Stem Cell Reports
Article
ISSCR
OPEN ACCESS

The chromatin accessibility landscape reveals distinct transcriptional regulation in the induction of human primordial germ cell-like cells from pluripotent stem cells

Xiaoman Wang,[1,3,12] Veeramohan Veerapandian,[2,3,12] Xinyan Yang,[3,12] Ke Song,[3,12] Xiaoheng Xu,[3] Manman Cui,[3] Weiyan Yuan,[3] Yaping Huang,[3] Xinyu Xia,[3] Zhaokai Yao,[3] Cong Wan,[3] Fang Luo,[3] Xiuling Song,[3] Xiaoru Wang,[3] Yi Zheng,[3] Andrew Paul Hutchins,[6] Ralf Jauch,[7] Meiyan Liang,[2] Chenhong Wang,[1] Zhaoting Liu,[3,*] Gang Chang,[4,12,*] and Xiao-Yang Zhao[3,5,8,9,10,11,*]

[1]Shenzhen Hospital of Southern Medical University, Shenzhen, Guangdong, China
[2]Shunde Hospital of Southern Medical University, Shunde, Guangdong, China
[3]Department of Developmental Biology, School of Basic Medical Sciences, Southern Medical University, Guangzhou, Guangdong, China
[4]Guangdong Provincial Key Laboratory of Regional Immunity and Diseases, Department of Biochemistry and Molecular Biology, Shenzhen University Health Science Center, Shenzhen, Guangdong, China
[5]Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory), Guangzhou, Guangdong, China
[6]Department of Biology, Southern University of Science and Technology, Shenzhen, Guangdong, China
[7]School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China
[8]State Key Laboratory of Organ Failure Research, Department of Developmental Biology, School of Basic Medical Sciences, Southern Medical University, Guangzhou, Guangdong, China
[9]Key Laboratory of Mental Health of the Ministry of Education, Guangdong-Hong Kong-Macao Greater Bay Area Center for Brain Science and Brain-Inspired Intelligence, Southern Medical University, Guangzhou, Guangdong, China
[10]Department of Obstetrics and Gynecology, Zhujiang Hospital, Southern Medical University, Guangzhou, Guangdong, China
[11]National Clinical Research Center for Kidney Disease, Guangzhou, China
[12]These authors contributed equally
*Correspondence: liuzhaoting@i.smu.edu.cn (Z.L.), changgang@szu.edu.cn (G.C.), zhaoxiaoyang@smu.edu.cn (X.-Y.Z.)
https://doi.org/10.1016/j.stemcr.2021.03.032

SUMMARY

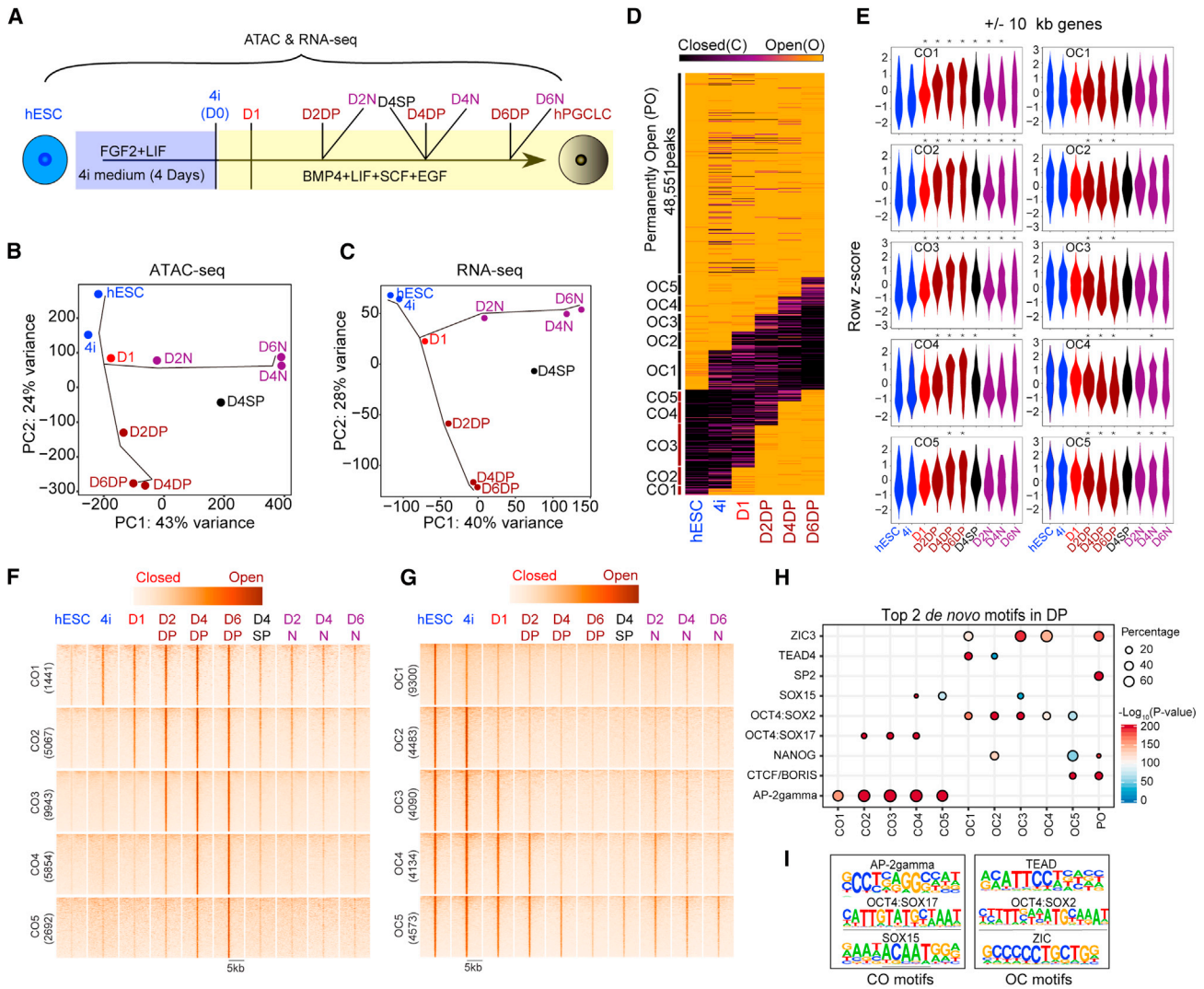
In vitro induction of human primordial germ cell-like cells (hPGCLCs) provides an ideal platform to recapitulate hPGC development. However, the detailed molecular mechanisms regulating the induction of hPGCLCs remain largely uncharacterized. Here, we profiled the chromatin accessibility and transcriptome dynamics throughout the process of hPGCLC induction. Genetic ablation of SOX15 indicated the crucial roles of SOX15 in the maintenance of hPGCLCs. Mechanistically, SOX15 exerted its roles via suppressing somatic gene expression and sustaining latent pluripotency. Notably, ETV5, a downstream regulator of SOX15, was also uncovered to be essential for hPGCLC maintenance. Finally, a stepwise switch of OCT4/SOX2, OCT4/SOX17, and OCT4/SOX15 binding motifs were found to be enriched in closed-to-open regions of human embryonic stem cells, and early- and late-stage hPGCLCs, respectively. Collectively, our data characterized the chromatin accessibility and transcriptome landscapes throughout hPGCLC induction and defined the SOX15-mediated regulatory networks underlying this process.

INTRODUCTION

The formation of human primordial germ cells (hPGCs) is critical for establishing the human germline and transmission of genetic information (Leitch et al., 2013). The recent development of *in vitro* differentiation protocols for human primordial germ cell-like cells (hPGCLCs) from human pluripotent stem cells (hPSCs) has minimized technical and ethical limitations inherent in using human tissues. This system has facilitated our understanding of hPGC biology, and might eventually provide a source of haploid germ cells for infertility treatments (Saitou and Miyauchi, 2016). However, the regulatory networks for germ cell and somatic lineage bifurcation are still unclear and the establishment of stable hPGCLCs and their further maturation remain challenging.

In mammals, primordial germ cells (PGCs) are specified from early embryonic cells through the sophisticated interactions between WNT and BMP pathways, which is highly conserved in humans, monkeys, pigs, and mice (Kobayashi et al., 2017). It has been reported that the transcription factors (TFs) BLIMP1 (PRDM1), TFAP2C, and PRDM14 are general regulators of PGC specification in both mice and humans (Irie et al., 2015; Sasaki et al., 2015; Sybirna et al., 2020). However, accumulating evidence indicated that the germ line specifications are actually quite different between humans and mice (Irie et al., 2015; Kobayashi et al., 2017; Kojima et al., 2017; Tang et al., 2015). For instance, the pluripotency factor SOX2 is essential for mouse PGC (mPGC) induction, but it is not expressed in human PGCs (Campolo et al., 2013; Perrett et al., 2008). Vice versa, SOX17 acts as a key regulator of initial induction of hPGCLCs, but is dispensable for that in mPGC specification (Irie et al., 2015; Kanai-Azuma et al., 2002).

The SOX family member SOX15, which shares a very similar HMG domain with SOX2 (Kamachi and Kondoh,



(legend on next page)

2013), is highly expressed in both hPGCs and mPGCs (Guo et al., 2015; Sarraj et al., 2003). Interestingly, the developmental defects due to Sox2 deficiency in mESCs can be rescued by overexpression of Sox15 (Niwa et al., 2016). Notably, a recent study uncovered the role of SOX15 in maintaining hPGCLC identity, but how SOX15 regulates hPGCLC induction is still unclear (Pierson Smela et al., 2019). Most SOX factors including SOX2, SOX17, and SOX15 bind to similar CATTGT-like DNA motifs (Hou et al., 2017; Maruyama et al., 2005). SOX2 and SOX15 also possess the ability to heterodimerize with OCT4 and bind a canonical SOXOCT motif composed of SOX and OCT half-sites (CATTGTCATGCAAAT-like) (Chang et al., 2017). The canonical SOXOCT motif is critical for the induction and maintenance of pluripotency in mice and humans (Aksoy et al., 2013a, 2013b; Jauch et al., 2011; Veerapandian et al., 2018). In addition, a recent study in seminoma cell lines revealed that the canonical SOXOCT motifs are bound by SOX17 to regulate pluripotency-related genes (Jostes et al., 2020). Therefore, it is speculated that OCT4/SOX17 or OCT4/SOX15 complexes exert overlapping regulatory roles in hPGCs or hPGCLCs.

In this study, we investigated the genome-wide chromatin changes and transcriptome dynamics in the process of hPGCLC induction via time course ATAC-seq (assay for transposase-accessible chromatin using sequencing) and RNA-seq (RNA sequencing) analyses. We obtained distinct patterns of CO/OC (closed-to-open/open-to-closed) loci that underlie the bifurcation of germline and non-germline lineage. The combined genetic ablation assay and integrated analysis of RNA-seq, ATAC-seq, and CUT&Tag-seq (cleavage under targets and tagmentation sequencing) demonstrated that SOX15 was crucial for the maintenance of hPGCLC identity by simultaneous somatic gene expression suppression and latent pluripotency preservation. ETV5, a downstream regulator of SOX15, was validated to be essential for hPGCLC maintenance. Moreover, in late-stage hPGCLCs, there was a switch toward utilization of an OCT4/SOX15, which was distinct from that in human embryonic stem cells (hESCs) (OCT4/SOX2) and early-stage hPGCLCs (OCT4/SOX17).

RESULTS

Chromatin accessibility and gene regulation dynamics during hPGCLC induction

To investigate the dynamic genome regulation during the induction of hPGCLCs from hESCs, we used a modified protocol based on a previous study (Mitsunaga et al., 2017) to obtain EpCAM$^+$/INTEGRIN$\alpha6^+$ (DP) and EpCAM$^-$/INTEGRIN$\alpha6^-$ (N) cells (Figure S1A). The PGC marker genes such as *TFAP2C* and *SOX17* were upregulated in EpCAM$^+$/INTEGRIN$\alpha6^+$ cells, while somatic genes such as *HOXA1* were upregulated in EpCAM$^-$/INTEGRIN$\alpha6^-$ cells (Figure S1B). We further confirmed the protein expression of OCT4, SOX17, and TFAP2C in embryoid bodies (EBs) at day 2, 4, and 6 via immunostaining (Figure S1C). We then performed a time course ATAC-seq and RNA-seq analysis throughout hPGCLC induction (Figure 1A). Principal-component analysis (PCA) revealed a cell-fate bifurcation between EpCAM$^+$/INTEGRIN$\alpha6^+$ and EpCAM$^-$/INTEGRIN$\alpha6^-$ cells along the trajectory of hPGCLC induction from day 1 (D1) onward (Figures 1B, 1C, S1D, and S1E).

We next used our ATAC-seq data to define the chromatin accessibility dynamics (Figures 1D and S1F) (Li et al., 2017). We defined the open chromatin peaks in each ATAC library using macs2 (Zhang et al., 2008) (Figures S1F and S1G) and grouped the open/closed regions as reported in previous studies (Li et al., 2017) (Figure 1D; Table S1). We could identify dynamically CO, OC, and permanently open (PO) regions. Many PO regions were enriched in the proximal promoters (Figure S1H). We then evaluated the gene expression patterns associated with the dynamic chromatin changes throughout hPGCLC induction, and

Figure 1. Chromatin accessibility and gene regulation dynamics during hPGCLC induction

(A) Schematic representation of time course ATAC-seq and RNA-seq library induction during the hPGCLC induction from hESCs. Day is represented as "D," EpCAM$^+$/INTEGRIN$\alpha6^+$ cells are represented as DP, EpCAM$^-$/INTEGRIN$\alpha6^-$ cells are represented as N, and EpCAM$^+$/INTEGRIN$\alpha6^-$ cells are represented as SP.

(B and C) PCA of ATAC-seq (B) and RNA-seq data (C). Cell types are labeled as described in (A) and two independent replicates are merged.

(D) Dynamically closed-to-open (CO), open-to-closed (OC), and permanently open (PO) chromatin regions are clustered and shown as a heatmap. CO, OC, and PO refer to closed in hESCs but open in D6DP, open in hESCs but closed in D6DP, and PO in both hESCs and D6DP, respectively.

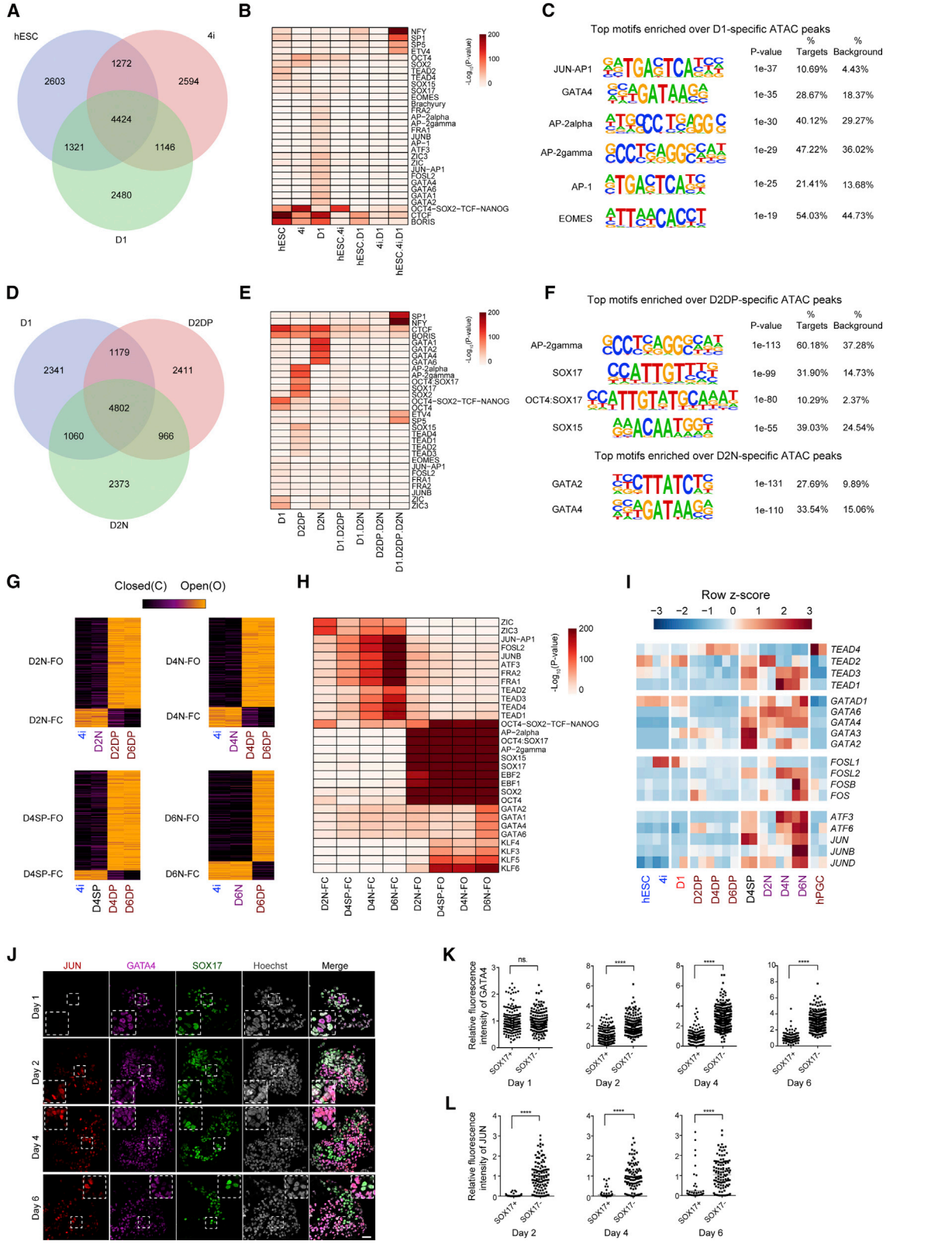
(E) Violin plots showing the expression levels of all genes with a TSS within 10 kb of an ATAC-seq peak for each CO or OC group. The Wilcoxon rank-sum test was performed. *$p < 0.01$.

(F and G) Heatmap showing the genome coverage of ATAC-seq signals on each CO (F) and OC (G) group.

(H) Bubble plot showing the top 2 *de novo* motifs in COs/OCs.

(I) Selected top ranked *de novo* motifs from CO (left) and OC (right).

(J) Representative genome coverage plots for ATAC-seq and RNA-seq signals for key germ cell genes.



(legend on next page)

observed a significant difference in gene expression patterns from D1 onward when compared with hESCs or 4i stage (Figure 1E). We then evaluated the OC1-5 and CO1-5 genomic regions in the N cells and found that N cells failed to close (FC) in OC3-5 and failed to robustly open in CO2-5 (Figures 1F and 1G).

To understand the mechanisms underlying the global chromatin dynamics, we measured the enrichment of TF binding motifs. AP-2gamma, OCT4:SOX17 (compressed SOXOCT motif) and single SOX15 motifs were enriched in the CO regions. In contrast, TEAD, OCT4:SOX2 (canonical SOXOCT motif), and ZIC motifs were enriched in the OC regions (Figure 1H). Interestingly, the compressed OCT4:SOX17 motif emerged in the chromatin regions that open early, while the single SOX15 motif was specially enriched at the sites that open later throughout PGCLC induction (CO4 and CO5; Figures 1H, 1I, S1I, and S1J). In addition, the DP cells exhibited expected stage-specific open chromatin signals at the TFAP2C, OCT4, SOX17, and BLIMP1 loci, and the corresponding transcripts were upregulated (Figure 1J). In summary, we comprehensively profiled the chromatin accessibility and transcriptome dynamics throughout hPGCLC induction, obtaining the specific CO/OC patterns and the enriched TF binding motifs.

Determination of the regulatory elements underlying cell-fate bifurcation of germline and non-germline lineages

Compared with EpCAM$^+$/INTEGRIN$\alpha6^+$ cells (DP cells committing to the germline lineage), EpCAM$^-$/INTEGRIN$\alpha6^-$ cells (N cells uncommitted to the germline lineage) were enriched with the binding motifs of representative somatic TFs, such as JUN-AP1, JUNB, and GATA motifs in the CO regions and pluripotency-associated TFs in the OC regions (Figures S2A–S2D). To elucidate the regulators accounting for the cell-fate bifurcation of germline and non-germline lineage, we first focused on the top 10k peaks from hESC, 4i, and D1 libraries and intersected the peaks from each library to obtain the specific and common peaks (Figure 2A). Interestingly, the top significantly enriched motifs over D1-specific peaks (2,480 peaks) are JUN-AP1/AP1, EOMES, AP-2alpha/AP-2gamma, and GATA binding motifs (Figures 2B, 2C, and S2E). It is well known that EOMES and TFAP2C (AP-2gamma) play critical roles in hPGCLC induction (Kojima et al., 2017); however, it is still unknown that, if the AP1 and GATA family TFs are essential for this process. We next compared the top 10 kb peaks from D1, D2DP, and D2N libraries and found that the D2DP-specific peaks were enriched with AP2, OCT4:SOX17, SOX17, and SOX15 motifs, while the D2N-specific peaks were enriched with GATA motifs (Figures 2D–2F and S2F).

