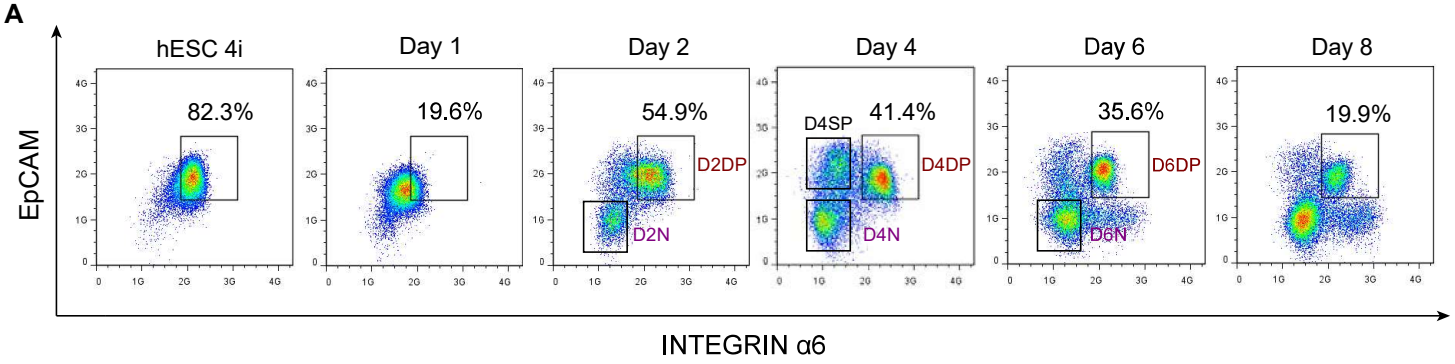
Stem Cell Reports, Volume *16*

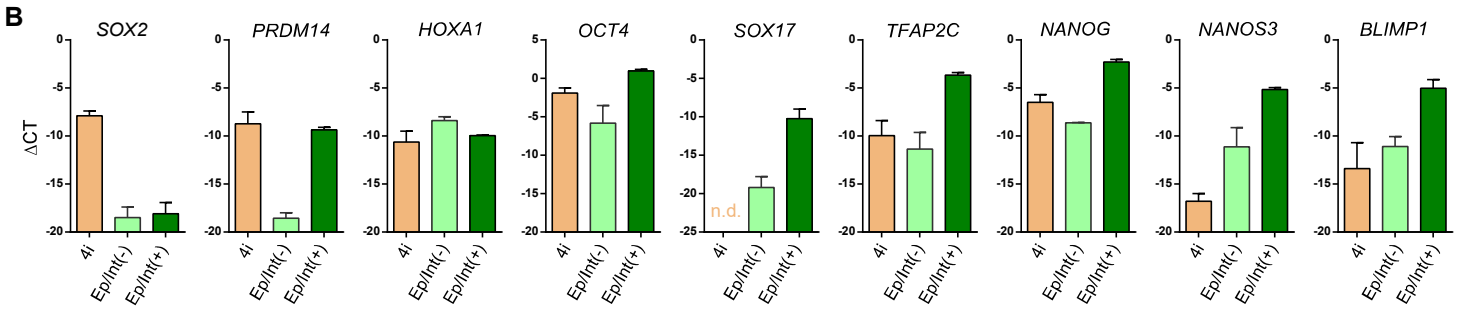
Supplemental Information

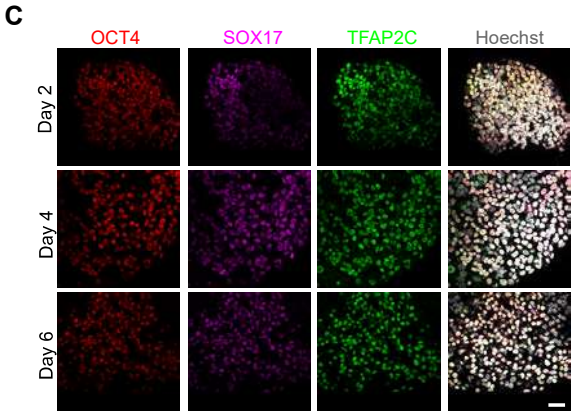
The chromatin accessibility landscape reveals distinct transcriptional regulation in the induction of human primordial germ cell-like cells from pluripotent stem cells

