

# Identification of a specific reprogramming-associated epigenetic signature in human induced pluripotent stem cells

Sergio Ruiz<sup>a,1</sup>, Dinh Diep<sup>b,c,1</sup>, Athurva Gore<sup>b</sup>, Athanasia D. Panopoulos<sup>a</sup>, Nuria Montserrat<sup>d</sup>, Nongluk Plongthongkum<sup>b</sup>, Sachin Kumar<sup>a</sup>, Ho-Lim Fung<sup>b</sup>, Alessandra Giorgetti<sup>d</sup>, Josipa Bilic<sup>d</sup>, Erika M. Batchelder<sup>a</sup>, Holm Zaehres<sup>e</sup>, Natalia G. Kan<sup>f</sup>, Hans Robert Schöler<sup>e</sup>, Mark Mercola<sup>f</sup>, Kun Zhang<sup>b,2</sup>, and Juan Carlos Izpisua Belmonte<sup>a,d,2</sup>

<sup>a</sup>Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037; <sup>b</sup>Department of Bioengineering, Institute for Genomic Medicine and Institute of Engineering in Medicine, and <sup>c</sup>Bioinformatics and Systems Biology Graduate Program, University of California at San Diego, La Jolla, CA 92093; <sup>d</sup>Center of Regenerative Medicine in Barcelona, 08003 Barcelona, Spain; <sup>e</sup>Department of Cell and Developmental Biology, Max Planck Institute for Molecular Biomedicine, 48149 Münster, Germany; and <sup>f</sup>Sandford-Burnham Medical Research Institute, La Jolla, CA 92037

Edited by Kathrin Plath, University of California, Los Angeles, CA, and accepted by the Editorial Board August 23, 2012 (received for review February 10, 2012)

**Generation of human induced pluripotent stem cells (hiPSCs) by the expression of specific transcription factors depends on successful epigenetic reprogramming to a pluripotent state. Although hiPSCs and human embryonic stem cells (hESCs) display a similar epigenome, recent reports demonstrated the persistence of specific epigenetic marks from the somatic cell type of origin and aberrant methylation patterns in hiPSCs. However, it remains unknown whether the use of different somatic cell sources, encompassing variable levels of selection pressure during reprogramming, influences the level of epigenetic aberrations in hiPSCs. In this work, we characterized the epigenomic integrity of 17 hiPSC lines derived from six different cell types with varied reprogramming efficiencies. We demonstrate that epigenetic aberrations are a general feature of the hiPSC state and are independent of the somatic cell source. Interestingly, we observe that the reprogramming efficiency of somatic cell lines inversely correlates with the amount of methylation change needed to acquire pluripotency. Additionally, we determine that both shared and line-specific epigenetic aberrations in hiPSCs can directly translate into changes in gene expression in both the pluripotent and differentiated states. Significantly, our analysis of different hiPSC lines from multiple cell types of origin allow us to identify a reprogramming-specific epigenetic signature comprised of nine aberrantly methylated genes that is able to segregate hESC and hiPSC lines regardless of the somatic cell source or differentiation state.**

Induction of pluripotency in human somatic cells is an inefficient process that can be achieved by the expression of defined transcription factors (1–5). This reprogramming process involves global epigenetic remodeling and overcoming similar roadblocks present during cell transformation, which might affect genomic and epigenomic integrity (6). In fact, several recent reports have shown that human induced pluripotent stem cells (hiPSCs) contain genetic and epigenetic aberrations throughout their genome compared with their parental somatic cell populations or to human embryonic stem cells (hESCs) (7–12). For example, the analysis of whole-genome DNA methylation profiles at single-nucleotide resolution in hiPSCs, their somatic cells of origin, and hESCs revealed the presence of more than 1,000 differentially methylated regions (DMRs) between hiPSCs and hESCs (11). Moreover, this analysis, and many others, demonstrated both the persistence of specific epigenetic marks from the somatic cell of origin (residual methylation) and the acquisition of unique methylation patterns in mouse iPSCs (miPSCs) and hiPSCs (11, 13–21). Interestingly, hiPSC lines also show incomplete reprogramming of non-CG methylation in regions proximal to telomeres and centromeres (11). Altogether, these epigenetic aberrations might explain some of the observed transcriptional variation between hESC and hiPSC lines (22–24). In one of the most comprehensive reports to date, Bock et al. (23) characterized a panel of 20 hESC and 12 hiPSC lines to demonstrate that despite their global similarity, a number of genes

in each pluripotent cell line deviated from the normal expected variation compared with the DNA methylation and gene expression levels observed in the other pluripotent cell lines. Interestingly, they reported that no apparent epigenetic deviation was unique to all hiPSC lines (23). Altogether, these findings demonstrate that hiPSCs contain epigenetic aberrations. However, a majority of these reports predominantly used fibroblast-derived hiPSC lines and, thus, it remains unknown whether the use of alternative somatic cell types with variable levels of selection pressure for reprogramming might result in hiPSC lines containing fewer (or perhaps none) of these epigenetic alterations. Furthermore, although it has been shown that aberrantly methylated CpG sites are transmitted to differentiated cells (11), it remains unclear whether these epigenetic aberrancies result in transcriptional variation after differentiation.

