**Stem Cell Reports, Volume *11***

## Supplemental Information

# Dynamics of Telomere Rejuvenation during Chemical Induction to Pluripotent Stem Cells

**Haifeng Fu, Cheng-lei Tian, Xiaoying Ye, Xiaoyan Sheng, Hua Wang, Yifei Liu, and Lin Liu**

**Supplemental Information**

# Dynamics of Telomere Rejuvenation During Chemical Induction to Pluripotent Stem Cells

Haifeng Fu, Xiaoying Ye, Cheng-lei Tian, Xiaoyan Sheng, Hua Wang, Yifei Liu and Lin Liu

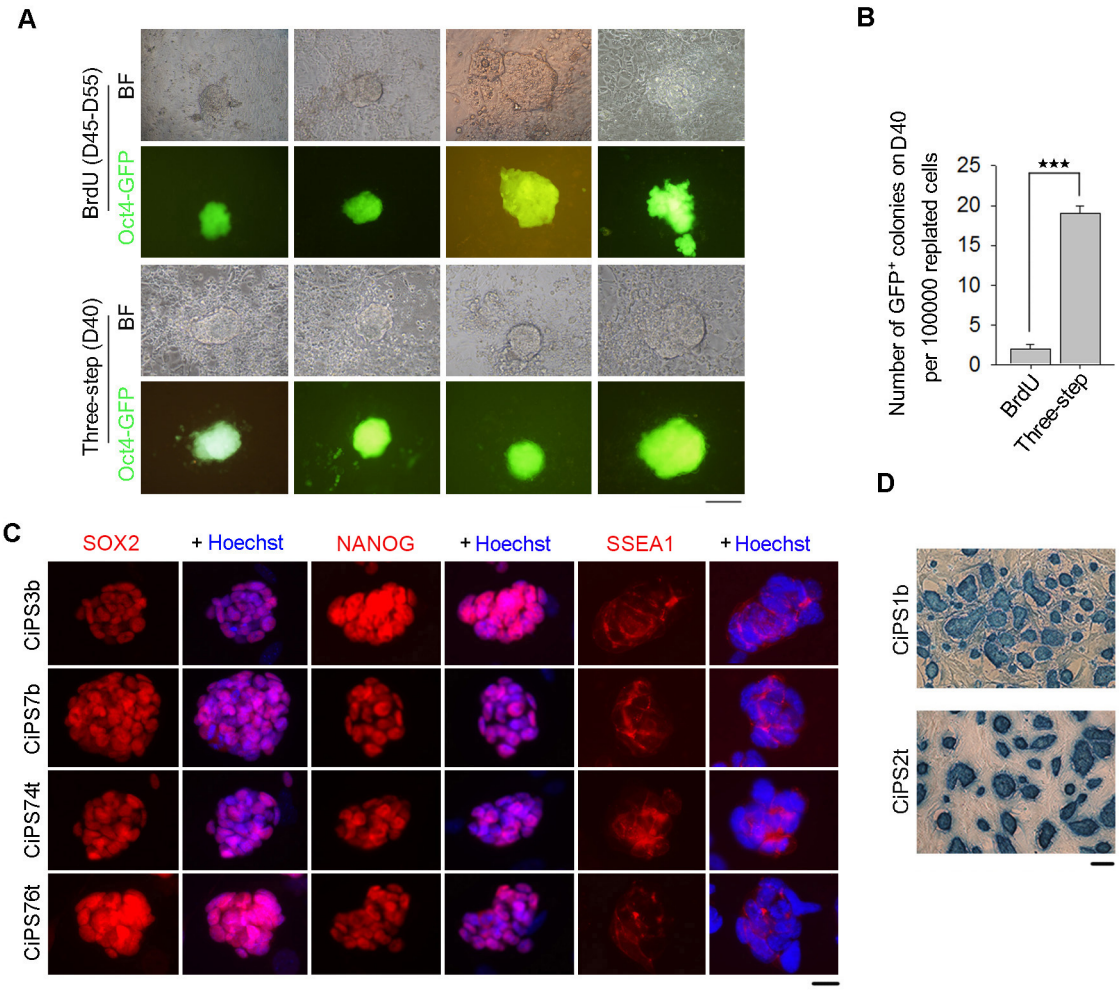



**Figure S1. Characterization of CiPSCs. Related to Figure 1.**

(A) Morphology of CiPSC primary colonies generated by BrdU and Three-step methods under bright field (BF) with phase contrast optics and Oct4-GFP fluorescence. Oct4-GFP fluorescence were visible in some cells at day 40 by BrdU method, but the colonies with clear boundary appeared later by BrdU method than did Three-step method. Scale bar=100 μm.

(B) Induction efficiency of colony formation based on number of Oct4-GFP fluorescence positive clones. Data represent mean ± SEM from three independent experiments. ***P<0.001.

(C) Immunofluorescence of pluripotent markers SOX2, NANOG and SSEA1 in various CiPSCs (b, BrdU method; t, Three-step method). Scale bar=20 μm.

(D) Alkaline phosphatase (AP) staining of CiPSCs at P5. Scale bar=100 μm.

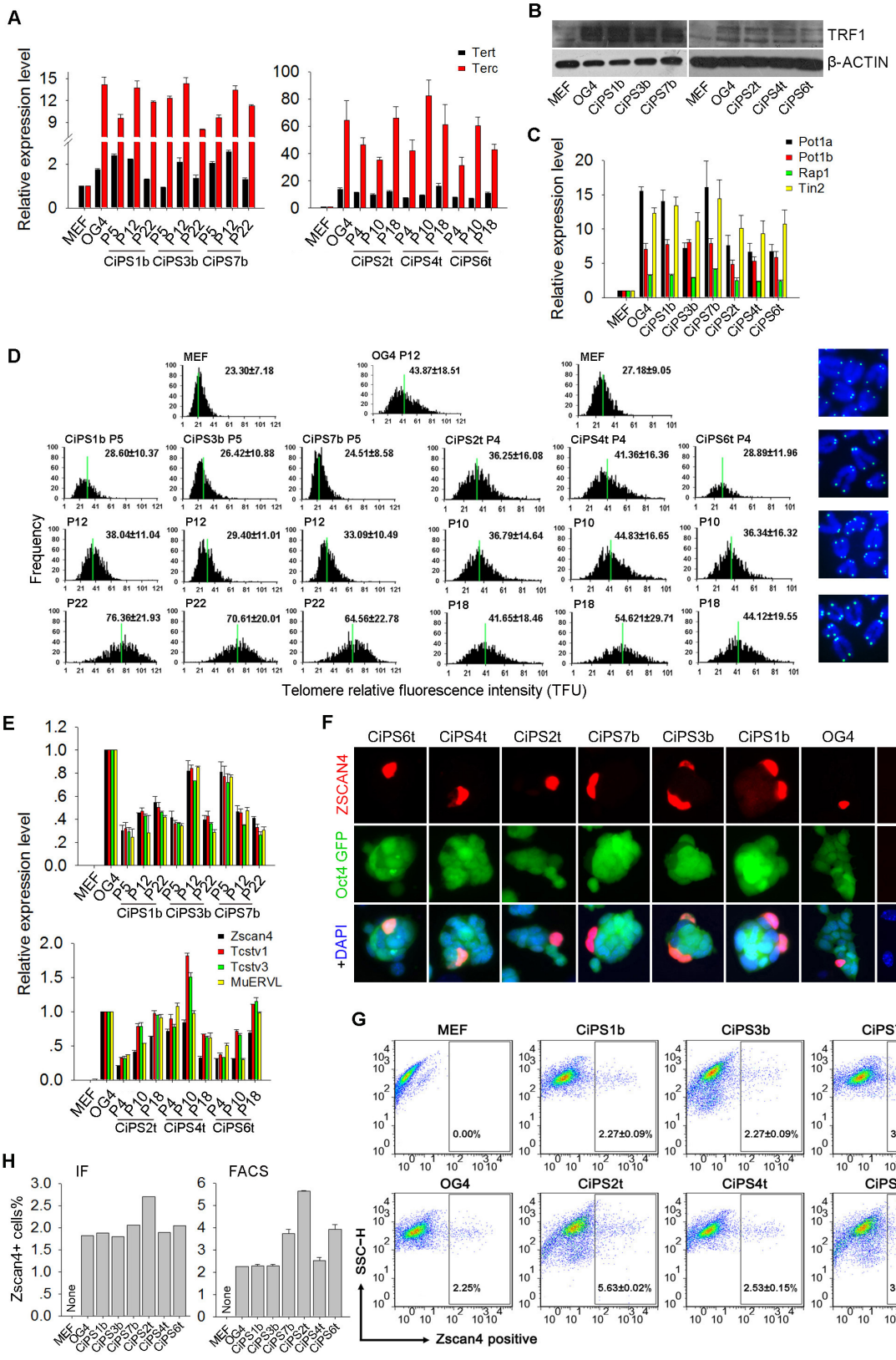

A
Relative expression level
Tert
Terc
MEF
OG4
CiPS1b
CiPS3b
CiPS7b
CiPS2t
CiPS4t
CiPS6t
B
TRF1
β-ACTIN
C
Relative expression level
Pot1a
Pot1b
Rap1
Tin2
D
Frequency
Telomere relative fluorescence intensity (TFU)
E
Relative expression level
Zscan4
Tcstv1
Tcstv3
MuERVL
F
ZSCAN4
Oct4 GFP
+DAPI
G
SSC-H
Zscan4 positive
H
IF
FACS
Zscan4+ cells%
None

**Figure S2. Telomere Rejuvenation and Expression of *Zscan4* and Two-Cell Genes in CiPSCs. Related to Figure 2.**

(A) Relative expression levels by qPCR of telomerase genes *Tert* and *Terc* in CiPSCs at various passages, compared with MEFs, and ESCs (OG4) also expressing Oct4-GFP fluorescence.

