HHS Public Access

Author manuscript

*Nat Biotechnol*. Author manuscript; available in PMC 2012 June 01.

Published in final edited form as:

*Nat Biotechnol*. ; 29(12): 1117–1119. doi:10.1038/nbt.2052.

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

**Kitai Kim**[1,3,4,8,*], **Rui Zhao**[1,3,4,*], **Akiko Doi**[5,*], **Kitwa Ng**[1,4,*], **Juli Unternaehrer**[1,3,4], **Patrick Cahan**[1,3,4], **Huo Hongguang**[1,4], **Yuin-Han Loh**[1,3,4], **Martin J. Aryee**[6], **M. William Lensch**[1,3,4], **Hu Li**[7], **James J. Collins**[7], **Andrew P. Feinberg**[5,#], and **George Q. Daley**[1,2,3,4,#]

[1]Stem 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, MA 02115, USA

[2]Division of Hematology, Brigham and Women's Hospital, Boston, MA 02115, USA

[3]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

[4]Harvard Stem Cell Institute, Cambridge, MA 02138, USA

[5]Center for Epigenetics and Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MA, 21205, USA

[6]Oncology Department, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD 21205, USA

[7]Department of Biomedical Engineering and Center for BioDynamics, Boston University, Boston, MA 02215, USA; Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA; Howard Hughes Medical Institute

## Abstract

We compared *bona-fide* human induced pluripotent stem cells (iPSC) derived from umbilical cord blood (CB) and neonatal keratinocytes (K). As a consequence of both incomplete erasure of tissue-specific methylation and aberrant *de novo* methylation, CB-iPSC and K-iPSC are distinct in genome-wide DNA methylation profiles and differentiation potential. Extended passage of some



# To Whom Correspondence should be Addressed: George Q. Daley, Phone: (617) 919-2013, Fax: (617) 730-0222, george.daley@childrens.harvard.edu; Andrew P. Feinberg, Phone: (410) 614-3489, Fax: (410) 614-9819, afeinberg@jhu.edu.

[8]Current Address: Department of Cancer Biology and Genetics, Center for Cell Engineering, Sloan Kettering Institute, New York, NY 10065, USA

*These authors made equal contributions

**Author contribution.** 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.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.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. K.K., R.Z., A.D., and K.N. contributed equally.

**Accession codes.** Gene expression microarray data and CHARM microarray data are deposited at Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under accession number GSE27224.

iPSC clones in culture didn't improve their epigenetic resemblance to ESC, implying that some human iPSC 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 due 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 induced pluripotent stem cells (iPSC), which functions to confer “epigenetic memory”, biasing the differentiation potential of iPSC towards lineages related to the donor cell[1, 2]. Here, we investigate whether epigenetic memory persists in human iPSC.

We reprogrammed neonatal CD34+ cells from umbilical cord blood 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; Supplementary Fig. 1). The resultant iPSC from at least 2 independent cord blood and 3 keratinocyte donors passed stringent pluripotency tests characteristically applied to human embryonic stem cells (ESC; Supplementary Methods). We then tested the potential of the CB-iPSC and K-iPSC to differentiate into keratinocytes[3]. In day 6 EBs, K-iPSC displayed 9.4-fold higher expression of the keratin 14 gene, a marker of early keratinocyte differentiation (Fig. 1b, Supplementary Fig. 2a), and quantitatively, K-iPSC yielded 23-fold more keratinocytes than CB-iPSC (Fig. 1c, Supplementary Fig. 2b), indicating that K-iPSC show enhanced keratinocyte potential relative to CB-iPSC. Next, we differentiated multiple independent clones of CB-iPSC and K-iPSC in methylcellulose to test hematopoietic potential. Despite their reduced capacity for keratinocyte differentiation, differentiating cultures of CB-iPSC produced an expected range of myeloid colony types, whereas K-iPSC surveyed from multiple donors yielded relatively few hematopoietic colonies (Fig. 1d). Multiple clones of CB-iPSC consistently yielded a greater frequency of hematopoietic colonies than multiple clones of K-iPSC, and more colonies than iPSCs isolated from adult CD34+ blood and keratinocytes (Supplementary Fig. 2a, b, c). However, neither CB-iPSC nor K-iPSC exhibited 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 ESC using Comprehensive High-throughput Array-based Relative Methylation (CHARM) analysis, which interrogates 5.2 million CpG sites, including virtually all CpG islands and shores[4]. 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, 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-iPSC are easily distinguished from K-iPSC (Fig. 2b). ESC cluster with

