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

3

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

6

7 ¹ CAS Key Laboratory of Genome Science and Information, Beijing Institute of
8 Genomics, Chinese Academy of Sciences, Beijing 100101, China

9 ² College of Life Sciences, University of Chinese Academy of Sciences, Beijing
10 100049, China

11 ³ Sino-Danish College, University of Chinese Academy of Sciences, Beijing 100190,
12 China

13 ⁴ Institute for Stem Cell and Regeneration, Chinese Academy of Sciences, Beijing
14 100101, China

15

16 [#] Equal contribution.

17 ^{*} Corresponding authors.

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

19

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

22

23 ^a ORCID: 0000-0001-9418-4848.

24 ^b ORCID: 0000-0002-0519-3935.

25 ^c ORCID: 0000-0002-9811-5302.

26 ^d ORCID: 0000-0001-9766-3334.

27 ^e ORCID: 0000-0002-8213-9166.

28 ^f ORCID: 0000-0003-0490-6507.

29 ^g ORCID: 0000-0002-6628-8620.

30

- 31 Total word counts: 6,288
- 32 Total figures: 6
- 33 Total supplementary figures: 7
- 34 Total supplementary tables: 7
- 35

36 **Abstract**

37 There is currently an imbalance between the supply and demand of functional red
 38 blood cells (RBCs) in clinical applications, and this imbalance can be addressed by
 39 regenerating RBCs with a variety of in vitro methods. Induced pluripotent stem cells
 40 (iPSCs) can address the low supply of cord blood and the ethical issues in embryonic
 41 stem cell research and provide a promising strategy to eliminate immune rejection.
 42 However, no complete single-cell level differentiation pathway exists for the
 43 iPSC-derived RBC differentiation system. In this study, we used iPSC line BC1 to
 44 establish an RBC regeneration system and used the 10× Genomics single-cell
 45 transcriptome platform to map the cell lineage and differentiation trajectories on day
 46 14 (D14) of the regeneration system. We found iPSC differentiation was not
 47 synchronized during embryoid body (EB) culture, and the D14 cells in the system
 48 mainly consisted of mesodermal and various blood cells, similar to yolk sac
 49 hematopoiesis. During asynchronous EB differentiation, iPSCs undergo three
 50 bifurcations before they enter erythroid differentiation, and the driver genes of each
 51 bifurcation were identified. The key roles of cell adhesion and estradiol in RBC
 52 regeneration were observed. This study provides systematically theoretical guidance
 53 for the optimization of the iPSC-derived RBC differentiation system.

54

55 **KEYWORDS:** scRNA-seq; iPSCs; Hematopoiesis; Erythropoiesis; Differentiation
 56 trajectory

57

58 Introduction

59 The substantial gap between the supply and demand of blood has always been a
60 serious problem in clinical practice [1]. Artificial blood is an important means of
61 solving this worldwide shortage, and obtaining functional red blood cells (RBCs) is
62 the key to artificial blood generation [2]. RBC regeneration *in vitro* has been an
63 important research direction in the blood research field for many years [3, 4]. In the
64 current study, the primitive materials that can be used for RBC regeneration are mainly
65 cord blood hematopoietic stem cells (HSCs) [5], embryonic stem cells (ESCs) [6, 7],
66 and induced pluripotent stem cells (iPSCs) [6, 8]. The regeneration technology of cord
67 blood HSCs is relatively mature, but there are problems that come with this technology,
68 such as difficulty in obtaining materials, significant differences among individuals,
69 high price of this technology, *etc.*, and ESCs are subject to ethical restrictions [9].
70 Human iPSCs can be differentiated into a variety of cell types *in vitro*, providing a
71 model for basic research and a source of clinically relevant cells [10-12]. Although the
72 derivation of iPSC lines for patients has now become a routine technique, how to
73 accurately differentiate iPSCs into the desired cell types is still a major challenge [13,
74 14].

