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

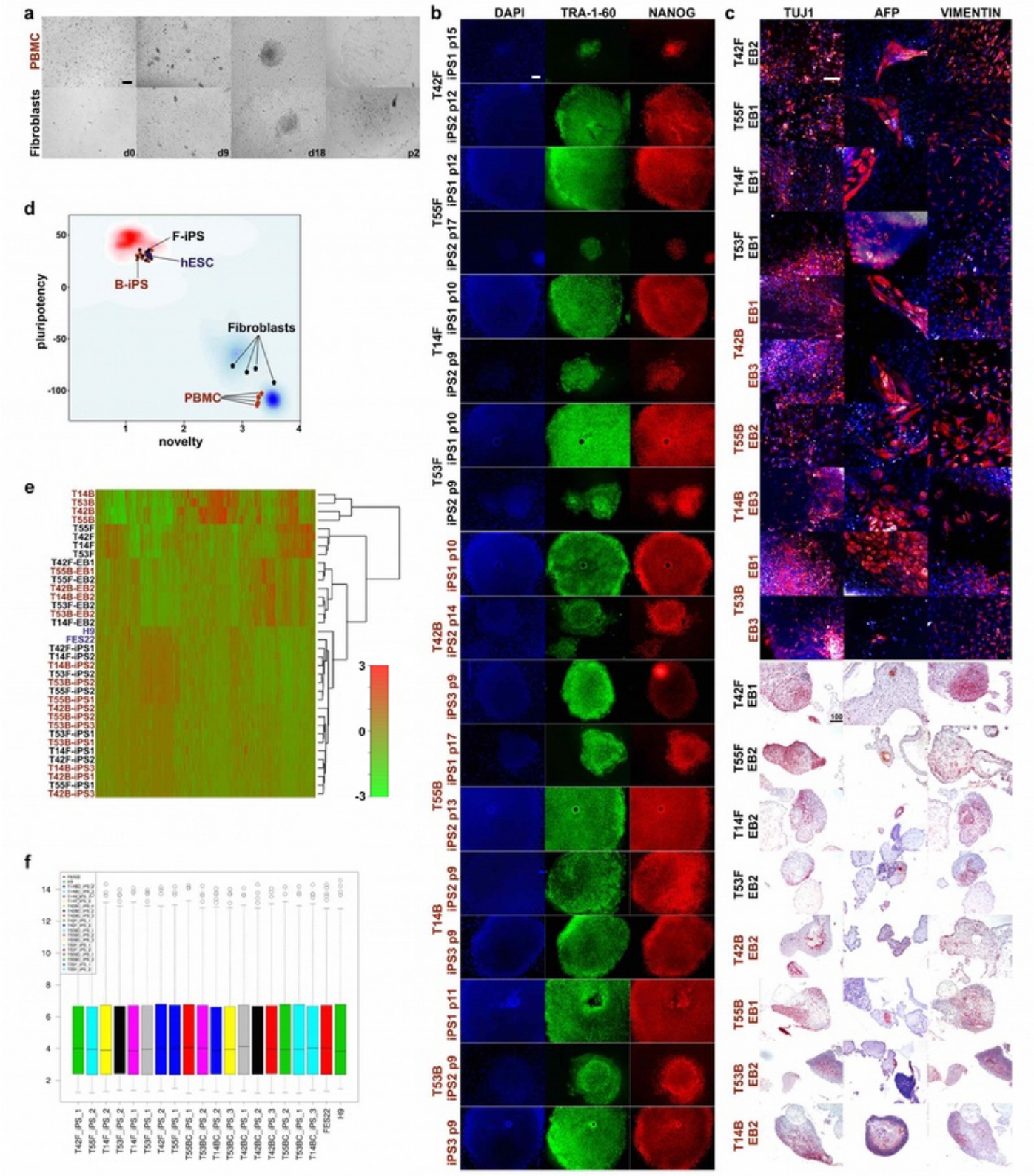


Figure S1, related to Figure 1. Characterization of iPSC lines.
(a) Representative microscopic images of emerging iPSC colonies (bright field) from PBMCs (T55B) and fibroblasts (T55F). **(b)** Expression of stem cell markers in all iPSC lines (n=18) derived from fibroblasts (F-iPS) and PBMCs (B-iPS). 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.