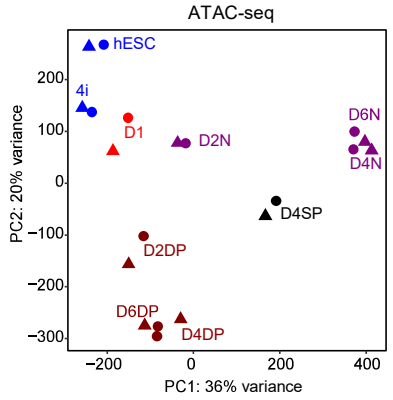
Xiaoman Wang, Veeramohan Veerapandian, Xinyan Yang, Ke Song, Xiaoheng Xu, Manman Cui, Weiyan Yuan, Yaping Huang, Xinyu Xia, Zhaokai Yao, Cong Wan, Fang Luo, Xiuling Song, Xiaoru Wang, Yi Zheng, Andrew Paul Hutchins, Ralf Jauch, Meiyan Liang, Chenhong Wang, Zhaoting Liu, Gang Chang, and Xiao-Yang Zhao

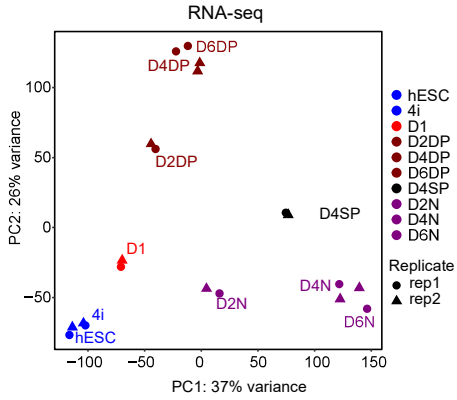
Figure S1

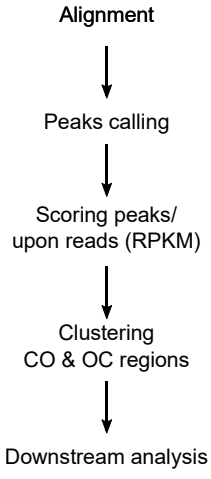


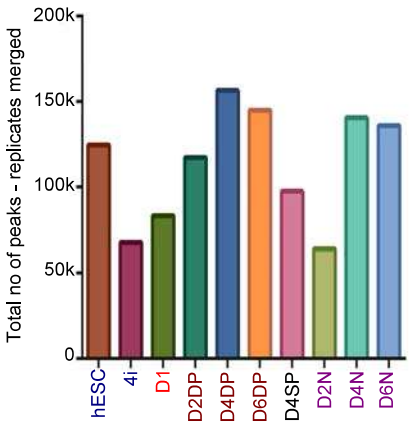


