DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Cell Type of Origin Influences the Molecular and Functional Properties of Mouse Induced Pluripotent Stem Cells

Citation
Polo, Jose M., Susanna Liu, Maria Eugenia Figueroa, Warakorn Kulalert, Sarah Eminli, Kah Yong Tan, Effie Apostolou, et al. 2010. Cell Type of Origin Influences the Molecular and Functional Properties of Mouse Induced Pluripotent Stem Cells. Nature Biotechnology 28, no. 8: 848–855.

Published Version
doi:10.1038/nbt.1667

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:12724030

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Cell of origin influences molecular and functional properties of murine induced pluripotent stem cells.

Jose M. Polo[1-4], Susanna Liu[5], Maria Eugenia Figueroa[6], Warakorn Kulalert[1-4], Sarah Eminli[1-4], Kah Yong Tan[1,4,8], Effie Apostolou[1-4], Matthias Stadtfeld[1-4], Yushan Li[6], Toshi Shioda[2], Sridaran Natesan[7], Amy J. Wagers[1,4,8], Ari Melnick[6], Todd Evans[5] and Konrad Hochedlinger[1-4, #]

[1] Howard Hughes Medical Institute and Department of Stem Cell and Regenerative Biology, Harvard University and Harvard Medical School, 7 Divinity Avenue, Cambridge, MA 02138, USA

[2] Massachusetts General Hospital Cancer Center, 149 13th Street, Charlestown, MA 02129, USA

[3] Massachusetts General Hospital Center for Regenerative Medicine, 185 Cambridge Street, Boston, MA 02114, USA

[4] Harvard Stem Cell Institute, 42 Church Street, Cambridge, MA 02138, USA

[5] Department of Surgery, Weill Cornell Medical College, New York, New York, NY 10065, USA.

[6] Department of Medicine, Hematology Oncology Division, Weill Cornell Medical College, New York, NY 10065, USA.

[7] Sanofi-Aventis Cambridge Genomics Center, 26 Landsdowne Street, Cambridge, Massachussetts 02139, USA

[8] Joslin Diabetes Center, 1 Joslin Place, Boston, MA 02215, USA

corresponding author (e-mail: khochedlinger@helix.mgh.harvard.edu)

Abstract

Induced pluripotent stem cells (iPSCs) have been derived from various somatic cell populations through ectopic expression of defined factors[1-10]. It remains unclear whether iPSCs generated from different cell types are molecularly and functionally similar. Here, we show that iPSCs obtained from fibroblasts, hematopoietic and myogenic cells exhibit distinct transcriptional and epigenetic patterns. Moreover, we demonstrate that cellular origin influences the *in vitro* differentiation potentials of iPSCs into embryoid bodies (EBs) and different hematopoietic cells. Importantly, continuous passaging of iPSCs largely attenuates these differences. Our results suggest that low-passage iPSCs retain a transient epigenetic memory of their somatic cells of origin, which manifests as differential gene expression and altered differentiation capacity. These observations might affect ongoing attempts to use iPSCs for disease modeling and also could be exploited for potential therapeutic applications to enhance differentiation into desired cell lineages.

iPSCs are usually obtained from fibroblasts after infection with viral constructs expressing the four transcription factors Oct4, Sox2, Klf4 and cMyc[8, 9]. In addition, other cell types, including blood[2, 4, 11], stomach and liver cells[1], keratinocytes[12, 13], melanocytes[14], pancreatic β cells[7] and neural progenitors[3, 15-17] have been reprogrammed into iPSCs. While these iPSC lines have been shown to be pluripotent and transcriptionally and epigenetically highly similar, recent studies detected substantial molecular and functional differences among iPSCs derived from distinctive cell types. For example, Yamanaka and colleagues showed that iPSCs produced from various fibroblasts, stomach and liver cells exhibit different propensities to form tumors in mice although the underlying molecular mechanisms remain elusive[18]. Another study identified persistent donor cell-specific gene expression patterns in iPSCs produced from different cell types, suggesting an influence of the somatic cell of origin on the molecular properties of resultant iPSCs[19]. Whether cellular origin also affected

the functional properties of iPSCs remained unexplored in that report. Of note, some of these studies may be confounded by the presence of different viral insertions in individual iPSC lines and by the fact that the analyzed iPSC lines were of different genetic background, which can affect both gene expression patterns[20] and the functionality[9, 21] of cells. Indeed, we have recently shown that many mouse iPSC lines derived from fibroblasts show aberrant silencing of a surprisingly small set of transcripts compared with genetically matched ESCs[22]. However, our study did not investigate whether additional cell of origin-specific differences may exist in iPSC lines derived from different cell types.

Patient-specific iPSCs are a valuable tool for the study and possibly treatment of numerous diseases[20, 23-26]. Thus, resolving the question of whether iPSCs produced from different cell types are molecularly and functionally equivalent is crucial for using these cells to model disease, which entails detecting subtle differences in the differentiation potentials or survival of patient-derived iPSCs[24, 27]. Furthermore, the identification of somatic cells that influence the differentiation capacities of resultant iPSCs into desired cell lineages could be useful in a therapeutic setting.

In order to assess if iPSCs derived from different somatic cell types are distinguishable, we compared here the transcriptional and epigenetic patterns, as well as the *in vitro* differentiation potentials, of iPSCs produced from four genetically identical adult mouse cell types that differed only in the lineage from which they were derived.

Results

Genetically matched iPSCs derived from different cell types.

Because the genetic background of ESCs can influence their transcriptional and functional behaviors, we employed a previously described “secondary system” to generate genetically identical iPSCs[2, 28] (Figure 1a). Briefly, iPSCs were generated from somatic cells using doxycycline-inducible lentiviruses expressing Oct4, Sox2, Klf4 and cMyc[29], and then clonally injected into blastocysts to produce isogenic chimeric mice. Thus, isolation of different cell types from these

chimeras and their subsequent exposure to doxycycline gave rise to iPSCs with the exact genetic makeup. In this study, we focused on iPSCs derived from tail tip-derived fibroblasts (TTFs), splenic B cells (B), bone marrow-derived granulocytes (Gra) and skeletal muscle precursors (SMPs)[30], which were continuously cultured for 2-3 weeks (passage 4 to 6) after picking. The pluripotency of some of these cell lines has been previously documented[2], or was analyzed in this study (Supplementary Table 1 and Supplementary Figure 1). Briefly, all cell lines grew at similar rates and independently of viral transgene expression (Supplementary Figure 2), upregulated the endogenous pluripotency genes *Nanog*, *Sox2* and *Oct4*, indicating successful epigenetic and transcriptional reprogramming (Supplementary Table 1). Moreover, all lines gave rise to differentiated teratomas and all tested lines supported the development of chimeric animals upon blastocyst injection, demonstrating their pluripotency (Supplementary Table 1). We therefore conclude that the cell lines analyzed here qualify as *bona fide* iPSC lines.

iPSCs produced from different cell types are transcriptionally distinguishable.

We first evaluated if iPSCs derived from defined somatic cell types retain gene expression patterns indicative of their cells of origin. Specifically, we assessed the expression of cell lineage-specific candidate genes in iPSCs derived from granulocytes (Gra-iPSCs) and SMPs (SMP-iPSCs). As expected, the SMP markers *Cxcr4* and *Integrin B1* and the granulocyte markers *lysozyme* and *Gr-1* were expressed at significantly higher levels in the somatic cells of origin than in resultant iPSCs (Supplementary Figure 3). Moreover, SMP-iPSCs expressed substantially higher levels of *Cxcr4* and *Itgb1* than Gra-iPSCs (Figure 1b) and Gra-iPSCs showed higher expression levels of *lysozyme* and *Gr-1* compared with SMP-iPSCs (Figure 1b). Together, these data suggest that iPSCs retain a transcriptional memory of their somatic cell of origin.

To test this notion globally, we compared and contrasted the transcriptional profiles of iPSC lines originating from SMPs (n=3) with those

derived from granulocytes (n=3), as well as expression profiles of iPSC lines originating from B cells (n=3) with those produced from TTFs (n=3). Of note, iPSCs were only compared with each other if they originated from the same chimeric mouse (B-iPSCs vs. TTF-iPSCs and Gra-iPSCs vs. SMP-iPSCs) (Figure 1a) in order to eliminate potential variability between different experiments and individual animals. All iPSC lines analyzed were between passage (p) 4 and 6. 1388 genes were differentially expressed (2 fold, corrected p=0.05) between SMP-iPSCs and Gra-iPSCs and 1090 genes between TTF-iPSCs and B-iPSCs (Supplementary Table 2). An analysis of the one hundred most differentially expressed genes across all samples indicated that iPSCs with the same cell of origin clustered together (Figure 1c). Consistent with this observation, unsupervised hierarchical clustering (Figure 1d) as well as principal component analysis (Supplementary Figure 4) of all genes, separated SMP-iPSCs and Gra-iPSCs, as well as TTF-iPSCs and B-iPSCs into different groups according to their cells of origin. Interestingly, Gene Ontology (GO) analysis of the one hundred most differentially expressed genes between SMP-iPSCs and Gra-iPSCs indicated an enrichment for genes belonging to the categories “myofibril” (7.6-fold enrichment), “contractile fiber” (7.3-fold enrichment) and “muscle development” (5.9-fold enrichment) as well as “B cell activation” (6.8-fold enrichment) and “leukocyte activation” (3.7-fold enrichment). Together, these results show that genetically identical iPSCs obtained from four different somatic cell types are distinguishable from each other using genome-wide transcriptional analyses, further supporting the notion that the donor cell type influences the overall gene expression pattern of resultant iPSCs.

In order to determine the effect on gene expression patterns of deriving iPSCs from different animals in independent experiments, we next compared the expression profiles of Gra-iPSCs derived from chimera #1 (n=3) with additional Gra-iPSCs derived from chimera #2 (n=3), as well as with SMP-iPSCs produced from chimera #1 and TTF-iPSCs produced from chimera #2 (see Figure 1a). Hierarchical clustering separated Gra-iPSCs according to their origin from different animals, suggesting a significant contribution of this experimental

variable to gene expression patterns (Supplementary Figure 5). However, when adding the expression data from TTF-iPSCs and SMP-iPSCs into the analysis, we found that differences due to cell of origin were stronger than differences caused by variations in experimental conditions or animals. These data reinforce the observation that iPSCs derived from different somatic cell types are transcriptionally discernible, even when they originate from different animals.

