iScience, Volume *26*

Supplemental information

TET1 facilitates specification of early human lineages including germ cells

Fei-Man Hsu, Qiu Ya Wu, Emily B. Fabyanic, Alex Wei, Hao Wu, and Amander T. Clark

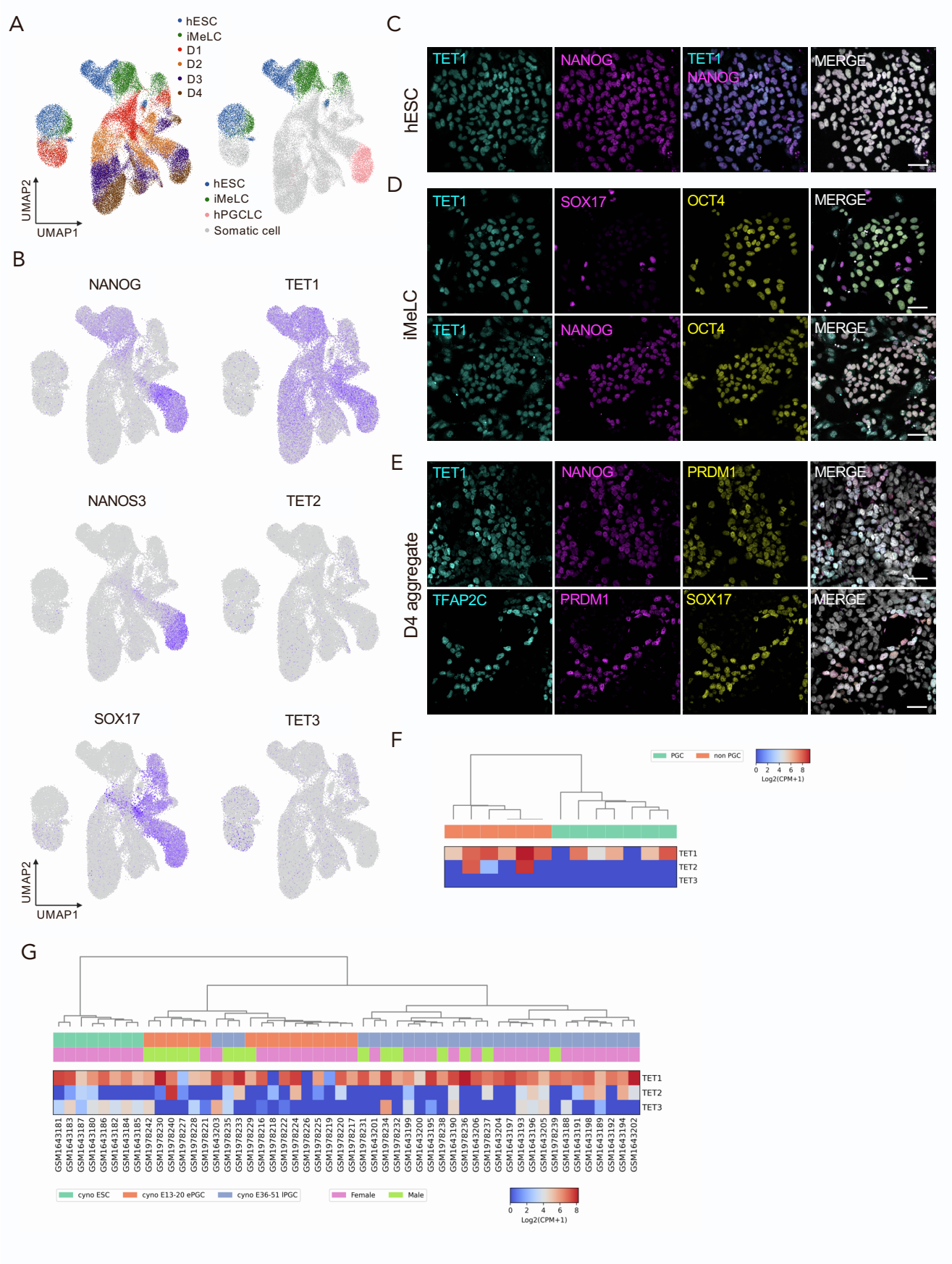
A
hESC
iMeLC
D1
D2
D3
D4
UMAP2
UMAP1
hESC
iMeLC
hPGCLC
Somatic cell
B
NANOG
TET1
NANOS3
TET2
SOX17
TET3
UMAP2
UMAP1
C
hESC
TET1
NANOG
TET1
NANOG
MERGE
D
iMeLC
TET1
SOX17
OCT4
MERGE
TET1
NANOG
OCT4
MERGE
E
D4 aggregate
TET1
NANOG
PRDM1
MERGE
TFAP2C
PRDM1
SOX17
MERGE
F
PGC
non PGC
0 2 4 6 8
Log2(CPM+1)
TET1
TET2
TET3
G
TET1
TET2
TET3
GSM1643181
GSM1643183
GSM1643187
GSM1643180
GSM1643186
GSM1643182
GSM1643184
GSM1643185
GSM1978242
GSM1978230
GSM1978240
GSM1978227
GSM1978228
GSM1978221
GSM1643203
GSM1978235
GSM1978233
GSM1978229
GSM1978216
GSM1978218
GSM1978222
GSM1978224
GSM1978226
GSM1978225
GSM1978219
GSM1978220
GSM1978217
GSM1643201
GSM1978231
GSM1978234
GSM1978232
GSM1643199
GSM1643200
GSM1643195
GSM1978238
GSM1643190
GSM1978236
GSM1643206
GSM1978237
GSM1643204
GSM1643197
GSM1643193
GSM1643196
GSM1643205
GSM1978239
GSM1643188
GSM1643191
GSM1643198
GSM1643189
GSM1643192
GSM1643194
GSM1643202
cyno ESC
cyno E13-20 ePGC
cyno E36-51 IPGC
Female
Male
0 2 4 6 8
Log2(CPM+1)

