Two-factor reprogramming of somatic cells to pluripotent stem cells reveals partial functional redundancy of Sox2 and Klf4

A Nemajerova*,[1], SY Kim[2], O Petrenko[1] and UM Moll[1]

Ectopic expression of defined sets of transcription factors in somatic cells enables them to adopt the qualities of pluripotency. Mouse embryonic fibroblasts (MEFs) are the classic target cell used to elucidate the core principles of nuclear reprogramming. However, their phenotypic and functional heterogeneity represents a major hurdle for mechanistic studies aimed at defining the molecular nature of cellular plasticity. We show that reducing the complexity of MEFs by flow cytometry allows the isolation of discrete cell subpopulations that can be efficiently reprogrammed to pluripotency with fewer genes. Using these FACS-sorted cells, we performed a systematic side-by-side analysis of the reprogramming efficiency with different two- and three-factor combinations of Oct4, Sox2 and Klf4. We show that introduction of exogenous Oct4 with either Sox2 or Klf4 does not directly convert MEFs to a pluripotent state. Instead, each combination of factors disrupts the normal cellular homeostasis and establishes transient states characterized by the concurrent expression of mixed lineage markers. These cells convert into induced pluripotent stem cells in a stochastic fashion. Our data suggest that there is a partial functional redundancy between Sox2 and Klf4 in the disruption of cellular homeostasis and activation of regulatory networks that define pluripotency.
Cell Death and Differentiation (2012) **19,** 1268–1276; doi:10.1038/cdd.2012.45; published online 27 April 2012

Reprogramming somatic cells into induced pluripotent stem (iPS) cells represents a valuable resource for the development of general and patient-specific therapies. The crucial issues facing the clinical translation of iPS technology are (i) choosing the most appropriate cell type for induction of pluripotency; (ii) improving the reprogramming efficiency; and (iii) gaining a mechanistic understanding of the reprogramming process. To date, multiple cell types have been used for iPS production, including embryonic and adult fibroblasts, adipocytes, cardiomyocytes, epithelial, vascular, hematopoietic and neuronal progenitors.[1,2] Reprogramming has been achieved through the expression of various combinations of transcription factors, most often consisting of Oct4, Klf4, Sox2 and c-Myc (OKSM).[3–6] An important observation from these studies is that the type of somatic cells chosen has a significant effect on the efficiency of iPS generation and the extent of manipulation required. Thus, less-differentiated cells reprogram more efficiently than differentiated cells. Moreover, partially differentiated cells, for example, neural stem cells, can be reprogrammed with fewer factors.[7–10]

It is accepted that the ectopic expression of pluripotency factors establishes an embryonic-like transcriptome that is sustained by the subsequent reactivation of endogenous gene-regulatory networks.[1,2] However, determining the order of critical events that take place during reprogramming and their timing has proven difficult, in part because the process is neither efficient (less than 0.1%), nor rapid (about 2–3 weeks),

nor homogeneous in terms of transgene integration. The advent of a 'secondary' reprogrammable system based on doxycycline-inducible lentiviruses circumvented some of these problems.[11–13] The system was further improved by integration of a single inducible cassette with four reprogramming factors into mouse genome.[14,15] Although the reprogramming efficiency of cells from such secondary mice is higher (1–10%), several important questions regarding the mechanistic underpinning of the process remain unanswered: Why do only a small fraction of cells reprogram, while others don't? Which cell types can be effectively reprogrammed into a pluripotent state? Is reprogramming a fundamentally deterministic process that occurs in a well-defined temporal order, or is it driven by random chains of stochastic events triggered by the aberrant levels of transcription factors?[16–18]

To date, the majority of mechanistic studies of iPS cells have used mouse embryonic fibroblasts (MEFs) as the starting cell population, and it was assumed that MEFs are rather well-defined and fairly homogeneous. The advantages of fibroblasts are that they are relatively easy to derive and reprogram. However, it has been shown that fibroblasts represent a dynamic population of cells, exhibiting functional heterogeneity among and within tissues.[19,20] Moreover, embryo-derived fibroblasts are expected to contain a wide spectrum of progenitor cells with diverse properties that are able to expand prior to terminal differentiation. The transcriptional profiles of partially reprogrammed cells, which exhibit

similar features regardless of the starting cell type,[21–24] may support the notion of a stepwise process with well-defined intermediate stages. However, the phenotypic complexity of MEFs still has to be viewed as a confounding variable that needs to be taken into account when evaluating high-throughput expression data from initially heterogeneous populations.

In this study, we sought to address these questions and determine whether different two- and three-factor combinations of Oct4, Sox2 and Klf4 induce the dedifferentiation of MEFs into a unique progenitor-like state before they acquire pluripotency. We also sought to prospectively identify and isolate the most reprogrammable MEF populations. Our results indicate that forced expression of Oct4 with either Sox2 or Klf4 does not directly convert MEFs to a pluripotent stem cell state. Instead, each combination of factors disrupts the normal cellular homeostasis and establishes transient or metastable states characterized by the concurrent expression of mixed lineage markers. These cells convert into iPS cells in a stochastic fashion.

Results

Efficient induction of iPS cells with two factors using FACS (fluorescence-activated cell sorting)-sorted MEFs and an optimized protocol. One of the biggest hurdles in reprogramming somatic cells into pluripotent cells is the fact that the process is inefficient and often incomplete. To overcome this limitation, we generated REBNA episomal retroviral vectors[25] expressing human Myc, Klf4, Oct4 and Sox2. With this optimized system, we were able to reprogram into iPS cells up to 2% of freshly isolated MEFs using 4F (OSKM) or 3F (OSK) combinations. We next assessed whether the complexity of MEFs may influence the levels of reprogramming. To this end, E12.5–E13.5 MEFs were FACS-sorted into four phenotypically distinct subpopulations based on the surface expression of Thy1 (CD90) and Sca1 (stem cell antigen 1) markers. Previous studies demonstrated that Thy1 expression defines lipofibroblastic or myofibroblastic lineage,[26] while Sca1 is a marker specifying murine mesenchymal and epithelial progenitor cells.[27] Our microarray analyses confirmed that Sca1-single-positive cells (SP) express genes broadly involved in embryonic and tissue morphogenesis, thus indicating their immature origin (Supplemental Table 1), whereas Thy1/Sca1-double-negative (DN) cells selectively express markers of myofibroblast-committed progenitors (Supplemental Table 2). Notably, the Thy1-positive fractions (composed of Thy1-single positive (Thy1-SP) and Thy1/Sca1 double-positive (DP) cells) represent the majority of MEFs (>70%), while Thy1-negative (Sca1-SP and DN) cells are a minority (10–15% each, Figure 1a).

