ISCI, Volume *11*

Supplemental Information

Derivation of Haploid Trophoblast Stem Cells via Conversion *In Vitro*

Keli Peng, Xu Li, Congyu Wu, Yuna Wang, Jian Yu, Jinxin Zhang, Qian Gao, Wenhao Zhang, Qian Zhang, Yong Fan, Yang Yu, and Ling Shuai

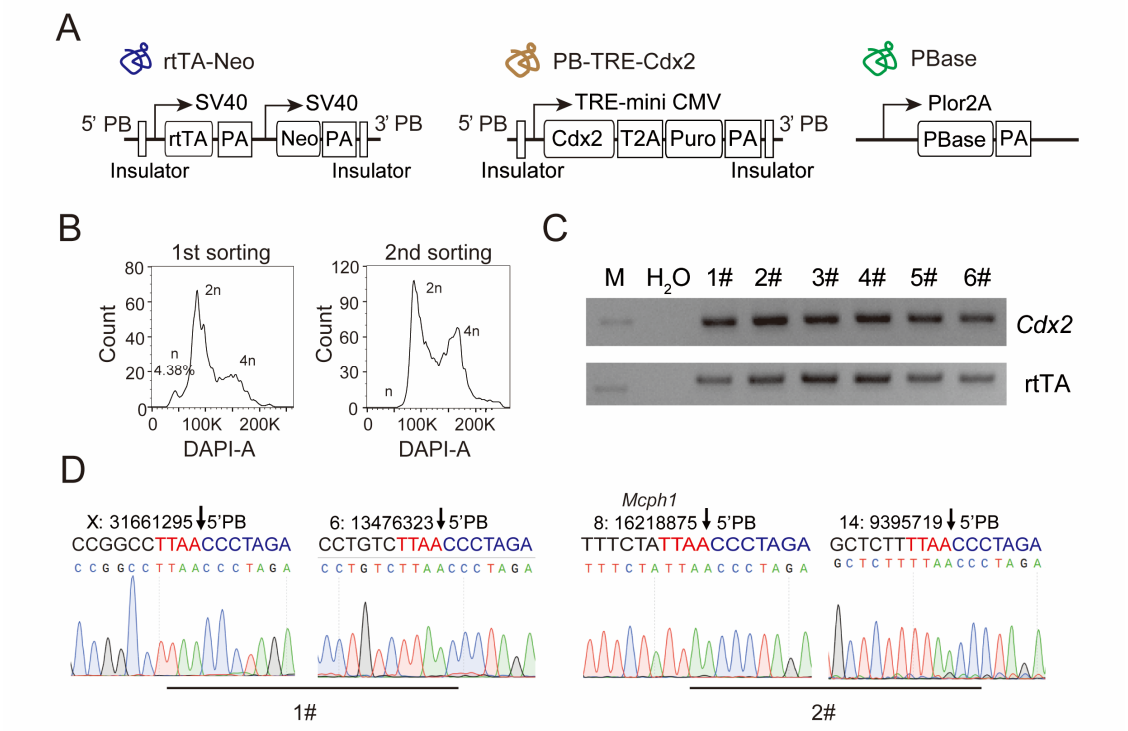

Figure S1 related to Figure 1

(A) Schematic overview of the vectors used for overexpression of *Cdx2*. There are three vectors: vector 1 has the *rtTA* and *Neo* selection genes; vector 2 has the *Cdx2* coding region and *Puromycin* resistance gene driven by TRE with a minimal CMV promoter; vector 3 is used as a PBase expression vector.

(B) DNA content analysis of the cells on Day 11 (first sorting) and Day 23 (second sorting) during conversion.

(C) Genotype of the six subclones from OE-Cdx2 haESCs with *Cdx2* and rtTA indicated.

(D) Sequencing traits of insertion sites of PB in #1 and #2.

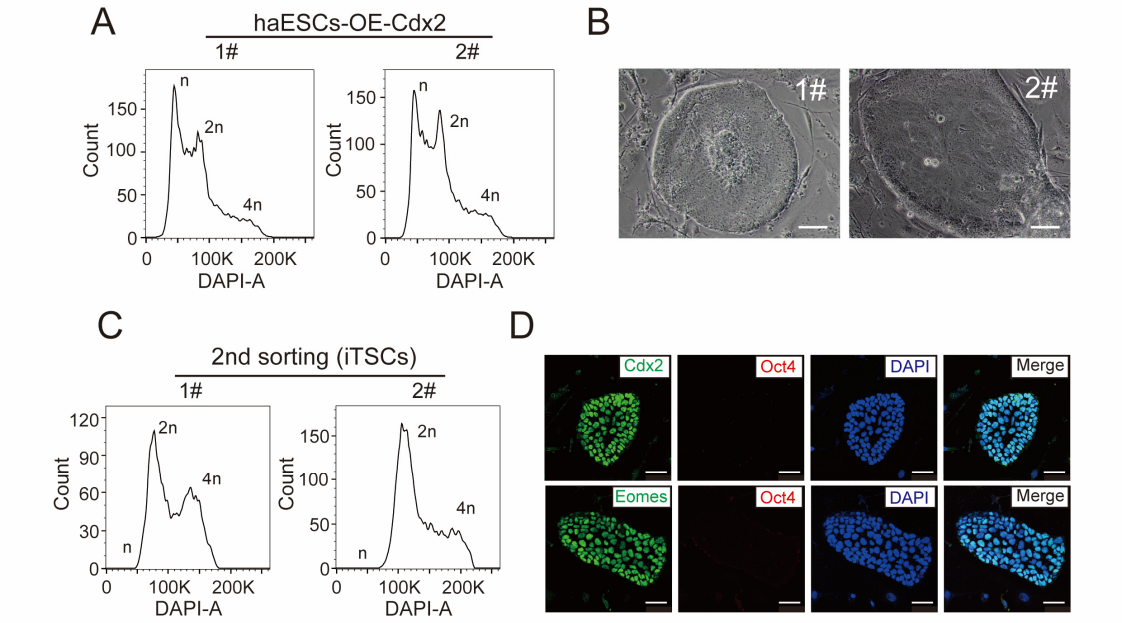

Figure S2 related to Figure 1

(A) DNA content analysis of two stable haploid transgenic OE-Cdx2 cell lines, #1 and #2.

(B) The morphology of iTSCs from subclones #1 and #2. Scale bar, 100 μm.

(C) DNA content analysis of the cells converted from #1 and #2 on Day 23 (second sorting).

