

# Epigenetic variation between human induced pluripotent stem cell lines is an indicator of differentiation capacity

Masatoshi Nishizawa<sup>1,2</sup>, Kazuhisa Chonabayashi<sup>1,2</sup>, Masaki Nomura<sup>1</sup>, Azusa Tanaka<sup>1</sup>, Masahiro, Nakamura<sup>1</sup>, Azusa Inagaki<sup>1</sup>, Misato Nishikawa<sup>1</sup>, Ikue Takei<sup>1</sup>, Akiko Oishi<sup>1</sup>, Koji Tanabe<sup>1</sup>, Mari Ohnuki<sup>1</sup>, Hidaka Yokota<sup>1</sup>, Michiyo Koyanagi-Aoi<sup>1</sup>, Keisuke Okita<sup>1</sup>, Akira Watanabe<sup>1,3</sup>, Akifumi Takaori-Kondo<sup>2</sup>, Shinya Yamanaka<sup>1,3,4</sup>, and Yoshinori Yoshida<sup>1,\*</sup>

<sup>1</sup>Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto 606-8507, Japan

<sup>2</sup>Department of Hematology and Oncology, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan

<sup>3</sup>Institute for Integrated Cell-Material Sciences, Kyoto University, Kyoto 606-8501, Japan

<sup>4</sup>Gladstone Institute of Cardiovascular Disease, San Francisco, California 94158, USA

\*Correspondence: [yoshinor@cira.kyoto-u.ac.jp](mailto:yoshinor@cira.kyoto-u.ac.jp)

**Running Title Hematopoietic Differentiation from Human iPS/ESC Lines**

## Summary

Variation in the differentiation capacity of induced pluripotent stem cells (iPSCs) to specific lineages is a significant concern for their use in clinical applications and disease modeling. To identify factors that affect differentiation capacity, we performed integration analyses between hematopoietic differentiation performance and molecular signatures such as gene expression, DNA methylation, and chromatin status, using 35 human iPSC lines and 4 ESC lines. Our analyses revealed that hematopoietic commitment of PSCs to hematopoietic precursors correlates with IGF2 expression level, which in turn depends on signaling-dependent chromatin accessibility at mesendodermal genes. Maturation capacity for conversion of PSC-derived hematopoietic precursors to mature blood associates with the amount and pattern of DNA methylation acquired during reprogramming. Our study therefore provides insight into the molecular features that determine the differential capacities seen among human iPSCs lines, and through the predictive potential of this information, highlight a way to select optimal iPSCs for clinical applications.

**Key words:** Human induced pluripotent stem cell, Hematopoietic differentiation, mRNA microarray, DNA methylation array, Reprograming-associated DNA methylation, Chromatin Accessibility, Insulin-like growth factor 2

1

## 2 **Introduction**

3       The potential of human induced pluripotent stem cells (iPSCs) (Takahashi et al.,  
4   2007; Takahashi and Yamanaka, 2006) to differentiate into many cell types makes them  
5   a promising source for regeneration therapy, drug screening, and pathogenetic study.

6   However, this potential is compromised by the large variation that exists among human  
7   iPSC lines in their differentiation capacity to specific lineages. One of the factors  
8   attributed to this variation is the residual DNA methylation signatures that remain after  
9   the reprogramming process to pluripotency, a phenomenon known as somatic memory  
10   (Kim et al., 2010; Kim et al., 2011; Lister et al., 2011; Ohi et al., 2011; Polo et al., 2010).

11   Another factor which could cause variation among iPSC lines is aberrations during the  
12   reprogramming process, with aberrations in DNA methylation being most investigated  
13   (Huang et al., 2014a; Koyanagi-Aoi et al., 2013; Lister et al., 2011; Nazor et al., 2012;  
14   Ruiz et al., 2012; Stadtfeld et al., 2010). Finally, a third factor is genetic differences  
15   among individual donors of human PSCs (Kajiwara et al., 2012). While all these  
16   theories deserve consideration, overall, reports that have investigated variations in  
17   human iPSC lines used sample numbers that were too small for reliable conclusions.

18       In the present study, we assessed the hematopoietic differentiation capacity of  
19   35 human iPSC lines from 4 different parental tissues and 4 embryonic stem cell (ESC)

lines by evaluating both the early phase (commitment phase) and late phase (maturation phase) of hematopoietic differentiation. In addition, we investigated gene expression, DNA methylation, and open chromatin accessibility, and performed correlation analysis of the differentiation capacities and these molecular signatures. The analysis revealed that in iPSCs, commitment capacity was associated with insulin-like growth factor 2 (IGF2) gene expression, which was under the control of fibroblast growth factor (FGF)/Activin signaling-dependent chromatin accessibility. In contrast, maturation capacity was associated with aberrant DNA methylation acquired during the reprogramming process. Our study provides important insights on the molecular mechanism of differentiation capacity and ways to select optimal iPSCs for clinical applications.

## **Results**

### **Hematopoietic Commitment Capacity is Associated with Expression Levels of IGF2 in Undifferentiated PSCs, but Not with Original Cell Type**

To investigate the differentiation capacity among human iPSCs from different origins, we first assessed the hematopoietic differentiation capacities of 35 iPSC lines derived from 4 types of somatic tissues (human dermal fibroblasts (HDF), hematopoietic cells such as cord blood (CB) and peripheral blood (PB), dental pulp

(DP) cells, and keratinocytes (Kera)) from 15 donors and of 4 ESC lines (Figure 1A, S1A-C, Table S1) in early phase (commitment capacity) and late phase (maturation capacity). For differentiation, we utilized a previously reported in vitro differentiation system with some modification (Grigoriadis et al., 2010).

We assessed commitment capacity by the number of generated CD43+ hematopoietic cells on day 15 (Figure 1B, S1D). The median percentage of CD43+ cells from the 39 lines was 72.7% (Figure S1E), indicating efficient differentiation of our protocol compared to previous reports (Grigoriadis et al., 2010). In addition, the analysis revealed no significant association between commitment capacity and iPSC origin (Figure 1C, S1F).

To identify the factors that influence commitment capacity, we compared gene expressions in the undifferentiated state between PSCs with high commitment capacity (Productive lines, n = 12) and those with low capacity (Non-Productive lines, n = 12) by microarray, identifying IGF2 as the most differentially expressed gene with significance (Figure 1D-E, S1G, Table S2). Additionally, this correlation held in iPSC lines generated by Sendai virus vectors (Sendai iPSC) (Figure S1H, Table S1).

On the other hand, DNA methylation analysis using Infinium methylation beads array 450K (Infinium 450K array) did not reveal any probes that had statistically significant methylation differences between the two groups (Figure S1I). In accordance

with mRNA expression, immunofluorescence staining with antibodies against IGF2 and OCT3/4 showed that Productive iPSCs were stained for IGF2. In addition, iPSC colonies with the tendency to differentiate had the highest IGF2 expression even though IGF2-positive cells were restricted to OCT3/4-positive cells (Figure S1J).

To investigate the role of IGF2 during the commitment phase, we investigated the function of the IGF2 gene by adding recombinant IGF2 protein (Figure S1K) or by inducing IGF2 overexpression in iPSCs (Figure 1F), and found that activation of the IGF2 signal promoted commitment capacity. In addition, commitment capacity was reduced by antibodies that blocked the IGF1 receptor, the principal receptor of IGF2 during embryogenesis (Figure S1L), or by knockdown of IGF2 (Figure 1G, S1M). To explore how IGF2 promotes hematopoietic commitment, we performed interim analysis on day 4, finding that IGF2 increased the expression level of itself and of mesendodermal genes, such as T and GATA6, and the hematopoietic gene HOXB4 (Figure S1N). In addition, we confirmed that Productive PSC lines expressed significantly higher levels of IGF2 and T than did Non-Productive PSC lines and tended to have higher expressions of GATA6 and HOXB4 (Figure S1O), suggesting that IGF2 promotes hematopoiesis through the up-regulation of these genes.

### **Network Analysis Revealed Co-regulatory Gene Network of IGF2**

Because preliminary correlation analysis using the gene expression of PSCs in

1 our laboratory and other laboratories revealed a correlation of IGF2 with  
2 mesendoderm/hematopoietic genes and pluripotent genes (Figure S2A, Table S3), we  
3 performed an unsupervised network analysis, unsigned Weighted Gene Co-expression  
4 Network Analysis (WGCNA) (Langfelder and Horvath, 2008), using the mRNA  
5 microarray data of the 39 PSC lines (Figure 2A, S2B). Unsigned WGCNA enabled us to  
6 explore positive and negative gene regulations. Based on the WGCNA, we grouped all  
7 the genes into 50 modules (Figure 2A, S2B). The IGF2-related gene module (hereafter  
8 called Module A) included mesendodermal genes, such as GATA6, CDX2, and BMP4,  
9 as positive regulators (Module A differentiation genes, 563 probes, 493 genes), and  
10 pluripotent genes, such as PRDM14, as negative regulators (Module A suppression  
11 genes 54 probes, 49 genes) (Figure 2B-D, S2B-C, Table S4). Module A differentiation  
12 genes were up-regulated immediately upon mesendodermal/hematopoietic  
13 differentiation on day 2-4, but not upon neural differentiation (Figure 2E, S2D-E). In  
14 addition, we found that the analysis enabled us to determine the correlation of each  
15 WGCNA module with commitment capacity, finding that Module A genes were  
16 associated with commitment capacity significantly as a whole (Figure 2B), whereas  
17 other identified gene modules of known importance in the differentiation of iPSCs, such  
18 as Module B (HERV-H-related) (Koyanagi-Aoi et al., 2013), Module C (Imprinting of  
19 chromosome 14q-related) (Stadtfield et al., 2010), and Module D (TGF $\beta$ /Nodal

1 signal-related) (Ramos-Mejia et al., 2010), were not. Module A was composed of highly  
2 variable genes among human PSC lines, as too were Modules B and C (Figure S2F),  
3 indicating that Module A genes largely contributed to variation in the gene expressions  
4 of undifferentiated human iPSCs. Consistently, the forced expression of IGF2 in iPSCs  
5 up-regulated Module A differentiation genes in iPSCs (Figure S2G).

### 6 **IGF2 and Co-regulated Genes are under Control of FGF/Activin Signaling** 7 **through Change in Chromatin Structure**

8 Next, we investigated the regulatory mechanism of IGF2 and Module A genes,  
9 finding that Module A differentiation genes were down-regulated upon adding basic  
10 FGF (bFGF) or Activin A to iPS/ESCs (Figure 2F, S3A). Principal component analysis  
11 (PCA) of the expression of module A genes revealed IGF2-high PSCs have a distinct  
12 module A gene expression profile to that of IGF2-low PSCs and similar to that of cells  
13 after 2 or 4 days of hematopoietic differentiation (Figure 2G). Additionally, the addition  
14 of bFGF or Activin A during the culture of iPS/ESCs converted the IGF2-high iPSC  
15 lines into an IGF2-low iPSC line-like state (Figure 2G, S3B). These findings indicate  
16 that Module A genes are under the control of FGF/Activin signaling.

17 It is known that the imprinting control region (ICR) of H19 regulates the  
18 expression of H19 and IGF2 in a reciprocal manner and that some iPS/ESC lines harbor  
19 gain of imprinting at H19-ICRs, which results in the biallelic expression of IGF2 and

1 decreased expression of H19 (Pick et al., 2009; Rugg-Gunn et al., 2005). We therefore  
2 investigated the methylation level of H19-ICRs (Figure 2H-I, S3C-D). The analysis  
3 revealed that 9 of the 39 lines harbored gain of imprinting there (Figure 2I), however,  
4 the IGF2 gene expression level did not correlate with the methylation status of  
5 H19-ICRs (Figure 2H-I, S3C). In addition, IGF2 expression was positively correlated  
6 with H19 gene expression in PSCs (Figure S3E). These findings suggested that the gain  
7 of imprinting of H19-ICRs in some iPSC lines was not the cause of the variable  
8 expression of IGF2 genes. DNA methylation analysis showed that the regulatory  
9 regions of IGF2 and GATA6 were hypomethylated in both IGF2-high and -low lines  
10 (Figure S3F) and that the methylation profiles of the promoter regions of Module A  
11 differentiation genes in IGF2-high and -low lines were similar (Figure S3G), suggesting  
12 that the DNA methylation status is also not the cause of the variation. To understand the  
13 regulatory mechanism of IGF2 and its related genes, we investigated open chromatin  
14 accessibility by ATAC-seq (Buenrostro et al., 2013), finding an open chromatin status in  
15 IGF2-high iPSCs compared with IGF2-low iPSCs in IGF2 and mesendodermal genes,  
16 such as GATA6 (Figure 2J). An increase in bFGF signaling closed the sites (Figure 2K),  
17 suggesting that IGF2 and mesendodermal genes are regulated by the chromatin structure,  
18 which is dependent on FGF/Activin signaling (Figure 2L).

## 19 **Hematopoietic Maturation Capacity is Associated with Distinct DNA Methylation**

## Patterns

Hematopoietic maturation capacity was assessed by performing colony forming unit (CFU) assays of PSC-derived hematopoietic precursor cells (HPCs) (CD34+CD38-CD43+lineage marker-; hereafter called CD34+38-43+lin-; Figure 3A-B, S4A-B) and by investigating the differentiation efficiency into erythroblasts (Figure 3C-F) and megakaryocytes (Figure 3G-H). The number of mixed colonies in the CFU assay was highly correlated with the erythroblast and megakaryocyte differentiation capacity (Figure 3E, H), suggesting this number is a surrogate marker of hematopoietic maturation capacity. Accordingly, we designated “maturation-competent lines” (MC lines, n = 18) and “maturation-deficient lines” (MD lines, n = 18) based on the number of mixed colonies in the CFU assay (Figure 3B). We also tested the T cell differentiation of iPSC-derived HPCs, finding those derived from MC-iPSCs had good differentiation capability but those derived from MD-iPSCs did not (Figure S4C-D). Interestingly, there was no positive correlation between commitment capacity and maturation capacity (Figure S4E).

To investigate the molecular signatures associated with maturation capacity, we compared the transcription profile between MC and MD lines in the undifferentiated state, finding lower gene expressions of TCERG1L and FAM19A5 in the MD lines (Figure 3I, Table S2), which was due to hypermethylation in these genes (Figure S4F).

1 These genes have been reported as methylated in some iPSC lines (Huang et al., 2014a;  
2 Koyanagi-Aoi et al., 2013; Ruiz et al., 2012). To further analyze the association between  
3 CpG loci specifically methylated in iPSCs (hereafter called iPS-DMRs) and maturation  
4 capacity, we identified iPS-DMRs (396 probes, 84 genes) by comparing DNA  
5 methylation between human iPSCs (n = 56) and ESCs (n = 20) (Figure 3J, Table S5-6).  
6 PCA with iPS-DMRs showed that iPSCs with similar methylation profiles to ESCs had  
7 higher maturation capacity than iPSC lines with distinct methylation profiles, with few  
8 exceptions (Figure 3K, S4G).

### 9 **Hypermethylation of MD-high DMRs is Acquired during Reprogramming** 10 **in MD-iPSC Lines**

11 To further analyze the association between the DNA methylation of PSC lines  
12 and maturation capacity, we compared methylation profiles between MD-PSC and  
13 MC-PSC lines, finding differently methylated regions (MD-high DMRs: 277 probes, 80  
14 genes; MC-high DMRs: 69 probes, 25 genes) (Figure 4A, S4H, Table S5-6). In  
15 MD-high DMRs, the top-ranked differentially methylated gene was an  
16 erythroid/megakaryocyte-specific gene: TRIM58 (Figure S4I-J, Table S5). Next, we  
17 affirmed the maturation capacity of iPSCs can be predicted prospectively by the status  
18 of MD-high DMRs using nine additional iPSC lines (Figure 4B, Table S1). Similarly, an  
19 association between DNA methylation and maturation capacity was observed in Sendai

iPSCs (Figure S4K), indicating that MD-high DMRs have a crucial impact on maturation deficiency in iPSCs generated by diverse types of reprogramming methods. In contrast, the profile of MC-high DMRs did not enable us to predict maturation capacity (Figure S4L).

Next, to investigate whether the status of MD-high DMRs is inherited from the original cells or acquired during the reprogramming, we traced the methylation levels of MD-high DMRs from the original cells to iPSCs. MD-high DMRs, including the TRIM58 gene, were not methylated in any parental tissues, but only MD iPSC lines acquired hypermethylation in MD-high DMRs during the reprogramming, whereas MC iPSC lines did not (Figure 4C-E). An analysis of intermediate cells expressing the TRA 1-60 antigen during reprogramming from HDF to pluripotency (Ohnuki et al., 2014) revealed that these aberrant DNA methylations developed in the late reprogramming phase (Figure 4F). Furthermore, we compared the methylation levels of MD-high DMRs after differentiation into HPCs, finding them retained in PSC-derived HPCs (Figure 4G, S5A).

Moreover, both HDF-iPSC and DP-iPSC lines were classified as MC-iPSCs and MD-iPSCs based on maturation capacity as well, and MC-iPSCs showed hypomethylated MD-high DMRs compared with MD-iPSCs among both HDF-iPSCs and DP-iPSCs (Figure 4H-I). Furthermore, we analyzed five sets of iPSC lines

generated from the same somatic samples of the same donors with different maturation capacity (Figure S5B), finding that MC-iPSCs had lower methylation levels of MD-high DMRs than MC-iPSCs in all sets despite the same genetic background and cell type origin. These findings suggest that the acquisition of aberrant hypermethylation in MD-high DMRs during reprogramming is a crucial factor in maturation capacity, irrespective of the cell type origin or donor.

We found that HDF-iPSCs generated by episomal vectors (episomal HDF-iPSCs) and those by Sendai virus vectors (Sendai HDF-iPSCs) are various in the extent of aberrant DNA methylation (from low to high) and maturation capacity (maturation-competent to maturation deficient), while most retro HDF-iPSC lines have high aberrant DNA methylation and attenuated maturation capacity (Figure S5C-F). This is the case also for iPSCs generated by the same reprogramming factors (Figure S5D). These data suggest the possibility that differences in delivery methods might influence methylation patterns and maturation capacity.

In addition, we observed the tendency that iPSCs from hematopoietic cells (Blood-iPSCs) had higher capacity for hematopoietic maturation than iPSCs from other tissues (Figure 4J), although the differences when comparing only episomal iPSCs were statistically insignificant (Figure S5G). This finding seemingly supports a somatic memory-induced differentiation preference, as reported previously (Kim et al., 2010;

Kim et al., 2011). However, Blood-iPSCs tended to acquire less DNA methylation in MD-high DMRs in contrast to iPSCs from other tissues (Figure 4C-E), therefore, fewer aberrant DNA methylations in Blood-iPSCs could be the cause of the high maturation capacity. In addition, HDF-iPSC lines and DP-iPSC lines with hypomethylated MD-high DMRs showed comparative differentiation capacity to Blood-iPSCs (Figure 4H-I), raising the question about the influence of somatic memory on maturation capacity. Furthermore, an earlier paper showed that Blood-iPSC lines have higher differentiation capacity into hepatocytes than do HDF-iPSCs or DP-iPSCs (Kajiwara et al., 2012), suggesting a high quality of Blood-iPSCs due to mechanisms other than somatic memory.

Taken together, the above findings strongly suggest that the high maturation capacity of iPSCs is primarily determined by aberrant DNA methylation, although there is the possibility that somatic memory can modify the high capacity in Blood-iPSC lines.

#### **Erythroid- and Megakaryocyte- Related Genes Are Aberrantly Methylated in MD-iPSCs and Derived HPCs**

Next, we checked the gene expression profiles of HPCs from MC-iPSCs and MD-iPSCs, finding that representative transcription factors of hematopoietic stem cells such as RUNX1 are up-regulated in MD-iPSC lines comparable to MC-iPSC lines

(Figure S6A), suggesting that MD lines differentiate into HPCs but exhibit maturation defects after commitment. In addition, in MD-high DMR-associated genes, erythroid- and megakaryocyte-related genes, including TRIM58 and ZFPM1, were significantly enriched (Figure 5A-B, S6B-D) (Fujiwara et al., 2009; Tijssen et al., 2011).

### **Disruption of TRIM58 Expression Impairs Hematopoietic Maturation Capacity**

Next, we investigated whether the most differentially methylated gene, TRIM58, affected maturation capacity. Although TRIM58 is associated with the number of platelets or size of red blood cells according to genome-wide association studies (Ganesh et al., 2009; Gieger et al., 2011) and with terminal erythropoiesis (Thom et al., 2014), whether the diversity of its methylation levels among cell lines can affect hematopoietic maturation capacity from iPSCs remains unknown. The gene expression analysis revealed that erythroblasts from MD-iPSCs have less TRIM58 expression (~10-fold less) than MC-iPSCs due to abnormal DNA methylation of the TRIM58 gene (Figure 5C). To evaluate the function of TRIM58 on maturation capacity, we established TRIM58 knockout (KO) iPSC lines by inducing frameshift mutations in the first exon of TRIM58 using CRISPR/Cas9 technology (Ran et al., 2013) (Figure 5D-E). Knockout of TRIM58 reduced maturation capacity, such as the capacity to form mixed colonies (Figure 5F), and differentiation into erythroid cells (Figure 5G) and megakaryocytes (Figure 5H). In addition, we forced TRIM58 gene expression in a MD-iPSC line

(409B2, episomal HDF-iPSC line), confirming it partially restored erythroid (Figure 5I) and megakaryocyte differentiation capacity (Figure 5J). These results suggest that acquired DNA methylation in the TRIM58 gene caused attenuated maturation capacity in MD-iPSCs. Although TRIM58 did not fully restore the maturation capacity of MD-iPSCs, the fact that MD-iPSC lines harbored multiple aberrant DNA methylations in erythroid- and megakaryocyte-related genes concomitantly (Figure S6E) suggests that an accumulation of abnormalities in these genes resulted in impaired hematopoietic maturation capacity.