Freshly sorted MEFs were then seeded at equal densities and transduced with different combinations of reprogramming factors. The transduction efficiencies of the sorted MEF fractions were equal, based on the levels of green fluorescent protein (GFP) expression (data not shown) and ectopic factor expression (Figure 1b). However, the frequency of iPS colonies produced by each subpopulation differed

(Figure 1c). Thus, OSKM- or OSK-transduced DN and Sca1-SP cells yielded significantly more iPS colonies than the corresponding Thy1-SP or DP fractions (Figure 1c and Supplemental Table 3). We found that Myc was fully dispensable for the reprogramming of DN and Sca1-SP cells (Figure 1c). Moreover, DN and Sca1-SP cells transduced with 2F combinations (OK or OS) also yielded significantly higher numbers of iPS colonies compared to Thy1-SP and DP fractions (Figure 1c). We calculated a reprogramming efficiency of up to 2% for OSK-transduced, and ~0.2% for OK- and OS-transduced DN and Sca1-SP cells, respectively. Colonies of iPS cells first appeared in OK- and 3F-transduced MEFs (4 to 5 days post infection), followed by OS-transduced cells (9 to 10 days post infection, Figures 1d and e). We expanded over 100 independent OK and OS colonies for further analysis. Genotyping of these iPS clones with transgene-specific primers verified the integration of the Oct4 and either Klf4 or Sox2 transgenes in OK and OS infections, respectively (Supplemental Figure 1). None of the clones examined (n = 23) contained unwarranted integrated transgenes (e.g., Myc), thereby ruling out the possibility of cross-contamination (Supplemental Figure 1). Together, these data indicate that DN and Sca1-SP cells represent superior cell types for reprogramming and are accessible by fewer genetic manipulations.

Two-factor-derived iPS cells are pluripotent. Most of our 2F colonies showed well-defined ES cell-like morphology with sharp colony borders. The established iPS cells were alkaline phosphatase-positive (100%), predominantly negative for Thy1 (90–100%), Sca1 (80–90%), CD34 (100%) and CD133 (95–100%), and did express variable levels of ES cell markers CDH1 (100%), EpCAM (100%), Kit (60–90%) and SSEA1 (50–60%) (Figure 2a and data not shown). The observed fractions of KIT-, Sca1- and SSEA1-negative iPS cells likely reflect heterogeneous expression of these markers, as previously reported for undifferentiated wild-type ES cells.[28,29] In a separate set of experiments, we generated 2F iPS cells from FACS-sorted MEFs containing the Oct4-GFP reporter gene. These iPS cells homogeneously expressed the GFP, confirming their undifferentiated status (Figure 2b). Following subcutaneous injection into nude mice, teratomas were formed by all clonal iPS cell lines tested (n = 20). Histological analysis revealed tissue structures indicative of all three germ layers (Figure 2c). We successfully obtained adult chimeric mice with independently produced OK (1 out of 1) and OS (1 out of 2) iPS cell lines (Figures 2d and e). The degree of chimerism, as assessed by coat color contribution, was wide-spread, ranging from 50 to 90% in individual animals. Taken together, these *in vivo* results indicate that 2F-derived iPS cells are pluripotent.

Gene expression profiles of two-factor-transduced MEFs. We next examined the expression profiles of each Thy1/Sca1 subpopulation transduced with the reprogramming factors. Analysis of global gene expression patterns demonstrated a strong correlation between transcriptomes of OS-infected cells and those of OK-infected cells from the Thy1-negative group (Figure 3a, top panels). In contrast, the
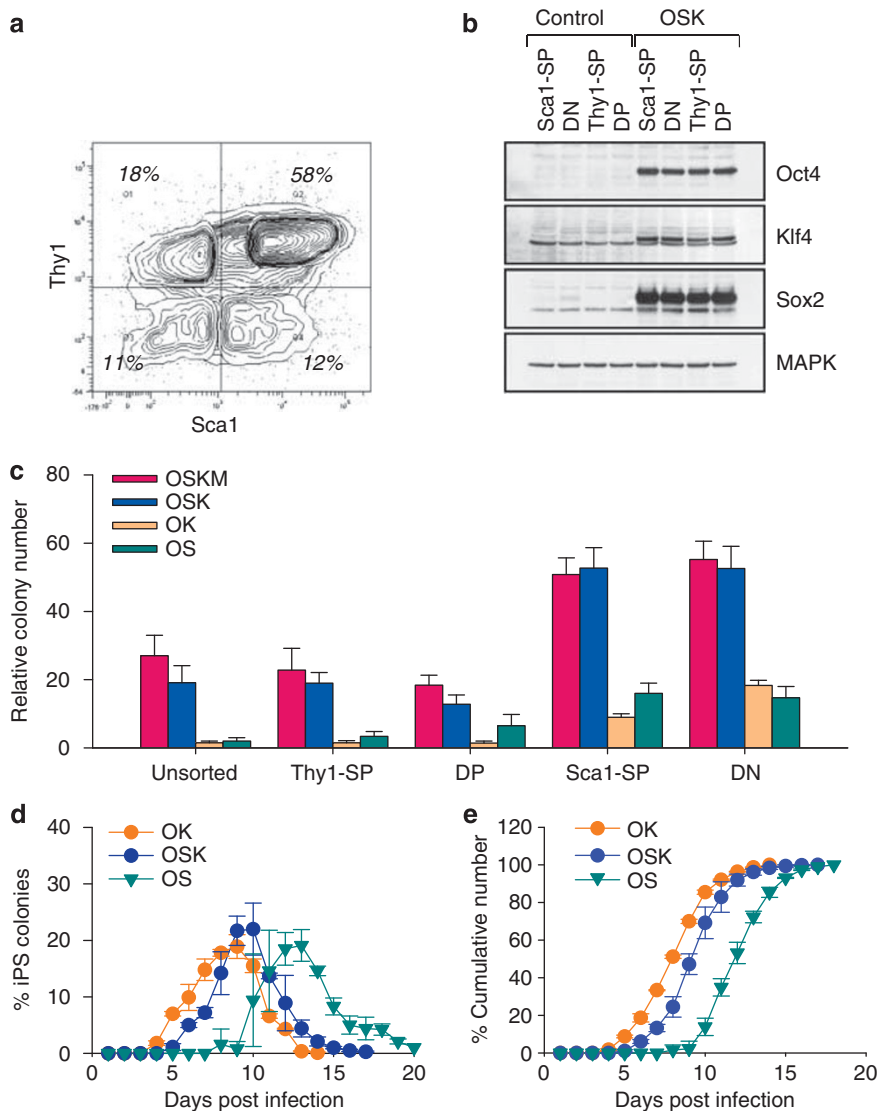
Figure 1 Efficient induction of iPS cells with two factors. (**a**) Flow-cytomeric analysis of Thy1 and Sca1 expression by wild-type MEFs at passage 2. (**b**) Immunoblot analysis of Oct4, Klf4 and Sox2 expression by sorted MEFs transduced with control or OSK retroviruses. MAPK is a loading control. (**c**) Relative efficiency of iPS cell generation from total unsorted MEFs or Thy1-SP, DP, Sca1-SP and DN fractions transduced with the indicated retroviruses. The results represent the average of 10 experiments. Values are normalized relative to unsorted MEFs transduced with OK viruses. The error bars correspond to standard error. (**d** and **e**) Reprogramming kinetics of DN cells measured as the percentage of newly produced iPS colonies per day (**d**) or the percentage of cumulative iPS colonies per day (**e**). Cells were transduced with the indicated retroviruses. Colonies consisting of ≥20 cells were counted daily. The results represent an average of two experiments performed in duplicate. The error bars correspond to standard deviation