(B and C) Expression of shelterin complex TRF1 protein or genes by Western blot or qPCR analysis.

(D) Histogram displays distribution of relative telomere length shown as TFU by telomere quantitative fluorescence *in situ* hybridization (Q-FISH). The medium telomere length (green bars) is shown as mean ± SD above each panel. On the right next to the cell line are telomere FISH images.

(E) RT-qPCR analysis of two-cell genes *Zscan4*, *Tcstv1*, *Tcstv3* and *MuEVRL* in CiPSCs following passage. MEFs and ESCs served as negative and positive controls.

(F) Immunofluorescence of ZSCAN4 in CiPS cell lines and ESCs OG4 at P15. MEFs as controls do not express ZSCAN4 and Oct4-GFP fluorescence. Scale bar=10 μm.

(G) Analysis of Zscan4$^{+}$ cells in CiPSC and ESC cultures at P15 by flow cytometry.

(H) Quantification and proportion of Zscan4$^{+}$ by immunofluorescence (IF, n=400 cells counted) or by FACS.

Data represent mean ± SEM from three independent experiments.

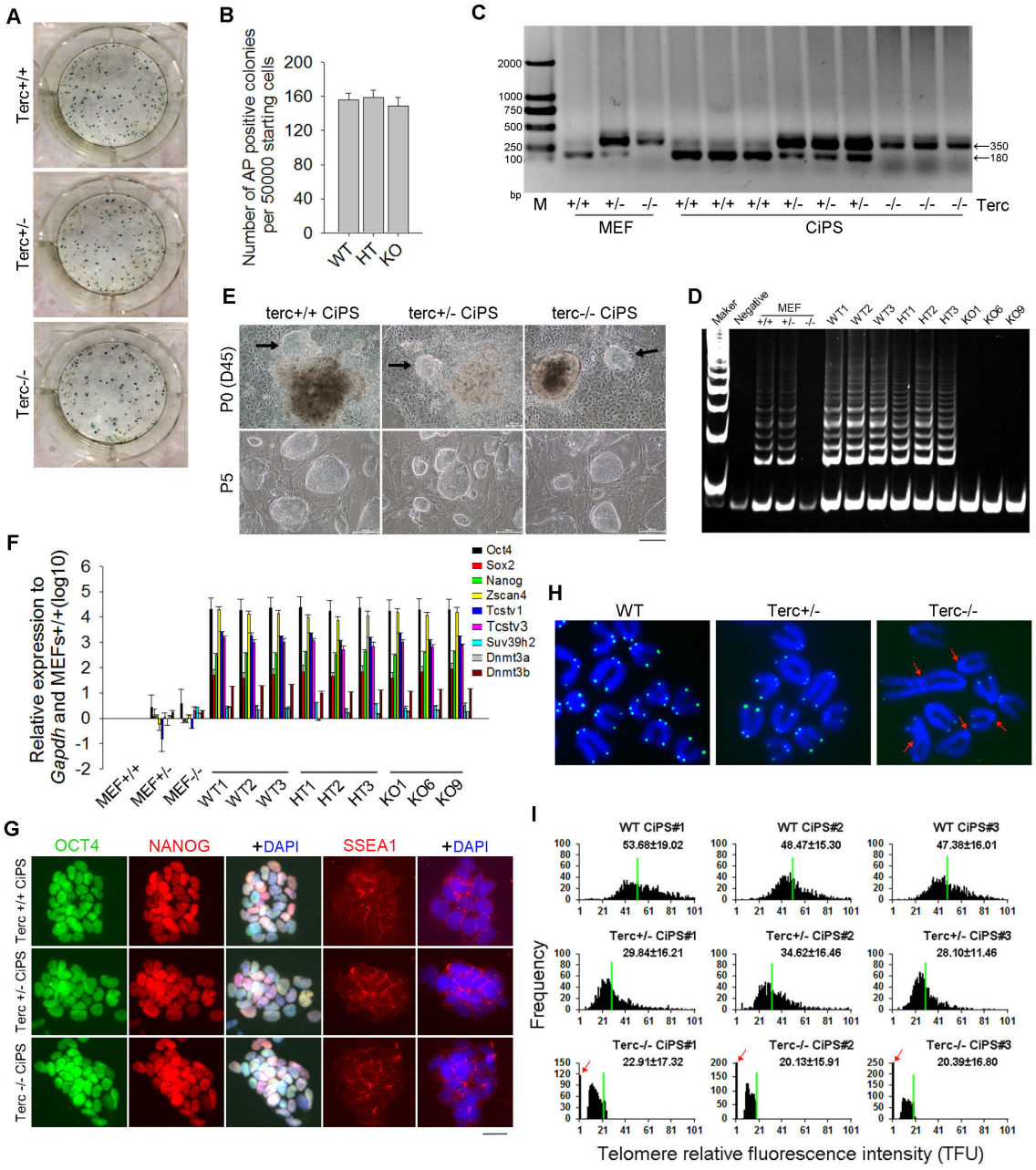



**Figure S3. Telomerase Is Dispensable for CiPSC Induction but Required for Telomere Elongation during Passages after CiPSC Formation. Related to Figure 2.**

(A) Representative images showing clonal formation based alkaline phosphatase (AP) activity staining. CiPSCs were generated using the three-step method described in Experimental Procedures.

(B) Number of AP positive colonies does not differ among telomerase *Terc* knockout (KO, *Terc*$^{-/-}$), heterozygous (HT, *Terc*$^{+/-}$) and Wild-type (WT, *Terc*$^{+/+}$) cells. Data represent mean ± SEM from three independent experiments. P>0.05.

(C) Genotyping of *Terc* KO (*Terc*$^{-/-}$), heterozygous (HT, *Terc*$^{+/-}$) and Wild-type (*Terc*$^{+/+}$) CiPSC lines and their progenitor MEFs.

(D) Telomerase activity by TRAP assay, confirming that *Terc*$^{-/-}$ cells do not express

telomerase activity.
(E) Typical morphology under phase contrast of primary colonies and formed CiPSCs at passage 5. Scale bar=100 μm.
(F) Relative expression level by qPCR of genes for pluripotency, two-cell genes and methyltransferase genes. Data represent mean ± SEM from three independent experiments.
(G) Immunofluorescence of OCT4, NANOG and SSEA1. Scale bar=20 μm.
(H and I) Telomere Q-FISH images (H) and Histogram displaying distribution of relative telomere length shown as TFU (I). The medium telomere length (green bars) is shown as mean ± SD above each panel. Red arrows indicate chromosome fusion or telomere-signal free ends indicative of telomere loss in *Terc*$^{-/-}$ CiPSCs. CiPSCs at passage 15 were analyzed for telomere lengths estimated by Q-FISH.