In this work, we characterize at single nucleotide resolution the methylation profile of 17 hiPSC lines derived from six different somatic cell types with varied reprogramming efficiencies. Our results show that, independent of the somatic cell source used for reprogramming, all hiPSC lines analyzed contain abnormal epigenetic patterns. We determine that a majority of these aberrantly methylated CpG sites are transmitted to differentiated cells and can be associated with changes in gene expression after differentiation. Importantly, we identify a reprogramming-associated epigenetic signature comprised of nine aberrantly methylated genes that can segregate hESC and hiPSC lines both in the pluripotent state and after differentiation. These observations will contribute to a deeper understanding of the reprogramming process and underscore the need for a rigorous evaluation of the epigenetic integrity of hiPSC lines.

## Results

**Reprogramming Efficiency Inversely Correlates with the Percentage of Epigenetic Modifications Observed After Reprogramming.** To gain insight into the epigenetic integrity of hiPSCs, we performed targeted bisulfite sequencing with padlock probes (25, 26) to analyze the methylomes of 17 hiPSC lines, their 6 somatic cell

Author contributions: S.R., D.D., A. Gore, K.Z., and J.C.I.B. designed research; S.R., D.D., A. Gore, A.D.P., N.M., N.P., H.-L.F., A. Giorgetti, J.B., E.M.B., N.G.K., and M.M. performed research; H.Z. and H.R.S. contributed new reagents/analytic tools; S.K. analyzed data; and S.R., D.D., A. Gore, A.D.P., K.Z., and J.C.I.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. K.P. is a guest editor invited by the Editorial Board.

The array as well as the methylation data reported in this paper have been deposited in the Gene Expression Omnibus (GEO database), [www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo) (accession nos. GSE39210 and GSE40372).

<sup>1</sup>S.R. and D.D. contributed equally to this work.

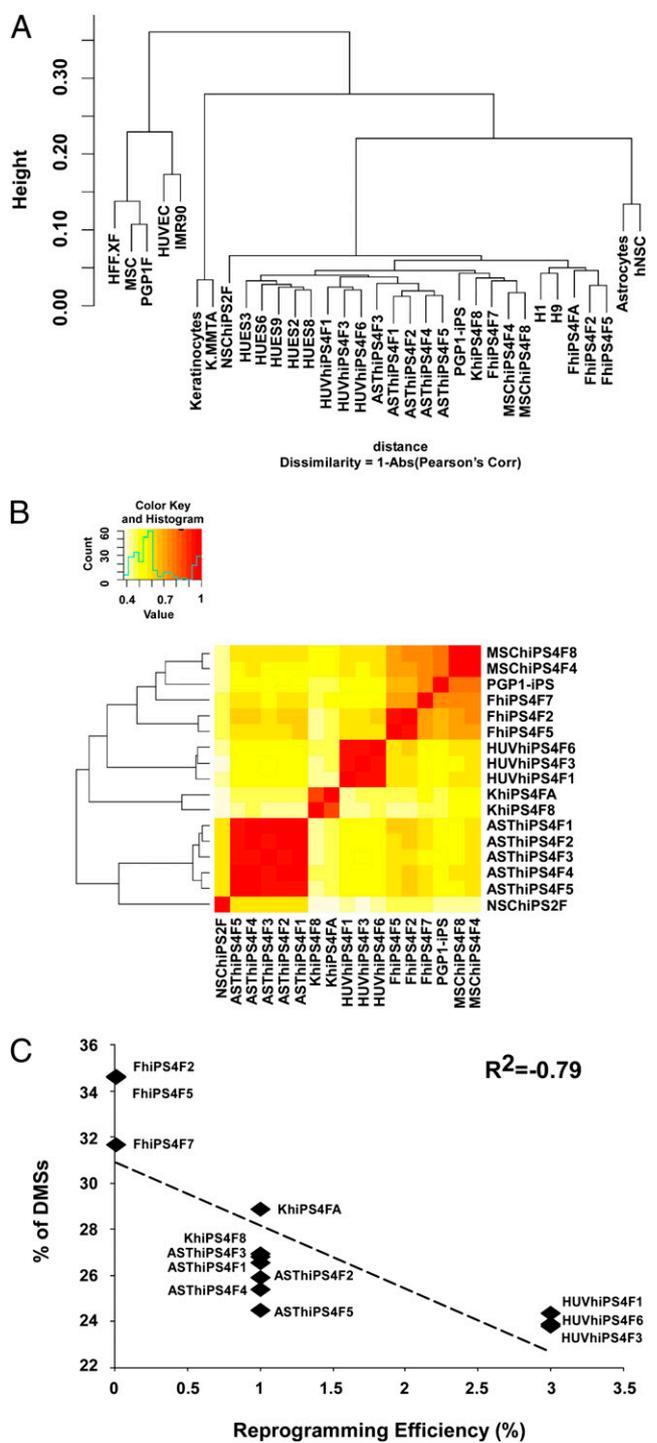
<sup>2</sup>To whom correspondence may be addressed. E-mail: [belmonte@salk.edu](mailto:belmonte@salk.edu) or [kzhang@bioeng.ucsd.edu](mailto:kzhang@bioeng.ucsd.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1202352109/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1202352109/-DCSupplemental).

