bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder.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.

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

Zijuan Xin [1,2,#], Wei Zhang[1,2,#], Shangjin Gong[1,2], Junwei Zhu[1,5], Yanming Li[1], Zhaojun Zhang[1,3,4,5*], Xiangdong Fang [1,2,3,4,5,*]

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

[5] *Beijing Key Laboratory of Genome and Precision Medicine Technologies, Beijing, 100101, China*

[#] Equal contribution.

[*] Corresponding authors.

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

Running title: *Xin ZJ et al / scRNA-seq analysis of iPSC-derived RBC differentiation system*

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder.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.

Abstract

There is an imbalance between the supply and demand of functional red blood cells (RBCs) in clinical applications. This imbalance can be addressed by regenerating RBCs using several *in vitro* methods. Induced pluripotent stem cells (iPSCs) can handle 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 a RBCs regeneration system. The 10× genomics single-cell transcriptome platform was used to map the cell lineage and differentiation trajectories on day 14 of the regeneration system. We observed that iPSCs differentiation was not synchronized during embryoid body (EB) culture. The cells (day 14) mainly consisted of mesodermal and various blood cells, similar to the yolk sac hematopoiesis. We identified six cell classifications and characterized the regulatory transcription factors (TFs) networks and cell-cell contacts underlying the system. iPSCs undergo two transformations during the differentiation trajectory, accompanied by the dynamic expression of cell adhesion molecules and estrogen-responsive genes. We identified different stages of erythroid cells such as burst-forming unit erythroid (BFU-E) and orthochromatic erythroblasts (ortho-E) and found that the regulation of TFs (e.g., TFDP1 and FOXO3) is erythroid-stage specific. Immune erythroid cells were identified in our system. This study provides systematic theoretical guidance for optimizing the iPSCs-derived RBCs differentiation system, and this system is a useful model for simulating *in vivo* hematopoietic development and differentiation.

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

Introduction

The substantial gap between the supply and demand for blood has always been a serious

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

clinical practice problem [1]. Artificial blood is an essential 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; however, some problems come with this technology, such as difficulty obtaining materials, significant differences among individuals, the high price of this technology, and ESCs are subject to ethical restrictions [9]. Human iPSCs can be differentiated into various 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 can differentiate into the 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 the 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

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

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 the differentiation and activation of HSCs, it is necessary to go through the cell fate determination program regulated by TFs GATA1, KLF1, and other TFs to enter the erythroid differentiation process [29]. The generation of RBCs is an elaborate 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] using the spin EB method to regenerate hematopoietic stem progenitor cells (HSPCs) *in vitro* and then induced the differentiation of RBCs from iPSCs. However, the differentiation efficiency and denucleation rate are low, and adult globin expression is limited; these drawbacks limit the clinical availability [37]. Simultaneously, iPSC differentiation is an elaborate 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 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* [39, 40]. By single-cell sequencing of embryoid bodies, researchers found that naïve H9 ESCs have a stronger hematopoietic capacity than primer H9 ESCs [41] and illuminated the heterogeneity of pluripotent

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

stem cell-derived endothelial cell differentiation [42]. Single-cell sequencing of day 29 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 [43]. The lack of a complete single-cell transcriptome map, as iPSCs differentiate into RBCs, does not allow scientists to guide iPSCs to 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 two transformations, during the first transformation, EHT, withdraw from pluripotency into lateral plate mesoderm and differentiate into hematopoietic endothelial cells and then into HSPCs, accompanied by cell adhesion molecules expression fluctuation. During the second transformation, HSPCs differentiate into erythroid progenitor cells and erythroid differentiation, estradiol promotes erythroid progenitor cell proliferation and inhibits erythroid differentiation. This study has an 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 d of EB culture [8, 44, 45] (**Figure 1**A). The established culture system can effectively produce SCs. Cell count analysis showed that the number of SCs on day 14 was approximately 10-fold higher than the number of iPSCs on day 0 (Figure 1B). Flow cytometry analysis showed that approximately 25.9% ± 3.05% of the SCs expressed cell surface marker CD34, approximately 27.08% ± 1.63% of the SCs expressed cell surface marker CD45, and approximately 22.9% of the SCs expressed both CD34 and CD45, indicating that

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

we collected a certain percentage of HSPCs from the SCs (Figure 1C). To detect hematopoietic differentiation's potency, we performed colony-forming unit (CFU) assays on the collected SCs from day 14. After two weeks of culture, differential blood lineage colonies were observed and counted under the microscope (Figure 1D and E). It suggested that the SCs have the potential to differentiate into various blood lineage cells. Subsequently, we attempted to induce HSPCs in SCs toward erythroid differentiation using serum-free medium (SFM) [8]. By assessing changes in expression of the cell surface markers CD71 and CD235a at different time points (Figure 1F), we found that on day 14 of the EB stage, the proportion of CD71 and CD235a double-positive cells in SCs exceeded 30%, indicating that a portion of cells in the system entered the erythroid differentiation stage. On day 28, more than 95% of cells expressed both erythrocyte markers, CD71 and CD235a (Figure 1F). qPCR assay showed that the expression of *HBB* and *HBG* increased significantly during the entire system differentiation (Figure 1G). Wright-Giemsa staining analysis showed that most cells were at the terminal erythroid differentiation stage and some cells were denucleated RBCs (Figure 1H). In summary, we successfully constructed an integrated *ex vivo* hiPSCs derived erythroid differentiation system by spin EB technology using the iPSC cell line BC1, as well as obtained and induced HSPCs to enter the erythroid differentiation process, eventually producing denucleated RBCs. Though we have constructed an erythroid cell differentiation system, we still do not understand what kind of molecular changes of the system's cells have undergone during the series of processes from pluripotent stem cells to erythroid cells, and what factors regulate the process of molecular changes experienced by cells. Therefore, to clarify the above problem, we collected cells that contain various cell types on day 14 at the middle time point of the differentiation system and performed single-cell transcriptome sequencing.

Cell composition of the iPSC-derived RBC differentiation system

To construct a scRNA-seq library, we isolated and selected EB cells (EBCs) and SCs on day 14, and constructed a scRNA-seq library based on 10× Genomics platform. After rigorous quality control analysis of the data by Seurat and the removal of double cells

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

droplets using R package DoubletFinder [46, 47], we obtained 3660 cells from batch 1 and 6538 cells from batch 2, given 10,198 high cells for downstream analysis (Figure S1A). On average, we detected 3523 informative genes and 17,825 unique molecular identifiers in each cell (Table S1). The correlation analysis of the expression pattern indicated that our 2 batch samples were positively correlated (Spearman coefficient = 0.904, $P < 0.001$) (Figure S1B). We then clustered cells based on Seurat's K-nearest neighbor method, annotated cells based on R package SingleR [48] and published datasets [49, 50], and visualized cells based on uniform manifold approximation and projection (UMAP) [51] (**Figure 2**A and B; Figure S2 and S3).

We classified cells into 6 categories from the cells belonging to starting cells or coming from different germ layers. iPS cells and ESCs were defined as starting cells (n = 1705); astrocytes, neuroepithelial cells, and neurons were classified as ectoderm-derived cells [52, 53] (n = 1529); epithelial and mesangial cells were classified as intermediate mesoderm-derived cells [54, 55] (n = 413); adipocytes, fibroblasts, myocytes, mesenchymal stem, skeletal muscle, smooth muscle, and tissue stem cells were classified as paraxial mesoderm-derived cells [56, 57] (n = 1934); endothelial cells and pericytes were classified as lateral plate mesoderm-derived cells [58] (n = 1561); $CD34^+$ HSPC, pro-myelocyte, myelocyte, macrophages, monocytes, neutrophils, $CD8^+$ T cells, burst-forming unit erythroid (BFU-E), colony-forming unit erythroid (CFU-E), erythrocytes, and orthochromatic erythroblasts (ortho-E) were classified as hematopoietic cells (n = 3056) (Figure 2C; Figure S4A).