### **Influence of Residual DNA Methylations Inherited from iPSC Origins**

As described above, there is the possibility that differences in maturation capacity were caused by residual DNA methylation inherited from the iPSC origin, i.e. somatic memory. To clarify this point, we investigated to what extent differences in DNA methylation in the original cells were passed to iPSCs by analyzing the DNA methylation of iPSCs generated from PB and HDF from a single donor and the DNA methylation of the corresponding original cells (Figure 6A). We identified a small number of epigenetic signatures passed from the parental cells to their iPSCs (hereafter called residual DMRs; ~0.1% of all Infinium 450K array probes on autosomes, 448 (PB-iPSC > HDF-iPSC) and 52 (HDF-iPSC > PB-iPSC) out of 437722 probes) (Figure 6B, Table S7), and found that most iPSC lines retained this residual signature (Figure

S7A). The residual DNA methylations seemed to affect the gene expression profiles of iPSCs and their derivatives (Figure S7B-C). To investigate the association between residual DMRs and maturation capacity, we compared the methylation levels of residual DMRs in iPS/ESC lines, finding that the methylation levels of residual DMRs in MC-HDF-iPSCs were much closer to those of HDF lines or MD-HDF-iPSCs than Blood-iPSCs or parental blood cells. This observation indicates that MC-HDF-iPSCs inherit the methylation signatures of HDF at the same level as MD-HDF-iPSCs (Figure 6C). Furthermore, PCA with the methylation of residual DMRs revealed that 2 HDF-iPSC lines have smaller amounts of residual methylation signatures, however, these two lines were MD lines because of the hypermethylation of MD-high DMRs despite their smaller residual methylation (Figure 6D). These findings indicated that the methylation levels of residual DMRs were not correlated with the maturation capacity of HDF-iPSCs, suggesting somatic memory had only a small effect on maturation capacity, especially compared with MD-high DMRs.

Next, we assessed the maturation capacity of iPSC lines at low passage number, because iPSCs reportedly retain somatic memory at around passage (p) = mid-10s, but thereafter lose it (Kim et al., 2011). However, we found that MC-HDF-iPSCs were already maturation-competent at p10 or 11 and exhibited almost the same maturation capacity as Blood-iPSCs (Figure S7D), which refuted the possibility that somatic

memory caused maturation deficiency in MC-HDF-iPSCs at early passage.

Taking these findings together, we attributed the difference in hematopoietic maturation capacity among iPSC lines primarily to de novo aberrant DNA methylation (Figure 6E).

### **Transplantation of iPSC-derived HPCs in immunodeficient mice**

One of the most important functions of hematopoietic precursor cells is the capacity to reconstitute the hematopoietic system. We therefore transplanted HPCs from MC-iPSCs and MD-iPSCs (control CD34<sup>+</sup> CB cells), however, neither MC-iPSC-derived HPCs nor MD-iPSC-derived HPCs engrafted in immunodeficient NOD/LtSz-scid IL2R $\gamma$ Null (NSG) mice despite the good engraftment of CD34<sup>+</sup> CB cells (14.8-62.0%) 3 months after transplantation (Figure S7E), which is consistent with earlier studies that showed iPSC-derived HPCs are devoid of engraftment capacity.

### **Influence of Aberrant DNA methylations on Hepatic Differentiation**

Next, to investigate whether aberrant DNA methylations of iPSCs affect differentiation capacity into other lineage cells, we reanalyzed the data of a previous study that investigated variations in hepatic differentiation capacity (Kajiwara et al., 2012). This analysis revealed that hepatic differentiation-deficient lines showed hypermethylated MD-high DMRs (Figure 6F), suggesting that aberrant DNA methylation also impairs differentiation into hepatocytes.

1

## 2 **Discussion**

3         In the present study, a separate assessment of the differentiation capacity in  
4         early phase (commitment capacity) and late phase (maturation capacity) allowed us to  
5         identify key factors associated with each phase (Figure 7A). Our findings suggest that  
6         transcriptome and epigenomic analysis are promising methods to predict the  
7         differentiation capacity of undifferentiated iPSCs (Figure 7B).

8         We found that IGF2 is responsible for the variation in hematopoietic  
9         commitment capacity among iPSC lines. IGF2 is known to play a pivotal role in  
10        embryonic growth, is associated with body size (DeChiara et al., 1990), and is  
11        specifically expressed in mesodermal or hepatic tissues but not in ectoderm during  
12        embryogenesis (Stylianopoulou et al., 1988). Regarding in vitro differentiation, IGF2 is  
13        known to enhance differentiation efficiency in mesodermal and hematopoietic cells  
14        (Hiroyama et al., 2006; Morali et al., 2000).

15        Although our findings revealed a role of the IGF2 gene in variable  
16        hematopoietic commitment capacity and its regulatory mechanism, the reason why or  
17        when these variations develop still remains unclear. Tracking analysis of the expression  
18        levels of IGF2 and H19 during reprogramming revealed that these two genes are highly  
19        up-regulated at the mid-reprogramming phase and thereafter suppressed (Figure S7F),

1 suggesting the transient up-regulation of these genes might relate to the variations.

2       Regarding maturation capacity, the association of de novo aberrant DNA  
3 methylations was found crucial, but why only some iPSC lines developed aberrant DNA  
4 methylation during reprogramming, even when the same genetic background or  
5 reprogramming method was used, is unclear. In particular, why Blood-iPSCs are  
6 unlikely to acquire aberrant DNA methylations compared with iPSCs from other origins  
7 deserves further study. Our methylation analysis revealed that blood cells (PB or CB)  
8 were in a hypermethylated state (~70% methylated) that almost matched the level in  
9 iPS/ESCs and is unlike the level in other parental tissues (~45% methylated) (Figure  
10 7C). In addition, clustering analysis revealed that the methylation profiles of blood cells  
11 (PB or CB) are much closer to those of ESCs than other somatic tissues (Figure S1C).  
12 For successful reprogramming, about 3 times more CpGs need de novo methylation in  
13 non-blood tissues than in blood (Figure 7D). Although MD-high DMRs were not  
14 methylated in any parental tissues (Figure 4C), non-blood tissues acquired more  
15 abnormal methylation in MD-high DMRs than did blood during reprogramming (Figure  
16 7E), and interestingly, the number of abnormally methylated CpGs in each tissue during  
17 reprogramming was almost proportional to the number of total CpGs methylated during  
18 reprogramming in the corresponding iPSCs (Figure 6H-I). From these observations, we  
19 speculated that the larger quantity of methylation required in non-blood origins risks

1 more aberrant DNA methylation in iPSCs, presumably through dysregulation of the  
2 methylation process (Figure 7F).

3 Another possible explanation for the variation of aberrant DNA methylation  
4 levels might come from DNA methyltransferase (DNMT) family genes. An earlier  
5 report showed that iPSCs generated from DNMT3b-deficient fibroblasts did not acquire  
6 aberrant DNA methylation in genes such as TCERG1L or FAM19A5 (Huang et al.,  
7 2014b). Therefore, we investigated the expression level of DNMT family genes, but did  
8 not observe any differences between MC-iPSC and MD-iPSC lines, Blood-iPSCs and  
9 HDF-iPSCs (Figure S7G), or different parental cells (Figure S7H).

10 Another study showed that decreased methylation levels in hematopoietic  
11 precursor cells skewed differentiation into myeloid lineage cells and disturbed  
12 differentiation into lymphocytes (Broske et al., 2009). We confirmed that MC-iPSCs  
13 have favorable lymphoid differentiation capacity, but that MD-iPSCs do not (Figure  
14 S4B-C), suggesting the differences observed in the present study are not the result of a  
15 skewed differentiation propensity towards myeloid lineage from lymphoid lineage.

16 In conclusion, we demonstrated the molecular mechanisms underlying clonal  
17 differences in the hematopoietic differentiation of iPSCs. We believe our findings will  
18 facilitate better understanding of the reprogramming mechanism and also optimal  
19 selection of iPSC lines for biomedical applications.

1

## 2 **Experimental Procedures**

### 3 **Human iPSC/ESC Lines and Cell Culture**

4 All human iPSC/ESC lines used in the present study are described in Table S1. Human  
5 PSCs were maintained on mitomycin C-treated SNL feeder cells in Primate ES cell  
6 medium (ReproCELL) containing 4 ng/mL recombinant bFGF (Wako), as described  
7 previously (Kajiwara et al., 2012). Before the differentiation experiments, PSCs were  
8 cultured on matrigel-coated dishes for one passage to eliminate feeder cells.

### 9 **Hematopoietic Differentiation and Assays**

10 Hematopoietic differentiation was performed as previously described with minor  
11 modifications (Grigoriadis et al., 2010). Briefly, PSCs were dissociated into small  
12 clusters seeded on a low-binding dish to form embryoid bodies and then differentiated  
13 in suspension culture. The commitment capacity was assessed by counting the number  
14 of generated CD43<sup>+</sup> cells on day 15 by flow cytometry. For CFU assays, we seeded a  
15 total of 5000 FACS-sorted hematopoietic precursor cells (CD34<sup>+</sup>38<sup>-</sup>43<sup>+</sup>lin<sup>-</sup>) into  
16 methylcellulose-based media (MethoCult H4435, Stemcell Technologies). After 14 days,  
17 the number of colonies was counted. Erythroid differentiation capacity was assessed on  
18 day 35 by counting the number of CD235a<sup>+</sup> cells by flow cytometry. To assess  
19 megakaryocyte differentiation, we seeded a total of 5000 FACS-sorted

CD34<sup>+</sup>38<sup>-</sup>43<sup>+</sup>lin<sup>-</sup> cells into a collagen based medium (MegaCult™-C Complete kit with Cytokines) and counted the number of CD41a<sup>+</sup> colonies according to the manufacturer's instructions. Detailed protocols are described in Supplemental Experimental Procedure.

#### **Analysis of DNA Methylation Status**

Analysis of the DNA methylation status was performed by Infinium 450K array or Methyl-seq (Agilent Technologies) unless stated otherwise. Details are described in Supplemental Experimental Procedures.

#### **Gene Expression Microarray Analysis**

Gene expression microarray analysis was performed using a SurePrint G3 Human Gene Expression 8 × 60K kit (G4851A; Agilent Technologies). Details are described in Supplemental Experimental Procedures.

#### **Statistical Analysis**

Statistical analysis was performed by GeneSpring GX 12.1, SPSS Statistics21 (IBM) or GraphPad Prism 6 (GraphPad Software).

#### **Author Contributions**

M.Nishizawa, A.T.-K., S.Y., and Y.Y. designed research; M.Nishizawa, K.C, M.Nomura, A.T., M.Nakamura, A.I., M.Nishikawa, I.T., A.O., M.O., M.K.-A., and A.W.

performed research; K.T., Y.H., and K.O. established iPSC lines, M.Nishizawa and M.Nomura analyzed data; and M.Nishizawa, S.Y., and Y.Y. wrote the paper.

#### **Acknowledgement**

We greatly appreciate members of the Yamanaka lab and the Department of Hematology and Oncology, Kyoto University, for critical discussions. We thank Akitsu Hotta and Knut Woltjen for providing the vectors; Katsuhiko Ito and Hiroshi Kawamoto for technical support; Peter Karagiannis and Hiroshi Kawamoto for reading the manuscript; Megumi Narita, Yoshiko Sato, and Yasuko Matsumura for technical assistance; and Tomohisa Kato, Junya Toguchida, Haruhisa Inoue, Jun Takahashi, and Ryosuke Takahashi for providing somatic cell samples. This work was supported by Grants-in-Aid for Scientific Research from the Japanese Society for the Promotion of Science (JSPS) and the Ministry of Education, Culture, Sports, Science, and Technology (MEXT); a grant from the Leading Project of the MEXT; a grant from the Funding Program for World-Leading Innovative Research and Development on Science and Technology (First Program) of the JSPS; Health Labour Sciences Research Grant of The Ministry of Health Labour and Welfare; a grant from Research Center Network for Realization of Regenerative Medicine; and iPS Cell Research Fund. M. N is a research fellow of JSPS. The authors declare the following conflicts of interest. Kyoto University

has filed a patent application broadly relevant to this work with S. Y., Y. Y., and M. N. as the investigators of record listed on the patent application; S.Y. is an unsalaried member of scientific advisory boards for iPS Academia Japan.

## Accession Numbers

All mRNA microarray data, Infinium 450K array data, Methyl-seq data, ATAC-seq data, and CTCF-ChIP-seq data have been deposited in the Gene Expression Omnibus (GEO) data base (GSE60924 and GSE74967).

## References

- Broske, A.M., Vockentanz, L., Kharazi, S., Huska, M.R., Mancini, E., Scheller, M., Kuhl, C., Enns, A., Prinz, M., Jaenisch, R., *et al.* (2009). DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nat Genet* *41*, 1207-1215.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* *10*, 1213-1218.
- DeChiara, T.M., Efstratiadis, A., and Robertson, E.J. (1990). A growth-deficiency

1 phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted  
2 by targeting. *Nature* 345, 78-80.

3 Fujiwara, T., O'Geen, H., Keles, S., Blahnik, K., Linnemann, A.K., Kang, Y.A., Choi, K.,  
4 Farnham, P.J., and Bresnick, E.H. (2009). Discovering hematopoietic mechanisms  
5 through genome-wide analysis of GATA factor chromatin occupancy. *Mol Cell* 36,  
6 667-681.

7 Ganesh, S.K., Zakai, N.A., van Rooij, F.J., Soranzo, N., Smith, A.V., Nalls, M.A., Chen,  
8 M.H., Kottgen, A., Glazer, N.L., Dehghan, A., *et al.* (2009). Multiple loci influence  
9 erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet* 41, 1191-1198.

10 Gieger, C., Radhakrishnan, A., Cvejic, A., Tang, W., Porcu, E., Pistis, G.,  
11 Serbanovic-Canic, J., Elling, U., Goodall, A.H., Labrune, Y., *et al.* (2011). New gene  
12 functions in megakaryopoiesis and platelet formation. *Nature* 480, 201-208.

13 Grigoriadis, A.E., Kennedy, M., Bozec, A., Brunton, F., Stenbeck, G., Park, I.H.,  
14 Wagner, E.F., and Keller, G.M. (2010). Directed differentiation of hematopoietic  
15 precursors and functional osteoclasts from human ES and iPS cells. *Blood*.

16 Hiroyama, T., Miharada, K., Aoki, N., Fujioka, T., Sudo, K., Danjo, I., Nagasawa, T.,  
17 and Nakamura, Y. (2006). Long-lasting in vitro hematopoiesis derived from primate  
18 embryonic stem cells. *Exp Hematol* 34, 760-769.

19 Huang, K., Shen, Y., Xue, Z., Bibikova, M., April, C., Liu, Z., Cheng, L., Nagy, A.,

1 Pellegrini, M., Fan, J.B., *et al.* (2014a). A panel of CpG methylation sites distinguishes  
 2 human embryonic stem cells and induced pluripotent stem cells. *Stem Cell Reports* 2,  
 3 36-43.  
 4 Huang, K., Wu, Z., Liu, Z., Hu, G., Yu, J., Chang, K.H., Kim, K.P., Le, T., Faull, K.F.,  
 5 Rao, N., *et al.* (2014b). Selective demethylation and altered gene expression are  
 6 associated with ICF syndrome in human-induced pluripotent stem cells and  
 7 mesenchymal stem cells. *Hum Mol Genet* 23, 6448-6457.  
 8 Kajiwara, M., Aoi, T., Okita, K., Takahashi, R., Inoue, H., Takayama, N., Endo, H., Eto,  
 9 K., Toguchida, J., Uemoto, S., *et al.* (2012). Donor-dependent variations in hepatic  
 10 differentiation from human-induced pluripotent stem cells. *Proc Natl Acad Sci U S A*  
 11 109, 12538-12543.  
 12 Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H.,  
 13 Ehrlich, L.I., *et al.* (2010). Epigenetic memory in induced pluripotent stem cells.  
 14 *Nature*.  
 15 Kim, K., Zhao, R., Doi, A., Ng, K., Unternaehrer, J., Cahan, P., Huo, H., Loh, Y.H.,  
 16 Aryee, M.J., Lensch, M.W., *et al.* (2011). Donor cell type can influence the epigenome  
 17 and differentiation potential of human induced pluripotent stem cells. *Nat Biotechnol* 29,  
 18 1117-1119.  
 19 Koyanagi-Aoi, M., Ohnuki, M., Takahashi, K., Okita, K., Noma, H., Sawamura, Y.,

1 Teramoto, I., Narita, M., Sato, Y., Ichisaka, T., *et al.* (2013). Differentiation-defective  
2 phenotypes revealed by large-scale analyses of human pluripotent stem cells. *Proc Natl*  
3 *Acad Sci U S A*.

4 Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation  
5 network analysis. *BMC bioinformatics* 9, 559.

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

10 Morali, O.G., Jouneau, A., McLaughlin, K.J., Thiery, J.P., and Larue, L. (2000). IGF-II  
11 promotes mesoderm formation. *Dev Biol* 227, 133-145.

12 Nazor, K.L., Altun, G., Lynch, C., Tran, H., Harness, J.V., Slavin, I., Garitaonandia, I.,  
13 Muller, F.J., Wang, Y.C., Boscolo, F.S., *et al.* (2012). Recurrent variations in DNA  
14 methylation in human pluripotent stem cells and their differentiated derivatives. *Cell*  
15 *Stem Cell* 10, 620-634.

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

19 Ohnuki, M., Tanabe, K., Sutou, K., Teramoto, I., Sawamura, Y., Narita, M., Nakamura,

1 M., Tokunaga, Y., Watanabe, A., Yamanaka, S., *et al.* (2014). Dynamic regulation of  
2 human endogenous retroviruses mediates factor-induced reprogramming and  
3 differentiation potential. *Proc Natl Acad Sci U S A* *111*, 12426-12431.

4 Pick, M., Stelzer, Y., Bar-Nur, O., Mayshar, Y., Eden, A., and Benvenisty, N. (2009).  
5 Clone and Gene Specific Aberrations of Parental Imprinting in Human Induced  
6 Pluripotent Stem Cells. *Stem Cells*.

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

11 Ramos-Mejia, V., Melen, G.J., Sanchez, L., Gutierrez-Aranda, I., Ligerio, G., Cortes,  
12 J.L., Real, P.J., Bueno, C., and Menendez, P. (2010). Nodal/Activin signaling predicts  
13 human pluripotent stem cell lines prone to differentiate toward the hematopoietic  
14 lineage. *Mol Ther* *18*, 2173-2181.

15 Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott,  
16 D.A., Inoue, A., Matoba, S., Zhang, Y., *et al.* (2013). Double nicking by RNA-guided  
17 CRISPR Cas9 for enhanced genome editing specificity. *Cell* *154*, 1380-1389.

18 Rugg-Gunn, P.J., Ferguson-Smith, A.C., and Pedersen, R.A. (2005). Epigenetic status of  
19 human embryonic stem cells. *Nat Genet* *37*, 585-587.

1 Ruiz, S., Diep, D., Gore, A., Panopoulos, A.D., Montserrat, N., Plongthongkum, N.,  
2 Kumar, S., Fung, H.L., Giorgetti, A., Bilic, J., *et al.* (2012). Identification of a specific  
3 reprogramming-associated epigenetic signature in human induced pluripotent stem cells.  
4 *Proc Natl Acad Sci U S A* 109, 16196-16201.

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

8 Stylianopoulou, F., Efstratiadis, A., Herbert, J., and Pintar, J. (1988). Pattern of the  
9 insulin-like growth factor II gene expression during rat embryogenesis. *Development*  
10 103, 497-506.

11 Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and  
12 Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts  
13 by defined factors. *Cell* 131, 861-872.

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

16 Thom, C.S., Traxler, E.A., Khandros, E., Nickas, J.M., Zhou, O.Y., Lazarus, J.E., Silva,  
17 A.P., Prabhu, D., Yao, Y., Aribéana, C., *et al.* (2014). Trim58 degrades Dynein and  
18 regulates terminal erythropoiesis. *Dev Cell* 30, 688-700.

19 Tijssen, M.R., Cvejic, A., Joshi, A., Hannah, R.L., Ferreira, R., Forrai, A., Bellissimo,

1 D.C., Oram, S.H., Smethurst, P.A., Wilson, N.K., *et al.* (2011). Genome-wide analysis  
2 of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes  
3 identifies hematopoietic regulators. *Dev Cell* 20, 597-609.

