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 Aimiuwu,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. Lowry2,5,6,7,*

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
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 (Hagedooperling 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 over-expression 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 embryodermal counterparts, but it remains unclear whether the small percentage of genes that are differentially expressed between iPSCs and ESCs has not been completely elucidated. Every study suggests that iPSCs are “nearly identical” to their embryodermal counterparts, but it remains unclear whether the small percentage of genes that are differentially expressed between iPSCs and ESCs is shared among different iPS 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 log2 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.

(B) GO analysis

<table>
<thead>
<tr>
<th>Category</th>
<th>e-hiPSC &gt; hESC (3,119 genes)</th>
<th>e-hiPSC &lt; hESC (28 genes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mitochondrion</td>
<td>25.28</td>
<td>13.85</td>
</tr>
<tr>
<td>Transcription</td>
<td>15.3</td>
<td>15.95</td>
</tr>
<tr>
<td>RNA processing</td>
<td>11.85</td>
<td>7.1</td>
</tr>
<tr>
<td>DNA repair</td>
<td>11.02</td>
<td>4.76</td>
</tr>
<tr>
<td>Mitosis</td>
<td>6.03</td>
<td>3.45</td>
</tr>
<tr>
<td>Wnt/β-catenin repeat</td>
<td>7.37</td>
<td></td>
</tr>
<tr>
<td>Methyltransferase</td>
<td>6.06</td>
<td></td>
</tr>
</tbody>
</table>

*P-value = 0

(B') Graphical representation of GO analysis

(C) Additional heatmap

<table>
<thead>
<tr>
<th>Category</th>
<th>e-hiPSC &gt; hESC (84 genes)</th>
<th>e-hiPSC &lt; hESC (7.6)</th>
</tr>
</thead>
</table>

*P-value < 10^-9

(C') Graphical representation of additional heatmap

(D) Venn diagram

3,629 genes

318 genes

hESC vs. e-hiPSC

318 genes

542 genes

hESC vs. l-hiPSC

*P-value < 10^-9

(D') Graphical representation of Venn diagram

P-value < 10^-10

(D'') Graphical representation of additional Venn diagram
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 cells types, a detailed genome-wide expression analysis was carried out between three hESC lines that we routinely maintain in the lab (HSF1, H9, and CSES4) 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...
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 hPSCs.

**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 iPSCs 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 iPSCs 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 iPSC 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).
Expression Signatures Distinguish iPSCs from hESCs

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 in our experiments were tested whether expression differences between miPSC and their embryonic equivalents as determined by a Student’s t test (p < 0.05) and have at least a 1.5-fold change. 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^{-13}), suggesting that the iPSC state is remarkably robust across species (Figure 3B; Table S6).

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^{-13}); between the Chin, Maherali, and Yu signatures, 106 genes were shared (p < 10^{-9}); between Chin, Soldner, and Yu, 48 genes were shared (p < 10^{-3}). Among all the experiments of all four laboratories, 15 genes are differentially expressed between early-passage hiPSCs and hESCs (p < 10^{-5}–4; 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 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^{-13}), 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 factors.
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 log2 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 log2 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 log2 expression ratio between mESCs and miPSCs for miPSC signature genes ordered according to decreasing ratios. For these genes, the binding strength (−log10pXbar) 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^{-5}). 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).
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...
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
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.

