

Induced Pluripotent Stem Cells and Embryonic Stem Cells Are Distinguished by Gene Expression Signatures

Mark H. Chin,¹ Mike J. Mason,^{1,8} Wei Xie,¹ Stefano Volinia,¹⁰ Mike Singer,¹¹ Cory Peterson,³ Gayane Ambartsumyan,² Otaren Aimuwu,² Laura Richter,² Jin Zhang,⁴ Ivan Khvorostov,⁴ Vanessa Ott,¹¹ Michael Grunstein,¹ Neta Lavon,⁹ Nissim Benvenisty,⁹ Carlo M. Croce,¹⁰ Amander T. Clark,^{2,5,6,7} Tim Baxter,¹¹ April D. Pyle,^{3,5,6,7} Mike A. Teitell,^{4,5,6,7} Matteo Pelegrini,^{2,6} Kathrin Plath,^{1,5,6,7,*} and William E. Lowry^{2,5,6,7,*}

¹Department of Biological Chemistry, David Geffen School of Medicine

²Department of Molecular, Cell and Developmental Biology

³Department of Microbiology, Immunology, and Molecular Genetics

⁴Departments of Pathology and Pediatrics, David Geffen School of Medicine

⁵Broad Stem Cell Center

⁶Molecular Biology Institute

⁷Jonsson Comprehensive Cancer Center

⁸Department of Statistics

University of California, Los Angeles, CA 90095, USA

⁹Department of Genetics, Hebrew University, Jerusalem 91904, Israel

¹⁰Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University, Columbus, OH 43210, USA

¹¹Roche NimbleGen, Inc., Madison, WI 53719, USA

*Correspondence: kplath@mednet.ucla.edu (K.P.), blowry@ucla.edu (W.E.L.)

DOI 10.1016/j.stem.2009.06.008

SUMMARY

Induced pluripotent stem cells (iPSCs) outwardly appear to be indistinguishable from embryonic stem cells (ESCs). A study of gene expression profiles of mouse and human ESCs and iPSCs suggests that, while iPSCs are quite similar to their embryonic counterparts, a recurrent gene expression signature appears in iPSCs regardless of their origin or the method by which they were generated. Upon extended culture, hiPSCs adopt a gene expression profile more similar to hESCs; however, they still retain a gene expression signature unique from hESCs that extends to miRNA expression. Genome-wide data suggested that the iPSC signature gene expression differences are due to differential promoter binding by the reprogramming factors. High-resolution array profiling demonstrated that there is no common specific subkaryotypic alteration that is required for reprogramming and that reprogramming does not lead to genomic instability. Together, these data suggest that iPSCs should be considered a unique subtype of pluripotent cell.

INTRODUCTION

Embryonic stem cells (ESCs) are in vitro representations of the inner cell mass of developing embryos (Gokhale and Andrews, 2006) and therefore present a valuable tool for regenerative medicine and serve as models of embryonic development in vitro (Keller, 2005). Induced pluripotent stem cells (iPSCs)

are not derived from embryos but are in vitro constructs thought to mimic ESCs (Hochedlinger and Plath, 2009; Nishikawa et al., 2008; Takahashi and Yamanaka, 2006). Therefore, a number of issues must be addressed before iPSC technology can be applied to regenerative medicine or in vitro modeling of disease or development. Are iPSCs as good as ESCs at replicating the state of bona fide embryonic cells? Do iPSCs generate differentiated progeny as efficiently as ESCs? Do the methods employed to generate iPSCs confound their use in a clinical or experimental setting? These questions should be at the forefront when considering whether iPSCs will serve as useful models of human development and disease. However, before these questions can be answered, it is critical to understand any molecular differences between iPSCs and ESCs in their undifferentiated state.

Even though many groups have now shown that both human (h) and mouse (m) somatic cells can be reprogrammed by overexpression of variable sets of a few transcription factors to what appears to be an embryonic state (Lowry et al., 2008; Maherali et al., 2007; Wernig et al., 2007; Okita et al., 2007; Park et al., 2008; Takahashi et al., 2007; Takahashi and Yamanaka, 2006; Yu et al., 2007), the degree of molecular similarity between iPSCs and ESCs has not been completely elucidated. Every study suggests that iPSCs are “nearly identical” to their embryo-derived counterparts, but it remains unclear whether the small percentage of genes that are differentially expressed between iPSCs and ESCs are shared among different iPSC lines and whether this difference is biologically significant. Careful study is warranted to discern whether these small differences observed between iPSC and ESC lines are particular to individual experiments or whether reprogramming of somatic cells generates a state that is common among iPSCs and unique from ESCs. Because of the methods used to reprogram somatic cells to an embryonic state, iPSCs could possess significant

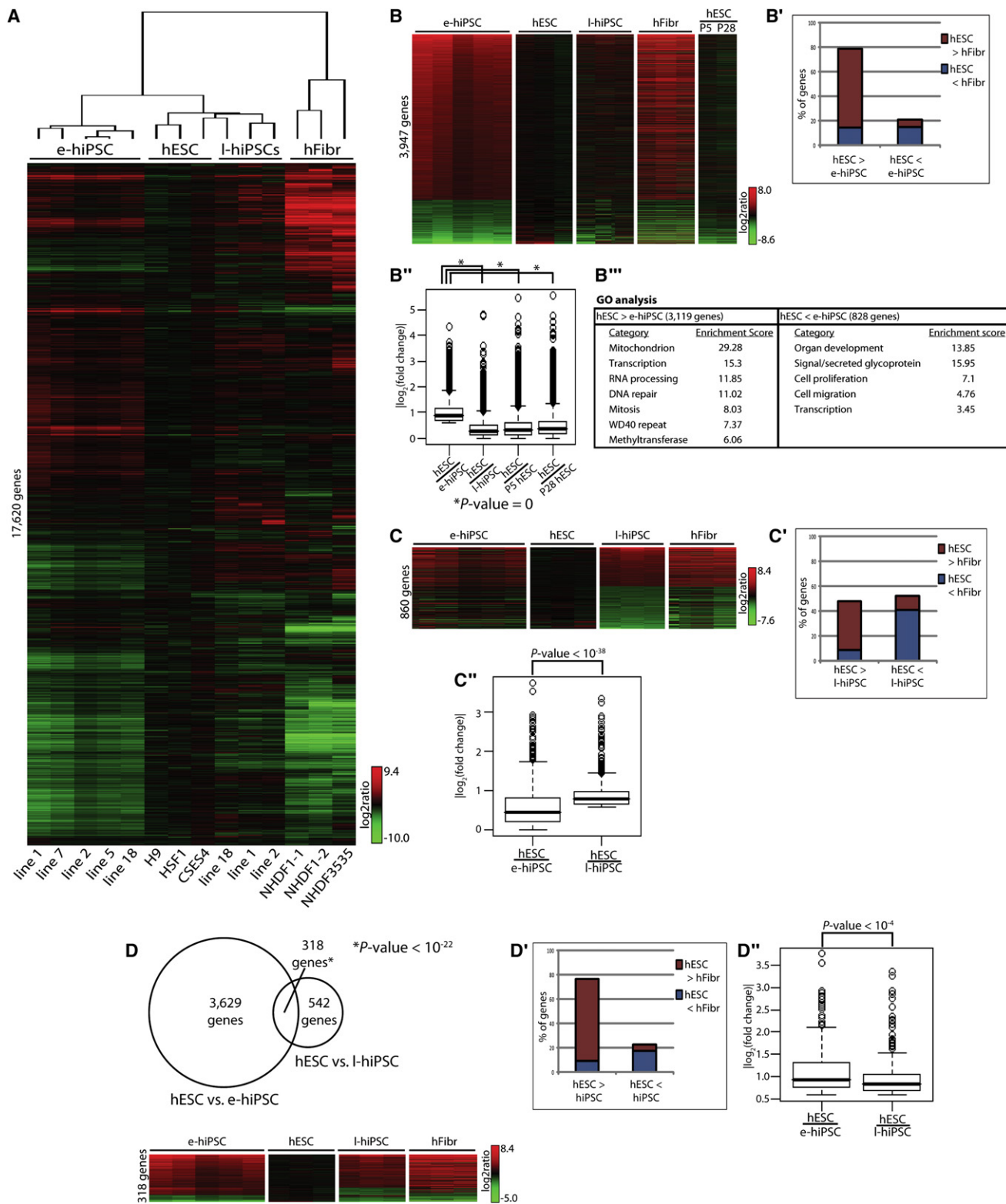


Figure 1. Expression Profiling Demonstrates Differences between hiPSCs and hESCs

(A) Unsupervised hierarchical clustering of global gene expression data in human fibroblasts (hFibr), early-passage hiPSCs (e-hiPSC), late-passage hiPSCs (l-hiPSC), and hESCs. Expression values are presented as the \log_2 ratio of the given gene divided by the average of the ESC lines (all subsequent expression heatmaps are presented in this manner). Individual cell lines used are indicated below the heatmap.

differences at various molecular levels, including the following: genomic integrity; epigenetic stability; noncoding, and perhaps even coding, RNA expression. To date, no one has described the full extent of differences between iPSCs and ESCs, and whether these differences are shared among reprogrammed lines derived by various methods and labs.

Here, we applied genome-wide methods to compare mouse and human iPSCs with ESCs by array CGH, to uncover subkaryotypic genome alterations; coding RNA profiling, to uncover gene expression changes; miRNA profiling, to determine changes in expression of small noncoding RNAs; and histone modification profiling, to determine whether epigenetic changes correlate with gene expression differences. The sum of these analyses uncovers a novel gene expression signature that is unique from ESCs and shared among iPSC lines generated from different species and in different reprogramming experiments. Whether the iPSC signature described here plays a functional role in self-renewal or differentiation warrants extensive further investigation.

