

Sox2 expression effects on direct reprogramming efficiency as determined by alternative somatic cell fate

Shinpei Yamaguchi^{a,b}, Kunio Hirano^a, Shogo Nagata^a and Takashi Tada^{a,b*}

^aStem Cell Engineering, Institute for Frontier Medical Sciences, Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507 Japan ^bJST, CREST, 4-1-8 Honcho Kawaguchi, Saitama, 332-0012, Japan

Key words: Induced pluripotent stem cell, Transcription factor, Sox2, Reprogramming

RUNNING TITLE: Negative effects of excess Sox2 on reprogramming

*Corresponding author: Laboratory of Stem Cell Engineering, Stem Cell Research Center, Institute for Frontier Medical Sciences, Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507 Japan
Tel: +81-75-751-4102; Fax: +81-75-751-4102; E-mail: ttada@frontier.kyoto-u.ac.jp

Abstract

Induced pluripotent stem cells (iPSCs) are generated by directly reprogramming somatic cells by forcing them to express the exogenous transcription factors, *Oct4*, *Sox2*, *Klf4* and *c-Myc* (OSKM). These cells could potentially be used in clinical applications and basic research. Here, we explored the molecular role of *Sox2* by generating iPSCs that expressed *Sox2* at various levels. Low *Sox2* (LS) expression increased the efficiency of generating partially reprogrammed iPSCs in combination with OKM. Notably, we detected a significant increase in the number of fully reprogrammed iPSCs with three factors of OK and LS. LS expression was linked with the reduced expression of ectoderm and mesoderm marker genes. This indicates that cell differentiation into the ectoderm and mesoderm lineages was impeded during reprogramming. The quality of the iPSCs that was generated by using OK and LS was comparable to that of iPSCs that were produced *via* conventional OSK as seen by pluripotent marker gene expression and chimera formation. We conclude that *Sox2* plays a crucial role in a dose-dependent manner in directly reprogramming of somatic cells to iPSCs.

Introduction

Reprogramming somatic cells into pluripotent embryonic stem cell (ESC)-like cells is achieved by the transduction using a defined set of transcription factors: *Oct4* (*Pou5f1*), *Sox2*, *Klf4* and *c-Myc* (OSKM) [1,2]. Induced pluripotent stem cells (iPSCs) can also be established in the absence of a transduced *c-Myc* (OSK) [3] or by using other combinations of factors [4,5]. Although the generating iPSCs is technically simple and ethically acceptable, transgene integration is problematic for clinical applications. Recently, several groups have reported on the generating iPSCs without having to use viral vectors [6,7]. However, a major issue with their approaches is the low efficiency of producing iPSCs. Uncovering the function of reprogramming factors would be helpful for improving the efficiency of direct reprogramming.

Two reprogramming factors, *Oct4* and *Sox2*, are core regulatory factors that function cooperatively with *Nanog* in both human and mouse ESCs [8,9]. *Oct4* has been studied at length and is recognized as being essential for the early development and maintenance of pluripotency. *Oct4*-deficient embryos fail to form an inner cell mass [10]. Reduced *Oct4* expression in ESCs causes them to differentiate into the trophectoderm, while doubly high expression induces differentiation into the primitive endoderm and mesoderm [11]. *Sox2*, which is a high-mobility group DNA-binding domain transcription factor, is also essential for early embryogenesis in mice [12]. The twenty genes that comprise the mouse Sox-family are classified into ten subgroups,

whereby *Sox2* is categorized into group B1 [13]. *Sox2* could be substituted for using *Sox1* or *Sox3* to generate iPSCs [3], even if low expression of these two genes occurs in ESCs. *Sox2* is expressed in the developing central nervous system and maintains the properties of neural precursor cells [14]. *Oct4* and *Sox2* cooperatively bind to and activate the transcription of genes that bear Octamer and Sox elements, such as *Oct4*, *Sox2*, *Nanog*, *Fgf4*, *Utf1* and *Lefty1* [15-20]. Similar to *Oct4*, *Sox2* depletion and overexpression promotes differentiation. A two-fold or less increase in *Sox2* reduces the transcription of *Oct4/Sox2* target genes in ESCs. This results in differentiation into various lineages, including the ectoderm and mesoderm [21-23], while the knockdown of *Sox2* induces differentiation into endoderm and trophectoderm lineages [23-25]. Therefore, precisely regulating the expression of the transcription factors is crucial to maintain the pluripotent state of ESCs. However, the effects of quantitative changes in reprogramming factors when inducing pluripotent stem cells remain poorly understood.

Here, we examined the suppressive effects of excessive *Sox2* on direct reprogramming. Decreased *Sox2* virus led to a marked increase in the number of *Oct4*-GFP-positive cells. However, resulting cell fates were different depending on whether the transduction occurred with or without *c-Myc* (OSKM and OSK). For OKM and reduced *Sox2*, the majority of the GFP-positive colonies was trapped in a partially reprogrammed state, while for OK and reduced *Sox2*, most of the cells turned into fully reprogrammed iPSCs. A significant increase in the number of differentiation

markers, which depended on the level of *Sox2*, was indicative that excessive *Sox2* expression promotes differentiation into the mesoderm or ectoderm in parallel with reprogramming. Thus, our data is supportive that the appropriate expression of *Sox2* is critical for efficiently generating iPSCs.

Results

Expression control through the amount of virus

To investigate the effect of *Sox2* expression on the reprogramming process, mouse embryonic fibroblasts (MEFs) were infected with various amounts of *Sox2* virus. We used two lines of experiments, one experiment which utilized the mock virus (the empty retroviral pMX vector), and the other which did not, to understand the efficiency of viral infection. The results of quantitative Western blotting analysis were indicative that the amount of Sox2 protein corresponded to the amount of Sox2 virus that was present in OSKM and OSK during viral infection (Fig. 1). Interestingly, no obvious difference in Sox2 expression was detected between the experiment that entailed using the mock virus and that which did not. Furthermore, an increase in Sox2 to 200% resulted in a decrease in Oct4 (Fig. 1). This is suggestive that excessive viral Oct4 and Sox2 proteins interacted with one another during transcription, translation or degradation regulation. It is possible that this Oct4 and Sox2 interaction may conceal the difference in the total amount of virus between the experiments that used the mock

virus compared to which did not. Therefore, we performed OSKM and OSK infection without mock.

Reprogramming by OKM plus low Sox2

MEFs from *Oct4-GFP/Neo-LacZ* (*Rosa26*) double-transgenic mouse embryos were prepared to evaluate the efficiency of reprogramming. GFP expression is driven by the *Oct4* promoter [26], while LacZ is ubiquitously expressed in the double-transgenic cell [27]. MEFs were split into several groups and infected with OSKM at various concentrations of *Sox2* virus. Kinetic analysis of GFP-positive cells on days 6, 8 and 10 after infection was indicative that the increased reprogramming efficiency was most effective at the 20% *Sox2* (LS) (Fig. 2A, B). The number of GFP-positive cells had increased about five times for OKM and LS (OKM+LS) when compared to conventional OSKM by using fluorescence-activated cell sorting (FACS) analysis (Supplementary Fig. 2A). For 10% and 200% *Sox2*, the reprogramming efficiency decreased compared to conventional OSKM on day 10 (Fig. 2B). To examine whether the increase of GFP-positive cells improved the generation efficiency to fully reprogrammed iPSCs, which were defined by the stable GFP expression and maintenance of undifferentiated state, each GFP-positive clone that was picked up was cultured for a month. Notably, fifteen fully reprogrammed clones were established out of seventy-two clones (20.8%) for conventional OSKM, while nine out of ninety-two

clones (9.4%) were established for OKM+LS (Fig. 2C). The OKM+LS infection resulted in a prominent increase in GFP-positive partially reprogrammed iPSCs. However, there was no significant increase in the number of fully reprogrammed iPSCs. In total, the reprogramming efficiency for MEFs being changed into fully reprogrammed iPSCs was not significantly improved.

The expression of lineage and pluripotent marker genes was then examined using quantitative reverse transcription-PCR (qPCR) to characterize differentiating cells and GFP-positive cells that were generated by reprogramming with OKM+LS. For pluripotent marker genes, *Fgf4*, *Utf1*, *Rex1*, *Eras* and *Sall4* expression was high for OKM+LS rather than for OSKM (Supplementary Fig. 1A-C). For the lineage marker genes, mesoderm marker expression including that of *Flk1*, *Gremlin* and *Myh2*, and ectoderm marker expression including that of *Nestin*, *Sox13*, *Sox21* and *CryM* was significantly low for OKM+LS compared to OSKM on day 3 (Supplementary Fig. 1D). However, the expression of endoderm and trophectoderm makers was not. A similar situation was noted for on days 7 and 12 after infection (Supplementary Fig. 1E). Notably, partially reprogrammed iPSCs in OKM+LS (Supplementary Fig. 2B) were characterized as having low expression of *Sox2*, *Nanog* and *Fgf4*, and high *c-Myc*, *Sox17* and *Desmin* expression (Supplementary Fig. 2C). Taking these findings into consideration, excessive *Sox2* expression for conventional OSKM induces differentiation of reprogramming cells into the ectoderm and mesoderm lineages, at

least in part, in a *Sox2*-dependent manner.