Figure S1. TET1 expression profiles during hPGCLC induction in UCLA2, related to Figure 1.

(A) UMAPs of scRNA-seq data from [1] showing UCLA2 hESCs, iMeLCs and day (D)1 to D4 of aggregate differentiation (n= 48,820 cells displayed)
(B) UMAPs from (A) displaying dynamic expression of the pluripotency gene (NANOG), the hPGC/hPGCLC-specific gene (NANOS3), the hPGC/hPGCLC and endoderm gene (SOX17) as well as the TET family mRNAs (*TET1, TET2* and *TET3*).
(C) Representative IF images of TET1 in undifferentiated NANOG+ hESCs. Scale bar, 40 μm.
(D) Representative IF images of TET1 in NANOG+ but not SOX17+ iMeLCs. Scale bar, 40 μm.
(E) Representative IF images of TET1 in NANOG+/PRDM1+ hPGCLCs at D4 of aggregate differentiation. PRDM1+ hPGCLCs are also positive for TFAP2C and SOX17. Scale bar, 40 μm.
(F) TET1 is expressed in CS9 hPGC.
(G) TET1 is expressed in cyPGC.

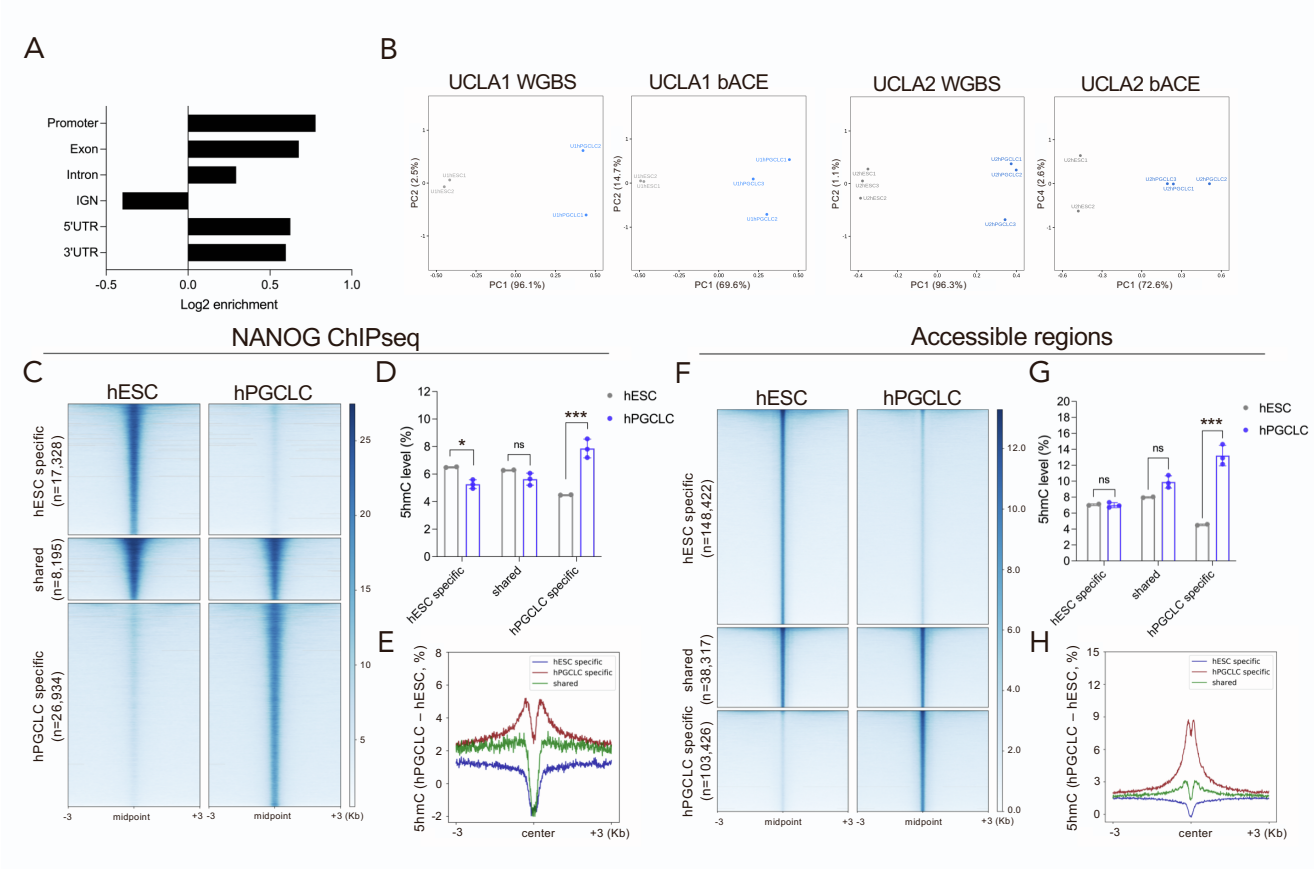

Figure S2. 5hmC is enriched at hPGCLC-specific TF binding sites and accessible regions, related to Figure 2.

(A) DhMR is enriched in gene-centric regions.
(B) PCA of WGBS and bACEseq in hESCs and hPGCLCs.
(C) Heatmap of NANOG binding regions in hESCs and hPGCLCs (hPGCLC-specific binding refers to specific binding in hPGCLCs relative to hESCs).
(D) Percentage 5hmC at genomic sites in (C) reveals that 5hmC is especially enriched at hPGCLC-specific NANOG binding regions in hPGCLCs. Data are represented as mean +/- SEM. The Student's *t*-test was performed to determine the significances with *$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$.
(E) Metaplot of Δ5hmC from regions in (C).
(F) Heatmap of accessible chromatin in hESCs and hPGCLCs (hPGCLC-specific refers to increased accessibility in hPGCLCs when compared to hESCs).
(G) Percentage 5hmC at genomic sites in (F) reveals that 5hmC is enriched at hPGCLC-specific accessible regions in hPGCLCs. Data are represented as mean +/- SEM. The Student's *t*-test was performed to determine the significances with *$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$.
(H) Metaplot of Δ5hmC from regions in (F).