4

## Figure Legends

### Figure 1 Hematopoietic Commitment Capacity of Human iPSC lines is Associated with IGF2 Expression

(A) Schematic of the experimental design. (B) Commitment capacity was assessed by the number of generated CD43<sup>+</sup> hematopoietic cells on day 15. Colors indicate the cell type origin of the iPSCs or ESCs. Error bars indicate SD (n = 3-4). A representative flow cytometric analysis is shown on the left. (C) The commitment capacity did not associate with the original cell type of the iPSCs (n = 35) or with ESCs (n = 4) significantly. Statistical analysis was performed by ANOVA. Red lines indicate median values. (D) A scatter plot comparing gene expressions between Productive PSC lines (n = 12) and Non-Productive PSC lines (n = 12) in the undifferentiated state by microarray analysis. Red dots indicate differentially expressed genes with significance (37 microarray probes, 32 genes; fold change (FC) > 2 and false discovery rate (FDR) < 0.05). **See also Table S2.** (E) The expression level of the IGF2 gene in Productive and Non-Productive PSCs by microarray analysis. \*\*\*\* p < 1×10<sup>-3</sup> (t-test). (F) Number of CD43<sup>+</sup> cells on day 15 with forced expression of the IGF2 gene during days 0-1. Data are shown as FC. \*p < 0.05 (t-test). (G) Number of CD43<sup>+</sup> cells on day 15 with knockdown of the IGF2 gene. Data are shown as FC. \*p < 0.05 (ANOVA and Dunnett's test). siRNA against IGF2 and control siRNA were transfected on day -2. Error bars are

SD. See also Figure S1.

## **Figure 2 Network Analysis by WGCNA Revealed Regulatory Mechanism of Commitment Capacity and IGF2 Gene**

(A) Commitment capacity-related gene modules according to unsigned WGCNA. (B) Representative WGCNA modules and their representative genes. Correlation of WGCNA modules with commitment capacity was assessed by calculating the correlation coefficient between the module eigengene and commitment capacity of 39 PSC lines using Pearson's method. (C) The interaction network of representative Module A genes (transcription factors and IGF2, H19, and BMP4) constructed using the data of unsigned WGCNA. Module A differentiation and suppression genes were defined as positively or negatively correlated with commitment capacity, respectively. Blue and red balls indicate genes with Module A differentiation and suppression genes, respectively. Line thickness indicates the connection strength between genes. (D) Heat-map showing eigengene values of Modules A-D in the 39 iPS/ESC lines and corresponding interpretation. Eigengenes were used to represent the gene expression profiles of each module in each cell line. (E) Expression levels of Module A differentiation genes and IGF2 at the undifferentiated state, cells on day 2 (D2 Hema) and on day 4 (D4 Hema) after hematopoietic differentiation, and cells on day 4 (D4

1 Neuron) after neural differentiation. Relative expression to undifferentiated iPSCs is  
 2 shown (log2 scale). (F) Expression levels of Module A differentiation genes and IGF2 at  
 3 different concentrations of bFGF (0, 4, 20, and 100 ng/ml) in iPS/ESCs (average of one  
 4 iPS and one ESC line). The relative expression to that under normal ESC culture  
 5 condition (bFGF: 4 ng/ml) is shown (log2 scale). (G) PCA of IGF2-high iPSCs (n = 10),  
 6 IGF2-low PSCs (n = 10), and iPS/ESCs (average of one iPS and one ESC line) cultured  
 7 at different concentrations of bFGF (0, 4, 20, and 100 ng/ml), and D2 Hema, D4 Hema,  
 8 and D4 Neuron with Module A genes. (H) Average methylation levels of IGF2-high  
 9 iPSC lines (n = 6) and IGF2-low iPSC lines (n = 6) based on Methyl-seq (upper) and  
 10 chromatin immunoprecipitation with massive parallel sequencing using a CTCF  
 11 antibody (CTCF-ChIP-seq) of an ESC line in H19-ICRs (lower). Raw data of each iPSC  
 12 line are presented in Figure S3C. (I) Methylation levels of 38 CpGs investigated by  
 13 Infinium450K array at H19-ICRs. H19-ICRs were defined by CTCF-ChIP-seq (Figure  
 14 S3C). \* indicates iPSC lines with gain of imprinting in H19-ICRs (median value of the  
 15 methylation > 0.75). (J,K) ATAC-seq analysis of a (J) IGF2-high iPSC line (7KE2),  
 16 IGF2-low iPSC line (Cira29), and (K) ESC line (khES3) with 4 ng/ml bFGF and 100  
 17 ng/ml bFGF at the loci of IGF2, GATA6, and REX1 (pluripotent gene: control). (L)  
 18 Conceptual view of commitment capacity. (G, I) Cell lines were grouped into three  
 19 groups, IGF2-high (n = 10), -int (n = 19), and -low iPS/ESC lines (n = 10), according to

IGF2 expression.

**See also Figures S2-3.**

**Figure 3 Hematopoietic Maturation Capacity is Associated with Distinct DNA Methylation Patterns of Undifferentiated PSCs**

(A) Gating strategy of PSC-derived HPCs (CD34+38-43+lin-) for the CFU assay and

(B) the result of a CFU assay of 39 iPS/ESC lines. Error bars indicate SD (n = 2-4).

Colors indicate the cell type origin of the iPSCs or ESCs. (C-F) Erythroblast

differentiation. (C) Representative flow cytometric analysis. (D) The pellet (top) and

May Grunwald/Giemsa staining (bottom) of magnetically-sorted glycophorin A (+)

erythroblasts on day 35. (E) The number of generated glycophorin A+ cells derived

from 16 PSC lines (n = 1-2). (F) Heat map of expression levels of erythrocyte-related

genes in iPS/ESC lines and their derived glycophorin A+ erythroblasts analyzed by

microarray. (G,H) Megakaryocyte differentiation using MegaCult™-C Complete kit

with Cytokines (Stemcell Technologies). (G) Flow cytometric gating strategy for

megakaryocyte differentiation and alkaline phosphatase (AP) staining of CD41a (+)

cells using AP-conjugated CD41a antibodies. (H) The results of megakaryocyte

differentiation of 16 PSC lines (n = 1-2). (I) Scatter plot of gene expressions of MC (n =

18) and MD lines (n = 18) in the undifferentiated state by microarray analysis. Red dots

1 indicate differentially expressed genes (165 microarray probes, 107 genes; FC > 2, FDR  
2 < 0.05). See also Table S2. (J) iPS-DMRs (396 probes, 84 genes, red dots; FDR < 0.05  
3 and  $\Delta$  methylation > 0.2) were identified by comparing DNA methylation profiles  
4 between iPSC (n = 56) and ESC lines (n = 20) using Infinium 450K array. Data from  
5 the Infinium 450K array and a public database (GSE30654) were used for the analysis.  
6 See also Table S5. (K) PCA of iPSC (n = 33) and ESC lines (n = 4) with iPS-DMRs.  
7 Shapes indicate the origin of iPSCs. iPSC lines are colored by maturation capacity. (E,  
8 H) Statistical comparisons were performed by t-test. Error bars are SD. **See also Figure**  
9 **S4.**

10  
11 **Figure 4 Aberrant DNA Methylation in MD-high DMRs was Acquired during**  
12 **Reprogramming**

13 (A) MD-high DMRs (277 probes, 80 genes, red dots; FDR < 0.05 and  $\Delta$  methylation >  
14 0.2) were identified by comparing DNA methylation profiles between MD-PSC (n = 18)  
15 and MC-PSC lines (n = 18) on autosomes by Infinium 450K array. (B, left) Prospective  
16 prediction of maturation capacity of 9 iPSC lines. The iPSC lines were divided into two  
17 groups (predicted MD lines and predicted MC lines) according to the methylation levels  
18 of MD-high DMRs and DMRs in TRIM58 genes (10 CpGs). MD-high DMRs: white  
19 lines and black boxes indicate the median and quartile; TRIM58: middle line indicates

median. (B, right) Subsequent validation of the prediction using CFU assays of HPCs from these lines. (C) Methylation levels of MD-high DMRs in parental lines of iPSCs, MD-PSCs, MC-PSCs, MD-HDF-iPSCs, MC-HDF-iPSCs, and iPSCs grouped by cell type origin and ESCs. \*\*\*  $p < 0.001$  (Mann-Whitney's U test or Kruskal-Wallis test). (D) DNA methylation levels of TRIM58 from parental lines to iPSCs ( $n = 32$ ). Infinium 450K array probes: TRIM58, cg10983544. (E) Methylation analysis of the TRIM58 gene by Methyl-seq in MD-HDF-iPSCs ( $n = 9$ ), MC-HDF-iPSCs ( $n = 2$ ), Blood-iPSCs ( $n = 8$ ), MD-Kera-iPSCs ( $n = 2$ ), and parental tissues (a HDF, a PB, and a Kera line) indicates the acquisition of hypermethylation in the TRIM58 promoter region during reprogramming in MD-HDF-iPSCs and MD-Kera-iPSCs, but not in Blood-iPSCs or MC-HDF-iPSCs. (F) Methylation levels of MD-high DMRs in intermediate cells during reprogramming from HDF to iPSCs on the designated day after the introduction of reprogramming factors and in an iPSC line (TIG120-4F1: MD-iPSC line) derived from the same HDF line. (G) Scatter plot of methylation levels of MD-high DMRs in HPCs from MC-PSC ( $n = 4$ ) and MD-PSC lines ( $n = 6$ ). (H, I) CFU assay of HDF-derived iPSCs (H, left) and DP-derived iPSC lines (I, left). Scatter plot of methylation levels of MD-high DMRs in MC and MD lines of HDF-iPSCs (H, right) and DP-iPSCs (I, right) are shown. (J) The number of mixed colonies in the CFU assay of iPS/ESC-derived HPCs (iPSCs  $n = 44$ , ESCs  $n = 4$ ). iPSC lines were grouped by cell type origin. The

results of Figure 3B and 4B are included. Red lines indicate median values. \*  $p < 0.05$  (Kruskal-Wallis test). **See also Figure S4-6.**

## **Figure 5 Aberrant DNA Methylation in Erythroid/Megakaryocyte-related Genes Impairs Hematopoietic Maturation**

(A, B) Erythroid-related genes (A) and megakaryocyte-related genes (B) are enriched in genes with MD-high DMRs. Erythroid- and megakaryocyte-related genes are defined as those expressed in iPSC-derived erythroblasts and megakaryocytes, respectively, at least three-fold more than in iPSCs or iPSC-derived HPCs. See also Figure S6B and C.

Statistical analysis was performed by the chi-square test. (C) The expression level of the TRIM58 gene in iPS/ESCs, HPCs, and erythroblasts derived from MC-HDF-iPSCs, MD-HDF-iPSCs, MC-PB-iPSCs, MC-CB-iPSCs, and ESCs. (D) Generation of TRIM58-KO iPSC lines by inducing a frame-shift mutation in the 1<sup>st</sup> exon of TRIM58 with the CRISPR/Cas9 system and validated by Sanger sequencing. Three TRIM58-KO lines were generated with different guide RNAs. (E) Western blotting analysis of erythroblasts differentiated from a wild type iPSC line and three KO lines using an antibody against TRIM58. (F-H) Assessment of the maturation capacity of WT-iPSCs and TRIM58-KO lines by (F) CFU assay, (G) erythroid differentiation, and (H) megakaryocyte differentiation. (I, J) Forced expression of TRIM58 in a MC-iPSC line

(409B2) enhanced erythroid differentiation capacity (I) and megakaryocyte differentiation capacity (J). Data are shown as fold change (G, I). (F-H) \*\*  $p < 0.01$ , \*  $p < 0.05$ : ANOVA and Dunnett's test;  $n = 3$ . (I-J) \*  $p < 0.05$ : t-test. (I)  $n = 4$ . (J)  $n = 3$ . Error bars are SD.

## **Figure 6 Residual DMRs are not the Determinant Factor in Hematopoietic Maturation Capacity**

(A) Strategy for the identification of residual DMRs. (B) Venn diagrams showing DMRs between a parental PB line and a parental HDF line and between a PB-iPSC line and a HDF-iPSC line from a single donor at p10 and p20. Red circles: PB-iPSC > HDF-iPSC; blue circles: HDF-iPSC > PB-iPSC; orange circles: PB > HDF; and green circles: HDF > PB. The yellow areas indicate inherited DMRs from parental lines to iPSCs (Residual DMRs). A difference in methylation > 0.2 was defined as significant. (C) Methylation levels of residual DMRs highly methylated in PB-iPSCs (448 Infinium 450K array probes, left) and in HDF-iPSCs (52 Infinium 450K array probes, right) on autosomes in MD-HDF-iPSCs, MC-HDF-iPSCs, Blood-iPSCs, parental lines (PB and HDF), and ESCs. (D) PCA of MD-HDF-iPSCs, MC-HDF-iPSCs, ESCs, Blood-iPSCs, DP-iPSCs, and Kera-iPSCs with methylation levels of HDF-high residual DMRs (left) and Blood-high residual DMRs (right). (E) Conceptual view of acquired aberrant DNA

methylation impairing maturation capacity. (F) Methylation levels of MD-high DMRs in hepatic differentiation-deficient iPSC lines (n = 15) and hepatic differentiation-competent iPSC lines (n = 7). \*\*\*  $p < 1 \times 10^{-3}$  (Mann-Whitney's U test). Hepatic maturation-deficient and -competent iPSC lines were defined by the criteria of albumin secretion (maturation deficient < 500 ng/ml/24 h, maturation competent > 1000 ng/ml/24 h) in a report by Kajiwara et al. 26 out of 39 PSC lines used in this study were overlapped with those used in the Kajiwara et al. study. **See also Figure S7.**

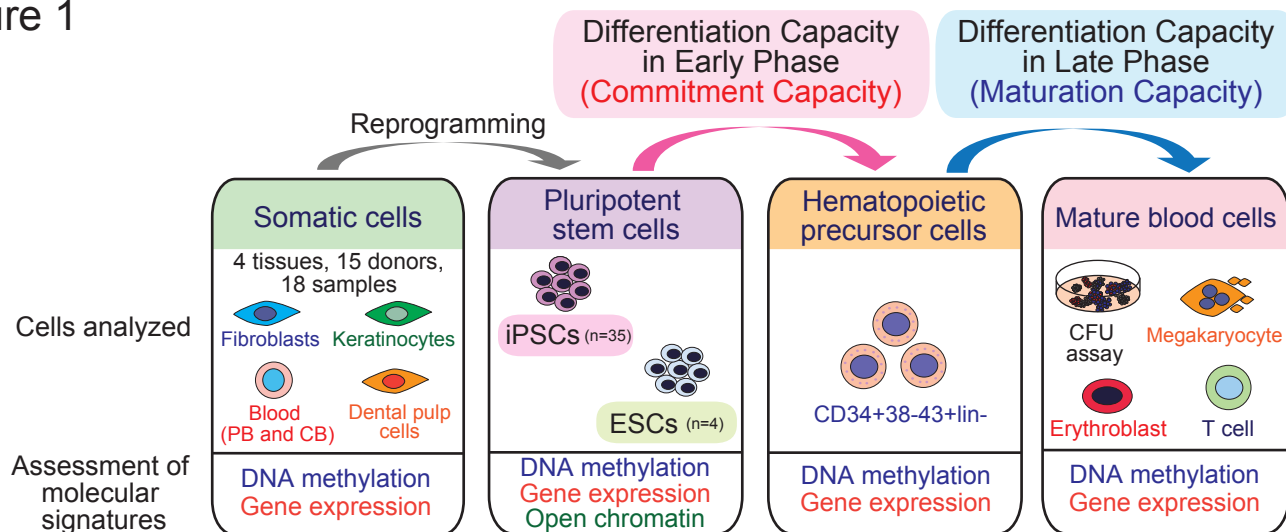
**Figure 7 Factors Responsible for the Diversity in the Capacity of Human iPSCs for Hematopoietic Commitment and Maturation.**

(A) Conceptual view of this study. (B) Receiver Operating Characteristic (ROC) curve analysis for commitment capacity (left) and maturation capacity (right). Statistical analysis was performed by calculating the area under the curve (AUC), \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ . (C) Methylation levels of all Infinium 450K probes on autosomes in parental tissues, ESCs, and iPSCs showed hypermethylated states in PB, CB, and ESCs compared with other somatic tissues. White lines indicate median values. (D) Number of CpGs that had changed their methylation levels from the original tissues to the corresponding iPSCs (CpGs:  $\Delta\beta$  value > 0.2). Blue: methylation, red: demethylation. Analysis by Infinium 450K array. (E) Number of CpGs in which the

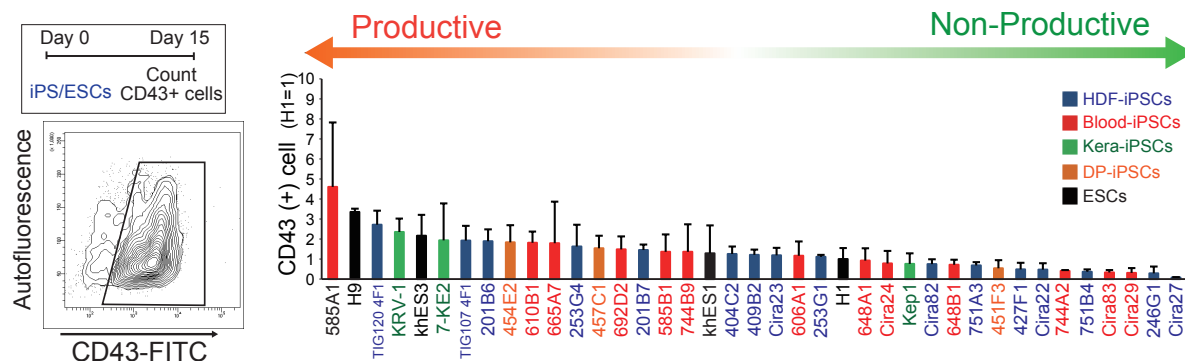
- 1 methylation levels were increased from the tissues to the corresponding iPSCs in
- 2 MD-high DMRs ( $\Delta\beta$  value > 0.2). Analysis by Infinium 450K array. (F) Conceptual
- 3 model of fewer aberrant DNA methylations in Blood-iPSCs.
- 4