Three fully reprogrammed OKM+LS iPSC lines (#1~#3) that were characterized using *Oct4*-GFP expression (Supplementary Fig.3A), pluripotent marker gene expression (Supplementary Fig. 3B), and pluripotent marker protein expression (Supplementary Fig. 3C) were pluripotent. This was revealed by contribution to the three primary layers and germ cells of an E12.5 chimeric embryo, for which two chimeras out of five embryos were obtained with the #2 iPSC line (Supplementary Fig. 3D).

Reprogramming by OK plus low Sox2

The improved reprogramming efficiency for changing MEFs into fully reprogrammed iPSCs was unclear with OKM+LS. Therefore, we examined the reprogramming efficiency in OK and various concentrations of *Sox2* to exclude the effect of *c-Myc*. The number of GFP-positive colonies with *Sox2* at 200%, 100%, 50%, 20% and 10% was counted (Fig. 2D). Excessive (200%) and extremely low (10%) *Sox2* resulted in no significant improvement in the reprogramming efficiency. For OK+20% *Sox2* (LS), the number of GFP-positive colonies was 3.3 times higher than for OSK on day 10 of infection. Notably, a similar tendency was detected consecutively on days 12, 14 and 16 (Fig. 2D). The rate for establishing fully reprogrammed iPSCs for OK+LS at 57% was comparable to that for conventional OSK at 52% (Fig. 2E). No increase in the

number of partially reprogrammed cells was observed for OK+LS (data not shown).

Taking these findings into consideration together, the net generation efficiency of fully reprogrammed iPSCs for OK+LS infection was approximately three times higher than for conventional OSK infection.

Alternative cell fates induced by OK plus LS

Pluripotent and lineage marker gene expression was analyzed using qPCR on day 3 after infection to explore the cell fate on reprogramming by using OK+LS. Among the pluripotent marker genes, expression of the reprogramming factors, total *Oct4* and *Klf4*, was markedly high, while that for total *Sox2* was low for OK+LS (Fig. 3A). The *Oct4/Sox2* target genes, *Nanog*, *Lefty1*, *Fgf4* and *Utf1* were highly expressed for OK+LS (Fig. 3B). In particular, *Fgf4* expression was about seven times higher for OK+LS than for OSK. For other pluripotent markers, *Rex1*, *Eras*, *Sall4* and *Ronin* were highly expressed for OK+LS (Fig. 3C).

Among the lineage marker gene expression for OK+LS was roughly similar to that for OSK (Fig. 3D). Specifically, the ectoderm markers, *Sox13*, *Sox21* and *CryM*, and the mesoderm marker, *Myh2* were significantly underexpressed. Endoderm marker, *Sox17* was highly expressed for OK+LS compared to OSK. Next, temporal expression was analyzed using qPCR on days 3, 7 and 16 after infection. The expression of *Myh2*, *Sox21* and *CryM* was extremely low for OK+LS compared to OSK on days 3, 7 and 16

after infection. However, no significant changes were detected for *Desmin* and *Gremlin* (Fig. 3E). These data are supportive that induced cell differentiation into the ectoderm and mesoderm lineages was impeded for OK+LS. Consequently, the decreased generation efficiency of partial iPSCs contributed to the increased generation efficiency of fully reprogrammed iPSCs for OK+LS.

Quality of iPSCs generated through OK plus low Sox2

The quality of fully reprogrammed iPSCs that were obtained using OK+LS was tested. The analyses included examining the following: colony morphology; uniform *Oct4*-GFP expression (Fig. 4A); expression of pluripotent marker gene expression: *Nanog*, *Rex1*, *Stella*, *Lefty1*, *Eras*, *Gdf3*, *Fgf4*, *Utf1* and *Sall4* RT-PCR analyses (Fig. 4B); and immunocytochemical analyses of *Nanog*, *Oct4*, *Sox2*, *SSEA-1*, *Stella*, and E-cadherin (Fig. 4C) in all #1~3 OK+LS-iPSC lines. The resulting data were indicative that the iPSCs with OK+LS were comparable to iPSCs with conventional OSK *in vitro*.

Three LacZ-positive iPSC lines, which have the normal karyotype $2n=40$ and *Gtl2* expression [28], were selected from twenty-two iPSC cell lines that were examined to determine pluripotency of the OK+LS-iPSCs *in vivo*. They were then microinjected into blastocysts. A contribution to the three germ layers of morphologically normal E15.5 or E16.5 chimeric embryos was detected by using X-galactosidase (Gal) staining for all three OK+LS-iPSC lines. One chimera out of four embryos for the

iPSC line #1, one out of eighteen for line #2, and seven out of eleven for line #3 were observed (Fig. 4D). X-gal-positive germ cells that were found in the testicular tubules of E16.5 chimeric embryos were indicative of the germ cell competence of OK+LS-iPSCs (Fig. 4D). Therefore, the OK+LS iPSCs were comparable to conventional iPSCs in terms of pluripotency.

Discussion

A simple method for decreasing *Sox2* virus improved the efficiency of *Oct4*-GFP-positive cells to emerge for OKM+LS and OK+LS. This is indicative of the attenuating effects on the reprogramming efficiency of overexpressed *Sox2* for conventional OKSM and OKS. The increased reprogramming efficiency for LS resulted from the decreased cell differentiation of MEFs into ectoderm and mesoderm lineages as observed by using qPCR analyses. The increased number of GFP-positive cells for OKM+LS was caused by the increase in partially reprogrammed iPSCs. Improved reprogramming efficiency of somatic cells into fully reprogrammed iPSCs was significant for OK+LS. These data are indicative that 1) the reprogramming efficiency of MEFs into iPSCs was significantly increased by low level *Sox2* expression in a dose-dependent manner; 2) improved reprogramming efficiency for reprogramming into iPSCs by using OK+LS was associated with impeding cell differentiation of MEFs into the ectoderm and mesoderm lineages; and 3) the

generation efficiency of partial iPSCs decreased through reprogramming with OK+LS.

Partial iPSCs were dominantly induced with OKM+LS but not with OK+LS. *c-Myc* plays a role in controlling the transcription activity of numerous genes that are related to dedifferentiation and proliferation [29,30]. The strong transforming activity of *c-Myc* induces clonal expansion of the reprogramming cells for OKM+LS (Fig. 5A). Thus, partial iPSC generation that induced with *c-Myc* masked the positive effect of LS. In the case of the iPSC induction without *c-Myc*, the improved effect of LS on the reprogramming efficiency became more apparent by the excluding the partial iPSCs (Fig. 5B).

In mice, neural stem cells (NSCs) that express endogenous *Sox2* are successfully reprogrammed into iPSCs only by using OKM infection [31,32]. Interestingly, the reprogramming efficiency of NSCs was decreased by addition of *Sox2* into OKM, such that excessive *Sox2* expression interfered with reprogramming. In humans, reprogramming efficiency is improved by increasing the amount of *Oct4*-expressing lentivirus three-fold [33]. Furthermore, increased *Sox2* virus down-regulated the efficiency of colony formation, while decreased *Sox2* enhanced this efficiency. Exogenous *Sox2* expression is essential for acquiring pluripotency among somatic cells through reprogramming in humans and the mouse [1,34]. These data are consistent with our findings. Thus, an appropriate level of *Sox2* expression is important to induce the efficient reprogramming.

Sox2 plays an essential role in mouse development. For example, it is involved in patterning of the anterior foregut, maintaining postnatal NSCs, neurogenesis in the brain, and the eye formation [14,35,36]. Interestingly, *Sox2* functions in a dose-dependent manner *in vivo*. In the endoderm, the region that highly expresses *Sox2* differentiates into the pharynx and esophagus. In the region where it is underexpressed, it becomes the trachea [35]. In the ectoderm, the heterozygotic loss of *Sox2* does not induced an obvious phenotype. However, a <40% reduction in *Sox2* induces aberrant neural progenitor differentiation, which results in microphthalmia [36]. Thus, *Sox2* plays a key role in differentiation into the ectoderm, mesoderm and endoderm lineages in a dose-dependent manner *in vivo*. It seems that dose-dependent effects of *Sox2* are common for *in vivo* development and *in vitro* reprogramming. However, a few changes in endoderm-related gene expression were detected for OKM+LS and OK+LS. In ESCs, forced up-regulation of *Sox2* results in no significant changes in the expression of endoderm-related genes [22]. This is suggestive that a dose-dependent effect of *Sox2* on the induction of endoderm lineage differentiation may not be obvious for direct reprogramming *in vitro*.