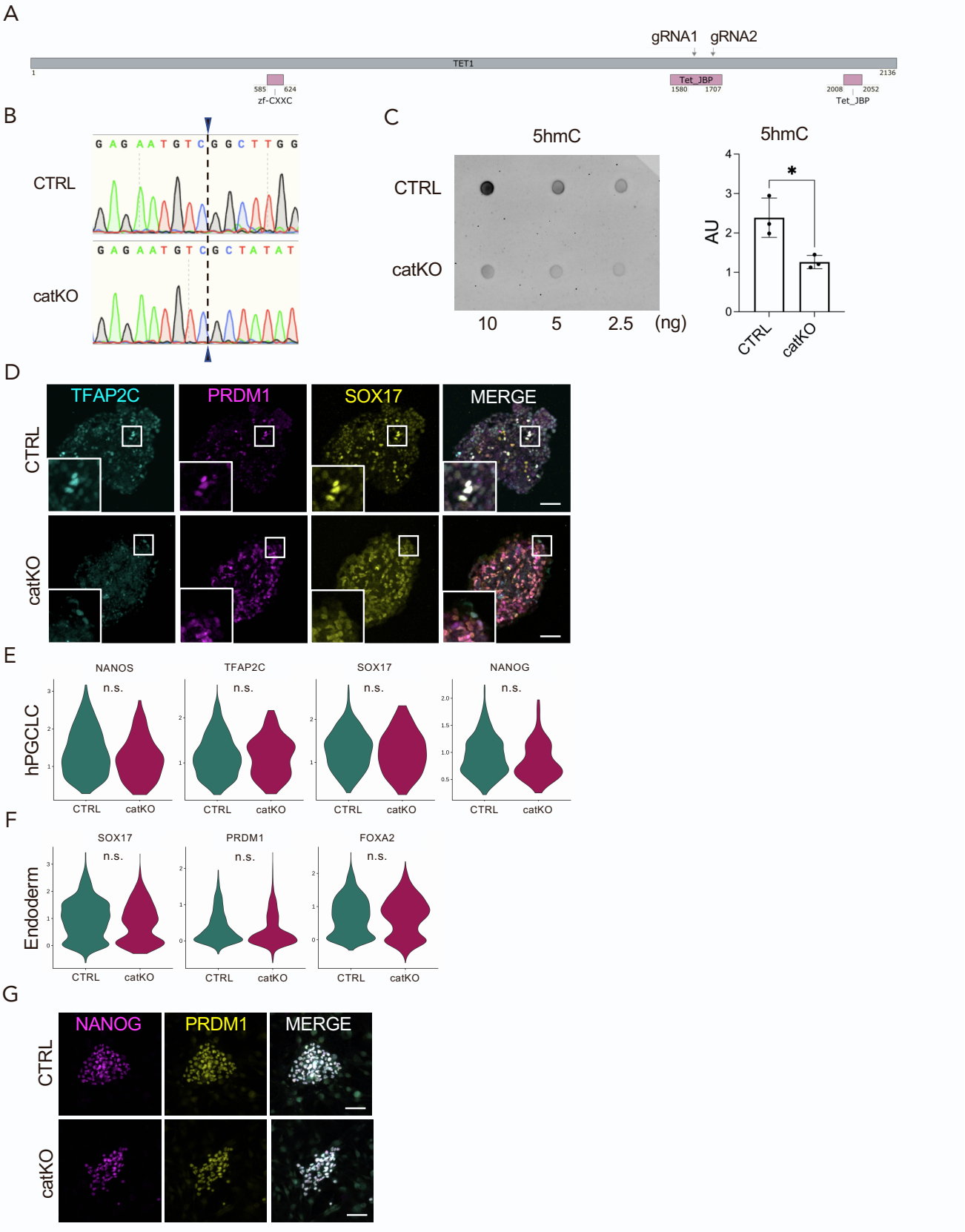
A
gRNA1
gRNA2
TET1
1
2136
585
624
zf-CXXC
Tet_JBP
1580
1707
2008
2052
Tet_JBP
B
CTRL
catKO
C
5hmC
CTRL
catKO
10
5
2.5
(ng)
5hmC
AU
CTRL
catKO
D
TFAP2C
PRDM1
SOX17
MERGE
CTRL
catKO
E
NANOS
TFAP2C
SOX17
NANOG
n.s.
hPGCLC
CTRL
catKO
F
SOX17
PRDM1
FOXA2
n.s.
Endoderm
CTRL
catKO
G
NANOG
PRDM1
MERGE
CTRL
catKO

Figure S3. Details of CRISPR/Cas9 knockout and hPGCLC and Endoderm-like cell identity with differentiation, related to Figure 3.

(A) A pair of guide RNAs were designed to target the TET1 catalytic domain.
(B) Representative Sanger sequencing results from CTRL and catKO.
(C) 5hmC dot blot in CTRL and catKO hESC genomic DNA. Data are represented as mean +/- SEM. The Student's *t*-test was performed to determine the significances with $*p < 0.05$, $**p < 0.01$ and $***p < 0.001$.
(D) Representation IF images showing endoderm bias begins at D2 in catKO aggregate. Scale bar, 80μm.
(E) TET1 catKO mutant hPGCLCs show comparable gene expression. The Student's *t*-test was performed to determine the significances with $*p < 0.05$.
(F) TET1 catKO mutant endoderm-like cells show comparable gene expression. The Student's *t*-test was performed to determine the significances with $*p < 0.05$.
(G) ITGA6/EPCAM hPGCLCs were isolated at D4 from CTRL and catKO aggregates and cultured in extended culture for 21 days (D4C21). Cultured hPGCLCs from CTRL and catKO express NANOG and PRDM1. Scale bar, 50μm.

