Mapping Human Pluripotent Stem Cell Derived Erythroid Differentiation by Single-Cell Transcriptome Analysis

Zijuan Xin [1,2,#,a], Wei Zhang[1,2,#,b], Shangjin Gong[1,2,c], Junwei Zhu[1,d], Yanming Li[1,e], Zhaojun Zhang[1,3,4,*,f], Xiangdong Fang [1,2,3,4,*,g]

[1] *CAS Key Laboratory of Genome Science and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China*

[2] *College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*

[3] *Sino-Danish College, University of Chinese Academy of Sciences, Beijing 100190, China*

[4] *Institute for Stem Cell and Regeneration, Chinese Academy of Sciences, Beijing 100101, China*

[#] Equal contribution.

[*] Corresponding authors.

E-mail: fangxd@big.ac.cn (Fang X), zhangzhaojun@big.ac.cn (Zhang Z)

Running title: *Xin Z and Zhang W / scRNA-seq Analysis of iPSC-derived RBC differentiation system*

[a] ORCID: 0000-0001-9418-4848.

[b] ORCID: 0000-0002-0519-3935.

[c] ORCID: 0000-0002-9811-5302.

[d] ORCID: 0000-0001-9766-3334.

[e] ORCID: 0000-0002-8213-9166.

[f] ORCID: 0000-0003-0490-6507.

[g] ORCID: 0000-0002-6628-8620.

Total word counts: 6,288

Total figures: 6

Total supplementary figures: 7

Total supplementary tables: 7

Abstract

There is currently an imbalance between the supply and demand of functional red blood cells (RBCs) in clinical applications, and this imbalance can be addressed by regenerating RBCs with a variety of in vitro methods. Induced pluripotent stem cells (iPSCs) can address the low supply of cord blood and the ethical issues in embryonic stem cell research and provide a promising strategy to eliminate immune rejection. However, no complete single-cell level differentiation pathway exists for the iPSC-derived RBC differentiation system. In this study, we used iPSC line BC1 to establish an RBC regeneration system and used the 10× Genomics single-cell transcriptome platform to map the cell lineage and differentiation trajectories on day 14 (D14) of the regeneration system. We found iPSC differentiation was not synchronized during embryoid body (EB) culture, and the D14 cells in the system mainly consisted of mesodermal and various blood cells, similar to yolk sac hematopoiesis. During asynchronous EB differentiation, iPSCs undergo three bifurcations before they enter erythroid differentiation, and the driver genes of each bifurcation were identified. The key roles of cell adhesion and estradiol in RBC regeneration were observed. This study provides systematically theoretical guidance for the optimization of the iPSC-derived RBC differentiation system.

KEYWORDS: scRNA-seq; iPSCs; Hematopoiesis; Erythropoiesis; Differentiation trajectory

Introduction

The substantial gap between the supply and demand of blood has always been a serious problem in clinical practice [1]. Artificial blood is an important means of solving this worldwide shortage, and obtaining functional red blood cells (RBCs) is the key to artificial blood generation [2]. RBC regeneration *in vitro* has been an important research direction in the blood research field for many years [3, 4]. In the current study, the primitive materials that can be used for RBC regeneration are mainly cord blood hematopoietic stem cells (HSCs) [5], embryonic stem cells (ESCs) [6, 7], and induced pluripotent stem cells (iPSCs) [6, 8]. The regeneration technology of cord blood HSCs is relatively mature, but there are problems that come with this technology, such as difficulty in obtaining materials, significant differences among individuals, high price of this technology, *etc*., and ESCs are subject to ethical restrictions [9]. Human iPSCs can be differentiated into a variety of cell types *in vitro*, providing a model for basic research and a source of clinically relevant cells [10-12]. Although the derivation of iPSC lines for patients has now become a routine technique, how to accurately differentiate iPSCs into the desired cell types is still a major challenge [13, 14].

Decades of research have shown that *in vitro* hematopoietic differentiation is closely related to in vivo development [15]. Hematopoiesis originates from the FLK1+ (KDR) lateral mesoderm, and some of the specialized cells can be induced to form vascular hemangioblasts by the transcription factor (TF) ETV2 [16, 17]. These cells have the ability to differentiate into blood and vascular precursor cells, during which ETV2 is regulated by the BMP and WNT signaling pathways [18, 19]. At this time, the primitive RBCs that express embryonic globin (ε-globin; HBE) are produced, which is the first wave of hematopoiesis in the yolk sac [20]. Hemangioblasts differentiate into endothelial cells and specialize in hematopoietic endothelial cells that produce erythroid-myeloid progenitors (EMPs). EMPs can differentiate into most of the myeloid and into definitive RBCs that can simultaneously express embryonic (ε-globin; HBE), fetal (γ-globin, HBG1 and HBG2) and adult (β-globin, HBB) globin;

this differentiation represents the second wave of hematopoiesis in the yolk sac [21]. HSCs can also be directly produced through endothelial-hematopoietic transition (EHT) induced by GATA2 and RUNX1 [22-24]. This occurs in the aorta-gonad-mesonephros (AGM) region that maintains lifelong hematopoietic differentiation and the hematopoietic cycle [25], and these cells can differentiate into definitive RBCs that express only adult globin and are regulated by signaling pathways such as VEGF and HIF [26, 27]. Hematopoietic differentiation is the differentiation process of HSCs to form various types of blood cells [28]. After differentiation and activation of HSCs, it is necessary to go through the cell fate determination program regulated by GATA1, KLF1 and other TFs to enter the erythroid differentiation process [29]. The generation of RBCs is a complex multistep process involving the differentiation of early erythroid progenitor cells into definitive RBCs, which require spatiotemporal specific completion of globin synthesis and assembly [30, 31], iron metabolism [32, 33], heme synthesis [34], cell denucleation and other processes, eventually forming intact functional RBCs [35]. However, the molecular and cellular mechanisms involved in these processes are still poorly understood.

Based on the existing molecular mechanisms of hematopoietic development and erythroid differentiation, scientists simulate embryonic hematopoiesis [36] used the spin EB method to regenerate HSPCs *in vitro*, and then induced the differentiation of RBCs from iPSCs. However, the differentiation efficiency and denucleation rate of the existing system is low, and the expression of adult globin is limited; these drawbacks limit the clinical availability [37]. At the same time, iPSC differentiation is a complex process, and flow cytometry and immunostaining have been used to determine the cell types during iPSC differentiation culture. However, these methods are limited by the number of fluorescent probes used, and it is not possible to solve important problems, such as the cell composition and differentiation path of the iPSC differentiation system, under high-resolution conditions. Recently, single-cell transcriptome technology has been able to efficiently capture and identify rare and transient cell types, determine the spatial or temporal localization of cells, and reconstitute gene regulatory networks, helping scientists understand the mechanisms by which development and cell fate are

determined [38]. In recent years, there have been many single-cell hematopoiesis studies *in vitro*. By single-cell sequencing of embryoid bodies, researchers found that naïve H9 ESCs have a stronger hematopoietic capacity than primer H9 ESCs [39] and illuminated the heterogeneity of pluripotent stem cell-derived endothelial cell differentiation [40]. Single-cell sequencing of D29 erythrocytes belonging to the iPSC-derived erythroid differentiation system revealed that cells expressing β-globin showed reduced transcripts encoding ribosomal proteins and increased expression of ubiquitin-proteasome system members [41]. The lack of a complete single-cell transcriptome map as iPSCs differentiate into RBCs does not allow scientists to systematically guide iPSCs to efficiently produce functional RBCs. Therefore, in this study, we established an iPSC-derived erythroid differentiation system and obtained a dynamic transcriptional map of cell differentiation through high-resolution single-cell transcriptomics sequencing. The cell map of the *in vitro* iPSC-derived RBC differentiation system and the *in vitro* differentiation trajectory of the intact iPSCs to RBCs were mapped for the first time. iPSCs undergo three bifurcations before they enter erythroid differentiation: the bifurcation between hematopoiesis and angiogenesis, the bifurcation between myeloid and lymphoid lineage, and the bifurcation between granulocyte-mononuclear and megakaryocyte-erythroid lineage. Cells undergo two transformations: during transformation one, cells actively express cell adhesion molecules, while during transformation two, cells downregulate these adhesion molecules. During erythroid differentiation, estradiol promotes erythroid progenitor cell proliferation and inhibits erythroid differentiation. This study has important guiding significance for analyzing the basic fate of cells and the gene network structure, as well as for designing more effective hematopoietic and erythropoiesis strategies for regenerative medicine.

Results

iPSC cell line-derived HSPC production and erythropoiesis

Depending on the spin EB culture method, we used the iPSC cell line BC1 to collect the suspension cells (SCs) through the cell strainer in the system after 14 days of culture [42, 43] (**Figure 1**A). The established culture system can effectively produce SCs. The cell count analysis showed that the number of SCs on day 14 (D14) was approximately 10-fold higher than the number of iPSCs on D0 (Figure 1B). Flow cytometry analysis showed that approximately 25.9±3.05% of the SCs expressed CD34, and approximately 27.08±1.63% of the SCs expressed CD45, indicating that we collected a certain percentage of HSPCs from the SCs (Figure 1C). We performed colony-forming unit (CFU) assays on the collected SCs on D14. After 14 days of culture, blood lineage colonies were observed under the microscope, indicating that the SCs have the potential to differentiate into various blood lineage cells; and the CFU-GEMM colony showed a slight rust red color, indicating that the erythroid cell colonies begin to express hemoglobin (Figure 1D). Subsequently, we attempted to induce HSPCs in SCs for erythroid differentiation. By assessing the changes in the expression of the cell surface markers CD71 and CD235a at different time points (Figure 1E), we found that on D14 of the EB stage, the proportion of CD71 and CD235a double-positive cells in SCs exceeded 30%, indicating that some cells in the system entered the erythroid differentiation stage. On D28, more than 95% of cells expressed both CD71 and CD235a (Figure 1E). Through Wright-Giemsa staining analysis, we observed that most of the cells were at the end of erythroid differentiation, and some cells were denucleated RBCs. In summary, we successfully obtained HSPCs by spin EB technology used the iPSC cell line BC1 and induced HSPCs to enter the erythroid differentiation process, eventually producing denucleated RBCs.

Analysis of the cell composition of the iPSC-derived RBC differentiation system

In this system, D14 is the intermediate time point. At this time, the sample contains both EBs and SCs, and the SCs contain a certain proportion of CD34+, CD45+ and CD71+ CD235a+ cells, including various cell types; in addition, the cell number is

large and the differentiation system is stable. Therefore, to establish the cell lineage and developmental trajectory of the iPSC-derived RBC differentiation system, we selected EB cells (EBCs) and SCs on D14 of culture and sequenced them using a 10× Genomics platform. After sequencing and data processing, we detected 470,711,092 reads and 4,441 cells, with an average of 105,992 reads per cell; these reads corresponded to 5,888 unique molecular identifier (UMI) counts and 1,644 genes. After quality control analysis of the data by Seurat [44], we finally obtained 3,215 cells with high-quality transcripts, including 704 SCs and 2,511 EBCs (Figure S1A). The t-SNE analysis at the sample level revealed that the cells were mainly divided into upper and lower parts. EBCs were mainly distributed in the upper part, and the corresponding highly expressed genes were mainly related to mesoderm differentiation. SCs were mainly distributed in the lower part, and the corresponding highly expressed genes mainly regulate hematopoietic differentiation and immune function (**Figure 2**A, S1B and C). We also performed bulk RNA-seq on SCs cultured at D14 and the iPSC line BC1. After comparing them with published HSPC transcriptome data from human cord blood, we found that SCs are significantly different from iPSCs, and the hematopoietic lineage-related gene expression patterns of which are similar to HSPCs and exhibit HSPC-like characteristics (Figure S1D and E). Subsequently, the intersection of the top 2,000 genes in EBCs and SCs was selected for further principal component analysis (PCA) and t-SNE analysis. The results showed that the differentiation process of cells was not synchronized. Seurat analysis divided our samples into ten clusters, including three progenitor cells clusters (cluster 9 mainly include pluripotent stem cells and high gene expression of *ZFP42*; cluster 1, 2 mainly include mesoderm cells and high gene expression of *PDGFRA*); CD33 cells cluster (cluster 0, high gene expression of *MNDA* and *S100A8*); mesenchymal cells (cluster 3, high gene expression of *LUM* and *DCN*); HSPCs (cluster 4, high gene expression of *RUNX1* and *MYB*); hemangioblasts (cluster 5, high gene expression of *KDR* and *CD34*) [45]; vascular endothelial cells (VECs) (cluster 6, high gene expression of *MALAT1* and *WSB1*); megakaryocyte and erythroid cells

(MEs) (cluster 7, high gene expression of *GATA1* and *HBA1*) [46]; and CD14 cells (cluster 8, high gene expression of *LYZ* and *CD14*) [47] (Figure 2B, C and S2C). The CD33 cells cluster had the largest number of cells, 695, which accounted for 22% of the total cells in the sample, and the progenitor cells-9 cluster had the fewest cells, only 115, accounting for 4%, indicating that the D14 cells in the system almost entered the differentiation process. Combined with the sample distribution results, SCs were mainly distributed in blood cell clusters, such as clusters 0, 4, 7, and 8, and EBCs were scattered and distributed in almost every cluster (Figure S2B); however, the MEs cluster was almost entirely made of SCs, which may be because the RBCs had low adhesion and were more easily released from the EBs (Figure 2A and B).

To further determine the function of the cell cluster, we performed functional annotations based on the top 50 genes expressed in each cluster. Gene annotation analysis revealed that progenitor cells-9 cluster genes were particularly involved in the function of embryonic morphogenesis (P = 1.7E-05), and the results also indicated that the progenitor cells-9 cluster had the differentiation potential to form each germ layer cell, such as kidney, podosoma, epithelial cells and sensorium. The progenitor cells-1 cluster was characterized by gastrula cells, and highly expressed genes were enriched in gastrulation (P = 6.01E-03). Gene annotation analysis results of the progenitor cells-1 cluster showed that these cells respond to growth factor stimulation and highly expressed the genes related to organ growth and tissue morphogenesis. Genes that were highly expressed in the progenitor cells-2 cluster were enriched in mesoderm formation (P = 1.55E-05), which was related to cardiac development and muscle formation, indicating that the progenitor cell-2 cluster was lateral mesoderm-like mesoderm. Genes that were highly expressed in the mesenchymal cells cluster were enriched in the connective tissue development pathway (P = 1.40E-06). Hemangioblasts cluster had the characteristics of hematopoietic endothelium (HE), and highly expressed genes were enriched in the pathways of endothelial development (P = 1.02E-08) and endothelial cell migration (P = 2.04E-04), suggesting that the hemangioblasts cluster simultaneously undergoes EHT and highly expresses cell adhesion-related genes. Hemangioblasts further

differentiated into VECs and HSPCs. Genes that were highly expressed in VECs were mainly enriched in the function of blood vessel development (*P* = 3.94E-03). Genes that were highly expressed in HSPCs were enriched in the pathways of lymphocyte differentiation (*P* = 3.94E-03), erythrocyte differentiation (*P* =9.58E-03), and myeloid cell differentiation (*P* = 1.49E-03), suggesting that HSPCs have the potential to differentiate into myeloid, lymphoid and erythroid cells. Genes that were highly expressed in the CD33 cells, MEs and CD14 cells clusters were obviously enriched in the pathways of granulocyte activation (*P* = 5.32E-35), oxygen transport (*P* = 2.18E-08), and phagocytosis (*P* = 9.33E-05), respectively, and these clusters all differentiated from HSPCs (Figure S3).

A higher resolution clustering of the scRNA-seq data

To further determine the process of cell differentiation, we performed a higher resolution clustering of the scRNA-seq data, and the results showed that clusters 0, 1, 4, and 8 had subclusters. Among them, the CD33 cells cluster had three subclusters, which were recorded as 0a, 0b, and 0c. Compared with clusters 0b and 0c, cluster 0a had increased *CD14* and *CTSS* gene expression, indicating that cluster 0a was a monocyte-like cells cluster [47]. Compared with clusters 0a and 0c, cluster 0b had increased *ELANE* and *MPO* gene expression, suggesting that cluster 0b was a granulocyte-macrophage progenitor (GMP)-like cells cluster. Compared with clusters 0a and 0b, cluster 0c had increased *NFC1* and *DEFA3* gene expression and was a granulocyte-like cells cluster (**Figure 3**A and B). Cluster 1 had two subclusters that were marked as clusters 1a and 1b. Cluster 1a had high expression levels of the *PITX2* and *PRRX1* genes, indicating that this cluster contained mesodermal cells; compared with cluster 1a, cluster 1b showed more stem cell characteristics with high expression of the *PDGFRB* and *ID3* genes [48, 49] (Figure 3A and C). The HSPCs cluster was divided into 2 subclusters that were marked as clusters 4a and 4b. Cluster 4a was characterized as a common myeloid progenitor (CMP)-like cluster and had the potential for myeloid cell differentiation and had high *PRG3* and *PRG2* gene expression; cluster 4b was characterized as a myeloid-lymphoid progenitor (MLP)-like cluster with high *EGR1* and *TRDC* gene expression [50] and was

enriched in the pathways related to lymphocyte differentiation [50]. In addition, the CD14 cells cluster was also divided into two clusters, namely, the macrophages cluster (cluster 8a, high gene expression of *CD36* and *CD163*) and the dendritic cells cluster (cluster 8b, high gene expression of *S100A8*) [51] (Figure S4). These results indicated that the cells in our system originated from cluster 9, differentiated into the KDR+ lateral mesoderm and formed endothelial progenitor cells (EPCs) under BPM4 factor induction; then by VEGF induction, endothelial cells undergo EHT, allowing the cells to differentiate into HSPCs. Then, HSPCs differentiate into various blood cells, mainly myeloid cells, similar to yolk sac hematopoiesis [45, 52, 53] (Figure S5A).

Dynamic regulation of cell adhesion molecules in hematopoietic development

To explain how endothelial characteristics lost during the formation of HSPCs, we performed differentially expressed gene (DEG) analysis between HSPCs and hemangioblasts and found that with the formation of HSPCs, the expression levels of TFs, such as GATA2 and TAL1, were significantly increased, and the expression levels of endothelial characteristic genes, such as *TIE1* and *KDR*, were significantly decreased (Figure 3D). Then, 466 genes that were significantly higher in HSPCs than in hemangioblasts were intersected with the 1,639 human TFs retrieved from the TF database [54], and 24 TFs were obtained. An interactive network analysis of these 24 TFs revealed that 4 of them were involved in erythroid differentiation, 4 TFs were involved in lymphoid differentiation, 8 TFs were involved in myeloid differentiation, and the others were potentially involved in hematopoietic differentiation (Figure 3E). At the same time, compared with HSPCs, the functional annotations of the top 50 genes that were highly expressed in hemangioblasts showed significant enrichment in hematopoietic differentiation pathways, such as the myeloid and lymphoid differentiation pathways. The top 50 highly expressed genes in the hemangioblasts were significantly enriched in the pathways of endothelial development and PI3K-AKT signaling, and a number of genes were also found to be enriched in pathways associated with cell adhesion; these genes were most highly expressed in hemangioblasts (Figure 3F and S5C). Making use of these genes for interaction

network analysis, the inner circle on the interaction network map showed genes enriched in the pathway, and the outer circle was genes that interacted with the inner circle genes. We can see that the outer circle genes were characteristic of HE genes, such as *KDR*, *NOTCH4*, and *ERG*, indicating that the outer circle genes were necessary for the formation of hemangioblasts but were significantly lost during hematopoietic differentiation [55, 56]. In addition, we found from the bulk RNA-seq data that relative to iPSCs, SCs also downregulated the expression of cell adhesion-related genes (Figure S5E), which may be the direct cause of SCs separation from EBs.