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

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

103 Based on the existing molecular mechanisms of hematopoietic development and
 104 erythroid differentiation, scientists simulate embryonic hematopoiesis [36] used the
 105 spin EB method to regenerate HSPCs *in vitro*, and then induced the differentiation of
 106 RBCs from iPSCs. However, the differentiation efficiency and denucleation rate of the
 107 existing system is low, and the expression of adult globin is limited; these drawbacks
 108 limit the clinical availability [37]. At the same time, iPSC differentiation is a complex
 109 process, and flow cytometry and immunostaining have been used to determine the cell
 110 types during iPSC differentiation culture. However, these methods are limited by the
 111 number of fluorescent probes used, and it is not possible to solve important problems,
 112 such as the cell composition and differentiation path of the iPSC differentiation system,
 113 under high-resolution conditions. Recently, single-cell transcriptome technology has
 114 been able to efficiently capture and identify rare and transient cell types, determine the
 115 spatial or temporal localization of cells, and reconstitute gene regulatory networks,
 116 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.

142

143 **Results**

144 **iPSC cell line-derived HSPC production and erythropoiesis**

145 Depending on the spin EB culture method, we used the iPSC cell line BC1 to collect
 146 the suspension cells (SCs) through the cell strainer in the system after 14 days of culture
 147 [42, 43] (**Figure 1A**). The established culture system can effectively produce SCs. The
 148 cell count analysis showed that the number of SCs on day 14 (D14) was approximately
 149 10-fold higher than the number of iPSCs on D0 (Figure 1B). Flow cytometry analysis
 150 showed that approximately $25.9 \pm 3.05\%$ of the SCs expressed CD34, and
 151 approximately $27.08 \pm 1.63\%$ of the SCs expressed CD45, indicating that we collected a
 152 certain percentage of HSPCs from the SCs (Figure 1C). We performed colony-forming
 153 unit (CFU) assays on the collected SCs on D14. After 14 days of culture, blood lineage
 154 colonies were observed under the microscope, indicating that the SCs have the potential
 155 to differentiate into various blood lineage cells; and the CFU-GEMM colony showed a
 156 slight rust red color, indicating that the erythroid cell colonies begin to express
 157 hemoglobin (Figure 1D). Subsequently, we attempted to induce HSPCs in SCs for
 158 erythroid differentiation. By assessing the changes in the expression of the cell surface
 159 markers CD71 and CD235a at different time points (Figure 1E), we found that on D14
 160 of the EB stage, the proportion of CD71 and CD235a double-positive cells in SCs
 161 exceeded 30%, indicating that some cells in the system entered the erythroid
 162 differentiation stage. On D28, more than 95% of cells expressed both CD71 and
 163 CD235a (Figure 1E). Through Wright-Giemsa staining analysis, we observed that most
 164 of the cells were at the end of erythroid differentiation, and some cells were denucleated
 165 RBCs. In summary, we successfully obtained HSPCs by spin EB technology used the
 166 iPSC cell line BC1 and induced HSPCs to enter the erythroid differentiation process,
 167 eventually producing denucleated RBCs.

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

169 In this system, D14 is the intermediate time point. At this time, the sample contains
 170 both EBs and SCs, and the SCs contain a certain proportion of CD34+, CD45+ and
 171 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 2A**, 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 *SI00A8*); 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, podosome, 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 3A 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 *SI00A8*) [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 4A 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 ($P = 6.31\text{E-}34$) and heart development ($P = 2.14\text{E-}28$). The EPCs cluster could further differentiate into blood vessel and blood cells, and the corresponding genes were significantly enriched in blood vessel development ($P = 1.35\text{E-}46$) and regulation of the MAPK cascade ($P = 6.92\text{E-}21$). Many genes were also significantly enriched in the pathways related to cell adhesion, such as KEGG pathway focal adhesion ($P = 3.31\text{E-}22$). And the blood cells underwent hematopoietic differentiation, the corresponding genes were enriched in myeloid leukocyte activation ($P = 1.20\text{E-}97$), hemostasis ($P = 2.69\text{E-}28$), cytokine production ($P = 2.04\text{E-}24$) 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