To examine the failure to commit to the germline lineage, we determined the loci that failed to open (FO) and FC in the mid and late stages, in which all EpCAM$^-$/INTEGRIN$\alpha6^-$ and EpCAM$^+$/INTEGRIN$\alpha6^-$ (SP) cells were compared with the EpCAM$^+$/INTEGRIN$\alpha6^+$ cells (Figure 2G). Of note, the FO loci were significantly enriched with AP2, OCT4, OCT4:SOX17, SOX17, SOX15, and EBF motifs, while the FC loci were enriched with AP1 and TEAD motifs (Figure 2H). Consistently, somatic lineage genes such as GATA, AP1, and TEAD family members were upregulated in N cells (Figure 2I). Next, we found that the day 1 EBs exhibited a higher proportion of GATA4-positive cells than that of SOX17-positive cells by immunostaining (Figures 2J, S2G, and S2H). Notably, there was no difference for GATA4 expression between SOX17-positive and -negative cells in day 1 EBs, while the SOX17-positive cells exhibited significantly lower GATA4 expression than that of SOX17-negative cells from day 2 (Figures 2J and 2K). Although the AP1-JUN motifs were enriched in D1-specific open regions, JUN protein cannot be

Figure 2. Determination of the regulatory elements underlying cell-fate bifurcation of germline and non-germline lineages

(A) Venn diagram showing the top 10k common and specific peaks in hESCs, 4i cells, and day 1 cells.

(B) Heatmap showing TF motifs significantly enriched in the common and specific peaks in hESCs, 4i cells, and day 1 cells defined in (A).

(C) The top binding motifs enriched in day 1-specific peaks.

(D) Venn diagram showing the top 10k common and specific peaks in day 1, D2DP, and D2N cells.

(E) Heatmap showing TF motifs significantly enriched for common and specific peaks in day 1, D2DP, and D2N cells defined in (D).

(F) The top binding motifs enriched in D2DP-specific and D2N-specific peaks.

(G) Failed-to-open (FO) and failed-to-close (FC) peaks for D2N, D4N, D4SP, and D6N cells compared with D6DP cells are shown. These peaks are derived from peaks open in 4i cells yet closed in D6DP cells or peaks closed in 4i cells yet open in D6DP cells. (H) Heatmap showing TF motifs significantly enriched in FO and FC peaks in D2N, D4N, D4SP, and D6N cells.

(I) Gene expression of TEAD, GATA family, and AP1 family TFs among DP and N cells as well as 7-week hPGCs (Irie et al., 2015).

(J) Immunofluorescence of JUN, GATA4, and SOX17 in EBs from day 1 to day 6, Scale bar, 50 μm. The dotted boxes enclose representative zoomed images.

(K and L) Quantification of relative fluorescence intensity of GATA4 (K) or JUN (L) for SOX17-positive and -negative cells measured by the ImageJ software. Eight slides of immunostaining from three independent experiments were used. Two-tailed Student's t test was performed, ****$p < 0.0001$.

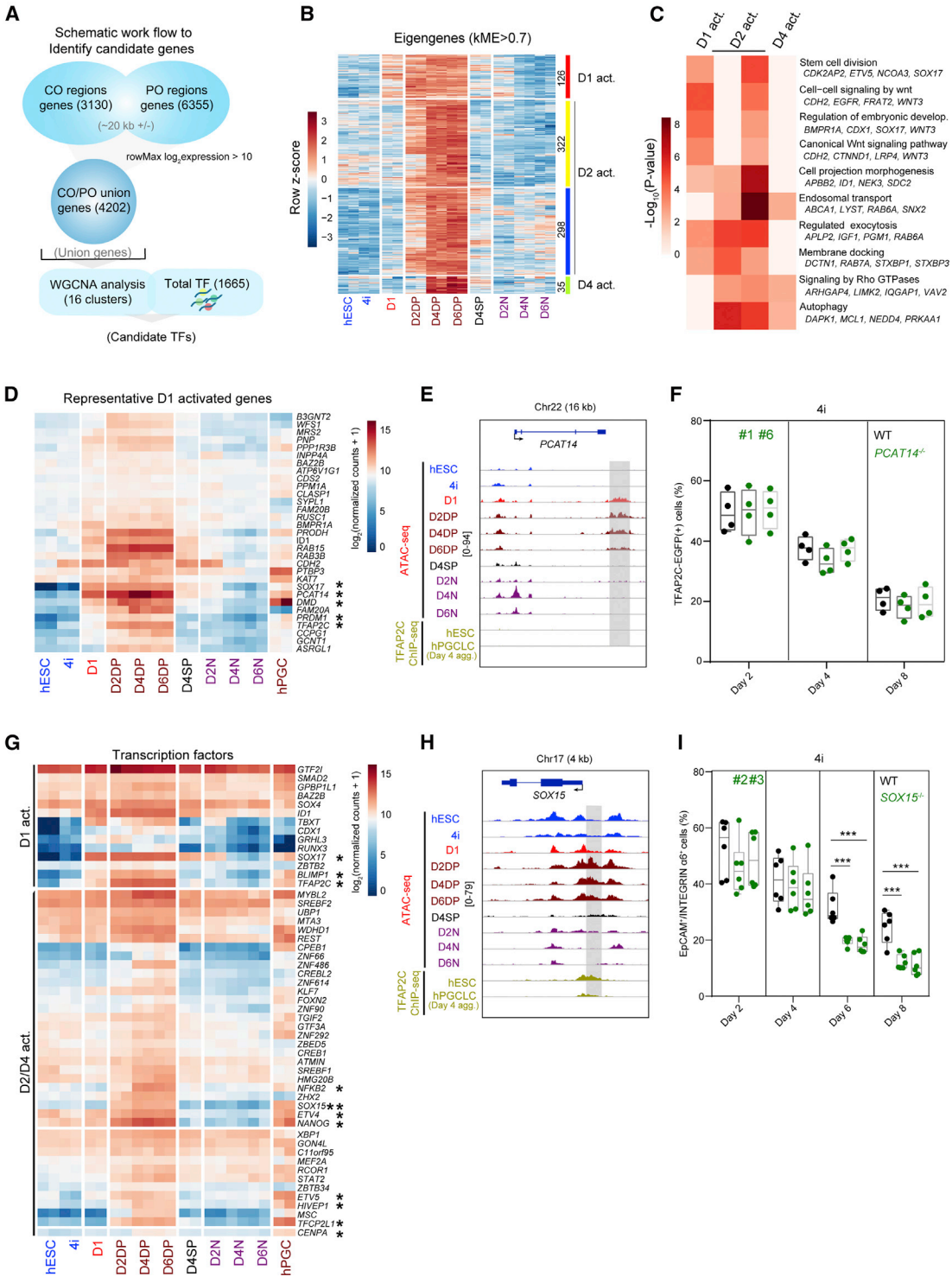


Figure 3. Transcriptional determinants during the induction of hPGCLCs from hESCs

(A) Schematic representation of candidate genes identification.

(B) Heatmap showing the expression of selected modules in which genes are specifically expressed in DP cells. Module eigengenes score (kME score > 0.7) was used to set the threshold to obtain candidate genes. The red, yellow/blue, and green modules were assigned to day 1-activated (D1 act.), day 2-activated (D2 act.), and day 4-activated (D4 act.) groups.

(legend continued on next page)

detected until day 2 (Figure 2J). Interestingly, a mutually exclusive expression pattern between JUN and SOX17 was identified (Figures 2J and 2L), suggesting the potential antagonism between JUN and SOX17.

Transcriptional determinants during the induction of hPGCLCs from hESCs

In contrast to EpCAM$^-$/INTEGRINα6$^-$ (N) cells, the EpCAM$^+$/INTEGRINα6$^+$ (DP) cells were on a trajectory toward gonadal hPGCs (Figure S3A). To evaluate the key transcriptional regulators that control hPGCLCs, we annotated all genes with a transcription start site (TSS) within 20 kb of the CO1-5 regions (3,130 genes) and constituently opened PO regions (6,355 genes) from Figure 1D. From this, we obtained 4,202 union genes from the CO and PO regions after removing lowly expressed genes (Figure 3A). We clustered the genes using weighted gene correlation network analysis (WGCNA) (Langfelder and Horvath, 2008) and obtained 16 modules across hPGCLC induction (Figure S3B; Table S2). These modules showed distinct patterns of gene expression and gene ontology (GO) (Figures 3B, 3C, and S3C–S3F). Based on the expression patterns, we assigned the WGCNA-identified modules: red as day 1 activated (D1 act.), yellow/blue as day 2 activated (D2 act.), and green as day 4 activated (D4 act.) (Figure 3B). GO analysis showed that nucleic acid metabolism-related terms were enriched in genes highly expressed from hESCs to D1 (Figures S3C and S3D), while genes in D1, D2, and D4 act. modules were related to terms such as "stem cells division" and "WNT signaling related" (Figure 3C; Table S2). Conversely, genes highly expressed in N cells were enriched in GO terms associated with somatic differentiation (Figures S3E and S3F; Table S2).

To find the critical genes involved in the induction of hPGCLCs, we first focused on the D1 act. genes that showed the similar expression patterns to *SOX17*. *PCAT14*, a long non-coding RNA, was activated from day 1 and exhibited high expression and specific open regions in hPGCLCs (Figures 3D, 3E, and S3G). But PCAT14 knockout (KO) exhibited no obvious effect on the induction of hPGCLCs (Figures 3F and S3H–S3J). Then we intersected genes in D1, D2, and D4 act. groups with a database of TFs (Hu et al., 2019) and 53 unique TFs were identified. In detail, *SOX17*, *BLIMP1*, and *TFAP2C* were activated at day 1, while *NFKB2*, *SOX15*, *ETV4*, *NANOG*, *ETV5*, *HIVEP1*, and *TFCP2L1* were activated from day 2 or day 4 (Figure 3G). All the TFs highlighted in Figure 3G showed significant expression levels in DP cells (hPGCLCs) like that in hPGCs, but were downregulated in N cells. Notably, there were specific open regions at the loci of SOX15 in DP cells from day 2 (Figure 3H), which were consistent with the gene expression pattern. In addition, the specific open regions of SOX15 genome loci showed the enrichment of TFAP2C peaks (Figure 3H). To confirm if SOX15 was essential for the induction of hPGCLCs from hESCs, we generated SOX15 KO hESC clones and confirmed the absence of the SOX15 protein (Figures S4A and S4B). The resulting SOX15 KOs were karyotypically normal and expressed pluripotency marker genes (Figures S4C and S4D). Interestingly, we found that the proportion of hPGCLCs was dramatically decreased on D6 and D8 in the SOX15 KO lines (Figures 3I, S4E, and S4F), indicating that SOX15 might be crucial for the maintenance of hPGCLC identity. In addition, genetic ablation of SOX15 also led to a decrease of EpCAM$^+$/INTEGRINα6$^+$ cells from D4 in the iMeLC system (Figures S4G–S4I). And SOX15 KO had no obvious effect on the cell-proliferation and apoptosis status of hPGCLCs (Figures S4J–S4M).

Absence of SOX15 in hPGCLCs derails the germline fate and initiates a somatic lineage program

To obtain a comprehensive insight into the roles of SOX15 throughout hPGCLC induction, we evaluated the impact of the SOX15 KO on the transcriptome via time course RNA-seq. Intriguingly, PCA showed that the divergence between the SOX15 wild type (WT) and KO started at D2 (Figure 4A). In support of this, the number of differentially expressed genes (DEGs) increased from day 2 onward

(C) Gene ontology (GO) analysis of the genes in the D1 act., D2 act., and D4 act. groups as defined in (B).
(D) Heatmap showing the expression pattern of representative D1 act. genes.
(E) Selected genomic views showing the ATAC-seq signals and TFAP2C chromatin immunoprecipitation sequencing (ChIP) signals (Chen et al., 2019) for *PCAT14* in the indicated samples. The specific open regions from day 1 are marked with a gray box.
(F) The percentages of TFAP2C-EGFP(+) cells of floating embryoids of WT (black) and *PCAT14* knockout (KO) lines (green) upon hPGCLC induction at the indicated days via the 4i method. Results of four independent experiments were shown (n = 4).
(G) Heatmap showing the overall expression of all TFs from the D1/D2/D4 act. modules. Key genes with relatively high expression in hPGCLCs and hPGCs are highlighted.
(H) Selected genomic views showing the ATAC-seq signals and TFAP2C ChIP signals (Chen et al., 2019) for SOX15 in the indicated samples. The specific open regions with TFAP2C binding are marked with a gray box.
(I) The percentages of EpCAM$^+$/INTEGRINα6$^+$ cells of floating embryoids from WT (black) and SOX15 KO lines (green) upon hPGCLC induction at the indicated days via the 4i method. Results of six independent experiments were shown (n = 6). Two-tailed Student's t test was performed, ***p < 0.001.

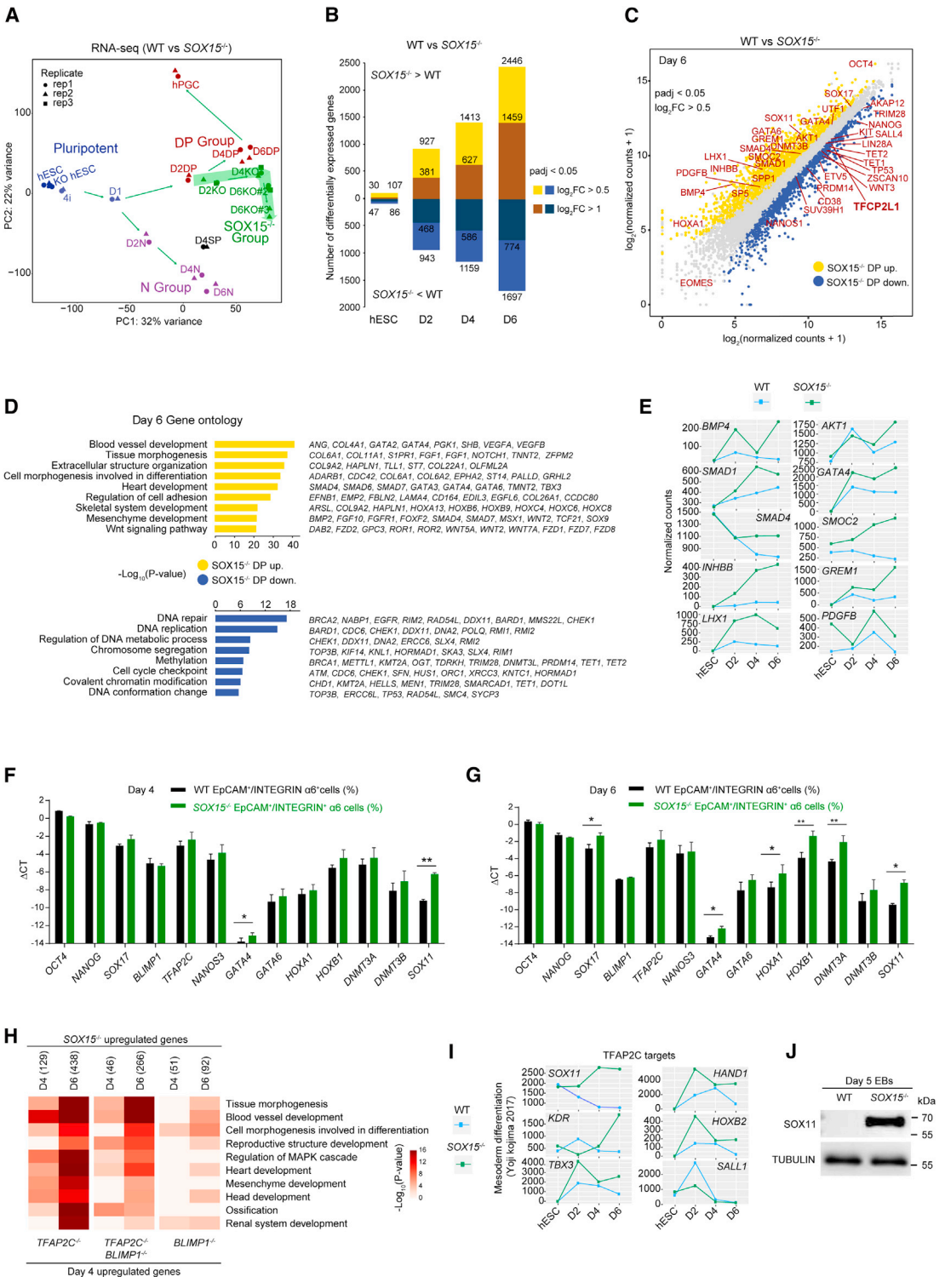


Figure 4. Absence of SOX15 in hPGCLCs derails the germline fate and initiates a somatic lineage program

(A) PCA of the RNA-seq data of WT and SOX15 KO samples. Cell types are indicated by different colors. The green color shows the diverted pathway of SOX15 KO cells. Results of three independent experiments were shown and the replicates are represented by triangles, squares, and circles.

(legend continued on next page)

(Figure 4B; Table S3). Among the late-stage (day 6) DEGs, we noticed that pluripotency genes were downregulated; however, a range of somatic genes were upregulated in *SOX15*$^{-/-}$ hPGCLCs relative to the control (Figures 4C–4E; Table S3). qPCR further validated the aberrant upregulation of somatic genes (*GATA4*, *GATA6*, *HOXA1*, and *HOXB1*) and the *de novo* DNA methyltransferases (*DNMT3A*) (Figures 4F and 4G). These data support the notion that SOX15 is essential for maintaining the germ cell identity of late-stage hPGCLCs.

Among the affected genes in *SOX15*$^{-/-}$ hPGCLCs, we found that several pluripotency-associated genes (such as *TFCP2L1*) were downregulated (Figure 4C). Thus, we next attempted to see if the naive stem cell-specific gene TFCP2L1 (Wang et al., 2019) was also involved in hPGCLC induction; however, TFCP2L1 KO did not impact the induction of hPGCLCs (Figures S5A–S5H).

SOX15 might act as a downstream regulator of TFAP2C

To identify the upstream regulator of SOX15, we first compared the expression of hPGC and pluripotency-associated marker genes in the *SOX15*$^{-/-}$, *SOX17*$^{-/-}$, *TFAP2C*$^{-/-}$, and *BLIMP1*$^{-/-}$ cells throughout hPGCLC induction (Kojima et al., 2017) (Figures S6A and S6B). We found that, at day 2, *SOX17*$^{-/-}$ cells exhibited the complete loss of expression of early hPGC and pluripotency-associated marker genes, while *TFAP2C*$^{-/-}$ cells maintained lower levels of *SOX17* and *BLIMP1* until day 2. Compared with *SOX17*$^{-/-}$ and *TFAP2C*$^{-/-}$ cells, the *BLIMP1*$^{-/-}$ cells expressed *SOX17* at similar levels to WT until day 2 and maintained higher levels of pluripotency and hPGC-associated markers until day 4 (Figure S6B). Notably, *SOX15* expression was not activated in both *SOX17*$^{-/-}$ or *TFAP2C*$^{-/-}$ cells, while it was indistinguishable in *BLIMP1*$^{-/-}$ cells. This implies that SOX15 might be the downstream regulator of SOX17 and TFAP2C but not BLIMP1. To further test whether SOX15 expression is dependent on SOX17 or TFAP2C, we analyzed the up- and downregulated genes in hPGCLCs using the RNA-seq data: *SOX15*$^{-/-}$ (day 2, day 4, or day 6), *SOX17*$^{-/-}$ (day 2), and *TFAP2C*$^{-/-}$ and *BLIMP1*$^{-/-}$ (day 4), compared with their controls (Kojima et al., 2017) (Figures S6C–S6H; Table S4). Interestingly, we observed that several canonical pathways such as ATF2 and AP1 were significantly enriched in commonly upregulated genes in *SOX15*$^{-/-}$ and *TFAP2C*$^{-/-}$ at day 4 and day 6, but not genes in *SOX15*$^{-/-}$ and *BLIMP1*$^{-/-}$ (Figure S6I). Furthermore, *SOX15*$^{-/-}$ cells and *TFAP2C*$^{-/-}$ cells shared many somatic lineage-related GO terms in the upregulated genes (Figure 4H). In addition, the downregulated genes shared with *TFAP2C*$^{-/-}$ or *BLIMP1*$^{-/-}$ had only a few significant associated GO terms (Figure S6J).