The molecular mechanisms that are involved in inducing of alternative cell fates by increasing and reducing the amount of *Sox2* during reprogramming remain unclear. An interesting finding was that *Oct4* changes its partner and targets genes in a dose-dependent manner in human ESCs [37]. Excessive expression of *Oct4* in human ESCs

leads to differentiation into a cardiac cells through switching of the binding element that is located to *Sox17* from the *Sox2* promoter. It is possible that switching protein interaction partners and binding affinity to regulatory elements of downstream genes, which changes cell fate, occurs in *Sox2* in a dose-dependent manner, similar to that of *Oct4*. The qPCR analyses that are discussed here were evaluated using RNA that was extracted from heterogeneous populations through reprogramming that was induced using defined factors. Further comparisons of expression profiles between colonies could be helpful for understanding the molecular events that underlie the dose-dependent effect of *Sox2*.

We conclude that the significantly increased generation efficiency of fully reprogrammed iPSCs *via* OK+LS resulted from efficient reprogramming into fully reprogramming iPSCs following a significant increase in *Oct4*-GFP-positive iPSCs. Optimizing the dose of reprogramming factors will be a key to realizing practical applications of recently developed integration-free and genetic modification-free reprogramming technologies.

Materials and methods

Cell culture

Mouse iPSCs or ESCs were maintained in DMEM/F12 (Dulbecco's modified Eagle's medium/Ham's F12, Wako Chemical, Osaka, Japan), which was supplemented with

15% fetal bovine serum (FBS; Gibco, Carlsbad, CA), 10^{-4} M 2-mercaptoethanol, and 400 units/ml recombinant LIF (Chemicon, Temecula, MA). The cells were maintained on MEFs feeder cells that had been inactivated with mitomycin C. To isolate double-transgenic MEFs, E12.5 embryos were minced with an 18-gauge needle after being obtained from GOD-18/delta PE/GFP (*Oct4*-GFP) transgenic females [26] that had been mated with 129/Rosa26 transgenic males [27]. The resulting cell suspension was plated and cultured with DMEM (Sigma Aldrich, St. Louis, MO) containing 10% FBS. MEFs were used during early passages.

iPSC induction from MEFs

Retroviral supernatants for *Oct4*, *Sox2*, *Klf4* and *c-Myc* were collected from PlatE packaging cells and stored at -80°C prior to being used for infection. Titers for *Oct4* and *Sox2* but not *Klf4* and *c-Myc* virus were determined by using immunocytochemical analyses of MEFs at 36 hours after infection. *Oct4* and *Sox2* virus were used at the same multiplicity of infection (MOI) of 3 for OSKM or OSK transduction. MEFs were seeded on gelatin-coated 12-well culture plates at 4×10^4 cells per well, and were infected with a mixture of stocked viral supernatant with $4 \mu\text{g/mL}$ polybrene (day 0) to determine the reprogramming efficiency. The medium was replaced with ESC medium on day 1. All GFP-positive colonies in each well were counted among days 4-12 for OSKM, and days 6-16 for OSK. Infected cells were reseeded on a feeder layer on day 4 and picked up on day 14 for OSKM or day 20 for OSK, to establish iPSCs.

Western blot hybridization

Whole cell lysate was extracted from MEFs on day 3 after infection. Total protein (5 μ g) was separated by using 10% SDS-polyacrylamide gel electrophoresis and transferred onto a nitrocellulose membrane (Millipore, Billerica, MA). It was subsequently reacted with anti-Oct4 (1:100 dilution), anti-Sox2 (1:2000), or anti-beta-actin (1:2000) antibodies at 4°C overnight. The membranes were then incubated with a secondary antibody (1/3000; GE Healthcare, Little Chalfont, Buckinghamshire, U.K.) for one hour and the signals were then detected by using the ECL Western blotting detection kit (GE Healthcare).

Immunocytochemistry

Cells that were fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS) for 15 minutes at room temperature were pre-treated with blocking solution at 4 °C overnight. The blocking solution was composed of PBS containing 3% BSA (Sigma), 2% skim milk (DIFCO, Detroit, MI), and 0.1% Triton X-100. The cells were then stained with fluorescence-conjugated secondary antibody (1/500; Invitrogen), following to immunoreaction with anti-Oct4 (1:50), anti-Nanog (1:1000), anti-Sox2 (1:1000), anti-Stella (1:500), anti-SSEA1 (1/1000), or anti-E-Cadherin (1/1000) at 4°C overnight.

Mouse chimeric embryos

OKM+LS- and OK+LS-iPSCs were microinjected into C57BL/6J or

(C57BL/6JxBDA)F2 blastocysts and then transferred into the uteruses of pseudo-pregnant ICR females. E12.5, E15.5 and E16.5 embryos were collected and stained according to X-Gal staining procedures [38].

Flow Cytometry

Cells were dissociated into single cells with 0.25% trypsin/ 0.4% EDTA/ PBS, washed once, re-suspended in PBS with 2% FBS, and analyzed using a FACS Vantage system (Becton Dickinson, Franklin Lakes, NJ).

Quantitative reverse transcription-PCR (qPCR)

Total RNA was extracted using TRIzol (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. DNase I (Roche Diagnostics, Basel, Switzerland)-treated RNA was reverse-transcribed by using random primers and Superscript III Reverse Transcriptase (Invitrogen). Amplification was performed using the Prism 7700 (Applied Biosystems, Foster City, CA) and Power SYBR Green PCR Master Mix (Applied Biosystems) according to the manufacturer's instructions with gene-specific primer sets. All reactions were carried out in duplicate and gene expression levels were normalized to *Gapdh*. Relative expression of each gene was quantified from threshold cycles for amplification using the $2^{\Delta\Delta Ct}$ method. The primer sets that were used are summarized in Supplementary Table 1.

Acknowledgements

We thank Dr S. Yamanaka for the anti-Sox2 antibody, and Dr. G. Kondoh and Miss H. Watanabe for the generation of chimeric embryos. This work was partly funded by grants from the Japan Society for the Promotion of Science, the Ministry of Education, Culture, Sports, Science and Technology, and the Japan Science and Technology Agency to Dr. T. Tada. Mr. K. Hirano is a research fellow at the Japan Society for the Promotion of Science.

References

- [1] K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors, *Cell*, 126 (2006) 663-676.
- [2] K. Hochedlinger, K. Plath, Epigenetic reprogramming and induced pluripotency, *Development*, 136 (2009) 509-523.
- [3] M. Nakagawa, M. Koyanagi, K. Tanabe, K. Takahashi, T. Ichisaka, T. Aoi, K. Okita, Y. Mochiduki, N. Takizawa, S. Yamanaka, Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts, *Nat. Biotechnol.*, 26 (2008) 101-106.

[4] B. Feng, J. Jiang, P. Kraus, J.H. Ng, J.C.D. Heng, Y.S. Chan, L.P. Yaw, W. Zhang, Y.H. Loh, J. Han, V.B. Vega, V. Cacheux-Rataboul, B. Lim, T. Lufkin, H.H. Ng, Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb, *Nat. Cell Biol.*, 11 (2009) 197-203.

[5] J.C.D. Heng, B. Feng, J. Han, J. Jiang, P. Kraus, J.H. Ng, Y.L. Orlov, M. Huss, L. Yang, T. Lufkin, B. Lim, H.H. Ng, The Nuclear Receptor Nr5a2 Can Replace Oct4 in the Reprogramming of Murine Somatic Cells to Pluripotent Cells, *Cell Stem Cell*, 6 (2010) 167-174

[6] H. Zhou, S. Wu, J. Y. Joo, S. Zhu, D.W. Han, T. Lin, S. Trauger, G. Bien, S. Yao, Y. Zhu, G. Siuzdak, H.R. Schöler, L. Duan, S. Ding, Generation of induced pluripotent stem cells using recombinant proteins, *Cell Stem Cell*, 4 (2009) 381-384.

[7] D. Kim, C.H. Kim, J.I. Moon, Y.-G. Chung, M.Y. Chang, B.S. Han, S. Ko, E. Yang, K. Y. Cha, R. Lanza, K.S. Kim, Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins, *Cell Stem Cell*, 4 (2009) 472-476.

[8] L.A. Boyer, T.I. Lee, M.F. Cole, S.E. Johnstone, S.S. Levine, J.P. Zucker, M.G.

