HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2020 January 02.

Published in final edited form as:

Cell Rep. 2019 December 24; 29(13): 4568–4582.e5. doi:10.1016/j.celrep.2019.11.083.

Human Primordial Germ Cells Are Specified from Lineage-Primed Progenitors

Di Chen[1,11], **Na Sun**[2,3,11], **Lei Hou**[2,3,11], **Rachel Kim**[4], **Jared Faith**[1], **Marianna Aslanyan**[1], **Yu Tao**[1], **Yi Zheng**[5], **Jianping Fu**[5,6,7], **Wanlu Liu**[8], **Manolis Kellis**[2,3,12,*], **Amander Clark**[1,4,9,10,12,*]

[1]Department of Molecular Cell and Developmental Biology, University of California, Los Angeles, Los Angeles, CA 90095, USA

[2]MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA 02139, USA

[3]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[4]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, Los Angeles, CA 90095, USA

[5]Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109, USA

[6]Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI 48109, USA

[7]Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109, USA

[8]Zhejiang University-University of Edinburgh Institute, Zhejiang University School of Medicine, 310058 Hangzhou, PR China

[9]Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA

[10]Jonsson Comprehensive Cancer Center, University of California, Los Angeles, Los Angeles, CA 90095, USA

[11]These authors contributed equally

[12]Lead Contact

SUMMARY

In vitro gametogenesis is the process of making germline cells from human pluripotent stem cells. The foundation of this model is the quality of the first progenitors called primordial germ cells

*Correspondence: manoli@mit.edu (M.K.), clarka@ucla.edu (A.C.).

AUTHOR CONTRIBUTIONS

D.C., N.S., L.H., and A.C. designed the experiments. D.C., M.A., and J.F. conducted the experiments. N.S. and L.H. analyzed the scRNA-seq data; W.L. analyzed the ChIP-seq data. Y.Z. and J.F. conducted the scRNA-seq of modeled human amnion cells. D.C., N.S., L.H., M.K., and A.C. interpreted the data, and D.C., N.S., L.H., and A.C. wrote the manuscript.

SUPPLEMENTAL INFORMATION

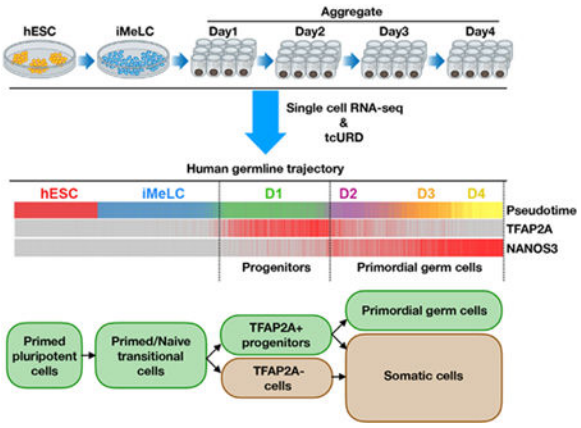
Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.11.083.

DECLARATION OF INTERESTS

The authors declare no competing interests.

(PGCs), which *in vivo* are specified during the peri-implantation window of human development. Here, we show that human PGC (hPGC) specification begins at day 12 post-fertilization. Using single-cell RNA sequencing of hPGC-like cells (hPGCLCs) differentiated from pluripotent stem cells, we discovered that hPGCLC specification involves resetting pluripotency toward a transitional state with shared characteristics between naive and primed pluripotency, followed by differentiation into lineage-primed TFAP2A^{+} progenitors. Applying the germline trajectory to *TFAP2C* mutants reveals that TFAP2C functions in the TFAP2A^{+} progenitors upstream of PRDM1 to regulate the expression of SOX17. This serves to protect hPGCLCs from crossing the Weismann's barrier to adopt somatic cell fates and, therefore, is an essential mechanism for successfully initiating *in vitro* gametogenesis.

Graphical Abstract



In Brief

Using genetics, genomics, and single-cell RNA-seq, Chen et al. characterize the human germline trajectory, revealing two pluripotent cell transitions during primordial germ cell specification. They reveal the identity of primordial germ cell progenitors and show that TFAP2C prevents gastrulation and amnion-like fate at the point of primordial germ cell specification.

INTRODUCTION

More than a century ago, the German biologist August Weismann proposed the germplasm theory of heredity in which he postulated that germ cells of animals contain an essential hereditary substance that passes from one generation to the next, now known to be DNA (Weismann, 1893). As a part of his theory, the concept of ''Weismann's barrier'' was also established (Weismann 1893). This is not a physical barrier but a model that proposes that hereditary information flows only through germline cells and not somatic cells of the body. This is important because without germline specification, the outcome is infertility, a disease that affects around 10% of humans of reproductive age. Understanding or possibly overcoming infertility requires a human model that recapitulates all stages of germline cell development, including how Weismann's barrier is first established with the specification of primordial germ cells (PGCs) and somatic cells in the embryo. Human pluripotent stem cells and the differentiation of PGC-like cells (PGCLCs) serve this purpose.

Specification of PGCs in animals involves two different strategies. One involves pre-formation, which occurs in the model organisms *Xenopus*, zebrafish, *Drosophila*, and *Caenorhabditis elegans* (Extavour and Akam, 2003; Williamson and Lehmann, 1996). In these organisms, rare cells of the newly fertilized embryo inherit molecules produced by the female germline (oocyte), bestowing upon these embryonic cells the maintenance of germline cell identity. In contrast, cells that do not inherit these maternal molecules from the oocyte after fertilization become somatic cells, thus irreversibly crossing Weismann's barrier and losing the capacity to contribute to the germline. For animals that specify germ cells, each generation by induction, such as mouse, pig, horse, rabbit, and human, Weismann's barrier encompasses an additional step, which involves an initial fate restriction from totipotency to pluripotency, and then around the time of embryo implantation and gastrulation, PGCs are induced from peri-implantation cell precursors (Magnúsdóttir and Surani, 2014; Tang et al., 2016). The identity of these precursors in humans is completely unknown. As PGC specification progresses, the remaining embryonic precursors cross Weismann's barrier to become somatic cells. In mouse, where mouse PGC (mPGC) induction is well studied, it has been reported that mPGCs are induced at embryonic day 6.5 (E6.5) by bone morphogenetic protein 4 (BMP4) signaling to the Wnt family member 3 (WNT3)-primed epiblast cells (Ohinata et al., 2009). The maintenance of germline identity downstream of BMP4 signaling involves the transcription factor network, including transcription factors (TFs) PRDM14, PRDM1 (also known as BLIMP1), and TFAP2C, with PRDM1 functioning upstream of TFAP2C to repress somatic cell fate and maintain mPGC identity (Magnúsdóttir and Surani, 2014; Magnúsdóttir et al., 2013; Nakaki et al., 2013; Ohinata et al., 2005; Weber et al., 2010; Yamaji et al., 2008).

Recent studies using the differentiation of human PGCLCs (hPGCLCs) from human pluripotent stem cells have revealed that the TF network required to specify and maintain human PGC (hPGC) fate is different from the mouse (Chen et al., 2017; Irie et al., 2015; Kojima et al., 2017, Sasaki et al., 2015). For example, SOX17 is required for hPGCLC specification, whereas in mouse it is not (Irie et al., 2015). In mPGCs, TFAP2C functions downstream of PRDM1 to repress somatic cell differentiation genes (Ohinata et al., 2005). However, in hPGCLC differentiation from *PRDM1* mutants, *TFAP2C* expression levels are unaffected in the PRDM1 mutant hPGCLCs relative to controls (Sasaki et al., 2015). One human-specific role for TFAP2C in hPGCLCs involves the opening of naive-specific enhancers and the acquisition of naive-like pluripotency (Chen et al., 2018). An alternate but not necessarily mutually exclusive role for TFAP2C is to repress somatic cell gene expression in *PRDM1*-positive hPGCLC precursors and, therefore, to keep cells within Weismann's barrier (Kojima et al., 2017). However, this result is predicated on the assumption that PRDM1 is expressed in hPGCLC precursors before TFAP2C, and also that PRDM1 is not expressed by the somatic cells of the aggregates. The only way to address this is to evaluate the germline and somatic cell trajectories from the pluripotent state through to the specification of hPGCLCs. Recently, single-cell RNA sequencing (scRNA-seq) has been widely used to map differentiation trajectories and to discover new cell types without known markers (Poirion et al., 2016; Stegle et al., 2015). Capitalizing on this approach to identify the precursors of hPGCs in the peri-implantation window, we performed scRNA-seq of more than 100,000 individual cells during hPGCLC specification to model the molecular events in

hPGC specification by using an adapted developmental trajectory detection algorithm named time constraint URD (tcURD). Through this work, we uncovered the human germline trajectory and discovered the identity of potential peri-implantation progenitors for hPGCs.

RESULTS

Tracing hPGCs during Peri-implantation Embryogenesis and in hESC-based hPGCLC Induction

Using human embryo attachment culture (Deglincerti et al., 2016; Shahbazi et al., 2016), we evaluated whether *ex vivo* hPGCs could be identified before primitive streak formation and, therefore, the feasibility of using the human embryo attachment culture model to identify hPGC progenitors. Under California state law, human embryo attachment culture is legally allowed up to day 12 post-fertilization, which is 2 days before the International Consensus of day 14 or before primitive streak formation. To trace *ex vivo* hPGCs, we first confirmed the co-expression of NANOG, SOX17, and TFAP2C (N/S/T) as being restricted to hPGCs (Gkountela et al., 2015; Guo et al., 2015; Tang et al., 2015) (Figures S1A and S1B). Next, we thawed a total of 127 day 5 or day 6 post-fertilization human blastocysts; twenty-four h after thawing, these blastocysts were considered day 6 or day 7, respectively. From this cohort, we performed whole-mount immunofluorescence followed by confocal imaging and 3D reconstruction of 78 human embryos from day 6 to day 12 (Figure 1A; Table S1). We discovered N/S/T triple-positive putative *ex vivo* hPGCs at day 12 post-fertilization in 2/26 embryos (Figures 1B and S1C; Video S1; Table S2). N/S/T triple-positive cells were not identified at any of the reminaing 52 embryos at either days 10–11 or days 6–7 (Figures S1D and S1E). Instead, in the pre-implantation blastocyst, the TFAP2C protein was expressed in all cells, whereas NANOG- and SOX17-positive cells were mutually exclusive, marking the pre-implantation epiblast (Pre-EPI) and hypoblast cells, respectively (Figure S1E). Taken together, these observations suggest that hPGC specification defined as N/S/T triple-positive cells in the peri-implantation embryo begins at around day 12 post-fertilization in humans before primitive streak formation.