To prove the cell classifications' results reliability, we identified the signature genes of each cell classification and calculated the Spearman coefficient between different cell classifications (Figure 2D; Figure S4B). It appears that the expression pattern of each cell classification was specific. We also performed cell cycle analysis based on the expression pattern of genes using Seurat and visualized the cell cycle analysis results into the UMAP dimension reduced map. We noted that among 10,198 cells, 2390 cells were in the S phase, 2076 cells were in the G2M phase, 5732 cells were in the G1 phase (Figure 2E). Specifically, we found that the cell cycle pattern was highly heterogeneous in different cell classifications, especially more than three-

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

quarters of the starting cells in the G2M/S phase (Pie chart in Figure 2E). Based on the signature genes of 6 cell classifications, we performed the Gene Ontology (GO) enrichment analysis using R package clusterProfiler to gain insight into the biological processes involved in the EB formation and cell differentiation process [59]. It is impressive that cell adhesion-related pathways were present in all cell classifications (Figure 2F).

ESAM affects hematopoiesis through cell-cell interactions

We identified that cell adhesion-related pathways may play an essential regulatory role during EB formation to hematopoiesis, and next analyzed the mode of cell adhesion between cells in our system. We used CellChat[60] to perform cell-cell contact analysis and evaluate the regulatory function of cell adhesion-related pathways in our system. We found that the contact score between the lateral plate mesoderm-derived cells, and other cells were the highest (**Figure 3**A), and the most significant cell-cell contact signal corresponding to the lateral plate mesoderm-derived cells was the contact between endothelial-specific adhesion molecule (ESAM) ligand and ESAM receptor (Figure 3B; Figure S5A and B). ESAM pathway is an autocrine signaling pathway, and ESAM was expressed and secreted by lateral plate mesoderm-derived cells, and then acted on itself and played a particular regulatory role in hematopoietic cells (Figure 3C and D). The lateral plate mesoderm-derived cells were mainly composed of endothelial cells (Figure 3D). Previous studies have shown that expression of the ESAM enhances the tight junctions between the arterial endothelium, which is not conducive for producing HSCs [61]. We performed bulk RNA-seq on SCs collected at day14 and the iPSC line BC1, and compared them with published HSPC transcriptome data from human cord blood [62]. We found that SCs are significantly different from iPSCs, and the hematopoietic lineage-related gene expression patterns are similar to HSPCs and exhibit HSPC-like characteristics (Figure S5C and D). Compared with BC1, SCs also down-regulated the expression of cell adhesion-related genes (Figure S5E), which may be the direct cause of SCs separation from EBs, and necessary demand for hematopoiesis.

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

Regulons networks in the iPSC-derived RBC differentiation system

To clarify regulons (i.e., TFs and their target genes) networks from EB formation to hematopoiesis in our system, we used the TFs included in the detected data to reduce dimension and mapped the SingleR cell annotation results to the new TFs dimension atlas. The R package, SCENIC, was used to calculate the area under the curve (AUC) scores of regulons [63]. The mapping data showed that the cell differentiation trajectory revealed by TFs is consistent with the clustering by signature genes (**Figure 4**A), and the same cell classification still clustered together. According to the AUC scores, regulons' activity has significant cell classification specificity (Figure 4B). Based on regulons' activity pattern, we constructed the interaction network of cell classification-specific regulons and their target genes based on Cytoscape. We observed that part of the target genes of TFs that were activated in the same cell classification was overlapped, suggesting that the regulation of TFs was synergistic in our system (Figure 4C). We performed the GO enrichment analysis for target genes of TFs by the clusterProfiler (Figure S6) and mapped the activation position of regulons in the UMAP atlas (Figure S7). Specifically, we observed that MECOM, ETS1, ELK3, ERG, and SOX18 showed high activities in lateral mesoderm-derived cell (Figure S6I, J, and S7), while GATA1, TAL1, RUNX1, CEBPD, MYB, and ELF1 were activated in hematopoietic cells (Figure S6K, L, and S7). CEBPD and MYB were mainly activated in myeloid cells and targeted to myeloid cell differentiation and cytokine production-related biological processes (Figure S6M, N, and S7). GATA1 and TAL1 were activated explicitly in erythroid cells, and participate in erythrocyte differentiation, homeostasis, Ras signaling transduction, and GTPase activity by regulating their target genes. At the same time, RUNX1 and ELF1 exhibited activities in both myeloid and erythroid cells and enriched in cell activation and leukocyte proliferation through their target genes.

To identify the TFs specific to the differentiation stage during hematopoietic differentiation, we calculated the regulon specificity score (RSS) for each cell type included in hematopoietic cells separately. We noticed that erythroid cells at different stages have a unique RSS pattern, suggesting that the regulation of TFs was erythroid-

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

stage specific (Figure S8). Especially in the erythrocytes cluster, we observed TFDP1 with the highest RSS that can heterodimerize with E2F proteins complex to control the transcriptional activity of numerous genes involved in cell cycle progression from G1 to S phase [64]. Wang et al. showed that the *kyo* mutant of TFDP1 could cause changes in the morphology and number of erythroid cells during the development of fish embryos, indicating that TFDP1 can be linked to the development of erythroid cells [65]. However, the regulatory role of TFDP1 in the development of mammalian erythroid cells has not been clarified. We speculate that TFDP1 may be related to cell number restriction during the maturation of erythroid cells.

Cell differentiation trajectory of the iPSC-derived RBC differentiation system

To further determine the cell differentiation trajectory during the differentiation of the entire system, we used Monocle2 to order single cells and construct an entire lineage differentiation trajectory with a tree-like structure [66] (**Figure 5**A). The results showed that cell differentiation originated from starting cells. After differentiation into different germ layers, two branches appeared. The hematopoiesis branch was mostly hematopoietic cells, and the non-hematopoiesis branch consisted of ectoderm, intermediate mesoderm, and other cells (Figure 5B; Figure S9A). We use a pseudo-time branch heatmap to show branch-specific expression patterns along two developmental pathways (hematopoiesis *vs*. non-hematopoiesis). Specifically, we identified 264 signature genes (cluster 1) for pre-branch, including 39 TF coding genes, such as *POU5F1* and *SOX2*, enriched by glycolysis/gluconeogenesis. Subsequently, pre-branch cells were driven by RUNX1, GATA2, and 22 other TFs to enter the hematopoiesis branch (cell fate 1), and 300 signature genes (cluster 2) were identified for hematopoiesis that were enriched in hematopoietic cell lineage pathway. Oppositely, pre-branch cells were driven by SOX7, SOX17, and 41 other TFs to enter non-hematopoiesis branch (cell fate 2), and 409 signature genes (cluster 2) were identified for non-hematopoiesis that were enriched in vascular smooth muscle contraction pathway (Figure 5C and D; Figure S9B). Several research teams have proved that RUNX1 and GATA2 mark the beginning of hematopoiesis [23, 67]. SOX7 activates

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

CDH5 promoter and hinders the binding of RUNX1 to DNA, and SOX17 mediates the repression of RUNX1 and GATA2, thereby hindering the transition to hematopoietic fate and maintaining the endothelial-arterial identity [67-69].

Meanwhile, Monocle2 divides cells into 3 states, pre-branch, hematopoiesis, and non-hematopoiesis branches corresponding to state 1, 2, and 3 (Figure 5E). Analysis of the expression trend of each branch signature genes in different state cells showed that each pseudo-time branch signature genes are highly expressed in its corresponding state cells (Figure 5F). The functional annotation analysis showed that state 1 cells (pre-branch) were enriched in the mitotic spindle, G2M checkpoint, mitotic prometaphase, indicating that it was in an active cell division state, and was regulated by P53, MYC, and mTOR pathways. State 2 cells (hematopoiesis branch) were enriched in heme metabolism, neutrophil degranulation, and other blood cell-related pathways, and are regulated by PI3K and IL6. State 3 cells (non-hematopoiesis branch) were enriched in angiogenesis, myogenesis, extracellular matrix organization, other mesodermal and mesenchymal cell-related pathways, and regulated by signal pathways, such as TGFβ and hypoxia (Figure 5G; Figure S9C).