351 HSPCs cluster driven by GATA2 and RUNX1 was further differentiated into a blood
352 lineage cells cluster under the induction of lineage-specific TFs, such as SPI1 and
353 SOX4 (**Figure 5A**). Functional annotation of genes in each cluster showed that the
354 genes in mesodermal cells were riched in skeletal system development ($P = 5.01E-21$)
355 and ossification ($P = 1.35E-13$). The blood vessel cluster genes were enriched in
356 angiogenesis ($P = 6.17E-09$), and the HSPCs cluster genes were enriched in ribosome
357 assembly ($P = 6.92E-26$) and *RUNX1* regulates transcription of genes involved in
358 differentiation of HSCs ($P = 3.55E-05$). The blood lineage cells cluster genes were
359 enriched in lymphocyte activation ($P = 2.34E-13$) and myeloid cell differentiation (P
360 $= 5.62E-09$).

361 To identify the driving factors for the differentiation of lymphoid and myeloid, we
362 constructed a differentiation trajectory for the HSPCs cluster and used 336 genes to
363 generate a pseudotime heatmap to show the gene expression dynamics of these
364 branches. Cells in the HSPCs cluster were divided into four clusters: multipotent
365 progenitors (MPPs), common lymphoid progenitors (CLPs), GMPs, and
366 megakaryocyte-erythroid progenitors (MEPs) (Figure 5B and C). In MPPs, the genes
367 *GATA2*, *RUNX1* and *CD34* were highly expressed, and functional annotation revealed
368 that the genes were significantly enriched in ribosome biogenesis ($P = 1.70E-14$) and
369 hematopoiesis ($P = 9.55E-04$). The CLPs cluster that driven by TFs such as IKZF1
370 and EGR1 was highly expressed *JAML*, *AIF1* and other lymphoid characteristic
371 genes, and functional annotation revealed a significant enrichment for lymphocyte
372 activation ($P = 1.95E-55$) and cytokine production ($P = 8.51E-32$). GMPs and MEPs
373 are both myeloid cells, under SPI1 and GATA1 induction, myeloid-specific genes
374 such as *LYZ*, *CLC*, *CD36* and *KLF1* were highly expressed. Functional annotation
375 revealed that the enriched genes were involved in myeloid leukocyte activation ($P =$
376 $1.58E-48$) and hemostasis ($P = 5.13E-25$) (Figure 5C, D and S6).

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

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

391 **Pseudotime trajectory of erythroid cells in the iPSC-derived RBC differentiation** 392 **system**

393 To accurately elucidate the molecular activities in erythroid differentiation, we
 394 performed pseudotime analysis of MEs cluster alone. Simultaneously, pseudotime
 395 heatmap analysis with 200 genes showed that cells could be divided into two
 396 subclusters, marked as MEPs and erythroid cells (Erys). The TF GATA1 was highly
 397 expressed in the MEPs cluster and Erys cluster highly expressed of the erythrocyte
 398 characteristic TF KLF1, while high expression of *GFL1B*, *NFE2*, *MAX* and other
 399 genes played an important role in the differentiation and maturation of RBCs (**Figure**
 400 **6A, B**). At the same time, the pseudotime trajectory analysis of the characterizing
 401 genes showed that in the process of erythroid differentiation, the expression levels of
 402 *CD34* and *TIE1* (endothelial marker genes) were downregulated and decreased
 403 rapidly during the differentiation process. *GATA1* was closely related to the early
 404 differentiation of erythroid cells and was rapidly upregulated during the
 405 differentiation process but was rapidly downregulated at the end of differentiation;
 406 consistent with previous studies, *TALI* and *GATA1* also showed similar trends [62].
 407 *KLF1* and *ALAS2* were also key factors during erythroid differentiation, and the
 408 expression level gradually increased with the differentiation process; *HBA1* and *HBB*
 409 encoded adult-type globin, *HBA1* was highly expressed throughout the erythroid
 410 differentiation process, and the expression levels of both *HBA1* and *HBB* gradually