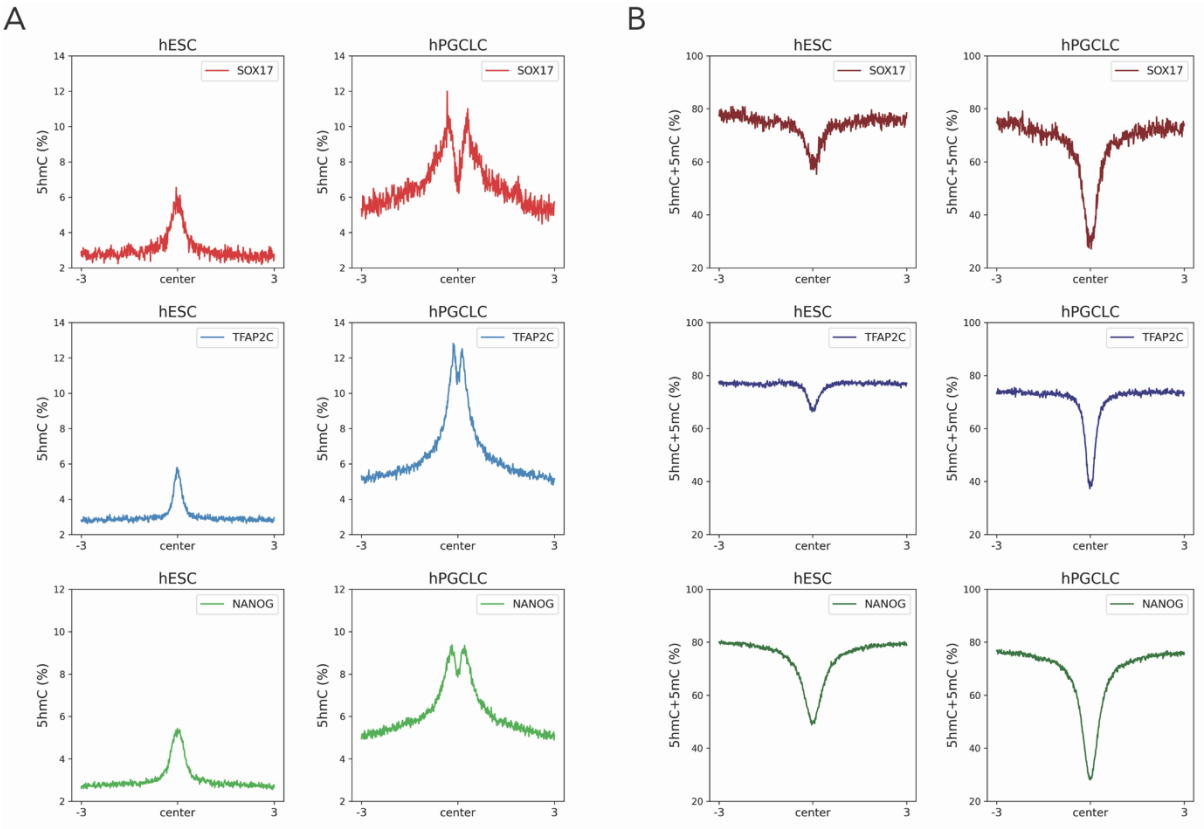

Figure S4. TET1 is less active in hESC at hPGCLC-specific TFBS, related to Figure 4.

(A) hPGCLC-specific TFBSs are slightly enriched in 5hmC in hESC, and gain even more 5hmC after hPGCLC induction.
(B) hPGCLC-specific TFBSs starts losing 5mC in hESC and are even depleted of 5mC after hPGCLC induction.

