Cell Research (2009) 19:296-306.

www.nature.com/cr

ORIGINAL ARTICLE

Generation of iPS cells using defined factors linked via the self-cleaving 2A sequences in a single open reading frame

Lijian Shao[1, *], Wei Feng[1, *], Yan Sun[1], Hao Bai[1], Jun Liu[1], Caroline Currie[1], Jaejung Kim[2], Rafael Gama[2], Zack Wang[1], Zhijian Qian[3], Lucy Liaw[1], Wen-Shu Wu[1]

[1]Center for Molecular Medicine, Maine Medical Center Research Institute, 81 Research Drive, Scarborough, Maine 04074, USA; [2]Functional Genomics Facility, Division of Biological Science, the University of Chicago, Chicago, Illinois 60637, USA; [3]Section of Hematology/Oncology and the Cancer Research Center, the University of Chicago, Chicago, Illinois 60637, USA

Generation of induced pluripotent stem (iPS) cells from somatic cells has been achieved successfully by simultaneous viral transduction of defined reprogramming transcription factors (TFs). However, the process requires multiple viral vectors for gene delivery. As a result, generated iPS cells harbor numerous viral integration sites in their genomes. This can increase the probability of gene mutagenesis and genomic instability, and present significant barriers to both research and clinical application studies of iPS cells. In this paper, we present a simple lentivirus reprogramming system in which defined factors are fused in-frame into a single open reading frame (ORF) via self-cleaving 2A sequences. A GFP marker is placed downstream of the transgene to enable tracking of transgene expression. We demonstrate that this polycistronic expression system efficiently generates iPS cells. The generated iPS cells have normal karyotypes and are similar to mouse embryonic stem cells in morphology and gene expression. Moreover, they can differentiate into cell types of the three embryonic germ layers in both *in vitro* and *in vivo* assays. Remarkably, most of these iPS cells only harbor a single copy of viral vector. This system provides a valuable tool for generation of iPS cells, and our data suggest that the balance of expression of transduced reprogramming TFs in each cell is essential for the reprogramming process. More importantly, when delivered by non-integrating gene-delivery systems, this re-engineered single ORF will facilitate efficient generation of human iPS cells free of genetic modifications.

Keywords: iPS cells, embryonic stem cells, self-cleaving 2A sequences, somatic cell reprogramming, polycistronic lentiviral vector

Cell Research (2009) **19**:296-306. doi: 10.1038/cr.2009.20; published online 24 February 2009

Introduction

Human embryonic stem (ES) cells have significant therapeutic potential for treatment of various diseases, but the generation of these cells raises ethical concerns. Therefore, generation of patient-specific (isogenic) pluripotent stem cells by somatic cell reprogramming approaches has been considered a viable solution. Reprogramming somatic cells into ES cell-like cells has been achieved by either transferring the somatic cell nuclei into enucleated eggs, or by fusing somatic cells with pluripotent ES cells [1]. However, these methodologies were limited by low efficiency and the requirement for fresh human oocytes, or by abnormal somatic/ES hybrid chromosomes.

Recently, reprogramming of murine and human somatic cells to pluripotent ES cell-like cells, termed induced pluripotent stem (iPS) cells, was first achieved successfully by simultaneous viral transduction of four transcription factors (TFs) together (*KLF4*, *OCT4*, *SOX2*, and *c-MYC* or *OCT4*, *SOX2*, *NANOG*, and *LIN28*) [2-6]. This technology does not require embryos or oocytes, thereby facilitates the generation of patient- and disease-specific

*These two authors contributed equally to this work.
Correspondence: Wen-Shu Wu
Tel: +1-207-886-8139; Fax: +1-207-886-8179
E-mail: wuw@mmc.org
Received 9 January 2009; revised 6 February 2009; accepted 6 February 2009; published online 24 February 2009

pluripotent stem cells that are valuable for personalized cell transplantation therapy without concern for immune rejection. Thus, this methodology effectively provides a potential alternative to the current source of ES cells. In addition, iPS cell technology has great application potential for understanding of disease mechanisms, drug screening, tissue engineering, and toxicology.

However, reprogramming by lentiviral/retroviral infection of defined TFs is inefficient (from 0.001% to 0.1 %) and requires very high transduction efficiency. Mouse embryo fibroblasts (MEFs) need at least 30% of retrovirus transduction efficiency [7] and an average of 15 different proviral copies [8] to be reprogrammed into iPS cells. In addition, although integrated TFs become transcriptionally silenced over time through de novo DNA methylation, they can be spontaneously reactivated during cell culture and differentiation. These problems associated with the current lentivirus/retrovirus-mediated reprogramming approaches raised safety issues for both basic research and clinical application [2-6]. Although virus-free mouse iPS cells were recently generated by adenovirus-mediated gene delivery and DNA transfection approaches [9, 10], efficiency of iPS cell generation is significantly lower (0.0006% - 0.0015%), compared with the retroviral or lentiviral infection approaches. Thus, lentivirus/retrovirus-mediated reprogramming methods are still major reprogramming approaches for generation of iPS cells, at least for basic research purpose.

Here, we devised a simple reprogramming system in which defined factors are in-frame fused into a single open reading frame (ORF) via self-cleaving 2A peptides [11], and are controlled by a CMV promoter in a lentivirus vector. Our data demonstrat that this polycistronic expression system efficiently reprograms somatic cells into iPS cells. The iPS cells generated by this system express stem cell markers and exhibit pluripotency as demonstrated by their ability to differentiate into cell types of the three embryonic germ layers in embryoid bodies (EBs) and teratomas, and by their high contribution to mouse chimeras. Notably, most of iPS cells generated by our system only contain a single copy of viral vector. Because we engineered defined TFs into a single ORF, we have simplified construction of non-integrating vectors encoding the defined reprogramming TFs, which should facilitate generation of human iPS cells free of genetic modifications.

Results

A polycistronic lentiviral expression vector for the generation of iPS cell lines

The reprogramming was achieved by simultaneous viral transduction of defined TFs (i.e. Oct4, c-Myc, Sox2, and Klf4) together into somatic cells. Therefore, each of the four TFs was randomly integrated into chromosomes, and expression of TFs in each cell was independent. Only those cells that harbor all of the viral vectors and have optimal expression of TFs are capable of reprogramming [8]. These reasons may account for low reprogramming efficiency of the current reprogramming approaches.

To test whether optimized expression of the defined TFs in each cell would improve reprogramming of somatic cells into iPS cells, we constructed a polycistronic lentiviral expression vector for optimized expression of four defined TFs (*KLF4*, *OCT4*, *SOX2*, and *c-MYC*) in which these four factors were fused as a fusion gene (*KOSM*) in a single ORF via self-cleaving 2A sequences [11], and this ORF was driven by a common CMV promoter (Figure 1A). In addition, a humanized GFP marker was cloned downstream of the *KOSM* gene that was separated by an internal ribosome entry site (IRES) to enable us to track transgene expression during the reprogramming process and differentiation of iPS cells. The self-cleaving 2A sequences derived from the foot-and-mouth disease virus are very small in size and can efficiently cleave polycistrons at specific site [11].