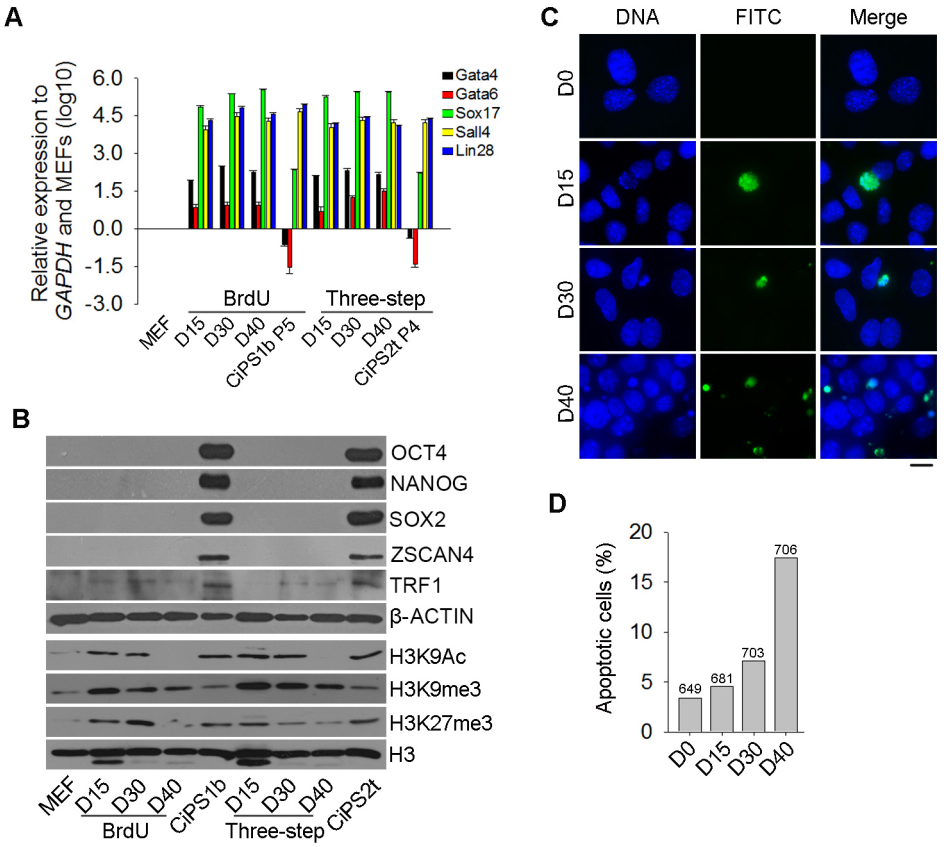



**Figure S4. Expression of Genes Related to XEN State and Pluripotency and Apoptosis during Chemical Induction. Related to Figure 4.**

(A) RT-qPCR analysis of relative expression levels of *Gata4, Gata6, Sox17, Sall4* and *Lin28* at various days during chemical induction using BrdU or Three-step method. Data represent mean ± SEM from three independent experiments.

(B) Western blot analysis of relative protein expression during induction. H3 and β-actin served as loading control.

(C) Representative images showing apoptotic cells by TUNEL assay. Apoptotic cells are revealed by fragmented nuclei and TUNEL signals (FITC green fluorescence). Scale bar=10 μm.

(D) Percentage of apoptosis cells on various days during chemical induction. Number of cells counted is shown above the bar. Data from three independent experiments.

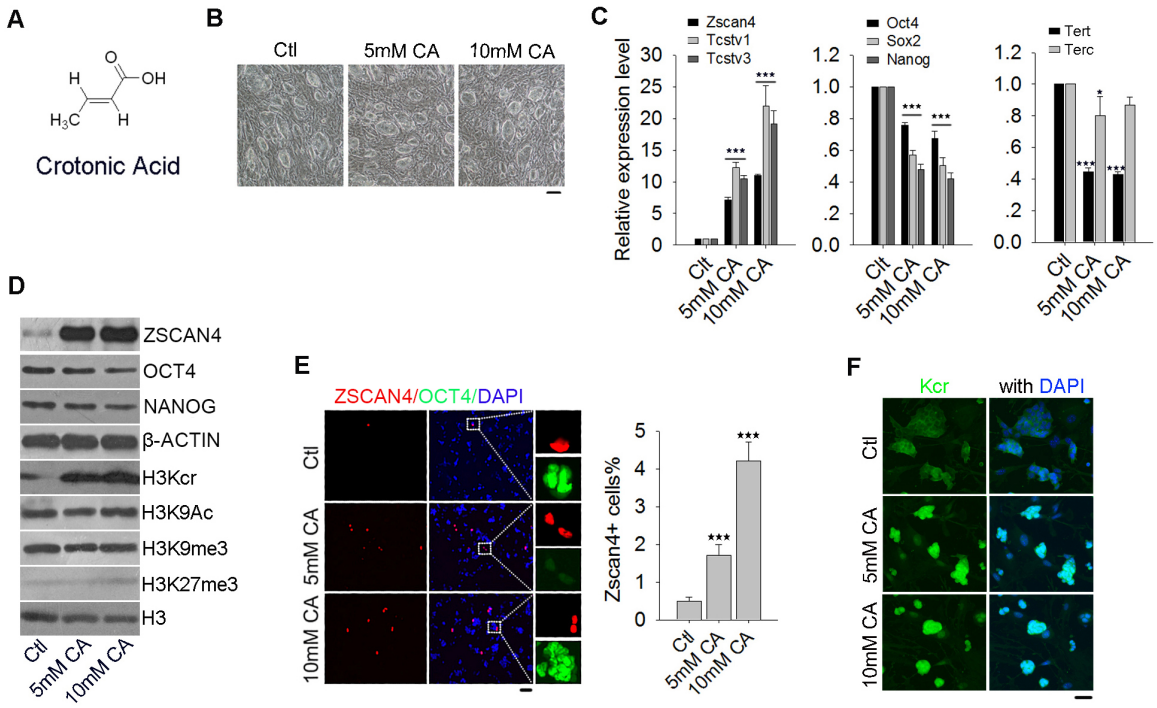



**Figure S5. Crotonic Acid Increases *Zscan4* Expression in mESCs. Related to Figure 5.**

(A) Chemical structures of crotonic acid (CA).

(B) Morphology of mouse ESCs (N33) following treatment with 5 or 10 mM crotonic acid (CA) for two passages. Scale bar=100 μm.

(C) Relative expression level by qPCR analysis of two-cell genes, pluripotent genes and telomerase genes of N33 ESCs treated with 5 or 10 mM crotonic acid for two passages. Data represent mean ± SEM from three independent experiments. *P<0.05; ***P<0.001, compared to controls (Ctl).

(D) Protein levels by Western blot of ZSCAN4, OCT4, NANOG, H3Kcr, H3K9me3, H3K9ac, and H3K27me3.

(E) Immunofluorescence of ZSCAN4 in ESCs and proportion of Zscan4$^{+}$ cells following treatment with crotonic acid. n=400 cells counted. ***P<0.001, compared to controls. Scale bar=10 μm. Data represent mean ± SEM from three independent experiments.

(F) Immunofluorescence of lysine crotonylation (Kcr) increases in N33 ESCs treated with 5 or 10 mM crotonic acid. Scale bar=50 μm.

