OCT4 Activity during Conversion of Human Intermediately Reprogrammed Stem Cells to iPS Cells through MET

Rika Teshigawara[1], Kunio Hirano[1], Shogo Nagata[1], Justin Ainscough[2] and Takashi Tada[1,*]

[1]Stem Cell Engineering, Institute for Frontier Medical Sciences, Kyoto University, 53 Kawahara-cho, Shogo-in, Sakyo-ku, Kyoto 606-8507 JAPAN.

[2]Department of Biology, University of York, York, YO10 5DD, UK.

[*]Correspondence: ttada@frontier.kyoto-u.ac.jp

Corresponding author: Takashi Tada, Department of Stem Cell Engineering, Institute for Frontier Medical Sciences, Kyoto University,
53 Kawahara-cho, Shogo-in, Sakyo-ku, Kyoto 606-8507 Japan
Tel & Fax: +81-75-751-4102
E-mail: ttada@frontier.kyoto-u.ac.jp

Summary

To facilitate understanding the mechanisms of somatic reprogramming to human induced Pluripotent Stem Cells (iPSCs), we have established intermediately Reprogrammed Stem Cells (iRSCs), human mesenchymal cells that express exogenous Oct4/Sox2/Klf4/c-Myc (OSKM) and endogenous SOX2/NANOG. iRSCs can be stably maintained at low density. At high density, however, they are induced to enter Mesenchymal-to-Epithelial Transition (MET), resulting in reprogramming to an iPSC state. Morphological changes through MET correlate with silencing of exogenous OSKM, and up-regulation of endogenous *OCT4*. A CRISPR/Cas9-mediated *GFP* knock-in visualized the temporal regulation of endogenous *OCT4* in cells converting from iRSC to iPSC state. *OCT4* activation coincident with OSKM silencing occurred prior to entering MET. Notably, *OCT4* instability was frequently observed in cells of developing post-MET colonies until a late stage (>200 cells), demonstrating that *OCT4*-activated post-MET cells switched from asymmetric to symmetric cell division in late stage reprogramming.

Introduction

Human induced pluripotent stem cells (iPSCs), first generated by the ectopic co-expression of four reprogramming factors, *Octamer-binding protein 4* (*Oct4*), *SRY-box containing gene 2* (*Sox2*), *Kruppel-like factor 4* (*Klf4*), and *c-Myelocytomatosis oncogene* (*c-Myc*) (OSKM) in dermal fibroblasts (Takahashi et al., 2007), hold great potential for application in autologous cell therapy, disease modeling, and drug discovery. Critical steps that currently hinder realization of this potential include low efficiency of somatic cell reprogramming and poor reproducibility of high quality iPSC generation. To help overcome these issues, novel approaches for iPSC production have been developed, however, the efficiency of reprogramming human cells remains less than 0.1% (Stadtfeld and Hochedlinger, 2010). Thus, iPSC generation is still an extremely time-consuming process. Thorough understanding of the mechanisms that modulate epigenetic reprogramming of somatic to pluripotent cells has important implications for downstream iPSC applications. However, the low efficiency and stochastic nature of reprogramming make dissecting the underlying mechanisms extremely difficult.

Success in this area has been largely restricted to investigations using mouse cells. In mice, it has been proposed that extinction of the somatic program and subsequent activation of endogenous pluripotency genes are key roadblocks in the process of reprogramming-to-iPSC (Stadtfeld and Hochedlinger, 2010). Somatic reprogramming-to-iPSC can be considered a gradual process with intermediate cell populations. Notably, somatic cell conversion to iPSC is accompanied by striking morphological changes as the cells transit from a mono-layer of adherent cells to multilayered epithelial cells, a process reminiscent of the Mesenchymal-to-Epithelial Transition (MET). This early step is crucial for successful reprogramming in the mouse(Li et al., 2010; Samavarchi-Tehrani et al., 2010). FACS-sorted intermediately reprogrammed cells, generated from doxycycline-inducible OSKM expressing somatic cells, were used for single cell gene expression analysis (Polo et al., 2012). This indicated that transcription factor-induced pluripotency elicited two waves of transcription, driven by *c-Myc* and *Klf4* (first wave), and *Oct4*/*Sox2*/*Klf4* (second wave). In the first wave, *Thy1*-negative/ stage-specific embryonic antigen-1 (SSEA1)-positive cells entered into MET, evidenced by down-regulation of *Snai1* and up-regulation of *E-Cadherin* at day 3 after OSKM induction. Subsequently, an *Oct4-GFP* reporter was up-regulated in the second wave at day 8-10 prior to formation of stable

iPSC colonies. In addition to SSEA1, three genes, *Epcam*, *c-Kit*, and *Pecam* were identified as markers for enrichment of reprogrammable cells at early (0-3 days) and late (day 9 onward) time points. Other detailed analyses in single cells at various stages during somatic reprogramming-to-iPSC demonstrated that expression of *Esrrb*, *Utf1*, *Lin28*, and *Dppa2* were better predictors for enrichment of cells progressing into iPSCs (Buganim et al., 2012).

In human, iPSCs have several distinct features that are different from mouse iPSCs. Human iPSCs form characteristic flat-shaped colonies, in contrast to bowl-shaped colonies formed by mouse iPSCs. The gene expression profile of human iPSCs resembles that of mouse Epiblast stem cells in a primed state of pluripotency rather than that of mouse iPSCs in a naïve state of pluripotency. Notably, human iPSCs are distinct from mouse iPSCs in their response to two kinase inhibitors (mitogen-activated protein/extracellular signal-regulated kinase inhibitor and glycogen synthase kinase 3 inhibitor) (Hirano et al., 2012). Furthermore, human somatic cells can be conventionally reprogrammed using human ES medium supplemented with knockout serum replacement and basic fibroblast growth factor (bFGF) but not using mouse ES medium supplemented with fetal bovine serum (FBS) and leukemia inhibitory factor (LIF), which is widely used in mouse iPSC generation. Collectively, these differences between human and mouse iPSCs indicate distinct processes of somatic cell reprogramming, even where the same transcription factor-induced reprogramming system is utilized.

Several types of mouse pre-iPSC lines, which were endogenous *Oct4*-negative, have been isolated (Chen et al., 2013; Theunissen et al., 2011). In human, partially reprogrammed iPSCs resumed reprogramming by up-regulation of *KLF4* (Nishimura et al., 2014), and pre-iPSC-like cell lines were established as cancer stem cell lines (Nagata et al., 2012). However, no intermediately reprogrammed stem cells (iRSC) capable of restarting reprogramming into iPSCs, have been established as stable cell lines. Recently, in mouse, global gene expression analyses of reprogramming colonies demonstrated that the pluripotent spectrum of somatic reprogramming encompasses multiple states (Tonge et al., 2014). Compact colony forming class (C-class) cells were *Nanog*-negative, while alternative fuzzy colony forming class (F-class) cells were *Nanog*-positive. It is therefore possible that somatic reprogramming encompasses an alternative intermediate pathway (*Nanog*-positive F-class), in addition to *Nanog*-negative C-class, even in human reprogramming.

Here we describe the isolation and characterization of clonally expandable, human intermediately reprogrammed stem cells (iRSCs), capable of resuming reprogramming to iPSC stage. GFP knock-in to the endogenous *OCT4* locus by CRISPR/Cas9-mediated genome editing enabled investigation of *OCT4-GFP* activation kinetics as the cells transit from iRSC to iPSC.

Results

Establishment of intermediately reprogrammed stem cell (iRSC) lines

To explore whether stable intermediately reprogrammed human cells could be isolated and maintained, DsRed-positive colonies were isolated 20-30 days after retroviral transduction of *Oct4*, *Sox2*, *Klf4*, *c-Myc*, and *DsRed* into Slc7a1-expressing human fetal lung fibroblasts (Takahashi et al., 2007), TIG1 and TIG3. These colonies were dissociated into single cells and expanded on Matrigel-coated culture dishes in MEF-conditioned medium in the absence of a feeder cell layer (Figure 1). Approximately one hundred colonies were generated per 10 cm culture dish. Two types of cells self-renewed and formed stable colonies, while the majority of cells appeared differentiated and/or disappeared following subculture. One type of colony closely resembled cancer stem cells (Nagata et al., 2012), the other retained phenotypic characteristics similar to those of mesenchymal cells, with large nuclei. These colonies were established at a frequency of approximately 5% of picked colonies, and could be stably maintained as scattered cells through subculture in MEF-conditioned medium for more than forty passages. Eleven such iRSC lines were independently established. Importantly, the cells retained normal karyotype through repeated passage (Figure S1). More importantly, the cells were observed to spontaneously progress into iPSCs at high frequency, when cultured for 3 or more days without subculture. Thus, reprogramming of the 'mesenchymal' cells into iPSCs was induced by simple change in culture condition to high density. These pre-iPSC mesenchymal cells are hereafter referred to as 'intermediately reprogrammed stem cells' (iRSCs), and have been stably maintained as established cell lines. Reprogramming of iRSC to iPSC was observed even after iRSCs were maintained for more than forty passages. These cells therefore appear to represent a stage in the somatic-to-iPSC reprogramming process at which the cells are paused in an intermediately reprogrammed state, awaiting further external cues.

Characterization of iRSCs

To characterize the properties of iRSCs, we first analyzed gene expression by RT-PCR and gene expression microarray. In iRSCs, exogenous *Oct4*, *Sox2*, *Klf4*, and *c-Myc* were expressed, while endogenous *OCT4* was silenced (Figure 2A). Pluripotent marker genes, *TDGF1* and *REX1* were repressed, while *NANOG* and endogenous *SOX2* were expressed. *NANOG* expression is likely driven by exogenous Oct4 and

Sox2 (Kuroda et al., 2005). Gene expression microarray analyses demonstrated that the global gene expression profile of iRSC is intermediate between TIG1 and iPSC (Figure 2B). The intermediate state was confirmed by comparative scatter plot analyses between iRSC and TIG1, and iRSC and iPSC (Figure 2C). Independently isolated iRSC lines showed similar global gene expression profile (Figure S2). Endogenous *NANOG* and *SOX2* were activated, while *OCT4* and *c-MYC* remained repressed. Somatic markers *FOXC1* and *MEIS2* were expressed in iRSCs. To further verify the intermediate state of iRSCs, gene expression was analyzed (Figure 2D). No expression of the pluripotent marker proteins, SSEA4, TRA1-60 and ECAD could be observed, while NANOG was readily detectable (Figure 2E). Consistent with the RT-PCR expression analysis, the promoter of *OCT4* was found to be hypermethylated in iRSC and TIG1, and hypomethylated in iPSCs (Figure 2F). These data support the notion that iRSCs are stably arrested in an intermediate state, between somatic cell and iPSC during the reprogramming process.

Reprogramming of iRSC to iPSC

To confirm that iRSCs are indeed intermediately reprogrammed cells capable of continued reprogramming into iPSCs at high frequency, DsRed-positive iRSCs were plated at high cell density (1.0×10^6) on Matrigel-coated 3.5 cm wells (day 0) and maintained in MEF-conditioned human iPSC medium for 3 days. DsRed-negative flat-shaped colonies consisting of 8 to 16 cells could be readily detected by day 3, with approximately 50-80 distinct colonies forming in each well by day 6 (Figure 3A). All of the colonies grew rapidly and developed as tightly connecting cells by day 6. At day 10, colony morphology closely resembled that of human iPSCs. Notably, due to the high frequency of colony formation, we were able to efficiently track, for the first time, reprogramming of iRSC to iPSC by Time-lapse imaging. This clearly showed that reprogramming iRSCs repressed DsRed marker gene expression at a very early stage, and then promptly divided into two flat cells with characteristic epithelial morphology (Movies S1-S3). The epithelial cells exhibited a cycling time of approximately 10 hours, and rapidly expanded to form packed iPSC colonies. To verify the observed switch from mesenchymal iRSC to iPSC, expression of mesenchymal and epithelial marker genes was analyzed by RT-PCR. Mesenchymal genes including *ZEB1* and *SNAI2* were expressed in iRSCs, but not epithelial genes such as *ECAD* and *EPCAM* (Figure 3B). Conversely, epithelial genes but not

mesenchymal genes were expressed in samples derived from iPSCs. These data demonstrate that MET occurred during iRSC to iPSC reprogramming. iPSCs generated from iRSCs had similar properties to iPSCs generated using conventional protocols, evidenced by expression of pluripotency marker proteins, OCT4, NANOG and ECAD, retention of normal karyotype, and retention of widespread differentiation potential to form various cell types in teratomas (Figure S3).