To further understand the molecular characteristics of cells that entered the hematopoiesis branch, we analyzed the difference between state 2 initial cells (the first 100 cells that entered the hematopoiesis branch) and state 3 initial cells (the first 100 cells that entered the non-hematopoiesis branch) and found that state 2 initial cells had a low expression of *COL5A1*, *COL6A3*, and other collagen genes, and a high expression of *AIF1*, *SPI1*, and other genes related to myeloid differentiation (Figure S9D). The highly expressed genes retrieved from the TRRUST database are regulated by 20 TFs such as SPI1 and GATA1 (Figure S9E). At the same time, these TFs and their target genes are significantly enriched in embryonic hemopoiesis, erythroid differentiation, and myeloid cell differentiation, suggesting that our system is similar to early embryos hematopoiesis and tend to produce erythroid and myeloid cells, and were regulated by Notch signaling pathway and estrogen (Figure S9F).

Transition from starting cell to hematopoietic progenitor cell

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

Previous studies have shown that HSPCs are differentiated from endothelial progenitor cells in the lateral mesoderm during embryonic development [17]. To determine how the transition from starting cells to hematopoietic cells, we conducted the pseudo-time analysis of ESCs, iPS cells, endothelial cells, and $CD34^+$ HSPCs. Most of the iPS cells and ESCs fell into the pre-branch and undergo bifurcation. The cell fate 1 branch was mostly $CD34^+$ HSPCs, and the cell fate two branch was mostly endothelial cells (**Figure 6**A and B; Figure S10A). Differential gene expression analysis was performed between the 2 branches to determine the characteristic genes in each fate cell. Interestingly, we identified hematopoietic endothelium (HE) gene set (cluster 5), including 10 TF coding genes expressed on both endothelial cells and HSPCs branch. The TF coding genes *LMO2* and *FLI1* in this gene set are the marker genes of HE during embryonic development [70, 71] . FOS, JUNB, and DAB2 are the downstream TFs of TGFβ, and TGFβ drives HE programming [72, 73] . FOS and MAP4K2 are the effect factors in the MAPK cascade that VEGF can directly regulate to establish arterial specifications of endothelial progenitor cells [74]. The gene set (cluster 5) was observed in cell adhesion molecules and tight junctions (Figure 6C).

HSPCs branch corresponds to state 2 cells, endothelial cells branch corresponds to state 3 cells, and HE gene set was highly expressed in state 2 and state 3 cells (Figure S10B and C). At the same time, we found that some $CD34^+$ HSPCs fell in state 3, which was defined as HECs. By comparing with $CD34^+$ HSPCs on state 2, we found that HECs highly expressed *KDR*, *SOX17*, and other endothelial characteristic genes and highly expressed genes were enriched in cell adhesion molecules and regulated by RAS, NOTCH, and VEGF (Figure 6D and E). We verified the expression of some cell adhesion molecules by qPCR assay combined with previous function annotation results for cell classifications marker genes and cell-cell contact pattern analysis. Results indicated that the expression of these molecules fluctuated during RBC regeneration; their expression is up-regulated during the HE formation [75] (EB day 4), while their expression is down-regulated in hematopoietic cells (SCs day 14) (Figure S10D).

Transition from HSPC to the erythroid cell

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

To clarify the transition from HSPCs to the erythroid cells, we conducted a pseudo-time analysis of hematopoietic cells. Hematopoietic cells are mostly composed of $CD34^+$ HSPCs, erythroid cells, monocyte-macrophages, and a small amount of lymphoid $CD8^+$ T cells (Figure S11A and B). The differentiation trajectory shows fewer cells on the pre-branch, and most of the cells are ordered to the erythropoiesis and non-erythropoiesis branches (**Figure 7**A and B). Specifically, some BFU-E, CFU-E cells, and $CD34^+$ HSPCs are distributed in the pre-branch corresponding to state 4. Non-erythroid cells (eosinophils, macrophages, monocytes, myelocytes, neutrophils, and pro-myelocytes) are mostly distributed in the non-erythropoiesis branch (cell fate 1) corresponding to state 3, and erythroid cells (BFU-E, CFU-E, erythrocytes, ortho-E) are mostly distributed in the erythropoiesis branch (cell fate 2) corresponding to state 1, 2, and 5 (Figure 7C and Figure S11C). Simultaneously, state 4 cells highly express MYB, CD34, CD45 (PTPRC), and CD44, which are yolk sac-derived myeloid-biased progenitors (YSMPs) markers, YSMPs are human yolk sac second wave hematopoietic progenitor cells. And state 4 cells almost no expression of *HOX6*， which is HSPC specific marker in the definitive hematopoietic AGM region [76], and EMP (mouse yolk sac second wave hematopoietic progenitor) marker *KIT* [77] (Figure 7D), suggesting that our system is similar to the second wave of hematopoiesis in the human yolk sac.

To further clarify the molecular regulation mode in the differentiation process of YSMP in the system, we performed a difference map between the 2 branches (erythropoiesis and non-erythropoiesis), YSMPs (pre-branch) were driven by TAL1, GATA1, and 12 other TFs enter the erythropoiesis branch (cell fate 2), oppositely, YSMPs were driven by CEBPD, MAF and other 22 TFs enter to non-erythropoiesis branch (cell fate 1). YSMPs (Pre-branch) identified 341 signature genes (cluster 2), enriched by oxidative phosphorylation. Interestingly, the gene set has a strong tendency for expression in erythropoiesis branch, which may be EPO driven in the system (Figure 7E and Figure S11D and E). Besides, we found that some of the erythroid cells will fall into the non-erythropoiesis branch, indicating other types of erythroid cell production in our system (Figure S11C).

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

Differentiation trajectory of erythroid cells

To accurately understand the molecular activities during erythroid differentiation, we took BFU-E, CFU-E, erythrocytes, ortho-E cells to perform a pseudo-time analysis (**Figure 8**A). The pseudo-time result was consistent with the SingleR cell annotation result: the erythroid progenitor cells were mostly distributed at the beginning of the differentiation trajectory, and the differentiated terminal cells were mainly at the end of the trajectory (Figure 8B; Figure S12A). Furthermore, cluster analysis of the top 1000 genes with the pseudo-time scoring significance was performed to obtain the highly expressed gene set cluster 1 of erythroid progenitor cells, including the GATA1 and 36 other TFs, which were enriched in the estrogen signaling pathway. The expression trend of genes enriched in the process of erythroid differentiation is mostly initially high and then decreased. The gene set cluster 2, which was highly expressed in the erythroid cells, includes 34 TFs, such as BCL11A, which enriched cholesterol metabolism, glycosaminoglycan binding, and oxygen carrier activity (Figure 8C; Figure S12B and C). The pseudo-time analysis divides cells into 3 states; the progenitor cells correspond to state 1 in which the cluster 1 gene set was highly expressed, the erythroid cells in state 2 and three in which the cluster 2 gene set was highly expressed (Figure 8D; Figure S12D). Specifically, state 1 (erythroid progenitor cells) responds more strongly to estrogen (Figure 8E). We tested the role of estrogen by adding estradiol during the induction and differentiation process of SC cells and observed that expression of *KLF1* slightly increased in the erythroid progenitor cells, but the expression of *ALAS2*, *GATA1*, and *KLF1* in the terminal erythroid differentiation was significantly inhibited (Figure S12E). Altogether, it suggested that estrogen promotes the proliferation of erythroid progenitor cells, but may inhibit erythroid differentiation (Figure S12E).

Interestingly, we found that state 2 cells highly express *CD74* and lowly express *EPOR*. Compared with state 3, state 2 lowly expresses genes (e.g., *HBD* and *HBE1*) that related to oxygen carrier activity, oxygen binding and highly expresses genes (e.g., *CD81*, *C1QBP*, and other myeloid characteristic genes) that related to MHC protein complex binding genes, which can be defined as immune erythroid cells (Figure 8F and

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

G; Figure S12F). Also, we performed a difference analysis between ortho-E and other erythroid cells (Figure S12G). Ortho-E cells had low expression of genes that are related to the ribosome (e.g., *RPS16* and *RPS13*), oxidative phosphorylation, and cell cycle. In contrast, having high expression genes that associated with PI3K-Akt and FoxO signaling pathway, among which *FOXO3* plays a crucial role in the enucleation of erythroid cells [78] (Figure 6H; Figure S12G and H).