To exclude the possibility that the observed gene expression differences were due to the specific secondary system used, we derived iPSCs from SMPs, granulocytes, B cells and peritoneal fibroblasts (PF) from reprogrammable mice[31] under identical culture conditions. Analysis of gene expression profiles of these lines at p4 showed clustering according to their cells of origin, with the exception of PF-derived iPSCs, which may be a consequence of the heterogeneity of the starting population. Collectively, these results corroborate the notion that iPSCs generated form different cell types exhibit distinct transcriptional patterns (Supplementary Figure 6).

iPSCs derived from different cell types exhibit discernible epigenetic patterns.

We next asked if the differential gene expression patterns we observed correlate with differences in epigenetic marks. To this end, we performed a genome-wide, restriction enzyme-based methylation analysis termed “HpaII tiny fragment Enrichment by Ligation-mediated PCR” (HELP) on the same samples we used for expression analysis. Unsupervised hierarchical clustering showed that Gra-iPSCs and SMP-iPSCs as well as B-iPSCs and TTF-iPSCs, which clustered separately in the transcriptional assays, were also distinguishable based on their methylation patterns (Figure 2a). Correspondence analysis of the same samples corroborated this finding (Figure 2b), indicating that the donor cell type not only affects the overall transcriptional pattern, but also, the promoter methylation pattern of resultant iPSCs.

Despite the separation of Gra-iPSCs from SMP-iPSCs and of TTF-iPSCs from B-iPSCs (Figure 2a, b) by hierarchical clustering, we detected few loci that

were differentially methylated with statistical significance using supervised analysis (69 genes between Gra-iPSCs and SMP-iPSCs and 0 genes between B-iPSCs and TTF-iPSCs), (Supplementary Table 3). To confirm this, we interrogated the DNA methylation status at the promoter regions of the previously analyzed markers *Cxcr4, Itgb1, Lysozyme* and *Gr-1* (Figure 1b) by employing EpiTYPER DNA methylation analyses, which quantifies gene-specific CpG methylation. As suggested by the HELP analysis, we failed to detect differences in the methylation levels of these candidate genes between SMP-iPSCs and Gra-iPSCs (Figure 2c), indicating that methylation differences are more subtle than the observed gene expression differences and raising the possibility that other chromatin marks may be responsible for the observed expression differences.

Indeed, we observed high levels of the activating marks H3Ac and H3K4me3 and low levels of the repressive marks H3K27me3 at the promoters of *Cxcr4* and *Itgb1* in SMPs and at the promoters of *lysozyme* and *Gr-1* in granulocytes, respectively, consistent with their abundant expression in these cell types (Figure 2d). Interestingly, SMP-iPSCs, which showed higher expression levels of *Cxcr4* and *Itgb1* than Gra-iPSCs (Figure 1b), were enriched for H3K4me3 compared with Gra-iPSCs at these two genes. A similar pattern was observed for the granulocyte specific genes in Gra-iPSCs compared with SMP-iPSCs, with *Gr-1* and *lysozyme* being elevated for H3K4me3 (Figure 2d). These data show that the observed expression differences among iPSCs derived from different cell types may be predominantly the consequence of differences in histone marks, further suggesting that iPSCs retain an epigenetic memory of their cells of origin.

iPSCs derived from different cell types have distinctive *in vitro* differentiation potentials.

Because the gene expression differences we observed among different iPSC lines affected genes known to be involved in the lineage-specific differentiation and function of the somatic cell types from which they were derived, we reasoned that these differences might affect their capacity to differentiate into defined cell

lineages. Thus, we evaluated the autonomous potential of the four types of iPSC lines by assessing their abilities to produce EBs, and subsequently erythrocyte progenitors, macrophages and mixed hematopoietic colonies, using established semi-quantitative differentiation protocols (Figure 3a). Most notably, TTF-iPSCs produced significantly smaller and fewer EBs compared with all the other iPSC lines (Figure 3b, c). Moreover, the EBs derived from TTF-iPSC generated relatively few erythrocyte, macrophage and mixed colony progenitors compared with B-iPSCs derived from the same animal, indicating striking differences in the differentiation potentials of these iPSCs (Figure 3d-g). In contrast, SMP-iPSCs and Gra-iPSCs showed equivalent abilities to produce EBs (Figure 3d-g). However, Gra-iPSCs gave rise to erythrocyte, macrophages and mixed colonies at higher efficiencies than SMP-iPSCs, suggesting a pattern of differentiation that reflects their cells of origin. Together, these data show that the cell type of origin may bias the differentiation potential of resultant iPSC lines.

Continuous passaging of iPSCs abrogates transcriptional, epigenetic and functional differences.

Previously published data suggest that early passage human iPSCs derived from fibroblasts are transcriptionally distinct from late passage iPSCs[32]. However, the effect of passaging on the functionality of iPSC was not examined in that study. We therefore wondered if continuous passaging of the various iPSC lines would eliminate the observed differences in gene expression and differentiation potential. For this analysis, we added to the B-iPSC/TTF-iPSC group, studied before, a new set of T cell- and granulocyte-derived iPSCs, which were all derived from chimera #2. These 12 iPSC lines were subjected to several additional rounds of passaging under identical culture conditions, and RNA was harvested at p10 and p16 for expression profiling. While unsupervised hierarchical clustering of these cell lines at low passage (p4) clearly separated each of the different iPSC lines according to their cells of origin (Figure 4a, left panel), unsupervised clustering of these iPSC lines at p10 showed that B-iPSC, TTF-iPSC and T-iPSC were indistinguishable from each other, whereas the Gra-

iPSC still clustered together (Figure 4a, middle panel). Further passaging of these cells until p16 entirely eliminated these differences (Figure 4a, right panel). Together, these data indicate that continuous cell division resolves transcriptional differences among iPSC lines. Consistent with this observation, the total number of differentially expressed genes between various pairs of iPSC lines derived from different cellular origins was reduced from around 500-2,000 in low passage cultures to only about 50 or even 0 in high passage cultures, further demonstrating that after extensive *in vitro* propagation, these iPSC lines have become very similar to each other (Figure 4b).

Analysis of the genes that changed upon passaging from p4 to p16 in Gra-iPSC, B-iPSC, TTF-iPSC showed 25% overlap with at least one of the other two groups of iPSC lines, suggesting that iPSCs seem to undergo some common changes during passaging, irrespective of their cell of origin (Figure 4c). GO analysis of these changes indicated a strong enrichment for developmental regulators. Moreover, the only GO cluster common to all three groups was ”organ development”, indicating that the passaging of iPSCs results in a change of differentiation-associated gene expression patterns (Figure 4c). The expression levels of the pluripotency genes Sox2 and Oct4, which are high already at low passage (Table 1), increased even further during the passaging process, supporting the notion that the pluripotency network becomes increasingly solidified during culture (Supplementary Figure 7), consistent with a previous report showing gradual upregulation of pluripotency-associated genes upon passaging of human iPSC lines[32].

To evaluate if the passaging of iPSCs attenuates the observed epigenetic differences, we performed HELP analysis on B-iPSCs and TTF-iPSCs at high passage. In contrast to low passage iPSCs, hierarchical unsupervised clustering analysis was no longer able to separate the iPSCs based on their cells of origin (Figure 4d). Accordingly, the methylation levels of histones at candidate genes became indistinguishable between Gra-iPSCs and SMP-iPSCs (Supplementary Figure 8). Notably, several of the analyzed loci showed an enrichment for both H3K4me3 and H3K27me3, indicative of bivalent domains that are characteristic

of pluripotent stem cells[33]. Thus, continuous passaging leads to an equilibration of the epigenetic differences detected in low-passage iPSCs.

Two possible mechanisms could account for the observed loss of epigenetic and transcriptional memory with increased passage number, (i) passive replication-dependent loss of somatic marks in the majority of iPSCs or (ii) selection of rare pre-existing fully reprogrammed cells over time. Since the selection model predicts that these rare clones have a growth or survival advantage, we would expect to see an impairment in growth rates of bulk iPSC cultures at low passage compared with iPSC cultures at high passage, which we did not observe (Supplementary Figure 9a). We also failed to detect significant differences when examining the growth rates of single cell clones established from low and high passage iPSC lines by using a colorimetric assay (XTT assay) that detects metabolic activity (Supplementary Figure 10) as well as by measuring the increase in cell numbers on three consecutive days (Supplementary Figure 11 and 12). Likewise, an analysis of the colony formation efficiency of single cell sorted iPSCs at low and high passage did not yield detectable differences (Supplementary Figure 9b). Collectively, these data argue against the presence of rare subclones that become selected over time but are rather consistent with the notion that all iPSC lines gradually resolve transcriptional and epigenetic differences with increased passaging. However, our results do not exclude a combined model involving passive resolution of epigenetic marks as well as selection of multiple clones.

Lastly, we sought to test if the similar transcriptional and epigenetic patterns of high passage iPSCs derived from distinct cells of origin would translate into an equalization of their differentiation potentials. We first performed an EB formation assays at different passages for TTF-iPSCs and B-iPSCs, which showed a strong difference at low passage. TTF-iPSC gave rise to similarly-sized EBs as B-iPSC around p10-p12 (Supplementary Figure 13a,b) and were indistinguishable at p16 (Supplementary Figure 13c,d). Moreover, EBs derived from TTF-iPSCs and B-iPSCs at p16 differentiated into comparable numbers of erythrocyte (Figure 4e), macrophage (Figure 4f) and mixed colony progenitors

(Figure 4g), thus proving that continuous cellular passaging eliminates differences in the differentiation potentials of these iPSCs.

Discussion.