Given that less than 10% of human blastocysts at day 12 had evidence of N/S/T triple-positive hPGCs, it would be difficult to discriminate the identity of hPGC precursors in this system without using large numbers of intact human embryos consented to research. Therefore, we used the differentiation of human embryonic stem cells (hESCs) into hPGCLCs by generating disorganized 3D aggregates in the presence of BMP4 and other cytokines (Sasaki et al., 2015) (Figure 1C). We used the two-step method of differentiation starting from primed pluripotent stem cells and first differentiating through an incipient mesoderm-like cell (iMELC) intermediate, as this approach has been successfully used by different groups to generate hPGCLCs from primed hESCs and hiPSCs (Sasaki et al.,2015; Chen et al., 2017; Chen et al., 2018; Yokobayashi et al., 2017; Kojima et al., 2017). Given the known variability in hPGCLC competency between independently derived pluripotent stem cell lines (Chen et al., 2017; Yokobayashi et al., 2017), we designed the experiment to evaluate hPGCLC specification in a highly germline competent hESC line (UCLA2) and a hESC line with significantly lower competency (UCLA1) (Figures 1D and 1E). These hESC lines were derived under identical conditions from individual blastocysts that were provided

by the same embryo donors (Diaz Perez et al., 2012). We discovered rare and yet reproducible N/S/T triple-positive hPGCLCs at day 1 of aggregate differentiation after exposure to BMP4 in both cell lines, with an increase in the percentage of triple-positive hPGCLCs emerging at day 2 in the highly competent UCLA2 line relative to UCLA1 (Figure 1F).

To better characterize the dynamics of germline specification, we applied scRNA-seq with 10x Genomics to hESCs, iMeLCs (incipient mesoderm-like cells), and aggregate cells at days 1–4 with two biological replicates for both UCLA1 and UCLA2 cell lines (Zheng et al., 2017) (Table S3). In total, we sequenced 85,309 cells with high quality to uncover the germline trajectory and explored the progenitors involved in establishing germline cell fate (Table S4). To identify the cell populations at each time point, we used canonical correlation analysis (CCA) for batch correction of the two biological replicates at each time point within each cell line and applied uniform manifold approximation and projection (UMAP) to visualize the distribution of single cells throughout the entire differentiation time course (Figures 1G, 1H, and S1F). Specified hPGCLCs were indicated by the germline-specific gene *NANOS3* (Gkountela et al., 2013; Butler et al., 2018; McInnes et al., 2018) (Figures 1G and 1H). Consistently, more *NANOS3*-positive hPGCLCs were present in UCLA2 than UCLA1. Interestingly, we observed clear branches of connected cell clusters following the temporal order of PGCLC differentiation in the 2D maps from both UCLA1 and UCLA2 (Figures 1G and 1H).

Mapping the Human Germline Trajectory by scRNA-seq

To precisely map the human germline trajectory, we first identified cells quadruple-positive for *NANOG*, *SOX17*, *TFAP2C*, and the germline-specific gene *NANOS3* (with log-transformed-normalized unique molecular identifier (UMI) counts of >0.5 for *NANOG*, *SOX17*, *TFAP2C*, and *NANOS3*) in UCLA1 and UCLA2 as hPGCLCs (Figures 2A, 2B, S2A, and S2B). The emergence of distinct quadruple cells was identified at day 1 and persistent during day 2 to day 4 differentiation (Figures 2A and 2B). These observations indicate that hPGCLC specification occurs within the first 24–48 h of BMP exposure, and UCLA2 responds more rapidly and efficiently relative to UCLA1.

To identify differentiation trajectories unbiasedly without using previously known lineage markers, we applied the computational trajectory detection method URD to the scRNA-seq data (Farrell et al., 2018). URD is mainly built on connecting cells with nearest neighbors in terms of transcriptome similarity under the assumption that the transcriptome of a cell changes progressively during differentiation. It has been reported that hPGCs and hPGCLCs regain pluripotency after specification (Chen et al., 2018), which may connect the specified hPGCLCs to hESCs directly. To avoid these possible artificial connections, we adapted URD with a constraint that cells could only be connected to its neighbors from the same, the previous, or the next time points (STAR Methods), and we applied this time constraint URD (tcURD) to predict germline trajectory (Figuress 2C, 2D, S2A, and S2B). This analysis revealed that in the less-competent UCLA1 hESC line, a subset of iMeLCs (yellow) had already exited from the branched cellular trajectory, to beome cells equivalent to visceral/yolk sac endoderm (Figure 2C). However, in UCLA2, all iMeLCs retain competency to

differentiate toward either hPGCLCs or somatic cells including endoderm as well as other peri-implantation lineages (Figure 2D). Therefore, unlike UCLA1, at the time of BMP4 exposure, more UCLA2 iMeLCs are still within Weismann's barrier and are, therefore, competent for hPGCLC specification. Notably, PRDM1 is not restricted only to germline trajectory but is also present in endoderm cells (Figures S2C and S2D).

To map the somatic lineages differentiating together with the hPGCLCs in the 3D disorganized aggregates, we compared our differentiation trajectories to the cynomolgus (cyno) macaque per-implantation embryo scRNA-seq datasets (Nakamura et al., 2016) and amnion-like cells (AMLCs) from 3D embryo models (Zheng et al., 2019). Basically, we identified the signature genes for each annotated cell type in cyno scRNA-seq data and AMLCs, followed by calculating the signature score in each human cell across the germline trajectory to score the ends of the branches at day 4 (STAR Methods). This analysis revealed that the somatic cells in the disorganized aggregates include yolk sac endoderm/primitive endoderm and a variety of somatic cell types equivalent to trophoblast (TE), gastrulating cells (Gast), amnion, and extra-embryonic mesenchyme (EXMC) (Figures 2C and 2D). This result indicates that the 3D disorganized aggregates from independently derived hESC lines have the potential to differentiate into a variety of peri-implantation cell types together with the germline cells in the presence of BMP4.

Given there was only one germline specification trajectory along with other somatic lineages (Figures 2C, 2D, S2C, and S2D) in both UCLA1 and UCLA2, we used the trajectory from each cell line to trace the peri-implantation embryonic progenitors to the human germline. Figures S2

hPGCLCs Are Specified through a Pluripotent State with Both Naive and Primed Characteristics

Next, we focused only on the germline trajectory. Using pseudotime analysis of the cells extracted from the germline trajectory, we first explored the dynamics of hPGC specification genes (*NANOG*, *SOX17*, *TFAP2C*, and *NANOS3*) together with the pluripotent TFs *SOX2* and *OCT4* (Also known as *POU5F1*). *OCT4* is expressed in all cells of the trajectory. *NANOG* is also expressed throughout the germline trajectory with notable downregulation toward the end of day 1, followed by a subsequent upregulation at day 2 during hPGCLC specification. *SOX2* repression in the germline trajectory begins at the time of iMeLC induction and is more pronounced at the end of day 1, concomitant with upregulation of the germline specifier *SOX17*. *TFAP2C* expression was dynamic in the germline trajectories of UCLA1 and UCLA2, being upregulated during day 1 in UCLA1 and UCLA2 (Figures 3A and S3A).

To map the *in vitro* germline trajectory back to *in vivo* embryogenesis, we again used the scRNA-seq data of cyno embryos (Nakamura et al., 2016), with a focus on the pluripotent stages of development, including inner cell mass (ICM), Pre-EPIs, post-implantation early epiblasts (PostE-EPIs), and post-implantation late epiblasts (PostL-EPIs). As previously reported, we discovered that undifferentiated UCLA2 (Figure 3B) and UCLA1 (Figure S3B) hESCs highly expressed the signatures of the cyno PostE-EPI and PostL-EPI with primed pluripotency (Nakamura et al., 2016). However, to our surprise, the iMeLCs within the

germline trajectory of UCLA2 and UCLA1 reset their pluripotency status to a state that correlates with both cyno ICM and Pre-EPI (naive) and PostE-EPI and PostL-EPI (primed) (Figures 3B and S3B). This result suggests that in 24 h of exposure to the cytokines in iMeLC media, the hESCs initiate a partial reversion toward naive-like pluripotency states found in the pre-implantation primate embryo while retaining primed state characteristics. To confirm the pluripotent state in germline trajectory iMeLCs, we also defined naive and primed signatures by using scRNA-seq data of human naive and primed hESCs (Messmer et al., 2019). Consistently, iMeLCs in the germline trajectory showed signatures of both naive and primed hESCs (Figures 3C and S3C). Thus, we termed the transitional pluripotent state of iMeLCs in germline trajectory as a ‘‘germinal pluripotent state,’’ as they are competent to give rise to hPGCs in addition to somatic lineages (Figure 2C and 2D). Furthermore, we discovered TFs that are enriched in germinal pluripotent cells which are expressed at lower levels in naive and primed pluripotent states (Figure 3D). To find the *in vivo* counterparts of embryonic cells in this transitional germinal pluripotent state, we applied scRNA-seq to day 12 human embryos. We identified 13 cells annotated as epiblasts based on the cyno embryonic cells (Nakamura et al., 2016) (Figure 3E). Within these 13 cells, we discovered 4 cells that show signatures of primed hESCs, naive hESCs, and iMeLCs in germline trajectory, indicating they are possible *in vivo* embryonic cells with germinal pluripoteny (Figure 3F).

Human PGC Specification Transitions through a *TFAP2A*$^{+}$ Progenitor with Both Amnion and Gastrulation Signatures

There are currently two hypotheses for the origin of hPGCs in amniotes with bilaminar embryonic discs. Using the non-human primate cyno macaque as a model, cynoPGCs are identified in the amnion at E11 (Sasaki et al., 2016). However, in the porcine (p), porcine PGCs (pPGCs) were first identified in pre-mesendoderm cells in the region of the posterior primitive streak (Kobayashi et al., 2017). To further map the human germline trajectory back to *in vivo* embryogenesis, we next tracked the putative amnion-related genes (*TFAP2A*, *GATA3*, and *CDX2*) in the germline trajectory (Shao et al., 2017a, 2017b). Strikingly, *TFAP2A* is upregulated for approximately 24 h during day 1, together with *GATA3* and *CDX2* (Figures 4A and 4B). Repression of *TFAP2A*, *GATA3*, and *CDX2* genes then mark the beginning of hPGCLC specification. Genes related to gastrulation, *EOMES* and *T* (also known as *BRACHYURY*), were highly and transiently upregulated at the beginning of day 1 in the UCLA2 germline trajectory. In contrast, the less-competent UCLA1 line exhibited only modest upregulation of *EOMES* and *T* (Figures 4A and 4B). Notably, *TFAP2A* is also expressed in gastrulating cells during cyno embryogenesis (Figure S4A). These observations suggest that hPGC specification originates from a transient population of *TFAP2A*-expressing progenitors that exhibit both amnion- and gastrulation-like TF identity, with cells exhibiting more *TFAP2A*$^{+}$ gastrulation-like identity in the earlier stages of UCLA2 relative to UCLA1 germline trajectories. Given that clear clusters of *NANOG*/*SOX17*/*TFAP2C*/*NANOS3* quadruple-positive cells are first shown at day 2 (Figures 2A and 2B), we termed the day 1 cells in germline trajectory as germline progenitors.