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 focused on bone marrow, peripheral blood, cord blood, and other systems [79, 80] . Here, we used the EBs culture method verified by many laboratories to construct the RBC regeneration system derived from iPSCs [8, 44, 45, 81]. The erythroid differentiation efficiency of iPSCs generated from different type of tissues might be different, however, these iPSCs cell lines can differentiate towards erythroid cells under appropriate inducing conditions [82]. After multiple induction and differentiation experiments under the same conditions, it is clear that HSPCs in the SCs produced by the EBs can stably enter the erythroid differentiation process (data not shown). After clarifying the repeatability and stability of the system, we performed a single-cell transcriptomics study on the RBC regeneration system derived from iPSCs, in which over 10, 000 high-quality cells of the regeneration system from two batches were analyzed. The two batches of cells are positively correlated, and the mesoderm and hematopoietic cells are dominantly present in each batch. To avoid the in-between cell types in the regeneration system, we applied four datasets that are integrated by the R package SingleR, including Human Primary Cell Atlas reference dataset, BLUEPRINT and Encode reference dataset, Database of Immune Cell Expression reference dataset and Novershtern hematopoietic reference dataset, and two public erythroid cell reference datasets, to carefully characterize the cell types in the iPSC erythroid differentiation system. The cell map and cell differentiation trajectory of the system

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

were mapped for the first time, and the whole transcriptome dynamics from iPSCs to the erythroid lineage were delineated.

This study observed that iPSCs went through two key transformations during EB formation to hematopoiesis and erythropoiesis. The first transformation was EHT, in which iPSCs differentiated into the germ layer under the induction of BMP4 and bFGF to form endothelial cells in the lateral plate mesoderm. Endothelial cells had strong cell-cell contact between themselves and other cells through the ESAM in the EHT. Under the induction of VEGF, HE was formed with the high expression of *FLI*, *LMO2*, and the cell adhesion molecules, and HE is converted to HSPCs with the up-regulated expression of *GATA2* under the influence of *RUNX1*, and the down-regulated expression of cell adhesion molecules. Reports showed that cell adhesion plays an essential regulatory function in mammalian embryonic development. Furthermore, cell-cell interactions were extremely indispensable in the information and the function of erythroblast islands, the niche where macrophages and immature erythrocytes were gathered and contacted to help erythrocytes maturation and enucleation during erythropoiesis [83–85].

The second transformation was HSPCs differentiation into erythroid cells. The HSPCs produced in the system were similar to $CD34^+$, $CD45^+$, and $CD44^+$ YSMP [86], and they tend to differentiate into erythroid cells under the induction of EPO. YSMP enters the erythroid differentiation pathway driven by GATA1. During this period, the expression of estrogen target genes was up-regulated at the beginning and then decreased. A recent study revealed that estrogen could promote self-renewal of HSCs, expand splenic HSCs, and promote erythropoiesis during pregnancy [87]. A small amount of immune $CD74^+$ erythroid cells are produced. Immune erythrocytes and their surface markers have been discovered and identified in recent years [88, 89] . This study found for the first time that the *in vitro* red blood cell regeneration system can also produce such cells, which provides a new direction for the clinical availability of RBCs regenerated *in vitro*. Also, we found that TFDP1 may be related to the restriction of cell number during the maturation of erythroid cells [65]. It was identified that *FOXO3* play a vital role in the denucleation of erythroid cells, and the high expression of *UPK1A*-

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

AS1 and *TRAPPC3L* in ortho-E may also regulate RBC denucleation-related genes.

Our results show that iPSCs go through two essential transformations to form erythroid cells in our system, accompanied by a complex TFs network and the dynamic expression of cell adhesion molecules and estrogen-responsive genes. It suggests that we can add cell adhesion activators before entering the first transformation, but add cell adhesion inhibitors during the transformation and inhibit endothelial TFs (e.g., SOX7) from increasing the production of HSPC in the system. Subsequently, we could activate erythroid TFs (e.g., GATA1) and add estradiol during the second transformation to increase the production of erythroid cells. This study provided a powerful theoretical guide for optimizing our iPSC-derived RBC regeneration system.

In summary, using the entire iPSC-derived RBC differentiation system in which both EB cells and SCs are included, we systematically characterized the molecular characteristics and differentiation trajectory of various cell types in the system. We provided essential factors driving iPSC differentiation into erythroid cells. However, technically it is not enough to improve the *in vitro* iPSC-derived RBC regeneration system at only one stage, and the RBC regeneration system needs to utilize different iPSC lines to prepare regenerative RBC of various blood types to suffice the needs of clinical blood transfusion under different blood types or different disease conditions in the future. Deciphering the RBC regeneration systems at different stages started from SCs in the EB culture system, we might identify more useful factors and pathways regulating adult globin expression and denucleation, which are the bottlenecks in this field. Moreover, we might characterize the real key regulators controlling RBC production process by comparing the *in vitro* iPSC-derived RBC regeneration system under physiological conditions (e.g., bone marrow). With the development of single-cell sequencing technology, we will be able to excavate more potential mechanisms regulating RBC development and maturation and obtain more mature regenerated functional RBCs *in vitro* for clinical applications in the future.

Materials and methods

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

iPSC and EB culture and erythropoiesis

iPSCs (sourced from Linzhao Cheng's lab) were cultured in vitronectin (Invitrogen, A14700, Carlsbad, CA, USA)-coated culture dishes using essential 8 medium (Gibco, A1517001, Carlsbad, CA, USA). When the cells reached 70% to 80% confluence (usually around 3 to 4 days), they were digested to SCs using 0.5 mM EDTA (Invitrogen, 15575020, Carlsbad, CA, USA) and passaged according to a ratio of 1 to 4.

EB culture was performed on 0-11 d, as previously described [8]. From day 11 to 14 of the EB culture, TPO was replaced by 3 U/mL erythropoietin (EPO) (PeproTech，100-64, Cranbury, NJ, USA). On day 14, SCs were collected using a 70 μm cell strainer. Subsequently, the SC solution was centrifuged at 300 *g* for 5 minutes to remove the supernatant and SFM containing 50 μg/ml stem cell factor (PeproTech, 300-07, Cranbury, NJ, USA), 100 μg/ml IL-3 (PeproTech, 200-03, Cranbury, NJ, USA), 2 U/ml EPO, and 1% penicillin/streptomycin (P/S) (Life Technologies, 15140122, Carlsbad, CA, USA) 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% CO_2, and the medium was changed every three days. On day 20, 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% CO_2, and the medium was changed every three days.

Colony formation unit assay

Incubation of day 14 SCs was performed using MethoCult™ H4434 classic methylcellulose medium for human cells (StemCell, 04434, Vancouver, B.C., Canada). Each well of a 12-well plate contained around 10,000 cells and 0.6 ml of the 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, approximately 1×10^5 cells were collected from the culture, centrifuged for five

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

minutes at 300 *g* to remove the supernatant, and washed twice with 2 ml 1× Dulbecco's phosphate-buffered saline (DPBS) (Gibco, 14190136, Carlsbad, CA, USA) solution containing 2% fetal bovine serum (Gibco, 16000044, Carlsbad, CA, USA) and 2 mM EDTA. After centrifugation the cells were resuspended in 50 μl DPBS buffer, and the antibodies (0.1 μl anti-CD71 and 0.5 μl anti-CD235a) were added, followed by incubation at 4°C for 10 min in the dark. The cells were washed twice by adding 2 ml of 1× DPBS buffer after incubation and resuspension 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 (v.7.6, Three Star). The antibodies (Miltenyi Biotec, Bergisch Gladbach, Germany) used in this study included anti-human CD34-FITC (130-098-142), anti-human CD45-PE (170-081-061), anti-human CD71-PE (130-099-219), and anti-human CD235a-APC (130-118-493).

