Cell Stem Cell
Matters Arising Response

Molecular Analyses of Human Induced Pluripotent Stem Cells and Embryonic Stem Cells

Mark H. Chin,[1] Matteo Pellegrini,[2] Kathrin Plath,[1,*] and William E. Lowry[2,*]
[1]Department of Biological Chemistry, David Geffen School of Medicine
[2]Department of Molecular, Cell and Developmental Biology
University of California Los Angeles, Los Angeles, CA 90095, USA
*Correspondence: kplath@mednet.ucla.edu (K.P.), blowry@ucla.edu (W.E.L.)
DOI 10.1016/j.stem.2010.06.019

SUMMARY

Recent work from our group and others has argued that human induced pluripotent stem cells (hiPSCs) generated by the introduction of four viruses bearing reprogramming factors differ from human embryonic stem cells (hESCs) at the level of gene expression (Chin et al., 2009). Many of the differences seen were common across independent labs and, at least to some extent, are thought to be a result of residual expression of donor cell-specific genes (Chin et al., 2009; Ghosh et al., 2010; Marchetto et al., 2009). Two new reports reanalyze similar expression data sets as those used in Chin et al. (2009) and come to different conclusions (Newman and Cooper, 2010; Guenther et al., 2010). We compare various approaches to perform gene expression meta-analysis that all support our original conclusions and present new data to demonstrate that polycistronic delivery of the reprogramming factors and extended culture brings hiPSCs transcriptionally closer to hESCs.

INTRODUCTION

The recent advent of reprogramming human somatic cells to a pluripotent state has led to the examination of human induced pluripotent stem cells (hiPSCs) and their relation to human embryonic stem cells (hESCs). Our study (Chin et al., 2009) demonstrated that hiPSCs generated up until that time had statistically significant differences in gene expression compared to available hESCs. Furthermore, many of these differences, particularly those found in early passage iPSC lines, were conserved across various studies and species, suggesting that these two cell types, although very similar to each other, were nonetheless transcriptionally distinguishable. Soon after, two additional studies came to similar conclusions using hiPSCs derived from fibroblasts, neural tissue, adipocytes, and keratinocytes (Ghosh et al., 2010; Marchetto et al., 2009). These studies went further to propose that residual misexpression from the cell type of origin is responsible for the observed differences in gene expression between hiPSCs and hESCs (Ghosh et al., 2010; Marchetto et al., 2009). Similarly, consistent differences in miRNA expression between hiPSCs and hESCs were also reported (Chin et al., 2009; Wilson et al., 2009).

hiPSCs and hESCs have also been compared at the level of the epigenome. In Chin et al. (2009), we showed by chromatin immunoprecipitation in combination with microarrays (ChIP-chip) that histone H3K27 and H3K4 trimethylation levels at promoter regions were not distinguishable between hiPSCs and hESCs. Hawkins et al., and now Guenther et al., confirmed that trimethylation of H3K27 was very similar between hESCs and hiPSCs by ChIP in combination with massive parallel sequencing (ChIP-Seq) (Hawkins et al., 2010). On the other hand, it was shown that trimethylation on histone H3K9 is significantly different between hESCs and hiPSCs and appears to contribute to the differences in gene expression between hESCs and hiPSCs seen at early passage (Hawkins et al., 2010).

Along this line, two groups also demonstrated that significant differences can be detected between hESCs and hiPSCs at the level of DNA methylation (Deng et al., 2009; Doi et al., 2009), further indicating that epigenetic modifications could be differentially affected in the reprogramming process. Furthermore, when hESCs and hiPSCs from patients with Fragile X (FX) Syndrome were generated, only the FX-hESCs displayed an active *FMR1* locus, whereas reprogramming failed to reactivate this locus (Urbach et al., 2010). Finally, recent work from our lab suggests that reprogramming female human fibroblasts under conventional culture conditions fails to reactivate the somatically silent X chromosome (Tchieu et al., 2010), whereas other work suggests that hESC lines that capture this primitive state can be generated (Lengner et al., 2010). In summary, there is considerable evidence that hiPSCs and hESCs are epigenetically and transcriptionally different at early passage.