To investigate the switch in expression from the exogenous to endogenous (*Oct4-OCT4* and *Sox2-SOX2*) genes, expression was analyzed by Q-PCR in MET. Following the observed down-regulation of DsRed reporter expression immediately prior to transformation from mesenchymal to epithelial cell (Movie S1 and S3), exogenous *Oct4* and *Sox2* were silenced by day 3 (Figure 3C). No expression of exogenous reprogramming genes was detected in iRSCs at day 6, or in established iPSCs. In addition to silencing of the exogenous genes, expression of pluripotent marker proteins, SSEA4 and TRA1-60 at day 3, and ECAD at day 6 was observed by immunocytochemistry (Figure 3D).

In conclusion, onset of MET in iRSCs is induced by culture at high cell density, coupled with silencing of exogenous reprogramming genes, and continued progression toward iPSC generation.

Visualization of Endogenous OCT4 in Reprogramming

To monitor endogenous OCT4 expression in living cells, *OCT4*-2A-*GFP*-*Lox*P-*PGK*-*Puro*-*Lox*P was integrated at the end of the exon 5 of *OCT4* using CRISPR/Cas9 (Figure 4). Two of sixty Puromycin-selected clones expressed GFP after complete reprogramming to iPSC. No DNA mutations were detected at the GFP-integrated allele of two clones. One iRSC clone had an eight-base deletion at the counterpart of GFP-integrated allele, while another clone was intact. Therefore, the OCT4-GFP (OG)-iRSC clone heterozygous for GFP with an intact wild-type allele was used for the following experiments. To prevent potential negative influence of the *PGK* promoter-driven Puromycin-resistance gene in *OCT4* expression, *PGK-Puromycin* was eliminated from the clone by transient expression of *Cre* recombinase (Figure 4). Finally, strong GFP expression was detected in iRSC-derived iPSC cytoplasm in two out of eight clones picked. The GFP integration site was again verified by genomic PCR with two independent clones (Figure 4). The PCR product (1071bp) of OCT4-lane1 (pink) contains the junction between the left arm and intron3, while that

(498bp) of OCT4-lane3 (green) shows integration of 2A-GFP. OCT4-lane2 (blue) demonstrates heterozygous integration of GFP as a 1764bp product for the targeted allele, and a 921bp product for the wild-type allele. To validate the integration site more precisely, DNA sequence analysis was performed around intron 3 and exon 5 of *OCT4* (Figure S4).

To report on endogenous OCT4 activity in iPSC, GFP expression was assessed at days 2, 4, 6, and 9 following resumption of reprogramming of OG-iRSC culture at high density (Figure 5A). MET, exogenous DsRed silencing, and GFP activation were coincident, suggesting that endogenous OCT4 activation can be appropriately reported through tagged GFP. Importantly, this provides a sensitive marker for reprogramming of OG-iRSCs into iPSCs.

Endogenous OCT4 activation prior to entry into MET

To explore activation timing of endogenous OCT4 in reprogramming of iRSCs into iPSCs, GFP expression was observed two day after iRSC culture at the high density. GFP expression was frequently detected in isolated DsRed-negative cells that retained mesenchymal cell phenotype (Figure 5B). GFP expression was also assessed by immunocytochemistry using anti-GFP antibody 24 hours after high-density culture, prior to build up of sufficient observable GFP protein (Figure 5B). DsRed-negative mesenchymal iRSCs were antibody positive at this early time point, indicating that activation of endogenous OCT4 occurs before entry into MET during iRSC to iPSC reprogramming. To further assess the reciprocity of GFP and DsRed expression, we performed detailed microscopic observation using the anti-GFP antibody. Thirty-eight (88%) of forty-three GFP-positive iRSCs were negative for DsRed, while five (12%) were faintly positive (Figure 5C). No DsRed and GFP-negative cell was detected at pre-MET. In general, DsRed expression was rapidly replaced by GFP, suggesting that exogenous OSKM silencing and endogenous OCT4 activation are tightly linked events that precede or are concurrent with initiation of MET.

Instability of OCT4 in Post-MET Reprogramming

To explore whether iRSCs can be reprogrammed into iPSCs as elite cells following endogenous OCT4 activation, GFP-expression was closely monitored in early post-MET colonies (Figure 6A). Cells in colonies were negative for DsRed and positive for GFP 1-2 days after entry into MET. Remarkably, a proportion of the

DsRed-negative cells within post-MET colonies became GFP-negative (endogenous OCT4-negative) 3-9 days after MET (Figure 6A and 6B). Time-lapse imaging demonstrated reversion of some GFP-positive post-MET cells to GFP-negative after cell division, indicating that OCT4 silencing may be a cell cycle-dependent event (Movie S4). Consistent with this, it has been suggested that cell cycle-dependent gene regulation may occur late in reprogramming (Tanaka et al., 2015). Emergence of GFP-negative post-MET cells was clearly detectable within colonies consisting of 25-200 cells. To verify the apparent instability of OCT4 during this critical early stage of the reprogramming process, post-MET colonies were immunostained with anti-GFP antibody (Figure S5). Some cells became detectable as GFP-positive, suggesting low-level expression below the sensitivity of GFP protein alone. Other cells, however, remained clearly GFP-negative. Both GFP-positive and negative cells showed DsRed-negative epithelial morphology, indicating that MET had occurred. The emergence frequency of GFP-negative/DsRed-negative cells over the time course of early reprogramming is summarized in Figure 6B. GFP-negative/DsRed-negative cell-containing colonies were frequently detected in colonies composed groups of 8-200 cells (Day 3-10). In contrast, colonies composed of more than 200 cells contained fewer GFP-negative/DsRed-negative cells, suggesting stabilization of endogenous OCT4 expression. The differential of GFP intensity between GFP-negative and positive cells at Day 3 was striking, with little evidence of cells expressing intermediate levels (Figure 6C).

To determine whether exogenous OSKM were silenced in GFP-negative cells as well as GFP-positive cells in post-MET colonies, total RNA of 10-100 cells micro-dissected from colonies were Q-PCR amplified with exogenous OSKM and DsRed-specific primer sets. All genes were found to be silenced to the same extent in GFP-positive and GFP-negative cells (Figure S6A). Lack of ectopic expression of OSKM in post-MET GFP-negative and positive cells suggests that the observed instability of endogenous *OCT4* is a characteristic and possibly regulated phenomenon in early post-MET colonies. This is likely to hold me true, even for development of OSKM-integration-free iPSCs.

To examine the fate of GFP-negative cells, endogenous *OCT4*, *SOX2*, *KLF4*, *c-MYC*, and *NANOG* transcription was examined by Q-PCR (Figure S6B). *OCT4* and *NANOG* were significantly decreased in GFP-negative cells relative to GFP-positive cells, while *SOX2*, *KLF4* and *c-MYC* were expressed at similar or higher level. Thus,

GFP-negative cells are evidently committed to differentiate down particular cell lineages. Further Q-PCR analyses with *PAX6* and *NESTIN*-specific primer sets demonstrated that neural differentiation was initiated in some cells of GFP-negative cells (Figure S6C), in which *OCT4* and *NANOG* were down-regulated.

To examine whether GFP-negative cells can convert to GFP-positive cells, ten independent GFP-negative 8-25-cell colonies identified 3-9 days after MET induction were assessed every 12 hours for the following 7 days. In all ten colonies, no induction of GFP was detected (Figure S7), indicating that GFP-negative cells entered a state out of recoverable pluripotency with induction of differentiation.

Collectively, the data suggests that activation of OCT4 in early post-MET cells is insufficient for determining the cell fate in reprogramming toward iPSCs. Several rounds of cell cycle after entry into MET may be required for establishment of a stable pluripotency molecular network, which enables symmetric cell divisions (Figure 6B).

Discussion

We have established human iRSC lines that are stably paused prior to entry into MET. These novel intermediately reprogrammed cells resume conversion to iPSC, when cultured at high density. iRSCs maintain a mesenchyme-like morphology with a unique gene expression profile characterized by transcription of exogenous OSKM, and endogenous *SOX2* and *NANOG*. Endogenous *OCT4* remains silent. Dramatic morphological change in MET, an early reprogramming event in iRSC-to-iPSC conversion, coincides with silencing of exogenous genes and activation of endogenous *OCT4*. Furthermore, visualization of endogenous *OCT4* through CRISPR/Cas9-mediated genome editing in iRSCs revealed *OCT4* activation occurs before entry into MET. In the early development stage of post-MET colonies, instability of *OCT4* expression resulted in commitment to cell differentiation. Thus, endogenous *OCT4* activation is a key event, but insufficient signature in reprogramming of iRSCs into iPSCs.

These newly established iRSC lines are the first intermediately reprogrammed cells able to resume reprogramming under defined culture conditions. In mice, it was thought that intermediately reprogrammed cells with silenced somatic genes, active SSEA1, and potential for conversion to iPSC state, are transient (Hochedlinger and Plath, 2009). Conversely, partially reprogrammed cells that express reprogramming transgenes and activating proliferation genes, but not endogenous pluripotency genes, have been established as stable cell lines. However, these lack potential for conversion to iPSC (Hochedlinger and Plath, 2009). Thus, establishment of human iRSCs as stable cell lines was unexpected. Importantly, our iRSCs are marked by expression of core endogenous pluripotency factors, SOX2 and NANOG, in addition to exogenous OSKM, but not endogenous OCT4. Expression of *SOX2* and *NANOG* in a pre-MET state in humans differs from expression of *Nanog* post-MET for iPSC maturation in mice (Brambrink et al., 2008; Wernig et al., 2008).

Resumption of reprogramming by activation of endogenous *OCT4* occurred simultaneously with silencing of OSKM. Notably, all OSKM transgenes integrated into the human genome were expressed in iRSCs. It is possible that retention of exogenous OSKM expression is required for iRSC stability in an intermediately reprogrammed state. If so, the iRSC state may be transient in integration-free iPSC generation systems (Stadtfeld and Hochedlinger, 2010). A similar situation was reported for human naïve-like iPSCs derived by doxycycline-induced temporal

expression of OSKM and Nanog(Buecker et al., 2010), suggesting that expression of OSKM functions to trap cells at an intermediately reprogrammed stage. Consistent with the idea, low level expression of Klf4 expression paused reprogramming into mouse iPSCs (Nishimura et al., 2014). Furthermore, partially reprogrammed human iPSCs, but not cell lines, emerged under similar conditions (Nishimura et al., 2014). It is interesting whether reprogramming efficiency of human iRSC to iPSC can be improved by temporal elevation of *KLF4*.

It remains unclear whether the reprogramming pathway from somatic cell to iPSC is fixed. Recently, it was demonstrated in the mouse that the pluripotency spectrum can encompass multiple, unique cell states, including an alternative somatic reprogramming path to iPSC through a Nanog-positive transient state, in addition to the preconceived Nanog-negative transient state (Tonge et al., 2014). In this context, human iRSCs, in which *NANOG* was precociously expressed prior to activation of endogenous *OCT4*, could exhibited a similar alternative reprogramming pathway. Interestingly, we found secondary conversion of isolated iRSCs that failed to convert during the first round experiment, suggesting requirement for an extra-cellular signal generated within a cellular micro-environment to trigger conversion to iPSC status (Data not shown). This elusive signal has not been identified yet. Therefore, human iRSCs can potentially provide new and rapid ways to unraveling the molecular mechanisms of alternative paths for somatic reprogramming to iPSC.

It has been debated whether reprogramming entails a hierarchic or stochastic process (Yamanaka, 2009). It is proposed that once OSKM are silenced and endogenous OCT4 activated in a stochastic manner, and then further reprogramming processes are likely to progress in a hierarchical manner (Buganim et al., 2012). By contrast to it, it has been proposed that endogenous *Oct4* activation is insufficient for progression of subsequent events in mouse somatic reprogramming(Greder et al., 2012; Silva et al., 2008). Notably, in human iRSC to iPSC reprogramming, time-lapse analyses of endogenous OCT4 activity clearly showed that GFP (OCT4)-positive cells may divide asymmetrically to form GFP-positive and negative cells soon after OCT4 activation. Thus, early OCT4-positive cells have epigenetic heterogeneity and instability through cell division before acquiring full pluripotency whereby a parent cell divides to form two daughter cells with the same pluripotent identity (Movie S4). Given that reprogramming is essentially a process of rewinding normal embryonic development, it is not unreasonable to expect switch from stem cells having traits of

asymmetric cell division to those having the property of symmetric cell division (Figure 7). We suggest that instability of endogenous OCT4 is linked to the cell characteristics of symmetric or asymmetric division.

