Published in final edited form as:

Nat Cell Biol. 2011 May ; 13(5): 541–549. doi:10.1038/ncb2239.

Incomplete DNA methylation underlies a transcriptional memory of the somatic cell in human iPS cells

Yuki Ohi^{1,2,+}, Han Qin^{1,2}, Chibo Hong³, Laure Blouin^{1,2}, Jose M. Polo^{4,5}, Tingxia Guo^{2,6}, Zhongxia Qi⁷, Sara L. Downey³, Philip D. Manos^{5,8}, Derrick J. Rossi^{5,8,9}, Jingwei Yu⁷, Matthias Hebrok^{2,6}, Konrad Hochedlinger^{4,5}, Joseph F. Costello³, Jun S. Song^{10,*,+}, and Miguel Ramalho-Santos^{1,2,*}

¹Departments of Ob/Gyn and Pathology, Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research and Center for Reproductive Sciences, University of California San Francisco, 513 Parnassus Ave, San Francisco, California 94143-0525, USA

²Diabetes Center, University of California, San Francisco, California 94143-0534, USA

³Department of Neurosurgery, Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, 1450 Third Street, San Francisco, California 94158, USA

⁴Department of Medicine, Howard Hughes Medical Institute, Cancer Center and Center for Regenerative Medicine, Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

⁵Department of Stem Cell and Regenerative Biology, Harvard Stem Cell Institute, Harvard University, Cambridge, Massachusetts 02138, USA

⁶Department of Medicine, University of California San Francisco, 513 Parnassus Ave, San Francisco, California 94143-0540, USA

⁷Department of Laboratory Medicine, University of California San Francisco, 185 Berry Street, San Francisco, California 94107, USA

⁸Stem Cell Program, Children's Hospital Boston, Boston, Massachusetts 02115, USA

⁹Immune Disease Institute, Program in Cellular and Molecular Medicine, Department of Pathology, Children's Hospital Boston, Harvard Medical School, Boston, Massachusetts 02115, USA

*Correspondence should be addressed to M.R.-S. (mrsantos@diabetes.ucsf.edu) or J.S.S. (SongJ@humgen.ucsf.edu).

+These authors contributed equally to this work.

AUTHOR CONTRIBUTIONS

Y.O., J.S.S. and M.R.-S. conceived the project. J.M.P., K.H., P.D.M. and D.J.R. provided reagents. Z.Q. and J.Y. provided assistance with data analysis. C.H. and S.L.D. performed the bisulfite sequencing analysis under supervision of J.F.C. T.G. performed the targeted differentiation to endoderm analysis under supervision of M.H. J.S.S. performed all of the bioinformatic analyses. Y.O., H.Q. and M.R.-S. designed and Y.O. and H.Q. carried out all other experiments with technical assistance from L.B. Y.O., J.S.S. and M.R.-S. wrote the manuscript with input from the other authors.

COMPETING FINANCIAL INTERESTS

The authors declare that they have no competing financial interests.

ACCESSION NUMBERS

The microarray data are available from Gene Expression Omnibus under access number GSE23034.

¹⁰Institute for Human Genetics, Department of Epidemiology and Biostatistics, and Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, 513 Parnassus Ave, San Francisco, California 94143, USA

Abstract

Human induced pluripotent stem (iPS) cells are remarkably similar to embryonic stem (ES) cells, but recent reports suggest that there may be important differences between them. We performed a systematic comparison of human iPS cells generated from hepatocytes (representative of endoderm), skin fibroblasts (mesoderm) and melanocytes (ectoderm). All low passage iPS cells analyzed retain a transcriptional memory of the original cells. The persistent expression of somatic genes can be partially explained by incomplete promoter DNA methylation. This epigenetic mechanism underlies a robust form of memory that can be found in iPS cells generated by multiple laboratories using different methods, including RNA transfection. Incompletely silenced genes tend to be isolated from other genes that are repressed during reprogramming, indicating that recruitment of the silencing machinery may be inefficient at isolated genes. Knockdown of the incompletely reprogrammed gene C9orf64 reduces the efficiency of human iPS cell generation, suggesting that somatic memory genes may be functionally relevant during reprogramming.

Human iPS cells can be derived from differentiated cells by activation of key transcription factors and hold enormous promise in regenerative medicine¹. While iPS cells are remarkably similar to ES cells, there may be important differences between them. Human iPS cells have been suggested to be less efficient than ES cells in targeted differentiation to neural and blood lineages^{2, 3}. Transcriptional differences have also been described and proposed to represent a persistent memory of the original somatic cells in iPS cells⁴⁻⁶. However, it has recently been countered that the transcriptional differences observed may largely be due to lab-specific batch effects^{7, 8}.

The current confusion surrounding this question derives from the poor overlap between gene sets attributed to somatic cell memory in different studies, and from a lack of correlation between gene expression and epigenetic information. Transcriptional differences between human iPS cells and ES cells could not be explained by differences in histone modification patterns^{4, 7}. Recent studies have identified differences in DNA methylation between iPS and ES cells in both mouse and human⁹⁻¹⁴. Mouse iPS cells have been shown to retain a DNA methylation memory of the original somatic cell that may bias iPS cell differentiation towards lineages related to that cell^{12, 14}. However, the DNA methylation differences found between iPS cells and ES cells were largely not demonstrated to correlate with gene expression differences⁹⁻¹⁴. A further limitation stems from the fact that iPS cells generated in different laboratories by different methodologies are often used for comparison^{4, 5}. In addition, the vast majority of human iPS cells analyzed to date, including in two very recent studies of genome-wide DNA methylation^{9, 13}, are derived from fibroblasts, thus limiting the evaluation of a potential memory of the original somatic cell in iPS cells.

We report here a systematic comparison of human iPS cells generated from different somatic cell types. Importantly, all iPS cells analyzed by transcriptional profiling were generated with the same methodology and analyzed in parallel. Our data allow us to

distinguish different types of somatic cell memory in human iPS cells, which can be partially explained by incomplete promoter DNA methylation. We find that the somatic memory gene C9orf64 regulates the efficiency of iPS cell generation, and that incompletely silenced genes tend to be isolated from other genes destined to be silenced during reprogramming.

RESULTS

Generation of human iPS cells from somatic cells representative of all 3 embryonic germ layers

We used a doxycycline-inducible lentivirus transgene system^{15, 16} to generate iPS cells (Supplementary Fig. S1). To have a broad range of starting differentiated states, somatic cells representative of the 3 embryonic germ layers were reprogrammed to iPS cells: adult hepatocytes (Hep) for endoderm, newborn foreskin fibroblasts (Fib) for mesoderm, and adult melanocytes (Mel) for ectoderm (Supplementary Fig. S1). The Mel-iPS cell lines have been previously described¹⁷. iPS cell pluripotency was extensively validated, including colony morphology, growth rate, marker expression, transgene independence, formation of Embryoid Bodies (EBs) and development of teratomas (Fig. 1 and Supplementary Fig. S2 and S4a)¹⁷. Integration analysis indicates that all iPS cell lines used are independent clones (data not shown). We focused our analysis in this study on low passage iPS cells (below passage 20), because they are expected to be more informative about the molecular mechanisms that underlie reprogramming.