To establish the direct relationship between SOX15 and TFAP2C as well as BLIMP1, we examined the target genes of BLIMP1, BLIMP1/TFAP2C, and TFAP2C, respectively (Kojima et al., 2017). We found that the targets of BLIMP1 and BLIMP1/TFAP2C were not affected in *SOX15*$^{-/-}$ cells (Figure S6K); however, TFAP2C target genes associated with mesoderm differentiation were significantly upregulated in *SOX15*$^{-/-}$ hPGCLCs (Figure 4I). Notably, chromatin immunoprecipitation sequencing analysis showed that TFAP2C can bind to several proximal elements at the SOX15 locus, supporting the notion that TFAP2C might be an upstream regulator of SOX15 (Figure 3H) (Chen et al., 2019). Western blot results further demonstrated that SOX11, a shared marker gene of mesoderm and ectoderm lineages, was upregulated in *SOX15*$^{-/-}$ cells (Figure 4J). Together, these results suggest that SOX15 might act as a downstream regulator of TFAP2C to exert its functions.

The suppression of somatic gene expression mediated by SOX15 is associated with chromatin accessibility

To determine how SOX15 exerts its roles in somatic gene expression suppression during hPGCLC induction, we performed ATAC-seq in WT and *SOX15*$^{-/-}$ hPGCLCs. PCA analysis demonstrated that *SOX15*$^{-/-}$ hPGCLCs cells diverged from the hPGCLC trajectory from day 4 onward

(B) Bar plot showing the number of differentially expressed genes (DEGs) during the induction of hPGCLCs from *SOX15*$^{-/-}$ hESCs (p_{adj} < 0.05, $\log_2$ fold change [FC] > 0.5 or > 1).
(C) Scatterplot showing the DEGs in *SOX15*$^{-/-}$ DP cells at day 6. The *SOX15*$^{-/-}$ upregulated and downregulated genes are color coded ($\log_2$ fold change > 0.5).
(D) GO terms enriched in DEGs in *SOX15*$^{-/-}$ DP cells ($\log_2$ fold change > 0.5).
(E) Line plots showing gene expression dynamics of the indicated genes.
(F and G) qPCR of the indicated genes in EpCAM$^+$/INTEGRINα6$^+$ cells of day 4 embryoids (F) and day 6 embryoids (G) derived from WT and *SOX15*$^{-/-}$ hESCs, respectively. Relative expression levels are shown normalized to *GAPDH*. Error bars indicate mean ± SD from three independent replicates. Two-tailed Student's t test was performed, *p < 0.05, **p < 0.01.
(H) Heatmap showing the GO terms enriched in the upregulated genes (day 4 and day 6) in *SOX15*$^{-/-}$ cells shared with *BLIMP1*$^{-/-}$ or *TFAP2C*$^{-/-}$ cells (day 4). The gene numbers here are from Figure S6.
(I) Line plots showing the gene expression of the downstream genes regulated by TFAP2C.
(J) Western blot analysis of SOX11 protein in day 5 *SOX15*$^{-/-}$ embryoids relative to the control. Tubulin was used as an inner control.

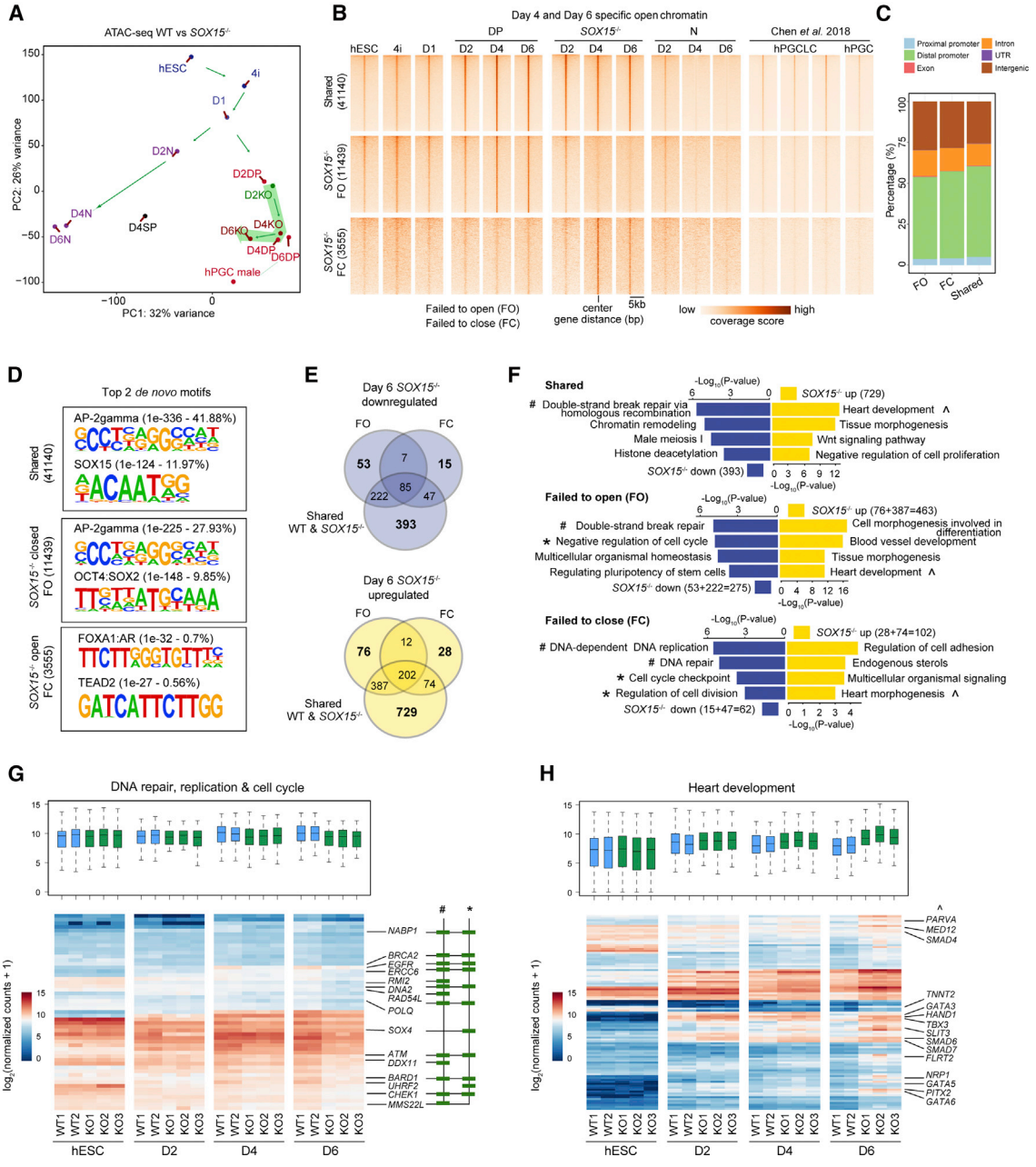


Figure 5. The suppression of somatic gene expression mediated by SOX15 is associated with chromatin accessibility

(A) PCA plot showing ATAC-seq analysis of the hPGCLC induction under the normal and SOX15 KO states. Two independent replicates are merged.

(B) Heatmap of ATAC-seq signals in the indicated samples over shared-open chromatin regions constituting 41,140 peaks, *SOX15*$^{-/-}$ FO regions constituting 11,439 peaks, and *SOX15*$^{-/-}$ FC regions constituting 3,555 peaks.

(C) Bar plot showing the percentage of genomic features from FO, FC, and shared regions.

D) Top 2 *de novo* motifs from shared, FO, and FC genomic regions. The name of the motifs with respective p value and percentage are shown on each motif.

(E) Venn diagram showing the intersection of nearby genes from FO, FC, or shared regions that shared with the downregulated and up-regulated genes in day 6 *SOX15*$^{-/-}$ cells, respectively. Log_2 fold change > 0.5.

(legend continued on next page)

(Figure 5A). We also combined the D4 and D6 ATAC-seq data and clustered them into three categories, shared-open, FO, and FC (Figure 5B). The open chromatin regions of shared, FC, and FO groups were highly enriched around the distal and intergenic regions (Figure 5C). Of note, we discovered the top TF binding motifs of each category (shared-open, AP2 and SOX15; FC, FOXA1:AR and TEAD2; FO, OCT4:SOX2) (Figure 5D).

We next extracted the day 6 DEGs around the shared, FO, and FC open chromatin regions, respectively. There were 53 downregulated genes and 76 upregulated genes for FO, 15 downregulated genes and 28 upregulated genes for FC, and 393 downregulated genes and 729 upregulated genes for shared-open regions (Figure 5E). Among the unique genes for shared-open regions, we observed that the heart development (i.e., somatic mesoderm)-related genes were upregulated, while double-strand break repair via homologous recombination-related genes were downregulated in *SOX15*$^{-/-}$ hPGCLCs (Figure 5F; Table S5). This result indicates that the genes nearby were still affected by the absence of SOX15, albeit no change of these shared-open chromatin regions. Due to the limited number of DEGs, we combined the common genes near shared regions of FO or FC regions for further analysis. GO analysis of genes near the FO and FC regions revealed that the downregulated genes were enriched in DNA replication and pluripotency-related GO terms, while the upregulated genes were enriched in heart development-related GO terms (Figure 5F; Table S5). In addition, we combined the genes in similar GO terms of shared, FC, as well as FO groups and found that the downregulated or upregulated genes of these GO terms in *SOX15*$^{-/-}$ cells exhibited differential expression patterns throughout hPGCLC induction (Figures 5G and 5H). Overall, loss of SOX15 in hPGCLCs disturbed the genes near the unchanged chromatin open regions or resulted in aberrant chromatin changes, both of which might further induce the observed cell-fate bifurcation to somatic lineages.

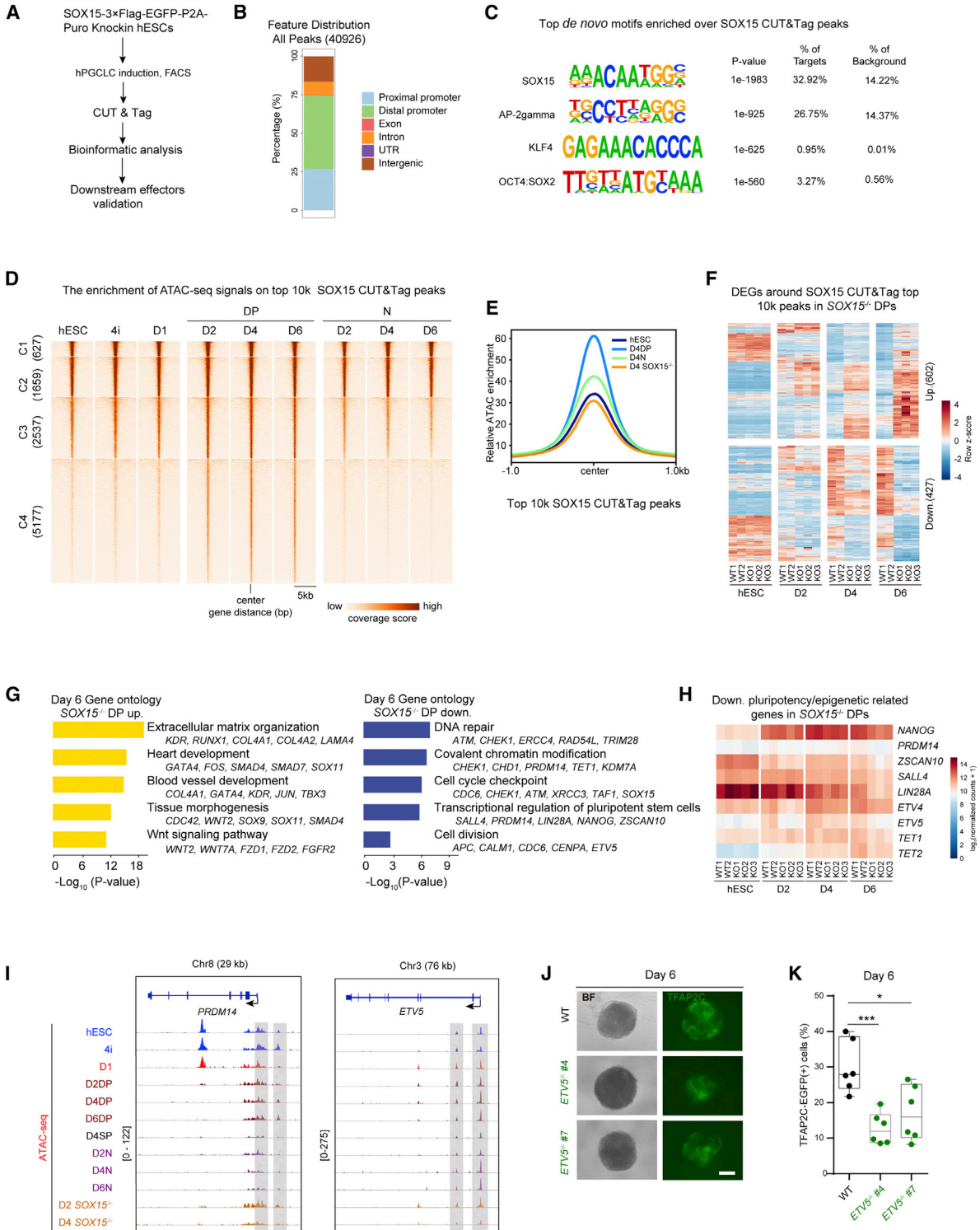
SOX15 exerts its function in hPGCLC maintenance by directly suppressing somatic gene expression and sustaining latent pluripotency

To find the target genes bound by SOX15, we first established SOX15-3×Flag-P2A-EGFP-Puro knockin cell lines and obtained day 4 EpCAM^{+}/INTEGRINα6^{+} (DP) cells to perform CUT&Tag assays (Figures 6A, S7A, and S7B) (Kaya-Okur et al., 2020). SOX15 peaks were mainly enriched around the proximal and distal promoter regions (Figure 6B). Next, we performed *de novo* motif search on SOX15 peaks. Interestingly, we found AP2-gamma, KLF4, and OCT4:SOX2 on the top enriched motif list, indicating that the AP2-gamma (TFAP2C) and KLF4 might also bind near the SOX15-bound regions (Figure 6C). For the extinguishment of SOX2 in hPGCLCs, the OCT4:SOX2 motifs regions should be bound by OCT4/SOX17 or/and OCT4/SOX15. By overlaying the enrichment of ATAC-seq signals on top 10k SOX15 peaks, we found four different signals clusters. Interestingly, cluster 4 (5,177 regions) showed a stronger signal specific to DP cells (Figure 6D). Moreover, the chromatin accessibility status near SOX15 peaks were dramatically decreased in D4 SOX15 KO DP cells (Figure 6E). Surprisingly, the relative ATAC signals on D2 for SOX15 peaks were highly enriched in both WT and SOX15 KO DP cells; however, the ATAC signal for SOX15 peaks were only remained enriched in WT DP cells on D4 and D6 (Figure S7C). These results indicate that SOX15 might exert its function by regulating chromatin accessibility and thereby target gene expression.

Then, we searched for D6 DEGs around the top 10k SOX15 peaks. About 602 upregulated genes and 427 downregulated genes were obtained (Figure 6F). GO analysis of these genes revealed that the upregulated genes were enriched in the terms associated with somatic lineage differentiation, while the downregulated genes were enriched for the DNA repair and pluripotency-related terms (Figure 6G; Table S6). Moreover, SOX15 peaks were detected at the proximal regulatory elements of several pluripotency-related genes, such as *PRDM14*, *NANOG*, *ETV4*, and *ETV5* (Figures 6G and S7D) (Kalkan et al., 2019; Murakami et al., 2016; Sybirna et al., 2020). These results suggest that SOX15 might be involved in maintaining the latent pluripotency of hPGCLCs (Leitch and Smith, 2013). In support of this, the regulatory elements bound by SOX15 of these genes showed decreased ATAC signals in day 4/6 SOX15 KO DP cells compared with that in WT DP cells, which were consistent with the downregulated expression of these genes (Figures 6G–6I and S7D). These results indicated that SOX15 exerted its functions in maintaining the identity of hPGCLCs through dual effect-simultaneous suppression of somatic gene expression and the retention of latent pluripotency.

Given the fact that PRDM14 and NANOG are implicated in the induction of PGCLCs (Murakami et al., 2016; Sybirna et al., 2020), we then investigated whether ETV5 was also involved in the maintenance of hPGCLCs by acting as a direct target of SOX15. To this end, the expression

(F) GO analysis for upregulated and downregulated genes nearby shared, FO, and FO regions, respectively, as shown in (E).
(G and H) Boxplots (with the median and 25th and 75th percentiles) and heatmap showing the expression patterns of specific genes representing the GO terms of DNA repair, DNA replication, and cell cycle (G), or heart development (H). The symbols #, *, and ^ represent the GO terms shown in (F) and key genes are indicated.



(legend on next page)

pattern of *ETV5* was first evaluated, and we found that *ETV5* was downregulated in D4 SOX15 KO DP cells (Figure S7E). Then, *ETV5* hESC KO clones (TFAP2C-EGFP knockin) were generated and the absence of the ETV5 protein was confirmed (Figures S7F and S7G). The resulting *ETV5* KOs can be induced into hPGCLCs with decreased ratio of hPGCLCs compared with WT control (Figures 6J, 6K, S7G, and S7H). These data proved that ETV5, which acted as a downstream regulator of SOX15, was essential for hPGCLC maintenance.

A stepwise OCT4:SOX motifs switch throughout hPGCLC induction

To further study the stage-specific role of SOXs and OCT4/SOXs in the induction of hPGCLCs, we performed a focused analysis of SOX motifs in open chromatin. First, we defined peaks in DP (top 10k peaks from day 2/4/6 DP libraries), N (peaks from day 2/4/6 N libraries), and E (peaks from early stage: hESCs, 4i, and day 1 libraries) groups and intersected the peaks to obtain specific peaks in each group (Figure 7A). The DP-specific ATAC signals (4,049 peaks) were enriched in hPGCLCs and gonadal hPGCs (Chen et al., 2018), while the N-specific ATAC signals (6,968 peaks) were not found in hPGCLCs and hPGCs (Figure S7I). These DP-specific regions showed an enrichment of known single SOX or OCTSOX motifs (Figures 7B and 7C). Notably, SOX2 and OCT4:SOX2 motifs (canonical SOXOCT motifs) were enriched in DP-specific regions (Figures 7B and 7C), which was not consistent with the absence of SOX2 in hPGCLCs (Figure 7D). This prompted us to ask if the SOX2 and OCT4:SOX2 motifs sites in DP group were engaged by SOX17 or SOX15 to form an OCT4/SOX17 or OCT4/SOX15 heterodimer. Notably, co-immunoprecipitation results in HEK293 cells showed that there was an interaction between OCT4 and SOX15 or SOX17 (Figures S7J and S7K). In addition, overexpression of Sox15 can rescue the defects that result from the absence of Sox2 in mESCs (Niwa et al., 2016). It is known that SOX17 heterodimerize with OCT4 to bind a compressed motif (OCT4:SOX17), which lacks a single base pair between the SOX and OCT half-sites compared with the canonical motifs (OCT4:SOX2) (Figure 7C), while SOX15 can heterodimerize with OCT4 on canonical elements (Chang et al., 2017), albeit there is no direct evidence to demonstrate the presence of OCT4:SOX15 motifs so far. Molecular modeling results further showed that human OCT4-SOX15-DNA complex shared a similar overall fold with mouse OCT4-SOX2-DNA complex (Figure 7E). Based on this evidence, the OCT4:SOX2 motifs enriched in the DP-specific group and SOX15 CUT&Tag peaks (Figure 6C) were most likely to be OCT4:SOX15 motifs.

To determine if the predicted OCT4:SOX15 motifs was functionally relevant, we first extracted 1,595, 68, and 3 ATAC-seq peaks including OCT4:SOX15 motifs sites in the shared, FO, and FC groups, respectively (Figure S7L). Next, we searched the DEGs around the predicted OCT4:SOX15 binding motif sites in *SOX15*$^{-/-}$ DP cells. GO analysis of genes around the shared regions showed that the 123 upregulated genes in *SOX15*$^{-/-}$ DP cells were enriched in terms such as "extracellular matrix organization," while the 66 downregulated genes were enriched in terms such as "cell fate commitment" (Figures 7F and 7G; Table S6). Notably, the downregulated genes included *PRDM14* and *NANOG*, which are critical for the latent pluripotency of germline.

