriched in the open chromatin at each developmental stage. In primed hESCs, we discovered enrichment for transcription factor motifs corresponding to OCT4, SOX, TEAD, and NANOG (Figure S2A). In iMeLCs we discovered motifs for GATA, TCF, TEAD and SOX corresponding to transcription factor families known to be involved in gastrulation (Figure S2A).

In order to identify germline cell-specific open chromatin (hPGCLCs and hPGCs), we focused on peaks that were hPGCLC specific, hPGC specific, or hPGCLC/PGC intersect (enriched in both). We found that AP2 motifs were strongly enriched in all three types of germline cell-specific open chromatin (Figure 2A). Notably, these germline cell-specific peaks were not open in somatic tissues, including embryonic heart, liver, lung, or skin, and were not open in hESCs or iMeLCs (Figures 2A and S2B). In order to confirm that the germline cell-specific open chromatin was also open in additional hPGC samples, we made four new ATAC-seq libraries from TNAP/cKIT hPGCs isolated by FACS (67d testes and 59d, 91d, and 101d ovaries). The germline cell-specific open chromatin is also present in all these different stages of hPGC development (Figure S2C). Taken together, we identified a signature of germline cell-specific open chromatin in six independently sorted hPGC samples from 59 to 101d of human development. Critically, motifs corresponding to the AP2 family of transcription factors were among the most highly enriched within these regions, as well as motifs corresponding to OCT4, SOX, KLF, NANOG, and GATA families.

In order to identify transcription factor candidates for the motifs identified by ATAC-seq, we used previously published RNA-seq data of hESCs, iMeLCs, hPGCLCs, and hPGCs (Chen et al., 2017) (Figure 2B). Members of the SOX family, specifically SOX17, are known to be required for hPGCLC formation (Irie et al., 2015; Kojima et al., 2017), and motifs for this family were significantly enriched in the open chromatin of hPGCLCs/hPGCs. Cluster analysis suggests that in addition to SOX17, other SOX family members, including SOX15, SOX13, and SOX4, may also be involved in hPGCLC/hPGC formation. The KLF family consists of 17 members, with *KLF4*, *KLF16*, *KLF11*, and *KLF13* exhibiting increased expression in hPGCs and hPGCLCs relative to hESCs and/or iMeLCs. In contrast, *KLF3*, *KLF5*, *KLF6*, *KLF7*, *KLF8*, *KLF10*, and *KLF12* are expressed in all samples. The GATA family has six members, with *GATA2*, *GATA4*, and *GATA6* being the only family members expressed in hPGCLCs and/or hPGCs. Finally, there are five AP2 family members with *TFAP2C* and *TFAP2A* both expressed in hPGCLCs and hPGCs; however, *TFAP2C* is more highly expressed. Additional motifs that emerged in this analysis include *NANOG* and *OCT4* (also called *POU5F1*), which are expressed in all germline cells as well as hESCs and iMeLCs; *TFCP2L1*, which is upregulated in hPGCLCs and hPGCs; and *EOMES*, which is not expressed in hPGCLCs or hPGCs but is expressed in hESCs and iMeLCs, where it regulates hPGCLC competency (Chen et al., 2017; Kojima et al., 2017).

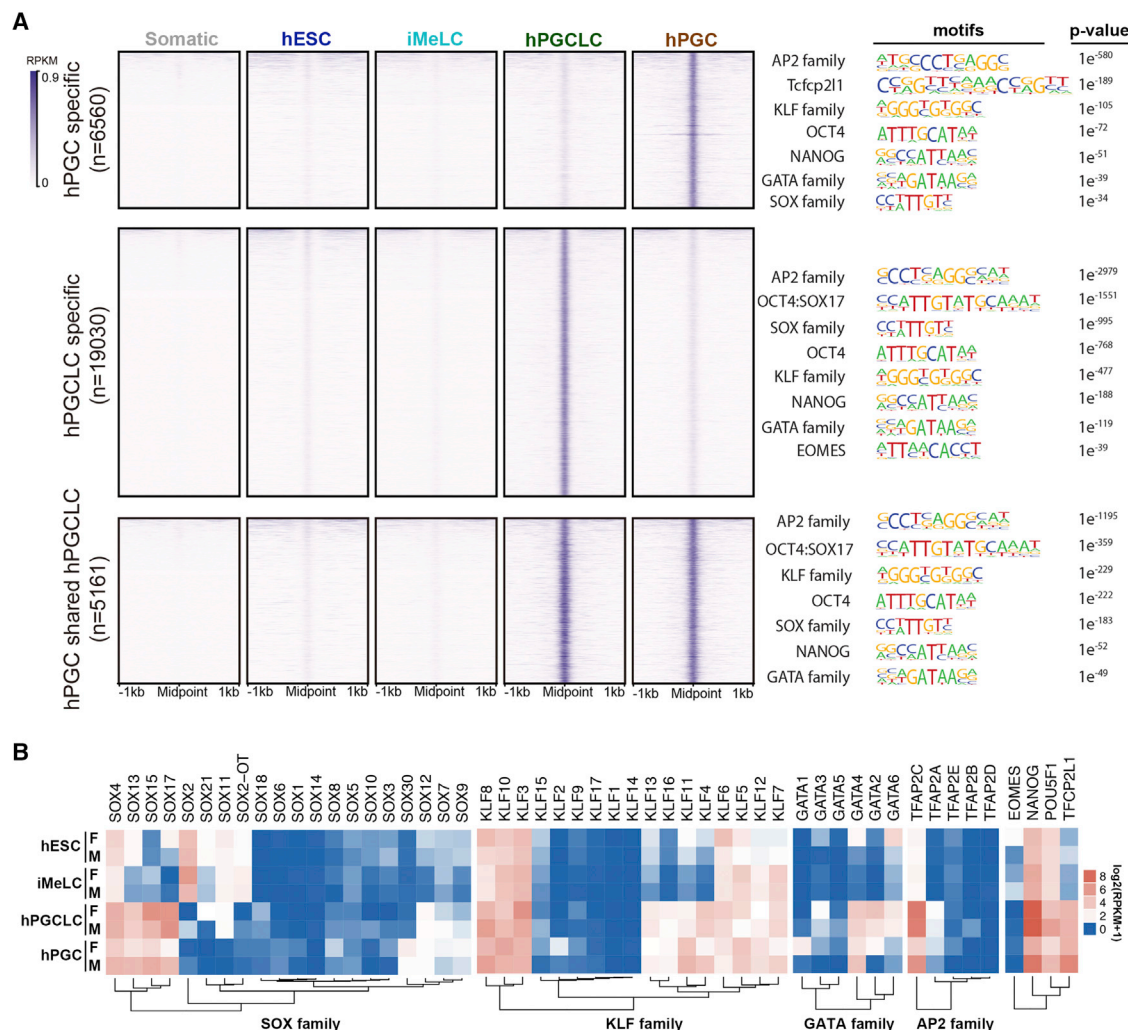


Figure 2. Transcription Factor Motifs Enriched in Open Chromatin of Human Germline Cells

(A) Heatmap of ATAC-seq signals in embryonic somatic tissues, hESCs, iMeLCs, hPGCLCs, and hPGCs over germline cell-specific open chromatin regions (defined as enriched in hPGCLCs, hPGCs, or both) and corresponding transcription factor motifs enriched for those regions.

(B) Heatmap of gene expression levels in hESCs, iMeLCs, hPGCLCs, and hPGCs for transcription factor family members with motifs identified as being enriched in germline cell-specific open chromatin. F, female; M, male.

See also Figure S2.

Human Germline Cells Exhibit Transcriptome and Open Chromatin Profiles that Resemble Ground-State Naive Pluripotent Stem Cells

The transcription factors TFAP2C, KLF4, and TFAP2L1 are all known markers of human ground-state naive pluripotency, a state that is achieved by culturing cells in 5i/L/FA or t2iLGo media (Pastor et al., 2018; Takashima et al., 2014; Theunissen et al., 2014). The term latent pluripotency is used to describe “hidden” pluripotency in mouse PGCs (mPGCs) because mPGC-expressed genes are also expressed by mESCs and because mPGCs have the capacity to undergo culture-induced reversion to pluripotent stem cells called mouse embryonic germ cells (mEGCs) *in vitro* without the need for exogenous reprogramming factors (Leitch and Smith, 2013). To identify the similarities and differences between human germline cells and the primed- and

ground-state naive pluripotency where KLF4, TFAP2L1 and TFAP2C are all highly upregulated, we performed principal component analysis (PCA) comparing all variable genes across the RNA-seq datasets of ground-state naive UCLA1 hESCs (5i/L/FA) (Pastor et al., 2018), primed UCLA1 and UCLA2 hESCs and iMeLCs, hPGCLCs (made from UCLA1 and UCLA2), and hPGCs (Chen et al., 2017) (Figure 3A). This analysis shows that hPGCLCs and hPGCs cluster in principal component 1 (PC1) together with ground-state naive pluripotent stem cells (Figure 3A). Consistently, additional ground-state-naive- rather than primed-state-expressed genes are also expressed in hPGCLCs and hPGCs, including *KLF5*, *DPPA3*, and *TBX3* (Figure 3B). Together, these observations suggest that some components of the ground-state naive pluripotency are re-established in newly specified human germline cells. However, it is

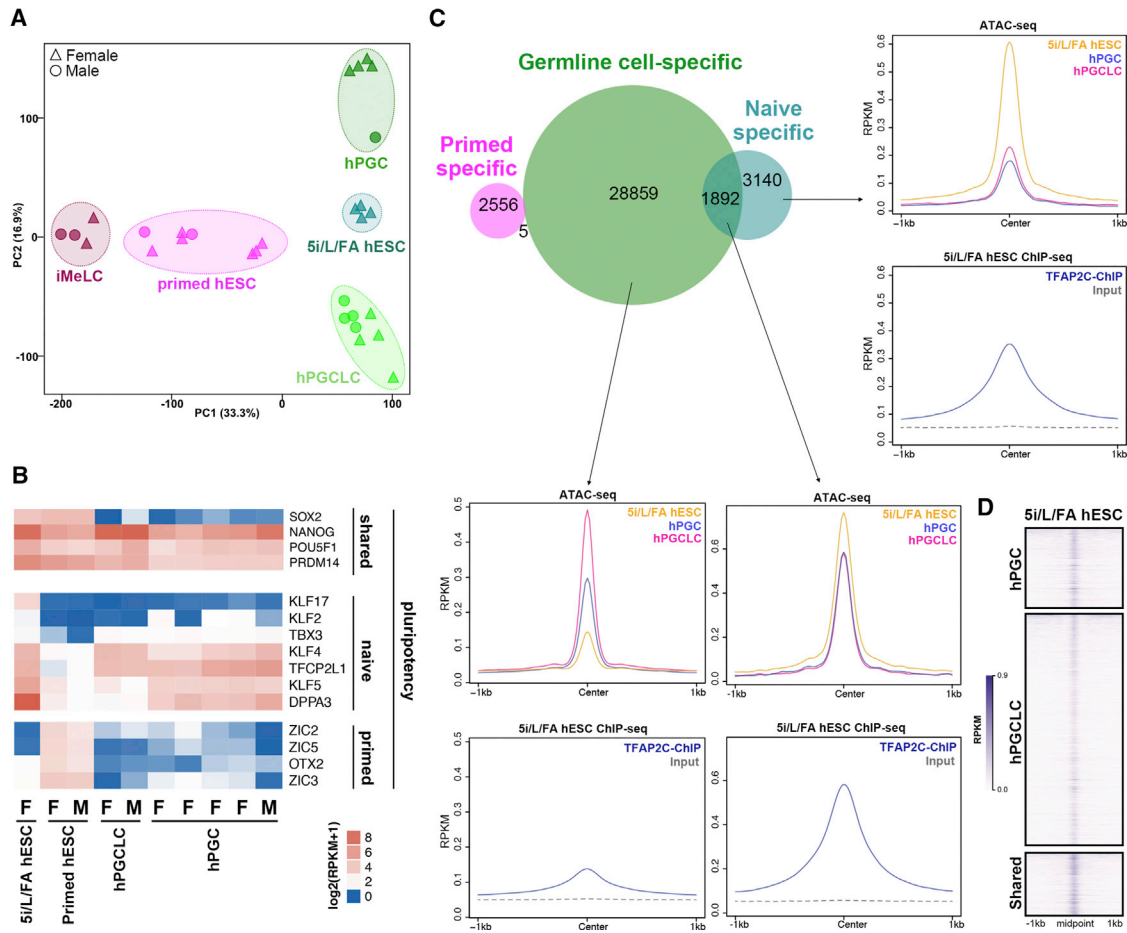


Figure 3. Reacquisition of Ground-State Naive Pluripotency in Human Germline Cells

(A) Principal component analysis (PCA) of transcriptomes of ground state naive hESCs cultured in 5i/L/FA media, primed hESCs, iMeLCs, hPGCLCs, and hPGCs. Gene expression analysis was based on the RNA-seq data from Pastor et al. (2016) (5i/L/FA ground-state naive hESCs) and Chen et al. (2017) (primed hESCs, iMeLCs, hPGCLCs, and hPGCs).