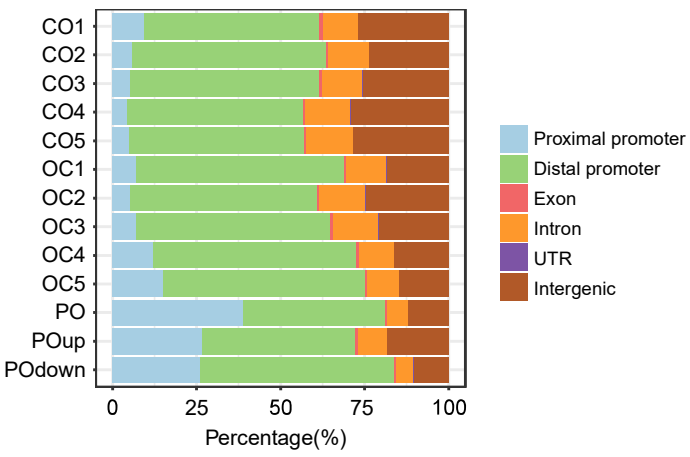












I

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

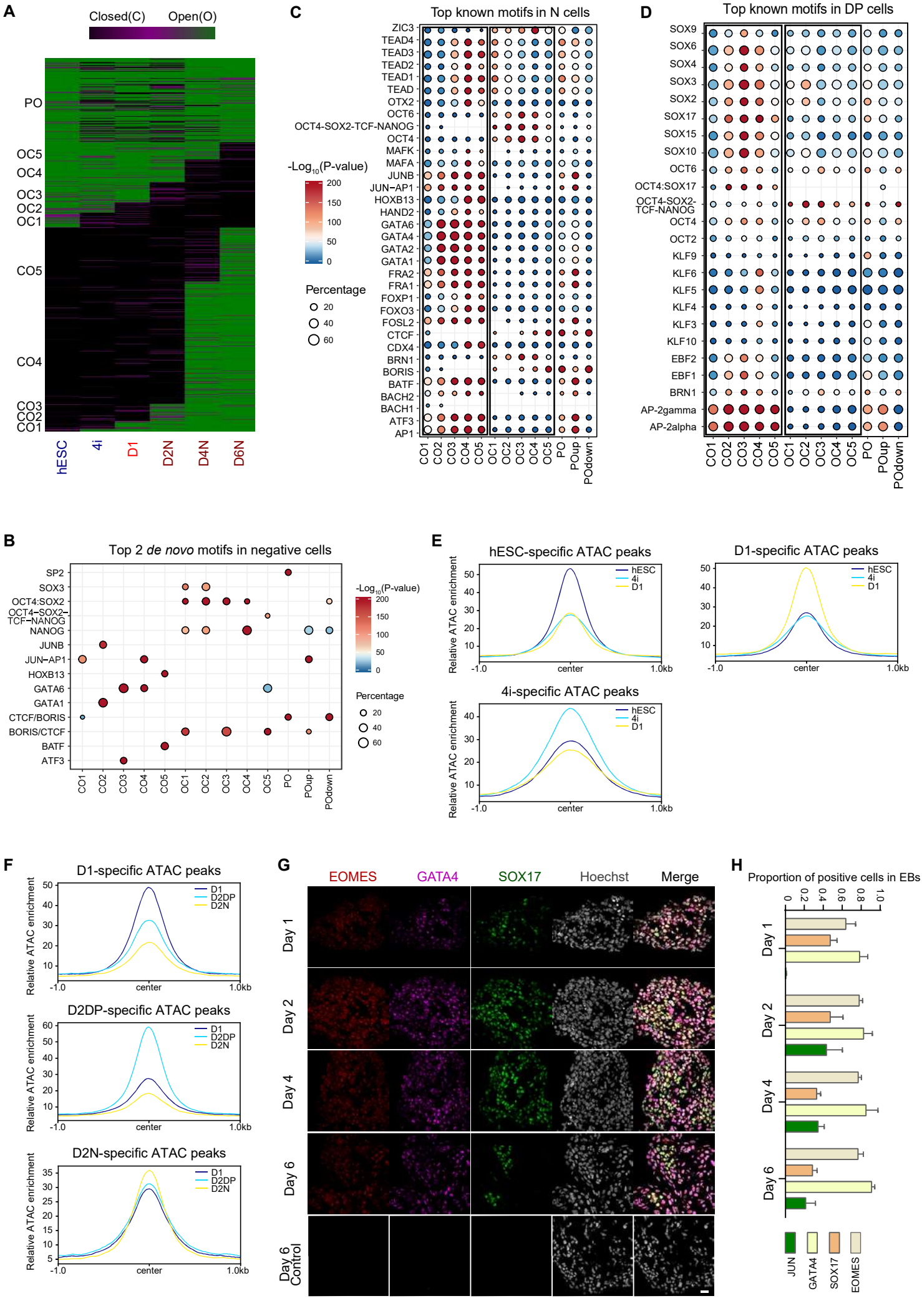

A
Closed(C)
Open(O)
PO
OC5
OC4
OC3
OC2
OC1
CO5
CO4
CO3
CO2
CO1
hESC
4i
D1
D2N
D4N
D6N
C
Top known motifs in N cells
ZIC3
TEAD4
TEAD3
TEAD2
TEAD1
TEAD
OTX2
OCT6
OCT4-SOX2-TCF-NANOG
OCT4
MAFK
MAFA
JUNB
JUN-AP1
HOXB13
HAND2
GATA6
GATA4
GATA2
GATA1
FRA2
FRA1
FOXP1
FOXO3
FOSL2
CTCF
CDX4
BRN1
BORIS
BATF
BACH2
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

