

Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells

Kitai Kim^{1-3,9,10}, Rui Zhao^{1-3,10}, Akiko Doi^{4,10}, Kitwa Ng^{1,3,10}, Juli Unternaehrer¹⁻³, Patrick Cahan¹⁻³, Huo Hongguang^{1,3}, Yuin-Han Loh¹⁻³, Martin J Aryee⁵, M William Lensch¹⁻³, Hu Li⁶, James J Collins^{6,7}, Andrew P Feinberg⁴ & George Q Daley^{1-3,8}

We compared bona fide human induced pluripotent stem cells (iPSCs) derived from umbilical cord blood (CB) cells and neonatal keratinocytes (K). As a consequence of both incomplete erasure of tissue-specific methylation and aberrant *de novo* methylation, CB-iPSCs and K-iPSCs were distinct in genome-wide DNA methylation profiles and differentiation potential. Extended passage of some iPSC clones in culture did not improve their epigenetic resemblance to embryonic stem cells, implying that some human iPSCs retain a residual 'epigenetic memory' of their tissue of origin.

Somatic cells within an organism share the same genomic sequence, but when development occurs from the single-cell zygote, subsequent generations of cells acquire differential patterns of gene expression owing to alterations in chromatin structure and chemical modifications of the DNA, such as cytosine methylation. Reprogramming of somatic cells to pluripotency reverses this process of cell specification through epigenetic modification, and entails erasure of tissue-specific DNA methylation and re-establishment of the embryonic methylome. We have detected residual and aberrant tissue-specific DNA methylation in mouse iPSCs, which functions to confer epigenetic memory, biasing the differentiation potential of iPSCs toward lineages related to the donor cell^{1,2}. Here we investigate whether epigenetic memory persists in human iPSCs.

We reprogrammed neonatal CD34⁺ cells from umbilical cord blood cells and foreskin keratinocytes, representative cell types of distinct embryonic germ layers, mesoderm and ectoderm, whose

differentiation potential can be readily assayed *in vitro* (Fig. 1a and Supplementary Fig. 1). The resultant iPSCs from at least two independent cord blood and three keratinocyte donors passed stringent pluripotency tests characteristically applied to human embryonic stem cells (ESCs) (Supplementary Methods). We then tested the potential of the CB-iPSCs and K-iPSCs to differentiate into keratinocytes³. In day-6 embryoid bodies, K-iPSCs displayed 9.4-fold higher expression of the gene encoding keratin 14, a marker of early keratinocyte differentiation (Fig. 1b and Supplementary Fig. 2a), and quantitatively, K-iPSCs yielded 23-fold more keratinocytes than CB-iPSCs (Fig. 1c and Supplementary Fig. 2b), indicating that K-iPSCs show enhanced keratinocyte potential relative to CB-iPSCs. Next, we differentiated multiple independent clones of CB-iPSCs and K-iPSCs in methylcellulose to test hematopoietic potential. Despite their reduced capacity for keratinocyte differentiation, differentiating cultures of CB-iPSCs produced an expected range of myeloid colony types, whereas K-iPSCs surveyed from multiple donors yielded relatively few hematopoietic colonies (Fig. 1d). Multiple clones of CB-iPSCs consistently yielded a greater frequency of hematopoietic colonies than multiple clones of K-iPSCs, and more colonies than iPSCs isolated from adult CD34⁺ blood and keratinocytes (Supplementary Fig. 2a–c). However, neither CB-iPSCs nor K-iPSCs showed a difference in tissue differentiation potential into definitive endoderm (Supplementary Fig. 2d).

To analyze genome-wide DNA methylation patterns, we compared multiple CB-iPSC and K-iPSC lines and their parental somatic cells to human ESCs using comprehensive, high-throughput, array-based, relative methylation (CHARM) analysis, which interrogates 5.2 million CpG sites, including virtually all CpG islands and shores⁴. We determined the number of differentially methylated regions (DMRs) in pair-wise comparisons, using a threshold area cutoff of 2, corresponding to an approximate 5% false discovery rate (FDR)⁵ (Fig. 2a and Supplementary Table 1). We confirmed the results of CHARM analysis by bisulfite pyrosequencing of multiple loci (Supplementary Fig. 3).