RESULTS

Distinct Gene Expression Signatures Are Associated with hiPSCs at Different Passages

To determine whether gene expression differences observed between ESCs and iPSCs are stochastic or indicative of differences between these pluripotent cell types, a detailed genome-wide expression analysis was carried out between three hESC lines that we routinely maintain in the lab (HSF1, H9, and CSE4) and hiPSC clones at different passages (Table S1, a summary of cell lines and passages used in this study, is available online). The hiPSC clones used here were obtained in a single fibroblast reprogramming experiment published previously (Lowry et al., 2008) through retroviral expression of *OCT4*, *SOX2*, *NANOG*, *KLF4*, and *C-MYC*. Five hiPSC clones (#1, 2, 5, 7, and 18), two of which had integrated the *NANOG* virus in addition to the other four factors (clones 1 and 5), were expanded for further analysis of pluripotency, including teratoma formation and in vitro differentiation (Karumbayaram et al., 2009; Lowry et al., 2008; Park et al., 2009). These clones were all profiled at early passage (passages [p] 5–9) and clones 1, 2, and 18 were also analyzed at late passage (p54–61). Unsupervised hierarchical clustering of the expression data across hESCs, early- and late-passage hiPSCs, and fibroblasts highlighted interesting patterns of gene expression between these cell types (Figure 1A). First, even though hiPSCs are considered highly similar to hESCs, they are more similar to each other than to hESCs, as shown previously (Lowry

et al., 2008). Second, late-passage hiPSCs cluster more closely with hESCs than with early-passage hiPSCs. In agreement with these findings, Pearson correlation analysis also demonstrated that the gene expression profile of late-passage hiPSCs is more closely related to hESCs than to early-passage hiPSCs using Fisher's z' -transformation comparison of correlations ($z = 0$, Figure S1).

A Unique Expression Signature for Early-Passage hiPSCs

Analyzing the expression differences between hESC lines and our early-passage hiPSC lines, we found 3,947 (out of 17,620) genes that are significantly different between all hiPSC lines and hESC lines as determined by a Student's t test ($p < 0.05$) and requiring an at least a 1.5-fold expression change between hiPSCs and hESCs (Figure 1B; termed early-passage hiPSC signature genes; Table S2). Since these expression differences to hESCs are shared among all five independent hiPSC clones, the data suggest that hiPSCs represent a common cell type that is similar to but distinct from hESCs. Within this expression signature, 79% of the genes are expressed at a lower level in iPSCs than ESCs (Figure 1B'). Gene Ontology analysis suggests that these genes have a role in basic processes (energy production, RNA processing, DNA repair, mitosis), while genes related to differentiation (organ development and signal/secreted glycoprotein) are more abundantly expressed in hiPSCs than ESCs (Figure 1B'').

These findings suggest that hiPSCs have not efficiently silenced the expression pattern of the somatic cell from which they are derived and failed to induce genes important for undifferentiated, highly proliferative hESCs. Indeed, a classification of the early-passage hiPSC signature genes according to their expression difference between fibroblasts and hESCs shows that 82% of the genes that are expressed at a higher level in hESCs versus hiPSCs are also more highly expressed in hESCs versus fibroblasts (Figure 1B'), indicating that an important difference between hESC and early-passage hiPSCs is the lack of the complete induction of these genes. When analyzing the genes with more abundant transcripts in early-passage hiPSC than hESCs, 71% appear to be inefficiently silenced from the fibroblast state (Figure 1B'). The remaining smaller portion of early hiPSC signature genes can be explained by excessive induction of an ESC-specific expression program or suppression of the fibroblast pattern (Figure 1B').

While the expression differences between early-passage hiPSC and hESC lines appear to be reprogramming dependent, one obvious explanation for the difference could be that we

(B) e-hiPSC signature genes. As in (A) except for the 3947 genes significantly different between e-hiPSC and hESC based on two criteria, Student's t test ($p < 0.05$) and at least a 1.5-fold change. Genes are ordered according to the decreasing average expression ratio between hESCs and e-hiPSCs. Expression data for passage 5 and 28 hESC (P5 and P28) for this set of genes are added to the right. (B') e-hiPSC genes were divided into those expressed at higher levels in hESCs than e-hiPSCs and vice versa. Each of these two groups was further subclassified into two groups, either more highly expressed in hESCs than fibroblasts (red) or more highly expressed in fibroblasts than hESCs (blue). (B'') Boxplots of the absolute value of the \log_2 fold change between hESC and the following: e-hiPSC, I-hiPSC, P5 or P28 hESC (*all $p = 0$). (B''') Gene ontology (GO) analysis for e-hiPSC signature genes upon division into those for which hESC expression was greater than e-hiPSC and vice versa. Only significant GO-terms with an enrichment value >3 ($p < 0.001$) are presented. (C) I-hiPSC signature genes. As in (B) except for the 860 genes significantly different between hESCs and late-passage hiPSCs. (C') Barplot similar to (B') using I-hiPSC signature genes. (C'') Boxplot similar to (B'') for I-hiPSC signature genes. (D) Common hiPSC signature genes were determined as the overlap between early- and late-passage signature genes from (B) and (C), respectively, as shown in the Venn diagram. The expression heatmap below is presented for the 318 genes in the overlap as in (B). (D') Barplot similar to (B') using common hiPSC signature genes. (D'') Boxplot similar to (B'') for common hiPSC signature genes.

compared early-passage hiPSCs with hESCs at higher passage (p37, 41, and 51) since the availability of hESCs at early passage is limited. Thus, the distinct expression pattern of early-passage hiPSCs versus late-passage hESCs could simply be due to differences induced by extended culturing. To estimate the contribution of culture-induced transcriptional changes, early (p5)- and middle (p28)-passage hESCs were obtained, profiled, and compared to our cell lines. This analysis suggested that the vast majority of the genes consistently differentially expressed between early-passage hiPSCs and hESCs do not differ dramatically between early-, middle-, and late-passage hESCs (Figure 1B, far right, and Figure S2). Together, these data indicate that the early-passage hiPSC signature is not a common feature of low-passage pluripotent stem cells but is specific to hiPSCs.

Expression Differences between Early- and Late-Passage hiPSCs

Upon extended passaging, the gene expression profile of hiPSCs appears to become more similar to hESCs (Figures 1A, 1B, and S1). In agreement with this conclusion, late-passage hiPSCs have a significantly decreased amplitude of expression differences for early-passage hiPSC signature genes (Figures 1B' and S3A). As expected, the same was true when comparing early-, middle-, and late-passage hESCs (Figure 1B''). Looking at 48 genes that are specifically expressed in hESCs (taken from Lowry et al., 2008), it is clear that hESC signature genes are expressed at lower levels in all early hiPSC lines but recover after extended culture to a level commensurate with that found in hESCs (Figure S4). These data indicate that many of the expression differences that occur between early-passage hiPSCs and hESCs get resolved upon extended passaging.

However, Figure 1A shows that late-passage iPSC still differ from ESCs. The differential expression between late-passage hiPSCs and hESCs of some of these genes was validated at the protein level (Figure S5). We therefore defined genes that are differentially expressed more than 1.5-fold between late-passage hiPSCs and hESCs and found 860 genes that fit these criteria (Figure 1C; termed late-passage hiPSC signature genes; Student's *t* test, $p < 0.05$; Table S3). Gene ontology analysis failed to uncover enrichment for any particular functional category among late-passage hiPSC signature genes, in agreement with the finding that at late passage the majority of expression differences of hiPSCs with hESCs that exist at early passage are resolved. Comparing fibroblast, hESC, and hiPSC expression, we found that 80% of the late-passage hiPSC signature can be attributed to inefficient silencing of the fibroblast expression pattern and lack of full induction of hESC-specific genes, similar to what was found for the early-passage hiPSC gene expression signature (Figure 1C'). In agreement with this notion, 318 genes (37%) are shared between early- and late-passage hiPSCs versus hESCs (Figure 1D; Table S4). This enduring (also termed common) hiPSC signature is clearly a result of differences between cells generated by the reprogramming process versus those derived from human embryos and does not appear to differ dramatically in expression between early and later passages of hiPSC (Figures 1D' and S3C). Nearly all of the genes in this group insufficiently induce hESC-specific genes and suppress fibroblast-specific genes (Figure 1D').

Furthermore, the common hiPSC signature genes exhibit the most dramatic change in gene expression between fibroblasts and hESCs among all signature expression groups (Figure S7). Surprisingly, many late-passage hiPSC signature genes are more similarly expressed between early hiPSCs and hESCs than in late hiPSCs (Figures 1C, 1C'', and S3B). This is consistent with the notion that an overall readjustment of the expression signature occurs upon passaging of hiPSCs, rather than simply closing in on the ESC expression. Taken together, the comparison of hiPSC and hESC expression patterns indicates that (1) at early passage hiPSC lines are incompletely reset to a hESC-like expression pattern and (2) even at late passage differences between hESCs and hiPSCs persist and reflect an imperfect resetting of somatic cell expression to an ESC-like state.

To exclude the possibility that gene expression differences between hESCs and hiPSCs at late passage could be due to differential proliferation of hiPSCs and hESCs, cell-cycle analysis was performed by FACS. This analysis demonstrated that late-passage hiPSCs and hESCs do not proceed through the cell cycle at different rates, and thus the late hiPSC signature is not due to varying proliferation capacity (Figure S6).