(B) Heatmap showing the expression of pluripotency genes in 5i/L/FA ground-state naive hESCs, primed hESCs, hPGCLCs, and hPGCs. The five hPGCs samples are 89d female, 103d female, 89d female, 89d female, and 59d male from left to right. F, female; M, male.

(C) Venn diagram showing the overlap of germline cell-specific ATAC-seq regions with naive-specific and primed-specific regions identified by Pastor et al. (2018). Metaplot of the ATAC-seq signals in 5i/L/FA ground-state naive hESCs, hPGCLCs, and hPGCs and TFAP2C ChIP-seq signals in naive hESCs over regions defined from the Venn diagram.

(D) Heatmap showing the ground-state naive hESCs ATAC-seq signals (Pastor et al., 2018) over hPGC-specific, hPGCLC-specific, and hPGC/hPGCLC-shared peaks.

See also Figure S3.

important to highlight that not all components of the naive pluripotent ground-state transcriptional program are reacquired by hPGCLCs or hPGCs (Figure 3B). For example, the naive ground-state transcription factor KLF17 (Guo et al., 2016) is not expressed in either hPGCLCs or hPGCs, and the common pluripotent transcription factor SOX2 is extinguished in human germline cells but is highly expressed in the naive state (Figure 3B). These observations indicate that when differentiated from primed pluripotent stem cells *in vitro* and also during differentiation *in vivo*, hPGCLCs and hPGCs acquire a transcriptome profile that shifts toward ground-state naive pluripotency.

TFAP2C functions in naive hPSCs to regulate the opening of naive-specific enhancers, and this opening is required for the

establishment and maintenance of ground-state naive pluripotent stem cells in 5i/L/FA and 2iLGo (Pastor et al., 2018). To determine which of the naive-specific enhancers are open in hPGCLCs and hPGCs, we compared the 30,751 total unique peaks in hPGCLCs/hPGCs to the ATAC-seq peaks previously defined (Pastor et al., 2018) as specific to the ground state naive (5,032) or the primed state (2,561) of pluripotency (Figure 3C). This analysis first shows that naive-specific ATAC-seq peaks exhibit more overlap with the open chromatin of hPGCs and hPGCLCs than primed-specific peaks (Figure 3C). Specifically, 38% (1,892 out of 5,032; $p < 0.05$, hypergeometric test) of ground-state naive specific peaks overlapped with germline cell-specific open chromatin whereas only 5 out of 2,561

primed-specific peaks overlapped (Figures 3C, S3A, and S3B). Using chromatin immunoprecipitation sequencing (ChIP-seq) for TFAP2C in 5i/L/FA cultured naive hPSCs (Pastor et al., 2018), we found that the 1,892 hPGC and hPGCLC overlapping peaks are also bound by TFAP2C in 5i/L/FA cultured cells (Figure 3C). Repeating this analysis with a subset of naive-specific enhancers (1,560) that are directly bound by TFAP2C and whose opening in primed to naive reversion functionally depends upon TFAP2C (Pastor et al., 2018), we discovered that the overlap now increases to 51% of naive specific peaks (800 out of 1,560; $p < 0.05$, hypergeometric test) (Figure S3C). Thus, we conclude that a significant fraction of naive-specific peaks previously defined as being enriched in enhancer epigenetic marks re-open in hPGCLCs differentiated from primed hESCs and are also maintained in hPGCs isolated from the embryo.

Analysis of the 1,892 peaks enriched in the naive-specific and germline cell-specific intersect group revealed a significant enrichment of AP2 motifs, as expected (Figure S3B). However, we also noted an even greater statistical enrichment of AP2 motifs in the germline cell-specific open chromatin outside of the ground-state naive-specific enhancers (Figure S3B), suggesting that AP2 transcription factors play additional roles outside of regulating naive-specific enhancers. To address this, we compared all ATAC-seq peaks identified in 5i/L/FA ground-state naive UCLA1 hESCs to the 30,751 total peaks identified in hPGCLCs and hPGCs. This analysis revealed that the regions of open chromatin unique to hPGCs and hPGCLCs are also open in 5i/L/FA cultured ground-state naive pluripotent stem cells (Figure 3D). Collectively, these data suggest that AP2 sites and most likely TFAP2C are significantly enriched in the open chromatin of hPGCLCs and hPGCs, and that a large fraction of previously defined naive-specific enhancers that are bound by TFAP2C in the ground-state naive pluripotency are also open in the chromatin of hPGCLCs and hPGCs.

TFAP2C Is Required for hPGCLC Induction and Expression of the Ground-State Naive Pluripotency Marker KLF4

Previous studies established that TFAP2C is required for hPGCLC differentiation from hiPSCs (Kojima et al., 2017). To confirm this result in hESCs, we used two independent mutant clones of *TFAP2C*^{-/-} UCLA1 hESCs generated by CRISPR/Cas9 (Pastor et al., 2018) that do not generate TFAP2C protein (Figure S4A). Using these null mutant hESC lines, we found that hPGCLC formation is completely ablated relative to controls (Figures 4A and 4B). To determine if TFAP2C is required for somatic cell differentiation, we injected control and *TFAP2C*^{-/-} hESCs lines into immune-deficient mice to create teratomas and discovered that *TFAP2C*^{-/-} hESCs from both null mutant hESC subclones are capable of teratoma formation similar to controls (Figure S4B). This suggests that TFAP2C is not necessary for exit from primed pluripotency and somatic cell differentiation per se, instead having a specific effect on the specification of hPGCLCs.

Given the discovery that the transcriptome and open chromatin of hPGCLCs and hPGCs resembles ground-state naive pluripotent stem cells, we next tracked the expression of the ground-state naive pluripotent transcription factor KLF4 together

with OCT4 as a marker of putative hPGCLCs (Figures 4C and S4C). In control aggregates, most cells are OCT4 positive at day 1 of differentiation, whereas KLF4 protein was not expressed by any cells within the aggregates. By day 2, KLF4 expression is detected in a subset of OCT4-positive cells, whereas at days 3 and 4 of aggregate differentiation, OCT4 and KLF4 protein expression almost completely overlap. In contrast in the *TFAP2C*^{-/-} aggregates, KLF4 protein was never expressed, and OCT4 protein diminished to near background levels between days 2 and 3 (Figures 4C and S4C). Taken together, our data suggest that in the absence of TFAP2C, the ground-state naive pluripotent transcription factor KLF4 is not expressed during aggregate differentiation, and therefore, we were next interested in whether KLF4 could be a target of TFAP2C.

To address this, we analyzed the ATAC-seq signals around the *KLF4* locus in hPGCLCs and hPGCs to identify differentially open peaks that also contain AP2 sites. We identified a new peak of open chromatin ~50 kb distal to the *KLF4* locus, which is open in hPGCLCs and hPGCs, but not hESCs, iMeLCs, or somatic cells (Figures 4D and 4E). We termed this region the KLF4 element (KE). To determine whether TFAP2C binds to the KE, we carried out ChIP-qPCR with anti-TFAP2C antibodies using day 4 aggregates generated from the H1 hESC line and discovered that TFAP2C binds to the KE (Figure 4F), thus suggesting that TFAP2C may act in part to regulate the expression of KLF4 during aggregate differentiation.

TFAP2C Regulates Germline Cell Formation Partially through the OCT4 NE

Given the unique extinction of OCT4 protein expression between days 2 and 3 of aggregate differentiation in *TFAP2C*^{-/-} cells (Figures 4C and S4C), we were next interested in evaluating OCT4 expression during aggregate differentiation and how OCT4 correlates with the specification of hPGCLCs. To achieve this, we utilized the H1 hESC line in which an *IRES-GFP* reporter is knocked into the coding region of *OCT4* before the stop codon using homologous recombination (Gkoutela et al., 2013). We call this hESC line OCT4-GFP. Using this tool, we show that GFP^{bright} cells are localized exclusively to the ITGA6/EPCAM double-positive hPGCLC population at day 4 of aggregate differentiation, whereas ITGA6/EPCAM^{dim/-} cells are GFP negative (Figure 5A). Using this gate as a reference, we tracked the emergence of GFP^{bright} hPGCLCs over the first 4 days of aggregate differentiation and show that GFP^{bright} hPGCLCs emerge between days 2.5 and 3.0 (Figures S5A and 5B). The GFP^{bright} hPGCLCs can subsequently be maintained for at least 8 days of aggregate differentiation (Figure S5A). The switch from dim to bright between days 2 and 3 of aggregate differentiation and the loss of OCT4 protein in *TFAP2C*^{-/-} aggregates within the same time frame (Figures 4C and S4C) may suggest that the regulation of OCT4 gene expression changes as germline cells are specified in the aggregates.

Previous studies using the mouse as a model discovered that the *Oct4* locus (also called *Pou5f1*) is regulated by alternate enhancers. Specifically, the *Oct4* distal enhancer (DE) regulates *Oct4* expression in the mouse inner cell mass (ICM) and mouse PGCs (mPGCs), whereas the *Oct4* proximal enhancer (PE) regulates *Oct4* expression in the post-implantation epiblast (Choi

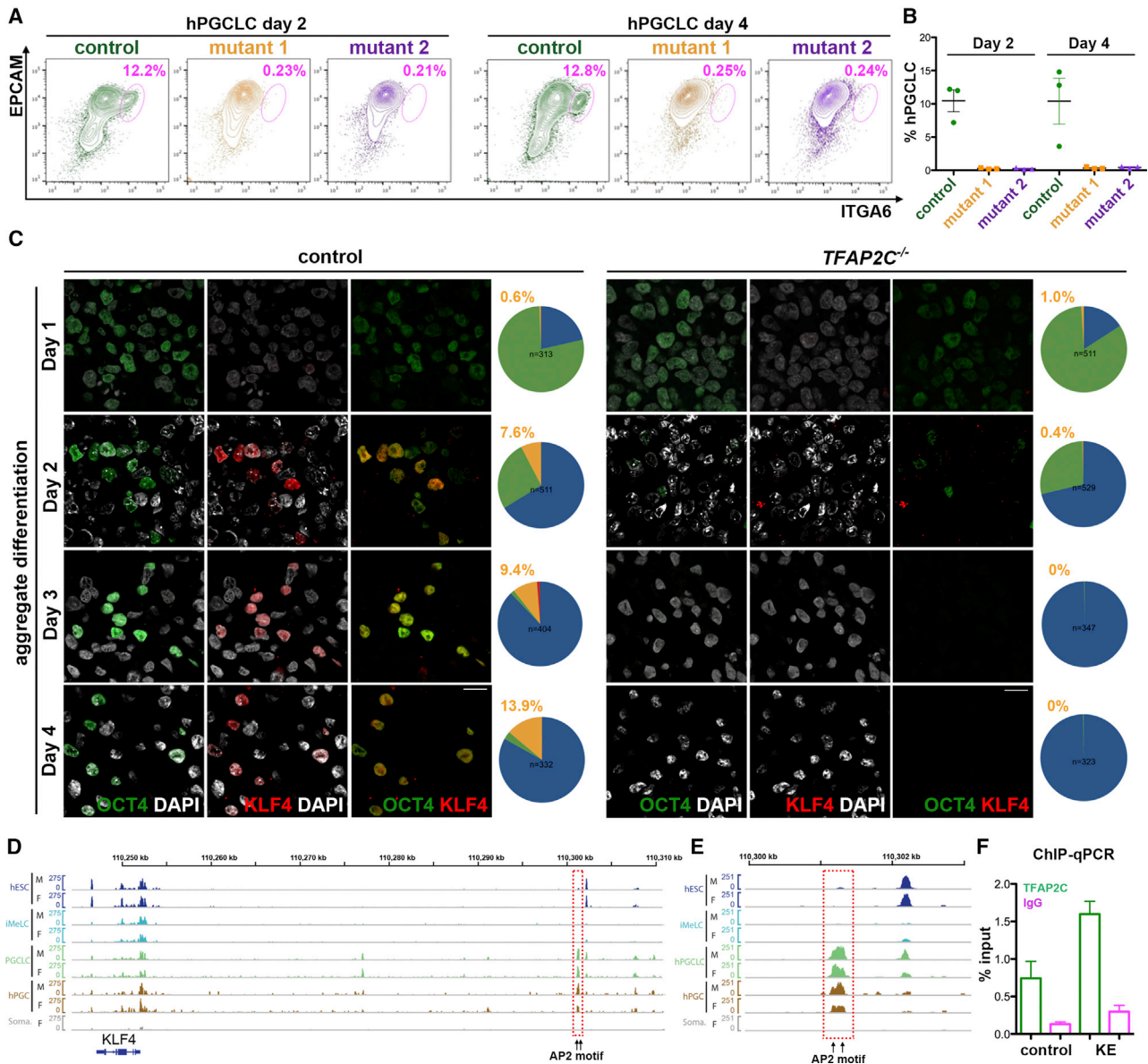


Figure 4. TFAP2C Is Required for hPGCLC Specification and Expression of the Naive Transcription Factor KLF4

(A) Flow cytometry showing the induction of hPGCLCs at days 2 and 4 of aggregate differentiation using control and *TFAP2C* mutant hESCs.