(D) Immunofluorescence staining of Oct4 (TRITC channel), Eomes (FITC channel) and Cdx2 (FITC channel) in iTSCs. DNA is stained with DAPI. Scale bar, 50 μm.

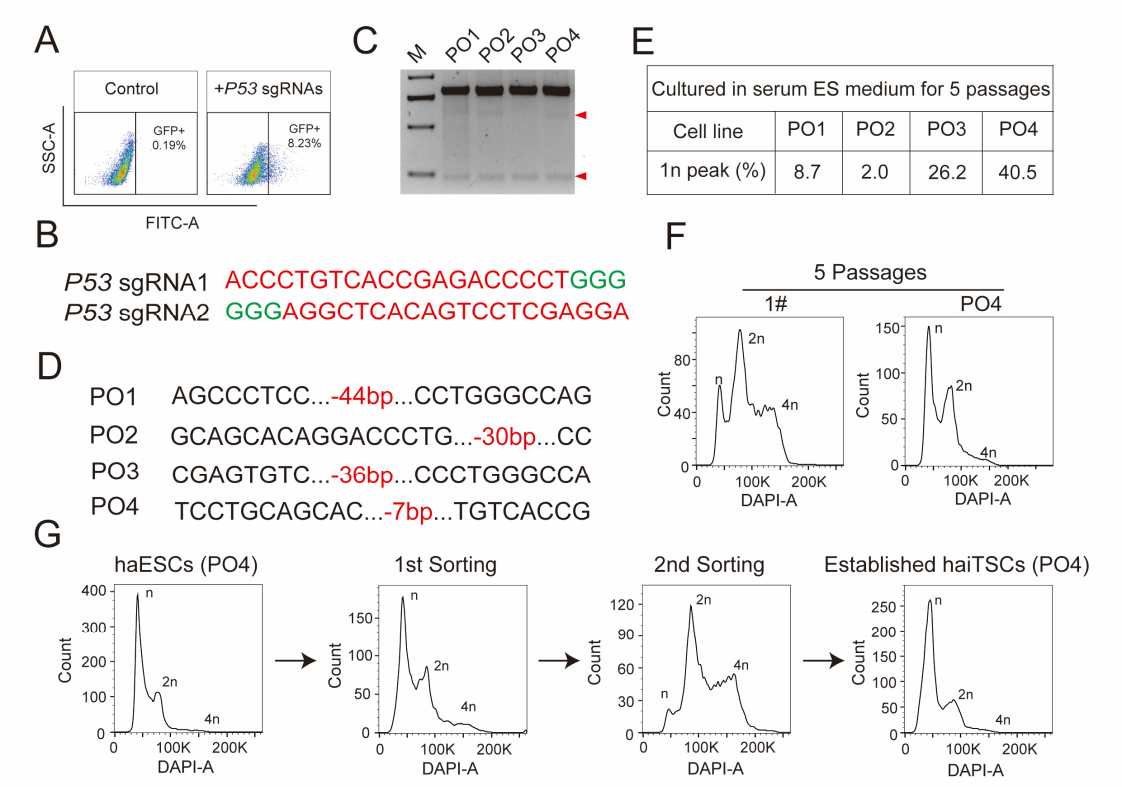

Figure S3 related to Figure 2

(A) Percentage of Cas9-GFP-positive cells in #1 by FACS 2 days after transfection.

(B) sgRNA sequences used for P53 knockout.

(C) *P53*-deleted genotypes in subclones PO1, PO2 and PO3.

(D) Percentages of the 1n peaks of four P53 knockout subclones, PO1, PO2, PO3 and PO4, cultured in serum ES medium for 5 passages.

(E) T7ENI cleavage analysis of the *P53* knockout cell lines PO1, PO2, PO3 and PO4. Cleaved products (red triangle) indicate the presence of mutations.

(F) DNA content analysis of #1 and PO4 cultured in 2i/L medium.

(G) DNA content analysis of cells during conversion on Day 0, Day 11, and Day 23 and final established haiTSCs. The percentages of the 1n (G0/G1) peak were 69.4%, 53.2%, 6.59%, and 70.4%, respectively.

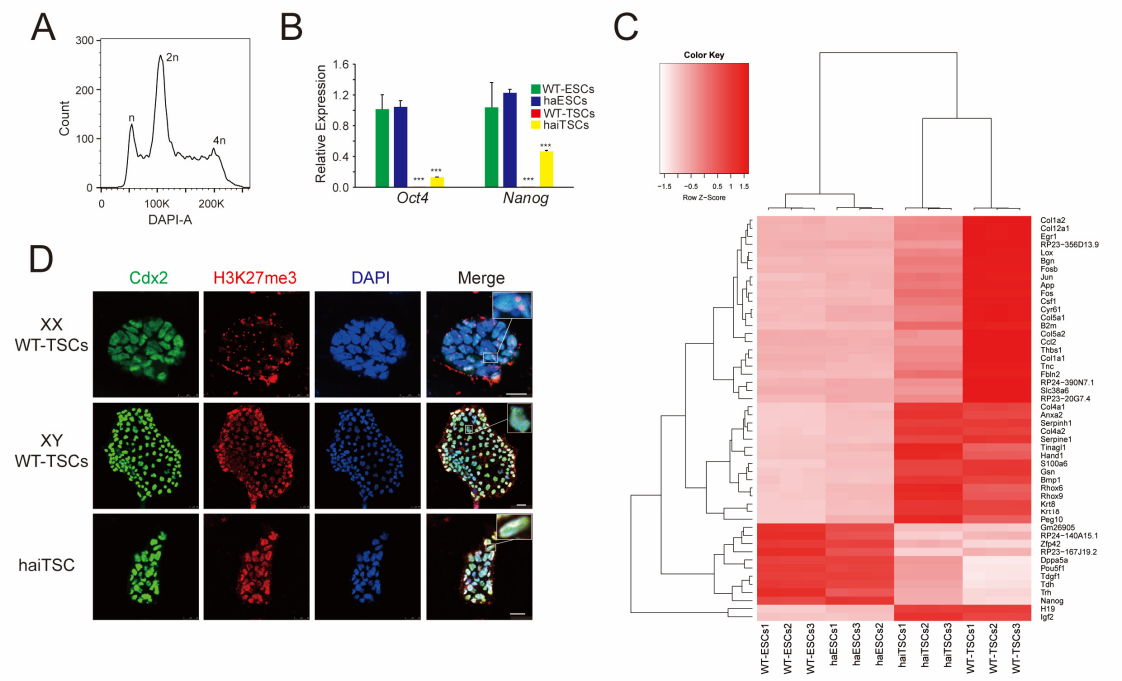

Figure S4 related to Figure 3

(A) DNA content analysis of haiTSCs after CDCP1 antibody sorting and expansion. The percentage of the 1n (G0/G1) peak was 15.3%.

