Supplementary information

Supplementary Materials and Methods:

Construction of human OCT4 promoter–EGFP lentiviral vector

The human OCT4 promoter was amplified from -3917 to +55 by polymerase chain reaction (PCR) with forward primer 5′-CTC GAG TTC CCA TGT CAA GTA AGT GGG GTG G-3′ and reverse primer 5′-GGA TTC CGA GAA GGC AAA ATC TGA AGC CAG G-3′ from genomic DNA of H9 ESCs (1). The PCR products were cloned into Topo TA vectors (Invitrogen) and sequences were checked bi-directionally. The fragments were digested at X*ho*I and B*amH*I sites, which were included in the primer sequences previously and subsequently cloned into the same cutting site of pSIN18.cPPT.hEF1αP.EGFP.WPRE (a gift from Dr. Nissim Reubinoff, Hadassah University Medical Center, Jerusalem, Israel) after the removal of hEF1α promoter. The recombinant DNA procedures involved were performed according to the National Institutes of Health guidelines.

Lentiviral production and generation of transgenic hESC line

Viruses were produced by the co-transfection of three plasmids, pSIN18.cPPT.pOCT4.EGFP.WPRE, CMVR8.91 and VSVG, into 293T cells using the calcium phosphate precipitation method (2). Cells were washed with PBS five hours after transfection, and 10 ml of medium containing 10% FBS was added back to the cells. Medium containing viral particles was collected at 48 hours and 72 hours. Viruses were then filtered through 0.45 μm filters and concentrated by ultracentrifugation at 25,000 RPM at 4°C for 2.5 hours. Concentrated viruses were used directly for transduction or stored at -80°C.

Procedures used for hESC transduction were similar to those described by Gropp *et al*. 2003 (3). In general, about 20-30 ESC colonies were mechanically sliced and transferred onto a 24-well plate in 200 μl ESC medium. This was followed by the addition of concentrated viral particles (around 2×10^8) in the presence of 5μg/ml Polybrene (Sigma). ES cell clumps and viral particles were incubated at 37°C for 4 hours. Transduced cell clumps were washed with ES culture medium 2-3 times before being transferred onto a new dish plated with a MEF feeder layer. The infected hESCs with OCT4-EGFP expression were manually enriched several passages to ~95% by mechanical slicing under a dissection fluorescence microscope (MZ16F, Leica Microsystems).

Immunofluorescence (IF) staining and confocal microscopy

Cells or cryo-sections were first washed by PBS three times and fixed with 4% paraformaldehyde in PBS for 20 minutes. This was followed by permeabilization and blocking with 0.2% Triton X-100 (Sigma) and 2% goat serum or donkey serum (Jackson Immunoresearch Laboratories) for 30 minutes. Cells or

sections were later incubated with primary antibodies (See Antibodies section) overnight at 4°C. The next day, cells were washed three times with PBS, incubated with secondary antibodies for 40 minutes, and stained with DAPI for 10 minutes in the dark. Cells were washed with PBS three times again before being mounted with Vectashield Mounting Medium (Vector Laboratories). IF staining of cells was observed on a confocal microscope (Leica TCS SP2 Microsystems) or an upright fluorescence microscope (Axiplan 2, Zeisis).

Meiotic spreads and meiotic marker staining

Meiotic spreads were performed according to Nicholas *et al.* 2009 (4). Slides were then blocked with 4% goat serum and incubated with meiotic markers SYCP1 (AB15087) and SYCP3 (AB15093) (Abcam) at 4°C overnight. The subsequent washes, secondary antibody staining and mounting were carried out as described above.

Flow cytometry and cell sorting

Cells were dissociated using trypsin-EDTA (Invitrogen) with physical pipetting to give a single cell suspension. For flow cytometry analysis, cells were prepared as previously described (5). Control cells were treated with non-specific isotype-matched antibodies. Secondary goat anti-cy5 IgG antibody (Jackson Laboratories) was used for detection. At least 10,000 cells were acquired for each sample using a FACS Callibur cytometer and analyzed with CELL-QUEST software (BD Bioscience). For cell sorting, cells were harvested and stained with directed conjugated antibodies (See Antibodies section) and 7-Amino-actinomycin D (7-AAD, BD Bioscience). A BD FACSVantage flow cytometer was used for cell sorting (Cytometry Research LLC).

Quantitative RT-PCR

Total RNA was extracted from sorted cells using an RNeasy Mini Kit and treated with DNaseI (Qiagen). cDNA was synthesized from 2 μg of total RNA using high-capacity cDNA reverse transcription kits (Applied Biosystems). Quantitative RT-PCR (Q-RT-PCR) was performed with power SYBRGreen PCR mater mix (Applied Biosystems) and signals were detected with an Applied Biosystems 7300 Real-time PCR system. Results were normalized using GAPDH and analyzed based on relative quantification (ΔΔ-Ct method). Specific primer sequences used are shown in Supplementary Table 1.