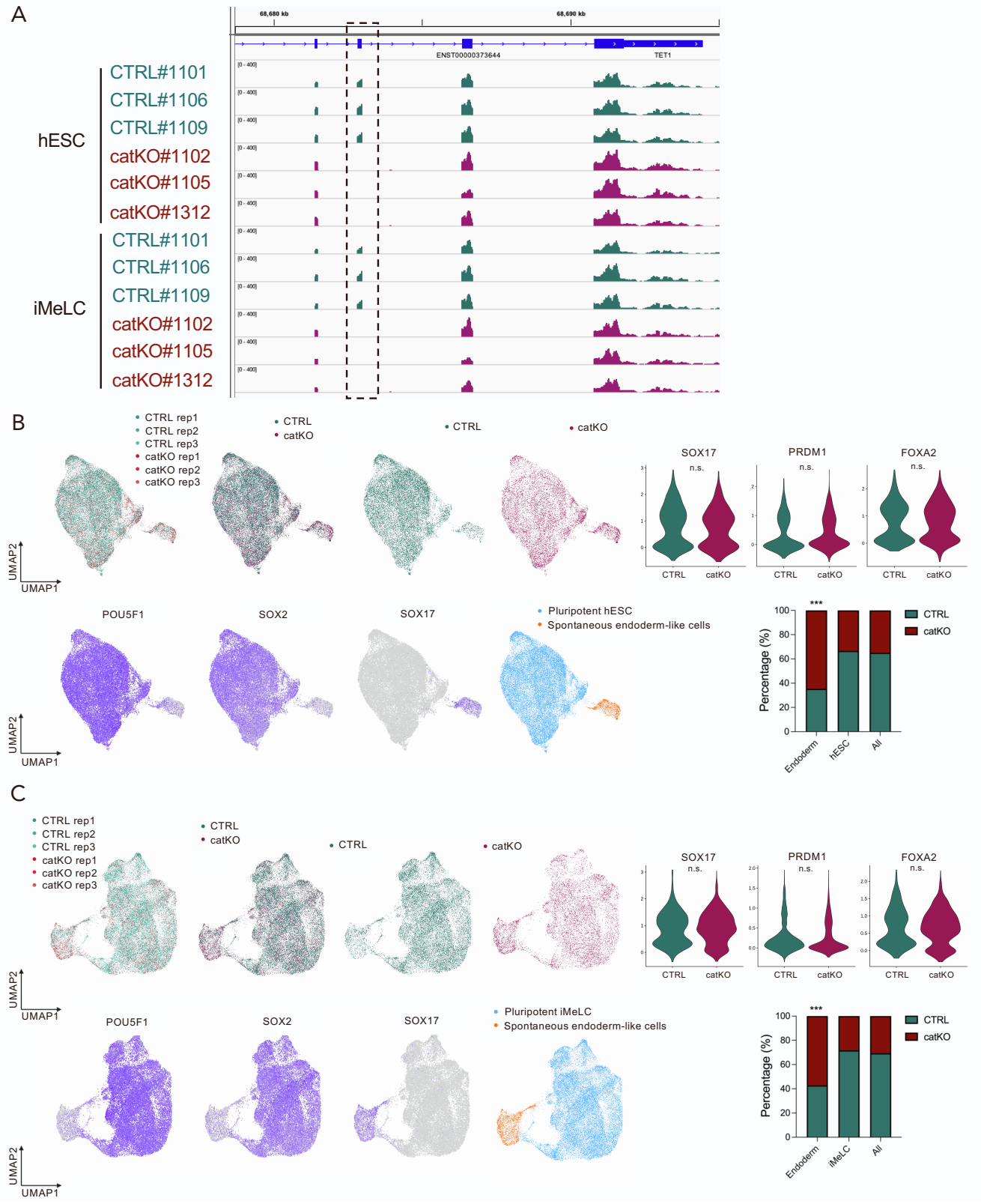
A
hESC
CTRL#1101
CTRL#1106
CTRL#1109
catKO#1102
catKO#1105
catKO#1312
iMeLC
CTRL#1101
CTRL#1106
CTRL#1109
catKO#1102
catKO#1105
catKO#1312
68,680 kb
68,690 kb
ENST00000373644
TET1
B
CTRL rep1
CTRL rep2
CTRL rep3
catKO rep1
catKO rep2
catKO rep3
CTRL
catKO
UMAP1
UMAP2
SOX17
PRDM1
FOXA2
n.s.
CTRL
catKO
POU5F1
SOX2
SOX17
Pluripotent hESC
Spontaneous endoderm-like cells
***
Percentage (%)
Endoderm
hESC
All
C
Pluripotent iMeLC
Spontaneous endoderm-like cells
iMeLC

Figure S5. TET1 catalytic activity restrains spontaneous endoderm differentiation in hESCs and iMeLCs, related to Figure 5.

(A) Genome browser track of bulk RNAseq showing deletion in TET1 catalytic domain.
(B) UMAP of scRNAseq data from CTRL and catKO hESCs generated using 10X Genomics. (n=3 independent replicates of CTRL and TET1 catKO), n= 29,433 number of cells displayed. Top shows overlap of CTRL and TET1 catKO hESCs in the UMAP. Bottom shows Pluripotency genes *POU5F1* and *SOX2*. The violin plots show the *SOX17, PRDM1* and *FOXA2* expression of the spontaneous endoderm-like cells cluster displayed on the UMAP, and the Student's *t*-test was performed to determine the significances with $*p < 0.05$. The stacked bar chart shows the composition of endoderm-like cells is significantly increase in catKO, and Fisher's exact test was used to determine the significance with $*p < 0.05$, $**p < 0.01$ and $***p < 0.001$.
(C) UMAP of scRNAseq from CTRL and catKO iMeLC generated with 10X Genomics. (n=3 independent replicates of CTRL and TET1 catKO), n= 28,025 number of cells displayed. Top shows overlay and deconvolution of CTRL and TET1 catKO iMeLCs in the UMAP. Bottom shows Pluripotency genes *POU5F1* and *SOX2*. The violin plots show the endoderm gene *SOX17, PRDM1* and *FOXA2* expression in the spontaneous endoderm-like cells cluster displayed on the UMAP, and the Student's *t*-test was performed to determine the significances with $*p < 0.05$. The stacked bar chart shows the composition of endoderm-like cells is significantly increased in catKO, and Fisher's exact test was used to determine the significance with $*p < 0.05$, $**p < 0.01$ and $***p < 0.001$.