Conservation of the hiPSC Expression Signature across Independent Reprogramming Experiments

Next, we determined if expression signatures observed between established hESC lines and hiPSCs from our lab also occur in reprogramming experiments by different labs to establish whether these differences are shared among reprogrammed lines derived by various methods and labs. To this end, we performed a similar analysis as described above with data available from other laboratories (NIH, Gene Expression Omnibus) and compared the overlapping signatures with those signatures derived from our early and late hiPSCs. In Maherali et al. (2008), neonatal foreskin fibroblasts were reprogrammed to the iPSC state by expressing *OCT4*, *SOX2*, *KLF4*, *NANOG*, and *C-MYC* using tetracycline-inducible lentiviral vectors (Maherali et al., 2008). Gene expression profiling from this experiment revealed 1653 genes at least 1.5-fold differentially expressed when comparing their hiPSCs to their hESCs (Figure 2). Of these, 618 overlapped with the 3947 early-passage hiPSC signature genes found in our hiPSC clones ($p < 10^{-47}$; Figure 2).

The same analysis was performed with data from Soldner et al., who reprogrammed dermal fibroblasts obtained from patients with Parkinson's disease using a single doxycycline-inducible lentivirus carrying either four (*OCT4*, *SOX2*, *c-MYC*, and *KLF4*) or three (*OCT4*, *SOX2*, and *KLF4*) reprogramming factors (Soldner et al., 2009). Importantly, in this study, the reprogramming factors were removed after establishment of hiPSC lines because the viral sequences encoding the factors were Cre-recombinase excisable. We found a 1.5-fold differential expression of 899 genes between their hiPSCs and their hESCs before excision of the reprogramming factors (2lox hiPSCs). Of these genes, 329 overlapped with our early-passage hiPSC signature ($p < 10^{-22}$; Figure 2). Following Cre-mediated depletion of the factors and subcloning of the iPSCs (1lox hiPSCs), 553 genes remained differentially expressed following our criteria, and 222 of these genes overlapped our early-passage hiPSCs ($p < 10^{-20}$).

		Chin, et al.		Maherali, et al.	Soldner, et al.		Yu, et al.	
		Early hiPSC	Late hiPSC	hiPSC	2lox hiPSC	1lox hiPSC	Episomal hiPSC	Subcloned hiPSC
Chin, et al.	Early hiPSC	3,947 total genes						
	Late hiPSC	318 genes $P < 10^{-22}$	860 total genes					
	hiPSC	618 genes $P < 10^{-47}$	126 genes $P < 10^{-6}$	1,653 total genes				
Soldner, et al.	2lox hiPSC	329 genes $P < 10^{-22}$	71 genes $P < 10^{-4}$	210 genes $P < 10^{-36}$	899 total genes			
	1lox hiPSC	222 genes $P < 10^{-20}$	49 genes $P < 10^{-4}$	133 genes $P < 10^{-24}$	150 genes $P < 10^{-68}$	553 total genes		
	Episomal hiPSC	306 genes $P < 10^{-17}$	77 genes $P < 10^{-6}$	259 genes $P < 10^{-68}$	120 genes $P < 10^{-23}$	92 genes $P < 10^{-24}$	874 total genes	
Yu, et al.	Subcloned hiPSC	174 genes $P < 10^{-12}$	56 genes $P < 10^{-9}$	143 genes $P < 10^{-37}$	60 genes $P < 10^{-10}$	49 genes $P < 10^{-12}$	343 genes $P = 0$	476 total genes

Yu et al. reprogrammed neonatal foreskin fibroblasts using nonintegrating episomal vectors encoding *OCT4*, *SOX2*, *NANOG*, *c-Myc*, *KLF4*, *LIN28*, and *SV40LT* (episomal hiPSC) (Yu et al., 2009). Upon continuous passaging, the episomal vectors are lost and hiPSC subclones without any ectopic DNA could be isolated (subcloned hiPSCs). An analysis of their expression data again revealed a set of genes that are differentially expressed between hESCs and hiPSCs and a highly significant overlap of these differentially expressed genes with those found differentially expressed between our hiPSCs and hESCs (Figure 2). This finding was particularly relevant not only because the Yu et al. lines never experienced integration, but also because the combination of reprogramming factors used differed slightly from that in other reprogramming experiments.

These analyses described the degree of similarity of differential expression between hESCs and hiPSCs generated in independent experiments. To determine the extent by which the same genes are differentially regulated in the same direction among independent experiments, a similar analysis was performed with the added requirement that direction of the expression change between hESCs and hiPSCs must be conserved in both experiments being compared. These data suggest that many of the genes shown in Figure 2 to be differentially expressed in multiple experiments were also changed in the same direction (Figure S8).

Further analysis to demonstrate the degree of overlap between any three hiPSC signatures also suggests a highly significant overlap. Between the Chin, Maherali, and Soldner signatures, 79 genes were shared ($p < 10^{-44}$); between the Chin, Maherali, and Yu signatures, 106 genes were shared ($p < 10^{-96}$); between Chin, Soldner, and Yu, 48 genes were shared ($p < 10^{-34}$). Among all the experiments of all four laboratories, 15 genes are differentially expressed between early-passaged hiPSCs and hESCs ($p < 10^{-54}$; Table S5). The highly significant overlap between each of all four of these completely independent reprogramming experiments suggests that the hiPSC state is not stochastic. Confirming this conclusion, a gene ontology analysis of the genes differentially expressed in the experiments from the four groups again

Figure 2. Differential Expression Patterns between hESC and hiPSC Are Conserved among Independent Reprogramming Experiments

iPSC signature genes defined as genes differentially expressed between hiPSC and hESC lines (Student's *t* test [$p < 0.05$] and at least a 1.5-fold change) were obtained from Figure 1 of this study (Chin et al.) and from additional published reprogramming experiments. The matrix summarizes the overlap of hiPSC signature genes between the different experiments. The values on the diagonal designate the total number of genes identified as significantly different in expression between hESCs and the indicated hiPSCs. The intersection of the rows and columns give the number of genes that are in common between the two respective experiments and the corresponding significance (*p* value) as determined using Fisher's exact test. For the Soldner et al. experiment (Soldner et al., 2009), data were analyzed before (2lox) and after (1lox) excision of the reprogramming factors. In Yu et al. (2009), iPSCs were generated with episomal vectors and analyzed before (episomal) and after subcloning. Genomic integrations were not detected for any of these subclones. The Maherali iPSCs (Maherali et al., 2008) were reprogrammed with integrating lentiviruses.

suggest that the signatures that arise in each reprogramming experiment share a functional similarity (Figure S9).

Conservation of an iPSC Expression Signature between Mouse and Human Reprogramming Experiments

To determine if the early hiPSC phenotype is specific to human reprogramming or a general iPSC phenomenon, a comparison of mouse iPSCs and ESCs was performed. Hierarchical clustering of mESCs and different miPSC cell lines that were obtained in a fibroblast reprogramming experiment with retrovirally delivered *Oct4*, *Sox2*, *Klf4*, and *c-Myc* was performed. This analysis demonstrated that, although highly similar, miPSCs and mESCs also differ in their expression (Figure 3A). As with the human reprogramming data, the sample tree of the hierarchical clustering revealed that mESCs and miPSCs cluster separately. Specifically, 1388 genes significantly differ in expression levels between miPSCs and their embryonic equivalents as determined by a Student's *t* test ($p < 0.05$) and have at least a 1.5-fold difference. Many of these genes are functionally involved in transcriptional regulation and organ development (Figure S9B), as observed with human iPSCs signature genes. To further assess the coregulation of genes in mouse and human iPSC reprogramming experiments, the subsequent analysis was limited to only those with identifiable homologs between mouse and human transcriptomes (HomoloGene database Release 63). Twenty-nine percent of the trimmed-down miPSC signature genes were also differentially expressed in our early hiPSCs ($p < 10^{-7}$), suggesting that the iPSC state is remarkably robust across species (Figure 3B; Table S6).

Similar to our observation with the human iPSC signature genes, the majority of miPSC signature genes appeared to be ESC-specific genes that were insufficiently induced and fibroblast-specific genes that were not repressed completely (Figure 3C). To determine whether differential regulation of target genes by the reprogramming factors themselves could drive the differential expression of genes between iPSCs and ESCs, we tested whether expression differences between miPSC and mESCs correlate with binding differences of the reprogramming