Figure 1

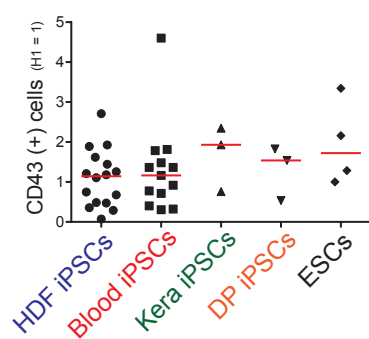
A



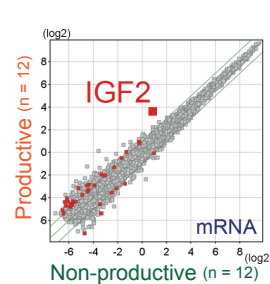
B



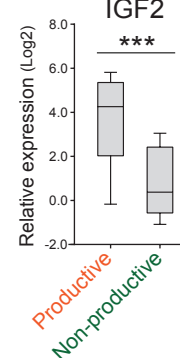
C



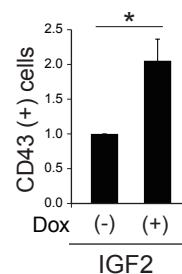
D



E



F



G

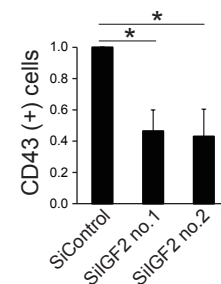


Figure 2

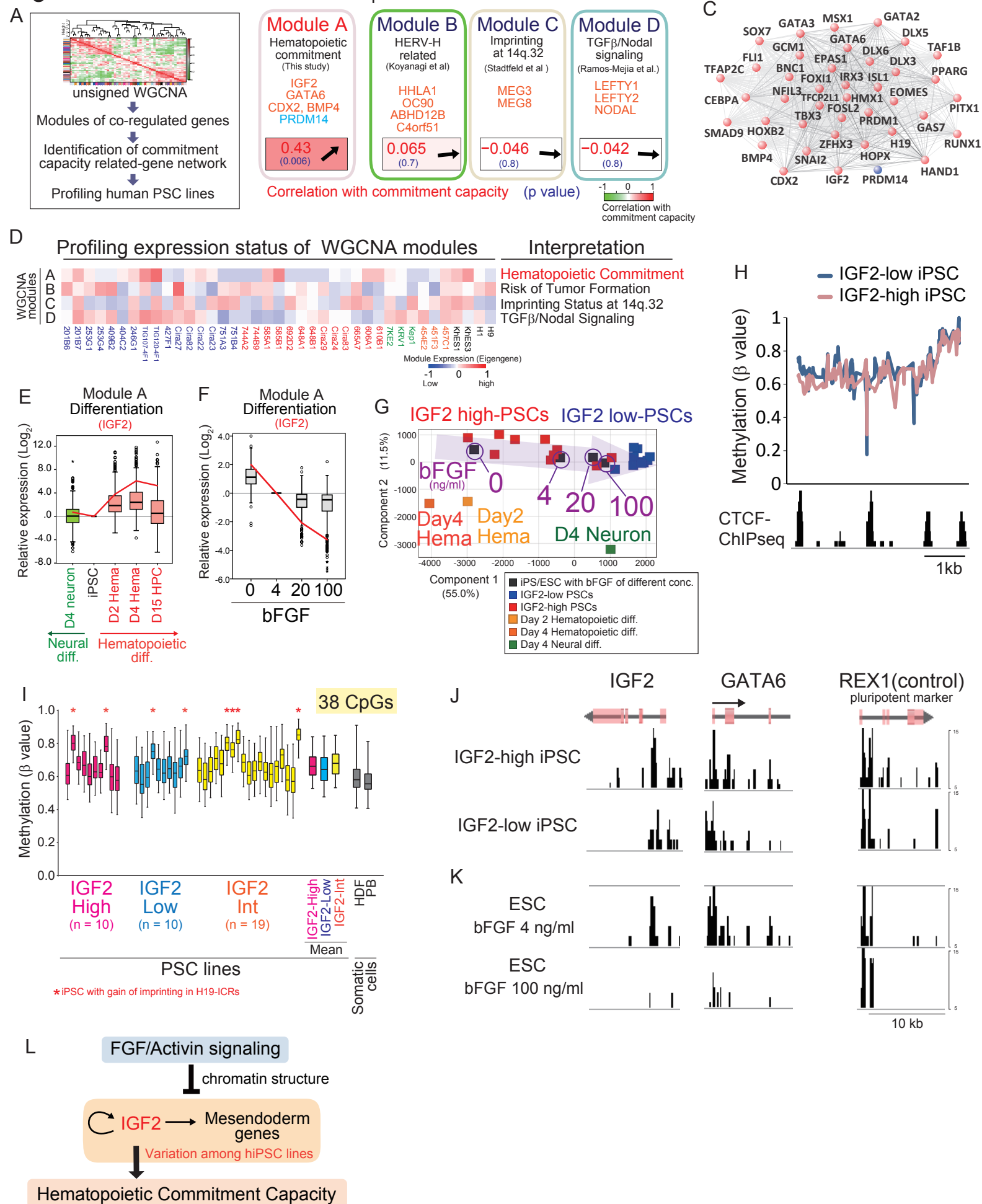


Figure 3

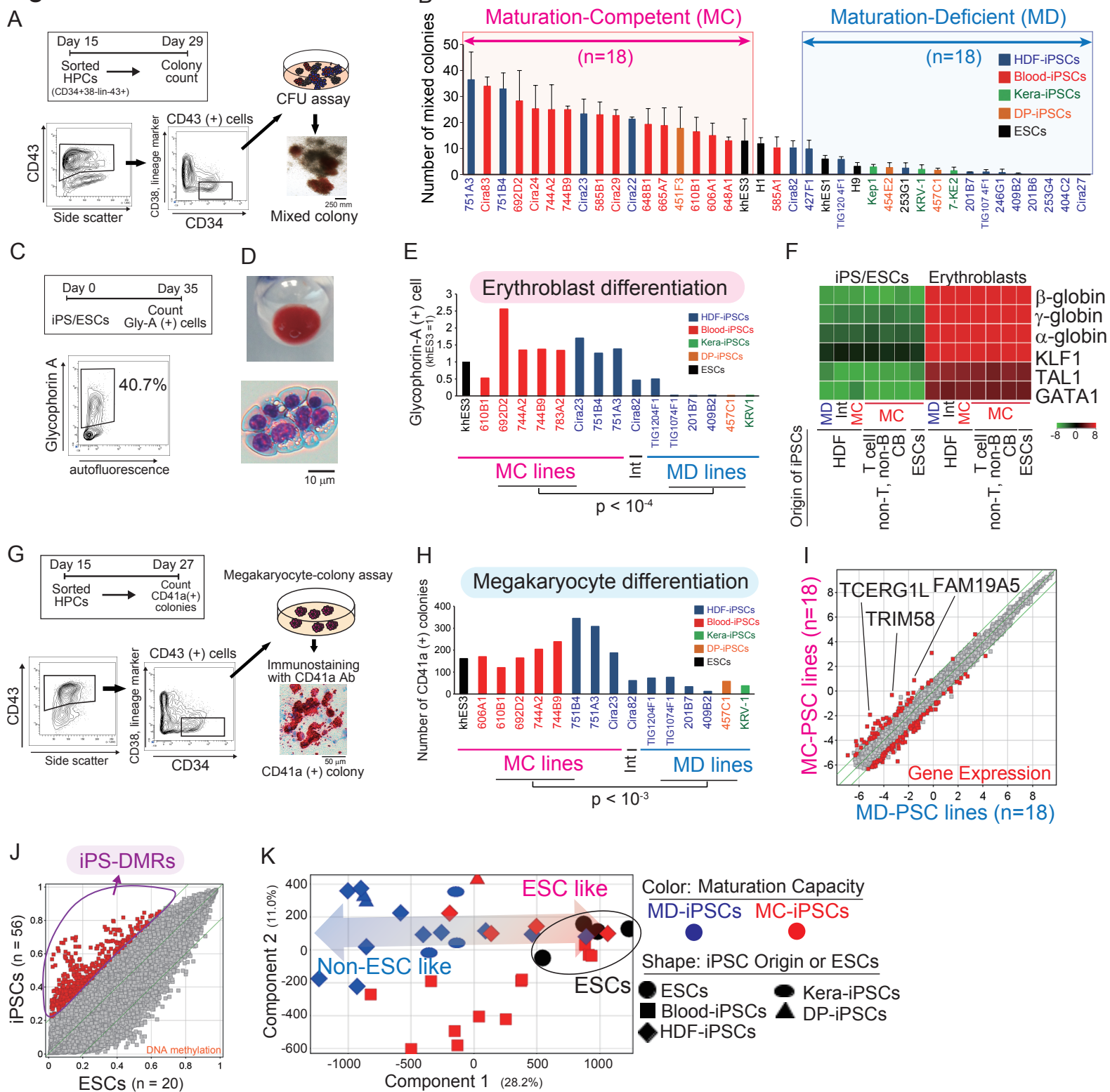


Figure 4

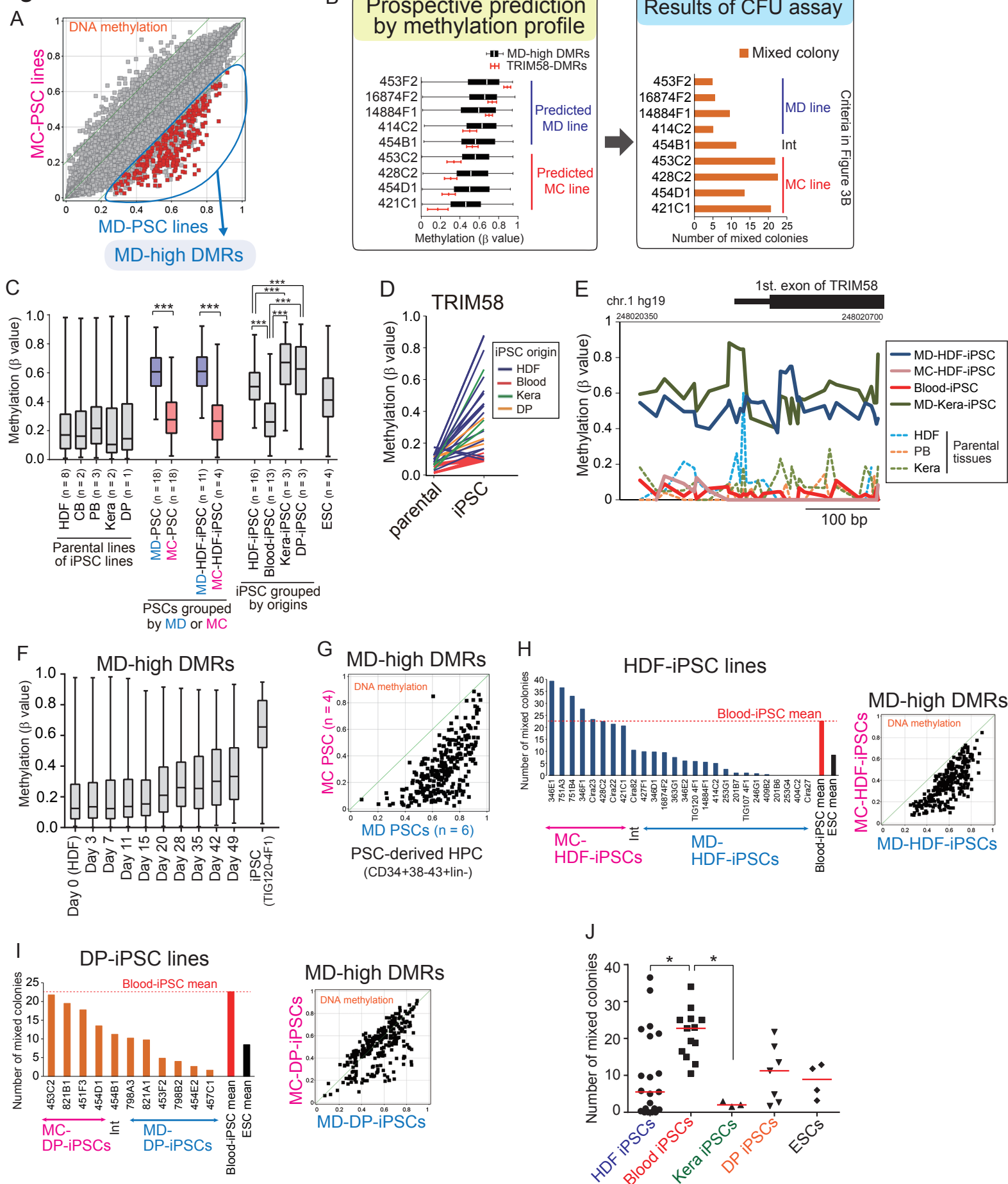


Figure 5

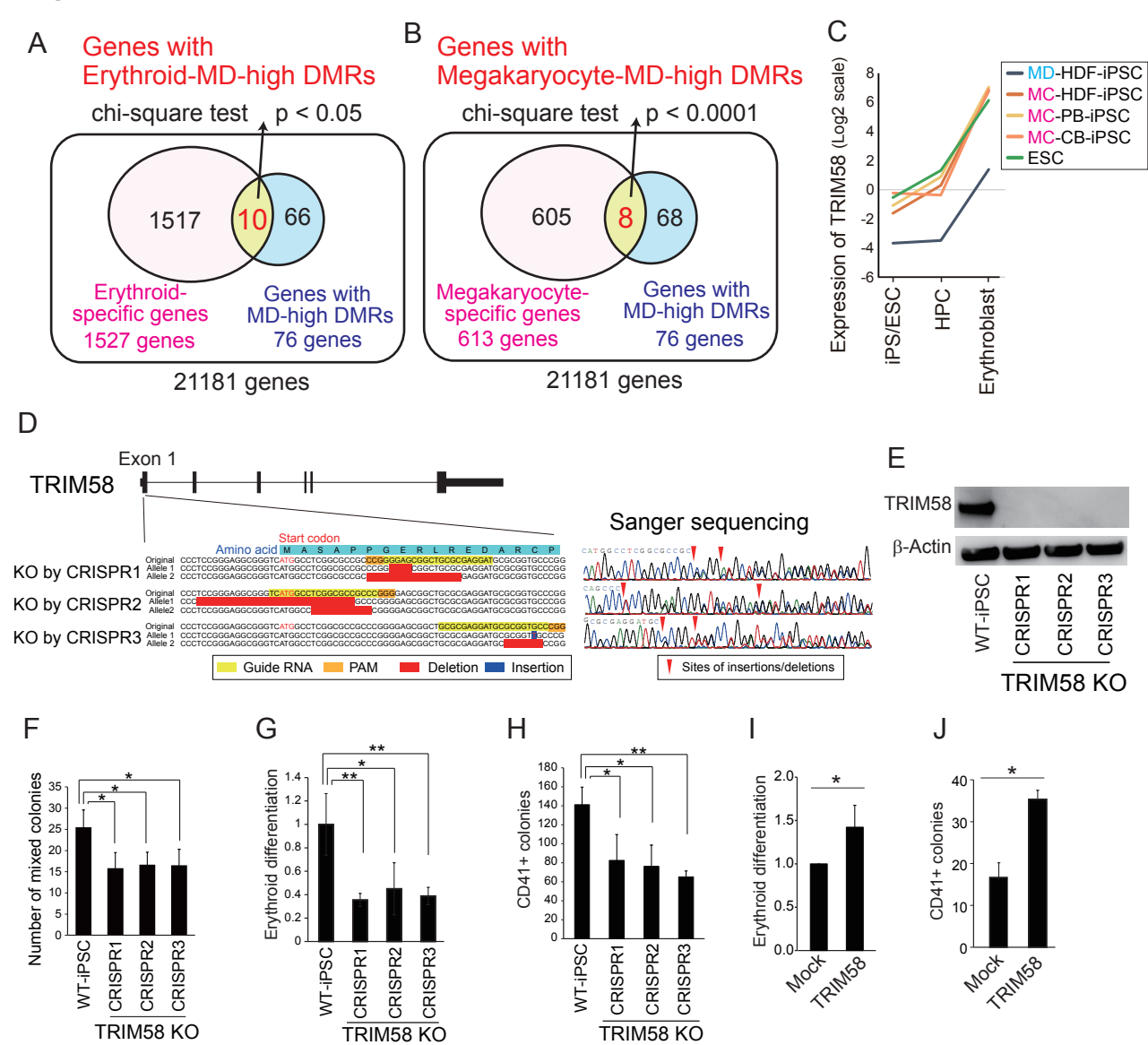


Figure 6

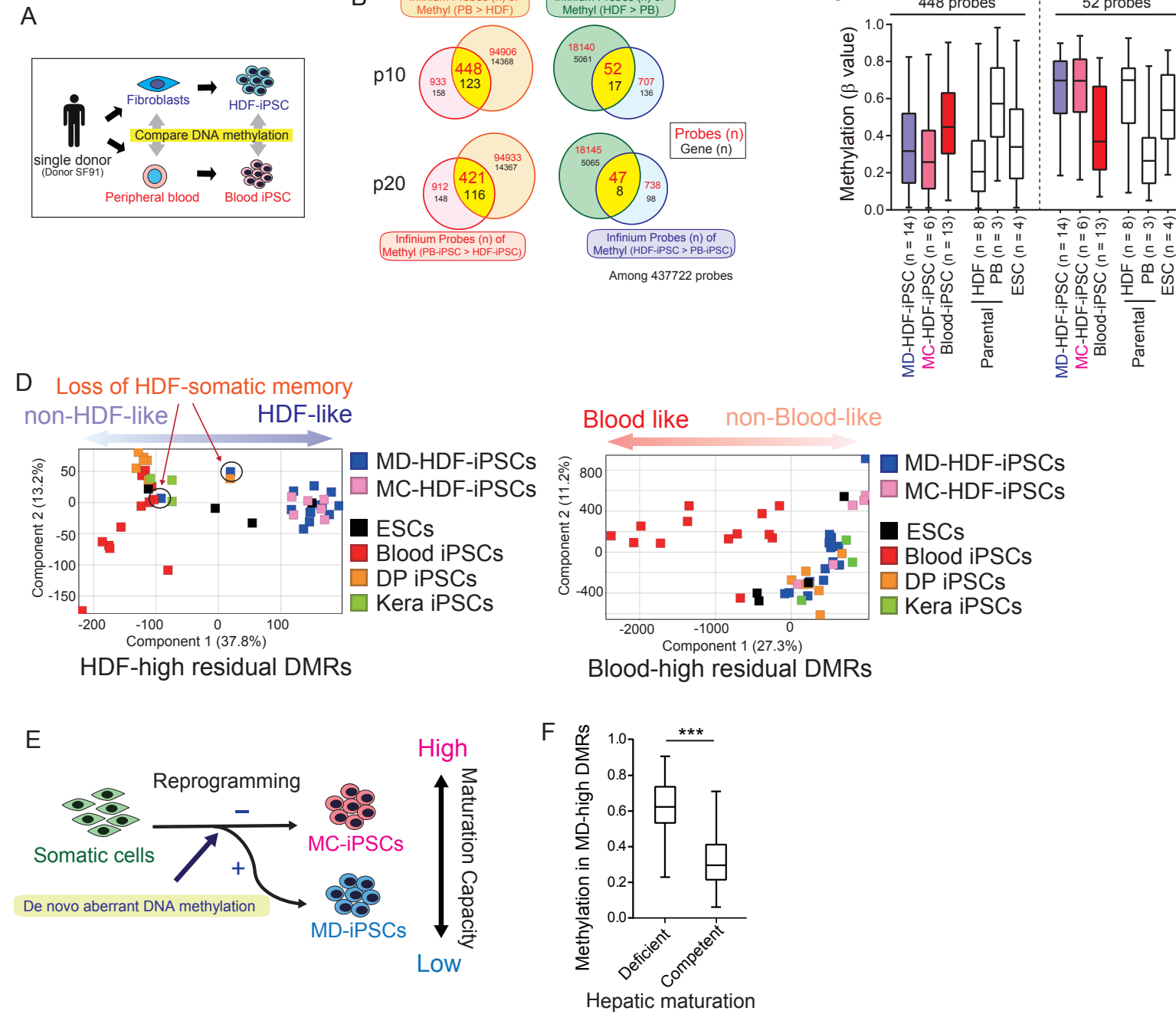
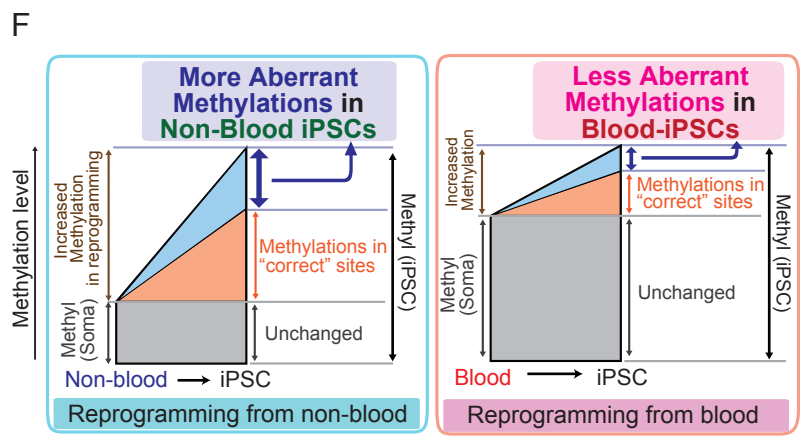
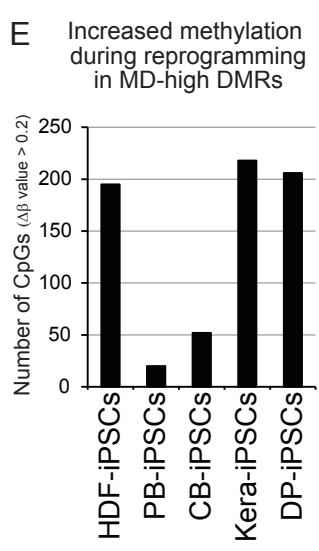
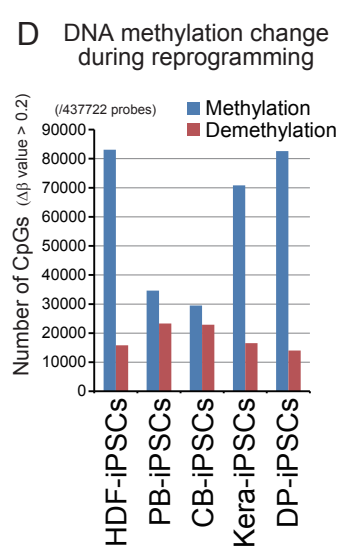
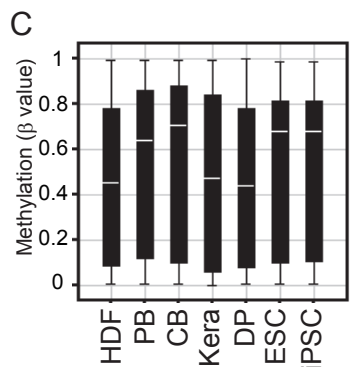
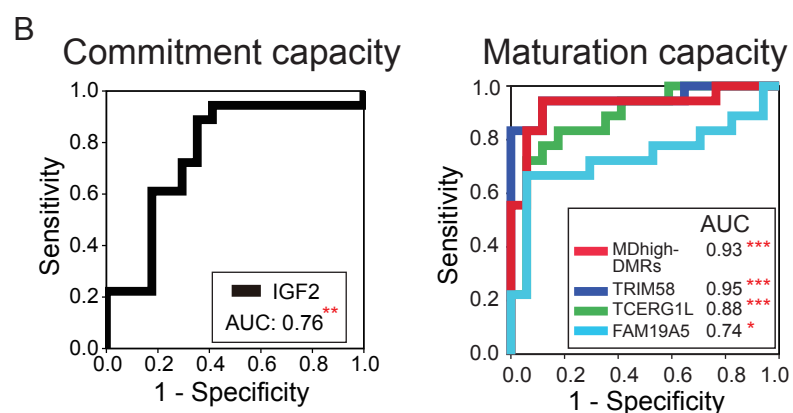
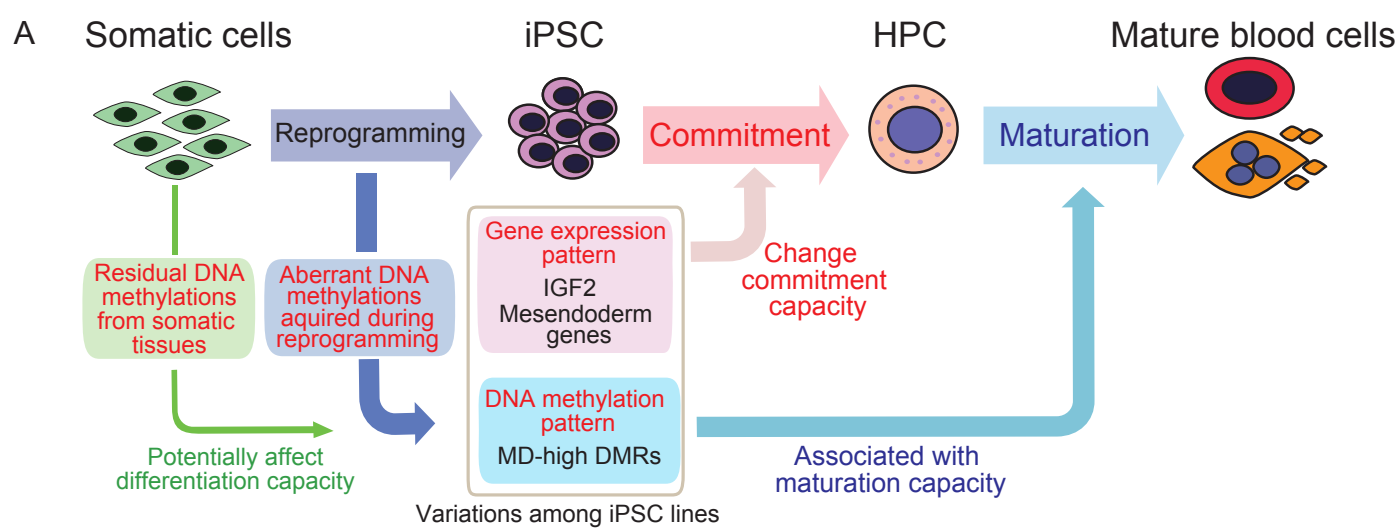
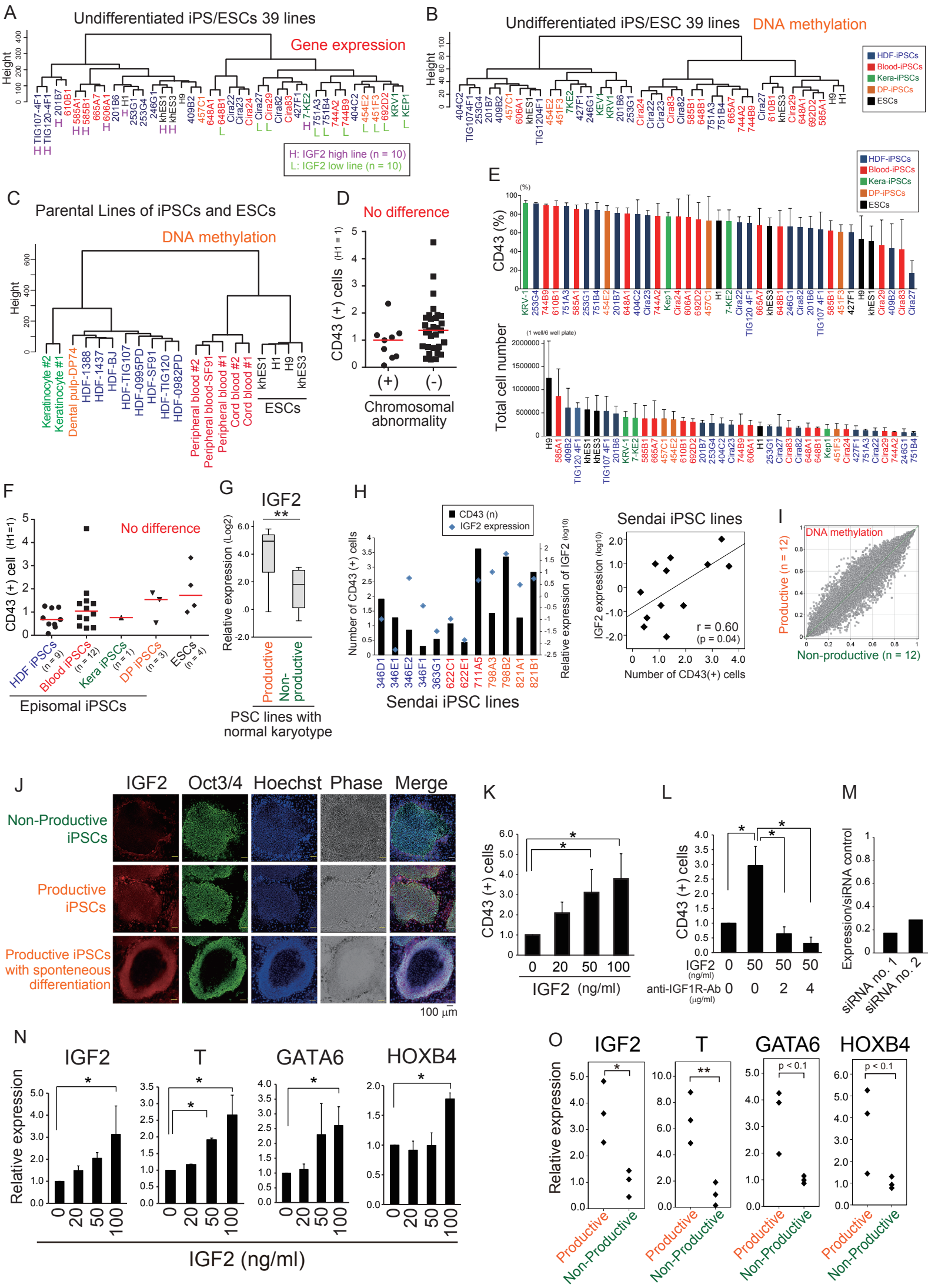