Unsupervised hierarchical clustering using the 1,000 most variable probes across samples revealed that CB-iPSCs are easily distinguished from K-iPSCs (Fig. 2b). ESCs cluster with both CB-iPSCs and K-iPSCs, suggesting that the separation between CB-iPSCs and K-iPSCs is not due to different levels of pluripotency, as confirmed by gene expression profiling and analysis of binding of the pluripotency core transcription factor to the DMRs that distinguish CB-iPSCs and K-iPSCs (Supplementary Methods, Supplementary Tables 4–6 and Supplementary Figures 7–15). Taken together, these data indicate that

¹Stem Cell Transplantation Program, Division of Pediatric Hematology/Oncology, Manton Center for Orphan Disease Research, Howard Hughes Medical Institute, Children's Hospital Boston and Dana-Farber Cancer Institute, Boston, Massachusetts, USA. ²Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, USA. ³Harvard Stem Cell Institute, Cambridge, Massachusetts, USA. ⁴Center for Epigenetics and Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. ⁵Oncology Department, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, Maryland, USA. ⁶Department of Biomedical Engineering and Center for BioDynamics, Boston University, Boston, Massachusetts, USA. ⁷Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts, USA.; Howard Hughes Medical Institute. ⁸Division of Hematology, Brigham and Women's Hospital, Boston, Massachusetts, USA. ⁹Present address: Department of Cancer Biology and Genetics, Center for Cell Engineering, Sloan-Kettering Institute, New York, New York, USA. ¹⁰These authors contributed equally to this work. Correspondence should be addressed to G.Q.D. (george.daley@childrens.harvard.edu) or A.P.F. (afeinberg@jhu.edu).

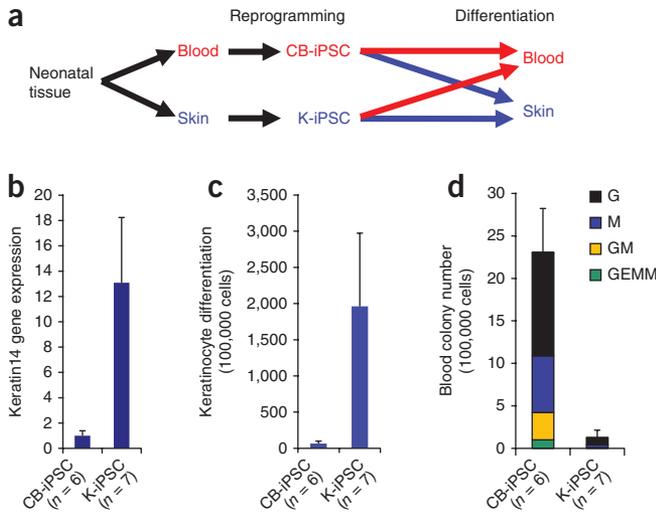


Figure 1 Derivation and differentiation of iPSCs from neonatal umbilical cord blood cells and foreskin fibroblasts. (a) Experimental schema. (b) Q-PCR of the keratinocyte marker K14 in day-6 embryoid bodies from CB-iPSCs ($n = 6$) and K-iPSCs ($n = 7$). Gene expression was normalized to actin, and shown as fold-difference relative to CB-iPSCs. (c) Numbers of keratinocytes differentiated from CB-iPSCs ($n = 6$) and K-iPSCs ($n = 7$). (d) Numbers of hematopoietic colony-forming cells in day-14 embryoid bodies differentiated from CB-iPSCs ($n = 6$) and K-iPSCs ($n = 7$). Error bar = s.d. G, granulocyte progenitor; M, macrophage progenitor; GM, granulocyte-macrophage progenitor; GEMM, granulocyte-erythrocyte-macrophage-megakaryocyte progenitor.

© 2011 Nature America, Inc. All rights reserved.

the methylation patterns of CB-iPSCs and K-iPSCs are considerably different and, despite fulfilling criteria for pluripotency, represent distinct epigenetic states as measured by CHARM.