Transcriptional profiling of iPS cells and ES cells

The expression levels in Hep, Fib, Mel, and the iPS cells derived from them were profiled in triplicates. In addition, three independent well-established ES cell lines, H1, H7 and H9, and their 8-day EBs were also profiled individually. All samples were analyzed using Affymetrix ST 1.0 microarrays (Supplementary Fig. S1). A hierarchical clustering of the data correctly classified the cell types as shown in Figure 2a. The three iPS cell types clustered together with the ES cells, forming a single branch of pluripotent cell samples. Figure 2b further shows that all somatic cells underwent extensive reprogramming toward an ES cell-like transcriptional profile.

iPS cells retain a transcriptional memory of the original somatic cell

We used the equal-variance *t*-statistic in order to find a global pattern of differential gene expression between iPS and ES cells. We plotted the gene expression differences between iPS cells and ES cells against the differences between the original somatic cells and ES cells and fitted LOESS regression curves to each plot (Fig. 3a and Supplementary Data 1; see Methods). We then performed bootstrap simulations to model noise in gene expression under the assumption that iPS and ES cells are truly identical and that their differences arise from random fluctuations. The actual regression curves lie well outside the intervals of simulated curves, revealing that genes that were highly expressed in somatic cells tend to be repressed but remain higher in iPS cells compared to ES cells, and conversely for genes lowly expressed in somatic cells (Fig. 3b). This pattern was observed for all three types of iPS cells analyzed (Fig. 3a,b; Supplementary Fig. S3a for Fib and Mel).

To find a confident set of differentially expressed genes, we used a robust statistical method, called Differential Expression via Distance Synthesis (DEDS), which combines t-test, moderated t-test, fold-change and Significance Analysis of Microarrays (SAM) into a summary statistic¹⁸. DEDS has been shown to out-perform the individual statistics on spike-in datasets, and its synthesis approach also makes it robust against the limitations of individual tests¹⁸. At 5% False Discovery Rate (FDR), this analysis confirmed that a very significant proportion (~50–60%) of the genes differentially expressed between iPS cells and ES cells represent a memory of the differential expression that already existed between the original somatic cells and ES cells (Fig. 3c, upper Venn diagrams). That is, a statistically significant ($10^{-6} > p > 10^{-16}$, Fisher's exact test) number of genes that were higher in iPS cells relative to ES cells resulted from incomplete silencing during reprogramming. Likewise, a statistically significant ($10^{-9} > p > 10^{-32}$, Fisher's exact test) number of genes that were lower in iPS cells relative to ES cells were the result of incomplete re-activation during reprogramming. No statistically significant overlap was found between genes that change in opposite directions in iPS cells and somatic cells, relative to ES cells (Fig. 3c, lower Venn diagrams). Our analysis thus demonstrates that iPS cells retain a transcriptional memory of the original somatic cells.

We next examined whether transcriptional memory in iPS cells is cell type-specific or associated with multiple differentiated states. In support of a cell type-specific transcriptional memory, $\sim 8 \pm 2\%$ of the genes differentially expressed between an iPS cell type and ES cells were already differentially expressed specifically in the original somatic cell (but not the other somatic cells), relative to ES cells (Supplementary Fig. S3b). However, the majority of the genes differentially expressed between each iPS cell type and ES cells ($52 \pm 5\%$ of total), were found to also be differentially expressed in two or all three somatic cell types relative to ES cells, indicating that they may represent a memory of a general differentiated state. Finally, $\sim 24 \pm 2\%$ of genes differentially expressed between each iPS cell type and ES cells were not differentially expressed between the original somatic cells and ES cells, and thus reflect aberrant transcriptional reprogramming (Supplementary Fig. S3b).

In addition, we do not find evidence of persistent expression of master transcriptional regulators of specific cell types. Microphthalmia-associated transcription factor (MITF) is a master regulator of melanocyte differentiation¹⁹ and regulates a class of melanocyte-specific genes. Our data show that MITF and its target genes TYR and TRPM1 were successfully suppressed in Mel-iPS cells to levels similar to ES cells (DEDS q -value = 0.3 for MITF). Similarly, the hepatocyte nuclear factor (HNF) transcription factors and their target genes highly expressed in hepatocytes²⁰ were reprogrammed in Hep-iPS to the ES cell state (the minimum DEDS q -value for HNFs was 0.2). These findings indicate that key lineage-specifying transcription factors do not seem to play a major role in establishing a persistent somatic transcriptional memory in iPS cells.

Finally, we found no evidence that Hep-iPS cells are more efficient than Fib-iPS cells in targeted differentiation towards endoderm at both the mRNA and protein level (Supplementary Fig. S4b–d and S5). We cannot exclude that differentiation biases towards the somatic cell type of origin might be observed using other targeted differentiation assays,

as has been described in mouse iPS cells^{12, 14}. Taken together, our data suggest that low passage human iPS cells retain a transcriptional memory of the somatic cells, with common as well as cell-specific components.

DNA methylation can partially explain somatic gene expression in iPS cells

We next analyzed available data on genome-wide DNA methylation in ES cells and fibroblasts²¹. We noticed that the top incompletely silenced genes in iPS cells, such as C9orf64, TRIM4 and COMT, showed preponderant promoter DNA methylation only in H1 ES cells and not in IMR90 lung fibroblasts (Supplementary Fig. S6a and S7). In order to perform an unbiased assessment of the contribution of differential DNA methylation to the observed differential expression between iPS and ES cells, the CpG islands of all genes higher in each iPS cell type relative to ES cells were examined for cytosines differentially methylated between IMR90 and H1²¹. We found that genes incompletely repressed in Fib-iPS cells showed a strong trend to be DNA methylated at their promoters in H1 ES cells but not IMR90 fibroblasts (Fig. 4a): the Pearson correlation coefficient between the log expression fold-change Fib-iPS/ES and CpG_{ES>IMR90} was 0.80 ($R^2 = 0.64$ for 12 RefSeq genes with DEDS q -value < 0.05). Strikingly, a similar correlation was found for Hep-iPS (Pearson Correlation = 0.37 for 56 RefSeq genes with DEDS q -value < 0.05) and for Mel-iPS (Pearson Correlation = 0.74 for 14 RefSeq genes with DEDS q -value < 0.05, Methods and Supplementary Fig. S6b). Figure 4b shows that the correlation remains high if we consider only those genes that were differentially expressed in all three iPS cells compared with ES cells, indicating that the contribution of DNA methylation to expression variation is not cell type-dependent. A similar analysis using CpG shores, 2 kb-long flanking regions of CpG islands that have previously been associated with incomplete reprogramming²², yielded only a weak explanation of $R^2 = 0.02$ for the observed variance in differential expression. Our data thus indicate that incomplete establishment of new promoter CpG DNA methylation may occur during reprogramming.