(B) Two independent *TFAP2C* mutant lines made from UCLA1 hESCs were used. hPGCLCs correspond to ITGA6/EPCAM double-positive cells. Three biological replicates were examined. Error bars represent SEM.

(C) Expression of KLF4 in OCT4-positive hPGCLCs from day 1 through day 4 of aggregate differentiation in control and *TFAP2C* mutant UCLA1 hESC lines. The percentages of OCT4 and KLF4 double-positive cells are quantified at each stage and are represented by the orange color in the pie chart. Green, OCT4 single-positive cells; red, KLF4 single-positive cells; blue, DAPI-positive cells but negative for OCT4 and KLF4. Scale bars, 15 μ m. The counting of all cell types is shown in Figure S4C.

(D and E) Screenshot of the ATAC-seq signal near *KLF4* (D). Red dashed box indicates a putative DNA regulatory element, which is closed in primed hESCs and iMeLCs but open in hPGCLCs and hPGCs. This DNA region is termed as the KLF4 element (KE). The germline cell-specific KE region with two AP2-binding sites (indicated by black arrows) is highlighted in (E).

(F) ChIP-qPCR of KE using anti-*TFAP2C* antibodies in day 4 aggregates from UCLA1. Immunoglobulin G (IgG) was used as ChIP control. Two biological replicates for ChIP and two technical replicates for qPCR were performed. Control is a genomic region at the *OCT4* locus without AP2-binding site. Error bars represent SEM.

See also Figure S4.

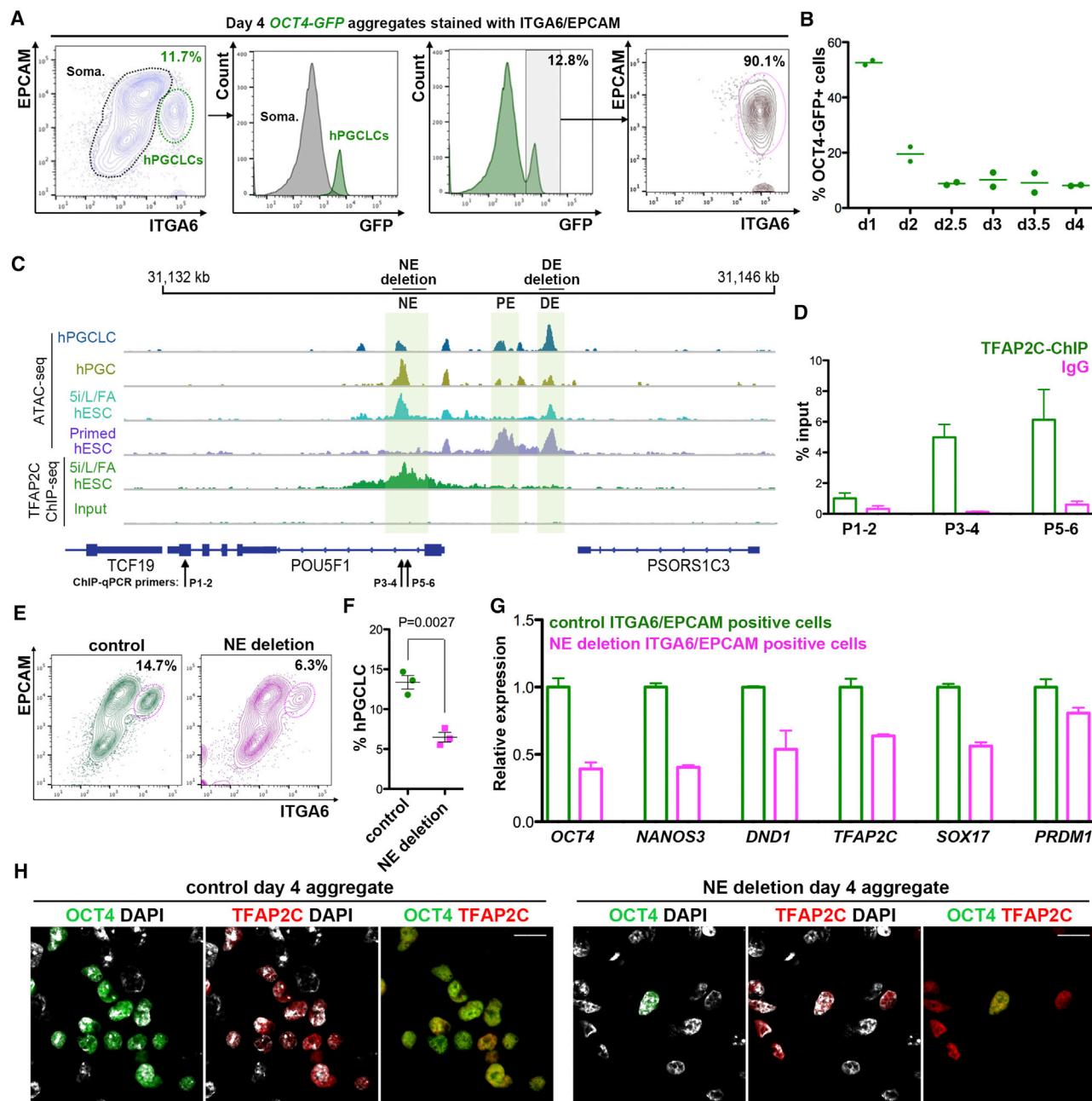


Figure 5. The *OCT4* NE Is Involved in hPGCLC Formation

(A) Flow cytometry of aggregates at day 4 of differentiation from H1 hESC line genetically modified to express GFP from the *OCT4* locus (*OCT4-GFP*). hPGCLCs (ITGA6/EPCAM double-positive cells) are also positive for GFP. In contrast, non-hPGCLCs are GFP negative. Most *OCT4-GFP*-positive cells are positive for ITGA6/EPCAM. Three biological replicates were performed.

(B) Summary of *OCT4-GFP*-positive cells during aggregate differentiation from days 1 to 4. The GFP-positive gate was set according to the GFP gate from (A). Two biological replicates were performed.

(C) Screenshot of ATAC-seq and TFAP2C ChIP-seq signals showing three enhancers at the *POU5F1* locus (encoding *OCT4*). Shaded boxes highlight the naive enhancer (NE), proximal enhancer (PE), and distal enhancer (DE) at the *POU5F1* locus. DE deletion and NE deletion indicate genomic regions that were deleted by CRISPR/Cas9-mediated genome editing. Primers for ChIP-qPCR (P1-2, P3-4, P5-6) of NE and control regions are shown.

(D) ChIP-qPCR using anti-TFAP2C antibodies in day 4 aggregates from the UCLA1 line. IgG was used as a ChIP control. Two biological replicates for ChIP and two technical replicates for qPCR were performed. Primer (P) locations at the *POU5F1* locus are shown in (C). Error bars represent SEM.

(E) Flow cytometry of control and *OCT4* NE deletion day 4 aggregates from the UCLA1 hESCs. hPGCLCs correspond to the ITGA6/EPCAM double-positive cells ($n = 3$ biological replicates).

(F) Quantification of hPGCLC percentages from (E). t test was applied. Error bars represent SEM.

(legend continued on next page)

et al., 2016; Ovitt and Schöler, 1998; Yeom et al., 1996). In hESCs, deleting the PE while simultaneously targeting the *OCT4* locus with a GFP reporter is a successful strategy for identifying ground-state naive pluripotent stem cells cultured in 5i/L/FA (Theunissen et al., 2014). Recently the NE was identified at the *OCT4* locus, which is bound by TFAP2C. This enhancer is not found in mouse naive ESCs cultured in 2i + LIF (Pastor et al., 2018). Critically, this NE is required to establish the ground state naive pluripotency from primed pluripotent stem cells (Pastor et al., 2018). Based on this result, we hypothesize that the TFAP2C-bound NE may also be involved in regulating *OCT4* expression during human germline cell development.

First, to determine whether the NE is also open in human germline cells, we compared ATAC-seq peaks for the NE in hPGCLCs and hPGCs to this region in naive and primed hESCs (Figure 5C). These data show that the PE, DE, and NE are all open in hPGCLCs at day 4 of aggregate differentiation, with the NE being more open in hPGCs than the DE and PE (Figure 5C). This observation raises the possibility that restriction of *OCT4* expression between days 2 and 3 of aggregate differentiation may be due to NE and/or the DE enhancer activation at the *OCT4* locus. Given that the NE is bound by TFAP2C in ground-state naive pluripotent stem cells, whereas the DE is not (Figure 5C), we next confirmed that the *OCT4* NE is also a target of TFAP2C during hPGCLC differentiation. To do this, we performed ChIP-qPCR on day 4 aggregates containing hPGCLCs and discovered that TFAP2C is bound to the NE, whereas it is not bound to a genomic region that does not contain AP2 sites (Figure 5D). Taken together, these experiments suggest that *OCT4* regulation during hPGCLC differentiation may involve enhancer activation at the DE as well as TFAP2C-bound enhancer activation at the NE.

In order to evaluate the role of the DE in hPGCLC formation, we generated an *OCT4* DE deletion in the UCLA1 hESC line and discovered that deleting the DE had no effect on hPGCLC differentiation (Figures S5B and S5C). To examine hPGCLC identity in the ITGA6/EPCAM double-positive hPGCLC population, we examined germline cell gene expression by real-time PCR and found comparable expression between mutant and control hPGCLCs isolated by FACS (Figure S5D). This result suggests that the DE is not a major regulatory feature involved in the specification of hPGCLCs.

In order to evaluate the role of the NE in hPGCLC specification, we first used the NE-deletion mutant generated in the UCLA1 hESC line using CRISPR/Cas9 (Figure 5C) (Pastor et al., 2018). Using the NE-deletion and control hESCs, we discovered that the percentage of hPGCLCs at day 2 of aggregate differentiation was comparable in both groups (Figure S5E), whereas the percentage of hPGCLCs was significantly decreased relative to controls at day 4 (Figures 5E and 5F). To confirm the mutant phenotype in another hESC line, we made NE-deletion mutants in the H1 hESC line. Consistently, we discovered that the per-

centage of hPGCLCs was also decreased at day 4 of aggregate differentiation in the NE deletion relative to control (Figures S5F and S5G). To evaluate germline identity, we collected control and NE-deletion hPGCLCs using FACS for ITGA6/EPCAM and examined germline cell gene expression by real-time PCR at day 4. The result shows that *OCT4* RNA expression is decreased by approximately half in the NE-deletion hPGCLCs at day 4, and this was accompanied by a decrease in the expression of diagnostic germ cell genes *NANOS3*, *DND1*, *TFAP2C*, *SOX17*, and *PRDM1* (Figure 5G). We also performed immunofluorescence and found that *OCT4*-positive cells were reduced in the NE-deletion aggregates at day 4 of differentiation, whereas TFAP2C single-positive cells were still detected in the aggregates (Figure 5H). Collectively, these data suggest that one of the mechanisms by which TFAP2C regulates human germline cell formation is opening naive-specific enhancers, with one of these enhancers corresponding to the NE at the *OCT4* locus.

DISCUSSION

Human germline cell specification is a critical biological event during early embryogenesis and is absolutely required for generating functional gametes at reproductive age. Although several transcription factors have been identified in the induction of germline cell fate in humans, systematic identification of transcription factors governing the precise stepwise sequence of events in human germline cell formation is lacking. In the current study, we identified and analyzed open chromatin specific to prenatal human germline cells using ATAC-seq and RNA-seq. Notably, SOX family transcription-factor-binding motifs were identified consistent with critical function of *SOX17* for human germline cell specification (Irie et al., 2015). In this study, we focused on the function of TFAP2C, which is a member of the AP2 family. The function of other transcription factor families, such as KLF and GATA, requires further analysis.

TFAP2C is required for hPGCLC formation, but the mechanism by which TFAP2C regulates hPGCLC development remains largely unknown (Kojima et al., 2017). In the mouse, *Tfap2c* functions downstream of *Prdm1* to reinforce germline cell identity after PGC specification (Magnúsdóttir et al., 2013; Magnúsdóttir and Surani, 2014; Nakaki et al., 2013). Using the mPGCLC differentiation model, *Tfap2c*^{-/-} mESCs are able to generate mPGCLCs; however, the resulting transcriptional program of the *Tfap2c*^{-/-} mPGCLCs is abnormal, including reduced expression of germline cell genes and upregulation of somatic cell genes (Weber et al., 2010). The role of TFAP2C in human germline cell formation was recently addressed by Kojima et al. (2017), and TFAP2C was shown to be required for germline cell formation. However, overexpression of TFAP2C was not sufficient to induce PGC formation in the absence of growth factors (Kobayashi et al., 2017). Although many genes are abnormally expressed in *PRDM1*-reporter-positive aggregate cells in the

(G) Expression of germ cell genes in ITGA6/EPCAM double-positive hPGCLCs from control and *OCT4* NE deletion samples at day 4 of aggregate differentiation from UCLA1 hESCs (n = 2 biological replicates). Error bars represent SEM.

