Stem Cell Reports
Article

ISSCR

OPEN ACCESS

Genetic Variability Overrides the Impact of Parental Cell Type and Determines iPSC Differentiation Potential

Aija Kyttälä,[1] Roksana Moraghebi,[2] Cristina Valensisi,[3,4] Johannes Kettunen,[1,5,6,13] Colin Andrus,[3] Kalyan Kumar Pasumarthy,[4] Mahito Nakanishi,[7] Ken Nishimura,[7,8] Manami Ohtaka,[7] Jere Weltner,[9] Ben Van Handel,[10] Olavi Parkkonen,[11] Juha Sinisalo,[11] Anu Jalanko,[1] R. David Hawkins,[3,4] Niels-Bjarne Woods,[2] Timo Otonkoski,[9,12,*] and Ras Trokovic[9,*]

[1]Genomics and Biomarkers Unit, National Institute for Health and Welfare (THL), THL Biobank, 00290 Helsinki, Finland
[2]Department of Molecular Medicine and Gene Therapy, Lund Stem Cell Center, Lund University, BMC A12, 221 84 Lund, Sweden
[3]Division of Medical Genetics, Departments of Medicine and Genome Sciences, University of Washington, Seattle, WA 98195-7720, USA
[4]Turku Centre for Biotechnology, Turku 20520, Finland
[5]Computational Medicine, Institute of Health Sciences, University of Oulu, Oulu 90014, Finland
[6]NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio 70210, Finland
[7]Biotechnology Research Institute for Drug Discovery, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8565, Japan
[8]Laboratory of Gene Regulation, Faculty of Medicine, University of Tsukuba, Tsukuba, Ibaraki 305-8575, Japan
[9]Research Programs Unit, Molecular Neurology and Biomedicum Stem Cell Centre, University of Helsinki, 00290 Helsinki, Finland
[10]Novogenix Laboratories, LLC, Los Angeles, CA 90033, USA
[11]Heart and Lung Center, Helsinki University Central Hospital and University of Helsinki, 00029 HUS Helsinki, Finland
[12]Children's Hospital, University of Helsinki and Helsinki University Central Hospital, 00029 HUS Helsinki, Finland
[13]Biocenter Oulu, University of Oulu, 90014 Oulu, Finland
*Correspondence: timo.otonkoski@helsinki.fi (T.O.), ras.trokovic@helsinki.fi (R.T.)
http://dx.doi.org/10.1016/j.stemcr.2015.12.009

SUMMARY

Reports on the retention of somatic cell memory in induced pluripotent stem cells (iPSCs) have complicated the selection of the optimal cell type for the generation of iPSC biobanks. To address this issue we compared transcriptomic, epigenetic, and differentiation propensities of genetically matched human iPSCs derived from fibroblasts and blood, two tissues of the most practical relevance for biobanking. Our results show that iPSC lines derived from the same donor are highly similar to each other. However, genetic variation imparts a donor-specific expression and methylation profile in reprogrammed cells that leads to variable functional capacities of iPSC lines. Our results suggest that integration-free, bona fide iPSC lines from fibroblasts and blood can be combined in repositories to form biobanks. Due to the impact of genetic variation on iPSC differentiation, biobanks should contain cells from large numbers of donors.

INTRODUCTION

Although cell-fate decisions are fairly stable in vivo, somatic cells can be reprogrammed back into pluripotency in vitro by ectopic expression of defined transcription factors (Takahashi and Yamanaka, 2006). Successful reprogramming requires complete erasure of somatic cell memory and establishment of a pluripotent stem cell epigenetic landscape (Nashun et al., 2015). Fibroblasts and peripheral blood mononuclear cells (PBMCs) are commonly used for reprogramming (Santostefano et al., 2015). Induced pluripotent stem cells (iPSCs) are known to be epigenetically similar to human embryonic stem cells (hESCs) (Guenther et al., 2010; Maherali et al., 2007), although several reports have suggested retention of epigenetic memory related to the cell of origin (Bar-Nur et al., 2011; Kim et al., 2010, 2011; Ohi et al., 2011; Polo et al., 2010). This phenomenon can have functional consequences by influencing iPSC differentiation propensity and biasing it toward the cell type of origin at the expense of other lineages (Bar-Nur et al., 2011; Kim et al., 2010; Polo et al., 2010). However, conflicting studies have shown that variations in directed differentiation (Kajiwara et al., 2012) and transcriptional heterogeneity (Rouhani et al., 2014) between iPSC lines were ascribed to the genetic background of the donor.

iPSC biobanks can provide powerful material for modeling human diseases and regenerative cell therapies. However, the absence of systematic molecular and functional studies of iPSC lines generated from different genetic backgrounds and cell types of origin has hampered reprogramming efforts for large-scale biobanking purposes. In particular, the omission of blood cells prevents leveraging the resources of numerous biorepositories that have collected blood cells for human genetic, metabolic, and related studies. In this study we examined whether comparable iPSC line collections can be established from fibroblasts and blood. To address issues of donor genetic background and cell type of origin, we produced genetically matched iPSC lines from fibroblasts and blood from several donors and thoroughly investigated their

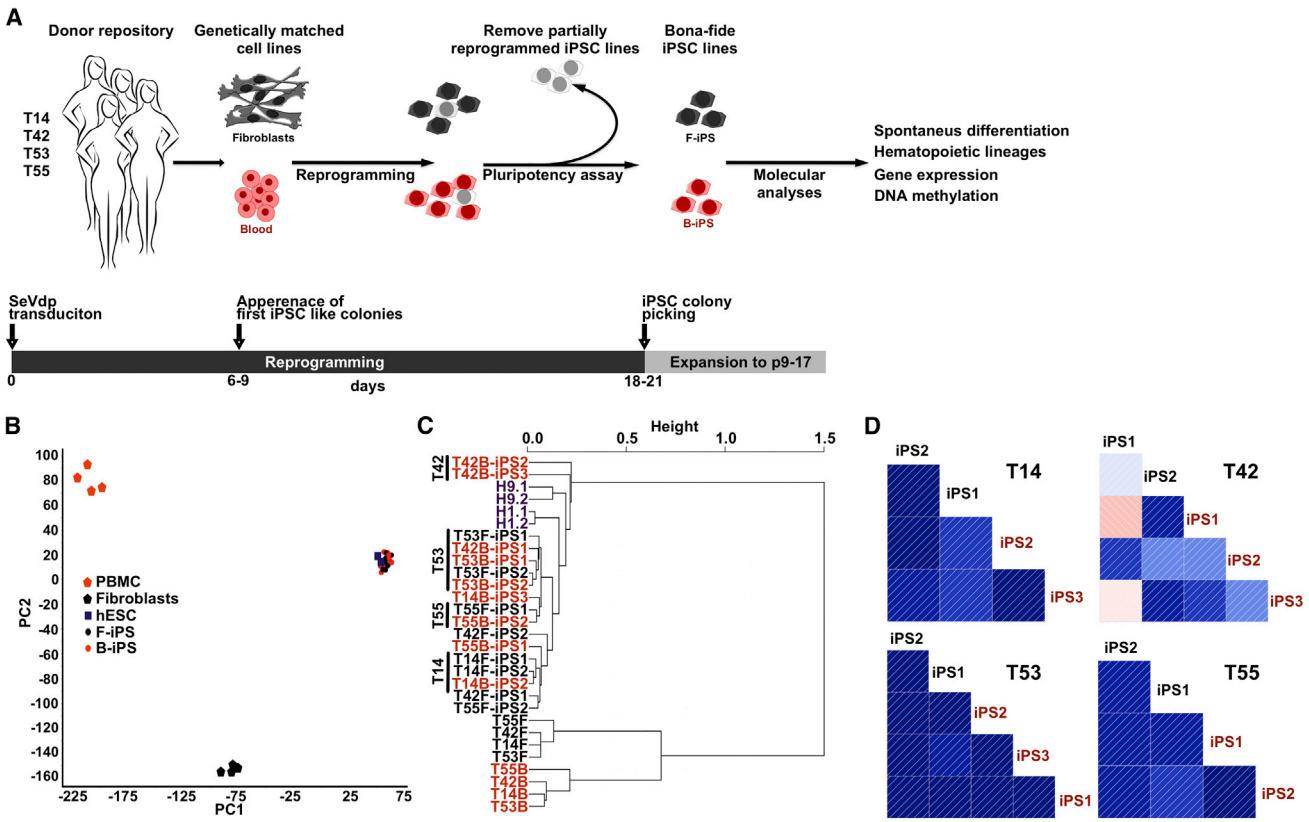


Figure 1. iPSCs Derived from Fibroblasts and Blood Cells Are Transcriptionally Similar

(A) Schematic representation of the study. Genetically matched iPSC were produced from fibroblasts and blood cells (PBMCs) from four female donors (T14, T42, T53, T55) using Sendai virus (SeVdp) mediated reprogramming. The cell lines used in this study are listed in Table 1. Fibroblast-derived iPSCs (F-iPS) and blood-derived iPSCs (B-iPS) are shown in black and red, respectively, throughout the figures.

(B) PCA of global gene expression data of genetically matched F-iPSCs (n = 8) and B-iPSCs (n = 10) derived from four donors, two human embryonic stem cell lines (hESC), and somatic cells of origin (n = 4/4). Characterization of iPSC lines is presented in Figure S1.

(C) Unsupervised hierarchical clustering of global DNA methylation profiles in genetically matched F-iPSC (n = 8) and B-iPSC (n = 10) lines, somatic cells of origin (n = 4/4), and two hESC lines (H1, H9) performed on a single-nucleotide level using reduced representation bisulfite sequencing.

(D) Pairwise correlation of the genetically matched F- and B-iPSC lines for each donor (T14, n = 4; T42, n = 5; T53, n = 5; T55, n = 4) after local-pooled-error test. The entire list of genes for each donor is presented in Table S1. The direction of the correlation is visualized using thin lines inside boxes, and the magnitude of correlation using the colors. Darker color corresponds to the higher correlation. Isogenic iPSC lines derived from donor T42 display the lowest correlation.

transcriptional and epigenetic status, as well as their spontaneous and multi-lineage hematopoietic differentiation potential.

RESULTS

Global Analysis of iPSC Lines Generated from Genetically Matched Fibroblasts and Blood

Variation between iPSC lines has been attributed to many factors, such as cell type of origin, donor, culture conditions, and reprogramming method. To perform unambiguous studies on retention of cell-type memory, we generated isogenic iPSC lines from fibroblasts (F-iPSCs) and PBMCs (B-iPSCs) by Sendai virus-mediated reprogramming under standardized conditions (Figure 1A and Table 1) (Nishimura et al., 2011; Trokovic et al., 2014). To reduce gender-associated variation, only female donors were selected for the study. All iPSC lines expressed stem cell markers and showed morphology and growth characteristics similar to those of hESCs, and were propagated up to passage 9–17 (Figures S1A and S1B; Table 1). All iPSC lines were able to spontaneously differentiate into three embryonic germ layers in embryoid bodies (Figure S1C). To avoid the confounding effects of partially reprogrammed cells, only cell lines identified as bona fide iPSCs by PluriTest

Table 1. Samples Used in the Study

Donor	Sex	Cell Line/Clone Number	p	Cell Type	Donor Cell Type	Array/No	RRBS	EB Array/No
T14	F	T14F		Fib		yes/1	yes	
		T14BC		PBMC		yes/1	yes	
		T14F_iPS.1	10	iPSC	Fib	yes/2	yes	yes/1
		T14F_iPS.2	9	iPSC	Fib	yes/1	yes	
		T14BC_iPS.2	9	iPSC	PBMC	yes/1	yes	yes/1
		T14BC_iPS.3	9	iPSC	PBMC	yes/2	yes	
T42	F	T42F		Fib		yes/1	yes	
		T42BC		PBMC		yes/1	yes	
		T42F_iPS.1	15	iPSC	Fib	yes/1	yes	yes/1
		T42F_iPS.2	12	iPSC	Fib	yes/2	yes	
		T42BC_iPS.1	10	iPSC	PBMC	yes/2	yes	
		T42BC_iPS.2	14	iPSC	PBMC	yes/2	yes	yes/1
		T42BC_iPS.3	9	iPSC	PBMC	yes/2	yes	
T53	F	T53F		Fib		yes/2	yes	
		T53BC		PBMC		yes/2	yes	
		T53F_iPS.1	10	iPSC	Fib	yes/2	yes	
		T53F_iPS.2	9	iPSC	Fib	yes/1	yes	yes/1
		T53BC_iPS.1	11	iPSC	PBMC	yes/2	yes	
		T53BC_iPS.2	9	iPSC	PBMC	yes/1	yes	yes/1
		T53BC_iPS.3	9	iPSC	PBMC	yes/2		
T55	F	T55F		Fib		yes/2	yes	
		T55BC		PBMC		yes/2	yes	
		T55F_iPS.1	12	iPSC	Fib	yes/2	yes	
		T55F_iPS.2	17	iPSC	Fib	yes/1	yes	yes/1
		T55BC_iPS.1	17	iPSC	PBMC	yes/1	yes	yes/1
		T55BC_iPS.2	13	iPSC	PBMC	yes/2	yes	
H9	F	H9	46	hESC		yes/2	yes	yes/1
FES22	M	FES22	56	hESC		yes/1		yes/1

F, female; M, male; p, passage; Fib, fibroblast; RRBS, reduced representation bisulfite sequencing.