Given the vastly different circumstances by which hiPSCs and hESCs are generated, it is not surprising that these various analyses deem them to have distinct molecular properties. Perhaps what is most surprising is in fact the high degree to which transcription factor-induced reprogramming is able to reconstruct the pluripotent state. A key question then becomes whether any of the gene expression differences seen between hiPSCs and hESCs are functionally relevant. Interestingly, it was shown that different combinations of reprogramming factors lead to mouse (m) iPSCs with different developmental potential (Han et al., 2010). Furthermore, recent advances with mouse reprogramming suggests that miPSC lines can be derived that possess the same functional characteristics as mESCs in their capacity to generate mice via tetraploid (4N)

complementation (Boland et al., 2009; Kang et al., 2009), but that most miPSC lines, even from the same starting population, do not support 4N complementation. Several new lines of evidence point to the activity of a few imprinted genes and miRNAs to be at least partially responsible for the differences in pluripotency (Liu et al., 2010; Stadtfeld et al., 2010). Specifically, cell lines that support 4N complementation express the Dlk1-Dio3 imprinted cluster, whereas those that fail in this approach do not (Stadtfeld et al., 2010). Thus, misexpression of just a few genes can functionally distinguish miPSCs from most miPSCs and mESCs, arguing that even very small differences can have profound consequences.

Correlating transcriptional differences between hiPSCs and hESCs with functional differences is considerably more difficult than in the mouse system because the powerful 4N complementation assay cannot be applied and quantitative assays to assess pluripotency are difficult to establish. It is conceivable that all the hiPSCs generated to date do fall into functional categories that have yet to be defined, pending the discovery of pluripotency assays that are more quantitative than teratoma formation. Given that we don't have these defined functional subclassifications of hiPSCs and hESCs, it is not yet possible to mine human cell reprogramming expression data for differences that correlate with functional outputs. It was also suggested that genetic background contributes significantly to gene expression changes between mESCs and miPSCs (Stadtfeld et al., 2010), but whether this also contributes to differences between hESCs and hiPSCs is still unknown. However, when comparing expression differences between early passage hiPSCs and hESCs of the reprogramming experiment from one lab with that from a different lab, we and others consistently found a significant proportion of genes to be differentially expressed between these two cell types (Chin et al., 2009; Ghosh et al., 2010; Marchetto et al., 2009).

We also reported that the overlap of expression differences decreases as more independent reprogramming experiments from different labs were compared (Chin et al., 2009). Regarding the question of consistent differences between hiPSCs and hESCs, two groups now suggest that when many lines of hiPSCs and hESCs from different labs are compared, consistent differences between them are largely lost and that most expression differences between hiPSCs and hESCs are lab specific or stochastic in nature (Newman and Cooper, 2010 and Guenther et al., 2010). Both sets of authors performed their own meta-analysis using similar data sets as presented in Chin et al. (2009), and the latter went further to derive additional data from new hiPSCs and compared them with a larger group of hESCs. Both groups took issue with the meta-analysis methods used in Chin et al. (2009), and we would like to take this opportunity to discuss the meta-analyses employed to start a conversation on "best practice" for meta-analysis of gene expression of hESCs and hiPSCs. We present a reanalysis of data from Chin et al. with additional statistical methods to demonstrate that the conclusions made in Chin et al were in fact appropriate, and we highlight reasons for the discrepancy in interpretations between Newman and Cooper, Guenther et al., and Chin et al. Furthermore, we present data with new hiPSCs to demonstrate that reprogramming methods may affect the kinetics of this process and reconfirm that extended culturing brings hiPSCs closer to hESCs.