correlation was lower when the respective OS- *versus* OK-infections were compared (Figure 3a, bottom panels). As expected, microarray analysis revealed that OK- *versus* OS-mediated reprogramming elicited dissimilar effects on the expression of many endogenous transcription factors. Thus, combined expression of Oct4 and Klf4 altered the levels of ~120 endogenous transcription factors, while the combined expression of Oct4 and Sox2 caused changes in the levels of ~90 endogenous transcription factors compared with the respective uninfected controls (Figures 3b and c and Supplemental Figures 2a and b). Only a small number of endogenous transcription factors were similarly induced or repressed in OK- and OS-transduced cells (Supplemental Figure 2c). Notably, gene set enrichment analysis[30] showed

that the OS combination of factors was more effective in inducing genes associated with pluripotency (i.e., the previously identified Core, Myc and ESC-like gene modules[31,32]), consistent with the notion that these two proteins bind to DNA together (Supplemental Figure 3). However, expression of either OS or OK also resulted in the concurrent induction of different or even conflicting lineage-specific markers (Figure 3d). Thus, infection with OS viruses led to upregulation of mesenchymal markers *ESM1* (endothelial cell-specific molecule 1), *MSC* (musculin) and *SOSTDC1* (sclerostin domain containing 1, inhibitor of Wnt-mediated signaling), but also neuronal genes, such as *NES* (Nestin), *RELN* (reelin), *SOX21* and *TNN* (tenascin N). Likewise, cells infected with OK viruses upregulated mesenchyme-specific
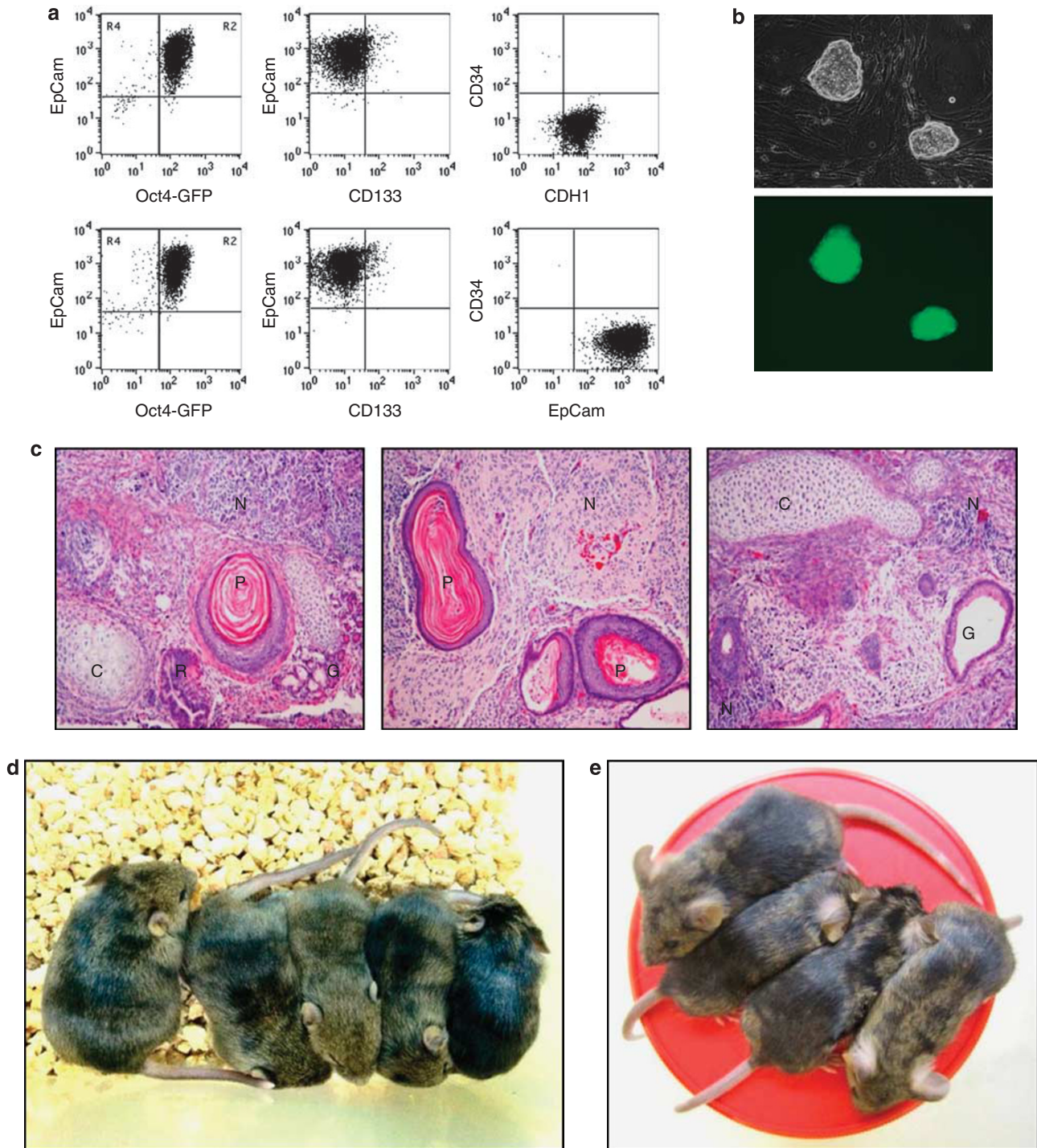
Figure 2 Assessment of the quality of two-factor-produced iPS cells. (**a**) FACS analysis of two independent iPS clones derived from OS-transduced DN Oct4-GFP MEFs. For the two panels on the right, each clone was gated on GFP-positive cells (R2). (**b**) GFP-positive ES-like colonies derived from OS-transduced Oct4-GFP-expressing MEFs. (**c**) Teratoma formation by 2F iPS cell lines. Hematoxylin and eosin staining identified mature tissue derivatives of all three germ layers: squamous epithelium with keratotic pearl (P, ectoderm); neuroepithelial cells (N, ectoderm); secretory epithelium (R, ciliated respiratory or G, glandular structures) (endoderm); and mature cartilage embedded in fibrovascular stroma (C, mesoderm). (**d** and **e**) Chimeric mice with agouti coat color originating from an OS iPS clone (**d**) or OK iPS clone (**e**)