Based on these results, we established a model that supports a stepwise switch of OCT/SOX heterodimerization preferences, from OCT4/SOX2 in pluripotent cells, to OCT4/SOX17 in early-stage cells, and then to a putative OCT4/SOX15 binding module in the late stage

Figure 6. SOX15 exerts its function in hPGCLC maintenance by directly suppressing somatic gene expression and sustaining latent pluripotency

(A) Schematic representation of the SOX15 CUT&Tag analysis workflow in hPGCLCs.
(B) Bar plot showing the percentage of genomic feature distribution of SOX15 peaks.
(C) The top binding motifs enriched in SOX15 peaks.
(D) Heatmap of ATAC-seq signals in the indicated samples over the top 10k SOX15 peaks.
(E) Pileup of the ATAC-seq signals over the top 10k SOX15 peaks regions in the indicated cells.
(F) Heatmap showing the expression patterns of upregulated or downregulated genes around the top 10k SOX15 peaks in day 6 SOX15 KO DP cells.
(G) GO analysis for the upregulated or downregulated genes as described in (F).
(H) Heatmap showing the expression patterns of downregulated pluripotency-related genes in SOX15 KO DP cells.
(I) Selected genomic views showing the ATAC-seq signals, TFAP2C ChIP signals (Chen et al., 2019), and SOX15 signals at the *PRDM14* and *ETV5* genome loci in the indicated samples. The specific open regions with SOX15 signals and decreased ATAC-seq signals from day 4 KO DP cells compared with those in DP cells are marked with a gray box.
(J) Bright-field (BF) and fluorescence (TFAP2C-EGFP) images of floating embryoids from WT and *ETV5*$^{-/-}$ lines at day 6. Scale bar, 200 μm.
(K) The percentages of TFAP2C-EGFP(+) cells of floating embryoids from day 6 WT (black) and ETV5 KO lines (green) upon hPGCLC induction via the 4i method. Results of six independent experiments are shown (n = 6). Two-tailed Student's t test was performed, *p < 0.05, ***p < 0.001.

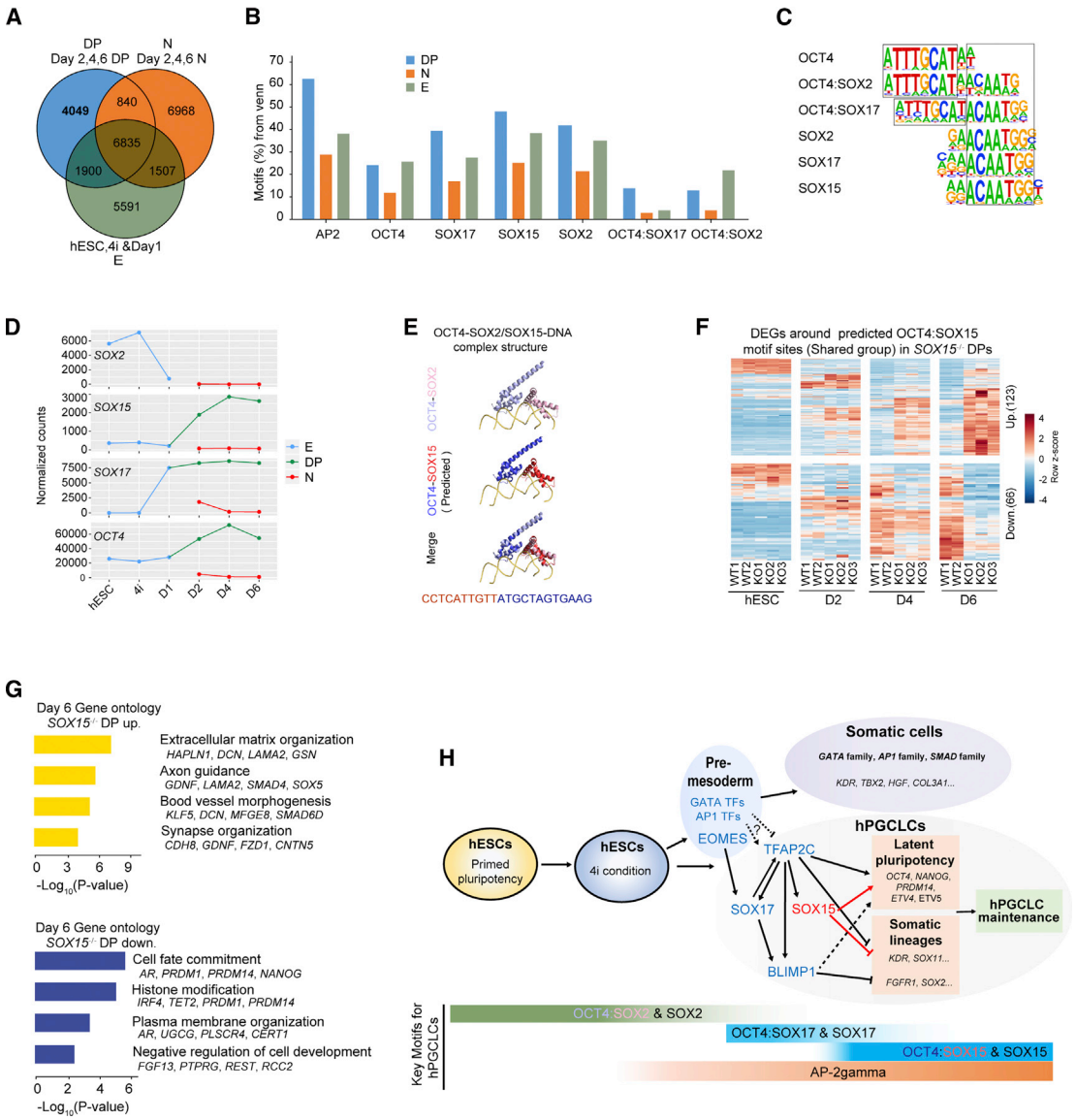


Figure 7. A stepwise OCT4:SOX motifs switch throughout hPGCLC induction

(A) Venn diagram showing the common and different peaks in DP, N, and E groups.

(B) Bar plot representing the percentage of known OCTSOX motifs enriched in DP/N/E-specific open chromatin regions.

(C) Known SOX and OCT motifs with respective TF binding sequences.

(D) Line plot showing the gene expression of *SOX2*, *SOX15*, *SOX17*, and *OCT4* in E (hESC, 4i, day 1), DP, and N cells.

(E) Ribbon diagrams showing the similarity between the structure of known mouse OCT4-SOX2-DNA complex (PDB: 6HT5) and the predicted modeled structure of human OCT4-SOX15-DNA complex.

(F) Heatmap showing the expression pattern of upregulated or downregulated genes around putative OCT4:SOX15 binding motif sites, which belong to the "Shared" group as described in Figure 5B, in day 6 SOX15 KO DP cells.

(G) GO analysis for the upregulated or downregulated genes as described in (F).

(H) Schematic showing the roles of SOX15, ETV5, and the key motifs during the induction of hPGCLCs.

(Figure 7H). This model describes the critical roles of SOX15 in the maintenance of hPGCLC identity via suppressing somatic gene expression and sustaining latent pluripotency.

DISCUSSION

Here, time course ATAC-seq and RNA-seq analyses were performed to resolve the dynamics of genome regulation

in both hPGCLCs and non-hPGCLCs. In addition, we identified the involvement of SOX15 in maintaining the identity of hPGCLCs. Further studies showed that SOX15 exerted its functions in hPGCLCs by suppression of somatic gene expression and retainment of latent pluripotency. Among the SOX15-mediated regulatory networks underlying latent pluripotency preservation, ETV5 was revealed to be critical for hPGCLC maintenance by acting as a downstream regulator of SOX15. Finally, a stepwise OCT4:SOX motifs switch was uncovered to have potential functions throughout hPGCLC induction. Based on our data and the accumulated evidence, we propose a model that SOX15 is involved in facilitating the establishment of hPGCLC regulatory networks (Figure 7H).

The analysis for chromatin dynamics of both hPGCLCs and non-hPGCLCs from hESCs revealed that several TF motifs as "accelerators" (AP2, OCT4:SOX17, and SOX15) or potential "suppressors" (GATA, AP1, and TEAD) of hPGCLC induction. However, it is noteworthy that GATA and AP1 motifs are not only enriched in non-hPGCLCs (Figures S2B and S2C), but also in the regions over D1-specific peaks, in which the EOMES motif is also enriched (Figure 2C). Therefore, it would be appealing to validate the functions of GATA and AP1 in the induction of hPGCLCs, which might provide new insights into the cell-fate bifurcation of germline and somatic lineage.

SOX17 and TFAP2C exert their functions in hPGCLC induction in an interdependent manner, and TFAP2C has a decisive role in the somatic lineage suppression to maintain the hPGCLC identity (Kobayashi et al., 2017; Kojima et al., 2017); growing evidence shows that TFAP2C is involved in the activation of OCT4 naive enhancers and the prevention of hPGCLCs from somatic lineages (Chen et al., 2018, 2019; Pastor et al., 2018). Consistent with these findings, our genome-wide analysis revealed that the hPGCLCs were enriched with TFAP2C motif elements as well as SOX17, SOX15, and OCT4/SOX motif elements, coinciding with the suppression of the somatic transcriptome. Moreover, we found that removal of SOX15 destabilizes hPGCLCs after day 4. A recent study demonstrates that the absence of SOX15 derails the germline fate of hPGCLCs and reactivation of SOX15 could rescue the hPGCLC identity in the *SOX15*$^{-/-}$ cell line (Pierson Smela et al., 2019); however, the detailed mechanisms of SOX15 in hPGCLCs are still unclear. Combined with ATAC-seq and CUT&Tag-seq analysis, we discovered that SOX15 played critical roles in the maintenance of hPGCLC identity by suppression of somatic gene expression and retainment of latent pluripotency.

In this study, a stepwise switch of the OCT4:SOX motif is uncovered throughout hPGCLC induction, in which OCT4/SOX2, OCT4/SOX17, and predicted OCT4/SOX15 motifs are enriched in open regions of hESCs, and early- and late-stage hPGCLCs, respectively. Further analysis demonstrated that the predicted OCT4/SOX15 binding motif is most likely to be functionally relevant, as exemplified by the involvement in the suppression of somatic gene expression. Previous studies reveal that the proper downregulation of SOX2 in the initial induction of hPGCLCs is possibly dependent on EOMES, but not SOX17 (Kojima et al., 2017). Coincident with the suppression of SOX2, the emergence of SOX17 expression from the early stage is mainly controlled by EOMES (Kojima et al., 2017). In this regard, it would be interesting to know the mechanisms regulating the shift from OCT4/SOX2 (pluripotent cells) to OCT4/SOX17 (early-stage hPGCLCs) and then to OCT4/SOX15 (mid- to late-stage hPGCLCs).

Collectively, this work characterizes the chromatin accessibility and transcriptome dynamics from hESCs to hPGCLCs or to non-hPGCLCs, providing novel insights into *in vitro* human germ cell induction, as exemplified by the critical role of SOX15 in the maintenance of hPGCLC identity by suppressing somatic gene expression and retaining latent pluripotency.

EXPERIMENTAL PROCEDURES

Induction of 4i hESCs and hPGCLCs

hPGCLCs were generated from hESCs based on a previously reported protocol (Mitsunaga et al., 2017) with slight modifications. Further information is provided in the supplemental experimental procedures.

Statistical analysis

Statistical analyses were performed using GraphPrism 6.0 software. All values are depicted as the mean ± SD. The statistical parameters, such as statistical analysis, n values, and statistical significance, are shown in the figure legends. Statistical significance is presented in the figures as *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, and not significant (ns, $p > 0.05$) (Student's t test) unless stated otherwise. The other statistical tests for DEG analysis, GO analysis, and motif discovery are implemented as part of the respective computational framework of the above websites and tools.

Data and code availability

The accession number for the ATAC-seq, RNA-seq and CUT&Tag-seq data reported in this paper is Gene Expression Omnibus (GEO): GSE143345.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.stemcr.2021.03.032.

AUTHOR CONTRIBUTIONS

X.-Y.Z., G.C., and Z.T.L. conceived and designed the experiments. X.M.W., Z.T.L., K.S., X.Y.X., M.M.C., X.H.X., C.W., W.Y.Y., Z.K.Y.,

X.R.W., and Y.Z. conducted the experiments. V.V., Y.X.Y., X.L.S., and F.L. performed all bioinformatics analysis. V.V., X.M.W., Z.T.L., G.C., and X.-Y.Z. wrote the manuscript. A.P.H., R.J., M.Y.L., and C.H.W. helped with data interpretation and manuscript reviewing. X.-Y.Z. supervised the project.

ACKNOWLEDGMENTS

We are grateful to Dr. Yong Fan for providing us with the human ESC line Fy-hES-3. This work was supported by the National Key R&D Program of China (2017YFA0105001 to X.-Y.Z., 2016YFC1000606 to X.-Y.Z), the National Natural Science Foundation of China (31671544 to X.-Y.Z., 31601208 to Z.T.L., 31970787 to G.C.., 31700676 to F.L., 32000579 to X.M.W.), the Key Research & Development Program of Guangzhou Regenerative Medicine and Health Guangdong Laboratory (2018GZR110104002 to X.-Y.Z.), Guangzhou Science And Technology project key project topic (201904020031 to X.-Y.Z.), the Natural Science Foundation of Guangdong Province (2019A1515010446 to G.C., 2017A030313098 to Z.T.L., 2016A030313604 to F.L.), the Clinical Innovation Research Program of Guangzhou Regenerative Medicine and Health Guangdong Laboratory (2018GZR0201003 to F.-F.H.), the Outstanding Scholar Program of Guangzhou Regenerative Medicine and Health Guangdong Laboratory (2018GZR110102004 to F.-F.H.), Guangdong Provincial Science and Technology Program (2019B030301009), the Natural Science Foundation of Shenzhen (JCYJ20180305163311448 to G.C.), and the China Postdoctoral Science Foundation (2020M672707 to X.M.W.).

Received: September 10, 2020
Revised: March 30, 2021
Accepted: March 31, 2021
Published: April 29, 2021

REFERENCES

Aksoy, I., Jauch, R., Chen, J., Dyla, M., Divakar, U., Bogu, G.K., Teo, R., Leng Ng, C.K., Herath, W., Lili, S., et al. (2013a). Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. EMBO J. *32*, 938–953.

Aksoy, I., Jauch, R., Eras, V., Chng, W.B., Chen, J., Divakar, U., Ng, C.K., Kolatkar, P.R., and Stanton, L.W. (2013b). Sox transcription factors require selective interactions with Oct4 and specific transactivation functions to mediate reprogramming. Stem Cells *31*, 2632–2646.

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

Chang, Y.K., Srivastava, Y., Hu, C., Joyce, A., Yang, X., Zuo, Z., Havranek, J.J., Stormo, G.D., and Jauch, R. (2017). Quantitative profiling of selective Sox/POU pairing on hundreds of sequences in parallel by Coop-seq. Nucleic Acids Res. *45*, 832–845.

Chen, D., Liu, W., Zimmerman, J., Pastor, W.A., Kim, R., Hosohama, L., Ho, J., Aslanyan, M., Gell, J.J., Jacobsen, S.E., et al. (2018). The TFAP2C-regulated OCT4 naive enhancer is involved in human germline formation. Cell Rep *25*, 3591–3602.e5.

Chen, D., Sun, N., Hou, L., Kim, R., Faith, J., Aslanyan, M., Tao, Y., Zheng, Y., Fu, J., Liu, W., et al. (2019). Human primordial germ cells are specified from lineage-primed progenitors. Cell Rep *29*, 4568–4582.e5.

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.

Hou, L., Srivastava, Y., and Jauch, R. (2017). Molecular basis for the genome engagement by Sox proteins. Semin. Cell Dev Biol *63*, 2–12.

Hu, H., Miao, Y.R., Jia, L.H., Yu, Q.Y., Zhang, Q., and Guo, A.Y. (2019). AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. Nucleic Acids Res. *47*, D33–D38.

Irie, N., Weinberger, L., Tang, W.W., 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.

Jauch, R., Aksoy, I., Hutchins, A.P., Ng, C.K., Tian, X.F., Chen, J., Palasingam, P., Robson, P., Stanton, L.W., and Kolatkar, P.R. (2011). Conversion of Sox17 into a pluripotency reprogramming factor by reengineering its association with Oct4 on DNA. Stem Cells *29*, 940–951.

Jostes, S.V., Fellermeyer, M., Arevalo, L., Merges, G.E., Kristiansen, G., Nettersheim, D., and Schorle, H. (2020). Unique and redundant roles of SOX2 and SOX17 in regulating the germ cell tumor fate. Int. J. Cancer *146*, 1592–1605.

Kalkan, T., Bornelov, S., Mulas, C., Diamanti, E., Lohoff, T., Ralser, M., Middelkamp, S., Lombard, P., Nichols, J., and Smith, A. (2019). Complementary activity of ETV5, RBPJ, and TCF3 drives formative transition from naive pluripotency. Cell Stem Cell *24*, 785–801.e7.

Kamachi, Y., and Kondoh, H. (2013). Sox proteins: regulators of cell fate specification and differentiation. Development *140*, 4129–4144.

Kanai-Azuma, M., Kanai, Y., Gad, J.M., Tajima, Y., Taya, C., Kurohmaru, M., Sanai, Y., Yonekawa, H., Yazaki, K., Tam, P.P., et al. (2002). Depletion of definitive gut endoderm in Sox17-null mutant mice. Development *129*, 2367–2379.

Kaya-Okur, H.S., Janssens, D.H., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2020). Efficient low-cost chromatin profiling with CUT&Tag. Nat. Protoc. *15*, 3264–3283.

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.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

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

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

Li, D., Liu, J., Yang, X., Zhou, C., Guo, J., Wu, C., Qin, Y., Guo, L., He, J., Yu, S., et al. (2017). Chromatin accessibility dynamics during iPSC reprogramming. Cell Stem Cell *21*, 819–833.e6.

Maruyama, M., Ichisaka, T., Nakagawa, M., and Yamanaka, S. (2005). Differential roles for Sox15 and Sox2 in transcriptional control in mouse embryonic stem cells. J. Biol. Chem. *280*, 24371–24379.

Mitsunaga, S., Odajima, J., Yawata, S., Shioda, K., Owa, C., Isselbacher, K.J., Hanna, J.H., and Shioda, T. (2017). Relevance of iPSC-derived human PGC-like cells at the surface of embryoid bodies to prechemotaxis migrating PGCs. Proc. Natl. Acad. Sci. U S A. *114*, E9913–E9922.

Murakami, K., Gunesdogan, U., Zylicz, J.J., Tang, W.W.C., Sengupta, R., Kobayashi, T., Kim, S., Butler, R., Dietmann, S., and Surani, M.A. (2016). NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers. Nature *529*, 403–407.

Niwa, H., Nakamura, A., Urata, M., Shirae-Kurabayashi, M., Kuraku, S., Russell, S., and Ohtsuka, S. (2016). The evolutionally-conserved function of group B1 Sox family members confers the unique role of Sox2 in mouse ES cells. BMC Evol. Biol. *16*, 173.

Pastor, W.A., Liu, W., Chen, D., Ho, J., Kim, R., Hunt, T.J., Lukianchikov, A., Liu, X., Polo, J.M., Jacobsen, S.E., et al. (2018). TFAP2C regulates transcription in human naive pluripotency by opening enhancers. Nat. Cell Biol *20*, 553–564.

Perrett, R.M., Turnpenny, L., Eckert, J.J., O'Shea, M., Sonne, S.B., Cameron, I.T., Wilson, D.I., Rajpert-De Meyts, E., and Hanley, N.A. (2008). The early human germ cell lineage does not express SOX2 during in vivo development or upon in vitro culture. Biol. Reprod. *78*, 852–858.

Pierson Smela, M., Sybirna, A., Wong, F.C.K., and Surani, M.A. (2019). Testing the role of SOX15 in human primordial germ cell fate. Wellcome Open Res. *4*, 122.

Saitou, M., and Miyauchi, H. (2016). Gametogenesis from pluripotent stem cells. Cell Stem Cell *18*, 721–735.

Sarraj, M.A., Wilmore, H.P., McClive, P.J., and Sinclair, A.H. (2003). Sox15 is up regulated in the embryonic mouse testis. Gene Expr. Patterns *3*, 413–417.