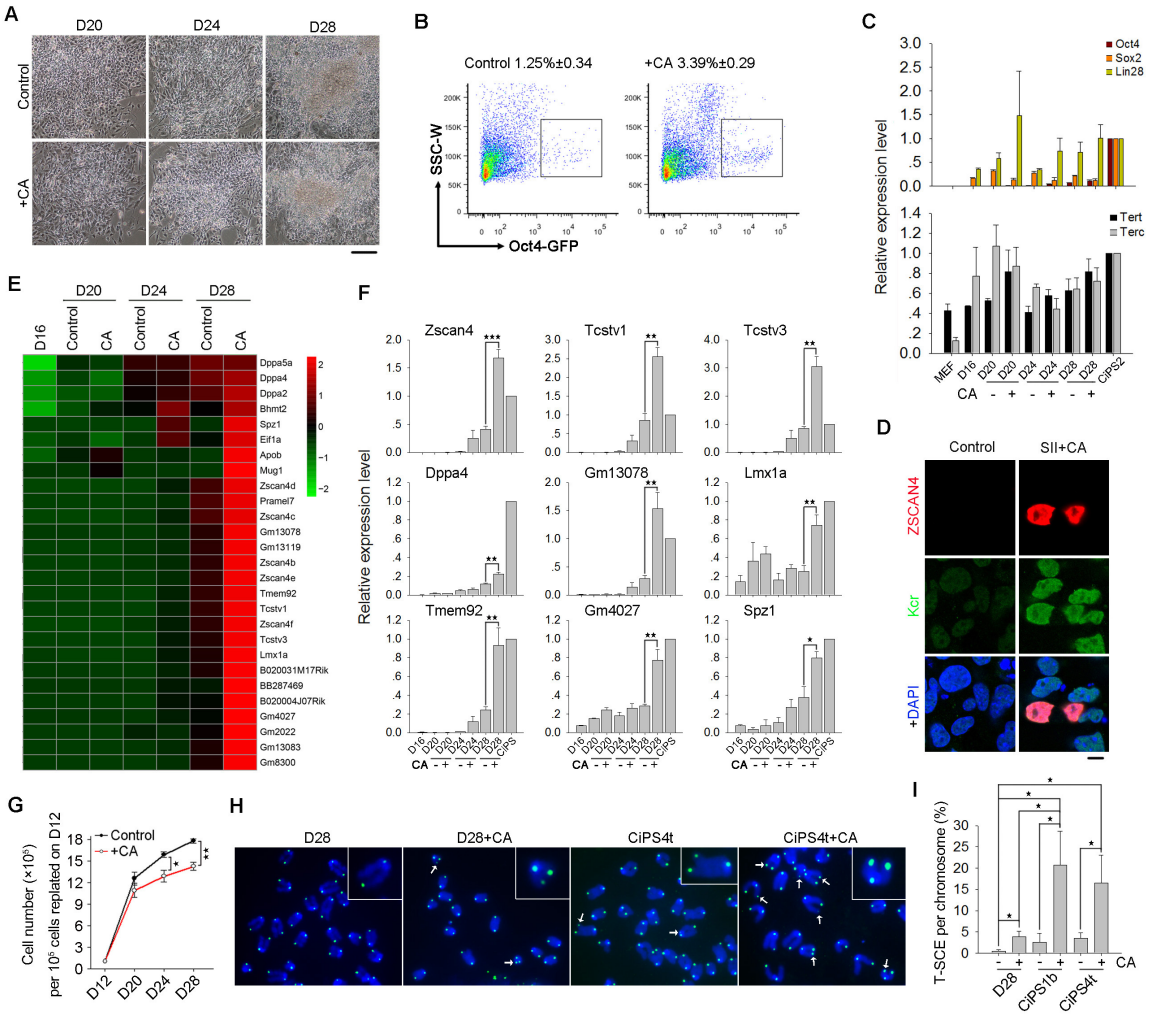



**Figure S6. Crotonic Acid Activates Two-Cell Genes and Enhances CiPSC Generation. Related to Figure 6.**

(A) Morphology of chemical induction on D20, D24 and D28 following crotonic acid (CA) addition at Stage II, compared with controls without the treatment. Scale bar=100 μm.

(B) Flow cytometry analysis of induction efficiency at D40 based on endogenous Oct4-GFP fluorescence.

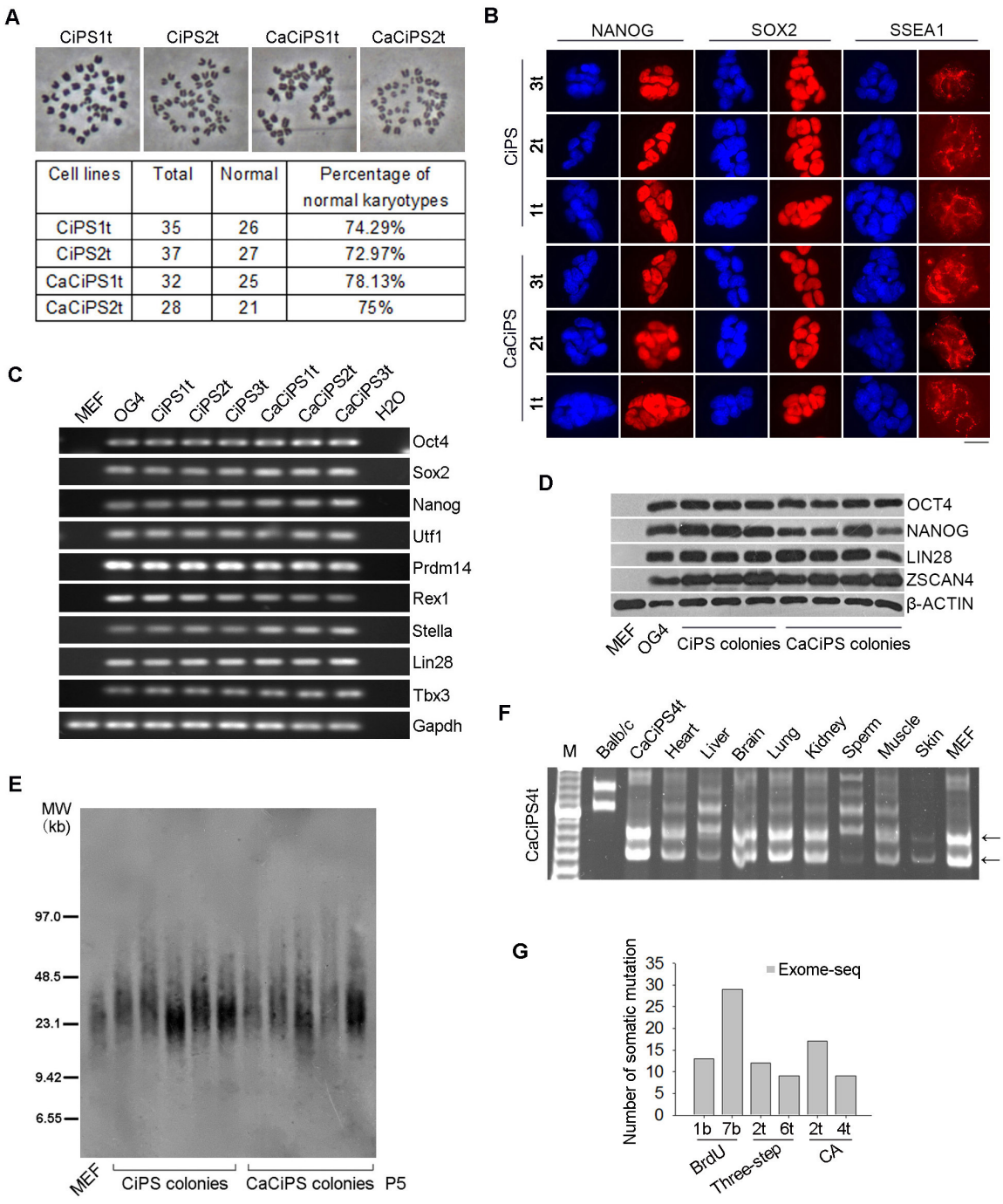
(C) Relative expression levels by qPCR of pluripotent genes *Oct4*, *Sox2*, and *Lin28* and telomerase genes *Tert* and *Terc* on D20, D24 and D28 of cells treated with or without crotonic acid. Data represent mean ± SEM from three independent experiments.

(D) Confocal microscopy showing co-expression of lysine crotonylation (Kcr) and Zscan4 on D28 of induction. Scale bar=10 μm.

(E) Heatmap displaying two-cell genes highly expressed in crotonic acid-treated cells on day 24 and particularly on day 28 by RNA-seq, compared with controls. The two-cell genes were listed in published data (Macfarlan, T.S., *et al.* Nature 2012; 487, 57-63).

(F) qPCR analysis validating the RNA-seq data showing increased expression of two-cell genes by crotonic acid in E. Data represent mean ± SEM from three

independent experiments.
(G) Reduced proliferation of chemical reprogramming cells after CA treatment. $10^5$ cells were re-plated on D12, and cells were counted at three time points. Data represent mean ± SEM from three independent experiments.
(H) Representative micrographs showing T-SCE (white arrows) by CO-FISH assay.
(I) T-SCE per chromosome in D28 reprogramming cells and two CiPS cell lines without or with CA treatment. At least 15 metaphase spreads were counted for frequency of T-SCE. Data from two independent experiments. Pairwise comparisons for statistical significance were made by t-tests.
*P<0.05, **P<0.01, ***P<0.001.




| Cell lines | Total | Normal | Percentage of normal karyotypes |
|---|---|---|---|
| CiPS1t | 35 | 26 | 74.29% |
| CiPS2t | 37 | 27 | 72.97% |
| CaCiPS1t | 32 | 25 | 78.13% |
| CaCiPS2t | 28 | 21 | 75% |

**Figure S7. Characterization of CaCiPSCs Obtained by Addition with Crotonic Acid at Stage II during Induction. Related to Figure 7.**

(A) Chromosome spread and karyotype analysis of CiPSCs as controls and CaCiPSCs. Majority of spread exhibit normal ploidy.

(B and C) Expression of pluripotency marker genes by immunofluorescence (B) and conventional RT-PCR (C). Scale bar=20 μm.

(D) Protein levels by Western blot of OCT4, SOX2, LIN28 and ZSCAN4 in CiPSCs and CaCiPSCs compared with MEFs and ESCs (OG4).

(E) Telomere length distribution shown as TRF by Southern blot analysis of CiPSCs and CaCiPSCs at passage 5. MEFs served as control.

(F) Microsatellite genotyping analysis showing contributions of CaCiPSCs to various tissues of the chimera. Four-eight-cell embryos from Balb/c mice were used as recipients. Arrows indicate donor MEFs, the derived CaCiPSCs and contributions in the chimera.
(G) Number of somatic mutations in coding regions in CiPSCs by whole exome sequencing analysis (Data detailed in Table S4 and S5).