Western blot analysis

The protein was extracted from collected cells by using M-PER mammalian protein extraction reagent (PIERCE). Protein concentration was measured (Bio-Rad protein assay, BioRad) and denatured at a 1:4 ratio with 5x Laemmli buffer at 100 °C for 5 min, then loaded on to a 10% SDS-PAGE gel. The SDS-PAGE gels were run at 120 V for 2 hours and transferred to a PVDF membrane for 1 hours at 15 V

by a semi-dry transfer device (Bio-Rad). Transferred blots were blocked in 10% non-fat milk for 1 hours at room temperature, then incubated with primary antibodies SMAD 1/5/8, phospho-SMAD 1/5/8, β-CATENIN (Cell Signalling), phospho-β-CATENIN (Abcam) and β-ACTIN (Sigma) overnight at 4°C. After two quick rinses and three washes for 10 min in PBS-T (PBS, pH7.4 with 0.1% Tween 20), the secondary antibodies anti-rabbit and mouse HRP conjugated (Sigma) were applied and incubated for 1 hours, and then washed three times with PBS-T for 10 min. SuperSignal West Femoto Maximum Sensitivity substrate (Thermo) was used for detecting the signal. Image J (NCBI) was used for quantifying the western hybridization signal.

DNA content analysis

Single cell suspensions were collected as described above and fixed with cold absolute ethanol for at least 1 hours at 4°C. The fixed cells were then incubated with 0.5 ml propidium iodide (PI) staining solution (0.1% Triton X-100/PBS, 0.2 mg/ml RNaseA, 0.02 mg/ml PI) for 3 hours at 4°C and subjected to FACS analysis.

Teratoma formation

Approximately 1×10^7 ES cells with undifferentiated morphology were injected into the rear leg muscles or testes of 5-8 week old NOD-SCID mice. Teratomas were allowed to develop for 10-12 weeks after injection before being excised and fixed with 4% paraformaldehyde overnight at 4°C. The fixed teratomas were then transferred to 30% sucrose for 48 hours before embedding for cryosectioning. Samples were cut at 10-20 μm in thickness and transferred to Poly-D Lysine-coated slides. Sections were subject to histochemistry staining. All the animal experiments were approved by the Animal Care and Use Committee of Academia Sinica and performed in accordance with the Institutional Animal Care and Use Committee guidelines of Academia Sinica.

Antibodies

Primary antibodies used for IF were as follows: STELLA (MAB2566), VASA (AF2030), SOX17 (AF1924), T (Brachyury) (AF2085), EpCAM (AF960) (R&D); DAZL (MCA2336) (Serotec); PAX6 (DSHB), NESTIN (MAB353), TRA 1-60 (MAB4360), SOX1 (AB15766), SSEA4 (MAB4304) (Millipore); GDF9 (sc-7407), OCT4 (sc-5279) (Santa Cruz); GFP (ab13970) and NANOG (ab21624) (Abcam); Alpha-actinin (A7811) (Sigma). Secondary antibodies used were goat anti-mouse cy3, goat anti-rabbit cy3 (Jackson Laboratories), donkey anti-goat 350, 594, donkey anti-rabbit 594, donkey anti-mouse 488 and (Molecular Probe, Invitrogen). For flow cytometry, APC mouse anti-Human CD326 (EpCAM) (324208) and PE anti-mose/human SSEA-1 (MC-480) (Biolegend) were applied.

References

1. Gerrard, L., Zhao, D., Clark, A. J., and Cui, W. (2005) *Stem Cells* **23**, 124-133
2. Dull, T., Zufferey, R., Kelly, M., Mandel, R. J., Nguyen, M., Trono, D., and Naldini, L. (1998) *J Virol* **72**, 8463-8471
3. Gropp, M., Itsykson, P., Singer, O., Ben-Hur, T., Reinhartz, E., Galun, E., and Reubinoff, B. E. (2003) *Mol Ther* **7**, 281-287
4. Nicholas, C. R., Haston, K. M., Grewall, A. K., Longacre, T. A., and Reijo Pera, R. A. (2009) *Hum Mol Genet* **18**, 4376-4389
5. Chen, H. F., Chuang, C. Y., Lee, W. C., Huang, H. P., Wu, H. C., Ho, H. N., Chen, Y. J., and Kuo, H. C. (2011) *Stem Cell Rev* **7**, 722-735