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

434 Our results showed that iPSCs entered erythroid differentiation, undergo two
 435 transformations, exit pluripotency, and high expression of *ETV2* and *KDR* under
 436 BMP4 and VEGF stimulation to induce cells to differentiate into hemangioblasts; this
 437 process is accompanied by cell adhesion activation [45, 55]. In addition, HSPCs enter
 438 the erythroid differentiation process accompanied by the loss of cell adhesion and the
 439 fluctuation of estradiol target genes expression. During this process, the HSPCs would
 440 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

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

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

480 By studying the differentiation trajectory of the iPSC-derived RBC differentiation
481 system, we found that the iPSCs differentiated enter hematopoietic lineage under the
482 driving force of *ETV2* and underwent three bifurcations, then enter erythroid
483 differentiation process. During the process of the bifurcation of hematopoiesis and
484 angioblasts, the cells enter the hematopoietic lineage driven by TFs, such as GATA2
485 and RUNX1; in the process of bifurcation of the myeloid and lymphoid lines, the cells
486 enter the myeloid lineage driven by TF SPI1; and during the process the bifurcation of
487 the GMPs and MEPs, the cells enter the megakaryocyte-erythroid lineage driven by
488 GATA1, and then erythrocytes are formed under the driving force of KLF1. These
489 series of fate-determining TFs can effectively guide the regulation of iPSC-derived
490 RBC differentiation process and provide estimable advice for cell regeneration in
491 other blood lineages. In addition, pseudotime clustering analysis also suggested that
492 the pluripotent stem cells to erythroid cell production underwent two major
493 transformations, including *ETV2*-directed withdrawal of iPSC pluripotency into the
494 hematopoietic lineage, followed by GATA1-directed HSPC erythroid differentiation.
495 We noted that cell adhesion molecules are actively expressed during the
496 transformation one process but are downregulated during transformation two, possibly
497 related to cell adhesion for the formation of the gastrula [55]. We found that during
498 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 μ m 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 μ g/ml stem cell factor (SCF) (PeproTech), 100 μ g/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 5×10^5 /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% CO₂, 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 2×10^6 , and the total volume of the medium per well was 2 ml. The cells were cultured at 37 °C under 5% CO₂, 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, $\sim 1 \times 10^5$ 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

$\sim 1 \times 10^5$ 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

557 45-degree angle to spread the cells evenly. The pushed film was placed in a 37 °C
 558 incubator for 1 to 3 hours. Then, 200 µL of Wright-Giemsa staining A solution
 559 (BASO, BA-4017) was placed onto the cell-containing slide, the slide was placed at
 560 room temperature for 1 minute and was immediately covered with 400 µL of B
 561 solution. After the slide was incubated at room temperature for 8 minutes, the slides
 562 were gently rinsed with fluid distilled water along the area around the slide without
 563 cells. The slides were rinsed for an additional 1 minute until the water ran clear and
 564 transparent. Then, the slides were placed at room temperature until dry, and a
 565 microscope was used to observe and record cell morphology.