Wright-Giemsa staining

Approximately 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 DPBS buffer. The cells were placed onto one of two clean slides, and the second slide was used to push the first to 30°-45° to spread the cells evenly. The pushed film was placed in a 37°C incubator for one to three hours. Then, 200 μL of Wright-Giemsa staining A solution (BASO, BA-4017, Zhuhai Guangdong, China) was placed onto the cell-containing slide, and the slide was placed at room temperature for one minute. It 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, day four EBs, day 14 SCs, and day 20 erythroid cells, and the total RNA was reverse transcribed into cDNA using the

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

PrimeScript RT Reagent Kit with gDNA Eraser (TaKaRa, RR047A, Kusatsu, Shiga, Japan). 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 Table S2.

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

Day 14 EBs were dissociated with Accutase solution (Sigma, A6964, St. Louis, Missouri, USA). Single-cell suspensions at 800 cells/μL were subjected to chromium 10× genomics library construction. Raw gene expression matrices generated by CellRanger (v.3.0.0) were imported into Seurat (v.3.2.2) [46]. After removing low quality cells and genes (percentage of mitochondria reads ≥ 25%, unique genes ≤ 200 in a cell, and genes expressed less than 5 cells), the left cells were annotated by SingleR and classified according to annotation results [48]. TFs regulatory network was constructed by SCENIC (v.1.1.2) [63], and cell contact patterns were constructed by CellChat (v.0.0.2) [60]. Particular clusters were imported into Monocle2 [66] for pseudo-time 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 to generate a gene expression matrix, which was then imported into R Package DEGseq to identify DEGs. Heatmap analysis and PCA were performed by pheatmap and PRINCOMP, respectively. We used clusterProfiler (v.3.18.0) to perform functional annotation analysis [59]. All data analyses were completed in R (v4.0.0) environment. The list of human TFs was downloaded from http://humantfs.ccbr.utoronto.ca/download.php. The Circular network was visualized by GeneMANIA [90].

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

CRediT author statement

Xinzi Juan: Investigation, Data curation, Formal analysis, Validation, Visualization,

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

Writing - original draft, Writing - review & editing. **Wei Zhang:** Investigation, Data curation, Formal analysis, Validation, Visualization, Writing - original draft, Writing - review & editing. **Shangjin Gong:** Formal analysis. **Junwei Zhu:** Validation. **Yanming Li:** Project administration. **Zhaojun Zhang:** Conceptualization, Resources, Writing - original draft, Writing - review & editing, Supervision. **Xiangdong Fang:** Conceptualization, Resources, Writing - original draft, Writing - review & editing, Supervision.

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, 2017YFC0907400, and 2018YFC0910700), and the National Natural Science Foundation of China (Grant Nos. 81670109, 81870097, 81700097, 81700116, and 82070114). We were grateful to Linzhao Cheng for agreeing to use iPSC line BC1 and Qianfei Wang for providing the iPSCs. We thank the editor Suzanne N in enago editing service who provided assistance for the manuscript.

ORCID

ORCID 0000-0001-9418-4848 (Xin, Z)

ORCID 0000-0002-0519-3935 (Zhang, W)

ORCID 0000-0002-9811-5302 (Gong, S)

ORCID 0000-0001-9766-3334 (Zhu, J)

ORCID 0000-0002-8213-9166 (Li, Y)

ORCID 0000-0003-0490-6507 (Zhang, Z)

ORCID 0000-0002-6628-8620 (Fang, X)

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

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+PDHFRalpha+ primitive mesoderm. Blood 2011;118:6975–86.

[18] Kirmizitas A, Meiklejohn S, Ciau-Uitz A, Stephenson R, Patient R. Dissecting BMP signaling input

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

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.

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

[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] Karamitros D, Stoilova B, Aboukhalil Z, Hamey F, Reinisch A, Samitsch M, et al. Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. Nat Immunol 2018;19:85–97.

[40] Pellin D, Loperfido M, Baricordi C, Wolock SL, Montepeloso A, Weinberg OK, et al. A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. Nat Commun 2019;10:2395.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, et al. Comprehensive integration of single-cell data. Cell 2019;177:1888–902 e21.

[47] McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. Cell Syst 2019;8:329–37 e4.

[48] Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol 2019;20:163–72.

[49] Hu J, Liu J, Xue F, Halverson G, Reid M, Guo A, et al. Isolation and functional characterization of human erythroblasts at distinct stages: Implications for understanding of normal and disordered erythropoiesis in vivo. Blood 2013;121:3246–53.

[50] Li J, Hale J, Bhagia P, Xue F, Chen L, Jaffray J, et al. Isolation and transcriptome analyses of human erythroid progenitors: BFU-E and CFU-E. Blood 2014;124:3636–45.

[51] Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using umap. Nat Biotechnol 2018.

[52] Verkhratsky A, Nedergaard M. Physiology of astroglia. Physiol Rev 2018;98:239–389.

[53] Horie R, Hazbun A, Chen K, Cao C, Levine M, Horie T. Shared evolutionary origin of vertebrate neural crest and cranial placodes. Nature 2018;560:228–32.

[54] Lam AQ, Freedman BS, Morizane R, Lerou PH, Valerius MT, Bonventre JV. Rapid and efficient differentiation of human pluripotent stem cells into intermediate mesoderm that forms tubules expressing kidney proximal tubular markers. J Am Soc Nephrol 2014;25:1211–25.

[55] Araoka T, Mae S, Kurose Y, Uesugi M, Ohta A, Yamanaka S, et al. Efficient and rapid induction of human iPSCs/ESCs into nephrogenic intermediate mesoderm using small molecule-based differentiation

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

methods. PLoS One 2014;9:e84881.

[56] Chal J, Al Tanoury Z, Oginuma M, Moncuquet P, Gobert B, Miyanari A, et al. Recapitulating early development of mouse musculoskeletal precursors of the paraxial mesoderm in vitro. Development 2018;145.

[57] Buckingham M, Rigby PW. Gene regulatory networks and transcriptional mechanisms that control myogenesis. Dev Cell 2014;28:225–38.

[58] Slukvin, II, Kumar A. The mesenchymoangioblast, mesodermal precursor for mesenchymal and endothelial cells. Cell Mol Life Sci 2018;75:3507–20.

[59] Yu G, Wang LG, Han Y, He QY. Clusterprofiler: An R package for comparing biological themes among gene clusters. OMICS 2012;16:284–7.

[60] Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun 2021;12:1088.

[61] Zhang C, Lv J, He Q, Wang S, Gao Y, Meng A, et al. Inhibition of endothelial ERK signalling by Smad1/5 is essential for haematopoietic stem cell emergence. Nat Commun 2014;5:3431.

[62] Ding N, Xi J, Li Y, Xie X, Shi J, Zhang Z, et al. Global transcriptome analysis for identification of interactions between coding and noncoding RNAs during human erythroid differentiation. Front Med 2016;10:297–310.

[63] Aibar S, Gonzalez-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: Single-cell regulatory network inference and clustering. Nat Methods 2017;14:1083–6.

[64] Girling R, Partridge JF, Bandara LR, Burden N, Totty NF, Hsuan JJ, et al. A new component of the transcription factor DRTF1/E2F. Nature 1993;362:83–7.

[65] Taimatsu K, Takubo K, Maruyama K, Suda T, Kudo A. Proliferation following tetraploidization regulates the size and number of erythrocytes in the blood flow during medaka development, as revealed by the abnormal karyotype of erythrocytes in the medaka TFDP1 mutant. Dev Dyn 2015;244:651–68.

[66] Qiu X, Hill A, Packer J, Lin D, Ma YA, Trapnell C. Single-cell mrna quantification and differential analysis with Census. Nat Methods 2017;14:309–15.

[67] Costa G, Mazan A, Gandillet A, Pearson S, Lacaud G, Kouskoff V. SOX7 regulates the expression of VE-cadherin in the haemogenic endothelium at the onset of haematopoietic development. Development 2012;139:1587–98.

[68] Clarke RL, Yzaguirre AD, Yashiro-Ohtani Y, Bondue A, Blanpain C, Pear WS, et al. The expression of Sox17 identifies and regulates haemogenic endothelium. Nat Cell Biol 2013;15:502–10.