Supplementary Table 1: Primer sets used for Quantitative PCR

Primer Name	Sequence (5′ to 3′)
OCT4-F	TGGGCTCGAGAAGGATGTG
OCT4-R	CTTGATCGCTTGCCCTTCTG
NANOG-F	CAGCAGATGCAAGAACTCTCCA
NANOG-R	CATTGCTATTCTTCGGCCAGT
STELLA-F	CGGGCTACCTGGTAGCAATTT
STELLA-R	TCCAGGATCTTTGAAACGTAGCA
DAZL-F	ACAGCAGCCCCAGAAGTG
DAZL-R	TGGAGTTTCAGGATTTGCAGT
VASA-F	TCTTCCTTCTACCATTGATGAATATGTT
VASA-R	CTGCCAGTATTCCCACAACGA
SYCP3-F	ATGGAAACTCAGCAGCAAGAGATAG
SYCP3-R	AAGAGTCATCAGAATAACATGGATTGAA
SOX17-F	TTCGTGTGCAAGCCTGAGAT
SOX17-R	GTGTGTAACACTGCTTCTGGCC
MIXL1-F	AACGAAATGTCTGAAGCCCCA
MIXL1-R	TCCTCCCATGAGTCCAGCTTT
BRACHYURY(T)-F	CCAATGAGATGATCGTGACCA
BRACHYURY(T)-R	ATTCCCCGTTCACGTACTTCC
GSC-F	CGGAGAAGTGGAACAAGACGT
GSC-R	GCGTGTGCAAGAAAGTAGCATC
PAX6-F	GATAACATACCAAGCGTGTCATCAATA
PAX6-R	TGCGCCCATCTGTTGC
SOX1-F	TGGCATCTAGGTCTTGGCTCA
SOX1-R	GCACGAAGCACCTGCAATAAG
GAPDH-F	CGGGAAACTGTGGCGTGATG
GAPDH-R	TG TGGAGGAGTGGGTGTCGCTGTT

Supplementary Table 2 (A): Summary of the correlation of percentage of $OCT4^{+}/EpCAM^{+}$ cells and various cytokines/growth factors treatments on differentiating OCT4-EGFP hESC line H1. **(B):** Summary of the correlation of percentage of $OCT4^{+}/EpCAM^{+}$ cells and various cytokines/growth factors treatments on differentiating OCT4-EGFP hESC line H9.

(A)

	FBS	W	B4	B7+8b	B4+7+8b	B4+W	B4+W+C+D	2-step
FBS	-	0.828484	0.063343	0.075892	0.611814	0.004879	0.017406	0.004081
W	0.828484	-	0.483446	0.00219	0.788155	0.002038	0.008590	0.007316
B4	0.063343	0.483446	-	0.064546	0.696641	0.005768	0.025558	0.002917
B7+8b	0.075891	0.002189	0.064546	-	0.095287	0.002048	0.006312	0.005469
B4+7+8b	0.611813	0.788154	0.095287	0.095287	-	0.007892	0.011631	0.014283
B4+W	0.004879	0.002037	0.005768	0.002048	0.007892	-	0.028290	0.038792
B4+W+C+D	0.017406	0.008590	0.025558	0.006312	0.011632	0.028290	-	0.532499
2-step	0.004081	0.007316	0.002917	0.005469	0.014283	0.038792	0.532498	-

(B)

	FBS	W	B4	B7+8b	B4+7+8b	B4+W	B4+W+C+D	2-step
FBS	-	0.026838	0.277030	0.013367	0.160588	0.026888	0.178475	0.003127
W	0.026838	-	0.151706	0.032492	0.561187	0.026073	0.138814	0.004361
B4	0.277030	0.151706	-	0.011386	0.750226	0.032642	0.140979	0.007882
B7+8b	0.013367	0.032492	0.011386	-	0.029248	0.024065	0.085724	0.004118
B4+7+8b	0.160588	0.561187	0.750226	0.029248	-	0.025179	0.133773	0.002823
B4+W	0.026888	0.026073	0.032642	0.024065	0.025179	-	0.253391	0.122089
B4+W+C+D	0.178475	0.138814	0.140979	0.085724	0.133773	0.253391	-	0.444217
2-step	0.003127	0.004361	0.007882	0.004118	0.002823	0.122089	0.444217	-

Supplementary Figure Legends:

Supplementary Figure 1: Generation of OCT4-EGFP reporter human ES cells. **(A)** Schematic representation of the OCT4 promoter-EGFP lentiviral vector. **(B)** The pluripotent markers, NANOG, OCT4 and Tra-1-60 stained positively in OCT4-EGFP reporter hESCs. Scale bars = 100 μm. **(C)** H&E stain revealed that teratomas derived from OCT4-EGFP hESCs contained ectodermal, mesodermal, and endodermal cell types. **(D)** The H9 OCT4-EGFP derived neural progenitor and cardiac like cells stained positively for SOX1 and Alpha-actinin, respectively; but negatively for EGFP. Scale bars = 100 μm.

Supplementary Figure 2: Correlation between OCT4-EGFP transgene and endogenous OCT4 expression in undifferentiated and differentiated OCT4-EGFP hESCs. **(A)** FACS analysis of the OCT4 antibody correlated with OCT4-EGFP$^+$ cells at ESC stage and IVD day 20. The average percentage of each group of cells was indicated in the corner. The parental H9 hESCs and their differentiated cells were stained with control isotype antibody for gating determination. **(B)** The percentage of OCT4 positive staining cells within total OCT4-EGFP$^+$ cells at ESC stage and IVD day 20 is shown in the histogram. All FACS data are presented as average ± SD of three experiments. **(C)** OCT4-EGFP$^+$ cells isolated from differentiating hESCs were not able to give rise to AP positive ESC colonies in the ESC culture condition. (**Left panel**) 1×10^5 of the OCT4-EGFP hESCs were dissociated into single cells by trypsin digestion and seeded onto MEF in ESC culture medium. AP positive ESC colonies emerged 6 days post cell seeding (3 experiments). (**Right panel**) An equal number of FACS sorted OCT4-EGFP$^+$ cells at IVD day 12 was cultured under the same culture conditions for 14 days but could not form AP positive ESC colonies (3 experiments).