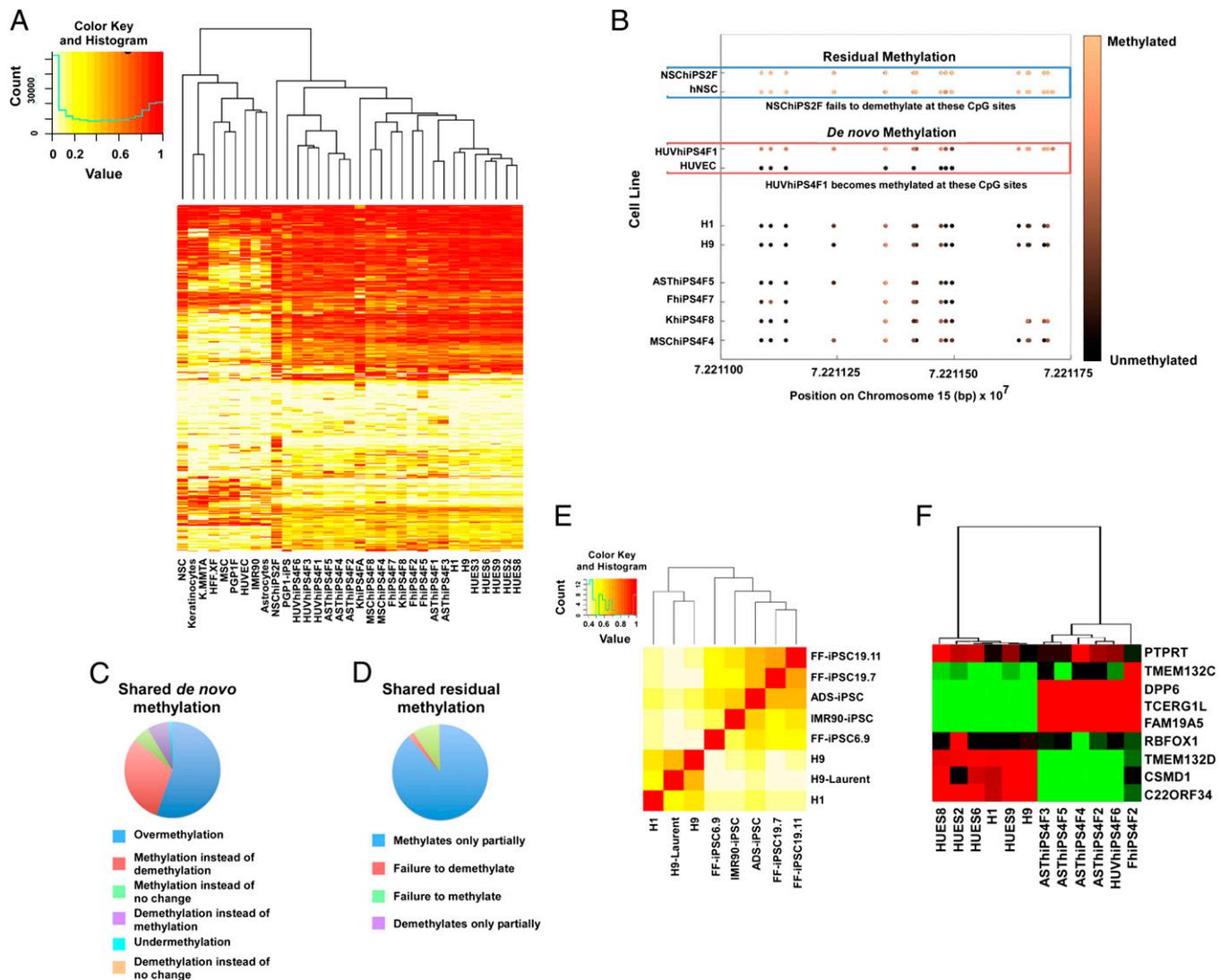
types of origin, and 7 hESC lines (*SI Text*, Fig. S1, and Table S1). We designed and synthesized a set of 330,000 synthetic probes targeting ~140,000 genomic regions known to be differentially methylated across different cell types (12, 27, 28) and additional functional regions. We determined the absolute methylation levels for an average of ~529,000 CpG sites per sample (Table S1). Although only ~1% of the human genome was covered by this assay, these preselected CpG sites were more than twice as informative as typical sites in CpG islands characterized by using lower resolution sequencing or in previously used bisulfite sequencing methods (Table S2). Unbiased hierarchical clustering of global methylation levels demonstrated a clear segregation of somatic cells and pluripotent cells (Fig. 1A). We also observed that hiPSC lines originating from the same somatic cell type tended to cluster together in subgroups (Fig. 1A and B), which, as reported (11, 13–21), supports the existence of residual methylation from somatic cells of origin in hiPSCs.

We analyzed the number of differentially methylated CpG sites (DMSs) in each hiPSC line by comparing each cell line to its direct somatic cell source of origin (Table S2). We observed that between 23% and 37% of CpG sites analyzed underwent a change in methylation state, with mesenchymal stem cells (MSCs) and fibroblasts requiring the most dramatic epigenetic change following reprogramming and neural stem cells (NSCs) requiring the least (Table S2). Interestingly, the percentage of DMSs after reprogramming correlated inversely with reprogramming efficiency, with cell sources undergoing the fewest epigenetic modifications reprogramming at higher efficiency (Fig. 1C). Moreover, we confirmed previous findings (11) and determined that, independent of somatic cell source, the global change in methylation observed after reprogramming is toward a more methylated state (Fig. S2A). Next, we investigated whether different somatic cell sources shared a core set of DMSs that might be essential to epigenetically reprogram to a pluripotent state. In fact, we observed that ~5,700 DMSs were shared among all hiPSC lines (Fig. S2B). Analysis of Gene Ontology for genes that could potentially be regulated by these DMSs revealed that genes with hypomethylated DMSs appeared to be enriched for cell signaling, protein refolding, cell metabolism, and neuronal development, whereas genes with hypermethylated DMSs appeared to be enriched for cell-cell adhesion and receptor behavior (Dataset S1).