In mice, MET occurs early in reprogramming of MEFs to iPSCs, characterized by up-regulation of epithelial genes, *E-Cadherin* and *Epcam*, and down-regulation of mesenchymal genes, *Snail* and *N-Cadherin* (Polo and Hochedlinger, 2010). In humans, we find that MET occurs at a later stage. We also show that, in humans, activation of endogenous *OCT4* occurs with entry into MET, while, in mice, endogenous *Oct4* is activated much (Li et al., 2010; Polo et al., 2012; Samavarchi-Tehrani et al., 2010)later. Together, the data suggest that MET is a checkpoint for entry into a primed state of pluripotency, while activation of endogenous OCT4/Oct4 commits the cell to continued reprogramming through activation of OCT4/Oct4-induced pluripotency molecular networks. In this context, in mice, the first wave of reprogramming driven by *c-Myc* and *Klf4* might induce MET-specific epigenetic and transcriptional changes rendering the cell in a primed state of pluripotency (Li et al., 2010). The second wave induced by *Oct4*, *Sox2*, and *Klf4* occurs much later and drives the cell into a naïve state of pluripotency(Polo et al., 2012). Thus, mouse somatic cell reprogramming-to-iPSC progresses from a primed to naïve state in chronological order. In humans, MET and second wave events occur simultaneously, late in the process. This implies that, in mice, the primed state is generated at a much earlier stage, with additional steps required prior to activation of endogenous *Oct4*. It is possible that, in humans, further epigenetic modification events are required between MET and activation of endogenous *OCT4* for generation of naïve iPSCs. We suggest that human iRSCs are a powerful tool for dissecting these latter events of reprogramming.

A key event in iRSC reprogramming-to-iPSC is entry to MET concomitant with gene silencing of exogenous OSKM and activation of endogenous *OCT4*. Moloney murine leukemia virus (MoMLV)-derived vectors (pMXs) were used for human iRSC generation. Negative regulatory elements in MoMLV have been mapped to long terminal repeats (LTRs) (Hotta and Ellis, 2008). One retroviral silencer (TRIM28) that binds to the primer-binding site of MoMLV plays an important role in transcriptional silencing (Wolf and Goff, 2007). Interestingly, Trim28 and Cnot3 were identified as transcriptional regulators that uniquely co-occupy putative gene promoters with c-Myc and Zfx, but not other pluripotency-associated factors,

including Oct4, Sox2, and Nanog (Hu et al., 2009). Therefore, the large Factor A complex that includes Trim28 might mediate silencing of exogenous OSKM without influencing endogenous *Oct4*. In mice, it has been shown that Sall4 is a master regulator of expression dosage of *Sall4* and *Oct4*, through a transcriptional regulation feedback loop (Yang et al., 2010). In humans, if similar events occur after silencing of exogenous OSKM by the action of Factor A complex, up-regulation of endogenous OCT4 might be induced as a direct consequence of exogenous *Oct4* silencing through an autocrine manner.

Many studies have shown that multiple epigenetic pathways are involved in activation of endogenous *OCT4*, especially changes in histone modification and DNA methylation. The observed DNA hypo-methylation at the endogenous *OCT4* promoter during transit from iRSC to iPSC might be a direct consequence of down-regulation of exogenous *Oct4*. It would be interesting to determine whether a repressive complex containing exogenous Oct4 can binds to the promoter region of endogenous *OCT4*.

With respect to putative clinical applications of somatic cell-reprogramming technologies, the use of transcription-factor-mediated direct lineage conversion or intermediate reprogramming without reversion to a fully pluripotent state has gained significant interest. Human iRSCs might represent a suitable cell source for directed conversion without full reprogramming, and for dissecting the molecular mechanisms involved in these alternative approaches. Furthermore, it has been reported that potential for genetic manipulation of human iPSCs and ESCs, desirable for repair of gene mutations and deficiencies, is limited due to the difficulty of single cell subcloning by dissociation-induced apoptosis(Cheng et al., 2012; Ohgushi and Sasai, 2011). Even using the Rho-associated kinase (ROCK) inhibitor, Y-27632 to protect against apoptosis (Watanabe et al., 2007), substantial effort is required to effectively subclone after gene transfer. In contrast, our mesenchyme-like iRSCs are readily expandable from a single cell after gene transfer using conventional delivery systems. Genome-edited clonal iRSCs are reprogrammable to iPSC state through simple modification of culture conditions. Thus, genome-edited iPSCs can be readily generated for future research and clinical application, as verified by the generation of OG-iRSCs, and OG-iRSC-derived iPSCs. Although OSKM-integrated iRSCs will not be suitable for direct use in regenerative therapy, they can provide an effective and powerful tool for investigating the causes of genetic diseases, and for screening of drugs following with disease modeling by genome editing.

Materials and Methods

Cell culture

Human fetal lung fibroblasts, TIG1 and TIG3 (JCRB Cell Bank, Japan) expressing *Slc7a1* (Takahashi et al., 2007) were cultured in Dulbecco's Modified Eagle's Medium (DMEM) (Sigma-Aldrich, USA) containing 10% FBS, and infected with *Oct4*, *Sox2*, *Klf4*, *c-Myc* and *DsRed* retroviruses. After reseeding onto inactivated MEF feeders 4 days after infection, the culture medium was changed to iPSC medium (DMEM/F12 HAM (Sigma-Aldrich, USA) supplemented with 20% KSR (Invitrogen, USA), L-glutamine, non-essential amino acids, 2-mercaptoethanol, and 10 ng/mL bFGF (Peprotech, USA)) at day 5. For establishment of iRSCs, self-renewable colonies were picked between days 20-30, and subcultured onto feeder-free Matrigel-coated dishes with MEF-conditioned iPSC medium (CM). iRSCs were stably maintained in CM through subculture at a density of 1.0 x 10^6 cells per Matrigel-coated 10 cm dish by conventional trypsin dissociation (Invitrogen). For conversion of iRSCs to iPSCs, 1.0 x 10^6 cells were plated onto feeder-free Matrigel-coated 3.5 cm dishes, and cultured for 6~10 days with CM. For analyzing efficiency of somatic reprogramming-to-iPSC, intermediately reprogrammed cells were dissociated at day 30 with the sequential treatment of CTK (0.25% Trypsin, 0.1 mg/ml Collagenase IV (Invitrogen), 20% KSR, and 1 mM MgCl in PBS) at 37˚C for 5 min, and trypsin at 37˚C for 5 min. 2.5 x 10^5 dissociated cells were plated onto feeder-free Matrigel-coated 3.5 cm dishes.

Genome editing

To visualize the expression of endogenous OCT4, "OCT4-2A-eGFP-PGK-Puro" plasmid (Hockemeyer et al., 2011) (Addgene, USA) was introduced into iRSCs.
CRISPR/Cas9 "pX330" (pX330-U6-Chimeric_BB-CBh-hSpCas9) (Cong et al., 2013) (Addgene) was used for higher efficiency of gene targeting. Target sequence oligos were inserted into pX330 following the protocol: www.genome-engineering.org. iRSCs (1.0×10^7) were trypsinized, washed with DMEM three times and suspended in 1mL DMEM. The mixture of OCT4-2A-eGFP-PGK-Puro (100μg) and pX330 (100μg) was transfected into iRSCs by electroporation (Gene Pulser Xcell System (Bio-Rad, USA); 250V, 500 μF, 4mm cuvette (BTX Electroporation Cuvette plus 640 (Harvard Apparatus, USA)). After electroporation, the cells were seeded onto two feeder-free Matrigel-coated 10 cm dishes, and cultured for 4 days at low density.

Following selection with 0.5 μg/ml Puromycin emerging resistant colonies were picked at day 5-6. Two days later, puromycin-selection (0.5 μg/ml) was performed again for purification of OCT4-2A-eGFP-PGK-Puro-transfected cell clones.
Subsequently, to excise the PGK-Puro cassette, OCT4-2A-eGFP-PGK-Puro-targeted iRSCs were treated with transient expression of Cre plasmid (800 ng) using Lipofectamin2000 (Invitrogen) according to the manufacture's protocol.

Generation of single cell-derived cDNA libraries

cDNA libraries from single cells were prepared as described previous (Saitou et al., 2002). Single cells were collected from DsRed-positive iRSCs, and from DsRed-negative cells at days 3, 6, and 10 from high-density cultures using a glass needle. Each single cell was transferred into a single tube containing lysis buffer, followed by sequential linker-ligation, reverse-transcription and cDNA amplification. The linear amplification of cDNA libraries was evaluated by spike RNAs (Data not shown).

RT-PCR

Total RNA was isolated from cultured cells using TRIzol (Invitrogen, USA), and cDNA synthesized with Superscript III reverse transcriptase following the manufacturer's instructions (Invitrogen). PCR was performed with Ex-Taq (Takara, Japan) using gene-specific primers. Q-PCR analyses were performed using Power SYBR Green PCR Master Mix according to the manufacturer's instructions (Applied Biosystems, USA). Samples containing10-100 cells were treated with CellAmp™ Direct RNA Prep Kit for RT-PCR (Takara) after picking and washing with PBS at 300g for 5 min. Subsequently, these samples were analyzed by Q-PCR with One Step SYBR® PrimeScript™ RT-PCR Kit II (Takara) according to the manufacturer's instructions. Primer sequences used are summarized in Table S1.

Bisulfite modification of DNA sequencing

Genomic DNA (1 μg) was bisulfite-treated with Imprint DNA Modification Kit following the manufacturer's instruction (SIGMA, USA). The promoter region of the human *OCT4* gene was PCR-amplified with specific primers (Table S1).

Microarray

For microarray analyses, 1 μg total RNA or amplified cDNA library was labeled according to manufacturer's instruction and hybridized to the human genome U133 Plus 2.0 Array (Affymetrix, USA). Raw data were normalized by the MAS 5.0 method using the bioconductor package on R program (http://www.r-project.org/). Heat maps of gene expression profile were visualized by MeV program (http://www.tm4.org/mev/). Hierarchical cluster was calculated by Pearson's correlation coefficient (r) and visualized by the pvclust package on R program. For scatter plot analyses, raw data were normalized by Robust Multichip Average method. Scatter plots were prepared using bioconductor package on R program, with the location of each gene indicated manually.

Immunocytochemistry

Cultured cells were fixed with 4% PFA (paraformaldehyde)/ PBS for 10 minutes at room temperature and pre-treated with blocking solution (3% BSA, 2% skimmed milk (DIFCO, USA) in PBST) at 4˚C overnight. The cells were immuno-reacted with primary antibodies (Table S2) overnight at 4˚C, and incubated with secondary antibody in blocking buffer for 1 hour.

Genomic PCR

Isolated genomic DNA of OG-iRSCs was analyzed by genomic PCR. The PCR products were amplified using GoTaq (Promega, USA). Annealing temperature and cycle numbers were set following the method of Stepdown PCR (Hecker and Roux, 1996). Primer sequences used are summarized in Table S1.

DNA sequencing

PCR products including GFP sequence or CRISPR/Cas9 target sequence into OG-iRSC genome were inserted pGEM-T (Easy) Vector (Promega) through TA cloning. The universal primers, T7 and SP6, and BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) were used for sequencing following the manufacturer's protocol. All samples were analyzed by Applied Biosystems 3130/3130xl Genetic Analyzer.

Teratoma

Teratomas, generated 6-8 weeks after transplantation of a cell suspension of 5.0 x 10^5 iRSCs or iPSCs under the kidney capsule of immunodeficient SCID mice (CLEA, Japan), were paraffin-embedded, sectioned, and stained with haematoxylin and eosin as described previously (Nagata et al., 2012).

Chromosome analysis

Cells were pre-treated with Colcemid (0.3 μg/ml), and trypsinized. The cell suspension was treated with 0.075M KCl for 8 minutes, and then fixed using 3:1 methanol:acetic acid. Chromosome spreads were prepared using an air-drying method, and stained with DAPI, and mounted with Slowfade Gold Antifade Reagnet (Invitrogen).