Supplementary Figure 3: IF analysis for OCT4-EGFP and germ cell markers, VASA, STELLA, DAZL from figure 2C. Scale bars = 100 μm.

Supplementary Figure 4: Enrichment of meiotic cell population by selection of OCT4-EGFP differentiated cells formed *in vitro*. **(A-F)** Differentiated H9 OCT4-EGFP$^+$ cells at IVD day 20-30 demonstrate partially elongated immunofluorescence staining for meiotic markers SYCP3 (A-D) and SYCP1 (E-F). **(G-H)** Immunofluorescence staining of complete SYCP3 and SYCP1 chromosomal alignment shown in germ cells from mouse neonatal ovaries. Scale Bars = 20 μm. **(I)** Bar chart showing percentage of SYCP3$^+$ meiotic cells in the cell population manually selected or un-selected by OCT4-EGFP fluorescence. A total of 100 meiotic spreads were counted in each categorized sample.

Supplementary Figure 5: Expression of EpCAM in pluripotent and differentiated H9 and NTU1 hESCs. **(A)** Immunofluorescence staining revealed that EpCAM was colocalized with pluripotent

markers OCT4, SSEA4, and TRA-1-60 in undifferentiated hESCs. Scale Bars =100 μm. (B) Percentage of EpCAM positive cells decreased along with the progression of *in vitro* differentiation.

Supplementary Figure 6: Comparison of expression level of germ cell enriched genes for cells treated with BMP4 and WNT3A at different time points during *in vitro* differentiation from H9 OCT4-EGFP ESCs. BMP4 and W NT3A were added at IVD day 0 (the first day of hESC differentiation), day 2, day 5 and day 10. Q-RT-PCR was performed at IVD day 15 to analysis the expression of STELLA, DAZL and VASA. The relative expression of each gene was first normalized to GAPDH and then represented as fold change relative to those of the control group (cells in differentiation medium without BMP4+WNT3A) (1 fold). Data represented as mean ± SD (n = 3).

Supplementary Figure 7: Percentages of OCT4-EGFP $^{+}$/EpCAM $^{+}$ cells from IVD day 15 of H1and H9 OCT4-EGFP ESCs in 15% FBS differentiation medium supplemented with BMP4/WNT3A and their antagonist. B4+W: Differentiation medium supplemented with BMP4 and WNT3A. B4+W/NOGGIN: Cells cultured in differentiation medium were supplemented with BMP4 and WNT3A for 4 days, then switched to medium containing NOGGIN. B4+W/DKK1: Cells cultured in differentiation medium were supplemented with BMP4 and WNT3A for 4 days, then switched to medium containing DKK1. Data represented as mean ± SD (n = 3); * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The p values were calculated between each corresponding cell line.