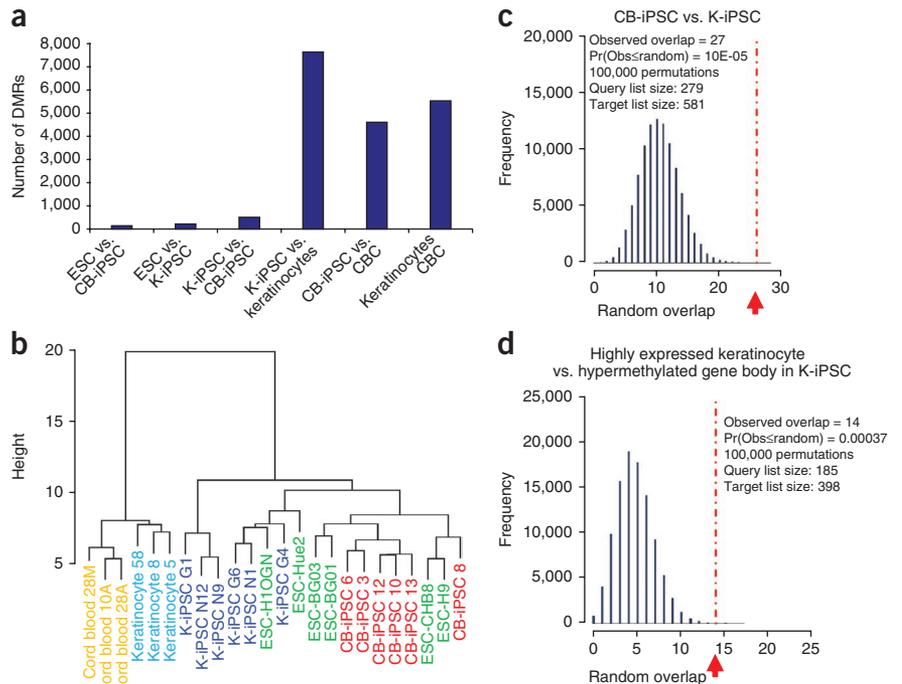
The set of genes identified by microarray analysis as differentially expressed between cord blood cells and keratinocytes includes regulators of cell identity (Supplementary Fig. 4). We found that 581 of the 1,519 differentially expressed genes are located in or near DMRs that distinguish cord blood cells from keratinocytes (1.73-fold more than expected by chance; $P < 10^{-5}$). We also found that 27 of these 581 genes are in or near DMRs that distinguish CB-iPSCs and K-iPSCs (Supplementary Table 2), a 2.55-fold enrichment over that predicted by chance ($P < 10^{-4}$, Fig. 2c). This enrichment persists even when pluripotency-related genes defined by compiled microarray data⁶ are removed from the gene set (Supplementary Fig. 5a). Further, 17 of the 581 genes overlap DMRs that distinguish CB-iPSCs from ESCs, a fourfold enrichment over that predicted by chance (Supplementary Fig. 5b).

Several lines of evidence support a direct mechanistic link between differential methylation and biased lineage differentiation of iPSC lines. First, a literature survey of the genes associated with the top ten DMRs that

distinguish CB-iPSCs from K-iPSCs indicates that five are associated with hematopoiesis, and four with epithelial cell phenotypes (Supplementary Table 3). Second, K-iPSCs maintained gene-body methylation, a phenomenon generally correlated with gene expression⁷, for numerous keratinocyte-associated genes. Of 185 gene bodies hypermethylated in K-iPSCs compared to CB-iPSCs, 14 correspond to tissue DMR-associated genes that are expressed in keratinocytes, a threefold enrichment over that predicted by chance (Fig. 2d). Conversely, of 19 gene bodies hypomethylated in CB-iPSCs compared to ESCs, four correspond to DMR-associated genes highly expressed in keratinocytes ($P = 0.021$), including the keratinocyte-specific transcription factor *RIPK4*. Taken together, our analysis supports the notion that the reprogramming process can leave residual methylation marks associated with both expression and repression of tissue-specific genes.