To verify that the *KOSM* fusion gene product can be processed efficiently into individual proteins, we transfected the expression vector pLentG-KOSM into 293T cells, and the correct size of each protein was confirmed by western blot analysis and compared with each protein translated from each individual expression vector (Figure 1B). Next, we prepared the lentiviruses from this vector and infected MEFs. We found that the lentiviruses carrying the *KOSM* fusion gene were efficiently transduced into MEFs, and that the GFP marker was clearly visualized by microscopy and flow cytometry (Figure1C and 1D).

Generation of iPS cells using defined TFs in a single ORF

To assess whether ectopic expression of the *KOSM* fusion gene can efficiently induce iPS cells, we introduced the *KOSM* fusion gene into MEFs via a lentivirus vector. ES cell-like cell colonies appeared from 3.15% of the infected cells (GFP^+) 6 to 8 days after viral infection (Figure 2A, upper panel), as expected. To study these ES cell-like cell colonies in more detail, we picked up 24 colonies at day 15 after viral infection, and expanded them for further analysis. We performed alkaline phosphatase (AP) activity staining, and found that 10 of these colonies (42%) were positive for AP (Figure 2A, lower panel). Immunofluorescence staining analysis further indicated that 8 of these colonies were positive for ES cell

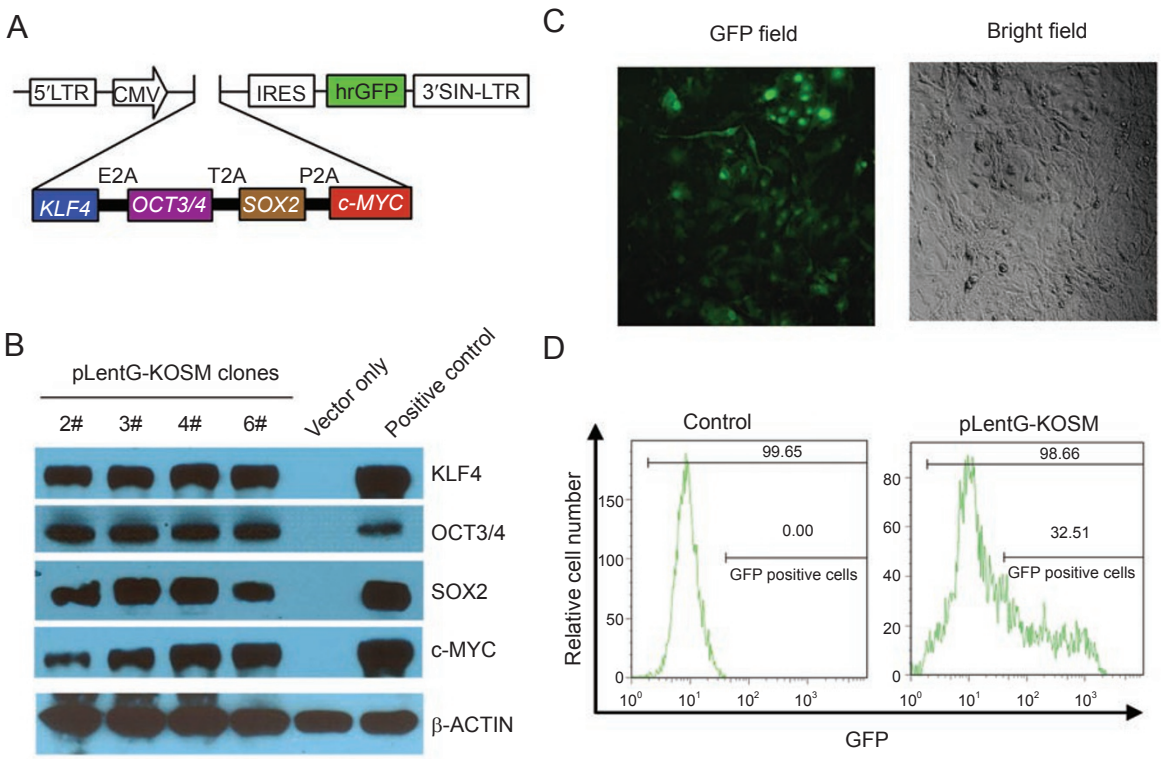

Figure 1 Generation of a lentiviral polycistronic expression vector for reprogramming. **(A)** Schematic representation of lentiviral expression vector (pLentG) and 2A-linked fusion gene (*KOSM*). Four defined TFs (*KLF4*, *OCT3/4*, *SOX2*, and *c-MYC*) were fused in-frame via 2A sequences and coexpressed as a single ORF. CMV promoter was used to drive the 2A-linked cassette and GFP marker, which were separated by an internal ribosome entry site (IRES) sequence. **(B)** Expression of *KOSM* fusion gene in 293T cells. pLentG-KOSM vector was transiently transfected into 293T cells. Correct processing of each factor was confirmed by western blot analysis. Expression vectors containing *KLF4*, *OCT3/4*, *SOX2*, and *c-MYC* cDNA were included as positive control. **(C)** Representation of GFP image in MEFs. GFP fluorescence was observed 3 days after infection with lentiviruses containing the *KOSM* fusion gene. **(D)** Representation of flow cytometry analysis of GFP-expressing MEFs 3 days after infection.

markers: OCT4, SOX2, and NANOG (Figure 2B). In addition, we synthesized cDNA from iPS cells and confirmed gene expression of multiple pluripotency markers in these cell colonies by RT-PCR (Figure 2C). Based on these data, we calculated the reprogramming efficiency and determined that 1.04 ± 0.03 % of infected MEFs (GFP^+) were reprogrammed to ES cell-like cell colonies that express transcripts of multiple pluripotency markers and show positive staining for ES cell markers: AP, OCT4, SOX2, and NANOG.

To further understand how similar the generated iPS cells were to mouse ES cells, we analyzed global gene-expression profiles of mouse iPS cells (clone 3) and ES cells by using mouse expression arrays (Agilent whole mouse genome oligo microarray). Scatter plot analysis demonstrated a tight correlation in gene expression between iPS cells and mouse ES cells (Figure 2D). The linear coefficient of determination (γ^2, the square of the correlation coefficient) between iPS cells and mouse ES cells was approximately 0.99, indicating that the generated iPS cells were similar to mouse ES cells in global gene expression.