Cell differentiation trajectory of the iPSC-derived RBC differentiation system

To further explain the differentiation trajectory of cells and the key factors affecting cell differentiation, we performed a pseudotime analysis of scRNA-seq data using R package Monocle [57]. The results also suggested that cell differentiation originated from the progenitor cell clusters from hemangioblasts then experienced the first bifurcation hemangioblasts. At the first bifurcation, cells differentiated into VECs and HSPCs; then, at the second bifurcation, HSPCs differentiated into lymphoid cells (Lyms) (small part cells) and CMPs (most part cells), and at the third bifurcation, CMPs differentiated into MEs and granulocytes and monocytes (GMs) (**Figure 4**A and B). Pseudotime heatmap analysis with 743 genes showed that all cells were divided into three clusters, marked as early mesoderm cells, EPCs, and blood cells. We compared the genes corresponding to the three clusters with the human TF database. The *TBX2* and *PITX1* genes corresponding to the early mesodermal cells cluster were involved in the differentiation of mesoderm, which guides mesodermal organ formation and somite formation (Figure 4C and D). The *ETV2* and *ERG* genes corresponding to the EPCs cluster could guide the cells to enter the hematopoietic lineage, and the cluster also significantly expressed endothelial characteristic genes, such as *TIE1* and *KDR* (Figure 4C and E). The TFs, including GATA2, corresponding to blood cells cluster marked the initiation of hematopoiesis and then formed various blood lineage cells under the guidance of EGR1, SPI1, GATA1, *etc.*, and the cluster also significantly expressed *F2R*, *HBA1* (the characteristic genes of

megakaryocyte-erythroid cells) and *LYZ*, *MNDA*, and *ELANE* (the characteristic genes of granulocyte-macrophage cells) (Figure 4C and F). In addition, pseudotime analysis showed that the expression of the MEP characteristic genes were close to the expression time of the HE characteristic gene *CD34*, and this phenomenon was similar to the gene expression in early yolk sac hematopoiesis [58, 59] (Figure 4E and F).

Functional annotation of the genes corresponding to each cluster indicated that the early mesoderm cells cluster had the potential to differentiate into various mesoderm organs, and the corresponding pathways were embryonic morphogenesis (P = 6.31E-34) and heart development (P = 2.14E-28). The EPCs cluster could further differentiate into blood vessel and blood cells, and the corresponding genes were significantly enriched in blood vessel development (P = 1.35E-46) and regulation of the MAPK cascade (P = 6.92E-21). Many genes were also significantly enriched in the pathways related to cell adhesion, such as KEGG pathway focal adhesion (P = 3.31E-22). And the blood cells underwent hematopoietic differentiation, the corresponding genes were enriched in myeloid leukocyte activation (P = 1.20E-97), hemostasis (P = 2.69E-28), cytokine production (P = 2.04E-24) and other pathways (Figure S6). The above results suggest that the system underwent two key transformations during the asynchronous differentiation process: (1) cells exited pluripotent stem cells during germ layer differentiation, entered the lateral mesoderm and were driven by the *ETV2* gene to differentiate into hematopoietic lineage [45]; (2) HE cells entered hematopoietic differentiation under the guidance of GATA2 to form various blood cells [60].

Fate-determining TFs drive cell bifurcation

To identify the key factors that drove cells into each bifurcation, we used a branched heatmap to show the gene expression dynamics of these bifurcations. At the first bifurcation, 606 genes divided cells into four clusters: mesodermal cells cluster, blood vessel cluster, HSPCs cluster and blood lineage cells cluster. Driven by the SOX18 TF, the blood vessel cluster entered the vascular branch, and the direction of cell differentiation was controlled by the SOX7 transcription suppressor factor [61]. The

HSPCs cluster driven by GATA2 and RUNX1 was further differentiated into a blood lineage cells cluster under the induction of lineage-specific TFs, such as SPI1 and SOX4 (**Figure 5**A). Functional annotation of genes in each cluster showed that the genes in mesodermal cells were riched in skeletal system development (*P* = 5.01E-21) and ossification (*P* = 1.35E-13). The blood vessel cluster genes were enriched in angiogenesis (*P* = 6.17E-09), and the HSPCs cluster genes were enriched in ribosome assembly (*P* = 6.92E-26) and *RUNX1* regulates transcription of genes involved in differentiation of HSCs (*P* = 3.55E-05). The blood lineage cells cluster genes were enriched in lymphocyte activation (*P* = 2.34E-13) and myeloid cell differentiation (*P* = 5.62E-09).

To identify the driving factors for the differentiation of lymphoid and myeloid, we constructed a differentiation trajectory for the HSPCs cluster and used 336 genes to generate a pseudotime heatmap to show the gene expression dynamics of these branches. Cells in the HSPCs cluster were divided into four clusters: multipotent progenitors (MPPs), common lymphoid progenitors (CLPs), GMPs, and megakaryocyte-erythroid progenitors (MEPs) (Figure 5B and C). In MPPs, the genes *GATA2*, *RUNX1* and *CD34* were highly expressed, and functional annotation revealed that the genes were significantly enriched in ribosome biogenesis (*P* = 1.70E-14) and hematopoiesis (*P* = 9.55E-04). The CLPs cluster that driven by TFs such as IKZF1 and EGR1 was highly expressed *JAML*, *AIF1* and other lymphoid characteristic genes, and functional annotation revealed a significant enrichment for lymphocyte activation (*P* = 1.95E-55) and cytokine production (*P* = 8.51E-32). GMPs and MEPs are both myeloid cells, under SPI1 and GATA1 induction, myeloid-specific genes such as *LYZ*, *CLC*, *CD36* and *KLF1* were highly expressed. Functional annotation revealed that the enriched genes were involved in myeloid leukocyte activation (*P* = 1.58E-48) and hemostasis (*P* = 5.13E-25) (Figure 5C, D and S6).

To identify the driving factors for the differentiation of GMPs and MEPs, we constructed a differentiation trajectory for the CD33 cells cluster, MEs cluster and CD14 cells cluster and used 633 genes to perform a pseudotime heatmap to show the gene expression dynamics of these cells. Cells were divided into three clusters: MEs,

CMPs and GMs. MEs driven by GATA1 were highly expressed in the RBC-specific gene *HBA1* and in the platelet-specific gene *GP9* [29]. Functional annotation of this cluster was significantly enriched in platelet activation (P = 8.51E-24) and regulation of ion transport (P = 1.95E-10). CMPs highly expressed *SPI1* and *MYB*, and functionally annotated genes from this cluster were significantly enriched in response to interleukin-1 (P = 2.14E-04) and the CMYB pathway (P = 3.80E-04). GMs driven by MAFB had high expression of the granulocyte-specific gene *MNDA* and monocyte-specific cell marker *CD14* (Figure 5E-G). Functional annotation of this cluster was significantly enriched in myeloid leukocyte activation (P = 1.23E-30) and regulation of the inflammatory response (P = 6.76E-14) (Figure S6).

Pseudotime trajectory of erythroid cells in the iPSC-derived RBC differentiation system

To accurately elucidate the molecular activities in erythroid differentiation, we performed pseudotime analysis of MEs cluster alone. Simultaneously, pseudotime heatmap analysis with 200 genes showed that cells could be divided into two subclusters, marked as MEPs and erythroid cells (Erys). The TF GATA1 was highly expressed in the MEPs cluster and Erys cluster highly expressed of the erythrocyte characteristic TF KLF1, while high expression of *GFL1B*, *NFE2*, *MAX* and other genes played an important role in the differentiation and maturation of RBCs (**Figure 6**A, B). At the same time, the pseudotime trajectory analysis of the characterizing genes showed that in the process of erythroid differentiation, the expression levels of *CD34* and *TIE1* (endothelial marker genes) were downregulated and decreased rapidly during the differentiation process. *GATA1* was closely related to the early differentiation of erythroid cells and was rapidly upregulated during the differentiation process but was rapidly downregulated at the end of differentiation; consistent with previous studies, *TAL1* and *GATA1* also showed similar trends [62]. *KLF1* and *ALAS2* were also key factors during erythroid differentiation, and the expression level gradually increased with the differentiation process; *HBA1* and *HBB* encoded adult-type globin, *HBA1* was highly expressed throughout the erythroid differentiation process, and the expression levels of both *HBA1* and *HBB* gradually

increased. *FTH1* and *FTL* encoded ferritin, which were highly expressed throughout the differentiation process and were gradually upregulated throughout the erythroid differentiation process, which may be related to the synthesis of heme (Figure 6C). Functional annotation results revelated that genes corresponding to MEPs were enriched for the functions of embryonic hemopoiesis (P = 2.95E-05) and response to estradiol (P = 7.72E-04). Functional annotation results of genes corresponding to Erys showed significant enrichment in oxygen transport (P = 5.84E-13), erythrocyte differentiation (P = 1.73E-07) and other RBC characteristic pathways. The target gene of estradiol is highly expressed at the beginning of the erythroid differentiation process, and then the expression level is reduced. The results indicate that estradiol may promote the proliferation of erythroid progenitor cells with erythroid differentiation but simultaneously inhibits erythroid differentiation (Figure 6E). The interaction network analysis of the corresponding genes in the O_2/CO_2 exchange in erythrocytes, heme metabolic process and iron ion homeostasis pathways showed that *ALAS2*, the hub gene in the interaction network, played a key role in the process of erythroid differentiation and maturation [63] (Figure 6F). In addition, we mapped the TFs of these 200 genes and the genes enriched in the pathways involved in erythroid differentiation (Figure S7B and C). In addition, the β-chain globin expression profile and the qPCR results corresponding to the erythroid cells in the data showed that the erythroid cells produced by our iPSC system are mainly derived from yolk sac hematopoiesis, and interestingly, during the hemangioblast formation period of D4 EBs [64], RBCs expressed *HBG*, which indicates that RBC production occurs before HSPCs are formed during embryo culture (Figure S7D).

Our results showed that iPSCs entered erythroid differentiation, undergo two transformations, exit pluripotency, and high expression of *ETV2* and *KDR* under BMP4 and VEGF stimulation to induce cells to differentiate into hemangioblasts; this process is accompanied by cell adhesion activation [45, 55]. In addition, HSPCs enter the erythroid differentiation process accompanied by the loss of cell adhesion and the fluctuation of estradiol target genes expression. During this process, the HSPCs would be induced by IKZF1 to undergo lymphoid differentiation, and induced by SPI1 to

undergo myeloid differentiation; subsequently CMPs would be induced by MAFB to enter granulocyte-mononuclear differentiation, and induced by GATA1 to enter megakaryocyte-erythroid differentiation; lastly, MEPs would be induced by KLF1 to enter erythroid differentiation (Figure 6G). We confirmed the expression trend of cell adhesion-related genes (*KDR*, *FN1*, *CD34*, *CDH5*, *CD99* and *ESAM*) in EB and SC differentiation (Figure S7E, left) and estradiol target genes (*ALAS2*, *GATA1* and *KLF1*) in erythroid differentiation (Figure S7E, right) by qPCR analysis. The experimental results were consistent with the data analysis.

Discussion

With the rapid development of single-cell technology, single-cell transcriptome studies on the development of the hematopoietic lineage have been extensively reported, and the known hematopoietic theory has been revised and supplemented, but most studies were focused on bone marrow, peripheral blood, cord blood and other systems [65, 66]. In this study, we performed a single-cell transcriptomics study on the RBC regeneration system derived from iPSCs. The cell map and cell differentiation trajectory of the system were mapped for the first time, and the whole transcriptome dynamics from iPSCs to the erythroid lineage were delineated.

Through the analysis of the cell map, we found that in the process of hematopoiesis, there are mainly mesoderm and various blood cells, which means that cell differentiation is not synchronized. The process of RBC regeneration by iPSCs through the embryoid stage is similar to yolk sac hematopoiesis, and these results are consistent with those of previous research [20, 21]. The D14 erythrocytes are composed of the primitive RBCs produced by the first wave of hematopoiesis and the definitive RBCs produced by the second wave of yolk sac hematopoiesis. Similar to the recent single-cell sequencing results of D29 iPS-derived erythroid differentiation, it was shown that the globin expression profile of iPSC-derived definitive erythrocytes corresponds to the globin expression profile of yolk sac erythroid progenitor cells [41]. Both results suggest that iPSC-derived erythropoiesis is

accompanied by globin gene expression switching, and this system can be used as a favorable model for studying the globin conversion mechanism [30].

According to the results of cell mapping analysis, cell differentiation originated from iPSCs. Under the action of the cytokine BMP4, iPSCs form KDR+ lateral mesoderm by means of germ layer differentiation and produce EPCs. Subsequently, under the action of VEGF, the EPCs are transdifferentiated to form HE cells, and HSPCs are released. Then, HSPCs are further induced to differentiate into blood lineage cells. Our system can form various blood cells by hierarchical induction, especially RBCs, and this type of blood cell formation is very close to the embryonic hematopoiesis model *in vivo*, indicating that our system is a good model for simulating *in vivo* hematopoietic differentiation and development [67].

By studying the differentiation trajectory of the iPSC-derived RBC differentiation system, we found that the iPSCs differentiated enter hematopoietic lineage under the driving force of *ETV2* and underwent three bifurcations, then enter erythroid differentiation process. During the process of the bifurcation of hematopoiesis and angioblasts, the cells enter the hematopoietic lineage driven by TFs, such as GATA2 and RUNX1; in the process of bifurcation of the myeloid and lymphoid lines, the cells enter the myeloid lineage driven by TF SPI1; and during the process the bifurcation of the GMPs and MEPs, the cells enter the megakaryocyte-erythroid lineage driven by GATA1, and then erythrocytes are formed under the driving force of KLF1. These series of fate-determining TFs can effectivly guide the regulation of iPSC-derived RBC differentiation process and provide estimable advice for cell regeneration in other blood lineages. In addition, pseudotime clustering analysis also suggested that the pluripotent stem cells to erythroid cell production underwent two major transformations, including *ETV2*-directed withdrawal of iPSC pluripotency into the hematopoietic lineage, followed by GATA1-directed HSPC erythroid differentiation. We noted that cell adhesion molecules are actively expressed during the transformation one process but are downregulated during transformation two, possibly related to cell adhesion for the formation of the gastrula [55]. We found that during iPSC-derived erythroid differentiation, estradiol promoted the proliferation of

erythroid progenitor cells and inhibited erythroid differentiation. We were advised that we could add a cell adhesion activator to stimulate the production of hemangioblasts during the first four days of culturing EBs and then could add estradiol to the medium 6 days before inducing SC differentiation to increase the efficiency of RBC regeneration. In addition, a number of TFs that play a key role in the differentiation and maturation of RBCs, such as *NEF2* and *MAX*, have also been discovered. This information provides a powerful theoretical guide for optimizing our iPSC-derived RBC regeneration system. With the continuous development of single-cell technology, we will be able to excavate more potential mechanisms regulating RBC development and maturation and will be able to achieve the development and differentiation of RBCs in our system, thereby obtaining more mature regenerated functional RBCs *in vitro* for use in clinical applications.

Materials and methods

iPSC and EB culture and erythropoiesis

iPSCs, which the original source was Linzhao Cheng's lab, were cultured in vitronectin (Life Technologies)-coated culture dishes with Essential 8 medium (Gibco). When the cells reached 70% to 80% confluence (usually around 3~4 days), they were digested to suspension cells using 0.5 mM EDTA (Invitrogen) and passaged according to a ratio of 1 to 4.

EB culture was performed during D0 to D11 as previously described [68]. From D11 to D14 of EB culture, TPO was replaced by 3 U/ml erythropoietin (EPO) (PeproTech). On D14, SCs were collected by a 70 μm cell strainer. Subsequently, the SC solution was centrifuged at 300 g for 5 minutes to remove the supernatant, and serum-free medium (SFM) containing 50 μg/ml stem cell factor (SCF) (PeproTech), 100 μg/ml IL-3 (PeproTech), 2 U/ml EPO and 1% penicillin/streptomycin (P/S) (Life Technologies) was used to resuspend the SC precipitates. The cell concentration was adjusted to 5 x 10^5/ml and cultured in a 6-well plate at a total volume of 2 ml per well. All cells were cultured for 6 days at 37 °C under 5% CO_2, and the medium was

changed every three days. On D20, the medium was replaced with SFM containing 2 U/ml EPO and 1% P/S, and the cell concentration was adjusted appropriately. The number of cells in each well did not exceed 2 x 10^6, and the total volume of the medium per well was 2 ml. The cells were cultured at 37 °C under 5% CO_2, and the medium was changed every three days.

Colony formation unit assay

Incubation of D14 SCs used MethoCult™ H4434 Classic Methylcellulose Medium for Human Cells (StemCell). Each well of a 12-well plate contained 8,000-10,000 cells and 0.6 ml of medium. Other operations were performed as described in the instructions. Cells were incubated for 12-14 days, and the clones were observed under the microscope.

Flow cytometry analysis

First, ~1 × 10^5 cells were collected from the culture, centrifuged at 300 g for 5 minutes to remove the medium, and washed twice with 2 ml 1× Dulbecco's phosphate buffered saline (DPBS) (Gibco) solution containing 2% FBS (Gibco) and 2 mM EDTA. After centrifugation at 300 g for 5 minutes to remove the supernatant, the cells were resuspended in 50 μl DPBS buffer, and the antibodies (0.1 μl for anti-CD71 and 0.5 μl for anti-CD235a) were added, followed by incubation at 4 °C for 10 minutes in the dark. The cells were washed twice by adding 2 ml of 1× DPBS buffer after the incubation and resuspended in 200 μl of DPBS buffer to prepare a cell suspension for the assay. The BD FACSAria II instrument was used for flow cytometry analysis, and the data analysis was performed with Flowjo software (Version 7.6, Three Star). The antibodies (Miltenyi Biotec) used in this study included anti-human CD34-FITC, anti-human CD45-PE, anti-human CD71-PE, and anti-human CD235a-APC.

Wright-Giemsa staining

~1 × 10^5 cells were taken, and the medium was removed by centrifugation at 300 g for 5 minutes; the cells were resuspended in 20 μl of DPBS buffer. The cells were placed onto one of two clean slides, and the other slide was used to push the slide at a 30 to

45-degree angle to spread the cells evenly. The pushed film was placed in a 37 °C incubator for 1 to 3 hours. Then, 200 μL of Wright-Giemsa staining A solution (BASO, BA-4017) was placed onto the cell-containing slide, the slide was placed at room temperature for 1 minute and was immediately covered with 400 μL of B solution. After the slide was incubated at room temperature for 8 minutes, the slides were gently rinsed with fluid distilled water along the area around the slide without cells. The slides were rinsed for an additional 1 minute until the water ran clear and transparent. Then, the slides were placed at room temperature until dry, and a microscope was used to observe and record cell morphology.

Total RNA extraction and qPCR

Total RNA was extracted from the iPSC line BC1, D4 EBs, D14 SCs, and D20 erythroid cells, and the total RNA was reverse transcribed into cDNA using the PrimeScript RT Reagent Kit with gDNA Eraser (TaKaRa). The diluted cDNA was used as a template for qPCR. The instrument used for qPCR was the Bio-Red CFX96 Real-Time PCR Detection System. All primers used in qPCR are listed in Supplementary Table S7.

scRNA-seq data analysis and bulk RNA-seq data analysis

D14 EBs were dissociated with Accutase solution (Sigma). Single-cell suspensions at 800 cells/μL were subjected to Chromium 10× Genomics library construction. Raw gene expression matrices generated by CellRanger (version 2.0.0) were imported into Seurat (version 2.4) [44]. After removing genes expressed in fewer than 5 cells, in cells with more than 5% mitochondria reads or in cells with less than 200 unique genes, two Seurat objects were integrated and clustered into 10 and 14 clusters based on fifteen principal components with the resolution parameter set at 0.2 and 0.8. Each cluster was identified according to the marker genes found by the Seurat “FindAllMarkers” function. Particular clusters were imported into Monocle2 [57] for pseudotime ordering. For bulk RNA-seq data, FASTQC and Trimmomatic were used to remove adaptor sequences and low-quality reads. Then, clean data were mapped to the reference genome (GRCh38) by HISAT2 [69] to generate a gene expression matrix, which was then imported into DEGseq [70] R Package to identify DEGs.

Heatmap analysis and PCA were performed by pheatmap and PRINCOMP, respectively. All data analyses were completed in the R 3.5.2 environment. We used Metascape [71] to perform functional annotation analysis. Human TF sequences were downloaded from http://humantfs.ccbr.utoronto.ca/download.php. The network was visualized by GeneMANIA [72].