(B) Expression level analysis of pluripotent marker genes (*Oct4* and *Nanog*) in haiTSCs, WT-TSCs, haESCs and WT-ESCs by qPCR. t test, ***$p < 0.001$. Data are represented as mean±SEM.

(C) Heatmap of the top 50 most variable genes across all four cell types. Data are first scaled by row and then represented using hierarchical clustering for both samples and genes.

(D) Immunofluorescence staining of Cdx2 (FITC channel) and H3K27me3 (TRITC channel) in female WT-TSCs, male WT-TSCs and haiTSCs. Female WT-TSCs (XX) and male (XY) WT-TSCs are used as controls. DNA is stained with DAPI. Scale bar, 50 μm.

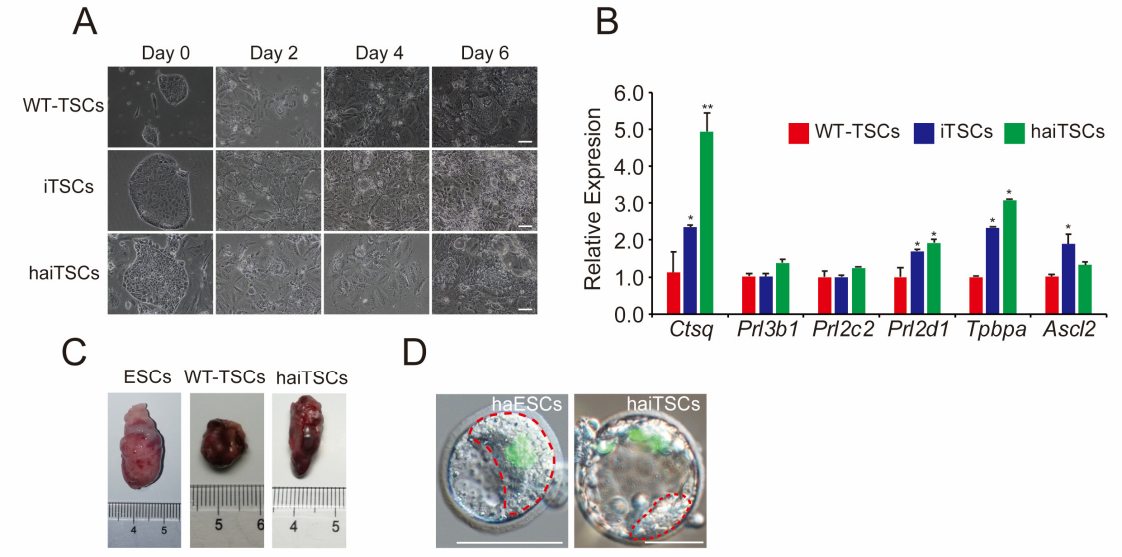

Figure S5 related to Figure 3

(A) Brightfield images of haiTSCs grown in differentiation medium (TS medium without F4H) on Day 0, Day 2, Day 4 and Day 6. Scale bar, 50 μm.

(B) The expression levels of trophoblast lineage-specific gene markers (Ctsq, Prl3b1, Prl2c2, Prl3d1, Tpbpa and Ascl2) in 4-day differentiated cells from WT-TSCs, iTSCs (diploid) and haiTSCs. t test, *p < 0.05, **p < 0.01. Data are represented as mean±SEM.

(C) Teratomas derived from ESCs and hemorrhagic lesions derived from WT-TSCs and haiTSCs.

(D) Reconstructed chimeric blastocysts, with separate contributions of GFP-haESCs (left) and GFP-haiTSCs (right). The red dashed line indicates the area of the inner cell mass. Scale bar, 40 μm.

A

NO.	Insertion Gene	Position	Chromosome	PB orientation
1	NA	Intergenic region	11	NA
2	Ptprd	Intron	4	Reverse
3	NA	Intergenic region	13	NA
4	Bach2	Intron	4	Forward
5	Csf3	Extron	11	Reverse
6	NA	Intergenic region	1	NA
7	Kif1b	Intron	4	Forward
8	Btk	Intron	X	Forward
9	Mllt3	Intron	4	Forward
10	NA	Intergenic region	X	NA
11	Gm27943	Extron	1	Reverse
12	NA	Intergenic region	4	NA
13	P4htm	Intron	17	Forward
14	Glb1l	Extron	1	Reverse
15	NA	Intergenic region	10	NA
16	Runx1	Intron	16	Reverse
17	Mamld1	Intron	X	Forward
18	Fgf14	Intron	1	Reverse
19	NA	Intergenic region	9	NA
20	Zfp236	Intron	18	Forward
21	NA	Intergenic region	8	NA
22	NA	Intergenic region	4	NA
23	NA	Intergenic region	11	NA
24	Btbd11	Intron	10	Forward
25	NA	Intergenic region	11	NA
26	NA	Intergenic region	6	NA
27	Usp7	Intron	16	Reverse
28	Cstf1	Intron	2	Reverse
29	Man1c1	Intron	4	Forward
30	P4htm	Intron	9	Forward
31	NA	Intergenic region	13	NA
32	NA	Intergenic region	4	NA
33	NA	Intergenic region	6	NA
34	NA	Intergenic region	16	NA
35	NA	Intergenic region	2	NA
36	Arntl	Intron	7	Forward
37	Asb6	Intron	2	Forward
38	NA	Intergenic region	10	Forward
39	Slco1a6	Intron	6	Reverse
40	Crtac1	Intron	19	Reverse
41	NA	Intergenic region	6	NA
42	NA	Intergenic region	18	NA
43	NA	Intergenic region	9	NA
44	NA	Intergenic region	5	NA
45	NA	Intergenic region	X	NA
46	Zmynd8	Intron	2	Forward
47	NA	Intergenic region	11	NA
48	Epb41l4a	Intron	2	Forward
49	NA	Intergenic region	1	NA
50	NA	Intergenic region	11	NA