Stability of OCT4 expression

During reprogramming from OG-iRSCs to iPSCs, OCT4-GFP positive and negative cells in colonies were counted manually. Colonies consisting of more than 95% GFP-positive cells were defined “GFP-positive”. iPSC colonies were classified into 7 categories (1, 2-8, 8-25, 25-50, 50-100, 100-200, 200-) based on the number of composing cells. At least 30 colonies in each category were analyzed. Standard error was calculated from average of emergence frequency of GFP positive colonies in each category.

GFP expression level in cells

GFP intensity in GFP-positive and negative cells at Day 3 was measured with Adobe Photoshop CS6. GFP intensity in a small circle (15 μm diameter/cell) was calculated in each cell. GFP intensity in DsRed-positive cells around the DsRed-negative colonies were estimated as background. Stably GFP-positive cells at Day 10 were positive control.

Time-lapse imaging

To observe the sequential changes from iRSC into iPSCs in Timelapse imaging, we seed the cells into Laminin521 (Biolamina, Sweden) -coated 3.5 cm glass-bottom dishes (Matsunami Glass, Japan). Samples were set in FLUOVIEW (FV10i-w)

(Olympus, Japan) and 10-50 fields were selected randomly and recorded with image processing software (FV10-ASW 4.1).

ACKNOWLEDGMENTS

This work is supported by the MEXT (23390067).

R.T. contributed to CRISPR and iRSC experiments, K.H. and S.N. to iRSC experiments, J.A. to discussion and writing, and T.T. to experimental plan, iRSC experiments, discussion, and writing.

References

Brambrink, T., Foreman, R., Welstead, G. G., Lengner, C. J., Wernig, M., Suh, H. and Jaenisch, R. (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* **2**, 151–159.

Buecker, C., Chen, H. H., Polo, J. M., Daheron, L., Bu, L., Barakat, T. S., Okwieka, P., Porter, A., Gribnau, J., Hochedlinger, K., et al. (2010). A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells. *Cell Stem Cell* **6**, 535–546.

Buganim, Y., Faddah, D. A., Cheng, A. W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S. L., van Oudenaarden, A. and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209–1222.

Chen, J., Liu, H., Liu, J., Qi, J., Wei, B., Yang, J., Liang, H., Chen, Y., Chen, J., Wu, Y., et al. (2013). H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat. Genet.* **45**, 34–42.

Cheng, L.-T., Sun, L.-T. and Tada, T. (2012). Genome editing in induced pluripotent stem cells. *Genes to cells* **17**, 431–438.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. a, et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Sci. New York NY* **339**, 819–823.

Greder, L. V, Gupta, S., Li, S., Abedin, M. J., Sajini, A., Segal, Y., Slack, J. M. W. and Dutton, J. R. (2012). Analysis of endogenous Oct4 activation during induced pluripotent stem cell reprogramming using an inducible Oct4 lineage label. *Stem Cells* **30**, 2596–2601.

Hecker, K. H. and Roux, H. (1996). High and low annealing temperatures increase both specificity and yield in touchdown and stepdown PCR. *Biotechniques* **20**, 478–485.

Hirano, K., Nagata, S., Yamaguchi, S., Nakagawa, M., Okita, K., Kotera, H., Ainscough, J. and Tada, T. (2012). Human and mouse induced pluripotent stem cells are differentially reprogrammed in response to kinase inhibitors. *Stem Cells Dev.* **21**, 1287–1298.

Hochedlinger, K. and Plath, K. (2009). Epigenetic reprogramming and induced pluripotency. *Development* **136**, 509–523.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C. S., Gao, Q., Cassady, J. P., Cost, G. J., Zhang, L., Santiago, Y., Miller, J. C., et al. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. *Nat. Biotechnol.* **29**, 731–734.

Hotta, A. and Ellis, J. (2008). Retroviral vector silencing during iPS cell induction: an epigenetic beacon that signals distinct pluripotent states. *J. Cell. Biochem.* **105**, 940–948.

Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S. H. and Elledge, S. J. (2009). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev.* **23**, 837–848.

Kuroda, T., Tada, M., Kubota, H., Kimura, H., Hatano, S., Suemori, H., Nakatsuji, N. and Tada, T. (2005). Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. *Mol. Cell. Biol.* **25**, 2475–2485.

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., et al. (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7**, 51–63.

Nagata, S., Hirano, K., Kanemori, M., Sun, L. T. and Tada, T. (2012). Self-renewal and pluripotency acquired through somatic reprogramming to human cancer stem cells. *PLoS One* **7**, e48699.

Nishimura, K., Kato, T., Chen, C., Oinam, L., Shiomitsu, E., Ayakawa, D., Ohtaka, M., Fukuda, A., Nakanishi, M. and Hisatake, K. (2014). Manipulation of KLF4 expression generates iPSCs paused at successive stages of reprogramming. *Stem Cell Reports* **3**, 915–929.

Ohgushi, M. and Sasai, Y. (2011). Lonely death dance of human pluripotent stem cells: ROCKing between metastable cell states. *Trends Cell Biol.* **21**, 274–282.

Polo, J. M. and Hochedlinger, K. (2010). When fibroblasts MET iPSCs. *Cell Stem Cell* **7**, 5–6.

Polo, J. M., Anderssen, E., Walsh, R. M., Schwarz, B. a., Nefzger, C. M., Lim, S. M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., et al. (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151**, 1617–1632.

Saitou, M., Barton, S. C. and Surani, M. A. (2002). A molecular programme for the specification of germ cell fate in mice. *Nature* **418**, 293–300.

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H., Beyer, T. A., Datti, A., Woltjen, K., Nagy, A. and Wrana, J. L. (2010). Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7**, 64–77.

Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T. W. and Smith, A. (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol.* **6**, e253.

Stadtfeld, M. and Hochedlinger, K. (2010). Induced pluripotency: history , mechanisms , and applications. *Genes Dev.* **24**, 2239–2263.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872.

Tanaka, Y., Hysolli, E., Su, J., Xiang, Y., Kim, K.-Y., Zhong, M., Li, Y., Heydari, K., Euskirchen, G., Snyder, M. P., et al. (2015). Transcriptome signature and regulation in human somatic cell reprogramming. *Stem Cell Reports* **4**, 1125–1139.

Theunissen, T. W., van Oosten, A. L., Castelo-Branco, G., Hall, J., Smith, A. and Silva, J. C. R. (2011). Nanog overcomes reprogramming barriers and induces pluripotency in minimal conditions. *Curr. Biol.* **21**, 65–71.

Tonge, P. D., Corso, A. J., Monetti, C., Hussein, S. M. I., Puri, M. C., Michael, I. P., Li, M., Lee, D.-S., Mar, J. C., Cloonan, N., et al. (2014). Divergent reprogramming routes lead to alternative stem-cell states. *Nature* **516**, 192–197.

Watanabe, K., Ueno, M., Kamiya, D., Nishiyama, A., Matsumura, M., Wataya, T., Takahashi, J. B., Nishikawa, S., Nishikawa, S., Muguruma, K., et al. (2007). A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nat. Biotechnol.* **25**, 681–686.

Wernig, M., Lengner, C. J., Hanna, J., Lodato, M. a, Steine, E., Foreman, R., Staerk, J., Markoulaki, S. and Jaenisch, R. (2008). A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nat. Biotechnol.* **26**, 916–924.

Wolf, D. and Goff, S. P. (2007). TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* **131**, 46–57.

Yamanaka, S. (2009). Elite and stochastic models for induced pluripotent stem cell generation. *Nature* **460**, 49–52.

Yang, J., Gao, C., Chai, L. and Ma, Y. (2010). A novel SALL4/OCT4 transcriptional feedback network for pluripotency of embryonic stem cells. *PLoS One* **5**, e10766.

Figures

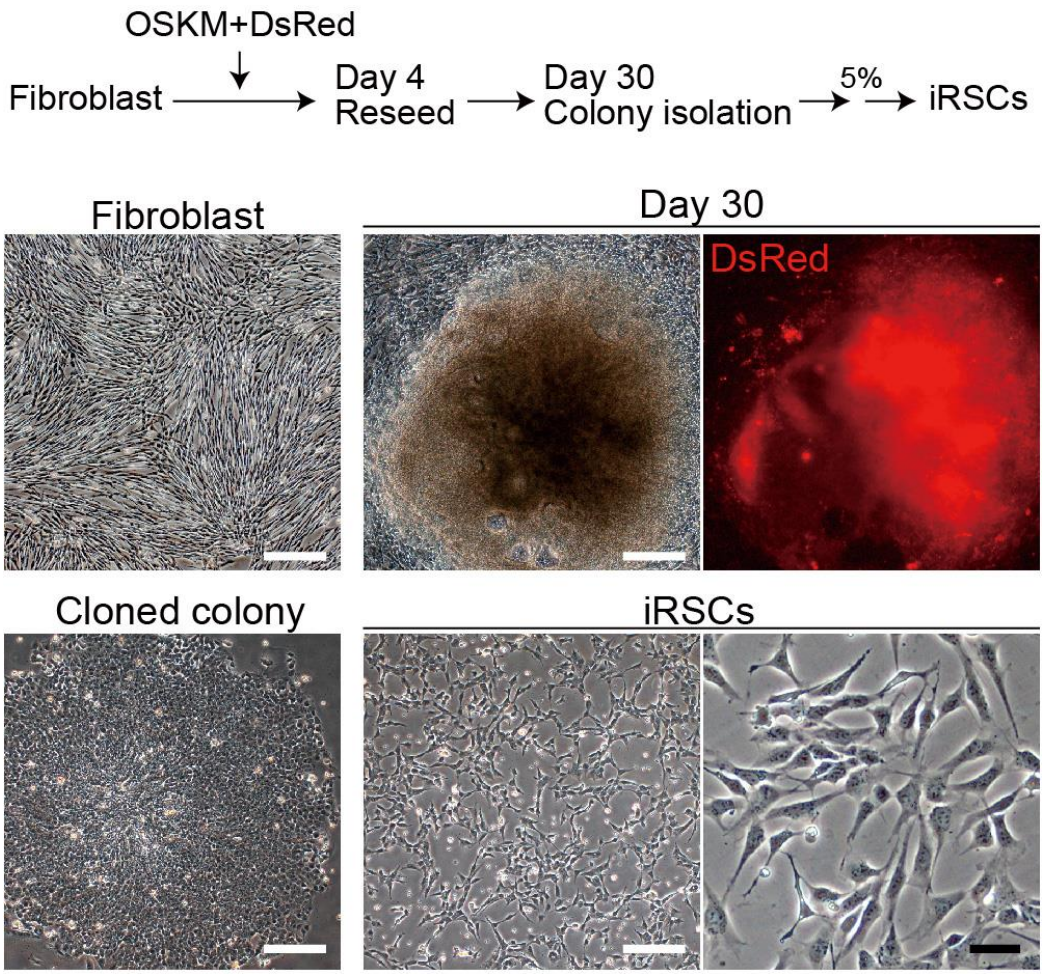


Figure 1. Method for Establishment of Intermediately Reprogrammed Stem Cells (iRSCs).

Human iRSCs derived from TIG1 fibroblasts by retroviral transduction of OSKM: Oct4, Sox2, Klf4, and c-Myc. DsRed is the reporter gene of exogenous OSKM. White scale bar: 300 μm. Black scale bar: 50 μm.

See also Figure S1.