Guenther, R.M. Kumar, H.L. Murray, R.G. Jenner, D.K. Gifford, D.A. Melton, R. Jaenisch, R.A. Young, Core transcriptional regulatory circuitry in human embryonic stem cells, *Cell*, 122 (2005) 947-956.

[9] Y.-H. Loh, Q. Wu, J.L. Chew, V.B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K. Y. Wong, K.W. Sung, C.W.H. Lee, X.D. Zhao, K.P. Chiu, L. Lipovich, V.A. Kuznetsov, P. Robson, L.W. Stanton, C.L. Wei, Y. Ruan, B. Lim, H.H. Ng, The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells, *Nat. Genet.*, 38 (2006) 431-440.

[10] J. Nichols, B. Zevnik, K. Anastassiadis, H. Niwa, D. Klewe-Nebenius, I. Chambers, H. Schöler, A. Smith, Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4, *Cell*, 95 (1998) 379-391.

[11] H. Niwa, J. Miyazaki, A.G. Smith, Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells, *Nat. Genet.*, 24 (2000) 372-376.

[12] A.A. Avilion, S.K. Nicolis, L.H. Pevny, L. Perez, N. Vivian, R. Lovell-Badge,

Multipotent cell lineages in early mouse development depend on SOX2 function, *Genes & Development*, 17 (2003) 126-140.

[13] G.E. Schepers, R.D. Teasdale, P. Koopman, Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families, *Developmental Cell*, 3 (2002) 167-170.

[14] L.H. Pevny, S.K. Nicolis, Sox2 roles in neural stem cells, *Int. J. Biochem. Cell Biol.*, 42 (2010) 421-424.

[15] S. Okumura-Nakanishi, M. Saito, H. Niwa, F. Ishikawa, Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells, *J. Biol. Chem.*, 280 (2005) 5307-5317.

[16] M. Tomioka, M. Nishimoto, S. Miyagi, T. Katayanagi, N. Fukui, H. Niwa, M. Muramatsu, A. Okuda, Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex, *Nucleic Acids Res.*, 30 (2002) 3202-3213.

[17] T. Kuroda, M. Tada, H. Kubota, H. Kimura, S. Y. Hatano, H. Suemori, N. Nakatsuji, T. Tada, Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression, *Molecular and Cellular Biology*, 25 (2005)

2475-2485.

[18] L.R. Johnson, K.A. Lamb, Q. Gao, T.K. Nowling, A. Rizzino, Role of the transcription factor Sox-2 in the expression of the FGF-4 gene in embryonal carcinoma cells, *Mol. Reprod. Dev.*, 50 (1998) 377-386.

[19] M. Nishimoto, A. Fukushima, A. Okuda, M. Muramatsu, The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2, *Mol. Cell. Biol.*, 19 (1999) 5453-5465.

[20] Y. Nakatake, N. Fukui, Y. Iwamatsu, S. Masui, K. Takahashi, R. Yagi, K. Yagi, J.I. Miyazaki, R. Matoba, M.S.H. Ko, H. Niwa, Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells, *Mol. Cell. Biol.*, 26 (2006) 7772-7782.

[21] B. Boer, J. Kopp, S. Mallanna, M. Desler, H. Chakravarthy, P.J. Wilder, C. Bernadt, A. Rizzino, Elevating the levels of Sox2 in embryonal carcinoma cells and embryonic stem cells inhibits the expression of Sox2:Oct-3/4 target genes, *Nucleic Acids Res.*, 35 (2007) 1773-1786.

[22] J.L. Kopp, B.D. Ormsbee, M. Desler, A. Rizzino, Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells, *Stem cells*, 26 (2008) 903-911.

[23] K. Adachi, H. Suemori, S.Y. Yasuda, N. Nakatsuji, E. Kawase, Role of SOX2 in maintaining pluripotency of human embryonic stem cells, *Genes Cells*, 15 (2010) 455-470.

[24] J.L. Chew, Y.H. Loh, W. Zhang, X. Chen, W.L. Tam, L.S. Yeap, P. Li, Y.S. Ang, B. Lim, P. Robson, H.H. Ng, Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells, *Mol. Cell. Biol.*, 25 (2005) 6031-6046.

[25] H. Fong, K.A. Hohenstein, P.J. Donovan, Regulation of self-renewal and pluripotency by Sox2 in human embryonic stem cells, *Stem cells*, 26 (2008) 1931-1938.

[26] T. Yoshimizu, N. Sugiyama, M. De Felice, Y.I. Yeom, K. Ohbo, K. Masuko, M. Obinata, K. Abe, H.R. Schöler, Y. Matsui, Germline-specific expression of the Oct-

4/green fluorescent protein (GFP) transgene in mice, *Dev. Growth Differ.*, 41 (1999) 675-684.

[27] G. Friedrich, P. Soriano, Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice, *Genes Dev.*, 5 (1991) 1513-1523.

[28] M. Stadtfeld, E. Apostolou, H. Akutsu, A. Fukuda, P. Follett, S. Natesan, T. Kono, T. Shioda, K. Hochedlinger, Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells, *Nature* 465 (2010), 175-181.

[29] P.S. Knoepfler, Why myc? An unexpected ingredient in the stem cell cocktail, *Cell Stem Cell*, 2 (2008) 18-21.

[30] R. Sridharan, J. Tchieu, M.J. Mason, R. Yachechko, E. Kuoy, S. Horvath, Q. Zhou, K. Plath, Role of the murine reprogramming factors in the induction of pluripotency, *Cell*, 136 (2009) 364-377.

[31] J.C. Silva, O. Barrandon, J. Nichols, J. Kawaguchi, T.W. Theunissen, A. Smith, Promotion of reprogramming to ground state pluripotency by signal inhibition, *PLoS*

Biol., 6 (2008) e253.

[32] S. Eminli, J. Utikal, K. Arnold, R. Jaenisch, K. Hochedlinger, Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression, *Stem cells*, 26 (2008) 2467-2474.

[33] E.P. Papapetrou, M.J. Tomishima, S.M. Chambers, Y. Mica, E. Reed, J. Menon, V. Tabar, Q. Mo, L. Studer, M. Sadelain, Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation, *Proc. Natl. Acad. Sci. U.S.A.*, 106 (2009) 12759-12764.

[34] W.E. Lowry, L. Richter, R. Yachechko, A.D. Pyle, J. Tchieu, R. Sridharan, A.T. Clark, K. Plath, Generation of human induced pluripotent stem cells from dermal fibroblasts, *Proc. Natl. Acad. Sci. U.S.A.*, 105 (2008) 2883-2888.

[35] J. Que, T. Okubo, J.R. Goldenring, K.T. Nam, R. Kurotani, E.E. Morrisey, O. Taranova, L.H. Pevny, B.L.M. Hogan, Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm, *Development*, 134 (2007) 2521-2531.

[36] O.V. Taranova, S.T. Magness, B.M. Fagan, Y. Wu, N. Surzenko, S.R. Hutton, L.H. Pevny, SOX2 is a dose-dependent regulator of retinal neural progenitor competence, *Genes Dev.*, 20 (2006) 1187-1202.

[37] S. Stefanovic, N. Abboud, S. Désilets, D. Nury, C. Cowan, M. Pucéat, Interplay of Oct4 with Sox2 and Sox17: a molecular switch from stem cell pluripotency to specifying a cardiac fate, *J. Cell Biol.*, 186 (2009) 665-673.

[38] M. Tada, T. Tada, L. Lefebvre, S.C. Barton, M.A. Surani, Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells, *EMBO J.*, 16 (1997) 6510-6520.

Figure legends

Figure 1. Expression of Sox2 in virus dilution series. The expression of Sox2, Oct4, and β -Actin (control) in mouse embryonic fibroblasts on day 3 after OSKM (*Oct4*; *Sox2*; *Klf4*; *c-Myc*) or OSK infection was analyzed by using Western blotting hybridization. The ratio of the band intensity of Sox2 or Oct4 to β -Actin was calculated using Image J software (NIH).