To better infer the relationships of the D1 progenitor cells to the peri-implantation embryo, we utilized the scRNA-seq dataset of the recently published organized 3D human embryo

model, with a focus on the AMLCs and day 2 (D2) hPGCLCs. The 3D embryo model was created by exposing the H9 hESC line to BMP4 for 48 h by using microfluidics, which results in the generation of posteriorized embryonic-like sacs with AMLCs, pre-gastrula cells, and hPGCLCs (Zheng et al., 2019). For gastrulation cell identity, we utilized the annotated gastrulating cells from cyno peri-implantation embryo (Sasaki et al., 2016). (Figures 4C and S3B). This analysis reveals that *TFAP2A*$^{+}$ day 1 progenitors are strongly enriched in cells with gastrulation identity throughout day 1 and that AMLC identity is also particularly strong at the end of day 1 just prior to *N*/*S*/*T* and *NANOS3* positive hPGCLC specification (Figures 4C and S4B). Notably the D2 hPGCLCs from the 3D embryo model strongly resemble the day 2 to day 3 hPGCLCs in the germline trajectory, whereas the early cynoPGCs (ecPGCs) resemble day 4 hPGCLCs generated in the aggregates (Figures 4C and S4B). Given that the hPGCLCs from the 3D embryo model at 48 h map the closest to the day 2–day 3 hPGCLCs in the germline trajectory from the 3D-disorganized aggregates, this result suggests that regardless of whether 3D-modeled embryos or 3D-disorganized aggregates are used to generate hPGCLCs, the identity of the hPGCLC 48 h (or 2 days) after BMP4 exposure is very similar. Given that the current approach for creating 3D-modeled embryos can only be sustained for 48–72 h before the system collapses, the disorganized aggregates provide a slightly extended opportunity to consolidate hPGCLC fate to one that is closer to ecPGCs found *in vivo*. (Figures 4D and S4C).

TFAP2C Functions Upstream of SOX17 in Human Germline Specification

Next, we focused on the cells in the germline trajectory to decipher the gene regulatory network associated with the establishment of hPGC fate. To do this, we first identified germline trajectory differentially expressed genes (gtDEGs) as the union of pseudotime correlated genes, and DEGs across stages in the germline trajectory (STAR Methods). About 3,000 gtDEGs were detected and grouped into 10 and 12 clusters in UCLA1 and UCLA2, respectively, based on their expression profiles across the trajectory (Figures 5A and S5A). A clear wave of dynamic gene expression was observed in both UCLA1 and UCLA2 from undifferentiated hESCs to hPGCLCs at day 4 (Figures 5A and S5A). To understand the regulatory mechanism, we predicted TFs based on the motif enrichment at the promoter regions of genes in each cluster as upstream regulators (Figures 5A and S5A; STAR Methods). From this analysis, we identified expected TFs known to be responsible for hPGCLC specification, including *SOX17*, *TFAP2C*, *PRDM1*, and *EOMES* (Figures 5A and S5A). In addition, we identified *TFAP2A* at day 1–2 of aggregate differentiation as well as the TFs *ID3*, *ID4*, and *TBX3* (Figures 5A and S5A) and others. Notably, germline trajectory TFs shared by UCLA1 and UCLA2 include germline related (*TFAP2C*, *SOX17*, and *PRDM1*), pluripotency related (*NANOG*, *POU5F1*, and *KLF4*), amnion/gastrulation related (*TFAP2A*, *CDX2*, and *GATA3*), and BMP4 signaling related (*ID1*, *ID2*, *ID3*, *SMAD6*, *MSX1*, and *MSX2*) (Figures 5A and S5A; Table S5).

An analysis of the TF regulatory network at the time of hPGCLC specification indicates that *SOX17* may be a putative target of TFAP2C (Figures 5B and S5B–S5D). To test this, we compared chromatin accessibility measured by ATAC-seq (Chen et al., 2018) and TFAP2C binding measured by chromatin immunoprecipitation sequencing (ChIP-seq) in hESCs, iMeLCs, and day 4 aggregates. To identify potential distal regulatory elements, we also

performed H3K27ac ChIP-seq in day 4 hPGCLCs isolated using ITGA6/EPCAM (Figure 5C). We discovered that TFAP2C is bound to the *SOX17* promoter in hESCs, iMeLCs, and day 4 cells. However, *SOX17* mRNA is only upregulated in hPGCLCs (Figure 5C). This observation suggests that the TFAP2C promoter binding at the *SOX17* locus is not sufficient to induce the dynamic upregulation of *SOX17* at the time of hPGCLC specification. Integrating the H3K27ac and TFAP2C ChIP-seq data, we propose that although TFAP2C is always bound to the *SOX17* promoter, it is the binding of TFAP2C to distal regulatory sites that are also coordinately enriched in H3K27ac in hPGCLCs to enable *SOX17* expression at the point of hPGCLC specification (Figure 5C).

TFAP2C Safeguards Germ Cell Fate in the Germline Trajectory

TFAP2C is critical for hPGCLC formation but is not necessary for hESC self-renewal or the differentiation of iMeLCs (Chen et al., 2018; Kojima et al., 2017). To evaluate the role of TFAP2C in the germline trajectory, we performed scRNA-seq of *TFAP2C*$^{-/-}$ cells at 6 time points (hESCs, iMeLCs, aggregates at days 1 to 4, 19,808 cells sequenced) and compared this to the single cells from UCLA1 wild-type cells, which is the genetic background used to make the *TFAP2C* mutant (Figure 6A). As shown in UMAP, most of cells from *TFAP2C*$^{-/-}$ are aligned with wild-type cells across the six differentiation time points (Figures 6B and 6C). In contrast, the germline cluster marked by *NANOS3* is composed almost entirely of wild-type cells (Figure 6C and 6D). The rare *TFAP2C*$^{-/-}$ cells in the hPGCLC cluster defined by UMAP were analyzed further and compared to wild-type UCLA1 hPGCLCs. This analysis showed that *TFAP2C*$^{-/-}$ hPGCLCs failed to repress genes associated with somatic cell embryonic morphogenesis (Figure S6A), suggesting that TFAP2C functions as predicted in specified hPGCLCs to prevent cells from exiting the germline to adopt a somatic cell fate (Kojima et al., 2017).

To evaluate whether TFAP2C functions earlier in the germline trajectory, we applied tcURD analysis to the merged wild-type and *TFAP2C*$^{-/-}$ data to identify blades that were highly enriched in wild-type cells (blue) or mutant cells (red) (Figures 6E and 6F). These data show a very distinct blade (red dashed line box in Figure 6F) composed of mostly *TFAP2C*$^{-/-}$ cells, that branches from the common day 1 progenitor (blue dashed line box in Figure 6F) to hPGCLCs (labeled as hPGCLC in Figure 6F). This *TFAP2C*$^{-/-}$ blade is negative for the hPGCLC markers *NANOS3* and *SOX17* (Figures 6G and 6H) and instead expresses somatic genes consistent with posterior primitive streak progenitors, including *MIXL1*, *T*, *HOXB5, HAND1*, *GATA6*, *FOXF1*, and *MESP2*. (Figures 6I, 6J, and S6C). To determine the consequence of failing to express *TFAP2C* in the TFAP2A common progenitors, we identified differentially expressed genes between wild-type and the *TFAP2C* mutant cells in the progenitor population, and discovered that *TFAP2C*$^{-/-}$ cells fail to upregulate *SOX17* and downregulate *SOX2* (Figure S6D). In contrast, the endoderm trajectory was composed of equivalent percentages of both wild-type and *TFAP2C* mutant cells, and these cells expressesed *SOX17*, indicating that the regulation of SOX17 expression by TFAP2C is unique to the germline at this stage in development (Figure 6H). To examine pluripotency and hPGC identity in wild-type and *TFAP2C*$^{-/-}$ cells, we evaluated the signatures from cyno Pre-EPI, PostL-EPI, and ecPGCs in day 1 to day 4 cells of the germline trajectory. No difference was detected in day 1 wild-type and *TFAP2C*$^{-/-}$ cells; however, starting from day

2, *TFAP2C*$^{-/-}$ cells were lower in pre-EPI pluripotency and ePGC identity but higher in PostL-EPI identity (Figure 6K). Collectively, our data indicate that TFAP2C functions to repress somatic cell fate in the newly specified hPGCLCs as first shown by (Kojima et al., 2017). In addition, we uncovered new roles, including preventing the *TFAP2A*$^{+}$ common progenitor cells from crossing Weismann's barrier to become somatic cells, uniquely regulating SOX17 expression at the point of hPGCLC specification from the *TFAP2A*$^{+}$ progenitors, and finally re-wiring pluripotency in the newly specified hPGCLCs at day 2 of aggregate differentiation to a naive-like state similar to primate pre-EPI pluripotency.

DISCUSSION

Specification of hPGCs in the embryo is critical for the generation of functional gametes in adults. Therefore, understanding the origin and specification of human germline cells is one of the most important events in our fundamental understanding of life, as it defines our ability to have children. The Weissman's barrier was coined to explain DNA transmission exclusively through germline cells (Weismann 1893). However, in species that differentiate PGCs by induction, the immediate progenitors of the germline differ across the species, and as a result, the TF network required for PGC specification has also diverged.

Human embryo attachment culture starting with intact human embryos donated to research suggests that hPGC specification, although rare, begins around day 12 of embryo development prior to the formation of primitive streak and gastrulation. This is consistent with a recent study using OCT4 and SOX17 as hPGC markers (Popovic et al., 2019). Using a different approach, we identified hPGCs/hPGCLCs as being triple positive for N/S/T, and by using the *in vitro* differentiation of hPGCLCs from hESCs, we showed that when cells co-express *N*/*S*/*T*, they also express the hPGC gene *NANOS3*. In contrast to the identity of specified hPGCs/hPGCLCs that have well-characterized molecular markers, nothing is known for the peri-implantation progenitor cells in the germline trajectory. In this study, we overcame the challenge by taking advantage of the differentiation time course of hPGCLC specification with scRNA-seq and tcURD-based trajectory analysis to trace the trajectory from hESCs to hPGCLCs. Different from most single-cell genomics trajectory tracing studies, we also managed to compare the trajectories from three genetically different cell lines, namely, high germline competency UCLA2, low germline competency UCLA1, and *TFAP2C*$^{-/-}$ *UCLA1,*which provides us a unique opportunity not only to trace germ cell development and characterize the progenitors before germ cell fate is determined but also to compare trajectories from different cell lines to pinpoint the potential regulatory circuits shared and specific from three scenarios in a stage-specific manner. Our strategy could be applied to other complicated biological trajectories without known markers for each stage

Within the germline trajectory after hESCs are induced to differentiate, our data indicate that the immediate pluripotent cell precursors exhibit an identity that can be annotated as simultaneously pre-implantation (naive) and post-implantation (primed) pluripotency. Indeed, transitional pre-implantation embryonic cellular states were recently reported with the re-analysis of the scRNA-seq data from Petropoulos et al. (2016) that identified cells containing both pre-EPI and primitive endoderm identity (Stirparo et al., 2018). Similarly, there is remarkable similarity between ICM and pre-EPI, with only a small number of genes

discriminating these different cell types (Blakeley et al., 2015; Petropoulos et al., 2016; Stirparo et al., 2018). Therefore, it could be conceivable that along the germline trajectory previously unannotated transitional states of pluripotency exist, that share pluripotent identities with both naive and primed pluripotent states. Given that this state has not been annotated before, we are calling this transitional naive/primed state of pluripotency ‘‘germinal pluripotency,’’ which is conducive to hPGC specification, amnion differentiation, and posterior primitive streak/mesoderm/extra embryonic mesenchyme and trophoblast development. This is critical because hPGCLCs are induced from primed state pluripotent stem cells, but our data indicate that these primed cells are reverted to a transitional state of pluripotency to initiate hPGCLC induction.