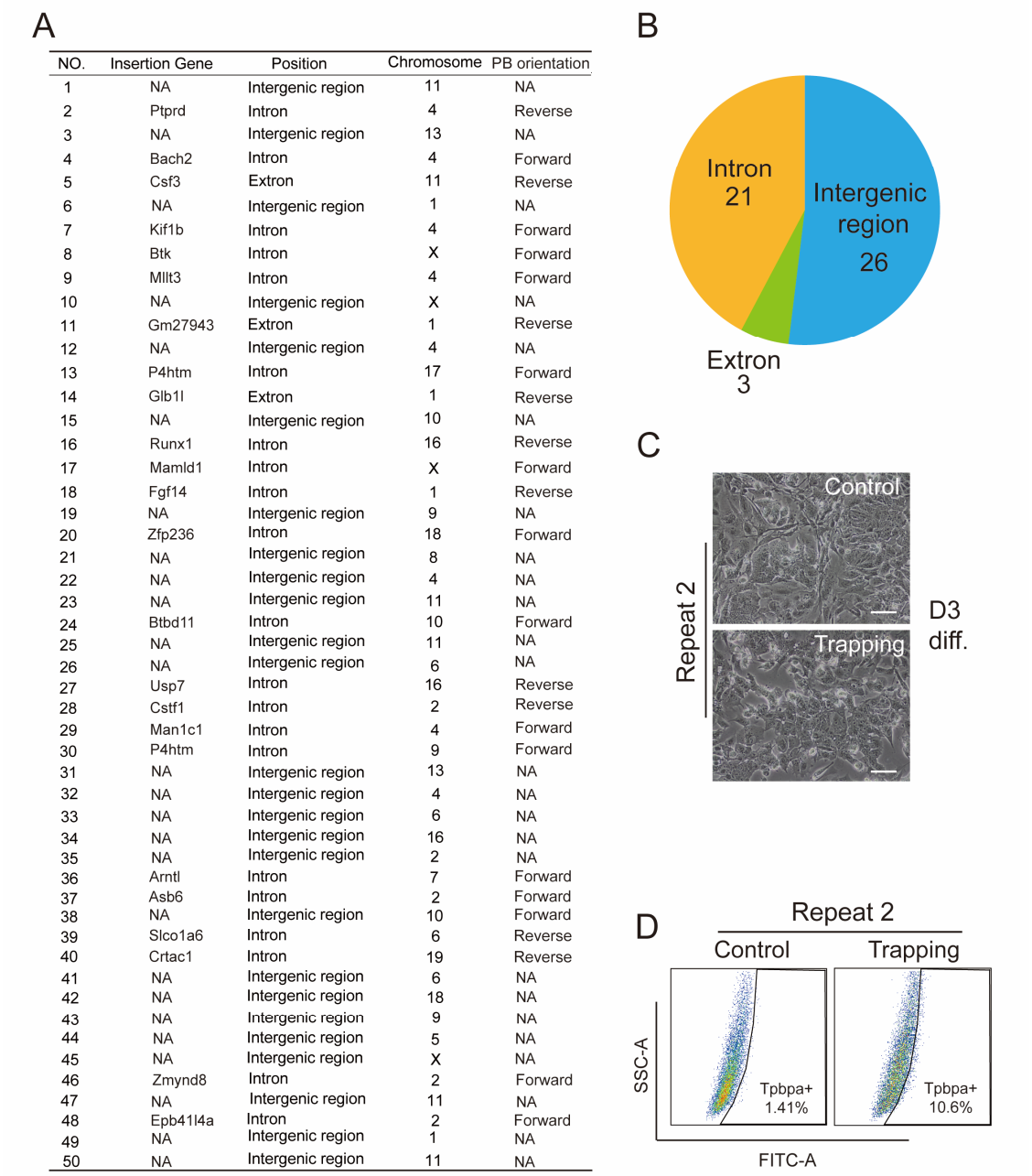

Figure S6 related to Figure 4

(A) Detected insertion sites of randomly picked subclones by Splinkerette PCR.

(B) Summary of the detected integrated sites.

(C) Brightfield images of differentiated cells from mutated (trapping) and nonmutated haiTSCs (control) in a single repeat. Scale bar, 100 μm.

(D) FACS analysis of Tpbpa-positive cells in differentiated cells from the trapping and control groups in the same repeat.