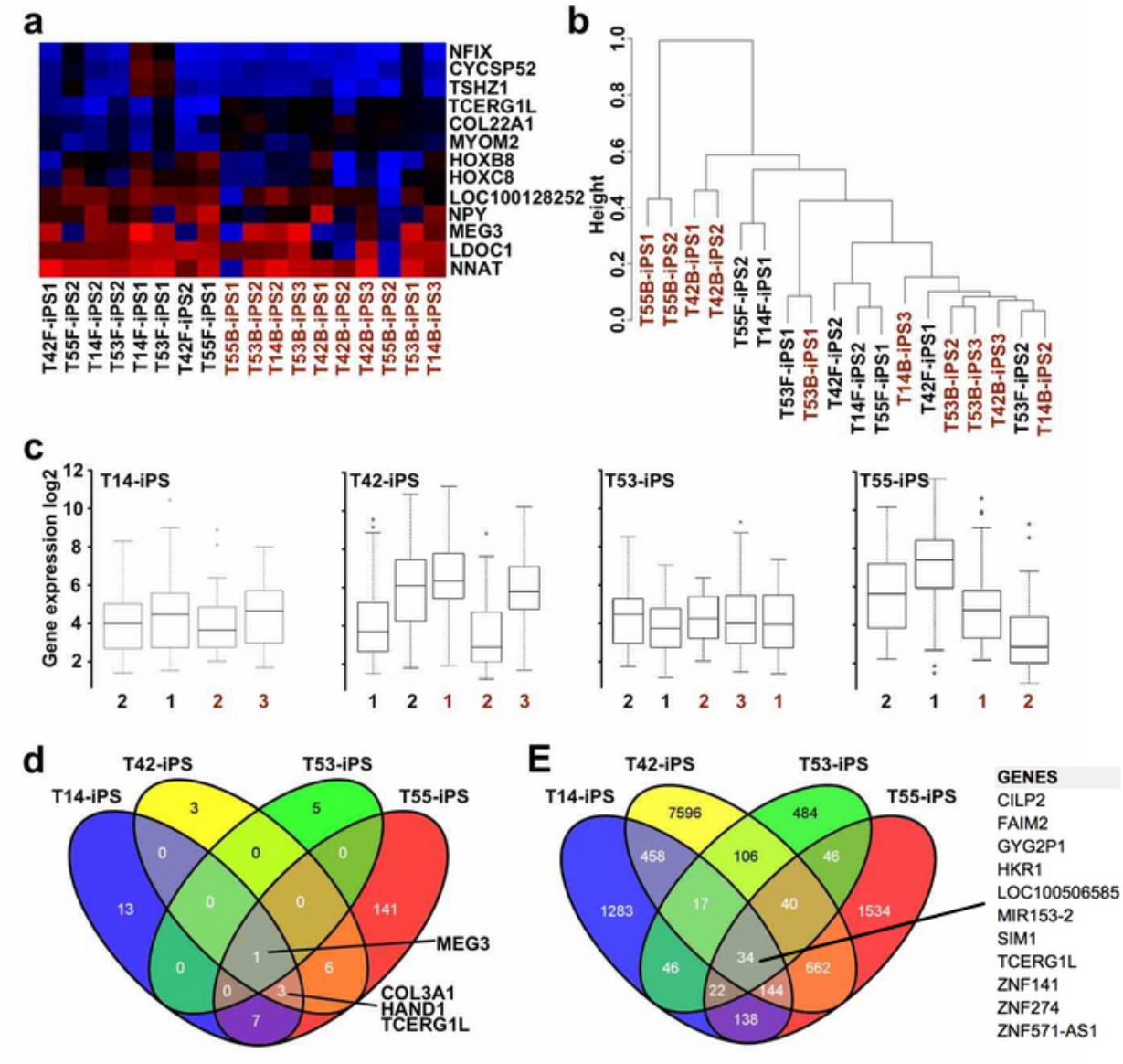


Figure S2, related to Figure 1 and 2. Analysis of source-cell-specific differences in F- and B-iPSCs.
(a) Hierarchical clustering of 13 differentially expressed genes between F- (n=8) and B-iPSC (n=10) groups (genes are listed in Table S2). Fold change (FC) cut-off>1. **(b)** Unsupervised hierarchical 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.