566 **Total RNA extraction and qPCR**

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

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

574 D14 EBs were dissociated with Accutase solution (Sigma). Single-cell suspensions at
 575 800 cells/µL were subjected to Chromium 10× Genomics library construction. Raw
 576 gene expression matrices generated by CellRanger (version 2.0.0) were imported into
 577 Seurat (version 2.4) [44]. After removing genes expressed in fewer than 5 cells, in
 578 cells with more than 5% mitochondria reads or in cells with less than 200 unique
 579 genes, two Seurat objects were integrated and clustered into 10 and 14 clusters based
 580 on fifteen principal components with the resolution parameter set at 0.2 and 0.8. Each
 581 cluster was identified according to the marker genes found by the Seurat
 582 “FindAllMarkers” function. Particular clusters were imported into Monocle2 [57] for
 583 pseudotime ordering. For bulk RNA-seq data, FASTQC and Trimmomatic were used
 584 to remove adaptor sequences and low-quality reads. Then, clean data were mapped to
 585 the reference genome (GRCh38) by HISAT2 [69] to generate a gene expression
 586 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.ccb.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

658 corresponds to in vivo waves of hematopoietic cell generation. *Stem Cell Reports*
659 2018;10:151-65.

660 [16] Koyano-Nakagawa N, Kweon J, Iacovino M, Shi X, Rasmussen TL, Borges L, et
661 al. Etv2 is expressed in the yolk sac hematopoietic and endothelial progenitors and
662 regulates lmo2 gene expression. *Stem Cells* 2012;30:1611-23.

663 [17] Kataoka H, Hayashi M, Nakagawa R, Tanaka Y, Izumi N, Nishikawa S, et al.
664 Etv2/er71 induces vascular mesoderm from flk1+pdgfralpha+ primitive mesoderm.
665 *Blood* 2011;118:6975-86.

666 [18] Kirmizitas A, Meiklejohn S, Ciau-Uitz A, Stephenson R, Patient R. Dissecting
667 bmp signaling input into the gene regulatory networks driving specification of the
668 blood stem cell lineage. *Proc Natl Acad Sci U S A* 2017;114:5814-21.

669 [19] Hubner K, Grassme KS, Rao J, Wenke NK, Zimmer CL, Korte L, et al. Wnt
670 signaling positively regulates endothelial cell fate specification in the flil1a-positive
671 progenitor population via lef1. *Dev Biol* 2017;430:142-55.

672 [20] Palis J. Primitive and definitive erythropoiesis in mammals. *Front Physiol*
673 2014;5:3.

674 [21] McGrath KE, Frame JM, Fromm GJ, Koniski AD, Kingsley PD, Little J, et al. A
675 transient definitive erythroid lineage with unique regulation of the beta-globin locus in
676 the mammalian embryo. *Blood* 2011;117:4600-8.

677 [22] Lie ALM, Marinopoulou E, Li Y, Patel R, Stefanska M, Bonifer C, et al. Runx1
678 positively regulates a cell adhesion and migration program in murine hemogenic
679 endothelium prior to blood emergence. *Blood* 2014;124:e11-20.

680 [23] de Pater E, Kaimakis P, Vink CS, Yokomizo T, Yamada-Inagawa T, van der
681 Linden R, et al. Gata2 is required for hsc generation and survival. *J Exp Med*
682 2013;210:2843-50.

683 [24] Huang K, Du J, Ma N, Liu J, Wu P, Dong X, et al. Gata2(-/-) human escs
684 undergo attenuated endothelial to hematopoietic transition and thereafter granulocyte
685 commitment. *Cell Regen (Lond)* 2015;4:4.

686 [25] Ivanovs A, Rybtsov S, Welch L, Anderson RA, Turner ML, Medvinsky A.
687 Highly potent human hematopoietic stem cells first emerge in the intraembryonic
688 aorta-gonad-mesonephros region. *J Exp Med* 2011;208:2417-27.

689 [26] Stratman AN, Davis MJ, Davis GE. Vegf and fgf prime vascular tube
690 morphogenesis and sprouting directed by hematopoietic stem cell cytokines. *Blood*
691 2011;117:3709-19.

692 [27] Guitart AV, Subramani C, Armesilla-Diaz A, Smith G, Sepulveda C, Gezer D, et
693 al. Hif-2alpha is not essential for cell-autonomous hematopoietic stem cell
694 maintenance. *Blood* 2013;122:1741-5.