Our study shows that genetically matched iPSCs retain a transient transcriptional and epigenetic memory of their cell of origin at low passage, which can substantially affect their differentiation capacities into EBs and different hematopoietic cell types (Figure 5). These molecular and functional differences are lost upon continuous passaging, however, indicating that complete reprogramming is a gradual process that continues beyond the acquisition of a *bona fide* iPSC state as measured by the activation of endogenous pluripotency genes, viral transgene-independent growth and the ability of iPSCs to differentiate into cell types of all three germlayers. Of note, the low passage iPSCs described here are different from "partially reprogrammed iPSCs"[34, 35], which depend on the continuous expression of viral transgenes and fail to activate and demethylate pluripotency genes or contribute to the formation of viable chimeras (Figure 5).

The mechanism by which passaging eliminates the molecular and functional differences between iPSCs of different origins remains to be determined. Three key observations argue against the possibility of selective expansion of a rare subset of fully-resolved iPSCs: (i) both low passage and high passage iPSCs had similar proliferation rates; (ii) there was little variability in the growth rate of single cell iPSC clones; and (iii) the number of passages required to resolve cell-of-origin differences was dependent upon the starting cell type. These observations thus suggest that the consolidation of the pluripotent transcriptional network upon passaging is a slow process, potentially facilitated by a positive feedback mechanism that gradually resolves the residual cell of origin-specific epigenetic marks and transcriptional patterns. In accordance with this idea is the finding that telomeres become gradually elongated with increased passage number of iPSCs[36]. Our results are also consistent with the previous observation that cloned embryos often retain donor cell-specific transcriptional

patterns and fail to efficiently activate embryonic genes over many cell divisions [37-40], suggesting possible similarities in the mechanisms of reprogramming by nuclear transfer and induced pluripotency.

The present results may help to explain some of the previously reported differences between ESCs and iPSCs[41, 42]. Most of these studies compared high passage ESC lines with iPSC lines of undefined, but presumably lower, passage that may not yet have reached an ESC-equivalent ground state. It should be informative to revisit these studies with genetically matched, transgene-free high passage iPSCs to see if this abrogates the seen gene expression and differentiation differences.

The observed trend of early-passage iPSC lines to differentiate preferentially into the cell lineage of origin could be used in potential clinical settings to produce certain somatic cell types that have thus far been difficult to obtain from ESCs. However, these data also serve as a cautionary note for ongoing attempts to recapitulate disease phenotypes *in vitro* using patient-specific low-passage iPSC lines, as the continued epigenetic, transcriptional and functional "maturation" of these iPSCs could confound the data obtained from these cells. Further elucidation of the molecular indicators of fully reprogrammed iPSCs should help in the establishment of standardized iPSC lines that can be compared with confidence in ongoing basic biological and drug discovery approaches.

Acknowledgments

We thank Nimet Maherali and Ryan Walsh for helpful suggestions and critical reading of the manuscript, Ben Wittner for statistical advice, J. LaVecchio G. Buruzula, Kat Folz-Donahue and Laura Prickett for expert cell sorting and Kathryn Coser for technical assistance. J.P. was supported by an MGH ECOR fellowship, E.A. by a Jane Coffin Childs fellowship, M.S. by a Schering fellowship and K.Y.T by the Agency of Science, Technology and Research Singapore. Support to A.M. was from the Lymphoma Society, SCOR#7132-08; to T.E. from NIH grant HL056182 and NYSTEM; to A.J.W. in part from the Burroughs-Welcome Fund, Harvard Stem Cell Institute, Peabody Foundation, and NIH 1 DP2 OD004345-01, and the Joslin Diabetes Center DERC (P30DK036836); to K.H. from HHMI, the NIH Director's Innovator Award and the Harvard Stem Cell Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. Aoi, T. et al. Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science* **321**, 699-702 (2008).
2. Eminli, S. et al. Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. *Nat Genet* **41**, 968-976 (2009).
3. Eminli, S., Utikal, J., Arnold, K., Jaenisch, R. & Hochedlinger, K. Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression. *Stem Cells* **26**, 2467-2474 (2008).
4. Hanna, J. et al. Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133**, 250-264 (2008).
5. Lowry, W.E. et al. Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc Natl Acad Sci U S A* **105**, 2883-2888 (2008).
6. Park, I.H. et al. Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141-146 (2008).
7. Stadtfeld, M., Brennand, K. & Hochedlinger, K. Reprogramming of pancreatic beta cells into induced pluripotent stem cells. *Curr Biol* **18**, 890-894 (2008).
8. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872 (2007).
9. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).
10. Yu, J. et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-1920 (2007).
11. Loh, Y.H. et al. Generation of induced pluripotent stem cells from human blood. *Blood* **113**, 5476-5479 (2009).
12. Aasen, T. et al. Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. *Nat Biotechnol* **26**, 1276-1284 (2008).
13. Maherali, N. et al. A high-efficiency system for the generation and study of human induced pluripotent stem cells. *Cell Stem Cell* **3**, 340-345 (2008).
14. Utikal, J., Maherali, N., Kulalert, W. & Hochedlinger, K. Sox2 is dispensable for the reprogramming of melanocytes and melanoma cells into induced pluripotent stem cells. *J Cell Sci* **122**, 3502-3510 (2009).
15. Kim, J.B. et al. Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. *Nature* **454**, 646-650 (2008).
16. Shi, Y. et al. A combined chemical and genetic approach for the generation of induced pluripotent stem cells. *Cell Stem Cell* **2**, 525-528 (2008).
17. Silva, J. et al. Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol* **6**, e253 (2008).
18. Miura, K. et al. Variation in the safety of induced pluripotent stem cell lines. *Nat Biotechnol* **27**, 743-745 (2009).
19. Ghosh, Z. et al. Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human

embryonic stem cells. *PloS one* **5**, e8975.
20. Soldner, F. et al. Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell* **136**, 964-977 (2009).
21. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313-317 (2007).
22. Stadtfeld, M. et al. Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature* **465**, 175-181.
23. Dimos, J.T. et al. Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science* **321**, 1218-1221 (2008).
24. Ebert, A.D. et al. Induced pluripotent stem cells from a spinal muscular atrophy patient. *Nature* **doi:10.1038/nature07677** (2009).
25. Park, I.H. et al. Disease-specific induced pluripotent stem cells. *Cell* **134**, 877-886 (2008).
26. Saha, K. & Jaenisch, R. Technical challenges in using human induced pluripotent stem cells to model disease. *Cell Stem Cell* **5**, 584-595 (2009).
27. Lee, G. et al. Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature* **461**, 402-406 (2009).
28. Wernig, M. et al. A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nat Biotechnol* **26**, 916-924 (2008).
29. Stadtfeld, M., Maherali, N., Breault, D.T. & Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2**, 230-240 (2008).
30. Cerletti, M. et al. Highly efficient, functional engraftment of skeletal muscle stem cells in dystrophic muscles. *Cell* **134**, 37-47 (2008).
31. Stadtfeld, M., Maherali, N., Borkent, M. & Hochedlinger, K. A reprogrammable mouse strain from gene-targeted embryonic stem cells. *Nat Methods* **7**, 53-55.
32. Chin, M.H. et al. Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* **5**, 111-123 (2009).
33. Bernstein, B.E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326 (2006).
34. Mikkelsen, T.S. et al. Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49-55 (2008).
35. Sridharan, R. et al. Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136**, 364-377 (2009).
36. Marion, R.M. et al. Telomeres acquire embryonic stem cell characteristics in induced pluripotent stem cells. *Cell Stem Cell* **4**, 141-154 (2009).
37. Boiani, M., Eckardt, S., Scholer, H.R. & McLaughlin, K.J. Oct4 distribution and level in mouse clones: consequences for pluripotency. *Genes Dev* **16**, 1209-1219. (2002).
38. Bortvin, A. et al. Incomplete reactivation of Oct4-related genes in mouse embryos cloned from somatic nuclei. *Development* **130**, 1673-1680 (2003).

39. Ng, R.K. & Gurdon, J.B. Epigenetic memory of active gene transcription is inherited through somatic cell nuclear transfer. *Proc Natl Acad Sci U S A* **102**, 1957-1962 (2005).
40. Ng, R.K. & Gurdon, J.B. Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. *Nat Cell Biol* **10**, 102-109 (2008).
41. Feng, Q. et al. Hemangioblastic Derivatives from Human Induced Pluripotent Stem Cells Exhibit Limited Expansion and Early Senescence. *Stem Cells*.
42. Hu, B.Y. et al. Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *Proc Natl Acad Sci U S A* **107**, 4335-4340.
43. Conboy, I.M., Conboy, M.J., Smythe, G.M. & Rando, T.A. Notch-mediated restoration of regenerative potential to aged muscle. *Science* **302**, 1575-1577 (2003).
44. Sherwood, R.I. et al. Isolation of adult mouse myogenic progenitors: functional heterogeneity of cells within and engrafting skeletal muscle. *Cell* **119**, 543-554 (2004).
45. Cheshier, S.H., Morrison, S.J., Liao, X. & Weissman, I.L. In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proc Natl Acad Sci U S A* **96**, 3120-3125 (1999).
46. Figueroa, M.E., Melnick, A. & Greally, J.M. Genome-wide determination of DNA methylation by Hpa II tiny fragment enrichment by ligation-mediated PCR (HELP) for the study of acute leukemias. *Methods Mol Biol* **538**, 395-407 (2009).
47. Selzer, R.R. et al. Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosomes Cancer* **44**, 305-319 (2005).
48. Thompson, R.F. et al. An analytical pipeline for genomic representations used for cytosine methylation studies. *Bioinformatics* **24**, 1161-1167 (2008).
49. Edgar, R., Domrachev, M. & Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207-210 (2002).
50. Culhane, A.C., Thioulouse, J., Perriere, G. & Higgins, D.G. MADE4: an R package for multivariate analysis of gene expression data. *Bioinformatics* **21**, 2789-2790 (2005).
51. Ehrich, M. et al. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci U S A* **102**, 15785-15790 (2005).

Figure legends

Figure 1. iPSCs derived from different cell types are transcriptionally distinguishable.