(H) Immunofluorescence of *OCT4* (green) and TFAP2C (red) in control and *OCT4* NE deletion aggregates at day 4 of differentiation from UCLA1 hESCs (n = 2 biological replicates). hPGCLCs correspond to *OCT4*/TFAP2C double-positive cells. Scale bars, 15 μ m.

See also Figure S5.

absence of TFAP2C, the mechanism by which TFAP2C regulates hPGC formation remains to be elucidated.

In a recent study, we showed that TFAP2C is required to open and maintain naive-specific enhancers during reversion from primed- to ground-state naive pluripotency (Pastor et al., 2018). In the current study, we discovered that 38% of enhancers identified as being naive specific and 51% of the TFAP2C-dependent naive-specific enhancers are open in human germline cells, suggesting that one of the roles for TFAP2C in the germline may be to promote the expression of genes traditionally associated with naive ground-state pluripotency. Outside of the naive-specific enhancers, we also show that AP2 motifs are enriched in thousands of additional regions of germline-specific open chromatin in the genome, suggesting that AP2 family members and TFAP2C play a complex and critical role in human germline cell development.

The overlap of TFAP2C-bound naive-specific enhancers with the earliest stages of human germline cell development as modeled with hPGCLCs raises the question of why human germline cells acquire a transcriptome and open chromatin state resembling ground-state naive pluripotency when differentiated from the primed state of pluripotency. Ground-state naive pluripotency *in vitro* in mouse and humans is also associated with a globally demethylated genome (Leitch et al., 2013a; Pastor et al., 2016; Takashima et al., 2014), and hPGCs *in vivo* are globally demethylated (Gkoutela et al., 2015; Tang et al., 2015). During early embryogenesis in mammals, there are two waves of DNA methylation reprogramming: one soon after fertilization and the other in PGCs (Monk, 2015). Epigenetic reprogramming is hypothesized to remove epialleles acquired in the preceding developmental events (e.g., epialleles that are generated during gametogenesis prior to fertilization and epialleles that are formed with primed epiblast differentiation prior to gastrulation). The acquisition of ground-state naive pluripotency during both stages of global DNA methylation reprogramming may be a checkpoint that enables germline development to progress. It is interesting that in the hPGCLC model, complete DNA methylation reprogramming has not been established by day 4 of aggregate differentiation (Irie et al., 2015; Sasaki et al., 2015; von Meyenn et al., 2016), yet the ground-state naive pluripotency marker KLF4 is expressed in hPGCLCs. This suggests that the switch from a primed pluripotency state toward one that resembles the naive ground state in human germline cells precedes DNA methylation reprogramming. Another possibility is that once germline cells acquire transcriptome and chromatin state resembling ground-state naive pluripotency, they are protected from differentiation cues so as to maintain germline cell identity. This finding is supported by the observation that human ground-state naive pluripotent stem cells do not easily respond to differentiation cues in order to generate teratomas in immunocompromised mice and require re-priming to differentiate effectively into embryoid bodies (Liu et al., 2017).

A major question for the function of TFAP2C during human germline cell formation is how TFAP2C mediates the switch of primed pluripotency to a state resembling ground-state naive. One mechanism reported here is to regulate the NE activation at the *OCT4* locus between days 2 and 3 of hPGCLC differentiation. In the mouse, the *Oct4* locus has two well-characterized

enhancers referred to as the DE and PE. Using transgenic mice, it was revealed that the DE was utilized for *Oct4* expression in the ICM and PGCs, whereas the PE was utilized for *Oct4* expression in the mouse post-implantation epiblast (Choi et al., 2016; Ovitt and Schöler, 1998; Yeom et al., 1996). There is no TFAP2C-bound NE enhancer in rodents, and the genomic sequence in this region is less conserved (Pastor et al., 2018). In the current study, we show that the NE functions to regulate *OCT4* expression between days 2 and 3 of aggregate differentiation and that deletion of the NE affects the maintenance of human germline cell identity. Given that the NE at the human *OCT4* locus contains three AP2 sites and a KLF site, it is possible that the regulation of the NE involves the combinatorial binding of both TFAP2C and possibly a KLF family member, although this remains to be determined.

Taken together, our data suggest a model for hPGC formation where TFAP2C functions to regulate the transition from primed-state pluripotency to a pluripotent state resembling ground-state naive by regulating the opening of naive-specific enhancers, with one functional example being the NE at the *OCT4* locus (Figure S6). Given that the naive-specific enhancers make up only a small fraction of human germline cell open chromatin enriched in AP2 motifs, our data indicate that a large number of other transcriptional units will also be regulated by TFAP2C and that these may also have critical roles in human germline cell development. Future studies will be critical to determine these other roles for TFAP2C.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Human fetal samples
 - Human ESC culture
- METHODS DETAILS
 - hPGCLC induction
 - Flow cytometry and fluorescence activated cell sorting
 - ATAC-seq
 - Real-time quantitative PCR
 - hESC mutants made by CRISPR/Cas9
 - Generation of teratomas
 - Immunofluorescence
 - Western blot
 - ChIP-qPCR
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Experimental design
 - Replicates and data pooling
 - RNA-seq data analysis
 - ATAC-seq data analysis
 - ChIP-seq data analysis
 - Motif Analysis
 - Principal component analysis (Figure 3A)
 - Code Availability
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and three tables and can be found with this article online at <https://doi.org/10.1016/j.celrep.2018.12.011>.

ACKNOWLEDGMENTS

The authors would like to thank Felicia Codrea, Jessica Scholes, and Jeffery Calimlim for FACS, Jinghua Tang for banking and culturing of the UCLA hESC lines, and Steven Peckman from the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research (BSCRC) for critical assistance with human subject and embryonic stem cell review. This work is supported by the NIH/NICHD (grant R01 HD079546 to A.T.C.). D.C. and W.A.P. acknowledge the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research for supporting the UCLA Training Program. Human fetal tissue research is supported by an NIH grant (5R24HD000836-53) to Ian Glass at the University of Washington Birth Defects laboratory. Human conceptus tissue requests can be sent to bdrl@u.washington.edu. S.E.J. is an investigator of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

D.C., W.L., and A.T.C. designed the experiments. D.C. and J.Z. conducted the experiments. W.L. analyzed the data. L.H., R.K., and J.G. conducted the teratoma assay. M.A. conducted western blot. W.P., J.H., and D.C. designed and made the *TFAP2C* mutant and *OCT4* enhancer deletion hESC lines. D.C., W.L., S.E.J., and A.T.C. interpreted the data and D.C., W.L., and A.T.C. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 12, 2018

Revised: November 15, 2018

Accepted: December 3, 2018

Published: December 26, 2018

REFERENCES

Adey, A., Morrison, H.G., Asan, X., Kitzman, J.O., Turner, E.H., Stackhouse, B., MacKenzie, A.P., Caruccio, N.C., Zhang, X., and Shendure, J. (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.* **11**, R119.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169.

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

Chen, D., Liu, W., Lukianchikov, A., Hancock, G.V., Zimmerman, J., Lowe, M.G., Kim, R., Galic, Z., Irie, N., Surani, M.A., et al. (2017). Germline competency of human embryonic stem cells depends on eomesodermin. *Biol. Reprod.* **97**, 850–861.

Choi, H.W., Joo, J.Y., Hong, Y.J., Kim, J.S., Song, H., Lee, J.W., Wu, G., Schöler, H.R., and Do, J.T. (2016). Distinct enhancer activity of Oct4 in naive and primed mouse pluripotency. *Stem Cell Reports* **7**, 911–926.

Clark, A.T., Gkoutela, S., Chen, D., Liu, W., Sosa, E., Sukhwani, M., Hennebold, J.D., and Orwig, K.E. (2017). Primate primordial germ cells acquire trans-plantation potential by Carnegie stage 23. *Stem Cell Reports* **9**, 329–341.

Diaz Perez, S.V., Kim, R., Li, Z., Marquez, V.E., Patel, S., Plath, K., and Clark, A.T. (2012). Derivation of new human embryonic stem cell lines reveals rapid epigenetic progression in vitro that can be prevented by chemical modification of chromatin. *Hum. Mol. Genet.* **21**, 751–764.

Extavour, C.G., and Akam, M. (2003). Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* **130**, 5869–5884.

Favorov, A., Mularoni, L., Cope, L.M., Medvedeva, Y., Mironov, A.A., Makeev, V.J., and Wheelan, S.J. (2012). Exploring massive, genome scale datasets with the GenometriCorr package. *PLoS Comput. Biol.* **8**, e1002529.

Gkoutela, S., Li, Z., Vincent, J.J., Zhang, K.X., Chen, A., Pellegrini, M., and Clark, A.T. (2013). The ontogeny of cKIT+ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation. *Nat. Cell Biol.* **15**, 113–122.

Gkoutela, 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.

Gu, W., Gaeta, X., Sahakyan, A., Chan, A.B., Hong, C.S., Kim, R., Braas, D., Plath, K., Lowry, W.E., and Christofk, H.R. (2016). Glycolytic metabolism plays a functional role in regulating human pluripotent stem cell state. *Cell Stem Cell* **19**, 476–490.

Guo, G., von Meyenn, F., Santos, F., Chen, Y., Reik, W., Bertone, P., Smith, A., and Nichols, J. (2016). Naive pluripotent stem cells derived directly from isolated cells of the human inner cell mass. *Stem Cell Reports* **6**, 437–446.

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532.

Hayashi, K., Ogushi, S., Kurimoto, K., Shimamoto, S., Ohta, H., and Saitou, M. (2012). Offspring from oocytes derived from in vitro primordial germ cell-like cells in mice. *Science* **338**, 971–975.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589.

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.

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

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.

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

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. (2013a). Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20**, 311–316.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.

Liu, X., Nefzger, C.M., Rossello, F.J., Chen, J., Knaupp, A.S., Firas, J., Ford, E., Pflueger, J., Paynter, J.M., Chy, H.S., et al. (2017). Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming. *Nat. Methods* **14**, 1055–1062.

Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* **141**, 245–252.

Magnúsdóttir, E., Dietmann, S., Murakami, K., Günesdogan, U., Tang, F., Bao, S., Diamanti, E., Lao, K., Gottgens, B., and Azim Surani, M. (2013). A tripertite