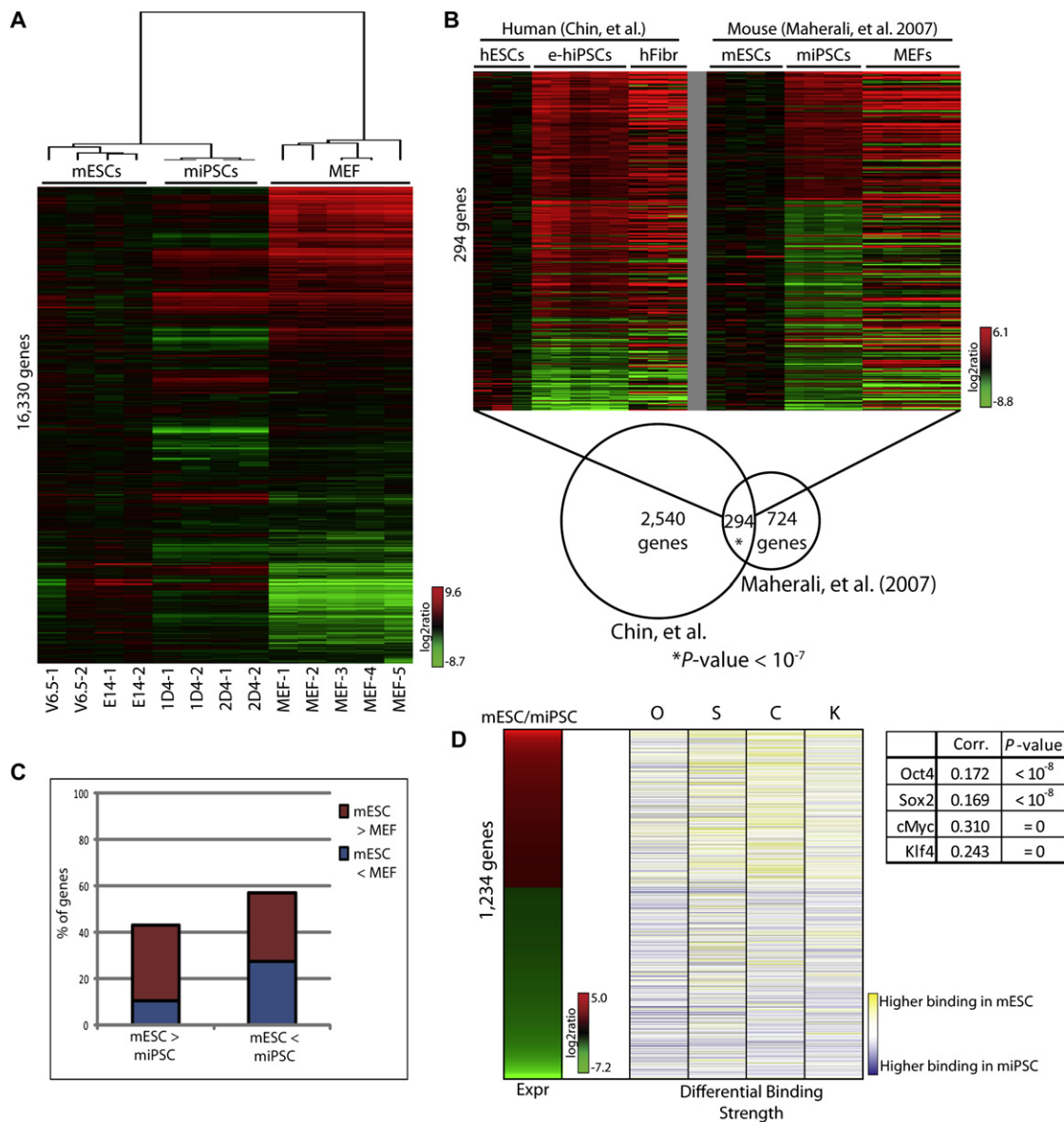


Figure 3. Reprogramming of Human and Mouse Fibroblasts Results in Conserved Expression Differences between ESC and iPSC

(A) Unsupervised hierarchical clustering of global gene expression data from mouse (m) ESCs, miPSCs, and mouse embryonic fibroblasts (MEFs) obtained from the Maherali et al. data set (Maherali et al., 2007). The lines included are biological replicates of V6.5 and E14 mESC lines and of 1D4 and 2DA iPSC lines, respectively. All \log_2 ratios are relative to averaged expression in mESCs.

(B) Comparison of iPSC signatures between human and mouse reprogramming experiments. Human early-passage iPSC signature genes as defined in Figure 1B of this study (Chin et al.) and mouse iPSC signature genes defined with the Student's *t* test ($p < 0.05$) and an at least 1.5-fold change between miPSCs and mESCs were further annotated to only include genes that have homologous partners across the two species according to the HomoloGene database, resulting in 2834 genes for the early hiPSC signature and 1018 genes for the miPSC signature. The Venn diagram shows the overlap between these two groups of genes, and significance was determined using the Fisher's exact test. The heatmap above displays the expression values for the 294 iPSC signature genes conserved between mouse and human reprogramming experiments across the different cell types and species, as \log_2 ratio relative to the average ESC expression for each species.

(C) Similar to Figure 1B', miPSC signature genes were divided into those expressed at higher levels in mESCs than miPSCs and vice versa. Each of these two groups was further subclassified into two groups, either more highly expressed in mESCs than fibroblasts (red color) or those more highly expressed in fibroblasts than hESCs (blue color).

(D) The heatmap depicts the \log_2 expression ratio between mESCs and miPSCs for miPSC signature genes ordered according to decreasing ratios. For these genes, the binding strength ($-\log_{10}P\text{-value}$) of each reprogramming factor (Oct4 [O], Klf4 [K], Sox2 [S], or c-Myc [C]; data obtained from Sridharan et al., 2009) in miPSCs was subtracted from the binding strength in mESCs. Higher binding strength in mESCs is represented in yellow and in miPSCs in blue. The intensity of the colors increases as differential binding increases. The Pearson correlation of the differential binding strength relative to differential gene expression is given in the attached table.

factors between the two cell types. This analysis took advantage of genome-wide location data of the target genes of c-Myc, Klf4, Sox2, and Oct4 proteins in the mESCs and miPSC lines that were used for the expression analysis described here (Sridharan et al., 2009). We previously reported that binding patterns of the reprogramming factors are highly similar in iPSCs and ESCs but that subtle differences exist, which we did not analyze further. Reanalysis of these minor differences in binding demonstrates that the promoter regions of those genes that are expressed at a higher level in mESCs than miPSCs are correlated with stronger binding by each of the reprogramming factors, particularly by c-Myc and Klf4 (Figure 3D). Conversely, the promoter regions of those genes that are expressed at a higher level in miPSCs are correlated with stronger binding by the reprogramming factors in miPSCs (Figure 3D).

To determine whether the iPSC signature is specific to reprogramming with fibroblasts as opposed to other cell types, this type of analysis was extended to iPSCs generated from mouse B cells by a different lab (Mikkelsen et al., 2008). As was shown with fibroblast-derived miPSCs, miPSC lines made from B cells also display a common group of genes differentially expressed compared to mESC lines (Figure S10). The high degree of overlap of B cell miPSC signatures with fibroblast miPSC lines suggests that iPSC gene expression signatures arise regardless of the cell type of origin (522 genes, $p < 10^{-43}$). Furthermore, a significant portion of the B cell miPSC signature genes are also found to be differentially expressed in our early-passage human iPSCs (729 genes, $p < 10^{-4}$). Taken together, early iPSCs possess a conserved gene expression signature that is shared regardless of the lab of origin, species, or cell type from which they were derived.

Global Reprogramming of Histone H3K27 Promoter Methylation in Late-Passage hiPSCs

Perhaps as intriguing as the finding that all early hiPSCs appear to share a common gene expression signature that sets them apart from hESCs is the fact that this signature disappears after extended culturing, albeit not completely. To further define at the molecular level how similar late-passage hiPSCs are to hESCs at similar passage, the state of histone H3 lysine 27 (K27) trimethylation was analyzed, since to date the genome-wide chromatin structure of hiPSCs has not been probed. This chromatin modification, established through Polycomb group proteins, is repressive in nature and plays an essential role in the regulation of the expression of many developmentally important genes (Cao and Zhang, 2004).

Genome-wide location analysis for histone H3K27 trimethylation in human fibroblast lines, two hESCs, and two hiPSC lines at late passage (p56, 71 for hiPSCs and p69, 64 for hESCs; Table S1) was performed using chromatin immunoprecipitation followed by hybridization to a human promoter array covering regions from -5.5 kb upstream to $+2.5$ kb downstream of the transcriptional start sites for about 17,000 genes. The overall pattern of H3K27 trimethylation at promoters was very similar among all the pluripotent stem cell lines tested and different from the fibroblasts from which the hiPSCs were derived (data not shown). When focusing on the promoter regions that are differentially methylated at H3K27 between hESCs and fibroblasts (see Experimental Procedures), hESCs and hiPSCs are

nearly identical in their methylation pattern (Figure 4A). Specifically, of the 978 genes that were identified as being different between hESCs and fibroblasts at high stringency ($p < 0.05$), 97% carried a methylation pattern virtually identical in hiPSCs and hESCs (ESC-like promoter regions in hiPSCs [E]). Pairwise correlation analysis verified this conclusion for this set of genes (Figure S11). Only 1% of the 978 genes were methylated in a more fibroblast-like pattern (F class promoter regions), and the remaining 2% of the loci were classified as neutral (N), as the differences were too small to be significant. The distribution remained highly similar when the stringency was lowered to include a larger set of genes and is highly significant, as confirmed by a random permutation test (Figure S12). Genes that were not differentially methylated between hESC and fibroblasts showed little or no difference in methylation pattern in hiPSCs, indicating that the hiPSCs had not acquired a completely novel epigenetic identity. As expected, there was a nearly perfect inverse correlation between H3K27 trimethylation of promoters and expression of these genes in hESCs, hiPSCs, and fibroblasts (Figure 4A).

Only 40 genes of the 860 late-passage hiPSC signature genes and 21 genes of the enduring iPSC signature genes were found to be differentially methylated in their promoter regions at H3K27 between fibroblasts and ESCs. However, their methylation pattern in iPSCs is reset to the ESC state (Figures 4B and S13). These data suggest that the aberrant expression of genes in late-passage hiPSCs compared to hESCs is not the result of differential H3K27 methylation between hESCs and hiPSCs. Interestingly, early-passage miPSC are also already completely reset in their histone H3K27 methylation patterns to the ESC state as determined previously (Maherali et al., 2007). Together, these results indicate that the H3K27 methylation state of the fibroblast genome is reset almost completely to an ESC state in iPSCs, suggesting that the early and late hiPSC gene expression signatures probably do not arise as a result of faulty resetting of H3K27 trimethylation during reprogramming, even though subtle differences in methylation patterns could still exist. In agreement with the conclusion that histone H3K27 trimethylation is not a histone mark that is aberrantly reset upon reprogramming, we found that the promoter regions of early, late, or common hiPSC signature genes undergo the same changes in H3K27 trimethylation between hESC and fibroblasts as genes that are equally expressed between hESC and hiPSC (Figure 4C). A similar observation is true for H3K4 trimethylation (Figure 4D). While there is no global correlation between these H3 modifications and hESC/hiPSC expression differences, the promoter regions of an established set of hESC-specific genes showed a much different pattern of histone methylation in hESCs relative to fibroblasts for both the repressive and active histone marks (Figures 4C, 4D, and S14), in agreement with previously published findings (Maherali et al., 2007; Sridharan et al., 2009).

miRNA Expression Signature of the hiPSC State

It has been clearly shown that various types of cells differ not only in the expression of their coding genes, but also in their noncoding genes. To determine whether miRNAs are expressed at a hESC-like level in hiPSCs, expression profiling of all known miRNAs was performed on hESCs, late-passage hiPSCs, and the fibroblasts from which they were derived (Table S1).