695 [28] Upadhaya S, Sawai CM, Papalexi E, Rashidfarrokhi A, Jang G, Chattopadhyay
696 P, et al. Kinetics of adult hematopoietic stem cell differentiation in vivo. *J Exp Med*
697 2018;215:2815-32.

698 [29] Palii CG, Cheng Q, Gillespie MA, Shannon P, Mazurczyk M, Napolitani G, et al.
699 Single-cell proteomics reveal that quantitative changes in co-expressed
700 lineage-specific transcription factors determine cell fate. *Cell Stem Cell*
701 2019;24:812-20.

702 [30] Qiu C, Olivier EN, Velho M, Bouhassira EE. Globin switches in yolk sac-like
703 primitive and fetal-like definitive red blood cells produced from human embryonic
704 stem cells. *Blood* 2008;111:2400-8.

705 [31] Yu X, Azzo A, Bilinovich SM, Li X, Dozmorov M, Kurita R, et al. Disruption of
706 the mbd2-nurd complex but not mbd3-nurd induces high level hbf expression in
707 human erythroid cells. *Haematologica* 2019.

708 [32] Chou AC, Broun GO, Jr., Fitch CD. Abnormalities of iron metabolism and
709 erythropoiesis in vitamin e-deficient rabbits. *Blood* 1978;52:187-95.

710 [33] Anderson GJ, Powell LW, Halliday JW. Transferrin receptor distribution and
711 regulation in the rat small intestine. Effect of iron stores and erythropoiesis.
712 *Gastroenterology* 1990;98:576-85.

713 [34] Rutherford TR, Weatherall DJ. Deficient heme synthesis as the cause of
714 noninducibility of hemoglobin synthesis in a friend erythroleukemia cell line. *Cell*
715 1979;16:415-23.

716 [35] Ji P, Murata-Hori M, Lodish HF. Formation of mammalian erythrocytes:
717 Chromatin condensation and enucleation. *Trends Cell Biol* 2011;21:409-15.

718 [36] Murry CE, Keller G. Differentiation of embryonic stem cells to clinically
719 relevant populations: Lessons from embryonic development. *Cell* 2008;132:661-80.

720 [37] Mazurier C, Douay L, Lapillonne H. Red blood cells from induced pluripotent
721 stem cells: Hurdles and developments. *Curr Opin Hematol* 2011;18:249-53.

722 [38] Tang F, Lao K, Surani MA. Development and applications of single-cell
723 transcriptome analysis. *Nat Methods* 2011;8:S6-11.

724 [39] Han X, Chen H, Huang D, Chen H, Fei L, Cheng C, et al. Mapping human
725 pluripotent stem cell differentiation pathways using high throughput single-cell
726 rna-sequencing. *Genome Biol* 2018;19:47.

727 [40] Swiers G, Baumann C, O'Rourke J, Giannoulatou E, Taylor S, Joshi A, et al.
728 Early dynamic fate changes in haemogenic endothelium characterized at the
729 single-cell level. *Nat Commun* 2013;4:2924.

730 [41] Vanuytsel K, Matte T, Leung A, Naing ZH, Morrison T, Chui DHK, et al.
731 Induced pluripotent stem cell-based mapping of beta-globin expression throughout
732 human erythropoietic development. *Blood Adv* 2018;2:1998-2011.

733 [42] Cheng L, Hansen NF, Zhao L, Du Y, Zou C, Donovan FX, et al. Low incidence
734 of DNA sequence variation in human induced pluripotent stem cells generated by
735 nonintegrating plasmid expression. *Cell Stem Cell* 2012;10:337-44.

736 [43] Chou BK, Mali P, Huang X, Ye Z, Dowey SN, Resar LM, et al. Efficient human
737 ips cell derivation by a non-integrating plasmid from blood cells with unique
738 epigenetic and gene expression signatures. *Cell Res* 2011;21:518-29.