(a) Flow-chart explaining the derivation and analysis of genetically matched iPSCs from different cell types. Secondary iPSCs were first injected into blastocyts to generate chimeric mice, from which the indicated somatic cell types were isolated. Exposure of these cells to doxycycline (dox) then gave rise to iPSCs. **(b)** Quantification of the expression levels of *Cxcr4, Itgb1, GR-1* and *Lysozyme*, by quantitative PCR in iPSCs derived from SMPs (SMP-iPSCs), in red, and granulocytes (Gra-iPSCs), in grey. The values were normalized to GAPDH expression, the error bars depict the S.E.M. (n=3). **(c)** Heatmap showing top 104 genes with highest variance in their expression levels. Left panel, SMP-iPSCs and Gra-iPSCs derived from chimera #1. Right panel, TTF-iPSCs and B-iPSCs derived from chimera #2. **(d)** Hierarchical unsupervised clustering of iPSC expression profiles using the correlation distance and the Ward method. SMP-iPSCs and Gra-iPSCs were derived from chimera #1 (left panel), TTF-iPSCs and B-iPSCs originate from chimera #2 (right panel). Chi #1, chimera #1; chi #2, chimera #2.

Figure 2. iPSCs derived from different cell types exhibit discernible epigenetic signatures.

(a) Hierarchical unsupervised clustering analysis of HELP genome-wide methylation data from indicated iPSC lines. **(b)** Correspondence analysis of SMP-iPSCs and Gra-iPSCs (left panel) from chimera #1, TTF-iPSCs and B-iPSCs (right panel) from chimera #2. **(c)** Graphic representation of DNA methylation quantification of specific CpGs (circles) in the promoter regions of the indicated candidate genes using EpiTYPER DNA methylation analyses. Yellow indicates 0% methylation and blue 100% methylation. **(d)** Chromatin inmunoprecipitation (ChIP) for H3 pan-acetylated (H3Ac, in blue), H3K4 trimethylated (H3K4me3, in green), H3K27 trimethylated (H3K27me3, in red) and

isotype control (IgG, in light blue) of Granulocytes (Gra), SMPs, Gra-iPSCs and SMP-iPSCs. Chi #1, chimera #1; chi #2, chimera #2. The error bars depict the S.E.M. (n=3)

Figure 3. iPSCs derived from different cell types have distinctive *in vitro* differentiation potentials.

(a) Experimental outline. iPSCs were first differentiated into embryoid bodies (EBs). At day 6, EBs were dissociated and plated in conditions to favor differentiation into erythrocyte progenitors (eryP), macrophage and mixed hematopoietic colonies. **(b)** Images showing EBs derived from B-iPSCs, TTF-iPSCs, Gra-iPSCs and SMP-iPSCs at same magnification. **(c)** Quantification of EBs sizes derived from B-iPSCs (green bar), TTF-iPSCs (blue bar), Gra-iPSCs (grey bar) and SMP-iPSCs (red bar); the diameter of the EBs was measured using arbitrary units (AU). The error bars depict the S.E.M. (n=30) **(d)** Representative images of erythrocyte progenitors (eryPs), macrophage colonies and mixed hematopoietic colonies. **(e-g)** Quantification of *in vitro* differentiation potentials of the different iPSCs into **(e)** EryPs **(f)** macrophage colonies and **(g)** mixed hematopoietic colonies. Chi #1, chimera #1; chi #2, chimera #2. The error bars depict the S.E.M. (n=12)

Figure 4. Continuous passaging of iPSCs abrogates transcriptional, epigenetic and functional differences.

(a) Hierarchical unsupervised clustering of expression profiles from B-iPSCs, T-iPSCs, TTF-iPSCs and Gra-iPSCs from chimera #2. Left panel shows clustering analysis of all iPSC samples at passage (p) 4; middle panel at p10 and right panel depicts same samples at p16. **(b)** Number of differentially expressed genes between pairs of iPSC samples used in (a); iPSCs at p4 are shown in blue bars, iPSCs at p10 are shown in orange bars and iPSCs at p16 are shown in red bars. The number of differently expressed genes between iPSCs was calculated using a pairwise analysis (2 fold), with t-test p value = 0.05 with Bejamini and Hochberg correction (n=3). **(c)** Venn diagram and gene ontology (GO) analysis showing

overlap of genes that change from p4 to p16 in Gra-IPSCs, TTF-iPSCs and B-iPSCs. Red line marks functional GO cluster of genes shared between all three iPSC groups. Black line marks functional GO cluster of genes shared by at least two of the iPSC groups. Functional ontology cluster analysis was performed using DAVIS algorithm. **(d)** Hierarchical unsupervised clustering using HELP genome-wide methylation profiles of B-iPSCs and TTF-iPSs at p16. **(e-g)** Quantification of *in vitro* differentiation potentials of B-iPSCs and TTF-iPSCs at p16 into **(e)** EryPs, **(f)** macrophage colonies, and **(g)** mixed hematopoietic colonies. The error bars depict the S.E.M. (n=9)

Figure 5. Model explaining the presented data.

iPSCs derived from different somatic cell types retain a transient epigenetic and transcriptional memory of their cell type of origin at low passage, despite acquiring pluripotent gene expression, transgene-independent growth and the ability to contribute to tissues in chimeras. Continuous passaging resolves these differences, giving rise to iPSCs that are molecularly and functionally indistinguishable. Note the difference between low passage iPSCs and partially reprogrammed cells, which require continuous viral transgene expression, and fail to activate endogenous pluripotency genes or support the development of viable mice.

Materials and Methods

Generation of iPSC lines

iPSC lines were generated as described previously[2]. Briefly, iPSC-derived somatic cells were isolated from chimeras by FACS, plated on feeders in the presence of cytokines in ESC culture conditions. Resultant iPSC colonies were picked and expanded in the absence of doxycycline and used for subsequent analyses.

SMP isolation

Myofiber-associated cells were prepared from intact limb muscles (EDL, gastrocnemius, quadriceps, soleus, TA, and triceps brachii) as described previously[43] [44]. Briefly, intact mouse limb muscles were digested with collagenase II to dissociate individual myofibers. These were triturated and digested with collagenase II and dispase to release myofiber-associated cells. The myofiber-associated cells were next ubfractionated by FACS, using the following marker profiles for each population: 1) SMPs: CD45−Sca-1−Mac-1−CXCR4+β1-integrin+ ; 2) Myoblast-containing
population: CD45−Sca-1−Mac-1−CXCR4- ; 3) Sca1+ mesenchymal cells: D45−Sca-1+Mac-1-. After the initial sort, cells were resorted by FACS using the same gating profile to increase the purity of the obtained population[45].

Blastocyst injections

For blastocyst injections female BDF1 mice were superovulated by intraperitoneal injection of PMS and hCG and mated to BDF1 stud males. Zygotes were isolated from females with a vaginal plug 24 hour after hCG injection. Zygotes for 2n injections were *in vitro* cultured for 3 days in vitro in KSOM media, blastocysts were identified, injected with ESCs or iPSCs and transferred into pseudopregnant recipient females.

Teratoma formation

iPS cells were harvested by trypsinization, preplated onto untreated culture plates to remove feeders as well as differentiating cells and injected into flanks of NOD/SCID mice, using ~ 5 million cells per injections. The mice were sacrificed 3-5 weeks after injection, teratomas dissected out and processed for histological analysis.

Cellular growth assays

To measure clonal growth potential of iPSCs, SSEA1 positve cells from the different iPSC lines were sorted into 96 wells-plates by FACS (BD). After 7 days the presence of iPSC colonies was scored based on morphology. To establish growth rates, the different bulk iPSCs lines or derivative subclones were plated in 6 gelatinized wells of a 12-well plates and each day the number of cells was counted in duplicate using a Countess cell counter (Invitrogen). For colorimetric measurement of growth, iPSCs lines were subcloned into 96 well-plates and after 7 days, the cells were exposed to XTT (TOX-2) (Sigma) reagent overnight and the absorbance at 450nm measured with a multiwell plate reader (Molecular Devices).

Cell culture

ESCs and iPSCs were cultured in ESC medium (DMEM with 15% FBS, L-Glutamin, penicillin-streptomycin, non-essential amino acids, b-mercaptoethanol and 1000 U/ml LIF) on irradiated feeder cells. Tail-tip fibroblast (TTF) cultures were established by trypsin digestion of tail-tip biopsies taken from newborn (3-8 days of age) chimeric mice produced by blastocyst injection of iPSCs.

RNA isolation

ESCs and iPSCs grown on 35mm dishes were harvested when they reached about 50% confluency and preplated on non-gelatinized T25 flasks for 45 minutes to remove feeder cells. Cells were spun down and the pellet used for isolation of total RNA using the miRNeasy Mini Kit (QIAGEN) without DNase digestion. RNA was eluted from the columns using 50 ml RNase-free water or TE

buffer, pH7.5 (10 mM Tris-HCl and 0.1 mM EDTA) and quantified using a Nanodrop (Nanodrop Technologies).

Quantitative PCR

cDNA was produced with the First Strand cDNA Synthesis Kit (Roche) using 1 mg of total RNA input. Real-time quantitative PCR reactions were set up in triplicate using 5 ml of cDNA (1:100 dilution) with the Brilliant II SYBR Green QPCR Master Mix (Stratagene) and run on a Mx3000P QPCR System (Stratagene). Primer sequences are listed in Supplementary Table 4.

mRNA profiling

Total RNA samples (RIN > 9) were subjected to transcriptomal analyses using Affymetrix HTMG- 430A mRNA expression microarray as previously described. All microarray data is available from the GEO repository (GSE22043).

Statistical analyses

Hierarchical clustering was performed using the GeneSifter software (Geospiza, Seattle). We used correlation distance and subsequent clustering using Ward's method. The differentially expressed genes (2 fold) were calculated using a t-test (p=0.05) with Benjamini and Hochberg correction. Principal component analysis was performed using the GeneSifter software. Gene ontology analisis ws performed using the DAVID software[22]. Using the classification stringency set to "High".