both CB-iPSC and K-iPSC, suggesting that the separation between CB-iPSC and K-iPSC 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-iPSC and K-iPSC (Supplementary Methods). Taken together, these data indicate that the methylation patterns of CB-iPSC and K-iPSC are considerably different, and which 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 and keratinocytes include 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 CB and keratinocytes (1.73-fold more than expected by chance; $p<10^{-5}$). We also found that 27 of these 581 genes in or near DMRs that distinguish CB-iPSC and K-iPSC (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 data[6] are removed from the gene set (Supplementary Fig. 5a). Further, 17 of the 581 genes overlap DMRs that distinguish CB-iPSC from ESC, a 4-fold 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 10 DMRs that distinguish CB-iPSC from K-iPSC indicates that 5 are associated with hematopoiesis, and 4 with epithelial cell phenotypes (Supplementary Table 3). Second, K-iPSC maintained gene-body methylation, a phenomenon generally correlated with gene expression[7], for numerous keratinocyte-associated genes. Of 185 gene bodies hypermethylated in K-iPSC versus CB-iPSC, 14 correspond to tissue DMR-associated genes that are expressed in keratinocytes, a 3-fold enrichment over that predicted by chance (Fig. 2d). Conversely, of 19 gene bodies hypomethylated in CB-iPSC *vs.* ESC, four correspond to DMR-associated genes highly expressed in keratinocytes (P-value=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-iPSC and K-iPSC (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, keratinocyte, and ESC. 75 residual DMRs were shared with cord blood or keratinocytes, 28 DMRs were specific for ESC, and 267 DMRs were newly generated during reprogramming, reminiscent of our prior observation of frequent *de novo* DMRs in mouse iPSC of fibroblast cell origin from aged donors[1]. 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 pluripotent stem cell lines, epigenetic memory can be erased over time by extended culture[2] or ectopic gene expression[1]. 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 the ability of passaging to erase epigenetic memory does not appear applicable to all clones. *HOXD8* was one of 27 DMR-associated genes in the somatic cell types that was more highly methylated in keratinocytes and K-iPSC relative to cord blood and CB-iPSC, respectively (Supplementary Fig. 3a). Interestingly, 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* expression[8] (Supplementary Fig. 6d). Because *HOX* genes exert major effects on blood differentiation, we hypothesized that expression of *HOXD8* might be playing a role in inhibiting blood-forming potential in early passage K-iPSC in the clone N9. Consistent with a negative effect of *HOXD8* expression on blood development, we observed that ectopic expression of *HOXD8* in CB-iPSC reduced blood forming potential (Supplementary Fig. 6e), as did restoring high levels of *HOXD8* in later passages of a K-iPSC clone N9 (Supplementary Fig. 6f). In contrast, shRNA-mediated *HOXD8* knockdown in the K-iPSC clone G6 significantly 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 ESC (Supplementary Fig. 6h); rather, extended passage may have locus-specific effects that influence the differential potential of iPSC.

Prior studies comparing human iPSC to ESC have found that iPSC are differentially methylated[5, 7, 9] and exhibit significant reprogramming variability[10, 11, 12], and recently a transcriptional memory of somatic cells in human iPSC has been reported[13]. 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-iPSC and K-iPSC retain residual epigenetic marks on a small number of genes, which nonetheless appear sufficient to skew differentiation potential towards the tissue of origin. Similarly, even in murine iPSC characterized by the more stringent assays of pluripotency available in mice (e.g., blastocyst chimerism and germ line transmissibility), we detected methylation signatures diagnostic of the tissue of origin and documented a preference to differentiate along the lineage of the donor cell[1]. The epigenetic signatures of the tissue of origin that are retained in iPSC reflect the technical limitations of reprogramming. Although most differences among iPSC, or between iPSC and ESC, 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 iPSC 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, prior to 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 of iPSC, especially as directed differentiation of pluripotent cells to tissues of interest remains a challenge for the field.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

G.Q.D was funded by 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), LLS (3567-07), and Cooley's Anemia Foundation. A.P.F was funded by NIH grants R37CA054358 and P50HG003233.