We next performed bisulfite sequencing analysis of promoter CpG methylation for 4 of the top somatic genes whose expression persists in iPS cells, C9orf64, TRIM4, COMT and CSRP1 (Fig. 4c). Consistent with the high expression levels of C9orf64, TRIM4 and COMT in somatic and iPS cells (Supplementary Table 1), the promoters of these 3 genes were depleted of CpG methylation in these cell types, but heavily methylated in ES cells (Fig. 4c and Supplementary Data 2). Consistent with the pattern of gene expression, CSRP1 displayed greater variability but also showed the trend of being most methylated in ES cells, intermediately methylated in all iPS cells and least methylated in the somatic cells. We validated differential methylation using 4 other independent human ES cell lines and 4 other independent iPS cell lines, including iPS cells generated with different methods such as RNA transfection (Supplementary Fig. S6c). In addition, we found that C9orf64, TRIM4 and COMT were also insufficiently methylated in 6 late passage iPS cell lines compared to 5 late passage ES cell lines (all above passage 30), which were analyzed in a recent study (Fig. 4d)⁹. These data indicate that the hypomethylated state of somatic cell genes can persist and correlate with expression in human iPS cells.

Meta-analysis confirms recurrent transcriptional memory associated with DNA methylation

We next sought to determine whether the genes associated with somatic cell memory in our data showed similar expression trends in other published datasets. A pooled analysis of 8 different studies^{4, 11, 23–28} comparing human iPS cells and ES cells revealed that the most incompletely silenced genes in our data, C9orf64 and TRIM4, are within the top differentially expressed genes in these other studies, with an expression ~4-fold higher in iPS than in ES cells (see Methods). We also compared our data to two recent studies that report large datasets comparing iPS cells to ES cells^{7, 29}. Guenther et al.⁷ profiled 7 different ES cell lines and 14 fibroblast-derived iPS cell lines, 6 of which had been treated to excise the reprogramming factors from the genome. Warren et al.²⁹ used synthetic mRNAs to reprogram four different types of fibroblasts and also profiled H1 and H9 ES cell lines. We first pooled together the two datasets using meta-DEDS³⁰, again synthesizing the aforementioned four statistical tests. At 5% FDR, 37 genes are higher in our Fib-iPS cells relative to ES cells, and 10 of them had higher DNA methylation levels in ES cells. 68% of these 37 genes are also higher in the pooled Guenther/Warren iPS cells compared to ES cells (Fig. 5a, “combined”, Fisher test $p = 7.4 \times 10^{-12}$ for the overlap). Strikingly, 9 out of the 10 differentially methylated genes were significantly higher in those iPS cells (Fig. 5a, Fisher test $p = 8.3 \times 10^{-7}$ for the overlap). Not only was the overlap between the genes significant, but their expression levels relative to ES cells also correlated well with our data (Fig. 5b). To test the robustness of this meta-analysis, we also analyzed the Guenther and Warren datasets separately: 5 out of our 10 genes (Fisher test $p = 5.8 \times 10^{-10}$ for the overlap) were also higher in the Guenther iPS cells. 7 out of our 10 genes (Fisher test $p = 1.0 \times 10^{-5}$ for the overlap) were also higher in the Warren iPS cells. Finally, even at the more stringent cutoff of 0% FDR estimated by mDEDS, 6 out of 10 genes (C9orf64, TSPYL5, TRIM4, IQCA1, DNAJC15, CAT, Fisher test $p = 2.1 \times 10^{-12}$ for the overlap) are still found higher in the pooled iPS cells.

We directly assessed a correlation between transcription and DNA methylation in the pooled datasets (Fig. 5c). We performed the expression/DNA methylation regression analysis described earlier (Fig. 4a,b) with the pooled Guenther and Warren data at 0% meta-DEDS FDR. Figure 5c shows that the log (iPS/ES) fold-changes correlate significantly with promoter DNA methylation levels in H1 ES cells (Pearson correlation=0.58, t-distribution p -value= 9.9×10^{-4}), similar to what we had observed for our data (Fig. 4a,b). These results provide an independent validation of our findings that differences in DNA methylation at certain somatic cell genes may underlie their expression in low passage iPS cells, independent of lab-specific variability and reprogramming methods.

The incompletely reprogrammed gene C9orf64 regulates the efficiency of iPS cell generation

We tested whether the expression of incompletely reprogrammed genes in iPS cells is spurious or has any relevance for reprogramming. We performed RNAi for the top incompletely reprogrammed gene, C9orf64, in the context of iPS cell generation. We found that RNAi against C9orf64 during generation of human iPS cells, using 3 independent shRNAs, significantly decreased the total number of Tra1-81+ colonies, compared to infection with the 4 factors alone or together with a non-targeting shRNA control (Fig. 6a).

The C9orf64 knockdown phenotype could be rescued by over-expression of an RNAi-immune cDNA (Supplementary Fig. S8). C9orf64 inhibition did not substantially reduce total cell numbers during the first 10 days of reprogramming, prior to the appearance of colonies (Fig. 6c). These results indicate that C9orf64 is required for efficient iPS cell generation, although its mode of action remains to be determined.

Proximity in the genome affects efficiency of gene silencing in iPS cells

We next sought to gain insight into the mechanisms that underlie persistent expression of somatic genes in iPS cells. DNA methyltransferases (DNMTs) were detected at equivalent levels in iPS cells and ES cells (Fig. 7a), suggesting that the differential methylation observed between iPS cells and ES cells cannot be attributed to insufficient DNMT levels. There is no correlation between the density of promoter CpGs and the extent to which somatic genes are silenced (data not shown). Interestingly, we found a non-random pattern in the genomic locations of incompletely silenced genes: they tend to be isolated from other genes that undergo silencing upon reprogramming (Fig. 7b). These findings suggest that the recruitment of the silencing machinery, including DNA methyltransferases, may be inefficient or delayed at certain somatic genes that are “left behind” due to their isolation.