Data availability

The single-cell RNA-seq of D14 EBCs and SCs, and the bulk RNA-seq of BC1 and D14 SCs is available in the Genome Sequence Archive (GSA) http://bigd.big.ac.cn/gsa/s/ZVZPnz5q

Authors' contributions

XF and ZZ designed and guided the study. WZ and JZ completed the experiments. ZX and SG completed the data analysis. YL supported the experiments. ZX and WZ drafted the manuscript, XF and ZZ edited the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Acknowledgments

This research was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA16010602), the National Key Research and Development Program of China (Grant Nos. 2016YFC0901700, 2016YFC0901603, 2017YFC0907400, 2018YFC0910700, 2017YFC0907403 and 2018YFC0910402) and the National Natural Science Foundation of China (Grant Nos. 81670109, 81870097, 81700097 and 81700116). We were grateful to Linzhao Cheng for agreeing to use iPSC line BC1 and Qianfei Wang for providing the iPSCs.

References

[1] Williamson LM, Devine DV. Challenges in the management of the blood supply. The Lancet 2013;381:1866-75.

[2] Moradi S, Jahanian-Najafabadi A, Roudkenar MH. Artificial blood substitutes: First steps on the long route to clinical utility. Clin Med Insights Blood Disord 2016;9:33-41.

[3] Batta K, Menegatti S, Garcia-Alegria E, Florkowska M, Lacaud G, Kouskoff V. Concise review: Recent advances in the in vitro derivation of blood cell populations. Stem Cells Transl Med 2016;5:1330-7.

[4] Christaki EE, Politou M, Antonelou M, Athanasopoulos A, Simantirakis E, Seghatchian J, et al. Ex vivo generation of transfusable red blood cells from various stem cell sources: A concise revisit of where we are now. Transfus Apher Sci 2019;58:108-12.

[5] Lopez-Yrigoyen M, Yang CT, Fidanza A, Cassetta L, Taylor AH, McCahill A, et al. Genetic programming of macrophages generates an in vitro model for the human erythroid island niche. Nat Commun 2019;10:881.

[6] Shen J, Zhu Y, Lyu C, Feng Z, Lyu S, Zhao Y, et al. Sequential cellular niches control the generation of enucleated erythrocytes from human pluripotent stem cells. Haematologica 2019.

[7] Choi KD, Vodyanik MA, Togarrati PP, Suknuntha K, Kumar A, Samarjeet F, et al. Identification of the hemogenic endothelial progenitor and its direct precursor in human pluripotent stem cell differentiation cultures. Cell Rep 2012;2:553-67.

[8] Liu Y, Wang Y, Gao Y, Forbes JA, Qayyum R, Becker L, et al. Efficient generation of megakaryocytes from human induced pluripotent stem cells using food and drug administration-approved pharmacological reagents. Stem Cells Transl Med 2015;4:309-19.

[9] Sun S, Peng Y, Liu J. Research advances in erythrocyte regeneration sources and methods in vitro. Cell Regen (Lond) 2018;7:45-9.

[10] Park YJ, Cha S, Park YS. Regenerative applications using tooth derived stem cells in other than tooth regeneration: A literature review. Stem Cells Int 2016;2016:9305986.

[11] Jiang L, Jones S, Jia X. Stem cell transplantation for peripheral nerve regeneration: Current options and opportunities. Int J Mol Sci 2017;18.

[12] Illich DJ, Demir N, Stojkovic M, Scheer M, Rothamel D, Neugebauer J, et al. Concise review: Induced pluripotent stem cells and lineage reprogramming: Prospects for bone regeneration. Stem Cells 2011;29:555-63.

[13] Wu SM, Hochedlinger K. Harnessing the potential of induced pluripotent stem cells for regenerative medicine. Nat Cell Biol 2011;13:497-505.

[14] Bilic J, Izpisua Belmonte JC. Concise review: Induced pluripotent stem cells versus embryonic stem cells: Close enough or yet too far apart? Stem Cells 2012;30:33-41.

[15] Kauts ML, Rodriguez-Seoane C, Kaimakis P, Mendes SC, Cortes-Lavaud X, Hill U, et al. In vitro differentiation of gata2 and ly6a reporter embryonic stem cells

corresponds to in vivo waves of hematopoietic cell generation. Stem Cell Reports 2018;10:151-65.
[16] Koyano-Nakagawa N, Kweon J, Iacovino M, Shi X, Rasmussen TL, Borges L, et al. Etv2 is expressed in the yolk sac hematopoietic and endothelial progenitors and regulates lmo2 gene expression. Stem Cells 2012;30:1611-23.
[17] Kataoka H, Hayashi M, Nakagawa R, Tanaka Y, Izumi N, Nishikawa S, et al. Etv2/er71 induces vascular mesoderm from flk1+pdgfralpha+ primitive mesoderm. Blood 2011;118:6975-86.
[18] Kirmizitas A, Meiklejohn S, Ciau-Uitz A, Stephenson R, Patient R. Dissecting bmp signaling input into the gene regulatory networks driving specification of the blood stem cell lineage. Proc Natl Acad Sci U S A 2017;114:5814-21.
[19] Hubner K, Grassme KS, Rao J, Wenke NK, Zimmer CL, Korte L, et al. Wnt signaling positively regulates endothelial cell fate specification in the fli1a-positive progenitor population via lef1. Dev Biol 2017;430:142-55.
[20] Palis J. Primitive and definitive erythropoiesis in mammals. Front Physiol 2014;5:3.
[21] McGrath KE, Frame JM, Fromm GJ, Koniski AD, Kingsley PD, Little J, et al. A transient definitive erythroid lineage with unique regulation of the beta-globin locus in the mammalian embryo. Blood 2011;117:4600-8.
[22] Lie ALM, Marinopoulou E, Li Y, Patel R, Stefanska M, Bonifer C, et al. Runx1 positively regulates a cell adhesion and migration program in murine hemogenic endothelium prior to blood emergence. Blood 2014;124:e11-20.
[23] de Pater E, Kaimakis P, Vink CS, Yokomizo T, Yamada-Inagawa T, van der Linden R, et al. Gata2 is required for hsc generation and survival. J Exp Med 2013;210:2843-50.
[24] Huang K, Du J, Ma N, Liu J, Wu P, Dong X, et al. Gata2(-/-) human escs undergo attenuated endothelial to hematopoietic transition and thereafter granulocyte commitment. Cell Regen (Lond) 2015;4:4.
[25] Ivanovs A, Rybtsov S, Welch L, Anderson RA, Turner ML, Medvinsky A. Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. J Exp Med 2011;208:2417-27.
[26] Stratman AN, Davis MJ, Davis GE. Vegf and fgf prime vascular tube morphogenesis and sprouting directed by hematopoietic stem cell cytokines. Blood 2011;117:3709-19.
[27] Guitart AV, Subramani C, Armesilla-Diaz A, Smith G, Sepulveda C, Gezer D, et al. Hif-2alpha is not essential for cell-autonomous hematopoietic stem cell maintenance. Blood 2013;122:1741-5.
[28] Upadhaya S, Sawai CM, Papalexi E, Rashidfarrokhi A, Jang G, Chattopadhyay P, et al. Kinetics of adult hematopoietic stem cell differentiation in vivo. J Exp Med 2018;215:2815-32.
[29] Palii CG, Cheng Q, Gillespie MA, Shannon P, Mazurczyk M, Napolitani G, et al. Single-cell proteomics reveal that quantitative changes in co-expressed lineage-specific transcription factors determine cell fate. Cell Stem Cell 2019;24:812-20.

[30] Qiu C, Olivier EN, Velho M, Bouhassira EE. Globin switches in yolk sac-like primitive and fetal-like definitive red blood cells produced from human embryonic stem cells. Blood 2008;111:2400-8.
[31] Yu X, Azzo A, Bilinovich SM, Li X, Dozmorov M, Kurita R, et al. Disruption of the mbd2-nurd complex but not mbd3-nurd induces high level hbf expression in human erythroid cells. Haematologica 2019.
[32] Chou AC, Broun GO, Jr., Fitch CD. Abnormalities of iron metabolism and erythropoiesis in vitamin e-deficient rabbits. Blood 1978;52:187-95.
[33] Anderson GJ, Powell LW, Halliday JW. Transferrin receptor distribution and regulation in the rat small intestine. Effect of iron stores and erythropoiesis. Gastroenterology 1990;98:576-85.
[34] Rutherford TR, Weatherall DJ. Deficient heme synthesis as the cause of noninducibility of hemoglobin synthesis in a friend erythroleukemia cell line. Cell 1979;16:415-23.
[35] Ji P, Murata-Hori M, Lodish HF. Formation of mammalian erythrocytes: Chromatin condensation and enucleation. Trends Cell Biol 2011;21:409-15.
[36] Murry CE, Keller G. Differentiation of embryonic stem cells to clinically relevant populations: Lessons from embryonic development. Cell 2008;132:661-80.
[37] Mazurier C, Douay L, Lapillonne H. Red blood cells from induced pluripotent stem cells: Hurdles and developments. Curr Opin Hematol 2011;18:249-53.
[38] Tang F, Lao K, Surani MA. Development and applications of single-cell transcriptome analysis. Nat Methods 2011;8:S6-11.
[39] Han X, Chen H, Huang D, Chen H, Fei L, Cheng C, et al. Mapping human pluripotent stem cell differentiation pathways using high throughput single-cell rna-sequencing. Genome Biol 2018;19:47.
[40] Swiers G, Baumann C, O'Rourke J, Giannoulatou E, Taylor S, Joshi A, et al. Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. Nat Commun 2013;4:2924.
[41] Vanuytsel K, Matte T, Leung A, Naing ZH, Morrison T, Chui DHK, et al. Induced pluripotent stem cell-based mapping of beta-globin expression throughout human erythropoietic development. Blood Adv 2018;2:1998-2011.
[42] Cheng L, Hansen NF, Zhao L, Du Y, Zou C, Donovan FX, et al. Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. Cell Stem Cell 2012;10:337-44.
[43] Chou BK, Mali P, Huang X, Ye Z, Dowey SN, Resar LM, et al. Efficient human ips cell derivation by a non-integrating plasmid from blood cells with unique epigenetic and gene expression signatures. Cell Res 2011;21:518-29.
[44] Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol 2015;33:495-502.
[45] Koyano-Nakagawa N, Garry DJ. Etv2 as an essential regulator of mesodermal lineage development. Cardiovasc Res 2017;113:1294-306.
[46] Tusi BK, Wolock SL, Weinreb C, Hwang Y, Hidalgo D, Zilionis R, et al. Population snapshots predict early haematopoietic and erythroid hierarchies. Nature 2018;555:54-60.

[47] Stansfield BK, Ingram DA. Clinical significance of monocyte heterogeneity. Clin Transl Med 2015;4:5.
[48] Scialdone A, Tanaka Y, Jawaid W, Moignard V, Wilson NK, Macaulay IC, et al. Resolving early mesoderm diversification through single-cell expression profiling. Nature 2016;535:289-93.
[49] Pijuan-Sala B, Griffiths JA, Guibentif C, Hiscock TW, Jawaid W, Calero-Nieto FJ, et al. A single-cell molecular map of mouse gastrulation and early organogenesis. Nature 2019;566:490-5.
[50] Min IM, Pietramaggiori G, Kim FS, Passegue E, Stevenson KE, Wagers AJ. The transcription factor egr1 controls both the proliferation and localization of hematopoietic stem cells. Cell Stem Cell 2008;2:380-91.
[51] Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell rna-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 2017;356.
[52] Nostro MC, Cheng X, Keller GM, Gadue P. Wnt, activin, and bmp signaling regulate distinct stages in the developmental pathway from embryonic stem cells to blood. Cell Stem Cell 2008;2:60-71.
[53] Metcalf D. Hematopoietic cytokines. Blood 2008;111:485-91.
[54] Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The human transcription factors. Cell 2018;172:650-65.
[55] Adams JC, Watt FM. Regulation of development and differentiation by the extracellular matrix. Development 1993;117:1183-98.
[56] Angelos MG, Abrahante JE, Blum RH, Kaufman DS. Single cell resolution of human hematoendothelial cells defines transcriptional signatures of hemogenic endothelium. Stem Cells 2018;36:206-17.
[57] Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 2014;32:381-6.
[58] Dzierzak E, Philipsen S. Erythropoiesis: Development and differentiation. Cold Spring Harb Perspect Med 2013;3:a011601.
[59] Palis J, Robertson S, Kennedy M, Wall C, Keller G. Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. Development 1999;126:5073-84.
[60] Gao X, Johnson KD, Chang YI, Boyer ME, Dewey CN, Zhang J, et al. Gata2 cis-element is required for hematopoietic stem cell generation in the mammalian embryo. J Exp Med 2013;210:2833-42.
[61] Downes M, Koopman P. Sox18 and the transcriptional regulation of blood vessel development. Trends Cardiovasc Med 2001;11:318-24.
[62] Rylski M, Welch JJ, Chen YY, Letting DL, Diehl JA, Chodosh LA, et al. Gata-1-mediated proliferation arrest during erythroid maturation. Mol Cell Biol 2003;23:5031-42.
[63] To-Figueras J, Ducamp S, Clayton J, Badenas C, Delaby C, Ged C, et al. Alas2 acts as a modifier gene in patients with congenital erythropoietic porphyria. Blood 2011;118:1443-51.

[64] Fehling HJ, Lacaud G, Kubo A, Kennedy M, Robertson S, Keller G, et al. Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. Development 2003;130:4217-27.

[65] Dzierzak E, Bigas A. Blood development: Hematopoietic stem cell dependence and independence. Cell Stem Cell 2018;22:639-51.

[66] Wilson NK, Göttgens B. Single-cell sequencing in normal and malignant hematopoiesis. HemaSphere 2018;2.

[67] Yvernogeau L, Gautier R, Khoury H, Menegatti S, Schmidt M, Gilles JF, et al. An in vitro model of hemogenic endothelium commitment and hematopoietic production. Development 2016;143:1302-12.

[68] Liu Y, Wang Y, Gao Y, Forbes JA, Qayyum R, Becker L, et al. Efficient generation of megakaryocytes from human induced pluripotent stem cells using food and drug administration-approved pharmacological reagents. STEM CELLS Translational Medicine 2015;4:309-19.

[69] Kim D, Langmead B, Salzberg SL. Hisat: A fast spliced aligner with low memory requirements. Nat Methods 2015;12:357-60.

[70] Wang L, Feng Z, Wang X, Wang X, Zhang X. Degseq: An r package for identifying differentially expressed genes from rna-seq data. Bioinformatics 2010;26:136-8.

[71] Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 2019;10:1523.

[72] Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The genemania prediction server: Biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res 2010;38:W214-20.

Figure legends

Figure 1 Identification of iPSC-derived HSPCs and erythroid cells.

A. Schematic diagram of the protocol used to generate HSPCs and erythroid cells from iPSCs *in vitro*. **B.** Histogram of iPSC counts on D0 and collected SCs on D14 per round-bottom 96-well plate (error bar = mean ± sd). **C.** Percentage of CD34+ and CD45+ cells in the total SCs (mean ± sd). **D.** Representative image of CFU-GEMM grown from collected SCs on D14. **E.** Representative results of CD71+ and CD235a+ cells expressed on D14, D17, D23 and D28 during the erythroid differentiation of SCs. F. Representative image of erythroid cells on D28.

Figure 2 Cell lineage map of the iPSC-derived RBC regeneration system.

A. t-SNE plot of single-cell sample profiles. **B**. t-SNE plot of single-cell cluster distribution. **C**. Heatmap shown the expression pattern of the top 50 DEGs in each cell cluster; all genes used are listed in Supplementary Table S2. **D**. The dotplot of marker genes of each cluster.

Figure 3 Subcluster analysis and DEG analysis.

A . t-SNE plots of single-cell subcluster distribution. **B**. Heatmap shown the DEG pattern of the subcluster from the CD33 cells cluster. Top 50 DEGs of each subcluster were shown, and genes were listed in Supplementary Table S4. **C**. Heatmap shown the DEG pattern of subclusters from progenitor cells-1 cluster. Top 50 DEGs of each subcluster were shown, and genes were listed in Supplementary Table S4. **D**. Volcanic plot of DEGs from DEG analysis between HSPCs and hemangioblasts ($P < 0.05$, $-1 < \text{Log2 FC} < 1$). **E**. GO analysis of DEGs from DEG analysis between HSPCs and hemangioblasts. Each functional term was shown in Supplementary Table S3. **F**. Network shown the high expression of TFs in HSPCs.

Figure 4 Cell pseudotime trajectory of the iPSC-derived RBC regeneration system.

A. Differentiation trajectory of iPSC-derived erythrocyte regeneration system by monocle. **B**. Heatmap shown the gene expression dynamics during iPSC

differentiation. Genes were listed in Supplementary Table S3. **C-E**. Pseudotime analysis of marker genes of each cluster from early mesodermal cells cluster (**C**), HE cells cluster (**D**) and blood cells cluster (**E**).

Figure 5 Driving branching TFs of cell pseudotime trajectory.

A. Branched heatmap of differentiation branch point 1 from Figure 4a. Genes were listed in Supplementary Table S5. **B**. Differentiation trajectory of HSPCs. **C**. Heatmap shown the gene expression dynamics in HSPCs. Genes were listed in Supplementary Table S5. **D**. Pseudotime analysis of the marker gene of each cluster in C. **E**. Differentiation trajectory of cells from the CD33 cells cluster, MEs cluster and CD14 cells cluster in Figure 2B. **F**. Heatmap shown the gene expression dynamics in cells from E. Genes were listed in Supplementary Table S5. **G**. Pseudotime analysis of the marker genes of each cluster in F.

Figure 6 Differentiation trajectory from MEPs to erythroid cells.

A. Differentiation trajectory of MEs. **B**. Heatmap shown the gene expression dynamics in MEs. Genes were listed in Supplementary Table S5. **C**. Pseudotime analysis of marker genes of each cluster in B. **D**. Functional annotation analysis of MEPs cluster in B. Each functional term was shown in Supplementary Table S5. **E**. Pseudotime analysis of genes in the response to the estradiol functional term from D. **F**. Network shown the genes involved in RBC development from the Erys cluster in B.

Supplementary material

Figure S1 Quality control for scRNA-seq data and bulk RNA-seq data analysis.

A. Distribution of genes, transcripts and mitochondrial read proportion detected in each cell. **B.** Heatmap shown the expression pattern of the top 50 DEGs in each single-cell sample. Genes were listed in Supplementary Table S1. **C.** Functional annotation analysis of the top 50 DEGs in each single-cell sample. Each functional term was shown in Supplementary Table S1. **D.** Heatmap of hematopoietic lineage

related gene expression patterns for four bulk RNA-seq datasets. **E.** PCA of four bulk RNA-seq datasets.

Figure S2 Overview of the 3,215 single cells from EBCs and SCs.

A. Distribution of cells in each cluster. **B.** Distribution of EBCs and SCs in each cluster. **C.** Identification of marker genes for each cluster.

Figure S3 Functional annotation analysis of the top 50 marker genes in each cluster, each functional term was shown in Table S3.

Figure S4 Subclusters of the CD33 cells cluster, progenitor cells-1 cluster, HSPCs cluster and CD14 cells cluster.

A, **B, E, F.** Violin plots shown the expression distributions of marker genes across subclusters: CD33 cells cluster (**A**) progenitor cells-1 cluster (**B**), HSPCs cluster (**E**) and CD14 cells cluster (**F**). Heatmaps shown the DEG pattern of each subcluster from the HSPCs cluster (**C**) and CD14 cells cluster (**D**). Gene sets of each cluster were listed in Supplementary Table S4.

Figure S5 iPSC differentiation process and DEG analysis results

A. The process of iPSCs as they differentiate into erythroid cells. **B.** Venn diagram shown the highly expressed TFs in HSPCs. **C.** t-SNE plots shown the cell adhesion molecules. **D.** Network shown the high expression of TFs in HSPCs. **E.** Functional annotation analysis of genes downregulation in SCs compared with that in iPSCs. Each functional term was shown in Supplementary Table S3.