Fig. S1

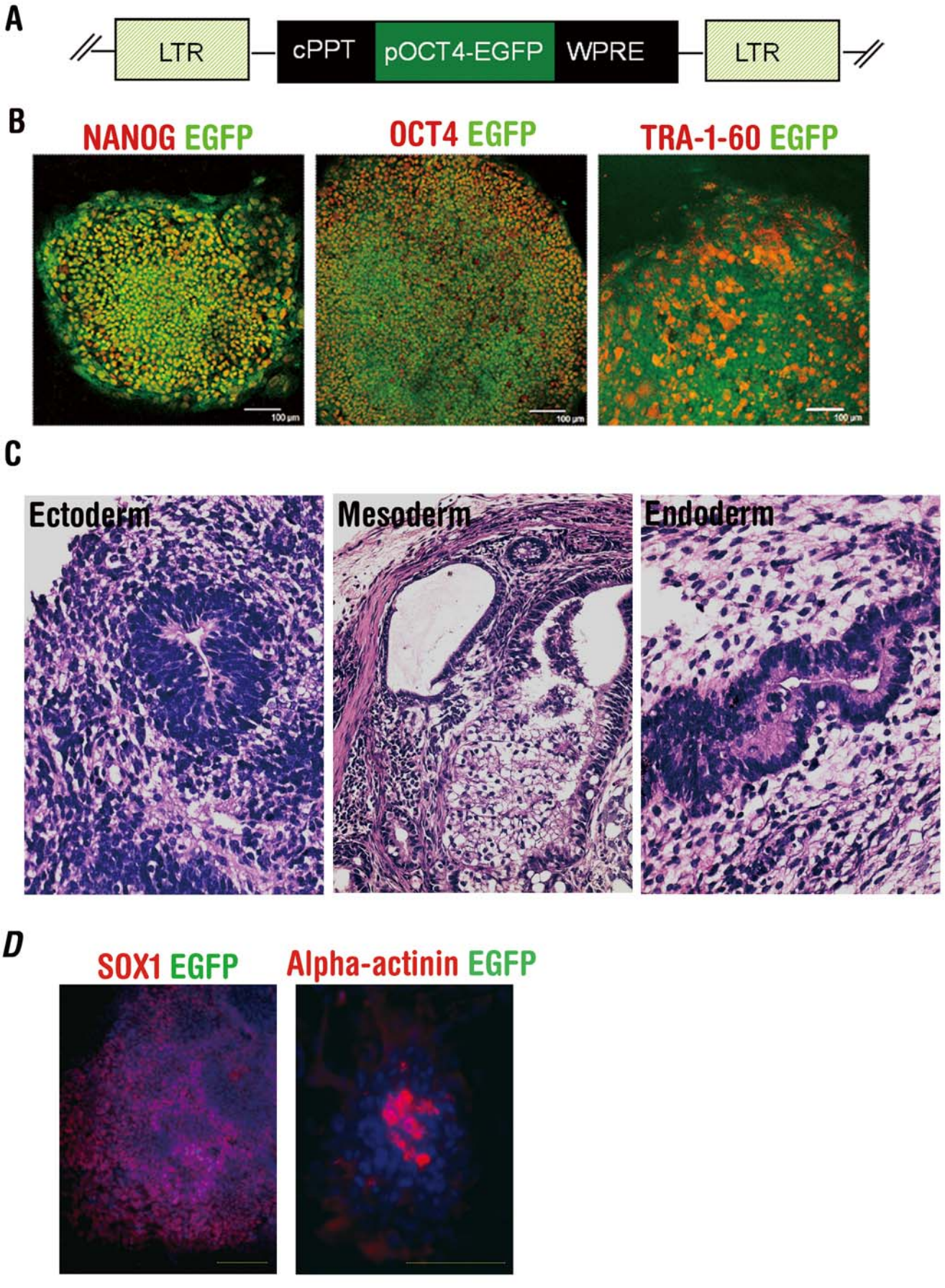

A
LTR
cPPT
pOCT4-EGFP
WPRE
LTR
B
NANOG EGFP
OCT4 EGFP
TRA-1-60 EGFP
100 μm
100 μm
100 μm
C
Ectoderm
Mesoderm
Endoderm
D
SOX1 EGFP
Alpha-actinin EGFP

Figure S2

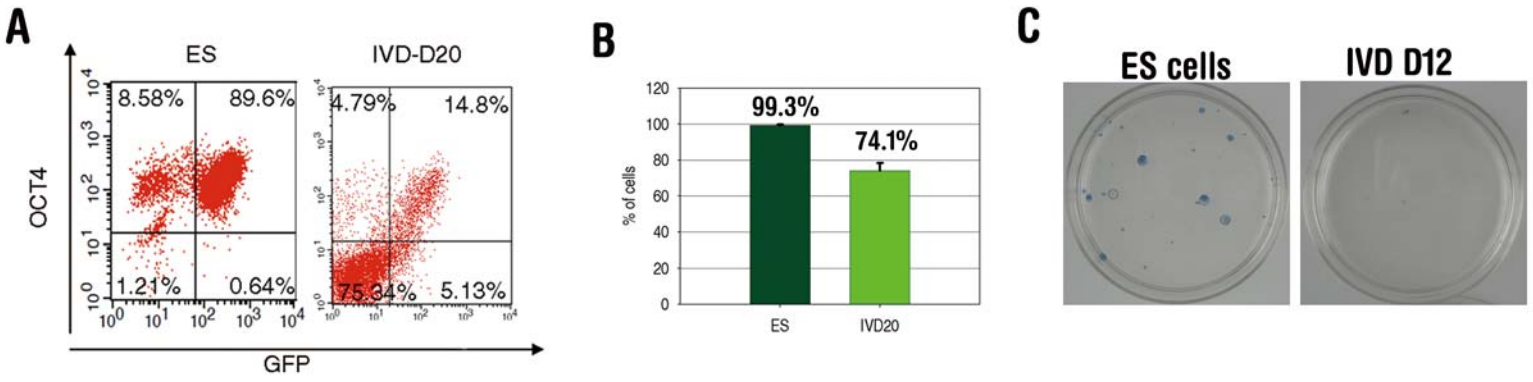

A
ES
IVD-D20
8.58%
89.6%
1.21%
0.64%
4.79%
14.8%
75.34%
5.13%
OCT4
GFP
B
% of cells
99.3%
74.1%
ES
IVD20
C
ES cells
IVD D12