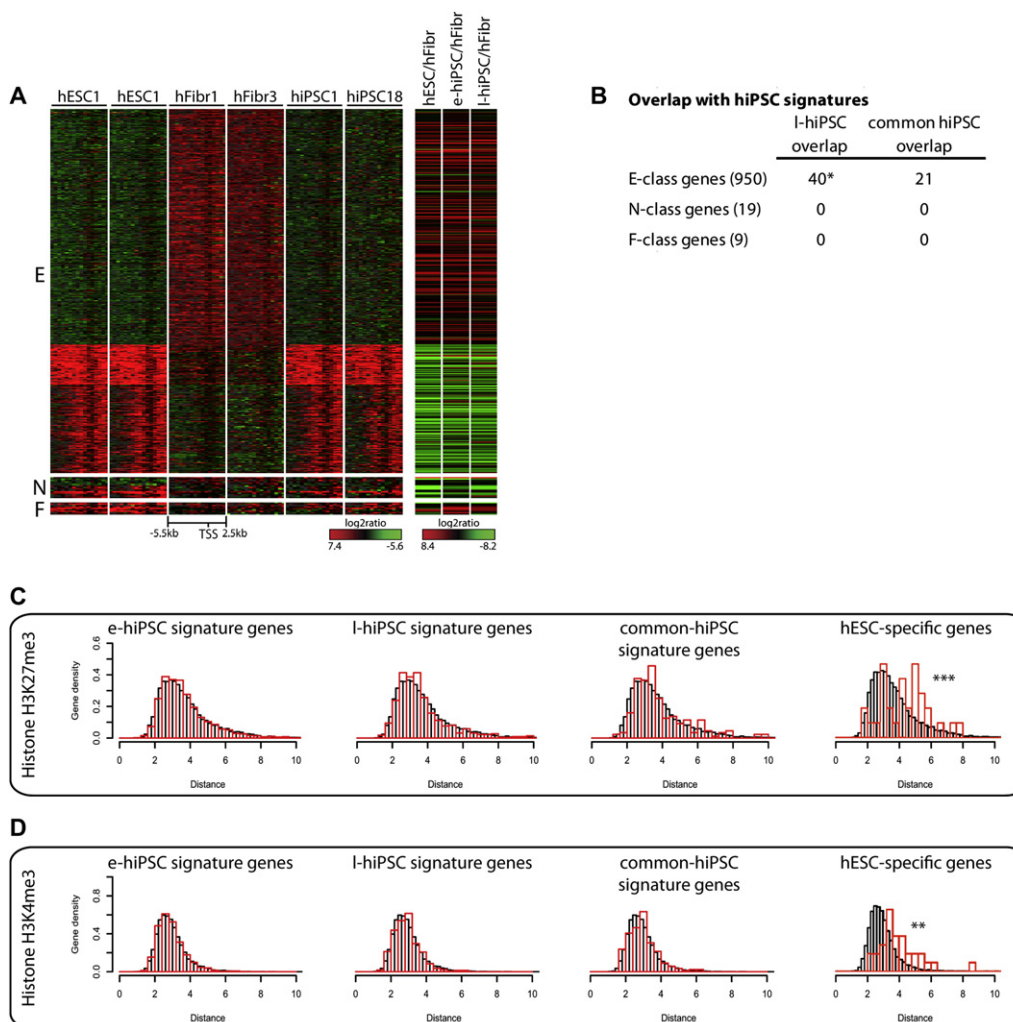


Figure 4. Histone H3 Methylation Analysis in hiPSC

(A) Tree-view representation of the hierarchical clustering of histone H3 K27 trimethylation in HSF1 (hESC1), HSF6 (hESC2), two human fibroblast lines (hFibr1 and -2), and late-passage hiPSCs (l-hiPSC1 and l-hiPSC18) across the promoter regions of all genes considered significantly differentially methylated between fibroblasts and hESCs ($p = 0.05$). Genes were classified as E (hESC-like, 950 genes), N (neutral, 19 genes), or F (fibroblast-like, 9 genes) based on the similarity of the methylation patterns in l-hiPSCs with hESCs or fibroblasts. For N and F class genes, the y axis is scaled 3× to make the methylation patterns visible. Each row represents the −5.5 kb to +2.5 kb region with respect to the transcription start site (TSS). The 8 kb promoter regions are divided into sixteen 500 bp regions displayed in pseudocolor based on the average log ratio of the IP to input probe signal intensity. Probes within a given 500 bp region are averaged. Dark gray coloring indicates missing values for enrichment due to the lack of probes. Expression levels for the represented genes are shown on the right for hESCs, e-hiPSCs, and l-hiPSCs relative to fibroblasts.

(B) Table showing the overlap between genes shown in (A) (which demonstrate differences in H3K27 methylation between hESCs and fibroblasts), grouped according to the methylation pattern in l-hiPSCs, and the late and common hiPSC expression signature genes as defined in Figures 1C and 1D (these genes are differentially expressed between l-hiPSCs and hESCs). * p value = 0.0063.

(C) Histograms detailing the differences of H3K27me3 patterns between hESC and fibroblasts, measured as the Euclidean distance across the sixteen 500 bp regions of the promoters for the indicated gene sets. The black outlined bars denote the distribution of Euclidean distances for all genes on the array (17,000), while the red outlined bars show the distribution for the indicated subset of signature genes.

(D) As in (C) but for histone H3K4 trimethylation. Note: hESC signature genes must undergo a significantly larger degree of change in both H3K4me3 (** p value = 0.005) and H3K27me3 (** $p < 10^{-7}$) than the global population of genes. None of the distributions for the other subsets of genes are significantly different from the global population.

Hierarchical clustering with the 105 miRNAs expressed in at least one cell type shows that there is little difference in miRNA expression among the pluripotent cells tested with hiPSCs and hESCs intermixed in the tree of the clustering. Conversely, all of the pluripotent cell lines have a vastly different miRNA profile than fibroblasts. Nevertheless, a few miRNAs were consistently

expressed differently between late hiPSCs and hESCs (Figure 5B). This finding was similar to data recently obtained by another group that also profiled the miRNA expression profile of different lines of a different set of hESCs and hiPSCs (highlighted miRNAs in Figure 5B [Wilson et al., 2009]), suggesting that a distinct miRNA pattern is highly reproducible between

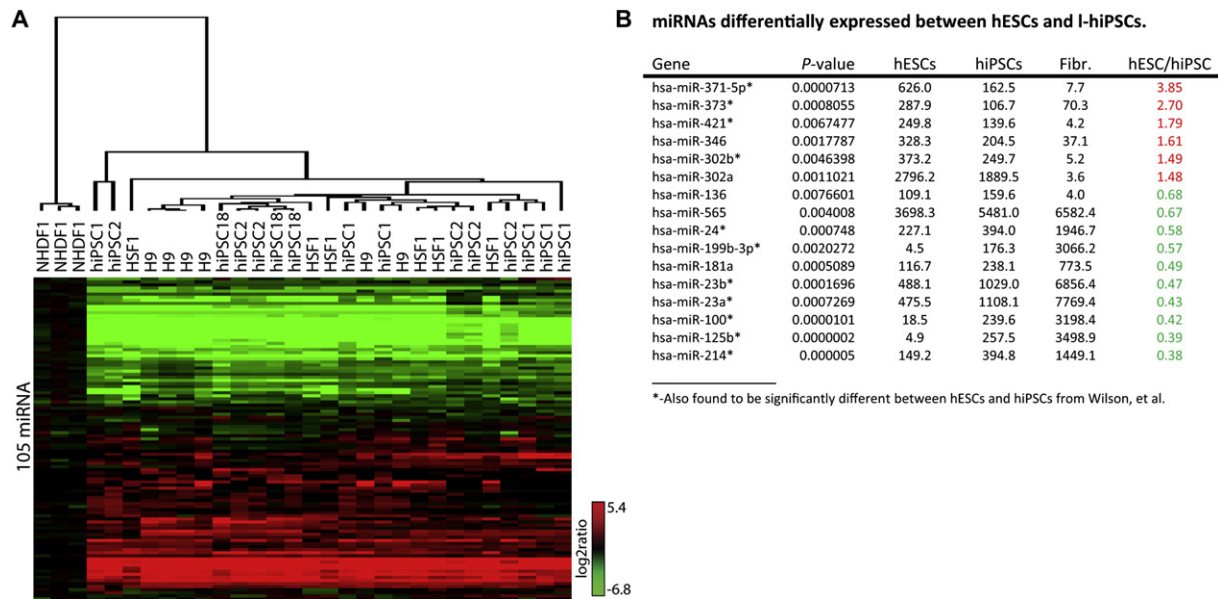


Figure 5. MicroRNA Expression Analysis in hESC, I-hiPSC, and Fibroblast Lines

(A) Hierarchical clustering of the expression data for the 105 microRNAs substantially expressed in either hESCs (H9 and HSF1), I-hiPSCs (lines 1, 2, and 18), or fibroblasts, given as log₂ ratios relative to expression in fibroblasts. Note: miRNA expression does not cluster the hESC lines distinctly from the hiPSC lines, whereas the fibroblasts, as expected, exhibit a different miRNA expression pattern.