To examine gene-silencing of the integrated transgene (KOSM) in iPS cells generated by this reprogramming system, we monitored GFP marker expression during the reprogramming process. Because the GFP marker and the KOSM fusion gene are driven by a common CMV promoter, GFP expression reflects expression of the KOSM transgene in iPS cells. By observing GFP expression in individual iPS cell colonies, we found that GFP expression was evident in ES cell-like cell colonies at day 6 after viral infection of the *KOSM* fusion gene (Figure 2E, upper panels), but almost undetectable at day 12 after infection (Figure 2E, lower panels) and throughout subsequent culturing (date not shown). To further confirm that the *KOSM* transgene was silenced in these

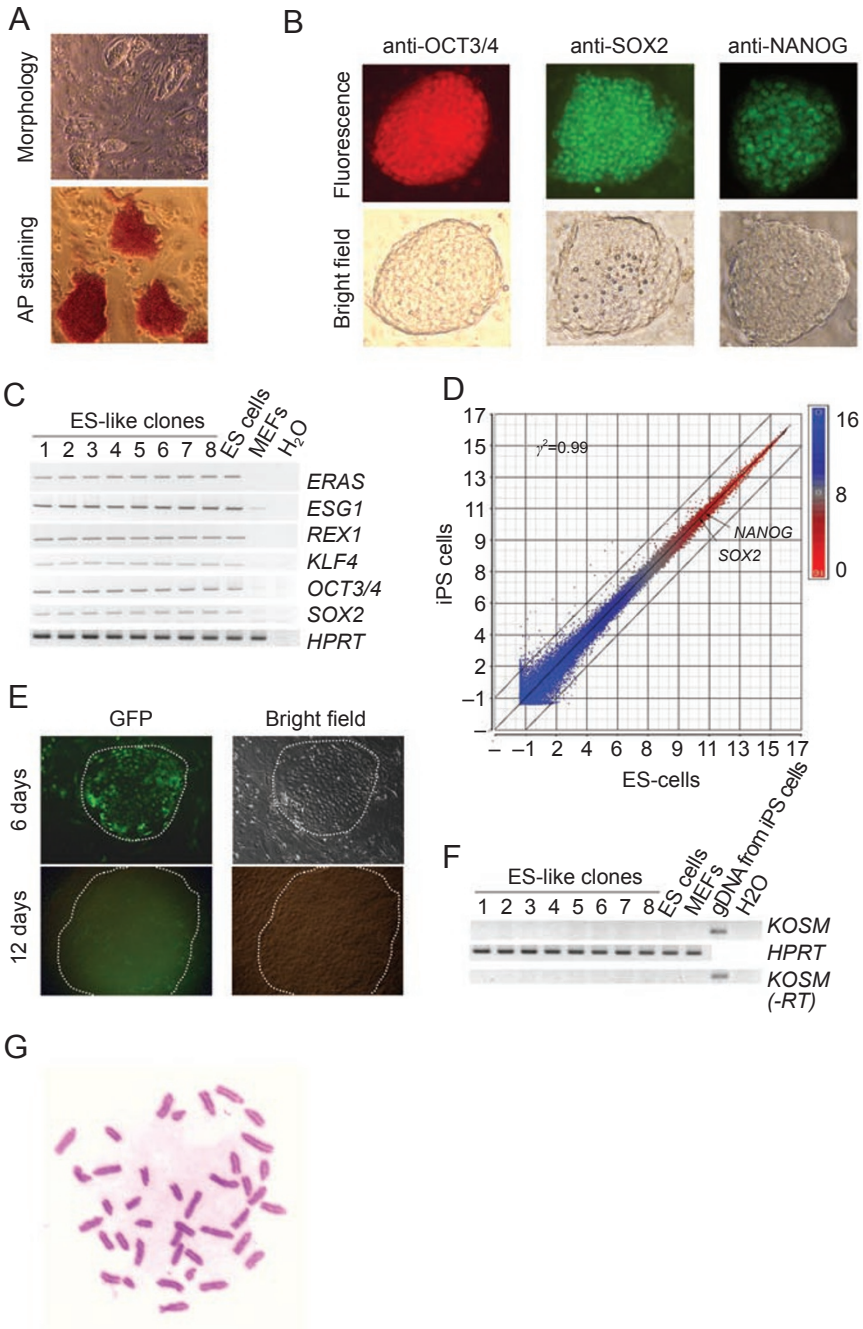

Figure 2 Characterization of induced iPS cell colonies generated from MEFs infected with the lentiviruses containing *KOSM* fusion gene. **(A)** Morphology (upper panel, 40× magnification) and AP staining (lower panel, 100× magnification) in induced ES cell-like cell colonies. **(B)** Staining of pluripotency markers (OCT3/4, SOX2, and NANOG) in induced cell colonies (200× magnification). **(C)** RT-PCR analysis of pluripotency marker gene expression in induced colonies. **(D)** Global gene-expression patterns were compared between mouse iPS cells (clone 3) and ES cells using mouse expression arrays. Arrows indicate expression levels of endogenous Nanog and Sox2. Lines indicate the diagonal and log2 fold changes between the two samples. **(E)** Fluorescence images of GFP for tracking expression of *KOSM* transgene during reprogramming (200× magnification). **(F)** RT-PCR analysis of *KOSM* transgene in the induced iPS cell colonies. cDNA samples were from (C), and we also synthesized cDNA from each iPS cell colony without RT. The genomic DNA (gDNA) from iPS cells transduced with lentiviruses containing the KOSM transgene served as control for PCR amplification of the KOSM fusion gene. *HPRT* amplification was the same as (C) and was included to serve as positive controls for RT-PCR experiment. The KOSM fusion gene was amplified from gDNA extracted from iPS cells, but not from cDNAs synthesized from iPS cells with or without RT. **(G)** Karyotyping analysis of a representative colony (1 000× magnification).

cell colonies, we examined the transcripts of the *KOSM* transgene in iPS cells. We used the same synthesized cDNAs, which were used in Figure 2C, and a pair of specific primers for the E2A linker between *KLF4* and *OCT3/4* sequences (Supplementary information, Table S1, for primer sequences). As expected, KOSM transgene was amplified by PCR from genomic DNA (gDNA) extracted from generated iPS cells (Figure 2F). In contrast, RT-PCR analysis indicated that the transcripts of the *KOSM* transgene were not detectable in these iPS cell colonies (Figure 2F). Together, these results confirmed that the *KOSM* transgene was indeed silenced in the iPS cell clones. In addition, we examined the karyotype of these induced cell colonies by chromosomal G-band analysis. We showed that these cells had a normal karyotype after being cultured for 5 passages (Figure 2G), suggesting that reprogramming of somatic cells by this system does not lead to chromosome abnormalities.