Figure S6 Functional annotation analysis for the heatmap genes in Figures 4 and 5. Each functional term was shown in Table S6.

Figure S7 Function of cell adhesion and estradiol during erythroid differentiation.

A. Functional annotation analysis of the Erys cluster in Figure 6B. Each functional term was shown in Supplementary Table S6**. B, C.** The network shown the TFs in Figure 6B (**B**) and genes of the erythrocyte differentiation functional term from A and Figure 6D (**C**). **D.** The β-chain globin expression profile. **E.** qPCR of cell adhesion molecules and the effects of estradiol throughout erythroid cell maturation.

Table S1 Heatmap genes for Figure S1B and functional term for Figure S1C

Table S2 Heatmap genes for Figure 2C

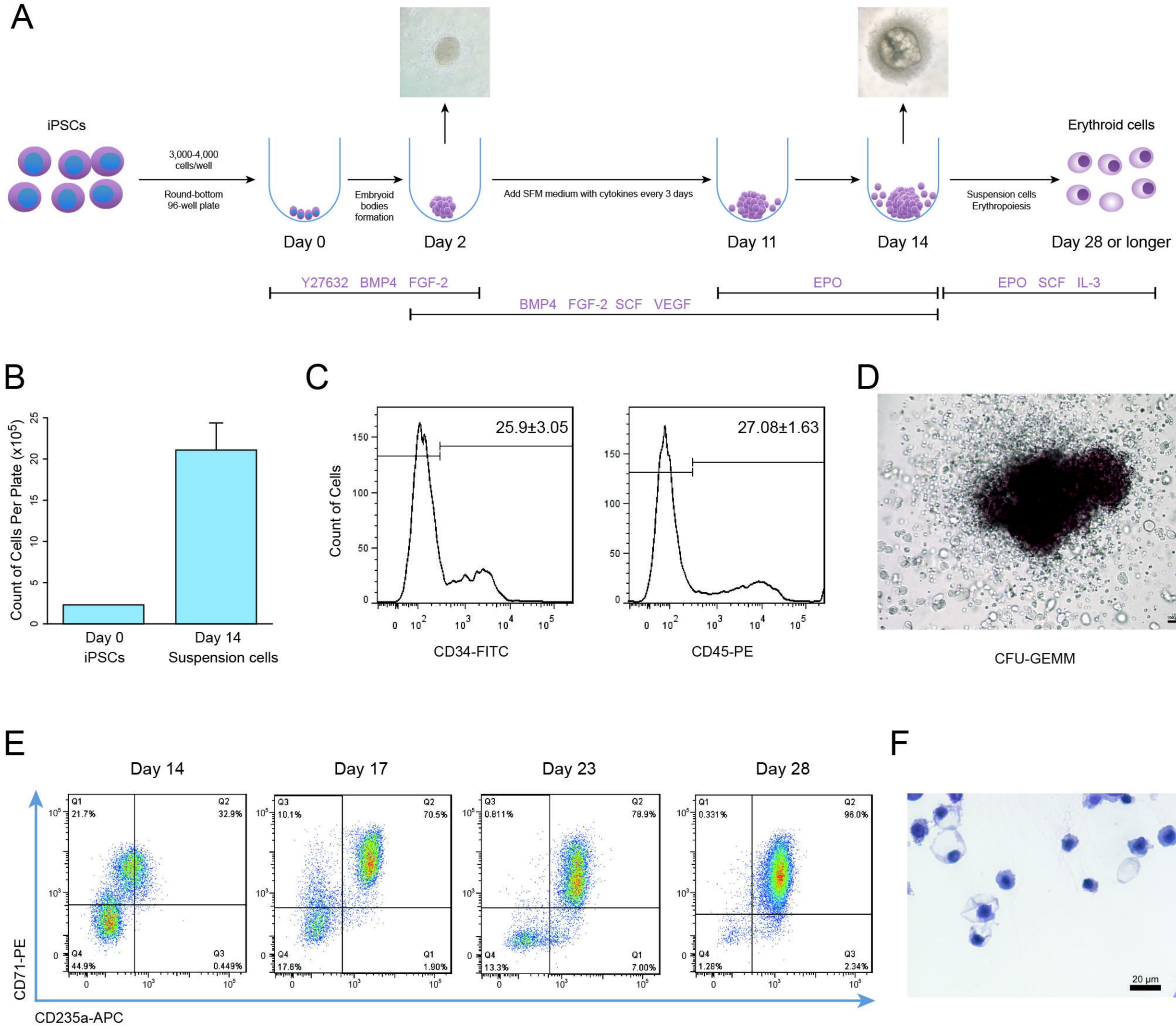
Table S3 Functional terms for Figure S3, 3F and S5E

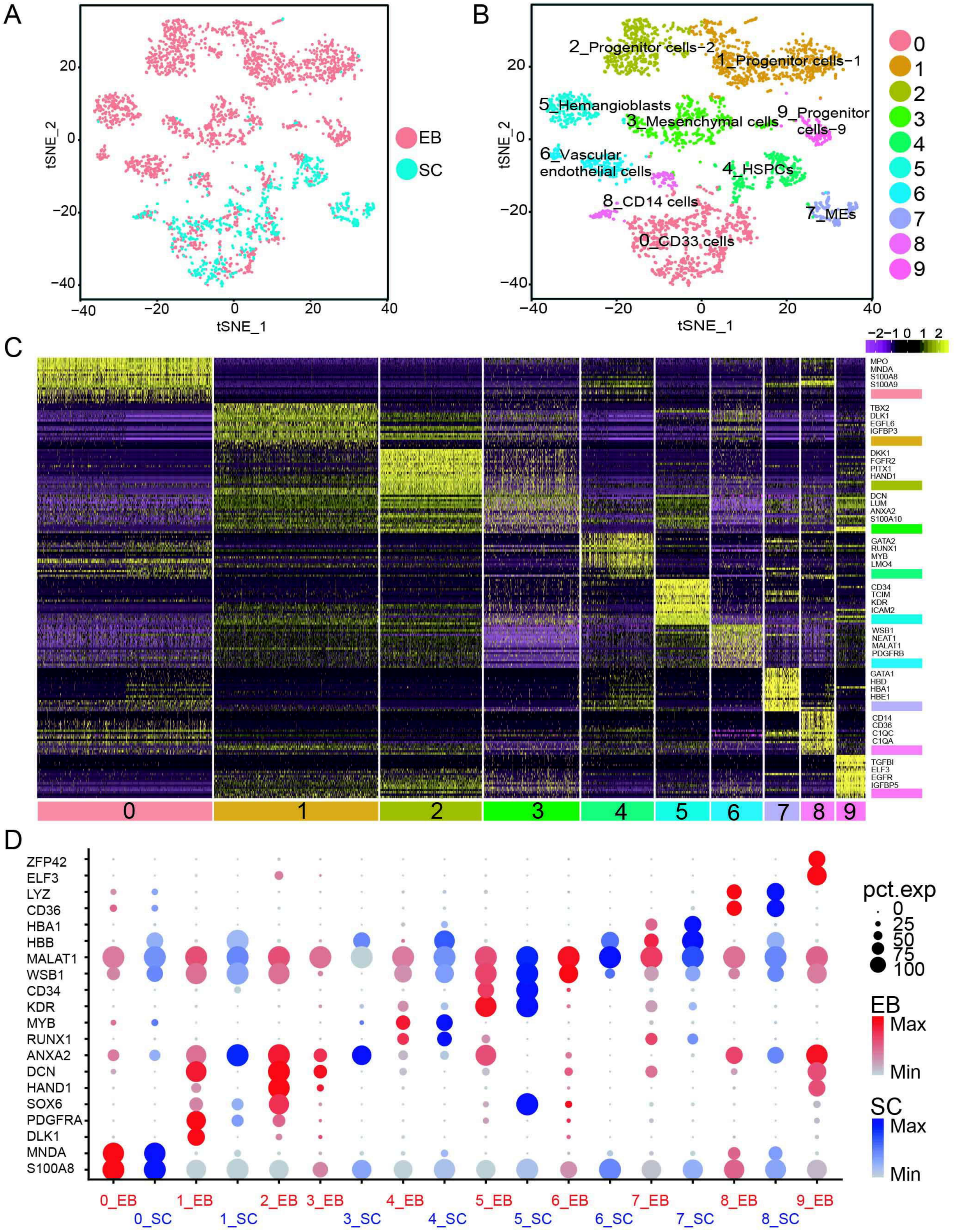
Table S4 Genes used for heatmap of subclusters 0a, 0b, 0c, 1a, 1b, 4a, 4b, 8a and 8b

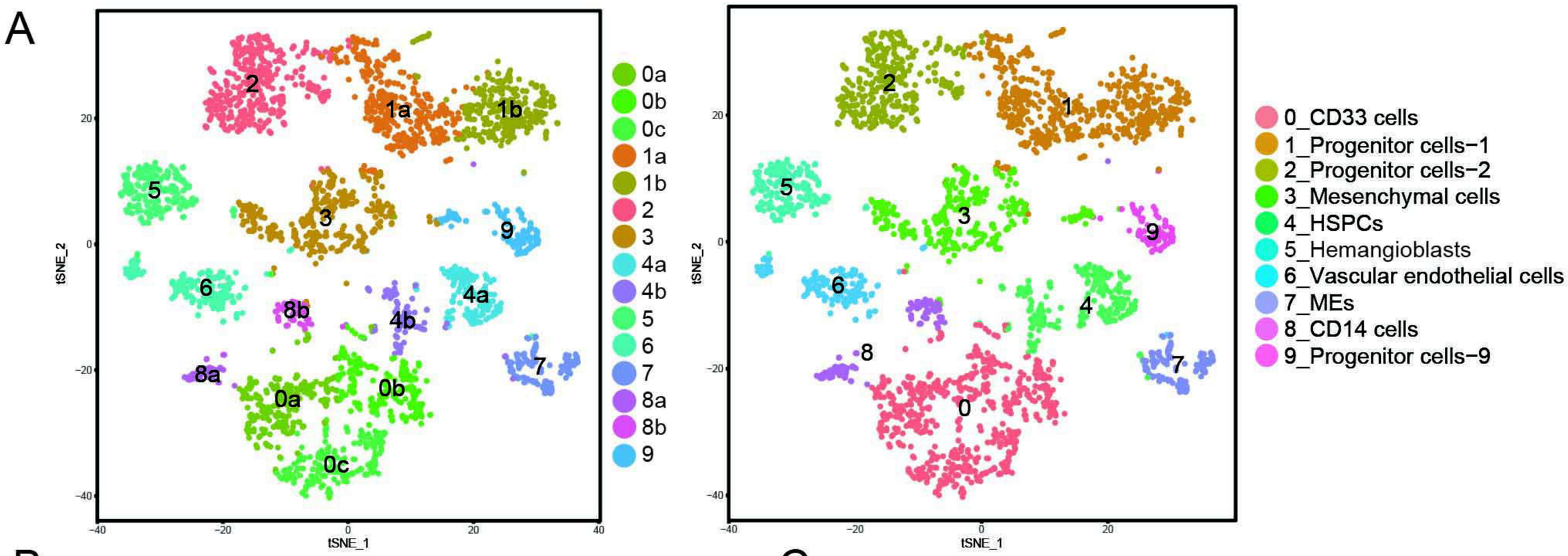
Table S5 Genes used for Figure 4C, 5A, 5C, 5F and 6B

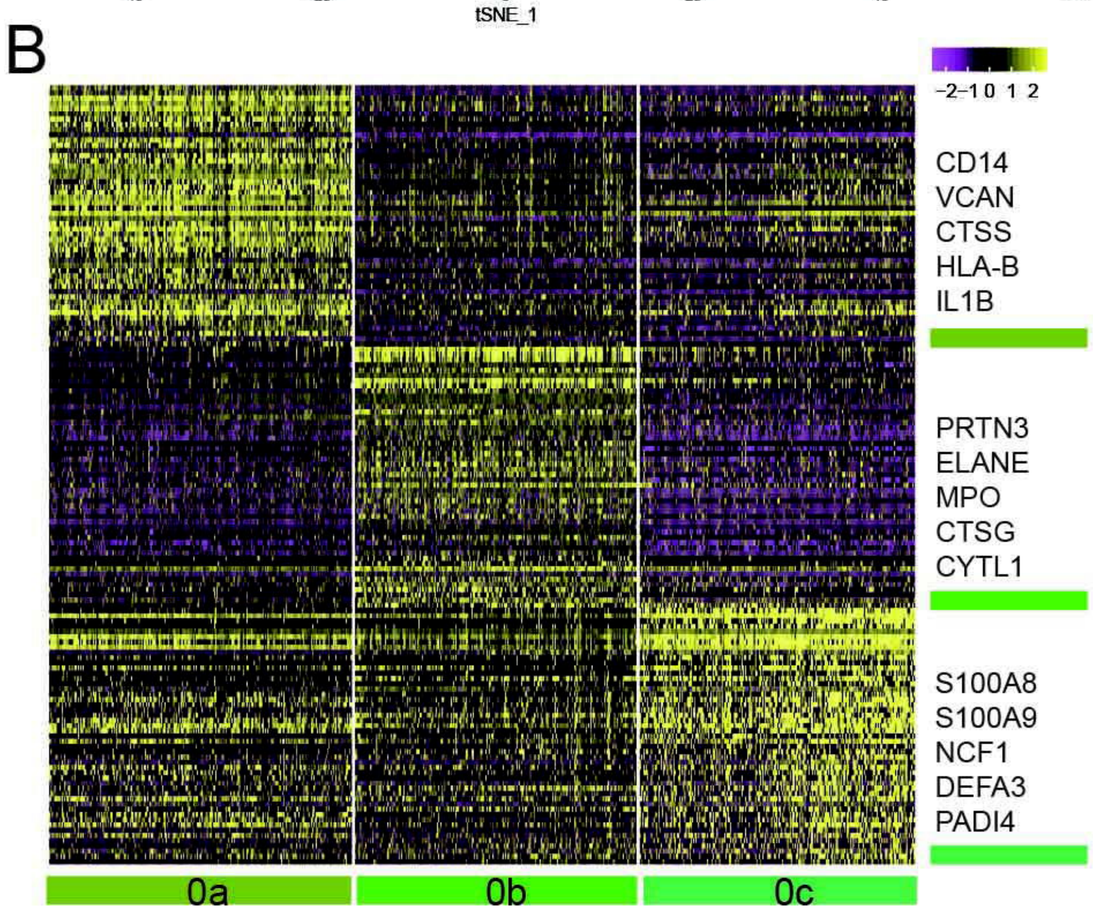
Table S6 Functional terms for Figure 5A, 5C, 5F and 6B

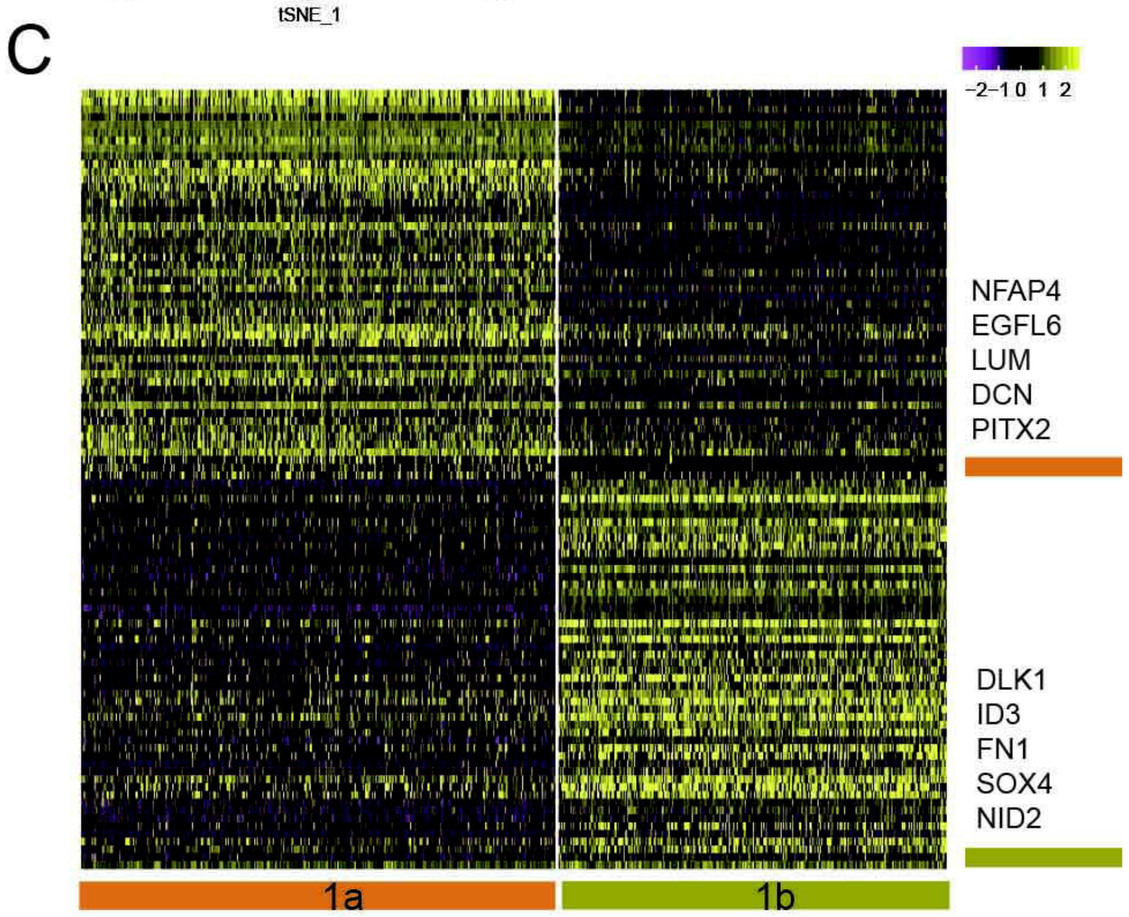
Table S7 qPCR primer for globin, cell adhesion and estradiol related genes