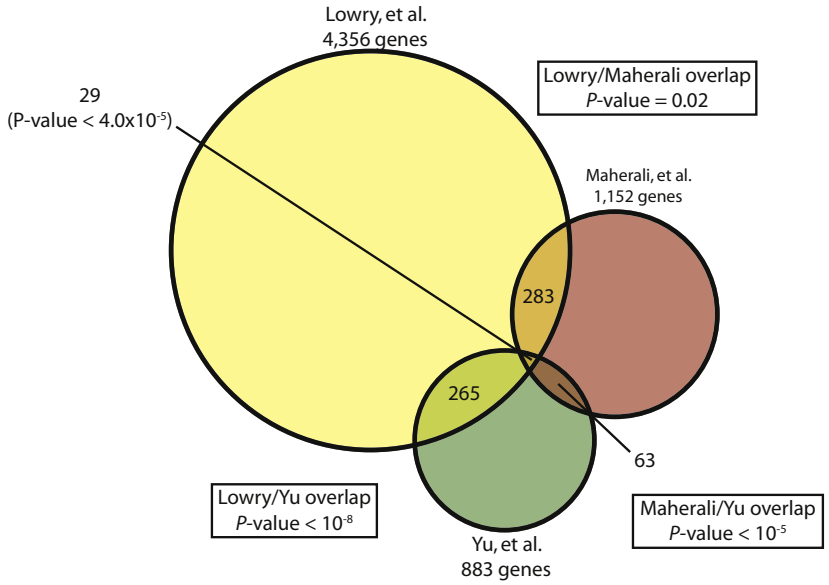
RESULTS AND DISCUSSION

In Chin et al. (2009), we sought to determine differences in gene expression between hESCs and early passage hiPSCs and to test the repeatability of the changes seen in among different independent reprogramming experiments from several labs (Hochedlinger, Thomson, Jaenisch, and our own). We chose to analyze the reprogramming experiments from different labs independently and applied our normalization and differential gene expression discovery algorithm to each reprogramming experiment separately. Thus, our normalization strategy was different from that used in Figure 4 of Guenther et al. (2010) and Figure 2 of Newman and Cooper (2010) in which both grouped all samples together and normalized them in a single event. In preparing Chin et al. (2009), we also found that, because hESCs and hiPSCs are quite similar to each other, clustering all data sets from different groups together appeared to cluster them according to lab of origin, not by cell type. Because our interest was to compare the differences between the hiPSCs and hESCs from each group, we believed that independent normalization was the appropriate strategy. Since each reprogramming experiment was treated independently, any measurement error from a single reprogramming experiment should have had no influence on data obtained from another. Given that we found significant overlap between the gene lists generated as differentially expressed between hESCs and hiPSCs among different reprogramming experiments, we argued that these differences were not entirely stochastic (Chin et al., 2009).

To increase the power of our analysis, we also filtered the expression data before generating lists of differentially expressed genes by collapsing the probe expression data for each gene. To this end, we started by only taking the "highest confidence" probe sets for a given RefSeq ID and eliminating probe sets that were defined by Affymetrix as potentially having more nonspecific binding, then averaged expression data from remaining probe sets for a given RefSeq to further reduce false positives. Importantly, by collapsing the probe sets to characterize each RefSeq ID, we were able to include expression data from mouse experiments, which required the use of Homologene database (NCBI), a function that does not work at the probe ID level. The extent to which this original approach contributed to differences in the interpretation of results between Newman and Cooper (2010), Guenther et al. (2010), and Chin et al. (2009) is unclear.

To then define significant gene expression differences between hiPSCs and hESCs, we employed the combination of a Student's t test ($p = 0.05$) and fold change (1.5×) in our prior study (Chin et al., 2009) rather than an approach based on t testing with false discovery correction (FDR). This approach was chosen because it was argued that the use of fold change with a p value cutoff filtering produces lists of differentially expressed genes in a more reproducible manner (Shi et al., 2008). However, it has also been proposed that multiple hypothesis testing may be the better method for eliminating noise in gene expression experiments (Chen and Sarkar, 2006; Grant et al., 2005; Shedden et al., 2005; Tsai et al., 2003). Therefore, we have now reanalyzed the same reprogramming data sets (from Hochedlinger, Thomson, and our own lab) that were

A Overlap requiring the same directionality of gene expression differences between hESCs and hiPSCs among various reprogramming experiments



B Overlap regardless of directionality of gene expression differences

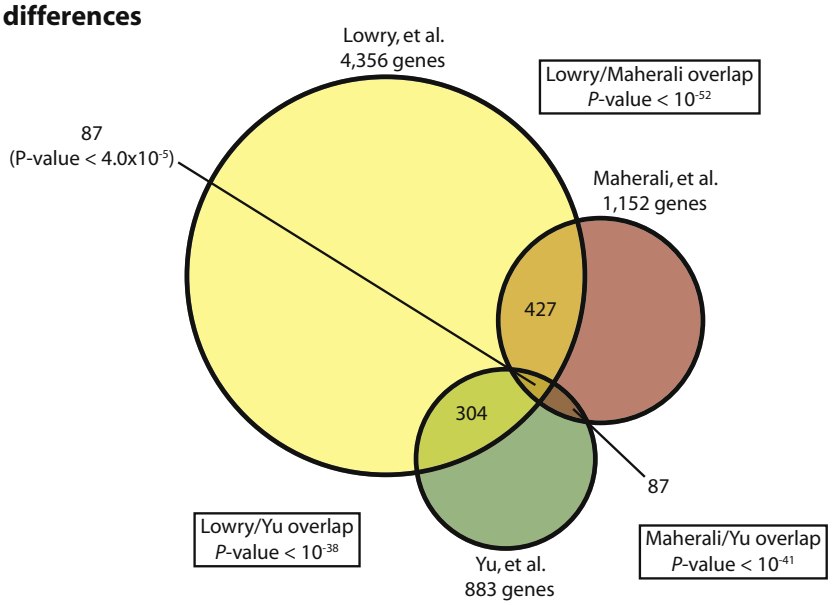


Figure 1. Overlap of Differentially Expressed Genes between hiPSCs and hESCs from Different Labs, Determined with the Bayesian T Test with FDR Correction and 1.5 Fold Cutoff

Differentially expressed genes between hiPSCs and hESCs were determined for each of the three indicated reprogramming experiments (Lowry et al., 2008; Maherali et al., 2008; Yu et al., 2009) with a Bayesian t test with a FDR < 0.05 along with a greater than 1.5 fold change requirement. In (A), overlap among the differentially expressed gene lists was only considered if they were differentially expressed in the same direction between hESCs and hiPSCs. In (B), directionality was not taken into consideration for the overlap. See also Figure S1. p values were determined as in Chin et al., (2009).