genes *CD34*, *LY6A* (Sca1), *MSC* and *TEK* (TEK tyrosine kinase), the neuronal genes *RELN* and *TNN*, but also the epithelial genes *KERA, KRT14, KRT16* and *KRTDAP* (keratocan, keratins 14, 16 and keratinocyte differentiation-associated protein, respectively) (Figure 3d).

In sum, these data indicate that cells transduced with OS-*versus* OK-expressing viruses differ in the molecular mechanisms underlying the induction of pluripotency. The observed alterations likely reflect the fact that, in addition to their roles in stem cell maintenance, Oct4, Sox2 and Klf4 are
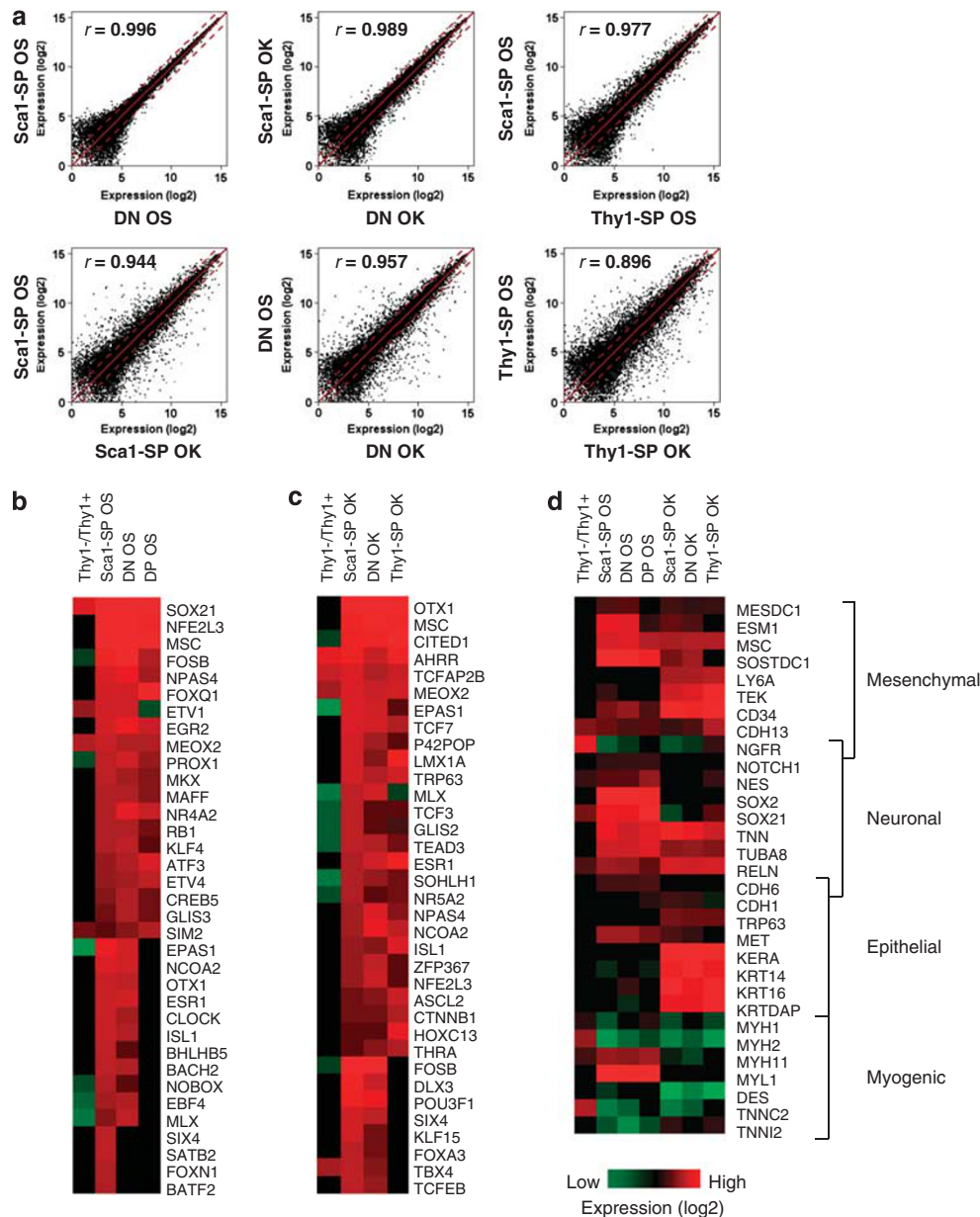
Figure 3 Gene expression profiles of two-factor-transduced MEFs. (**a**) Scatter plots of global gene expression patterns comparing Sca1-SP, DN and Thy1-SP cells 4 days post infection with the indicated retroviruses. The degree of correlation was assessed using Pearson's correlation coefficient (*r*). The central red line represents equal gene expression. The outer red lines indicate twofold different expression. (**b** and **c**) Heat maps from microarray analyses showing induction of endogenous transcription factors in Sca1-SP, DN, Thy1-SP and DP MEFs transduced with OS- (**b**) or OK-expressing retroviruses (**c**). Values are given relative to their own uninfected controls. Expression profiles of the corresponding uninfected Sca1-SP and DN (summarily called 'Thy1 − ') *versus* Thy1-SP and DP (summarily called 'Thy1 + ') cells are shown as controls. (**d**) Simultaneous induction of lineage-specific genes in OS- and OK-transduced Sca1-SP, DN and Thy1-SP MEFs. Values are given relative to their own uninfected controls

lineage specifiers and promote mesodermal, neural and epithelial differentiation, respectively.