Figure S3

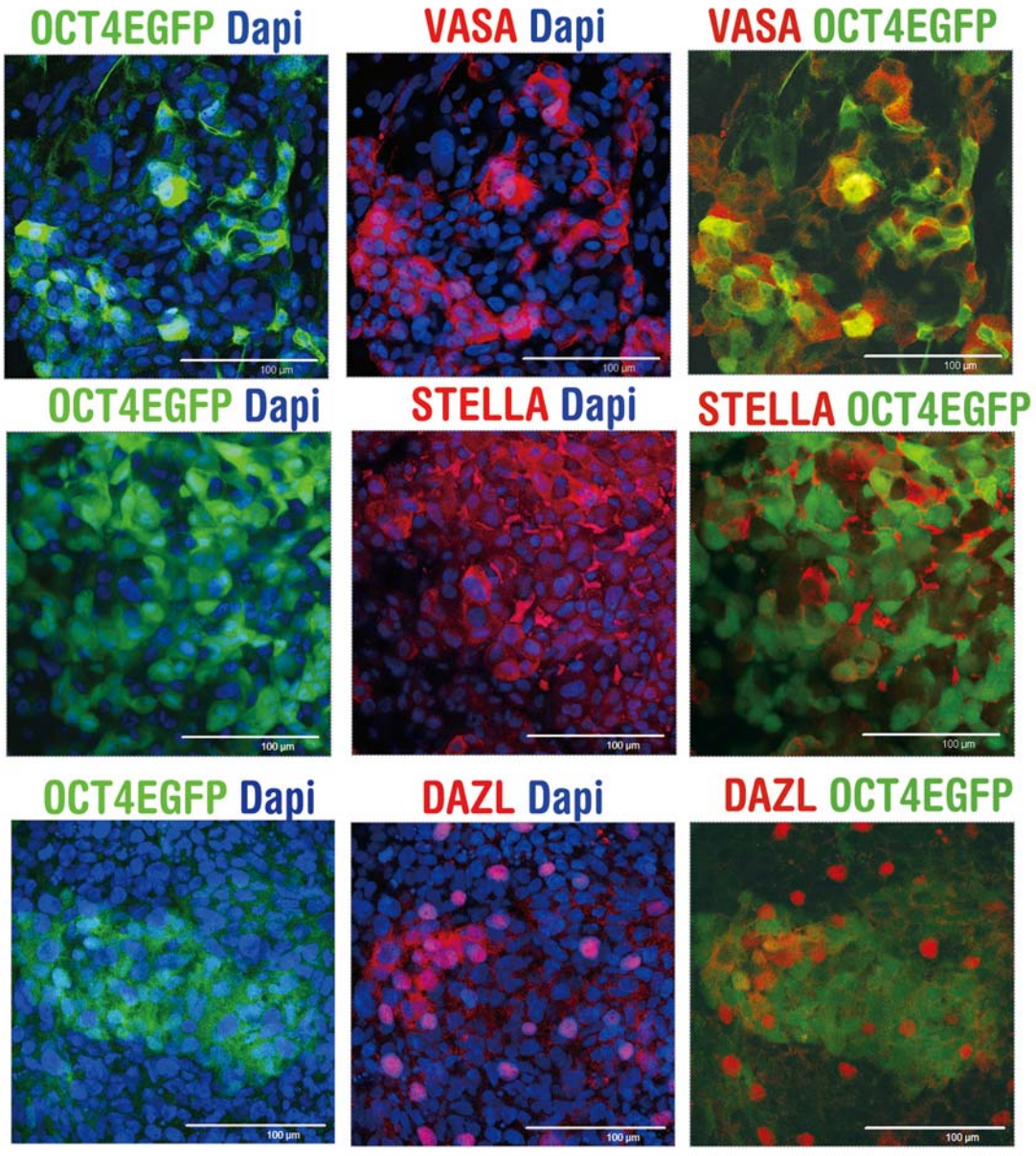

OCT4EGFP Dapi
VASA Dapi
VASA OCT4EGFP
OCT4EGFP Dapi
STELLA Dapi
STELLA OCT4EGFP
OCT4EGFP Dapi
DAZL Dapi
DAZL OCT4EGFP
100 µm