739 [44] Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of
740 single-cell gene expression data. *Nat Biotechnol* 2015;33:495-502.

741 [45] Koyano-Nakagawa N, Garry DJ. ETV2 as an essential regulator of mesodermal
742 lineage development. *Cardiovasc Res* 2017;113:1294-306.

743 [46] Tusi BK, Wolock SL, Weinreb C, Hwang Y, Hidalgo D, Zilionis R, et al.
744 Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature*
745 2018;555:54-60.

746 [47] Stansfield BK, Ingram DA. Clinical significance of monocyte heterogeneity. *Clin*
747 *Transl Med* 2015;4:5.

748 [48] Scialdone A, Tanaka Y, Jawaid W, Moignard V, Wilson NK, Macaulay IC, et al.
749 Resolving early mesoderm diversification through single-cell expression profiling.
750 *Nature* 2016;535:289-93.

751 [49] Pijuan-Sala B, Griffiths JA, Guibentif C, Hiscock TW, Jawaid W, Calero-Nieto
752 FJ, et al. A single-cell molecular map of mouse gastrulation and early organogenesis.
753 *Nature* 2019;566:490-5.

754 [50] Min IM, Pietramaggiore G, Kim FS, Passegue E, Stevenson KE, Wagers AJ. The
755 transcription factor *egr1* controls both the proliferation and localization of
756 hematopoietic stem cells. *Cell Stem Cell* 2008;2:380-91.

757 [51] Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al.
758 Single-cell rna-seq reveals new types of human blood dendritic cells, monocytes, and
759 progenitors. *Science* 2017;356.

760 [52] Nostro MC, Cheng X, Keller GM, Gadue P. Wnt, activin, and bmp signaling
761 regulate distinct stages in the developmental pathway from embryonic stem cells to
762 blood. *Cell Stem Cell* 2008;2:60-71.

763 [53] Metcalf D. Hematopoietic cytokines. *Blood* 2008;111:485-91.

764 [54] Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The human
765 transcription factors. *Cell* 2018;172:650-65.

766 [55] Adams JC, Watt FM. Regulation of development and differentiation by the
767 extracellular matrix. *Development* 1993;117:1183-98.

768 [56] Angelos MG, Abrahante JE, Blum RH, Kaufman DS. Single cell resolution of
769 human hematoendothelial cells defines transcriptional signatures of hemogenic
770 endothelium. *Stem Cells* 2018;36:206-17.

771 [57] Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The
772 dynamics and regulators of cell fate decisions are revealed by pseudotemporal
773 ordering of single cells. *Nat Biotechnol* 2014;32:381-6.

774 [58] Dzierzak E, Philipsen S. Erythropoiesis: Development and differentiation. *Cold*
775 *Spring Harb Perspect Med* 2013;3:a011601.

776 [59] Palis J, Robertson S, Kennedy M, Wall C, Keller G. Development of erythroid
777 and myeloid progenitors in the yolk sac and embryo proper of the mouse.
778 *Development* 1999;126:5073-84.

779 [60] Gao X, Johnson KD, Chang YI, Boyer ME, Dewey CN, Zhang J, et al. *Gata2*
780 *cis*-element is required for hematopoietic stem cell generation in the mammalian
781 embryo. *J Exp Med* 2013;210:2833-42.

782 [61] Downes M, Koopman P. *Sox18* and the transcriptional regulation of blood vessel
783 development. *Trends Cardiovasc Med* 2001;11:318-24.

784 [62] Rylski M, Welch JJ, Chen YY, Letting DL, Diehl JA, Chodosh LA, et al.
785 *Gata-1*-mediated proliferation arrest during erythroid maturation. *Mol Cell Biol*
786 2003;23:5031-42.

787 [63] To-Figueras J, Ducamp S, Clayton J, Badenas C, Delaby C, Ged C, et al. *Alas2*
788 acts as a modifier gene in patients with congenital erythropoietic porphyria. *Blood*
789 2011;118:1443-51.