Embryoid body (EB) formation

Prior to plating EBs, the iPSCs were depleted of MEFs by splitting the cells 1:3 onto gelatin-coated plates on each day, for 2 consecutive days. On the 3rd day (designated day 0), iPSCs were trypsinized and plated at a density of 5,000 cells/ml in Isocove's Modified Dulbecco's Medium (IMDM) with 15% FCS (Atlanta Biologicals), 10% protein-free hybridoma medium (PFHM-II; Gibco), 2 mM L-glutamine (Gibco), 200 µg/mL transferrin (Roche), 0.5 mM ascorbic acid (Sigma),

and 4.5 x 10–4 M monothioglycerol (MTG; Sigma). Differentiation was carried out in 60-mm ethylene oxide-treated Petri grade dishes (Parter Medical). The EBs were left to differentiate until day 6, upon which the cells were harvested to assay for hematopoietic colonies.

Hematopoietic colony formation assays

Day 6 EBs were collected by gravity, dissociated with trypsin and then passed several times through a 20 gauge needle to ensure dissociation. For the growth of hematopoietic progenitors, the cells were then seeded at a density of 100,000 cells/ml in IMDM containing 1% methylcellulose (Fluka Biochemika), 15% plasma-derived serum (PDS; Animal Technologies), 5% PFHM-II, and specific cytokines as follows: primitive erythrocytes (erythropoietin [EPO, 2 U/mL]); macrophages (IL-3 [10ng/ml], M-CSF [5 ng/mL]); megakaryocytes (IL-3 [10ng/ml], IL-11 [5 ng/mL], thrombopoietin [TPO, 5 ng/mL]); mixed colonies (SCF [5ng/ml], IL-3 [10 ng/mL], G-CSF [30 ng/mL], GM-CSF [10 ng/mL], IL-11 [5 ng/mL], IL-6 [5 ng/mL], TPO [5 ng/mL], and M-CSF [5 ng/mL]). All cytokines were purchased from R&D Systems. Primitive erythroctye colonies (eryPs) were counted on day 10 (4 days after EB harvest). Macrophage colonies were counted on day 13 (7 days after EB harvest). Mixed colonies were counted on day 14 (8 days after EB harvest) and consist of a layer of macrophages, a layer of granulocytes, and a central core of red erythroid cells. Statistical analysis was performed using the Krward software. P values were calculating using the non-parametric Wilkinson test.

HELP DNA methylation analysis

High-molecular-weight DNA was isolated from iPSCs using the PureGene kit from Qiagen (Valencia, CA) and the HELP (HpaII tiny fragment enrichment by ligation mediated PCR) assay was carried out as previously described1, 2. Briefly, One microgram of genomic DNA was digested overnight with either HpaII or MspI (NEB, Ipswich, MA). On the following day the reactions were extracted once with phenol-chloroform and resuspended in 11 µL of 10 mM Tris-HCl pH

8.0 and the digested DNA was used to set up an overnight ligation of the JHpaII adapter using T4 DNA ligase. The adapter-ligated DNA was used to carry out the PCR amplification of the HpaII and MspI-digested DNA as previously described[46]. All samples for microarray hybridization were processed at the Roche-NimbleGen Service Laboratory. Samples were labeled using Cy-labeled random primers (9mers) and then hybridized onto a mouse custom-designed oligonucleotide array (50-mers) covering 25,720 HpaII amplifiable fragments (HAF) (>50,000 CpGs), annotated to 15,465 unique gene symbols (Roche NimbleGen, Design name: 2006-10-26_MM5_HELP_Promoter Design ID=4803). HpaII amplifiable fragments are defined as genomic sequences contained between two flanking HpaII sites found within 200-2,000 bp from each other and is represented on the array by 15 individual probes, randomly distributed across the microarray slide. HAF were first re-aligned to the MM9 July 2007 build of the mouse genome and then annotated to the nearest transcription start site (TSS), allowing for a maximum distance of 5 kb from the TSS. Scanning was performed using a GenePix 4000B scanner (Axon Instruments) as previously described[47]. Quality control and data analysis of HELP microarrays was performed as described in Thompson et al[48].

Signal intensities at each HpaII amplifiable fragment were calculated as a robust (25% trimmed) mean of their component probe-level signal intensities. Any fragments found within the level of background MspI signal intensity, measured as 2.5 mean-absolute-differences (MAD) above the median of random probe signals, were categorized as “failed.” These “failed” loci therefore represent the population of fragments that did not amplify by PCR, whatever the biological (e.g. genomic deletions and other sequence errors) or experimental cause. On the other hand, “Methylated” loci were so designated when the level of HpaII signal intensity was similarly indistinguishable from background. PCR-amplifying fragments (those not flagged as either “methylated” or “failed”) were normalized using an intra-array quantile approach wherein HpaII/MspI ratios are aligned across density-dependent sliding windows of fragment size-sorted data. DNA methylation was therefore measured as the log2(HpaII/MspI) ratio, where

HpaII reflects the hypomethylated fraction of the genome and MspI represents the whole genome reference. Analysis of normalized data revealed the presence of a bimodal distribution. For each sample a cutoff was selected at the point that more clearly separated these two populations and the data were centered around this point. Each fragment was then categorized as either methylated, if the centered log HpaII/MspI ratio was less than zero, or hypomethylated if on the other hand the log ratio was greater than zero. All microarray data will be available from the GEO repository[49].

HELP data analysis

Statistical analysis was performed using R 2.9 and BioConductor[50]. Unsupervised hierarchical clustering of HELP data was performed using the subset of probe sets (n=3745) with standard deviation > 1 across all cases. We used 1- Pearson correlation distance, followed by a Lingoes transformation of the distance matrix to a Euclidean one and subsequent clustering using Ward's method. Correspondence analysis was performed using the BioConductor package MADE4. The top 100 genes whose methylation status varied the most across the different groups were identified as those with the greatest standard deviation across all samples.

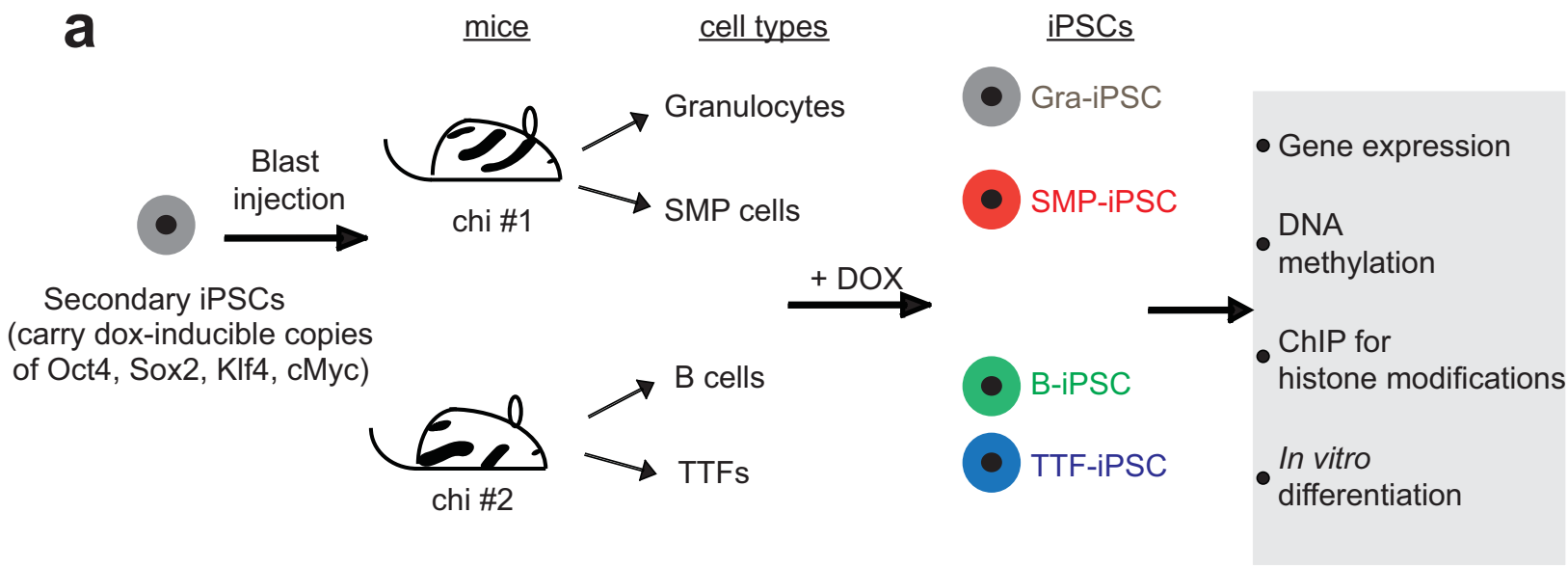
Quantitative DNA methylation analysis by MassARRAY EpiTyping