Figure 7  
Figure 7



# Supplemental information Supplementary Figure 1, Related to Figure 1

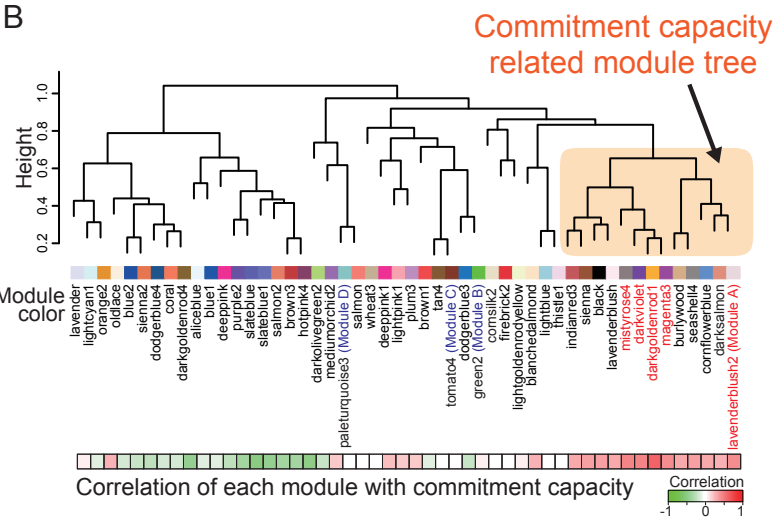
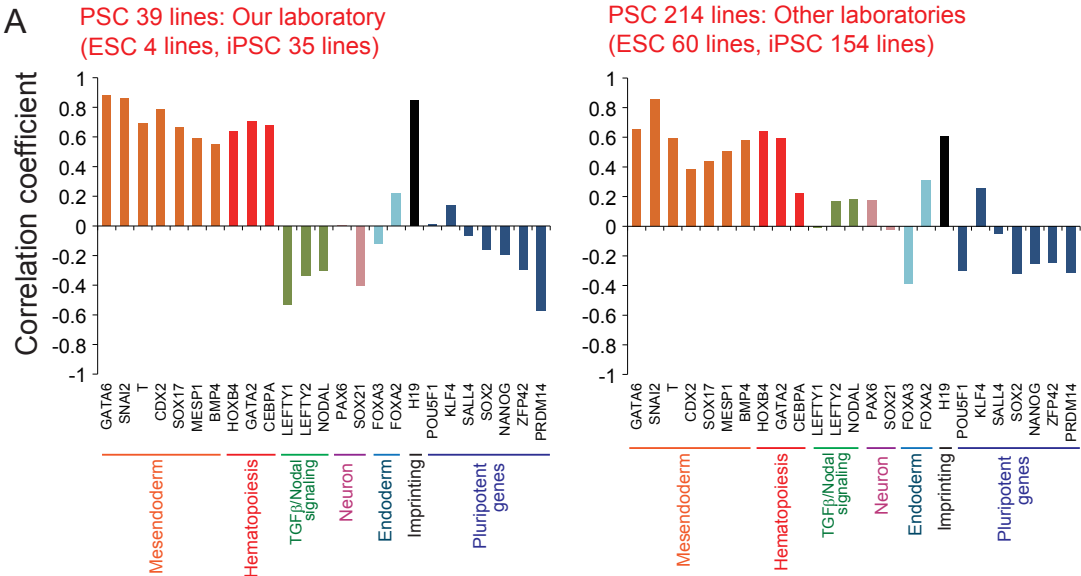


### **Supplementary Figure 1, related to Figure 1**

(A) (B) Hierarchical clustering of 39 iPSC/ESC lines using gene expression data from the microarray analysis (A) and DNA methylation data from Infinium450K array (B) on autosomes. PSC lines with high or low expression of IGF2 tend to be clustered separately (H: IGF2-high PSC lines,  $n = 10$ ; L: IGF2-low PSC lines,  $n = 10$ ). (C) Hierarchical clustering of the parental cell lines of iPSCs with DNA methylation data from Infinium 450K array on autosomes. 16 of 18 parental lines of the iPSC lines used in the differentiation experiments were analyzed. The analysis showed that hematopoietic cells such as PB or CB are much closer to ESCs in methylation profile than HDFs, Kera, or DP cells. (D) Comparison of commitment capacity between PSC lines with ( $n = 9$ ) and without chromosomal abnormality ( $n = 30$ ) revealed no difference between them. Statistical analysis was performed by t-test. Red lines indicate median values, related to Figure 1B. (E) The percentage of CD43<sup>+</sup> cells in 39 PSC lines (upper) and number of total generated cells differentiated from one well of a 6-well plate dish (lower) on day 15, related to Figure 1B. (F) Number of CD43<sup>+</sup> cells generated from episomal iPSC and ESC lines grouped by cell origin. There were no significant differences between iPSCs from the different origins. Statistical analysis was performed by ANOVA. (G) Comparison of IGF2 gene expression between productive PSC lines ( $n = 10$ ) and non-productive PSC lines ( $n = 10$ ) among PSCs with normal karyotypes revealed that high IGF2 expression was associated with favorable commitment capacity.  $** p < 0.01$  (t-test). (H) (left) Number of generated CD43<sup>+</sup> cells on day 15 ( $n = 2$ ) and IGF2 expression (log10) analyzed by quantitative polymerase chain reaction (qPCR) in 12 Sendai iPSC lines. (right) Pearson correlation analysis shows a positive correlation between the IGF2 expression and number of CD43<sup>+</sup> cells. IGF2 gene expression was normalized to the ACTB gene expression. The CD43<sup>+</sup> number was normalized to the number from the ES line (H1) in Figure 1B. The colors of the cell line names indicate the iPSC origin (blue: HDF-iPSCs, red: blood-iPSCs, and orange: DP-iPSCs). (I) Comparison between DNA methylation profiles of Productive ( $n = 12$ ) and Non-Productive PSC lines ( $n = 12$ ) showed no significantly differentially methylated probes between the two groups, suggesting commitment capacity was not associated with the DNA methylation profile. Significance was defined as  $p < 0.05$  calculated by Mann-Whitney's U test, FDR  $< 0.05$ , and difference in  $\beta$  value  $> 0.2$ . (J) Immunostaining with antibodies against IGF2 and OCT3/4, staining of Hoechst33342, and phase-contrast images of colonies of a Productive iPSC line (7KE2), a Non-Productive iPSC line (Cira29),

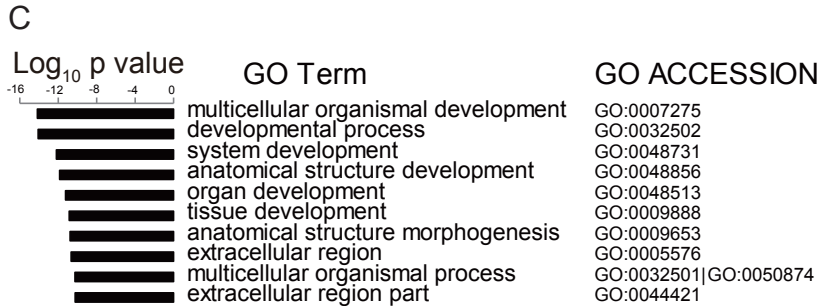
and a spontaneously differentiated colony of the 7KE2 line. (K) Number of generated CD43<sup>+</sup> cells on day 15 with the addition of IGF2 (day 0-15). \*  $p < 0.05$  (ANOVA and Dunnett's test;  $n = 3$ ). (L) Number of generated CD43<sup>+</sup> cells on day 15 with the addition of IGF2 and anti-IGF1 receptor antibody (day 0-15). \*  $p < 0.05$  (ANOVA and Tukey's posthoc test;  $n = 3$ ). (M) Knock-down efficiency of siRNA against IGF2 by qPCR. (N) qPCR analysis showed that the addition of IGF2 increased the expression level of IGF2, T, GATA6, and HOXB4 in hematopoietic differentiation on day 4. \*  $p < 0.05$  (ANOVA and Dunnett's test). IGF2 was administered on day 0-4. (O) The expression levels of IGF2, T, GATA6, and HOXB4 in 6 PSC lines (3 Productive PSC lines: 7KE2, TIG120-4F1, and khES3; and 3 Non-Productive iPSC lines: 751B4, Cira29, and 648B1) on day 4 by qPCR analysis. The relative expression was normalized to the mean expression of Non-Productive PSCs. \*\*  $p < 0.01$ , \*  $p < 0.05$  (t-test). Error bars are SD.

Supplementary Figure 2, Related to Figure2

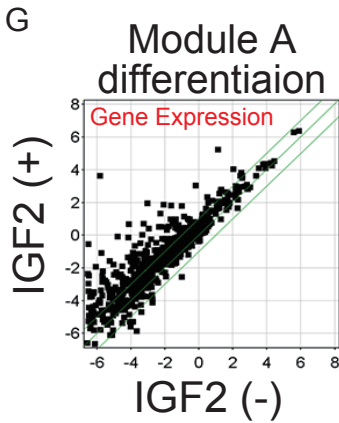
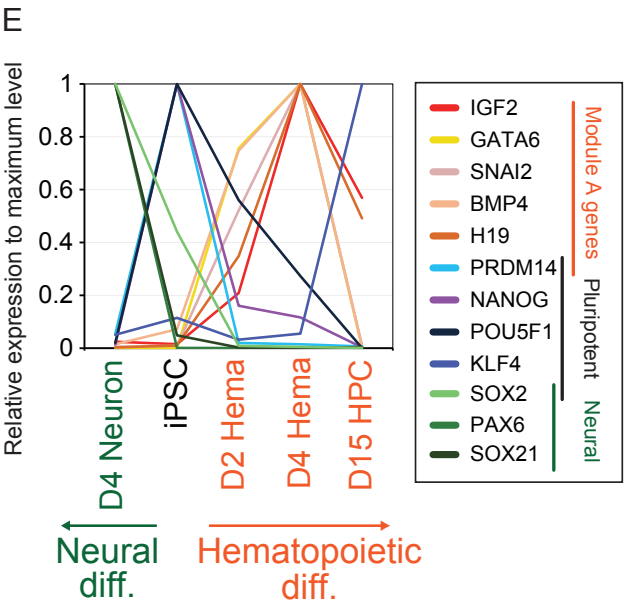
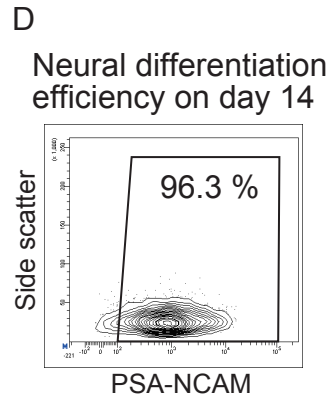
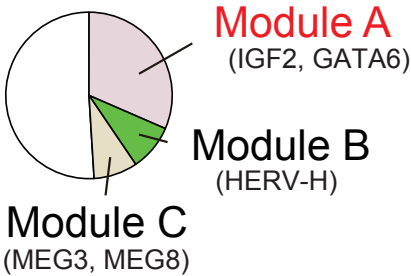


TOP5 WGCNA modules correlated with commitment

| Module color              | Correlation | p value | Representative genes            |
|---------------------------|-------------|---------|---------------------------------|
| darkgoldenrod1            | 0.62        | 0.00    | TGFB2                           |
| mistyrose4                | 0.48        | 0.00    | SOX17                           |
| darkviolet                | 0.46        | 0.00    |                                 |
| magenta3                  | 0.45        | 0.00    |                                 |
| lavenderblush2 (module A) | 0.43        | 0.01    | IGF2, GATA6, CDX2, BMP4, PRDM14 |



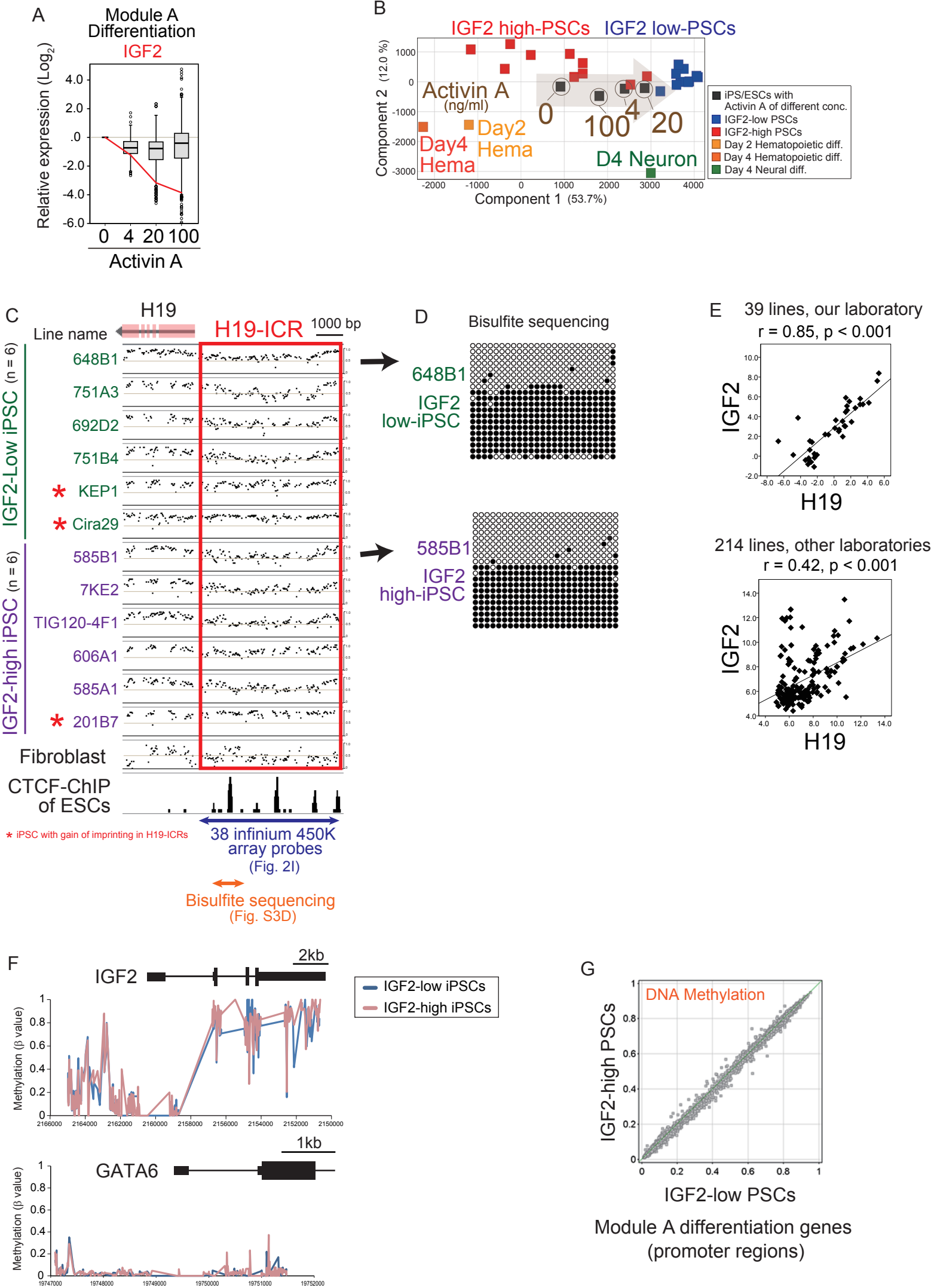
**F** Top 200 variable genes among human iPS/ESC lines



### **Supplementary Figure 2, related to Figure 2**

(A) Pearson's correlation coefficients of the expression level of IGF2 and other genes in undifferentiated iPSC/ESC lines using mRNA microarray data of our laboratory (n = 39, left) and other laboratories obtained from a public database (n = 214, right). Information of the microarray data of iPSC lines from other laboratories is listed in Table S3. (B) Hierarchical clustering of the 50 WGCNA modules using mRNA microarray data of 39 PSC lines (left). The top 5 modules that highly correlated with hematopoietic commitment capacity, their correlation coefficient with commitment capacity, p values, and representative genes (right). The correlation of each module with hematopoietic commitment capacity is shown as a heat map (bottom). This analysis provides a commitment capacity-related module tree (orange block). (C) Gene ontology (GO) analysis of Module A genes. The top 10 GO terms enriched in Module A genes are presented. (D) Neural differentiation efficiency on day 14 was assessed by flow cytometry using anti-PSA-NCAM antibody, related to Figure 2E, 2G, and S3B. (E) Gene expression profiles of iPSCs, D2Hema, D4Hema, and D4 neuron analyzed by microarray, related to Figures 2E, 2G, and S3B. (F) Module A genes accounted for 31.5% (63 out of 200 microarray probes) of the top 200 variable microarray probes among the 39 PSC lines. The top 200 variable genes were defined by calculating standard deviations. (G) A scatter plot showing forced expression of the IGF2 gene (IGF2 (+)) increased Module A differentiation gene expression. Expression was forced in iPSC line 744A2 by the doxycycline-inducible system for 7 days. Log<sub>2</sub> scale. Control: doxycycline-negative condition (IGF2 (-)). This result indicates that IGF2 overexpression recapitulates the expression pattern of IGF2-high iPSCs.