Table 1. Regions in hiPSCs with Genomic Abnormalities

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Cell Line Identified</th>
<th>Region</th>
<th>Size (Mb)</th>
<th>Z-Score</th>
<th>Mean (log2ratio) Fibroblast/hiPSC</th>
<th>No. of Genes</th>
<th>No. of miRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>7p11.2</td>
<td>hiPSC2</td>
<td>56808738-62112547</td>
<td>5.3 Mb</td>
<td>19.37</td>
<td>0.213</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8q12.3</td>
<td>hiPSC1</td>
<td>6203767-100955803</td>
<td>83.8 Mb</td>
<td>45.26</td>
<td>−0.185</td>
<td>148</td>
<td>3</td>
</tr>
<tr>
<td>9q34.3</td>
<td>hiPSC2</td>
<td>13771482-13962739</td>
<td>1.9 Mb</td>
<td>21.60</td>
<td>0.207</td>
<td>75</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>hiPSC18</td>
<td>135323818-139634573</td>
<td>4.3 Mb</td>
<td>21.38</td>
<td>0.138</td>
<td>102</td>
<td>1</td>
</tr>
<tr>
<td>10q21.3</td>
<td>hiPSC2</td>
<td>67642045-67930512</td>
<td>0.3 Mb</td>
<td>19.24</td>
<td>0.473</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>11p11.12</td>
<td>hiPSC2</td>
<td>48381190-55253892</td>
<td>6.9 Mb</td>
<td>19.24</td>
<td>0.149</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>12p13.31</td>
<td>hiPSC1</td>
<td>62119-11760632</td>
<td>11.7 Mb</td>
<td>45.32</td>
<td>−0.175</td>
<td>176</td>
<td>2</td>
</tr>
<tr>
<td>14q32.33</td>
<td>hiPSC2</td>
<td>103616693-106182502</td>
<td>2.6 Mb</td>
<td>19.71</td>
<td>0.163</td>
<td>27</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>hiPSC18</td>
<td>10335890-10622351</td>
<td>2.9 Mb</td>
<td>18.12</td>
<td>0.141</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>16p13.3</td>
<td>hiPSC2</td>
<td>7675-2208335</td>
<td>2.2 Mb</td>
<td>19.30</td>
<td>0.172</td>
<td>104</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>hiPSC18</td>
<td>53-2234072</td>
<td>2.2 Mb</td>
<td>20.49</td>
<td>0.181</td>
<td>107</td>
<td>2</td>
</tr>
<tr>
<td>19p13.3</td>
<td>hiPSC2</td>
<td>208435-2449396</td>
<td>2.2 Mb</td>
<td>24.08</td>
<td>0.213</td>
<td>86</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>hiPSC18</td>
<td>40285-3774141</td>
<td>3.7 Mb</td>
<td>24.35</td>
<td>0.170</td>
<td>121</td>
<td>2</td>
</tr>
<tr>
<td>20q13.33</td>
<td>hiPSC18</td>
<td>59860004-62429688</td>
<td>2.6 Mb</td>
<td>18.50</td>
<td>0.154</td>
<td>81</td>
<td>7</td>
</tr>
<tr>
<td>21q22.3</td>
<td>hiPSC1</td>
<td>38084180-46913738</td>
<td>8.8 Mb</td>
<td>73.43</td>
<td>0.320</td>
<td>144</td>
<td>0</td>
</tr>
<tr>
<td>22q13.32</td>
<td>hiPSC2</td>
<td>46997532-49534378</td>
<td>2.5 Mb</td>
<td>21.66</td>
<td>0.177</td>
<td>37</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>hiPSC18</td>
<td>40986242-49567312</td>
<td>8.6 Mb</td>
<td>22.54</td>
<td>0.100</td>
<td>103</td>
<td>3</td>
</tr>
<tr>
<td>Xq21.1</td>
<td>hiPSC1</td>
<td>77883731-90950867</td>
<td>13.1 Mb</td>
<td>20.61</td>
<td>0.080</td>
<td>30</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>hiPSC2</td>
<td>25192507-151731387</td>
<td>126.5 Mb</td>
<td>31.42</td>
<td>0.122</td>
<td>714</td>
<td>69</td>
</tr>
<tr>
<td>Yp11.2-q11.21</td>
<td>hiPSC2</td>
<td>2858143-20250487</td>
<td>17.4 Mb</td>
<td>28.19</td>
<td>0.120</td>
<td>27</td>
<td>0</td>
</tr>
</tbody>
</table>

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 feedback loops of gene regulation under the direction of the
Expression Signatures Distinguish iPSCs from hESCs

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, fibroblast 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 (Krutzenfeld 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

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 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.

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.
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; Karumadhy et al., 2007; 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**

**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 Maheralli et al. (2007) and Mikkelson 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 r 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 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.

**ACKNOWLEDGMENTS**

M.H.C. is supported by the USHHS Ruth L. Kirschstein Institutional (NRSA #T32 CA099056) and G.A. by NIH/NICHD 5K12 HD001281. S.V. is supported by Regione Emilia Romagna PRIRITT Biopharmnet. C.M.C. is supported by NIH-NCI. M.A.T. is supported by NIH and CIRM Grant R51-00313. M.G. is supported by Regione Emilia Romagna PRIRITT Biopharmnet. C.M.C. 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 OD001688-01), 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 (JCLA), 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