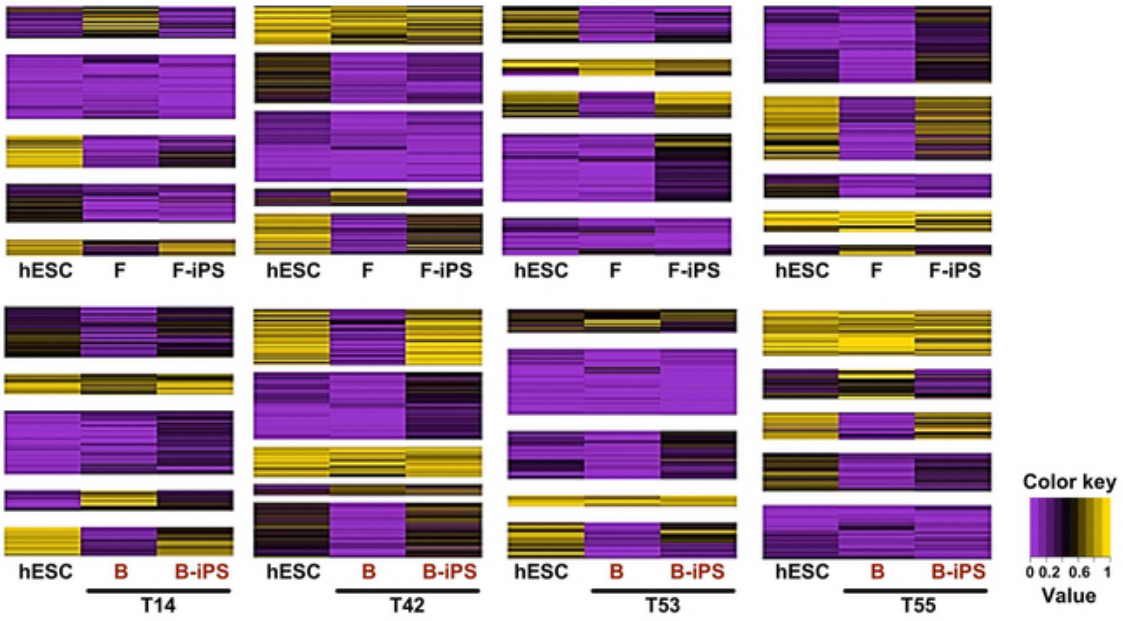


Figure S3, related to Figure 2. Abnormal methylation and source-cell-specific differences in isogenic iPSC lines.
K-means (K=5) clustering heatmap representation of the methylation level of the donor-unique DMCs in hESC (H9), iPSC lines and parental fibroblasts (F) and blood cells (B) per each isogenic 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.