[69] Lilly AJ, Costa G, Largeot A, Fadlullah MZ, Lie ALM, Lacaud G, et al. Interplay between SOX7 and RUNX1 regulates hemogenic endothelial fate in the yolk sac. Development 2016;143:4341–51.

[70] Liu F, Walmsley M, Rodaway A, Patient R. Fli1 acts at the top of the transcriptional network driving blood and endothelial development. Curr Biol 2008;18:1234–40.

[71] Stanulovic VS, Cauchy P, Assi SA, Hoogenkamp M. LMO2 is required for TAL1 DNA binding activity and initiation of definitive haematopoiesis at the haemangioblast stage. Nucleic Acids Res 2017;45:9874–88.

[72] Lam J, van den Bosch M, Wegrzyn J, Parker J, Ibrahim R, Slowski K, et al. miR-143/145 differentially regulate hematopoietic stem and progenitor activity through suppression of canonical TGFbeta signaling. Nat Commun 2018;9:2418.

[73] Monteiro R, Pinheiro P, Joseph N, Peterkin T, Koth J, Repapi E, et al. Transforming growth factor beta drives hemogenic endothelium programming and the transition to hematopoietic stem cells. Dev Cell 2016;38:358–70.

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

[74] Obier N, Cauchy P, Assi SA, Gilmour J, Lie ALM, Lichtinger M, et al. Cooperative binding of AP-1 and TEAD4 modulates the balance between vascular smooth muscle and hemogenic cell fate. Development 2016;143:4324–40.

[75] 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.

[76] Bian Z, Gong Y, Huang T, Lee CZW, Bian L, Bai Z, et al. Deciphering human macrophage development at single-cell resolution. Nature 2020;582:571–6.

[77] Dong J, Hu Y, Fan X, Wu X, Mao Y, Hu B, et al. Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hybrid state during mouse organogenesis. Genome Biol 2018;19:31.

[78] Liang R, Camprecios G, Kou Y, McGrath K, Nowak R, Catherman S, et al. A systems approach identifies essential FOXO3 functions at key steps of terminal erythropoiesis. PLoS Genet 2015;11:e1005526.

[79] Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, et al. Human haematopoietic stem cell lineage commitment is a continuous process. Nat Cell Biol 2017;19:271–81.

[80] Zheng S, Papalexi E, Butler A, Stephenson W, Satija R. Molecular transitions in early progenitors during human cord blood hematopoiesis. Mol Syst Biol 2018;14:e8041.

[81] Li Y, Jin C, Bai H, Gao Y, Sun S, Chen L, et al. Human NOTCH4 is a key target of RUNX1 in megakaryocytic differentiation. Blood 2018;131:191–201.

[82] Ebrahimi M, Forouzesh M, Raoufi S, Ramazii M, Ghaedrahmati F, Farzaneh M. Differentiation of human induced pluripotent stem cells into erythroid cells. Stem Cell Res Ther 2020;11:483.

[83] Chasis JA. Erythroblastic islands: Specialized microenvironmental niches for erythropoiesis. Curr Opin Hematol 2006;13:137–41.

[84] Anselmo A, Lauranzano E, Soldani C, Ploia C, Angioni R, D'Amico G, et al. Identification of a novel agrin-dependent pathway in cell signaling and adhesion within the erythroid niche. Cell Death Differ 2016;23:1322–30.

[85] Li W, Wang Y, Zhao H, Zhang H, Xu Y, Wang S, et al. Identification and transcriptome analysis of erythroblastic island macrophages. Blood 2019;134:480–91.

[86] Demirci S, Leonard A, Tisdale JF. Hematopoietic stem cells from pluripotent stem cells: Clinical potential, challenges, and future perspectives. Stem Cells Transl Med 2020.

[87] Nakada D, Oguro H, Levi BP, Ryan N, Kitano A, Saitoh Y, et al. Oestrogen increases haematopoietic stem-cell self-renewal in females and during pregnancy. Nature 2014;505:555–8.

[88] Xie X LM, Zhang Y, Wang B, Zhu C, Wang C, Li Q, et al. Single-cell transcriptomic landscape of human blood cells. National Science Review 2020.

[89] Anderson HL, Brodsky IE, Mangalmurti NS. The evolving erythrocyte: Red blood cells as modulators of innate immunity. J Immunol 2018;201:1343–51.

[90] 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

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

iPSCs *in vitro*. **B**. Histogram of iPSC counts on day 0 and collected SCs on day 14 per round-bottom 96-well plate (error bar = mean ± SD). **C**. Percentage of $CD34^+$ and $CD45^+$ cells in the total SCs (mean ± SD). **D**. Count of colonies that grew from SCs collected on day 14. **E**. The morphology of hematopoietic colony formation units grown from SCs. Scale bar = 1000 μm. **F**. Flow cytometry results showing cells expression of CD71 and CD235a on day 14, 17, 23, and 28 during the erythroid differentiation of SCs. **G**. qPCR of β-chain and γ-chain globin expression on day 4, day 14 + 3, and day 14 + 9 during the entire system differentiation. **H**. The morphology of erythroid cells on day 28. Scale bar = 20 μm.

Figure 2 Cell composition in the iPSCs-derived system

A. UMAP plot of cells labeled by batches. B. UMAP visualization of EBCs and SCs. Cells were labeled by SingleR annotation results. C. UMAP plot of cells labeled by cell classifications based on starting cells or derived from different germ layers. **D**. The dot plot of signature genes for each cell classification. **E**. UMAP visualization of cell cycle analyzed by Seurat. The pie chart showed the percentage of cells in different cell cycles in the starting cells. Phases of cells indicated by different colors, G1 in green, G2M in brown, and S in gray. **F**. GO term enrichment results of signature genes from each cell classification.

Figure 3 Cell-cell contacts analysis

A. Heatmap of cell-cell contact scores analyzed by CellChat. **B**. Contribution of each L-R pair in CDH5, CD34, CD40, ESAM, MADCAM, and PTPRM signaling pathway. **C**. The role of each classification cell in ESAM signaling pathway. **D**. UMAP plots of every single-cell *ESAM* expression level and cell counts bar plot of endothelial cells and pericytes in lateral plate mesoderm-derived cells.

Figure 4 Regulons regulation analysis

A. UMAP map of all cells based on human transcription factors, with colors indicating six cell classifications. **B**. Heatmap plot of regulons AUC patterns related to cell

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

classifications. The high AUC score regulons in each cell classification were listed on the right side of the plot. **C**. The network of the high AUC score regulons in each cell classification.

Figure 5 Cell pseudo-time trajectory of the iPSCs-derived RBCs regeneration system.

A. Developmental pseudo-time of iPSCs-derived erythrocytes regeneration system by Monocle2. **B**. The arrangement of different classification cells on pseudo-time trajectory. **C**. Heatmap showing the specific gene expression patterns in hematopoietic and non-hematopoietic branches. **D**. Functional annotation analysis of each cluster gene from **C**. **E**. The arrangement of different state cells on pseudo-time trajectory. **F**. The expression of trend changes between states of each cluster gene from **D**. **G**. GSVA analysis among different cell states.

Figure 6 Key factors driving hematopoiesis

A. Developmental pseudo-time of iPS cells, Embryonic stem cells, Endothelial cells, and $CD34^+$ HSPC by Monocle2. **B**. The arrangement of different type cells on pseudo-time trajectory. **C**. Heat map showing the specific gene expression patterns in endothelial cells and HSPCs branches. **D**. Volcanic plot of DEGs from gene differential expression analysis between HECs and HSPCs ($P < 0.05$, $|\log2\ FC| < 0.5$). **E**. KEGG analysis of HECs high expression gene.

Figure 7 Differentiation trajectory of hematopoietic cells

A. Developmental pseudo-time of hematopoietic cells by Monocle2. **B**. The arrangement of different type cells on hematopoietic pseudo-time trajectory. **C**. The arrangement of different state cells on pseudo-time trajectory and UMAP plot of hematopoietic cell states distribution. **D**. UMAP plots of every single cell showing the expression level of *MYB*, *CD34*, *PTPRC*, *CD44*, *HOXA6*, and *KIT*. **E**. Heat map showing the specific gene expression patterns in erythropoiesis and non-erythropoiesis branches.