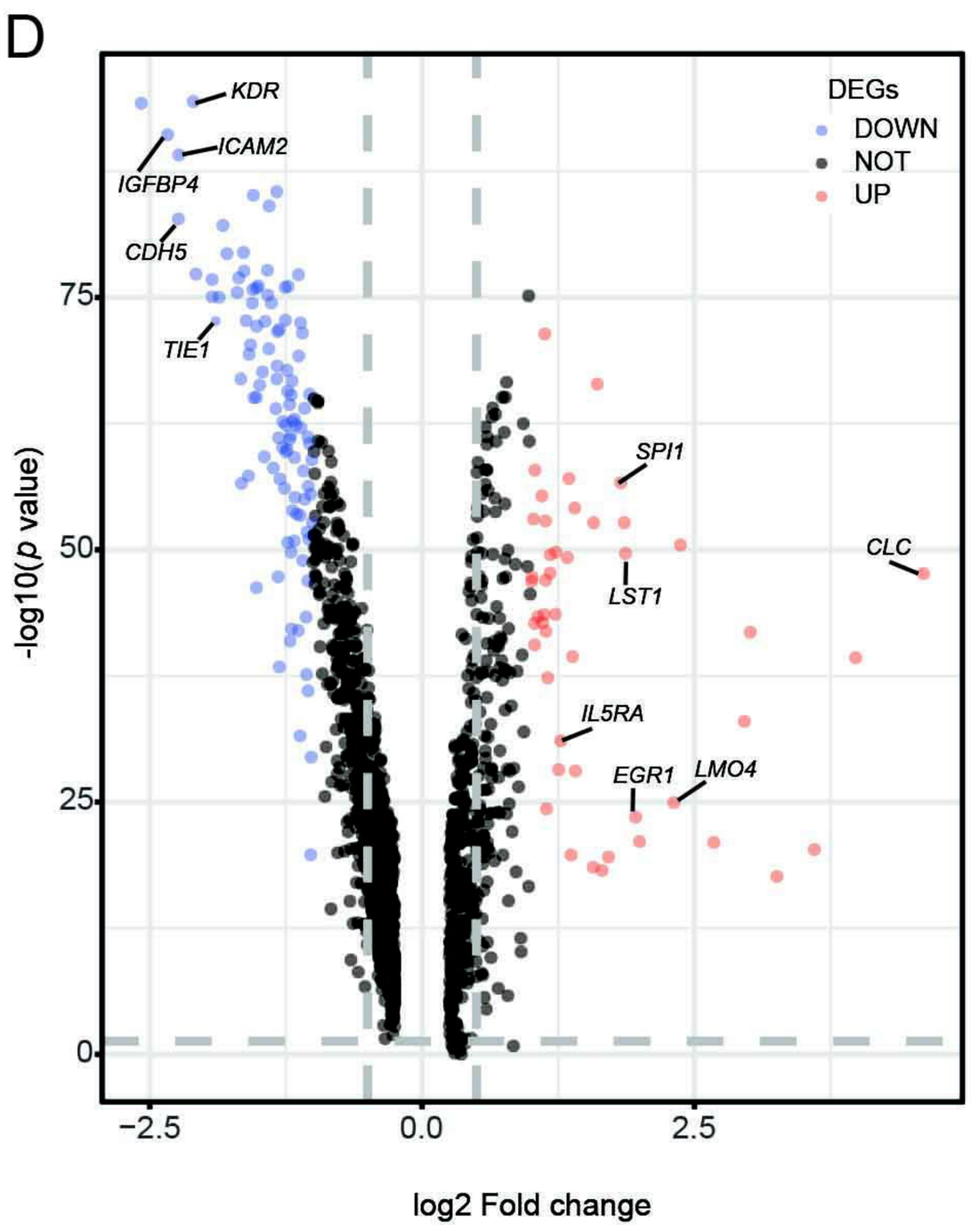

A
iPSCs
3,000-4,000 cells/well
Round-bottom 96-well plate
Embryoid bodies formation
Add SFM medium with cytokines every 3 days
Suspension cells Erythropoiesis
Erythroid cells
Day 0
Day 2
Day 11
Day 14
Day 28 or longer
Y27632 BMP4 FGF-2
BMP4 FGF-2 SCF VEGF
EPO
EPO SCF IL-3
B
Count of Cells Per Plate (x10^5)
Day 0 iPSCs
Day 14 Suspension cells
C
Count of Cells
25.9±3.05
27.08±1.63
CD34-FITC
CD45-PE
D
CFU-GEMM
E
Day 14
Day 17
Day 23
Day 28
Q1 21.7%
Q2 32.9%
Q4 44.9%
Q3 0.449%
Q3 10.1%
Q2 70.5%
Q4 17.6%
Q1 1.90%
Q3 0.811%
Q2 78.9%
Q4 13.3%
Q1 7.00%
Q1 0.331%
Q2 96.0%
Q4 1.28%
Q3 2.34%
CD71-PE
CD235a-APC
F
20 μm

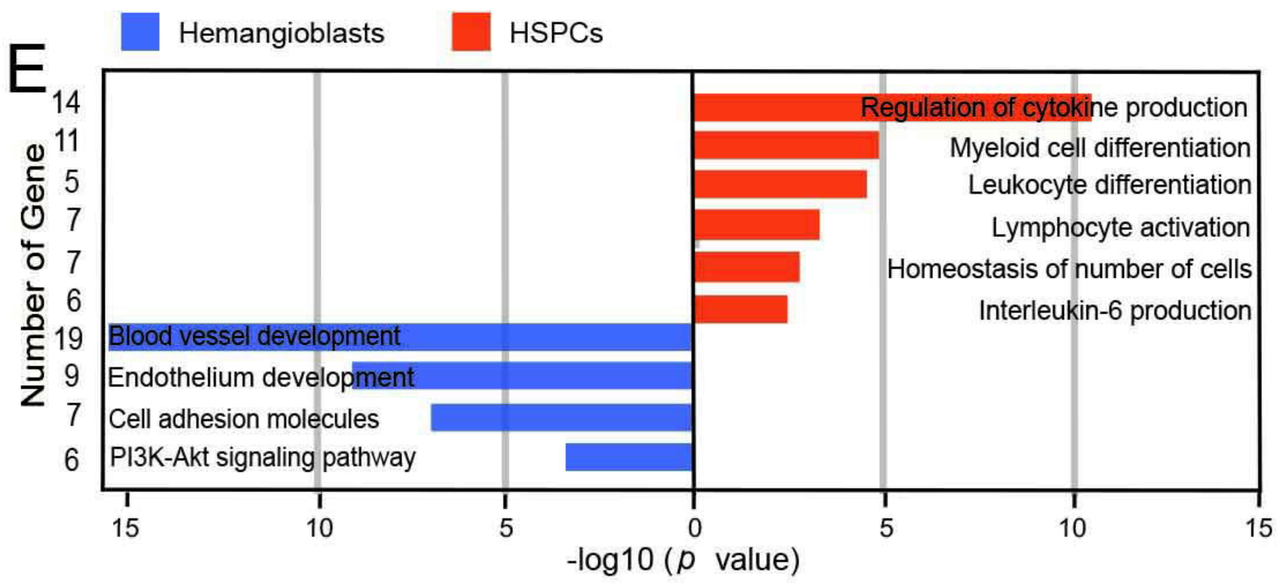

A
tSNE_2
tSNE_1
EB
SC
B
2_Progenitor cells-2
1_Progenitor cells-1
5_Hemangioblasts
3_Mesenchymal cells
9_Progenitor cells-9
6_Vascular endothelial cells
4_HSPCs
8_CD14 cells
7_MEs
0_CD33 cells
C
MPO
MNDA
S100A8
S100A9
TBX2
DLK1
EGFL6
IGFBP3
DKK1
FGFR2
PITX1
HAND1
DCN
LUM
ANXA2
S100A10
GATA2
RUNX1
MYB
LMO4
CD34
TCIM
KDR
ICAM2
WSB1
NEAT1
MALAT1
PDGFRB
GATA1
HBD
HBA1
HBE1
CD14
CD36
C1QC
C1QA
TGFBI
ELF3
EGFR
IGFBP5
D
ZFP42
ELF3
LYZ
CD36
HBA1
HBB
MALAT1
WSB1
CD34
KDR
MYB
RUNX1
ANXA2
DCN
HAND1
SOX6
PDGFRA
DLK1
MNDA
S100A8
0_EB
0_SC
1_EB
1_SC
2_EB
3_EB
3_SC
4_EB
4_SC
5_EB
5_SC
6_EB
6_SC
7_EB
7_SC
8_EB
8_SC
9_EB
pct.exp
EB
Max
Min
SC
Max
Min

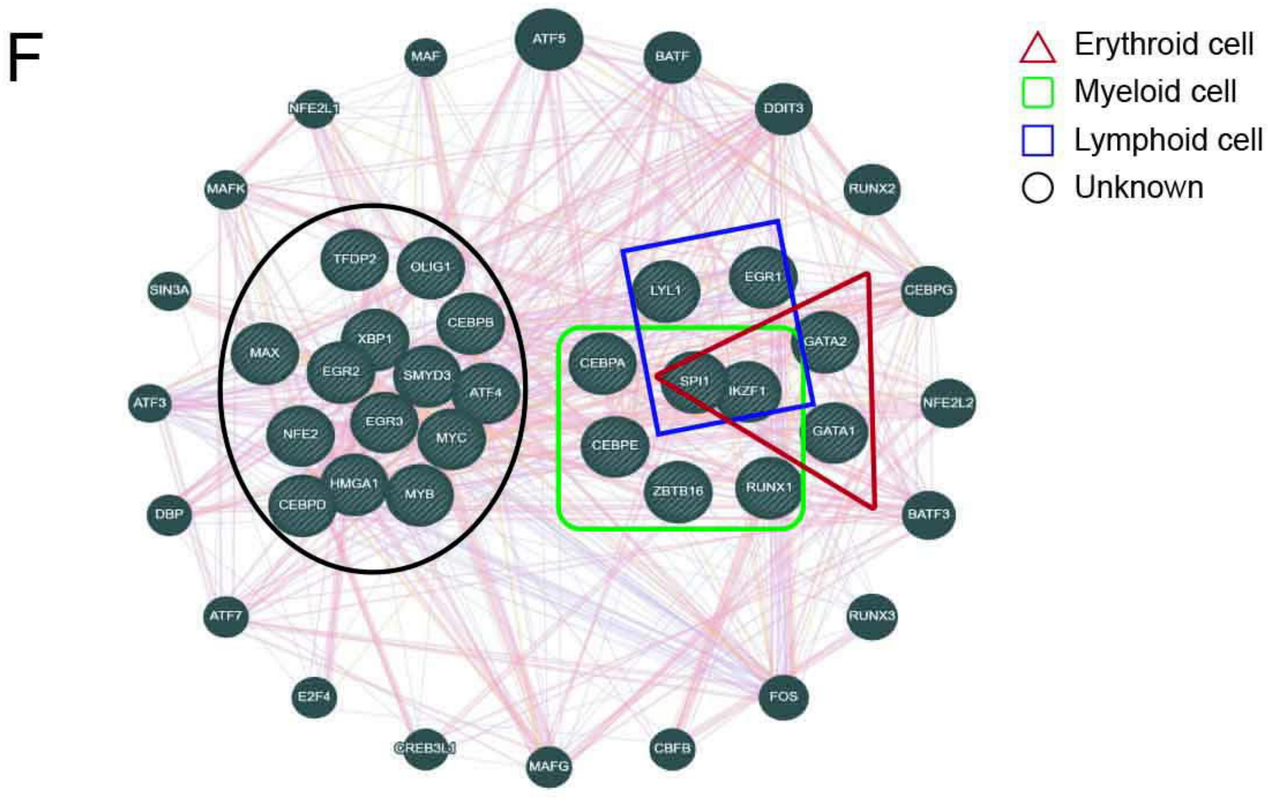

A
tSNE_1
tSNE_2
0a
0b
0c
1a
1b
2
3
4a
4b
5
6
7
8a
8b
9
0_CD33 cells
1_Progenitor cells−1
2_Progenitor cells−2
3_Mesenchymal cells
4_HSPCs
5_Hemangioblasts
6_Vascular endothelial cells
7_MEs
8_CD14 cells
9_Progenitor cells−9

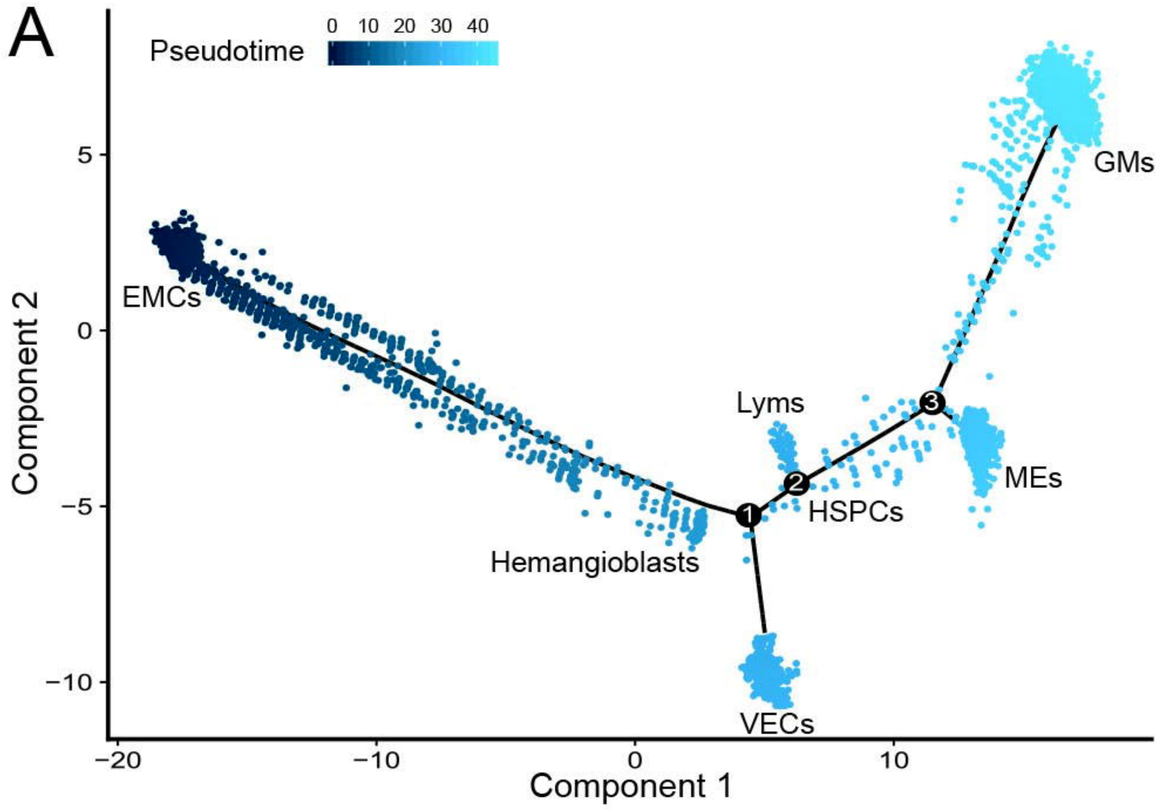

B
CD14
VCAN
CTSS
HLA-B
IL1B
PRTN3
ELANE
MPO
CTSG
CYTL1
S100A8
S100A9
NCF1
DEFA3
PADI4
0a
0b
0c

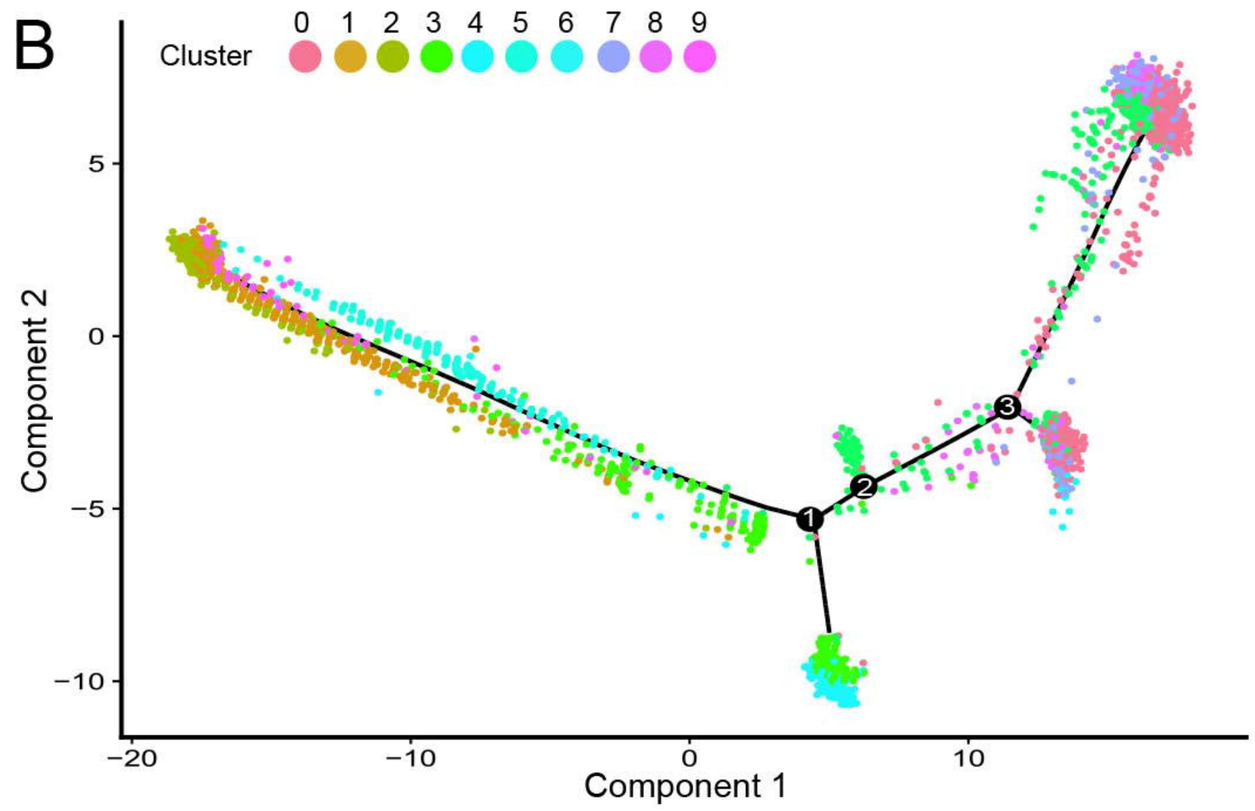

C
NFAP4
EGFL6
LUM
DCN
PITX2
DLK1
ID3
FN1
SOX4
NID2
1a
1b

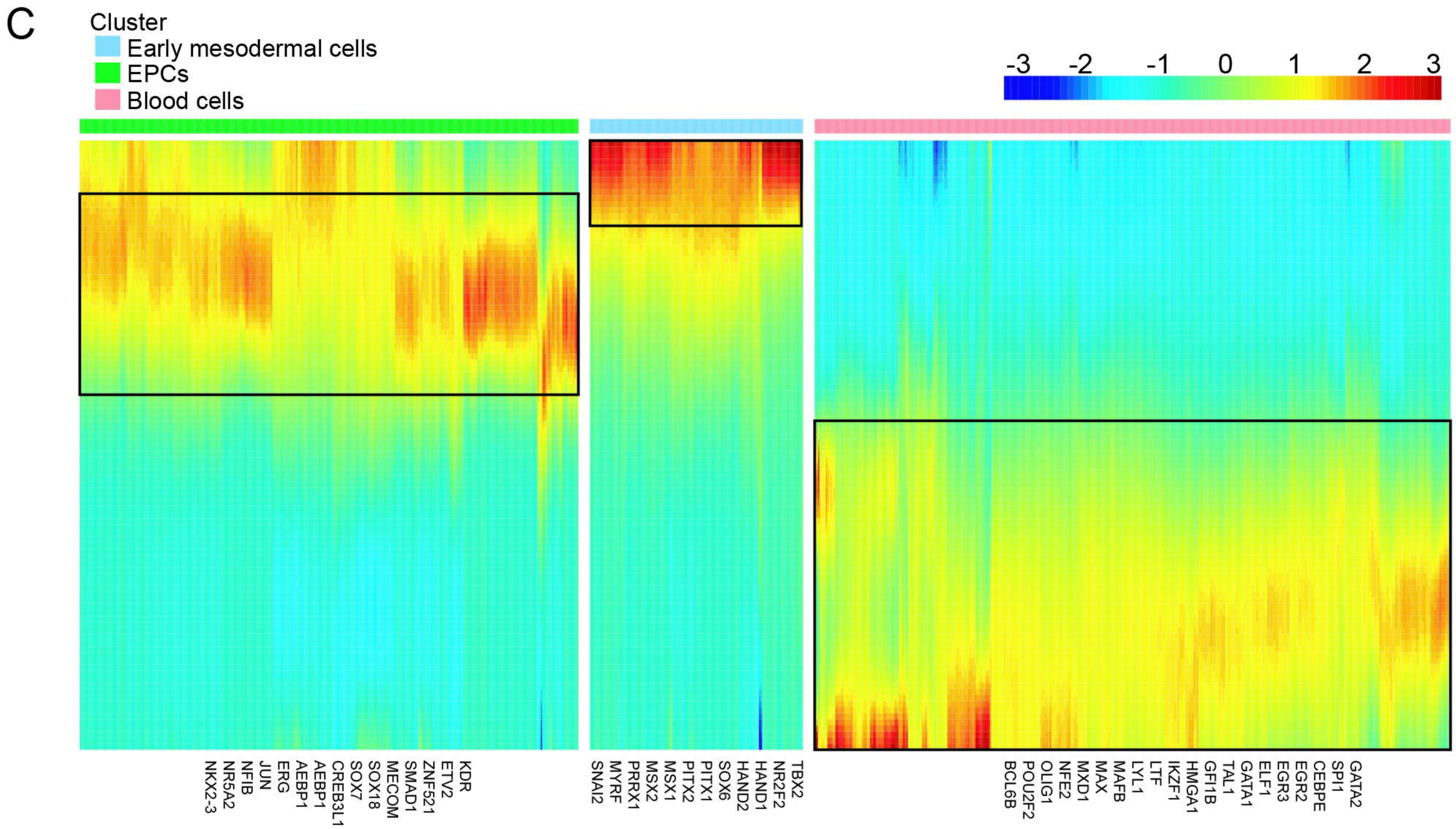

D
DEGs
DOWN
NOT
UP
KDR
ICAM2
IGFBP4
CDH5
TIE1
SPI1
LST1
CLC
IL5RA
EGR1
LMO4
-log10(p value)
log2 Fold change