DISCUSSION

Our data document how remarkably similar to human ES cells are iPS cells generated from different somatic cell types. Nevertheless, we find that iPS cells retain a residual transcriptional memory of the somatic cells, and provide data in support of inefficient promoter DNA methylation as the underlying mechanism. Many factors may contribute to variability in gene expression in human iPS cells, including genetic background, starting somatic cell, method used for reprogramming, culture conditions, passage number and batch effects in microarray studies. Some of these same factors may also affect ES cells and have complicated an analysis of the potential transcriptional differences between human iPS cells and ES cells⁴⁻⁸. The strength of our study resided in comparing human iPS cells generated from different somatic cell types using the same methodology and analyzed in parallel. Our use of gene expression and DNA methylation, rather than gene expression alone, allowed us to find evidence for somatic cell memory in other studies.

It has been shown that promoter DNA de-methylation, a pre-requisite for gene re-activation, can be inefficient during generation of iPS cells^{12, 31}. We report here that DNA methylation and silencing of somatic genes may also contribute to reprogramming (Fig. 8). A complex balance between DNA de-methylation and methylation is therefore likely to be critical for reprogramming. Our data suggest that care should be taken when using small molecules that promote DNA de-methylation in iPS cells, and that an evaluation of the DNA methylation status of somatic cell genes may be warranted in the validation of new human iPS cell lines.

It is important to point out that most of our findings pertain to low passage (<p20) human iPS cells, and that many of the differences relative to ES cells are expected to be attenuated, although possibly not completely abolished (see Fig. 4d), with extensive passaging^{4, 14}. The expression profile of ES cells, on the other hand, has been suggested to be relatively stable with passaging⁴. It will nevertheless be important to determine if variability between ES cell

lines, or any gene expression changes that ES cells may develop with continued culture, are also mediated by differential DNA methylation.

The C9orf64 RNAi data suggest that some somatic genes may continue to be expressed in low passage iPS cells because they play an active role during reprogramming. C9orf64 is a conserved gene of unknown function with no known protein domains. It is possible that it is required to stabilize an intermediate stage with characteristics of both the somatic and the reprogrammed state, although further studies will be required to address this.

Our data indicate that gene density can affect the efficiency with which genes are silenced. The proximity of multiple genes being repressed may synergize in recruiting the silencing machinery, whereas silencing may be inefficient or delayed in more isolated regions, where stochastic events thought to underlie the reprogramming process^{32, 33} may have a lower probability of occurring (Fig. 8). It will be of interest to determine how positional effects in the genome affect the efficiency of epigenetic and transcriptional reprogramming.

Interestingly, several of the somatic cell memory genes reported here have been associated with cancer. TSPYL5 is silenced and DNA methylated in a subset of cancers^{34–36}. C9orf64 is deleted in some cases of Acute Myeloid Leukemia³⁷, and its promoter region is methylated in some breast cancer cell lines³⁸. CSRP1 has been proposed to be a tumor suppressor silenced by DNA methylation in hepatocellular carcinoma³⁹. It is therefore possible that deletion or epigenetic silencing of genes associated with somatic cell memory may contribute to cancer progression. Indeed, our preliminary findings indicate that the incompletely silenced genes reported here show a significant trend for down-regulation during progression of hepatocellular carcinoma (data not shown). Our results prompt an evaluation of the role of somatic cell memory genes in cancer models.

METHODS

Lentivirus Production

The doxycycline-inducible lentiviral vectors and a lentiviral vector constitutively expressing rtTA used in our study have been previously described¹⁶. For virus production, 293T cells at 60–70% confluency were transfected in 10 cm plates with 4 µg of the lentiviral vectors together with 1 µg each of the packaging plasmids VSV-G, MDL-RRE and RSVr using Eugene 6 (Roche). Viral supernatants were harvested after 72 hours, filtered and concentrated with 1 ml of cold PEG-it Virus Precipitation Solution (System Biosciences) for every four volumes of virus. The virus supernatant and PEG-it mixture was incubated overnight at 4°C. The mixture was centrifuged at 1500 × g for 30 minutes at 4°C, resuspended in 100 µl cold phosphate-buffered saline (PBS), and stored at –80°C. Lentiviral infections were performed in 1 ml medium using 10 µl rtTA, 5 µl each of OCT4, SOX2, KLF4 and NANOG, and 2 µl cMYC for one well of a 6-well plate. 8 µg/ml polybrene (Sigma) was used for each infection.

Cell Culture and Human iPS Cell Generation

Human primary newborn foreskin (BJ) fibroblasts were obtained from ATCC (reference #: CRL-2522) and cultured in DMEM with 10% FBS, 1x glutamine, 1x nonessential amino

acids, 1x sodium pyruvate, 2x penicillin/streptomycin, and 0.06 mM β -mercaptoethanol (fibroblast medium). For lentiviral infections of fibroblasts, 50,000 cells were plated per well of a 6-well plate and infected overnight. The day after infection, the virus was removed and replaced with fresh fibroblast medium. 48 hours after infection, 1–6 well of infected cells was trypsinized and seeded onto irradiated mouse embryonic feeders (MEFs) in DMEM/F12 with 20% KSR, 0.5x glutamine, 1x nonessential amino acids, 2x penicillin/streptomycin, 0.1 mM β -mercaptoethanol, 10 ng/ml bFGF (hES cell medium) containing 2% FBS and 1 μ g/ml doxycycline in 10 cm plate format. The melanocytes were obtained from Promocell (reference #: C-12402). The NANOG transgene was not used for deriving melanocyte-iPS cells (only the dox-inducible 4 factors were used¹⁷).

Adult human primary hepatocytes were obtained from Lonza (reference #: CC-2703W6) and cultured in human hepatocyte growth medium (HCM, Lonza). Hepatocytes were received as non-proliferating monolayers of cells shipped in 6-well plate format. Upon arrival, shipping medium was replaced with fresh HCM and the cells were allowed to recover in a 5% CO₂, 37°C incubator for approximately 2 hours prior to infection. Virus infections were carried out in 1 ml HCM per well of a 6-well plate on 2 subsequent days. The day after the last infection, cells were mechanically dissociated into single cells and seeded in HCM onto irradiated MEFs in 10 cm plate format. The following day, cells were transferred to hES cell medium containing 1 μ g/ml doxycycline and fed with this medium daily until the appearance of hES cell-like colonies (up to 40 days). In all cases of human somatic cell reprogramming, live Tra1-81 staining was performed as previously published²⁶.