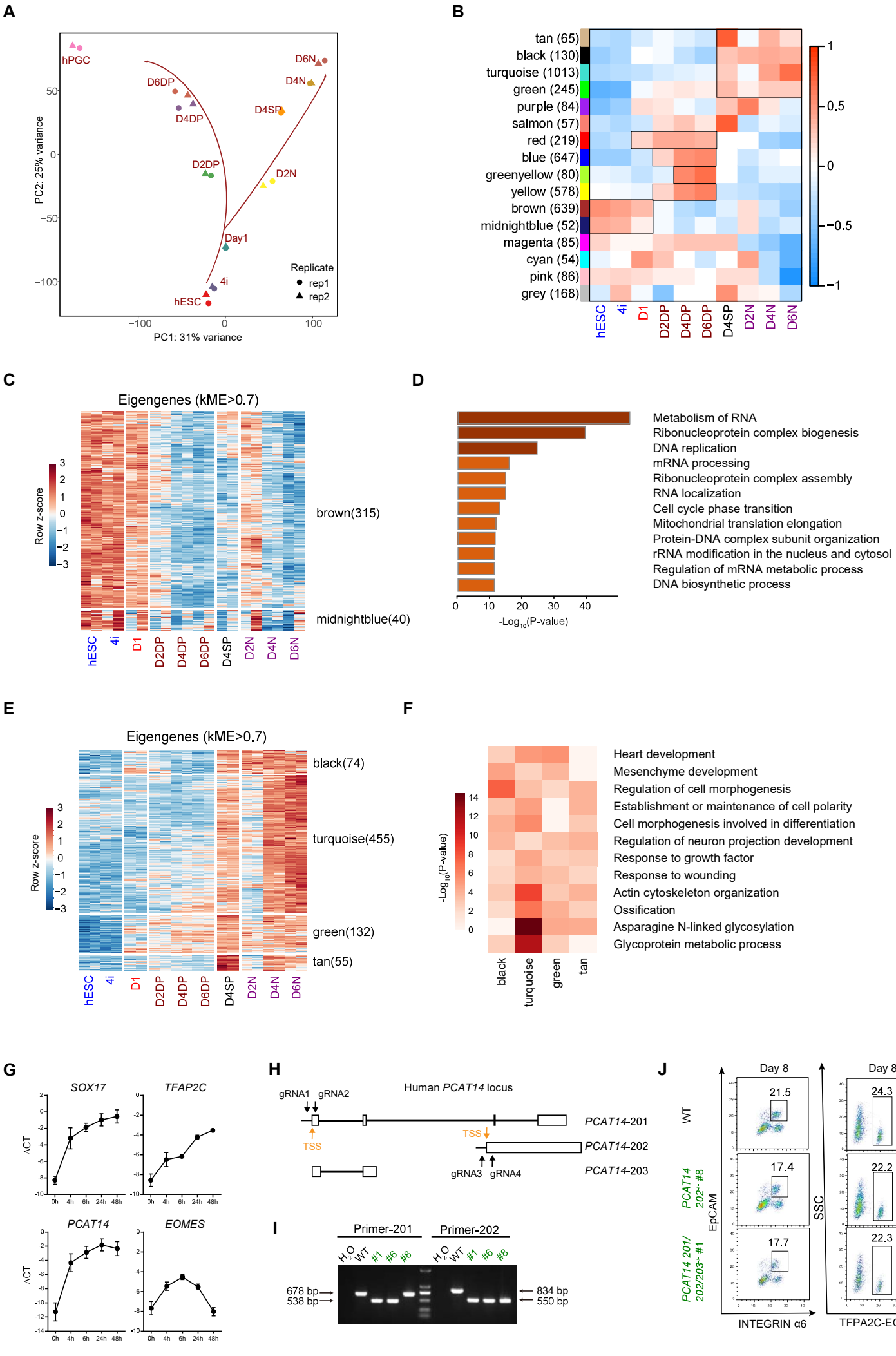

A
hPGC
D6DP
D6N
D4N
D4DP
D4SP
D2DP
D2N
Day1
4i
hESC
Replicate
rep1
rep2
PC2: 25% variance
PC1: 31% variance
B
tan (65)
black (130)
turquoise (1013)
green (245)
purple (84)
salmon (57)
red (219)
blue (647)
greenyellow (80)
yellow (578)
brown (639)
midnightblue (52)
magenta (85)
cyan (54)
pink (86)
grey (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