Figure 2. Increase in the number of *Oct4*-GFP-positive cells for direct reprogramming by using OKM (*Oct4*; *Klf4*; *c-Myc*) or OK with *Sox2* at low expression level (LS). **(A)** GFP-positive iPSC colonies on day 10 after infection of OSKM (*Sox2* 100%) and OKM+LS (*Sox2* 20%). The colonies are encircled by dotted lines. **(B)** Kinetic analysis of the number of *Oct4*-GFP-positive colonies with OSKM. The error bars that indicate S.E.M are hidden in the graph. 200% *Sox2*, S.E.M.= ± 0.0 (day 4), ± 0.0 (day 6), ± 3.6 (day 8), ± 4.0 (day10); 100% *Sox2*, S.E.M.= ± 0.0 (day 4), ± 0.8 (day 6), ± 11.8 (day 8), ± 14.1 (day10); 50% *Sox2*, S.E.M.= ± 0.0 (day 4), ± 3.1 (day 6), ± 18.7 (day 8), ± 22.5 (day10); 20% *Sox2*, S.E.M.= 0.0 (day 4), ± 3.8 (day 6), ± 17.8 (day 8), ± 28.0 (day10); 10% *Sox2*, S.E.M.= ± 0.0 (day 4), ± 3.7 (day 6), ± 19.9 (day 8), ± 16.5 (day10); 0% *Sox2*; S.E.M.= ± 0.0 (day 4), ± 0.0 (day 6), ± 0.9 (day 8), ± 0.2 (day10). **(C)** The generation frequency of fully reprogrammed iPSCs by OSKM and

OKM+LS transduction. **(D)** Kinetic analysis of the number of *Oct4*-GFP-positive colonies with OSK. The error bars that indicate S.E.M are hidden in the graph. 200% *Sox2*, S.E.M.= ± 0.0 (day 8), ± 0.0 (day 10), ± 0.0 (day 12), ± 0.1 (day14), ± 0.5 (day16); 100% *Sox2*, S.E.M.= ± 0.0 (day 8), ± 0.5 (day 10), ± 1.2 (day 12), ± 2.6 (day14), ± 4.8 (day16); 50% *Sox2*, S.E.M.= ± 0.0 (day 8), ± 1.3 (day 10), ± 3.7 (day 12), ± 10.7 (day14), ± 8.3 (day16); 20% *Sox2*, S.E.M.= 0.0 (day 8), ± 1.1 (day 10), ± 3.3 (day 12), ± 6.6 (day14), ± 8.0 (day16); 10% *Sox2*, S.E.M.= ± 0.0 (day 8), ± 0.6 (day 10), ± 2.2 (day 12), ± 4.8 (day14), ± 4.2 (day16); 0% *Sox2*, S.E.M.= ± 0.0 (day 8), ± 0.0 (day 10), ± 0.0 (day 12), ± 0.0 (day14), ± 0.1 (day16). **(E)** The generation frequency of fully reprogrammed iPSCs by OSK and OK+LS transduction.

Figure 3. Expression of pluripotent and lineage marker genes in somatic cells that had been reprogrammed with *Oct4*, *Klf4* (OK) and low level *Sox2* (LS) expression as determined by using quantitative reverse transcription-PCR (qPCR) analysis. **(A)** Relative expression of reprogramming factors on day 3 after infection (day 3). The primer sets were designed to amplify both endogenous and transgenic expressions. n=5. **(B-D)** The relative expression of *Oct4*/*Sox2* target genes on day 3 (n=5) **(B)**, stem cell markers on day 3 (n=5) **(C)**, and lineage-specific markers on day 3 (n=5) **(D)**. **(E)** Kinetic analyses of gene expression relative to MEFs in lineage-specific genes on days 3, 7, and 16. TE, trophectoderm. * $P < 0.05$. Error bars, S.E.M.

Figure 4. Pluripotency of induced pluripotent stem cells established by using *Oct4*, *Klf4*, and low *Sox2* (OK+LS-iPSCs). **(A)** Expression of Oct4-GFP in morphologically normal OK+LS-iPSCs. GFP expression was detected to be uniform in the cell colonies. **(B)** Expression of pluripotent marker genes in #1-3 OK+LS iPSC lines by using RT-PCR. *Gapdh* was used as a loading control. The primer sets that were used are summarized in Supplementary Table 1. **(C)** The expression of pluripotent marker proteins by using immunostaining. The expression of Nanog, Oct4, Sox2, SSEA1, Stella and E-cadherin (Ecad) was detected as red signals. Nuclei were visualized by using 4', 6-Diamidino-2-Phenylindole (DAPI). **(D)** Contribution of OK+LS- iPSCs derivatives in mouse E15.5 or E16.5 chimeric embryos. The iPSC derivatives were visualized as blue cells by using X-gal staining. Germ cells derived from iPSCs were detected as X-gal-positive cells in the testis of E15.5 #3 embryo (right panel).

Figure 5. Schematic representation of the reprogramming effects of *Sox2* at the low expression level (LS) with and without *c-Myc* on cell fate. **(A)** The reprogramming effect of *Oct4*, *Klf4*, *c-Myc* (OKM) and LS on cell fate. For conventional OSKM, the majority of transduction-mediated reprogramming cells undergo differentiation to ectoderm and mesoderm lineages. A limited number of cells are changed into fully

reprogrammed induced pluripotent stem cells (iPSCs). For OKM and LS, the limited number of cells underwent differentiation into the ectoderm and mesoderm lineage, while the number of partially reprogrammed iPSCs significantly increased compared to the OSKM. Furthermore, no increase in the number of fully reprogrammed iPSCs was detected. **(B)** The reprogramming effects of *Oct4*, *Klf4* (OK) and LS on cell fate. For conventional OSK, a number of infected cells do not reprogram or differentiate into the ectoderm and mesoderm lineages. However, a limited number of cells are changed into fully reprogrammed iPSCs. For OK and LS, the number of fully reprogrammed cells significantly increases instead of there being a decrease in the number of cells that differentiate into the ectoderm and mesoderm lineages.

Supplementary Figure 1. Expression of pluripotent and lineage marker genes in somatic cells that had been reprogrammed with *Oct4*, *Klf4*, *c-Myc* (OKM) and the low *Sox2* (LS) expression as determined by using quantitative reverse transcription-PCR (qPCR) analysis.

(A) Relative expression of reprogramming factors on day 3 after infection. Primer sets were designed to amplify endogenous and transgenic expressions (n=6). (B-D) Relative expression of *Oct4/Sox2* target genes on day 3 (n=6) (B), stem cell markers on day 3 (n=6) (C), and lineage markers on day 3 (n=6) (D). (E) Kinetic analyses of gene expression relative to MEFs for lineage marker genes on days 3, 7, and 12. TE, trophectoderm. * $P < 0.05$. Error bars, S.E.M.

Supplementary Figure 2. Characteristics of partially reprogrammed induced pluripotent stem cells (partial iPSCs) that were established by using *Oct4*, *Klf4*, *c-Myc* (OKM) and low *Sox2* (LS) expression. (A) The number of GFP-positive iPSCs on day 7 after OSKM and OKM+LS transduction by using FACS analysis. (B) Morphology of partial and fully reprogrammed iPSC colonies. *Oct4*-GFP is visualized as green. (C) Expression of pluripotent and lineage marker genes in partial iPSCs relative to ESCs. Error bars, S.E.M.

Supplementary Figure 3. Pluripotency of induced pluripotent stem cells that were

established by using *Oct4*, *Klf4*, *c-Myc*, and low *Sox2* (OKM+LS-iPSCs). **(A)** Expression of *Oct4*-GFP in morphologically normal OKM+LS-iPSCs. GFP expression was detected as being uniform in the cell colonies. **(B)** Expression of pluripotent marker genes in #1-3 OKM+LS iPSC lines by using RT-PCR. *Gapdh* was used as a loading control. The primer sets that were used are summarized in Supplementary Table 1. **(C)** Expression of pluripotent marker proteins by using immunostaining. The expression of Nanog, Sox2, Stella and SSEA1 was detected as red signals. The insets are used to show nuclei visualized by using 4', 6-Diamidino-2-Phenylindole (DAPI). **(D)** The contribution of OKM+LS- iPSCs derivatives in a mouse E12.5 #2 chimeric embryo. The iPSC derivatives were visualized as being blue cells by using X-gal staining. Germ cells derived from iPSCs were detected as *Oct4*-GFP-positive cells in the genital ridges (lower panels).