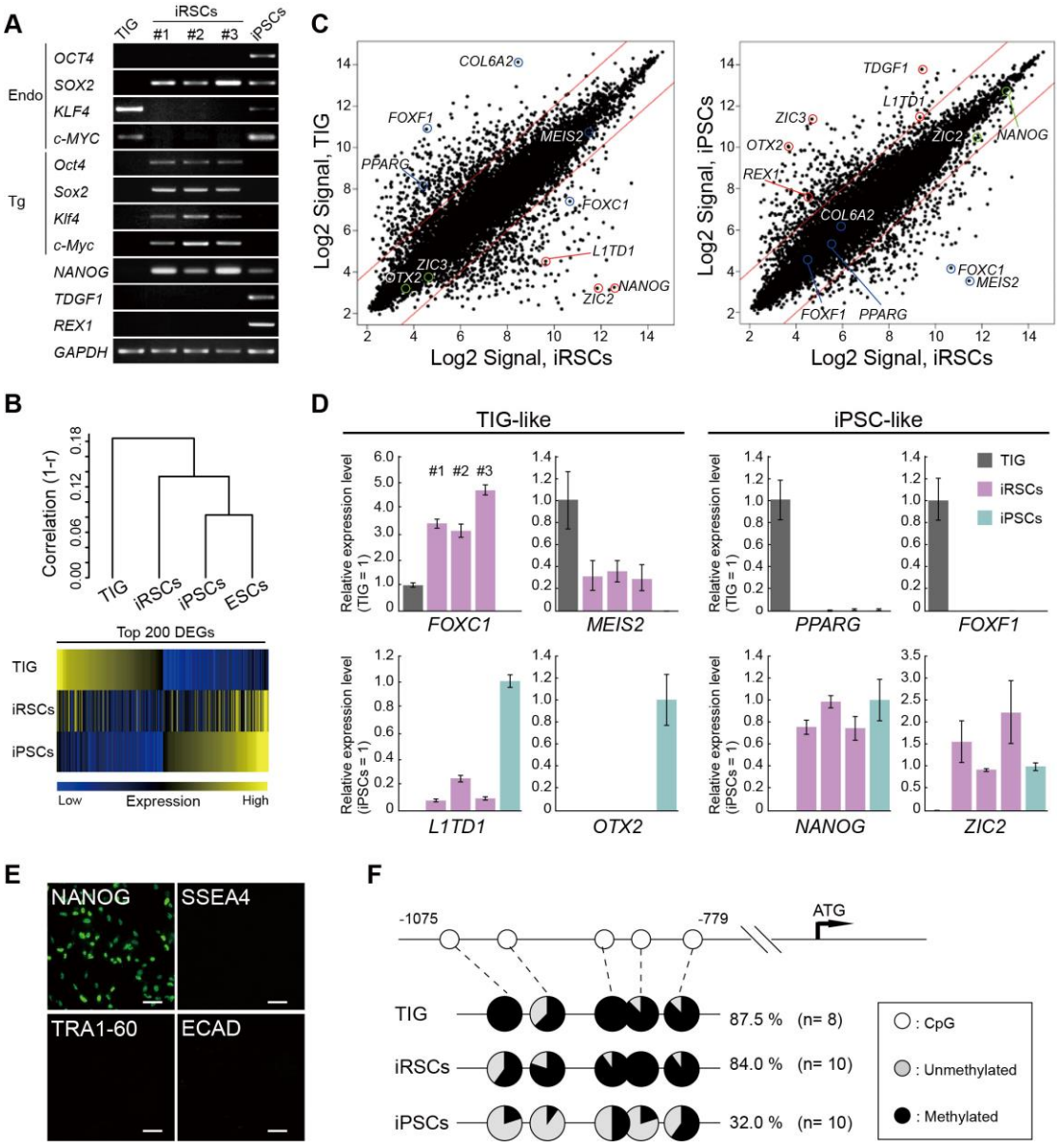


Figure 2. Characterization of Human iRSCs.

(A) Expression of pluripotency-associated genes. Endo: endogenous genes, Tg: transgenic genes. *NANOG*, *TDGF1*, and *REX1* are pluripotency genes. *GAPDH* is positive control.

(B) Hierarchical clustering (upper) and heatmap of global gene expression.

(C) Scatter plot of gene expression in TIG1 fibroblasts, iRSCs, and iPSCs. Red lines indicate 4-fold change in signal intensity. Blue circle; somatic genes, Red circle; pluripotency genes with greater than 4-fold change, Green circle; pluripotency genes with less than 4-fold change.

(D) Expression of pluripotency-associated genes in iRSCs.

(E) Immunocytochemistry of pluripotency-associated (NANOG, SSEA4, TRA1-60), and epithelial marker (ECAD) proteins. Immuno-positive cells are green. Scale bar: 100 μm.

(F) DNA methylation in promoter region of *OCT4* gene.

See also Figure S2.

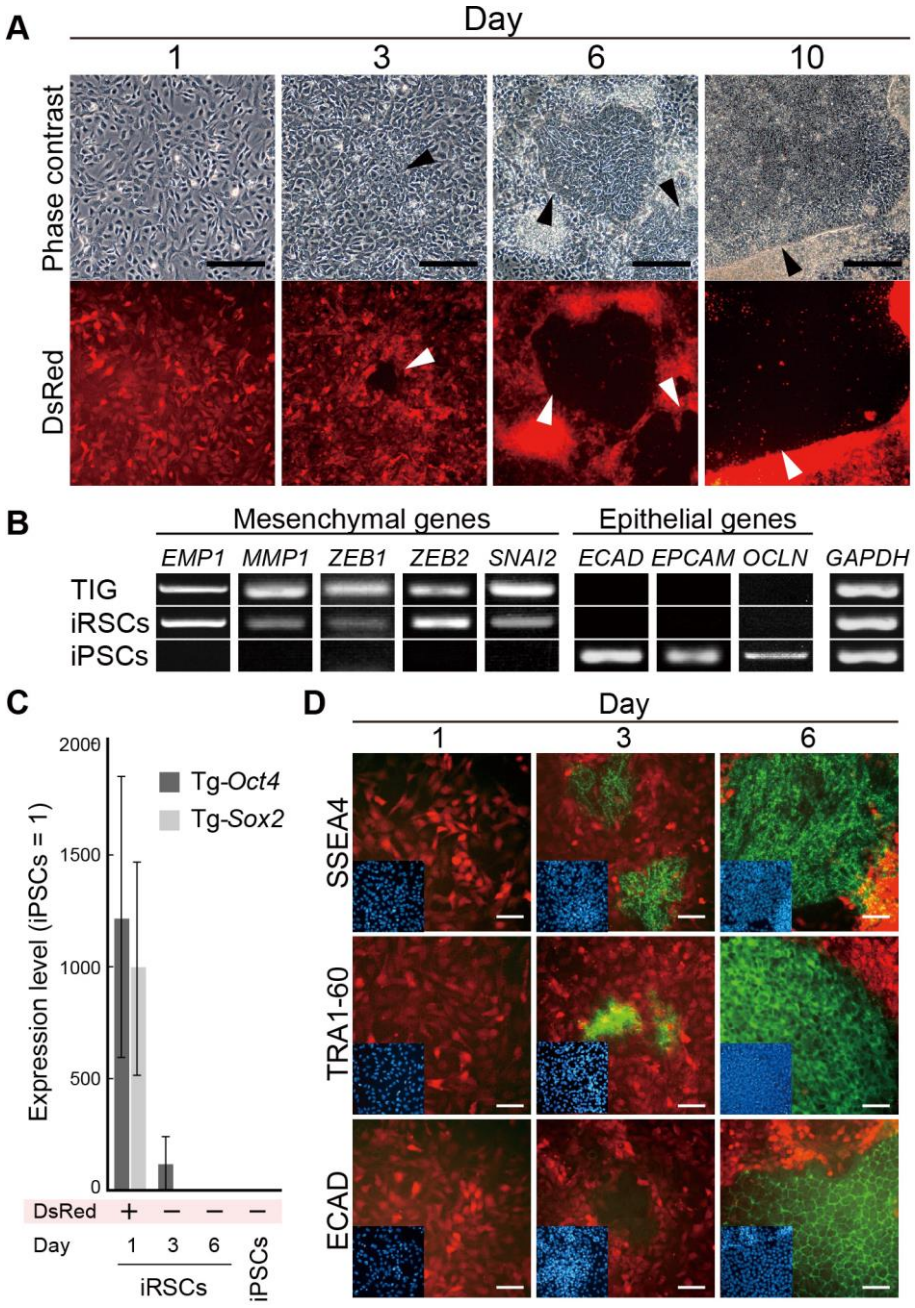


Figure 3. Conversion of iRSC to iPSC through Mesenchymal-to-Epithelial Transition (MET).

(A) Sequential change of iRSC to iPSC at day 1, 3, 6, and 10. Arrowheads: DsRed-negative iPSC colonies. Scale bar: 200 μm.

(B) Expression switches from mesenchymal to epithelial profile in MET.

(C) Transcriptional silencing of exogenous *Oct4* and *Sox2*. All reactions were performed in duplicate. Error bars represent the SEM of triplicates.

(D) Immunocytochemistry of pluripotency-associated (SSEA4, TRA1-60), and epithelial marker (ECAD) proteins. Immuno-positive cells are green. Scale bar: 50 μm. Red is DsRed. Blue (inset) is DNA.

See also Figure S3 and Movies S1-S3.

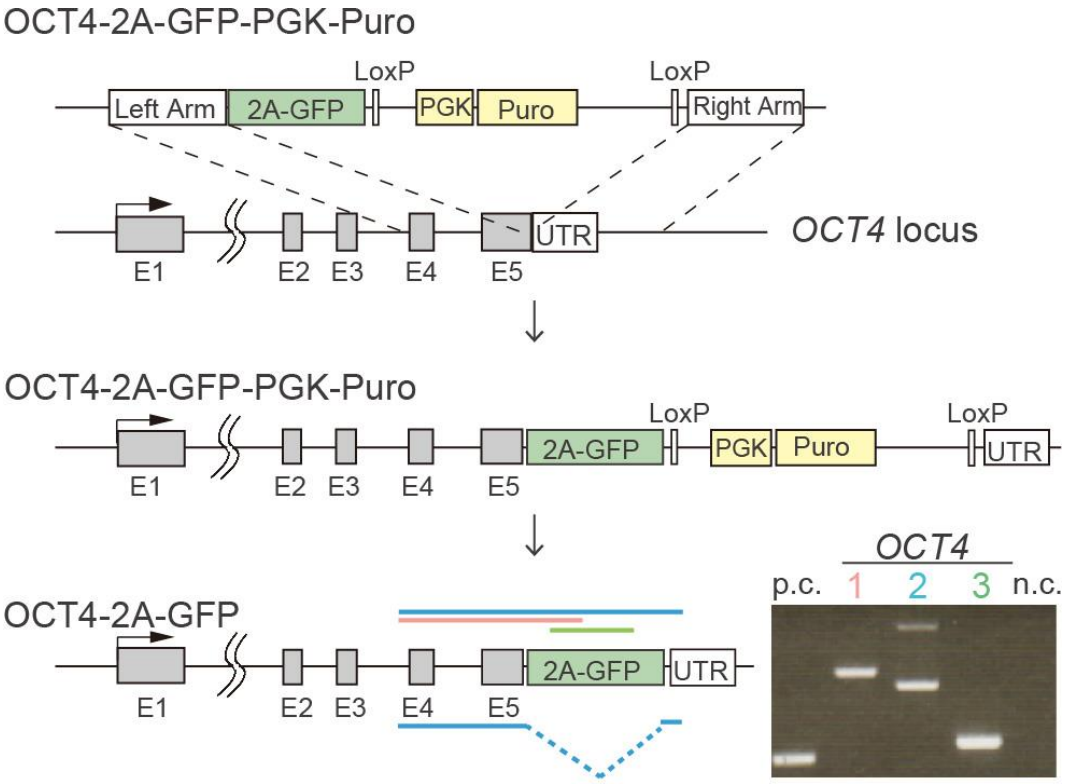


Figure 4. Genome editing for OG (OCT4-GFP) –iRSC.

Homologous recombination between the human *OCT4* and OCT4-2A-GFP-LoxP-PGK-Puro-LoxP was promoted by CRISPR/Cas9-mediation with a 20-base target sequence (5'-TCTCCCATGCATTCAAACTG-3'), including stop codon on Exon 5. The PGK-Puro cassette was eliminated after targeting by Cre-LoxP deletion. The image (bottom right) shows genomic PCR of OG-iRSC. The location of each PCR product is represented by the same color bar in the diagram of OCT-2A-GFP. E: Exon. Puro: Puromycin. p.c.: positive control. n.c.: negative control.

See also Figure S4.