Generation of iPS cells from Nestin-GFP MEFs. The experiments described in Figures 1 and 2 show that DN and Sca1-SP cells yield higher numbers of IPS colonies, suggesting that Thy1-SP and DP cells represent inferior cell types for nuclear reprogramming. In support of this, time-course analysis of sorted MEFs showed that Thy1-SP and DP fractions represent progressive differentiation stages

within MEF populations (Supplemental Figure 4). These results suggest that in the course of reprogramming of total unsorted MEFs, a significant proportion of 2F and 3F iPS clones might be generated from immature DN and Sca1-SP cells or their subsets thereof, whose reprogramming is more efficient than that of Thy1-SP and DP cells. To test this idea, we followed the reprogramming of embryonic fibroblasts derived from mice expressing GFP under the control of the endogenous Nestin promoter.[33] Nestin has been extensively studied as a marker for neural and mesenchymal stem/

progenitor cells.[33,34] Our microarray analyses revealed that both DN and Sca1-SP populations are enriched for Nestin-positive cells (Figure 3d and Supplemental Figure 5a). We surmised that these Nestin-positive cells might represent a virtually homogeneous subset of immature cells and thus be tangible for mechanistic studies of reprogramming.

As reported,[33] MEFs from Nestin-GFP embryos could be sorted into three populations based on the levels of GFP expression (GFP-bright, -dim and -negative, Figures 4a and b). In several independent E12.5 MEF cultures, the GFP-bright cells accounted for ∼2% of the total number of cells (Figure 4a). A fraction of these cells expressed NGFR, another surface marker that distinguishes Thy1-negative from Thy1-positive cells (Figure 4a and Supplemental Figure 5b). To test their respective reprogramming competence, each of the Nestin-GFP fractions was infected with 2F or 3F

Figure 4 Generation of iPS cells from Nestin-GFP MEFs. (**a**) FACS analysis of GFP and NGFR expression by E12.5 Nestin-GFP MEFs. GFP-negative (R1), -dim (R2) and -bright (R3 and R4) fractions are indicated. (**b**) Experimental design and analysis of iPS clones derived from Nestin-GFP MEFs. (**c**) Reprogramming efficiency of 3F- and 2F-transduced GFP-negative, -dim and -bright fractions. The error bars correspond to standard error. (**d**) Reprogramming efficiency of 3F- and 2F-transduced GFP-negative, GFP-bright NGFR-positive and GFP-bright NGFR-negative MEFs. (**e**) Induction of GFP expression 7 days post transduction with OSK, OK or OS viruses. Transduction with OSK induces strong expression of GFP (i.e., Nestin) in both GFP-dim and GFP-negative cells. A representative of two experiments is shown

retroviruses as a reference (Figure 4b). The iPS colonies were detected based on their ES cell-like morphology and alkaline phosphatase staining. Notably, the efficiency of 2F reprogramming of GFP-bright cells was at least 5–10 times higher than that of GFP-negative cells (Figure 4c). Furthermore, Nestin-SP cells were more reprogrammable compared to their NGFR-positive counterparts (Figure 4d). In addition, NGFR expression was virtually lost after transduction of GFP-bright cells with OK or OSK retroviruses (see below), whereas the Nestin-GFP transgene remained transcriptionally active regardless of which reprogramming factors were used (Figure 4e). Transduction with OSK retroviruses and, to a lesser extent, with OS or OK viruses also induced strong expression of GFP (i.e., Nestin) in both GFP-dim and GFP-negative cells (Figure 4e and Supplemental Figure 6). Thus, the acquired expression of Nestin by cells mirrors their ability to convert into iPS cells (Figure 4c).

We next assessed the temporal induction patterns of lineage-specific markers during the reprogramming process. To this end, GFP-bright, -dim and -negative cells were infected with 2F or 3F retroviruses. On days 1, 4, 7 and 11 post infection and after becoming stable iPS cell lines, they were stained with antibodies for the tissue- and lineage-specific surface proteins and scored by FACS on a single-cell basis (Figure 5a and data not shown). Consistent with our earlier results, this analysis again revealed that induction of reprogramming by OK versus OS viruses generates phenotypically dissimilar cells. Thus, uninfected GFP-bright cells were predominantly negative for Thy1 (80%) and Sca1 (70%), and showed no significant expression of CD34, CD133, KIT, SSEA1, CDH1 (E-cadherin) or EpCAM (Figure 5a). Transduction with OK or OSK viruses caused transient expression of Thy1, Sca1, CD34 and CD133, but early loss of NGFR (Figure 5a). Subsets of these cells concurrently expressed CD34 and CD133 (Figure 5a). In contrast, cells transduced with OS viruses retained significant NGFR expression but showed no induction of CD34 or CD133 even after 11 days in culture (Figure 5a). Only small fractions of these cells (1–5%) showed expression of KIT, SSEA1, CDH1 and EpCAM prior to their conversion into iPS cells (Figure 5a).