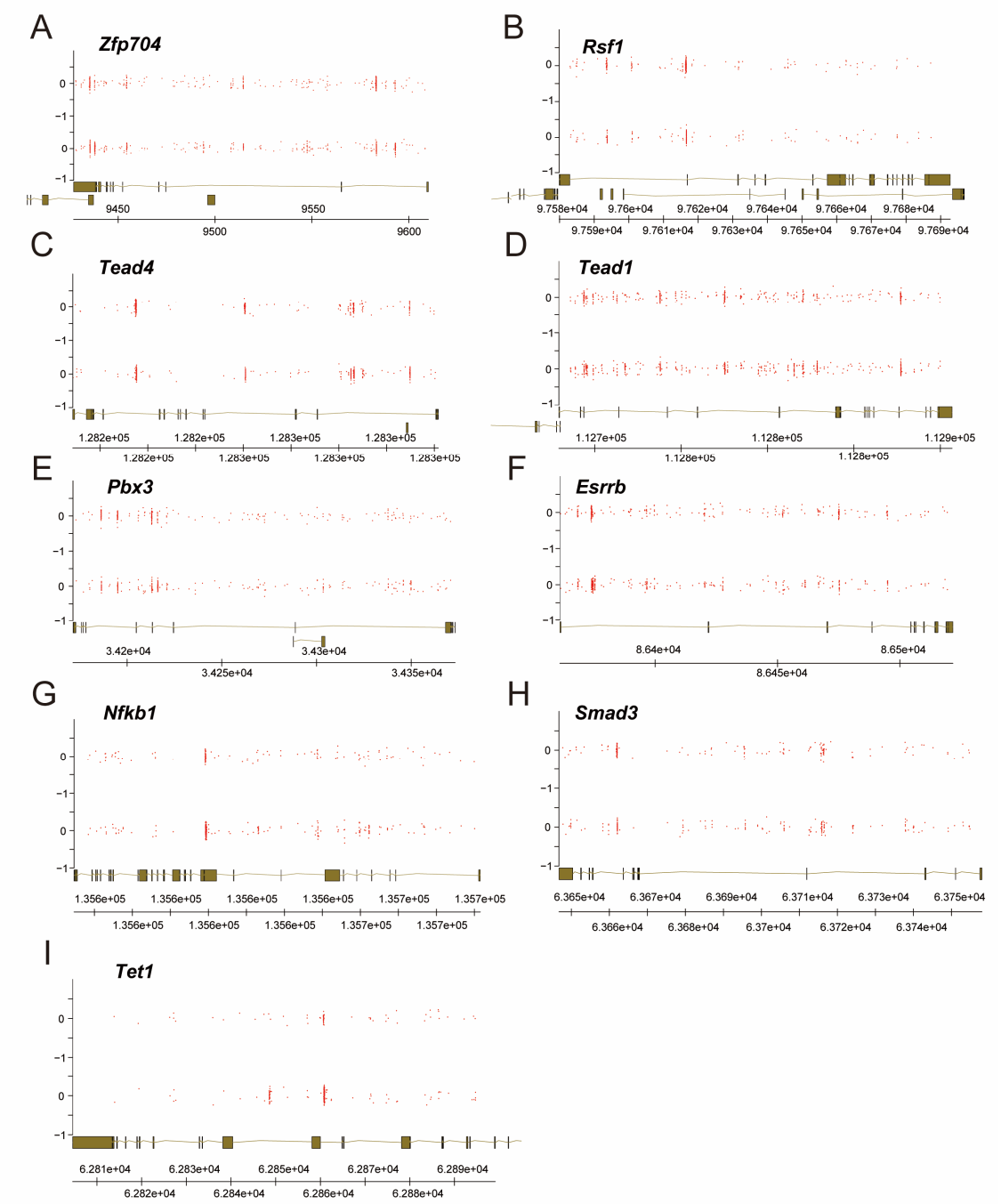

Figure S7 related to Figure 5

(A-I) Strand-specific coverage tracks of the genes *Zfp704*, *Rsf1*, *Tead4*, *Tead1*, *Pbx3*, *Esrrb*, *Nfkb1*, *Smad3* and *Tet1* for the selected library (red).

Table S1. Primer Sequences, related to Figures 1, 2, 3, 4 and 6.

Genotyping	**Target Name**	**Sequence 5'-3'**
	AD	F: TATGCCAAGTACGCCCCCTA
	AD	R: AGTAATTCCAGAGCGCCGTT
	OE	F: GGTGACGTGGAGGAGAATCC
	OE	R: CTCGTAGAAGGGGAGGTTGC
	Htra1	F: GGCCAATGGGCTTAACCGT
	Htra1	R: GACTGGTCGGGCTGAGTTG
Construct vector		
	rtTA-neo	F: CCTACTAGTCACGGATCCAGACATGATAAGATACATT
	rtTA-neo	R: TCACACGCGTCACGGTACCTTACTTAGTTACCCGGGG
	Cdx2	F:GAGCTCGGTACCCGGGGATCCATGTACGTGAGCTACCTTCTG
	Cdx2	R: ACGCGTCTGGGTGACAGTGGAGTT
	Puro	F: GTCACCCAGACGCGTGaGGGcaGaGGaaGtcttc
	Puro	R: GCTGCCACTGTTTCTTTAGG
qPCR		
	Gapdh	F: AGGTCGGTGTGAACGGATTTG
	Gapdh	R: TGTAGACCATGTAGTTGAGGTCA
	Oct4	F: GGATGGCATACTGTGGACCTC
	Oct4	R: TTTCATGTCCTGGGACTCCTCG
	Nanog	F: CCAGGGCTATCTGGTGAACG
	Nanog	R: CCCGAAGTTATGGAGCGGAG
	Rex1	F: CCCTCGACAGACTGACCCTAA
	Rex1	R: TCGGGGCTAATCTCACTTTCAT
	Cdx2	F: GTCCCTAGGAAGCCAAGTGAA
	Cdx2	R: TTGGCTCTGCGGTTCTGAAA
	Eomes	F: GGAAGTGACAGAGGACGGTG
	Eomes	R: TTGGCGAAGGGGTTATGGTC
	Elf5	F: TCTGCTGCGACCAGTACAAG
	Elf5	R: GGAGTAACCTTGCGAGCGAA
	Tfap2C	F: ATCCCTCACCTCTCCTCTCC
	Tfap2c	R: CCAGATGCGAGTAATGGTCGG
	Ctsq	F: AGGCTATGTGACTCGTGTGA
	Ctsq	R: GGCACCAGTCACAGGAAAAG
	Prl3b1	F: CCAGAAAACAGCGAGCAAGT
	Prl3b1	R: CCAGGCTTGTAAAATAGTGATGG
	Prl2c2	F: GCCGGCAGTTTGTCTCATAA
	Prl2c2	R: TGAGCCCGAGCACGTTAGAA

	Tpbpa	F: GCTATAGTCCCTGAAGCGCA
	Tpbpa	R: ACTCCACACTGCTTTTATGAGA
	Prl3d1	F: GCCGCAGATGTGTATAGGGA
	Prl3d1	R: AGGGGAAGTGTTCTGTCTGT
	P53	F: CATGAACCGCCGACCTATCC
	P53	R: GCAGTTCAGGGCAAAGGACT
	P21	F: CCCGAGAACGGTGGAACTTT
	P21	R: AGAGTGCAAGACAGCGACAA
	Reprimo	F: TCGCAGTCATGTGTGTGCTC
	Reprimo	R: GCCTCCGGTCCTTCACTAGG
	Cdk1	F: ACACACGAGGTAGTGACGC
	Cdk1	R: TCTGAGTCGCCGTGGAAAAG
	Ascl2	F: AAGCACACCTTGACTGGTACG
	Ascl2	R: AAGTGGACGTTTGCACCTTCA
	Gcm1	F: GCTCCACAGAGGAAGGCCGC
	Gcm1	R: GTTGGTGACCGGGAAGCCGC
Splinkerette PCR		
	Adaptor top	GTTCCCATGGTACTACTCATATAATACGACTCACTATAGG TGACAGCGAGCGCT
	Adaptor bottom	GCGCTCGCTGTCACCTATAGTGAGTCGTATTATAATTTTTT TTTCAAAAAAA
	Adaptor	R1: GTTCCCATGGTACTACTCATA
	PB5’	F1: GATATACAGACCGATAAAACACATGCGTCA
	PB3’	F1: GACGGATTCGCGCTATTTAGAAAGAGAG
	Adaptor	R2: TAATACGACTCACTATAGG
	PB5’	F2: ACGCATGATTATCTTTAACGTACGTCAC
	PB3’	F2: CATGCGTCAATTTTACGCAGACTATC
P53 knockout sgRNA		
	sgRNA1-1	CACCGAGGAGCTCCTGACACTCGGA
	sgRNA1-2	AAACTCCGAGTGTCAGGAGCTCCTC
	sgRNA2-1	CACCGACCCTGTCACCGAGACCCCT
	sgRNA2-2	AAACAGGGGTCTCGGTGACAGGGTC
Htra1 knockout sgRNA		
	sgRNA1-1	CACCGGCCGCGTCCGCGACGCGTG
	sgRNA1-2	AAACCACGCGTCGCGGACGCGGCC

Transparent Methods

Animal use and care

Specific pathogen-free (SPF) grade mice were purchased from Beijing Vital River Laboratory Animal Technology Co. Ltd. All animal procedures were performed under the ethical guidelines of Nankai University Animal Center. Female 129Sv/Jae mice were sacrificed to provide oocytes for the establishment of haESCs and haiTSCs.