With the exposure to BMP4, progenitors for hPGCLCs are specified from germinal pluripotent cells. The lineage-primed progenitors transiently expressed amnion and gastrulation markers, such as TFAP2A, GATA3, EOMES, and T. In the cyno, the earliest cynoPGCs are identified before primitive streak formation at day 11 in the amnion (Sasaki et al., 2016). Using microfluidics to create posteriorized 3D-modeled embryos, we could identify nascent hPGCLCs in both the amnion as well as the pre-gastrula region 48 h after BMP4 exposure (Zheng et al., 2019). BMP4 was essential to create the posteriorized 3D-modeled embryos with competency to specify hPGCLCs, and similarly, BMP4 is also essential to hPGC specification in the 3D-disorganized aggregates used in the current study (Sasaki et al., 2015). The tcURD analysis of the 3D-disorganized aggregates suggests that in addition to hPGCLCs, BMP4 stimulation creates cells with amnion, gastrulating trophoblast and extra embryonic mesenchyme cell identity in the aggregate. Interestingly, 48 h of BMP4 exposure in the posteriorized 3D embryo models leads to the differentiation of hPGCLCs with the highest similarity to the day 2–day 3 hPGCLCs from the disorganized aggregates. In contrast, the day 4 hPGCLCs from the disorganized aggregates were more closely related to the fully specified ecPGCs from the cyno embryo. Given that this relationship appeared to be time dependent, once the technology for maintaining organized 3D embryo models for longer periods of time is achieved, it will be valuable to determine whether the hPGCLCs from the 3D embryo models progresses closer to fully specified cynoPGC or day 4 hPGCLCs.

The advantage of the organized 3D embryo models over the disorganized 3D aggregate is that the site of nascent hPGCLC specification can be evaluated. In the 3D embryo models, hPGCLCs are identified in both the amnion as well as in the pre-gastrula region 48 h after BMP4 exposure. Whether hPGCLCs are first specified in the amnion before migrating into the pre-gastrula regions of the 3D embryo models remains to be determined. In the porcine and mouse peri-implantation embryos, PGCs are identified in the pre-mesendoderm around the time of primitive streak formation (Kobayashi et al., 2017), arguing for a pre-gastrula origin in these species. Without positional information, our germline trajectory using the disorganized aggregates argues for two sites of specification (amnion and gastrula), with hPGCLCs being specified from a flexible TFAP2A-positive embryonic progenitor that has the potential to give rise to multiple somatic fates.

Specification of hPGCs requires TFAP2C and SOX17 (Chen et al., 2018; Irie et al., 2015; Kobayashi et al., 2017; Kojima et al., 2017). However, of these two TFs, only SOX17 can

induce hPGCLC fate from pre-mesendoderm cells in the absence of growth factor signaling (Kobayashi et al., 2017). SOX17 is also a master regulator of endoderm differentiation. Therefore, the mechanism of action for SOX17 in hPGC specification verses endoderm identity must be controlled by a germline determinant. Our study demonstrates that prior to SOX17 upregulation in the germline trajectory, the immediate progenitors are restrained under Weismann's barrier and that TFAP2C acts at the point of hPGCLC specification to upregulate SOX17 in the hPGCLCs. Therefore, without TFAP2C to protect the germline integrity, the TFAP2A-positive progenitors ''cross'' Weismann's barrier to differentiate into gastrulating cells. In contrast, the formation of amnion is unaffected in the absence of TFAP2C. Thus, our data illustrate that the first role for TFAP2C in the germline is not to induce PGC fate per se but rather to safeguard Weismann's barrier at the point of hPGC specification and to induce SOX17 in the response to BMP4 from a TFAP2A-positive progenitor and to establish naive-like pluripotency in the germline (Figure 7).

STAR★METHODS

LEAD CONTACT AND MATERIALS AVAILABILITY

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). All unique/stable reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human pre-implantation embryo samples—Use of human embryos in this research project followed California State law. This requires review by the Institutional Review Board (IRB) and the human embryonic stem cell research oversight committee (ESCRO). These committees approve the process of informed consent, and experiments using human embryos for research purposes. Patients were not paid for participation, and all donors were informed that the embryos would be destroyed as part of the research study and that the donated embryos would not be used to make a baby. All research with human embryos in this study complied with the principles laid out in the International Society for Stem Cell Research. Human blastocysts at day (D6), D7 and D8 post fertilization were used in this study (See Table S1). The human blastocysts were thawed using Vit Kit-Thaw (Irvine Scientific) according to manufacturer protocol. After thawing, embryos were cultured overnight in 5%O2, 6% CO2 at 37oC, and the zona pellucida was removed with Tyrode's acidified solution (Irvine Scientific). Human embryo attachment culture was performed following the procedures of (Deglincerti et al., 2016; Shahbazi et al., 2016) using the commercially available medias IVC1 (*In Vitro* Culture 1) and IVC2 from Cell Guidance Systems.

Human Fetal Gonad samples—Use of human fetal tissue in this research project was reviewed by the UCLA IRB and the project was determined as being exempt under 45 CRF 46.102(f). 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. Human genital ridges were obtained from Professor Katja-Layland, Fraunhofer Institute for Interfacial Engineering and Biotechnology, Stuttgart, Germany, and Department of Women's Health, Research Institute for Women's Health, Tubingen, Germany. 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. The human gonad samples used in this study include week 4 female and week 5 male conceptus, and a day 53 female fetal ovary and a day 59 male fetal testis.

Human ESC culture—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), 100mM L-Glutamine (GIBCO, 25030-081), 1x MEM Non-Essential Amino Acids (NEAA) (GIBCO, 11140-050), 55mM 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. Mycoplasma test (Lonza, LT07-418) was performed every month. All experiments were approved by the UCLA Embryonic Stem Cell Research Oversight Committee.

METHOD DETAILS

hPGCLC induction—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-Strep-tomycin-Glutamine (GIBCO, 10378-016), 1mM sodium pyruvate (GIBCO, 11360-070), 50ng/mL Activin A (Peprotech, AF-120-14E), 3mM CHIR99021 (Stemgent,04-0004), 10mM 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 ultralow cell attachment U-bottom 96-well plates (Corning, 7007) at the density of 3,000 cells/well in 200ml/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), 10mM 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—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).

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 95C 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).

For staining of human embryos, attachment cultures were washed once with PBS and fixed in 4% PFA (Electron Microscopy Sciences) at room temperature for 20 min followed by 2 washes in PBS. Attachments were permeabilized in 0.5% Triton-X (Sigma) in PBS for 20 min at room temperature and then washed once in wash buffer 0.1% Tween-20 (Sigma) in PBS. Samples were blocked in 10% normal donkey serum (Jackson ImmunoResearch) in 0.1% Tween-20/PBS at room temperature and incubated with primary antibody diluted in block overnight at 4°C.

Images were taken using LSM 780 Confocal Instrument (Zeiss). The primary antibodies used for immunofluorescence in this study include: rabbit-anti-NANOG (Cell Signaling Technology, 4903, 1:100), mouse-anti-TFAP2C (Santa Cruz Biotechnology, sc12762, 1:100), goat-anti-SOX17 (Neuromics, GT15094, 1:100). The secondary antibodies used in this study were from Jackson ImmunoResearch Laboratories and Invitrogen. Samples were covered in drops of SlowFade Diamond Antifade mountant with DAPI (Invitrogen) and imaged on an LSM780 or LSM880 confocal microscope (Zeiss). All images were processed by Imaris software.

Single cell RNA-sequencing—Single cell suspension was made from hESCs, iMeLCs, day 1 to day 4 aggregates after trypsin treatment. Single cells were washed three to five times in PBS supplemented with 0.04% BSA. Single cell suspension was sequenced using 10X Genomics.

ChIP-sequencing—ChIP-seq of hESCs, iMeLCs, and day 4 aggregates with anti-TFAP2C antibodies was performed as previously described (Pastor et al., 2018). Two biological replicates were performed using about 500k cells. hESCs and iMeLCs were dissociated with 0.05% Trypsin-EDTA (GIBCO, 25300-054) for 5 minutes at 37°C, and aggregates were dissociated with 0.05% Trypsin-EDTA (GIBCO, 25300-054) for 10 minutes

at 37°C, followed by washing 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 wash 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 uL 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 μL of the supernatant was saved as input. 30 μL 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 mL Dilution Buffer and added to the samples. The samples with beads were rotated for 2 hours 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, 60uL of pre-washed Protein A Dynabeads was added to each sample and incubated at 4°C for 2 hours. 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 μL 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 1.5uL of 10mg/mL RNaseA (PureLink RNase A, Invitrogen 12091-021) for 30 min at 65°C and then with 10ul of 10mg/mL Proteinase K for 2 hours at 56°C. Samples were purified with MinElute PCR Purification Kit (QIAGEN, 28006). ChIP-seq libraries were made using Ovation Ultralow Systems V2 (NuGEN, 0344) according the instruction. For ChIP-seq of sorted day4 hPGCLCs using anti-H3K27ac antibodies, about 5000 sorted hPGCLCs were collected and ChIP was performed using Low Cell ChIP-Seq Kit (ACTIVE MOTIF, 53084).

QUANTIFICATION AND STATISTICAL ANALYSIS

Experimental design—Single cell RNA-seq libraries were made from tow biological replicates of UCLA1, UCLA2 and TFAP2C mutant hESCs, iMeLCs, and day 1 to day 4 aggregates. ChIP-seq libraries were made from two biological replicates of UCLA1 hESCs, iMeLCs, day4 aggregates with anti-TFAP2C antibodies and sorted hPGCLCs with anti-H3K27ac antibodies. No statistical calculation was used to estimate the sample size. Randomization/stratification/blinding were not applicable to these experiments. For flow cytometry, at least ten biological replicates were performed. For statistical analysis, mean and standard error of the mean and t test were calculated using Graphpad Prism software

Replicates and data pooling—To call ChIP-seq peaks or display ChIP-seq data in Figure 5, all reads from same condition were merged to increase coverage. For single cell RNA-seq, two biological replicates were merged after batch correction.