790 [64] Fehling HJ, Lacaud G, Kubo A, Kennedy M, Robertson S, Keller G, et al.
791 Tracking mesoderm induction and its specification to the hemangioblast during
792 embryonic stem cell differentiation. *Development* 2003;130:4217-27.
793 [65] Dzierzak E, Bigas A. Blood development: Hematopoietic stem cell dependence
794 and independence. *Cell Stem Cell* 2018;22:639-51.
795 [66] Wilson NK, Göttgens B. Single-cell sequencing in normal and malignant
796 hematopoiesis. *HemaSphere* 2018;2.
797 [67] Yvernogeau L, Gautier R, Khoury H, Menegatti S, Schmidt M, Gilles JF, et al.
798 An in vitro model of hemogenic endothelium commitment and hematopoietic
799 production. *Development* 2016;143:1302-12.
800 [68] Liu Y, Wang Y, Gao Y, Forbes JA, Qayyum R, Becker L, et al. Efficient
801 generation of megakaryocytes from human induced pluripotent stem cells using food
802 and drug administration-approved pharmacological reagents. *STEM CELLS*
803 *Translational Medicine* 2015;4:309-19.
804 [69] Kim D, Langmead B, Salzberg SL. Hisat: A fast spliced aligner with low
805 memory requirements. *Nat Methods* 2015;12:357-60.
806 [70] Wang L, Feng Z, Wang X, Wang X, Zhang X. Degseq: An r package for
807 identifying differentially expressed genes from rna-seq data. *Bioinformatics*
808 2010;26:136-8.
809 [71] Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al.
810 Metascape provides a biologist-oriented resource for the analysis of systems-level
811 datasets. *Nat Commun* 2019;10:1523.
812 [72] Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al.
813 The genemania prediction server: Biological network integration for gene
814 prioritization and predicting gene function. *Nucleic Acids Res* 2010;38:W214-20.
815
816

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 $\pm$ sd). **C.** Percentage of CD34+ and CD45+ cells in the total SCs (mean $\pm$ 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 ($P < 0.05$, $-1 < \text{Log}_2 \text{FC} < 1$). **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

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

881

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

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

885

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

888

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

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

896

897 **Figure S5 iPSC differentiation process and DEG analysis results**

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

903

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

906

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

909 **A.** Functional annotation analysis of the Erys cluster in Figure 6B. Each functional
910 term was shown in Supplementary Table S6. **B, C.** The network shown the TFs in
911 Figure 6B (**B**) and genes of the erythrocyte differentiation functional term from A and
912 Figure 6D (**C**). **D.** The β -chain globin expression profile. **E.** qPCR of cell adhesion
913 molecules and the effects of estradiol throughout erythroid cell maturation.
914
915 **Table S1** Heatmap genes for Figure S1B and functional term for Figure S1C
916
917 **Table S2** Heatmap genes for Figure 2C
918
919 **Table S3** Functional terms for Figure S3, 3F and S5E
920
921 **Table S4** Genes used for heatmap of subclusters 0a, 0b, 0c, 1a, 1b, 4a, 4b, 8a and 8b
922
923 **Table S5** Genes used for Figure 4C, 5A, 5C, 5F and 6B
924
925 **Table S6** Functional terms for Figure 5A, 5C, 5F and 6B
926
927 **Table S7** qPCR primer for globin, cell adhesion and estradiol related genes

A

B

C

D

E

F

A

B

C

D

E

F

A

B

C

Cluster 0 1 2 3 4 5 6 7 8 9

Number of Gene

$-\log_{10}(p \text{ value})$

$-\log_{10}(p \text{ value})$

bioRxiv preprint doi: <https://doi.org/10.1101/858777>; this version posted November 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bioRxiv preprint doi: <https://doi.org/10.1101/859777>; this version posted November 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.