Cell Culture

HaESCs were cultured as described previously (Li et al., 2012). TSC medium was made according to a previous report (Tanaka et al., 1998) with slight modification. Briefly, we used a combined TSC medium termed 70CM, containing 30% RPMI 1640 (Thermo) and supplemented with 20% FBS (BI), 2 mM L-glutamine (Sigma), 25 ng/ml human recombinant FGF4 (PeproTech) and 1 μg/ml heparin (Millipore), and 70% MEF conditioned medium with the same supplements. To culture WT-TSCs and stable haiTSCs in chemically defined medium, cells were grown on Matrigel (BD)-coated or feeder-coated dishes in TX medium as described previously (Kubaczka et al., 2014). For differentiation experiments, TS medium lacking FGF4 and heparin was used.

Vector construction

For vector 1, the *PiggyBac* Dual Promoter was digested by *BamHI* and *KpnI* (Thermo), and rtTA and Neo were inserted. For vector 2, the *PiggyBac* Dual promoter was digested by *XhoI* and *EcoRV*, and the TRE, Cdx2 coding sequence and Puro fragments were inserted. *P53*-targeting sgRNAs were created with the CRISPR design tool (www.crispr.mit.edu). The pSpCas9 (BB)-2A-GFP (PX458) vector was digested and dephosphorylated by *BbsI* and *FastAP* (Fermentas). Single-strand oligonucleotides were synthesized and phosphorylated by *T4PNK* (Takara Japan), and a pair of oligonucleotides was annealed and ligated into lined PX458 to generate knockout plasmids. All plasmids were purchased from Addgene (USA). All primers used are listed in Table S1.

Transfection

To obtain OE-Cdx2 cell lines, approximately 1×10^6 haESCs were electroporated with 6 μg of vector 1 or vector 2 and 2 μg of transposase vector by using an electroporator (Invitrogen) at 1,400 V, 10 μs with three pulses. For vector 1 selection, cells were treated with 250 μg/ml G418 (Thermo) for 7 days. Neomycin-resistant cells were purified for haploid cells. For deletion of *P53*, approximately 1×10^6 cells were transfected with 4 μg of sgRNA-Cas9 plasmid. Cas9-GFP-positive cells were sorted 36 hours after transfection by flow cytometry. For the gene trapping experiment, a combination of 10 μg of PBase plasmid and 30 μg of *piggyBac* plasmid (designs see Figure 4A) was electroporated into 1×10^7 haiTSCs using the same conditions as described above (Wang et al., 2018). Puromycin-resistant cells were harvested for Splinkerrette PCR.

Generation of induced TSCs from haESCs *in vitro*

OE-Cdx2 haESCs were seeded at a density of 1×10^5 per well of a six-well plate containing feeder cells and cultured in ESC medium for the first 24 hours. To initiate induction, cells were cultured in TSC media with doxycycline (1 μg/ml) and puromycin (1 μg/ml). Puromycin was withdrawn after 5 days of selection. The medium was refreshed every day until Day 11 to enrich haploid cells, as shown by Hoechst 33342 staining. Haploid cells were sorted according to a previous report (Shuai et al., 2014) a second time after another 12 days. Enriched haploid cells were expanded, and FACS was performed with the CDCP1 antibody as

described previously (Rugg-Gunn et al., 2012).

Immunostaining, AP staining, and karyotype Analysis

Before staining, cells were fixed with 4% paraformaldehyde. Permeabilization was performed with 0.3% Triton X-100 (Sigma) for 40 minutes, and then the cells were incubated with 1% BSA (Sigma) for 1 hour at room temperature. Samples were incubated with primary antibodies against Oct4 (Abcam), SSEA1 (Santa Cruz), Cdx2 (Abcam), Nanog (Abcam), Eomes (Abcam) and Tpbpa (Abcam) at 4℃ overnight. After three wash steps, the cells were stained with fluorescently coupled secondary antibodies at room temperature for 1 hour. After another three times wash steps, the nuclei were stained with DAPI (YEASEN) for 10 minutes at room temperature. For immunofluorescence: sections were permeabilized with 0.5% Triton X-100 in PBS and then blocked with 5% BSA for 1 h at RT. Slides were then incubated in primary antibody diluted in blocking solution at 4°C overnight. The following primary antibodies were used: Alpha (Ab13970, Abcam), Proliferin (sc-271891, Santa Cruz), GCM1 (sc-101173, Santa Cruz), and Anti-GFP (Ab104401, Abcam). Subsequently, sections were washed with PBST three times and incubated for 1 h at RT in the dark with secondary fluorescence antibodies. The nuclei were stained with Hoechst 33342 for 10 min at RT. AP staining of OE-Cdx2 haESCs was performed as described previously (Zhao et al., 2009). AP staining was performed according to the manufacturer's instructions for the alkaline phosphatase kit (Beyotime). The results were observed under an inverted microscope or laser-scanning confocal microscope. Karyotype analysis was performed following standard instructions.

Quantitative Real-time PCR

Total RNA was purified from cells using Trizol reagent (Thermo), while cDNA was obtained using a Prime Script™ RT Reagent Kit with gDNA Eraser (Takara). Quantitative PCRs were performed with an ABI QuantStudio™ 6 Flex machine using FS Universal SYBR Green Master (Roche). Relative expression levels were normalized to *Gapdh*. Averages and SD values were from three independent experiments. All the primers used are listed in Table S1.