## Table S1. Primers for RT-qPCR analysis

| Genes | Forward | Reverse |
|---|---|---|
| Oct4 | TTGGGCTAGAGAAGGATGTGGTT | GGAAAAGGGACTGAGTAGAGTGTGG |
| Sox2 | GCACATGAACGGCTGGAGCAACG | TGCTGCGAGTAGGACATGCTGTAGG |
| Nanog | TTGCTTACAAGGGTCTGCTACT | ACTGGTAGAAGAATCAGGGCT |
| Tert | ACTGGTGGAGATCATCTTTCTGGG | ACCTGAGGAGTCTGACATATTGGC |
| Terc | CATTAGCTGTGGGTTCTGGTCT | TCCTGCGCTGACGTTTGTTT |
| P53 | TGTTATGTGCACGTACTCTCCTCC | GTGCTGTGACTTCTTGTAGATGGC |
| P21 | AGACCTGTGAAGACAGGAATGGTC | AGCAGATCACCAGATTAACCCTCC |
| Sall4 | TGGCAGACGAGAAGTTCTTTC | TCCAACATTTATCCGAGCACAG |
| Zscan4 | AAATGCCTTATGTCTGTTCCCTATG | TGTGGTAATTCCTCAGGTGACGAT |
| Tcstv1 | TGAACCCTGATGCCTGCTAAGACT | AGATGGCTGCAAAGACACAACTGC |
| Tcstv3 | AGAAAGGGCTGGAACTTGTGACCT | AAAGCTCTTTGAAGCCATGCCCAG |
| Rex1 | ACGAGTGGCAGTTTCTTCTTGGGA | TATGACTCACTTCCAGGGGGCACT |
| Pot1b | TCCTCATACGAGGGAAGGTG | GATGCTGGGATCTGGAAAAA |
| Pot1a | AGCTTCACTCCTCAGGACCA | GGGTTCCATCCCATACCTTT |
| Rap1 | CCGCTACCTCAAGCACCTAC | CACTCCTCCTCAGGCAAGTC |
| Tin2 | GGAGTTTCTGCAGTCCTTGC | TCTGGACTCTGCTGGGAAGT |
| MuERV-L | CCCATCATGAGCTGGGTACT | CGTGCAGATCCATCAGTAAA |
| Sox17 | CTCGGGGATGTAAAGGTGAA | GCTTCTCTGCCAAGGTCAAC |
| GATA4 | TCTCACTATGGGCACAGCAG | GCGATGTCTGAGTGACAGGA |
| GATA6 | CAAAAGCTTGCTCCGGTAAC | TGAGGTGGTCGCTTGTGTAG |
| GAPDH | TCAACAGCAACTCCCACTCTTCCA | ACCACCCTGTTGCTGTAGCCGTAT |
| Lin28a | CCGCAGTTGTAGCACCTGTCT | GAAGAACATGCAGAAGCGAAGA |

## Primers for RT-PCR analysis

| Genes | Forward | Reverse |
|---|---|---|
| Oct4 | TTGGGCTAGAGAAGGATGTGGTT | GGAAAAGGGACTGAGTAGAGTGTGG |
| Sox2 | GCACATGAACGGCTGGAGCAACG | TGCTGCGAGTAGGACATGCTGTAGG |
| Nanog | TTGCTTACAAGGGTCTGCTACT | ACTGGTAGAAGAATCAGGGCT |
| Utf1 | TTCGCCGCCGCTCTACT | CAGGGGCAGGTTCGTCATT |
| Lin28a | CCGCAGTTGTAGCACCTGTCT | GAAGAACATGCAGAAGCGAAGA |
| Stella | CCCAATGAAGGACCCTGAAAC | AATGGCTCACTGTCCCGTTCA |
| Rex1 | ACGAGTGGCAGTTTCTTCTTGGGA | TATGACTCACTTCCAGGGGGCACT |
| Tbx3 | CCACCCGTTCCTCAATTTGAACAG | CGGAAGCCATTGATGGTAAAGCTG |
| Prdm14 | ACAGCCAAGCAATTTGCACTAC | TTACCTGGCATTTTCATTGCTC |
| GAPDH | TCAACAGCAACTCCCACTCTTCCA | ACCACCCTGTTGCTGTAGCCGTAT |

## Primers for ChIP-qPCR analysis

| Genes | Forward | Reverse |
|---|---|---|
| Actin | CGTGTGACAAAGCTAATGAGGCTG | CTAAGTTCAGTGTGCTGGGAGTCT |
| Zscan4-1 | GCATTATCTGTTCCTCTGGGTC | AACTCCTGTTCCTGGGTGGG |
| Zscan4-2 | TCCCTAGAATACAGTCCTCA | GTAGAATCCTTGATAGTGGG |
| Ch13_Subtel | GCACACTTGGTGGGCTAAGAAGATG | TTAAATCCTGACCAAAATGCCTGGC |
| MuERV-L | CAGAGCATTCACACTGGGGA | GTGAGCCTTCCAATTCCGGA |
| ERVK10c | TGTCAGCTGGCAAAAGAGTAAA | AAGACAGGGGAAGTCAGTTCAG |

## Table S2. Antibody details

| Antibody | Source | Cat. No. |
|---|---|---|
| NANOG | Abcam | ab80892 |
| OCT3/4 | Santa Cruz Biotechnology | sc-5279 |
| SOX2 | Millipore | AB5603 |
| SSEA1 | Millipore | MAB4301 |
| ZSCAN4 | Millipore | AB4340 |
| TCSTV1/3 | Custom-made | |
| LIN28a | CST | 3978S |
| βIII-TUBULIN | Chemicon | CBL412 |
| TRF1 | Alpha Dragrostic | TRF12-S |
| γH2AX | Millipore | 05-636 |
| AFP | Dako | DAK-N1501 |
| α-SMA | Abcam | ab5694-100 |
| NESTIN | Millipore | MAB353 |
| Pan anti-crotonyllysine | PTM Biolabs | PTM-502 |
| β-ACTIN | Abmart | P30002 |
| H3K9me3 | Abcam | ab8898 |
| H3K27me3 | Millipore | 07-449 |
| H3K9ac | Abcam | ab4441 |
| H3 | Abcam | ab1791 |

## Table S4. Sequencing Statistics for Mutation Discovery

| Cell Line | Quality-Filtered Sequences (bp) | 10 x Exome Coverage Sequences (bp) | No. of High-Quality Coding Variants | dbSNP Percentage | No. of Unique Coding Mutations |
|---|---|---|---|---|---|
| MEF | 5,521,756,147 | 38,115,851 | 18,243 | 87% | — |
| MEF+BrdU | 5,472,472,685 | 38,382,831 | 18,387 | 86% | 0 |
| CiPS1b | 5,614,716,507 | 37,886,101 | 12,942 | 90% | 13 |
| CiPS7b | 5,697,170,568 | 37,890,164 | 12,277 | 89% | 29 |
| MEF | 5,525,093,305 | 38,415,593 | 17,981 | 87% | — |
| CiPS2t | 5,177,923,246 | 38,037,210 | 12,667 | 92% | 12 |
| CiPS6t | 5,217,052,761 | 37,706,210 | 12,771 | 82% | 9 |
| CaCiPS2t | 5,059,846,928 | 36,645,405 | 17,909 | 86% | 17 |
| CaCiPS4t | 7,101,264,336 | 38,808,513 | 17,550 | 87% | 9 |

## Table S5. Summary of Somatic Mutations

| Cell Line | Mutated Genes |
|---|---|
| MEF+BrdU | n/a |
| CiPS1b | *Chd2*, *Sv2b*, *Vmn2r65*, *Xrra1*, *Olfr657*, *Olfr667*, *Olfr480*, *Vmn1r194*, *Shisa9*, *Dgcr8*, *Ufd1*, *Abi3bp*, *Tmem2* |
| CiPS7b | *Ttll4*, *Olfr1221*, *Olfr1288*,*Eif2ak4*, *Pak6*, *Errfi1*, *Grm3*, *Oasl1*, *Cttnbp2*, *Prss3*, *Lsr*, *Olfr598*, *Syt9*, *Cend1*,*Il12rb1*, *Agt*, *Slc44a2*, *Grik4*, *C1qtnf5_Mfrp*, *Atm*, *Aldh8a1*, *Ddx5*,*Sdk2*, *Aspscr1*, *Fam208b*,*Olfr1368*, *Nutm2*, *Csf2rb2*, *Vwa5b2* |
| CiPS2t | *Lamc1*, *Pramel7*, *Olfr1154*,*Fpgt*, *Qrich1*, *Usp32*,*Kif2b*, *Akap6*, *Mboat1*, *Arhgap39*, *Ano6*, *Pdzph1*, *Tigd3* |
| CiPS6t | *Chd6*, *Med8*, *Cyp2b19*, *Ltbp4*, *Map2k7*, *Cdk19*, *Hdac9*, *Calml3*, *Tmem44* |
| CaCiPS2t | *Mcm3*, *Gli2*, *Olfr1154*, *Gpcpd1*, *Nrd1*, *Capzb*, *Ccl25*, *Aktip*, *Cdh23*, *Mdm2*, *Gpatch8*, *Kif19a*, *Ppp1r13b*,*Bmp1*, *Dgkh*, *Trappc9*, *Atp9b* |
| CaCiPS4t | *Ttn*, *Abcc9*, *Vmn2r57*, *Olfr834*, *Vmn2r82*, *Gli3*, *Tspan17*, *Adgrv1*, *Krt72* |