Immunohistochemical Analysis

Human ES and iPS cells were fixed directly in culturing plates (for pluripotency marker analysis) or on glass cover slips (for the targeted differentiation analysis) with 4% paraformaldehyde, and permeabilized with 0.1% Triton X-100. Cells were then stained with primary antibodies against SSEA-3 (MAB4303, Millipore), SSEA-4 (MAB4304, Millipore), Tra1-60 (ab16288, Abcam), Tra1-81 (MAB4381, Millipore), FOXA2 (07633, Upstate), SOX17 (AF1924, R&D Systems) and HNF1b (AF3330, R&D Systems). Respective secondary antibodies were conjugated to either Alexa Fluor 594 or Alexa Fluor 488 (Invitrogen). Cell counting was done with Cellprofiler 2.0.

Quantitative PCR

RNA was isolated from cells using the RNeasy Mini RNA Isolation kit (Qiagen). cDNA was produced with the High-Capacity cDNA Reverse Transcription kit (Applied BioSystems) using random primers. Real-time quantitative PCR (qRT-PCR) reactions were performed in triplicate with the SYBR Green qRT-PCR Master Mix (Applied BioSystems) and run on an Applied BioSystems 7900HT Sequence Detection System. Primer sequences are listed in Supplementary Table 3.

Stochastic Differentiation to EBs

Human ES and iPS cells were lifted from feeder cells using a 1:1 ratio of Dispase/Collagenase IV mix (1 mg/ml each). 1 ml of the Dispase/Collagenase IV mixture was used for one well of a 6-well plate. Cells were then grown in suspension culture with Knockout

DMEM containing 20% FBS, 0.5x glutamine, 1x nonessential amino acids, and 0.1 mM β -mercaptoethanol. Embryoid bodies (EBs) were collected and analyzed at d8 for markers of the 3 embryonic germ layers.

Directed Differentiation to Endoderm

iPS and ES cells were differentiated towards endoderm using a published protocol⁴⁰ (Supplementary Fig. S4b). Two clones each of Hep-iPS cells and Fib-iPS cells and 2 lines of ES cells (H1 and H9) were used in this analysis. Cells were collected on d3 (definitive endoderm stage) and d6 (primitive gut tube stage) after differentiation and processed for either qRT-PCR or immunohistochemical analysis.

Teratoma Induction

Human ES and iPS cells were grown to 70–80% confluency in 6-well plate format and one entire plate-worth of cells was used to inject one immunocompromised SCID/Beige mouse subcutaneously into 2 sites near the hind flanks. Each 6-well plate worth of cells was pelleted and resuspended in 140 μ l of DMEM/F12 and immediately prior to injection, 60 μ l of Matrigel (BD Biosciences) was mixed with the cells for a total volume of 200 μ l. 100 μ l of the cell/Matrigel mix was injected into each site. Tumors developed after 6–12 weeks and were processed for histological analysis.

Expression Data Analysis

The Affymetrix ST 1.0 expression data were normalized together using RMA and the latest RefSeq probe mapping to the reference human genome^{41, 42}. To minimize redundancy, RefSeq probes corresponding to the same Gene Symbol were combined if they show no within-array variation for all 24 samples. This filtering process yielded a final list of 26,532 RefSeq genes. The equal-variance *t*-test was used to assess the significance of differential expression between groups. The expression profiles of the three ES cell lines were pooled together into one group. ANOVA was performed to find 453 genes that are significantly different among the 8 groups (Hep, Hep-iPS, Fib, Fib-iPS, Mel, Mel-iPS, ES, EB) at a *p*-value cutoff of 10^{-14} . Figure 2a shows the average-linkage clustering of the samples using those genes.

Bootstrap Simulation

Assuming the null hypothesis that the log expression levels for each gene are identically distributed in ES and iPS cells, we estimated a normal null distribution separately for each gene by using maximum likelihood on the pooled dataset of 3 iPS and 3 ES replicates. Six independent samples were then drawn from the normal distribution for each gene and grouped into 3 ES vs. 3 iPS; one complete parametric bootstrap simulation consisted of such re-sampling for all RefSeq genes on the microarray. A LOESS curve was fitted to *t*-test *p*-values for each bootstrap simulation. The entire process was repeated 1000 times, and Figure 3a shows the enveloping curves for the simulated LOESS regression.

Independent Confirmation of Incompletely Silenced Genes

We pooled together 24 iPS cell and 18 ES cell expression profiles from Gene Expression Omnibus (GSE18226, GSE14711, GSE9865, GSE16654, GSE6561, GSE7896, GSE9440, GSE15176). The data were normalized together using RMA and then corrected for potential batch effects using an empirical Bayes method⁴³.

Meta-analysis of published iPS cell expression profiles

The data from Guenther et al.⁷ (GSE23402) and Warren et al.²⁹ (GSE23583) were normalized together using RMA, as described above. We used the Bioconductor package DEDS¹⁸. We performed 2000 permutations and used 5% FDR as a cutoff for deciding differential expression. Meta-DEDS was used to pool together the two datasets, again applying 2000 permutations and 5% or 0% FDR.

CpG methylation analysis

We consider a CpG island to be associated with a gene if it contains the transcription start site (TSS) of the gene or if one of its edges lies within 2kb from the TSS of the gene. Using the DEDS method¹⁸, 64 RefSeq genes were found to be expressed at a higher level in Fibroblast iPS cells than ES cells at the q -value cutoff of 0.05. Among the 64 genes, 12 genes had differentially methylated cytosines between IMR90 and ES cells in their CpG islands located within 2kb.

For the 12 genes, we define $f = \log$ expression fold-change between Fib-iPS and ES. (Note that $f > 0$ if the expression is higher in the iPS cell).

Let $\text{CpG}_{\text{ES} > \text{IMR90}}$ = number of C with higher methylation in ES compared with IMR90.

Let $\text{CpG}_{\text{IMR90} > \text{ES}}$ = number of C with higher methylation in IMR90 compared with ES.

The Pearson correlation between f and $\text{CpG}_{\text{ES} > \text{IMR90}}$ in the corresponding CpG island is 0.80 and the p -value for the correlation is 1.9×10^{-3} . (The corresponding correlation and p -value are 0.37 and 5.1×10^{-3} for Hep-iPS and 0.74 and 2.5×10^{-3} for Mel-iPS). Six genes were differentially expressed in all iPS cells compared to ES cells at a DEDS q -value cutoff of 0.05 and had differentially methylated CpG islands between IMR90 and ES cells. The Pearson correlation between f and $\text{CpG}_{\text{ES} > \text{IMR90}}$ for those genes is 0.88 and p -value = 0.02.

A least squares linear regression model was fitted to the log differential expression fold changes with $\text{CpG}_{\text{ES} > \text{IMR90}}$ and $\text{CpG}_{\text{IMR90} > \text{ES}}$ as two predictors. Only $\text{CpG}_{\text{ES} > \text{IMR90}}$, and not $\text{CpG}_{\text{IMR90} > \text{ES}}$, contributed significantly to the model. The statistical package R was used for the computations.