used in Chin et al. (2009) by using a Bayesian t test to adjust for error variance, with an FDR < 0.05 and a 1.5 fold-change cutoff. This examination found a similar number of genes differentially expressed between hiPSCs and hESCs in each reprogramming experiment as described previously in Chin et al. (Figure S1A available online). Furthermore, the genes differentially expressed between hESCs and hiPSCs at early passage identified previously by p value and fold-change cutoff (Chin et al., 2009) and now by FDR and fold change were 70%–98% identical, indicating that, in this case, the use of p value and fold change provided a similar stringency to FDR/fold change (Figure S1A).

We also demonstrated in Chin et al. (Figure S8 in Chin et al., 2009) that even when taking directionality of expression differences between early passage hiPSCs and hESCs into account, there was still significant overlap between differential gene expression lists (hiPSCs versus hESCs) from various labs in over 80% of dual comparisons. Here, we have also reanalyzed the overlap of differential gene lists generated by using the Bayesian t test with FDR correction and 1.5 fold-change cutoff between the Hochedlinger (Maherali et al., 2008), Thompson (Yu et al., 2009), and our reprogramming experiments. We found that the data sets still have a significant number of gene expression changes in common even when directionality of gene expression differences was taken into account (Figures 1A and 1B). When determining the gene expression differences between hESCs and hiPSCs and comparing data across different labs or from independent experiments, it is of course more likely that those differences with conserved directionality are more likely to be functionally relevant. However, until functional data suggest otherwise, it is possible that even the differences whose direction is not conserved but still found consistently found in different labs are important.

We have now also performed a post hoc permutation analysis to estimate the FDR for differentially expressed genes between hESCs and early passage hiPSCs from different reprogramming experiments that were determined by p value and fold-change cutoff in Chin et al. (2009). Although the sample size with only few hESC and hiPSC lines was not large, this analysis

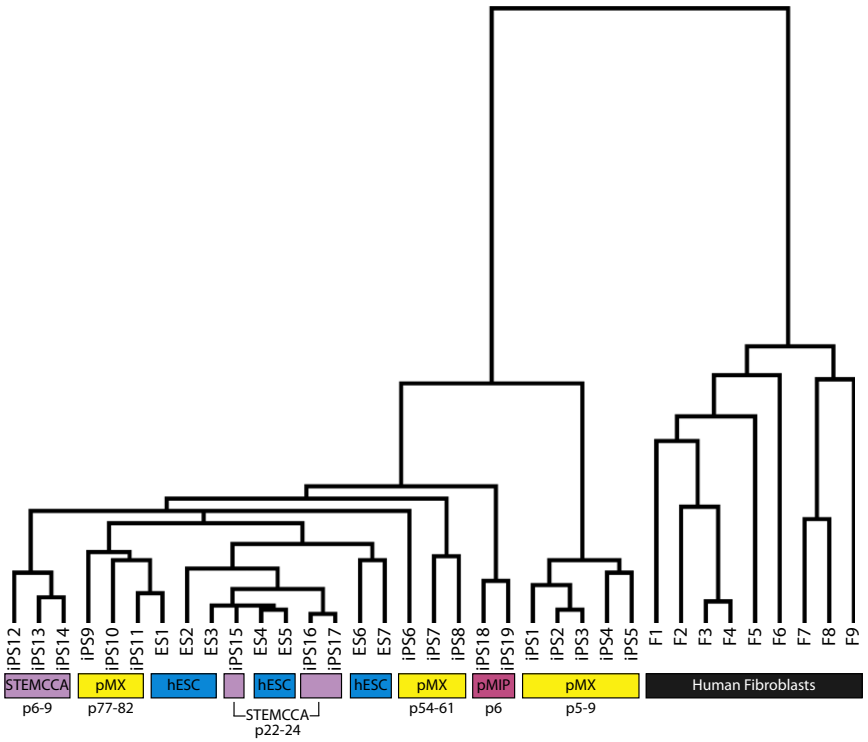


Figure 2. Variation in Gene Expression between hiPSCs and hESCs Generated in Our Laboratories with Different Technical Strategies

Hierarchical clustering of expression data obtained for hESCs (blue highlight), various hiPSCs at different passage (p) (yellow for single pMX-retroviral hiPSCs, pink for polycistronic pMIP retroviral hiPSCs, and purple for polycistronic STEMCCA lentiviral hiPSCs), and fibroblasts (F, black). The Supplemental Experimental Procedures section provides further identification for each cell line.

demonstrated that the statistical stringency provided by the use of p value and fold-change cutoff was equal to an approximated FDR of 2% for early passage hiPSCs versus hESCs (Figure S1B). Therefore, we conclude that the differentially expressed genes between early passage hiPSCs and hESCs and the overlap among different reprogramming experiments as determined in Chin et al. (2009) are reproducible when using different analysis tools.