In vitro differentiation capacity of iPS cells generated by the defined TFs in a single ORF

Although 8 of the selected 24 colonies were positive for AP and expressed other pluripotent markers (Figure 2A-2C), it remained questionable as to whether these induced colonies are true pluripotent iPS cells with full differentiation capacity. To address this question, we determined the differentiation potential of these cell colonies by using floating cultivation to form embryoid bodies (EBs) *in vitro*. These cell colonies usually formed ball-shaped EB structures after 9 days in suspension culture with differentiation medium (Figure 3A). Notably, all of the EBs derived from the iPS cells were negative for GFP expression (Figure 3A, right panel), indicating that the *KOSM* transgene remained silenced during EB formation. We transferred these EB-like structures derived from each cell colony to gelatin-coated cell culture plates. After another 7 days of cultivation, the cells were detected to be positive for α-smooth muscle actin (α-SMA, mesoderm) and albumin (endoderm) by immunofluorescence staining (Figure 3B, left and middle panels).

To induce neuronal differentiation in these induced cell colonies, we added all-trans retinoic acid (1 μM) in culture medium and continued the culture for an additional 7 days. By immunofluorescence staining analysis (Figure 3B, right panel), we detected cells positive for neuron-specific βIII-tubulin (a marker of ectoderm). By RT-PCR analysis, we further confirmed that these differentiated cells expressed transcripts for *AFP* (endoderm), *NESTIN* (ectoderm), and *α-SMA* (mesoderm). Moreover, the beating of the cardiac muscle was observed during differentiation of these induced cell colonies (Supplementary information, Video S1). Together, these results indicate that these ES cell-like cell colonies not only expressed pluripotency markers (Figure 2A-2C), but also could differentiate into ectoderm-, mesoderm-, and endoderm-derived three germ layers *in vitro* (Figure 3B and 3C).

In vivo pluripotency of iPS cells generated by defined factors in a single ORF

To assess *in vivo* pluripotency of the iPS cells generated by the polycistronic lentiviral expression vector, we injected subcutaneously iPS cells (clone 1, 2, 3, and 4) into the flanks of syngeneic C57/BL6 mice. Five weeks after injection, teratomas from these iPS cell clones became palpable. Histological examination demonstrated that the teratomas (from clone 1) contained various tissues derived from the three embryonic germ layers (Figure 4A), including neural tissues (ectoderm), cartilage (mesoderm), and gut-like epithelium (endoderm). In addition, the immunocytochemistry showed that sections from the teratomas were stained positive by antibodies recognizing three lineage-specific markers (Figure 4B): βIII-tubulin (neuron-specific, ectoderm), α-SMA (muscle-specific, mesoderm), and CK18 (epithelial-specific, endoderm). The positive cells stained by these three antibodies exhibit typical structures for neurons, smooth muscles, and epithelium, respectively. Furthermore, we injected iPS cells (clone 1, labelled with EGFP) into blastocysts and found that donor iPS cells contributed to embryo development of chimeric embryos (Figure 4C). In addition, we carefully examined the chimeric embryo and found that GFP expression is evident in brain and liver (Figure 4D) as well as many other tissues (data not shown), indicating that iPS cells generated by the KOSM fusion gene can contribute to different tissues in chimeric embryos. Overall, these results indicate that iPS cells generated by expression of the KOSM fusion gene have full capacity to differentiate into the three embryonic germ layers *in vivo*.

iPS cells generated by defined TFs in a single ORF harbor fewer copy numbers of viral vector and showed diminished reactivation of exogenous genes in their differentiated cells

One major problem with the current lentivirus/retrovirus-mediated reprogramming approach is the spontaneous reactivation of the transduced transgene in iPS cells during differentiation [2-6], which increases the risk of tumorigenesis and obviously hinders basic research and clinical application. Our data showed that GFP marker expression remained silenced in the EBs derived from the iPS cells generated by this polycistronic lentiviral expression vector (Figure 3A).

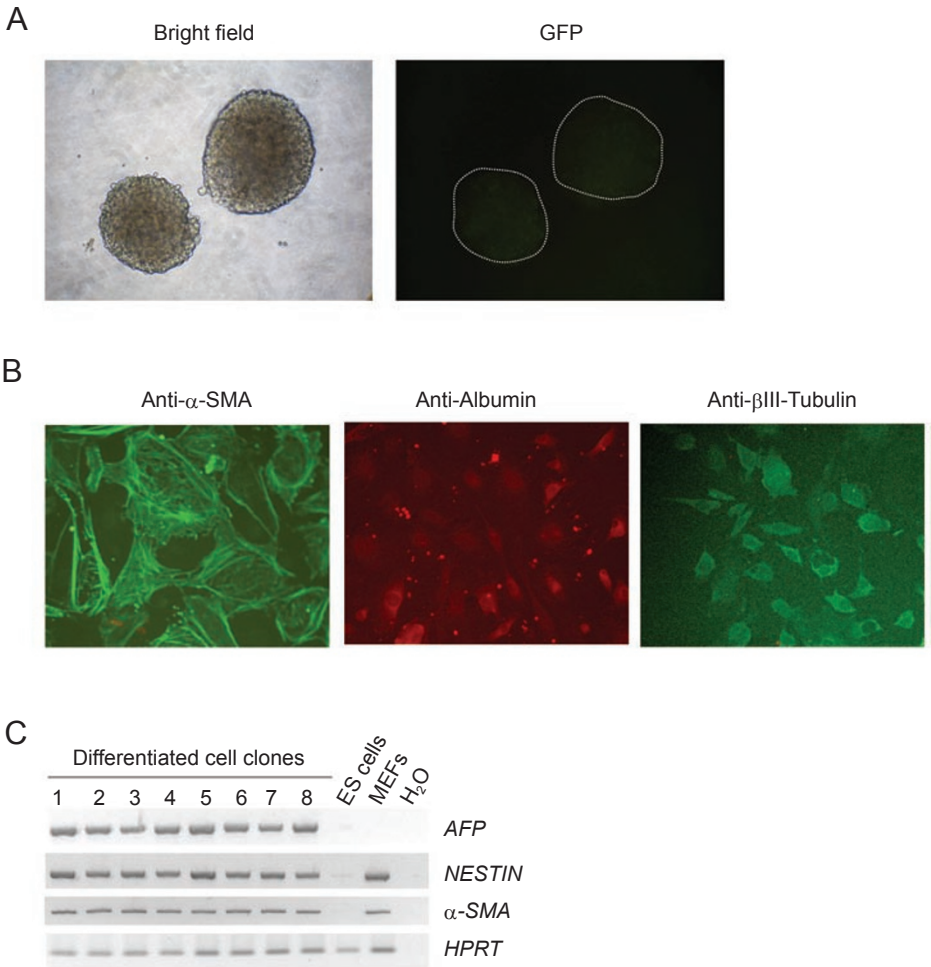

Figure 3 *In vitro* differentiation of generated iPS cells. **(A)** Silencing of GFP marker in embryoid bodies derived from iPS cell colonies (200× magnification). **(B)** Immunofluorescence staining of differentiated cells derived from iPS cells with antibodies specific for lineage-specific markers: α-SMA, Albumin, and βIII-Tubulin (200× magnification). **(C)** RT-PCR analysis of differentiation markers for three germ layers in differentiated cells derived from iPS cells.