Clonal Bisulfite Sequencing

Total genomic DNA underwent bisulfite conversion following an established protocol⁴⁴ with modification of: 95°C for 1 min, 50°C for 59 min for a total of 16 cycles. Regions of interest were amplified with PCR primers (Supplementary Table 2) and were subsequently cloned using pCR2.1/TOPO (Invitrogen). Individual bacterial colonies were subject to PCR using vector-specific primers and sequenced using an ABI 3700 automated DNA sequencer.

RNAi in reprogramming

Newborn foreskin fibroblasts were seeded at 30,000 cells per well of a 6-well plate the day before infection. Cells were infected with 0.5 μ l each of concentrated retroviruses (obtained from the Harvard Gene Therapy Initiative) leading to the overexpression of OCT4, SOX2 and KLF4 and 0.05 μ l in the case of cMYC, alone or in combination with 50 μ l of non-concentrated lentivirus for a non-targeting shRNA (ATCTCGCTTGGGCGAGAGTAAG), C9orf64 shRNA (3 independent shRNAs – shRNA1: CATGTTTGCTGATTATAGA; shRNA2: CTTTGATATTTAGAGAACA; shRNA3: GAGGTTATAGGAAATTGAT), or a p53 shRNA (GACTCCAGTGGTAATCTACT). Cells were infected in 1 ml hES cell medium (see “Cell Culture and Human iPS Cell Generation”) and 8 μ g/ml polybrene. Cells remained in the presence of virus for 48 hours and on the day after virus addition, 1 ml of fibroblast medium was added. 48 hours after infection, virus was removed and cells were cultured in ES cell medium. On d20-d28 after infection, live Tra1-81 staining was performed in order to identify fully reprogrammed iPS cell colonies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to thank Susan Fisher, Olga Genbachev, Andy Leavitt and Bruce Conklin for expert advice on culturing human ES cells, Deepa Subramanyam and Robert Blelloch for the Adult Fibroblast-iPS 1 cell line, Linda Ta, Alexander Williams and Alisha Holloway at the Gladstone Institutes, Jennifer Bolen at the Mouse Pathology Core Facility for expert assistance, Jochen Utikal for technical advice, and Jean Yang and Anna Campain for sharing their meta-DEDS code. We thank members of the Santos lab, Robert Blelloch, Holger Willenbring, Susan Fisher and Marica Grskovic for helpful discussions and critical reading of the manuscript. Work in the Santos lab is supported by CIRM, JDRF, an NIH Director’s New Innovator Award and the Helmsley Trust. Y.O. was partially supported by the UCSF Diabetes Center and a T32 grant from the NICHD to the UCSF Center for Reproductive Sciences. T.G. was partially supported by the Helmsley Trust. S.L.D. was partially supported by CIRM. J.S.S. was partially supported by the PhRMA Foundation.

References

1. Yamanaka S. A fresh look at iPS cells. *Cell*. 2009; 137:13–17. [PubMed: 19345179]
2. Feng Q, et al. Hemangioblastic derivatives from human induced pluripotent stem cells exhibit limited expansion and early senescence. *Stem Cells*. 2010; 28:704–712. [PubMed: 20155819]
3. Hu BY, et al. Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *Proc Natl Acad Sci U S A*. 2010; 107:4335–4340. [PubMed: 20160098]
4. Chin MH, et al. Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell*. 2009; 5:111–123. [PubMed: 19570518]
5. Ghosh Z, et al. Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. *PLoS One*. 2010; 5:e8975. [PubMed: 20126639]
6. Marchetto MC, et al. Transcriptional signature and memory retention of human-induced pluripotent stem cells. *PLoS One*. 2009; 4:e7076. [PubMed: 19763270]
7. Guenther MG, et al. Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. *Cell Stem Cell*. 2010; 7:249–257. [PubMed: 20682450]
8. Newman AM, Cooper JB. Lab-specific gene expression signatures in pluripotent stem cells. *Cell Stem Cell*. 2010; 7:258–262. [PubMed: 20682451]

9. Bock C, et al. Reference Maps of Human ES and iPS Cell Variation Enable High-Throughput Characterization of Pluripotent Cell Lines. *Cell*. 2011; 144:439–452. [PubMed: 21295703]
10. Deng J, et al. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat Biotechnol*. 2009; 27:353–360. [PubMed: 19330000]
11. Doi A, et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet*. 2009; 41:1350–1353. [PubMed: 19881528]
12. Kim K, et al. Epigenetic memory in induced pluripotent stem cells. *Nature*. 2010; 467:285–290. [PubMed: 20644535]
13. Lister R, et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature*. 2011; 471:68–73. [PubMed: 21289626]
14. Polo JM, et al. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat Biotechnol*. 2010; 28:848–855. [PubMed: 20644536]
15. Hockemeyer D, et al. A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. *Cell Stem Cell*. 2008; 3:346–353. [PubMed: 18786421]
16. Maherali N, et al. A high-efficiency system for the generation and study of human induced pluripotent stem cells. *Cell Stem Cell*. 2008; 3:340–345. [PubMed: 18786420]
17. Utikal J, Maherali N, Kulalert W, Hochedlinger K. Sox2 is dispensable for the reprogramming of melanocytes and melanoma cells into induced pluripotent stem cells. *J Cell Sci*. 2009; 122:3502–3510. [PubMed: 19723802]
18. Yang YH, Xiao Y, Segal MR. Identifying differentially expressed genes from microarray experiments via statistic synthesis. *Bioinformatics*. 2005; 21:1084–1093. [PubMed: 15513985]
19. Hemesath TJ, et al. microphthalmia, a critical factor in melanocyte development, defines a discrete transcription factor family. *Genes Dev*. 1994; 8:2770–2780. [PubMed: 7958932]
20. Odom DT, et al. Control of pancreas and liver gene expression by HNF transcription factors. *Science*. 2004; 303:1378–1381. [PubMed: 14988562]
21. Lister R, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009; 462:315–322. [PubMed: 19829295]
22. Irizarry RA, et al. Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res*. 2008; 18:780–790. [PubMed: 18316654]
23. Avery K, Avery S, Shepherd J, Heath PR, Moore H. Sphingosine-1-phosphate mediates transcriptional regulation of key targets associated with survival, proliferation, and pluripotency in human embryonic stem cells. *Stem Cells Dev*. 2008; 17:1195–1205. [PubMed: 18393631]
24. Baker DE, et al. Adaptation to culture of human embryonic stem cells and oncogenesis in vivo. *Nat Biotechnol*. 2007; 25:207–215. [PubMed: 17287758]
25. Li SS, et al. Target identification of microRNAs expressed highly in human embryonic stem cells. *J Cell Biochem*. 2009; 106:1020–1030. [PubMed: 19229866]
26. Lowry WE, et al. Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc Natl Acad Sci U S A*. 2008; 105:2883–2888. [PubMed: 18287077]
27. Soldner F, et al. Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell*. 2009; 136:964–977. [PubMed: 19269371]
28. Yu J, et al. Human induced pluripotent stem cells free of vector and transgene sequences. *Science*. 2009; 324:797–801. [PubMed: 19325077]
29. Warren L, et al. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell*. 2010; 7:618–630. [PubMed: 20888316]
30. Campain A, Yang YH. Comparison study of microarray meta-analysis methods. *BMC Bioinformatics*. 2010; 11:408. [PubMed: 20678237]
31. Mikkelsen TS, et al. Dissecting direct reprogramming through integrative genomic analysis. *Nature*. 2008; 454:49–55. [PubMed: 18509334]
32. Hanna J, et al. Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature*. 2009; 462:595–601. [PubMed: 19898493]