Western blotting

For western blotting experiments, protein samples were extracted by RIPA lysis solution (Solarbio) from ESCs and TSCs. Lysates were precleared by centrifugation for 5 minutes. Equal amounts of cell lysates were separated by SDS-PAGE and transferred to a polyvinylidene fluoride (PVDF) membrane (GE). The membranes were blocked in 5% nonfat dry milk for 1 hour, washed three times by TBST, and incubated with primary antibodies overnight at 4℃. Then, the membranes were washed and incubated with appropriate secondary antibodies for 1 hour at room temperature. Signals on the membrane were detected by using Western Blotting Detection Reagents (Engreen) and imaged. The primary antibodies used for western blotting included anti-Cdx2 (Abcam), anti-Eomes (Abcam) and anti-Gapdh (Abcam).

Bisulphite Sequencing

Genomic DNA of each sample was modified with the EpiTect Bisulfite Kit (Qiagen). Two-round nested PCR was performed to amplify the promoter region of each gene and PCR products were purified from agarose gel by Gel Extraction Kit (OMEGA).Primer sequences were as previous described for Oct4 (Blelloch et al., 2006), Nanog(Takahashi and Yamanaka, 2006) and Elf5(Ng et al., 2008).

Splinkerrette PCR

Splinkerrette PCR was used to find the insertion sites after *piggyBac* transposon transfection as previously described (Uren et al., 2009). Briefly, the genomic DNA purified from cells was digested by *Bsp143I*

(Thermo), and an adenine was added to the tails of fragments by *Taq* DNA Polymerase (Thermo). Purified products were linked to Splinkerrette adaptors for nest PCR. Nest PCR products were linked into pEASY Blunt simple vectors (Transgene) for Sanger sequencing. All primers and Splinkerrette adaptors used are listed in Table S1.

T7ENI assay

Following the manufacturer's recommended protocol (NEB), target DNA fragments containing sgRNA target sites were amplified from transfected cells and wild-type cell genomic DNA by Q5® Hot Start High-Fidelity DNA Polymerase (NEB). The two kinds of fragments were mixed together and annealed in Buffer 2.0 (NEB). T7 Endonuclease I was added into the mixture, and digestion was carried out for 1 hour at 37℃. The products were run on a 2% TAE gel (Biowest), stained with SYBR® Safe DNA gel stain (Thermo) and imaged on a gel imager (Bio-Rad).

TSC Transplantation and immunohistochemistry staining

TSC transplantations were performed as published (Kubaczka et al., 2014). Briefly, 1×10^6 haiTSCs were resuspended in 200 μl of 70CM containing FGF4/heparin and injected into the testis of 8-week-old male ICR mice. Approximately 21 days postinjection, hemorrhagic lesions were dissected from the testis, fixed in 4% paraformaldehyde overnight, embedded in paraffin, and sectioned. Sections were rehydrated and treated for antigen retrieval in sodium citrate buffer at a sub-boiling temperature. Next, endogenous peroxidase activity was quenched with 3% hydrogen peroxidase at room temperature (RT). For immunohistochemistry: sections were blocked with 5% BSA for 1 hour at RT. The following primary antibodies were used: Tfap2c (sc-12762; Santa Cruz) and CD31 (MA5-13188, Thermo Fisher). For immunofluorescence: sections were permeabilized with 0.5% Triton X-100 in PBS and then blocked with 5% BSA for 1 h at RT. All primary antibodies against Alpha (Ab13970, Abcam), Proliferin (sc-271891, Santa Cruz), GCM1 (sc-101173, Santa Cruz), and Anti-GFP (Ab104401, Abcam) were incubated with samples overnight at 4℃.

Before staining, cells were fixed with 4% paraformaldehyde. Permeabilization was performed with 0.3% Triton X-100 (Sigma) for 40 minutes and then incubated with 1% BSA (Sigma) for 1 hour at RT. Samples were incubated with primary antibodies against Oct4 (Abcam), SSEA1 (Santa Cruz), Cdx2 (Abcam), Nanog (Abcam), Eomes (Abcam) and Tpbpa (Abcam) at 4℃ overnight. After three times washing steps, the samples were stained with fluorescently coupled secondary antibodies at room temperature for 1 hour.

Diploid chimeric assay

Chimeric embryos were generated by injecting haiTSC-GFP cells into CD-1 mouse recipient 4-cell embryos according to a previous protocol (Li et al., 2012), using ES-GFP cells as a control treatment. Diploid 4-cell embryos were collected from the uterus of a 2.5 dpc female CD-1 mouse. Approximately 10-15 ES-GFP cells and FACS-purified G0/G1 phase haiTS-GFP cells were injected into each blastocyst. Then, chimeric embryos were cultured to the blastocyst stage in KSOM (Millipore) for subsequent analysis.

Genetic screening in haiTSCs

HaiTSCs were mutated with trapping vectors (Figure 4A) and treated with puromycin (1 μg/ml) for 1 day. Puromycin-resistant haiTSCs and nontransfected haiTSCs were differentiated in standard TS medium without heparin and FGF4 for 3 days. Cells positive for Tpbpa (Abcam, ab104401) among the differentiated cells were harvested by antibody staining and sorting according to the manufacturer's

instructions. In each repeat, we selected Tpbpa-positive cells in the significantly increasing group relative to the nontransfected group to perform high-throughput sequencing.

Analysis of RNA-seq data

Raw data were trimmed using customized scripts to remove low-quality bases (quality score < 5) or reads (low-quality bases > 50% of the read size). Then, the abundance of each transcript was counted using Kallisto (Bray et al., 2016) with Gencode M18 (Harrow et al., 2006) and further summarized for each gene using the R package tximport (Soneson et al., 2015). Genes covered by less than 2 reads across all samples or with more than 2 samples possessing zero reads were filtered out. Genes were normalized using relative-log-expression (RLE) from DESeq2 (Love et al., 2014). Correlations among samples were measured based on the top 50 most variable genes using cosine distance. To generate the MDS plot of all samples, data were first transformed using the function variance Stabilizing() and then the function plotPCA() from DESeq2 (Love et al., 2014). Plots were visualized by ggplot2 (Ginestet, 2011).

Analysis of *piggyBac* screening data

Reads were mapped to the genome assembly from UCSC (mm10) (Zimin et al., 2014) using bowtie2 (v2.2.3) (Langmead and Salzberg, 2012) under sensitive-local mode. Read distributions across different regions were summarized using RSeQC (v2.6.1) (Wang et al., 2012). VISITs (v0.22) (Yu and Ciaudo, 2017) was used to estimate the number of insertions for each gene using the annotation from Gencode (Harrow et al., 2006), excluding duplicated and multiple-hit reads.