- p transcription factor network regulates primordial germ cell specification in mice.
- Nat. Cell Biol.*
- 15**
- , 905–915.
- Monk, D. (2015). Germline-derived DNA methylation and early embryo epigenetic reprogramming: The selected survival of imprints. *Int. J. Biochem. Cell Biol.* **67**, 128–138.
- Nakaki, F., Hayashi, K., Ohta, H., Kurimoto, K., Yabuta, Y., and Saitou, M. (2013). Induction of mouse germ-cell fate by transcription factors in vitro. *Nature* **501**, 222–226.
- Nakamura, T., and Extavour, C.G. (2016). The transcriptional repressor Blimp-1 acts downstream of BMP signaling to generate primordial germ cells in the cricket *Gryllus bimaculatus*. *Development* **143**, 255–263.
- Nakamura, 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.
- Ovitt, C.E., and Schöler, H.R. (1998). The molecular biology of Oct-4 in the early mouse embryo. *Mol. Hum. Reprod.* **4**, 1021–1031.
- Pastor, W.A., Chen, D., Liu, W., Kim, R., Sahakyan, A., Lukianchikov, A., Plath, K., Jacobsen, S.E., and Clark, A.T. (2016). Naive human pluripotent cells feature a methylation landscape devoid of blastocyst or germline memory. *Cell Stem Cell* **18**, 323–329.
- Pastor, W.A., Liu, W., Chen, D., Ho, J., Kim, R., Hunt, T.J., Lukianchikov, A., Liu, X., Polo, J.M., Jacobsen, S.E., and Clark, A.T. (2018). TFAP2C regulates transcription in human naive pluripotency by opening enhancers. *Nat. Cell Biol.* **20**, 553–564.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842.
- Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308.
- Raz, E. (2003). Primordial germ-cell development: the zebrafish perspective. *Nat. Rev. Genet.* **4**, 690–700.
- 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.
- Sasaki, K., Nakamura, T., Okamoto, I., Yabuta, Y., Iwatani, C., Tsuchiya, H., Seita, Y., Nakamura, S., Shiraki, N., Takakuwa, T., et al. (2016). The germ cell fate of cynomolgus monkeys is specified in the nascent amnion. *Dev. Cell* **39**, 169–185.
- Schwager, E.E., Meng, Y., and Extavour, C.G. (2015). vasa and piwi are required for mitotic integrity in early embryogenesis in the spider *Parasteatoda tepidariorum*. *Dev. Biol.* **402**, 276–290.
- Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BME Genomics* **15**, 284.
- Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* **158**, 1254–1269.
- 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.
- Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* **15**, 471–487.
- Theunissen, T.W., Friedli, M., He, Y., Planet, E., O’Neil, R.C., Markoulaki, S., Pontis, J., Wang, H., Iouranova, A., Imbeault, M., et al. (2016). Molecular criteria for defining the naive human pluripotent state. *Cell Stem Cell* **19**, 502–515.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111.
- von Meyenn, F., Berrens, R.V., Andrews, S., Santos, F., Collier, A.J., Krueger, F., Osorno, R., Dean, W., Rugg-Gunn, P.J., and Reik, W. (2016). Comparative principles of DNA methylation reprogramming during human and mouse in vitro primordial germ cell specification. *Dev. Cell* **39**, 104–115.
- Weber, S., Eckert, D., Nettersheim, D., Gillis, A.J.M., Schäfer, S., Kuckenberg, P., Ehlermann, J., Werling, U., Biermann, K., Looijenga, L.H.J., and Schorle, H. (2010). Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance. *Biol. Reprod.* **82**, 214–223.
- Weinberger, L., Ayyash, M., Novershtern, N., and Hanna, J.H. (2016). Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat. Rev. Mol. Cell Biol.* **17**, 155–169.
- Williamson, A., and Lehmann, R. (1996). Germ cell development in *Drosophila*. *Annu. Rev. Cell Dev. Biol.* **12**, 365–391.
- Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hübner, K., and Schöler, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881–894.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Goat polyclonal anti-human OCT3/4	Santa Cruz Biotechnology	Cat#sc-8628; RRID: AB_653551
Rabbit monoclonal anti-human PRDM1	Cell Signaling Technology	Cat#9115S; RRID: AB_2169699
Mouse monoclonal anti-human PRDM1	R&D Systems	Cat#MAB36081; RRID: AB_10718104
Rabbit polyclonal anti-human TFAP2C	Santa Cruz Biotechnology	Cat#sc-8977; RRID: AB_2286995
Mouse monoclonal anti-TFAP2C	Santa Cruz Biotechnology	Cat#sc-12762; RRID: AB_667770
Goat anti-human SOX17	Neuromics	Cat#GT15094; RRID: AB_2195648
Goat polyclonal anti-human KLF4	R&D Systems	Cat#AF3640; RRID: AB_2130224
BV421 conjugated anti-human/mouse CD49f (ITGA6)	BioLegend	Cat#313624; RRID: AB_2562244
488 conjugated anti-human/mouse CD49f (ITGA6)	BioLegend	Cat#313608; RRID: AB_493635
Alexa Fluor 488-conjugated anti-human CD326 (EPCAM)	BioLegend	Cat#324210; RRID: AB_756084
APC-conjugated anti-human CD326 (EPCAM)	BioLegend	324208; RRID: AB_756082
PE-conjugated mouse monoclonal anti-human TNAP	BD PharMingen	Cat#561433; RRID: AB_10645791
APC-conjugated mouse monoclonal anti-human cKIT	BD PharMingen	Cat#550412; RRID: AB_398461
Rabbit monoclonal anti-human TFAP2C	Abcam	Cat#Ab76007; RRID: AB_1309954
Mouse monoclonal anti-Histone H3	Abcam	Cat#Ab10799; RRID: AB_470239
Bacterial and Virus Strains		
Stbl3 Chemically Competent <i>E. coli</i>	ThermoFisher	Cat#C737303
Chemicals, Peptides, and Recombinant Proteins		
CHIR99021	Stemgent	Cat# 04-0004
Y27632	Stemgent	Cat# 04-0012-10
Recombinant Human FGF basic Protein	R&D systems	Cat#233-FB
Recombinant Activin A	Peptotech	Cat# AF-120-14E
Recombinant human LIF	Millipore	Cat# LIF1005
Recombinant human BMP4	R&D systems	Cat#314-BP
Recombinant human EGF	R&D systems	Cat#236-EG
Critical Commercial Assays		
P3 Primary Cell 4D-Nucleofector® X Kit	Lonza	Cat#V4XP-3024
TaqMan Universal PCR Master Mix	Applied Biosystems	Cat#4304437
Quick-DNA Microprep Kit	ZYMO RESEARCH	Cat#D3021
One Shot Stbl3 Chemically Competent <i>E. coli</i>	ThermoFisher	Cat#C737303
Deposited Data		
ATAC-seq data	This paper	GEO: GSE120648
Experimental Models: Cell Lines		
UCLA1	Diaz Perez et al., 2012	N/A
UCLA2	Diaz Perez et al., 2012	N/A
H1 [OCT4-IRES-GFP]	Gkountela et al., 2013	N/A
UCLA1 <i>TFAP2C</i> ^{-/-} line 1	Pastor et al., 2018	N/A
UCLA1 <i>TFAP2C</i> ^{-/-} line 2	Pastor et al., 2018	N/A
UCLA1 <i>OCT4</i> NE deletion	Pastor et al., 2018	N/A
UCLA1 <i>OCT4</i> DE deletion	This paper	N/A
H1 OCT4-GFP <i>OCT4</i> NE deletion	This paper	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
See Table S2 for the TaqMan probes used in this study	This paper	N/A
See Table S2 for the gRNAs used in this study	This paper	N/A
See Table S2 for the primers used in this study	This paper	N/A
Recombinant DNA		
pSpCas9(BB)-2A-Puro (PX459) V2.0	Ran et al., 2013	Addgene, Cat#62988
Software and Algorithms		
BD FACSDIVA software	BD Biosciences	N/A
FLOWJO software	FLOWJO	N/A
GraphPad Prism software	GraphPad	N/A
Photoshop software	Adobe	N/A
Illustrator software	Adobe	N/A
R	https://www.R-project.org	N/A
HOMER Motif Analysis	http://homer.ucsd.edu/homer/motif/	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Amander Clark (clarka@ucla.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human fetal samples

Human prenatal testes and ovaries were acquired following elected termination and pathological evaluation after UCLA-IRB review which deemed the project exempt under 45 CRF 46.102(f). All prenatal gonads were obtained from the University of Washington Birth Defects Research Laboratory (BDRL), under the regulatory oversight of the University of Washington IRB approved Human Subjects protocol combined with a Certificate of Confidentiality from the Federal Government. BDRL collected the fetal testes and ovaries and shipped them overnight in HBSS with ice pack for immediate processing in Los Angeles. All consented material was donated anonymously and carried no personal identifiers. Developmental age was documented by BDRL as days post fertilization using a combination of prenatal intakes, foot length, Streeter's Stages and crown-rump length. All prenatal gonads documented with birth defect or chromosomal abnormality were excluded from this study.

Human ESC culture

Human ESC lines used in this study include UCLA1 (46, XX) ([Diaz Perez et al., 2012](#)), UCLA2 (46, XY) ([Diaz Perez et al., 2012](#)), and H1 OCT4-GFP (46, XY) ([Gkountela et al., 2013](#)). hESCs were cultured on mitomycin C-inactivated mouse embryonic fibroblasts (MEFs) in hESC media, which is composed of 20% knockout serum replacement (KSR) (GIBCO, 10828-028), 100 μ M L-Glutamine (GIBCO, 25030-081), 1x MEM Non-Essential Amino Acids (NEAA) (GIBCO, 11140-050), 55 μ M 2-Mercaptoethanol (GIBCO, 21985-023), 10ng/mL recombinant human FGF basic (R&D systems, 233-FB), 1x Penicillin-Streptomycin (GIBCO, 15140-122), and 50ng/mL primocin (InvivoGen, ant-pm-2) in DMEM/F12 media (GIBCO, 11330-032). The hESCs were split every 7 days using Collagenase type IV (GIBCO, 17104-019). All hESC lines used in this study are registered with the National Institute of Health Human Embryonic Stem Cell Registry and are available for research use with NIH funds. The hESC lines used in this study are listed in [Table S1](#). Mycoplasma test (Lonza, LT07-418) was performed every month to all cell lines used in this study. All experiments were approved by the UCLA Embryonic Stem Cell Research Oversight Committee.

METHODS DETAILS

hPGCLC induction

hPGCLCs were induced from primed hESCs as described before ([Chen et al., 2017](#)). hESCs were dissociated into single cells with 0.05% Trypsin-EDTA (GIBCO, 25300-054) and plated onto Human Plasma Fibronectin (Invitrogen, 33016-015)-coated 12-well-plates at the density of 200,000 cells/well in 2mL/well of iMeLC media, which is composed of 15% KSR (GIBCO, 10828-028), 1x NEAA (GIBCO, 11140-050), 0.1mM 2-Mercaptoethanol (GIBCO, 21985-023), 1x Penicillin-Streptomycin-Glutamine (GIBCO, 10378-016), 1mM sodium pyruvate (GIBCO, 11360-070), 50ng/mL Activin A (Peprotech, AF-120-14E), 3 μ M CHIR99021 (Stemgent,

04-0004), 10 μ M of ROCKi (Y27632, Stemgent, 04-0012-10), and 50ng/mL primocin in Glasgow's MEM (GMEM) (GIBCO, 11710-035). After 24 hr, iMeLCs were dissociated into single cells with 0.05% Trypsin-EDTA and plated into ultra-low cell attachment U-bottom 96-well plates (Corning, 7007) at the density of 3,000 cells/well in 200 μ L/well of hPGCLC media, which is composed of 15% KSR (GIBCO, 10828-028), 1x NEAA (GIBCO, 11140-050), 0.1mM 2-Mercaptoethanol (GIBCO, 21985-023), 1x Penicillin-Streptomycin-Glutamine (GIBCO, 10378-016), 1mM sodium pyruvate (GIBCO, 11360-070), 10ng/mL human LIF (Millipore, LIF1005), 200ng/mL human BMP4 (R&D systems, 314-BP), 50ng/mL human EGF (R&D systems, 236-EG), 10 μ M of ROCKi (Y27632, Stemgent, 04-0012-10), and 50ng/mL primocin in Glasgow's MEM (GMEM) (GIBCO, 11710-035).

Flow cytometry and fluorescence activated cell sorting

Human prenatal gonads or aggregates were dissociated with 0.25% trypsin (GIBCO, 25200-056) for 5 min or 0.05% Trypsin-EDTA (GIBCO, 25300-054) for 10 min at 37°C. The dissociated cells were stained with conjugated antibodies, washed with FACS buffer (1% BSA in PBS) and resuspended in FACS buffer with 7-AAD (BD PharMingen, 559925) as viability dye. The single cell suspension was analyzed or sorted for further experiments. The conjugated antibodies used in this study include: ITGA6 conjugated with BV421 (BioLegend, 313624, 1:60), ITGA6 conjugated with 488 (BioLegend, 313608, 1:60), EPCAM conjugated with 488 (BioLegend, 324210, 1:60), EPCAM conjugated with APC (BioLegend, 324208, 1:60), tissue nonspecific alkaline phosphatase (TNAP) conjugated with PE (BD PharMingen, 561433, 1:60), and cKIT conjugated with APC (BD PharMingen, 550412, 1:60).

ATAC-seq

ATAC-seq was performed using Nextera DNA library prep kit (Illumina, 15028212) as previously described (Pastor et al., 2018). Cells were collected in lysis buffer (10mM Tris pH7.4, 10mM NaCl, 3mM MgCl₂, 1% NP40) and spun at 500 g at 4°C for 10 min. The pellet was resuspended in the transposase reaction mix (25 μ L 2 \times Tagmentation buffer, 2.5 μ L transposase and 22.5 μ L nuclease-free water) and incubated at 37°C for 30 min. The samples were purified using MinElute PCR Purification Kit (QIAGEN, 28006) and amplified using 1 \times NEBnext PCR master mix (NEB, M0541S) and 1.25 μ M of custom Nextera PCR primers 1 and 2 with the following PCR conditions: 72°C for 5 min; 98°C for 30 s; and thermocycling at 98°C for 10 s, 63°C for 30 s and 72°C for 1 min. Samples were amplified for five cycles and 5 μ L of the PCR reaction was used to determine the required cycles of amplification by real-time PCR. The remaining 45 μ L reaction was amplified with the determined cycles and purified with MinElute PCR Purification Kit (QIAGEN, 28006) yielding a final library concentration of about 30 nM in 20 μ L. Libraries were subjected to pair-end 50bp sequencing on HiSeq 2000 or HiSeq 2500 sequencer with 4-6 indexed libraries per lane.