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

Figure 8 Pseudo-time trajectory of erythroid cells

A. Developmental process of erythroid cells by Monocle2. **B**. The arrangement of different type cells on erythroid cell pseudo-time trajectory. **C**. Heatmap showing the gene expression dynamics during erythroid differentiation. **D**. The arrangement of different state cells on erythroid cell pseudo-time trajectory. **E**. GSVA analysis between state one and three cells. **F**. UMAP plots of erythroid cell state distribution and *TFRC*, *CD74*, and *EPOR* expression level. **G**. GO analysis of DEGs between state one and three cells. **H**. KEGG Network constructed with high expression genes in ortho-E cells

Supplementary material

Figure S1 Comparison of two batches

A. Pie chart of the cell number percentage of the two batches. **B**. Correlation analysis of the two batches by the Spearman correlation coefficient of all detected genes.

Figure S2 Cell score heatmap of 32 cell types annotated by SingleR and the distribution of each cell type in the two sample batches. Batch 1 is represented in orange, and batch 2 is represented in blue.

Figure S3 The location of each cell type annotated by SingleR in the UMAP plot.

Figure S4 Each cell classification counts in batches and correlation analysis

A. The number of cells in each cell classification after cell-type classification (left) and the batch source of each cell classification (right). **B**. Heatmap plot of Spearman correlation coefficient between cell classifications.

Figure S5 Cell-cell contact involving bulk RNA-seq data analysis

A and **B**. The cell-cell contact patterns of iPSCs-derived erythrocyte regeneration system, cell patterns_outgoing (**A**), Cell patterns_incoming (**B**). **C**. Heatmap plot of hematopoietic lineage-related gene expression patterns for four bulk RNA-seq datasets.

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

Blue indicates low gene expression and red indicates high gene expression. **D**. PCA of four bulk RNA-seq datasets. The red arrow indicates the PCA position of the four bulk RNA-seq datasets. **E**. Gene function annotation analysis for low expression genes in SCs.

Figure S6 GO annotation of target genes in regulons

The GO enrichment results and heatmap plot displayed the relationship between biological processes and the target genes of TFs in Figure 4C. NANOG and POU5F1 that are activated in starting cells (**A** and **B**), SOX10, ONECUT2, and SIX1 that are activated in ectoderm-derived cells (**C** and **D**), FOXA3, CREB3, and TEAD1 that are activated in intermediated mesoderm-derived cells (**E** and **F**), HOXC6 and PRDM6 that performed functions in paraxial mesoderm-derived cells (**G** and **H**), MECOM, ETS1, ELK3, ERG, and SOX18 that performed functions in lateral mesoderm-derived cells (**I** and **J**), GATA1, TAL1, RUNX1, CEBPD, MYB, and ELF1 that are activated in hematopoietic cells (**K** and **L**). In contrast, CEBPD and MYB were specifically activated in myeloid cells (**M** and **N**) and GATA1 and TAL1 were specifically activated in the erythroid cells (**O** and **P**).

Figure S7 Distribution of regulons activity in TFs UMAP reduced dimension

Schematic diagram of the regions of 21 cell classifications specificity activation regulons on the UMAP map. Regulon activated cells were colored red, and regulon non-activated cells were colored gray.

Figure S8 RSS (regulon specificity score) pattern of each hematopoietic cell type

The top five regulons in each hematopoietic cell type were highlighted in red and labeled on the plot based on the RSS. The content in the label brackets is the number of target genes in the regulons. The regulons specificity score is shown on the y-axis.

Figure S9 Regulation of iPSCs differentiation direction

A. The arrangement of each classification cells on pseudo-time trajectory. **B**. *POU5F1*,

bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. 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.

RUNX1, and *SOX7* expression level on total cell pseudo-time trajectory. **C**. KEGG analysis of marker genes in each state cell. **D**. Volcanic plot of DEGs from gene differential expression analysis between state two and three initial cells ($P < 0.05$, |log2 FC| < 0.5). **E**. Identification of top20 TFs regulating high expression genes in state two initial cells. **F**. Network of TFs from E by GeneMANIA.

Figure S10 Dynamic expression of cell adhesion molecules during EHT

A. The arrangement of embryonic stem cells, iPS cells, endothelial cells, and $CD34^+$ HSPCs on pseudo-time trajectory. **B**. The arrangement of state cells on pseudo-time trajectory. **C**. The expression of trend changes between states of each cluster gene from Figure 5C. **D**. qPCR analysis of cell adhesion molecules showing fluctuation throughout EHT.

Figure S11 Hematopoietic cells differentiation trajectory

A. UMAP plot of hematopoietic cells distribution. **B**. Cell proportion pie chart of hematopoietic cells. **C**. The arrangement of each type of cells on hematopoietic pseudo-time trajectory. **D**. *GATA1*and *CEBPD* expression level on hematopoietic cells pseudo-time trajectory. **E**. The expression of trend changes between states of each cluster gene from Figure 7E.

Figure S12 Dynamic expression of estrogen target genes during erythroid differentiation

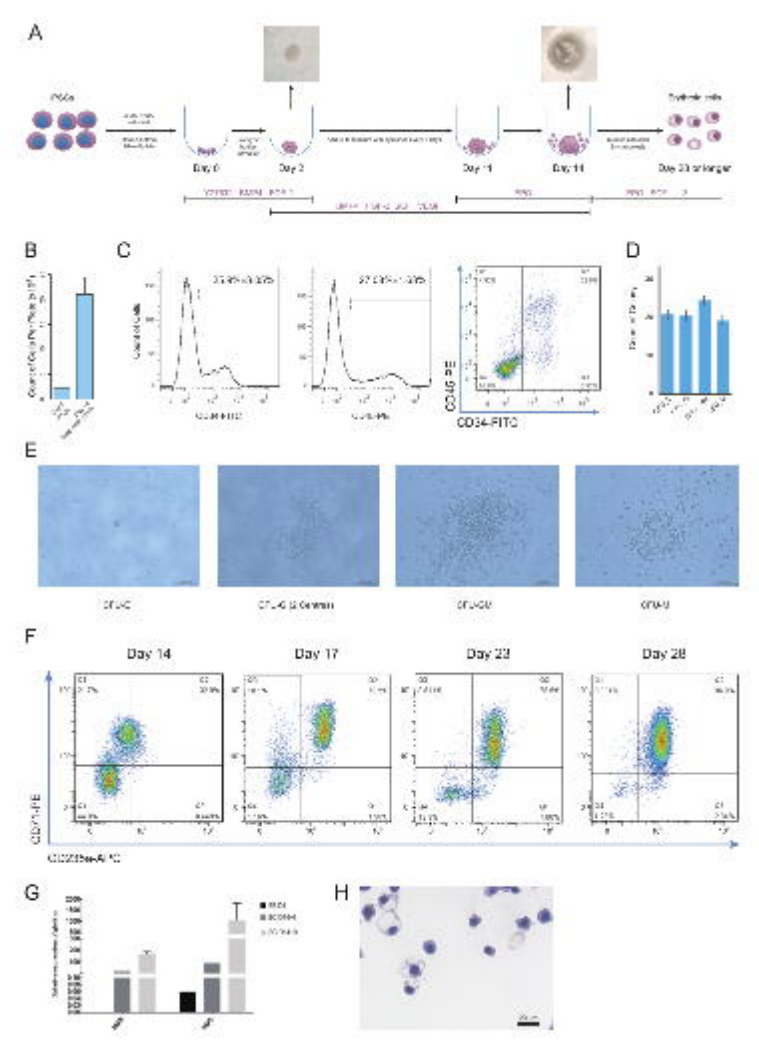
A. The arrangement of BFU-E, CFU-E, erythrocytes, ortho-E cells on erythroid pseudo-time trajectory. **B**. Pseudo-time analysis of genes in the estrogen signaling pathway KEGG term from Figure 6C. **C**. GO analysis of cluster 2 genes from Figure 8C. **D**. The expression of trend changes between states of each cluster gene from Figure 8C. **E**. qPCR analysis of the effects of estradiol during erythroid differentiation. **F**. Volcanic plot of DEGs from gene differential expression analysis between state one and three cells ($P < 0.05$, |log2 FC| < 0.5). **G**. Volcanic plot of DEGs from gene differential expression analysis between ortho-E and other erythroid cells (P < 0.05,