To further assess how often the transduced *KOSM* transgene was reactivated in differentiated cells derived from these iPS cell colonies, we analyzed the transcripts of *KOSM* in differentiated cells by RT-PCR by using primers specific for the E2A linker between *KLF4* and *OCT3/4* sequences. Although DNA band of expected size was amplified from the genomic DNA (gDNA) extracted from iPS cells transduced with lentiviruses containing the KOSM transgene, we could not detect the transcripts of the KOSM transgene in differentiated cells derived from iPS cells. This result indicates that the transduced *KOSM* fusion gene remained silenced in differentiated cells (Figure 5A), suggesting that our polycistronic vector not only improved reprogramming efficiency, but also alleviated spontaneous reactivation of the transduced transgene in differentiated cells derived from iPS cells,

Because the four reprogramming TFs were fused in-frame into a single ORF, one copy of infected KOSM transgene is theoretically enough to express the four reprogramming TFs. To investigate whether there are fewer proviral integration sites in the iPS cell colonies generated by our reprogramming system, we examined lentiviral integration sites in the 8 iPS cell colonies using inverse PCR [12] that was used in previous studies to examine proviral integration sites in iPS cells [7]. Our data revealed that a single integration site was found in 5 of 8 iPS cell colonies, and 2 sites were detected in the other 3 colonies. These results indicate that most of the iPS cell colonies contained a single copy of the lentiviral vector (Figure 5B), suggesting our approach is a relatively safer one for generation of iPS cells.

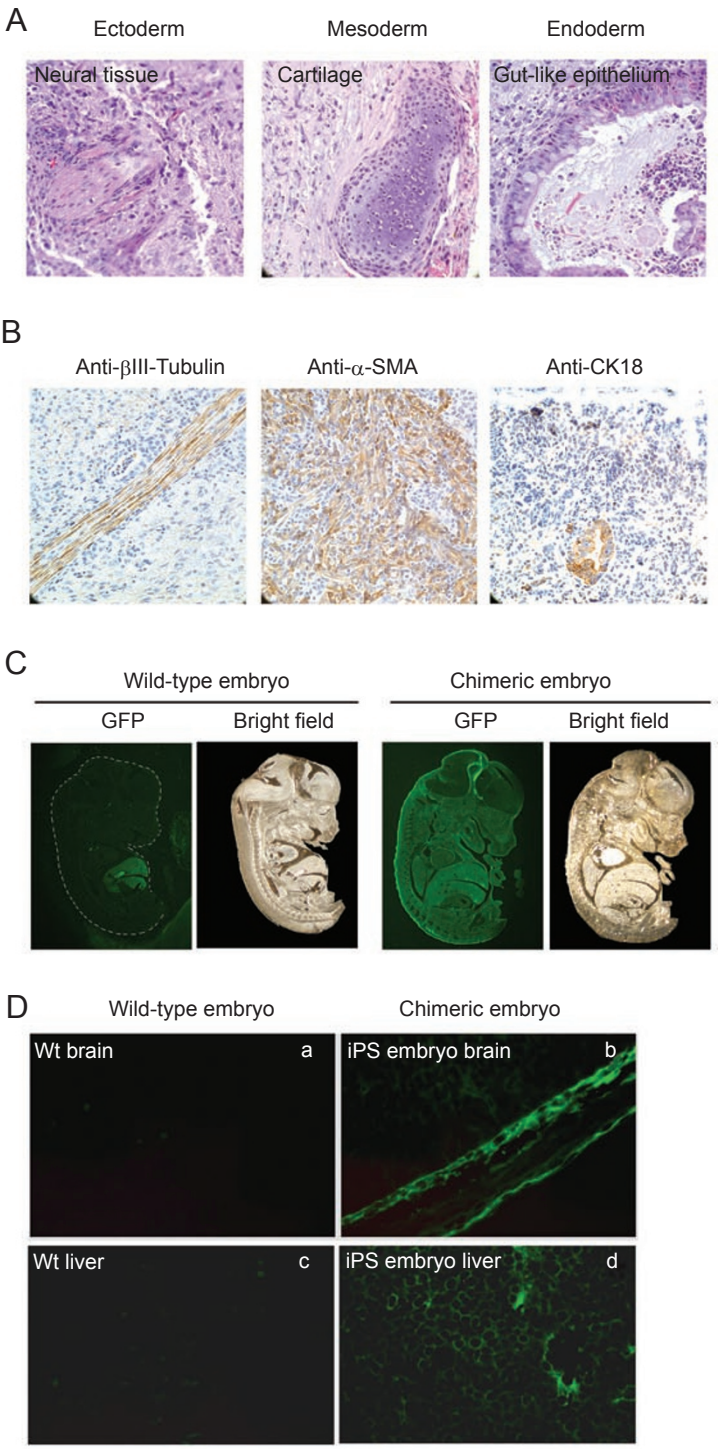

Figure 4 *In vivo* pluripotency of generated iPS cells. **(A)** Hematoxylin and eosin staining of teratoma derived from iPS cells (clone 1). Cells were injected subcutaneously into left and right flanks of syngeneic C57/BL6 mice. A teratoma formed 5 weeks after injection was dissected from the mouse and processed for staining (400× magnification). **(B)** Immunohistochemistry of teratoma. Sections of teratoma derived from iPS cells (clone 1) were stained with antibodies specific for βIII-Tubulin, α-SMA, and CK18. Positively stained cells (brown color) show typical structures for neurons, muscles, and epithelium, respectively (400× magnification). **(C)** High contribution of iPS cells in chimeric embryos. GFP expression of a chimeric embryo from donor iPS cells (from clone 1, labeled with GFP) was evident (right panels) compared to normal embryo (the noninjected control, in left panels). **(D)** Section of chimeric embryos showed contribution of iPS cells to brain (b) and liver (d). Sections from a wild-type embryo were included as negative control (a and c). Magnification is 400×.