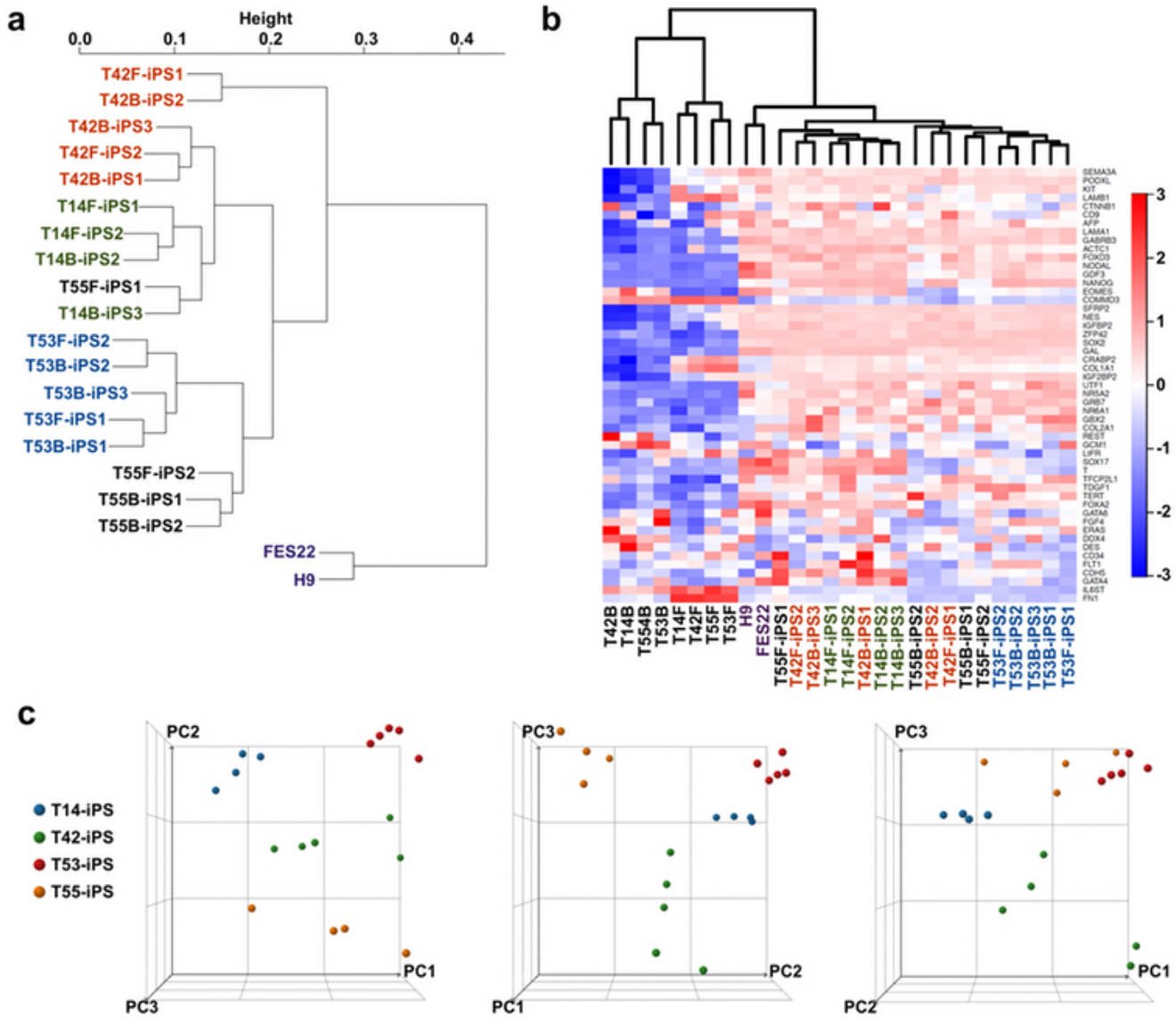


Figure S4, related to Figure 3. Analyses of differentially expressed genes between iPSC lines.
(a) Unsupervised hierarchical clustering of all iPSC samples using the 1,000 most variably expressed genes across all stem cell lines. The entire list of 1,000 genes can be found as Table S3. Note that in this 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