(B) Table showing miRNAs that were differentially expressed between the two hESC lines (12 replicates total) and three I-hiPSC lines (15 replicates total). Expression values for the given cell lines are given as well as the p value derived from the Student's t test. Asterisk indicates that miRNA was also found to be differentially expressed between hESC and hiPSC lines in Wilson et al. (2009).

different reprogramming experiments, and that hiPSCs have a miRNA signature that defines them as unique from hESCs.

An Analysis of the Genomic Stability of hiPSCs

A priori, the cause of the differential expression of genes between hiPSC and hESC could be that the reprogramming protocol itself requires or leads to genomic alterations. It has been suggested that, because reprogramming efficiency is low and because exogenous expression mediated by retrovirus requires genomic integration, reprogramming perhaps is accompanied by genomic alterations. With the advent of integration-free reprogramming, many of these concerns are probably not valid (Kaji et al., 2009; Soldner et al., 2009; Stadtfeld et al., 2008; Woltjen et al., 2009; Yu et al., 2009). Regardless, the genomic stability of both miPSCs and hiPSCs had not been examined after extended passaging by any technique more sensitive than karyotyping. Many groups, including ours, showed that reprogrammed lines usually have a normal karyotype (Lowry et al., 2008), but it has remained formally possible that subkaryotypic alterations accompany reprogramming. It is also possible that hiPSCs could have an unstable genome, prone to alteration due to some unknown byproduct of the reprogramming process. To date, no one has yet profiled iPSCs from any species to resolve these issues, which could prove critical in the application of iPSC technology to regenerative medicine.

To determine systematically whether our hiPSC lines contain genomic alterations that could possibly explain the differences in gene expression between hESCs and hiPSCs, array comparative genomic hybridization (aCGH) was performed on three hiPSC lines and the fibroblasts from which they were derived.

Using Human CGH Tiling Arrays (NimbleGen, Roche), a few subkaryotypic alterations were detected in each late-passage hiPSC line relative to the starting fibroblast line (Table 1; Figure S15). As confirmation of the validity of the approach, the duplication of part of chromosome 8 in the hiPSC line 1 identified by array CGH had already been discovered by karyotyping at p44 (Figure 6A). hiPSC line 1 must have acquired this duplication of part of chromosome 8 upon extended passaging, as it was not detected at p9 (Lowry et al., 2008).

Interestingly, none of the genomic alterations detected by aCGH appeared to be shared among all three hiPSC lines (Figure 6B; Table 1), leading to two conclusions: (1) no particular genomic alteration is required for reprogramming; (2) these genomic alterations cannot directly explain the early hiPSC signature because the signature strictly represents changes found in all three lines. Genes harbored in genomic regions that are altered in hiPSCs are significantly enriched for lipocalins and serine proteases (in hiPSC 18), tumor antigens (hiPSC 2), and lectins, keratins, and sensory transduction (hiPSC 1), with none of these functional classifications being conserved between two different hiPSC lines. Regardless, these analyses suggest that the genome of reprogrammed cells is both normal and highly stable even after at least 44 passages.

DISCUSSION

While there is still much to learn about the molecular details of the iPSC state, our data indicate that early- and late-passage hiPSCs are not identical to their embryo-derived counterparts. Many groups have generated iPSCs from both human and

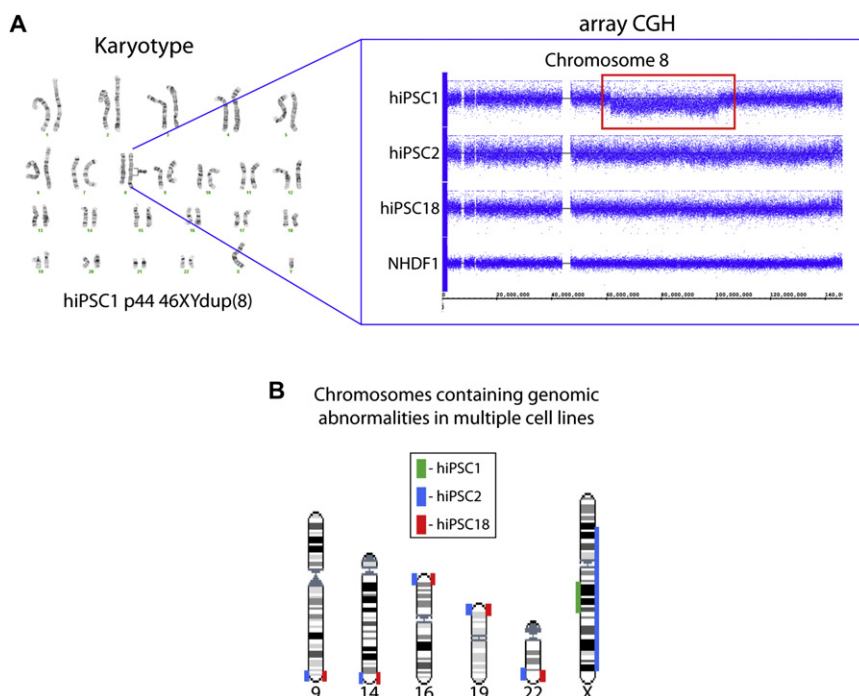
Table 1. Regions in hiPSCs with Genomic Abnormalities

Chromosome	Cell Line Identified	Region	Size	Z-Score	Mean (log2ratio) Fibroblast/ hiPSC	No. of Genes	No. of miRNA
7p11.2	hiPSC2	56808738-62112547	5.3 Mb	19.37	0.213	1	0
8q12.3	hiPSC1	62203767-100955803	83.8 Mb	45.26	-0.185	148	3
9q34.3	hiPSC2	137714842-139627239	1.9 Mb	21.60	0.207	75	1
	hiPSC18	135323818-139634573	4.3 Mb	21.38	0.138	102	1
10q21.3	hiPSC2	67642045-67930512	0.3 Mb	19.24	0.473	1	0
11p11.12	hiPSC2	48381190-55253892	6.9 Mb	19.24	0.149	16	0
12p13.31	hiPSC1	62119-11760562	11.7 Mb	45.32	-0.175	176	2
14q32.33	hiPSC2	103616693-106182502	2.6 Mb	19.71	0.163	27	1
	hiPSC18	103353890-106222351	2.9 Mb	18.12	0.141	30	1
16p13.3	hiPSC2	7675-2208335	2.2 Mb	19.30	0.172	104	2
	hiPSC18	53-2234072	2.2 Mb	20.49	0.181	107	2
19p13.3	hiPSC2	208435-2449396	2.2 Mb	24.08	0.213	86	2
	hiPSC18	40285-3774141	3.7 Mb	24.35	0.170	121	2
20q13.33	hiPSC18	59860004-62429688	2.6 Mb	18.50	0.154	81	7
21q22.3	hiPSC1	38084180-46913738	8.8 Mb	73.43	0.320	144	0
22q13.32	hiPSC2	46997532-49534378	2.5 Mb	21.66	0.177	37	0
	hiPSC18	40986242-49567312	8.6 Mb	22.54	0.100	103	3
Xq21.1	hiPSC1	77883731-90950867	13.1 Mb	20.61	0.080	30	2
	hiPSC2	25192507-151731387	126.5 Mb	31.42	0.122	714	69
Yp11.2-q11.21	hiPSC2	2858143-20250487	17.4 Mb	28.19	0.120	27	0

mouse somatic cells, and each group suggested that their iPSCs were “nearly” identical to the ESCs they used for comparison. Until now, it was not clear if the small differences observed in gene expression between iPSCs and ESCs were due to stochastic differences in each experiment, or whether all reprogrammed cells share a signature that distinguishes them from ESCs. Reanalyzing hiPSCs and miPSCs suggests that in fact

all iPSCs share a gene expression signature that defines the iPSC state as unique from that of ESCs.

The gene expression signature observed in early-passage hiPSCs seems to be partially corrected upon extended culturing in vitro, suggesting that perhaps some form of “reprogramming” continues in culture. This could be due to feed-forward or feed-back loops of gene regulation under the direction of the

**Figure 6. Genomic Abnormalities Are Not Conserved among I-hiPSC Lines**

(A) Karyotype analysis of the only genomic abnormality detected in any of our hiPSC lines. A clonal duplication on chromosome 8 was found in 19 of 20 metaphase spreads in hiPSC line 1 at passage 44. Array CGH also identified this region (Z-score = 45) solely in the hiPSC1 line (see Table 1), and the array CGH data for this region in all hiPSC lines and fibroblasts are given. The duplicated region in hiPSC1 “steps” down from the midline, indicated by the red box.

(B) A cartoon schematic depicting all overlapping genomic abnormalities among hiPSC lines that were determined by array CGH. Alterations in I-hiPSC1 are indicated with green bars, in I-hiPSC2 with blue bars, and in I-hiPSC18 with red bars. These six regions were the only ones found to be shared between any two hiPSC lines. No genomic aberrations were found in all three I-hiPSC lines.

endogenously expressed pluripotency genes (Jaenisch and Young, 2008). Moreover, since low-passage hESCs did not appear to share the early-passage hiPSC signature, it seems as though this extended reprogramming phase is not simply due to the time a pluripotent cell spent in culture, but something more specific to iPSCs. While late-passage hiPSCs appeared to be much more similar to their embryo-derived counterparts with regard to most of the transcriptome (including coding and microRNA), there is a group of genes and miRNAs that are differentially expressed compared to hESCs. For the most part, these differences reflect either an insufficient induction of ESC genes or insufficient suppression of fibroblast genes. Together, these findings suggest that the reprogramming process does not drive fibroblasts to a state identical to ESCs.