Single-cell RNA-seq data analysis—The reads from single cell RNA-seq were aligned to the GRCh38 human reference genome using the Cell Range v.2.2.0 pipeline with default parameters (10x Genomes). The generated cell-by-gene unique molecular identifier (UMI) count matrix was analyzed using the Seurat R package v.2.3.4. We only kept the cells expressing at least 200 genes and genes with expression in at least 3 cells. The cells were also filtered by the maximum of 8000 expressed genes and of 20% mitochondrial genes. The UMI counts were then normalized for each cell by the total expression, multiplied by 10000, and log-transformed. We used Seurat's default method to identify highly variable genes and scale data for regressing out variation from UMI and mitochondrial genes. The scaled data with variable genes was used to perform principal component analysis (PCA). The top 50 principal components were chosen for further analysis, including clustering to identify cell populations. UMAPs were calculated by RunUMAP function in Seurat package using top 50 PCs and min_dist = 0.75.

To do batch correction between two biological replicates, we used Seurat's canonical correlation analysis procedure with default parameter. Briefly, this analysis identifies vectors with correlation between datasets and then aligns values along these vectors to reduce the variation from batch. The top 30 canonical correlation vectors were used further for clustering and UMAP visualization.

Prediction of developmental trajectories using tcURD—The matrix of cell X top 30 CCA components after batch correction was used to build the tree of lineages by tcURD, which is based on URD package (v.1.0.2) in R (Farrell et al., 2018) and do pseudotime analysis afterward. We modified URD as tcURD where the diffusion map is calculated by destiny package from a sparse distance matrix, with 0 for most of element except those from nearest neighbors. Instead of calculating distance from top K nearest neighbors for each cell among all other cells, we put a time constraint that nearest neighbors could only come from samples at the same or nearby time point for this time series experiment so that we could avoid artificial connections of two cells from samples far away from each other in temporal order. The pipeline can be found on GitHub (https://github.com/kdyhl/singleCell_tcURD).

Annotation of cell lineage—We re-processed the single cell RNA-seq data in 390 single cells of cynomolgus embryos across 13 cell types (GEO: GSE74767). Specifically, we identified differentially expressed genes (DEGs) across 13 cell types and generated the average expression profiles for 13 cell types based on these DEGs. The average profiles for cells at the stage of hPGCLC day4 were also generated in the same way across the segments in the URD tree. We next calculated the Pearson Correlation Coefficient (PCC) to evaluate the similarity between each lineage in the URD tree and cell type in cyno embryo. The similar analysis was carried out to define the cell identity of clusters in *ex vivo* single cell data.

Signature analysis—We downloaded the processed single cell RNA-seq data of cynomolgus (cyno) embryos (GEO: GSE74767, GSE76267) and identified the highly expressed genes as signature for each cell type annotated in the original paper compared with cells from all different cell identities. Next, we calculated the average expression level of each signature set in single cells of germline trajectory and visualized by heatmap. We also downloaded the processed single cell RNA-seq data from naive and primed hESCs (GEO: GSE76970) and did the similar analysis. Similar analysis was performed in single cell RNA-seq data of human synthetic amnion-like cells (hsAM) and PGCLCs (hsPGCLC) derived by exposing the H9 hESC line to BMP4 for 48 hours using the synthetic embryonic-like sac model.

Identification of germline trajectory differentially expressed genes (gtDEGs)—We defined germline trajectory differentially expressed genes (gtDEGs) as below: (1) we identified the pseudotime correlation genes with absolute Pearson correlation coefficient (PCC) > 0.2 between pseudotime and gene expression level. (2) we called the differentially expression genes between any two time points across germline trajectory using FindAllMarkers function in Seurat package with default parameters. (3) the union of geneset (1) or (2) were defined as gtDEGs for further analysis.

Motif analysis—To predict the upstream regulators of gtDEGs, we identified the enrichment of transcription factor binding motifs in the promoter region (500bp upstream to 100bp downstream of transcription start site) using findMotifs.pl function in HOMER v4.9 software with default parameters. Those transcription factors were then filtered by gtDEGs.

Transcription factor (TF) network analysis—Among gtDEGs, we focused on the regulatory relationship between transcription factors. We build a TF-TF regulatory network based on the presence of transcription factor binding motif at promoter regions (500bp upstream to 100bp downstream of transcription start site). The network was visualized in Cytoscape v.3.7.1(Shannon et al., 2003). We annotated nodes and edges. The color of node represents the pseudotime of its highest expression levels from red to yellow. The size of node represents the degree (number of edges). The shape of node represents whether it belongs to upstream regulator (triangle: regulator; circle: non-regulator). The arrow of edge represents the direction of regulation.

ChIP-seq analysis—FastQC were used to check the quality of the sequencing reads. Then reads were mapped to the reference hg19 genome using bowtie by allowing maximal two mismatches per reads and take one best hit when the reads can be mapped to multiple places. PCR duplciates were removed using SAMtools and ChIPseq tracks were generated with deeptools.

Code Availability—Custom scripts used for demultiplexing NGS reads, calculating count matrix, visualization for single cell maps, generating differentiation trajectory and annotating cell lineages will be made available upon request.

DATA AND CODE AVAILABILITY

The accession number for the single cell RNA-seq data and ChIP-seq data of key cell types including hESCs, iMeLCs, day 1 to day4 aggregates, and sorted hPGCLCs reported in this paper is GEO: GSE140021. The pipeline for tcURD can be found on GitHub (https://github.com/kdyhl/singleCell_tcURD).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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 UCLA Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research (BSCRC) for critical assistance with human subject and embryonic stem cell research oversight review. The authors would also like to thank Professor Katja-Layland from the Fraunhofer Institute for Interfacial Engineering and Biotechnology, Stuttgart, Germany, and the Department of Women's Health, Research Institute for Women's Health, Tubingen, Germany for the early 4- to 5-week de-identified human genital ridge samples. This work is supported by funds from an anonymous donor, as well as NIH/NICHD R01 HD079546 (A.C.), NIH NCATS UCLA CTSI grant number UL1TR0001881, and the UCLA David Geffen School of Medicine Regenerative Medicine Theme Award (A.C.). All human pre-implantation embryo and human embryo attachment culture studies were performed using funds from the UCLA Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research Innovation Award. No NIH funds were used for research with human pre-implantation embryos. Human fetal tissue research is supported by a grant to Ian Glass at the University of Washington Birth Defects laboratory, 5R24HD000836-53. Human conceptus tissue requests can be made to bdrl@u.washington.edu.

REFERENCES

Blakeley P, Fogarty NME, Del Valle I, Wamaitha SE, Hu TX, Elder K, Snell P, Christie L, Robson P, and Niakan KK (2015). Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. Development 142, 3151–3165. [PubMed: 26293300]

Butler A, Hoffman P, Smibert P, Papalexi E, and Satija R (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol 36, 411–420. [PubMed: 29608179]

Chen D, Liu W, Lukianchikov A, Hancock GV, Zimmerman J, Lowe MG, Kim R, Galic Z, Irie N, Surani MA, et al. (2017). Germline competency of human embryonic stem cells depends on eomesodermin. Biol. Reprod 97, 850–861. [PubMed: 29091993]

Chen D, Liu W, Zimmerman J, Pastor WA, Kim R, Hosohama L, Ho J, Aslanyan M, Gell JJ, Jacobsen SE, and Clark AT (2018). The TFAP2C-Regulated OCT4 Naive Enhancer Is Involved in Human Germline Formation. Cell Rep 25, 3591–3602.e5. [PubMed: 30590035]

Deglincerti A, Croft GF, Pietila LN, Zernicka-Goetz M, Siggia ED, and Brivanlou AH (2016). Self-organization of the in vitro attached human embryo. Nature 533, 251–254. [PubMed: 27144363]

Diaz Perez SV, Kim R, Li Z, Marquez VE, Patel S, Plath K, and Clark AT (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. [PubMed: 22058289]

Extavour CG, and Akam M (2003). Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. Development 130, 5869–5884. [PubMed: 14597570]

Farrell JA, Wang Y, Riesenfeld SJ, Shekhar K, Regev A, and Schier AF (2018). Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science 360, eaar3131. [PubMed: 29700225]

Gkountela S, Li Z, Vincent JJ, Zhang KX, Chen A, Pellegrini M, and Clark AT (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. [PubMed: 23242216]

Gkountela S, Zhang KX, Shafiq TA, Liao W-W, Hargan-Calvopiña J, Chen P-Y, and Clark AT (2015). DNA Demethylation Dynamics in the Human Prenatal Germline. Cell 161, 1425–1436. [PubMed: 26004067]

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. [PubMed: 26046443]

Irie N, Weinberger L, Tang WWC, Kobayashi T, Viukov S, Manor YS, Dietmann S, Hanna JH, and Surani MA (2015). SOX17 is a critical specifier of human primordial germ cell fate. Cell 160, 253–268. [PubMed: 25543152]

Kobayashi T, Zhang H, Tang WWC, Irie N, Withey S, Klisch D, Sybirna A, Dietmann S, Contreras DA, Webb R, et al. (2017). Principles of early human development and germ cell program from conserved model systems. Nature 546, 416–420. [PubMed: 28607482]

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. [PubMed: 28985527]

Magnúsdóttir E, and Surani MA (2014). How to make a primordial germ cell. Development 141, 245–252. [PubMed: 24381195]

Magnúsdóttir E, Dietmann S, Murakami K, Günesdogan U, Tang F, Bao S, Diamanti E, Lao K, Gottgens B, and Azim Surani M (2013). A tripartite transcription factor network regulates primordial germ cell specification in mice. Nat. Cell Biol 15, 905–915. [PubMed: 23851488]

McInnes L, Healy J, and Melville J (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction arXiv, arXiv, 1802.03426. https://arxiv.org/abs/1802.03426.

Messmer T, von Meyenn F, Savino A, Santos F, Mohammed H, Lun ATL, Marioni JC, and Reik W (2019). Transcriptional heterogeneity in naive and primed human pluripotent stem cells at single-cell resolution. Cell Rep 26, 815–824.e4. [PubMed: 30673604]

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. [PubMed: 23913270]

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. [PubMed: 27556940]

Ohinata Y, Payer B, O'Carroll D, Ancelin K, Ono Y, Sano M, Barton SC, Obukhanych T, Nussenzweig M, Tarakhovsky A, et al. (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. Nature 436, 207–213. [PubMed: 15937476]

Ohinata Y, Ohta H, Shigeta M, Yamanaka K, Wakayama T, and Saitou M (2009). A signaling principle for the specification of the germ cell lineage in mice. Cell 137, 571–584. [PubMed: 19410550]

Pastor WA, Liu W, Chen D, Ho J, Kim R, Hunt TJ, Lukianchikov A, Liu X, Polo JM, Jacobsen SE, and Clark AT (2018). TFAP2C regulates transcription in human naive pluripotency by opening enhancers. Nature Cell Biology 20, 553–564. [PubMed: 29695788]

Petropoulos S, Edsgärd D, Reinius B, Deng Q, Panula SP, Codeluppi S, Plaza Reyes A, Linnarsson S, Sandberg R, and Lanner F (2016). Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. Cell 165, 1012–1026. [PubMed: 27062923]

Poirion OB, Zhu X, Ching T, and Garmire L (2016). Single-Cell Transcriptomics Bioinformatics and Computational Challenges. Front. Genet 7, 163. [PubMed: 27708664]