Real-time quantitative PCR

Sorted cells or cell pellets were lysed in 350 μ L of RLT buffer (QIAGEN) and RNA was extracted using RNeasy micro kit (QIAGEN, 74004). cDNA was synthesized using SuperScript® II Reverse Transcriptase (Invitrogen, 18064-014). Real-time quantitative PCR was performed using TaqMan® Universal PCR Master Mix (Applied Biosystems, 4304437) and the expression level of genes-of-interest was normalized to the expression of housekeeping gene *GAPDH*. The Taqman probes used in this study are listed in Table S2. Two biological replicates were performed for each experiment and two technical replicates were conducted for each biological replicate.

hESC mutants made by CRISPR/Cas9

To delete the OCT4 distal enhancer, a pair of guide RNA (gRNA) was designed using <http://zlab.bio/guide-design-resources> and cloned into PX459 vector (Ran et al., 2013). The OCT4 naive enhancer deletion was described previously (Pastor et al., 2018). 4 μ g of gRNA pair or 2 μ g of pMax-GFP was electroporated into 800,000 UCLA1 hES cells using P3 Primary Cell 4D-Nucleofector® X Kit according to the manufacturer's instructions (Lonza, V4XP-3024). 24 hr after nucleofection, cells were dissociated with Accutase (ThermoFisher Scientific, A1110501) and seeded on a 6-well-plate for 4 days before passaging into 10-cm²-dish in low density for screening. 96 individual colonies were picked after 9 days and expanded. Distal enhancer deletion candidate lines were screened for the presence of shorter bands due to deletion. To determine the precise mutations, genomic DNA was extracted from about 1 million cells using Quick-DNA Microprep Kit (ZYMO RESEARCH, D3021). 1 μ L of the genomic DNA was used as PCR template for genotyping using Phusion High-Fidelity DNA Polymerase (NEW ENGLAND BioLabs, M0530L). Genotyping primers were listed in Table S2. In order to characterize the mutant alleles, mutant bands were cloned into Blunt-PCR-Cloning vector using Zero Blunt PCR Cloning Kit (ThermoFisher, K270020). Ten white colonies were picked and sequenced to determine the precise deletion sites. The sequence of the pair of gRNA for deleting OCT4 distal enhancer are listed in Table S2.

Generation of teratomas

Surgery was performed following Institutional Approval for Appropriate Care and Use of Laboratory Animals by the UCLA Institutional Animal Care and Use Committee [Chancellor's Animal Research Committee (ARC)] (Animal Welfare Assurance number A3196-01). hESCs of control UCLA1, and OCT4 NE deletion mutant 1 and mutant 2 UCLA1 lines in matrigel (BD) were transplanted into the testis of adult SCID mice. Six to eight weeks after surgery, mice were euthanized and the tumors were removed and fixed in 4% formaldehyde overnight at room temperature. Tumor tissues were embedded in paraffin and cut as 5 μ m sections for hematoxylin and eosin staining.

Immunofluorescence

Slides of paraffin-embedded sections were deparaffinized by successive treatment with xylene and 100%, 95%, 70% and 50% ethanol. Antigen retrieval was performed by incubation with 10mM Tris pH 9.0, 1mM EDTA, 0.05% Tween at 95°C for 40 min. The slides were cooled and washed with 1xPBS (phosphate buffered saline) and 1xTBS (PBS + 0.2% Tween). The samples were permeabilized with 0.5% Triton X-100 in 1xPBS, then washed with 1xTBS and blocked with 10% donkey serum in 1xTBS. Primary antibody incubation was conducted overnight in 10% donkey serum. Samples were again washed with 1xTBS-tween and incubated with fluorescent secondary antibodies at 1:200 for 45 min, then washed and mounted using with ProLong Gold Antifade Mountant with DAPI (ThermoFisher). Images were taken using LSM 780 Confocal Instrument (Zeiss). The primary antibodies used for immunofluorescence in this study include: goat-anti-OCT4 (Santa Cruz Biotechnology, sc-8628x, 1:100), rabbit-anti-PRDM1 (Cell Signaling Technology, 9115, 1:100), mouse-anti-PRDM1 (R&D Systems, MAB36081, 1:100), rabbit-anti-TFAP2C (Santa Cruz Biotechnology, sc-8977, 1:100), mouse-anti-TFAP2C (Santa Cruz Biotechnology, sc12762, 1:100), goat-anti-SOX17 (Neuromics, GT15094, 1:100), KLF4 (R&D Systems, AF3640, 1:100). The secondary antibodies used in this study were all from Jackson ImmunoResearch Laboratories and DAPI was counterstained to indicate nuclei.

Western blot

Protein lysate of hESCs was prepared in RIPA buffer and NuPAGE® LDS sample loading buffer and denatured for 10 min at 98°C. Samples were run on an SDS- PAGE 10% Bis-Tris gel (ThermoFisher) 60–80 min at 100V, transferred at 100V for 60–70 min, and blocked with 5% non-fat dried milk in 0.1%TBST (0.1% Tween in Tris Buffer Saline solution) for 1 hr. Primary antibody was added to blocking buffer and incubated at room temperature for 1 hr. Secondary antibodies were added after washing with TBST. The Pierce™ ECL Western Blotting Substrate (ThermoScientific, 32109) was used on the membrane before film exposure. The primary antibodies used in this study include: rabbit-anti-TFAP2C (Abcam, 76007, 1:1000), mouse-anti-Histone 3 (Abcam, 10799, 1:8000). The secondary antibodies used include: Donkey-anti-goat HRP (Invitrogen, A15999, 1:2000), Sheep-anti-mouse HRP (GE Healthcare Life Sciences, NA931VS, 1:2000).

ChIP-qPCR

ChIP was performed as previously described (Pastor et al., 2018). Two biological replicates were performed using H1 OCT4-GFP hESC line and 4 plates of PGCLC aggregates were collected for each ChIP-qPCR. Aggregates were dissociated with 0.05% Trypsin-EDTA (GIBCO, 25300-054) for 10 min at 37°C, washed twice with PBS, fixed in 1% formaldehyde and flash frozen. After thawing, the cells were resuspended in 1 mL of Buffer 1 (10mM Tris-HCl pH8.0, 0.25% Triton X-100, 10mM EDTA, 0.5mM EGTA, 1x Protease Inhibitors (Roche), 1mM PMSF) and incubated at room temperature for 15 min on a rotator. Samples were spun at 4000 rpm for 5 min and the pellets were washed with 1 mL of Buffer 2 (10mM Tris-HCl pH8.0, 200mM NaCl, 10mM EDTA, 0.5mM EGTA, 1x Protease Inhibitors (Roche), 1mM PMSF) and resuspended in 650 μ L Buffer 3 (10mM Tris-HCl pH8.0, 10mM EDTA, 0.5mM EGTA, 1x Protease Inhibitors (Roche), 1mM PMSF) and sonicated with Covaris S2 with the following program: Intensity = 5; Cycles/burst = 200; Duty Cycle = 5%; 4 x (30" on/30" off/30" on/30" off). Sonicated lysate was centrifuged at 14,000rpm for 10 min at 4°C and the supernatant was collected into a new tube. 65 μ L of the supernatant was saved as input. 30 μ L Protein A Dynabeads (Invitrogen, 10001D) was washed with Dilution Buffer (16.7 mM Tris-HCl pH8.0, 0.01% SDS, 1.1% Triton X-100, 1.2mM EDTA, 167mM NaCl) three times and resuspended in 650 μ L Dilution Buffer and added to the samples. The samples with beads were rotated for 2 hr at 4°C and the beads were removed by magnetic rack. Each sample was split into two parts: one half for Rabbit-anti-TFAP2C (Santa Cruz Biotechnology, sc-8977), and the other half for Rabbit-IgG as control. Samples were incubated at 4°C overnight. On the second day, 60 μ L of pre-washed Protein A Dynabeads was added to each sample and incubated at 4°C for 2 hr. Samples were placed on magnetic rack to remove supernatant and the beads were washed twice with Buffer A (50mM HEPES pH7.9, 1% Triton X-100, 0.1% Deoxycholate, 1mM EDTA, 140mM NaCl), Buffer B (50mM HEPES pH7.9, 0.1% SDS, 1% Triton X-100, 0.1% Deoxycholate, 1mM EDTA, 500mM NaCl), TE and eluted with 150 μ L of Elution Buffer (50mM Tris-Cl pH8, 1mM EDTA, 1% SDS) and incubated at 65°C for 10 min. Samples were placed on magnetic rack and ChIP-samples were collected. Both ChIP-samples and input samples were heated overnight at 65°C. On the third day, samples were treated with 1.5 μ L of 10mg/mL RNaseA (PureLink RNase A, Invitrogen 12091-021) for 30 min at 65°C and then with 10 μ L of 10mg/mL Proteinase K for 2 hr at 56°C. Samples were purified with MinElute PCR Purification Kit (QIAGEN, 28006). qPCR was performed using TaqMan Universal PCR master mix (Applied Biosystems, 4304437). The primers for ChIP-qPCR are listed in Table S2.

QUANTIFICATION AND STATISTICAL ANALYSIS

Experimental design

ATAC-seq libraries were made from six biological replicates of human embryonic gonads and two biological replicates of male and female hESCs, iMeLCs, and hPGCLCs (except one for male hPGCLC). The assessment of the effect of gene knockout on hPGCLC induction was determined by using two independent knockout lines for each gene/enhancer (except for one line for OCT4 DE deletion) and parental control lines (UCLA1 or H1 OCT4-GFP). The hPGCLC induction from the knockout and control lines was performed in parallel, and was replicated three to six times (except one for day 2 aggregates induced from OCT4 NE deletion of UCLA1 hESC line). No statistical calculation was used to estimate the sample size. Randomization/stratification/blinding were not

applicable to these experiments. For realtime PCR, western blot and immunofluorescence, at least two biological replicates were performed.

Replicates and data pooling

To call ATAC-seq peaks or display of ATAC-seq data in figures, all reads from same condition were merged to increase coverage. RPKM from biological RNA-seq sample were merged for a given condition when comparing RPKM across different conditions (Figure 2B). For Principal component analysis (PCA), biological replicates for RNA-seq were considered separately (Figure 2B).

RNA-seq data analysis

For RNA-seq data in 5i/L/FA hESCs, previously published RNA-seq data (Pastor et al., 2016) were used in this study. For RNA-seq data of primed hESCs, hPGCLC and hPGC, previously published RNA-seq data (Chen et al., 2017) were used in this study. All raw RNA-seq reads were converted to gene expression level as described before (Pastor et al., 2016). Briefly, raw RNA-seq reads were then aligned to hg19 using Tophat (Trapnell et al., 2009) with ‘-g 1’ and ‘-no-coverage-search’ parameter. After alignment, read counts per gene were calculated using HTseq (Anders et al., 2015). Gene expression levels were determined by RPKM (reads per kilobase of exons per million aligned reads) with a customized R script. For heatmap of gene expression, $\log_2(\text{RPKM}+1)$ value were used as described in figure legends. The RNA-seq data used in this study is listed in Table S3.

ATAC-seq data analysis

ATAC-seq data were processed as previously described (Pastor et al., 2018). Briefly, after converting qseq file from sequencer to fastq format with customized script, paired-end 50bp ATAC-seq reads were mapped to hg19 genome using Bowtie with the parameters ‘-m 1’ and ‘-X 2000’. PCR duplicates were removed with Samtools (rmdup function) (Li et al., 2009). As described before, all reads aligned to the positive strands were offset by 4 bp and all reads aligned to the negative strands were offset by 5 bp due to Tn5 insertion (Adey et al., 2010). Since ATAC-seq is also able to capture long reads derived from nucleosome signals, we only focus on open chromatin reads pair with length less than 100bp. MACS2 callpeaks tools were used to identify ATAC-seq peaks in somatic tissue, hESC, iMeLC, hPGCLC and hPGC datasets (Zhang et al., 2008). In order to identify peaks specific to one condition among different conditions, bedtools multiinter option was used with Ryan Layers’s clustering algorithm applied (Quinlan and Hall, 2010). To calculate the overlap of ATAC-seq peaks in hESC, iMeLC, hPGCLC, hPGC, jaccard statistics were calculated using bedtools jaccard function (Favorov et al., 2012). ATAC-seq signals, heatmap and metaplot were visualized with ngs.plot (Shen et al., 2014). The ATAC-seq data used in this study is listed in Table S3.

ChIP-seq data analysis

Processed TFAP2C ChIP-seq data in Naive hESC from previously published (Pastor et al., 2018) were used in this paper (Table S3).

Motif Analysis

To identify enriched motifs in different peak sets, the 200bp region flanking peak summit were used. HOMER (findMotifGenome tool) (Heinz et al., 2010) were utilized with appropriate genome and default settings.

Principal component analysis (Figure 3A)

For principal component analysis (PCA) for RNA-seq data, RPKM for each sample were used as input. Variance of each gene’s RPKM in different samples were then calculated (rowVars function in R). PCA analysis (prcomp function in R) was performed on all genes across samples. PCA plots were then plotted with ggplot2 package in R (<http://ggplot2.tidyverse.org/>).

Code Availability

Custom scripts used for demultiplexing NGS reads, calculating RPKM, generating DNA methylation metaplots and comparing distribution of peaks to expression of nearby genes will be made available upon request.

DATA AND SOFTWARE AVAILABILITY

The accession number for the ATAC-seq data and RNA-seq data of key cell types including hESCs, iMeLCs, hPGCLCs, hPGCs, and embryonic somatic tissues reported in this paper are GEO: GSE120648 and GEO: GSE93126.