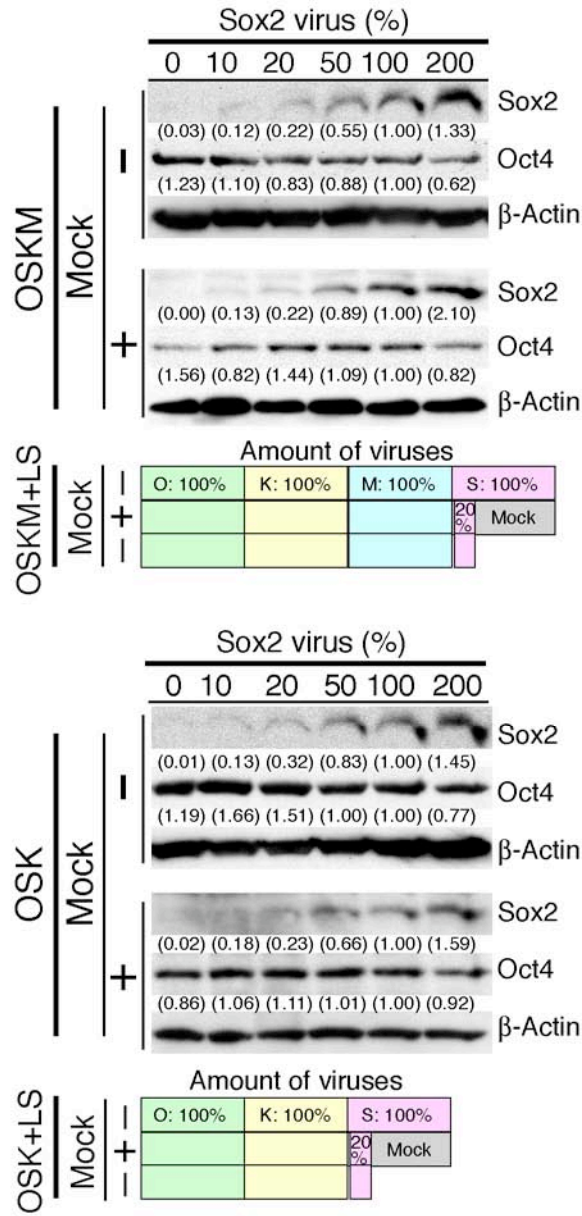


Figure 1

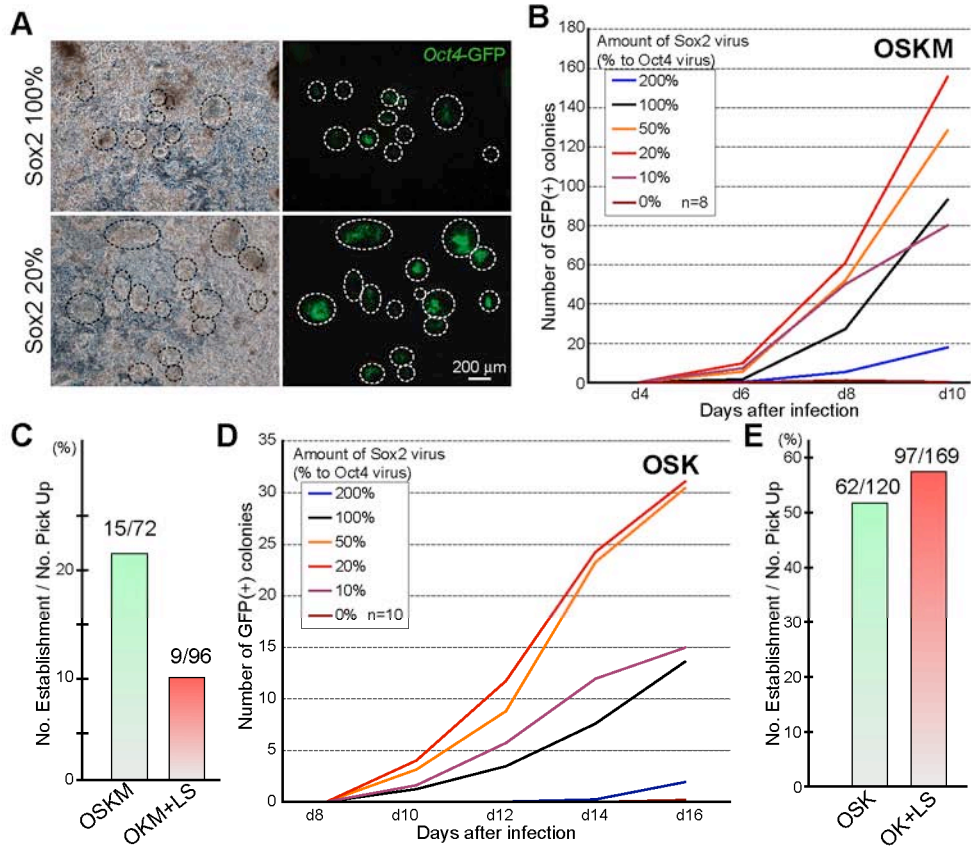


Figure 2

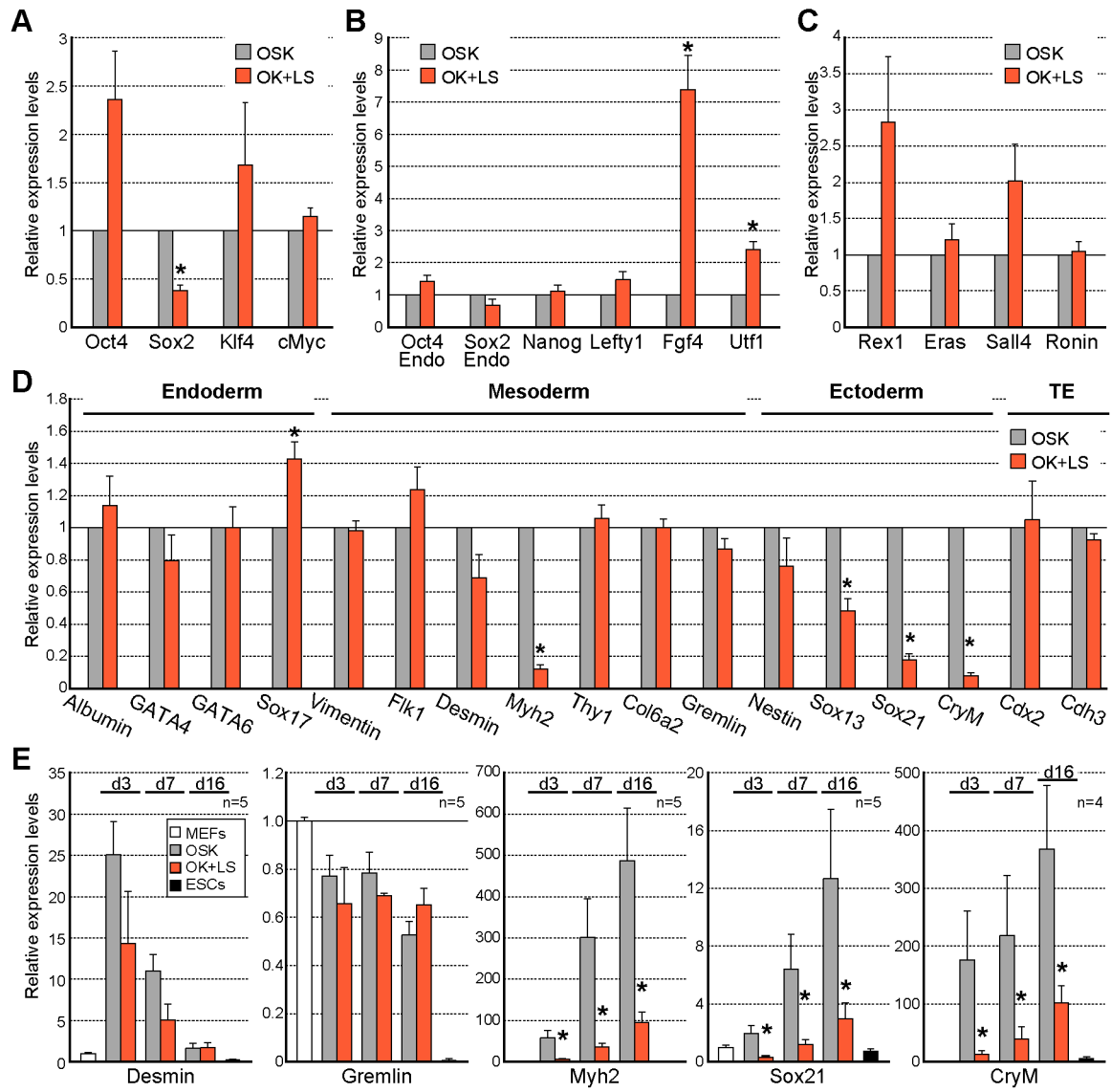