(Muller et al., 2011) were selected for further experiments (Figure S1D). To avoid batch effects in expression profiling (Leek et al., 2010), we distributed F- and B-iPSC lines across batches (Table 1). Global gene expression analysis of all cell lines showed that pluripotent stem cells (PSCs) clustered together and were clearly separated from their parental cell lines (Figures 1B and S1E). Expression analysis of genes located in X chromosome showed little variation between lines (Figure S1F), suggesting that our female iPSC lines retain an inactive X chromosome (Tchieu et al., 2010). Global DNA methylation analysis performed at a single-nucleotide level using reduced representation bisulfite sequencing (RRBS) (Meissner et al., 2005) also resulted in a clustering of PSCs (Figure 1C). Interestingly, both

global DNA methylation and gene expression analyses revealed a tendency of iPSC lines to cluster according to the donor rather than cell type of origin (Figures 1C and S1E).

Cell Type of Origin Contributes Minimally to iPSC Variability

To analyze differences in expression profiles resulting solely from the cell type of origin, we grouped iPSC lines according to their parental cell type into two groups (F- and B-iPSCs). The reproducibility-optimized test statistic (Elo et al., 2008) and significance analysis of microarrays (SAM) (Tusher et al., 2001) identified only two differentially expressed genes (*TCERG1L* and *COL22A1*) between the iPSC groups. Using a separate statistical test, two-group empirical Bayes method (BH, $p < 0.05$) (Smyth, 2004), we identified only *TCERG1L* as a gene expressed significantly higher in B-iPSCs than in F-iPSCs. To increase the power of analysis, we used an arbitrary fold-change (FC) cutoff of >1. We subsequently identified 13 differentially expressed genes between F- and B-iPSC groups. However, unsupervised hierarchical clustering of all iPSC lines did not separate them into two groups according to the cell type of origin (Figures S2A and S2B). This suggests that the detected differences between iPSC lines were not due to different tissues of origin.

To eliminate variability resulting from genetic background, we compared F- versus B-iPSCs for each donor. Using a local-pooled-error (LPE) test, a statistical test well suited for small sample sizes (Jain et al., 2003), we identified 24 (T14B- versus F-iPSCs), 13 (T42B- versus F-iPSCs), 6 (T53B- versus F-iPSCs), and 158 (T55B- versus F-iPSCs) differentially expressed genes between the isogenic iPSC lines (Figure 1D and Table S1). Of interest, we noticed that T42- and T55-derived iPSCs showed larger intra-line variability compared with T14- and T53-derived iPSC lines (Figure S2C). *MEG3* was the only common element for all four donors. In addition, *TCERG1L*, *COL3A1*, and *HAND1* were common between three donors (T14, T42, and T55) (Figure S2D). These data are in line with previous studies showing that the imprinted genes *MEG3* and *TCERG1L* are frequently differentially expressed between PSC lines (Lister et al., 2011; Stadtfeld et al., 2010; Wang et al., 2013). To increase the power of analysis we again used the arbitrary FC cutoff >1 and identified 134–435 differentially expressed genes between isogenic iPSC lines (Table S2). After examination of gene ontology (GO) terms for all groups, identified using LPE or FC >1 tests, we did not find terms enriched for hematopoietic processes that would suggest a cell type of origin bias. Together, genome-wide transcriptomic analysis results show that F- and B-iPSCs are highly similar to each other, suggesting that cell type of origin is not a major factor resulting in iPSC line variability.

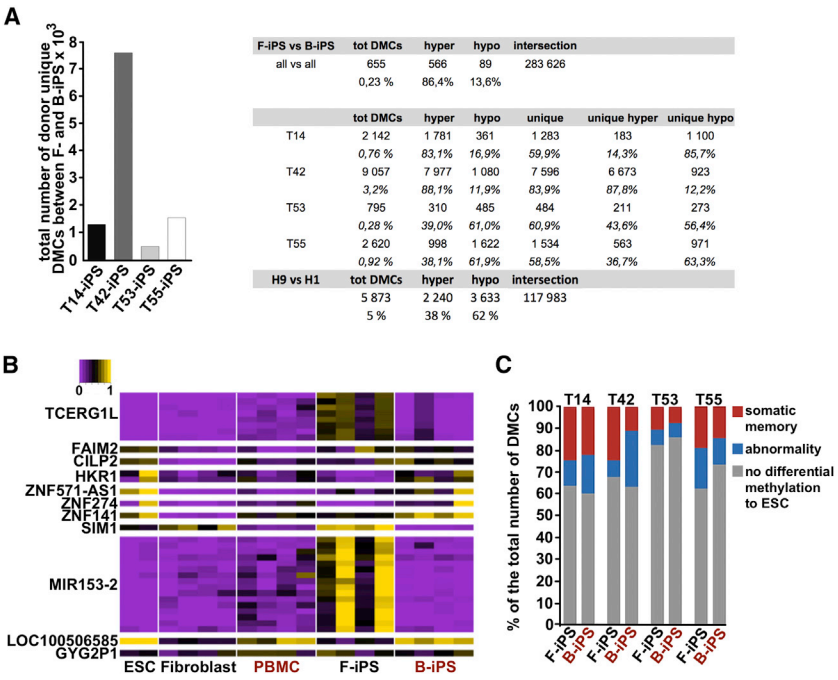
Epigenetic Differences of iPSC Lines Are Minimally Explained by Original Cell Type

In previous reports, different PSCs were shown to harbor unique CpG methylation profiles due to either residual somatic cell memory or aberrant methylation (Lister et al., 2011). To determine whether the epigenetic differences result from the cell type of origin, we grouped iPSC lines into F- and B-iPSCs and compared them using RRBS. In total we identified 655 differentially methylated CpG sites (DMCs) (0.23% of common CGs). Hypermethylated DMCs predominated in B-iPSCs (566 = 86.4% of DMCs) compared with F-iPSCs (Figure 2A).

To eliminate donor-derived variability in methylation profiles, we compared F- and B-iPSCs for each donor. We identified the largest number of DMCs for isogenic iPSCs derived from donor T42 (9,057 = 3.2%) followed by T55 (2,620 = 0.92%), T14 (2,142 = 0.76%), and T53 (795 = 0.28%) (Figure 2A). Of these, an average of 66% were donor-unique DMCs, with the exception of T42-derived F- and B-iPSC lines where 84% of the DMCs were unique for that genetic background (Figure 2A). We identified 34 DMCs common to all four donors (Figure S2E). Notably, eight of these common DMCs are in the *TCERG1L* locus (Lister et al., 2011; Wang et al., 2013) (Figure 2B). The fraction of CpGs differentially methylated between F- and B-iPSCs is lower than between two different hESC lines (H1 and H9) (Bock et al., 2011) (Figure 2A). This again suggested that F- and B-iPSCs are highly similar to each other. To quantify both abnormal methylation and somatic memory phenotypes we compared the methylation signatures of hESCs, isogenic iPSC lines, and somatic parental cells using k-means clustering (Figure S3). We found that the methylation profiles of 7%–25% of DMCs in iPSCs resembled those of the corresponding parental somatic cells (Figure 2C), and often exhibited a donor-specific signature of memory. On average, 70% of DMCs were similar to hESCs (Figure 2C), which is in line with previous reports (Bock et al., 2011).

Donor-Related Variability Influences Expression of Lineage Priming Genes in iPSC Lines

As isogenic iPSC lines showed a tendency to cluster in previous analyses (Figures 1C and S1E), we next selected the top 1,000 genes showing the largest variance in gene expression between all PSC lines (Table S3). Unsupervised hierarchical clustering using the 1,000 most variably expressed genes resulted in clustering of isogenic iPSC lines, supporting our previous observation that differences in gene expression were mostly donor dependent (Figure S4A). However, we also observed some variability



F-iPS vs B-iPS	tot DMCs	hyper	hypo	intersection		
all vs all	655	566	89	283 626		
	0,23 %	86,4%	13,6%			
	tot DMCs	hyper	hypo	unique	unique hyper	unique hypo
T14	2 142	1 781	361	1 283	183	1 100
	0,76 %	*83,1%*	*16,9%*	*59,9%*	*14,3%*	*85,7%*
T42	9 057	7 977	1 080	7 596	6 673	923
	3,2%	*88,1%*	*11,9%*	*83,9%*	*87,8%*	*12,2%*
T53	795	310	485	484	211	273
	0,28 %	*39,0%*	*61,0%*	*60,9%*	*43,6%*	*56,4%*
T55	2 620	998	1 622	1 534	563	971
	0,92 %	*38,1%*	*61,9%*	*58,5%*	*36,7%*	*63,3%*
H9 vs H1	tot DMCs	hyper	hypo	intersection		
	5 873	2 240	3 633	117 983		
	5 %	38 %	62 %			

Figure 2. Methylomic Analyses Demonstrate Minimal Contribution of Source-Cell-Specific Differences to iPSC Variability

(A) Total number of donor-unique, differentially methylated cytosines (DMCs) between genetically matched iPSC lines derived from fibroblasts and blood cells. Analyses were performed for each donor (T14, n = 4; T42, n = 5; T53, n = 5; T55, n = 4). The table shows the number of DMCs per pairwise comparison as well as the number of hypermethylated CpGs and hypomethylated CpGs in B-iPSCs with respect to F-iPSCs and H1 hESCs with respect to H9. RRBS data for H1 hESCs were obtained from the ENCODE project (Meissner et al., 2008).

(B) A heatmap representing methylation level of the 34 common DMCs across all pluripotent stem cells (n = 20) and somatic cell lines (n = 8). DMCs were annotated at the nearest transcription start site. ESC, embryonic stem cells; PBMC, blood cells; F-iPS, iPSCs derived from fibroblast; B-iPS, iPSCs derived from blood.