Cell Reports, Volume 25

Supplemental Information

The TFAP2C-Regulated *OCT4* Naive Enhancer

Is Involved in Human Germline Formation

Di Chen, Wanlu Liu, Jill Zimmerman, William A. Pastor, Rachel Kim, Linzi Hosohama, Jamie Ho, Marianna Aslanyan, Joanna J. Gell, Steven E. Jacobsen, and Amander T. Clark

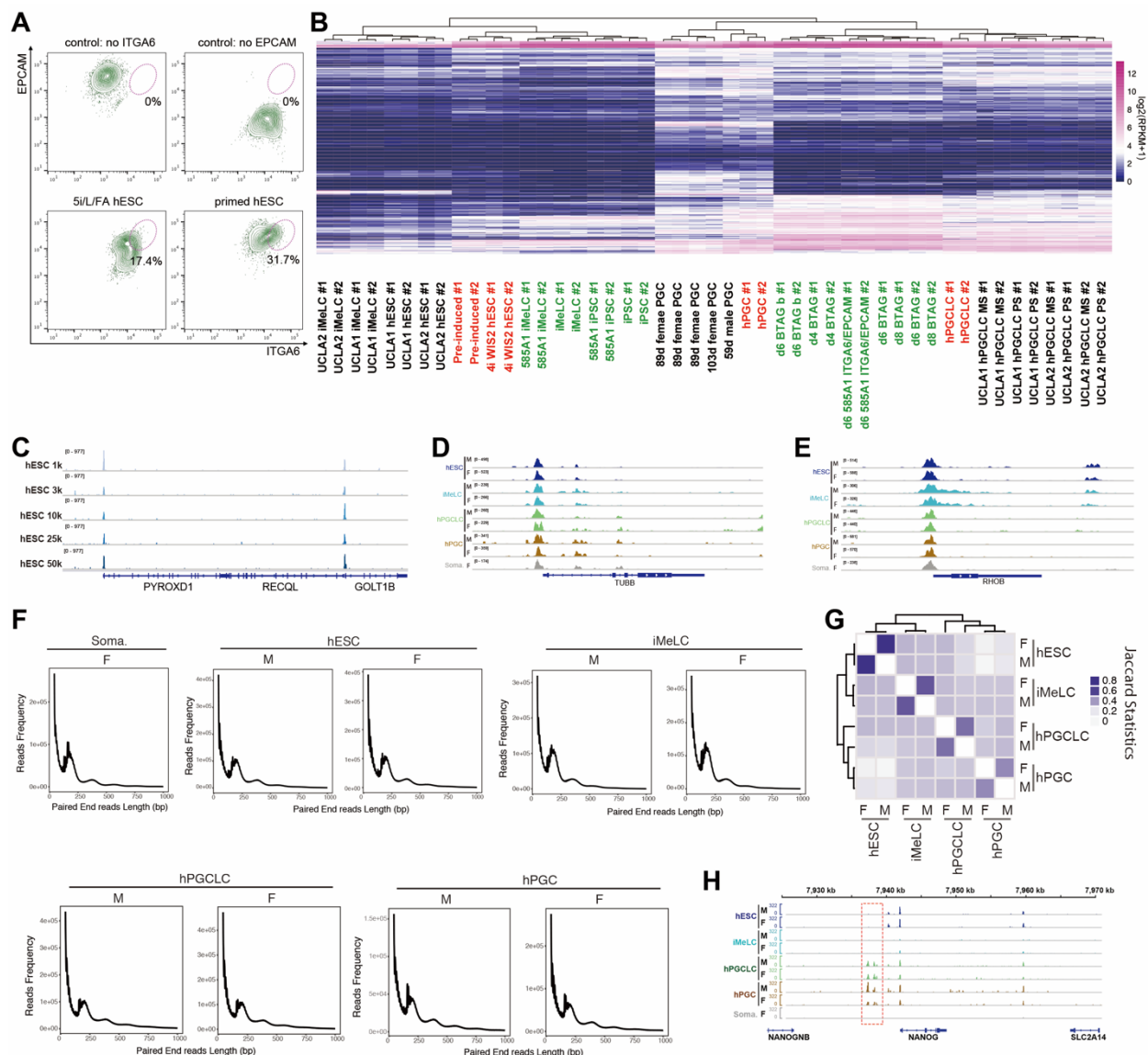


Figure S1. Comparison of human germ cells and ATAC-seq of primed hESCs, iMeLCs, hPGCLCs, and hPGCs. Related to Figure 1.

A: Flow cytometry for ITGA6/EPCAM in 5i/L/FA ground state naïve and primed states using the UCLA1 hESC line.

B: Comparison of RNA-seq datasets of human germ cells from different sources. Red indicates samples from Irie et al., 2015 (Irie et al., 2015); green indicates samples from Sasaki et al., 2015 (Sasaki et al., 2015); black indicates samples from Chen et al., 2017 (Chen et al., 2017). MS and PS indicate with and without SCF in aggregate media, respectively. The hPGCLCs analyzed in the current study were created in aggregates without SCF.

C: Screenshot of ATAC-seq signals from libraries made from 1,000, 3,000, 10,000, 25,000 and 50,000 hESCs.

D-E: Screenshot of ATAC-seq signals over *TUBB* (D), *RHOB* (E) in primed hESCs, iMeLCs, hPGCLCs, hPGCs, and embryonic somatic tissues.

F: Size distribution of ATAC-seq libraries made from embryonic somatic tissues (Soma.), primed hESCs, iMeLCs, hPGCLCs, and hPGCs.

G: Heat map of the overlap (measured by jaccard statistics) among ATAC-seq peaks from primed hESCs, iMeLCs, hPGCLCs, and hPGCs. hESCs and iMeLCs cluster together regardless of sex, and hPGCLCs and hPGCs cluster together regardless of sex.

H: Screenshot of ATAC-seq signals over *NANOG* locus in primed hESCs, iMeLCs, hPGCLCs, hPGCs, and embryonic somatic tissues.

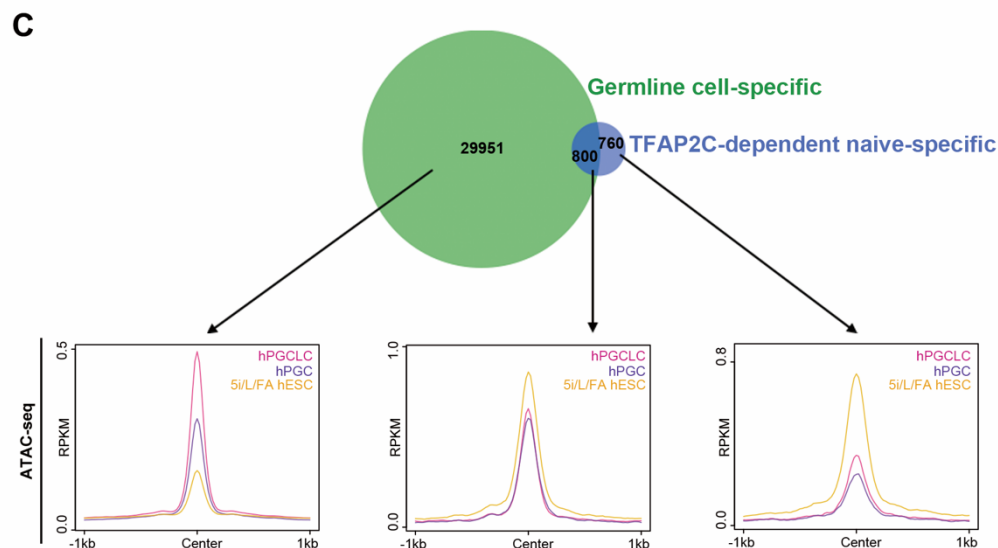
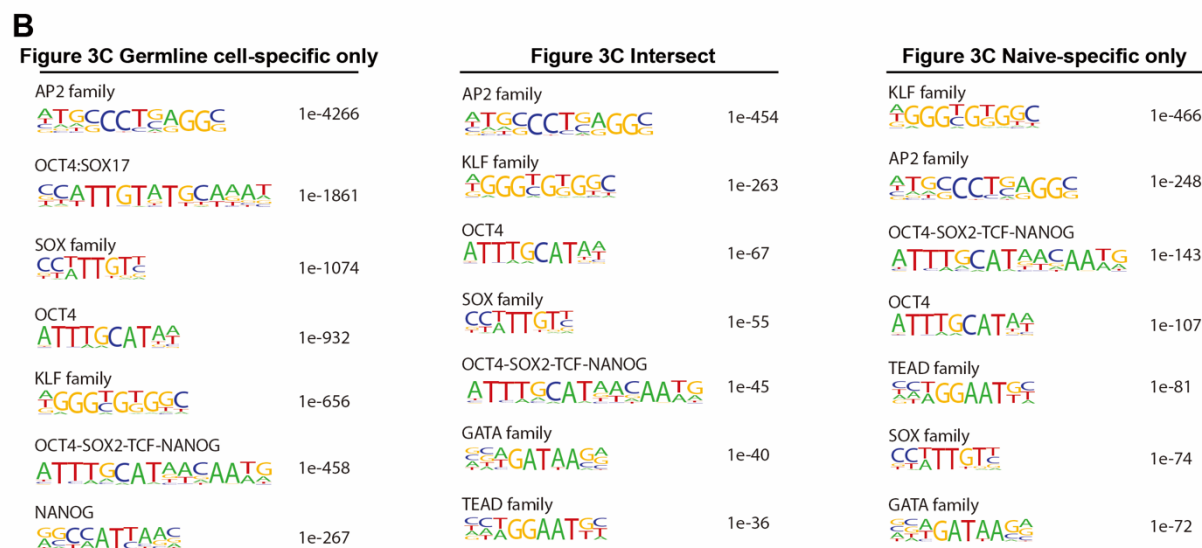
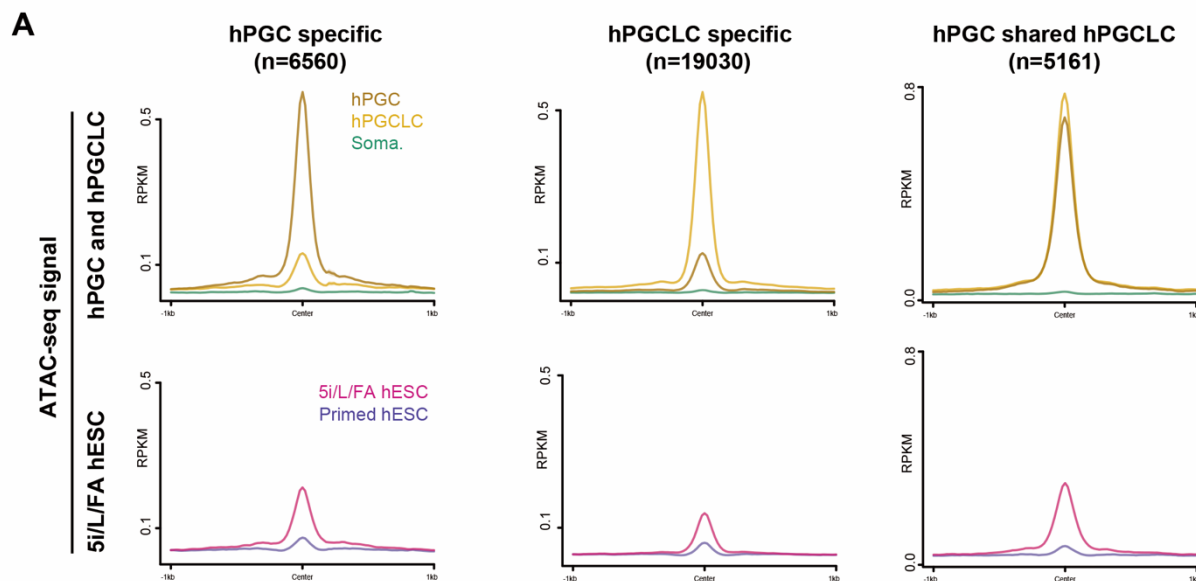


Figure S3. Latent ground state naïve pluripotency characteristics in human germline cells. Related to Figure 3.

A: ATAC-seq signals in hPGCs, hPGCLCs, somatic tissues and ground state naïve hESCs over hPGCLC-specific, hPGC-specific, and hPGCLC/hPGC-intersect open chromatin regions.

B: Motif analysis for germline cell-specific, naïve-specific, germline cell-naïve intersect regions in Figure 3C.

C: Venn diagram showing the overlap of germline cell-specific peaks with TFAP2C-dependent naïve-specific putative enhancers identified by Pastor et al., 2018 (Pastor et al., 2018). Metaplots show the ATAC-seq signals in hPGCLCs, hPGCs, and naïve hESCs over regions defined by the Venn diagram.