Reprogramming Is Not Perfect

It is not surprising that iPSCs are not perfectly identical to ESCs considering the vastly different set of circumstances by which they were generated. ESCs are derived from the inner cell mass of an embryo and are thought to undergo significant changes as they adapt to *in vitro* culture. However, mESCs can be placed back into a blastocyst and contribute to the resulting offspring even at 100%, suggesting that the changes induced by *in vitro* culture either are not fate changing or are reversible. Of course, it is far more difficult to compare hESCs to the cells of the inner cell mass from which they were derived in order to understand their origins, for technical and ethical reasons. Regardless, it is clear that iPSCs arise by a markedly different mechanism. iPSCs start out as fully determined somatic cells. These somatic cells possess nuclei that are almost completely refractory to reprogramming, as demonstrated by the low efficiency of cloning by somatic cell nuclear transfer (Gurdon and Melton, 2008; Markoulaki et al., 2008) or of reprogramming with the four Yamanaka factors (Takahashi and Yamanaka, 2006). Therefore, a drastic molecular change is presumably essential to reset the somatic nucleus to an embryonic/pluripotent state.

A great deal of effort is underway to understand the role each of the reprogramming factors plays during the process, beginning with documentation of the complete set of target genes at different stages (Sridharan et al., 2009). Considering all the changes to the transcriptome, epigenome, metabolome, and proteome that are likely required for reprogramming, it should come as no surprise that reprogramming somatic cells with four transcription factors does not perfectly recapitulate the state of ESCs. The data presented here describe the deficits of reprogramming with regards to just portions of the transcriptome and epigenome. It is likely that there are a number of other fundamental molecular characteristics that distinguish iPSCs from ESCs. Even though not tested extensively, one of the functional manifestations of these differences could be that miPSCs have not yet been shown to support the generation of adult mice that are completely derived from these cells.

Do Errors in Epigenetic Reprogramming Generate the iPSC State?

We next considered whether the iPSC state arises because of defective resetting and/or re-establishment of the epigenome that is thought to occur during reprogramming (Maherali et al., 2007; Takahashi et al., 2007; Wernig et al., 2007). Clearly, fibro-

blast and ESC epigenomes are maintained in very different states, ostensibly to help control gene expression, differentiation potential, self-renewal, etc. There are data to suggest that when fibroblasts are reprogrammed, the histone code is dramatically altered, whereby modifications that are known to correlate with gene silencing are removed from pluripotency genes and replaced by those that mark active genes and vice versa (Maherali et al., 2007). Here, we examined which promoters were associated with a histone mark that is well established to be linked to gene silencing in fibroblasts, hESCs, and hiPSCs. Overall, hiPSCs and hESCs had a very similar pattern of H3K27 trimethylation of promoter regions, and this pattern was strikingly different from fibroblasts. The promoters of the late hiPSC signature genes appeared to have a H3K27 trimethylation pattern similar to that found in hESCs. These data suggested that the late hiPSC signature does not arise as a result of aberrant resetting of these histone methylation marks. Of course, there are a multitude of various types and combinations of histone modifications, many of which are known to be associated with active or silenced genes, so any of these others might yet explain the presence of the late hiPSC expression signature. Recently, Gurdon and colleagues suggested, for example, that the histone variant H3.3 is a carrier of an epigenetic memory in frog cloning experiments (Ng and Gurdon, 2008).

Are hESCs and hiPSCs More Similar in Their Noncoding RNA Expression?

Recent data suggest that most cell types express a unique pattern of noncoding RNAs such as miRNAs (Laurent et al., 2008). miRNAs are known to suppress expression of their homologous target RNAs through the association with the RISC complex (RNA-induced silencing complex) (Tang, 2005). miRNA expression profiles are known to change as tissues develop and individual cells differentiate (Krutfeldt et al., 2006; Yi et al., 2006, 2008, 2009). Profiling the expression of miRNAs in undifferentiated hESCs, hiPSCs, and fibroblasts demonstrated a vast difference in expression of at least 100 miRNAs between these two pluripotent populations and fibroblasts. A handful of miRNAs are significantly different in expression between hESCs and hiPSCs. Most of these miRNAs were also described as differentially expressed between hESCs and hiPSCs in an independent experiment with independently derived hESCs and hiPSCs (Wilson et al., 2009). Importantly, hiPSCs in this study were derived by overexpression of the Thomson set of reprogramming factors replacing c-MYC and KLF4 with NANOG and LIN28. Since each miRNA is known to have multiple targets, it is formally possible that even the 10 to 12 miRNAs shown to be differentially expressed between hESCs and hiPSCs could explain the occurrence of the late hiPSC signature. However, because *in silico* miRNA target prediction has not been perfected, future efforts will be required to uncover the contribution of differential expression of these miRNAs to the late hiPSC signature. In any case, it is interesting that some of the miRNAs that are differentially expressed between hiPSC and hESCs include a group of ESC-specific miRNAs (Card et al., 2008). Furthermore, the miR-302 and miR-371/372/373 clusters encode the human homologs of the mouse 290–295 cluster, which are indicated as enhancers of the reprogramming process (Judson et al., 2009). Clearly, further study will be required to elucidate the

role of these and other noncoding RNAs in the reprogramming process and the maintenance of the iPSC state.

Consequences of the iPSC State

The results described here suggest that hiPSCs represent a unique type of pluripotent cell as defined by gene expression. What are the physiological consequences of the variance? To date, no one has described significant functional differences between hiPSCs and hESCs. Many groups have shown that hiPSCs are pluripotent by embryoid body and teratoma formation assays (Lowry et al., 2008; Park et al., 2008; Takahashi et al., 2007; Yu et al., 2007). Of course, several gold-standard assays of pluripotency used for mouse pluripotent cells cannot be performed with human equivalents (chimerism, germline transmission), so it is not possible to judge the relative pluripotency of hiPSCs and hESCs. Some groups have described small differences between hiPSCs and hESCs in their relative abilities to undergo directed differentiation (Choi et al., 2009; Karumbayaram et al., 2009; Zhang et al., 2009). However, because of the inherent biases among pluripotent cell lines to adopt particular fates, it is unclear whether there are any general differences between hiPSCs and hESCs in this regard (Osafune et al., 2008). Additionally, there are no published data to suggest that hiPSCs and hESCs function differently in the undifferentiated state. The molecular differences between iPSCs and ESCs described here should drive intense effort in the future aimed at uncovering any possible physiological consequences.

EXPERIMENTAL PROCEDURES

Tissue Culture

Cells were cultured as described in Lowry et al. (2008).

Gene Expression Analysis

Gene expression profiling was performed as described (Lowry et al., 2008). All human expression data from this experiment and those conducted in other labs were obtained with the HG-U133plus2 microarray platform (Affymetrix). Mouse expression data were extracted from Maherali et al. (2007) and Mikkelsen et al. (2008), both using the Mouse Expression Array 430 platform (Affymetrix). For analyses, the array data for fibroblasts, ESCs, and iPSCs were normalized independently for each experiment using Robust Multichip Analysis (RMA) in R (Bioconductor). Expression data for each gene were obtained from respective probe sets utilizing a hierarchical averaging algorithm. Specifically, exponent expression values were averaged for individual RefSeq identifiers based on the specificity of the probes assigned to each RefSeq. If multiple “_at” probes existed for RefSeq gene X, then those probes were averaged. If no specific probes existed for that RefSeq, then the next level “_a_at” probes were used. This filtering continued until the highest-confidence probes were chosen to represent each RefSeq, thereby ensuring that analysis was specific to each gene. The resulting human and mouse data sets contain 17,620 and 16,330 genes, respectively. 11,975 homologous genes were separated for direct comparison between the human and mouse data sets as curated by the Homologene database. All cell line correlations were a measure of Pearson's rho implemented in R. Significance of overlap between any two data sets was measured using Fisher's exact test. Significance of the overlap of the three human data sets was measured using simulation with replacement. Global array clustering was performed using Cluster 3.0 and presented using Java Treeview 1.1.1 with gene expression values presented as a log₂ ratio compared to averaged ESC expression. Class prediction was conducted using Student's t test combined with a requirement for a 1.5-fold change between the average of the cell lines being compared. Boxplots were created in R, and differences observed were assigned significance values using the Wilcoxon rank-sum test.

miRNA Expression Analysis

miRNA expression analysis was conducted as described (Zhang et al., 2008) using the Ohio State University Comprehensive Cancer Center (OSUCCC) miRNA Expression Bioarrays.

Histone Methylation Analysis

Genome-wide chromatin analysis was performed as described (Maherali et al., 2007).

aCGH Methods

Genomic DNA from cell lines with indicated passages was collected and purified using QIAGEN DNA kit (QIAGEN, Germany). Hybridization was conducted with Human CGH 2.1M Whole-Genome-Tiling v2.0D Array (NimbleGen), with a resolution of 5 kb over the entire human genome. Hybridization and raw data collection were performed as described in Selzer et al. (2005).