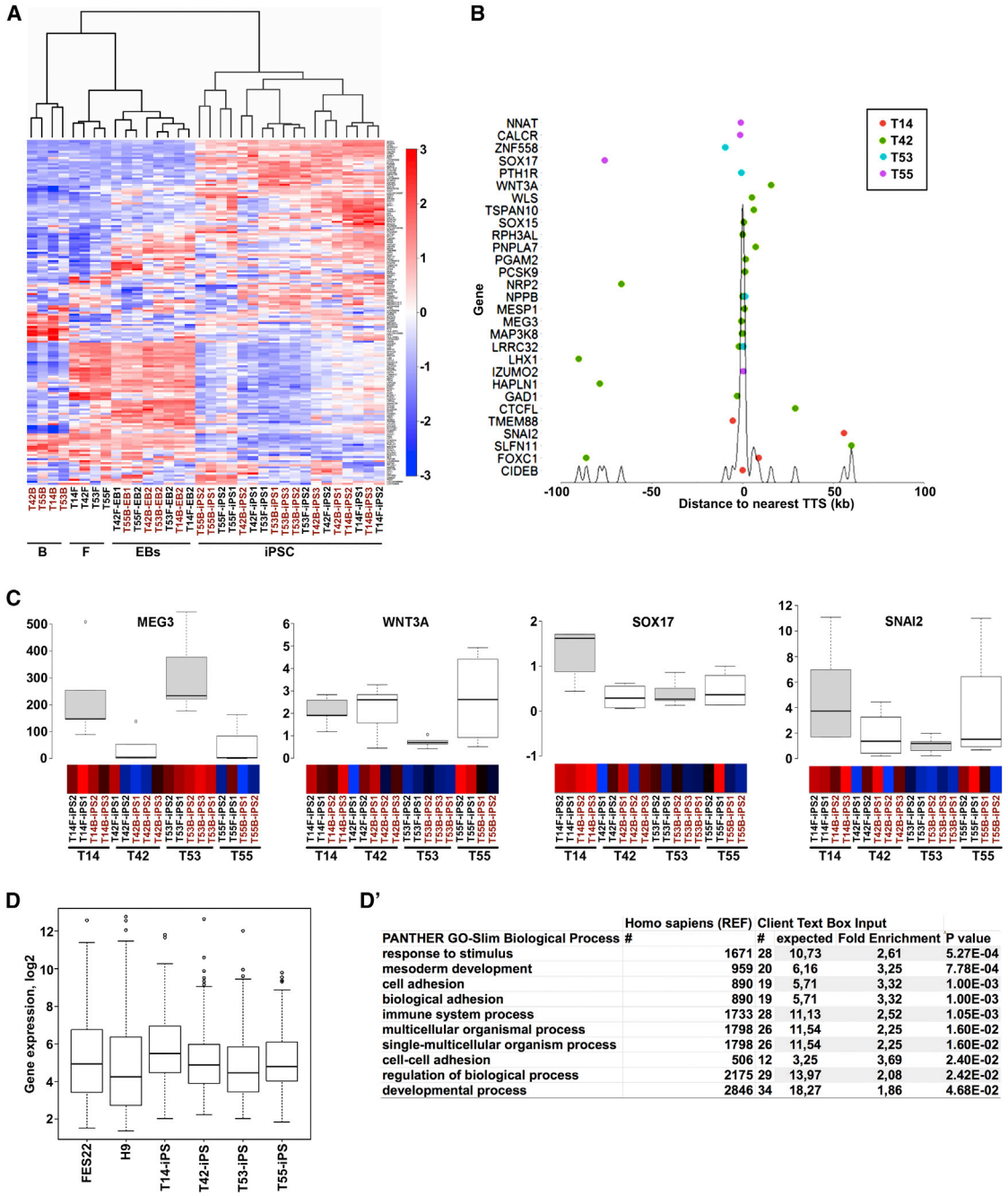
(C) Aberrant methylation and somatic cell memory in genetically matched F- and B-iPSC lines (T14, n = 4; T42, n = 5; T53, n = 5; T55, n = 4) compared with embryonic stem cells (ESC). The bar chart represents the percentage of the total number of DMCs. The list of values can be found in (A). Donors are indicated above the bars.

between isogenic iPSC lines. GO analysis of these 1,000 genes indicated enrichment of categories representative of developmental pathways (Table S4). The oPOSSUM algorithm (Ho Sui et al., 2007) was used to identify regulatory motif over-representation across the most differentially expressed genes. This indicated hits in transcription factors related to the maintenance and differentiation of PSCs (Table S5). To analyze this further, we focused on an independent panel of genes associated with PSCs and their early differentiation, selected by the International Stem Cell Initiative (Adewumi et al., 2007). Unsupervised hierarchical clustering of iPSC lines according to the expression of these genes resulted in clustering according to the donor, confirming our previous findings that donor-dependent characteristics influence expression of genes related to pluripotency and differentiation (Figure S4B).

To further study the statistical differences between isogenic iPSC lines, we grouped them based on the donor (T14, T42, T53, and T55) and performed SAM on the 1,000 most variably expressed genes across all PSC lines. This resulted in the identification of 167 differentially expressed genes between the isogenic iPSC groups from different donors (Table S6). Clustering of all samples according to these genes was visualized as an annotated heatmap (Figure 3A). Principal-component analysis (PCA) using this167-gene set confirmed donor-specific clustering (Figure S4C). Furthermore, when we annotated the DMCs by assigning them to the nearest transcription start site (TSS) (Table S7), we found that 29 of those 167 genes also overlapped with donor-specific DMCs. These DMCs were largely enriched in promoter regions (conventionally defined as ±1 kb from TSS), supporting the hypothesis that epigenetic differences reflect the transcriptional variation between iPSC lines derived from different donors (Figure 3B). qPCR analyses on selected genes (*MEG3*, *WNT3A*, *SOX17*, and *SNAI2*) replicated these results (Figure 3C). GO analysis of these 167 differentially expressed genes primarily indicated biological processes related to early development, which was confirmed using a PANTHER over-representation test (Figures 3D and 3D′). This suggests that donor-based variability has a major influence on the expression of genes related to pluripotency and lineage priming.

Spontaneous Differentiation Potential of F- and B-iPSCs

We then used embryoid body (EB) analysis to investigate the impact of somatic cell type- and donor-dependent characteristics on the spontaneous differentiation potential of F- and B-iPSCs. Correlation clustering of global gene expression showed no specific clustering of EBs (Figure S5A). Although it has been reported that iPSC lines preferentially differentiate into the lineage of the cell type of origin, we were unable to detect any differentiation bias



PANTHER GO-Slim Biological Process	Homo sapiens (REF) #	Client Text Box Input #	expected	Fold Enrichment	P value
response to stimulus	1671	28	10,73	2,61	5.27E-04
mesoderm development	959	20	6,16	3,25	7.78E-04
cell adhesion	890	19	5,71	3,32	1.00E-03
biological adhesion	890	19	5,71	3,32	1.00E-03
immune system process	1733	28	11,13	2,52	1.05E-03
multicellular organismal process	1798	26	11,54	2,25	1.60E-02
single-multicellular organism process	1798	26	11,54	2,25	1.60E-02
cell-cell adhesion	506	12	3,25	3,69	2.40E-02
regulation of biological process	2175	29	13,97	2,08	2.42E-02
developmental process	2846	34	18,27	1,86	4.68E-02

Figure 3. Donor-Dependent Variability Affects Expression of Genes Related to Lineage Priming in iPSCs

(A) Annotated heatmap showing expression of 167 genes across all cell lines (n = 34). The entire list of 167 genes can be found in Table S6. Individual cell lines used are indicated below the heatmap. The color bar on the right side demonstrates the log2 fold changes. B, blood cells; F, fibroblasts; EBs, embryoid bodies.

(B) Donor-unique methylation signatures. The plot shows the distance of the donor-unique DMCs from the nearest transcription start site (TTS) in a ±100-kb region. Genes (29) overlapping with donor-unique DMCs are listed. The density curve (black) shows the enrichment per position.

(C) Verification of the genome-wide expression analysis with qPCR on selected genes (from A and B). Centerlines show the medians of *MEG3*, *WNT3A*, *SOX17*, and *SNAI2* genes relative to human embryonic stem cell line (H9). Box limits indicate the 25th and 75th percentiles as determined by R software. Boxplots represent all iPSC lines for each donor (T14, n = 4; T42, n = 5; T53, n = 5; T55, n = 4; two technical replicates). Heatmaps from genome-wide expression analysis for each gene are shown below the boxplot.

(legend continued on next page)

toward any embryonic germ layer of specific lineage (data not shown) or the hematopoietic lineages in particular (Figure S5B) (gene list according to Bock et al., 2011). Both F- and B-iPSC lines are able to differentiate into derivatives of all three embryonic lineages despite the cell type of origin. To examine whether EBs showed donor-specific differentiation propensities, we analyzed the expression of 167 genes that separated iPSC lines from different donors (Table S6) and found that spontaneously differentiated EBs maintained this difference (Figure 3A). Together, these results indicate that the differences in gene expression at the iPSC stage are maintained through differentiation.

Hematopoietic Cell Lineage Differentiation of Genetically Matched F- and B-iPSC Lines

To determine the functional consequences of donor-related transcriptional and epigenetic differences, we differentiated iPSC lines toward the hematopoietic lineage using a previously described protocol (Ronn et al., 2015; Woods et al., 2011). We selected iPSC lines derived from two donors (T42 and T55) showing the largest transcriptional and epigenetic intra-line variability (Figure 1D). We used fluorescence-activated cell sorting (FACS) to identify frequencies of hematopoietic cells (CD45+), progenitors (CD45+ CD34+), and more developmentally immature hematopoietic cells (CD43+) (Figure 4A). There was no difference between isogenic F- and B-iPSC lines (Figure 4B). Hematopoietic and hematopoietic progenitor cells showed similar frequencies when we compared iPSC differentiation potential between two donors (Figure 4). However, donor T42 yielded significantly fewer mature hematopoietic cells than T55 (Figure 4C). Interestingly, microscopic analysis revealed fewer large hemoglobinized erythroid cell clusters in iPSC lines derived from donor T42 than from T55 (Figure 5A). FACS analysis using erythroblast markers (CD45− CD33 − glycophorin A [GPA]+, transferrin [CD71]+) confirmed the reduced erythroid potential of T42 iPSC lines compared with T55 (Figures 5B and 5C). To evaluate the functionality of these cells, we plated them into methylcellulose and measured their erythroid colony-forming potential. iPSC lines derived from donor T42 yielded fewer erythroid colonies than those from T55 (Figure 5D). These results demonstrate that iPSC lines derived from different donors can possess significant variability in lineage commitment potential irrespective of their cell source.

Characterization of Molecular Mechanisms Underlying Variable Erythroid Differentiation Potential

Next, we applied gene set enrichment analysis (GSEA) to examine the expression of functionally related genes. iPSC clones from donor T42 (F-iPS2 and B-iPS1) showed downregulation of genes (25 of 61 genes; false discovery rate [FDR] = 0.002, normalized enrichment score [NES] = −2.2084) associated with Diamond-Blackfan anemia (DBA), which is functionally characterized by diminished erythroid precursor cells (Figure 5E) (Gazda et al., 2006). The finding was reproducible with two additional clones from the same donor (17 of 61 genes; FDR = 0.001, NES = −2.2400) (Figures S6A and S6B). Altogether, 15 overlapping genes were identified between these two sets (Figure S6C). We also asked whether differentiation potential and the noted differential gene expression differences between the lines can be linked to changes in methylation. We conducted additional comparisons between (T42F-iPS2, T42B-iPS1) and (T55F-iPS2, T55B-iPS1) for DMCs of 25% or more, and annotated to the TSS of the nearest protein coding gene. We found several genes that are likely involved in the differentiation of the lines to contain DMCs (Table S8). For example, we found that CpGs associated with *NUCB2* and *RB1* were more highly methylated in cells derived from donor T55 than T42. Reciprocally, CpGs associated with the genes *CPSF6*, *IBTK*, and *PFDN4* were more highly methylated in T42 than in T55.

Interestingly, the same hematopoietic tendency was observed independently in EBs derived from donor T42 and T55. Global gene expression patterns showed 7- to 16-fold upregulation of embryonic and fetal hemoglobin gene expression (*HBE1*, *HBA2*, *HBG1*, and *HBG2*) in EBs derived from T42 in comparison with T55. Elevated fetal hemoglobin production has been associated with a reduced total number of circulating erythrocytes in DBA patients (Alter, 1979) and a group of disorders called hereditary persistence of fetal hemoglobin (Forget, 1998). Moreover, of the 66 genes present in the DBA gene set noted above, a total of eight of these were maintained as differentially expressed in the EB stage in donor T55 compared with T42: *CASP6*, *CNOT8*, *IFT74*, *LSM5*, *LYPLA1*, *PFDN4*, *PRSS2*, and *SPAST*. In addition, we observed 1- to 12-fold upregulation of megakaryocyte-specific genes (*FLI1*, *MPL*, *GP9/CD42a*, *CD36*, *ITGA2B/CD41*) and a 2- to 30-fold upregulation of myeloid lineage genes (*MPO*, *CSF1R*, *SPI1*, *CSF3R*) in EBs derived from T42 compared with T55 (Figures S6D–S6F).

(D) Boxplots showing the mean values of gene expression (log2) of 167 genes (shown in A and listed in Table S6) in human embryonic stem cell lines (FES22, H9), and iPSC lines derived from four different donors (T14, n = 4; T42, n = 5; T53, n = 5; T55, n = 4).
(D′) Gene ontology analysis for the 167 genes analyzed using the PANTHER over-representation test. Bonferroni correction for multiple testing was applied. Only results with $p < 0.05$ are displayed.