Figure 3

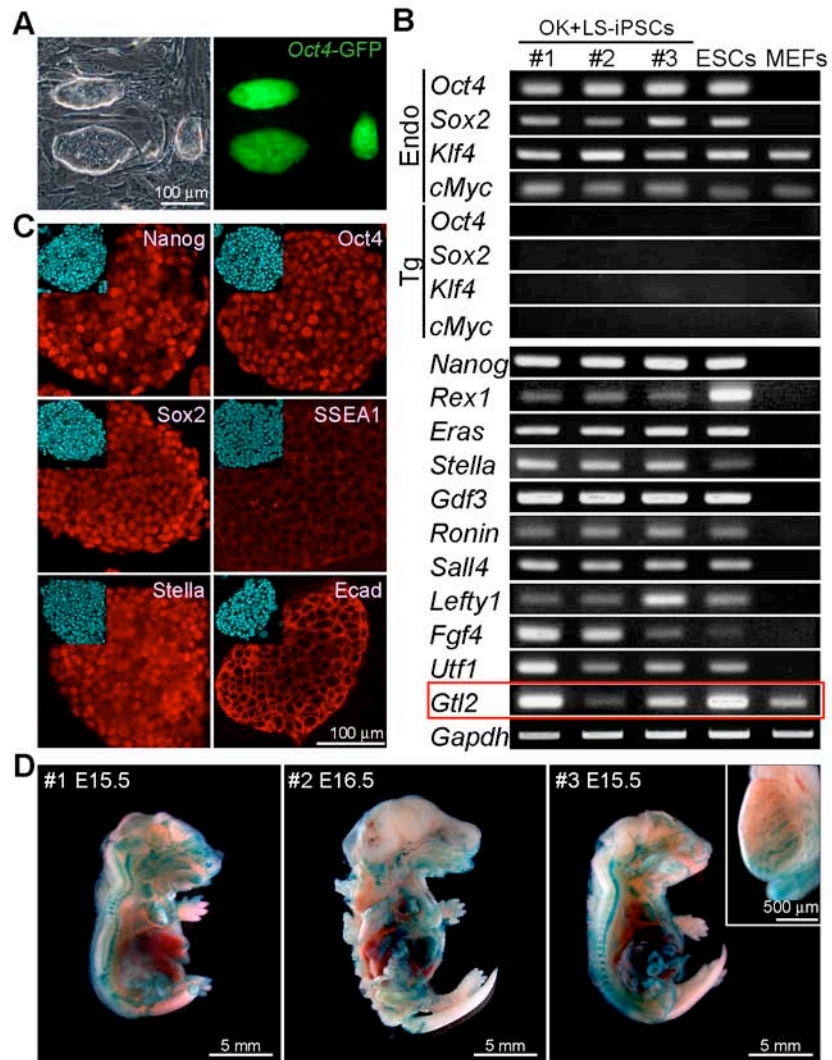


Figure 4

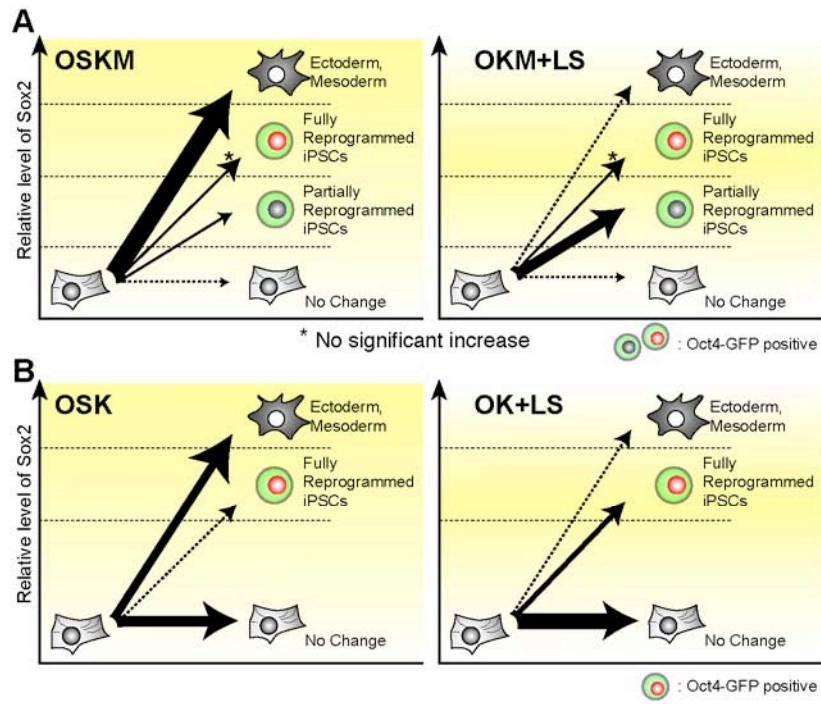
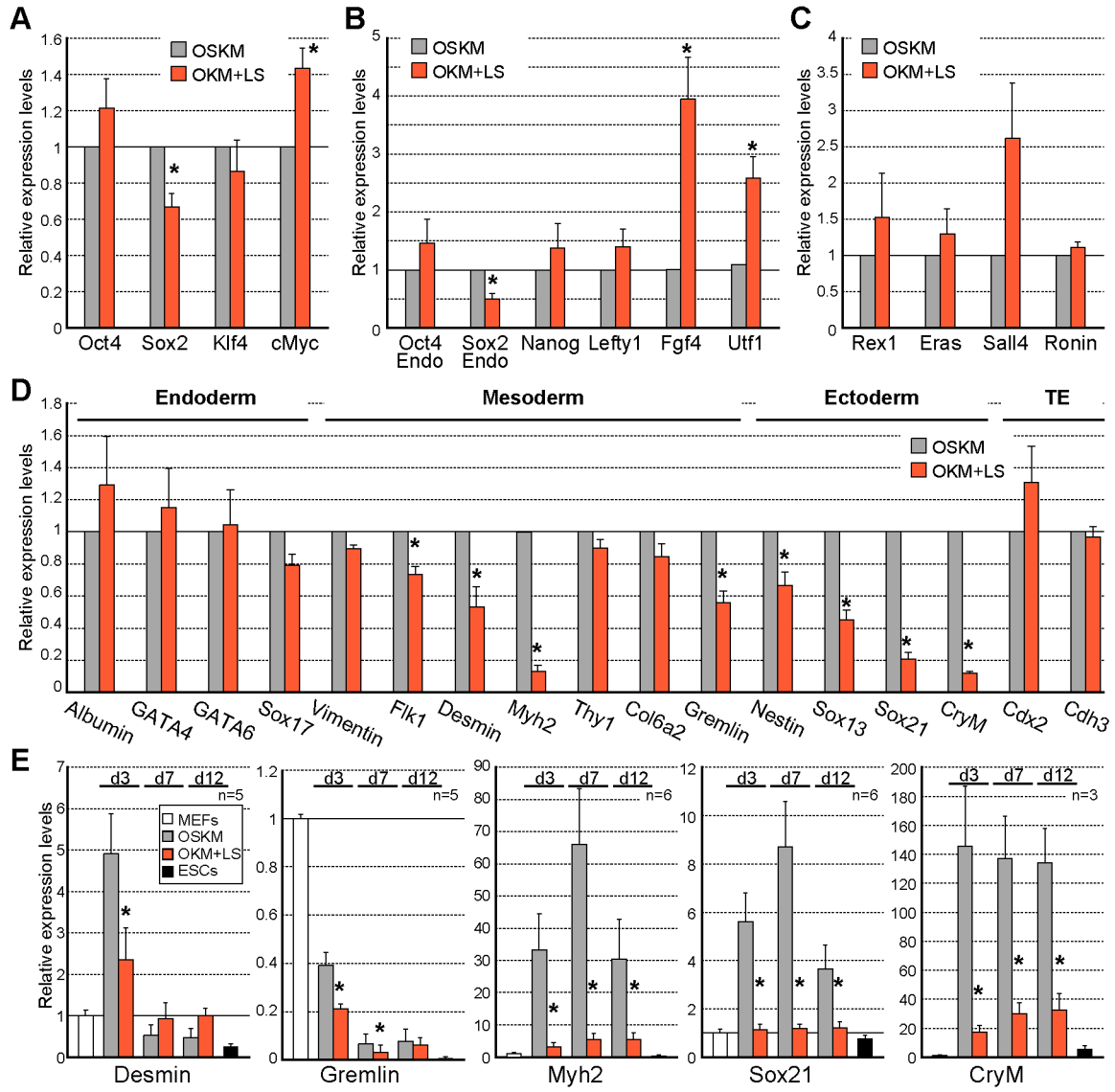