**hiPSC Lines Share a Core Set of Aberrantly Methylated Genes That Segregate Them from hESCs.** We compared the methylation state at each CpG site in individual hiPSC lines to that of their parental source and seven hESC lines. Using an algorithm based on the  $\chi^2$  test with multiple testing corrections, we identified sites where hiPSC lines carried a methylation pattern significantly different from hESC lines (*SI Materials and Methods*). hiPSC lines derived from the same somatic cell source carried similar, although not identical, aberrant methylation patterns and clustered together based on methylation level at aberrant sites (Fig. 2A). We categorized the aberrantly methylated CpG sites into two categories: residual methylation, where the CpG site in a hiPSC line retains the methylation level of its parental cell instead of reaching the level observed in hESCs (Fig. 2B), and de novo methylation, where the CpG site in a hiPSC line acquires a methylation state found neither in its somatic source nor in hESCs (Fig. 2B). We determined that the percentage of aberrant CpG sites varied between 0.92% and 3.82% across the hiPSC lines analyzed. Furthermore, the percentage of CpG sites that showed residual methylation or de novo methylation varied between 0.32% and 1.60% and 0.57% and 2.98%, respectively (Table 1). Although we did not find a direct correlation between the amount of aberrant methylation and reprogramming efficiency or somatic cell type, we noted that some cell types appeared to possess lower aberrant methylation levels (e.g., astrocyte-derived lines) compared with others (e.g., fibroblast-derived lines) (Table 1). We determined that most aberrantly methylated CpG sites showing de novo methylation were characterized by excessive methylation after reprogramming (Fig. 2C), whereas most aber-



**Fig. 1.** Identification and classification of the epigenetic changes occurring during cell reprogramming. (A) Hierarchical clustering of the indicated cell lines based on the methylation state of all characterized CpG sites. HUVEC, human umbilical vein endothelial cell; K-MMTA, keratinocyte cell line; MSC, mesenchymal stem cell; NSC, neural stem cell; PGP1F, HFF.XF, IMR90, fibroblasts lines. (B) Heatmap and ordered dendrogram for all hiPSC lines based on the level of relative change observed at each differentially methylated site compared with the values observed in each respective somatic cell of origin. Pearson's correlation values were used to generate a single distance metric. (C) Reprogramming efficiency of somatic cell lines estimated after hiPSCs generation by retroviral infection of OCT4, SOX2, KLF4, and cMYC inversely correlates with the percentage of differential methylation achieved in hiPSC lines. Note that amount of epigenetic reorganization required appears to be a barrier to reprogramming.  $R^2$ , Pearson's correlation value.



**Fig. 2.** Pluripotent cells can be segregated based on the methylation/gene expression level of nine genes. (A) Heatmap and hierarchical clustering results of the cell lines used in this study using methylation patterns at CpG sites containing aberrant methylation in at least one hiPSC line. Similar aberrant epigenetic patterns were observed in hiPSCs derived from common somatic sources, and these lines accordingly tend to cluster together. (B) Graphical representation of an example of residual methylation and de novo methylation located on chromosome 15 (*ISLR2* gene). Each circle corresponds to an individual CpG site and the level of methylation is represented in a colored pattern. In the example shown, NSChIP52F retains the epigenetic pattern of its somatic progenitor (hNSC), showing residual methylation. HUVhiPS4F1 takes on an epigenetic pattern not observed in its somatic progenitor or any of the other pluripotent lines, showing a hiPSC line-specific de novo methylation. Methylation levels of the same CpG sites in hESC and hiPSC lines were included for comparison. (C and D) Types of methylation errors leading to epigenetic aberrations. Most aberrantly methylated CpG sites associated to genes showing de novo methylation (C) and residual methylation (D) in all hiPSC lines are characterized by overmethylation or partial methylation, respectively. (E) Heatmap and ordered dendrogram for the hiPSC and hESC described lines (11) based on the level of relative change observed at CpG sites associated to our nine signature genes. Note that hESC and hiPSC lines segregated in two different groups. (F) Hierarchical clustering of six hiPSC (ASThiPS4F2, 3, 4, and 5, HUVhiPS4F6, and FhiPS4F2) and six hESC (H1, H9, HUES2, HUES6, HUES8, and HUES9) lines based on the gene expression level analyzed by real-time PCR of the nine common aberrantly methylated genes identified in hiPSC lines used in this study.

rantly methylated CpG sites associated with genes showing residual methylation were characterized by only partial methylation occurring after reprogramming (Fig. 2D).

To gain insight into potential functional consequences of these epigenetic aberrations, we linked each aberrant CpG site with its closest gene (*SI Materials and Methods*) and used this subset of genes for further analysis. Interestingly, we observed that a very small number of genes contained aberrant methylation patterns in nearly all hiPSC lines assayed in our study (16/17 hiPSC lines) regardless of somatic cell source (Dataset S2). We hypothesized that the nine genes (*PTPRT*, *TMEM132C*, *TMEM132D*, *TCERG1L*, *DPP6*, *FAM19A5*, *RBFOX1*, *CSMD1*, and *C22ORF34*) we identified might represent a core set of aberrantly methylated

genes that can systematically distinguish hiPSC and hESC lines. Thus, we performed unbiased hierarchical clustering based on the methylation status of CpG sites associated to this small subset of genes in previously published independent methylation datasets. Specifically, we first examined a set of whole-genome bisulfite sequencing data performed in three hESC and five hiPSC lines (11). We found that, similar to what we observed for our dataset, the methylation level of CpG sites associated to the nine genes was able to clearly segregate hESC and hiPSC lines into two distinct groups (Fig. 2E). Additionally, we used a recently published dataset that profiled the genome-wide DNA methylation level for more than 450,000 CpG sites in 19 hESC and 29 hiPSC lines (13) and observed that, despite the lower