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.

Sybirna, A., Tang, W.W.C., Pierson Smela, M., Dietmann, S., Gruhn, W.H., Brosh, R., and Surani, M.A. (2020). A critical role of PRDM14 in human primordial germ cell fate revealed by inducible degrons. Nat. Commun. *11*, 1282.

Tang, W.W., 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.

Veerapandian, V., Ackermann, J.O., Srivastava, Y., Malik, V., Weng, M., Yang, X., and Jauch, R. (2018). Directed evolution of reprogramming factors by cell selection and sequencing. Stem Cell Reports *11*, 593–606.

Wang, X., Wang, X., Zhang, S., Sun, H., Li, S., Ding, H., You, Y., Zhang, X., and Ye, S.D. (2019). The transcription factor TFCP2L1 induces expression of distinct target genes and promotes self-renewal of mouse and human embryonic stem cells. J. Biol. Chem. *294*, 6007–6016.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137.

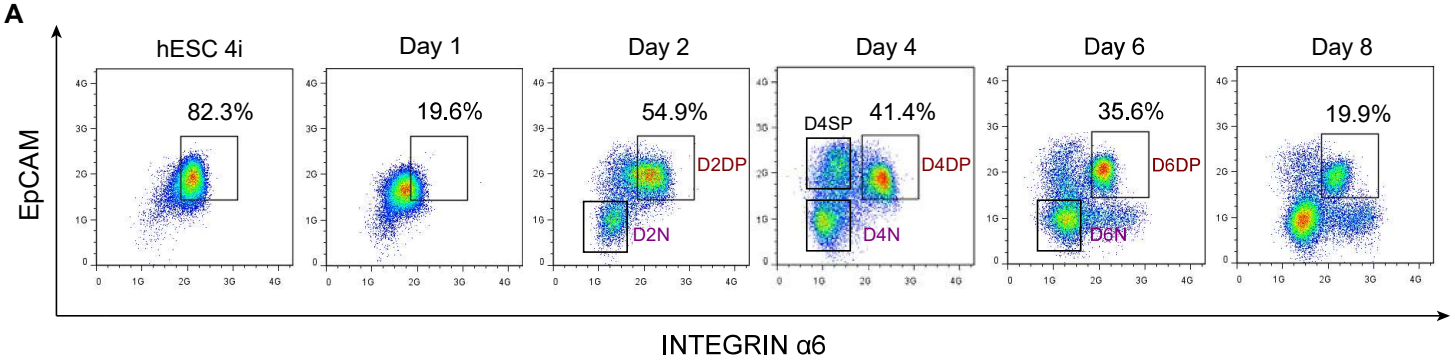
Stem Cell Reports, Volume *16*

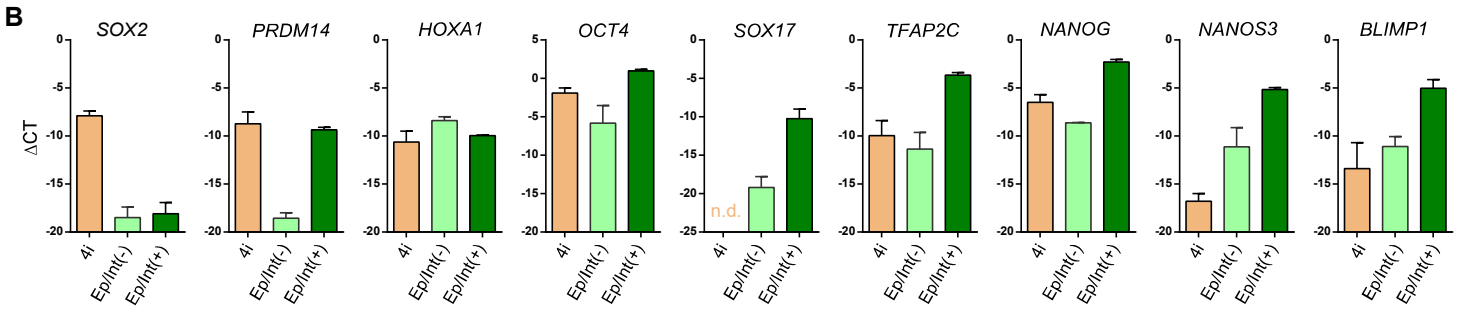
Supplemental Information

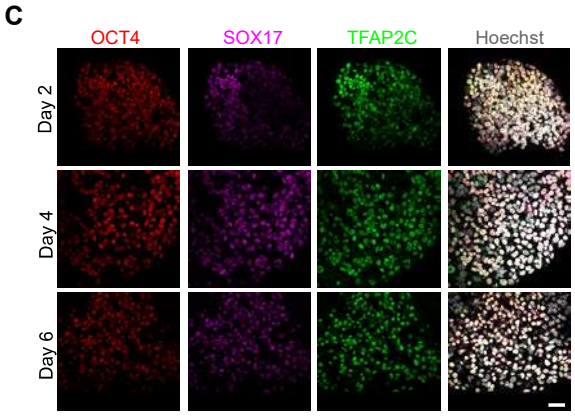
The chromatin accessibility landscape reveals distinct transcriptional regulation in the induction of human primordial germ cell-like cells from pluripotent stem cells

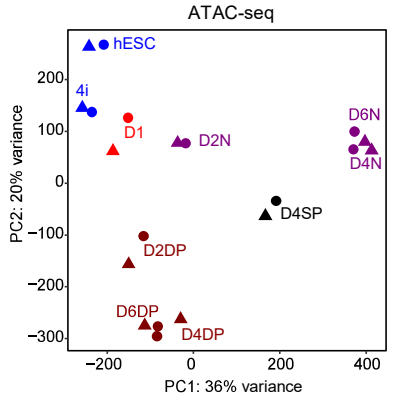
Xiaoman Wang, Veeramohan Veerapandian, Xinyan Yang, Ke Song, Xiaoheng Xu, Manman Cui, Weiyan Yuan, Yaping Huang, Xinyu Xia, Zhaokai Yao, Cong Wan, Fang Luo, Xiuling Song, Xiaoru Wang, Yi Zheng, Andrew Paul Hutchins, Ralf Jauch, Meiyan Liang, Chenhong Wang, Zhaoting Liu, Gang Chang, and Xiao-Yang Zhao

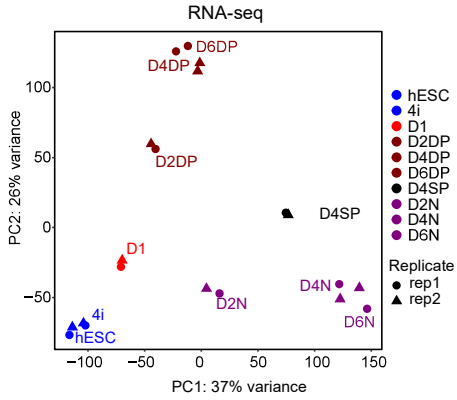
Figure S1

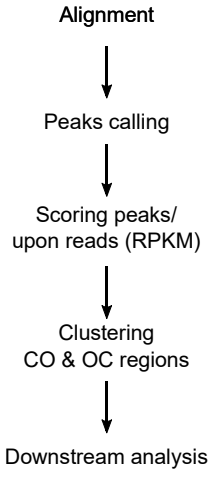


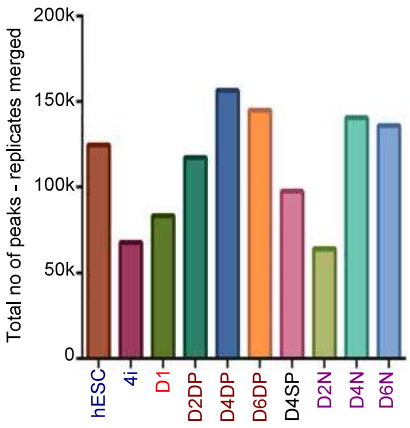


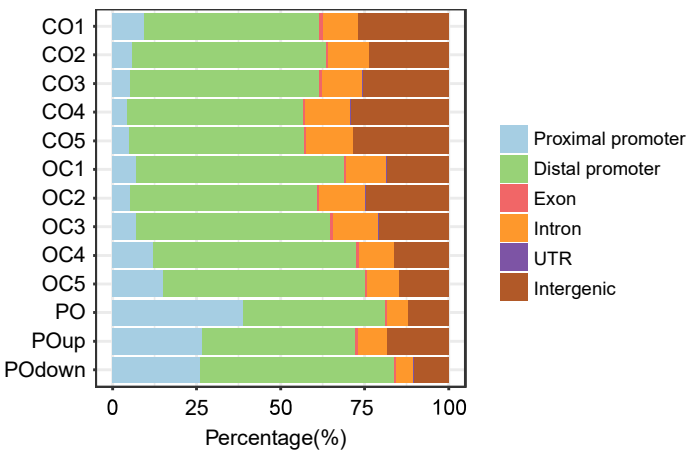












I

de novo motifs		% of Targets	% of Background	P-value	rank
CO1	AP-2gamma/TFAP2C	49.06%	18.95%	1e-145	1
CO2	OCT4:SOX17	10.22%	1.21%	1e-292	2
CO3	OCT4:SOX17	16.75%	4.76%	1e-427	2
CO4	SOX15	5.91%	0.48%	1e-245	2
CO5	SOX15	24.03%	10.52%	1e-88	2

J

de novo motifs		% of Targets	% of Background	P-value	rank
OC1	TEAD	23.13%	11.94%	1e-197	1
OC2	OCT4:SOX2	24.38%	6.2%	1e-334	1
OC3	OCT4:SOX2	21.17%	5.21%	1e-271	1
OC4	ZIC	62.80%	43.83%	1e-131	1
OC5	BORIS	16.12%	3.14%	1e-286	1

Figure S1. Quality Assessment of hPGCLC Induction, ATAC-seq and RNA-seq, Related to Figure 1

(A) FACS analysis of 4i cells and day 1–8 embryoids with EpCAM and INTEGRINα6 markers to detect hPGCLCs. DP, N and SP shown in Figure 1A are marked. DP, $EpCAM^{+}/INTEGRIN\alpha6^{+}$ cells; N, $EpCAM^{-}/INTEGRIN\alpha6^{-}$ cells; SP, $EpCAM^{+}/INTEGRIN\alpha6^{-}$ cells.

(B) Quantitative gene expression analysis of the indicated genes in 4i cells and $EpCAM^{+}/INTEGRIN\alpha6^{+}$ cells and $EpCAM^{-}/INTEGRIN\alpha6^{-}$ cells of day 4 embryoids. Relative expression levels are shown normalized to *GAPDH*. Error bars indicate mean ± SD from at least three independent biological replicates. Ep/Int(+): $EpCAM^{+}/INTEGRIN\alpha6^{+}$ cells, Ep/Int(-): $EpCAM^{-}/INTEGRIN\alpha6^{-}$ cells.

(C) Immunofluorescence of OCT4, SOX17 and TFAP2C in embryoids at day 2, 4 and 6. Scale bar, 100 μm.

(D, E) PCA of ATAC-seq (D) and RNA-seq (E) data of the indicated samples. The two independent replicates are represented as triangle and circle dots, respectively.

(F) Schematic representation of ATAC-seq analysis workflow.

(G) Bar plot showing the number of ATAC-seq peaks in all indicated samples.

(H) Bar chart showing the percentage of genomic feature distribution on CO, OC and PO chromatin regions.

(I, J) Top 2 *de novo* motif logos in CO and OC regions are highlighted with scores and ranking.

Figure S2

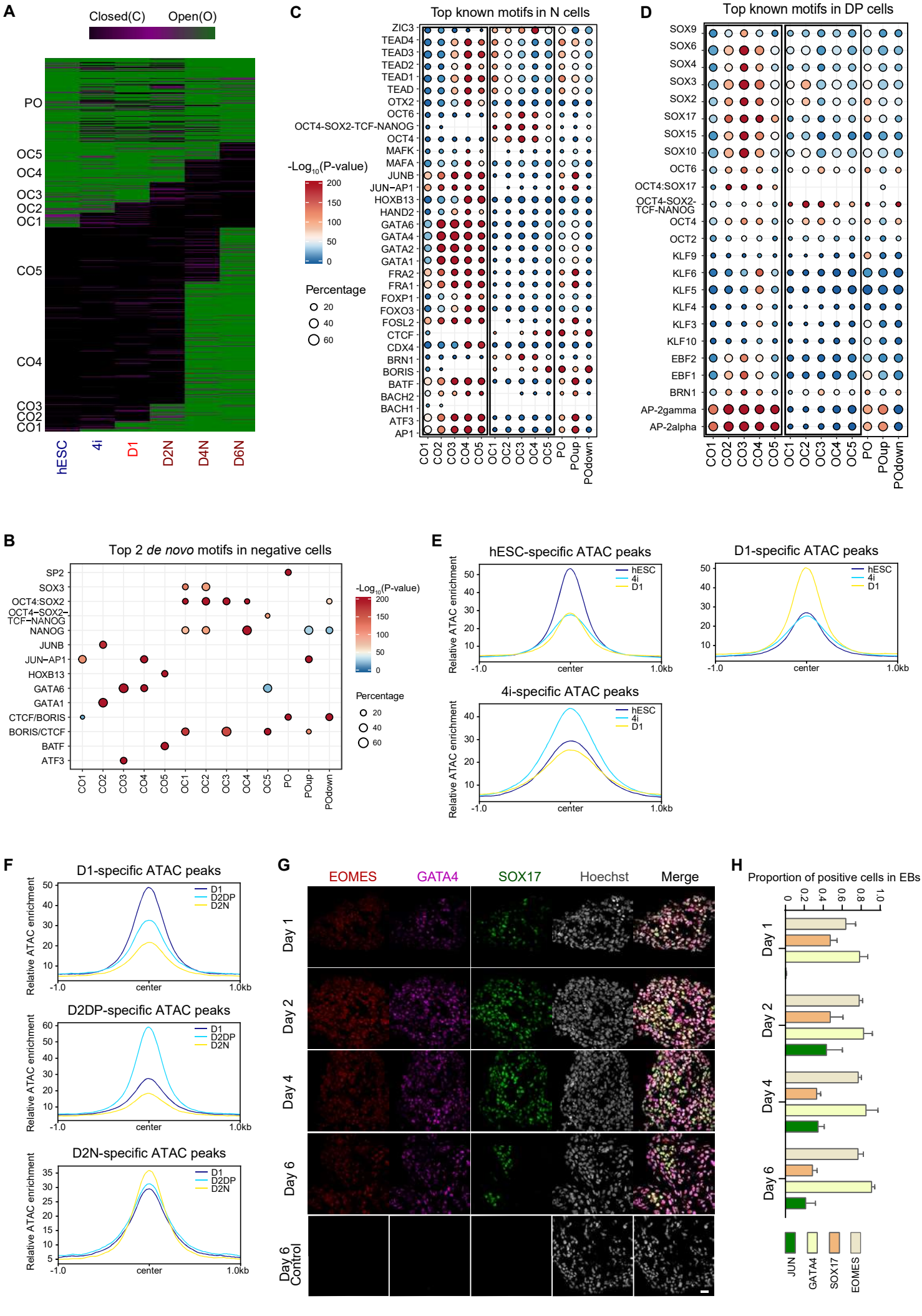

A
Closed(C)
Open(O)
PO
OC5
OC4
OC3
OC2
OC1
CO5
CO4
CO3
CO2
CO1
hESC
4i
D1
D2N
D4N
D6N
C
Top known motifs in N cells
ZIC3
TEAD4
TEAD3
TEAD2
TEAD1
TEAD
OTX2
OCT6
OCT4-SOX2-TCF-NANOG
OCT4
MAFK
MAFA
JUNB
JUN-AP1
HOXB13
HAND2
GATA6
GATA4
GATA2
GATA1
FRA2
FRA1
FOXP1
FOXO3
FOSL2
CTCF
CDX4
BRN1
BORIS
BATF
BACH2
BACH1
ATF3
AP1
-Log10(P-value)
200
150
100
50
0
Percentage
20
40
60
CO1 CO2 CO3 CO4 CO5 OC1 OC2 OC3 OC4 OC5 PO POup POdown
D
Top known motifs in DP cells
SOX9
SOX6
SOX4
SOX3
SOX2
SOX17
SOX15
SOX10
OCT6
OCT4:SOX17
OCT4-SOX2-TCF-NANOG
OCT4
OCT2
KLF9
KLF6
KLF5
KLF4
KLF3
KLF10
EBF2
EBF1
BRN1
AP-2gamma
AP-2alpha
CO1 CO2 CO3 CO4 CO5 OC1 OC2 OC3 OC4 OC5 PO POup POdown
B
Top 2 de novo motifs in negative cells
SP2
SOX3
OCT4:SOX2
OCT4-SOX2-TCF-NANOG
JUNB
JUN-AP1
HOXB13
GATA6
GATA1
CTCF/BORIS
BORIS/CTCF
BATF
ATF3
-Log10(P-value)
200
150
100
50
0
Percentage
20
40
60
CO1 CO2 CO3 CO4 CO5 OC1 OC2 OC3 OC4 OC5 PO POup POdown
E
hESC-specific ATAC peaks
D1-specific ATAC peaks
4i-specific ATAC peaks
Relative ATAC enrichment
hESC
4i
D1
-1.0
center
1.0kb
F
D1-specific ATAC peaks
D2DP-specific ATAC peaks
D2N-specific ATAC peaks
Relative ATAC enrichment
D1
D2DP
D2N
-1.0
center
1.0kb
G
EOMES
GATA4
SOX17
Hoechst
Merge
Day 1
Day 2
Day 4
Day 6
Day 6 Control
H
Proportion of positive cells in EBs
0
0.2
0.4
0.6
0.8
1.0
Day 1
Day 2
Day 4
Day 6
JUN
GATA4
SOX17
EOMES

Figure S2. Chromatin Dynamics from hESCs to Negative Cells during hPGCLC Induction, Related to Figure 1 and Figure 2

(A) Heatmap showing dynamically closed open (CO), open closed (OC) and permanently open (PO) chromatin groups. CO, OC and PO refer to closed in hESCs but open in D6N, open in hESCs but closed in D6N and permanently open in both hESCs and D6N, respectively. CO and OC are separated into 5 subgroups (CO1-5; OC1-5) based on the day when they changed from closed to open or open to closed.

(B) Bubble plot showing the Top 2 *de novo* motifs enriched in CO/OC/PO categories in $EpCAM^-$/$INTEGRIN\alpha6^-$ cells.

(C) Bubble plot showing the top known motifs enriched in CO/OC/PO categories in $EpCAM^-$/$INTEGRIN\alpha6^-$ cells.

(D) Bubble plot showing the top known motifs enriched in CO/OC/PO categories in $EpCAM^+$/$INTEGRIN\alpha6^+$ cells as described in Figure 1D. In panels B, C, D, the size of the bubble represents the percentage of respective motifs in each library and the significance of P-value are shown as gradient color code.

(E) Pileup of the ATAC-seq signals in hESCs, 4i and day 1 cells at the regions with specific peaks as shown in Figure 2A.

(F) Pileup of the ATAC-seq signals in D1, D2DP and D2N cells at the regions with specific peaks as shown in Figure 2D.

(G) Immunofluorescence of EOMES, GATA4 and SOX17 in EBs from day 1 to day 6, Scale bar, 50 μm. The day 6 EBs group incubating without primary antibodies as the control.

(H) The proportion of EOMES, JUN, GATA4 and SOX17 positive cells in EBs at day 1, day 2, day 4 and day 6. At least 5 slides of immunostaining from two independent experiments were used.