Supplementary figure 3, related to Figure 2



### **Supplementary Figure 3, Related to Figure 2.**

(A) Addition of Activin A decreased the expression of Module A genes in iPS/ESCs (average of one iPS and one ESC line) (Red: IGF2 gene). (B) PCA of IGF2-high iPSCs (n = 10), IGF2-low PSCs (n = 10), iPS/ESCs (average of one iPS and one ESC line) cultured with different concentrations of Activin A (0, 4, 20, and 100 ng/ml), D2 Hema, D4 Hema, and D4 Neuron with Module A genes showed the addition of Activin A turned IGF2-high iPS/ESCs into an IGF2-low PSC-like state. (C) Methylation levels of IGF2-high iPSC lines (n = 6), IGF2-low iPSC lines (n = 6), and a fibroblast line (control) analyzed by Methyl-seq (upper) and CTCF-ChIP-seq of an ESC line (khES1) (bottom) in H19-ICRs. \* indicates iPSC lines with gain of imprinting in H19-ICRs. According to this analysis, we re-defined H19-ICRs and used this new definition for the analyses in Figures 2H-I and S3D. (D) Conventional bisulfite sequencing of H19-ICRs of 2 iPSC lines. (E) Gene expression analysis of iPS/ESCs in our laboratory (n = 39) (upper) and those in other laboratories (n = 214) (lower) shows that the expression levels of IGF2 and H19 are linearly and positively correlated. Because H19-ICRs reciprocally regulate IGF2 and H19 genes, this finding suggests that another factor governs the expression level of IGF2. Correlation coefficient (r) was calculated by Pearson's method. (F) DNA methylation analysis by Methyl-seq showed promoter regions of IGF2 and GATA6 were unmethylated in IGF2-high iPSC (n = 6) and IGF2-low iPSC (n = 6) lines. The average methylation level of each group is shown. (G) Infinium 450K array revealed no significant difference in CpGs of the promoter regions of Module A differentiation genes between IGF2-high PSCs (n = 10) and IGF2-low PSCs (n = 10), indicating that the variation of these genes among lines is not caused by the DNA methylation status. Total 5044 CpGs were analyzed.

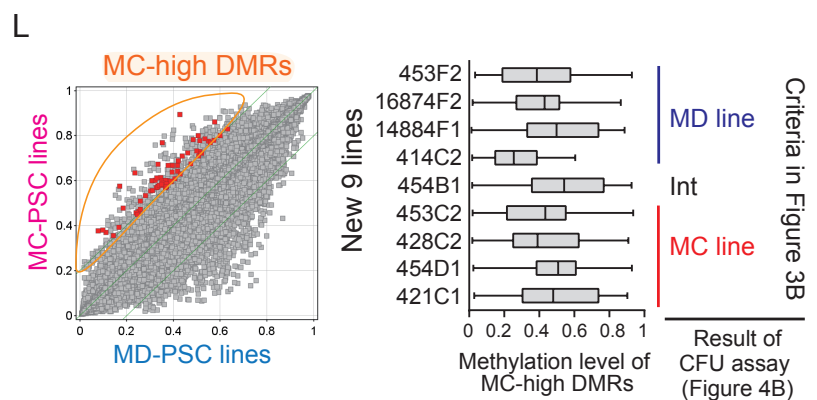
**A**

Number of Colonies

Legend:

- Mixed
- Erythroid (Erythroid + E burst colony)
- Myeloid (GM+G+M colony)

Cell Lines (X-axis): 751A3, Cira83, 751B4, 692D2, Cira24, 7444B, 7448B, Cira23, 565B1, Cira29, Cira22, 648B1, 665A7, 451F3, 610B1, 606A1, 648A1, khes3, H1, 585A1, Cira82, 427F1, khes1, TG129.4F1, H9, Kcp1, 454E2, 253G1, KR.V1, 457C1, 7.KE2, T01B7, TG107.4F1, 246 G1, 409B2.



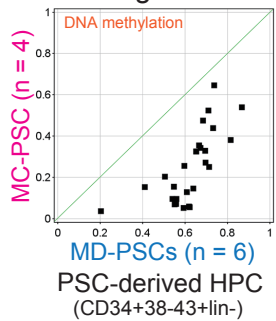
#### **Supplementary Figure 4, related to Figure 3-4**

(A) Result of the CFU assay, related to Figure 3B. Error bars are SD. (B) Comparison of maturation capacity between PSC lines with (n = 9) and without chromosomal abnormality (n = 30) revealed no differences. Red lines indicate median values. Statistical analysis was performed by Mann-Whitney's U test, related to Figure 3B. (C-D) T cell differentiation from iPS/ESC lines. We used a two-step differentiation method, namely differentiation from HPCs to progenitor T cells (pro-T) and from pro-T to CD4+CD8+ double positive T cells. (C) Gating strategy of CD7+CD45RA+ pro-T cells on T cell differentiation day 6-7 (left) and CD7+CD45RA+ percentage of a PB-iPSC line with no B or T cell rearrangement, PB-iPSC lines with T cell rearrangement, MC-HDF-iPSC lines, MD-HDF-iPSC lines, and an ESC line (H1) (n = 1-3). MD-HDF-iPSC lines showed very poor pro-T differentiation capacity, preventing their further differentiation into double positive T-cells. (D) Flow cytometric analysis of CD4+CD8+ double positive T cell differentiation from pro-T cells from a MC-HDF-iPSC line (751B4). (E) The scatter plot shows that the number of mixed colonies (maturation capacity) and of generated CD43+ cells in the CFU assay (commitment capacity) were not correlated. Correlation coefficient (r) was calculated by Pearson's method. (F) Methyl-seq analysis of TCERG1L and FAM19A5 genes in MD-iPSC lines (n = 11), MC-iPSC lines (n = 9), and parental lines (a HDF, a PB, and a Kera line) shows hypermethylation of the promoter regions of these genes in MD-iPSC lines, but no methylation in the parental tissues. (G) PCA of episomal iPSCs (n = 25) and ESCs (n = 4) with methylation profiles in iPS-DMRs shows that iPSCs with similar methylation profiles to ESCs had higher maturation capacity than iPSC lines with distinct methylation profiles. This tendency was more clearly observed when comparing episomal iPSC lines than when comparing both retro iPSCs and episomal iPSCs (Figure 3K). Shapes indicate the origin of the iPSCs. iPSCs are colored by maturation capacity. (H) Comparison of MD-iPSC lines with normal karyotypes (n = 12) and MC-iPSC lines with normal karyotypes (n = 13) showed differences in methylation levels of MD-high DMRs. \*\*\* p < 0.001 (Mann-Whitney's U test). (I) Methylation levels of TRIM58 in MD-PSC lines and MC-PSC lines obtained by Infinium 450K array (Infinium 450K array probes: cg10983544). \*\*\* p < 0.001 (Mann-Whitney's U test). (J) Validation of DMRs in TRIM58 using pyrosequencing analysis of bisulfite-converted DNA. DNA methylation analyses of 6 iPSC lines (3 MD-iPSC lines and 3 MC-iPSC lines) by pyrosequencing analysis and Infinium 450K array are presented. (K, left) Methylation levels of MD-high DMRs and

DMRs in the TRIM58 gene of 12 Sendai iPSC lines. (right) Numbers of mixed colonies in the CFU assay in these lines. These data suggest that the extent of aberrant DNA methylation levels can predict maturation capacity in Sendai iPSC lines. Colors of the cell line names indicate iPSC origins (blue: HDF-iPSCs, red: blood-iPSCs, and orange: DP-iPSCs). (L, left) MC-high DMRs are defined (69 probes, 25 genes, red dots; FDR < 0.05 and  $\Delta$  methylation > 0.2) (right) Analysis of methylation levels of MC-high MDRs and CFU assay of HPCs from newly assayed 9 iPSC lines revealed that methylation levels of MC-high DMRs cannot predict maturation capacity. Error bars are SD.

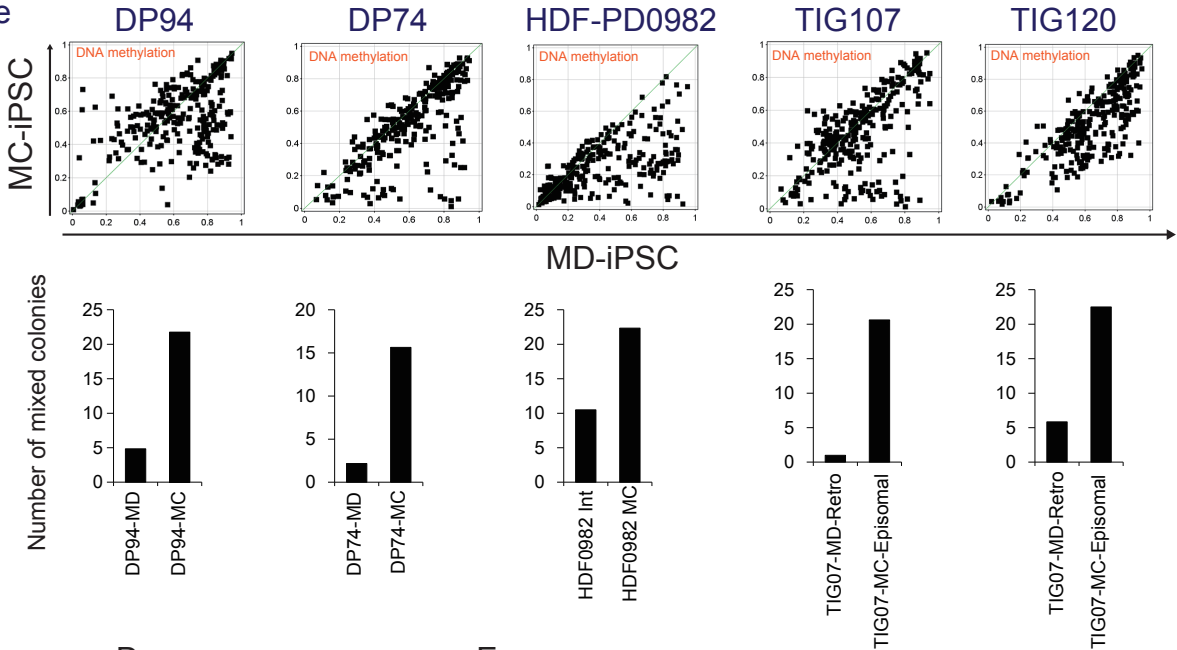
Supplementary Figure 5, Related to Figure 4

A Erythroid/Megakaryocyte  
MD-high DMRs

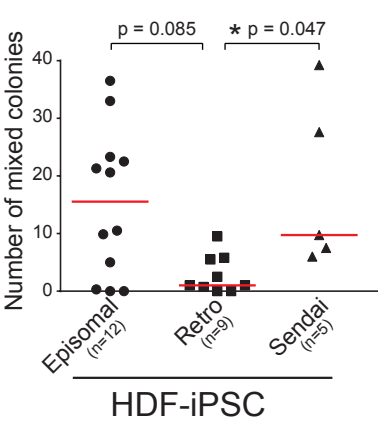


B Parental line name

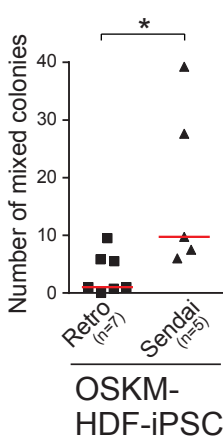
DNA methylation  
(MD-high DMRs)



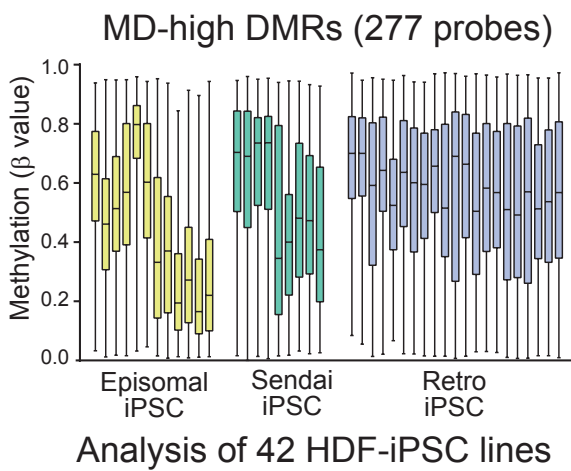
C



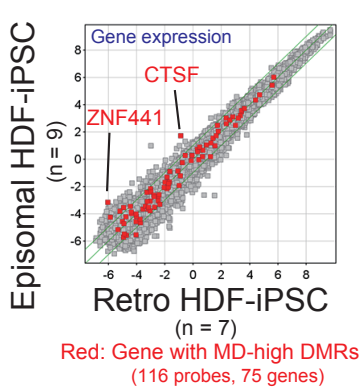
D



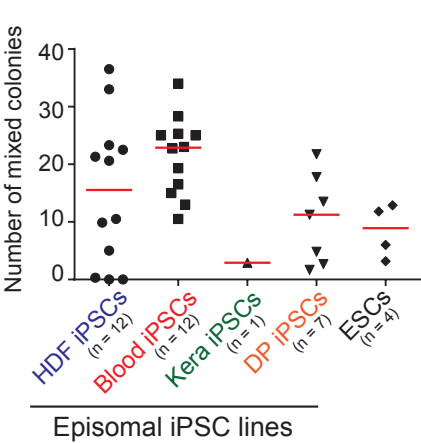
E



F



G

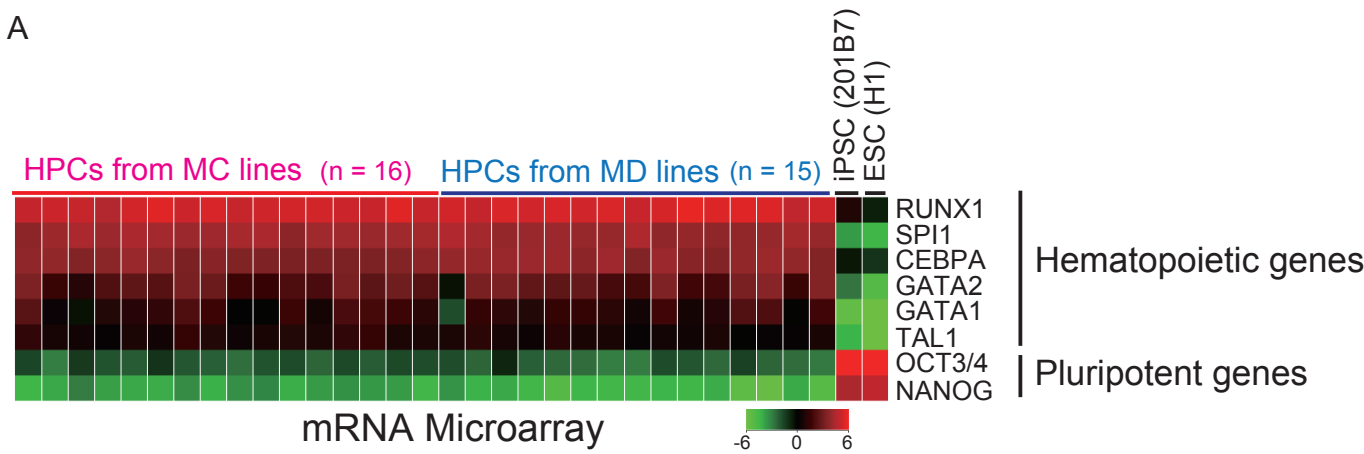


### **Supplementary Figure 5, related to Figure 4**

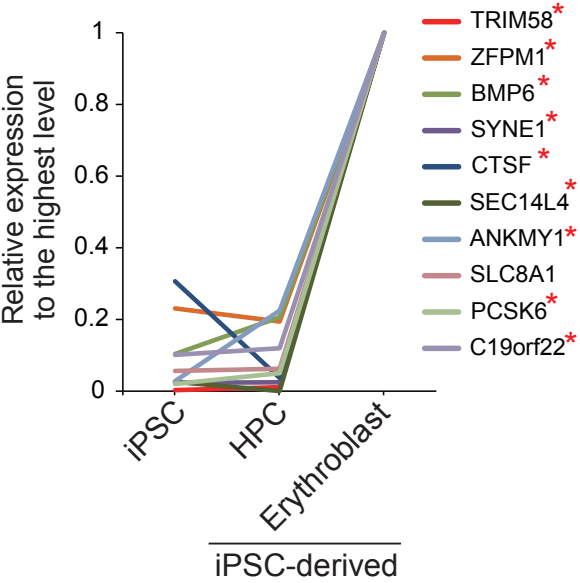
(A) Scatter plot of methylation levels of MD-high DMRs in erythroid/megakaryocyte-related genes between PSCs derived from MC-PSC ( $n = 4$ ) and MD-PSC lines ( $n = 6$ ), indicating that the difference in methylation profiles in MD-high DMRs in these genes was transmitted to their derived HPCs. (B) Comparison of methylation levels of MD-high DMRs (upper) and the number of mixed colonies in CFU assays (lower) of five sets of iPSC lines derived from the same tissues of single donors with different maturation capacities. In every set, the MD-iPSC line showed hypermethylation of MD-high DMRs compared with the corresponding MC-iPSC line. Three sets of iPSC lines (iPSCs from DP94, DP74, and HDF-PD0982) were generated using episomal vectors and 2 sets of iPSC lines (iPSC lines generated from HDF-TIG107 and HDF-TIG120) using retrovirus vectors or episomal vectors as indicated. (C) Comparison of number of mixed colonies in CFU assays among HDF-iPSCs generated with different reprogramming methods (episomal vectors:  $n = 12$ , retrovirus vectors:  $n = 9$ , and Sendai virus vectors:  $n = 5$ ). Statistical analysis was performed by Kruskal-Wallis test. \*:  $p < 0.05$ . (D) Comparison of the number of mixed colonies in CFU assays between HDF-iPSC lines generated with 4 reprogramming factors (OCT3/4, SOX2, KLF4, and C-MYC) and different reprogramming vectors (retrovirus vectors:  $n = 7$  and Sendai virus vectors:  $n = 5$ ) revealed that those generated with Sendai virus vectors had higher maturation capacity. \*:  $p < 0.05$  (Mann-Whitney's U test). (E) Methylation levels of MD-high DMRs in episomal HDF-iPSCs ( $n = 12$ ), Sendai-HDF-iPSCs ( $n = 9$ ), and retro HDF-iPSCs ( $n = 21$ ). All episomal HDF-iPSC lines, 9 out of 21 retro HDF-iPSC lines, and 5 out of 9 Sendai HDF-iPSC lines were generated in our lab; the other 12 retro HDF-iPSC lines and 4 Sendai HDF-iPSCs were created in other laboratories, and the corresponding methylation data were obtained from a public database (GEO accession: GSE31848, GSE53096, and GSE60225) (Nazor et al., 2012)(Ma et al., 2014; Nazor et al., 2014). (F) Scatter plot of gene expressions on autosomes between episomal HDF-iPSCs and retro HDF-iPSCs shows that only several genes with MD-high DMRs were differentially expressed. Genes with MD-high DMRs (75 genes, 116 probes) are highlighted red. (G) Mixed colonies of episomal iPSC lines ( $n = 32$ ) in the CFU assay were grouped by cell type origin and ESCs. Red lines indicate median values. There was no statistical difference between different origins by Kruskal-Wallis test, related to Figure 4J.