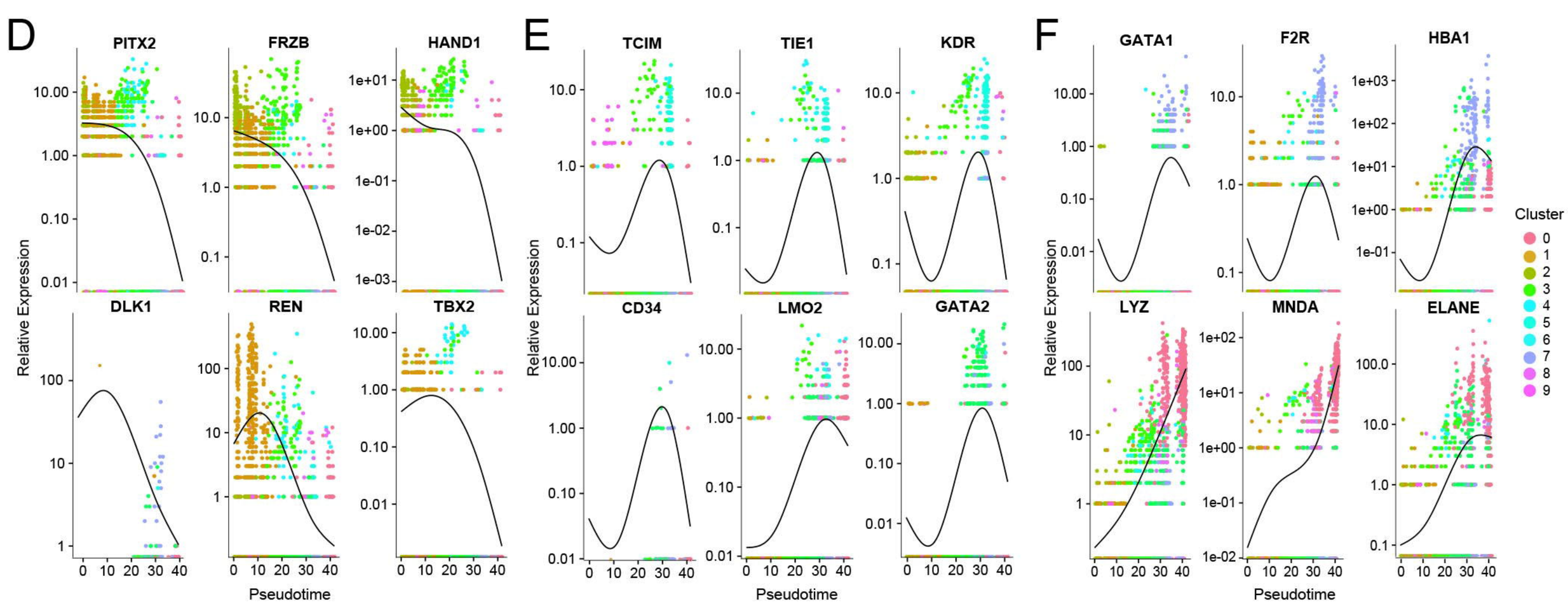

E
Hemangioblasts
HSPCs
Regulation of cytokine production
Myeloid cell differentiation
Leukocyte differentiation
Lymphocyte activation
Homeostasis of number of cells
Interleukin-6 production
Blood vessel development
Endothelium development
Cell adhesion molecules
PI3K-Akt signaling pathway
Number of Gene
-log10 (p value)

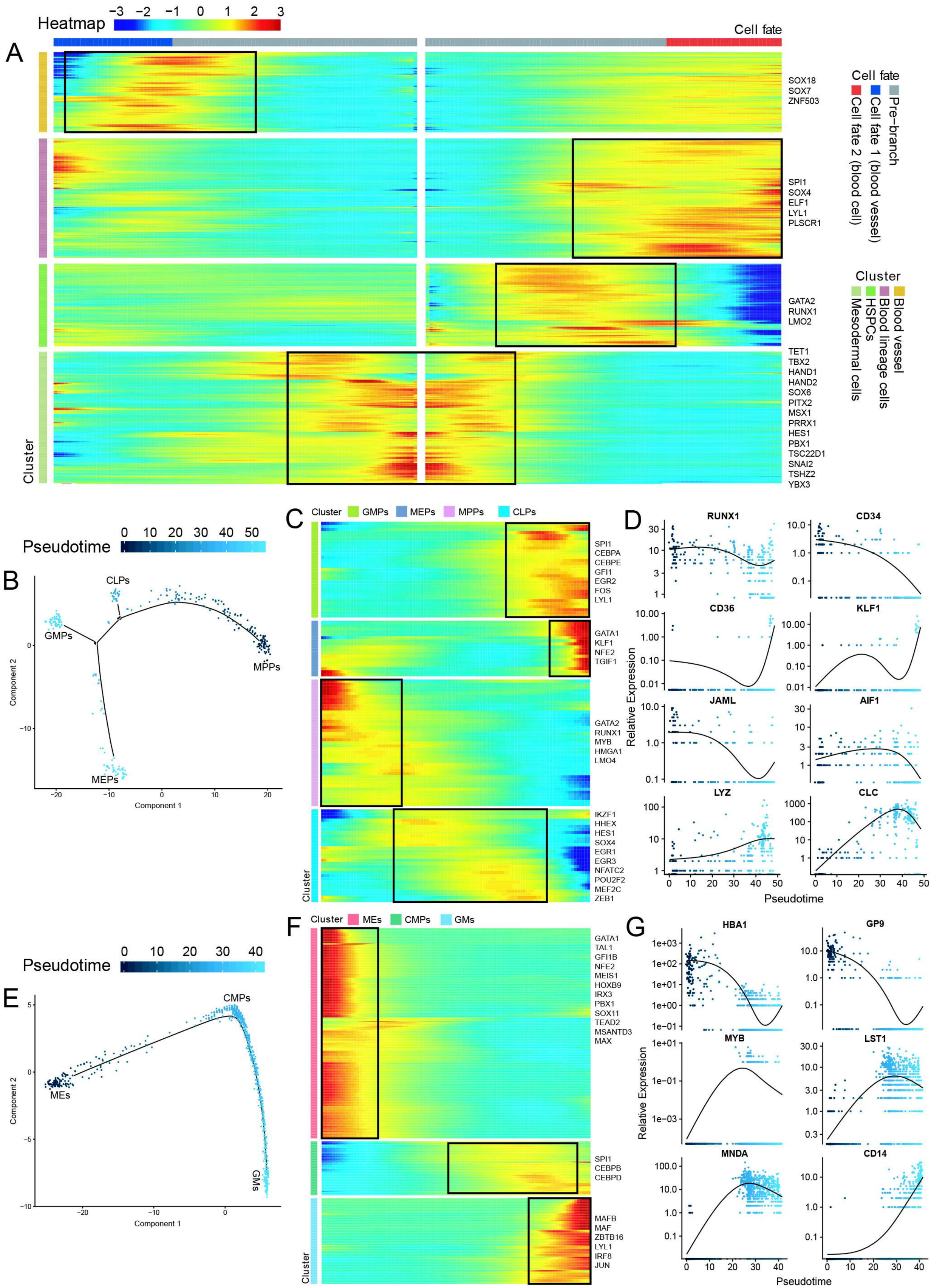

F
Erythroid cell
Myeloid cell
Lymphoid cell
Unknown

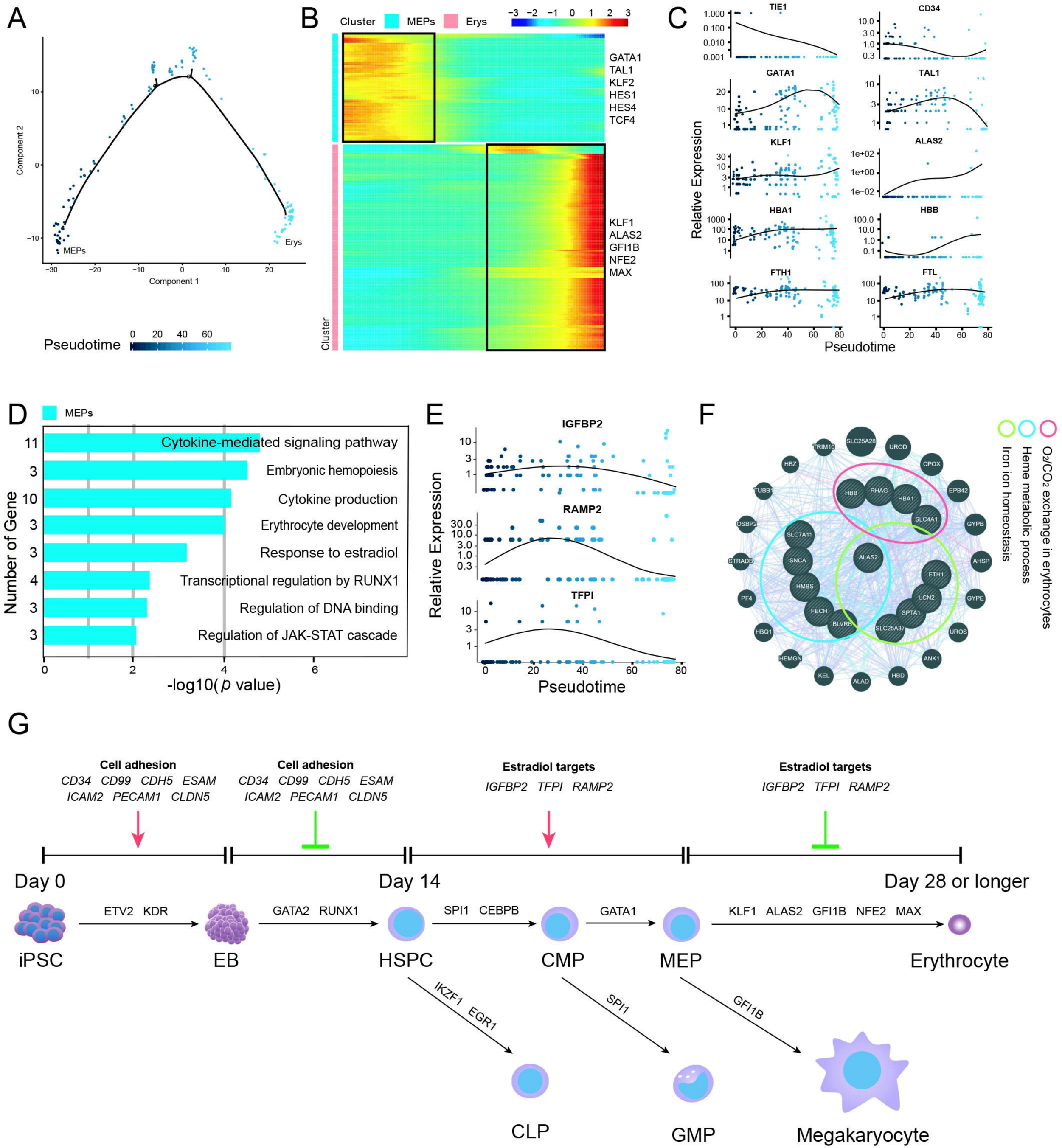

A
Pseudotime
0 10 20 30 40
EMCs
Hemangioblasts
Lyms
HSPCs
GMs
MEs
VECs
Component 1
Component 2

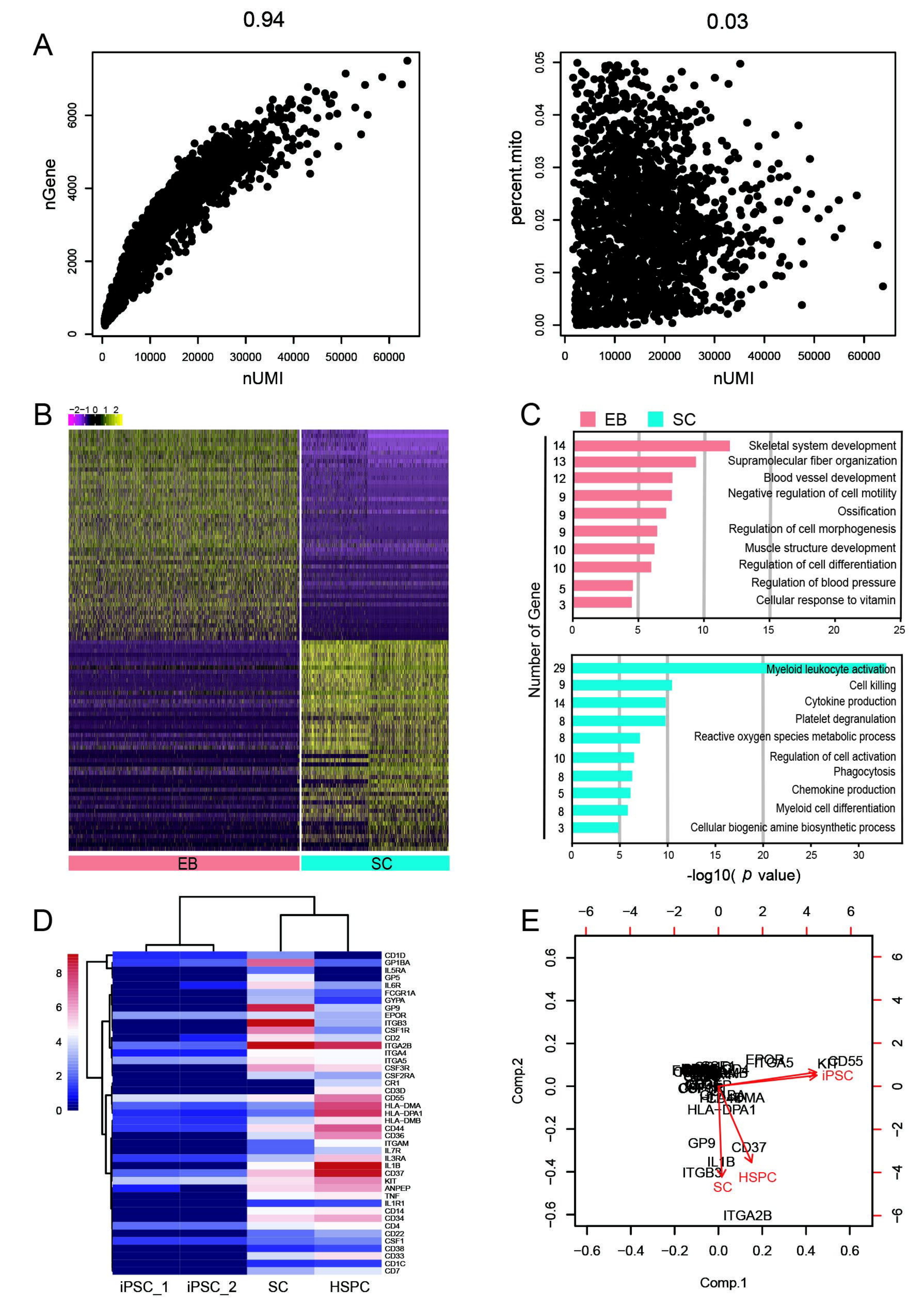

B
Cluster
0 1 2 3 4 5 6 7 8 9
Component 1
Component 2

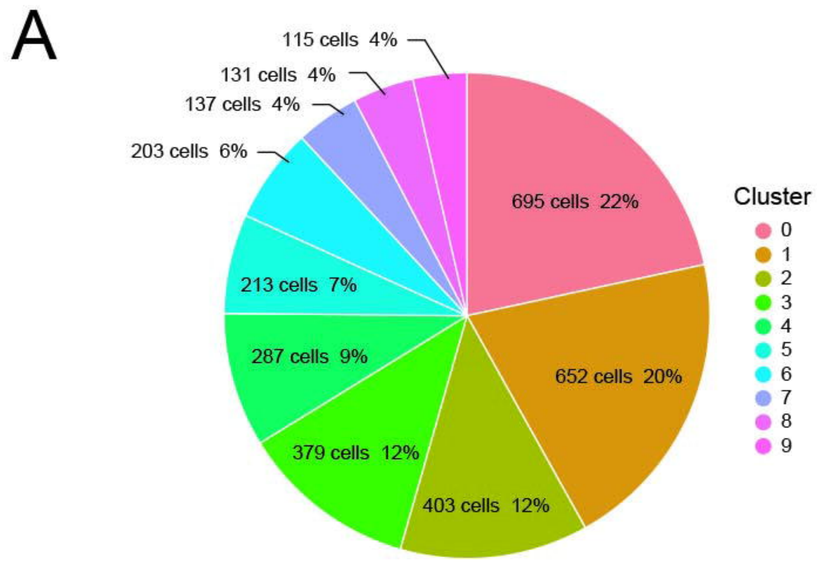

C
Cluster
Early mesodermal cells
EPCs
Blood cells
-3 -2 -1 0 1 2 3
NKX2-3
NR5A2
NFIB
JUN
ERG
AEBP1
AEBP1
CREB3L1
SOX7
SOX18
MECOM
SMAD1
ZNF521
ETV2
KDR
SNAI2
MYRF
PRRX1
MSX2
MSX1
PITX2
PITX1
SOX6
HAND2
HAND1
NR2F2
TBX2
BCL6B
POU2F2
OLIG1
NFE2
MXD1
MAX
MAFB
LYL1
LTF
IKZF1
HMGA1
GFI1B
TAL1
GATA1
ELF1
EGR3
EGR2
CEBPE
SPI1
GATA2

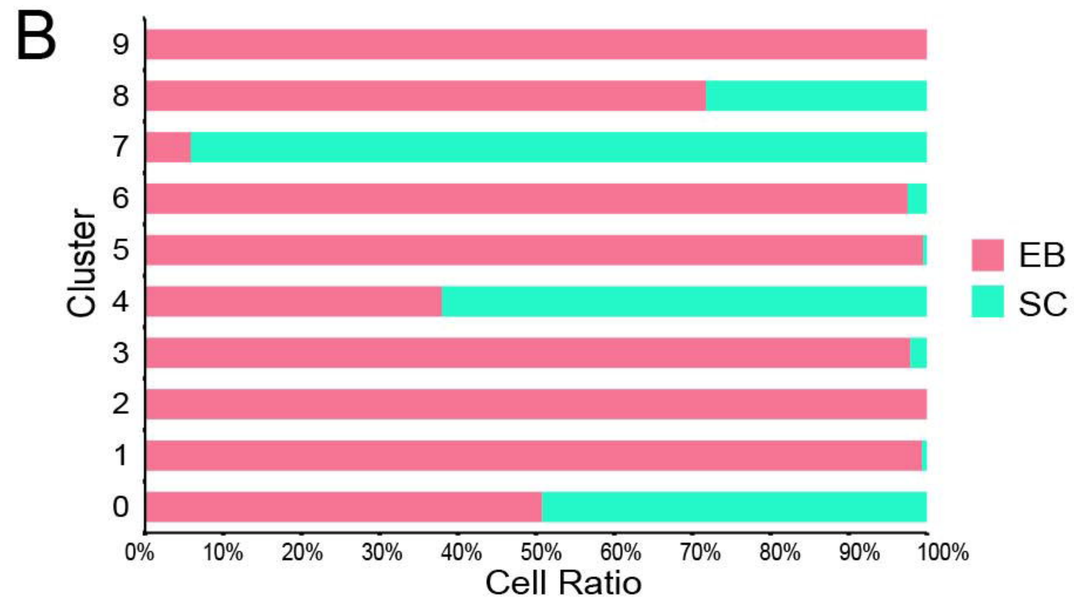

D
PITX2
FRZB
HAND1
DLK1
REN
TBX2
E
TCIM
TIE1
KDR
CD34
LMO2
GATA2
F
GATA1
F2R
HBA1
LYZ
MNDA
ELANE
Relative Expression
Pseudotime
Cluster
0 1 2 3 4 5 6 7 8 9