Table S1. WGBS and bACEseq libraries construct in this study, related to Figure 2.

Library type	Cell line	Sample	#Total reads	Mappability (%)
WGBS	UCLA1	hESC replicate2	664,005,942	97.2
WGBS	UCLA1	hESC replicate3	457,638,140	97.2
WGBS	UCLA1	D4hPGCLC replicate2	318,024,343	97.1
WGBS	UCLA1	D4hPGCLC replicate3	546,319,972	96.9
WGBS	UCLA2	hESC replicate1	810,926,832	97.2
WGBS	UCLA2	hESC replicate2	431,308,172	97.2
WGBS	UCLA2	hESC replicate3	612,300,195	97.2
WGBS	UCLA2	D4hPGCLC replicate1	795,694,334	96.6
WGBS	UCLA2	D4hPGCLC replicate2	588,486,763	97.1
WGBS	UCLA2	D4hPGCLC replicate3	502,871,088	97.0
bACEseq	UCLA1	hESC replicate1	1,328,431,585	96.3
bACEseq	UCLA1	hESC replicate3	839,571,111	96.7
bACEseq	UCLA1	D4hPGCLC replicate1	1,680,496,307	94.5
bACEseq	UCLA1	D4hPGCLC replicate2	717,046,904	96.1
bACEseq	UCLA1	D4hPGCLC replicate3	824,697,411	96.0
bACEseq	UCLA2	hESC replicate2	805,636,528	96.3
bACEseq	UCLA2	hESC replicate3	613,582,866	96.6
bACEseq	UCLA2	D4hPGCLC replicate1	686,857,892	96.1
bACEseq	UCLA2	D4hPGCLC replicate2	722,444,940	96.1
bACEseq	UCLA2	D4hPGCLC replicate3	630,833,981	96.8

Table S2. Details of 10X libraries created for this study, related to Figure 3.

Sample	#Total reads	Mappability (%)	#Valid cells	median UMIs
CTRL#1101 hESC	171,283,234	72.8	7587	6010
CTRL#1106 hESC	173,545,059	70.5	6834	6387
CTRL#1109 hESC	128,274,507	72.2	4754	6536
catKO#1102 hESC	165,061,174	58.2	2948	11596
catKO#1105 hESC	174,609,315	48.2	3437	8396
catKO#1312 hESC	132,432,829	46.4	3873	5713
CTRL#1101 iMeLC	172,479,910	76.3	2126	27789
CTRL#1106 iMeLC	171,565,256	65.5	8574	6393
CTRL#1109 iMeLC	177,802,407	73.8	8680	6599
catKO#1102 iMeLC	170,486,170	65.8	3547	14311
catKO#1105 iMeLC	174,911,961	42.9	2167	15279
catKO#1312 iMeLC	140,329,600	63.0	2895	13682
CTRL#1101 D4aggr	192,139,067	85.4	5313	3948
CTRL#1106 D4aggr	225,893,933	85.3	5347	4114
CTRL#1109 D4aggr	166,441,276	83.2	9675	2323
catKO#1102 D4aggr	169,078,959	82.4	6341	3079
catKO#1105 D4aggr	188,952,867	83.1	6579	3377
catKO#1312 D4aggr	208,218,684	83.6	4964	4405

Table S3. Details of bulk RNAseq libraries created for this study, related to Figure 5.