Western blot analysis confirmed these results. Thus, transduction of sorted MEFs with OS, OK or OSK viruses caused little or no changes in the expression levels of CDH1, EpCAM or KIT genes during initiation of the reprogramming event (Figure 5b). Whether the small amount of CDH1 protein induced in OK and OSK cells (but not in OS cells) is present at the cell surface rather than in the cytosol cannot be determined with certainty, given the low score by FACS (Figure 5a). Furthermore, quantitative reverse transcription-polymerase chain reaction (qRT-PCR) revealed that mRNAs for E cadherin, EpCAM and KIT were only partially induced compared to iPS cells (Figure 5c and Supplemental Figure 7). On the other hand, the induced expression of pluripotency transcription factors Sox2 and, to a lesser extent, Sall4 and Zic3 either preceded or closely paralleled the later induction of epithelial and ES cell markers (Figure 5c). Accordingly, OS- and OSK-transduced cells expressed total (endogenous and exogenous) Sox2 protein at levels similar to those present in iPS and wild-type ES cells (Figure 5b). In sum, a side-by-side comparison of 2F and 3F transductions demonstrates that
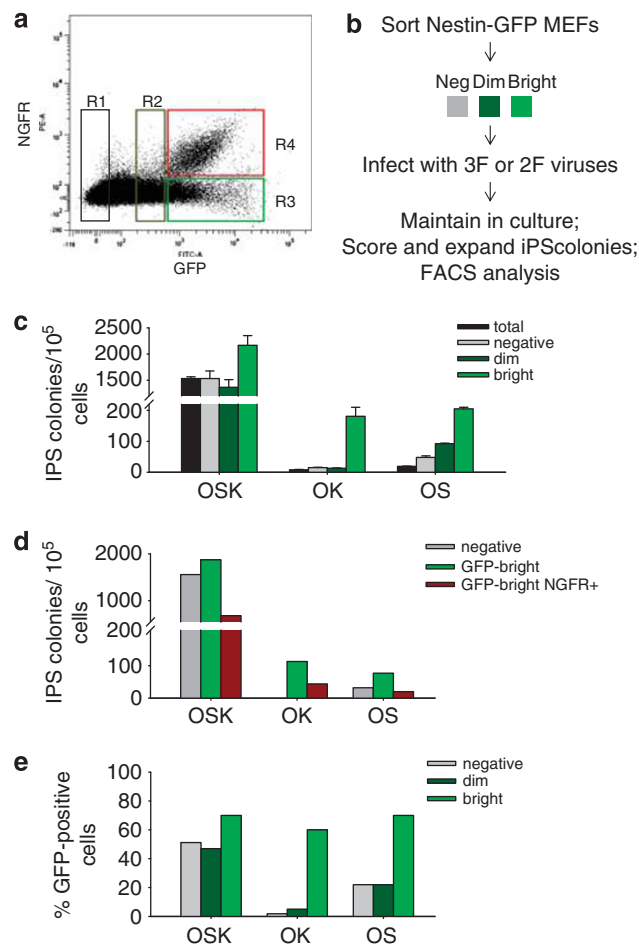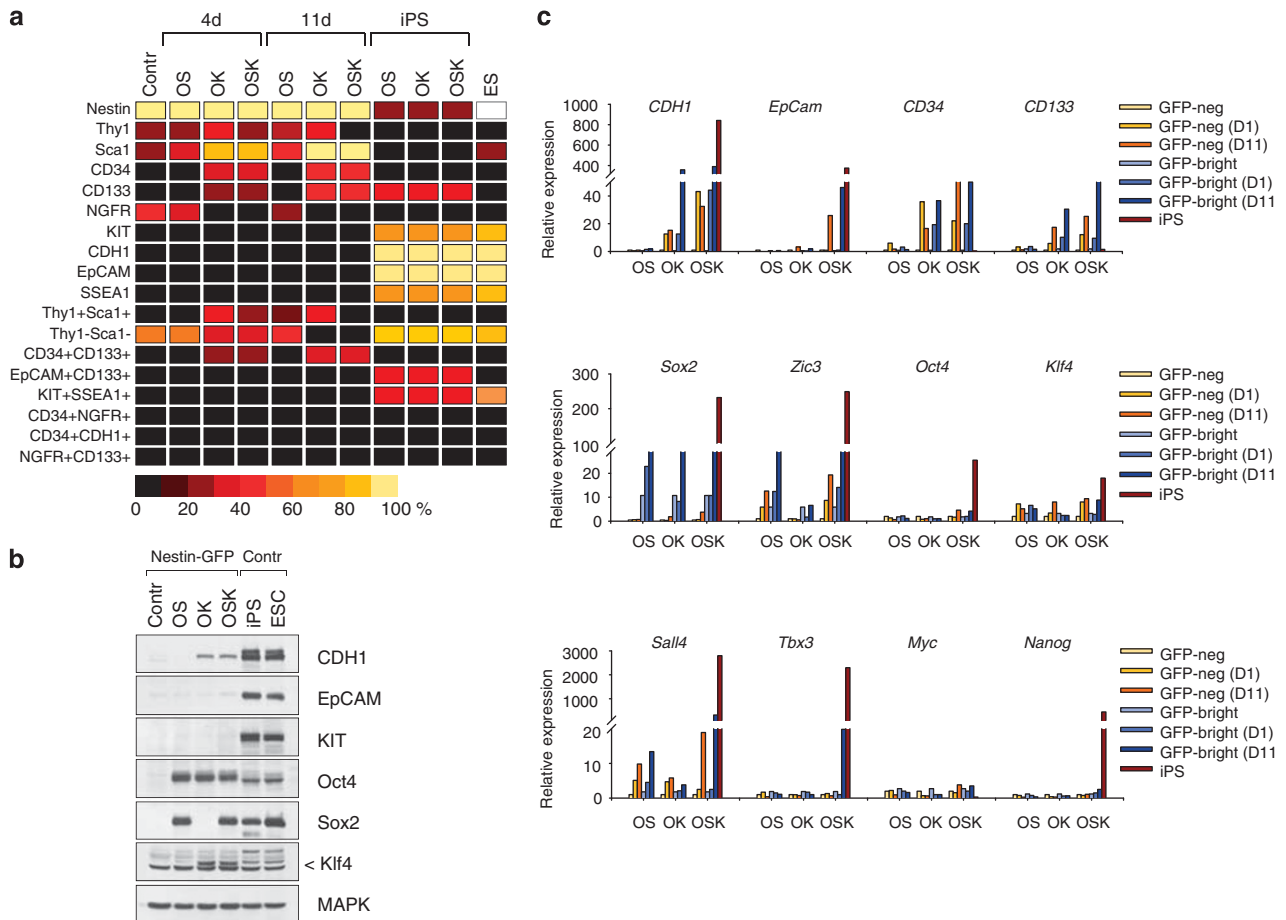
Figure 5 Multilineage pattern of gene expression parallels induction of pluripotency. (**a**) Schematic heat map showing the percentage of Nestin-GFP-bright cells with co-expression of lineage-specific genes as determined by FACS analysis. The analysis before (Control) and after transduction with OS, OK or OSK viruses is shown at the indicated time points. Expression profiles of iPS and WT ES cells are shown for comparison. Color scale indicates the percentage of cells that express the indicated markers. (**b**) Immunoblot analysis of Nestin-GFP-bright MEFs before (Control) and 7 days after transduction with OS, OK or OSK viruses. Expression profiles of iPS and WT ES cells are shown for comparison. MAPK is a control for equal loading. (**c**) Quantitative real-time reverse transcription–polymerase chain reaction (qRT-PCR) analysis of the indicated endogenous genes in Nestin-GFP-negative (yellow group) and Nestin-GFP-bright MEFs (blue group) before and 1 or 11 days after transduction with OS, OK or OSK viruses. Data were normalized to the corresponding HPRT value to obtain relative changes in the indicated mRNAs. Mouse-specific primers that did not cross-react with the exogenous human counterparts were used for Oct4, Sox2 and Klf4

forced expression of Oct4 with either Sox2 or Klf4 does not directly convert MEFs to a pluripotent state. Instead, each combination of factors disrupts the normal cellular homeostasis and establishes transient states characterized by the concurrent expression of mixed lineage markers. These cells convert into iPS cells in a stochastic fashion.