The original hiPSC lines described in Chin et al. (2009) and Lowry et al. (2008) were generated by the expression of the Yamanaka set of four transcription factors each from individual MMLV-based retroviruses (pMX vectors). It has since been shown that lentivirus is suitable for delivering the reprogramming factors (Brambrink et al., 2008; Maherali et al., 2008; Stadtfeld et al., 2008), and newer methods use polycistronic vectors to deliver all four reprogramming factors in one virus and were shown to be very effective at generating hiPSC lines (Sommer et al., 2008; Carey et al., 2009; Chang et al., 2009; Gonzalez et al., 2009; Shao et al., 2009). To further define gene expression differences due to reprogramming, we have now generated new hiPSC lines with various newer methods and under improved culturing conditions during the reprogramming process. We took advantage of polycistronic vectors and employed both a MMLV-based retrovirus and a lentivirus to deliver the reprogramming factors to a variety of fibroblast lines derived from both males and females (Tchieu et al., 2010) (see Supplemental Experimental Procedures for a summary of the lines used and their accession numbers). As shown in Figure 2, all early passage hiPSC lines generated with polycistronic reprogramming tools are more similar to hESCs than the hiPSCs made originally in our lab with four separate viruses each carrying one reprogramming factor. Together, these data imply that whereas previous methods to derive hiPSCs generated lines that could be distinguished from hESCs by gene expression as shown in three independent studies (Chin et al., 2009; Ghosh et al., 2010; Marchetto et al., 2009), newer methods and procedures appear to more faithfully reprogram fibroblasts to a pluripotent state.

As shown in Chin et al. (2009), hiPSC gene expression becomes much more similar to hESCs as a result of passaging as the gene expression profiles of our late passage pMX-hiPSCs 1, 2, and 18 converged with that of hESCs. Similarly, Soldner et al. demonstrated that although gene expression differences were detected between hiPSCs and hESCs, much of this difference was eliminated upon removal of the reprogramming factors by Cre-mediated excision (Soldner et al., 2009). Whether extended passaging due to subcloning played a role in this effect is unknown. Here, we show that our original pMX-hiPSC lines when passaged at least 77 times came even closer to hESCs than they were after ~55 passages (Figures 2 and 3). Furthermore, we show that extended passaging of the polycistronic lines brought them transcriptionally even closer to hESCs (Figure 3), suggesting that passaging can play a role despite the method used to reprogram.

It should be noted that the comparisons made in Figure 3 of Guenther et al. used data from late passage hiPSCs from Chin et al. but early passage hiPSC lines from Maherali et al. (2008) and Yu et al. (2009), which could shed light on the differing interpretations of Guenther et al. and Chin et al. (2009). This probably explains the low number of genes presented in Figure 3C in Guenther et al. (2010) for the Chin et al. data and why the Maherali and Yu data sets overlapped with each other but not with the Chin et al. data as they did in the original Chin et al. and in Figure 1A here.

Our new data are congruent with some new results from Guenther et al. (2010), showing that newly derived lines made by polycistronic vectors and passaged significantly in our lab do not consistently cluster separately from hESCs also grown in our lab. This finding does not suggest that these newly derived hiPSCs are identical to hESCs, only that there are fewer differences between them than found in previous hiPSCs lines