## EXPERIMENTAL PROCEDURES

### Mice and Cell Culture

Oct4-GFP (OG2) mice (B6/CBA, JAX stock #004654) that carry Oct4 distal promoter-driven GFP were purchased from Model Animal Research Center of Nanjing University. The transgenic GFP expression of the reporter is under the control of Oct4 promoter and distal enhancer, but the proximal enhancer region is deleted, so the Oct4-ΔPE-GFP transgenic mice have been used for establishing naïve mouse ESCs (Bao et al., 2009; Yeom et al., 1996; Tang et al., 2010). ESC line (OG4) expressing distal Oct4-GFP was derived from OG2 mice and characterized based on the method described (Huang et al. 2008, 2011), but with addition of 2i in the medium. N33 mESC line was derived from C57BL/6 mice (Huang et al. 2011).

Briefly, intact blastocysts were seeded on feeder layers of mitomycin C-treated MEF cells, prepared on 0.1% gelatin-treated four-well culture dishes, in ESC medium consisting of knockout DMEM, 20% knockout serum replacement, supplemented with 1000 units/ml mouse leukemia inhibitory factor, 0.1 mM NEAA, 1 mM L-glutamine, 0.1 mM β-mercaptoethanol, 50 IU/ml penicillin and streptomycin and 2i (3μM CHIR99021 and 1μM PD0325901). Half of the medium was changed daily. Approximately 10 days after seeding, ICM outgrowths were mechanically removed and digested with 0.25% trypsin-ETDA into small clumps, digestion stopped with trypsin inhibitor, and cell suspensions reseeded on fresh feeder cells. Stable ESC lines were routinely obtained after one or two passages. Then, all ESC lines were passaged and cultured in ES medium added with fetal bovine serum and 2i, instead of KSR-based ESC medium.

### Isolation of Mouse Embryonic Fibroblasts (MEFs)

MEFs were derived from E13.5 embryos isolated by caesarean section and washed in phosphate-buffered saline (PBS). Heads and visceral tissues were removed, and remaining tissue was washed in fresh PBS, then submerged in 0.05 mM trypsin/1 mM EDTA HBSS solution and incubated at 37°C for 10 min. Tissue was pipetted repeatedly to aid in tissue dissociation, then added to MEF medium containing 10% FBS and plated (passage 0).

### CiPSC Induction from Mouse Embryonic Fibroblasts

Mouse embryonic fibroblasts (MEFs) were isolated from OG2 mice. Isolated OG-MEF cells at early passages (up to passage 3) were used for chemical induction following the method described (Zhao et al., 2015). MEFs were seeded at a density of 50000 cells per well on 6 well plate. On the next day (day 0), the MEFs medium was replaced with chemical induction medium. On day 0-12, the induction medium contains 100 ng/ml bFGF, 0.5 mM VPA, 20 μM CHIR99021, 10 μM Repsox, 5 μM Tranylcypromine, 50 μM Forskolin, 0.05 μM AM580 and 5 μM EPZ004777. On day 12, the cells were trypsinized, harvested and then re-plated at 100000 cells per well of a 6 well plate. During day 12-16, concentrations of bFGF, CHIR99021, and

Forskolin were reduced to 25 ng/ml, 10 μM and 10 μM respectively. On day 16-28, the induction medium contains 25 ng/ml bFGF, 0.5 mM VPA, 10 μM CHIR99021, 10 μM Repsox, 5 μM Tranylcypromine, 10 μM Forskolin, 0.05 μM AM580, 0.05 μM DZNep, 0.5 μM 5-aza-dC, and 5 μM SGC0946. On day 28, the culture was transferred into N2B27-2iL medium with 3 μM CHIR99021, 1 μM PD0325901, and 1000 U/ml LIF. After another 8-12 days, Oct4 GFP-positive CiPSC primary colonies emerged and were then picked up for expansion and characterization.

For BrdU method, the induction medium contains the small-molecule cocktails: 0.5 mM VPA, 10 μM CHIR99021, 10 μM Repsox, 5 μM Tranylcypromine, 50 μM Forskolin, 1 μM TTNPB, and 8μM BrdU (Long et al., 2015). DZNep (50 nM) was added to the cell cultures on day 16. On day 28-32, the culture was transferred into N2B27-2iL medium. After additional 16-20 days, GFP-positive CiPSC colonies emerged. The medium was changed every 4 days.

In experiments by treatment with crotonic acid during induction, 6-8 mM crotonic acid (pH=7) was added into induction medium from day 16 to day 28 continuously. Oct4 GFP-positive CiPSC colonies emerged about on day 32.

**Embryoid Body Formation Test**

Embryoid body (EB) formation *in vitro* was performed as described previously (Mao et al., 2014). CiPSCs were removed off feeder cells twice based on their differences in the adherence to the bottom of dish. The CiPSCs were diluted to $4\times10^4$ per milliliter. Every 30 μl was pipetted to form a hanging drop on the cover of the 100-mm dish. EBs formed on day 4 and were transferred to 6 well plates for adherent culture. EBs were fixed for immunofluorescence staining using markers of three embryonic germ layers on day 15.

**Teratoma Formation Test**

$1\times10^6$ CiPSCs were injected subcutaneously into about 6-week-old immunodeficient nude mice. About 4 weeks after injection, the mice were humanely sacrificed, and the teratomas were excised, fixed in 4% paraformaldehyde, dehydrated in gradient ethanol, embedded in paraffin, and sectioned for histological examination by haematoxylin and eosin staining.

**Production of Chimeras and Genotyping**

Approximately 10-15 CiPSCs were injected into four or eight-cell embryos as hosts using a Piezo injector as described (Huang et al., 2008). Injected embryos were cultured overnight in $KSOM_{AA}$ medium. Blastocysts were transferred into uterine horns of 2.5 dpc surrogate mice. Pregnant females delivered pups naturally at about 19.5 dpc. Pups were identified initially by coat color. The contribution of CiPSCs to various tissues in chimeras was confirmed by standard DNA microsatellite genotyping analysis using D12Mit136 primers: 5′-TT AAT TTT GAG TGG GTT TGGC-3′ and 5′-TTG CTA CAT GTA CAC TGA TCT CCA-3′, about 147 bp.

**Immunofluorescence Microscopy**

Cells were washed twice in PBS, fixed in freshly prepared 3.7% paraformaldehyde for 30 min at 4°C, washed once in PBS and permeabilized in 0.1% Triton X-100 in blocking solution (3% goat serum plus 0.1% BSA in PBS) for 30 min at room temperature, then washed once in PBS, and left in blocking solution for 2 h. Cells were incubated overnight at 4°C with primary antibodies against OCT4 (sc5279, Santa Cruz), NANOG (ab80892, Abcam), SSEA-1 (MAB4301, Millipore), βIII-tubulin (CBL412, Chemicon), AFP (DAK-N1501, Dako), α-SMA (ab5694-100, Abcam), NESTIN (MAB353, Millipore), ZSCAN4 (AB4340, Millipore), γH2AX (05-636, Millipore), TRF1 (TRF12-S, Alpha Diagnostic). Anti-Pan-Kcr (PTM-502) was permeabilized and blocked in 5% BSA in PBS. Cells were washed three times (each for 15 min) with blocking solution, and incubated for 2 h with secondary antibodies at room temperature. Goat Anti-Mouse IgG (H+L) FITC (115-095-003, Jackson) and Goat Anti-Rabbit IgG (H+L) Alexa Fluor® 594 (111-585-003, Jackson), diluted 1:200 with blocking solution, were used. Samples were washed, and counterstained with 0.5 μg/ml Hoechst 33342 (H1398, MP) or DAPI in Vectashield mounting medium. Fluorescence was detected and imaged using Axio-Imager Z2 fluorescence microscope (Carl Zeiss).