33. Meissner A, Wernig M, Jaenisch R. Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nat Biotechnol.* 2007; 25:1177–1181. [PubMed: 17724450]
34. Chapman EJ, Kelly G, Knowles MA. Genes involved in differentiation, stem cell renewal, and tumorigenesis are modulated in telomerase-immortalized human urothelial cells. *Mol Cancer Res.* 2008; 6:1154–1168. [PubMed: 18644980]
35. Jung Y, Park J, Bang YJ, Kim TY. Gene silencing of TSPYL5 mediated by aberrant promoter methylation in gastric cancers. *Lab Invest.* 2008; 88:153–160. [PubMed: 18059362]
36. Kim TY, Zhong S, Fields CR, Kim JH, Robertson KD. Epigenomic profiling reveals novel and frequent targets of aberrant DNA methylation-mediated silencing in malignant glioma. *Cancer Res.* 2006; 66:7490–7501. [PubMed: 16885346]
37. Sweetser DA, et al. Delineation of the minimal commonly deleted segment and identification of candidate tumor-suppressor genes in del(9q) acute myeloid leukemia. *Genes Chromosomes Cancer.* 2005; 44:279–291. [PubMed: 16015647]
38. Cai LY, et al. Identification of PRTFDC1 silencing and aberrant promoter methylation of GPR150, ITGA8 and HOXD11 in ovarian cancers. *Life Sci.* 2007; 80:1458–1465. [PubMed: 17303177]
39. Hirasawa Y, et al. Methylation status of genes upregulated by demethylating agent 5-aza-2'-deoxycytidine in hepatocellular carcinoma. *Oncology.* 2006; 71:77–85. [PubMed: 17341888]
40. Kroon E, et al. Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. *Nat Biotechnol.* 2008; 26:443–452. [PubMed: 18288110]
41. Dai MH, et al. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Research.* 2005; 33:E175–175. [PubMed: 16284200]
42. Irizarry RA, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics.* 2003; 4:249–264. [PubMed: 12925520]
43. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007; 8:118–127. [PubMed: 16632515]
44. Grunau C, Clark SJ, Rosenthal A. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res.* 2001; 29:E65–65. [PubMed: 11433041]

Figure 1. Pluripotency validation for the Hepatocyte-iPS cells derived and used for the microarray studies

(a) The 3 Hep-iPS clones used in this analysis showed strong, positive staining for all human ES cell specific-markers analyzed by immunostaining. Scale bar represents 300 μ m. (b) All Hep-iPS clones showed high expression of endogenous pluripotency markers and negligible levels of transgene expression by qRT-PCR. Values were standardized to GAPDH and Ubb, then normalized to H9 ES cells (endogenous) or 5-factor infected hepatocytes + dox for 4 days (viral). Data are from triplicate reactions. Error bars represent standard deviations. (c) All Hep-iPS clones formed embryoid bodies in vitro when grown under non-attachment conditions. Shown here are d8 EBs alongside with control ES cell-derived EBs. Scale bar represents 200 μ m. (d) Pluripotency of the Hep-iPS cell clones was further confirmed by their ability to form teratomas in vivo, comprised of tissues derived from all 3 germ layers. (i) Neural tissue (ectoderm), (ii) Striated muscle and adipocytes (mesoderm), (iii) gut-like epithelium (endoderm). Also see Supplementary Fig. S2 for pluripotency validation of Fib-iPS cells used for the microarray analysis. Mel-iPS cells have previously been described¹⁷.

Figure 2. Multiple cell types undergo extensive transcriptional reprogramming to the human iPS cell state

(a) Average-linkage hierarchical clustering of the RMA-normalized expression profiles shows that the replicate data cluster together tightly, confirming the reproducibility of the experiments, and that the somatic cells have been successfully reprogrammed. (b) The box-plot of log expression fold changes for all RefSeq genes further shows that the iPS cells have been reprogrammed to closely resemble the transcriptional profiles of ES cells. The black center line represents the median. The upper and lower edges of the box represent the first and third quartiles, and they define the inter-quartile range. Outliers farther than 1.5 times the inter-quartile range from the box are shown as circles.

Figure 3. iPS cells retain a transcriptional memory of the original somatic cell

(a) LOESS curves fitted to the scatter plots of t -test $\log p$ -values: $-\log(p)$ and $\log(p)$ are plotted for fold changes greater than 1 and less than 1, respectively. The black line is a curve fitted to our data, and other curves are fitted to the 1000 bootstrap simulation datasets obtained by assuming identically distributed iPS and ES cell expression levels. The black line shows clear deviation from the null hypothesis $iPS=ES$ and thus reflects the trend that the transcriptional memory of the originating cell type is retained in low passage iPS cells: genes that were higher (or lower) in the somatic cell than in ES cells tend to be significantly repressed (or induced) during reprogramming, but nevertheless remain higher (or lower) in iPS cells than in ES cells. (b) Box-plots of expression levels for 191 genes that are higher in both iPS cells and somatic cells relative to ES cells (upper right corner genes in (a)) and 391 genes that are lower in both iPS cells and somatic cells relative to ES cells (lower left corner genes) at a t -test p -value cutoff of 0.01. The plots illustrate progressive convergence of somatic gene expression towards the ES cell state. (c) The top panel shows progressively reprogrammed genes (somatic > iPS > ES or somatic < iPS < ES), and the bottom panel shows over-reprogrammed genes (somatic < ES < iPS or somatic > ES > iPS). The p -values for the overlaps are from Fisher's exact test, and show significant overlaps only for progressively reprogrammed genes. The standard deviations indicate variation among the 3 cell types.