Figure S3

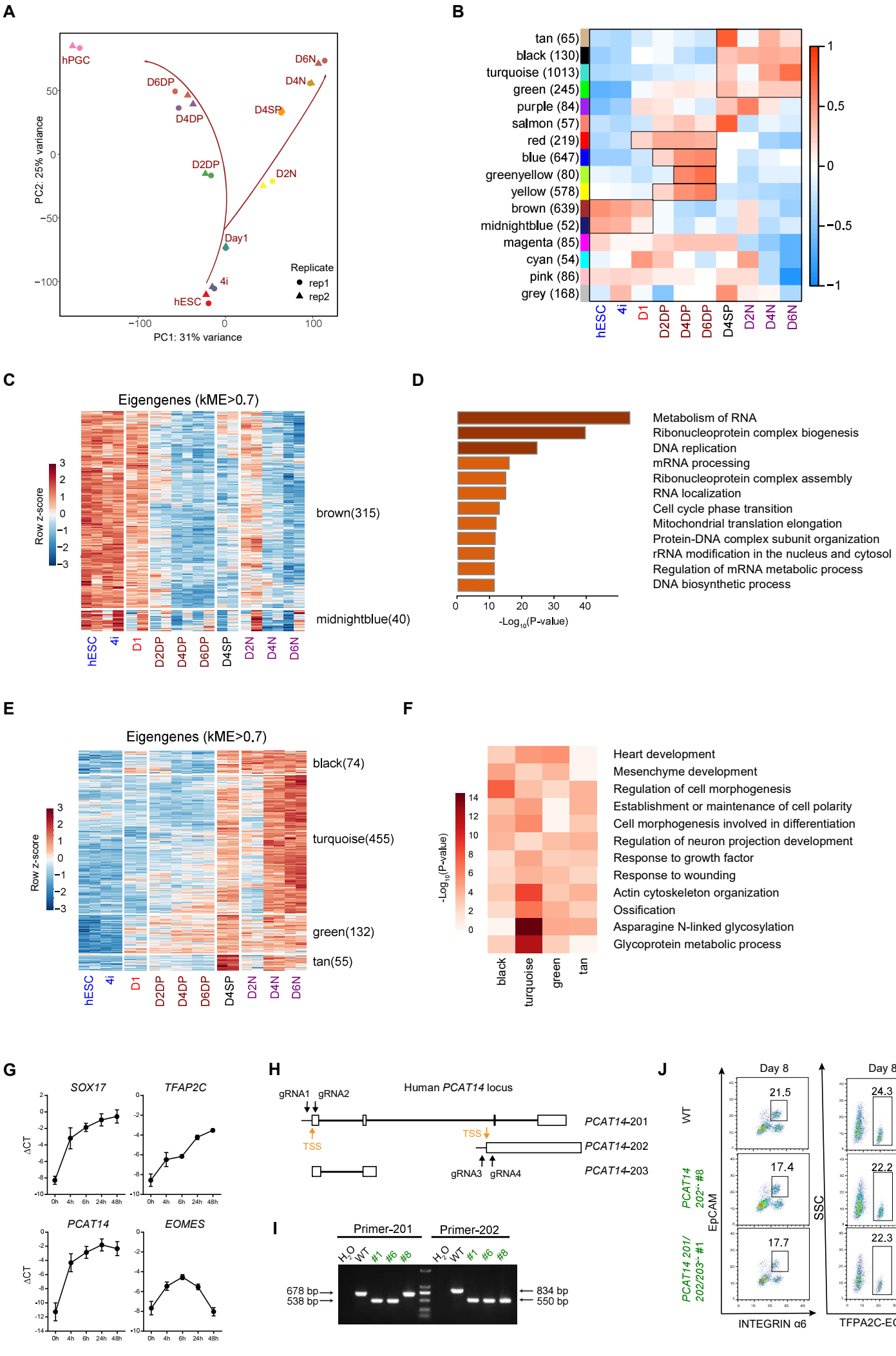

A
hPGC
D6DP
D6N
D4N
D4DP
D4SP
D2DP
D2N
Day1
4i
hESC
Replicate
rep1
rep2
PC2: 25% variance
PC1: 31% variance
B
tan (65)
black (130)
turquoise (1013)
green (245)
purple (84)
salmon (57)
red (219)
blue (647)
greenyellow (80)
yellow (578)
brown (639)
midnightblue (52)
magenta (85)
cyan (54)
pink (86)
grey (168)
hESC 4i D1 D2DP D4DP D6DP D4SP D2N D4N D6N
C
Eigengenes (kME>0.7)
Row z-score
brown(315)
midnightblue(40)
D
Metabolism of RNA
Ribonucleoprotein complex biogenesis
DNA replication
mRNA processing
Ribonucleoprotein complex assembly
RNA localization
Cell cycle phase transition
Mitochondrial translation elongation
Protein-DNA complex subunit organization
rRNA modification in the nucleus and cytosol
Regulation of mRNA metabolic process
DNA biosynthetic process
-Log10(P-value)
E
Eigengenes (kME>0.7)
Row z-score
black(74)
turquoise(455)
green(132)
tan(55)
F
Heart development
Mesenchyme development
Regulation of cell morphogenesis
Establishment or maintenance of cell polarity
Cell morphogenesis involved in differentiation
Regulation of neuron projection development
Response to growth factor
Response to wounding
Actin cytoskeleton organization
Ossification
Asparagine N-linked glycosylation
Glycoprotein metabolic process
-Log10(P-value)
black turquoise green tan
G
SOX17
TFAP2C
PCAT14
EOMES
ΔCT
0h 4h 6h 24h 48h
H
Human PCAT14 locus
gRNA1 gRNA2
TSS
PCAT14-201
PCAT14-202
PCAT14-203
gRNA3 gRNA4
I
Primer-201
Primer-202
H2O WT #1 #6 #8
678 bp
538 bp
834 bp
550 bp
J
Day 8
WT
PCAT14 202-/- #8
PCAT14 201/ 202/203-/- #1
21.5
17.4
17.7
24.3
22.2
22.3
EpCAM
INTEGRIN α6
SSC
TFPA2C-EGFP

Figure S3. Transcriptional Dynamics Associated with the Accessible Genome, Related to Figure 3

(A) PCA of RNA-seq data of the indicated samples including the dataset of hPGCs (Irie et al., 2015; Kanai-Azuma et al., 2002; Magnusdottir et al., 2013; Tang et al., 2015). Sample libraries are color coded and replicates are represented by triangle and circle. (B) Heatmap showing the gene expression patterns of 16 modules in the indicated samples. All 4202 genes as described in Figure 2A were used for hierarchical clustering via Weighted Correlation Network Analysis (WGCNA).

(C) Heatmap showing the gene expression of selected modules in which genes were highly expressed in hESCs, 4i cells and day 1 cells.

(D) Gene Ontology analysis of the genes as described in (C).

(E) Heatmap showing the gene expression of selected modules in which genes were highly expressed in EpCAM^{-}/INTEGRIN$\alpha 6^{-}$ cells.

(F) Gene Ontology analysis of the genes as described in (E).

(G) Quantitative gene expression analysis of the *PCAT14* and other key hPGCLC genes in embryoids at the indicated timepoints during the induction of hPGCLCs. Relative expression levels are shown normalized to *GAPDH*. Error bars indicate mean ± SD from two independent replicates.

(H) Targeting strategy of *PCAT14* knockout in hESCs with the designated guide RNA (red).

(I) Validation of the deletion of DNA fragments in hESCs of WT and *PCAT14*$^{-/-}$ lines via PCR.

(J) FACS analysis for EpCAM and INTEGRINα6 or TFAP2C-EGFP expression of day 8 embryoids derived from WT and PCAT14 knockout 4i hESCs.

Figure S4

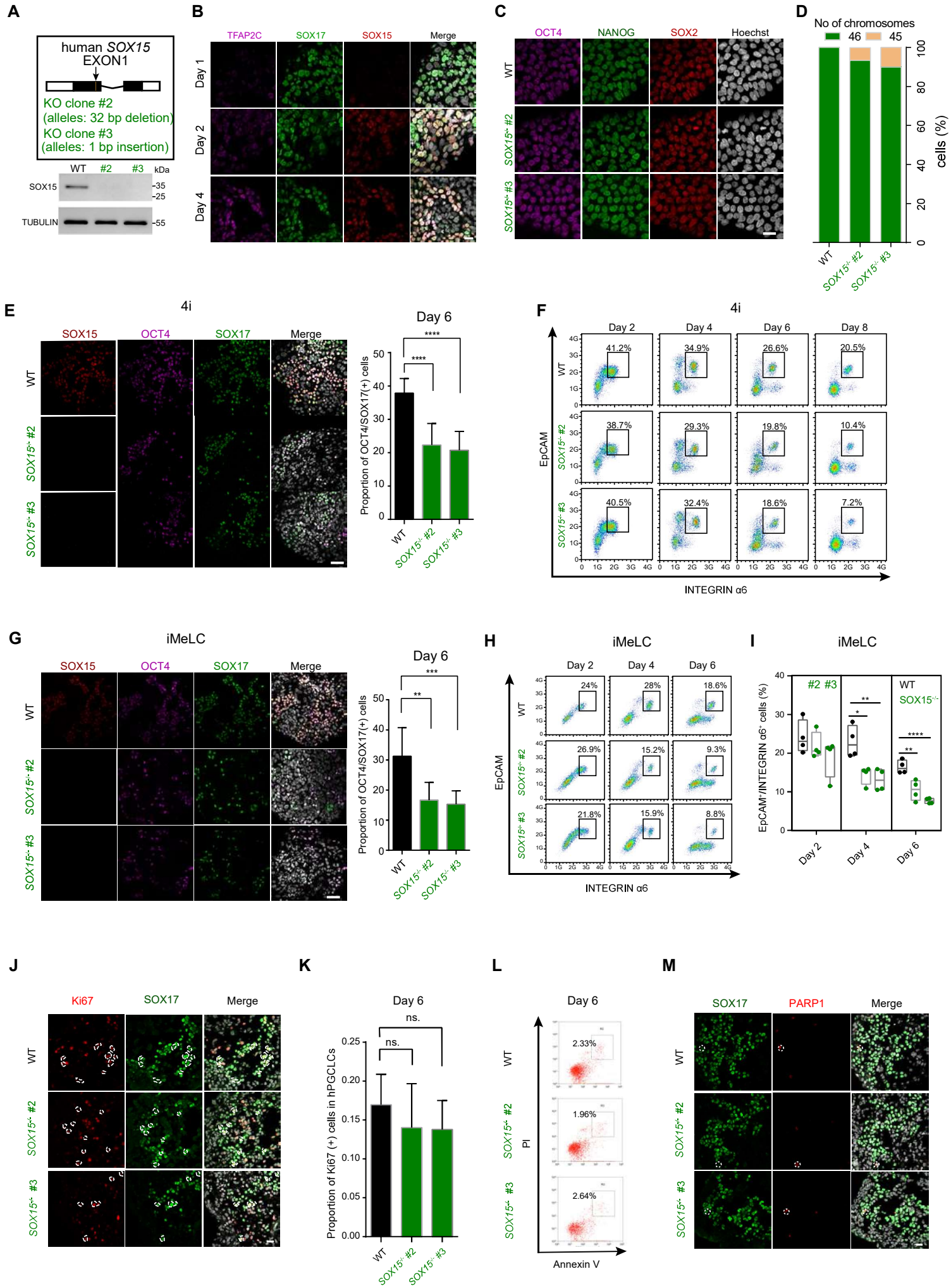

A
human SOX15 EXON1
KO clone #2 (alleles: 32 bp deletion)
KO clone #3 (alleles: 1 bp insertion)
WT #2 #3 kDa
SOX15
TUBULIN
-35
-25
-55
B
TFAP2C
SOX17
SOX15
Merge
Day 1
Day 2
Day 4
C
OCT4
NANOG
SOX2
Hoechst
WT
SOX15-/- #2
SOX15-/- #3
D
No of chromosomes
46
45
cells (%)
E
4i
SOX15
OCT4
SOX17
Merge
Day 6
Proportion of OCT4/SOX17(+) cells
F
4i
Day 2
Day 4
Day 6
Day 8
41.2%
34.9%
26.6%
20.5%
38.7%
29.3%
19.8%
10.4%
40.5%
32.4%
18.6%
7.2%
EpCAM
INTEGRIN α6
G
iMeLC
Day 6
H
iMeLC
24%
28%
18.6%
26.9%
15.2%
9.3%
21.8%
15.9%
8.8%
I
iMeLC
#2 #3
WT
SOX15-/-
EpCAM-/INTEGRIN α6- cells (%)
J
Ki67
SOX17
Merge
K
Day 6
ns.
Proportion of Ki67 (+) cells in hPGCLCs
L
Day 6
2.33%
1.96%
2.64%
PI
Annexin V
M
SOX17
PARP1
Merge

Figure S4. Genetic Ablation of SOX15 Impact the Efficiency of hPGCLC Induction, Related to Figure 3

(A) The position of guide RNA (black arrow) for SOX15 knockout in hESCs and the resulting two knockout lines with indicated deleted or inserted sequences. The lower panel shows the SOX15 protein expression in day 4 embryoids of WT and SOX15 KO lines by western blot. TUBULIN was used as the inner control.

(B) Immunofluorescence analysis of TFAP2C, SOX17 and SOX15 in day 1, day 2 and day 4 embryoids. Scale bar, 20 μm.

(C) Immunofluorescence analysis of OCT4, NANOG and SOX2 in WT and SOX15 knockout hESC lines. Scale bar, 20 μm.

(D) Karyotypes represented by the percentages of the indicated chromosome numbers of WT and *SOX15*$^{-/-}$ cell lines. The color-coding is as indicated.

(E) Immunofluorescence analysis of SOX15, OCT4 and SOX17 in day 6 WT and KO embryoids via 4i method (left) and the percentages of SOX17^{+}/OCT4^{+} cells (right) in embryoids. Scale bar, 100 μm. 8 slides of immunostaining from three independent experiments were used. Two-tailed Student's t test was performed, ****$P < 0.0001$.

(F) FACS analysis for the expression of EpCAM and INTEGRINα6 in embryoids derived from WT and knockout 4i hESCs upon hPGCLC induction at the indicated days.

(G) Immunofluorescence analysis of SOX15, OCT4 and SOX17 in day 6 WT and KO embryoids via iMeLC method (left) and the percentages of SOX17^{+}/OCT4^{+} cells (right) in embryoids. Scale bar, 100 μm. 8 slides of immunostaining from three independent experiments were used. Two-tailed Student's t test was performed, **$P < 0.01$, ***$P < 0.001$.

(H) FACS analysis for the expression of EpCAM and INTEGRINα6 in embryoids derived from WT and knockout iMeLCs upon hPGCLC induction at the indicated days.

(I) The percentages of EpCAM^{+}/INTEGRINα6^{+} cells in embryoids derived from WT (black) and knockout lines (green) upon hPGCLC induction at the indicated days via iMeLC method. Results of 4 independent experiments were shown (n = 4). Two-tailed Student's t test was performed, *$P < 0.05$, **$P < 0.01$, ****$P < 0.0001$.

(J-K) Immunofluorescence analysis of Ki67 and SOX17 in day 6 WT and SOX15 knockout embryoids, scale bar, 50 μm (J), and the proportion of Ki67 positive cells in SOX17^{+} hPGCLCs

(K) 8 slides of immunostaining from three independent experiments were used. Two-tailed Student's t test was performed, ns, not significant.

(L) FACS analysis of the apoptosis status in day 6 $EpCAM^{+}$/$INTEGRIN\alpha6^{+}$ cells derived from WT and SOX15 knockout 4i cells by staining with PI and Annexin V.

(M) Immunofluorescence analysis of PARP1 in D6 $SOX17^{+}$ hPGCLCs derived from WT and SOX15 knockout 4i hESCs, scale bar, 20 μm.

Figure S5

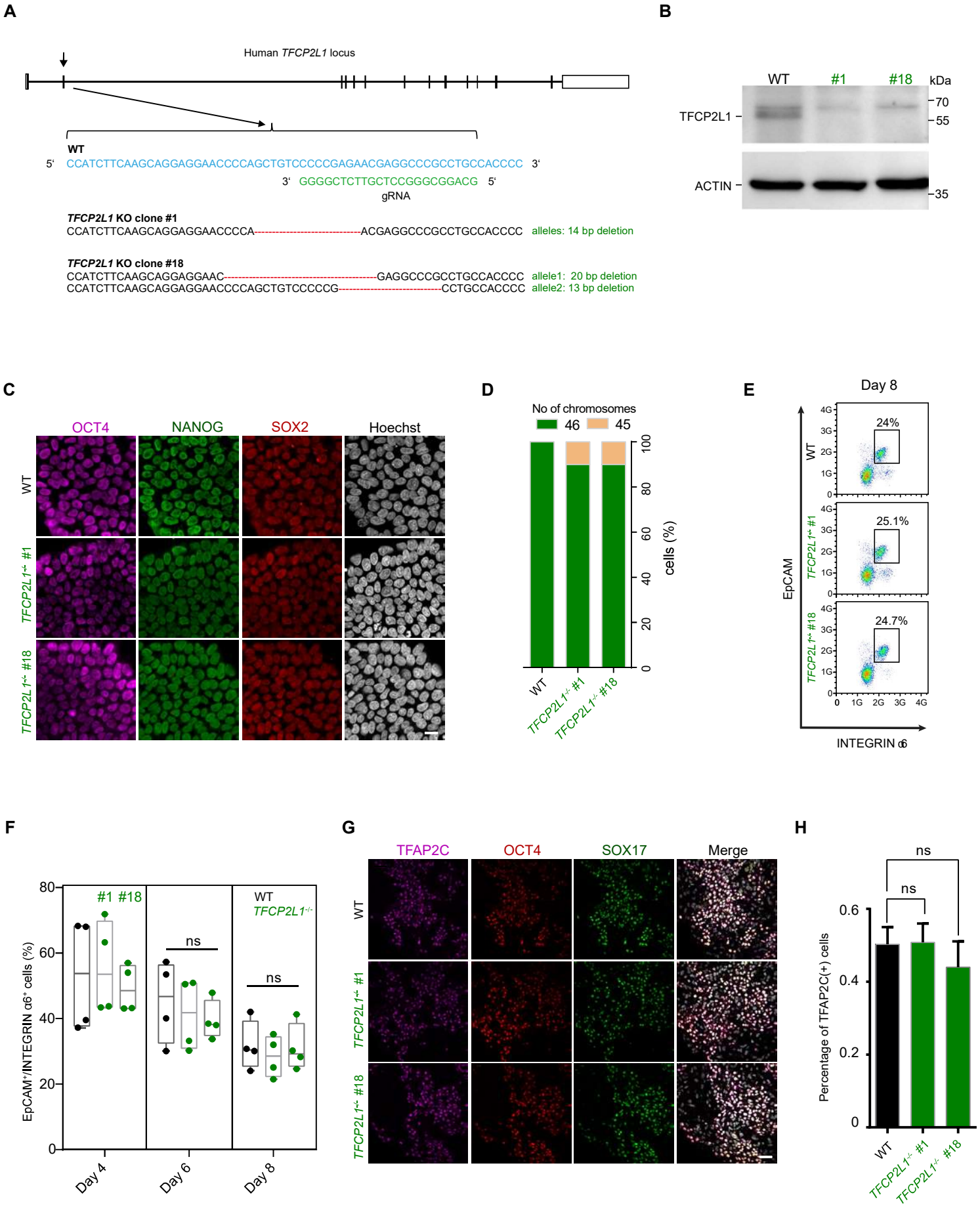

A
Human TFCP2L1 locus
WT
5' CCATCTTCAAGCAGGAGGAACCCCAGCTGTCCCCCGAGAACGAGGCCCGCCTGCCACCCC 3'
3' GGGGCTCTTGCTCCGGGCGGACG 5'
gRNA
TFCP2L1 KO clone #1
CCATCTTCAAGCAGGAGGAACCCCA----------ACGAGGCCCGCCTGCCACCCC alleles: 14 bp deletion
TFCP2L1 KO clone #18
CCATCTTCAAGCAGGAGGAAC----------GAGGCCCGCCTGCCACCCC allele1: 20 bp deletion
CCATCTTCAAGCAGGAGGAACCCCAGCTGTCCCCCG----------CCTGCCACCCC allele2: 13 bp deletion
B
WT
#1
#18
kDa
TFCP2L1
ACTIN
70
55
35
C
OCT4
NANOG
SOX2
Hoechst
WT
TFCP2L1-/- #1
TFCP2L1-/- #18
D
No of chromosomes
46
45
cells (%)
WT
TFCP2L1-/- #1
TFCP2L1-/- #18
E
Day 8
24%
25.1%
24.7%
EpCAM
INTEGRIN α6
F
#1 #18
WT
TFCP2L1-/-
ns
EpCAM+/INTEGRIN α6+ cells (%)
Day 4
Day 6
Day 8
G
TFAP2C
OCT4
SOX17
Merge
WT
TFCP2L1-/- #1
TFCP2L1-/- #18
H
ns
Percentage of TFAP2C(+) cells
WT
TFCP2L1-/- #1
TFCP2L1-/- #18

Figure S5. TFCP2L1 is Dispensable for hPGCLC Induction, Related to Figure 4

(A) Targeting strategy of TFCP2L1 knockout in hESCs with the designated guide RNA (green) and the resulting two *TFCP2L1*$^{-/-}$ lines with altered sequences.