**Fluorescence Microscopy of Teratoma Sections**

Briefly, after being deparaffinized, rehydration and wash in 0.01 M PBS (pH 7.2-7.4), sections were incubated with 3% $H_2O_2$ for 10 min at room temperature to block endogenous peroxidase, subjected to high pressure antigen recovery sequentially in 0.01 M citrate buffer (pH 6.0) for 3 min, incubated with blocking solution (5% goat serum and 0.1% BSA in PBS) for 2 h at room temperature, and then with the diluted primary antibodies overnight at 4 °C. The following primary antibodies were used for immunocytochemistry: NESTIN (1:200, MAB353, Millipore), SMA (1:200, ab5694, Abcam) and AFP (1:5, DAK-N150130, Dako). Blocking solution without the primary antibody served as negative control. After washing with PBS, sections were incubated with appropriate secondary antibodies (1:200, goat anti-mouse IgG (H+L) FITC (115-095-003, Jackson) or goat anti-rabbit IgG (H+L) AlexaFluor® 594 (111-585-003, Jackson). Nuclei were stained using Vectashield medium (Vector) added with DAPI and photographed with Zeiss Axio Imager Z2.

**Gene Expression Analysis by Quantitative Real-Time PCR**

Total RNA was purified using a RNA mini kit (Qiagen), treated with DNase I (Qiagen), and the cDNA was generated from 2μg RNA using Oligo (dT)18 primer (Takara) and M-MLV Reverse Transcriptase (Invitrogen). Real-time quantitative PCR reactions were set up in duplicate with the FS Universal SYBR Green Master (Roche) and carried out on an iCycler MyiQ2 Detection System (BIO-RAD). All reactions were carried out by amplifying target genes and internal control in the same plate. Each sample was repeated three times and normalized using GAPDH as the internal control. The amplification was performed for primary denaturation at 95°C for 10 min, then 40 cycles of denaturation at 95°C for 15 s, annealing and

elongation at 58°C for 1 min, and the last cycle under 55-95°C for dissociation curve. Relative quantitative evaluation of target gene was determined by comparing the threshold cycles. Primers were confirmed their specificity with dissociation curves. Most primers were designed using the IDT DNA website and listed in Table S1.

**Western Blot**

Cells were washed twice in PBS, collected, and lysed in cell lysis buffer on ice for 30 min and then sonicated for 1 min at 60 of amplitude with 2 s intervals. After centrifugation at 10000g, 4°C for 10 min, supernatant was transferred into new tubes. The concentration of the protein sample was measured by bicinchoninic acid, and then protein samples were boiled in SDS Sample Buffer at 99°C for 10 min. 10μg total protein of each cell extract was resolved by 10% or 12% Acr-Bis SDS-PAGE and transferred to polyvinylidene difluoride membranes (PVDF; Millipore). Nonspecific binding was blocked by incubation in 5% skim milk or 5% BSA in TBST at room temperature for 2 h. Blots were then probed with primary antibodies overnight incubation at 4°C with OCT4 (sc5279, Santa Cruz), NANOG (ab21603, Abcam), SOX2 (AB5603, Millipore), TCSTV1/3 (custom-made), ZSCAN4 (AB4340; Millipore), H3K9me3 (ab8898, Abcam), H3K9ac (04-1003, Millipore), H3K27me3 (07-449, Millipore), Pan-Kcr (PTM-502, PTM biolabs), H3 (ab1791, Abcam), and β-ACTIN (P30002, Abmart). Immunoreactive bands were then probed for 2 h at room temperature with the appropriate horseradish peroxidase (HRP)-conjugated secondary antibodies, anti-Rabbit IgG-HRP (GE Healthcare, NA934V), or goat anti-Mouse IgG (H + L)/HRP (ZSGB-BIO, ZB-2305). Protein bands were detected by Chemiluminescent HRP substrate (Millipore, WBKLS0500). Information for antibodies is listed in Table S2.

**Telomerase Activity Assay**

Telomerase activity was determined by the Stretch PCR method according to manufacturer's instruction using TeloChaser Telomerase assay kit (T0001, MD Biotechnology). About $2.5\times10^4$ cells from each sample were lysed. Lysis buffer served as negative controls. PCR products of cell lysate were separated on non-denaturing TBE-based 10% polyacrylamide gel electrophoresis and visualized by ethidium bromide staining.

**FACS Analysis**

For the FACS analysis of Zscan4 expression, CiPSCs were collected and washed with cold PBS, then fixed in cold 70% ethanol, permeabilized in 0.1% Triton X-100 in blocking solution (3% goat serum in PBS) for 30 min, washed three times and left in blocking solution for 2 h. Cells were incubated overnight at 4 °C with primary antibody against ZSCAN4, washed three times, and incubated for 2h with secondary antibody 594-goat anti-rabbit IgG (H+L) (111-585-003, Jackson) diluted 1:200 with blocking solution. Samples were washed three times with PBS and FACS analysis performed using a FACScan Flow Cytometer (BD Biosciences). For

analysis of endogenous Oct4-GFP, cells were collected and washed with cold PBS, subject to FACS analysis.

**Telomere Q-FISH**

Telomere length was estimated by telomere Q-FISH as described previously (Herrera et al., 1999; Poon et al., 1999). Telomeres were denatured at 80 ºC for 3 min and hybridized with FITC-labeled $(CCCTAA)_3$ peptide nucleic acid (PNA) probe at 0.5 μg/ml (Panagene, Korea). Chromosomes were stained with 0.5 μg/ml DAPI. Fluorescence from chromosomes and telomeres was digitally imaged on a Zeiss microscope with FITC/DAPI filters, using AxioCam and AxioVision software 4.6. Telomere length shown as telomere fluorescence intensity was integrated using the TFL-TELO program (a gift kindly provided by Peter Lansdorp, Terry Fox Laboratory).

**Telomere Restriction Fragment (TRF) Measurement**

The TRF analysis was performed using a commercial kit (TeloTAGGG Telomere Length Assay, catalog no. 12209136001, Roche Life Science), based on the method described previously (Sung et al., 2014). DNA was extracted from cells. A total of 3 μg DNA was digested overnight with MboI at 37°C and electrophoresed through 1% agarose gels in 0.5 × TBE at 14°C for 16 h at 6 V/cm with an initial pulse time of 1 s and end in 12 s using a CHEF Mapper pulsed field electrophoresis system (Bio-Rad). The gel was blotted and probed using reagents in the kit.

**Telomere Chromosome Orientation-Fluorescence *In Situ* Hybridization (CO-FISH)**

CO-FISH assay was performed as described previously with minor modification (Bailey et al., 2004). Subconfluent ES or CiPS cells were incubated with BrdU (10 μM) for 10-12 h, and MEF or day 28 reprogramming cells incubated with BrdU (10 μM) for 20-24 h. Nocodazole at 0.3 μg/ml was added for 3 h (and 5 h for day 28 reprogramming cells) prior to cell harvest, and metaphase spreads were prepared by a routine method. Chromosome slides were treated with RNase A, fixed with 4% formaldehyde, then stained with Hoechst 33258 (0.5 mg/ml), incubated in 2 × SSC (Invitrogen) for 15 min and exposed to 365 nm UV light (Stratalinker 1800UV irradiator) for 40 min. The BrdU-substituted DNA was digested with Exonuclease III (Takara). The slides were then dehydrated through ethanol series and air-dried. PNA-FISH was performed with FITC-OO-$(CCCTAA)_3$ (Panagene, F1009). Slides were hybridized, washed, dehydrated, mounted, and counter-stained with 1.25 μg/ml DAPI in VectaShield antifade medium. Digital images were captured using CCD camera on Zeiss Imager Z2 microscope. To analyze telomere sister chromatid exchange (T-SCE), one signal at each end of the chromosome was counted as no T-SCE, while two signals at both chromatids on one chromosome end were identified as one T-SCE. At least 15 metaphase spreads were counted for frequency of T-SCE. Pairwise comparisons for statistical significance were made by t-tests.

**C-circle Assay**

The protocol for c-circle amplification was slightly modified from the method described (Henson et al., 2017). Briefly, 400ng genomic DNA was digested with MboI restriction enzymes at 37°C overnight and purified by phenol-chloroform extraction. Genomic DNA from ALT positive (U2OS) cells and ALT negative (Hela) cells was used as a positive and negative control. DNA was diluted in double distilled water. Samples (10 µl) were combined with 10 µl 1×Φ29 Buffer (NEB, UK) containing BSA, 0.2 mM each dATP, dGTP, dTTP and incubated in the presence or absence of 5U ΦDNA polymerase (NEB, UK) at 30°C for 12 h and then at 65°C for 20 min. Added 200 µl 6×SSC to reaction products and dot-blotted onto a 6×SSC-soaked nylon membrane. DNA was cross-linked onto the membrane and hybridized with a DIG-labeled probe $(CCCTAA)_4$ to detect C-circle amplification products. Blots were washed, and exposed to Tanon 5200.