## References

1. Kim K, et al. Epigenetic memory in induced pluripotent stem cells. Nature. 2010
2. Polo JM, et al. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nature biotechnology. 2010; 28:848–855.
3. Metallo MC, Ji L, de Pablo JJ, Palecek SP. Directed Differentiation of Human Embryonic Stem Cells to Epidermal Progenitors. Epidermal Cells, Methods in Molecular Biology. 2009; 585:83–92. [PubMed: 19907998]
4. Irizarry RA, et al. Comprehensive high-throughput arrays for relative methylation (CHARM). Genome research. 2008; 18:780–790. [PubMed: 18316654]
5. Doi A, et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat Genet. 2009
6. Bhattacharya B, et al. Comparison of the gene expression profile of undifferentiated human embryonic stem cell lines and differentiating embryoid bodies. BMC Dev Biol. 2005; 5:22. [PubMed: 16207381]
7. Ball MP, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nature biotechnology. 2009; 27:361–368.
8. Soshnikova N, Montavon T, Leleu M, Galjart N, Duboule D. Functional analysis of CTCF during mammalian limb development. Developmental cell. 2010; 19:819–830. [PubMed: 21145498]
9. Stadtfeld M, et al. Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature. 2010; 465:175–181. [PubMed: 20418860]
10. Lister R, et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature. 2011 Published online on Feb 3, 2011.
11. Bock C, et al. Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell. 2011; 144:439–452. [PubMed: 21295703]
12. Boulting GL, et al. A functionally characterized test set of human induced pluripotent stem cells. Nature biotechnology. 2011; 29:279–286.
13. Ohi Y, et al. Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. Nat Cell Biol. 2011