To determine whether the 370 DMRs that distinguish CB-iPSCs and K-iPSCs (Supplementary Table 1) represented residual methylation left over from the tissue of origin or were instead *de novo* and potentially aberrant methylation signatures generated during reprogramming, we compared the state of these regions in cord blood cells, keratinocytes and ESCs. Seventy-five residual DMRs were found in cord blood cells and keratinocytes, 28 DMRs were specific for ESCs and 267 DMRs were newly generated during reprogramming, reminiscent of our prior observation of frequent *de novo* DMRs in mouse iPSCs of fibroblast cell origin from aged donors¹.

Figure 2 Analysis of methylation in CB-iPSCs, K-iPSCs, ESCs and somatic cells. (a) Numbers of differentially methylated regions (DMRs) between CB-iPSCs, K-iPSCs, ESCs, umbilical cord blood cells and cultured keratinocytes. DMRs were defined by an area cutoff of 2.0. CBC, cord blood cells. (b) Cluster dendrogram analysis using the top 1,000 most variable probes across all samples. (c,d) Gene enrichment analysis of DMRs. Blue histograms represent a probability distribution of the number of genes predicted to overlap DMRs by chance. Red vertical lines indicate the observed number of genes that overlap DMRs. (c) Genes differentially methylated between CB-iPSCs and K-iPSCs are enriched in DMR-associated genes (genes both differentially expressed and methylated between cord blood cells and keratinocytes). (d) Genes highly expressed in keratinocytes are enriched in DMRs that are both hypermethylated in K-iPSCs relative to CB-iPSCs and are located in gene bodies rather than promoters.



Using ESCs as a reference, both CB-iPSC and K-iPSC DMRs were highly enriched for tissue-of-origin DMRs ($P < 0.0001$).

In studies of murine induced pluripotent stem cell lines, epigenetic memory can be erased over time by extended culture² or ectopic gene expression¹. During extended culture of K-iPSC clones N9 and G6 and CB-iPSC clone 6, the K-iPSC N9 clone showed a gradual increase in blood-forming potential (Supplementary Fig. 6a), but erasure of epigenetic memory by passaging did not occur in all clones. *HOXD8* was one of 27 DMR-associated genes in the somatic cell types that was more highly methylated in keratinocytes and K-iPSCs relative to cord blood cells and CB-iPSCs, respectively (Supplementary Fig. 3a). Notably, at the higher passages of a K-iPSC clone (N9), the restored blood-forming potential coincided with a reduction of *HOXD8* expression and methylation (Supplementary Fig. 6b,c). The *HOXD8* DMR overlaps with a CTCF binding site, which may control *HOXD8* expression⁸ (Supplementary Fig. 6d). Because *HOX* genes exert major effects on blood cell differentiation, we hypothesized that expression of *HOXD8* might be playing a role in inhibiting blood-forming potential in early passage K-iPSCs in the clone N9. Consistent with a negative effect of *HOXD8* expression on blood development, we observed that ectopic expression of *HOXD8* in CB-iPSCs reduced blood-forming potential (Supplementary Fig. 6e), as did restoring high levels of *HOXD8* in later passages of K-iPSC clone N9 (Supplementary Fig. 6f). In contrast, short hairpin RNA-mediated *HOXD8* knockdown in the K-iPSC clone G6 substantially improved its hematopoietic capacity (Supplementary Fig. 6g). These data correlate differential methylation and expression of the *HOXD8* locus with blood-forming potential in iPSC lines. However, further CHARM analysis of early and late passages of the clone N9 indicates that extended tissue culture did not make the cells epigenetically closer to ESCs (Supplementary Fig. 6h); rather, extended passage may have had locus-specific effects that influence the differential potential of iPSCs.