Library type	Cell line	Sample	#Total read pairs	Mappability (%)
RNAseq	UCLA1	CTRL#1101 hESC	65,488,888	60.6
RNAseq	UCLA1	CTRL#1106 hESC	78,326,513	56.4
RNAseq	UCLA1	CTRL#1109 hESC	75,213,438	63.2
RNAseq	UCLA1	catKO#1102 hESC	46,951,012	68.3
RNAseq	UCLA1	catKO#1105 hESC	72,163,591	65.9
RNAseq	UCLA1	catKO#1312 hESC	69,858,296	65.3
RNAseq	UCLA1	CTRL#1101 iMeLC	129,947,695	55.8
RNAseq	UCLA1	CTRL#1106 iMeLC	115,548,571	61.7
RNAseq	UCLA1	CTRL#1109 iMeLC	134,637,954	61.5
RNAseq	UCLA1	catKO#1102 iMeLC	65,759,606	57.3
RNAseq	UCLA1	catKO#1105 iMeLC	143,230,121	49.8
RNAseq	UCLA1	catKO#1312 iMeLC	111,129,317	52.1

Table S4. Genomic datasets re-analyzed in this study, related to Figure 1, 2 and 4.

Library type	Protein	Dataset	Reference
ChIPseq	H3K27ac	GSE69646	Ji *et al.*, 2016 [2]
ChIPseq	H3K27ac	GSE140021	Chen *et al.*, 2019 [1]
ChIPseq	H3K4me3	GSE159654	Tang *et al.*, 2022 [3]
ChIPseq	TET1-V5	GSE169207	Dixon *et al.*, 2021 [4]
ChIPseq	TFAP2C	GSE140021	Chen *et al.*, 2019 [1]
ChIPseq	SOX17	GSE182218	Xiang *et al.*, 2022 [5]
ChIPseq	NANOG	GSE182218	Xiang *et al.*, 2022 [5]
ChIPseq	PRDM1	GSE159654	Tang *et al.*, 2022 [3]
ATACseq	NA	GSE120648	Chen *et al.*, 2018 [6]
Single cell RNAseq	NA	GSE140021	Chen *et al.*, 2019 [1]
Single cell RNAseq	NA	GSE76267	Sasaki *et al.*, 2016 [7]
Single cell RNAseq	NA	GSE74767	Nakamura *et al.*, 2016 [8]
Single cell RNAseq	NA	E-MTAB-6967	Tyser *et al.*, 2021 [9]

REFERENCES

1. 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 reports *29*, 4568-4582.e5. 10.1016/j.celrep.2019.11.083.

2. Ji, X., Dadon, D.B., Powell, B.E., Fan, Z.P., Borges-Rivera, D., Shachar, S., Weintraub, A.S., Hnisz, D., Pegoraro, G., Lee, T.I., et al. (2016). 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. Cell Stem Cell *18*, 262–275. 10.1016/j.stem.2015.11.007.

3. Tang, W.W.C., Castillo-Venzor, A., Gruhn, W.H., Kobayashi, T., Penfold, C.A., Morgan, M.D., Sun, D., Irie, N., and Surani, M.A. (2022). Sequential enhancer state remodelling defines human germline competence and specification. Nat Cell Biol, 1–13. 10.1038/s41556-022-00878-z.

4. Dixon, G., Pan, H., Yang, D., Rosen, B.P., Jashari, T., Verma, N., Pulecio, J., Caspi, I., Lee, K., Stransky, S., et al. (2021). QSER1 protects DNA methylation valleys from de novo methylation. Science *372*, eabd0875. 10.1126/science.abd0875.

5. Xiang, X., Tao, Y., DiRusso, J., Hsu, F.-M., Zhang, J., Xue, Z., Pontis, J., Trono, D., Liu, W., and Clark, A.T. (2022). Human reproduction is regulated by retrotransposons derived from ancient Hominidae-specific viral infections. Nat Commun *13*, 463. 10.1038/s41467-022-28105-1.

6. 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 reports *25*, 3591-3602.e5. 10.1016/j.celrep.2018.12.011.

7. Sasaki, K., Nakamura, T., Okamoto, I., Yabuta, Y., Iwatani, C., Tsuchiya, H., Seita, Y., Nakamura, S., Shiraki, N., Takakuwa, T., et al. (2016). The Germ Cell Fate of Cynomolgus Monkeys Is Specified in the Nascent Amnion. Developmental cell *39*, 169–185. 10.1016/j.devcel.2016.09.007.

8. Nakamura, T., Okamoto, I., Sasaki, K., Yabuta, Y., Iwatani, C., Tsuchiya, H., Seita, Y., Nakamura, S., Yamamoto, T., and Saitou, M. (2016). A developmental coordinate of pluripotency among mice, monkeys and humans. Nature *537*, 57–62. 10.1038/nature19096.

9. Tyser, R.C.V., Mahammadov, E., Nakanoh, S., Vallier, L., Scialdone, A., and Srinivas, S. (2021). Single-cell transcriptomic characterization of a gastrulating human embryo. Nature *600*, 285–289. 10.1038/s41586-021-04158-y.