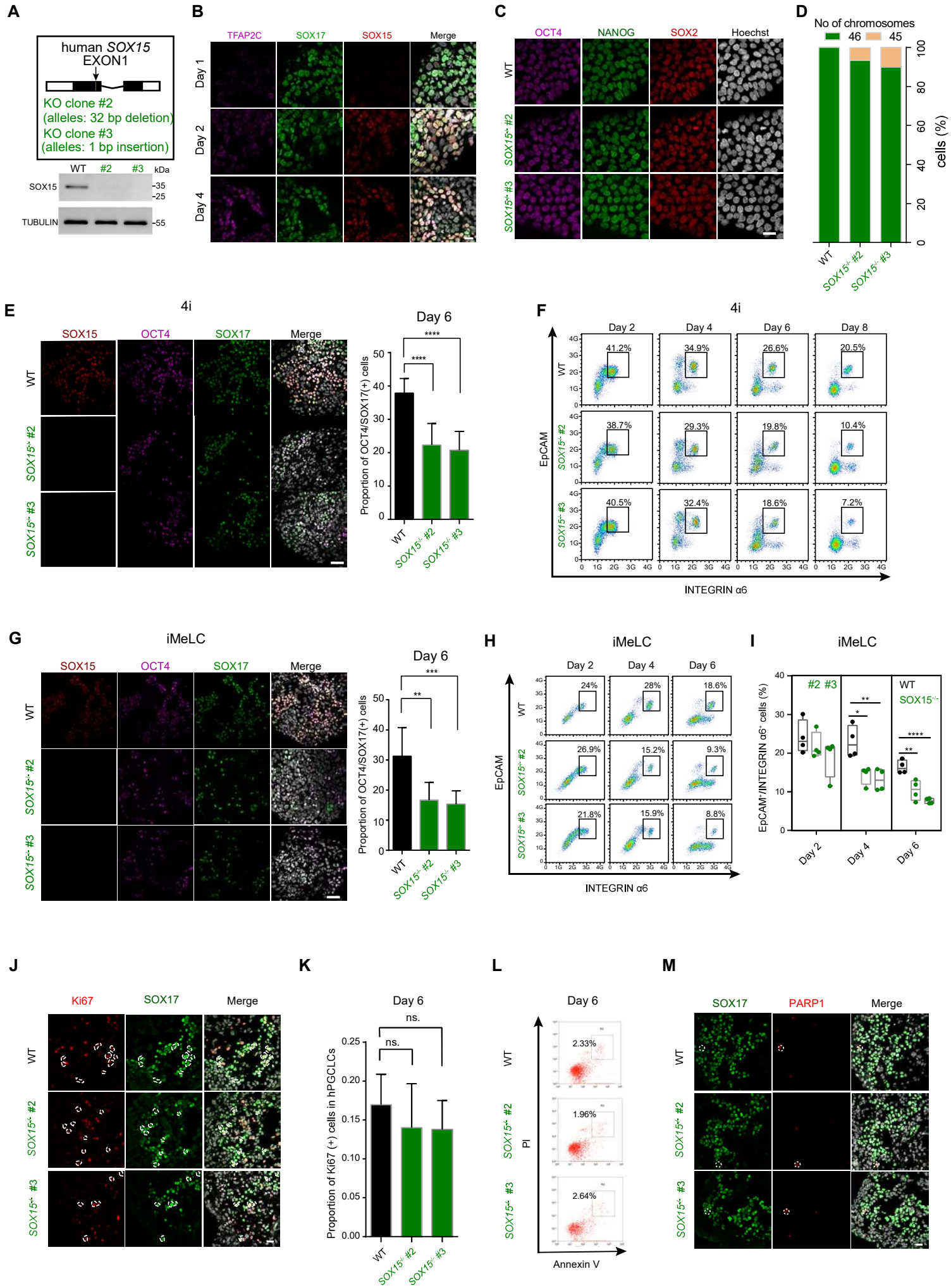

A
human SOX15 EXON1
KO clone #2 (alleles: 32 bp deletion)
KO clone #3 (alleles: 1 bp insertion)
WT #2 #3 kDa
SOX15
TUBULIN
-35
-25
-55
B
TFAP2C
SOX17
SOX15
Merge
Day 1
Day 2
Day 4
C
OCT4
NANOG
SOX2
Hoechst
WT
SOX15-/- #2
SOX15-/- #3
D
No of chromosomes
46
45
cells (%)
E
4i
SOX15
OCT4
SOX17
Merge
Day 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