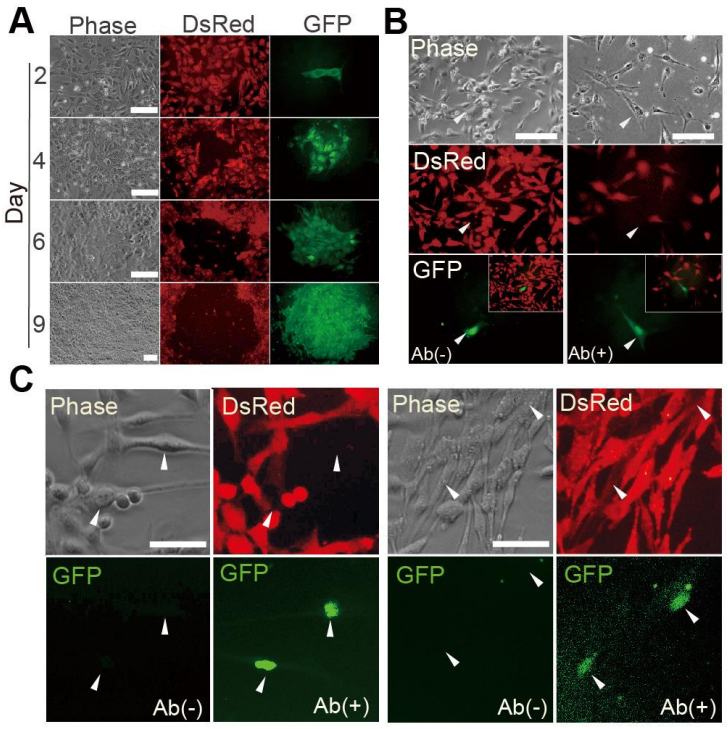


Figure 5. Reprogramming of OG-iRSC to iPSC.

(A) Visualization of OCT4 expression as GFP in reprogramming of OG-iRSC to iPSC at day 2, 4, 6, and 9. Phase: Phase contrast. Scale bar: 100 μm.

(B) Endogenous OCT4 activation prior to entering MET. Mesenchymal iRSCs were DsRed-negative and GFP-positive within 24 hours of high density culture. GFP was detected using an anti-GFP antibody at 24 hours (right panels). By 48 hours GFP was clearly detectable without antibody (left panels). Phase: Phase contrast. Ab: Antibody. (-): without antibody. (+): with antibody. Arrowheads: DsRed-negative and GFP-positive cells. Scale bar: 100 μm.

(C) Silencing of DsRed in GFP-positive iRSCs. Detection of GFP expression in DsRed-negative pre-MET iRSCs with anti-GFP antibody, even in GFP-negative iRSCs without anti-GFP antibody (left panels). About 10% GFP-positive cells with anti-GFP antibody were faintly positive for DsRed (right panels). Phase: Phase contrast. Scale bar: 50 μm. Ab: Antibody. (-): without antibody. (+): with antibody. Arrowheads: GFP-positive cells.

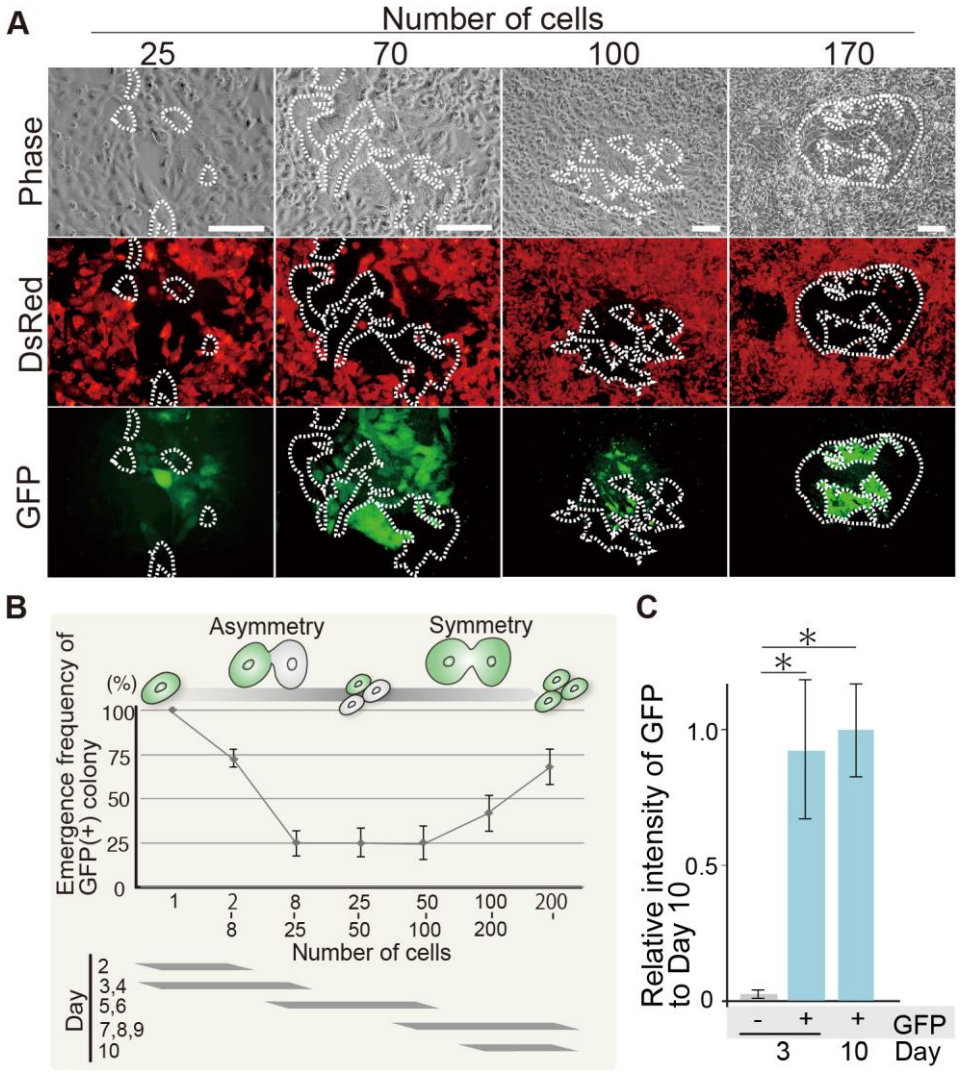


Figure 6. Instability of OCT4-GFP expression.

(A) Instability of OCT4 expression in post-MET iRSCs toward reprogramming into iPSCs. White dotted lines denote cells within reprogramming colonies that are DsRed-negative and GFP-negative without immunostaining. Phase: Phase contrast. Scale bar; 100 μm.

(B) Emergence frequency of GFP-positive colonies in early post-MET reprogramming. GFP(+) colony: more than 95% cells are GFP(+) in DsRed(-) colony. Cell division switched from asymmetric to symmetric in early post-MET iRSCs (upper diagram). Bottom gray bars show timing of switching from asymmetric to symmetric cell division after resumption of reprogramming.

(C) Comparison of relative intensity of GFP between GFP-positive and negative cells.

Relative GFP intensity to Day10 (GFP stable) GFP-positive cell shows in GFP-positive and negative cells at Day3 (GFP unstable). Data are represented as mean ± standard deviation, GFP-positive cells, GFP-negative cells at day 3; n=20. GFP-positive cells at day 10; n=30.

See also Figure S5, 6, 7 and Movies S4.

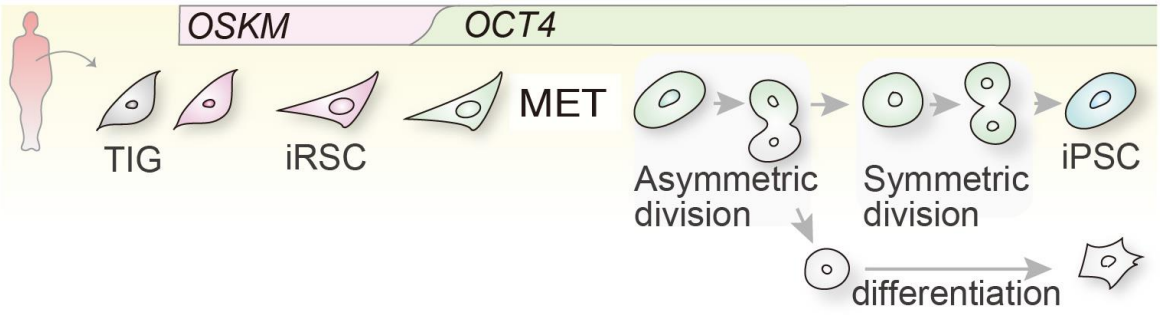


Figure 7. Route of human somatic reprogramming to iPSC.

In human somatic reprogramming, silencing of exogenous OSKM genes and activation of *OCT4* occur before the key event of MET. In early post-MET, asymmetric division of OCT4-positive cells generates OCT4-negative cells. iPSC phenotype is stabilized after establishment of symmetric division of OCT4-positive cells.

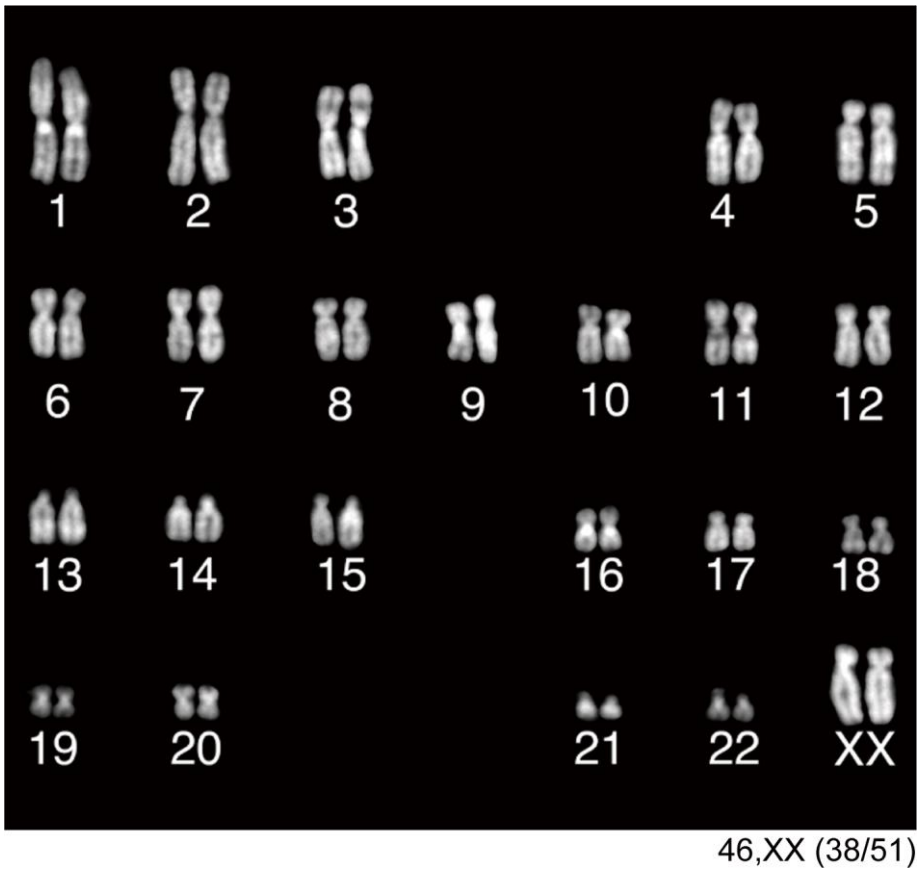


Figure S1. Representative normal karyotype of iRSC derived from female TIG1 fibroblasts, Related to Figure 1.

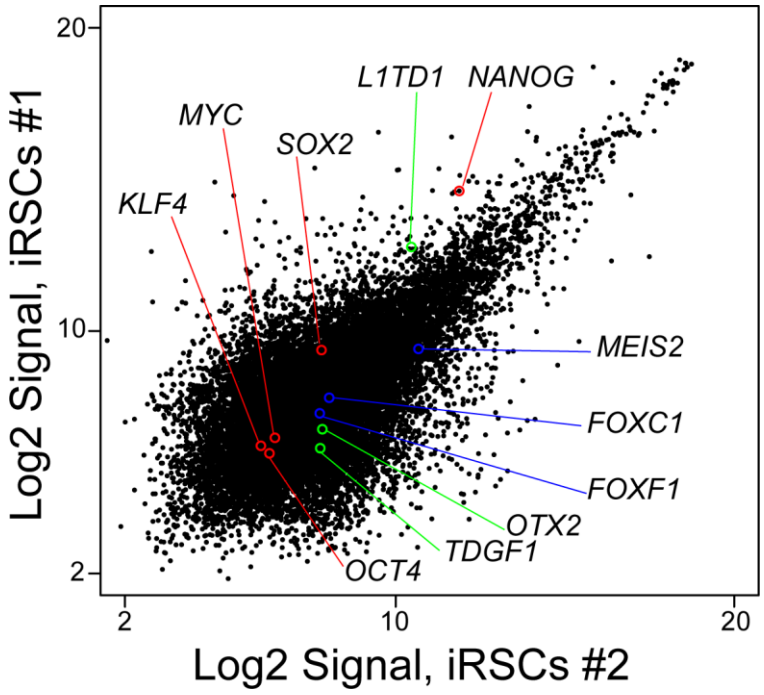


Figure S2. Comparison of global gene expression profile between iRSC lines. Related to Figure 2. Scatter plot of gene expression in iRSC-line #1 and #2. Red circle; endogenous OCT4, SOX2, KLF4, MYC and NANOG, Blue circle; somatic genes, Green circle; pluripotency genes.