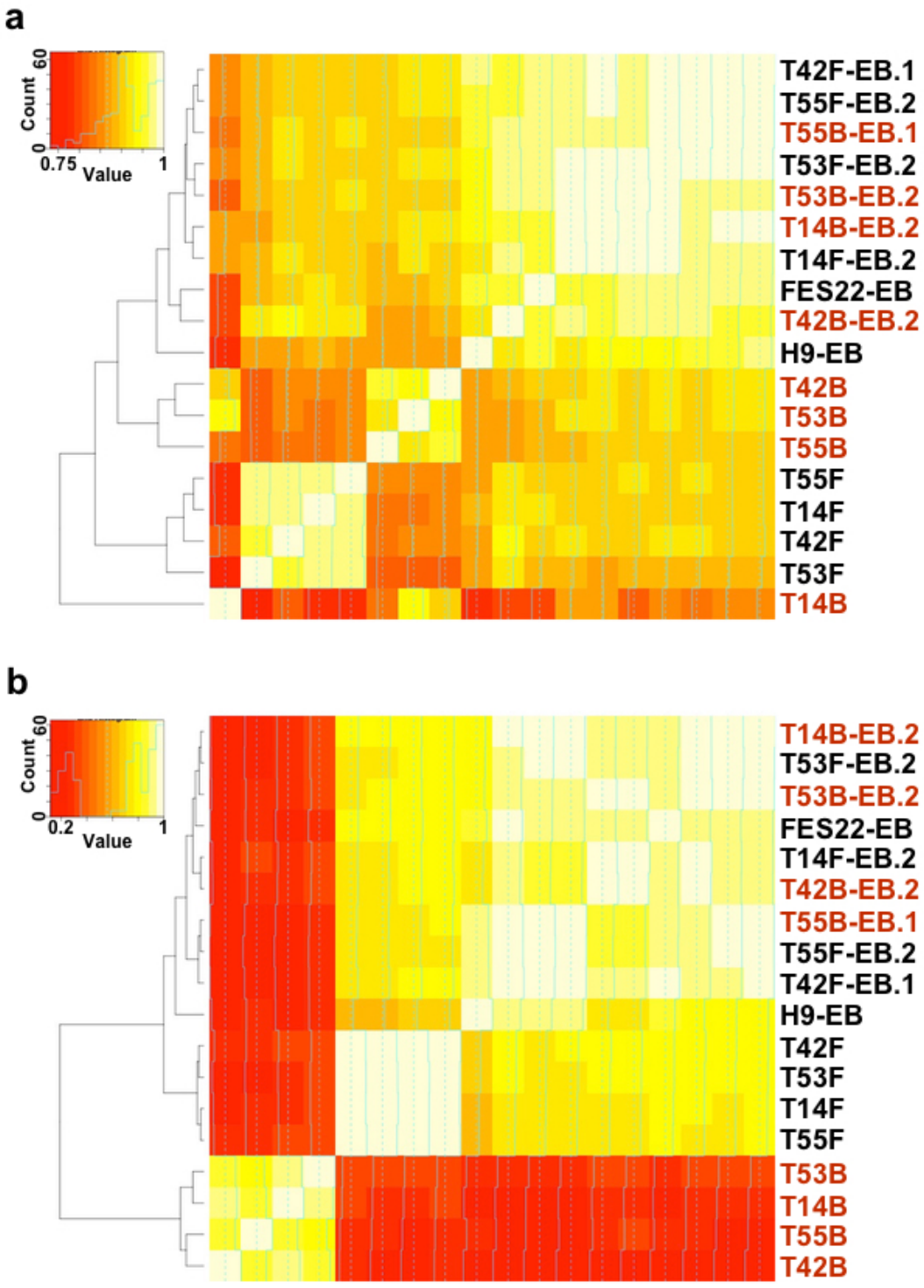
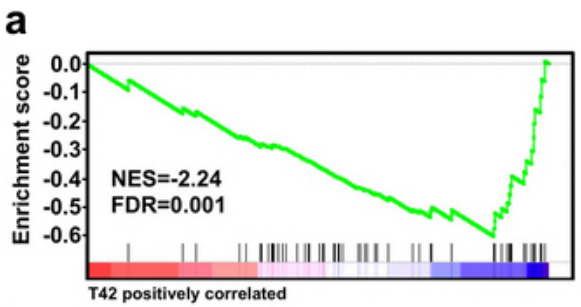
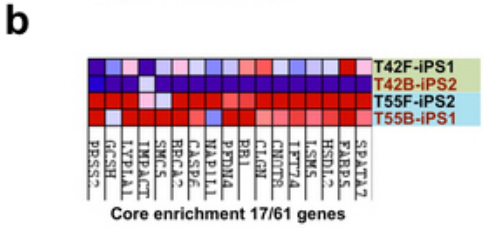
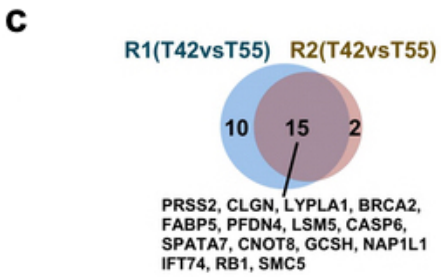


Figure S5, related to Figure 4. Gene expression analysis of spontaneously differentiated F- and B-iPSCs.
(a) Correlation heatmap clustering of global gene expression of F- (n=4) and B-iPSCs (n=4) derived embryoid bodies (EBs), hESCs derived EBs (H9 and FES22) and parental fibroblasts (F, n=4) and blood 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).