Figure S4

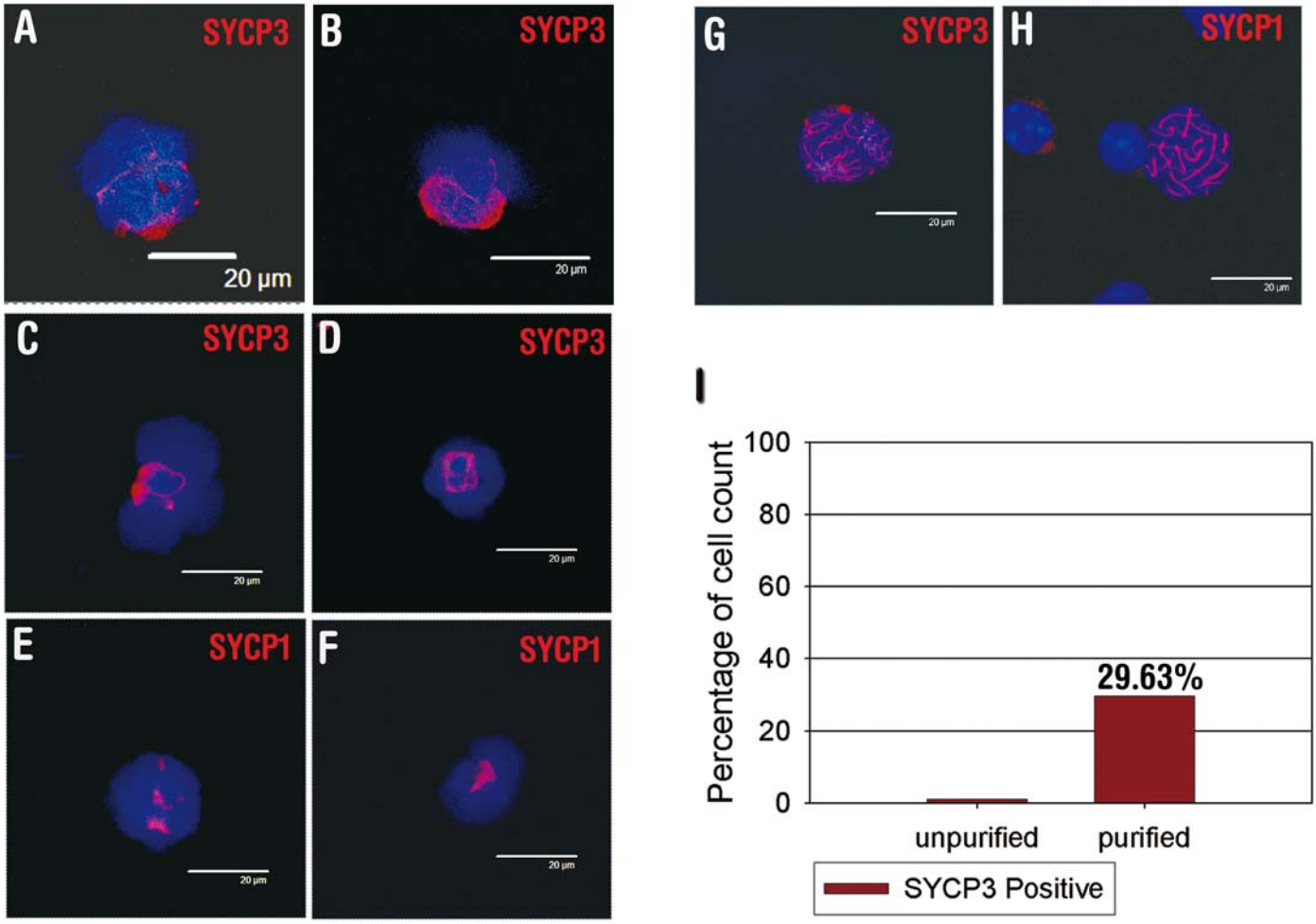
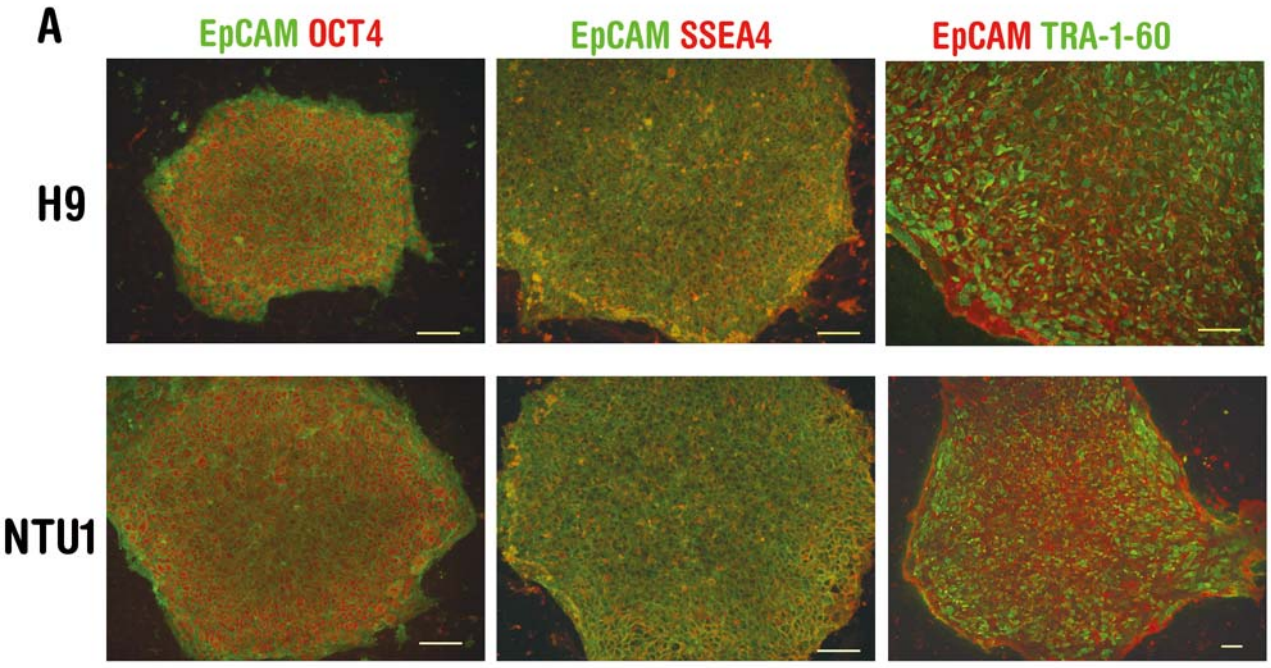