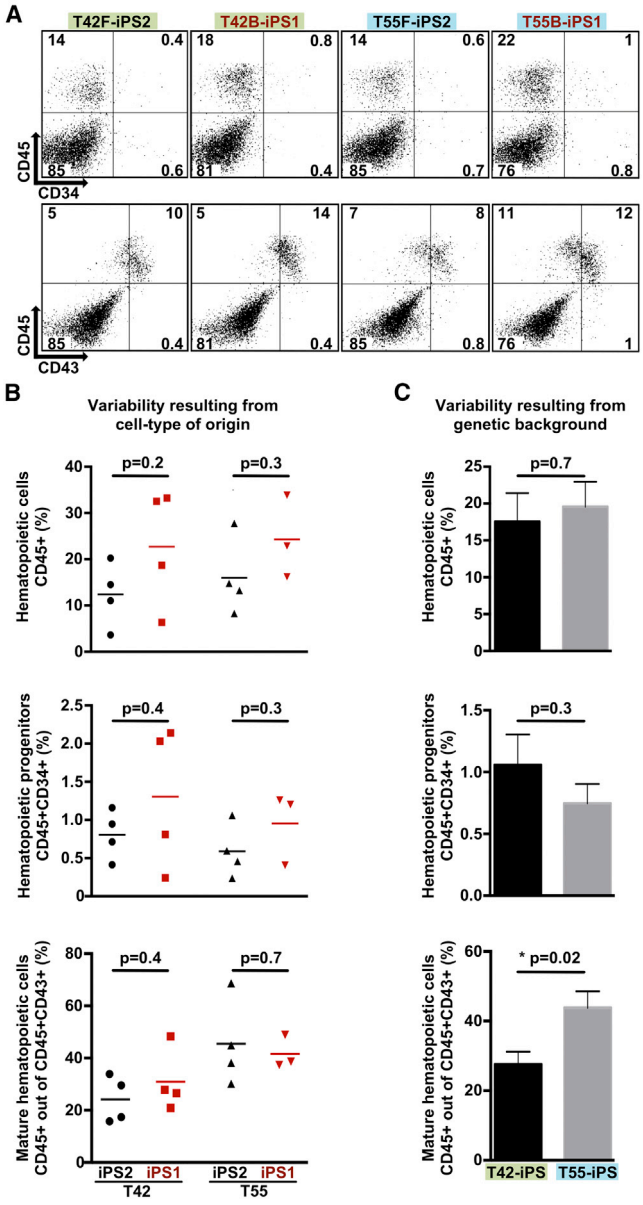


Figure 4. Genetically Matched F- and B-iPSCs Have Similar Hematopoietic Differentiation Capacity

(A) Representative FACS profiles of differentiated genetically matched F- and B-iPSC lines (F in black, B in red) from two donors. Green (T42) and blue (T55) colors mark iPSC lines derived from different donors. The plots show the percentage of hematopoietic cells (CD45+) and hematopoietic progenitors (CD45+ CD34+), immature (CD45+ CD43+), and more mature hematopoietic cells (CD45+ CD43−).

(B) Assessment of variability resulting from cell type of origin. Percentages of hematopoietic cells, hematopoietic progenitors, and mature hematopoietic cells are shown in scatterplots.

(C) Assessment of variability resulting from donor genetic background. Bar graphs show average percentage of hematopoietic cells, hematopoietic progenitors, and mature hematopoietic cells, differentiated from isogenic F- and B-iPSC lines. Donors are indicated in green (T42) and blue (T55). Values are mean ± SEM. Statistics performed by Student's t test, in three or four biological replicates.

DISCUSSION

Four major observations emerge from this study. First, we show that source-cell-specific differences are not retained to a significant extent in isogenic iPSC lines. This is in line with a recent report (Rouhani et al., 2014), but contrasts with earlier studies which observed disruptive retention of somatic cell memory in iPSC lines (Bar-Nur et al., 2011; Kim et al., 2010; Polo et al., 2010). Long-term culture has shown to be advantageous in erasing cell-type-specific memory (Kim et al., 2010; Polo et al., 2010), diminishing transcriptional differences between human iPSC and ESC lines (Chin et al., 2010), and eliminating genetic mosaicism (Hussein et al., 2011). In addition, we used the all-in-one type footprint free SeVdp-iPS system for generation of uniform iPSC lines (Nishimura et al., 2011). We performed molecular analyses at later passages and obtained highly similar molecular signatures for genetically matched iPSC lines derived from two different tissues, even though fibroblasts and PBMCs include multiple cell populations with distinct epigenetic and transcriptional landscapes (Sorrell and Caplan, 2004; Zhang and Huang, 2012).

Our data show that the majority of transcriptional and epigenetic signatures present in iPSCs are donor determined. This is well in line with recent studies that have suggested the influence of genetic background on transcription (Rouhani et al., 2014; Shao et al., 2013) and differentiation of iPSC lines (Kajiwara et al., 2012; Mills et al., 2013). Also, embryonic stem cells derived from individual donors are shown to maintain line-specific signatures and have distinct differentiation potentials (Bock et al., 2011; Chen et al., 2009; Osafune et al., 2008). Our present data point out that donor-dependent signatures specifically affect gene expressions involved in early embryonic lineage specification, resulting in variability between iPSC lines.

Third, we show that iPSC differentiation propensities are significantly biased by donor-dependent variability and not by cell type of origin. Earlier studies have focused on the relationship between genetic variability and molecular signatures of iPSC lines (Kajiwara et al., 2012; Rouhani et al., 2014; Shao et al., 2013). However, only limited information was available on the contribution of donor background to the functional differences of iPSC lines arising from the source-cell-specific differences. This was thoroughly addressed in the present study by combining global

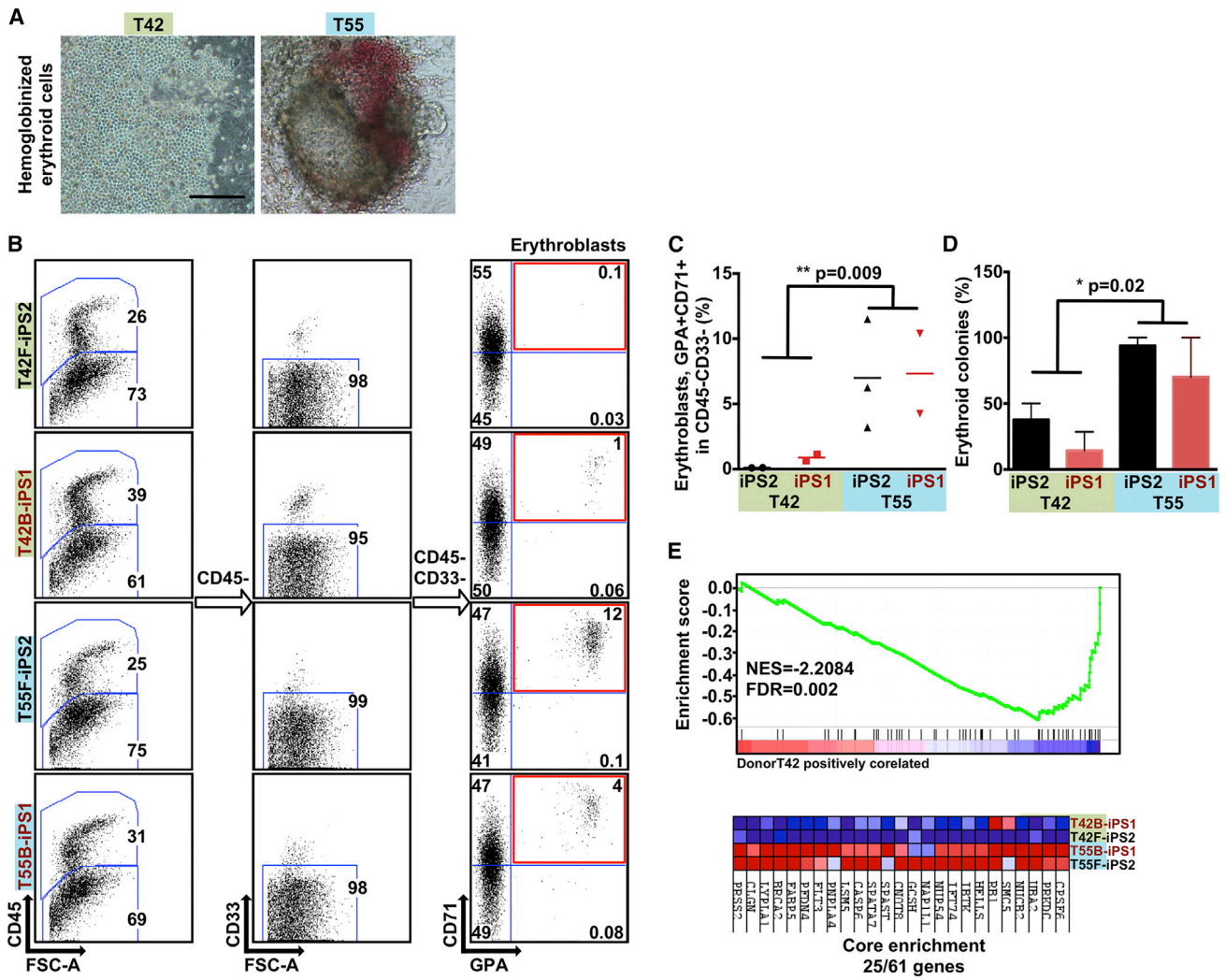


Figure 5. iPSCs Derived from Different Donors Show Variable Erythroid Lineage Differentiation Propensity

(A) Representative micrographs showing the hematopoietic cells differentiated from iPSCs with reduced hemoglobinized erythroid cells in iPSCs derived from T42 (left panel). Scale bar, 100 μm.

(B) Representative FACS plots of differentiated genetically matched F- and B-iPSC lines (n = 4) showing the percentage of erythroblasts (GPA+ CD71+ gated from CD45− CD33−).

(C) Scatter graphs presenting percentage of erythroblasts generated from the genetically matched F- and B-iPSC lines. Each dot represents an independent experiment.

(D) Percentage of erythroid colonies out of the total colony number generated from iPSC lines. Values are mean ± SEM. Statistics performed by Student's t test, in three or four biological replicates.

(E) Gene set enrichment plot and heatmap showing the genes constituting the core enrichment of overlap between multi-potent progenitors sorted from the bone marrow of DBA patients and iPSC lines derived from T42 cells. T42 iPSC lines show decreased expression levels of genes that are downregulated in Diamond-Blackfan anemia patients.

transcriptional and epigenetic analyses with spontaneous and targeted differentiation of isogenic iPSC lines. Although we selected iPSC lines from healthy donors showing the highest intra-individual variation in transcriptional and epigenetic analyses, we could not detect major differences in the differentiation potential of isogenic iPSC lines originating from fibroblasts and blood. However, we detected donor-dependent transcriptional differences in spontaneously differentiated EBs and significant variation in erythroid differentiation potential between iPSC lines derived from two healthy donors, regardless of the cell type of origin. Using GSEA we were able to associate the low erythroid-forming potential of the healthy donor to genes previously indicated in DBA. Moreover, the

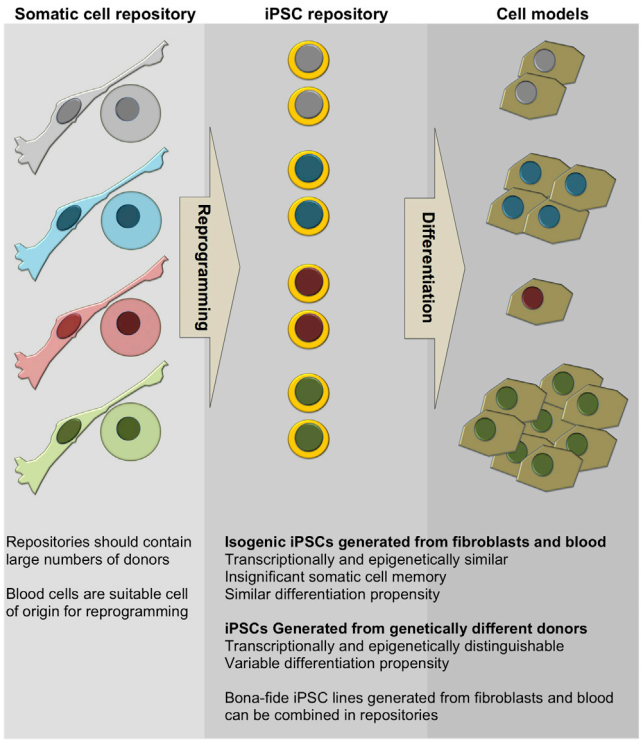


Figure 6. Implications for Biobanking

iPSCs generated from genetically different donors show transcriptional and epigenetic variation, which is reflected in variable differentiation propensities. iPSC lines generated from genetically matched fibroblasts and blood cells are molecularly and functionally similar, implying that iPSCs derived from different tissues can be combined in repositories.

differential expression of these genes in different donors at the iPSC stage was, at least in part, maintained through differentiation to EBs, providing further evidence in support of the donor-related differences affecting the differentiation potential of iPSC lines.