Data were first filtered by removing genes covered by < 2 reads, then normalized using relative-log-expression (RLE) from DESeq2 (v1.10.1) (Love et al., 2014). Log2-transformed fold changes were further calculated based on normalized data. The coverage tracks were generated using R package GenomeGraphs (v1.30) (Bullard, 2016). For enrichment analysis, gene sets were retrieved from the Gene Ontology (Ashburner et al., 2000) and KEGG databases (Kanehisa and Goto, 2000). The top 200 genes with the most abundant insertions were used. Fisher's exact test followed by the Benjamini-Hochberg correction (Benjamini and Hochberg, 1995) was used to generate the FDR.

Supplemental Reference

ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S., EPPIG, J. T., HARRIS, M. A., HILL, D. P., ISSEL-TARVER, L., KASARSKIS, A., LEWIS, S., MATESE, J. C., RICHARDSON, J. E., RINGWALD, M., RUBIN, G. M. & SHERLOCK, G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet,* 25**,** 25-9.

BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological,* 57**,** 289-300.

BLELLOCH, R., WANG, Z., MEISSNER, A., POLLARD, S., SMITH, A. & JAENISCH, R. 2006. Reprogramming efficiency following somatic cell nuclear transfer is influenced by the differentiation and methylation state of the donor nucleus. *Stem Cells,* 24**,** 2007-13.

BRAY, N. L., PIMENTEL, H., MELSTED, P. & PACHTER, L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol,* 34**,** 525-7.

BULLARD, S. D. A. J. 2016. GenomeGraphs: Plotting genomic information from Ensembl. R package version 1.32.0. *Bioconductor*.

GINESTET, C. 2011. ggplot2: elegant graphics for data analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.

HARROW, J., DENOEUD, F., FRANKISH, A., REYMOND, A., CHEN, C. K., CHRAST, J., LAGARDE, J., GILBERT, J. G., STOREY, R., SWARBRECK, D., ROSSIER, C., UCLA, C., HUBBARD, T., ANTONARAKIS, S. E. & GUIGO, R. 2006. GENCODE: producing a reference annotation for ENCODE. *Genome Biol,* 7 Suppl 1**,** S4 1-9.

KANEHISA, M. & GOTO, S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res,* 28**,** 27-30.

KUBACZKA, C., SENNER, C., ARAUZO-BRAVO, M. J., SHARMA, N., KUCKENBERG, P., BECKER, A., ZIMMER, A., BRUSTLE, O., PEITZ, M., HEMBERGER, M. & SCHORLE, H. 2014. Derivation and maintenance of murine trophoblast stem cells under defined conditions. *Stem Cell Reports,* 2**,** 232-42.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9**,** 357-9.

LI, W., SHUAI, L., WAN, H., DONG, M., WANG, M., SANG, L., FENG, C., LUO, G. Z., LI, T., LI, X., WANG, L., ZHENG, Q. Y., SHENG, C., WU, H. J., LIU, Z., LIU, L., WANG, L., WANG, X. J., ZHAO, X. Y. & ZHOU, Q. 2012. Androgenetic haploid embryonic stem cells produce live transgenic mice. *Nature,* 490**,** 407-11.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15**,** 550.

NG, R. K., DEAN, W., DAWSON, C., LUCIFERO, D., MADEJA, Z., REIK, W. & HEMBERGER, M. 2008. Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nat Cell Biol,* 10**,** 1280-90.

RUGG-GUNN, P. J., COX, B. J., LANNER, F., SHARMA, P., IGNATCHENKO, V., MCDONALD, A. C., GARNER, J., GRAMOLINI, A. O., ROSSANT, J. & KISLINGER, T. 2012. Cell-surface proteomics identifies lineage-specific markers of embryo-derived stem cells. *Dev Cell,* 22**,** 887-901.

SHUAI, L., LI, W., WAN, H., ZHAO, X. Y., WANG, L. & ZHOU, Q. 2014. Generation of Mammalian offspring by haploid embryonic stem cells microinjection. *Curr Protoc Stem Cell Biol,* 31**,** 1A 6 1-15.

SONESON, C., LOVE, M. I. & ROBINSON, M. D. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res,* 4**,** 1521.

TAKAHASHI, K. & YAMANAKA, S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell,* 126**,** 663-76.

TANAKA, S., KUNATH, T., HADJANTONAKIS, A. K., NAGY, A. & ROSSANT, J. 1998. Promotion of trophoblast stem cell proliferation by FGF4. *Science,* 282**,** 2072-5.

UREN, A. G., MIKKERS, H., KOOL, J., VAN DER WEYDEN, L., LUND, A. H., WILSON, C. H., RANCE, R., JONKERS, J., VAN LOHUIZEN, M., BERNS, A. & ADAMS, D. J. 2009. A high-throughput splinkerette-PCR method for the isolation and sequencing of retroviral insertion sites. *Nat Protoc,* 4**,** 789-98.

WANG, H., ZHANG, W., YU, J., WU, C., GAO, Q., LI, X., LI, Y., ZHANG, J., TIAN, Y., TAN, T., JI, W., LI, L., YU, Y. & SHUAI, L. 2018. Genetic screening and multipotency in rhesus monkey haploid neural progenitor cells. *Development,* 145.

WANG, L., WANG, S. & LI, W. 2012. RSeQC: quality control of RNA-seq experiments. *Bioinformatics,* 28**,** 2184-5.

YU, J. & CIAUDO, C. 2017. Vector Integration Sites Identification for Gene-Trap Screening in Mammalian Haploid Cells. *Scientific reports,* 7**,** 44736.

ZHAO, X. Y., LI, W., LV, Z., LIU, L., TONG, M., HAI, T., HAO, J., GUO, C. L., MA, Q. W., WANG, L., ZENG, F. Y. & ZHOU, Q. 2009. iPS cells produce viable mice through tetraploid complementation. *Nature,* 461**,** 86-U88.

ZIMIN, A. V., CORNISH, A. S., MAUDHOO, M. D., GIBBS, R. M., ZHANG, X., PANDEY, S., MEEHAN, D. T., WIPFLER, K., BOSINGER, S. E., JOHNSON, Z. P., THARP, G. K., MARCAIS, G., ROBERTS, M., FERGUSON, B., FOX, H. S., TREANGEN, T., SALZBERG, S. L., YORKE, J. A. & NORGREN, R. B., JR. 2014. A new rhesus macaque assembly and annotation for next-generation sequencing analyses. *Biol Direct,* 9**,** 20.