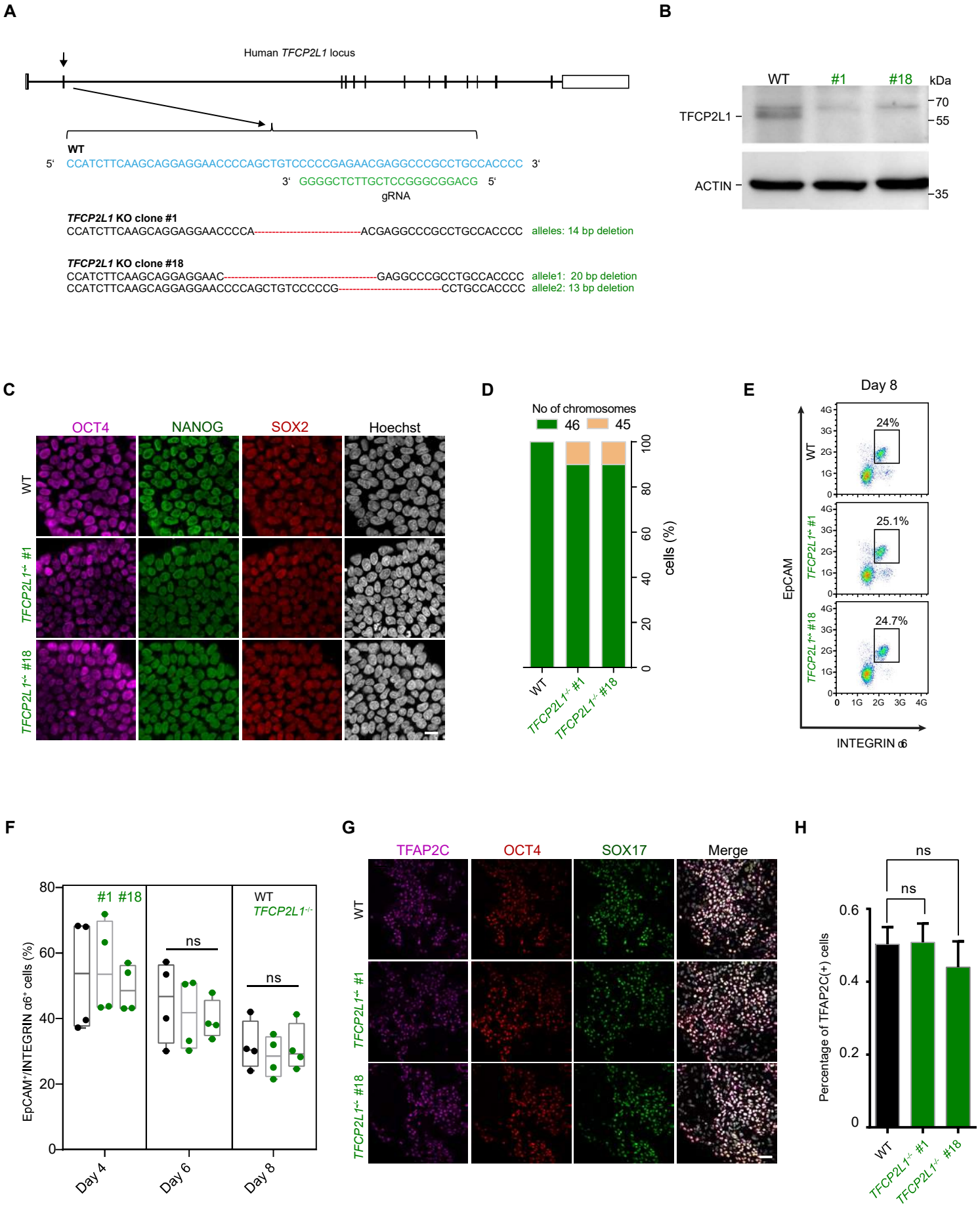

A
Human TFCP2L1 locus
WT
5' CCATCTTCAAGCAGGAGGAACCCCAGCTGTCCCCCGAGAACGAGGCCCGCCTGCCACCCC 3'
3' GGGGCTCTTGCTCCGGGCGGACG 5'
gRNA
TFCP2L1 KO clone #1
CCATCTTCAAGCAGGAGGAACCCCA----------ACGAGGCCCGCCTGCCACCCC alleles: 14 bp deletion
TFCP2L1 KO clone #18
CCATCTTCAAGCAGGAGGAAC----------GAGGCCCGCCTGCCACCCC allele1: 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