Prior studies comparing human iPSCs to ESCs have found that iPSCs are differentially methylated^{5,7,9} and exhibit significant reprogramming variability^{10–12}, and recently a transcriptional memory of somatic cells in human iPSCs has been reported¹³. However, a link has not yet been identified between disparate methylation signatures at multiple loci and altered differentiation potential. Here we find that methylation at loci important to tissue fate persists as a form of epigenetic memory even in human iPSCs that pass stringent criteria for pluripotency typically applied to human cells. CB-iPSCs and K-iPSCs retain residual epigenetic marks on a small number of genes, which nonetheless appear sufficient to skew differentiation potential toward the tissue of origin. Similarly, even in murine iPSCs characterized by the more stringent assays of pluripotency available in mice (e.g., blastocyst chimerism and germline transmissibility), we detected methylation signatures diagnostic of the tissue of origin and documented a preference to differentiate along the lineage of the donor cell¹. The epigenetic signatures of the tissue of origin that are retained in iPSCs reflect the technical limitations of reprogramming.

Although most differences among iPSCs, or between iPSCs and ESCs, represent random or stochastic differences, we observed in our experiments that residual epigenetic marks reflecting the tissue of origin can skew differentiation potential. Ultimately, refined methods of reprogramming may generate iPSCs that more closely approximate the epigenome of embryo-derived stem cells. Alternatively, more permissive differentiation conditions, either *in vitro* or *in vivo*, may overcome residual epigenetic barriers. Regardless, before clinical applications, the *in vivo* behavior of *in vitro*-differentiated cells derived from iPSCs of a wider range of donor tissues by various differentiation protocols and culture conditions will require substantial evaluation. A differentiation bias of existing iPSCs may be advantageous in certain research and therapeutic applications, especially as directed differentiation of pluripotent cells to tissues of interest remains a challenge for the field.

Accession codes. Gene expression microarray data and CHARM microarray data are deposited at Gene Expression Omnibus under accession number GSE27224.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

G.Q.D. was funded by US National Institutes of Health (NIH) grants DK70055 and DK59279, special funds received by the NIH under the American Recovery and Reinvestment Act (RC2-HL102815). K.K. was supported by NIH (K99HL093212-01), Leukemia and Lymphoma Society (3567-07), and Cooley's Anemia Foundation. A.P.F. was funded by NIH grants R37CA054358 and P50HG003233.

AUTHOR CONTRIBUTIONS

K.K., R.Z., K.N. and G.Q.D. conceived the experimental plan. K.K., R.Z., A.D., K.N., J.U., H.H., M.W.L., Y.-H.L. and H.L. performed the experiments. K.K., A.D., P.C. and M.J.A. performed data analysis. A.D., M.J.A. and A.P.F. performed CHARM and guided analysis of methylation. K.K., R.Z., A.D., K.N., J.U., P.C., J.J.C., M.W.L., A.P.F. and G.Q.D. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/nbt/index.html>.

Published online at <http://www.nature.com/nbt/index.html>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Kim, K. *et al. Nature* **467**, 285–290 (2010).
- Polo, J.M. *et al. Nat. Biotechnol.* **28**, 848–855 (2010).
- Metallo, M.C., Ji, L., de Pablo, J.J. & Palecek, S.P. *Methods Mol. Biol.* **585**, 83–92 (2010).
- Irizarry, R.A. *et al. Genome Res.* **18**, 780–790 (2008).
- Doi, A. *et al. Nat. Genet.* **41**, 1350–1353 (2009).
- Bhattacharya, B. *et al. BMC Dev. Biol.* **5**, 22 (2005).
- Ball, M.P. *et al. Nat. Biotechnol.* **27**, 361–368 (2009).
- Soshnikova, N., Montavon, T., Leleu, M., Galjart, N. & Duboule, D. *Dev. Cell* **19**, 819–830 (2010).
- Stadtfeld, M. *et al. Nature* **465**, 175–181 (2010).
- Lister, R. *et al. Nature* **471**, 68–73 (2011).
- Bock, C. *et al. Cell* **144**, 439–452 (2011).
- Boulting, G.L. *et al. Nat. Biotechnol.* **29**, 279–286 (2011).
- Ohi, Y. *et al. Nat. Cell Biol.* **13**, 541–549 (2011).