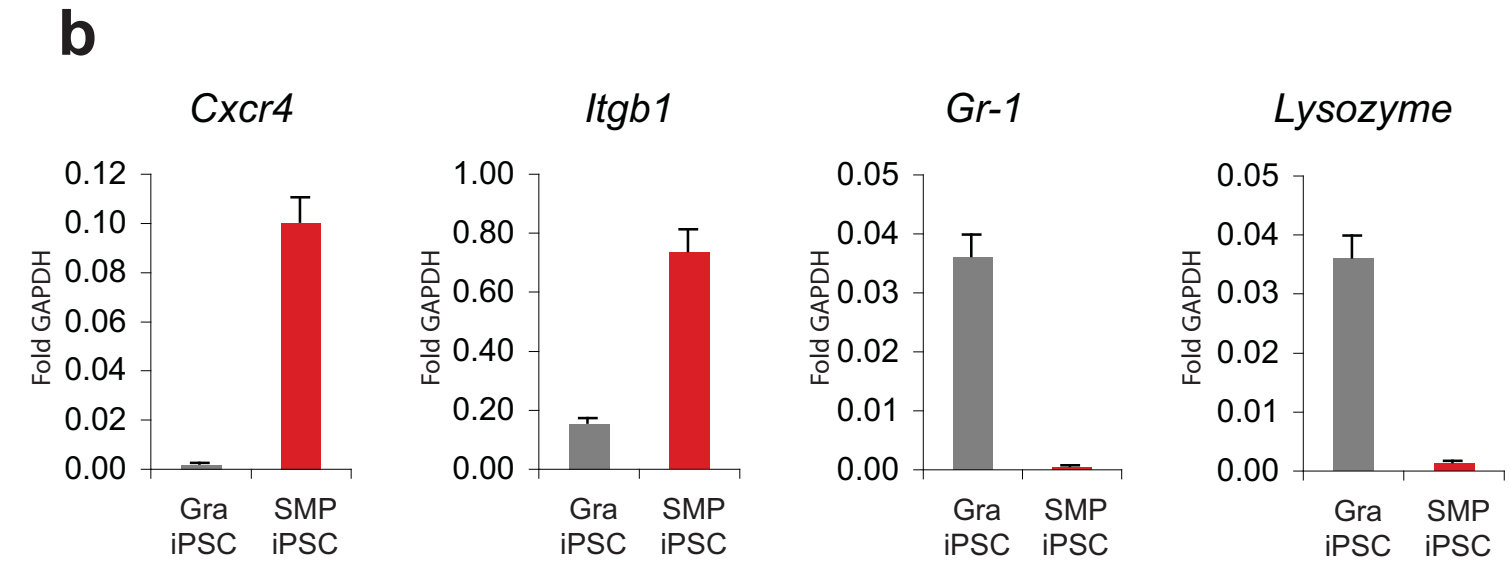
Validation of HELP findings was performed by MALDI-TOF mass spectrometry using EpiTyper by MassARRAY (Sequenom, CA) on bisulfite-converted DNA following manufacturer's instructions[51], but using the Fast Start High Fidelity Taq polymerase from Roche for the PCR amplification of the bisulfite converted DNA. MassArray primers were designed to cover the promoter regions of the indicated genes. (Primer sequences available as Supplementary Table 5).

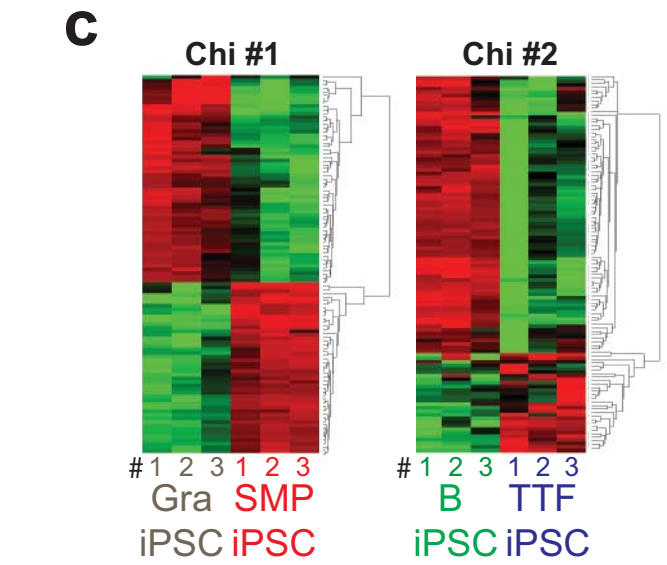
Chromatin immunoprecipitation (ChIP)

Cells were fixed in 1% formaldehyde for 10 minutes, quenched with glycine and washed 3 times with PBS. Cells were then resuspended in lysis buffer and

sonicated 10 x 30 sec in a Bioruptor (Diagenode, Philadelphia, PA) to shear the chromatin to an average length of 600 bp. Supernatants were precleared using protein-A agarose beads (Roche, Mannheim, Germany) and 10% input was collected. Immunoprecipitations were performed using polyclonal antibodies to H3K4trimethylated, H3K27trimethylated , H3 pan-acetylation and normal rabbit serum (Upstate, Temucula, CA). DNA-protein complexes were pulled-down using Protein-A agarose beads and washed. DNA was recoverd by overnight incubation at 65°C to reverse cross-links and purified using QIAquick PCR purification columns (Qiagen, Maryland). Enrichment of the modified histones in different genes was detected by quantitative real time PCR using the primers in the Supplementary Table 4.


a
mice
cell types
iPSCs
Blast injection
Secondary iPSCs (carry dox-inducible copies of Oct4, Sox2, Klf4, cMyc)
chi #1
chi #2
Granulocytes
SMP cells
B cells
TTFs
+ DOX
Gra-iPSC
SMP-iPSC
B-iPSC
TTF-iPSC
Gene expression
DNA methylation
ChIP for histone modifications
In vitro differentiation


b
Cxcr4
Itgb1
Gr-1
Lysozyme
Fold GAPDH
Gra iPSC
SMP iPSC


c
Chi #1
Chi #2
Gra iPSC
SMP iPSC
B iPSC
TTF iPSC

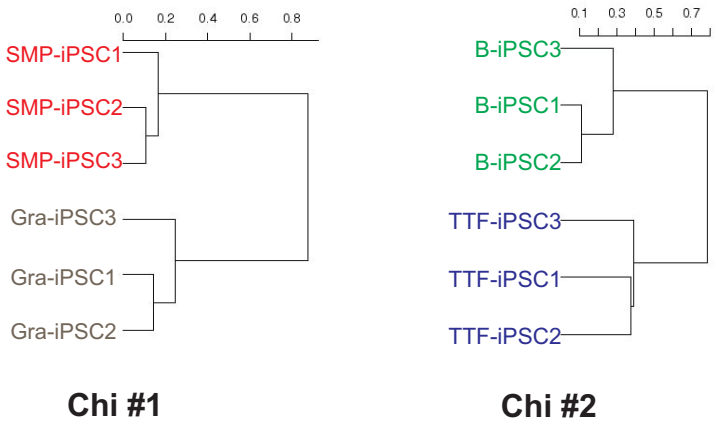

d
SMP-iPSC1
SMP-iPSC2
SMP-iPSC3
Gra-iPSC3
Gra-iPSC1
Gra-iPSC2
B-iPSC3
B-iPSC1
B-iPSC2
TTF-iPSC3
TTF-iPSC1
TTF-iPSC2
Chi #1
Chi #2