**Table 1. Summary of CpG sites containing residual methylation and de novo methylation in targeted regions**

Cell line	Testable sites	% aberrant	% memory	% mutation	No. of genes potentially affected by memory	No. of genes potentially affected by mutation
ASThiPS4F4	434388	1.02	0.45	0.57	191	182
ASThiPS4F5	437266	0.92	0.35	0.58	211	186
ASThiPS4F1	404245	1.30	0.35	0.96	189	310
ASThiPS4F2	380656	1.16	0.41	0.75	171	243
ASThiPS4F3	343025	2.07	0.41	1.65	219	616
FhiPS4F7	340395	2.53	1.27	1.25	487	591
HUVhiPS4F1	374103	1.33	0.38	0.95	200	474
HUVhiPS4F3	392482	1.41	0.42	0.99	251	588
HUVhiPS4F6	433768	1.29	0.32	0.97	190	455
FhiPS4F2	354763	1.62	0.52	1.10	292	213
FhiPS4F5	296451	2.47	0.62	1.85	362	682
KhiPS4F8	396085	2.60	0.82	1.78	586	1040
KhiPS4FA	270126	2.41	0.46	1.95	288	831
MSChiPS4F4	437957	2.34	0.96	1.39	560	462
MSChiPS4F8	429575	2.85	1.60	1.25	896	552
NSChiPS2F	327308	3.82	0.84	2.98	538	1912
PGP1-iPS	437433	2.63	1.47	1.16	997	703

Aberrantly (residual and de novo) methylated CpG sites were classified as such when showing at least a 0.2 change in absolute methylation level and considered to have methylation levels from different underlying distributions by the  $\chi^2$  test (with Benjamini-Hochberg multiple testing correction, FDR = 0.01). Genes potentially affected by aberrantly methylated CpG sites were defined as described in *SI Materials and Methods*.

resolution, a similar clustering analysis clearly segregated all but two hiPSC lines from hESC lines (Fig. S3A).

Next, we investigated whether our core set of aberrantly methylated genes showed differential gene expression in hiPSC lines compared with hESC lines by performing real-time PCR analysis on RNA obtained from six hiPSC lines and six hESC lines (a description of the primers used in this study can be found in Table S3). An unbiased hierarchical clustering of the real-time PCR data results examining the gene expression of the nine shared aberrantly methylated genes demonstrated a clear segregation between hiPSC and hESC lines (Fig. 2F). Furthermore, to determine the global relevance of these findings, we also performed a similar unbiased hierarchical clustering by using previously reported independent datasets containing a variety of hESC and hiPSC lines (a total of 12 datasets). Overall, when examining the expression of these nine genes, we determined that although clear outliers and different subgroups among hiPSC lines were detected, a majority of the dataset clusters showed separation between hiPSC and hESC lines (Fig. S3B and C and Dataset S2). These combined results suggest the existence of shared epigenetic aberrancies associated to a small subset of genes in hiPSC lines. The validation of these aberrancies by using our data and data from independent laboratories strongly corroborates the strength of our findings.

**Aberrant Methylation at CpG Sites Is Transmitted During hiPSC Differentiation, Resulting in Transcriptional Changes Compared with Differentiated hESCs.**