Supplementary Figure 6, Related to Figure 4

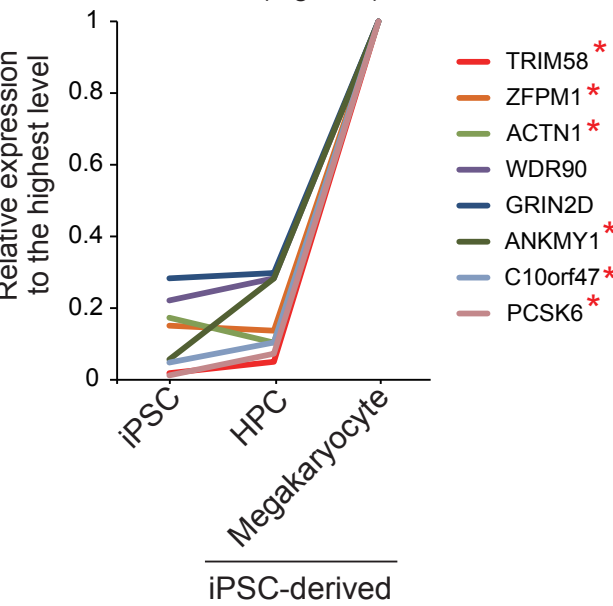
A



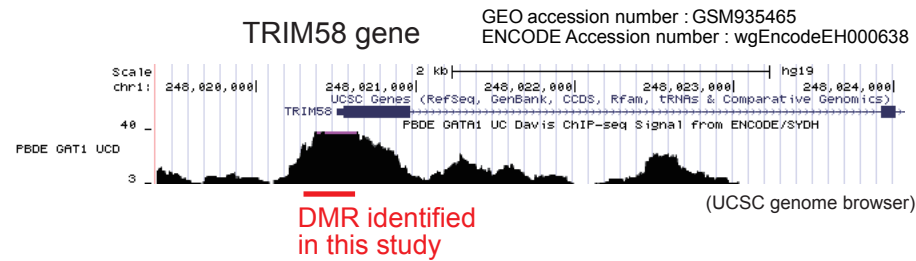
B Genes with erythroid-MD-high DMRs (10 genes)



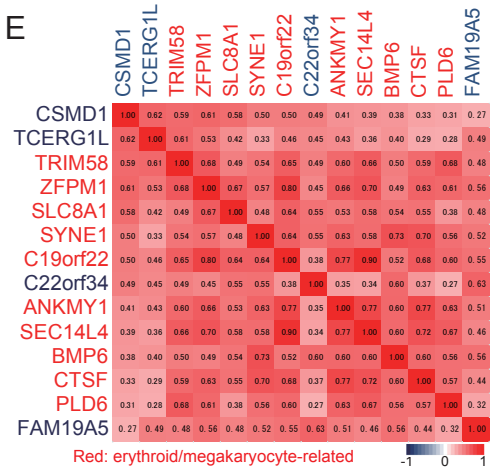
C Genes with megakaryocyte-MD-high DMRs (8 genes)



D ChIP-sequencing with GATA1 antibody using peripheral blood-derived erythroblasts



E

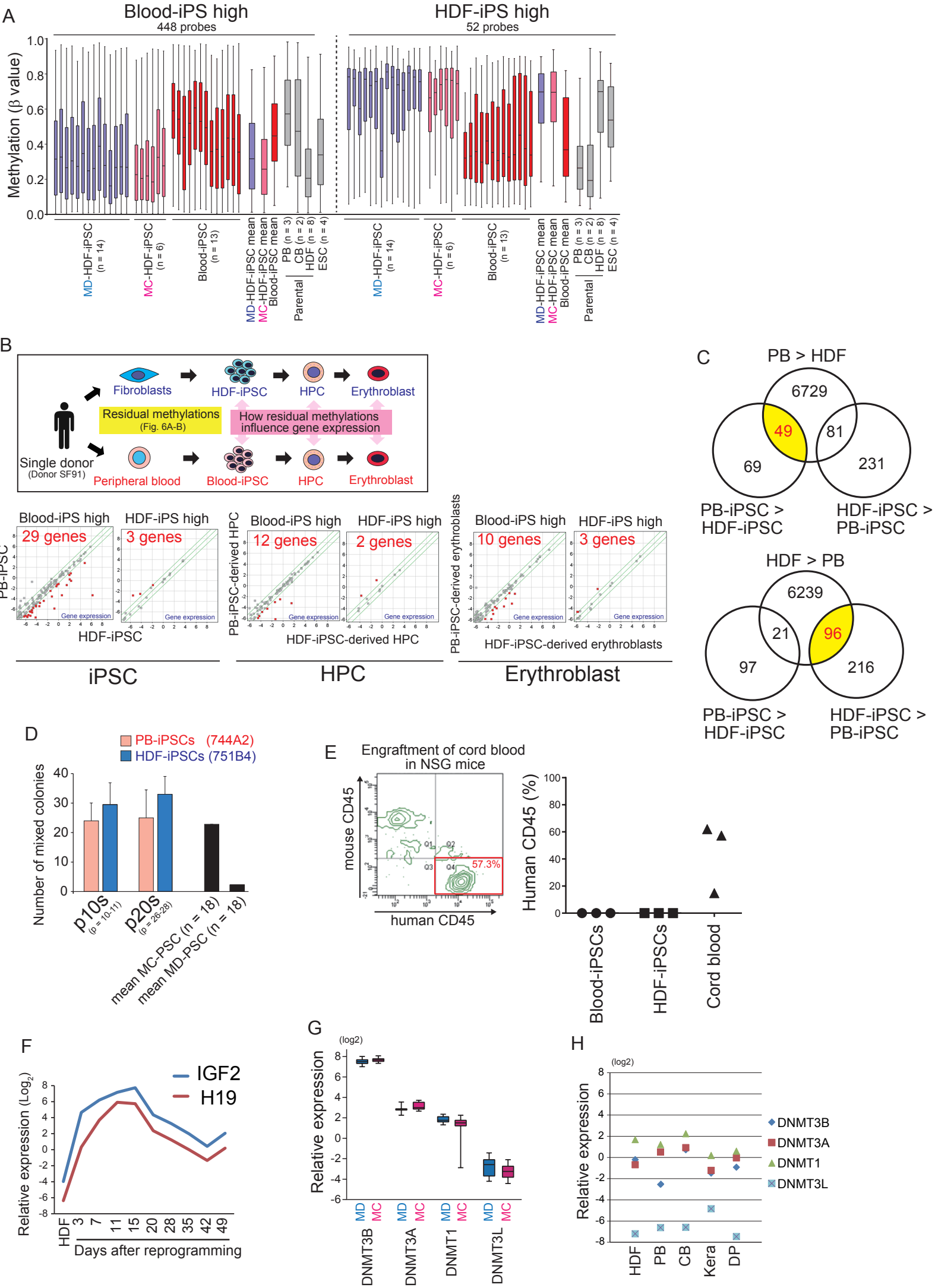


Spearman's correlation coefficient between DMRs (39 PCS lines)

### **Supplementary Figure 6, related to Figure 5**

(A) Heat map showing expression levels of representative hematopoietic stem cell marker genes and pluripotent genes in CD34<sup>+</sup>38<sup>-</sup>43<sup>+</sup> lin<sup>-</sup> HPCs differentiated from MC (n = 16) and MD lines (n = 15), and undifferentiated iPSC/ESC lines. (B) The expression levels of erythroid-related genes with MD-high DMRs in iPSCs, iPSC-derived HPCs, and iPSC-derived glycophorin A (+) erythroblasts were analyzed by mRNA microarray using the 751B4 iPSC line and its derivatives. (C) The expression levels of megakaryocyte-related genes with MD-high DMRs in iPSCs, iPSC-derived HPCs, and iPSC-derived megakaryocytes were analyzed by mRNA microarray using the 692D2 iPSC line and its derivatives. The expression data of 692D2 iPSC-derived megakaryocytes were obtained from an open database (GSM1309050) (Nakamura et al., 2014). Genes with asterisks indicate the presence of binding sites of the GATA1 based on ChIP-sequencing analysis from previous reports (Fujiwara et al., 2009; Tijssen et al., 2011) or data from public databases (ENCODE project; GEO accession number: GSM935465). Importantly, most erythroid/megakaryocyte-related genes with MD-high DMRs were direct targets of GATA1. (D) ChIP sequencing of PB-derived erythroblasts with GATA1 antibody in the ENCODE project (GEO accession number: GSM935465, ENCODE accession number: wgEncodeEH000638) showed MD-high DMRs in the TRIM58 gene are GATA1-binding sites. (E) Heat map showing Spearman's correlation coefficient of methylation levels of MD-high DMRs revealed that some iPSC lines harbored aberrant DNA methylation in multiple loci in MD-high DMRs. The genes highlighted red are erythroid/megakaryocyte-related genes.

Supplementary Figure 7, related to Figure 6 and Discussion



### Supplementary Figure 7, related to Figure 6 and Discussion

(A) Box plots show methylation levels of residual DMRs highly methylated in PB-iPSCs (448 Infinium 450K array probes, left) and HDF-iPSCs (52 Infinium 450K array probes, right) in MD-HDF-iPSCs, MC-HDF-iPSCs, Blood-iPSCs, parental lines (HDF, PB, and CB), and ESCs. (B) Comparison of expression levels of genes with residual DNA methylations (defined in Figure 6A-B) in the single donor (donor SF91)-derived iPSCs, HPCs (CD34+38-43+lin-), and magnetically sorted glycophorin-A<sup>+</sup> erythroblasts. Genes in which hypo/hyper-methylation of residual DNA methylations increased/suppressed expression ( $FC > 2$ ) are colored red and listed in Table S7. This result indicates that residual DNA methylations from parental tissues affected the gene expression profiles of iPSCs and iPSC-derived cells after differentiation. (C) Venn diagrams show differentially expressed microarray probes between parental HDF lines ( $n = 4$ ) and PB lines ( $n = 3$ ) ( $FC > 2$ ) and between HDF-iPSC lines ( $n = 16$ ) and PB-iPSC lines ( $n = 10$ ) ( $FC > 2$ ). The yellow areas indicate genes with PB > HDF and PB-iPSC > HDF-iPSC (49 probes, 31 genes) or those genes with PB < HDF and PB-iPSC < HDF-iPSC (96 probes, 80 genes). A total of 27584 microarray probes were analyzed. These genes are listed in Table S7. (D) Number of mixed colonies in the CFU assay of a PB-iPSC line (744A2) and HDF-iPSC line (751B4) derived from a single donor at p10s (p10-11) and p20s (p26-28), mean MC-PSCs, and mean MD-PSCs indicate that the MC-HDF-iPSC line (751B4) was maturation-competent from p10s and that there were no differences in maturation capacity between the two lines at p10s or p20s. (E) Engraftment rate of iPSC-derived HPCs and CB cells in NSG mice. (left) Representative FACS analysis of NSG mice transplanted with CB. (right) Engraftment rates. Engraftment of CD34+43<sup>+</sup> HPCs derived from MD-HDF-iPSCs (409B2), those from MC-Blood-iPSCs (692D2), and CD34<sup>+</sup> CB cells in sublethally-irradiated NSG mice (1.5 Gy) was assessed by the proportion of human CD45<sup>+</sup> cells in mouse PB 3 months after transplantation.  $1-4 \times 10^5$  iPSC-derived HPCs or CB cells were transplanted intravenously ( $n = 3$ ). (F) Change in the expression level of IGF2 and H19 during reprogramming from HDF based on data from another paper (GEO accession number: GSE54848) (Ohnuki et al., 2014) ( $\log_2$  scale). (G) Expression level of DNMT family genes in MC- ( $n = 18$ ) and MD-iPSC lines ( $n = 18$ ) analyzed by microarray ( $\log_2$  scale). (H) Expression of DNMT family genes in parental somatic cells analyzed by microarray ( $\log_2$  scale). Error bars are SD.

## **Supplementary Tables**

**Table S1**      **Human iPS/ESC lines used in Differentiation Experiments, related to Figure 1**

**Table S2**      **Differentially Expressed Genes between Productive and Non-Productive PSC Lines and between MC-PSC and MD-PSC Lines, related to Figure 1 and Figure 3**

**Table S3**      **List of Gene Expression Array Data of Human iPSCs and ESCs in Other Laboratories used for Analyses in Figures S2A and S3E, related to Figure 2**

**Table S4**      **List of Module A Genes in WGCNA of 39 iPS/ESC Lines, related to Figure 2**

**Table S5**      **List of iPS-DMRs, MD-High DMRs, and MC-High DMRs. List of iPS/ESC Lines used to Identify iPS-DMRs, related to Figure 3 and Figure 4**

**Table S6**      **Overlaps between Aberrant DNA Methylations in the Present and Other Studies, related to Figure 3 and Figure 4**

**Table S7**      **List of Residual DMRs, Residual DMRs that Affect Gene expression**

**of iPSCs and iPSC-Derivatives, and Genes Differentially Expressed  
between HDF and PB and between HDF-iPSC Lines and PB-iPSC  
Lines, related to Figure 6**

## **Supplemental Experimental Procedures**

### **Generation of Human iPSCs**

Detailed information of the iPSC lines used is listed in Table S1. iPSCs derived from human dermal fibroblast (HDF) were generated by retrovirus vectors with 3 or 4 reprogramming factors (OCT3/4, KLF4, SOX2,  $\pm$  c-MYC) (Nakagawa et al., 2008; Takahashi et al., 2007) or by episomal plasmid vectors with 6 to 7 reprogramming factors (OCT3/4, KLF4, SOX2, LIN28, L-MYC/C-MYC,  $\pm$  NANOG, and  $\pm$  short hairpin RNA against p53(shp53)) as reported previously (Okita et al., 2011). Dental pulp (DP)-iPSCs were generated with episomal vectors and 6 or 7 factors as reported previously (Okita et al., 2011). Keratinocytes (Kera)-iPSC lines were generated by retrovirus vectors (7KE2 and KRV1) with 4 factors or by episomal plasmid vectors (KEP1) with 6 reprogramming factors as described above. Keratinocytes were purchased from Lonza or ATCC. iPSCs derived from cord blood (CB) or peripheral (PB) using episomal vectors were created as reported previously (Okita et al., 2013). Sendai iPSC lines were generated from CB, PB, HDF, or DP cells using Sendai virus vectors carrying OCT3/4, SOX2, KLF4, and c-MYC as described previously (Fusaki et al., 2009; Seki et al., 2010). The generation of iPSCs from healthy donors or patients with Parkinson's disease (Cira22, Cira23, Cira24, Cira27,

Cira29, Cira82, and Cira83) was approved by the ethical committee of the Department of Medicine and Graduate School of Medicine, Kyoto University, and informed consent was obtained from the donors. All iPSC lines in the initial assessment (Figure 1B and 3B) were confirmed to have teratoma-forming ability. Histological analysis showed that the teratomas had tissues of all three germ layers, including the neural epithelium, pigment epithelium, cartilage, and gut-like epithelium. The animal experiments for the teratoma formation of human iPS/ESCs were performed following the code of ethics of the Animal Research Committee of Kyoto University. Karyotyping of all iPS/ESC lines used in the differentiation experiments was performed (Table S1).

### **RNA Extraction and Quantitative RT-PCR Analysis**

RNA extraction and reverse transcription were performed as previously reported (Kajiwara et al., 2012). Briefly, we lysed the cells with QIAzol Lysis Reagent (QIAGEN). RNA was extracted using miRNeasy Mini kit (QIAGEN) according to the manufacturer's protocol. Reverse transcription was performed with the oligo dT20 primer and ReverTra Ace- $\alpha$ -kit (Toyobo). TaqMan probes, TaqMan Gene Expression Master Mix, and StepOne Real-Time PCR System (Applied Biosystems) were used to investigate the expression levels of each gene. Expression levels were normalized against the expression level of an internal

housekeeping gene (ACTB). The qPCR Taqman probes used in this study were as follows.

IGF2: Hs00171254\_m1, GATA6 Hs00232018\_m1, T: Hs00610081\_m1, HOXB4: Hs00256884\_m1, and ACTB: Hs99999903\_m1.

### **Gene Expression Microarray Analysis**

Gene expression microarray analysis was performed using SurePrint G3 Human Gene Expression 8 × 60K kit (G4851A; Agilent technologies). The data were analyzed using GeneSpring GX 12.1 software (Agilent Technologies), and normalization of the expression data was performed as reported previously (Koyanagi-Aoi et al., 2013). We utilized only the probes detected in at least 10 iPS/ESC lines (out of 39) for the analysis. The number of microarray probes used for the analysis was 27584.

### **DNA Extraction and Bisulfite Conversion of DNA**

Genomic DNA was extracted with DNeasy Blood & Tissue kit (QIAGEN). Bisulfite conversion of genomic DNA was performed using EZ DNA Methylation-Gold kit (Zymo Research) as reported previously (Koyanagi-Aoi et al., 2013).

### **Analysis of DNA Methylation Status by Infinium 450K Methylation Beads Array**

Data processing was performed by Genome-Studio software (Illumina) and GeneSpring GX 12.1. Beta-values (methylated allele intensity/(unmethylated allele intensity +

methyated allele intensity)) were used to measure the methylation levels of CpG loci (0: unmethylated and 1: methylated). Quantile normalization was performed for the analysis of PSC lines. Regarding the analysis of original cells or PSC-derivatives, we did not perform quantile normalization because of the large variation in the distribution of methylation levels of the probes among samples. We excluded probes in which the target CpG was located within 10 bps of single nucleotide polymorphisms (SNPs) from the analyses to exclude bias by SNPs on the assays. We also excluded probes on sex chromosomes from the analyses to exclude a bias of sex differences among lines. The number of Infinium 450K array probes used for the analysis was 437722.

### **Hematopoietic Differentiation and Assays**

Hematopoietic differentiation was performed as previously described with some modification (Grigoriadis et al., 2010). PSCs were dissociated into small clusters seeded on a low-binding dish to form embryoid bodies and differentiated in suspension culture. Differentiation medium consisted of StemPro-34 (Invitrogen), 2 mM L-glutamine (Invitrogen),  $4 \times 10^{-4}$  M monothioglycerol (Sigma-Aldrich), 150 µg/ml transferrin (Roche), and 50 µg/mL ascorbic acid (Sigma-Aldrich). Cytokines or a chemical compound were utilized as follows during differentiation: on day 0-1, human BMP4 (hBMP4; 10 ng/mL, R

& D Systems) and Rock-inhibitor (Y27632; 10  $\mu$ M, Wako); on day 1-4, hBMP4 (10 ng/mL) and human basic fibroblast growth factor (hbFGF; 5 ng/mL, R & D Systems); on day 4-8, human vascular endothelial growth factor (hVEGF; 10 ng/mL, R & D Systems), hbFGF (1 ng/mL), human interleukin-6 (hIL-6; 10 ng/mL, R & D Systems), hIL-3 (40 ng/mL, R & D Systems), hIL-11 (5 ng/mL, R & D Systems), human stem cell factor (hSCF; 100 ng/mL, R & D Systems), and human Fms-like tyrosine kinase 3 ligand (hFlt3L; 100 ng/mL, R & D Systems); and on Day 8-20, hVEGF (10 ng/mL), hIL-6 (10 ng/mL), hIL-3 (40 ng/mL), hIL-11 (5 ng/mL), hSCF (100 ng/mL), hFlt3L (100 ng/mL), human erythropoietin (hEPO; 4U/mL, Calbiochem), and human thrombopoietin (hTPO; 50 ng/mL, R & D Systems). For erythroid differentiation, we differentiated PSCs by the aforementioned method for 20 days and changed the medium to erythroid differentiation medium, which contained hEPO (4U/ml), hSCF (100 ng/mL), and hIL-3 (5 ng/mL). On day 28, hIL-3 was discontinued. The efficiency of hematopoietic differentiation in early phase (commitment capacity) was assessed by counting the number of generated CD43<sup>+</sup> cells from 80% confluent PSCs in a 6-well culture plate with TC10 automated cell counter (Bio-Rad) and a flow cytometer (FACS ARIA II: Becton-Dickinson) on day 15. Erythroid differentiation efficiency was assessed by counting the number of generated glycophorin

(+) cells from 80% confluent PSCs in a 6-well culture plate with the flow cytometer on day 35. For the gene expression analysis of iPS/ESC-derived erythroblasts, erythroblasts were purified by CD235a (Glycophorin A)-MicroBeads (Miltenyi Biotec, 120-000-308) and autoMACS (Miltenyi Biotec) on day 35, and were subject to analysis. For CFU assays, we seeded a total of 5000 FACS-sorted hematopoietic precursor cells (CD34<sup>+</sup>38<sup>+</sup>43<sup>+</sup>lin<sup>-</sup>) into methylcellulose-based media (MethoCult H4435, Stemcell Technologies). After 14 days, the number of colonies was counted. Classification of each colony was based on the morphology and flow cytometric analysis. Mixed colonies were defined as those that contained several lineage cells, usually myeloid and erythroid cells. The definition of a colony was a cell cluster consisting of more than 20 cells. For megakaryocyte differentiation, we seeded a total of 5000 FACS-sorted CD34<sup>+</sup>38<sup>+</sup>43<sup>+</sup>lin<sup>-</sup> cells into collagen based medium (MegaCult™-C Complete kit with Cytokines). After 12 days, we performed immunostaining with anti-CD41a antibody according to the manufacturer's instructions and counted the number of CD41a positive colonies. For T cell differentiation, we used a co-culture system using OP9-DL1 stromal feeder cells with some modification of a previous report (Vizcardo et al., 2013). Briefly, fluorescence-activated cell sorting (FACS) sorted-CD34<sup>+</sup>38<sup>+</sup>lin<sup>-</sup>43<sup>+</sup> iPS/ESC-derived HPCs on day 15 were seeded onto OP9-DL1

stromal cells and cultured with  $\alpha$ -MEM (Invitrogen) with 20% FCS, human IL-7 (5 ng/ml, Invitrogen), hFlt-3L (5 ng/ml), and hSCF (10 ng/ml). To assess pro-T differentiation capacity, CD7+CD45RA+ cells were analyzed by flow cytometry on day 21 or 22. For further differentiation into CD4+CD8+ double positive T cell differentiation, CD7+CD45RA+CD5-CD34+ cells were sorted by FACS ARIA II on day 21 or 22, and the cells were transferred onto new dishes layered with OP9-DL1 cells and cultured with the medium above. Flow cytometric analysis was performed on day 43 with anti-CD4 and -CD8 antibodies. To assess the effect of exogenous IGF2 on hematopoietic commitment, recombinant human IGF2 (Peprotech) was used.