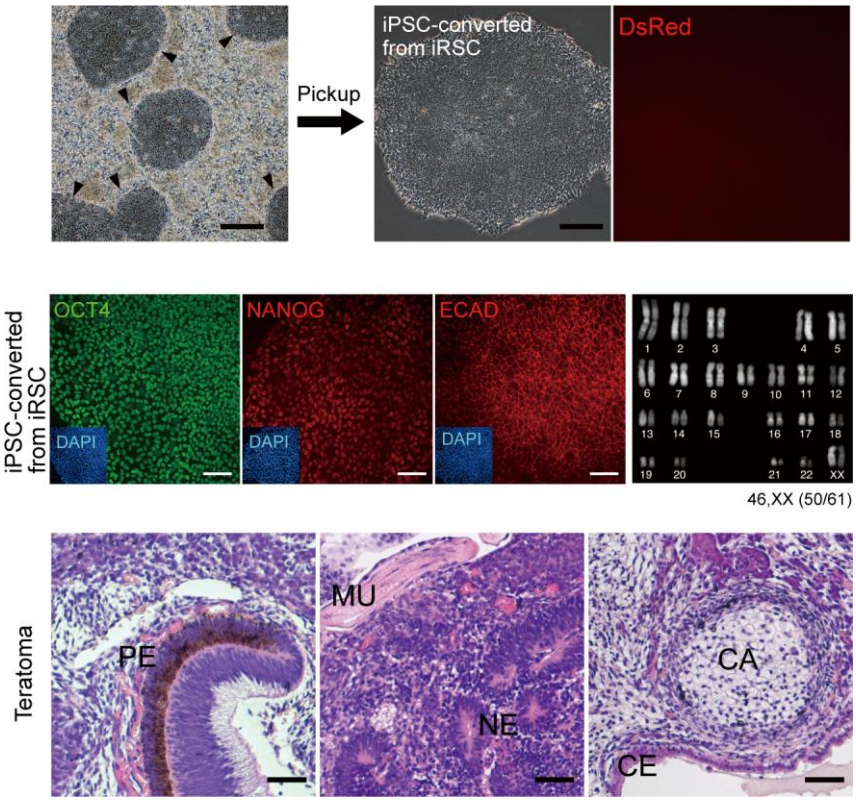


Figure S3. Properties of iPSC-derived iRSCs, Related to Figure 3.

Several colonies of iPSCs (black arrow heads) are observed in iRSCs at day 10 after culture at high cell density. Isolated iPSC colony is negative for the exogenous DsRed reporter gene. Scale bar: 300μm (upper row). Expression of pluripotency-associated proteins is visualized by immuno-staining. Scale bar: 100μm (middle row). iRSC-converted iPSC has normal karyotype, 2n=46, XX (right in middle row). Teratomas are generated by transplantation of iRSCs into kidney capsule. Scale bar: 50μm. PE; pigmented epithelial cells, MU; muscle, NE; neuronal ectoderm, CA; cartilage, CE; cilia-epithelial cell.

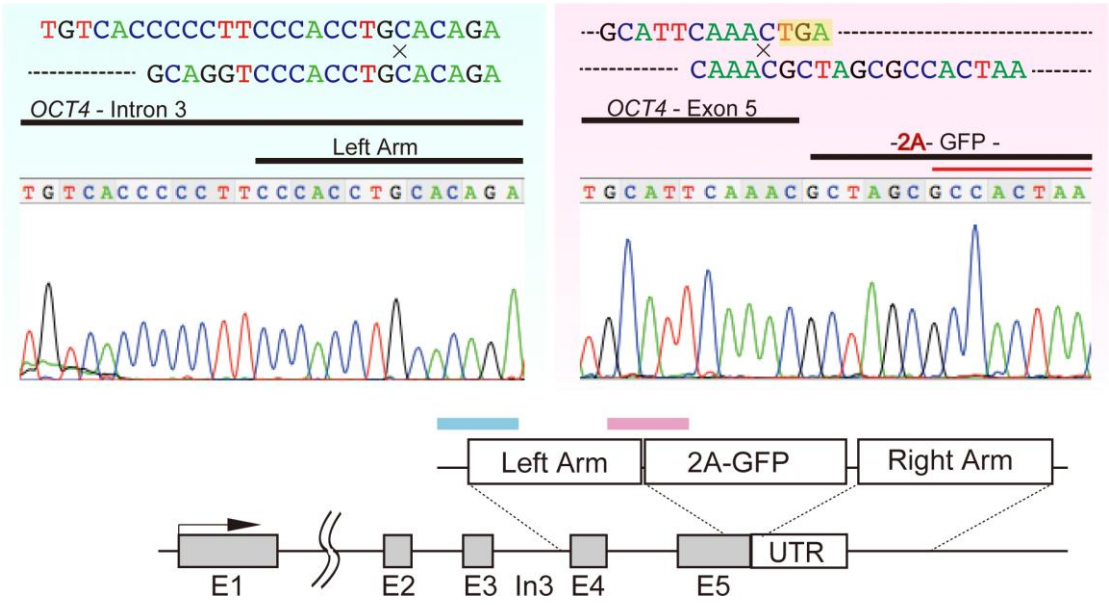


Figure S4. Correct insertion of OCT4-GFP. Related to Figure 4.

DNA sequence analysis shows that OCT4-GFP was inserted into the correct site. Left (blue) figure corresponds to the blue bar (around intron 3 of OCT4) and Right (pink) figure corresponds to the pink bar (around of 2A-GFP) in the bottom diagram. E: Exon. In: Intron.

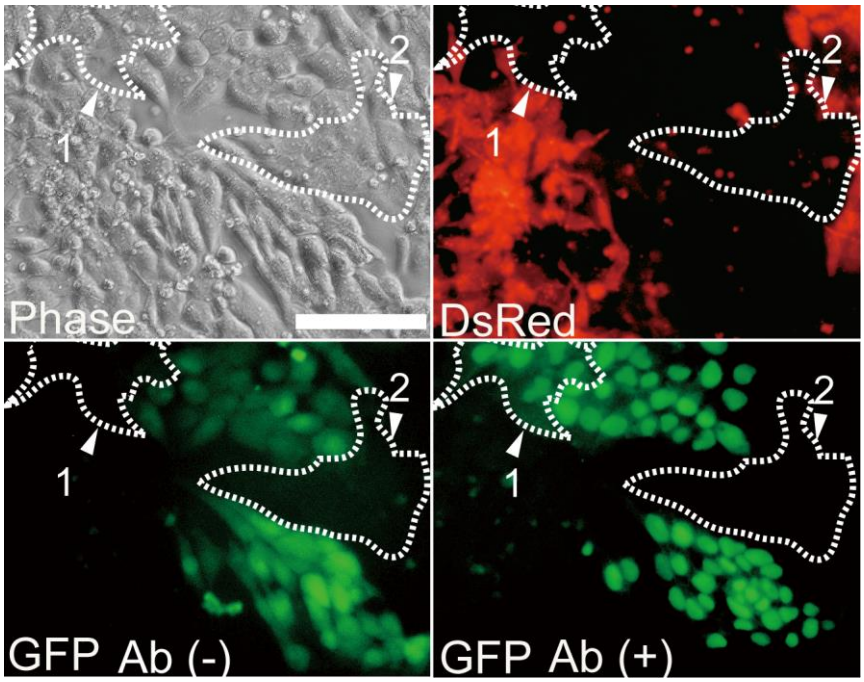


Figure S5. Instability of OCT4-GFP detected with anti-GFP antibody. Related to Figure 6.
DsRed, GFP without antibody, and GFP with antibody were detected in the same microscopic field five days after high-density culture of OG-iRSCs. White dotted lines (regions 1 and 2) surround DsRed-negative reprogramming OG-iRSC colonies. Cells in region 1, which were GFP negative without antibody became GFP positive with antibody, while cells in region 2 were GFP negative even with antibody, verifying the emergence of OCT4-negative post-MET iRSCs. Phase: Phase contrast. Scale bar: 100μm. Ab: Antibody. (-) No antibody. (+) with antibody.

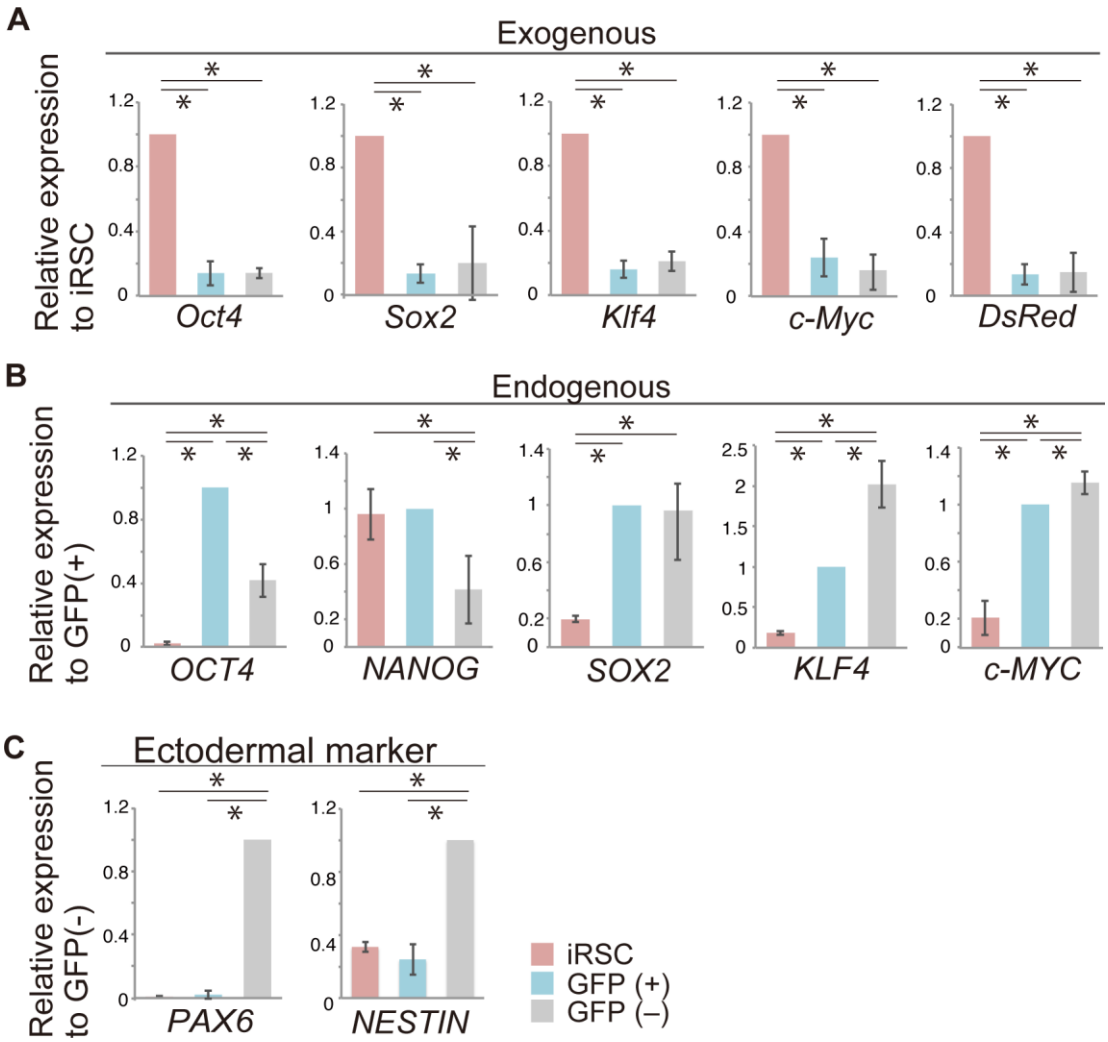


Figure S6. Gene expression in DsRed(-)/GFP(-) post-MET cells. Related to Figure 6.