Finally, an important practical implication of our data is that both fibroblasts and blood can be used for the generation of comparable iPSC lines for large-scale biobanking purposes (Figure 6). Our results demonstrate that variability among small cohorts of iPSCs may lead to erroneous conclusions (Sandoe and Eggan, 2013). Because of the inherent differences resulting from the donor-dependent variability, it seems obvious that relatively large cohorts of iPSC lines from different donors, rather than several isogenic clones from a few donors, would be needed to obtain reliable results concerning the impact of donor-specific variants. However, the fact that intra-individual clonal variation is still present after careful technical standardization and iPSC characterization also suggests that genetically matched clones from the donor should be available in biobanks.

EXPERIMENTAL PROCEDURES

Ethical Statement

Skin biopsies and blood samples were collected from Takotsubo cardiomyopathy patients with written consent permitted by the Ethical Committee for Internal Diseases of the Hospital District of Helsinki and Uusimaa (permit no: 352/13/03/01/2009). In brief, Takotsubo cardiomyopathy mimics acute myocardial infarction with similar symptoms and findings, but without coronary artery disease. Patients develop transient congestive heart failure under emotional or physical stress, but recover fully. Takotsubo occurs almost exclusively among post-menopausal women. All donors were females without any hematological medical condition.

Reprogramming, Cell Culture, and Fingerprinting

Fibroblasts were grown from skin biopsies under glass plates in DMEM + 20% fetal calf serum and antibiotics. Mononuclear cells were extracted freshly from blood by Ficoll-extraction method. PBMCs (1×10^5 to 1×10^6) and fibroblasts were transduced with SeVdp as previously described (Nishimura et al., 2011; Trokovic et al., 2014). Cells were plated on mitomycin C-treated murine embryonic fibroblasts (3.75×10^5 feeder cells/well) on a six-well plate in hES medium: DMEM/F12 with GlutaMAX, supplemented with 20% KO-serum replacement, 0.1 mM β-mercaptoethanol, 1% non-essential amino acids (all from Life Technologies), and 6 ng/ml basic fibroblast growth factor (bFGF; Sigma). For feeder-free cultures iPSC lines were grown on Matrigel (growth factor reduced, BD Biosciences) in StemPro (Life Technologies) or in Essential-8 medium (E8, Life Technologies). The donor identity of all iPSC lines was confirmed by microsatellite marker analyses.

Characterization of iPSC Lines

iPSC lines were characterized for expression of stem cell markers by RT-PCR and immunofluorescence microscopy as previously described (Trokovic et al., 2013, 2014). For immunofluorescence analyses we used stem cell-specific antibodies against TRA-1-60 (Millipore, MAB4360), NANOG (Cell Signaling, D73G4), SSEA4 (Millipore, MAB4304), and OCT4 (Santa Cruz Biotechnology, sc-9081). Cell nuclei were stained with DAPI (Vectashield, Vector).

The raw expression data of all iPSC clones was analyzed by PluriTest (http://www.pluritest.org/) and only the clones that successfully passed the test were selected for further analyses (Muller et al., 2011).

EBs were grown from each iPSC clone in low-attachment plates (Corning) in hES medium without bFGF. After 21 days, total RNA was extracted for expression analyses using an AllPrep DNA/RNA/Protein kit (Qiagen). Germ layer-specific expression of EBs was analyzed using antibodies which recognize β-III-tubulin (R&D Systems, MAB1195), AFP (DAKO, A0008), and vimentin (Dako, M0725).

Gene Expression Profiling

Total genomic DNA/RNA was extracted from all cells using AllPrep DNA/RNA/Protein kit (Qiagen) according to the manufacturer's instructions. Gene expression analysis was performed at the Institute for Molecular Medicine Finland (FIMM) Technology Center,

University of Helsinki, using Illumina Human-12 v4 Expression BeadChips. For further analyses, data were processed by removal of background and quantile normalization. The data analysis was performed with R statistics and Chipster (Kallio et al., 2011). All iPSC lines analyzed were between passages (p) 9 and 17. Genes were filtered on the basis of their SD (2 SDs = 95%).

Comparative analysis between F- and B-iPSCs was performed using LPE test. p-Value cutoff was <0.05 and the FDR was controlled by adjusting the p value with Benjamini-Hochberg. An interactive tool for comparing lists with Venn diagrams was done at http://bioinfogp.cnb.csic.es/tools/venny/index.html. Gene lists were analyzed with DAVID (Huang da et al., 2009). For pathway analysis we used the PANTHER Classification System (Thomas et al., 2003). Microarray data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-3825.

Microarray data were confirmed using qPCR. cDNAs for qPCR reactions were produced from the extracted RNA samples (2 μg) by Maxima First Strand cDNA synthesis kit (Thermo Scientific). qPCR was performed with EvaGreen qPCR mix (Solis Biodyne) in Corbett RotorGene. Primers used NANOG F/R (CTC AGC CTC CAG CAG ATG C/TA GAT TTC ATT CTC TGG TTC TGG); SOX17 F/R (CCG AGT TGA GCA AGA TGC TG/T GCA TGT GCT GCA CGC GCA); SNAI2 F/R (GGT TGC TTC AAG GAC ACA TTA G/TT GAC CTG TCT GCA AAT GCT C); MEG3 F/R (AAG GAC CAC CTC CTC TCC AT/A GGA AAC CGT GCT CCT AGT G); and WNT3A (GCC CCA CTC GGA TAC TTC T/GG CAT GAT CTC CAC GTA GT).

DNA Methylation Analysis

RRBS was performed on 500 ng of genomic DNA for all cell lines as previously described (Gu et al., 2010). All libraries have been sequenced to an average of 1,500,000 individual CpG per sample at a genomic coverage of 5× or higher. Raw reads were mapped using Bismark (Krueger and Andrews, 2011), and methylation calling was performed using a custom method. Methylation analysis, unsupervised hierarchical clustering, PCA, and pairwise comparisons were performed using MethylKit. Differential methylation was calculated using Fisher's exact test. A cutoff of 25% was applied based on a difference in methylation level. p Values were adjusted using SLIM method and a cutoff of FDR $\leq$ 0.01 was applied. Genomic annotation was performed using Homer annotation (Heinz et al., 2010). Data analysis was performed with the R statistics package (http://www.r-project.org/). RRBS data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-3859.

Differentiation of Human iPSCs into Hematopoietic Cells

Hematopoietic cell differentiation was performed as previously described (Ronn et al., 2015; Woods et al., 2011). Micrographs of cells from the cultures were taken at day 22 prior to whole well harvest. Floating and individualized cells were pooled, washed, and divided into two samples. One sample was used for assessment of hematopoietic lineage markers using FACS. The other sample was plated into methylcellulose (MethoCult H4435, StemCell Technologies) for colony-forming unit assay. The cells for FACS analysis were stained for mouse anti-human CD45/CD43/CD34/CD71 antibodies (BD Pharmingen; CD71APC, cat. no. 551374, clone M-A712. CD45 FITC or PE cat. no. 555482, 555483, clone HI30. CD43FITC cat. no. 555475, clone 1G10. CD34APC cat. no. 555824, clone 581), CD33/CD235a (GPA) antibodies (eBioscience, CD235a(Gly-A) cat. no. 12-9987-82, clone HIR2(GA-R2). CD33PE-Cy7 cat. no. 25-0338-42 clone WM-53(WM53)) and analyzed by a FACSCanto II flow cytometer.

Microarray data were evaluated at the level of gene sets to define and quantitate trends in gene expression similar to published data. Ranked gene lists were created and submitted to the online public repository provided by the BROAD Institute for Gene Set Enrichment Analysis (GSEATo) (Subramanian et al., 2005). Venn diagram analysis was performed using Microsoft PowerPoint tool software.

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and eight tables and can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2015.12.009.

AUTHOR CONTRIBUTIONS

A.K., R.T., T.O., N.-B.W., and D.H. conceived the study and designed the experiments. O.P. and J.S. provided samples for the study. M.N., K.N., and M.O. established and provided Sendai virus vector for reprogramming experiments. C.V., C.A., K.K.P., and D.H. performed epigenetic analyses. R.M. and N.-B.W. performed differentiation and functional analyses. B.V.H., J.K., and J.W. helped in the bioinformatics analyses. A.K. and R.T. performed all other experiments and analyzed the data. A.J., T.O., N.-B.W., and D.H. provided funding for the work. All authors contributed to writing of the manuscript.

ACKNOWLEDGMENTS

This work was supported by the Academy of Finland (Grant No. 141482 and No. 257157), the Sigrid Juselius Foundation, Research Funds of the Helsinki University Central Hospital, The Swedish Research Council, AFA Insurance (Sweden), Lund University Medical Faculty, the HematoLinné Program Grant, Stem Therapy Program Grant, and the Crafoord Foundation (N.-B.W.). J.K. was supported through funds from the Academy of Finland (grant No. 283045). We thank Ms. Anne Nyberg for excellent technical assistance. We thank participants of the Takotsubo study, and Dr. Robert Leigh for constructive comments and proofreading of the manuscript. Roksana Moraghebi and Cristina Valensisi contributed equally to the study.

Received: September 14, 2015
Revised: December 8, 2015
Accepted: December 14, 2015
Published: January 14, 2016

REFERENCES

Adewumi, O., Aflatoonian, B., Ahrlund-Richter, L., Amit, M., Andrews, P.W., Beighton, G., Bello, P.A., Benvenisty, N., Berry, L.S., Bevan, S., et al. (2007). Characterization of human embryonic

stem cell lines by the International Stem Cell Initiative. Nat. Biotechnol. *25*, 803–816.

Alter, B.P. (1979). Fetal erythropoiesis in stress hematopoiesis. Exp. Hematol. *7* (*Suppl 5*), 200–209.

Bar-Nur, O., Russ, H.A., Efrat, S., and Benvenisty, N. (2011). Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells. Cell Stem Cell *9*, 17–23.

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H., et al. (2011). Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell *144*, 439–452.

Chen, A.E., Egli, D., Niakan, K., Deng, J., Akutsu, H., Yamaki, M., Cowan, C., Fitz-Gerald, C., Zhang, K., Melton, D.A., et al. (2009). Optimal timing of inner cell mass isolation increases the efficiency of human embryonic stem cell derivation and allows generation of sibling cell lines. Cell Stem Cell *4*, 103–106.

Chin, M.H., Pellegrini, M., Plath, K., and Lowry, W.E. (2010). Molecular analyses of human induced pluripotent stem cells and embryonic stem cells. Cell Stem Cell *7*, 263–269.

Elo, L.L., Filen, S., Lahesmaa, R., and Aittokallio, T. (2008). Reproducibility-optimized test statistic for ranking genes in microarray studies. IEEE/ACM Trans. Comput. Biol. Bioinform. *5*, 423–431.

Forget, B.G. (1998). Molecular basis of hereditary persistence of fetal hemoglobin. Ann. N. Y. Acad. Sci. *850*, 38–44.

Gazda, H.T., Kho, A.T., Sanoudou, D., Zaucha, J.M., Kohane, I.S., Sieff, C.A., and Beggs, A.H. (2006). Defective ribosomal protein gene expression alters transcription, translation, apoptosis, and oncogenic pathways in Diamond-Blackfan anemia. Stem Cells *24*, 2034–2044.