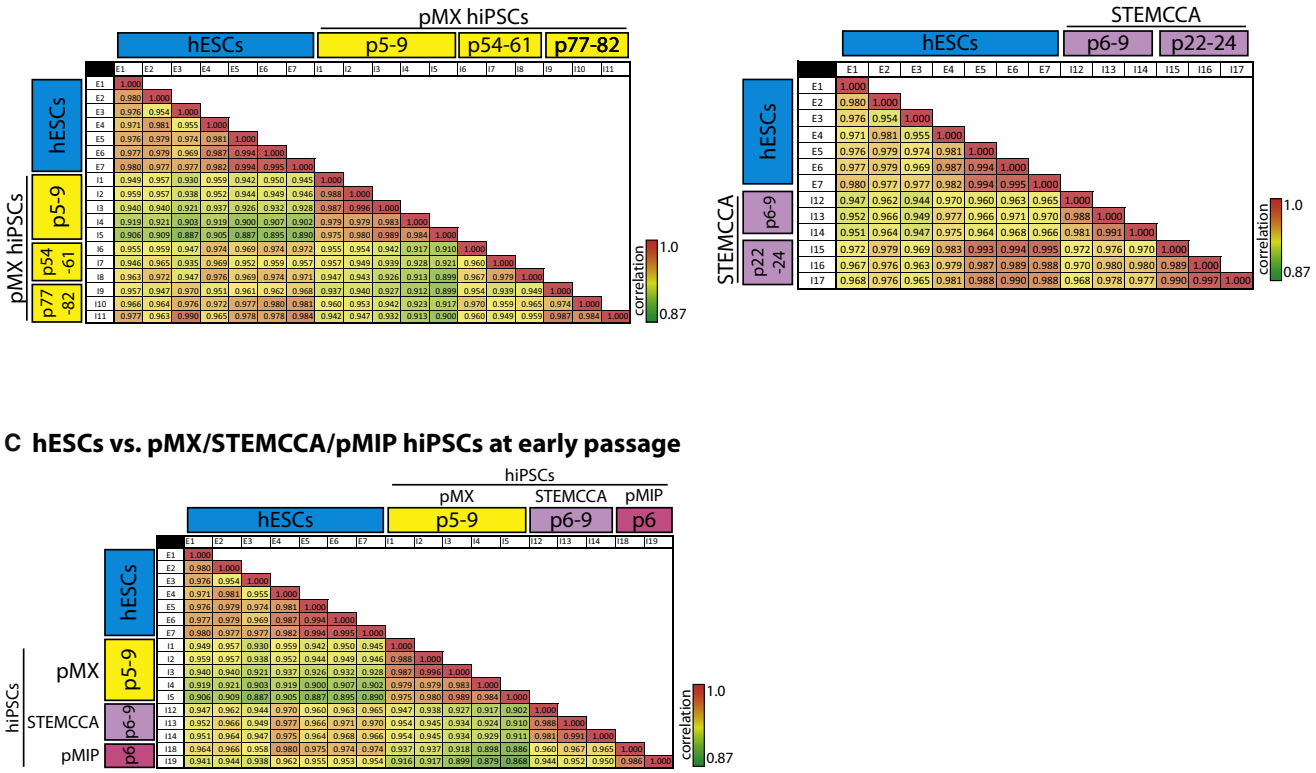


Figure 3. Extended Passaging Brings the Gene Expression Signature of hiPSCs Closer to that of hESCs

(A) Correlation table comparing normalized and filtered expression data for hESC lines to hiPSCs generated with four separate pMX retroviruses, which were profiled at different passages (p) as indicated in the yellow highlight. Values present the Pearson's correlation coefficient between two cell lines.

(B) Correlation table comparing hESC lines to hiPSCs generated with the polycistronic STEMCCA lentivirus at different passages.

(C) Correlation table comparing hESCs to hiPSCs at early passage derived from the three unique reprogramming experiments: 4-factor pMX-hiPSCs (yellow), polycistronic STEMCCA lentivirus (purple), and polycistronic pMIP retrovirus (pink).

made with different methods. The Guenther et al. (2010) data set shows that comparing hiPSCs and hESCs leaves just four genes with consistent gene expression differences. Although this small number of differences might be irrelevant, it is worth noting that just one genetic locus containing several noncoding transcripts and miRNAs appears to distinguish different functional classes of miPSCs (Stadtfeld et al., 2010).

In light of these findings, it is important to consider passage number, reprogramming technology, and genetic background when comparing pluripotent cells from various sources. However, it remains unclear what drives the transition of hiPSCs closer to hESCs. Our data suggest that with improved technology, one can probe deeper to find expression differences between hiPSCs and hESCs not related to lab-specific differences.

EXPERIMENTAL PROCEDURES

Tissue Culture and Reprogramming Methods

Cells were cultured and reprogrammed as described in Lowry et al. (2008) and Tchieu et al. (2010). A summary of the cell lines used in this study is given in the Supplemental Experimental Procedures. All cells were grown under a protocol approved by the Embryonic Stem Cell Research Oversight (ESCRO) committee at UCLA.

Gene Expression Analysis

The gene expression profiles of several new hiPSC and hESC as well as fibroblast lines was determined by Affymetrix Arrays as described in Chin et al. (2009) and are summarized in the Supplemental Experimental Procedures with their NCBI Gene Expression Omnibus (GEO) database accession numbers. The gene expression data was normalized with Robust Multichip Analysis (RMA) in R. Probe set data were then collapsed to RefSeqIDs with the highest confidence probesets as in Chin et al. (2009). A Bayesian t test was implemented so that differentially expressed genes between hiPSC and hESC lines could be determined (Baldi and Long, 2001; Sharov et al., 2005). Genes with an FDR < 0.05 and a fold change greater than 1.5-fold were called significantly differentially expressed. Significance between any two data sets was determined with the hypergeometric test, and for three data sets the significance was measured with a simulation with replacement.