### **Flow Cytometric Analysis and Cell Sorting**

For the flow cytometric analysis or the cell sorting of PSC-derived hematopoietic cells, embryoid bodies containing hematopoietic cells were dispersed by trypsin and passed through a 21 gauge needle 7 times. The cells were then subject to flow cytometric analysis or cell sorting using FACS ARIA II. Dead cells were removed by 7-AAD (BD Pharmingen).

### **Analysis of Intermediate Cells during Reprogramming**

To analyze the methylation levels of MD-high DMRs, DNA methylation data from

intermediate cells during reprogramming from HDF to iPSCs in a previous paper (Ohnuki et al., 2014) were analyzed (GEO accession number: GSE54848). In our reprogramming setting using retroviral vectors carrying 4 reprogramming factors (OCT3/4, SOX2, KLF4, and C-Myc), the first small iPSC colonies became visible almost from day 20, and ESC-like colonies became clearly visible from day 25. Thereafter, the number of iPSC colonies was still increasing until around day 40.

### **Antibodies**

Antibodies used in the flow cytometric analysis included CD34-PE (BD Biosciences 348057), CD43-FITC (BD Biosciences, 555475), CD38-APC (BD Biosciences, 555462), CD235a-APC (BD Bioscience, 551336), Human Hematopoietic Lineage APC cocktail (eBioscience, 22-7776-72), CD4-APC (BD Biosciences, 340443), CD8-PE (BD Biosciences, 340046), CD45RA-Brilliant Violet421 (BioLegend, 304130), CD7-PE (BD Biosciences, 340581), CD5-FITC (BD Biosciences, 555342), CD34-APC (BD Biosciences, 340441), mouse anti-human CD45- BUV395 (BD Horizon, 563792), rat anti-mouse CD45-BV421 (BD Horizon, 563890), and PSA-NCAM (Miltenyi Biotec, 130-093-273). Anti-CD221 (IGF1 receptor) mouse monoclonal antibody (1H7 clone, Thermo Scientific) was used for blocking the IGF1 receptor.

## **Hierarchical Clustering Analysis and Principal Component Analysis**

Hierarchical clustering of PSC lines by Infinium450K array and mRNA microarray data was performed using hclust in R with Euclidian distance and Ward's linkage. Principal component analysis was performed by GeneSpring GX 12.1.

## **ROC curve analysis**

To perform ROC curve analysis, we classified 35 iPSC lines in the first assessment of differentiation capacity into productive (n = 17) and non-productive lines (n = 18) in early differentiation and maturation competent (n = 17) and maturation deficient lines (n = 18) in late differentiation. The AUC values were calculated, and the significances (p values) were calculated by testing the null hypothesis that AUC values were 0.5.

## **Analysis of DNA Methylation Status by Pyrosequencing**

Pyrosequencing was performed as reported previously (Kajiwara et al., 2012). Bisulfite-treated DNA was amplified by the PyroMark PCR kit (QIAGEN). Pyrosequencing was conducted using the PyroMarkQ96 ID system (QIAGEN) according to standard procedures. Primers for the analysis of TRIM58 were as follows: primer set 1

|                             |                               |           |
|-----------------------------|-------------------------------|-----------|
| (Forward:                   | GGAGGGATTTTAGTTAGAAATGTTTAGT, | Reverse:  |
| ACACCCCATAAAAAACCACCACTCAA, |                               | Sequence: |

AAAAAACCACCACTCAA), primer set 2 (Forward:  
TTTAGGGAGGAGGGTTATGGTTTG, Reverse:  
CTCCAAAATACACCTAAAACAAAACTATA, Sequence:  
TGTTTGGATTTTTTGTAGGA), primer set 3 (Forward:  
GGTGTGTTTGGATTTTTTGTAG, Reverse: CACCCTCTCCACCAAACC, Sequence:  
GTATTTTAGAGTTTTG), and primer set 4 (Forward:  
GTTGGAGGGTTTGGTGGAGAG, Reverse: AAAAACCCTCACCTAATAACTACC,  
Sequence: CAAAACCACTCAAATCCT).

### **Definition of iPS-DMRs**

iPS-DMRs were identified by comparing the DNA methylation profiles of human ESC lines (n = 20) and of human iPSC lines (n = 56) as investigated by Infinium 450K array. Significance was defined as  $p < 0.05$  calculated by Mann-Whitney's U test, FDR < 0.05, and a difference in  $\beta$  values > 0.2. The list of human iPSC and ESC lines used are shown in Table S5. Infinium 450K array data of 40 iPSC and 5 ESC lines were acquired in our laboratory, and the data of another 16 iPSC and 15 ESC lines were downloaded from a public database (GEO accession number: GSE30654) (Nazor et al., 2012).

### **Definition of MD-high DMRs and MC-high DMRs**

MD-high DMRs and MC-high DMRs were identified by comparing the DNA methylation profiles of MC-PSC lines (n = 18) and of MD-PSC lines (n = 18) as investigated by Infinium 450K array. Significance was defined as  $p < 0.05$  calculated by Mann-Whitney's U test, FDR < 0.05, and a difference in  $\beta$  values > 0.2.

### **May Grunwald/Giemsa Staining**

May Grunwald/Giemsa staining was performed by the May-Grunwald stain solution (Wako) and Giemsa solution (Wako) according to the manufacturer's instructions.

### **Plasmid Construction (IGF2 and TRIM58 genes)**

The open reading frame (ORF) of the IGF2 and TRIM58 genes were amplified by PCR and subcloned into pENTR-d-topo vector (Life Technology). The ORF of IGF2 was transferred to a PiggyBac vector with the doxycycline-inducible expression system (PB-A103-gw-GB), and the ORF of TRIM58 was transferred to another PiggyBac vector (PB-CAG-gw-IP) with constitutive expression of a gene of interest by a gateway LR reaction according to the manufacturer's instructions.

### **Generation of Doxycycline-inducible IGF2-expressing iPSC Line**

The doxycycline-inducible IGF2-expression vector (PB-A103-IGF2-GB) was transfected into an iPSC line (744A2, PB-derived IGF2-low line) with transposase-expressing vector

using FugeneHD (Promega), and then the transfected cells were selected using Blasticidin for a week. To induce IGF2 gene expression, doxycycline ranging from 0.04 to 1 ng/ml was added to the culture or differentiation medium.

### **Generation of TRIM58-Expressing iPSC Line and Control Line**

We transfected the TRIM58 gene-expressing vector (PB-CAG-TRIM58-IP), which enables constitutive expression of TRIM58, and mock vector (PB-CAG-IP) into 409B2 (MD-HDF-iPSC line) using FugeneHD (Promega), and then the transfected cells were selected using puromycin for a week.

### **Knockdown of IGF2 gene by RNA Interference in iPSCs**

Knockdown of the IGF2 gene was performed with Stealth RNAi against human IGF2 genes (no. 1: HSS179798, no.2: HSS179799; Life Technologies) and control siRNA (Med GC 12935-300; Life Technologies). These siRNAs were transfected into iPSCs using Stemfect RNA transfection kit according to the manufacturer's protocol (Stemgent). siRNAs were transfected on day -2 into iPSCs, which were cultured on a matrigel-coated dish with MEF conditioning medium supplemented with 4 ng/ml basic FGF. Two days after the transfection of siRNAs, the same number of iPSCs were aggregated using a 96-well plate dish to form embryoid bodies and subject to hematopoietic differentiation medium.

## **Neural Differentiation**

Neural differentiation of iPSCs was performed by the serum free embryoid body formation (SFEBq) method as reported previously with some modification (Koyanagi-Aoi et al., 2013) (Morizane et al., 2011). Briefly, after embryoid body formation of dissociated human iPS/ESCs, the cells were cultured for 14 days in neural differentiation medium in low binding culture dish (Lipidure-Coat Plate A-U96; NOF Corp). The neural differentiation medium consisted of DMEM/F-12 (Gibco), 1% MEM Non-Essential Amino Acids Solution (Gibco), 5% Knockout Serum Replacement (Gibco), 2 mM L-glutamine (Gibco), 0.1 mM 2-mercaptoethanol (Gibco), 10  $\mu$ M Y-27632, 2  $\mu$ M dorsomorphin (Sigma), and 10  $\mu$ M SB431542 (Sigma).

## **Establishment of TRIM58-KO iPSC lines by CRISPR/Cas9 Technology**

We established three TRIM58-KO iPSC lines by the CRISPR/Cas9 system using three different guide RNAs. The three guide RNAs were designed just downstream of the transcription start site of the TRIM58 gene in the first exon according to the website CRISPR DISIGN (<http://crispr.mit.edu>). The guide RNA sequences were as follows:

CRISPR #1: ATCCTCGCGCAGCCGCTCCC, CRISPR #2: GCATGGCCTGGGCGCCGCCC, CRISPR #3: GCGCGAGGATGCGCGGTGCC. Only in

CRISPR #2, the first base was changed from T to G. We transfected CRISPR/Cas9 vector and guide RNA-expressing vector into 610B1 (MC-Blood-iPSC line) as reported previously (Li et al., 2015) and picked up single cell-derived iPSC clones. Knockout of the TRIM58 gene was confirmed by Sanger sequencing and Western blotting. The knockout efficiency of each guide RNA was 1 in 6 clones (16.7%).

### **Western Blotting**

Erythroblasts differentiated from 610B1 iPSC line and 3 TRIM58-KO lines were lysed by Mammalian Protein Extraction Reagent (M-PER, Life Technologies). 40 µg of protein in each lane were separated by electrophoresis using NuPAGE 3-8% Tris-Acetate Gel (Life Technologies) and transferred to nitrocellulose membranes using iBlot (Life Technologies) for 7 minutes. After blocking with TBS-T supplemented with 1% skim milk, the membranes were incubated with primary antibodies: TRIM58 goat polyclonal antibody against C-terminus of TRIM58 (1:100, sc-165772, Santa Cruz) or anti-β-actin (1:1000, A5441, Sigma) for 1 h at room temperature. After washing with TBS-T, the membranes were stained with horseradish peroxidase (HRP)-conjugated secondary antibodies: donkey anti-goat IgG-HRP (Santa Cruz Biotechnology Inc.; 1:3000) or anti-mouse IgG-HRP (1:3000, #7076, Cell Signaling) for 1 h at room temperature. Signals were visualized by

Immobilon Western chemiluminescent HRP substrate (Millipore) and LAS3000 imaging system (FUJIFILM).

### **Conventional Bisulfite Sequencing of H19-ICRs**

The bisulfite-converted DNA in H19-ICR was amplified by PCR as reported previously (Ulaner et al., 2003), and the PCR product was sub-cloned into pCR2.1-TOPO vector (Life Technologies) and sequenced by Sanger sequencing. The sequences were analyzed using QUMA software (<http://quma.cdb.riken.jp/top/index.html>). The primers used for the assay were as follows: Forward (H19-5962BT): TGTGAAGGTTGGGAGATGGGA, Reverse (H19-6413BT): CCCAAACCATAACACTAAAACCCTC, and sequencing primer (T7 primer): AATACGACTCACTATAGGG.

### **WGCNA**

We applied Weighted Gene Co-expression Network Analysis (Langfelder and Horvath, 2008) to the gene expression data of 39 iPS/ESC lines to build an unsigned weighted gene co-expression network. The parameter  $\beta$ , which controls the scale-free topology, was determined to be 12. Gene modules consisting of co-regulated genes were obtained by clustering genes using the unsigned topological overlap measure (TOM) matrix, where the hierarchical tree was cut by the dynamic tree cut algorithm (Langfelder et al., 2008) with

the deepSplit parameter 4. We obtained 50 gene modules in total. The character of each module was represented by the module eigengene, which is the first principal component of the module and could be interpreted as a weighted average of gene expression values. The module eigengenes were used to calculate the correlation of the modules with hematopoietic commitment capacity (CD43 (n)) (Figure 2B).

### **Transplantation of iPSC-derived HPCs or CB in NSG mice**

NOD/LtSz-scid IL2R $\gamma$ Null (NSG) mice (Charles River Japan) were used for the transplantation experiments. CD34<sup>+</sup>43<sup>+</sup> Blood/HDF-iPSC-derived HPCs ( $1-4 \times 10^5$  cells, n = 3 each) or CD34<sup>+</sup> CB cells purchased from Life Technologies ( $1-4 \times 10^5$  cells, n = 3) were injected intravenously into sublethally (1.5 Gy) irradiated NSG mice. Engraftment was assessed by flow cytometry using anti-human CD45 antibodies and anti-mouse CD45 antibodies 3 months after transplantation. All animal experiments were performed in accordance to institutional guidelines.

### **Methyl-seq Analysis**

One  $\mu$ g of input genomic DNA was fragmented to an average size range of 150 bp using Covaris E210 (Covaris Inc.). End-repair, 3'end-adenylation, ligation with methylated adaptor and target capture by hybridization were conducted using SureSelectXT

Methyl-Seq Regent kit (Agilent Technologies) according to the manufacturer's instructions.

Bisulfite conversion of the libraries was Performed using EZ DNA Methylation Gold kit (Zymo Research). The libraries were amplified by PCR for 8 cycles and subjected to ligation with indices for multiplex barcoding by second PCR for 6 cycles. The libraries were sequenced with the TruSeq SBS kit v3 and TruSeq PE Cluster kit v3 (Illumina) in 101-cycle Paired-End mode of HiSeq 2500 (Illumina).

### **Open Chromatin Assay**

Open chromatin regions were identified by ATAC-seq (Buenrostro et al., 2013). Briefly, 10,000 cells were lysed and subjected to nuclear extraction using cold lysis buffer (10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% IGEPAL CA-630 and 10 mM Tris-HCl, pH 7.4). Nuclei were then pelleted by centrifugation for 15 min at 500 g and 4 °C. The pellets were re-suspended in 50 µl reaction buffer containing 2.5 µl of Tn5 transposase and 25 µl of TD buffer (Nextera DNA Sample Prep kit, Illumina) and incubated at 37 °C for 30 min. The reaction was terminated by purification of the samples using PCR purification kit (Qiagen). Tagmented DNA libraries were amplified by two sequential 12-cycle PCR using indexing primers included in the Nextera Index kit and NEBnext High Fidelity PCR master mix. The Illumina libraries were sequenced with the TruSeq SBS kit v3 and TruSeq PE Cluster kit v3

in 63-cycle Paired-End mode of HiSeq 2500.

### **CTCF-ChIP-seq**

Human ESCs (khES1) were cross-linked with 1% formaldehyde. The cross-linked cells were sonicated to shear DNA. DNA-protein complexes were incubated with Protein G beads (GE Healthcare) and an anti-CTCF polyclonal antibody (ab070303, Abcam). After washing the beads, immunoprecipitants were eluted with 30  $\mu$ L Tris buffer. The sequencing library preparation including barcoding was conducted using a Mondrian SP (NuGEN Technologies) and the Ovation SP Ultralow Library system (NuGEN Technologies). The Illumina libraries were sequenced with TruSeq SBS kit v3 and TruSeq SR Cluster kit v3 in 51-cycle Single-Read mode of HiSeq 2500.

### **Data Analysis of Massively Parallel Sequencing**

Sequence reads of ATAC-seq and ChIP-seq were extracted in FASTQ format using BCL2FASTQ Conversion Software 1.8.4 in the CASAVA 1.8.2 pipeline. The sequenced short reads were mapped to the human reference genome hg19 using BWA 0.7.12 (Li and Durbin, 2009). Sequence alignment and methylation calling of Methyl-seq data were conducted using Bismark software (version 0.14.1) (Krueger and Andrews, 2011) and Bowtie software (version 2.1.0) (Langmead and Salzberg, 2012). Data were visualized by

GenomeJack 2.1 (Mitsubishi Space Software).

## References

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10, 1213-1218.

Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., and Hasegawa, M. (2009). Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proc Jpn Acad Ser B Phys Biol Sci* 85, 348-362.

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

Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* 24, 719-720.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357-359.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with

Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760.

Li, H.L., Fujimoto, N., Sasakawa, N., Shirai, S., Ohkame, T., Sakuma, T., Tanaka, M., Amano, N., Watanabe, A., Sakurai, H., *et al.* (2015). Precise correction of the dystrophin gene in duchenne muscular dystrophy patient induced pluripotent stem cells by TALEN and CRISPR-Cas9. *Stem Cell Reports* 4, 143-154.

Ma, H., Morey, R., O'Neil, R.C., He, Y., Daughtry, B., Schultz, M.D., Hariharan, M., Nery, J.R., Castanon, R., Sabatini, K., *et al.* (2014). Abnormalities in human pluripotent cells due to reprogramming mechanisms. *Nature* 511, 177-183.

Morizane, A., Doi, D., Kikuchi, T., Nishimura, K., and Takahashi, J. (2011). Small-molecule inhibitors of bone morphogenic protein and activin/nodal signals promote highly efficient neural induction from human pluripotent stem cells. *J Neurosci Res* 89, 117-126.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol* 26, 101-106.

Nakamura, S., Takayama, N., Hirata, S., Seo, H., Endo, H., Ochi, K., Fujita, K., Koike, T., Harimoto, K., Dohda, T., *et al.* (2014). Expandable megakaryocyte cell lines enable

clinically applicable generation of platelets from human induced pluripotent stem cells. *Cell Stem Cell* 14, 535-548.

Nazor, K.L., Boland, M.J., Bibikova, M., Klotzle, B., Yu, M., Glenn-Pratola, V.L., Schell, J.P., Coleman, R.L., Cabral-da-Silva, M.C., Schmidt, U., *et al.* (2014). Application of a low cost array-based technique - TAB-Array - for quantifying and mapping both 5mC and 5hmC at single base resolution in human pluripotent stem cells. *Genomics* 104, 358-367.

Okita, K., Matsumura, Y., Sato, Y., Okada, A., Morizane, A., Okamoto, S., Hong, H., Nakagawa, M., Tanabe, K., Tezuka, K., *et al.* (2011). A more efficient method to generate integration-free human iPS cells. *Nat Methods* 8, 409-412.

Okita, K., Yamakawa, T., Matsumura, Y., Sato, Y., Amano, N., Watanabe, A., Goshima, N., and Yamanaka, S. (2013). An efficient nonviral method to generate integration-free human-induced pluripotent stem cells from cord blood and peripheral blood cells. *Stem Cells* 31, 458-466.

Seki, T., Yuasa, S., Oda, M., Egashira, T., Yae, K., Kusumoto, D., Nakata, H., Tohyama, S., Hashimoto, H., Kodaira, M., *et al.* (2010). Generation of induced pluripotent stem cells from human terminally differentiated circulating T cells. *Cell Stem Cell* 7, 11-14.

Ulaner, G.A., Vu, T.H., Li, T., Hu, J.F., Yao, X.M., Yang, Y., Gorlick, R., Meyers, P., Healey,

J., Ladanyi, M., *et al.* (2003). Loss of imprinting of IGF2 and H19 in osteosarcoma is accompanied by reciprocal methylation changes of a CTCF-binding site. *Hum Mol Genet* 12, 535-549.

Vizcardo, R., Masuda, K., Yamada, D., Ikawa, T., Shimizu, K., Fujii, S., Koseki, H., and Kawamoto, H. (2013). Regeneration of human tumor antigen-specific T cells from iPSCs derived from mature CD8(+) T cells. *Cell Stem Cell* 12, 31-36.