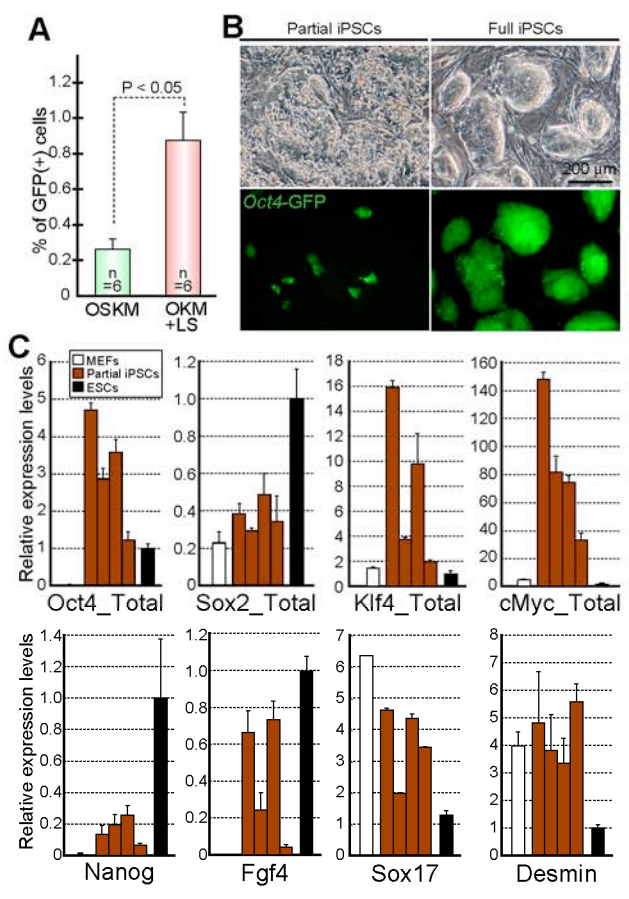
Figure 5

Supplementary Table 1 Primer sequences for quantitative reverse transcription-PCR analysis and reverse transcription-PCR analysis

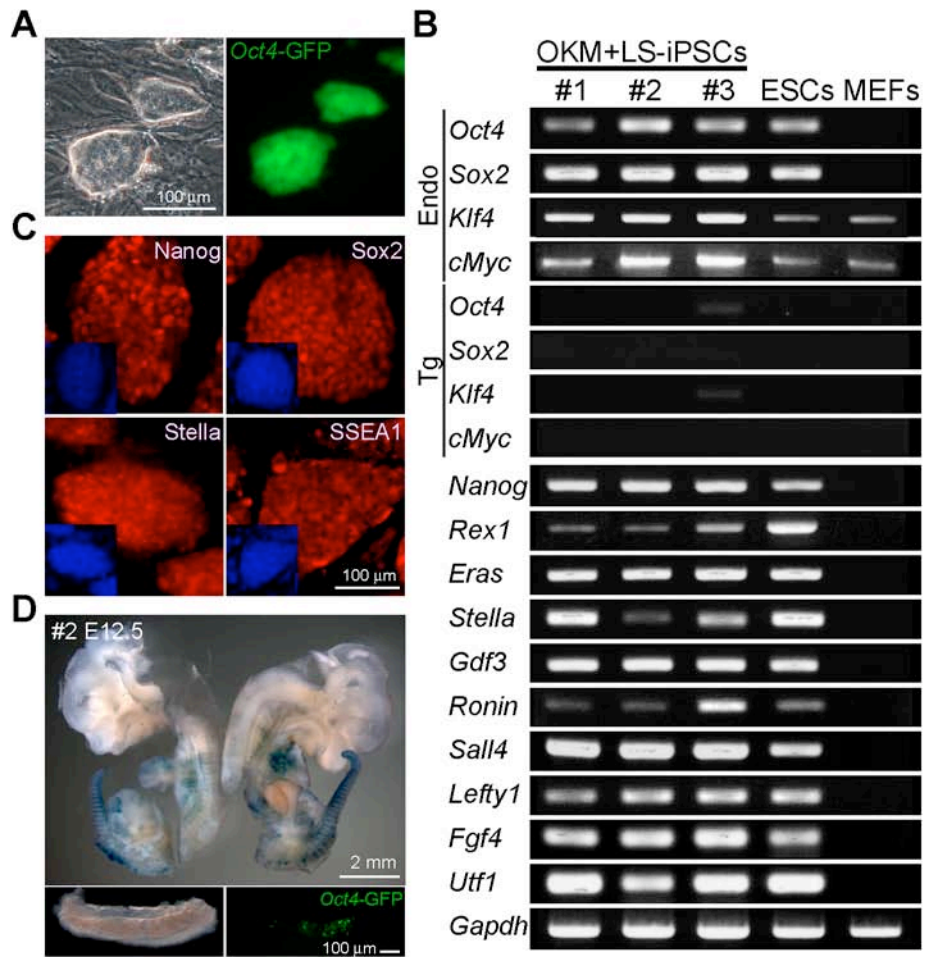
Gene name	Forward Primer (5' - 3')	Reverse Primer (5' - 3')
<i>Oct4</i> (endogenous)	TCTTTCCACCAGGCCCGGCTC	TGCGGGCGGACATGGGGAGATCC
<i>Sox2</i> (endogenous)	TAGAGATAGACTCCGGGCGATGA	TTGCCTTAAACAAGACCACGAAA
<i>Klf4</i> (endogenous)	GCGAACTCACACAGGCGAGAAACC	TCGCTTCCTCTTCCGACACA
<i>cMyc</i> (endogenous)	CAGAGGAGGAACGAGCTGAAGCGC	AAGTTTGAGGCAGTAAAATTATGGCTGAAGC
<i>Oct4</i> (transgene)	CCCATGGTGGTGGTACGGGAATTC	AGTTGCTTTCCACTCGTGCT
<i>Sox2</i> (transgene)	CCCATGGTGGTGGTACGGGAATTC	TCTCGGTCTCGGACAAAAGT
<i>Klf4</i> (transgene)	CCCATGGTGGTGGTACGGGAATTC	GTCGTTGAACTCCTCGGTCT
<i>cMyc</i> (transgene)	CTCCTGGCAAAGGTCAGAG	GACATGGCCTGCCCGTTATTATT
<i>Nanog</i>	CTTTCACCTATTAAGGTGCTTGC	TGGCATCGGTTTCATCATGGTAC
<i>Lefty-1</i>	CTCACCTGCTAGTGTTCCGG	GAGCCTAGAATCGATAAGGG
<i>Fgf4</i>	TCGGTGTGCCTTTCTTTACC	ACCTTCATGGTAGGCGACAC
<i>Utf1</i>	TTACGAGCACCGACACTCTG	GGCCCAGAACTGTTGAGATG
<i>Rex1</i>	GACATCATGAATGAACAAAAAATG	CCTTCAGCATTCTTCCCTG
<i>Eras</i>	CAAAGATGCTGGCAGGCAGCTACC	GACAAGCAGGGCAAAGGCTTCCTC
<i>Sall4</i>	ATCCGAGCACAGCCCACCTTTGTC	ACGCTGGTGTACTGGTTCAGGAG
<i>Ronin</i>	GCCTCAGAGCTAGAGGCTGCTACG	TGGAAGGAGTCACGAATTCTGCAG
<i>Albumin</i>	CGCCCATCGGTATAATGATTT	ATTTGGCATGCTCATCGTATG
<i>GATA4</i>	CCCTACCCAGCCTACATGG	ACATATCGAGATTGGGGTGTCT
<i>GATA6</i>	GCAATGCATGCGGTCTCTAC	CTCTTGGTAGCACCAGCTCA
<i>Sox17</i>	GATGCGGGATACGCCAGTG	CCACCACCTCGCCTTTAC
<i>Vimentin</i>	AAAGTGGAATCCTTGCAGG	ACGCTTTCATACTGCTGGC
<i>Flk1</i>	CCCTTCAGATTACTTGCAGG	GAAGATACTGTCACCACCGC
<i>Desmin</i>	TTGGGGTCGCTGCGGTCTAGCC	GGTCGTCTATCAGGTTGTCACG
<i>Myh2</i>	GGCGGAAAGAAGCTACCAT	CCCATAATGCATCACAGCAC
<i>Thy1</i>	GGAACTCTTGGCACCATGAAC	GGGACACCTGCAAGACTGAGA
<i>Col6a2</i>	CCACCACTGAAAGGAACAACAA	TCCAACACGAAATACACGTTGAC
<i>Gremlin2</i>	GAAGCATTGCCGGTGCAT	TCCTGCCCATGAGTTCAC
<i>Nestin</i>	CCCTGAAGTCGAGGAGCTG	CTGCTGCACCTCTAAGCGA
<i>Sox13</i>	TACTCACAGACTGATCCCCAG	TCTACAAGTGCAGTTTCACCC
<i>Sox21</i>	CCTAAGATGCACAACTCGGA	CTTGTAGTCGGGATGCTCC
<i>CryM</i>	GTGCTTCTCTTTGATCCCAG	GCTTCAACAGCTTTGTGGC
<i>Cdx2</i>	AGACAAATACCGGGTGGTGTA	CCAGCTCACTTTTCCTCTGA
<i>Cdh3</i>	ACATCATTATGCTGACTAGGGG	TATGGTGAAGACTCCTTCGG
<i>Gtl2</i>	TTGCACATTCCTGTGGGAC	AAGCACCATGAGCCACTAGG



Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3