Popovic M, Bialecka M, Gomes Fernandes M, Taelman J, Van Der Jeught M, De Sutter P, Heindryckx B, and Chuva De Sousa Lopes SM (2019). Human blastocyst outgrowths recapitulate primordial germ cell specification events. Mol. Hum. Reprod 25, 519–526. [PubMed: 31211841]

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. [PubMed: 26189426]

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. [PubMed: 27720607]

Shahbazi MN, Jedrusik A, Vuoristo S, Recher G, Hupalowska A, Bolton V, Fogarty NNM, Campbell A, Devito L, Ilic D, et al. (2016). Self-organization of the human embryo in the absence of maternal tissues. Nat. Cell Biol 18, 700–708. [PubMed: 27144686]

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13, 2498–2504. [PubMed: 14597658]

Shao Y, Taniguchi K, Gurdziel K, Townshend RF, Xue X, Yong KMA, Sang J, Spence JR, Gumucio DL, and Fu J (2017a). Self-organized amniogenesis by human pluripotent stem cells in a biomimetic implantation-like niche. Nat. Mater 16, 419–425. [PubMed: 27941807]

Shao Y, Taniguchi K, Townshend RF, Miki T, Gumucio DL, and Fu J (2017b). A pluripotent stem cell-based model for post-implantation human amniotic sac development. Nat. Commun 8, 208. [PubMed: 28785084]

Stegle O, Teichmann SA, and Marioni JC (2015). Computational and analytical challenges in single-cell transcriptomics. Nat. Rev. Genet 16, 133–145. [PubMed: 25628217]

Stirparo GG, Boroviak T, Guo G, Nichols J, Smith A, and Bertone P (2018). Integrated analysis of single-cell embryo data yields a unified transcriptome signature for the human pre-implantation epiblast. Development 145, dev158501. [PubMed: 29361568]

Tang WWC, Dietmann S, Irie N, Leitch HG, Floros VI, Bradshaw CR, Hackett JA, Chinnery PF, and Surani MA (2015). A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. Cell 161, 1453–1467. [PubMed: 26046444]

Tang WWC, Kobayashi T, Irie N, Dietmann S, and Surani MA (2016). Specification and epigenetic programming of the human germ line. Nat. Rev. Genet 17, 585–600. [PubMed: 27573372]

Weber S, Eckert D, Nettersheim D, Gillis AJM, Schäfer S, Kuckenberg P, Ehlermann J, Werling U, Biermann K, Looijenga LHJ, and Schorle H (2010). Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance. Biol. Reprod 82, 214–223. [PubMed: 19776388]

Weismann A (1893). The Germ-Plasm: A Theory of Heredity (Scribner)

Williamson and Lehmann. (1996). Germ cell development in Drosophila. Annu. Rev. Cell Dev. Biol 12, 365–391. [PubMed: 8970731]

Yamaji M, Seki Y, Kurimoto K, Yabuta Y, Yuasa M, Shigeta M, Yamanaka K, Ohinata Y, and Saitou M (2008). Critical function of Prdm14 for the establishment of the germ cell lineage in mice. Nat. Genet 40, 1016–1022. [PubMed: 18622394]

Yokobayashi S, Okita K, Nakagawa M, Nakamura T, Yabuta Y, Yamamoto T, and Saitou M (2017). Clonal variation of human induced pluripotent stem cells for induction into the germ cell fate. Biol. Reprod 96, 1154–1166. [PubMed: 28453617]

Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun 8, 14049. [PubMed: 28091601]

Zheng Y, Xue X, Shao Y, Wang S, Esfahani SN, Li Z, Muncie JM, Lakins JN, Weaver VM, Gumucio DL, and Fu J (2019). Controlled modelling of human epiblast and amnion development using stem cells. Nature 573, 421–425. [PubMed: 31511693]

Highlights

- Human germline cell specification begins from a transitional pluripotent state
- Human primordial germ cells are specified from lineage-primed progenitors
- Lineage-primed TFAP2A progenitors have gastrulating and amnion cell identity
- TFAP2C regulates SOX17 at the point of human primordial germ cell specification

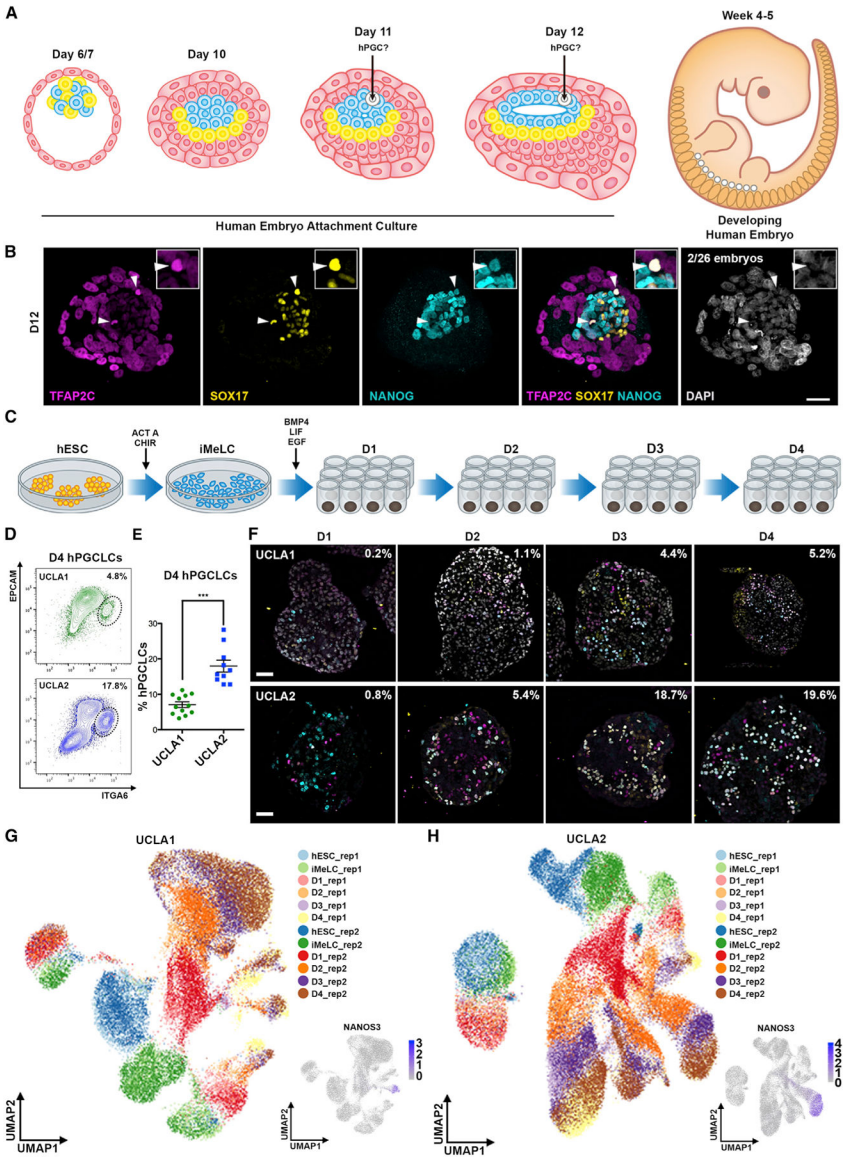


Figure 1. Tracking the Specification of Human PGCs

(A) Schematic illustration of human embryos used in this study. Frozen embryos at day 5 or day 6 post-fertilization were thawed and allowed to recover for 24 h to become day 6 or day 7, staining or culturing until day 12 as human embryo attachment cultures. Human embryos at weeks 4–5 were stained as positive controls for identifying hPGCs.

(B) Human embryo attachment culture at day 12 stained for TFAP2C (purple), SOX17 (yellow), and NANOG (cyan). Two TFAP2C/SOX17/NANOG triple-positive cells are highlighted by a white arrowhead, one of which is enlarged in inset. The number in the 4′,6-diamidino-2-phenylindole (DAPI) (gray) panel indicates triple-positive cells were identified in two day 12 embryos from a total of 26 embryos examined at this time point.

(C) Schematic illustration of hPGCLC differentiation. The UCLA1 and UCLA2 hESC lines are induced to differentiate into iMeLCs for 24 h, followed by differentiation as 3D

disorganized aggregates for 4 days. Samples were collected at all six time points from each hESC line in biological duplicate for 10x Genomics scRNA-seq.

(D) Identification of ITGA6/EPCAM double-positive hPGCLCs by flow cytometry from UCLA1 and UCLA2.

(E) Percentage of ITGA6/EPCAM double-positive hPGCLCs isolated from UCLA1 and UCLA2 at day 4 of differentiation. ***, indicates statistically significant difference comparing more than 10 biological replicates of UCLA1 and UCLA2 at day 4 of differentiation (shown is mean and SEM)

(F) Immunofluorescence of TFAP2C (purple), SOX17 (yellow), and NANOG (cyan) in day 1–4 aggregates from UCLA1 (top) and UCLA2 (bottom). The percentage of TFAP2C/SOX17/NANOG triple-positive cells at each day is shown. Percentages were calculated from five or more biological replicates at each day.

(G and H) UMAP display of all cells in the UCLA1 (G) and UCLA2 (H) dataset in biological duplicate (n = 12 samples for each cell line). The expression of *NANOS3* indicates hPGCLCs.

Scale bars: 50 μM. See also Figure S1.

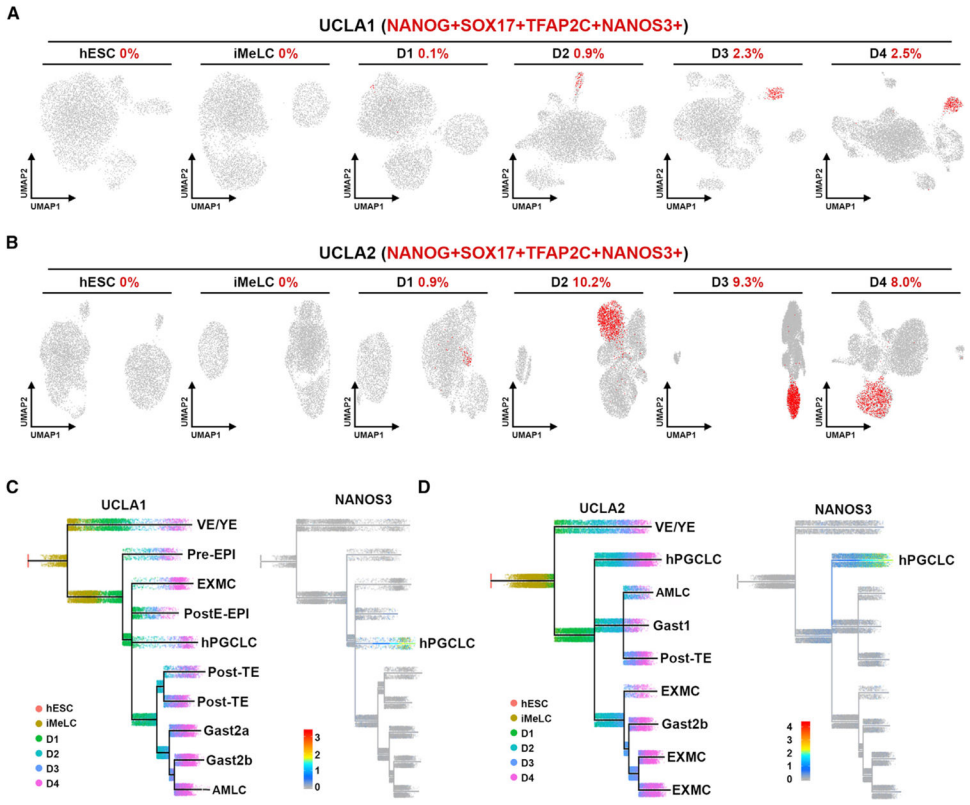


Figure 2. Highly Germline-Competent hESC Lines Specify More hPGCs within 48 h of BMP4 Exposure

(A and B) The emergence of *TFAP2C*/*SOX17*/*NANOG*/*NANOS3* quadruple-positive cells in UCLA1 (A) and UCLA2 (B) by scRNA-seq. UMAP was applied to visualize all single cells at each stage. The percentage of *TFAP2C*/*SOX17*/*NANOG*/*NANOS3* quadruple-positive cells (red) is shown. Cutoff: log-transformed-normalized UMI counts of >0.5 for *TFAP2C*, *SOX17*, *NANOG*, and *NANOS3*.