Discussion

Recent studies in fibroblasts suggest that reprogramming to pluripotency follows a well-organized temporal sequence of molecular and cellular events termed initiation, maturation and stabilization.[29,35,36] This sequence was proposed to entail a rate-limiting step of mesenchymal-to-epithelial transition (MET) that marks initiation.[23,24] According to this model, the acquisition of epithelial features is accompanied by the loss of differentiated cell characteristics and massive downregulation of somatic genes.[23,24,29,35] A subset of cells will then initiate the embryonic transcriptome, and, finally, upregulated pluripotency genes establish an ES cell-like expression

program.[23,24,29,35] Although this concept is appealing, the question of why the majority of cells fail to reprogram persists.

Several insights emerge from our study. We performed a systematic molecular and functional dissection of primary MEFs. First, we show that MEFs, independent of strain background, are heterogeneous with respect to reprogramming competency and can be FACS-sorted into distinct subpopulations based on differential expression of cell surface markers. Importantly, these subpopulations yield significantly different efficiencies in iPS generation with different combinations of reprogramming factors. Specifically, Thy1-SP and DP cells, which constitute the vast majority of the unsorted MEF population, represent a poor target for iPS production, whereas the preexisting DN and Sca1-SP subpopulations represent the more reprogrammable cell types. Thus, Nestin-GFP-positive cells, which represent a distinct subpopulation of DN and Sca1-SP cells, can be effectively reprogrammed with only two factors, Oct4/Klf4 and Oct4/Sox2.

Second, we find that, compared to OKSM fibroblasts, our three-partite system (OS, OK and OSK) enables a more comprehensive and unbiased view of the reprogramming process. We show that Sox2 stands out as the most responsive endogenous pluripotency factor during early reprogramming stages. Moreover, the sustained expression of Sox2 precedes the acquisition of epithelial and ES cell-like phenotypes in cells that undergo reprogramming. We find it plausible that a subset of Nestin-GFP-positive cells expresses sufficiently high levels of endogenous Sox2, and thus requires fewer exogenous factors. However, it is unlikely that these cells may still be heterogeneous and contain multiple subpopulations, one responding to overexpression of Sox2 and the other responding to Klf4. Given that Klf4 is required for the activation of epithelial gene expression during somatic cell reprogramming,[23,24] we speculate that a proportion of Sox2-overexpressing cells are able to bypass an early Klf4-dependent phase. Thus, notwithstanding the differences in microarray profiling, our data suggest that there is a partial functional redundancy between Sox2 and Klf4 in the disruption of cellular homeostasis and activation of pluripotency-related gene networks. The partial overlap between these factors could explain why Oct4/Sox2 and Oct4/Klf4-mediated reprogramming had the same degree of efficiency ($\sim 0.2\%$), but was an order of magnitude higher with three factors combined.

One important question is whether in our system all transduced genes were expressed at levels sufficient to achieve the optimal reprogramming efficiency. We addressed this in the following way. The distribution of retroviral-mediated gene expression is essentially stochastic.[37] Simple calculations show that for 2% of cells to be effectively reprogrammed with three factors, at least 27% of them would have contained an optimal amount of each of these factors (i.e., $0.27^3 = 0.02$). Hence, $\sim 7\%$ of cells (i.e., 0.27^2) would have expressed optimal levels of each of the two factors, Oct4/Klf4 or Oct4/Sox2. Given that 0.2% of cells became reprogrammed, this corresponds to a reprogramming efficiency of $\sim 3\%$ of cells with optimal expression of both factors (i.e., $0.2 \times 100/7$). Thus, we conclude that the expression levels of exogenous factors did not have a rate-limiting role during iPS induction in this system.

In sum, our results support a version of the stochastic model in which cells of different degrees of differentiation have the potential to generate iPS cells, albeit with very different kinetics and the requirement of three rather than two factors for reprogramming of more differentiated cells. The main difference between undifferentiated and differentiated cells, in this respect, is that the former are less stable at high levels of internal or external stimuli.[38,39] Our results infer that regardless of the cell type the net result of stochastic transitions is the establishment of a precarious balance between 'supply' and 'demand' for transcription factors that have explicit lineage-specifying activities.[39] Based on this, we conclude that forced expression of Oct4 with either Sox2 or Klf4 does not directly convert MEFs to a pluripotent stem cell state. Instead, each combination of factors disrupts the normal cellular homeostasis and establishes transient states characterized by the concurrent expression of mixed lineage markers (Supplemental Figure 8). Viewed in this light, the MET observed in four-factor reprogramming might be just one of many possible transient stages.

Materials and Methods

All cells were cultured on 0.1% gelatinized tissue culture plates. Wild-type mice (Jackson Laboratory, Bar Harbor, ME, USA) were maintained on the 129/Sv background. Oct4-GFP mice (Jackson Laboratory) and Nestin-GFP transgenic mice[33] were maintained on a mixed 129/B6 genetic background. MEFs were derived from E12.5–E13.5 embryos using standard procedures. iPS induction was performed according to Li *et al.*[40] Retroviral vectors encoding human Oct4, Sox2, Klf4 and c-Myc were transfected into packaging Phoenix E cells using Lipofectamine reagent (Invitrogen, Grand Island, NY, USA). Viral supernatants were collected and filtered through a 0.45-μm cellulose acetate filter. MEFs were plated at a density of 2×10^5 cells per 6-cm plate and incubated with the viral supernatant overnight. After four successive infections, cells were switched to knockout serum replacement (KSR) medium consisting of DMEM/F12 supplemented with 10% KSR (Invitrogen), $1 \times$ nonessential amino acids, $1 \times$ Glutamine, $1 \times$ Pen/Strep, 0.1 μM β-mercaptoethanol and 1000 U/ml leukemia inhibitory factor (Millipore, Billerica, MA, USA). The transduction efficiency of MEFs was determined by extrapolation from infection with a GFP-expressing control virus, which was conducted in parallel. In order to improve transduction efficiencies, the proviral parts of pMX vectors contained within the corresponding BspH1 sites were subcloned into episomal vector REBNA.[25] Molecular, histological, microarray and other analyses were performed as described in Supplemental Experimental Procedures, Figures and Tables.