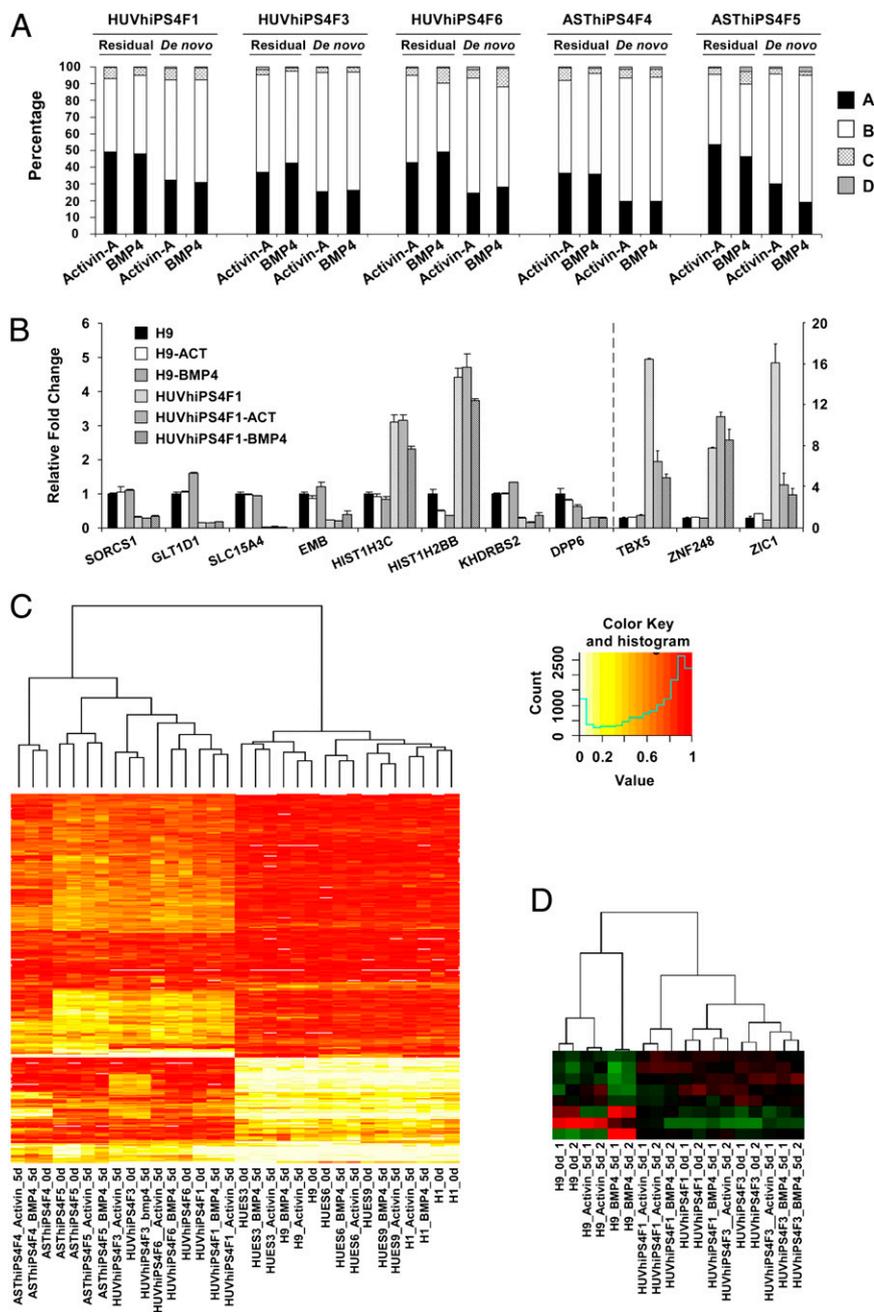
To further test whether the aberrant methylation and gene expression levels observed in hiPSC lines were maintained after loss of the pluripotent state, we differentiated five hESC lines and five hiPSC lines toward two different germ cell layers, endoderm and trophoectoderm, by using Activin-A and BMP4, respectively. We then performed targeted bisulfite sequencing to analyze the methylomes of the hESC and hiPSC lines in their pluripotent and differentiated states (Table S1). In addition, the gene expression levels of H9, HUVhiPS4F1 and HUVhiPS4F3 were profiled in duplicate by using Affymetrix ST 1.0 microarrays. Between 0.3% and 1% of CpG sites were aberrantly methylated in the hiPSC lines with respect to hESC lines (Dataset S3). We first investigated whether these epigenetic aberrations resulted in changes in gene expression in undifferentiated cells. We observed that between 3% and 7% of genes linked to these aberrantly methylated sites showed differential gene

expression in hiPSC lines compared with hESC lines (Fig. S4A and Dataset S3). Additionally, we tested the expression of five genes with line-specific epigenetic de novo methylation in HUVhiPS4F1 and observed that these genes also showed differential gene expression compared with other hiPSC or hESC lines (Fig. S4B). Taken together, these results indicate that some epigenetic aberrations are associated with changes in gene expression levels.

We next analyzed the methylation status of the aberrant CpG sites in both hiPSCs and hESCs after each differentiation protocol. The CpG sites were classified based on their postdifferentiation methylation status into four categories (see Fig. 3 for a detailed description; Dataset S3). We observed that ~20–50% of the aberrantly methylated CpG sites detected in hiPSC lines remained aberrant after differentiation into either of the two separate cell lineages (Fig. 3A). Importantly, we observed that a subset of genes associated with these CpG sites showing differential gene expression level in undifferentiated hiPSCs compared with hESCs still remain in that condition regardless of differentiation protocol (Fig. 3B, Fig. S4C, and Dataset S3). Finally, to further validate the potential of the identified hiPSC-specific epigenetic signature described above, we clustered the pluripotent cells and their differentiated progenies based on both the methylation level and transcriptional abundance of the nine signature genes (Fig. 3C and D and Fig. S4D). Interestingly, the samples segregated based on whether the progenitor line was a hiPSC or hESC, and clustered by specific cell line but not by differentiation protocol. Altogether, these data suggest that the methylation and gene expression levels of the aberrantly methylated genes in hiPSC lines still segregate hESCs and hiPSCs even after differentiation toward independent germ cell layers.

**Discussion**

In this work, we have used an expanded bisulfite padlock probe set to interrogate the methylation level of targeted CpG sites identified to carry differential methylation in various cell states regardless of CpG density (12, 26–28). This unique approach identified genes linked to individual aberrantly methylated CpG sites that are not necessarily located in CpG-enriched genomic regions. Our results show that epigenetic aberrations occur in hiPSCs regardless of the somatic cell type of origin. We demonstrated that aberrant epigenetic patterns in hiPSC lines influence gene expression and could explain functional diversity