SUPPLEMENTAL INFORMATION

Supplemental Information includes one figure and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.stem.2010.06.019.

ACKNOWLEDGMENTS

We would like to thank David Casero for his statistical advice and assistance with implementation. M.H.C. is supported by a Ruth L. Kirschstein Institutional Training Grant (NRSA # T32CA009056), and K.P. was supported by the NIH Director's Young Innovator Award (DP2OD001686) and a CIRM Young

Investigator Award (RN1-00564). This work was also supported by the CIRM New Cell Line grant (Jerome Zack, PI). W.E.L. holds the Maria Rowena Ross Chair in Cell Biology and Biochemistry and is supported by The March of Dimes.

Received: May 5, 2010
Revised: June 15, 2010
Accepted: June 29, 2010
Published: August 5, 2010

REFERENCES

Baldi, P., and Long, A.D. (2001). A Bayesian framework for the analysis of microarray expression data: Regularized t -test and statistical inferences of gene changes. Bioinformatics *17*, 509–519.

Boland, M.J., Hazen, J.L., Nazor, K.L., Rodriguez, A.R., Gifford, W., Martin, G., Kupriyanov, S., and Baldwin, K.K. (2009). Adult mice generated from induced pluripotent stem cells. Nature *461*, 91–94.

Brambrink, T., Foreman, R., Welstead, G.G., Lengner, C.J., Wernig, M., Suh, H., and Jaenisch, R. (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell Stem Cell *2*, 151–159.

Carey, B.W., Markoulaki, S., Hanna, J., Saha, K., Gao, Q., Mitalipova, M., and Jaenisch, R. (2009). Reprogramming of murine and human somatic cells using a single polycistronic vector. Proc. Natl. Acad. Sci. USA *106*, 157–162.

Chang, C.W., Lai, Y.S., Pawlik, K.M., Liu, K., Sun, C.W., Li, C., Schoeb, T.R., and Townes, T.M. (2009). Polycistronic lentiviral vector for "hit and run" reprogramming of adult skin fibroblasts to induced pluripotent stem cells. Stem Cells *27*, 1042–1049.

Chen, J., and Sarkar, S.K. (2006). A Bayesian determination of threshold for identifying differentially expressed genes in microarray experiments. Stat. Med. *25*, 3174–3189.

Chin, M.H., Mason, M.J., Xie, W., Volinia, S., Singer, M., Peterson, C., Ambartsumyan, G., Aimiuwu, O., Richter, L., Zhang, J., et al. (2009). Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. Cell Stem Cell *5*, 111–123.

Deng, J., Shoemaker, R., Xie, B., Gore, A., LeProust, E.M., Antosiewicz-Bourget, J., Egli, D., Maherali, N., Park, I.H., Yu, J., et al. (2009). Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. Nat. Biotechnol. *27*, 353–360.

Doi, A., Park, I.H., Wen, B., Murakami, P., Aryee, M.J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S., et al. (2009). Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat. Genet. *41*, 1350–1353.

Ghosh, Z., Wilson, K.D., Wu, Y., Hu, S., Quertermous, T., and Wu, J.C. (2010). Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. PLoS ONE *5*, e8975.

Gonzalez, F., Barragan Monasterio, M., Tiscornia, G., Montserrat Pulido, N., Vassena, R., Batlle Morera, L., Rodriguez Piza, I., and Izpisua Belmonte, J.C. (2009). Generation of mouse-induced pluripotent stem cells by transient expression of a single nonviral polycistronic vector. Proc. Natl. Acad. Sci. USA *106*, 8918–8922.

Grant, G.R., Liu, J., and Stoeckert, C.J., Jr. (2005). A practical false discovery rate approach to identifying patterns of differential expression in microarray data. Bioinformatics *21*, 2684–2690.

Guenther, M.G., Frampton, G.M., Soldner, F., Hockemeyer, D., Mitalipova, M., Jaenisch, R., and Young, R.A. (2010). Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell Stem Cell *7*, this issue, 249–257.

Han, J., Yuan, P., Yang, H., Zhang, J., Soh, B.S., Li, P., Lim, S.L., Cao, S., Tay, J., Orlov, Y.L., et al. (2010). Tbx3 improves the germ-line competency of induced pluripotent stem cells. Nature *463*, 1096–1100.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., et al. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell *6*, 479–491.

Kang, L., Wang, J., Zhang, Y., Kou, Z., and Gao, S. (2009). iPS cells can support full-term development of tetraploid blastocyst-complemented embryos. Cell Stem Cell *5*, 135–138.

Lengner, C.J., Gimelbrant, A.A., Erwin, J.A., Cheng, A.W., Guenther, M.G., Welstead, G.G., Alagappan, R., Frampton, G.M., Xu, P., Muffat, J., et al. (2010). Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. Cell *141*, 872–883.