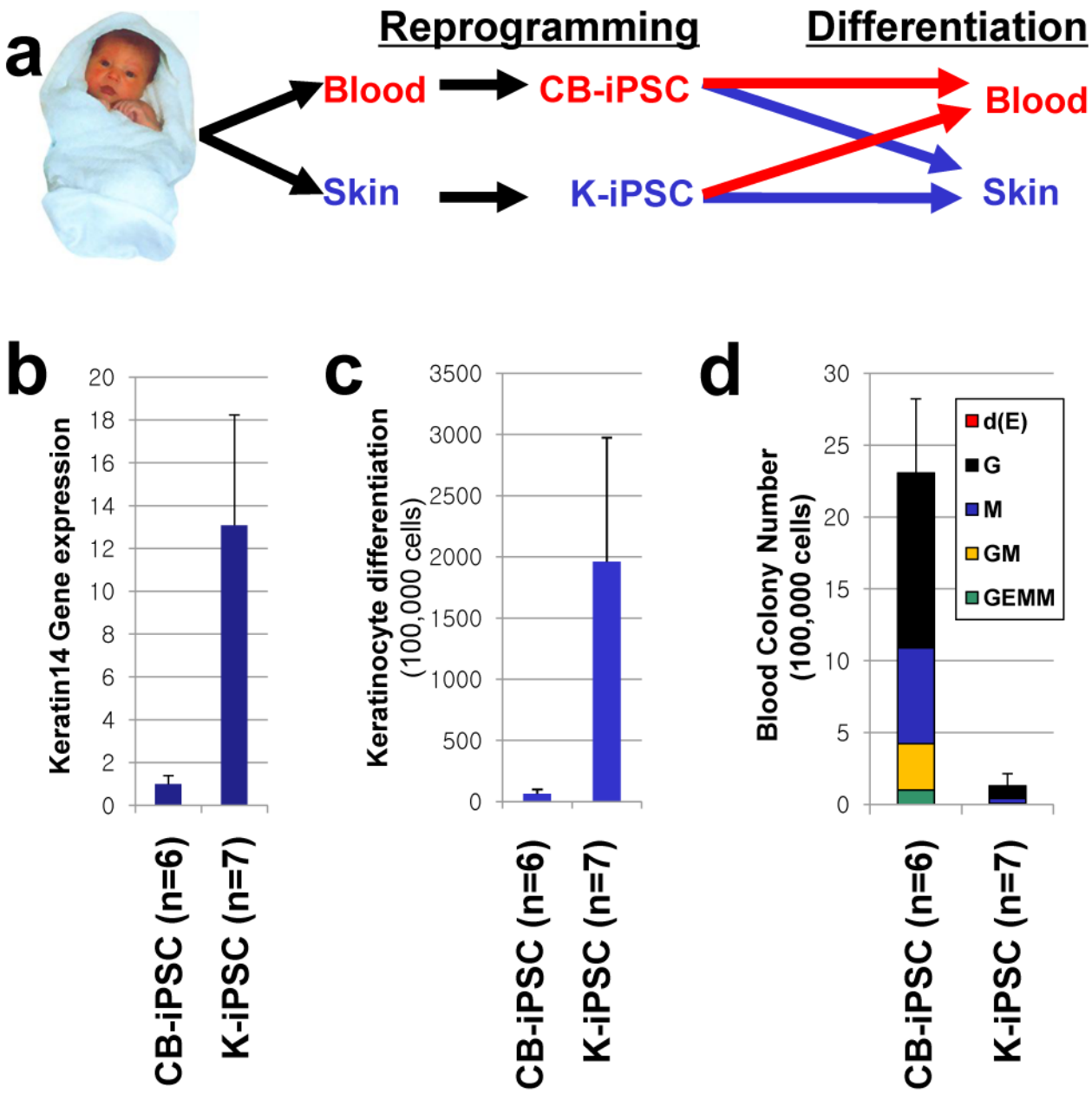
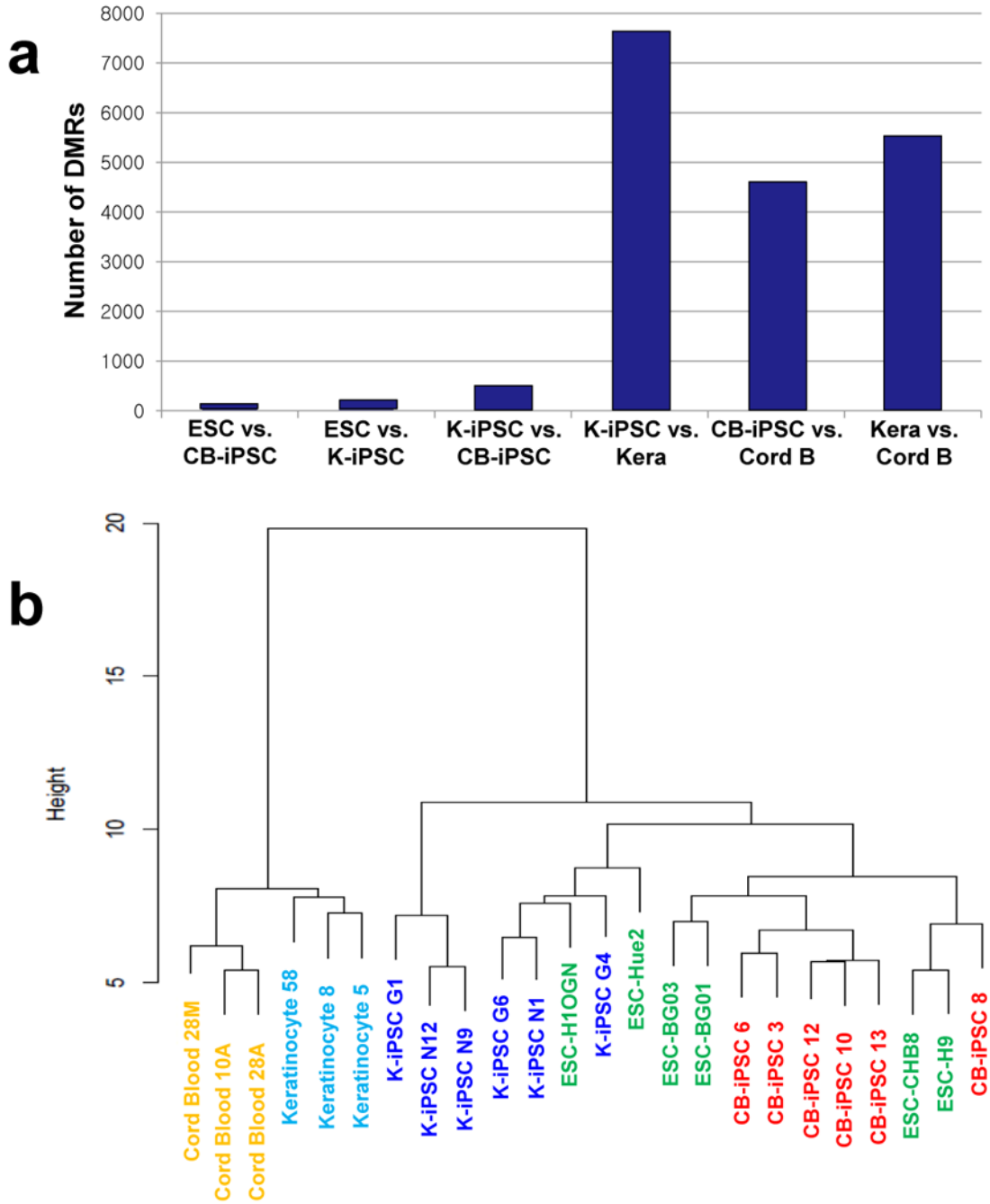