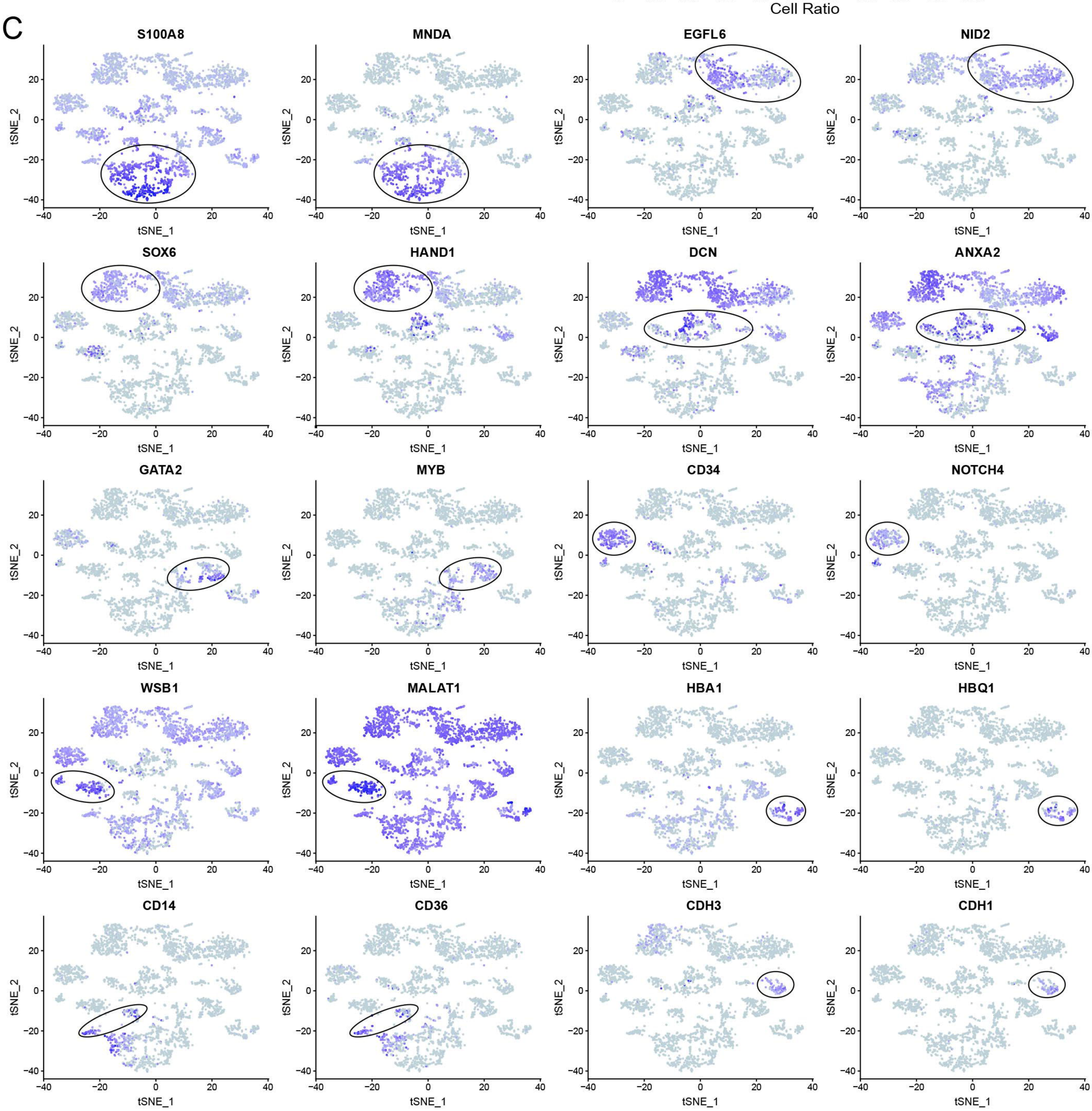

A
Heatmap
−3 −2 −1 0 1 2 3
Cell fate
SOX18
SOX7
ZNF503
SPI1
SOX4
ELF1
LYL1
PLSCR1
GATA2
RUNX1
LMO2
TET1
TBX2
HAND1
HAND2
SOX6
PITX2
MSX1
PRRX1
HES1
PBX1
TSC22D1
SNAI2
TSHZ2
YBX3
Cluster
Cell fate
Pre−branch
Cell fate 1 (blood vessel)
Cell fate 2 (blood cell)
Cluster
Blood vessel
Blood lineage cells
HSPCs
Mesodermal cells
Pseudotime
0 10 20 30 40 50
B
CLPs
GMPs
MPPs
MEPs
Component 1
Component 2
C
Cluster GMPs MEPs MPPs CLPs
SPI1
CEBPA
CEBPE
GFI1
EGR2
FOS
LYL1
GATA1
KLF1
NFE2
TGIF1
GATA2
RUNX1
MYB
HMGA1
LMO4
IKZF1
HHEX
HES1
SOX4
EGR1
EGR3
NFATC2
POU2F2
MEF2C
ZEB1
D
RUNX1
CD34
CD36
KLF1
JAML
AIF1
LYZ
CLC
Relative Expression
Pseudotime
Pseudotime
0 10 20 30 40
E
CMPs
MEs
GMs
Component 1
Component 2
F
Cluster MEs CMPs GMs
GATA1
TAL1
GFI1B
NFE2
MEIS1
HOXB9
IRX3
PBX1
SOX11
TEAD2
MSANTD3
MAX
SPI1
CEBPB
CEBPD
MAFB
MAF
ZBTB16
LYL1
IRF8
JUN
G
HBA1
GP9
MYB
LST1
MNDA
CD14
Relative Expression
Pseudotime

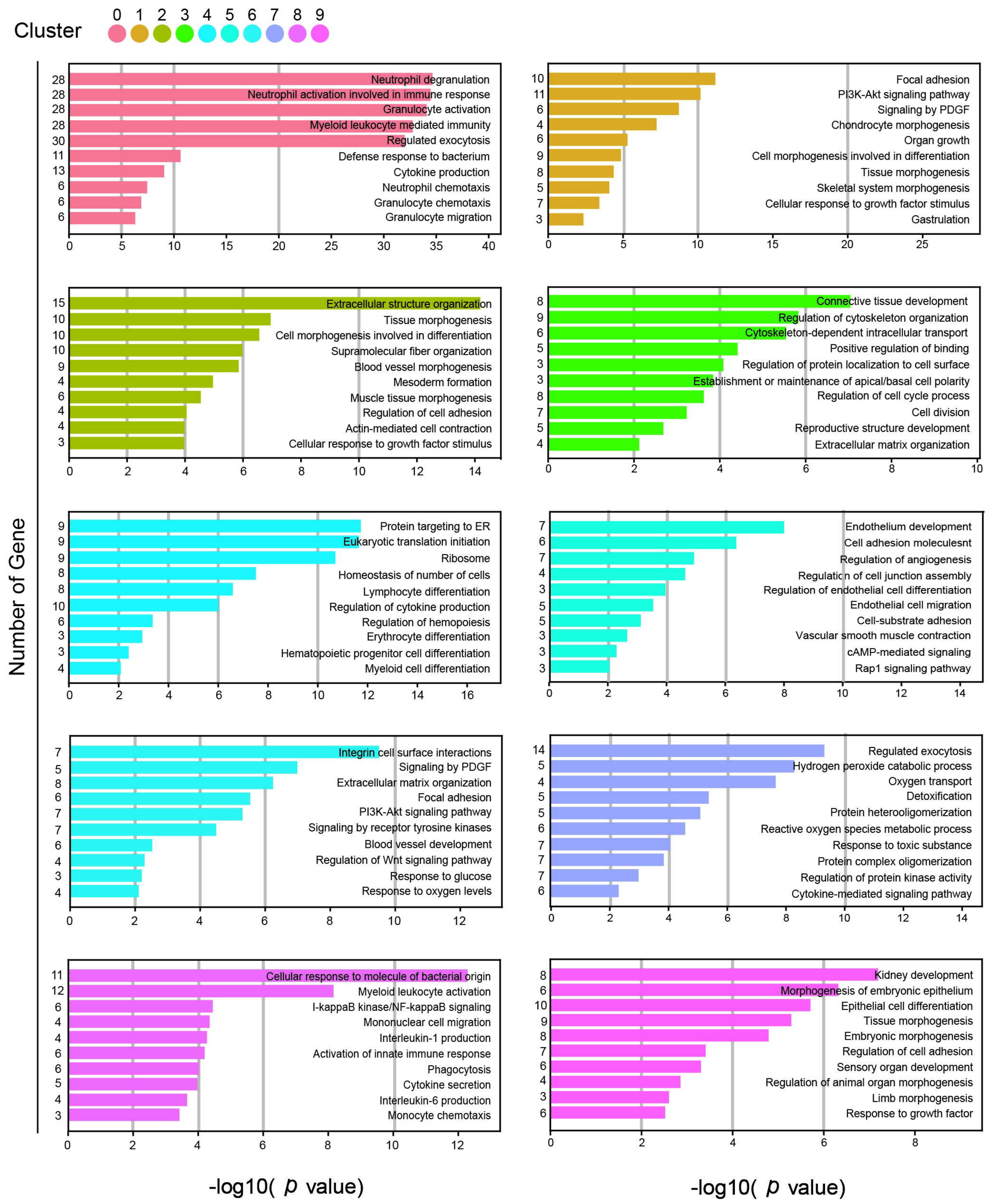

A
Component 2
Component 1
MEPs
Erys
Pseudotime
B
Cluster
MEPs
Erys
GATA1
TAL1
KLF2
HES1
HES4
TCF4
KLF1
ALAS2
GFI1B
NFE2
MAX
C
TIE1
CD34
GATA1
TAL1
KLF1
ALAS2
HBA1
HBB
FTH1
FTL
Relative Expression
Pseudotime
D
MEPs
Cytokine-mediated signaling pathway
Embryonic hemopoiesis
Cytokine production
Erythrocyte development
Response to estradiol
Transcriptional regulation by RUNX1
Regulation of DNA binding
Regulation of JAK-STAT cascade
Number of Gene
-log10(p value)
E
IGFBP2
RAMP2
TFPI
Relative Expression
Pseudotime
F
O2/CO2 exchange in erythrocytes
Heme metabolic process
Iron ion homeostasis
G
Cell adhesion
CD34 CD99 CDH5 ESAM
ICAM2 PECAM1 CLDN5
Cell adhesion
CD34 CD99 CDH5 ESAM
ICAM2 PECAM1 CLDN5
Estradiol targets
IGFBP2 TFPI RAMP2
Estradiol targets
IGFBP2 TFPI RAMP2
Day 0
Day 14
Day 28 or longer
ETV2 KDR
GATA2 RUNX1
SPI1 CEBPB
GATA1
KLF1 ALAS2 GFI1B NFE2 MAX
iPSC
EB
HSPC
CMP
MEP
Erythrocyte
IKZF1 EGR1
SPI1
GFI1B
CLP
GMP
Megakaryocyte

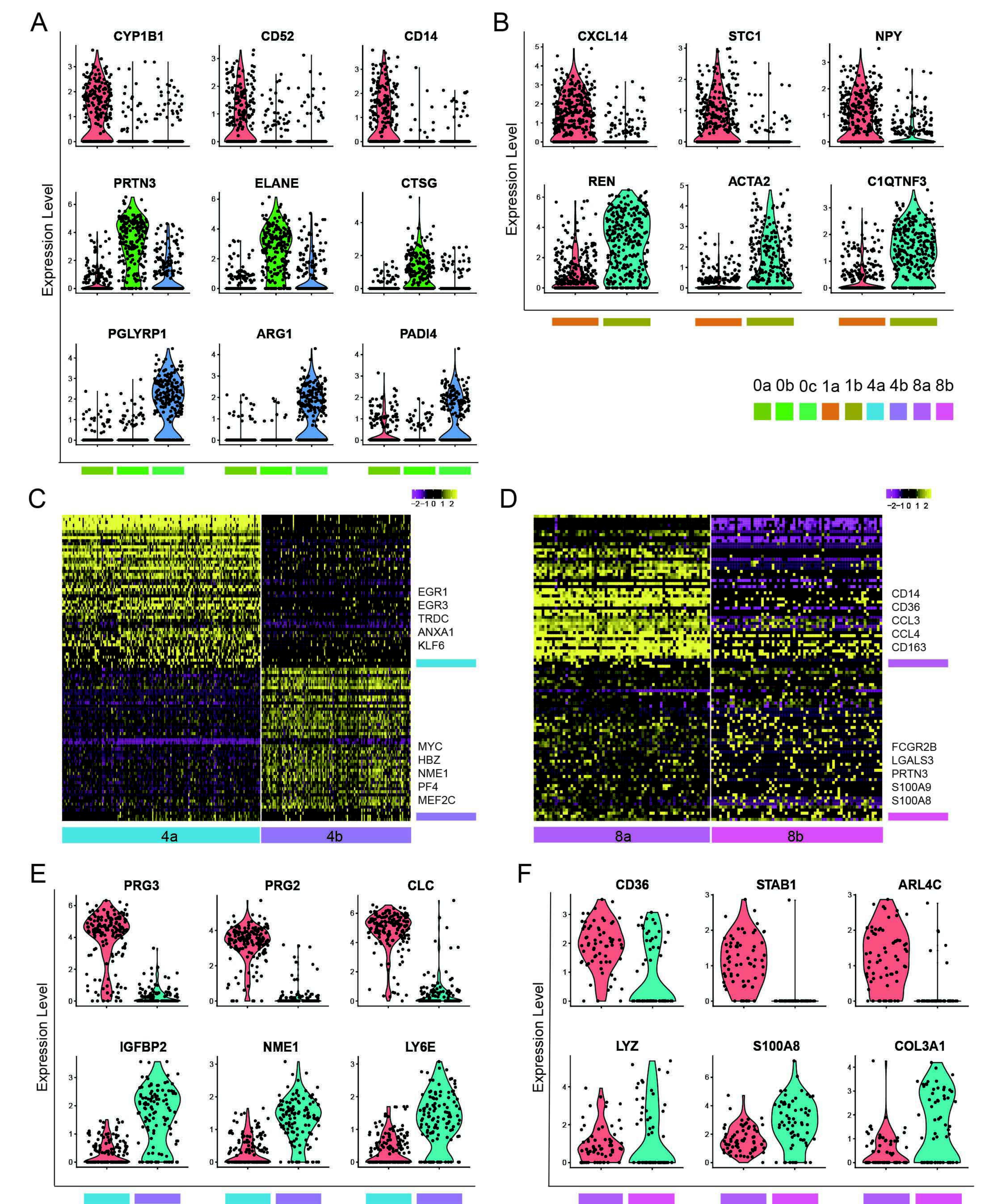

A
0.94
nGene
nUMI
0.03
percent.mito
nUMI
B
EB
SC
C
EB
SC
Skeletal system development
Supramolecular fiber organization
Blood vessel development
Negative regulation of cell motility
Ossification
Regulation of cell morphogenesis
Muscle structure development
Regulation of cell differentiation
Regulation of blood pressure
Cellular response to vitamin
Myeloid leukocyte activation
Cell killing
Cytokine production
Platelet degranulation
Reactive oxygen species metabolic process
Regulation of cell activation
Phagocytosis
Chemokine production
Myeloid cell differentiation
Cellular biogenic amine biosynthetic process
Number of Gene
-log10(p value)
D
iPSC_1
iPSC_2
SC
HSPC
E
Comp.1
Comp.2
iPSC
HSPC
SC
HLA-DPA1
GP9
CD37
IL1B
ITGB3
ITGA2B

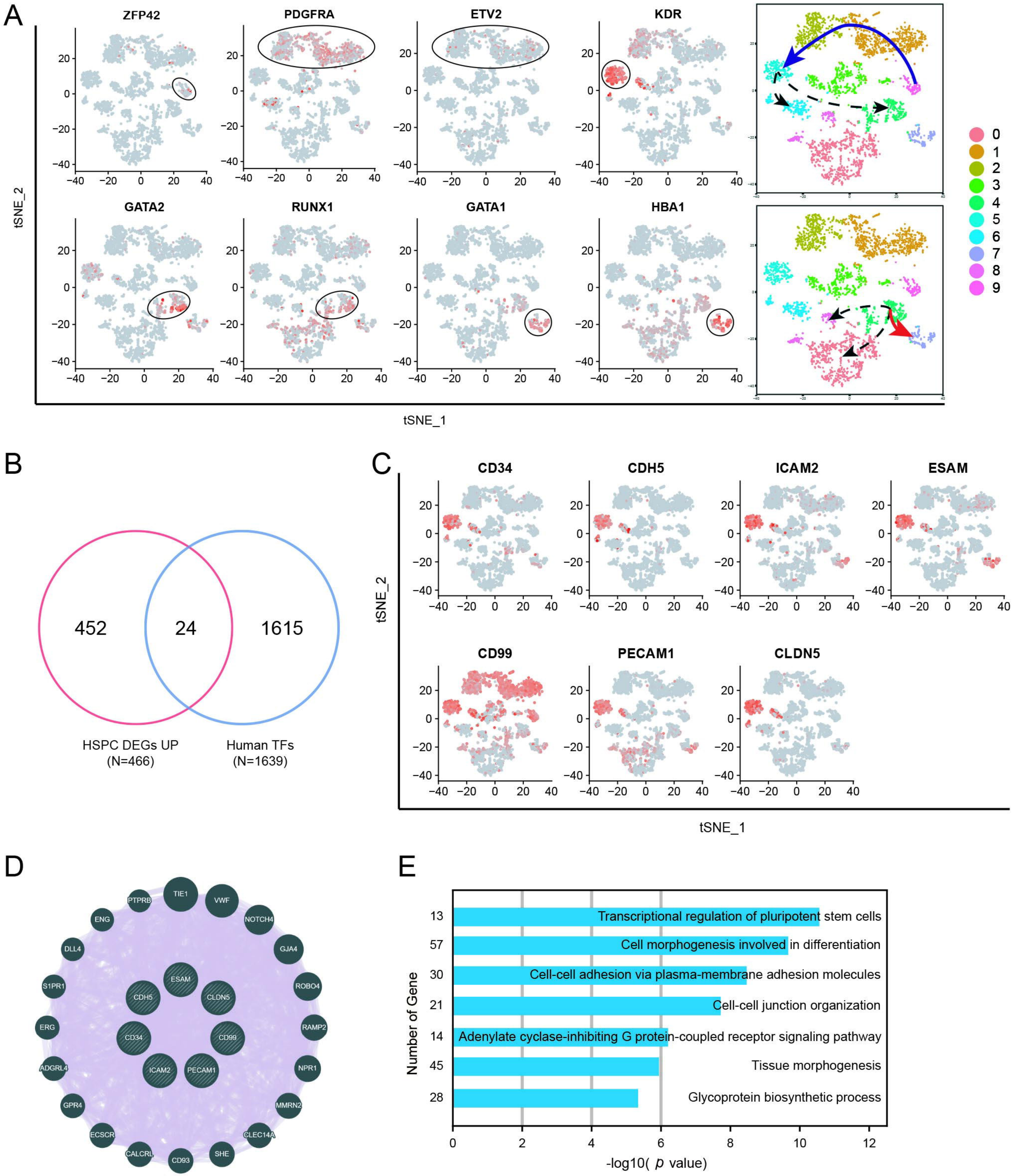

A
115 cells 4%
131 cells 4%
137 cells 4%
203 cells 6%
213 cells 7%
287 cells 9%
379 cells 12%
403 cells 12%
652 cells 20%
695 cells 22%
Cluster
0
1
2
3
4
5
6
7
8
9