Liu, L., Luo, G.Z., Yang, W., Zhao, X., Zheng, Q., Lv, Z., Li, W., Wu, H.J., Wang, L., Wang, X.J., and Zhou, Q. (2010). Activation of the imprinted Dlk1-Dio3 region correlates with pluripotency levels of mouse stem cells. J. Biol. Chem. *285*, 19483–19490.

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., 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.

Marchetto, M.C., Yeo, G.W., Kainohana, O., Marsala, M., Gage, F.H., and Muotri, A.R. (2009). Transcriptional signature and memory retention of human-induced pluripotent stem cells. PLoS ONE *4*, e7076.

Newman, A.M., and Cooper, J.B. (2010). Lab-specific gene expression signatures in pluripotent stem cells. Cell Stem Cell *7*, this issue, 258–262.

Shao, L., Feng, W., Sun, Y., Bai, H., Liu, J., Currie, C., Kim, J., Gama, R., Wang, Z., Qian, Z., et al. (2009). Generation of iPS cells using defined factors linked via the self-cleaving 2A sequences in a single open reading frame. Cell Res. *19*, 296–306.

Sharov, A.A., Dudekula, D.B., and Ko, M.S. (2005). A web-based tool for principal component and significance analysis of microarray data. Bioinformatics *21*, 2548–2549.

Shedden, K., Chen, W., Kuick, R., Ghosh, D., Macdonald, J., Cho, K.R., Giordano, T.J., Gruber, S.B., Fearon, E.R., Taylor, J.M., and Hanash, S. (2005). Comparison of seven methods for producing Affymetrix expression scores based on False Discovery Rates in disease profiling data. BMC Bioinformatics *6*, 26.

Shi, L., Jones, W.D., Jensen, R.V., Harris, S.C., Perkins, R.G., Goodsaid, F.M., Guo, L., Croner, L.J., Boysen, C., Fang, H., et al. (2008). The balance of reproducibility, sensitivity, and specificity of lists of differentially expressed genes in microarray studies. BMC Bioinformatics *9* (*Suppl 9*), S10.

Soldner, F., Hockemeyer, D., Beard, C., Gao, Q., Bell, G.W., Cook, E.G., Hargus, 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.

Sommer, C.A., Stadtfeld, M., Murphy, G.J., Hochedlinger, K., Kotton, D.N., and Mostoslavsky, G. (2008). iPS Cell generation using a single lentiviral stem cell cassette. Stem Cells *27*, 543–549.

Stadtfeld, M., Brennand, K., and Hochedlinger, K. (2008). Reprogramming of pancreatic beta cells into induced pluripotent stem cells. Curr. Biol. *18*, 890–894.

Stadtfeld, M., Apostolou, E., Akutsu, H., Fukuda, A., Follett, P., Natesan, S., Kono, T., Shioda, T., and Hochedlinger, K. (2010). Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature *465*, 175–181.

Tchieu, J., Kuoy, E., Chin, M.H., Trinh, H., Patterson, M., Sherman, S., Aimiuwu, O., Lingren, A., Zack, J.A., Clark, A., et al. (2010). Female human iPS cells retain an inactive X-chromosome. Cell Stem Cell *7*. in press.

Tsai, C.A., Hsueh, H.M., and Chen, J.J. (2003). Estimation of false discovery rates in multiple testing: Application to gene microarray data. Biometrics *59*, 1071–1081.

Urbach, A., Bar-Nur, O., Daley, G.Q., and Benvenisty, N. (2010). Differential modeling of fragile X syndrome by human embryonic stem cells and induced pluripotent stem cells. Cell Stem Cell *6*, 407–411.

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.

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, I.I., and Thomson, J.A. (2009). Human induced pluripotent stem cells free of vector and transgene sequences. Science *324*, 797–801.

Note Added in Proof

Two more groups have now shown that DNA methylation patterns can be used to molecularly distinguish iPSC lines both from ESCs and from iPSCs made from different cell types and that this memory can play a functional role during differentiation (Kim et al., 2010; Polo et al., 2010). Furthermore, one of these studies also demonstrated that the epigenetic differences between pluripotent stem cell lines disappear as the cells are passaged extensively (Polo et al., 2010), consistent with what was found for gene expression in Chin et al. (2009) and the data presented here.

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I.R., et al. (2010). Epigenetic memory in induced pluripotent stem cells. Nature., in press. Published online July 19, 2010. 10.1038/nature09342.

Polo, J.M., Liu, S., Figueroa, M.E., Kulalert, W., Eminli S., Tan, K.Y., Apostolou, E., Stadtfeld, M., Li, Y., Shioda, T., et al. (2010). Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat. Biotechnol., in press. Published online July 19, 2010. 10.1038/nbt.1667.