Gu, H., Bock, C., Mikkelsen, T.S., Jager, N., Smith, Z.D., Tomazou, E., Gnirke, A., Lander, E.S., and Meissner, A. (2010). Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution. Nat. Methods *7*, 133–136.

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

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Ho Sui, S.J., Fulton, D.L., Arenillas, D.J., Kwon, A.T., and Wasserman, W.W. (2007). oPOSSUM: integrated tools for analysis of regulatory motif over-representation. Nucleic Acids Res. *35*, W245–W252.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Hussein, S.M., Batada, N.N., Vuoristo, S., Ching, R.W., Autio, R., Narva, E., Ng, S., Sourour, M., Hamalainen, R., Olsson, C., et al. (2011). Copy number variation and selection during reprogramming to pluripotency. Nature *471*, 58–62.

Jain, N., Thatte, J., Braciale, T., Ley, K., O'Connell, M., and Lee, J.K. (2003). Local-pooled-error test for identifying differentially expressed genes with a small number of replicated microarrays. Bioinformatics *19*, 1945–1951.

Kajiwara, M., Aoi, T., Okita, K., Takahashi, R., Inoue, H., Takayama, N., Endo, H., Eto, K., Toguchida, J., Uemoto, S., et al. (2012). Donor-dependent variations in hepatic differentiation from human-induced pluripotent stem cells. Proc. Natl. Acad. Sci. USA *109*, 12538–12543.

Kallio, M.A., Tuimala, J.T., Hupponen, T., Klemela, P., Gentile, M., Scheinin, I., Koski, M., Kaki, J., and Korpelainen, E.I. (2011). Chipster: user-friendly analysis software for microarray and other high-throughput data. BMC Genomics *12*, 507.

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I., et al. (2010). Epigenetic memory in induced pluripotent stem cells. Nature *467*, 285–290.

Kim, K., Zhao, R., Doi, A., Ng, K., Unternaehrer, J., Cahan, P., Huo, H., Loh, Y.H., Aryee, M.J., Lensch, M.W., et al. (2011). Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. Nat. Biotechnol. *29*, 1117–1119.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.

Leek, J.T., Scharpf, R.B., Bravo, H.C., Simcha, D., Langmead, B., Johnson, W.E., Geman, D., Baggerly, K., and Irizarry, R.A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. Nat. Rev. Genet. *11*, 733–739.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., et al. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature *471*, 68–73.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., et al. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell *1*, 55–70.

Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. *33*, 5868–5877.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766–770.

Mills, J.A., Wang, K., Paluru, P., Ying, L., Lu, L., Galvao, A.M., Xu, D., Yao, Y., Sullivan, S.K., Sullivan, L.M., et al. (2013). Clonal genetic and hematopoietic heterogeneity among human-induced pluripotent stem cell lines. Blood *122*, 2047–2051.

Muller, F.J., Schuldt, B.M., Williams, R., Mason, D., Altun, G., Papapetrou, E.P., Danner, S., Goldmann, J.E., Herbst, A., Schmidt, N.O., et al. (2011). A bioinformatic assay for pluripotency in human cells. Nat. Methods *8*, 315–317.

Nashun, B., Hill, P.W., and Hajkova, P. (2015). Reprogramming of cell fate: epigenetic memory and the erasure of memories past. EMBO J. *34*, 1296–1308.

Nishimura, K., Sano, M., Ohtaka, M., Furuta, B., Umemura, Y., Nakajima, Y., Ikehara, Y., Kobayashi, T., Segawa, H., Takayasu, S., et al. (2011). Development of defective and persistent Sendai virus vector: a unique gene delivery/expression system ideal for cell reprogramming. J. Biol. Chem. *286*, 4760–4771.

Ohi, Y., Qin, H., Hong, C., Blouin, L., Polo, J.M., Guo, T., Qi, Z., Downey, S.L., Manos, P.D., Rossi, D.J., et al. (2011). Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. Nat. Cell Biol. *13*, 541–549.

Osafune, K., Caron, L., Borowiak, M., Martinez, R.J., Fitz-Gerald, C.S., Sato, Y., Cowan, C.A., Chien, K.R., and Melton, D.A. (2008). Marked differences in differentiation propensity among human embryonic stem cell lines. Nat. Biotechnol. *26*, 313–315.

Polo, J.M., Liu, S., Figueroa, M.E., Kulalert, W., Eminli, S., Tan, K.Y., Apostolou, E., Stadtfeld, M., Li, Y., Shioda, T., et al. (2010). Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat. Biotechnol. *28*, 848–855.

Ronn, R.E., Guibentif, C., Moraghebi, R., Chaves, P., Saxena, S., Garcia, B., and Woods, N.B. (2015). Retinoic acid regulates hematopoietic development from human pluripotent stem cells. Stem Cell Rep. *4*, 269–281.

Rouhani, F., Kumasaka, N., de Brito, M.C., Bradley, A., Vallier, L., and Gaffney, D. (2014). Genetic background drives transcriptional variation in human induced pluripotent stem cells. PLoS Genet. *10*, e1004432.

Sandoe, J., and Eggan, K. (2013). Opportunities and challenges of pluripotent stem cell neurodegenerative disease models. Nat. Neurosci. *16*, 780–789.

Santostefano, K.E., Hamazaki, T., Biel, N.M., Jin, S., Umezawa, A., and Terada, N. (2015). A practical guide to induced pluripotent stem cell research using patient samples. Lab. Invest. *95*, 4–13.

Shao, K., Koch, C., Gupta, M.K., Lin, Q., Lenz, M., Laufs, S., Denecke, B., Schmidt, M., Linke, M., Hennies, H.C., et al. (2013). Induced pluripotent mesenchymal stromal cell clones retain donor-derived differences in DNA methylation profiles. Mol. Ther. *21*, 240–250.

Smyth, G.K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. *3*, Article3.

Sorrell, J.M., and Caplan, A.I. (2004). Fibroblast heterogeneity: more than skin deep. J. Cell Sci. *117*, 667–675.

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

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Tchieu, J., Kuoy, E., Chin, M.H., Trinh, H., Patterson, M., Sherman, S.P., Aimiuwu, O., Lindgren, A., Hakimian, S., Zack, J.A., et al. (2010). Female human iPSCs retain an inactive X chromosome. Cell Stem Cell *7*, 329–342.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. *13*, 2129–2141.

Trokovic, R., Weltner, J., Manninen, T., Mikkola, M., Lundin, K., Hamalainen, R., Suomalainen, A., and Otonkoski, T. (2013). Small molecule inhibitors promote efficient generation of induced pluripotent stem cells from human skeletal myoblasts. Stem Cells Dev. *22*, 114–123.

Trokovic, R., Weltner, J., Nishimura, K., Ohtaka, M., Nakanishi, M., Salomaa, V., Jalanko, A., Otonkoski, T., and Kyttala, A. (2014). Advanced feeder-free generation of induced pluripotent stem cells directly from blood cells. Stem Cells Transl. Med. *3*, 1402–1409.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA *98*, 5116–5121.

Wang, T., Wu, H., Li, Y., Szulwach, K.E., Lin, L., Li, X., Chen, I.P., Goldlust, I.S., Chamberlain, S.J., Dodd, A., et al. (2013). Subtelomeric hotspots of aberrant 5-hydroxymethylcytosine-mediated epigenetic modifications during reprogramming to pluripotency. Nat. Cell Biol. *15*, 700–711.

Woods, N.B., Parker, A.S., Moraghebi, R., Lutz, M.K., Firth, A.L., Brennand, K.J., Berggren, W.T., Raya, A., Izpisua Belmonte, J.C., Gage, F.H., et al. (2011). Brief report: efficient generation of hematopoietic precursors and progenitors from human pluripotent stem cell lines. Stem Cells *29*, 1158–1164.

Zhang, M., and Huang, B. (2012). The multi-differentiation potential of peripheral blood mononuclear cells. Stem Cell Res. Ther. *3*, 48.

Stem Cell Reports, Volume *6*

Supplemental Information

Genetic Variability Overrides the Impact of Parental Cell Type and Determines iPSC Differentiation Potential

Aija Kyttälä, Roksana Moraghebi, Cristina Valensisi, Johannes Kettunen, Colin Andrus, Kalyan Kumar Pasumarthy, Mahito Nakanishi, Ken Nishimura, Manami Ohtaka, Jere Weltner, Ben Van Handel, Olavi Parkkonen, Juha Sinisalo, Anu Jalanko, R. David Hawkins, Niels-Bjarne Woods, Timo Otonkoski, and Ras Trokovic

Figure S 1, related to Figure 1. Characterization of iPSC lines

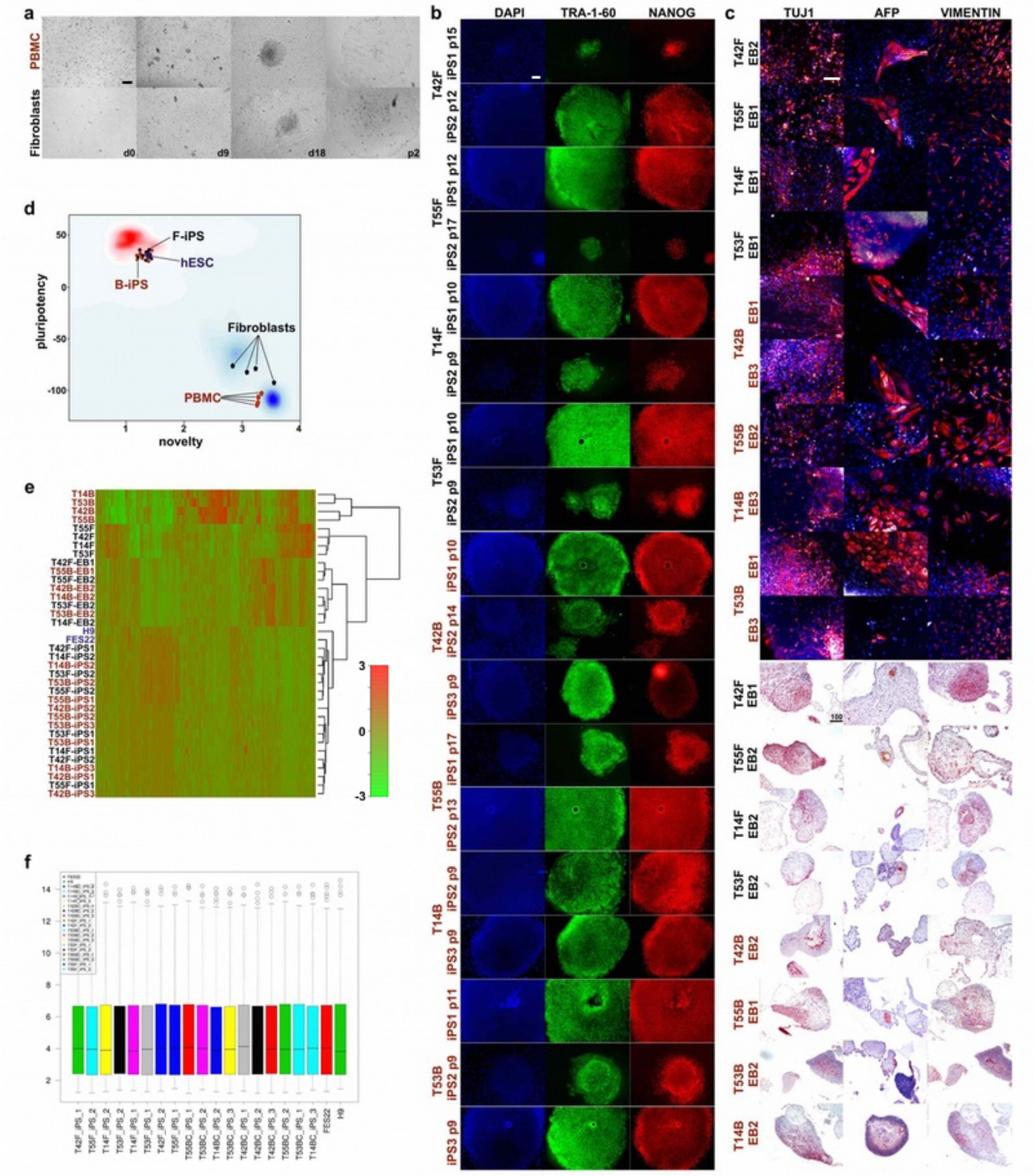


Figure S1, related to Figure 1. Characterization of iPSC lines.
(a) Representative microscopic images of emerging iPSC colonies (bright field) from PBMCs (T55B) and fibroblasts (T55F). **(b)** Expression of stem cell markers in all iPSC lines (n=18) derived from fibroblasts (F-iPS) and PBMCs (B-iPS). TRA-1-60 (green), NANOG (red). Nuclear staining DAPI (blue). **(c)** Differentiation of iPSC lines detected by immunostaining with ectodermal (TUJ1), endodermal (AFP), and mesodermal (VIMENTIN) lineage markers in the cells derived from embryoid bodies (EBs). Nuclear staining (DAPI). **(d)** Pluripotent transcriptional profile measured in PluriTest™. F-iPS (n=8; black dots), B-iPS (n=10; red dots), and hESC (n=2; purple squares) qualified as pluripotent as determined by the Pluripotency and Novelty scores. Donor PBMCs and fibroblasts are plotted in the previously referenced somatic cells (blue cloud). **(e)** A heatmap showing global gene expression levels in all samples (n=36). Samples and genes are clustered using Pearson correlation as a distance measure and average linkage for constructing the dendogram. **(f)** X-chromosome located gene expression boxplot of all pluripotent stem cell lines used in the study (n=20). Bars in a,b 200μm and in c 100μm.