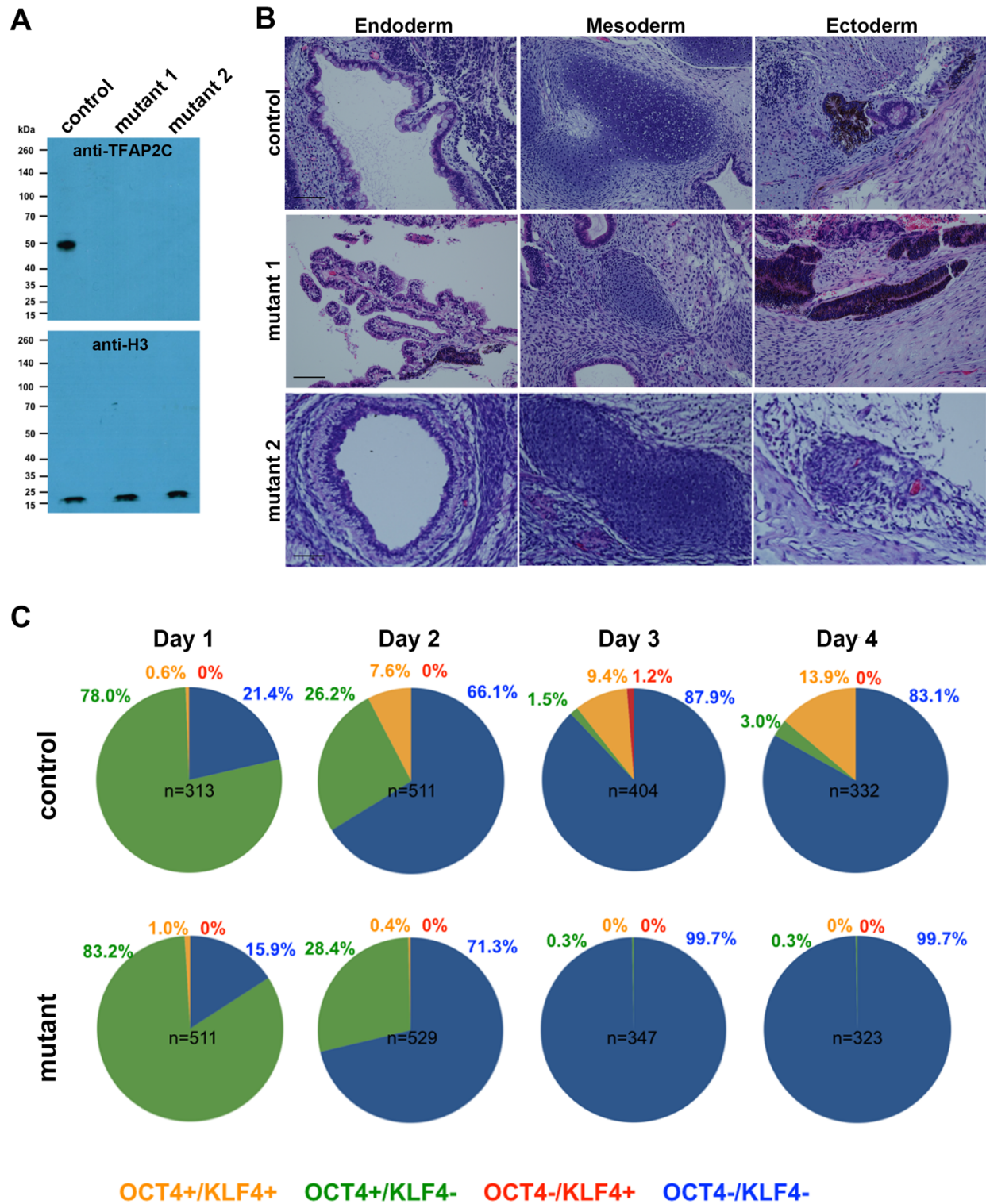


Figure S4. Characterization of *TFAP2C* mutant UCLA1 hESC lines. Related to Figure 4.

A: Western blot of TFAP2C in control and *TFAP2C* mutant 1 and mutant 2 primed hESCs. Histone H3 was used as loading control. Note that no TFAP2C protein is expressed in either *TFAP2C* mutant subline.

B: Teratoma formation of control and *TFAP2C* mutant 1 and mutant 2 UCLA1 hESCs. Representative endoderm, mesoderm, and ectoderm derivatives are shown. Scale bar: 100 μ m (n=8 biological replicates for control and n=4 biological replicates for each mutant).

C: Summary of hPGCLC induction from day 1 to day 4 from control *TFAP2C* mutant 1 UCLA1 hESCs. Four to seven biological replicates were used to count the OCT4 single positive (green), KLF4 signal positive (red), OCT4/KLF4 double positive (orange), OCT4/KLF4 double negative but DAPI positive (blue).

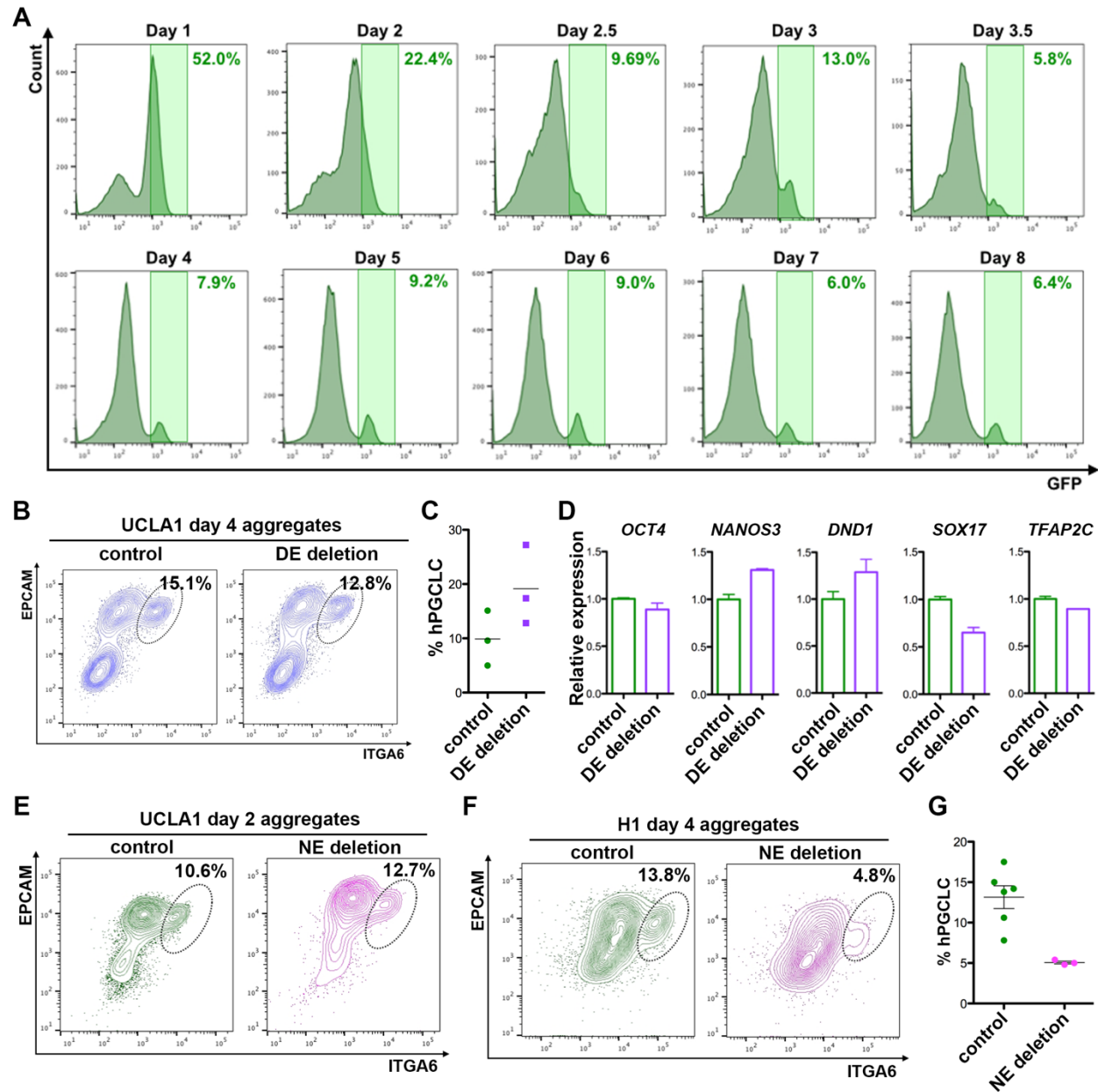


Figure S5. OCT4 expression dynamics during hPGCLC differentiation and analysis of the *OCT4* enhancers.

Related to Figure 5.

A: GFP dynamics during hPGCLC induction from day 1 to day 8 of aggregate differentiation using the H1 *OCT4*-GFP hESC line. Shaded box indicates the GFP positive gate used to discriminate hPGCLCs at day 4.

B-C: Flow cytometry showing percentage of hPGCLCs in control and *OCT4* DE deletion aggregates using the UCLA1 hESC line. hPGCLCs correspond to the ITGA6/EPCAM double positive cells (n= 2 biological replicates).

D: Expression of germ cell genes in ITGA6/EPCAM double positive hPGCLCs isolated from control and *OCT4* DE deletion aggregates at day 4 of differentiation (n= 2 biological replicates). Error bar represents SEM.

E: Flow cytometry showing percentage of hPGCLCs at day 2 of aggregate differentiation in control and *OCT4* NE deletion UCLA1 hESC line. hPGCLCs correspond to the ITGA6/EPCAM double positive cells.

F-G: Flow cytometry and quantification showing percentage of hPGCLCs in control and *OCT4* NE deletion aggregates at day 4 created from H1 hESC line. hPGCLCs correspond to the ITGA6/EPCAM double positive cells (n= 6 biological replicates for control and n=3 biological replicates for NE deletion). Error bar represents SEM in G.

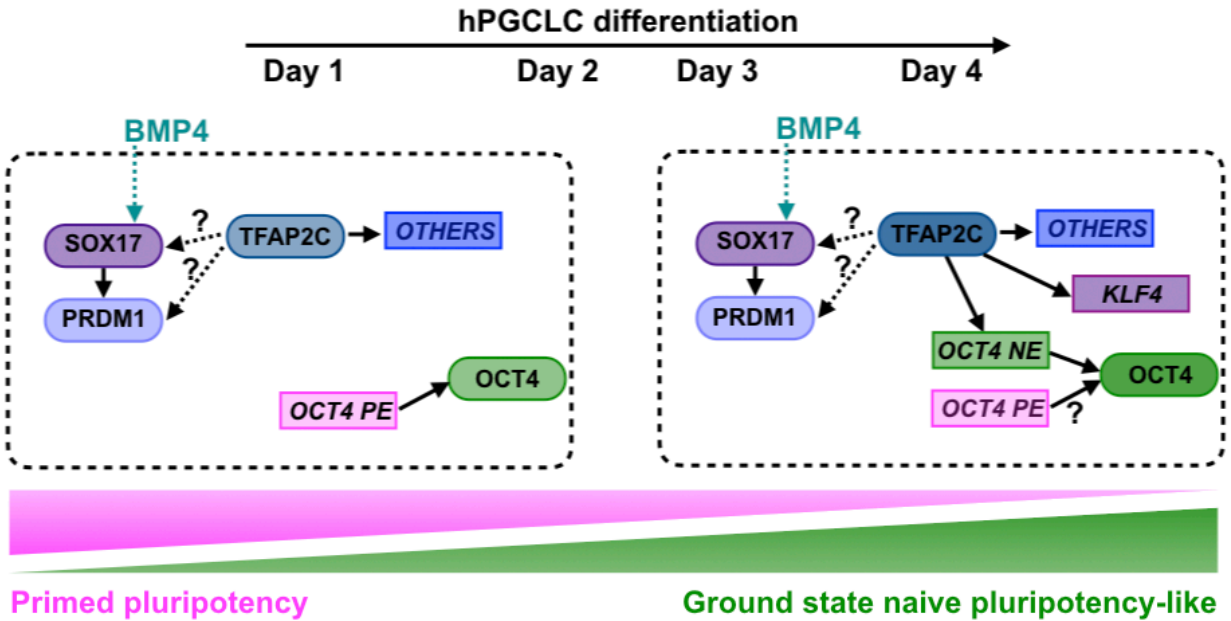


Figure S6. Model

With exposure to BMP4, we propose that hPGCLCs acquire a state that resembles ground state naïve pluripotency between day 2 to day 3 of aggregate differentiation through the binding of TFAP2C to the *OCT4* NE and other elements. One read out for the establishment of ground state naïve pluripotency-like state is the expression of KLF4. In addition, TFAP2C also regulates many other targets (designated as *OTHERS*) that could function together with or independent from other master regulators of hPGCLC development including SOX17 and PRDM1 as well as in maintenance of hPGCLC fate after day 3. *PE*: OCT4 proximal enhancer. *NE*: OCT4 naive enhancer.