Figure 1

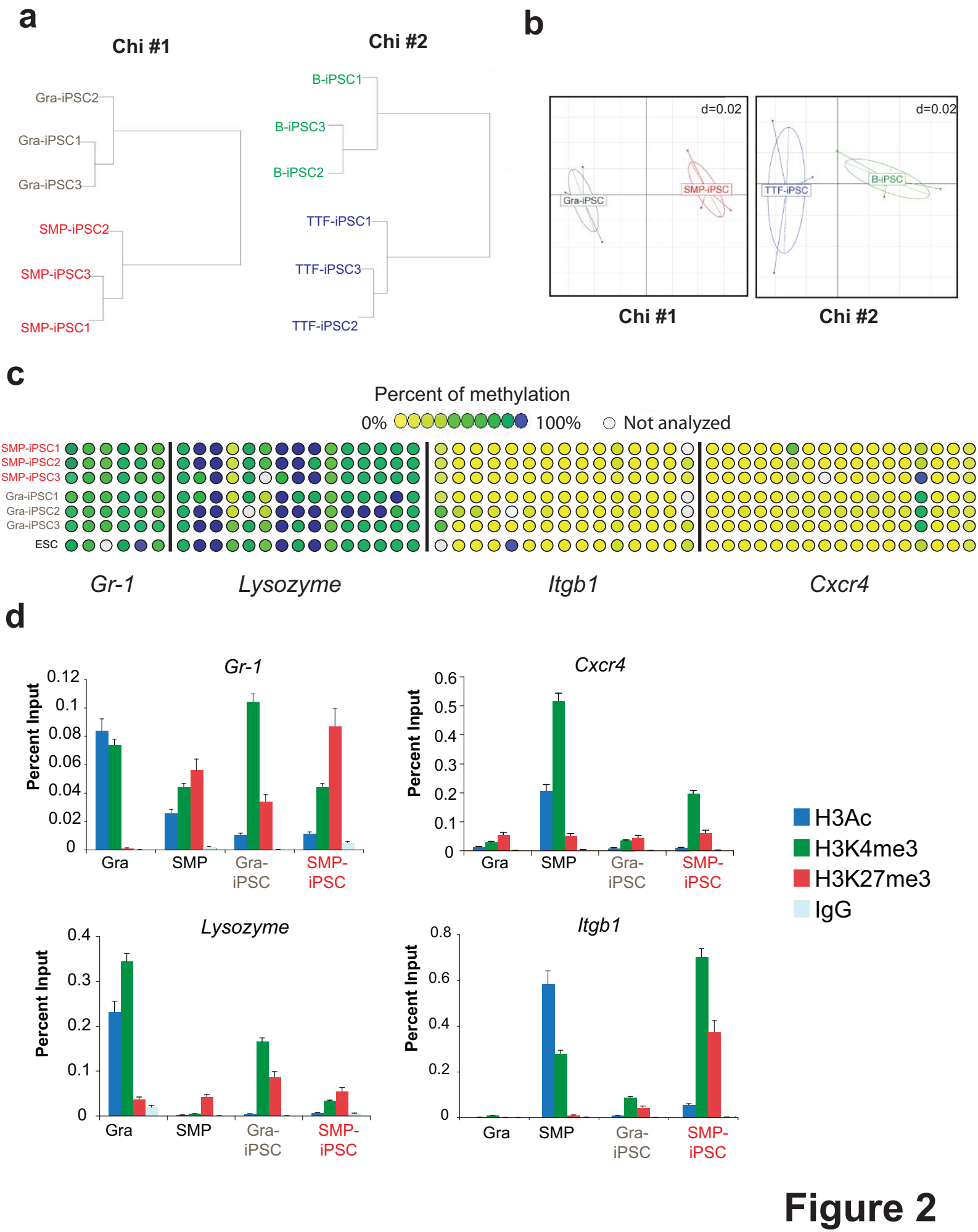
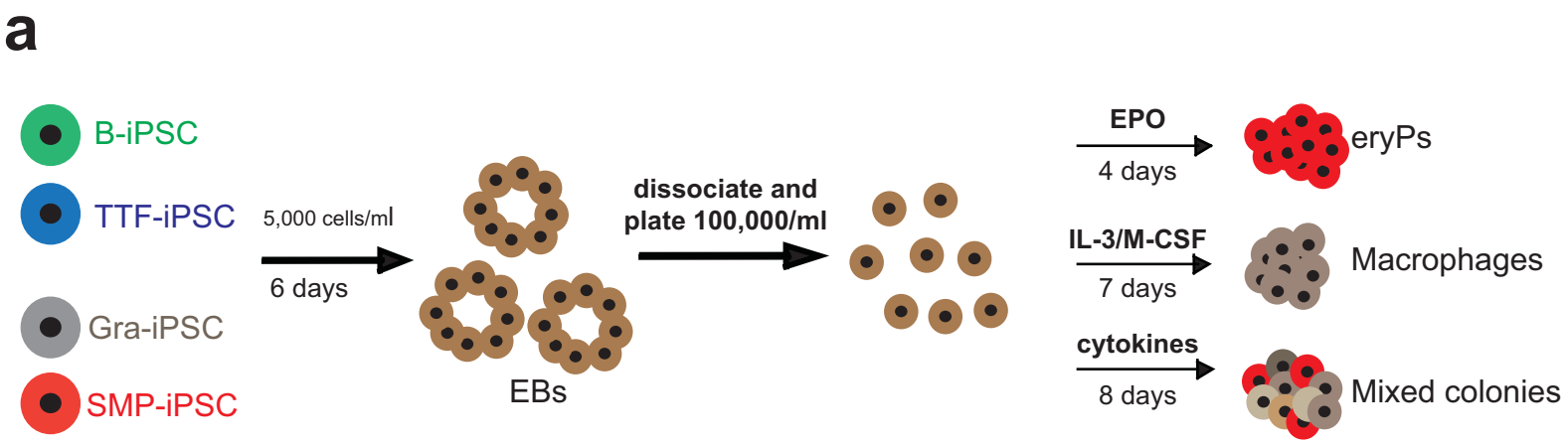
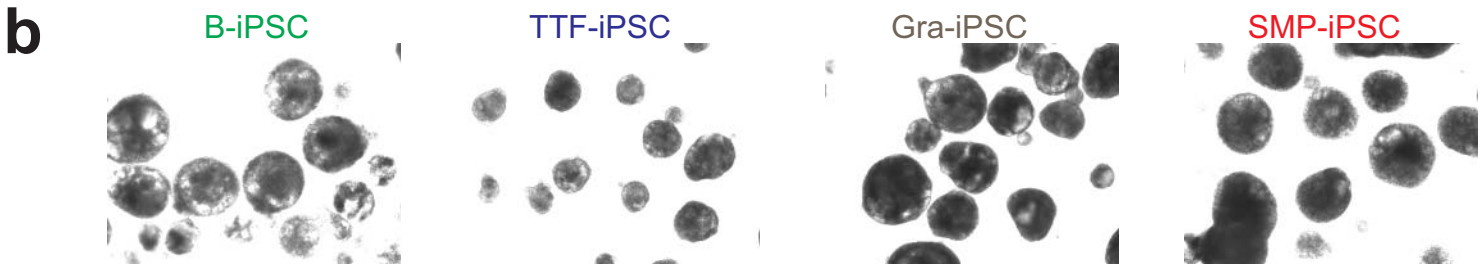
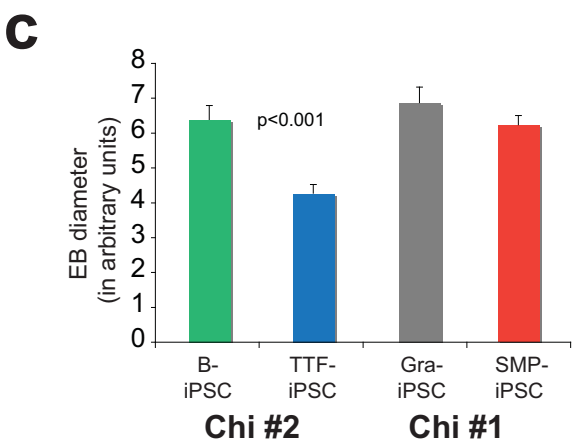
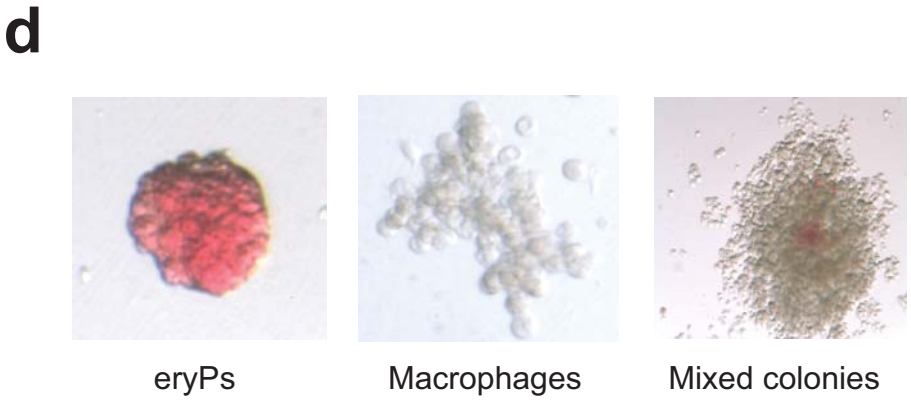

a
Chi #1
Chi #2
Gra-iPSC2
Gra-iPSC1
Gra-iPSC3
SMP-iPSC2
SMP-iPSC3
SMP-iPSC1
B-iPSC1
B-iPSC3
B-iPSC2
TTF-iPSC1
TTF-iPSC3
TTF-iPSC2
b
d=0.02
Gra-iPSC
SMP-iPSC
d=0.02
TTF-iPSC
B-iPSC
Chi #1
Chi #2
c
Percent of methylation
0%
100%
Not analyzed
SMP-iPSC1
SMP-iPSC2
SMP-iPSC3
Gra-iPSC1
Gra-iPSC2
Gra-iPSC3
ESC
Gr-1
Lysozyme
Itgb1
Cxcr4
d
Gr-1
Cxcr4
Lysozyme
Itgb1
Percent Input
Gra
SMP
Gra-iPSC
SMP-iPSC
H3Ac
H3K4me3
H3K27me3
IgG

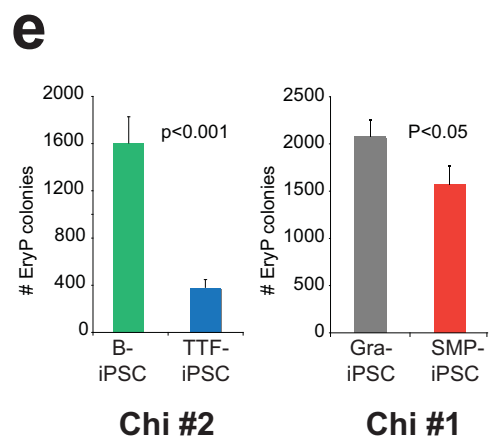
Figure 2

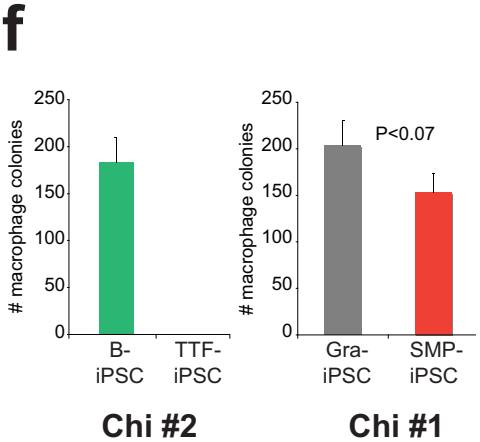

a
B-iPSC
TTF-iPSC
Gra-iPSC
SMP-iPSC
5,000 cells/ml
6 days
EBs
dissociate and
plate 100,000/ml
EPO
4 days
eryPs
IL-3/M-CSF
7 days
Macrophages
cytokines
8 days
Mixed colonies


b
B-iPSC
TTF-iPSC
Gra-iPSC
SMP-iPSC


c
EB diameter
(in arbitrary units)
p<0.001
B-iPSC
TTF-iPSC
Gra-iPSC
SMP-iPSC
Chi #2
Chi #1