**Fig. 3.** Reprogramming-associated epigenetic/transcriptional signatures segregate hiPSCs and hESCs after differentiation. (A) Percentage of aberrant CpG sites identified between hESC-derived lines and the corresponding hiPSC-derived lines classified in the following categories: aberrant methylation remains and is still aberrant compared with differentiated hESCs (A); aberrant methylation remains but is the same as the one found in differentiated hESCs (B); aberrant methylation is removed during differentiation reaching the level found in differentiated hESCs (C); and aberrant methylation changed to a new aberrant methylation state (D). (B) Genes with aberrantly methylated CpG sites and differential transcriptional abundance with at least a twofold cutoff were identified in the HUVhiPS4F1 cell line after comparison with H9 cells. Graph shows the relative fold change in the expression of genes still aberrantly methylated after differentiation between the differentiated HUVhiPS4F1 cell line and the differentiated hESC cell line. Note that differential expression was independent on whether Activin or BMP4-differentiated cells were analyzed. (C) Hierarchical clustering of hESC (H1, H9, HUES3, HUES6, and HUES9) and hiPSC (HUVhiPS4F1, HUVhiPS4F3, HUVhiPS4F6, ASThiPS4F4, and ASThiPS4F5) lines in their pluripotent and differentiated states based on the methylation level of the nine common aberrantly methylated genes identified in the hiPSC lines used in this study. (D) Hierarchical clustering of hESC (H9) and hiPSC (HUVhiPS4F1 and HUVhiPS4F3) lines in their pluripotent and differentiated states based on the gene expression level of the nine common aberrantly methylated genes identified in the hiPSC lines used in this study. Data were obtained from microarray analysis.

within hiPSC lines and between hiPSC and hESC lines (29–31). In fact, we observed the existence of genes aberrantly methylated and differentially expressed in hiPSC lines compared with hESCs that still remained in that condition after differentiation regardless of differentiation protocol.

The use of hiPSC lines derived from six different somatic cell types enabled us to narrow down a precise core set of genes that contained aberrant epigenetic patterns associated with the hiPSC state. This analysis led us to identify a reprogramming-associated epigenetic signature based on the methylation level of nine genes that could segregate hESC and hiPSC lines in both the pluripotent state and after differentiation. There have been many reports suggesting the existence of epigenetic and transcriptional differences between hiPSC and hESC lines (11–24). Interestingly, recently reported analysis using restricted representation bisulfite sequencing (RRBS) showed that although cell line-specific outliers at both the methylation and gene expression levels could be

identified, no apparent epigenetic deviation was unique to all hiPSC lines (23). However, the data presented therein (23) did not appear to target any of the aberrantly methylated CpG sites covered in our hiPSC-specific signature, because RRBS mainly focuses on the analysis of CpG islands (resulting in low coverage of genomic regions with low CpG density, including many functional elements such as enhancers). When we compared the lists of CpG sites associated with the nine genes characterized by our dataset to the Bock et al. dataset (23), there was almost no overlap between the two sets of analyzed CpG sites. In fact, in the Bock et al. dataset (23), only 1 CpG site of the ~600 we identified as aberrantly methylated CpG sites associated to the 9 genes was included in their analysis. Thus, when we clustered the pluripotent cell lines used in the Lister et al. dataset (which analyzed a near-complete selection of CpG sites genome-wide in an unbiased manner; ref. 11) based on the CpG sites that were analyzed by the Bock et al. dataset, no clear separation was observed

between hESC and hiPSC lines. However, when we clustered the hESC and hiPSC lines included in the Lister et al. dataset based on the CpG sites analyzed in our study, we found that we were able to segregate the two different pluripotent cell types. Furthermore, when we compared our data to an extensive set of genome-wide DNA methylation profiling of hESC and hiPSC lines, that had analyzed CpG sites that overlapped with our dataset (13), we were again able to separate these pluripotent cell lines based on our identified hiPSC-specific epigenetic signature. Altogether, these findings indicate that when characterizing the epigenetic differences between hiPSCs and hESCs, cautions must be taken to interpret the results when only a subset of genomic regions is investigated.

Furthermore, we also validated our reprogramming-associated epigenetic signature by using gene expression data from several previously reported datasets (refs. 13 and 32 and Dataset S2). We observed that a majority of independent clusters separated hiPSC and hESC lines, although clear outliers and different subgroups among hiPSC lines were detected. This result is not totally unexpected because it has been shown that gene expression levels in pluripotent cells are highly variable and depend on how pluripotent cells are generated or maintained (33). Moreover, Bock et al. (23) also reported the existence of genes in pluripotent cells that contained similar methylation levels but were associated to variable levels of gene expression. Therefore, we cannot exclude the possibility that some hiPSC lines might not segregate well from hESC lines when using the gene expression levels of these nine genes to cluster them.

Finally, although the genes *TMEM132D*, *FAM19A5*, and *TCERG1L* have been reported to be involved in neural processes, we did not identify any significant functional enrichment associated with the nine genes aberrantly methylated in hiPSC lines. Interestingly, Lister et al. (11) identified five of our nine genes (*TMEM132C*, *TMEM132D*, *FAM19A5*, *DPP6*, and *TCERG1L*) located within non-CG mega-DMRs as clear outliers in terms of gene expression compared with hESCs. In fact, up to half of their gene outliers located within non-CG mega-DMRs (11) were

observed aberrantly methylated in 14 of the 17 hiPSC used in this study (Dataset S3). Further studies will be needed to better clarify the role of non-CG mega-DMRs and their implication in the functional behavior of hiPSCs compared with hESCs.

Overall, the results shown here demonstrate the existence of intrinsic common reprogramming-associated epigenetic differences associated with the hiPSC state. We demonstrated that the epigenetic signature described in this work, based on the methylation level of nine genes, can segregate hiPSC and hESC lines in both the pluripotent state and after differentiation and could explain some of the functional differences between these two pluripotent cell types.