**Figure 1. Derivation and differentiation of iPSC from neonatal umbilical cord blood and foreskin fibroblasts**

(a) Experimental schema. (b) Q-PCR of the keratinocyte marker K14 in D6 EBs from CB-iPSC (n=6) and K-iPSC (n=7). Gene expression was normalized to Actin, and shown as fold difference relative to CB-iPSC. (c) Numbers of keratinocytes differentiated from CB-iPSC (n=6) and K-iPSC (n=7). (d) Numbers of hematopoietic colony forming cells in D14 EBs differentiated from CB-iPSC (n=6) and K-iPSC (n=7). Error bar = s.d.



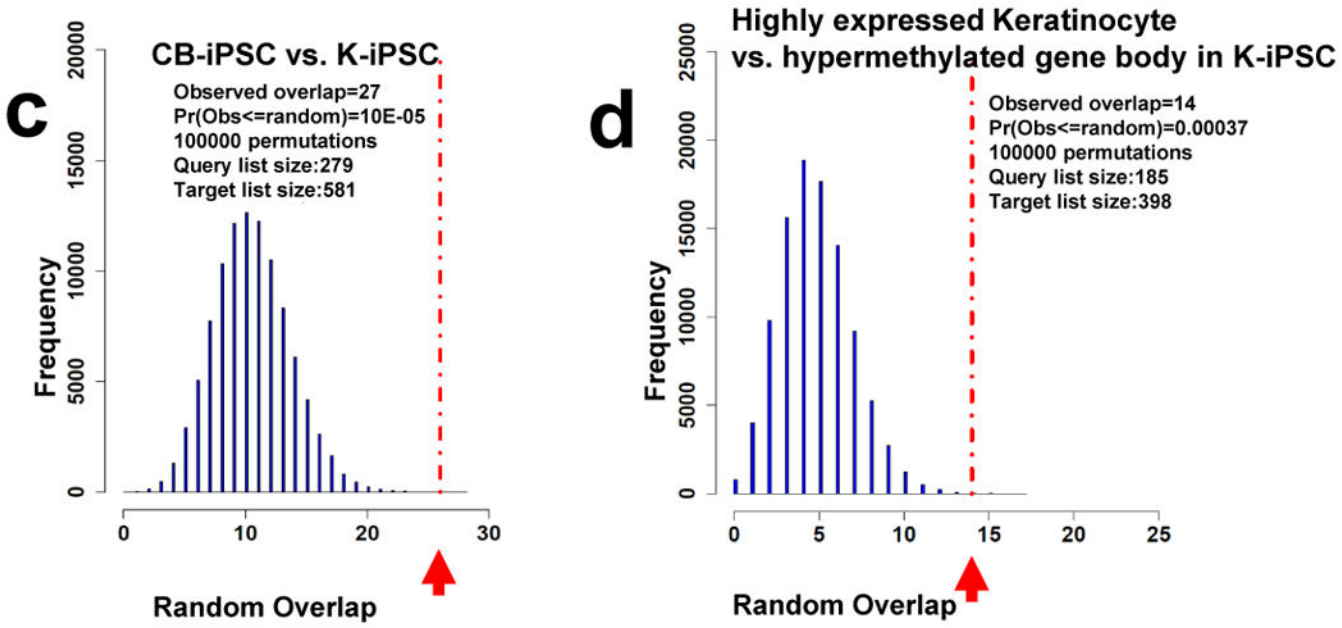



**Figure 2. Analysis of methylation in CB-iPSC, K-iPSC, ESC, and somatic cells**

(a) Numbers of differentially methylated regions (DMRs) between CB-iPSC, K-iPSC, ESC, umbilical cord blood, and cultured keratinocytes. DMRs were defined by an area cutoff of 2.0. (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-iPSC and K-iPSC are enriched in DMR-associated genes (genes both differentially expressed and methylated between cord blood and keratinocytes). (d) Genes highly expressed in keratinocytes are enriched in DMRs that are both hypermethylated in K-iPSC relative to CB-iPSC and are located in gene bodies rather than promoters.