(C and D) URD tree showing the developmental trajectories from hESC and iMeLCs through 4 days of aggregate differentiation from UCLA1 (C) and UCLA2 (D). The expression of *NANOS3* marks the end of the germline trajectory and establishment of hPGCLCs in (C) and (D). Cells with endoderm (Endo) identity are outside of the germline trajectory, indicating hPGCLCs do not originate from *SOX17*-positive endoderm. VE/YE, visceral/yolk-sac endoderm; EXMC, extra-embryonic mesenchyme; Pre-EPIs, pre-implantation epiblasts; PostE-EPI, post-implantation early epiblasts; Post-TE, post-implantation trophectoderm; Gast, gastrulating cells; AMLCs, amnion-like cells (Nakamura et al., 2016; Zheng et al., 2019).

See also Figure S2.

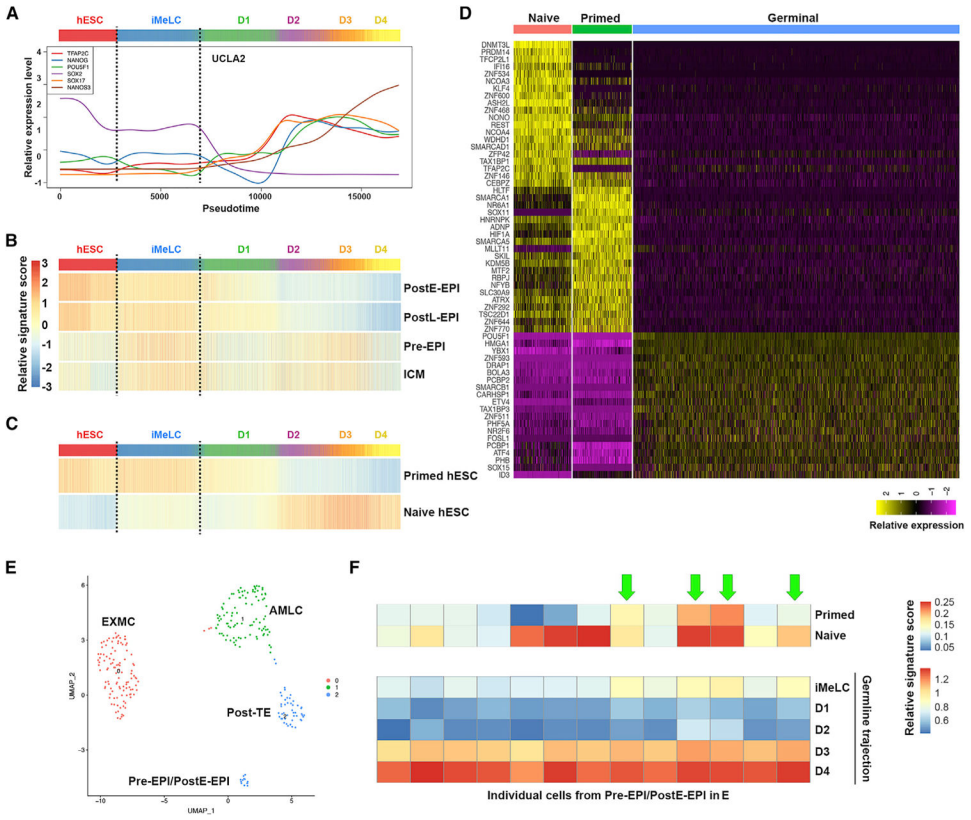


Figure 3. hPGCLC Specification through a Translational Germinal Pluripotent State

(A) The expression patterns of *TFAP2C*, *NANOG*, *SOX2*, *POU5F1*(*OCT4*), *SOX17*, and *NANOS3* within the germline trajectory of UCLA2. Dash lines outline iMeLCs in germline trajectory in (A)–(C).

(B) Relative average expression of signature genes from cynomolgus ICM (inner cell mass), Pre-EPIs, PostE-EPIs, and PostL-EPI (post-implantation late epiblast) (Nakamura et al., 2016) in germline trajectory of UCLA2.

(C) Relative average expression of signature genes from primed and naive hESCs (Messmer et al., 2019) in germline trajectory of UCLA2. Signature score bar is same as (B).

(D) Heatmap showing the expression of differentially expressed genes in naive, primed, and transitional germinal pluripotent cells.

(E) UMAP displaying the single cells from day 12 human embryonic cells.

(F) Relative expression of signature genes in the 13 epiblasts cells from (E) compared to primed and naive hESCs (Messmer et al., 2019) and cells from iMeLCs, day 1 to day 4 cells in germline trajectory.

See also Figure S3.

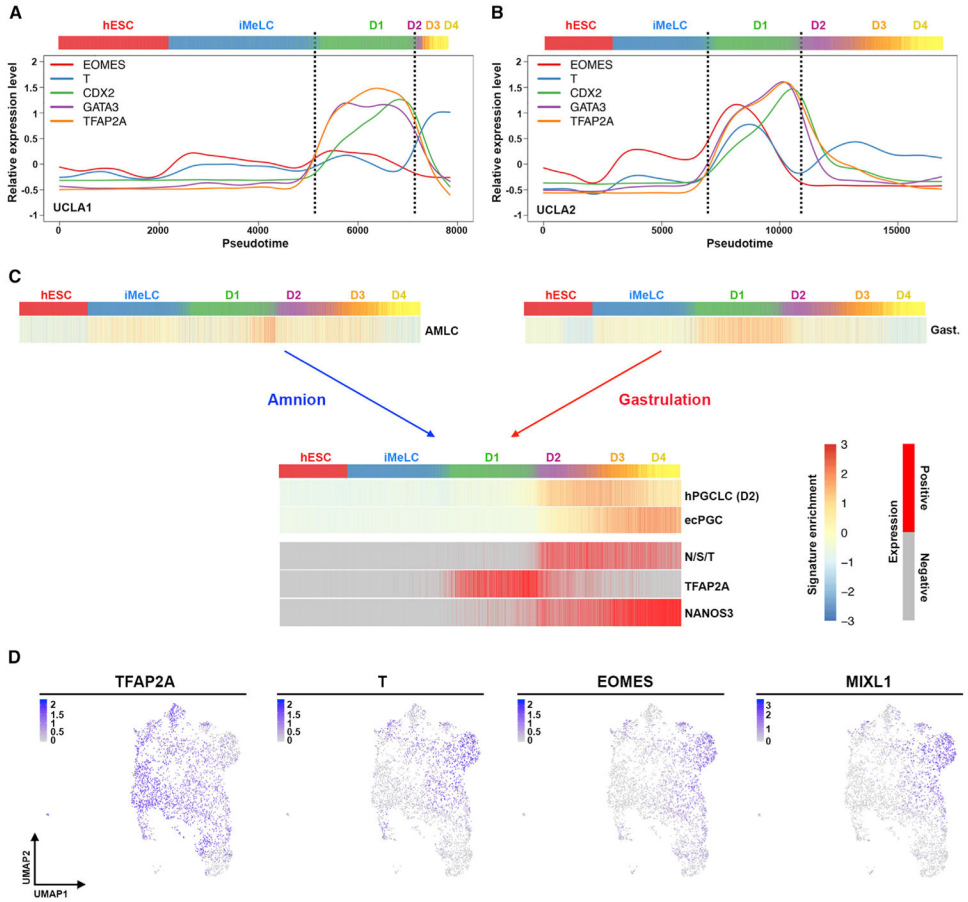


Figure 4. hPGCLC Specification Involves a Trajectory through a Transient TFAP2A-Positive State with a Signature of Both Amnion and Gastrulating Cell Fates

(A and B) The expression patterns of *EMOES*, *T*, *CDX2*, *GATA3*, and *TFAP2A* within germline trajectory of UCLA1 (A) and UCLA2 (B).

(C) Relative average expression of signature genes from cynomolgus Gast (gastrulating cells), ePGC (early PGCs) (Nakamura et al., 2016) and human AMLCs and hPGCLC at day 2 (Zheng et al., 2019) in germline trajectory of UCLA2.

(D) Expression of *TFAP2A*, *T*, *EOMES*, and *MIXL1* in day 1 progenitor population from UCLA2.

See also Figure S4.

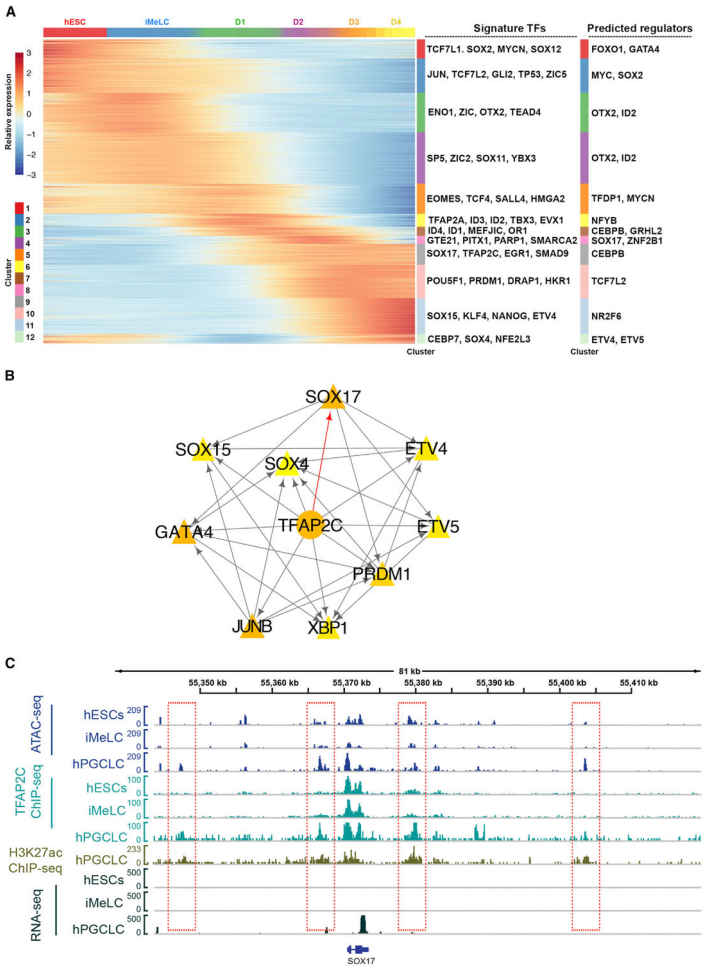


Figure 5. TFAP2C Acts Upstream of SOX17

(A) Heatmap showing the differentially expressed genes in the UCLA2 germline trajectory. Signature transcription factors (TFs) and predicted regulators are shown for each cluster.