d
eryPs
Macrophages
Mixed colonies


e
EryP colonies
p<0.001
P<0.05
B-iPSC
TTF-iPSC
Gra-iPSC
SMP-iPSC
Chi #2
Chi #1


f
macrophage colonies
P<0.07
B-iPSC
TTF-iPSC
Gra-iPSC
SMP-iPSC
Chi #2
Chi #1

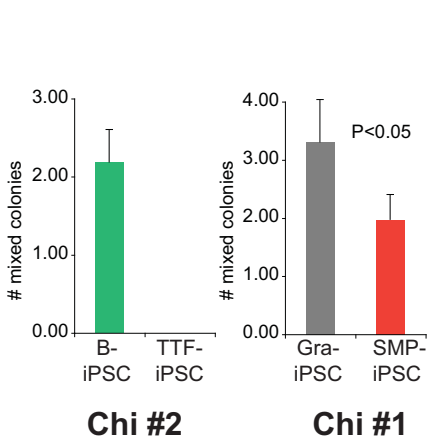

g
mixed colonies
P<0.05
B-iPSC
TTF-iPSC
Gra-iPSC
SMP-iPSC
Chi #2
Chi #1

Figure 3

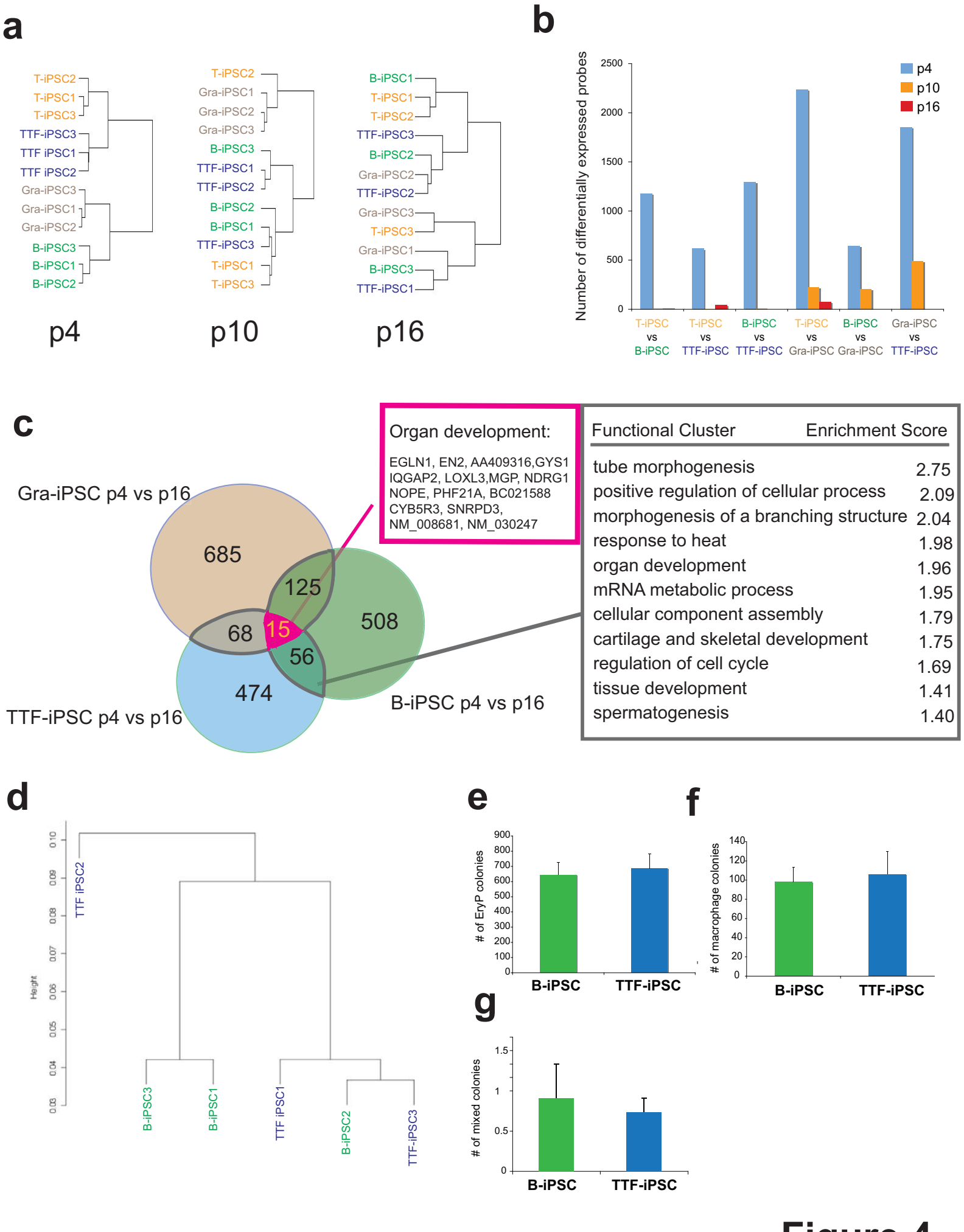

a
T-iPSC2
T-iPSC1
T-iPSC3
TTF-iPSC3
TTF iPSC1
TTF iPSC2
Gra-iPSC3
Gra-iPSC1
Gra-iPSC2
B-iPSC3
B-iPSC1
B-iPSC2
p4
T-iPSC2
Gra-iPSC1
Gra-iPSC2
Gra-iPSC3
B-iPSC3
TTF-iPSC1
TTF-iPSC2
B-iPSC2
B-iPSC1
TTF-iPSC3
T-iPSC1
T-iPSC3
p10
B-iPSC1
T-iPSC1
T-iPSC2
TTF-iPSC3
B-iPSC2
Gra-iPSC2
TTF-iPSC2
Gra-iPSC3
T-iPSC3
Gra-iPSC1
B-iPSC3
TTF-iPSC1
p16
b
Number of differentially expressed probes
2500
2000
1500
1000
500
0
p4
p10
p16
T-iPSC vs B-iPSC
T-iPSC vs TTF-iPSC
B-iPSC vs TTF-iPSC
T-iPSC vs Gra-iPSC
B-iPSC vs Gra-iPSC
Gra-iPSC vs TTF-iPSC
c
Gra-iPSC p4 vs p16
685
125
68
15
508
56
474
TTF-iPSC p4 vs p16
B-iPSC p4 vs p16
Organ development:
EGLN1, EN2, AA409316,GYS1
IQGAP2, LOXL3,MGP, NDRG1
NOPE, PHF21A, BC021588
CYB5R3, SNRPD3,
NM_008681, NM_030247
Functional Cluster
Enrichment Score
tube morphogenesis 2.75
positive regulation of cellular process 2.09
morphogenesis of a branching structure 2.04
response to heat 1.98
organ development 1.96
mRNA metabolic process 1.95
cellular component assembly 1.79
cartilage and skeletal development 1.75
regulation of cell cycle 1.69
tissue development 1.41
spermatogenesis 1.40
d
Height
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
TTF iPSC2
B-iPSC3
B-iPSC1
TTF iPSC1
B-iPSC2
TTF-iPSC3
e
# of EryP colonies
900
800
700
600
500
400
300
200
100
0
B-iPSC
TTF-iPSC
f
# of macrophage colonies
140
120
100
80
60
40
20
0
B-iPSC
TTF-iPSC
g
# of mixed colonies
1.5
1
0.5
0
B-iPSC
TTF-iPSC

Figure 4

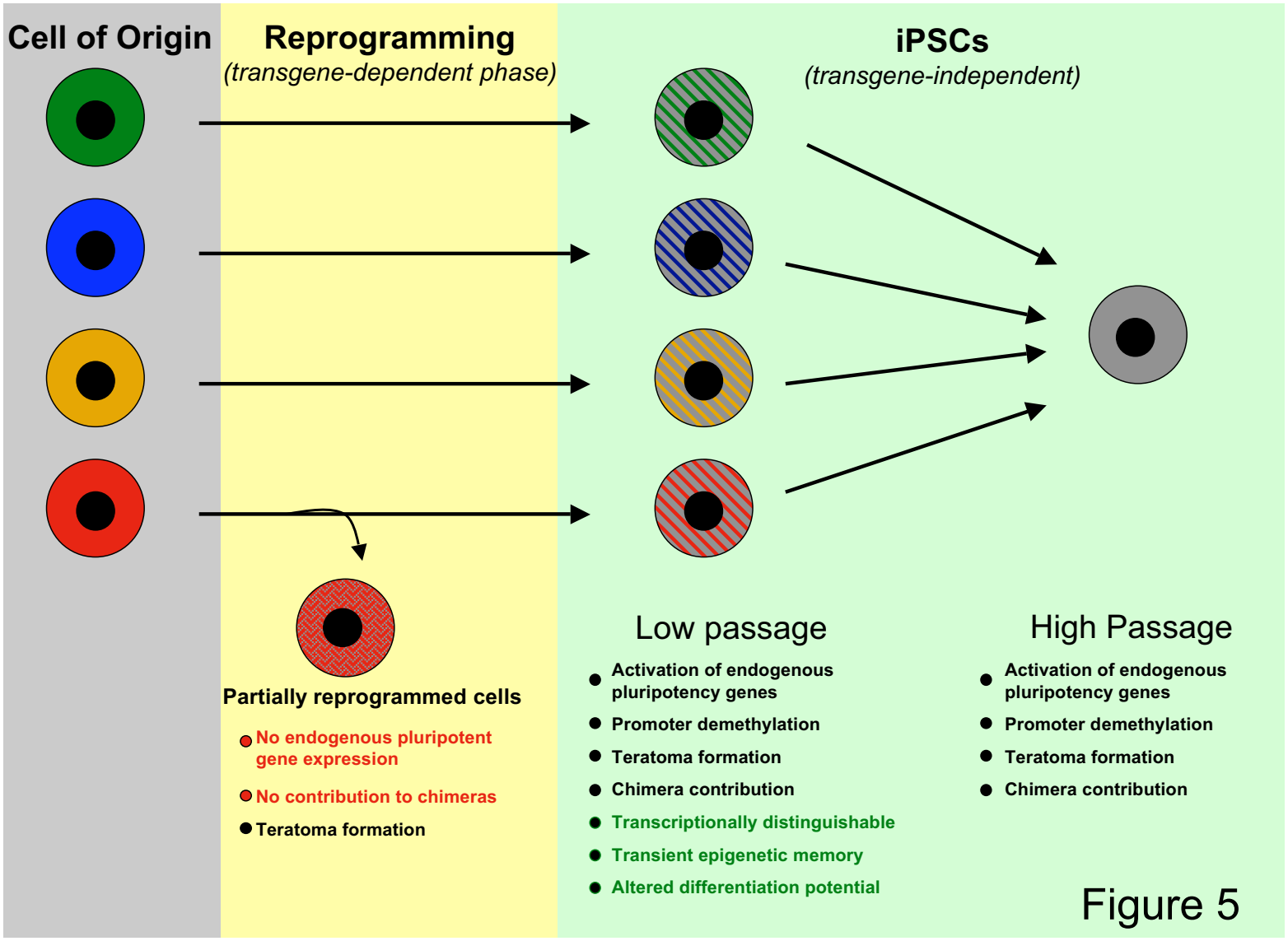

Cell of Origin
Reprogramming
(transgene-dependent phase)
iPSCs
(transgene-independent)
Partially reprogrammed cells
No endogenous pluripotent gene expression
No contribution to chimeras
Teratoma formation
Low passage
Activation of endogenous pluripotency genes
Promoter demethylation
Teratoma formation
Chimera contribution
Transcriptionally distinguishable
Transient epigenetic memory
Altered differentiation potential
High Passage
Activation of endogenous pluripotency genes
Promoter demethylation
Teratoma formation
Chimera contribution

Figure 5