(B) Western blot analysis of the expression of TFCP2L1 protein in day 4 embryoids of WT and *TFCP2L1*$^{-/-}$ lines. ACTIN was used as the inner control.

(C) Immunofluorescence analysis of OCT4, NANOG and SOX2 in WT and *TFCP2L1*$^{-/-}$ hESCs. Scale bar, 20 μm.

(D) Karyotypes represented by the percentages of the indicated chromosome numbers of WT and TFCP2L1 knockout lines. The color-coding is as indicated.

(E) FACS analysis for the expression of EpCAM and INTEGRINα6 in day 8 embryoids derived from WT and knockout 4i hESCs.

(F) The percentages of EpCAM^{+}/INTEGRINα6^{+} cells in the embryoids of WT (black) and TFCP2L1knockout lines (green) upon hPGCLC induction at the indicated days via the 4i method. Results of 4 independent experiments were shown (n = 4). Two-tailed Student's t test was performed, ns, not significant.

(G) Immunofluorescence analysis of TFAP2C, OCT4 and SOX17 in day 6 WT and TFCP2L1 KO embryoids via 4i method, scale bar, 100 μm.

(H) The percentage of TFAP2C positive cells in day 6 WT and TFCP2L1 KO embryoids (G), 8 slides of immunostaining from three independent experiments were used. Two-tailed Student's t test was performed, ns, not significant.

Figure S6

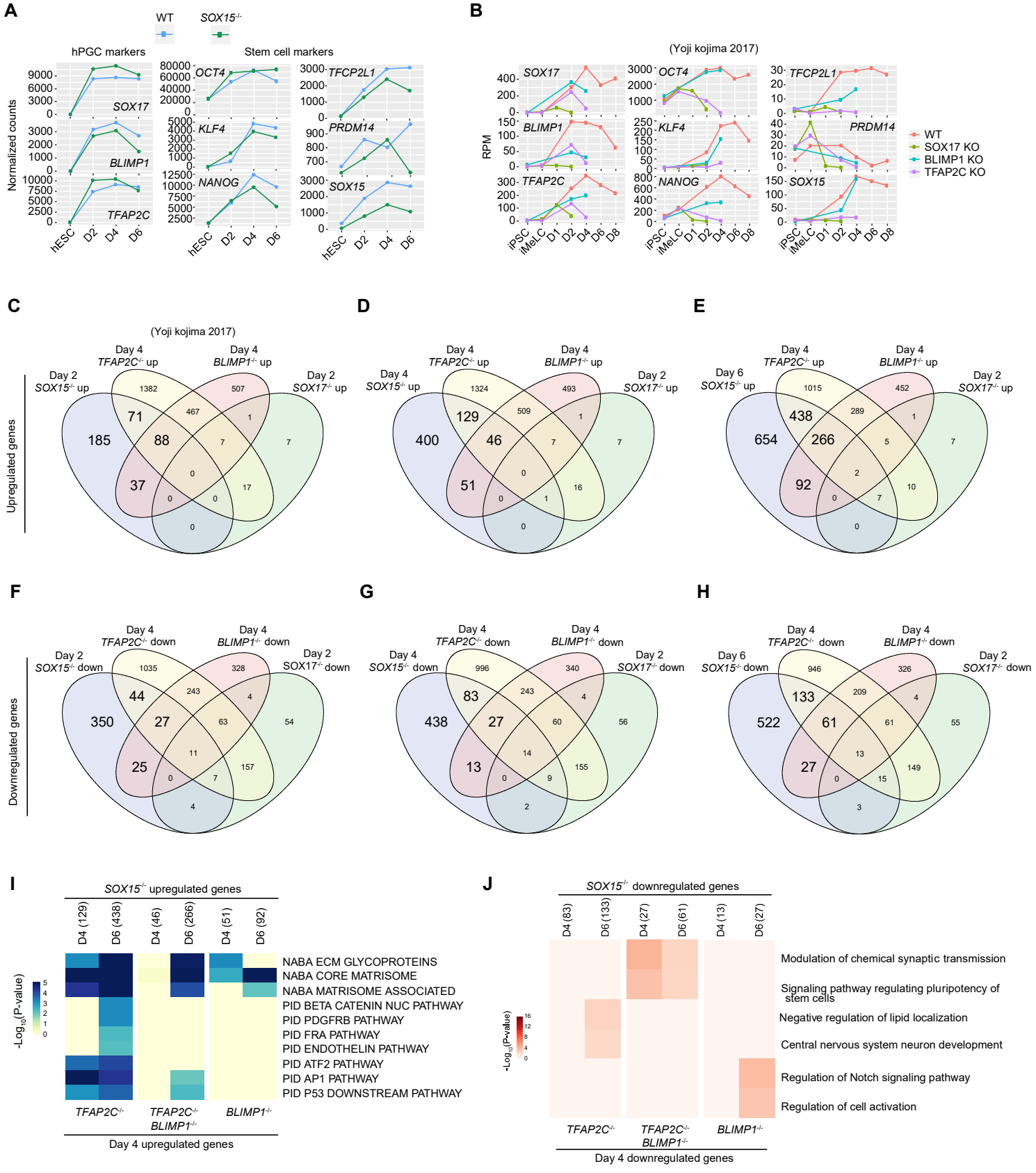

A
WT
SOX15-/-
hPGC markers
Stem cell markers
Normalized counts
SOX17
BLIMP1
TFAP2C
OCT4
KLF4
NANOG
TFCP2L1
PRDM14
SOX15
hESC
D2
D4
D6
B
(Yoji kojima 2017)
RPM
iPSC
iMeLC
D1
D8
WT
SOX17 KO
BLIMP1 KO
TFAP2C KO
C
(Yoji kojima 2017)
Upregulated genes
Day 2 SOX15-/- up
Day 4 TFAP2C-/- up
Day 4 BLIMP1-/- up
Day 2 SOX17-/- up
185
1382
507
71
467
1
88
7
7
0
37
0
0
17
0
D
Day 4 SOX15-/- up
Day 4 TFAP2C-/- up
Day 4 BLIMP1-/- up
Day 2 SOX17-/- up
400
1324
493
129
509
1
46
7
7
0
51
0
1
16
0
E
Day 6 SOX15-/- up
Day 4 TFAP2C-/- up
Day 4 BLIMP1-/- up
Day 2 SOX17-/- up
654
1015
452
438
289
1
266
5
7
2
92
0
7
10
0
F
Downregulated genes
Day 2 SOX15-/- down
Day 4 TFAP2C-/- down
Day 4 BLIMP1-/- down
Day 2 SOX17-/- down
350
1035
328
44
243
4
27
63
54
11
25
0
7
157
4
G
Day 4 SOX15-/- down
Day 4 TFAP2C-/- down
Day 4 BLIMP1-/- down
Day 2 SOX17-/- down
438
996
340
83
243
4
27
60
56
14
13
0
9
155
2
H
Day 6 SOX15-/- down
Day 4 TFAP2C-/- down
Day 4 BLIMP1-/- down
Day 2 SOX17-/- down
522
946
326
133
209
4
61
61
55
13
27
0
15
149
3
I
SOX15-/- upregulated genes
D4 (129)
D6 (438)
D4 (46)
D6 (266)
D4 (51)
D6 (92)
-Log10(P-value)
NABA ECM GLYCOPROTEINS
NABA CORE MATRISOME
NABA MATRISOME ASSOCIATED
PID BETA CATENIN NUC PATHWAY
PID PDGFRB PATHWAY
PID FRA PATHWAY
PID ENDOTHELIN PATHWAY
PID ATF2 PATHWAY
PID AP1 PATHWAY
PID P53 DOWNSTREAM PATHWAY
TFAP2C-/-
TFAP2C-/- BLIMP1-/-
BLIMP1-/-
Day 4 upregulated genes
J
SOX15-/- downregulated genes
D4 (83)
D6 (133)
D4 (27)
D6 (61)
D4 (13)
D6 (27)
-Log10(P-value)
Modulation of chemical synaptic transmission
Signaling pathway regulating pluripotency of stem cells
Negative regulation of lipid localization
Central nervous system neuron development
Regulation of Notch signaling pathway
Regulation of cell activation
TFAP2C-/-
TFAP2C-/- BLIMP1-/-
BLIMP1-/-
Day 4 downregulated genes
K
BLIMP1 targets
BLIMP1/TFAP2C targets
Normalized counts
Neural differentiation (Yoji kojima 2017)
Germ cell developement (Yoji kojima 2017)
FGFR1
ERBB2
SOX2
NANOS3
ITGA6
DNMT3B
hESC
D2
D4
D6
WT
SOX15-/-

Figure S6. SOX15 might Act as a Downstream Regulator of TFAP2C, Related to Figure 4

(A) Line plots showing gene expression of germ cell and stem cell markers. The blue line represents the WT hPGCLCs and green line represents the *SOX15*$^{-/-}$ hPGCLCs.

(B) Line plots showing gene expression of the indicated genes throughout hPGCLC induction in WT, SOX17 KO (*SOX17*$^{-/-}$), BLIMP1 KO (*BLIMP1*$^{-/-}$) and TFAP2C KO (*TFAP2C*$^{-/-}$) cells (Kojima et al., 2017).

(C, D, E) Venn diagram showing the upregulated genes in *SOX15*$^{-/-}$, *TFAP2C*$^{-/-}$, *BLIMP1*$^{-/-}$ and *SOX17*$^{-/-}$ cells. The upregulated genes at day 2 (C), day 4 (D) and day 6 (E) in *SOX15*$^{-/-}$ cells were intersected with those of *TFAP2C*$^{-/-}$, *BLIMP1*$^{-/-}$ (day 4) or *SOX17*$^{-/-}$ (day 2) cells. The genes highlighted as bold in panels C, D and E Venn are used for the analysis.

(F, G, H) Venn diagram showing the downregulated genes in *SOX15*$^{-/-}$, *TFAP2C*$^{-/-}$, *BLIMP1*$^{-/-}$ and *SOX17*$^{-/-}$ cells. The downregulated genes at day 2 (F), day 4 (G) and day 6 (H) in *SOX15*$^{-/-}$ cells were intersected with those of *TFAP2C*$^{-/-}$, *BLIMP1*$^{-/-}$ (day 4) or *SOX17*$^{-/-}$ (day 2) cells.

(I, J) Heatmap showing the canonical pathways enriched in the upregulated genes (I) or GO terms in the downregulated genes (J) (day 4 and day 6) in *SOX15*$^{-/-}$ cells shared with *BLIMP1*$^{-/-}$ or *TFAP2C*$^{-/-}$ cells (day 4). The DEGs from *SOX15*$^{-/-}$ cells, *BLIMP1*$^{-/-}$ and *TFAP2C*$^{-/-}$ cells are based on $\log_2$fold change > 1.

(K) Line plots showing the expression of downstream genes regulated by BLIMP1 alone or BLIMP1/TFAP2C.

Figure S7

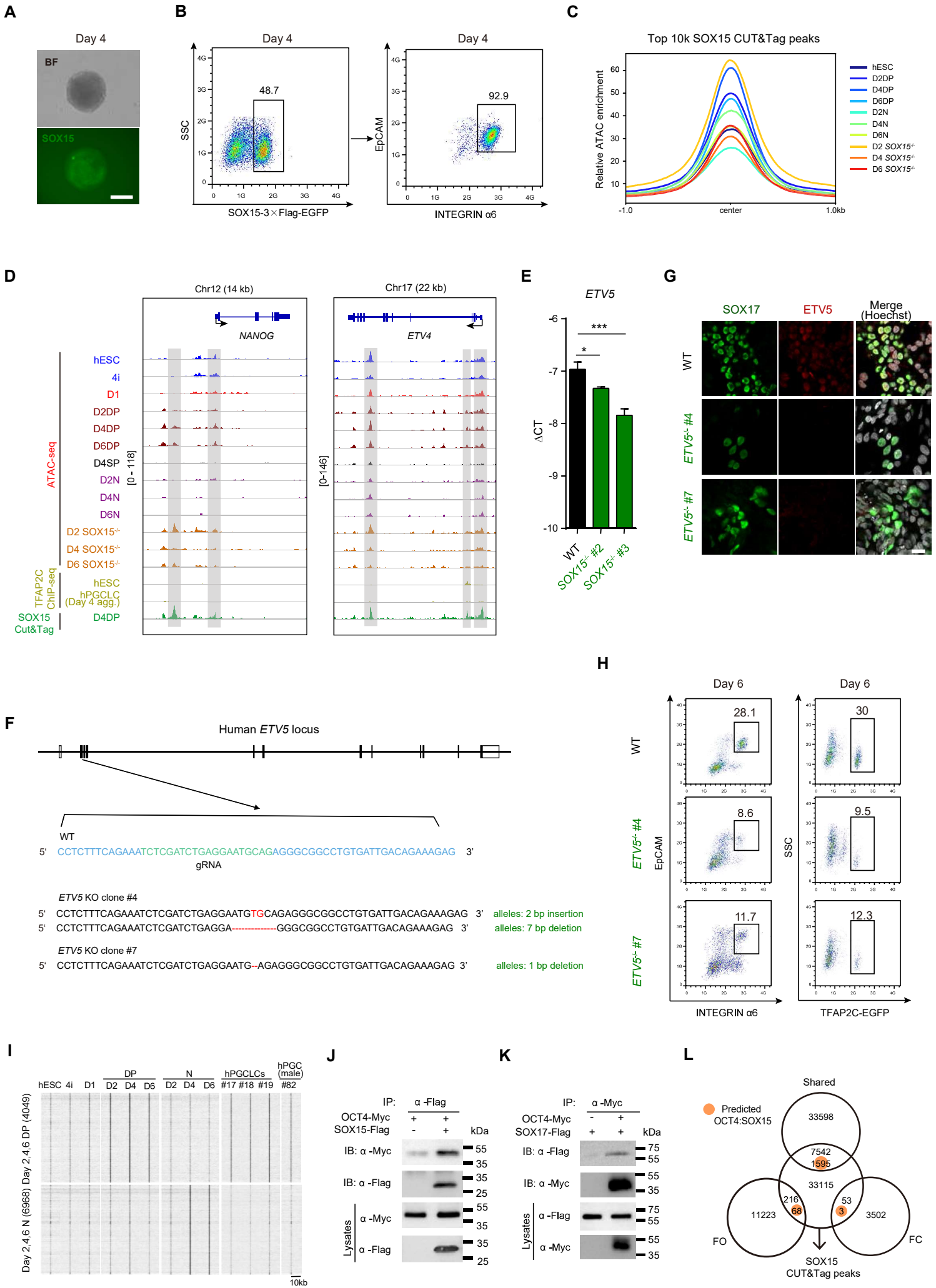

A
Day 4
BF
SOX15
B
Day 4
48.7
SSC
SOX15-3×Flag-EGFP
Day 4
92.9
EpCAM
INTEGRIN α6
C
Top 10k SOX15 CUT&Tag peaks
Relative ATAC enrichment
-1.0
center
1.0kb
hESC
D2DP
D4DP
D6DP
D2N
D4N
D6N
D2 SOX15-/-
D4 SOX15-/-
D6 SOX15-/-
D
Chr12 (14 kb)
NANOG
Chr17 (22 kb)
ETV4
ATAC-seq
hESC
4i
D1
D2DP
D4DP
D6DP
D4SP
D2N
D4N
D6N
D2 SOX15-/-
D4 SOX15-/-
D6 SOX15-/-
[0 - 118]
[0-146]
TFAP2C ChIP-seq
hESC
hPGCLC (Day 4 agg.)
SOX15 Cut&Tag
D4DP
E
ETV5
***
*
ΔCT
WT
SOX15-/- #2
SOX15-/- #3
G
SOX17
ETV5
Merge (Hoechst)
WT
ETV5-/- #4
ETV5-/- #7
F
Human ETV5 locus
WT
5' CCTCTTTCAGAAATCTCGATCTGAGGAATGCAGAGGGCGGCCTGTGATTGACAGAAAGAG 3'
gRNA
ETV5 KO clone #4
5' CCTCTTTCAGAAATCTCGATCTGAGGAATGTGCAGAGGGCGGCCTGTGATTGACAGAAAGAG 3' alleles: 2 bp insertion
5' CCTCTTTCAGAAATCTCGATCTGAGGA----------GGGCGGCCTGTGATTGACAGAAAGAG 3' alleles: 7 bp deletion
ETV5 KO clone #7
5' CCTCTTTCAGAAATCTCGATCTGAGGAATG--AGAGGGCGGCCTGTGATTGACAGAAAGAG 3' alleles: 1 bp deletion
H
Day 6
Day 6
WT
ETV5-/- #4
ETV5-/- #7
28.1
30
8.6
9.5
11.7
12.3
EpCAM
SSC
INTEGRIN α6
TFAP2C-EGFP
I
hESC 4i D1
DP D2 D4 D6
N D2 D4 D6
hPGCLCs #17 #18 #19
hPGC (male) #82
Day 2,4,6 DP (4049)
Day 2,4,6 N (6968)
10kb
J
IP: α -Flag
OCT4-Myc + +
SOX15-Flag - +
kDa
IB: α -Myc
IB: α -Flag
Lysates
α -Myc
α -Flag
K
IP: α -Myc
OCT4-Myc - +
SOX17-Flag + +
kDa
IB: α -Flag
IB: α -Myc
Lysates
α -Flag
α -Myc
L
Shared
Predicted OCT4:SOX15
33598
7542
1595
33115
216
68
53
3
11223
3502
FO
FC
SOX15 CUT&Tag peaks

Figure S7. ETV5 Acts as a Downstream Regulator of SOX15 in hPGCLC Maintenance and the interaction between OCT4 and SOX15 as well as SOX17, Related to Figure 6 and Figure 7

(A) Bright field (BF) and fluorescence (SOX15-EGFP) images of day 4 floating embryoids from SOX15-3×Flag-EGFP-Puro Knockin 4i hESCs. Scale bar, 200 μm.

(B) FACS analysis for the EGFP expression in day 4 embryoids derived from SOX15-3×Flag-P2A-EGFP-Puro Knockin 4i hESCs.

(C) Pileup of the ATAC-seq signals at the top 10K SOX15 CUT&Tag peaks regions in hESCs, WT EpCAM$^+$/INTEGRINα6$^+$ cells, WT EpCAM$^-$/INTEGRINα6$^-$ cells and SOX15 KO EpCAM$^+$/INTEGRINα6$^+$ cells.

(D) Selected genomic views showing the ATAC-seq signals, TFAP2C ChIP-seq signals (Chen et al., 2019) and SOX15 CUT&Tag signals at the *NANOG* and *ETV4* genome loci in the indicated samples. The specific open regions with SOX15 CUT&Tag signals and decreased ATAC-seq signals from day 4 SOX15 KO EpCAM$^+$/INTEGRINα6$^+$ cells compared to that in WT EpCAM$^+$/INTEGRINα6$^+$ cells are marked with a gray box.

(E) Quantitative gene expression analysis of the *ETV5* in EpCAM$^+$/INTEGRINα6$^+$ cells of day 4 embryoids derived from WT and *SOX15*$^{-/-}$ hESCs. Relative expression levels are shown normalized to *GAPDH*. Error bars indicate mean ± SD from three independent replicates. Two-tailed Student's t test was performed, $*P < 0.05$, $***P < 0.001$.

(F) Gene targeting strategy of *ETV5* knockout in hESCs with the designated guide RNA (green) and the resulting two *ETV5*$^{-/-}$ lines with altered sequences.

(G) Immunofluorescence analysis of SOX17 and ETV5 in day 6 WT and ETV5 KO embryoids. The KO embryoids show no ETV5 signals. Scale bar, 20 μm.

(H) FACS analysis for the expression of EpCAM/INTEGRINα6 or TFAP2C-EGFP in day 6 embryoids derived from WT and ETV5 KO hESCs via 4i method.

(I) Heatmap of ATAC-seq signals in the indicated samples at the DP specific (4049) and N specific (6968) open chromatin regions as described in Figure 7A. The hPGCLCs and hPGC ATAC-seq data (Chen et al., 2018) were included as the control.

(J-K) Co-immunoprecipitation analysis for the protein-protein interaction of SOX15 (J) or SOX17 (K) with OCT4 in HEK293T cells.

(L) The number of peaks near to the predicted OCT4:SOX15 motif sites included in Shared, FO and FC group as described in Figure 5, respectively.