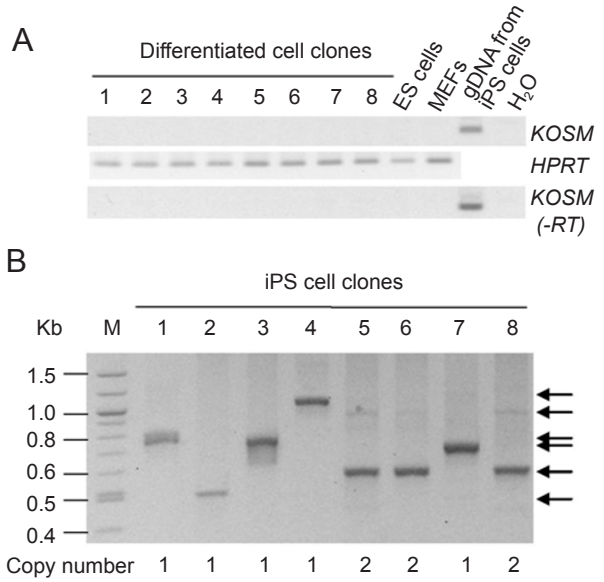

Figure 5 Analysis of transgene expression and copy number in differentiated cells derived from iPS cells. **(A)** RT-PCR analysis of *KOSM* transgene expression in differentiated cells. cDNA was synthesized from RNAs of iPS cells-derived differentiated cells with RT or without RT. Genomic DNA (gDNA) from iPS cells served as control for PCR amplification of *KOSM* gene. *HPRT* amplification served as positive controls for the RT-PCR experiment. *KOSM* fusion gene was amplified from gDNA extracted from iPS cells, but not from cDNAs synthesized from differentiated cells derived from iPS cells with or without RT. **(B)** Inverse PCR analysis of lentiviral integration sites in iPS cells. Genomic DNA was extracted from each of iPS cell clones and processed for inverse PCR analysis using 3′-LTR-specific primers. The number of lentiviral copy is shown at bottom. Arrowheads indicate amplified bands corresponding to 3′ LTR-flanking cellular DNA, whose size is dependent on the integration site. M: 100 bp DNA ladder.

Discussion

The current reprogramming approaches, which were accomplished by simultaneous lentiviral/retroviral infection of four defined TFs into somatic cells, encountered four major problems: i) low reprogramming efficiency; ii) multiple viral integrations in genomes of iPS cells, leading to an increased risk of genomic instability, gene mutagenesis, or both; iii) incomplete silencing of transduced exogenous genes in iPS cells; and iv) reactivation of transduced genes during iPS cell differentiation, thereby increasing cancer-causing potential. In addition, these approaches are inefficient and require very high virus titer. For instance, MEFs need at least 30% retrovirus transduction efficiency [7] and an average of 15 different viral copies [8] in order to be reprogrammed into iPS cells. In the present study, we were able to generate iPS cells from somatic cells by our new system when viral transduction efficiency was 1~5% (data not shown). Interestingly, we found that 63% (5/8) of the iPS cell colonies generated with our reprogramming system contained only one copy of viral vector, suggesting that delivery of four defined TFs by one vector (in a single ORF) is superior in the generation of iPS cells.

Notably, the reprogramming system that we devised here improved the reprogramming efficiency and the gene-silencing of exogenous TFs in iPS cells, and effectively diminished the reactivation of the transduced *KOSM* fusion gene in differentiated cells. We reasoned that the improved silencing of transduced exogenous gene in these iPS cells may be due to either less viral integration sites in genomes of cells, or the particular promoter used in our system, the CMV promoter. Previous studies have used retrovirus promoters and encountered problems with incomplete silencing of exogenous genes [2-6]. Therefore, it is possible that the CMV promoter is more susceptible to silencing when somatic cells are reprogrammed into pluripotent state [13]. Concurrent to our study, two other groups [14, 15] also reported single lentivirus vector systems for generation of iPS cells, presenting reprogramming approaches congruent to our approach in this study. Similar to our findings, the results from these studies indicate that a single copy of viral vector is sufficient to reprogram mouse and human somatic cells into iPS cells.

Recently, virus-free mouse iPS cells have been generated by adenovirus-mediated gene delivery and DNA transfection approaches [9, 10]. However, efficiency of iPS cell generation by these approaches has been significantly lower compared to the retroviral or lentiviral infection approaches. So far, ‘genetically clean’ human iPS cells have not yet been generated, possibly due to extremely low reprogramming efficiency of human somatic cells when defined reprogramming TFs are delivered by non-integrating vectors. Because a single copy of viral vector is present in most of the iPS cell clones generated by our system, it will be interest in future studies to test whether our new reprogramming system alone or perhaps together with small chemical molecules, which enhance reprogramming process [16, 17], could significantly improve generation of human iPS cells. Also, future studies should test whether delivering the *KOSM* fusion gene into somatic cells via adenoviral vectors could efficiently reprogram somatic cells into virus-free human iPS cells.

Materials and Methods

Plasmid construction

KLF4, *OCT3/4*, *SOX2*, and *c-MYC* were amplified by poly-

merase chain reaction (PCR) using Phusion™ High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA), and they were in-frame linked by E2A, T2A, and P2A sequence [11, 18, 19], respectively, as a single open reading frame (ORF) designated as *KOSM*. The *KOSM* fusion gene was then cloned into the lentivirus vector pLentG, and the resulting plasmid was designated as pLentG-KOSM. The complete sequence will be provided upon request.

Transfection and western blot analysis

One μg of pLentG-KOSM was transfected into 293T cells in a 6-well plate by Fugene HD (Roche). Forty-eight hours after transfection, the cells were washed with cold PBS buffer and lysed directly with RIPA lysis buffer, supplemented with protease inhibitor cocktail (Sigma). The cell lysates were separated by electrophoresis on 12% SDS polyacrylamide gel and transferred to a nitrocellulose membrane (Pierce). The blot was blocked with TBST (20 mM Tris-HCl, pH 7.6, 136 mM NaCl, and 0.05% Tween-20) containing 5% non-fat milk, and then incubated with primary antibody solution at 4 °C overnight. After washing with TBST, the membrane was incubated with horseradish peroxidase (HRP)-conjugated secondary antibody for one hour at room temperature. Signals were detected with the Immobilon Western Chemiluminescent HRP substrate (Pierce). Primary antibodies were: anti-OCT3/4 (1:2 000, #SC-5279, Santa Cruz, CA), anti-SOX2 (1:2 000, #2748s, Cell Signaling, Danvers, MA), anti-KLF4 (1:2 000, #SC-20691, Santa Cruz, CA), anti-c-MYC (1:2 000, #9402, Cell Signaling, Danvers, MA), anti-β-Actin (1:5 000, #A300-491A, Bethyl Laboratories, Montgomery, TX), anti-mouse IgG-HRP (1:5 000, #1858413, Pierce, Rockford, IL), anti-rabbit IgG-HRP (1:2 000, #1858415, Pierce, Rockford, IL).