(A) Expression levels of exogenous Oct4, Sox2, Klf4, c-Myc and DsRed in DsRed-negative/GFP-negative cells by Q-PCR. Data are represented as mean ± standard deviation, Oct4, Klf4, c-Myc, DsRed: n = 4. Sox2: n=8

(B) Expression levels of endogenous OCT4, NANOG, SOX2, KLF4 and c-MYC in DsRed-negative/GFP-negative cells by Q-PCR. Data are represented as mean ± standard deviation, OCT4, NANOG, SOX2, c-MYC: n=4. KLF4: GFP(+) n=6, GFP(-) n=5.

(C) Expression levels of ectodermal marker genes, PAX6 and NESTIN in DsRed-negative/GFP-negative cells by Q-PCR. Data are represented as mean ± standard deviation, n=4.

iRSC (red bar): OCT4-GFP iRSC, GFP(+) (blue bar): DsRed-negative/GFP-positive cells, GFP(-) (gray bar): DsRed-negative/GFP-negative cells.

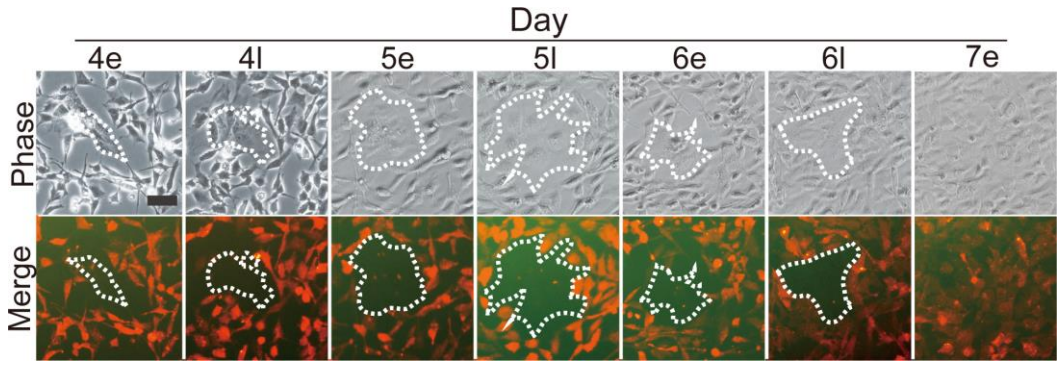
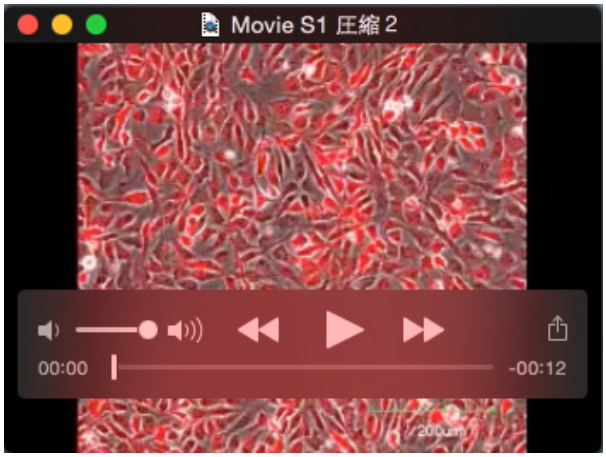


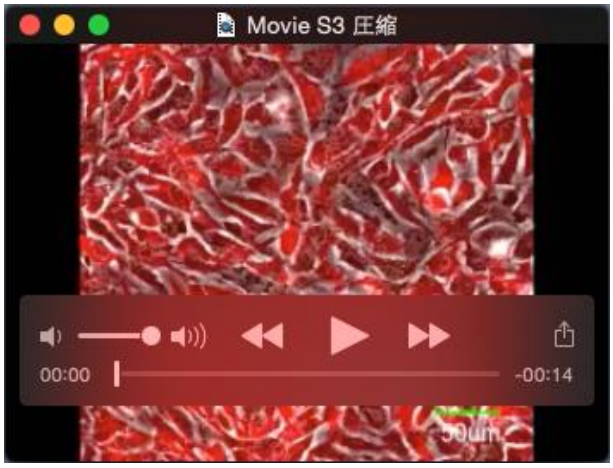
Figure S7. No conversion of OCT4-GFP(-) cells to (+) cells. Related to Figure 6.

Sequential change of the same OCT4-GFP-negative colony observed at 12 hourly intervals between day 4-7. Phase: phase contrast, Merge; merge of DsRed and GFP. Region enclosed with white dotted line: OCT4-GFP negative colony. Phase: Phase contrast. Scale bar: 50μm. e: early. l: late.

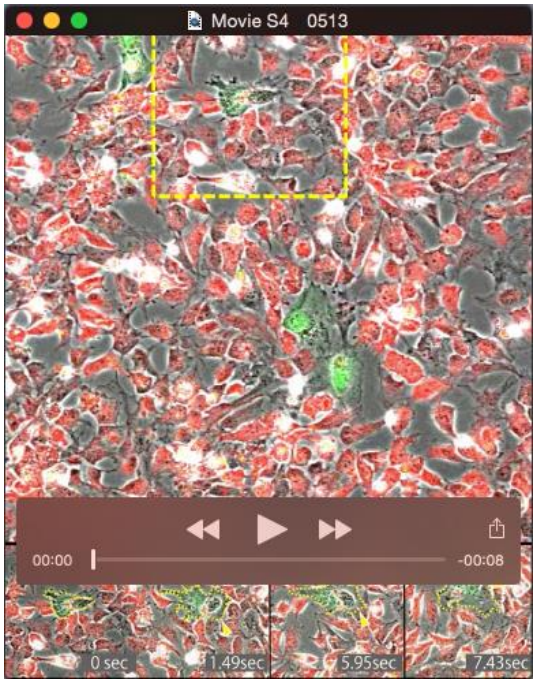


Movie 1. Conversion of iRSCs toward iPSCs between days 1-3, Related to Figure 3. Time-lapse images are captured from one day after reseeding, taken at a rate of 18 frames per 5 hours.

Movie 2. Conversion of iRSCs toward iPSCs between days 3-6, Related to Figure 3. Time-lapse images captured from three day after reseeding, taken at a rate of 18 frames per 5 hours.



Movie 3. High frame rate images in initial stages of iRSC-to-iPSC conversion during entry into MET, Related to Figure 3. Time-lapse images captured between days 1-2 at a rate of 11 frames per hour.



Movie 4. Generation of GFP-positive and negative cells through asymmetric cell division, Related to Figure 6. Time-lapse images are captured between days 2-4 at a rate of 3 frames per hour.

Table S1. Primers

	Target	Forward	Reverse
RT-PCR analysis	*OCT4*	GCACTGTACTCCTCGGTCCCTTTCCC	CTTCCCTCCAACCAGTTGCCCCAAAC
	SOX2	GGGAAATGGGAGGGGTGCAAAAGAGG	TTGCGTGAGTGTGGATGGGATTGGTG
	KLF4	ACTCGCCTTGCTGATTGTCT	GAACGTGGAGAAAGATGGGA
	c-Myc	CGGGCGGGCACTTTG	GGAGAGTCGCGTCCTTGCT
	Tg-Oct4	CCCATGGTGGTGGTACGGGAATTC	AGTTGCTTTCCACTCGTGCT
	Tg-Sox2	CCCATGGTGGTGGTACGGGAATTC	TCTCGGTCTCGGACAAAAGT
	Tg-Klf4	CCCATGGTGGTGGTACGGGAATTC	GTCGTTGAACTCCTCGGTCT
	Tg-c-Myc	CAGAGGAGGAACGAGCTGAAGCGC	GACATGGCCTGCCCGGTTATTATT
	NANOG	AAAGAATCTTCACCTATGCC	GAAGGAAGAGGAGAGACAGT
	TDGF1	CCGCCCCGACTGGGGTTTGT	AAGCAGGAGCAAGGCGTCCAG
	REX1	TTAGCTAGGCCTGGTTGCAT	GGGCTCTTGCTGTTATCCAG
	ECAD	GAGCTTGTCATTGAGCCTGGCA	TGGGCAAATGTGTTCAGCTCAGC
	EPCAM	AATGTGTGTGCGTGGGACGA	GGTAAAGCCAGTTTCAAGCTGC
	OCLN	TCACACCCCAGACGATGTCTTCA	GGGAGGCTGGTAGATCATCACA
	EMP1	GCTGTCCCTCATGGAGACCT	AAGTGGGATAGGCAGGGTCC
	MMP1	CGCTGGGAGCAAACACATCT	TTCATGAGCCGCAACACGAT
	ZEB1	TCTGACTCTCAGCTCCTGCACT	GCCAGGCACCCTGTTAGGCA
	ZEB2	TTGTTACCTTCGCTGTGAATTGAA	GGACACAGCCTACTAGCCCAA
	SNAI2	TTCAAATGCATACCACAAATGCAAT	AGTGGTTTGGTACTAATCATGAAGC
	GAPDH	CTTCTTTTGCGTCGCCAGCCGAG	CAGCCTTGACGGTGCCATGGAA
Quantitative RT-PC**R** **analysis**	*OCT4*	GAGTGAGAGGCAACCTGGAG	ACACTCGGACCACATCCTTC
	SOX2	TAAGTACTGGCGAACCATCT	AAATTACCAACGGTGTCAAC
	KLF4	TGCCAAGGGGGTGACTGGAAGT	TCTTCCCTCCCCCAACTCACGG
	c-Myc	CGGGCGGGCACTTTG	GGAGAGTCGCGTCCTTGCT
	NONOG	TGGGATTGGGAGGCTTTGCT	TGAAACACTCGGTGAAATCAGGG
	Tg-Oct4	ACTAGCATTGAGAACCGTGTG	GGTGTCCCTGTAGCCTCATAC
	Tg-Sox2	GCGCCCAGTAGACTGCACA	ACATGTGCGACAGGGGCAG
	Tg-Klf4	TCCCTAGAGGCCCATTTGAG	GGGGACTTGTGACTGCATCT
	Tg-c-Myc	TGTGGAGAAGAGGCAAACCCC	TCCAAGACGTTGTGTGTCCG
	Tg-DsRed	TACGTGAAGCACCCCGCCGA	GCCGCCGTCCTCGAAGTTCA
	FOXC1	TTCGAGTCACAGAGGATCGG	TAGTTCGGCTTTGAGGGTGT
	FOXF1	TCTCGCTCAACGAGTGCTTC	TCATGCTGTACATGGGCTTG
	L1TD1	TATACTGTTGGGGGAGGGCT	CTTGCCATCTTTTCCCGTGC
	MEIS2	CAGTGTAGCTTCACCTGGTACA	TGGGCTGTACTATTCTTCTTCTGG
	OTX2	GTATGGACTTGCTGCACCCC	AAACCATACCTGCACCCTCG
	PPARG	CCTGCAGGAGCAGAGCAAA	GCCCTCGGATATGAGAACCC
	ZIC2	CAGAACGGCTTCGTTGACTC	AAGTCCCGGGTGGAGTTGAA
	NESTIN	TCCAGGAACGGAAAATCAAG	GCCTCCTCATCCCCTACTTC
	PAX6	GTCCATCTTTGCTTGGGAAA	TAGCCAGGTTGCGAAGAACT
	GAPDH	CTGGCCAAGGTCATCCATGAC	CCATCCACAGTCTTCTGGGTG
Bisulfite sequencing analysis	*Bis-OCT4*	GAAGGGGAAGTAGGGATTAATTTT	CAACAACCATAAACACAATAACCAA
Genomic PCR	*OCT4*-lane1	GTCACAGACCCCTGTGATGC	AAGTCGTGCTGCTTCATGTG
	OCT4-lane2	GTCACAGACCCCTGTGATGC	CAACCAGTTGCCCCAAACTC
	OCT4-lane3	CGTAAACGGCCACAAGTTCA	GGGGTGTTCTGCTGGTAGTG
	GAPDH	GCCTCACTCCTTTTGCAGAC	TGAGCTTGACAAAGTGGTCG

Table S2. Antibodies for Immunocytochemistry

	Target	Dilution ratio	Manufacturer
Primary Antibodies	NANOG	1:200	ReproCELL, Japan
	SSEA4	1:500	Hybridoma Bank, USA
	TRA1-60	1:500	Millipore, USA
	ECAD	1:200	Takara, Japan
	OCT4	1:50	Santa Cruz Biotechnology, USA
	EGFP	1:1000	Nacalai, Japan
Secondary Antibody	Alexa 488	1:500	Molecular Probes, USA