Supplemental Tables

Table S1. The genome loci of peaks in CO1-CO5 and OC1-OC5 groups. Related to Figure 1

Table S2. WGCNA analysis of CO/PO union genes and GO analysis of genes in selected modules. Related to Figure 3 and Figure S3

Table S3. Differentially expressed genes between SOX15 KO cells and WT cells and GO analysis of genes upregulated/downregulated in SOX15 KO $EpCAM^{+}$/$INTEGRIN\alpha6^{+}$ compared to wild-type $EpCAM^{+}$/$INTEGRIN\alpha6^{+}$ at day 6. Related to Figure 4

Table S4. Co-upregulated/co-downregulated genes between *SOX15*$^{-/-}$ cells and *TFAP2C*$^{-/-}$ cells or *BLIMP1*$^{-/-}$ cells. Related to Figure 4 and Figure S6

Table S5. Day 6 *SOX15*$^{-/-}$ upregulated and downregulated genes near to shared, FO and FC regions and GO analysis. Related to Figure 5

Table S6. Day 6 *SOX15*$^{-/-}$ upregulated and downregulated genes near to SOX15 CUT&Tag top 10k peaks or SOX15 peaks including predicted OCT4:SOX15 binding sites and GO analysis. Related to Figure 6 and Figure 7

Table S7. Primers for qPCR used in this study

Supplemental Experimental Procedures

Culture of hESCs

The Fy-hES-3 and all KO cell lines were cultured in feeder-free medium (CELLAPY, CA1001500) on Matrigel (Corning, 354277). Cell media were changed daily and cells were passaged every 4 to 6 days using EDTA (CELLAPY, CA3001500).

Induction of 4i hESCs and hPGCLCs

hPGCLCs were generated from hESCs based on the protocol (Mitsunaga et al., 2017) with slight modifications. The hESCs on Matrigel were treated with TrypLE^{TM} Express enzyme to enable their dissociation into single cells. The 4i hESCs were induced by plating 7.0×10^4 hESCs per well of 12-well plate on mouse embryonic feeders (MEFs) in 4i medium containing KnockOut DMEM (ThermoFisher, 10829018), 20% Knockout Serum Replacement (ThermoFisher, 10828-028 or A3181501), 1% NEAA (ThermoFisher, 11140050), 1 mM sodium pyruvate (ThermoFisher, 11360070), 1% Glutmax (ThermoFisher, 35050061), and 0.055 mM 2-mercaptoethanol (ThermoFisher, 21985023), 20 ng/ml human LIF (Peprotech, 300-05-500), 8 ng/ml bFGF (SCI), 1 ng/ml TGF-β1 (Peprotech, 100-21), 3 mM CHIR99021 (TOCRIS, 4423), 1 mM PD0325901 (TOCRIS, 4192), 5 mM SB203580 (TOCRIS bioscience), and 5 mM SP600125 (TOCRIS). 10 μM of a ROCK inhibitor (R&D, 1254/10) was used for 24 h after the induction and then the medium without ROCK inhibitor was used. After 4 days of induction, the cells were dissociated with TrypLE and plated into ultra-low cell attachment U-bottom 96-well plates (Corning, 7007) at a density of 3,500–4,000 cells/well in 100 μl hPGCLC medium. hPGCLC medium is composed of GMEM (ThermoFisher, 11710-035), 15% KSR, 1% NEAA, 1 mM sodium pyruvate, 1% Glutmax, and 0.055 mM 2-mercaptoethanol (ThermoFisher), 300 ng/ml BMP4 (R&D Systems), 100 ng/ml SCF (Peprotech, 300-07), 50 ng/ml EGF (R&D Systems), 100 ng/ml human LIF (Peprotech, 300-05-500) and 10 mM ROCK inhibitor. The medium was not changed until the EBs were used.

Generation of Knockout hESC Lines

To knock out *SOX15* and *TFCP2L1* genes, guide RNAs (gRNA) were designed using

https://zlab.bio/guide-design-resources and cloned into pX330 vector. 10 μg pX330 constructs containing gRNA were electroporated into Fy-hES-3 cells using Neon^{TM} Transfection System (Thermofisher, MPK10096). Two days later, the top 1% EGFP positive cells were sorted by FACS and the sorted cells were picked manually into Matrigel-coated 96-well-plate at density of a single cell per well with mTeSR1 medium containing 10 μM ROCK inhibitor (R&D, Y27632). After 3 days, the medium was changed to fresh mTeSR1 with 2 μM Y27632 and one week later the cells were cultured in mTeSR1 without Y27632 until passage. Twelve to fifteen days after sorting, the survived clones were passaged into 24-well plates and half of the cells were harvested for genotyping. To determine the mutation sites, genomic DNA was extracted for sequencing. Human *SOX15* and *TFCP2L1* genes were targeted with the guide sequence GCTCCAGGCCTGGTCCTGTGAGG and GCAGGCGGGCCTCGTTCTCGGGG, respectively.

Generation of TFAP2C-p2A-EGFP-Puro Knockin, SOX15-3×Flag-p2A-EGFP Knockin hESC Lines, PCAT14 Knockout and ETV5 Knockout hESC Lines

The establishment of TFAP2C-p2A-EGFP and SOX15-3×Flag-p2A-EGFP-Puro hESCs were made as previously described (Sasaki et al., 2015) with slight modification. To construct the HMEJ donor for TFAP2C-p2A-EGFP knock in hESC lines, the homology arms flanking TFAP2C stop codon [left (5-prime) arm: 801 bp; right (3-prime) arm: 636 bp] were amplified by PCR using the primer pairs as listed in primers used in this study and sub-cloned into the T vector. The SOX15-3 × Flag-p2A-EGFP knock in HMEJ donor with the homology arms flanking SOX15 stop codon (left arm: 801 bp; right arm: 800 bp) were amplified by PCR using the primer pairs listed in primers and also sub-cloned into the T vector. The 3×Flag-p2A-EGFP fragments with CAG-puro cassettes flanked by LoxP sites were amplified by PCR, and then inserted in place of stop codon in T vector. The p2A-EGFP fragments with CAG-puro cassettes flanked by LoxP sites were amplified by PCR, and then inserted into the stop codon in T vector. We used pX330 (Addgene catalog no. 42230) to generate a single Cas-9-gRNA-EGFP vector. The CRISPR construct targeting the TFAP2C and SOX15 stop codon were generated as

described above with the following gRNA sequences: TGGAGAAAATGGAGAAACACAGG and ATGAGGGTTAGAGGTGGGTTAGG. The activities of the CRISPR were evaluated by T7E1 assay. All the plasmid constructs were extracted using the Plasmid Midi Kit (Qiagen, 12143) and verified by DNA sequencing. The method to electroporate plasmid into Fy-hES-3 hESCs was similar to that used in the generation of knockout SOX15/TFCP2L1 hESC lines. The PCAT14 and ETV5 knockout hESCs lines were also generated based on the TFAP2C-p2A-EGFP knockin hESCs using the same knockout strategy. Human *PCAT14* was targeted with the guide sequences: TTGTTCACATGTTTTCCTGC (gRNA1), CAAGTCTCTCGTTCCACCTG (gRNA2), GTCATGGGAGTTCCAGAAAA (gRNA3), AACAACATCTTACTGGTAAA (gRNA4) and *ETV5* was targeted with guide sequence TCTCGATCTGAGGAATGCAG.

Karyotype Analysis

Metaphase chromosomes from hESCs were harvested when the cells reached 60%~80% confluent density in 6-cm dish. Fresh medium with 250 ng/mL of Demecolcine (Sigma Aldrich, D1925) were used and the cells were incubated for 2~2.5 h. Then the cells were dissociated into single cells and treated with hypotonic solution (0.59 g KCl in 100 mL H_2O) at 37°C for 15~30 min. Subsequently, the cells were collected and 2 mL new hypotonic solution were added. Then the cell pellet was pipetted gently and mixed with 2 mL fixative (75% methanol and 25% acetic acid). Finally, the cells were resuspended in 500 μL~1 mL fixative and the spread cells were then stained with Giemsa. Karyotype images were obtained with microscope (Zeiss, Axio Imager.A2).

Western Blot

hPGCLC EBs samples were lysed and run on an SDS- PAGE gel. The primary antibodies used in this study: rabbit-anti-SOX15 (Abcam, ab55960), rabbit-anti-TFCP2L1 (R&D, AF5726), rabbit-anti-SOX11 (Abcam, ab134107), mouse-anti-ACTIN (Proteintech, 20536) and mouse-anti-TUBULIN (SUNGENE, KM9007). The secondary antibodies used in this study: anti-rabbit HRP (ZSJB-BIO, zb2301) and anti-mouse HRP (ZSJB-BIO, zb2305). The ECL kit (YEASON, 36208ES60) was used on the membrane before film exposure.

Cells Transfection

HEK293T cells were seeded at a density of 1×10^6 cells per 10 cm plate. When cell confluency was reached 90%, new culture media was replaced. HEK293T cells were transfected using 3 μL of Polyethylenimine (PEI) per 1 μg of plasmid DNA following the manufacturer's instructions.

Co-Immunoprecipitation

To detect the interaction of SOX15 or SOX17 with OCT4, HEK293T cells were transfected with 6 μg of SOX15-Flag or SOX17-Flag vector and 6 μg of OCT4-Myc vector per 10 cm plate. Cell were collected at 48 h after transfection and lysed in lysis buffer (25 mM of Tris pH 7.4, 150 mM of NaCl, 0.5% Triton X-100, 1 mM of EDTA pH 8.0) supplemented with protease cocktail (B14001, bimake). Cellular debris was cleared by centrifugation at 15,000 rpm for 10 min. For immunoprecipitation, cell lysates were incubated with anti-Flag beads (B26101, bimake) at 4°C overnight. Beads were washed three times by lysis buffer. For immunoblotting, beads in 2×sodium dodecyl sulfate (SDS) protein sample buffer were denatured at 95°C for 8 min and then were resolved by electrophoresis through a 10% SDS polyacrylamide gel.

Fluorescent Activated Cell Sorting (FACS)

Day 2-8 EBs were washed in PBS and dissociated with 0.05% (before Day 4) or 0.25% trypsin (after Day 6) for 5-20 min at 37°C. Dissociated cells were resuspended in FACS solution consisted of 2% (v/v) fetal bovine serum (FBS) in PBS. Samples were stained with APC-conjugated anti-human CD326 (EpCAM) antibody (Biolegend, 324208) and BV421-conjugated anti-human/mouse CD49f (INTEGRINα6) antibody (Biolegend, 313624) for 15 min at 4°C. Then the samples were loaded on a MoFlo XDP (Beckman Coulter) for FACS. PI and Annexin V (YEASEN, 40302ES50) were used to evaluate the apoptosis states of WT and SOX15 KO hPGCLCs.

Immunofluorescence

For immunofluorescence of EBs, two or three EBs were collected in 1.5 mL tubes and fixed in 4% paraformaldehyde in PBS for 1 h. After washed twice in PBS, the samples were permeabilized and blocked in blocking solution comprise of 2% bovine serum albumin and 0.2% Triton X-100 in PBS for 30 min at room temperature and then followed by incubation with primary antibodies diluted in blocking solution overnight at 4°C. Subsequently, the samples were washed with PBS for three times, then incubated with secondary antibodies and 10 μg/mL of Hoechst in blocking solution for 1 h at room temperature. After washed three times with PBS again, the EBs were fixed on the slides which were mounted by mounting medium (Solarbio, S2100) and low-temperature agar (YESEN, 10208ES60). For immunofluorescence of hESCs, the clones were grown on circular slides and fixed in 4% paraformaldehyde in PBS for 30 min. Images were taken by confocal laser scanning microscope (Carl Zeiss LSM 880). For immunofluorescence of cryosections slides, EBs induced from WT or hESCs were fixed in 4% paraformaldehyde in PBS for 1 h, then washed twice in PBS and incubated with 30% sucrose for 1 h at 4°C. Then the samples were embedded in OCT embedding matrix and stored at -80°C. Subsequently, samples were sliced into 8-μm cryosections by a cryostat (Leica, Heidelberger, Germany). Before immunofluorescence, slides with cryosections were air dried at room temperature for at least 15 min. The antibody incubation and following steps were similar to that described in immunofluorescence of EBs, The primary antibodies were listed as follows: Mouse anti-OCT4 (1:400, Santa Cruz, sc-5279), Rabbit anti-SOX2 (1:400, Abcam, Ab97959), Goat anti-Nanog (1:100, R&D Systems, AF1997), Goat anti-SOX17 (1:200, R&D Systems, AF1924), Rabbit anti-TFAP2C (1:400, Santa Cruz, sc-8977), Rabbit anti-SOX15 (1:200, Abcam, ab55960), Mouse anti-GATA4 (1:200, Santa Cruz, sc-25310), Rabbit anti-JUN (1:200, Abcam, ab32137), Rabbit anti-ETV5 (1:200, Proteintech, 13011-1-AP), Rabbit anti-Ki67 (1:400, Abcam, ab15580), Rabbit anti-PARP1 (1:400, Abcam, ab32064).

Quantitative PCR (q-PCR)

Total RNA was extracted from cells using Trizol (Invitrogen, 15596026) according to the manufacturer's instructions. cDNA was synthesized using HiScript QRT SuperMix for qPCR (Vazyme, R123-01). Quantitative PCR was performed using 2×PCR Master Mix (GenStar,

A301-10) and the expression level of genes-of-interest was normalized to the expression of *GAPDH* according to a previous study (Irie et al., 2015). The primer sequences used in this study are listed in Table S7. Error bars are mean ± SD from three independent experiments.

CUT&Tag

In order to study the distribution of SOX15 in hPGCLC, we used NovoNGS ® CUT&Tag 2.0 High-Sensitivity Kit (N259-YH01, Novoprotein) to capture SOX15-binding sites. The experimental process was performed according to the manufacturer's instructions. In brief, 1×10^5 day 4 SOX15-3 × Flag-p2A-EGFP hPGCLC were prepared and immobilized on concanavalin A beads. Beads are incubated with a Flag primary antibody (F1804, Sigma), followed by incubation with a secondary antibody anti-Mouse IgG (ab6708, Abcam). Beads were washed and incubated with pA-Tn5. Tn5 was activated by addition of $Mg2^+$ and incubated at 37°C for 1 h. Reactions were stopped by the addition of 10 μL 0.5M EDTA, 3 μL 10% SDS and 2.5 μL 20 mg/mL Proteinase K to each sample. DNA was extracted with phenol-chloroform and constructed CUT&Tag library according to the manufacturer's instructions. Library was quantified by Equalbit dsDNA HS Assay Kit (Vazyme, EQ111-01) using Qubit™ 4 Fluorometer (Invitrogen, Q33226). Libraries were subjected to paired-end 150 bp sequencing on NovaSeq platform at Novogene.

ATAC-seq Library Generation

ATAC-seq was performed using True Prep DNA Library Prep Kit V2 for Illumina (Vazyme, TD501). Cells were collected in PBS (2% BSA) and spun at 500 g at 4°C for 10 min. The pellet was resuspended in 50 μl lysis buffer and incubated at 4°C for 15 min and spun at 500 g at 4°C for 5 min. The supernatants were removed by carefully pipetting away from the pellets. For the transposition reaction, 10 μl 5×TTBL buffer, 5 μl TTE Mix V50 were combined and added to each pellet up to 50 μl. The samples were incubated at 37°C for 30 min followed by immediate purification using Beckman Beads. The PCR was set up in a 50 μl reaction volume using 24 μl of transposed DNA, 10 μl of 5×TAB, 5 μl PPM and 5 μl P5 and P7 primers in TruePrep Index Kit V2 for Illumina (Vazyme, TD202). PCR parameters were: 72°C for 5 min, 98°C for 30 s

and 15 cycles of 98°C for 10 s, 60°C for 30 s and 72°C for 30 s. The libraries were purified using QIAGEN MinElute PCR purification kit (QIAGEN, Cat#28004) followed by Agencourt AMPure XP beads (Beckman Coulter, A63880). Library fragments ranging from 200 to 700 bp were enriched and the final elution volume was 30 μl. Libraries were sequenced using pair-end 150 bp sequencing on an Illumina Hiseq XTEN platform at Novogene.

RNA Isolation and Library Generation

In order to construct the RNA libraries, total RNA was extracted using TRIzol™ Reagent (Invitrogen, 15596026). Total RNA (500-1000 ng) was reverse transcribed and amplified into cDNA using NEBNext Ultra™ II Directional RNA Library Prep Kit for Illumina (NEB, E7760L). RNA-seq libraries were generated with fragmented cDNA using KAPA Hyper Prep Kit (KAPABIOSYSTEMS, KK8505). Libraries were quantified by Equalbit dsDNA HS Assay Kit (Vazyme, EQ111-01) using Qubit™ 4 Fluorometer (Invitrogen, Q33226). Libraries were subjected to paired-end 150 bp sequencing on Illumina Hiseq XTEN platform at Novogene.

RNA-seq Data Analysis

The human transcriptome index was generated using the reference genome hg38 with Ensembl version 95 and aligned to hg38 transcriptome using RSEM integrated bowtie2 (Li and Dewey, 2011). Gene counts were calculated using RSEM and normalized for GC content using EDAseq. Low expressed gene were discarded by cutoff (>= 50). Differentially expressed genes were identified using DESeq2 (Love et al., 2014). The gene intersections were performed using R-package (VennDiagram). GO analysis was performed using the webtool Metascape (www.metascape.org).

ATAC-seq and CUT&Tag Data Analysis

ATAC-seq and CUT&Tag data were processed using similar data processing procedures. In brief, the total reads were trimmed using bbduk and trimmomatic, with the length cut-off 35 bp and aligned to hg38 using bowtie with the options (--very-sensitive --end-to-end). Then low-quality reads were removed using samtools with the option (-q 35). The mitochondrial sequences were removed using grep. Biological replicates were merged. MACS2 was used to

call narrow peaks with options (-g hs -f BAMPE -B --call-summits). Bigwigs were generated using bedtools and bedGraphToBigWig. Bedtools was used to calculate the genome coverage score of bam files on macs peaks. The genome coverage score was normalized to library size and the PCA was plotted. Deeptools and EA-seq were used to generate genome coverage heatmaps. The findMotifsGenome.pl program in Homer was used to find specific motifs. To define the open and closed regions, we used an approach from a previous study (Li et al., 2017) with some modifications. In brief, after obtaining all the ATAC-seq peaks by macs2, we merged the peaks of all samples as a superset of all peaks. Then we used glbase3 python package (Hutchins et al., 2014) to calculate the RPKM of normalized bigwig files of each sample on the superset of all peaks. After a series of threshold filtering, we set 16 as the threshold value to annotate open/closed regions. If the RPKM of sample is below this value, it is annotated as ‘closed’, otherwise it is annotated as ‘open’.

Supplemental References

Chen, D., Liu, W., Zimmerman, J., Pastor, W.A., Kim, R., Hosohama, L., Ho, J., Aslanyan, M., Gell, J.J., Jacobsen, S.E*., et al.* (2018). The TFAP2C-Regulated OCT4 Naive Enhancer Is Involved in Human Germline Formation. Cell Rep *25*, 3591-3602 e3595.

Chen, D., Sun, N., Hou, L., Kim, R., Faith, J., Aslanyan, M., Tao, Y., Zheng, Y., Fu, J., Liu, W*., et al.* (2019). Human Primordial Germ Cells Are Specified from Lineage-Primed Progenitors. Cell Rep *29*, 4568-4582 e4565.

Hutchins, A.P., Jauch, R., Dyla, M., and Miranda-Saavedra, D. (2014). glbase: a framework for combining, analyzing and displaying heterogeneous genomic and high-throughput sequencing data. Cell Regen *3*, 1.

Irie, N., Weinberger, L., Tang, W.W., 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.

Kanai-Azuma, M., Kanai, Y., Gad, J.M., Tajima, Y., Taya, C., Kurohmaru, M., Sanai, Y., Yonekawa, H., Yazaki, K., Tam, P.P*., et al.* (2002). Depletion of definitive gut endoderm in Sox17-null mutant mice. Development *129*, 2367-2379.

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 e515.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, D., Liu, J., Yang, X., Zhou, C., Guo, J., Wu, C., Qin, Y., Guo, L., He, J., Yu, S*., et al.* (2017). Chromatin Accessibility Dynamics during iPSC Reprogramming. Cell Stem Cell *21*, 819-833 e816.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Magnusdottir, E., Dietmann, S., Murakami, K., Gunesdogan, 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.

Mitsunaga, S., Odajima, J., Yawata, S., Shioda, K., Owa, C., Isselbacher, K.J., Hanna, J.H., and Shioda, T. (2017). Relevance of iPSC-derived human PGC-like cells at the surface of embryoid bodies to prechemotaxis migrating PGCs. Proc Natl Acad Sci U S A *114*, E9913-E9922.

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.

Tang, W.W., 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.