(B) Network showing the regulatory role of TFAP2C and its putative targets in the germline trajectory of UCLA2. Red arrows highlight the predicted direct regulation of SOX17 by TFAP2C.

(C) Tracks showing the ATAC-seq, TFAP2C ChIP-seq, H3K27ac ChIP-seq, and RNA-seq peaks around *SOX17* site at each stage.

See also Figure S5.

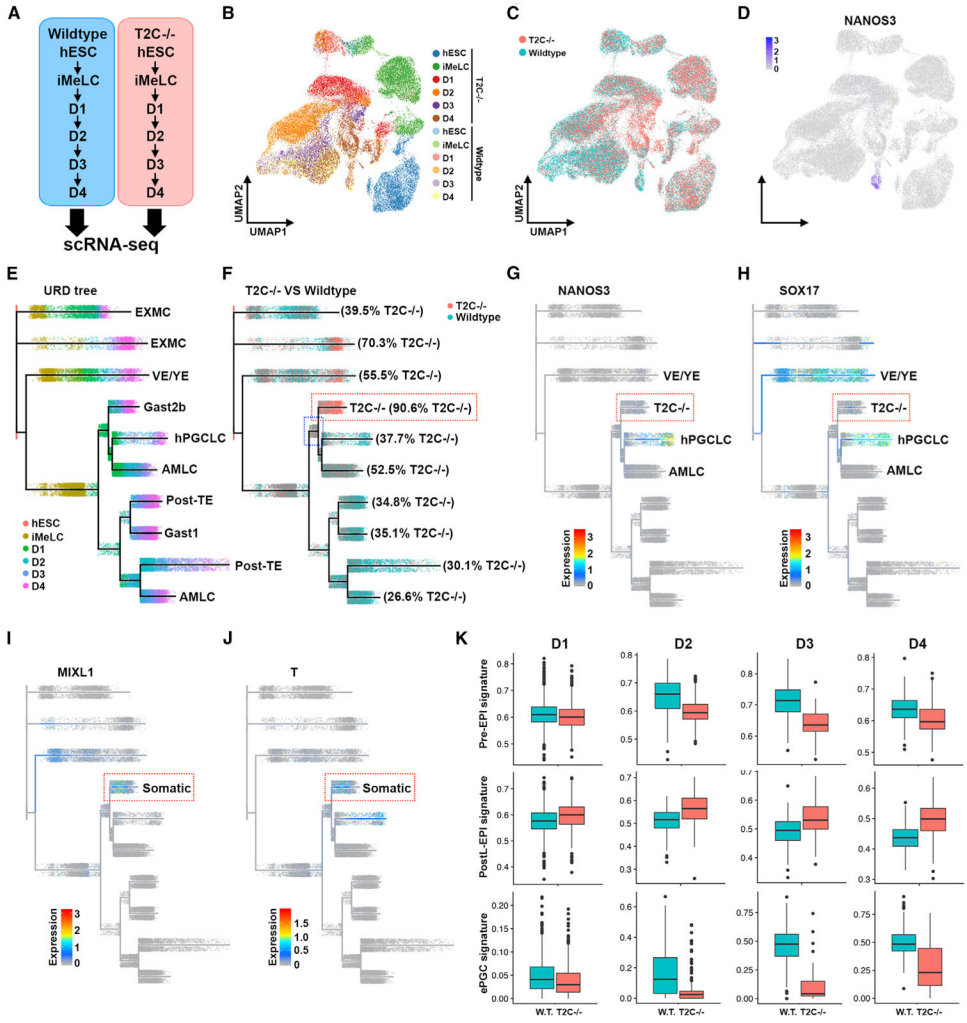


Figure 6. TFAP2C Mutant Cells Exit the Germline Trajectory to Become Primitive Streak or Amnion-like Somatic Cells

(A) Experimental design comparing scRNA-seq libraries of *TFAP2C* mutant (*T2C*$^{-/-}$) hESCs, iMeLCs, day 1, day 2, day 3, and day 4 to wild-type cells of the same genetic background (UCLA1).

(B) UMAP showing wild-type and *T2C*$^{-/-}$ single cells.

(C) Overlay from wild-type UCLA1 (blue) and *T2C*$^{-/-}$ cells (pink) from each of the six time points. Clusters that are largely pink are enriched in mutant cells; clusters that are largely blue are enriched in wild-type cells.

(D) *NANOS3* expression indicates hPGCLCs.

(E) URD trees showing the differentiation trajectories of all cells combined (wild-type and *T2C*$^{-/-}$ mutant).

(F) Overlay of *TFAP2C*$^{-/-}$ cells (pink) and wild-type UCLA1 (blue) cells in the URD trees illustrates cell trajectories enriched in either mutant (pink) or control cells (blue) cells. Percentages of *TFAP2C*$^{-/-}$ cells in each of the blades of the trajectories.

(G and H) *NANOS3* (G) and *SOX17* (H) highlight the germline trajectory ending in hPGCLCs (G) and hPGCLCs and VE/YE (H) Red box indicates the mutant-enriched trajectory that originates from the common progenitor with hPGCLCs.
(I and J) URD tree showing the expression of somatic (primitive-streak or early amnion genes) *MIXL1* (I) and *T* (J), in the *TFAP2C*$^{-/-}$ enriched blade.
(K) Comparison of cynomolgus Pre-EPI, PostL-EPI, and ePGCs (Nakamura et al., 2016) signatures in germline trajectory between wild-type and *TFAP2C* mutant cells.
See also Figure S5.

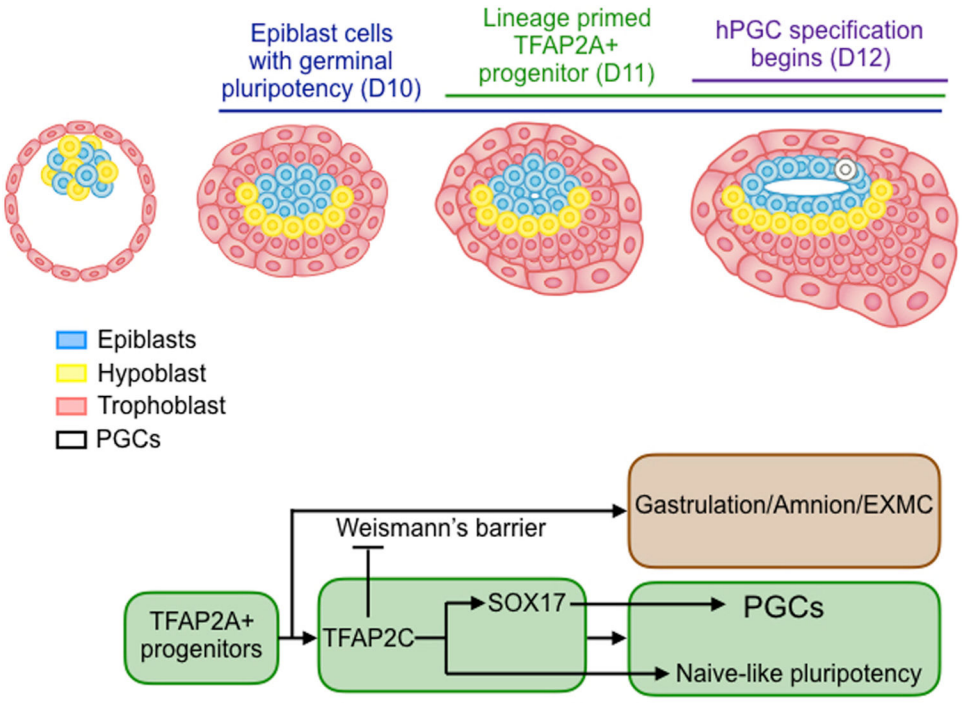


Figure 7. Proposed Model

Human PGCs are specified beginning at day 12. Based on the *in vitro* germline trajectory, we propose that hPGCs originate from a lineage-primed *TFAP2A*$^{+}$ progenitor at around day 11 which is specified from a transitional pluripotent state that shares characteristics with pre- and post-implantation epiblasts that we have called ‘‘germinal pluripotency.’’ *TFAP2A*$^{+}$ progenitors have the potential for hPGC specification and also for differentiation of somatic cells, including amnion and gastrulating cells. TFAP2C functions at the critical point of hPGC specification to directly regulate expression of the human germ cell fate determinant SOX17, while simultaneously preventing germline cells from crossing Weismann’s barrier to become somatic cells.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
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
Rabbit anti-human NANOG	Cell Signaling Technology	Cat#4903; RRID: AB_10829232
Rabbit anti-human-H3K27ac	Abcam	Cat#ab4729; RRID: AB_2118291
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	Cat#324208; RRID: AB_756082
Biological Samples		
Human pre-implantation embryos at day 6, day 7 and day 8 post-fertilization	Surplus embryos from *in vitro* fertilization clinics in the United States.	N.A.
Human week 4 female and week 5 male genital ridges	Fraunhofer Institute for Interfacial Engineering and Biotechnology, Germany	N.A.
Human day 53 female fetal ovary and day 59 male fetal testis	University of Washington Birth Defects Research Laboratory	N.A.
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	Peprotech	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		
Ovation Ultralow System V2	NuGEN	Cat#0344
Low Cell ChIP-Seq Kit	ACITVE MOTIF	Cat#53084
VIT KIT Thaw	Irvine Scientific	Cat#90137-SO
IVC medium	Cell Guidance Systems	Cat #M11-6
IVC medium	Cell Guidance Systems	Cat# M12-6
Deposited Data		
Single cell RNA-seq data	This paper	GSE140021
ChIP-seq data	This paper	GSE140021

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Single cell RNA-seq data	Zheng et al., 2019	GSE134571
Single cell RNA-seq data	Nakamura et al., 2016	GSE74767
Single cell RNA-seq data	Sasaki et al., 2016	GSE76267
ATAC-seq data	Chen et al., 2018	GSE120648
ChIP-seq data	Pastor et al., 2018	GSE101074
Experimental Models: Cell Lines		
UCLA1	Diaz Perez et al., 2012	N/A
UCLA2	Diaz Perez et al., 2012	N/A
UCLA1 *TFAP2C*$^{-/-}$ line 1	Pastor et al., 2018	N/A
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