Preparation of mouse embryo fibroblasts (MEFs)

The embryos were harvested from pregnant female (C57/BL6 background) mice on day 14~16 of gestation, and were removed from the extraembryonic membranes while the embryos were in a petri dish with PBS. The heads and all internal organs were removed and the embryo carcasses were washed with PBS buffer. The prepared individual embryo were transferred into a 6-well plate on ice with 0.25% trypsin, minced with scissors, and then left in cold room overnight. Twelve hours later, the minced embryoes were moved to a 37 °C bath for 20 min to activate trypsin, followed by addition of DMEM medium containing 10% FBS (heat-inactivated). The embryo pieces were vigorously disrupted until cloudy, and then spun down and aspirated off the medium. The MEFs were cultured in fresh DMEM-F12 cell culture medium at 37 °C.

Lentivirus production and generation of iPS cells from MEFs

The 293T cells were plated at 8×10^5 cells per 60 mm dish and incubated overnight. The next day, the cells were transfected with a mixture of DNA containing 2 μg of pLentG-KOSM, 1 μg of pCMV-VSVG, and 1.5 μg of psPAX2 (Addgene) by Fugene HD (Roche), according to the manufacturer's instruction. Twenty-four hours after transfection, the supernatant of transfected cells was collected and filtered through a 0.22 μm pore-size filter. For virus infection, MEFs (passage 3) were seeded in a 6-well plate at 2×10^4 cells per well at one day before transduction. The medium was replaced with virus-containing supernatant supplemented with 8 μg/ml polybrene (Sigma), and centrifuged at 900 g for 1 h. The cells were infected twice and incubated with fresh DMEM-F12 medium.

At day 3 post-infection, the medium was switched to the serum-free ESGRO medium (Millipore) and changed daily until induced colonies were picked up at day 15 post-infection. GFP ratio was determined by flow cytometry at day 3 after virus infection. Individual ES cell-like cell colony was monitored for GFP marker expression during reprogramming.

Immunofluorescence staining, Immunocytochemistry and alkaline phosphatase staining

For immunofluorescence staining of cells, cells were fixed with PBS containing 4% paraformaldehyde for 30 min at room temperature. After washing with PBS, the cells were treated with PBS containing 1% bovine serum albumin (BSA), and 0.1% Triton X-100 for one hour at room temperature, followed by incubation with primary and second antibodies.

For immunocytochemistry immunofluorescence staining of tissues, 5 μm paraffin-embedded sections were prepared from teratoma and mouse embryos, and incubated in citrate buffer (pH 6.0) at 92 °C for 20 min. The sections were then washed three times with phosphate-buffered saline (PBS). Staining of teratomas was performed with the avidin/biotin blocking kit and the M.O.M. Peroxidase Kit (Vector Laboratories) according to the manufacturer's guidelines. Sections of mouse embryos were stained with anti-GFP antibody (1:400, 3E6, A-11120, Invitrogen).

Following antibodies were used in this study: anti-OCT3/4 (1:50, #sc-5279, Santa Cruz), anti-SOX2 (1:100, #2748s, Cell Signaling), anti-NANOG (1:200, A300-398A-1, Bethyl Laboratories, Inc), anti-βIII-Tubulin (1:100, #CBL412, Millipore), anti-α-SMA (1:100, #CBL171, Upstate), anti-CK18 (1:500, #C-04, ab668, Abcam), anti-Albumin (1:50, #A90-134A , Bethyl Laboratories), or anti-GFP (1:400, 3E6, #A-11120, Invitrogen). Secondary antibodies: Texas Red-conjugated goat anti-mouse IgG (1: 100, #T-6390, Invitrogen) or FITC-conjugated goat anti-rabbit IgG (1:100, #81-6111, Invitrogen).

Alkaline phosphatase staining was performed by using the alkaline phosphatase detection kit (#SCR004, Millipore).

RNA extraction and RT-PCR

Total RNA was extracted by using Trizol reagent (Invitrogen) and treated with DNase I (Roche) to remove genomic DNA contamination. One μg of total RNA was used to synthesize cDNA by using qScrip cDNA supermix kit (Quanta Biosciences) and dT20 primer, according to the manufacturer's instructions. PCR was performed with gene-specific primers. The list of primers is included in the Supplementary information, Table S1.

Karyotyping analysis

The cells were cultured for 24 h in a CO_2 incubator, and treated with colcemid (0.07 μg/ml, final concentration) for four hours before harvesting. The cells were washed with PBS and then trypsin lysed and transferred into 15 ml tubes. The cells were centrifuged for 10 min at 500 g and the supernatant was removed and resuspended with 10 ml KCl solution (75 mM). The cell mixtures were incubated for 30 min in 37 °C water bath and then fixed by adding 2 ml fixative solution (methanol/acetic acid 3:1). The fixed cells

were washed at least two times with 10 ml of fixative solution before being applied onto chilled slides. The slides with chromosomes were dried and treated with 0.0025% trypsin for 5 min and stained with Giemsa (1:10) for 5 to 10 min.

Microarray and array analysis

For microarray experiment, RNA concentration and purity were determined by measurement of A 260, A260/A280 and A260/A230ratios using a NanoDrop-1000 Spectrophotometer (NanoDrop Technologies). RNA quality was checked by the Agilent 2100 Bioanalyzer (Agilent Technologies). All RNA samples had high integrity (RIN > 9.9) and showed no DNA contamination. For each sample, cDNA was generated from 600 ng of total RNA with Agilent Quick Amp kit (Agilent Technologies) according to manufacturer. cDNA was then amplified and labeled with cyanine 3-CTP using Agilent Quick Amp kit. Approximately 1.7 μg of cyanine 3-labeled cRNA was fragmented and hybridized to the Agilent Whole Mouse Genome Oligo Microarray 4x44k at 65 °C for 17 h in agilent hybridization oven, according to manufacturer's recommendation. Each sample was hybridized with 2 arrays in technical replicates. Hybridized arrays were washed and scanned on a GenePix 4000B scanner (Axon Instruments) using GenePix Pro 6.0 software (Axon Instruments) at 5 μm resolution.

The raw data were normalized (quantile normalization and $\log_2$ transformed) and filtered with Partek software (Partek) after extracting them with Feature Extraction software 9.5.1 (Agilent). 'Fold change' or '*P* value' were used for identification of differently expressed genes. Thresholds for selecting significant genes were set at $2 \leq$ relative fold difference ≤ -2, or *P* value of 0.001. Genes that met the criteria simultaneously were considered as significant changes. Pathway analysis was done with Ingenuity Pathway Analysis Software (Ingenuity Systems). The data include significant functions and significant pathways.

Calculation of reprogramming efficiency

We quantified the infected MEFs (GFP^+ cells) at day 3 post-infection. We also counted the total ES cell-like cell colonies, AP^+ colonies, $Oct4^+$ and $Sox2^+$ colonies. The total number of true pluripotent iPS cell colonies was determined based on the results of *in vitro* differentiation experiment.