Conflict of Interest
The authors declare no conflict of interest.

1. Hanna JH, Saha K, Jaenisch R. Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell* 2010; **143**: 508–525.
2. Stadtfeld M, Hochedlinger K. Induced pluripotency: history, mechanisms, and applications. *Genes Dev* 2010; **24**: 2239–2263.
3. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006; **126**: 663–676.
4. Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* 2007; **1**: 55–70.
5. Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, Hochedlinger K *et al.* In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* 2007; **448**: 318–324.
6. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* 2007; **318**: 1917–1920.
7. Kim JB, Zaehres H, Wu G, Gentile L, Ko K, Sebastiano V *et al.* Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. *Nature* 2008; **454**: 646–650.
8. Tsai SY, Clavel C, Kim S, Ang YS, Grisanti L, Lee DF *et al.* Oct4 and klf4 reprogram dermal papilla cells into induced pluripotent stem cells. *Stem Cells* 2010; **28**: 221–228.
9. Kim JB, Greber B, Arauzo-Bravo MJ, Meyer J, Park KI, Zaehres H *et al.* Direct reprogramming of human neural stem cells by OCT4. *Nature* 2009; **461**: 649–3.
10. Tsai SY, Bouwman BA, Ang YS, Kim SJ, Lee DF, Lemischka IR *et al.* Single transcription factor reprogramming of hair follicle dermal papilla cells to induced pluripotent stem cells. *Stem Cells* 2011; **29**: 964–971.
11. Wernig M, Lengner CJ, Hanna J, Lodato MA, Steine E, Foreman R *et al.* A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nat Biotechnol* 2008; **26**: 916–924.
12. Markoulaki S, Hanna J, Beard C, Carey BW, Cheng AW, Lengner CJ *et al.* Transgenic mice with defined combinations of drug-inducible reprogramming factors. *Nat Biotechnol* 2009; **27**: 169–171.
13. Eminli S, Utikal J, Arnold K, Jaenisch R, Hochedlinger K. Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression. *Stem Cells* 2008; **26**: 2467–2474.
14. Stadtfeld M, Maherali N, Borkent M, Hochedlinger K. A reprogrammable mouse strain from gene-targeted embryonic stem cells. *Nat Methods* 2010; **7**: 53–55.

15. Carey BW, Markoulaki S, Beard C, Hanna J, Jaenisch R. Single-gene transgenic mouse strains for reprogramming adult somatic cells. *Nat Methods* 2010; **7**: 56–59.

16. Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creyghton MP *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* 2009; **462**: 595–601.

17. Hanna J, Carey BW, Jaenisch R. Reprogramming of somatic cell identity. *Cold Spring Harb Symp Quant Biol* 2008; **73**: 147–155.

18. Nagy A, Nagy K. The mysteries of induced pluripotency: where will they lead? *Nat Methods* 2010; **7**: 22–24.

19. Koumas L, Smith TJ, Feldon S, Blumberg N, Phipps RP. Thy-1 expression in human fibroblast subsets defines myofibroblastic or lipofibroblastic phenotypes. *Am J Pathol* 2003; **163**: 1291–1300.

20. Phan SH. Biology of fibroblasts and myofibroblasts. *Proc Am Thorac Soc* 2008; **5**: 334–337.

21. Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* 2008; **454**: 49–55.

22. Sridharan R, Tchieu J, Mason MJ, Yachechko R, Kuoy E, Horvath S *et al.* Role of the murine reprogramming factors in the induction of pluripotency. *Cell* 2009; **136**: 364–377.

23. Li R, Liang J, Ni S, Zhou T, Qing X, Li H *et al.* A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* 2010; **7**: 51–63.

24. Samavarchi-Tehrani P, Golipour A, David L, Sung HK, Beyer TA, Datti A *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* 2010; **7**: 64–77.

25. Petrenko O, Beavis A, Klaine M, Kittappa R, Godin I, Lemischka IR. The molecular characterization of the fetal stem cell marker AA4. *Immunity* 1999; **10**: 691–700.

26. Sanders YY, Kumbla P, Hagood JS. Enhanced myofibroblastic differentiation and survival in Thy-1(-) lung fibroblasts. *Am J Respir Cell Mol Biol* 2007; **36**: 226–235.

27. Holmes C, Stanford WL. Concise review: stem cell antigen-1: expression, function, and enigma. *Stem Cells* 2007; **25**: 1339–1347.

28. Gu B, Zhang J, Wang W, Mo L, Zhou Y, Chen L *et al.* Global expression of cell surface proteins in embryonic stem cells. *PLoS One* 2010; **5**: e15795.

29. Stadtfeld M, Maherali N, Breault DT, Hochedlinger K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* 2008; **2**: 230–240.

30. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005; **102**: 15545–15550.

31. Wong DJ, Liu H, Ridky TW, Cassarino D, Segal E, Chang HY. Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell* 2008; **2**: 333–344.

32. Kim J, Woo AJ, Chu J, Snow JW, Fujiwara Y, Kim CG *et al.* A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell* 2010; **143**: 313–324.

33. Mignone JL, Kukekov V, Chiang AS, Steindler D, Enikolopov G. Neural stem and progenitor cells in nestin-GFP transgenic mice. *J Comp Neurol* 2004; **469**: 311–324.

34. Mendez-Ferrer S, Michurina TV, Ferraro F, Mazloom AR, Macarthur BD, Lira SA *et al.* Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature* 2010; **466**: 829–834.

35. Brambrink T, Foreman R, Welstead GG, Lengner CJ, Wernig M, Suh H *et al.* Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* 2008; **2**: 151–159.

36. Smith ZD, Nachman I, Regev A, Meissner A. Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. *Nat Biotechnol* 2010; **28**: 521–526.

37. Nash KL, Lever AM. Green fluorescent protein: green cells do not always indicate gene expression. *Gene Ther* 2004; **11**: 882–883.

38. MacArthur BD, Please CP, Oreffo RO. Stochasticity and the molecular mechanisms of induced pluripotency. *PLoS One* 2008; **3**: e3086.

39. Loh KM, Lim B. A precarious balance: pluripotency factors as lineage specifiers. *Cell Stem Cell* 2011; **8**: 363–369.

40. Li H, Collado M, Villasante A, Strati K, Ortega S, Canamero M *et al.* The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature* 2009; **460**: 1136–1139.

Supplementary Information accompanies the paper on Cell Death and Differentiation website (http://www.nature.com/cdd)