CGH Analysis Methods

Raw signal intensities for Cy3 and Cy5 were extracted from each array. Intensity values were averaged for the three replicate probes. Log ratios of the average values were generated for each of the two dyes. Each array was normalized by subtracting from each individual probe the mean log ratio values over all probes in the array. Regions were computed along the chromosome that had elevated average values, possibly representing copy number variation (CNV). All possible windows were computed within the chromosome, and for each window computed a Z-score. Because of computational limitations, each chromosome was segmented into pieces corresponding to 3000 probes, leading to a potential overestimation of the number of CNV regions if these span the boundaries across two chunks, since they would then be considered two separate CNVs. Based on random permutations of the array probes, we established that a Z-score of 18 for a region containing more than five probes provides a false-positive rate of less than 1%. The code was implemented in Matlab.

ACCESSION NUMBERS

Microarray and ChIP-chip array data are available at the NCBI Gene Expression Omnibus database under the accession numbers GSE12390, 7815, 14012, 9865, 14711, 15176, and 16654.

SUPPLEMENTAL DATA

Supplemental Data include six tables and 15 figures and can be found with this article online at [http://www.cell.com/cell-stem-cell/supplemental/S1934-5909\(09\)00292-6](http://www.cell.com/cell-stem-cell/supplemental/S1934-5909(09)00292-6).

ACKNOWLEDGMENTS

M.H.C. is supported by the USHHS Ruth L. Kirschstein Institutional (NRSA #T32 CA009056) and G.A. by NIH/NICHHD 5 K12 HD001281. S.V. is supported by Regione Emilia Romagna PRRITT Biopharmnet. C.M.C. is supported by NIH-NCI. M.A.T. is supported by NIH and CIRM Grant RS1-00313. M.G. is supported by NIH GM23674. N.B. is supported by the Legacy Heritage Fund. K.P. is supported by the V and Kimmel Scholar Foundations, the NIH Director's Young Innovator Award (DP2 OD001686-01), and a CIRM Young Investigator Award (RN1-00564-1). W.E.L. holds the Maria Rowena Ross Term Chair in Cell Biology and Biochemistry and is supported by CIRM #RS1-00259-1, and the Basil O'Connor Starter Scholar Award from The March of Dimes. This work was also supported by the CIRM New Cell Line grant to Jerome Zack (UCLA), W.E.L., and K.P. (RL1-00681-1). A.T.C., A.D.P., K.P., and M.A.T. are also supported by NIH P01 GM081621-01A1.

Received: March 28, 2009

Revised: May 30, 2009

Accepted: June 12, 2009

Published: July 1, 2009

REFERENCES

- Cao, R., and Zhang, Y. (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr. Opin. Genet. Dev.* 14, 155–164.
- Card, D.A., Hebbard, P.B., Li, L., Trotter, K.W., Komatsu, Y., Mishina, Y., and Archer, T.K. (2008). Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells. *Mol. Cell. Biol.* 28, 6426–6438.
- Choi, K.D., Yu, J., Smuga-Otto, K., Salvagiotto, G., Rehrauer, W., Vodyanik, M., Thomson, J., and Slukvin, I. (2009). Hematopoietic and endothelial differentiation of human induced pluripotent stem cells. *Stem Cells* 27, 559–567.
- Gokhale, P.J., and Andrews, P.W. (2006). A prospective on stem cell research. *Semin. Reprod. Med.* 24, 289–297.
- Gurdon, J.B., and Melton, D.A. (2008). Nuclear reprogramming in cells. *Science* 322, 1811–1815.
- Hochedlinger, K., and Plath, K. (2009). Epigenetic reprogramming and induced pluripotency. *Development* 136, 509–523.
- Jaenisch, R., and Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* 132, 567–582.
- Judson, R.L., Babiarz, J.E., Venere, M., and Blelloch, R. (2009). Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nat. Biotechnol.* 27, 459–461.
- Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P., and Woltjen, K. (2009). Virus-free induction of pluripotency and subsequent excision of reprogramming factors. *Nature* 458, 771–775.
- Karumbayaram, S., Novitskiy, B.G., Patterson, M., Umbach, J.A., Richter, L., Lindgren, A., Conway, A., Clark, A., Goldman, S.A., Plath, K., et al. (2009). Directed differentiation of human induced-pluripotent stem cells generates active motor neurons. *Stem Cells* 27, 806–811.
- Keller, G. (2005). Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev.* 19, 1129–1155.
- Krutzfeldt, J., Poy, M.N., and Stoffel, M. (2006). Strategies to determine the biological function of microRNAs. *Nat. Genet.* 38 (Suppl.), S14–S19.
- Laurent, L.C., Chen, J., Ulitsky, I., Mueller, F.J., Lu, C., Shamir, R., Fan, J.B., and Loring, J.F. (2008). Comprehensive microRNA profiling reveals a unique human embryonic stem cell signature dominated by a single seed sequence. *Stem Cells* 26, 1506–1516.
- Lowry, W.E., Richter, L., Yachechko, R., Pyle, A.D., Tchieu, J., Sridharan, R., Clark, A.T., and Plath, K. (2008). Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc. Natl. Acad. Sci. USA* 105, 2883–2888.
- Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., et al. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* 1, 55–70.
- Maherali, N., Ahfeldt, T., Rigamonti, A., Utikal, J., Cowan, C., and Hochedlinger, K. (2008). A high-efficiency system for the generation and study of human induced pluripotent stem cells. *Cell Stem Cell* 3, 340–345.
- Markoulaki, S., Meissner, A., and Jaenisch, R. (2008). Somatic cell nuclear transfer and derivation of embryonic stem cells in the mouse. *Methods* 45, 101–114.
- Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. *Nature* 454, 49–55.
- Ng, R.K., and Gurdon, J.B. (2008). Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. *Nat. Cell Biol.* 10, 102–109.
- Nishikawa, S., Goldstein, R.A., and Nierras, C.R. (2008). The promise of human induced pluripotent stem cells for research and therapy. *Nat. Rev. Mol. Cell Biol.* 9, 725–729.
- Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. *Nature* 448, 313–317.
- Osafune, K., Caron, L., Borowiak, M., Martinez, R.J., Fitz-Gerald, C.S., Sato, Y., Cowan, C.A., Chien, K.R., and Melton, D.A. (2008). Marked differences in differentiation propensity among human embryonic stem cell lines. *Nat. Biotechnol.* 26, 313–315.
- Park, I.H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* 451, 141–146.
- Park, T.S., Galic, Z., Conway, A.E., Lindgren, A., Van Handel, B.J., Magnusson, M., Richter, L., Teitell, M.A., Mikkola, H., Lowry, W.E., et al. (2009). Derivation of primordial germ cells from human embryonic and induced pluripotent stem cells is significantly improved by coculture with human fetal gonadal cells. *Stem Cells* 27, 783–795.
- Selzer, R.R., Richmond, T.A., Pofahl, N.J., Green, R.D., Eis, P.S., Nair, P., Brothman, A.R., and Stallings, R.L. (2005). Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosomes Cancer* 44, 305–319.
- Soldner, F., Hockemeyer, D., Beard, C., Gao, Q., Bell, G.W., Cook, E.G., Haragus, G., Blak, A., Cooper, O., Mitalipova, M., et al. (2009). Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell* 136, 964–977.
- Sridharan, R., Tchieu, J., Mason, M.J., Yachechko, R., Kuoy, E., Horvath, S., Zhou, Q., and Plath, K. (2009). Role of the murine reprogramming factors in the induction of pluripotency. *Cell* 136, 364–377.
- Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G., and Hochedlinger, K. (2008). Induced pluripotent stem cells generated without viral integration. *Science* 322, 945–949.
- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676.
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131, 861–872.
- Tang, G. (2005). siRNA and miRNA: an insight into RISCs. *Trends Biochem. Sci.* 30, 106–114.
- Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* 448, 318–324.
- Wilson, K.D., Venkatasubrahmanyam, S., Jia, F., Sun, N., Butte, A.J., and Wu, J.C. (2009). MicroRNA profiling of human-induced pluripotent stem cells. *Stem Cells Dev.* 18, 749–758.
- Woltjen, K., Michael, I.P., Mohseni, P., Desai, R., Mileikovsky, M., Hamalainen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M., et al. (2009). piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* 458, 766–770.
- Yi, R., O'Carroll, D., Pasolli, H.A., Zhang, Z., Dietrich, F.S., Tarakhovskiy, A., and Fuchs, E. (2006). Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. *Nat. Genet.* 38, 356–362.
- Yi, R., Poy, M.N., Stoffel, M., and Fuchs, E. (2008). A skin microRNA promotes differentiation by repressing 'stemness'. *Nature* 452, 225–229.
- Yi, R., Pasolli, H.A., Landthaler, M., Hafner, M., Ojo, T., Sheridan, R., Sander, C., O'Carroll, D., Stoffel, M., Tuschl, T., et al. (2009). DGCR8-dependent microRNA biogenesis is essential for skin development. *Proc. Natl. Acad. Sci. USA* 106, 498–502.
- Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* 318, 1917–1920.
- Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, I., and Thomson, J. (2009). Human induced pluripotent stem cells free of vector and transgene sequences. *Science* 324, 797–801.
- Zhang, L., Volinia, S., Bonome, T., Calin, G.A., Greshock, J., Yang, N., Liu, C.G., Giannakakis, A., Alexiou, P., Hasegawa, K., et al. (2008). Genomic and epigenetic alterations deregulate microRNA expression in human epithelial ovarian cancer. *Proc. Natl. Acad. Sci. USA* 105, 7004–7009.
- Zhang, J., Wilson, G.F., Soerens, A.G., Koonce, C.H., Yu, J., Palecek, S.P., Thomson, J.A., and Kamp, T.J. (2009). Functional cardiomyocytes derived from human induced pluripotent stem cells. *Circ. Res.* 104, e30–e41.