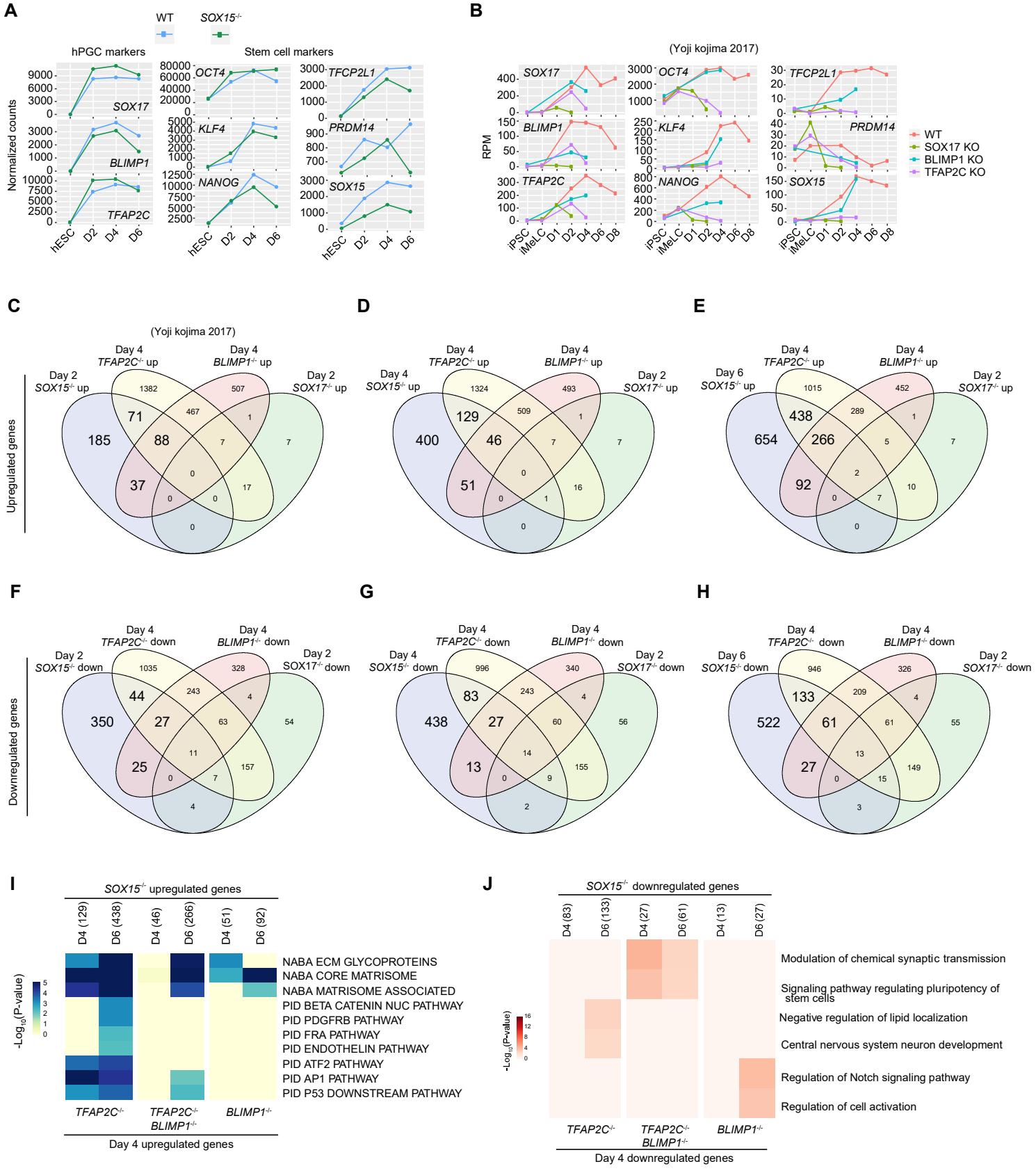

A
WT
SOX15-/-
hPGC markers
Stem cell markers
Normalized counts
SOX17
BLIMP1
TFAP2C
OCT4
KLF4
NANOG
TFCP2L1
PRDM14
SOX15
hESC
D2
D4
D6
B
(Yoji kojima 2017)
RPM
iPSC
iMeLC
D1
D8
WT
SOX17 KO
BLIMP1 KO
TFAP2C KO
C
(Yoji kojima 2017)
Upregulated genes
Day 2 SOX15-/- up
Day 4 TFAP2C-/- up
Day 4 BLIMP1-/- up
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