**Apoptosis by TUNEL Assay**

Apoptosis was revealed by catalytically incorporating fluorescein-12-dUTP at 3´-OH DNA ends with the Terminal Deoxynucleotidyl Transferase, Recombinant, enzyme (rTdT) using a commercial kit (TB235, Promega). Briefly, after being fixed in 4% paraformaldehyde, washed in PBS, and permeabilized in 0.2% Triton X-100, cells were incubated with rTdT buffer at 37°C for one hour, stained with DAPI for 15 min, washed, mounted in Vectashield, and immediately analyzed under a fluorescence microscope using a standard fluorescein filter set.

**Library Preparation and RNA-sequencing**

Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5x). First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase $H^-$). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3’ ends of DNA fragments, NEBNext Adaptor with hairpin loop structure were ligated to prepare for hybridization. To select cDNA fragments of preferentially 150~200 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then 3 µl USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37 °C for 15 min followed by 5 min at 95 °C before PCR. PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. At last, PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system.

Clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer’s instructions. After cluster generation, the library preparations

were sequenced on an Illumina Hiseq platform and 125 bp/150 bp paired-end reads were generated.

**Differential Gene Expression Analysis**

DESeq provide statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting P-values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate (Hochberg and Benjamini, 1990). Genes with an adjusted P-value <0.05 found by DESeq were assigned as differentially expressed. For DEGSeq without biological replicates: prior to differential gene expression analysis, for each sequenced library, the read counts were adjusted by edgeR program package through one scaling normalized factor. Differential expression analysis of two conditions was performed using the DEG-Seq R package (1.20.0). The P-values were adjusted using the Benjamin & Hochberg method(Hochberg and Benjamini, 1990). Corrected P-value of 0.005 and $\log_2$ (fold change) of 1 were set as the threshold for significantly differential expression.

Clustering and analysis of the early embryo development genes and two-cell genes provided in the heatmap was obtained using RNA-seq data published (Fan et al., 2015; Macfarlan et al., 2012). All 196 differentially expressed genes in our RNA-seq data (Table S3) were compared with genes expressed at various early development stages obtained by single cell RNA-seq data (Fan et al., 2015), using fisher's exact test (P adj$<$0.01). The 196 differentially expressed genes also were compared with the RNA-seq data from ESC cultures (Macfarlan et al., 2012). The calculated z score of selected genes was used for heatmap.

**ChIP-qPCR analysis**

Cells were harvested and fixed by freshly prepared 1% paraformaldehyde solution for 8 min at room temperature. Glycine was added to quench unreacted formaldehyde. The nuclei were extracted, lysed, and sonicated. DNA fragments were then enriched by immunoprecipitation with H3K9me3 (ab8898; Abcam) and HP1$\alpha$ (05-689; Millipore) antibody. The eluted protein:DNA complex was reverse-crosslinked at 65 °C overnight. DNA was recovered after proteinase and RNase A treatment. ChIP-enriched DNA was analyzed by qPCR using primers for *Zscan4*, *Chr13_subtel* and *ERVK10c* loci (Table S1). *$\beta$-actin* served as negative control.

**Whole-exome Sequencing**

Paired-end DNA library were prepared according to manufacturer's instructions (Agilent). Genomic DNAs (gDNA) from cell samples were sheared into 180~280 bp fragments by Covaris S220 sonicator. Ends of the gDNA fragments were repaired; 3′ ends were adenylated. Both ends of the gDNA fragments were ligated at the 3′ ends with paired-end adaptors (Illumina) with a single 'T' base overhang and purified using AMPure SPRI beads from Agencourt. The adaptor-modified gDNA

fragments were enriched by six cycles of PCR using SureSelect Primer and SureSelect ILM Indexing PreCapture PCR Reverse Primer. The concentration and the size distribution of the libraries were determined on an Agilent Bioanalyzer DNA 1000 chip. Whole exome capture was carried out using Agilent's SureSelect Mouse All Exon V1 Agilent 5190-4642. An amount of 0.5 μg prepared gDNA library was hybridized with capture library of biotinylated RNA baits for 5 min at 95 °C, 24 h at 65 °C. The captured DNA−RNA hybrids were recovered using Dynabeads MyOne Streptavidin T1 from Dynal. DNA was eluted from the beads and desalted using Qiagen MinElute PCR purification columns. The purified capture products were then amplified using the SureSelect ILM Indexing Post Capture Forward PCR Primer and PCR Primer Index 1 through Index 16 (Agilent) for 12 cycles. 50 Mb of DNA sequences of 221,784 exons from 24,306 genes were captured. After DNA quality assessment, captured DNA library were sequenced on Illumina Hiseq 4000 sequencing platform (Illumina) according to manufacturer's instructions for paired-end 150 bp reads (Novogene, Beijing). Libraries were loaded onto paired-end flow cells at concentrations of 14–15 pM to generate cluster densities of 800,000–900,000 per mm2 using Illumina cBot and HiSeq paired-end cluster kit version 1.

**Sequence Alignment and Variant Calling**

Variant calling was performed as previously described (Gore et al., 2011). Briefly, paired-end clean reads in FastQ format generated by the Illumina pipeline were aligned to the mouse reference genome mm10 by Burrows-Wheeler Aligner (BWA) to obtain the original mapping results stored in BAM format. SAMtools, Picard (http://broadinstitute.github.io/picard/) were used to sort BAM files and do duplicate marking to generate final BAM files for computation of the sequence coverage and depth. BCFtools was used to call variants, and VariantAnnotation was performed to do annotation for VCF (Variant Call Format). Variant positions were obtained at this step. RefSeq was applied for annotation to determine amino acid alternations. Variants obtained from previous steps were compared against SNPs present in the dbSNP to discard known SNPs. The consensus sequences in mouse CiPSC samples were then used to compare with progenitor MEF samples to find candidate novel mutations. Each heterozygous SNP identified in CiPSC lines that was not observed in the progenitor line was considered candidate mutations. We used DAVID to analyze the somatic-coding mutations and check for commonly mutated pathways. The mutations do not seem to be mutated in any common pathways.

**Statistical Analysis**

Data were analyzed by student's t-test or $\chi 2$ test or fisher's exact test for paired comparison, or by ANOVA and means compared by Fisher's protected least-significant difference (PLSD) for multiple comparisons using the StatView software from SAS Institute Inc. (Cary, NC). Significant differences were defined as *P< 0.05, **P< 0.01, or ***P< 0.001.

## REFERENCES


Bailey, S.M., Brenneman, M.A., and Goodwin, E.H. (2004). Frequent recombination in telomeric DNA may extend the proliferative life of telomerase-negative cells. Nucleic Acids Res *32*, 3743-3751.

Fan, X., Zhang, X., Wu, X., Guo, H., Hu, Y., Tang, F., and Huang, Y. (2015). Single-cell RNA-seq transcriptome analysis of linear and circular RNAs in mouse preimplantation embryos. Genome Biol *16*, 148.

Gore, A., Li, Z., Fung, H.L., Young, J.E., Agarwal, S., Antosiewicz-Bourget, J., Canto, I., Giorgetti, A., Israel, M.A., Kiskinis, E., et al. (2011). Somatic coding mutations in human induced pluripotent stem cells. Nature *471*, 63-67.

Henson, J.D., Lau, L.M., Koch, S., Martin La Rotta, N., Dagg, R.A., and Reddel, R.R. (2017). The C-Circle Assay for alternative-lengthening-of-telomeres activity. Methods *114*, 74-84.

Herrera, E., Samper, E., Martin-Caballero, J., Flores, J.M., Lee, H.W., and Blasco, M.A. (1999). Disease states associated with telomerase deficiency appear earlier in mice with short telomeres. EMBO J *18*, 2950-2960.

Hochberg, Y., and Benjamini, Y. (1990). More powerful procedures for multiple significance testing. Stat Med *9*, 811-818.

Huang, J., Deng, K., Wu, H., Liu, Z., Chen, Z., Cao, S., Zhou, L., Ye, X., Keefe, D.L., and Liu, L. (2008). Efficient production of mice from embryonic stem cells injected into four- or eight-cell embryos by piezo micromanipulation. Stem Cells *26*, 1883-1890.

Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature *487*, 57-63.

Mao, J., Zhang, Q., Ye, X., Liu, K., and Liu, L. (2014). Efficient induction of pluripotent stem cells from granulosa cells by Oct4 and Sox2. Stem Cells Dev *23*, 779-789.

Poon, S.S., Martens, U.M., Ward, R.K., and Lansdorp, P.M. (1999). Telomere length measurements using digital fluorescence microscopy. Cytometry *36*, 267-278.

Sung, L.Y., Chang, W.F., Zhang, Q., Liu, C.C., Liou, J.Y., Chang, C.C., Ou-Yang, H., Guo, R., Fu, H., Cheng, W.T.*, et al.* (2014). Telomere elongation and naive pluripotent stem cells achieved from telomerase haplo-insufficient cells by somatic cell nuclear transfer. Cell Rep *9*, 1603-1609.