Figure S 2, related to Figure 1 and 2. Analysis of cell type memory in F- and B-iPSCs.

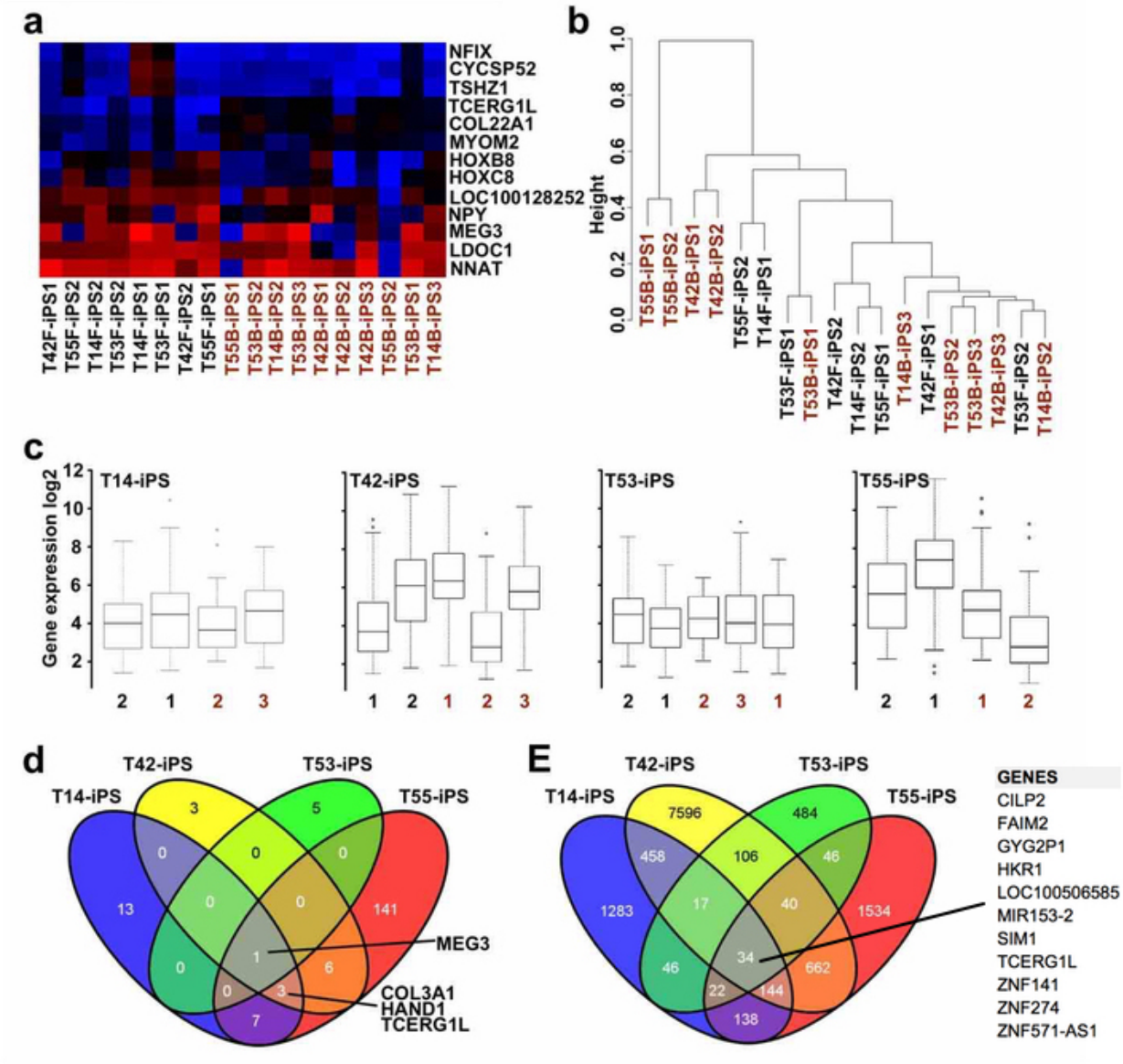


Figure S2, related to Figure 1 and 2. Analysis of source-cell-specific differences in F- and B-iPSCs.
(a) Hierarchical clustering of 13 differentially expressed genes between F- (n=8) and B-iPSC (n=10) groups (genes are listed in Table S2). Fold change (FC) cut-off>1. **(b)** Unsupervised hierarchical clustering of iPSC lines according to 13 genes in (a). **(c)** Boxplots showing the gene expression (log2) of isogenic F- and B-iPSC lines for each donor. Results are filtered using standard deviation (3SDs=99.7%). iPSC clone numbers are indicated below box plots for each donor (T14, n=4; T42, n=5; T53, n=5; T55, n=4). **(d)** Venn diagram representing overlap of differentially expressed genes between isogenic F- and B-iPSCs for each donor (T14, n=4; T42, n=5; T53, n=5; T55, n=4). The entire list of genes can be found as Table S1. **(e)** Venn diagram representing differentially methylated CpGs (DMCs) between isogenic F- and B-iPSCs for each donor (T14, n=4; T42, n=5; T53, n=5; T55, n=4). Genomic annotation of the 34 DMCs common to all the donor are listed.

Figure S 3, related to Figure 2. Abnormal methylation and somatic memory in isogenic iPSC lines.

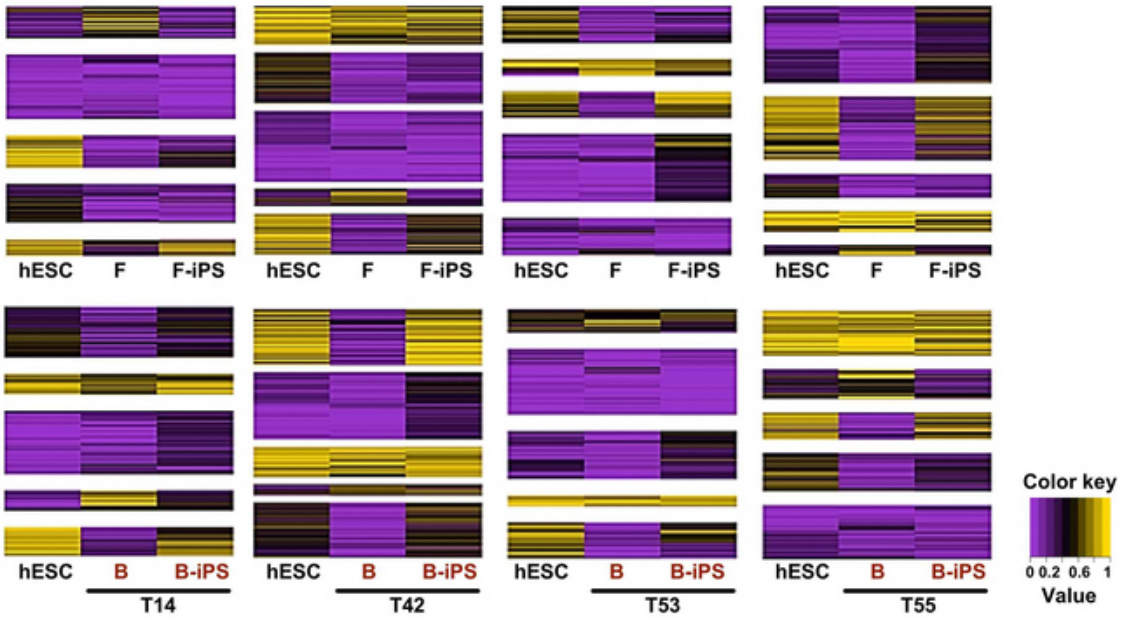


Figure S3, related to Figure 2. Abnormal methylation and source-cell-specific differences in isogenic iPSC lines.
K-means (K=5) clustering heatmap representation of the methylation level of the donor-unique DMCs in hESC (H9), iPSC lines and parental fibroblasts (F) and blood cells (B) per each isogenic group (T14, n=4; T42, n=5; T53, n=5; T55, n=4) (See Figure 2A). F-iPSC lines are shown in the top row; B-iPSC lines are shown at the bottom. The number of CpGs shown for donors T14, T42, T53 and T55 is 1,283; 7,596; 484 and 1,534 respectively.

Figure S 4, related to Figure 3.

Analyses of differentially expressed genes between iPSC lines.