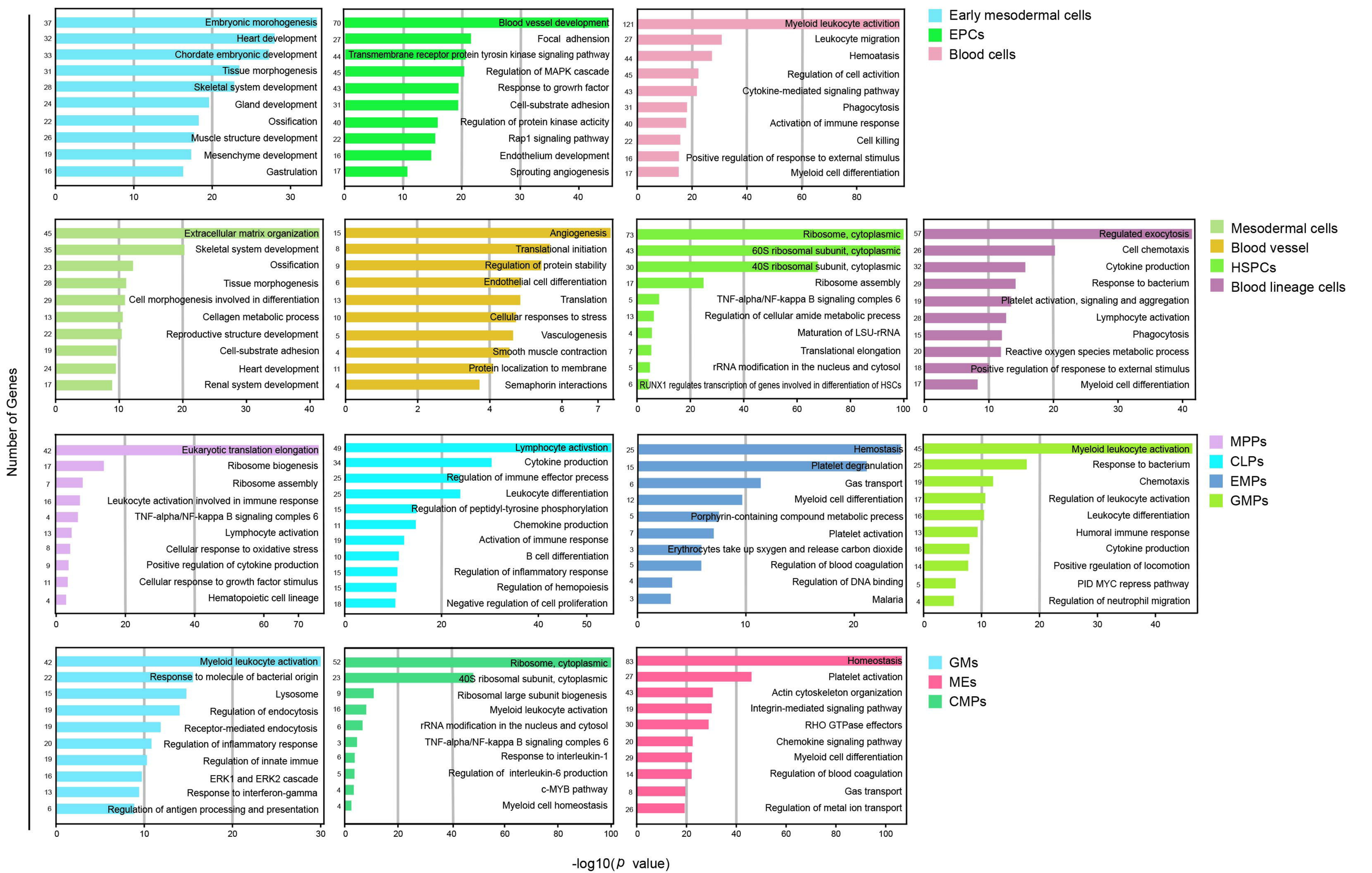

B
Cluster
9
8
7
6
5
4
3
2
1
0
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Cell Ratio
EB
SC

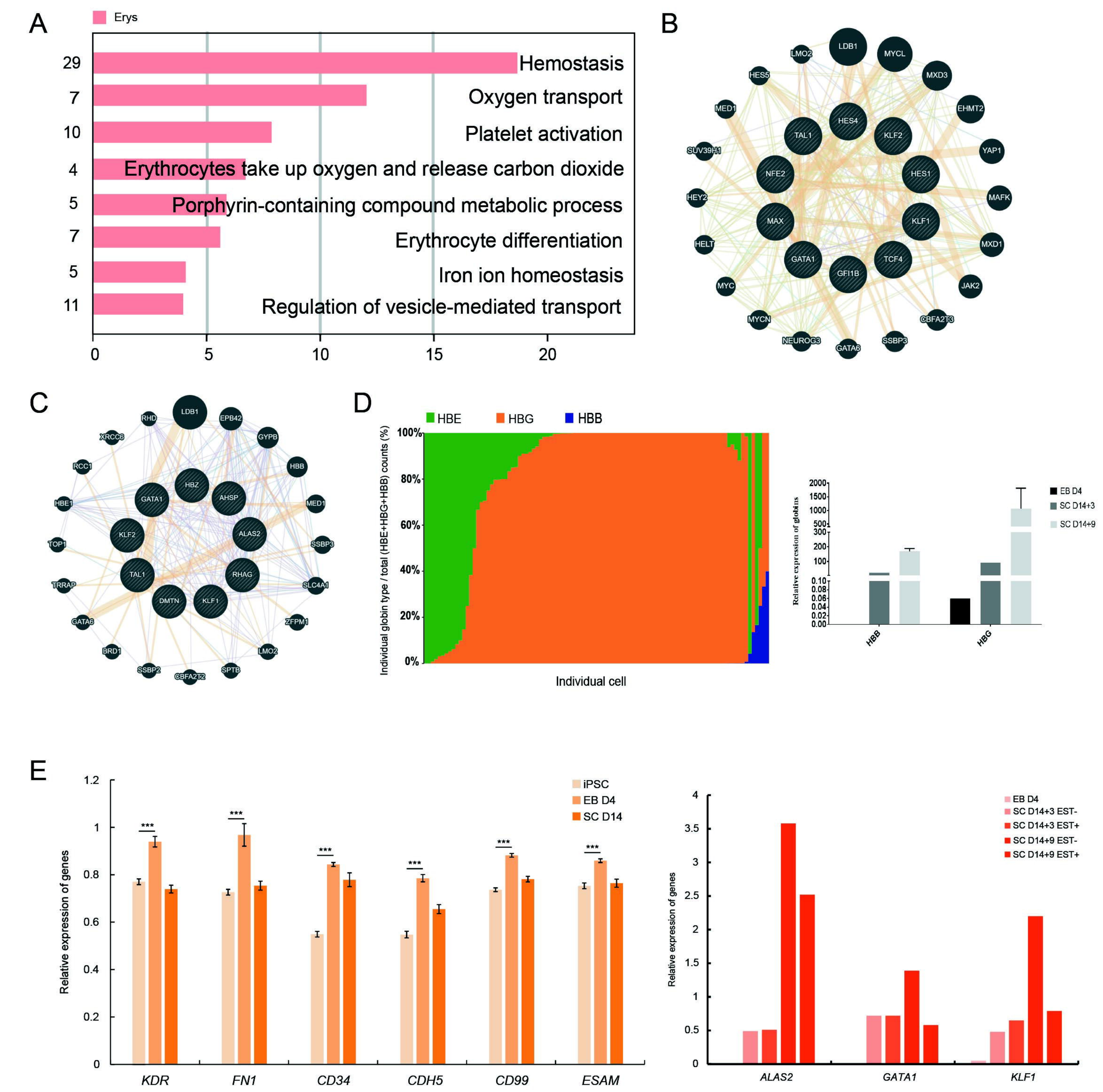

C
S100A8
MNDA
EGFL6
NID2
SOX6
HAND1
DCN
ANXA2
GATA2
MYB
CD34
NOTCH4
WSB1
MALAT1
HBA1
HBQ1
CD14
CD36
CDH3
CDH1
tSNE_1
tSNE_2


Cluster
0 1 2 3 4 5 6 7 8 9
Number of Gene
28 Neutrophil degranulation
28 Neutrophil activation involved in immune response
28 Granulocyte activation
28 Myeloid leukocyte mediated immunity
30 Regulated exocytosis
11 Defense response to bacterium
13 Cytokine production
6 Neutrophil chemotaxis
6 Granulocyte chemotaxis
6 Granulocyte migration
0 5 10 15 20 25 30 35 40
10 Focal adhesion
11 PI3K-Akt signaling pathway
6 Signaling by PDGF
4 Chondrocyte morphogenesis
6 Organ growth
9 Cell morphogenesis involved in differentiation
8 Tissue morphogenesis
5 Skeletal system morphogenesis
7 Cellular response to growth factor stimulus
3 Gastrulation
0 5 10 15 20 25
15 Extracellular structure organization
10 Tissue morphogenesis
10 Cell morphogenesis involved in differentiation
10 Supramolecular fiber organization
9 Blood vessel morphogenesis
4 Mesoderm formation
6 Muscle tissue morphogenesis
4 Regulation of cell adhesion
4 Actin-mediated cell contraction
3 Cellular response to growth factor stimulus
0 2 4 6 8 10 12 14
8 Connective tissue development
9 Regulation of cytoskeleton organization
6 Cytoskeleton-dependent intracellular transport
5 Positive regulation of binding
3 Regulation of protein localization to cell surface
3 Establishment or maintenance of apical/basal cell polarity
8 Regulation of cell cycle process
7 Cell division
5 Reproductive structure development
4 Extracellular matrix organization
0 2 4 6 8 10
9 Protein targeting to ER
9 Eukaryotic translation initiation
9 Ribosome
8 Homeostasis of number of cells
8 Lymphocyte differentiation
10 Regulation of cytokine production
6 Regulation of hemopoiesis
3 Erythrocyte differentiation
3 Hematopoietic progenitor cell differentiation
4 Myeloid cell differentiation
0 2 4 6 8 10 12 14 16
7 Endothelium development
6 Cell adhesion moleculesnt
7 Regulation of angiogenesis
4 Regulation of cell junction assembly
3 Regulation of endothelial cell differentiation
5 Endothelial cell migration
5 Cell-substrate adhesion
3 Vascular smooth muscle contraction
3 cAMP-mediated signaling
3 Rap1 signaling pathway
0 2 4 6 8 10 12 14
7 Integrin cell surface interactions
5 Signaling by PDGF
8 Extracellular matrix organization
6 Focal adhesion
7 PI3K-Akt signaling pathway
7 Signaling by receptor tyrosine kinases
6 Blood vessel development
4 Regulation of Wnt signaling pathway
3 Response to glucose
4 Response to oxygen levels
0 2 4 6 8 10 12
14 Regulated exocytosis
5 Hydrogen peroxide catabolic process
4 Oxygen transport
5 Detoxification
5 Protein heterooligomerization
6 Reactive oxygen species metabolic process
7 Response to toxic substance
7 Protein complex oligomerization
7 Regulation of protein kinase activity
6 Cytokine-mediated signaling pathway
0 2 4 6 8 10 12 14
11 Cellular response to molecule of bacterial origin
12 Myeloid leukocyte activation
6 I-kappaB kinase/NF-kappaB signaling
4 Mononuclear cell migration
4 Interleukin-1 production
6 Activation of innate immune response
6 Phagocytosis
5 Cytokine secretion
4 Interleukin-6 production
3 Monocyte chemotaxis
0 2 4 6 8 10 12
8 Kidney development
6 Morphogenesis of embryonic epithelium
10 Epithelial cell differentiation
9 Tissue morphogenesis
8 Embryonic morphogenesis
7 Regulation of cell adhesion
6 Sensory organ development
4 Regulation of animal organ morphogenesis
3 Limb morphogenesis
6 Response to growth factor
0 2 4 6 8
-log10(p value)
-log10(p value)


A
CYP1B1
CD52
CD14
PRTN3
ELANE
CTSG
PGLYRP1
ARG1
PADI4
Expression Level
B
CXCL14
STC1
NPY
REN
ACTA2
C1QTNF3
Expression Level
0a 0b 0c 1a 1b 4a 4b 8a 8b
C
EGR1
EGR3
TRDC
ANXA1
KLF6
MYC
HBZ
NME1
PF4
MEF2C
4a
4b
D
CD14
CD36
CCL3
CCL4
CD163
FCGR2B
LGALS3
PRTN3
S100A9
S100A8
8a
8b
E
PRG3
PRG2
CLC
IGFBP2
NME1
LY6E
Expression Level
F
CD36
STAB1
ARL4C
LYZ
S100A8
COL3A1
Expression Level


A
ZFP42
PDGFRA
ETV2
KDR
GATA2
RUNX1
GATA1
HBA1
tSNE_2
tSNE_1
0
1
2
3
4
5
6
7
8
9
B
452
24
1615
HSPC DEGs UP
(N=466)
Human TFs
(N=1639)
C
CD34
CDH5
ICAM2
ESAM
CD99
PECAM1
CLDN5
tSNE_2
tSNE_1
D
TIE1
VWF
NOTCH4
GJA4
ROBO4
RAMP2
NPR1
MMRN2
CLEC14A
SHE
CD93
CALCRL
ECSCR
GPR4
ADGRL4
ERG
S1PR1
DLL4
ENG
PTPRB
ESAM
CDH5
CLDN5
CD34
CD99
ICAM2
PECAM1
E
Number of Gene
13 Transcriptional regulation of pluripotent stem cells
57 Cell morphogenesis involved in differentiation
30 Cell-cell adhesion via plasma-membrane adhesion molecules
21 Cell-cell junction organization
14 Adenylate cyclase-inhibiting G protein-coupled receptor signaling pathway
45 Tissue morphogenesis
28 Glycoprotein biosynthetic process
-log10(p value)


Early mesodermal cells
EPCs
Blood cells
Mesodermal cells
Blood vessel
HSPCs
Blood lineage cells
MPPs
CLPs
EMPs
GMPs
GMs
MEs
CMPs
Number of Genes
-log10(p value)
37 Embryonic morohogenesis
32 Heart development
33 Chordate embryonic development
31 Tissue morphogenesis
28 Skeletal system development
24 Gland development
22 Ossification
26 Muscle structure development
19 Mesenchyme development
16 Gastrulation
70 Blood vessel development
27 Focal adhension
44 Transmembrane receptor protein tyrosin kinase signaling pathway
45 Regulation of MAPK cascade
43 Response to growrh factor
31 Cell-substrate adhesion
40 Regulation of protein kinase acticity
22 Rap1 signaling pathway
16 Endothelium development
17 Sprouting angiogenesis
121 Myeloid leukocyte activition
27 Leukocyte migration
44 Hemoatasis
45 Regulation of cell activition
43 Cytokine-mediated signaling pathway
31 Phagocytosis
40 Activation of immune response
22 Cell killing
16 Positive regulation of response to external stimulus
17 Myeloid cell differentiation
45 Extracellular matrix organization
35 Skeletal system development
23 Ossification
28 Tissue morphogenesis
29 Cell morphogenesis involved in differentiation
13 Cellagen metabolic process
22 Reproductive structure development
19 Cell-substrate adhesion
24 Heart development
17 Renal system development
15 Angiogenesis
8 Translational initiation
9 Regulation of protein stability
6 Endothelial cell differentiation
13 Translation
10 Cellular responses to stress
5 Vasculogenesis
4 Smooth muscle contraction
11 Protein localization to membrane
4 Semaphorin interactions
73 Ribosome, cytoplasmic
43 60S ribosomal subunit, cytoplasmic
30 40S ribosomal subunit, cytoplasmic
17 Ribosome assembly
5 TNF-alpha/NF-kappa B signaling comples 6
13 Regulation of cellular amide metabolic precess
4 Maturation of LSU-rRNA
7 Translational elongation
5 rRNA modification in the nucleus and cytosol
6 RUNX1 regulates transcription of genes involved in differentiation of HSCs
57 Regulated exocytosis
26 Cell chemotaxis
32 Cytokine production
29 Response to bacterium
19 Platelet activation, signaling and aggregation
28 Lymphocyte activation
15 Phagocytosis
20 Reactive oxygen species metabolic process
18 Positive regulation of responese to external stimulus
17 Myeloid cell differentiation
42 Eukaryotic translation elongation
17 Ribosome biogenesis
7 Ribosome assembly
16 Leukocyte activation involved in immune response
4 TNF-alpha/NF-kappa B signaling comples 6
13 Lymphocyte activation
8 Cellular response to oxidative stress
9 Positive regulation of cytokine production
11 Cellular response to growth factor stimulus
4 Hematopoietic cell lineage
49 Lymphocyte activstion
34 Cytokine production
25 Regulation of immune effector precess
25 Leukocyte differentiation
15 Regulation of peptidyl-tyrosine phosphorylation
11 Chemokine production
19 Activation of immune response
10 B cell differentiation
15 Regulation of inflammatory response
15 Regulation of hemopoiesis
18 Negative regulation of cell proliferation
25 Hemostasis
15 Platelet degranulation
6 Gas transport
12 Myeloid cell differentiation
5 Porphyrin-containing compound metabolic precess
7 Platelet activation
3 Erythrocytes take up sxygen and release carbon dioxide
5 Regulation of blood coagulation
4 Regulation of DNA binding
3 Malaria
45 Myeloid leukocyte activation
25 Response to bacterium
19 Chemotaxis
17 Regulation of leukocyte activation
16 Leukocyte differentiation
13 Humoral immune response
16 Cytokine production
14 Positive rgeulation of locomotion
5 PID MYC repress pathway
4 Regulation of neutrophil migration
42 Myeloid leukocyte activation
22 Response to molecule of bacterial origin
15 Lysosome
19 Regulation of endocytosis
19 Receptor-mediated endocytosis
20 Regulation of inflammatory response
19 Regulation of innate immue
16 ERK1 and ERK2 cascade
13 Response to interferon-gamma
6 Regulation of antigen processing and presentation
52 Ribosome, cytoplasmic
23 40S ribosomal subunit, cytoplasmic
9 Ribosomal large subunit biogenesis
16 Myeloid leukocyte activation
6 rRNA modification in the nucleus and cytosol
3 TNF-alpha/NF-kappa B signaling comples 6
6 Response to interleukin-1
5 Regulation of interleukin-6 production
4 c-MYB pathway
4 Myeloid cell homeostasis
83 Homeostasis
27 Platelet activation
43 Actin cytoskeleton organization
19 Integrin-mediated signaling pathway
30 RHO GTPase effectors
20 Chemokine signaling pathway
29 Myeloid cell differentiation
14 Regulation of blood coagulation
8 Gas transport
26 Regulation of metal ion transport


A
Erys
29 Hemostasis
7 Oxygen transport
10 Platelet activation
4 Erythrocytes take up oxygen and release carbon dioxide
5 Porphyrin-containing compound metabolic process
7 Erythrocyte differentiation
5 Iron ion homeostasis
11 Regulation of vesicle-mediated transport
0 5 10 15 20
B
LDB1 MYCL LMO2 MXD3 HES5 EHMT2 MED1 HES4 TAL1 KLF2 SUV39H1 YAP1 NFE2 HES1 HEY2 MAFK MAX KLF1 HELT MXD1 GATA1 TCF4 GFI1B MYC JAK2 MYCN CBFA2T3 NEUROG3 SSBP3 GATA6
C
LDB1 RHD EPB42 XRCC6 GYPB RCC1 HBB HBZ HBE1 GATA1 AHSP MED1 KLF2 ALAS2 TOP1 SSBP3 TAL1 RHAG TRRAP SLC4A1 DMTN KLF1 GATA6 ZFPM1 BRD1 LMO2 SSBP2 CBFA2T2 SPTB
D
HBE HBG HBB
Individual globin type / total (HBE+HBG+HBB) counts (%)
100% 80% 60% 40% 20% 0%
Individual cell
Relative expression of globins
2000 1500 1000 500 300 200 100 0.10 0.08 0.06 0.04 0.02 0.00
HBB HBG
EB D4
SC D14+3
SC D14+9
E
iPSC
EB D4
SC D14
Relative expression of genes
1.2 1 0.8 0.6 0.4 0.2 0
***
KDR FN1 CD34 CDH5 CD99 ESAM
EB D4
SC D14+3 EST-
SC D14+3 EST+
SC D14+9 EST-
SC D14+9 EST+
Relative expression of genes
4 3.5 3 2.5 2 1.5 1 0.5 0
ALAS2 GATA1 KLF1