A
SYCP3
20 μm
B
SYCP3
20 μm
C
SYCP3
20 μm
D
SYCP3
20 μm
E
SYCP1
20 μm
F
SYCP1
20 μm
G
SYCP3
20 μm
H
SYCP1
20 μm
I
Percentage of cell count
100
80
60
40
20
0
29.63%
unpurified
purified
SYCP3 Positive

Figure S5


A
EpCAM OCT4
EpCAM SSEA4
EpCAM TRA-1-60
H9
NTU1

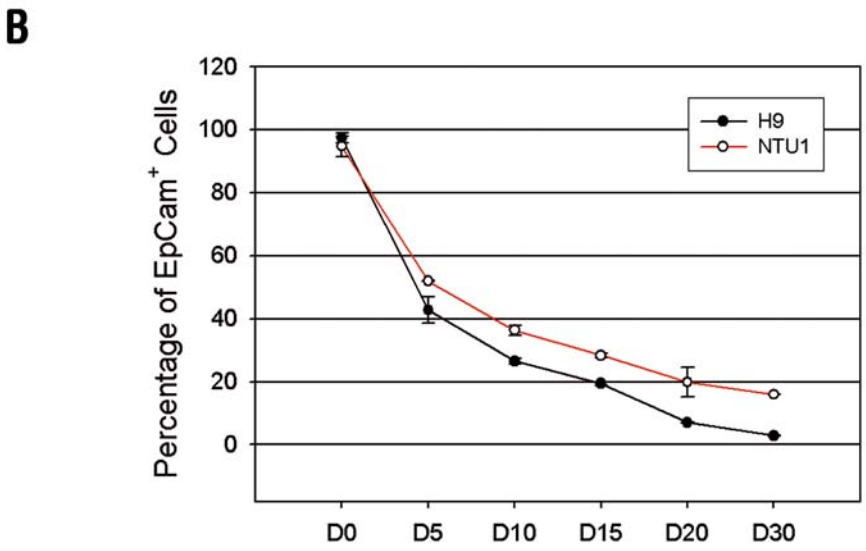
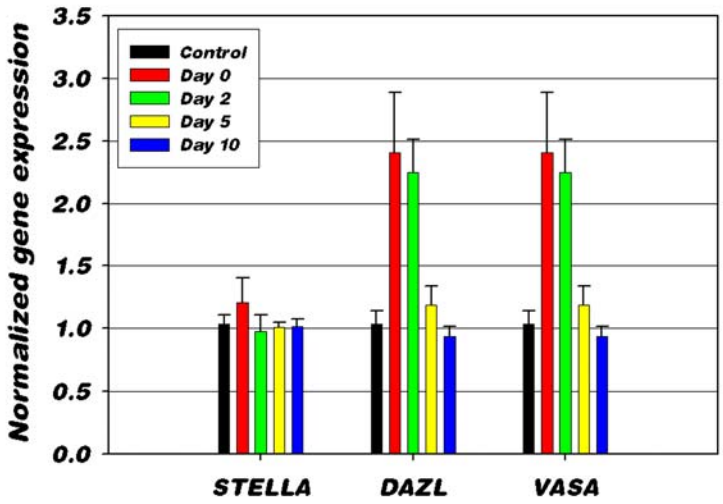
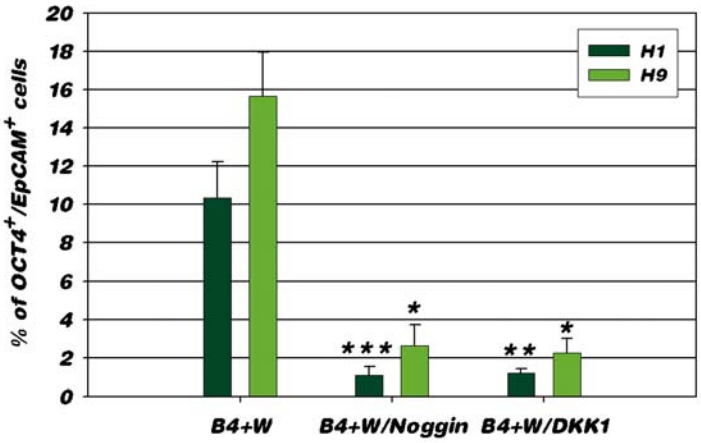

B
Percentage of EpCam+ Cells
120
100
80
60
40
20
0
H9
NTU1
D0
D5
D10
D15
D20
D30

Figure S6


Normalized gene expression
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Control
Day 0
Day 2
Day 5
Day 10
STELLA
DAZL
VASA

Fgure S7


20
18
16
14
12
10
8
6
4
2
0
% of OCT4+/EpCAM+ cells
H1
H9
***
*
**
*
B4+W
B4+W/Noggin
B4+W/DKK1