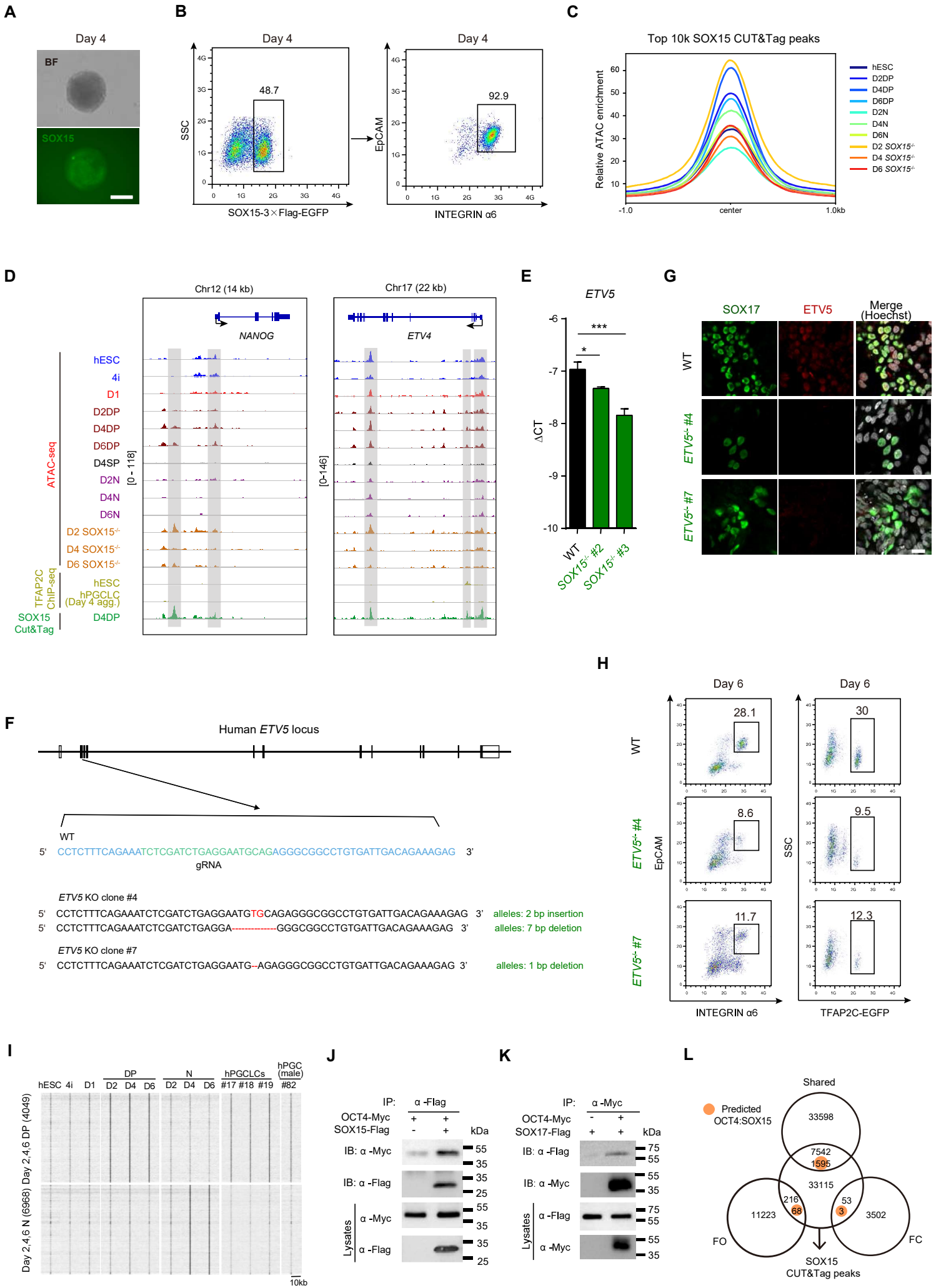

A
Day 4
BF
SOX15
B
Day 4
48.7
SSC
SOX15-3×Flag-EGFP
Day 4
92.9
EpCAM
INTEGRIN α6
C
Top 10k SOX15 CUT&Tag peaks
Relative ATAC enrichment
-1.0
center
1.0kb
hESC
D2DP
D4DP
D6DP
D2N
D4N
D6N
D2 SOX15-/-
D4 SOX15-/-
D6 SOX15-/-
D
Chr12 (14 kb)
NANOG
Chr17 (22 kb)
ETV4
ATAC-seq
hESC
4i
D1
D2DP
D4DP
D6DP
D4SP
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.