## Materials and Methods

**Cell Culture.** Human H9 (WA09), H1 (WA01), HUES2, HUES3, HUES6, HUES8, and HUES9 embryonic stem cell lines were obtained from WiCell Research Institute or Harvard University and maintained as described (34). Derived hiPSCs were cultured as described (34). IMR90 human fibroblasts (ATCC; CCL-186) and 293T cells were cultured in DMEM (Invitrogen) supplemented with 10% FBS and 0.1 mM nonessential amino acids. HUVEC cells were obtained from Lonza (C-2519A) and grown with EGM-2 media (Lonza) as recommended. MSCs were kindly provided by Cécile Volle (Sanofi-Aventis, Toulouse, France) and grown in  $\alpha$ -MEM (Invitrogen) containing 10% FBS (HyClone), penicillin/streptomycin, sodium pyruvate, nonessential amino acids, and L-glutamine (all from Invitrogen). Human keratinocytes were obtained and cultured as described (35).

Additional experimental and data analysis procedures are provided in *SI Materials and Methods*.

**ACKNOWLEDGMENTS.** We thank the J.C.I.B. and K.Z. laboratories for helpful discussions and to Dr. Travis Berggren, Margaret Lutz, and Veronica Modesto for their support at the Salk Institute-Stem Cell Core. S.R. was partially supported by Instituto de Salud Carlos III Grant CGCV-1335/07-3. A.G. was supported by the Focht-Powell Fellowship and a California Institute for Regenerative Medicine (CIRM) Predoctoral Fellowship. A.D.P. was partially supported by National Institutes of Health (NIH) Training Grant T32 CA009370. Work in this manuscript was supported by grants from Fundacion Cellex, Ministerio de Economía y Competitividad (MINECO), Sanofi, the G. Harold and Leila Y. Mathers Charitable Foundation, and The Leona M. and Harry B. Helmsley Charitable Trust, California Institute for Regenerative Medicine Grants RB3-05083 and TR1-01273, and NIH Grant R01 GM097253.

- Takahashi K, et al. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131:861–872.
- Yu J, et al. (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science* 318:1917–1920.
- Lowry WE, et al. (2008) Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc Natl Acad Sci USA* 105:2883–2888.
- Meissner A, Wernig M, Jaenisch R (2007) Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nat Biotechnol* 25:1177–1181.
- Park IH, et al. (2008) Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* 451:141–146.
- Daley GQ (2008) Common themes of dedifferentiation in somatic cell reprogramming and cancer. *Cold Spring Harb Symp Quant Biol* 73:171–174.
- Hussein SMI, et al. (2011) Copy number variation and selection during reprogramming to pluripotency. *Nature* 471:58–62.
- Laurent LC, et al. (2011) Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. *Cell Stem Cell* 8:106–118.
- Gore A, et al. (2011) Somatic coding mutations in human induced pluripotent stem cells. *Nature* 471:63–67.
- Mayshar Y, et al. (2010) Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. *Cell Stem Cell* 7:521–531.
- Lister R, et al. (2011) Hotspots of aberrant epigenetic reprogramming in human induced pluripotent stem cells. *Nature* 471:86–73.
- Doi A, 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.
- Nazor KL, et al. (2012) Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. *Cell Stem Cell* 10:620–634.
- Marchetto MC, et al. (2009) Transcriptional signature and memory retention of human-induced pluripotent stem cells. *PLoS ONE* 4:e7076.
- Ohi Y, et al. (2011) Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPSCs. *Nat Cell Biol* 13:541–549.
- Bar-Nur O, Russ HA, Efrat S, Benvenisty N (2011) Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells. *Cell Stem Cell* 9:17–23.
- Polo JM, et al. (2010) Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat Biotechnol* 28:848–855.
- Kim K, et al. (2010) Epigenetic memory in induced pluripotent stem cells. *Nature* 467:285–290.
- Kim K, et al. (2011) Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. *Nat Biotechnol* 29:1117–1119.
- Quattrocchi M, et al. (2011) Intrinsic cell memory reinforces myogenic commitment of pericyte-derived iPSCs. *J Pathol* 223:593–603.
- Hu Q, Friedrich AM, Johnson LV, Clegg DO (2010) Memory in induced pluripotent stem cells: Reprogrammed human retinal-pigmented epithelial cells show tendency for spontaneous redifferentiation. *Stem Cells* 28:1981–1991.
- Chin MH, et al. (2009) Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* 5:111–123.
- Bock C, et al. (2011) Reference Maps of human ES and iPSC cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* 144:439–452.
- Ghosh Z, et al. (2010) Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. *PLoS ONE* 5:e8975.
- Deng J, et al. (2009) Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat Biotechnol* 27:353–360.
- Diep D, et al. (2012) Library-free methylation sequencing with bisulfite padlock probes. *Nat Methods* 9:270–272.
- Irizarry RA, et al. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 41:178–186.
- Lister R, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462:315–322.
- Miura K, et al. (2009) Variation in the safety of induced pluripotent stem cell lines. *Nat Biotechnol* 27:743–745.
- Feng Q, et al. (2010) Hemangioblastic derivatives from human induced pluripotent stem cells exhibit limited expansion and early senescence. *Stem Cells* 28:704–712.
- Hu BY, et al. (2010) Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *Proc Natl Acad Sci USA* 107:4335–4340.
- Yu J, et al. (2009) Human induced pluripotent stem cells free of vector and transgene sequences. *Science* 324:797–801.
- Newman AM, Cooper JB (2010) Lab-specific gene expression signatures in pluripotent stem cells. *Cell Stem Cell* 7:258–262.
- Ruiz S, et al. (2011) A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity. *Curr Biol* 21:45–52.
- Aasen T, et al. (2008) Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. *Nat Biotechnol* 26:1276–1284.