Figure 4. DNA methylation can partially explain somatic gene expression in iPS cells

(a) The genes which maintain higher expression levels in Fib-iPS cells compared to ES cells tend to be also methylated at higher levels in H1 ES cells compared to the fibroblast cell line IMR90. The Pearson correlation coefficient between the log expression fold change and single-nucleotide resolution differences in CpG island methylation was 0.80 ($R^2=0.64$, p -value=0.002). (b) The correlation was 0.88 ($R^2=0.78$, p -value=0.02) for six genes which remain higher in all three iPS cell types compared to ES cells. $\text{CpG}_{\text{ES} > \text{IMR90}}$ is the number of cytosines in CpG islands with higher methylation in H1 than IMR90. (c) The overall DNA methylation of 4 of the top somatic genes whose expression persists in low passage iPS cells. DNA methylation was examined with bisulfite sequencing analysis in 3 types of somatic cells (hepatocytes, fibroblasts and melanocytes), 2 clones for each iPS cell type and H1 and H9 ES cells. The detailed bisulfite sequencing data for all samples can be found in Supplementary Data 2. (d) Higher passage iPS cells retain incomplete DNA methylation at somatic cell memory genes. CpG island methylation levels were examined for our validated somatic memory genes (Fig. 4c) in 5 ES cell lines and 6 iPS cell lines with passage number >30 (passage range 30–58, data from a recent study⁹). The boxplot shows the difference in methylation levels between the higher passage ES and iPS cells. One-sided Wilcoxon test p -values confirm that C9orf64, TRIM4 and COMT are still resistant to promoter DNA methylation (i.e., they are hypomethylated) in high passage iPS cells relative to high passage ES cells. No significant difference in DNA methylation was found for the more variable of the 4 genes, CSRP1.

Figure 5. Meta-analysis of DNA methylation-associated transcriptional memory in independent datasets

(a) 37 genes expressed at higher levels in our Fib-iPS cells relative to ES cells tend to show higher expression in the iPS cells generated by Guenther et al.⁷ and Warren et al.²⁹ but there is high variability when expression data alone are used (cyan box-plots). However, when we use 10 differentially expressed genes from our data that were also differentially DNA methylated in ES cells, a greater proportion show persistent higher expression in the iPS cells of the two datasets (yellow box-plots). (b) The heat-map shows the iPS/ES fold-change ranking of the 10 genes that are higher in our Fib-iPS cells and also methylated in ES cells. (The higher the rank, the greater the fold-change). Shown are the Spearman rank correlation coefficients of fold-changes between our data and those of Guenther et al.⁷ and Warren et al.²⁹ (c) 29 genes were significantly higher in iPS cells relative to ES cells in a pooled analysis of the Guenther and Warren datasets and also differentially methylated in ES cells²¹. Differential expression was determined by applying meta-DEDS analysis to the pooled dataset at a stringent cutoff of 0% FDR. The figure shows that the fold-change levels of those genes correlate significantly with DNA methylation levels ($p\text{-value} = 9.9 \times 10^{-4}$): the higher the fold change in iPS cells relative to ES cells, the higher the promoter DNA methylation in H1 ES cells relative to IMR90 fibroblasts.

Figure 6. The somatic cell memory gene C9orf64 is required for efficient generation of iPS cells
 (a) The number of Tra1-81+ iPS cell colonies was counted on d20 after infection of BJ foreskin fibroblasts with 4f alone (4f), 4f+non-targeting shRNA (4f+NTi), 4f+C9orf64 shRNA (3 different short hairpins targeting C9orf64 were independently tested, 4f+Ci1, 4f+Ci2 and 4f+Ci3) or 4f+p53 shRNA (4f+p53i). Infections were performed in duplicate. Knockdown of C9orf64 resulted in a significant reduction in the number of Tra1-81+ iPS cell colonies compared to 4f alone, 4f+NTi or 4f+p53i. (b) Reduction in the levels of C9orf64 expression achieved by each of the 3 shRNA constructs was confirmed by qRT-PCR. The 4f, 4f+NTi and 4f+p53i conditions showed no significant reduction in C9orf64 expression. Fib-iPS and H9 hES cells served as positive and negative controls for C9orf64 expression, respectively. p53 expression analysis further validated the specificity of the shRNAs. Values were standardized to GAPDH and Ubb, then normalized to uninfected BJ fibroblasts. Note log2 scale in Y-axis: e.g., -2 equals down 4x, -3 equals down 8x, etc. Data are from triplicate reactions. (c) Growth curves of fibroblasts infected with 4f, 4f+NTi, 4f+Ci1 and 4f+Ci2, counted on d0, d1, d4, d7 and d10 post-infection. Infections were performed in triplicate. C9orf64 RNAi did not substantially alter total cell numbers during

the first 10 days of reprogramming. In all panels data shown are representative of two independent experiments, and error bars represent standard deviations.

Figure 7. Proximity in the genome affects efficiency of gene silencing in iPS cells

(a) The heatmap shows the expression levels of 4 different DNA methyltransferases (DNMTs) in all of the cell types analyzed in our microarray study. The expression level of these DNMTs is relatively equivalent between iPS cells and ES cell controls, suggesting that any differential DNA methylation in iPS cells is not due to insufficient DNMT expression.

(b) We considered 62 silenced genes (“Reprogrammed”) and 5 genes whose expression persists in iPS cells (“Fib-iPS>ES”), with at least 10 cytosines methylated only in ES and also showing higher expression in Fib than ES cells. The “Reprogrammed” genes tend to have nearby genes that also require silencing, while the “Fib-iPS>ES” genes are more isolated. The one-sided Wilcoxon p-values for the difference in the number of nearby genes between the two groups are 0.054, 0.022 and 0.028 for the 20kb, 50kb and 100kb distance restrictions, respectively. The local density of genes, irrespective of expression status, was also slightly lower near genes whose expression persists in iPS cells.

Figure 8. Model for the role of DNA methylation in reprogramming to the human iPS cell state It has previously been shown that DNA de-methylation and re-activation of pluripotency genes is an essential component of reprogramming³¹ (top). In addition, incomplete de-methylation of genes silenced in the somatic cell, including developmental regulators of other lineages, has been shown to persist in mouse iPS cells and may affect their differentiation^{12, 14} (middle). We report here that differential methylation of somatic cell genes underlies their differential expression in human iPS cells (bottom “somatic” panel), and that somatic genes whose expression persists in low passage iPS cells tend to be isolated from other genes that undergo silencing. Clustering of genes requiring simultaneous repression may facilitate recruitment of the silencing machinery, including DNMTs, and regional DNA methylation (top “somatic” panel). Extensive passaging (pink arrows), may lead to further epigenetic silencing of somatic genes in human iPS cells. Arrows indicate active transcription, while hooks indicate repression.