In vitro differentiation of iPS cells

For embryoid bodies (EBs) formation, the induced colonies were expanded and were harvested by treating with Accutase (#SCR005, Millipore). The clumps of the cells were transferred to ultra-low 6-well plate in the ES cell medium (20% FBS, 2mM L-glutamine, 1×10^{-4} M nonessential amino acids, 1×10^{-4} M 2-mercaptoethanol, penicillin, and streptomycin, without LIF). The medium was changed every other day.

After 9 days of suspension in culture, the EBs were transferred to gelatin-coated plate and cultured in the same medium for another 7-9 days. For neural differentiation, all-trans retinoic acid (1 μM) was added into the medium at day 5 after suspension in culture and continued for 2 days. At day 7, the EBs were transferred to cell culture plate and cultured in the ES cell medium for another 6-7 days.

Teratoma formation

Five million iPS cells were collected from MEF feeder layers by collagenase IV and were resuspended in a mixture of DMEM culture medium and Matrigel (ratio of 1:1). The cell mixtures were injected subcutaneously into flanks of syngeneic mice. Tumors were dissected from mice 5-8 weeks after injection and paraffin sections were processed and stained with haematoxylin and eosin.

Generation and histological analysis of chimeras

Blastocysts were obtained through mating of female BDF1 (primed with hormone) and male BDF1 or C57BL/6J mice. Chimeras were produced by injecting iPS cells labeled with an EGFP expressing lentivirus into blastocysts, followed by implantation into pseudopregnant ICR mice. Chimeric embryos were dissected 10.5 days after injection and fixed. Paraffin-embedded sections (5 μm) were prepared from embryos and stained with anti-GFP antibody (1:400, 3E6, A-11120, Invitrogen).

Determination of lentiviral integration sites by inverse PCR

Lentiviral integration sites in iPS cell colonies were determined by inverse PCR according to the previous report [12]. Briefly, Genomic DNA (100 ng) was isolated from iPS cells using the Blood & Cell Culture Mini Kit (QIAGEN) and completely digested with *Taq* I restriction enzyme (clone #1, #3 − #8) or *Tsp509* I (clone #2) for 12 h at 65 °C. The fragments containing the 3′LTR with flanking DNA were purified using the QIAquick Nucleotide Removal Kit and self-ligated with T4 DNA ligase (10 unit) in 100 μl volume at 15 °C overnight. The self-ligated DNA was amplified by first PCR with the primers 3′LTR-1F (5′- TGGATGGTGCTACAAGCTAGTACCAGTTGAG) and 3′LTR-1R (5′-GGTCAGTGGATATCTGATCCCTG). Nested PCR was then carried out on 2 μl of the first PCR product with the primers 3′LTR-2F (5′-AGCCAATGAAGGAGAGAACACCCGCTTGTTACAC) and 3′LTR-2R (5′-GTGGTAGATCCACAGATCAAGGATATCTTG). The resulting PCR products were run on a 2% gel.

Acknowledgments

We thank Dr Liangyou Rui for providing the lentivirus expression vector pLentG and thank our colleagues at Maine Medical Center Research Institute for their critical review. We also thank the MMCRI Cell Separation & Analysis Core, the Histology Core, and the Mouse Transgenic Core for their services. This work was supported in part by an NIH grant from the National Center for Research Resources (P20 RR018789), USA. W.S.W was supported by a K01 award from the National Institute of Diabetes and Digestive and Kidney Disease (K01DK078180-01), USA.

References

1. Byrne JA. Generation of isogenic pluripotent stem cells. *Hum Mol Genet* 2008; **17**:R37- 41.
2. Okita K, Ichisaka T, Yamanaka S. Generation of germline-competent induced pluripotent stem cells. *Nature* 2007; **448**:313-317.
3. Wernig M, Meissner A, Foreman R, *et al*. *In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* 2007; **448**:318-324.
4. Takahashi K, Tanabe K, Ohnuki M, *et al*. Induction of pluripotent stem cells from adult human fibroblasts by defined factors.

Cell 2007; **131**:861-872.
5. Yu J, Vodyanik MA, Smuga-Otto K*, et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* 2007; **318**:1917-1920.
6. Park IH, Zhao R, West JA*, et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* 2008; **451**:141-146.
7. Aoi T, Yae K, Nakagawa M, Ichisaka T, *et al*. Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science* 2008; **321**:699-702
8. Wernig M, Lengner CJ, Hanna J*, et al.* A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nat Biotechnol* 2008; **26**:916-924.
9. Stadtfeld M, Nagaya M, Utikal J, Weir G, Hochedlinger K. Induced pluripotent stem cells generated without viral integration. *Science* 2008; **322**:945-949
10. Okita K, Nakagawa M, Hyenjong H, Ichisaka T, Yamanaka S. Generation of mouse induced pluripotent stem cells without viral vectors. *Science* 2008; **322**:949-953.
11. Szymczak AL, Workman CJ, Wang Y, *et al*. Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nat Biotechnol* 2004; **22**:589-594.
12. Nolta JA, Dao MA, Wells S, Smogorzewska EM, Kohn DB. Transduction of pluripotent human hematopoietic stem cells demonstrated by clonal analysis after engraftment in immune-deficient mice. *Proc Natl Acad Sci USA* 1996; **93**:2414-2419.
13. Chung S, Andersson T, Sonntag KC, Bjorklund L, Isacson O, Kim KS. Analysis of different promoter systems for efficient transgene expression in mouse embryonic stem cell lines. *Stem Cells* 2002; **20**:139-145.
14. Sommer CA, Stadtfeld M, Murphy GJ, Hochedlinger K, Kotton DN, Mostoslavsky G. iPS cell generation using a single lentiviral stem cell cassette. *Stem Cells* 2008 Dec 18. doi:10.1634/stemcells.2008-1075
15. Carey BW, Markoulaki S, Hanna J, *et al*. Reprogramming of murine and human somatic cells using a single polycistronic vector. *Proc Natl Acad Sci US*A 2009; **106**:157-162.
16. Huangfu D, Maehr R, Guo W, *et al*. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat Biotechnol* 2008; **26**:795-797.
17. Shi Y, Do JT, Desponts C, Hahm HS, Scholer HR, Ding S. A combined chemical and genetic approach for the generation of induced pluripotent stem cells. *Cell Stem Cell* 2008; **2**:525-528.
18. Holst J, Vignali KM, Burton AR, Vignali DA. Rapid analysis of T-cell selection *in vivo* using T cell-receptor retrogenic mice. *Nat Methods* 2006; **3**:191-197.
19. Fang J, Qian JJ, Yi S, *et al*. Stable antibody expression at therapeutic levels using the 2A peptide. *Nat Biotechnol* 2005; **23**:584-590.

(**Supplementary information** is linked to the online version of the paper on the Cell Research website.)