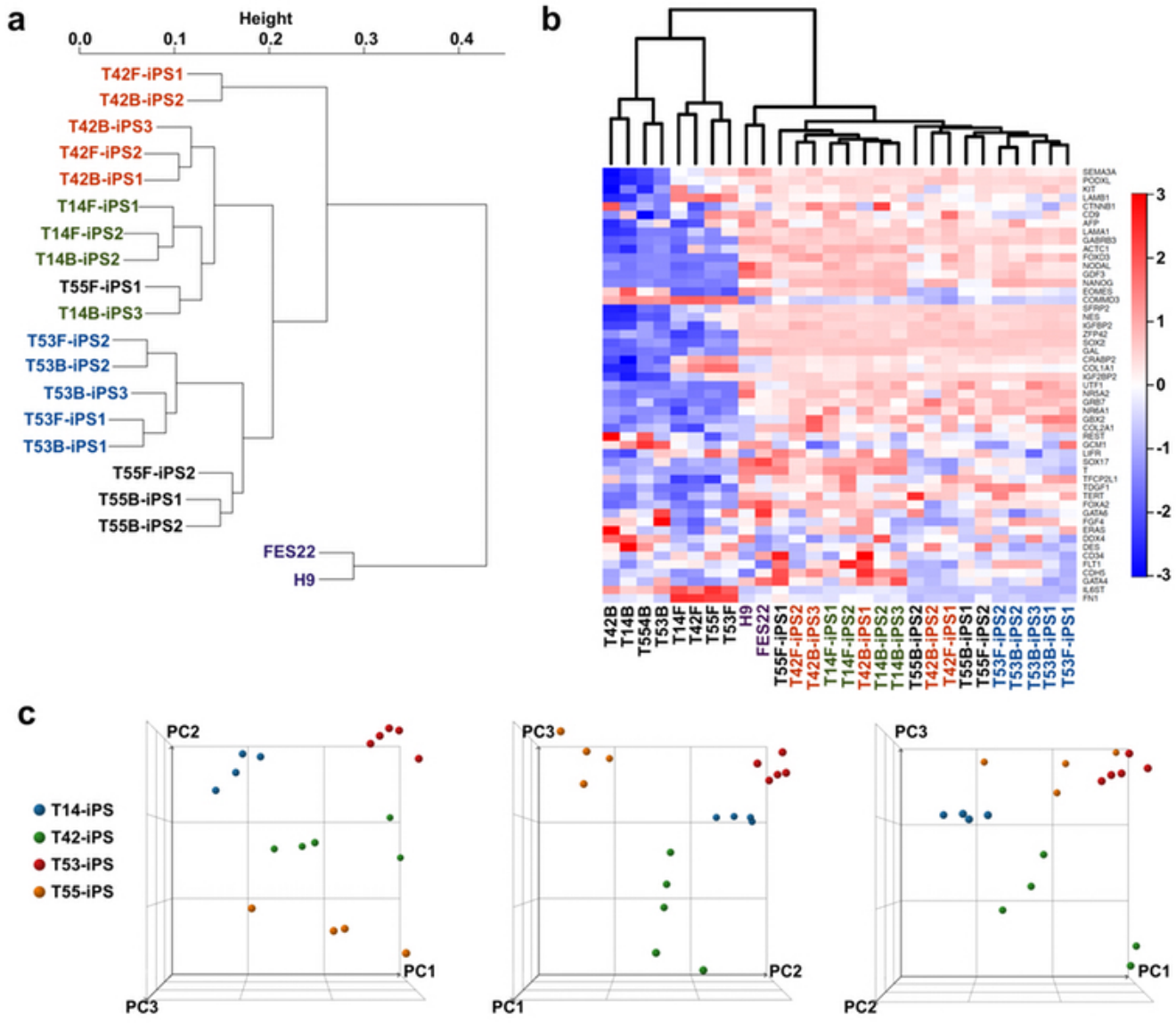


Figure S4, related to Figure 3. Analyses of differentially expressed genes between iPSC lines.
(a) Unsupervised hierarchical clustering of all iPSC samples using the 1,000 most variably expressed genes across all stem cell lines. The entire list of 1,000 genes can be found as Table S3. Note that in this and the next figure iPSC lines are exceptionally coloured according to the donor (T42, red; T14, green; T53, blue; and T55, black). **(b)** Annotated heatmap of expression of genes associated with stem cells and differentiation, selected by the International Stem Cell Initiative (Adewumi et al., 2007) in all pluripotent stem cells (n=20), blood cells (B, n=4), and fibroblasts (F, n=4). The colour bar on the right side demonstrates the log2 fold changes. **(c)** Correlation plots between first three principal components of the expression in 167 genes that showed the largest standard deviation for all iPSC lines (n=18) as described in Figure 3a, the entire list of 167 genes can be found as Table S6. The three first components explained a very large proportion of variance (70%) in the expression data cumulatively. The three plots show donor specific clustering but not clustering due to cell type. Colour codes for each donor is shown on the left.

Figure S5, related to Figure 4. Gene expression analysis of spontaneously differentiated F- and B-iPSCs

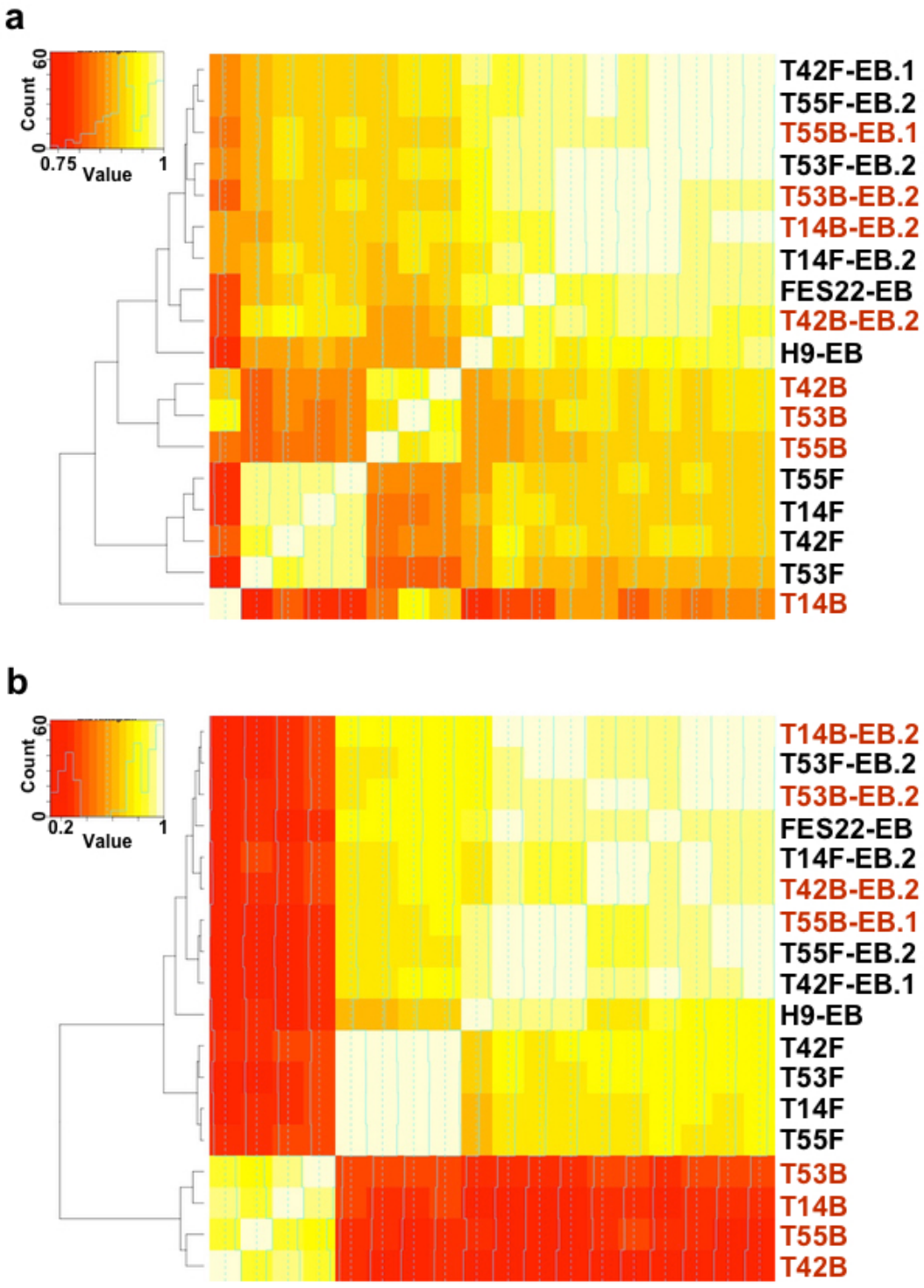
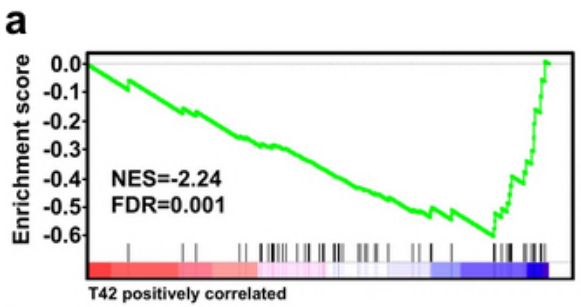
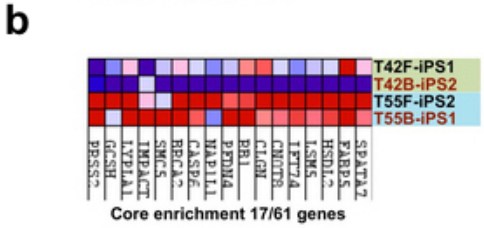
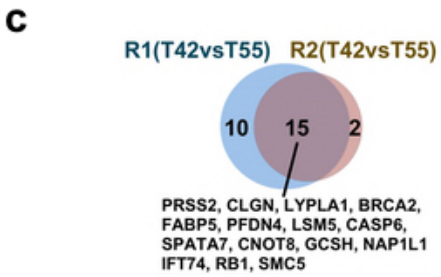


Figure S5, related to Figure 4. Gene expression analysis of spontaneously differentiated F- and B-iPSCs.
(a) Correlation heatmap clustering of global gene expression of F- (n=4) and B-iPSCs (n=4) derived embryoid bodies (EBs), hESCs derived EBs (H9 and FES22) and parental fibroblasts (F, n=4) and blood cells (B, n=4). **(b)** Correlation heatmap clustering of hematopoietic gene expression (Bock et al., 2011) of F- and B-iPSCs derived EBs. F-iPSC derived EBs are marked in black and B-iPSCs derived EBs in red colour.

Figure S 6, related to Figure 5. Gene set enrichment analysis (GSEA) of two additional iPSC lines derived from donors T42 and T55







d

Symbol	FoldChange_Abs	Pvalue
HBG1	16	0,024
HBG2	8,3	0,029
HBE1	7	0,007
NFE2	6,9	0,01
HBA2	6,6	0,002

e

Symbol	FoldChange_Abs	Pvalue
MPO	30,5	0,006
CSF1R	11,9	0,014
SPI1	8,6	0,004
CSF3R	2,9	0,028

f

Symbol	FoldChange_Abs	Pvalue
GP9 (CD42a)	12,6	0,017
CD36	11,3	0,003
MPL (CD110)	5,7	0,004
PPBP	4,6	0,015
VWF	3,5	0,036
FLI1	2,2	0,02
ITGA2B (CD41)	1,8	0,04

Figure S6, related to Figure 5. Gene set enrichment analysis (GSEA) of additional iPSC lines derived from the donors T42 and T55.
(a) Comparison of two additional iPSC lines derived from T42 cells (n=2) to the same iPSC lines derived from T55 cells (n=2) using GSEA also revealed significant depletion of DBA-associated genes. **(b)** Heat map showing the genes constituting the core enrichment of overlap between multipotent progenitors sorted from the bone marrow of DBA patients and two additional iPSC lines derived from T42 cells. **(c)** Venn diagram depicting the overlap (15 out of a possible 17 core enrichment genes) between the two comparisons. **(d)** Upregulated embryonic and fetal haemoglobin related genes in EBs T42 vs. T55 (n=4). **(e)** Upregulated myeloid lineage genes in EBs T42 vs. T55 (n=4). **(f)** Upregulated megakaryocyte lineage genes in EBs T42 vs. T55 (n=4).

Titles and Legends for supplemental tables submitted as separate excel files:

Table S1 (related to Figure 1). Differentially expressed genes between isogenic F-iPSC and B-iPSC lines. Local-pooled-error test (LPE) was used for the identification of 24 (T14B- vs F-iPSCs), 13 (T42B- vs F-iPSCs), 6 (T53B-vs F-iPSCs), and 158 (T55B- vs F-iPSCs) differentially expressed genes between the isogenic iPSC lines

Table S2 (related to Figure 1). Differentially expressed genes between isogenic F- and B-iPSC lines FC cut off >1

Table S3 (related to Figure 3 and Figure S4). The top 1,000 genes showing the largest variance in gene expression between all PSC lines

Table S4(related to Figure 3 and Figure S4). Gene Ontology analysis of 1,000 most differentially expressed genes between all PSC

Table S5 (related to Figure S4b). Analysis of regulatory motif over-representation across the most differentially expressed genes

Table S6 (related to Figure 3A and S4). **Identification of 167 differentially expressed genes between the isogenic iPSC groups**.
Isogenic iPSC lines were grouped based on the donor (T14, T42, T53, and T55). Significance Analysis of Microarrays (SAM) was performed on the 1,000 most variably expressed genes across all PSC lines. This resulted in the identification of 167 differentially expressed genes between the isogenic iPSC groups from different donors.

Table S7 (related to Figure 3B). Gene regions identified by nearest TSS.
Differential methylation was calculated in pairwise comparisons using Fisher's exact. A cut off of 25% was applied to the difference in methylation level. P-values were adjusted using SLIM method and a cut off of FDR <= 0.01 was applied. Donor-unique differential methylated CpGs (DMCs) were defined as the DMCs in the isogenic comparison F- vs B-iPSCs that were unique to each donor. The table lists the gene annotation based on the nearest transcription start site (TSS) for the donor-unique DMCs in isogenic F-iPSC vs B-iPSC comparisons.

Table S8 (related to Figure 5 and S6).. Gene regions identified by nearest TSS between T42 and T55 derived iPSC lines.
Comparisons between (T42F-iPS2, T42B-iPS1) and (T55F-iPS2, T55B-iPS1) for DMCs of 25% or more, and annotated to the nearest transcription start site (TSS).