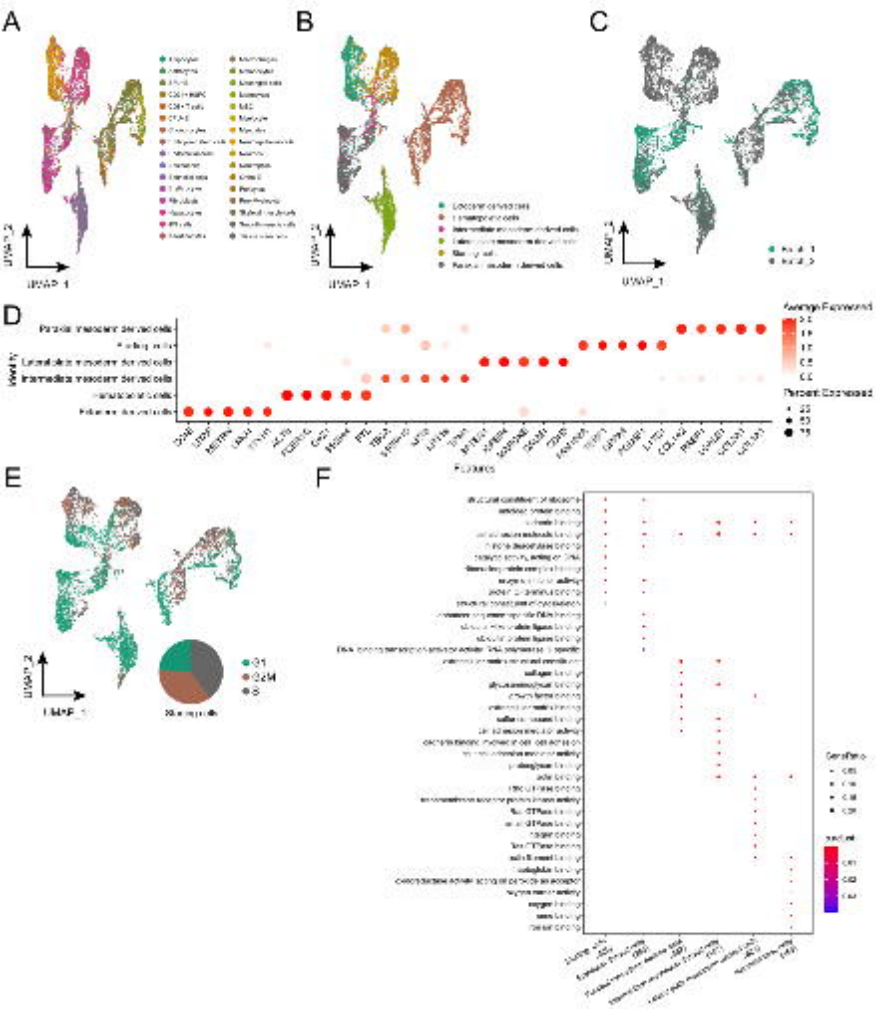
bioRxiv preprint doi: https://doi.org/10.1101/859777; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

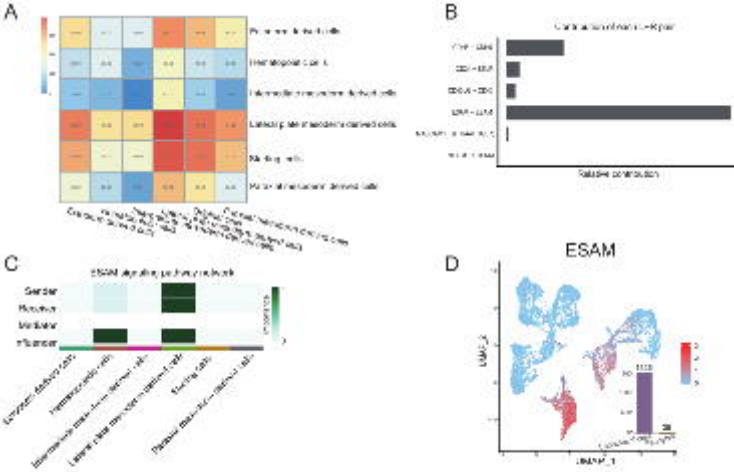
|log2 FC| < 0.5). **H**. KEGG analysis of DEGs from **G**.

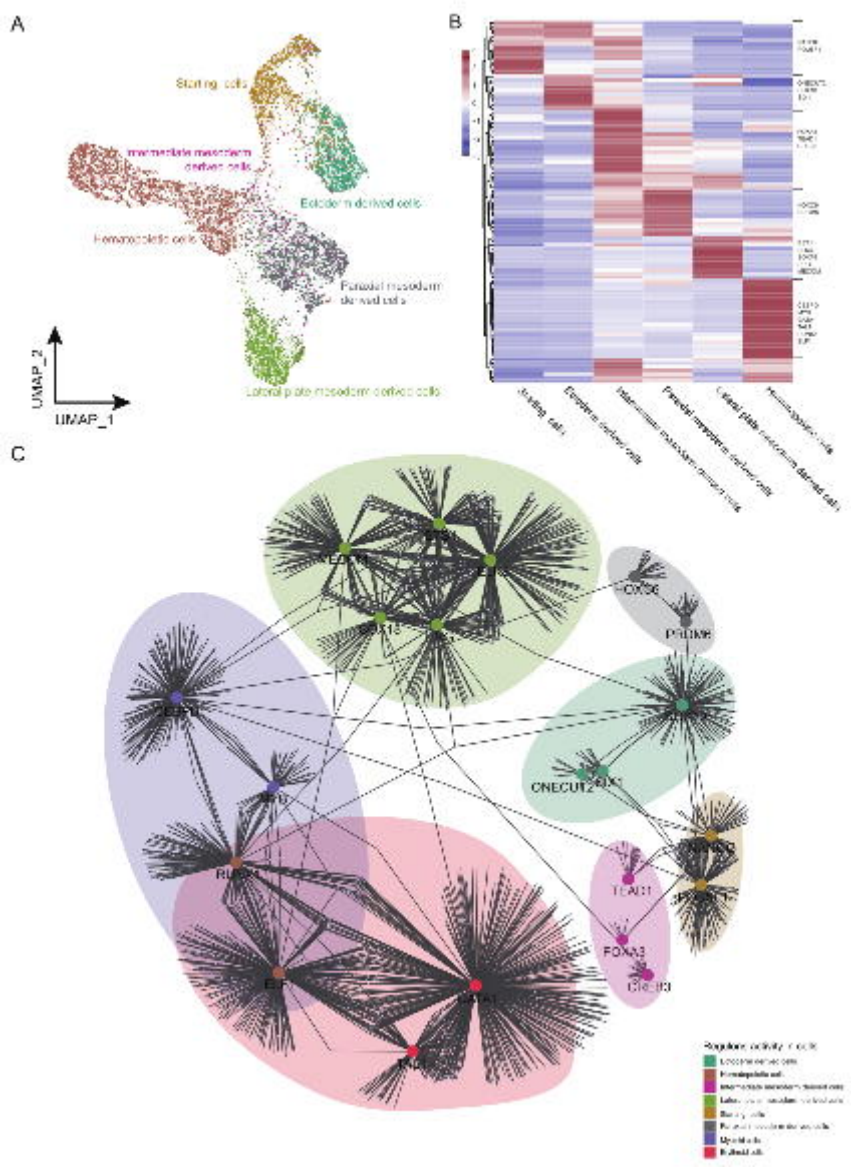
Table S1 Sample quality evaluation indicators and results

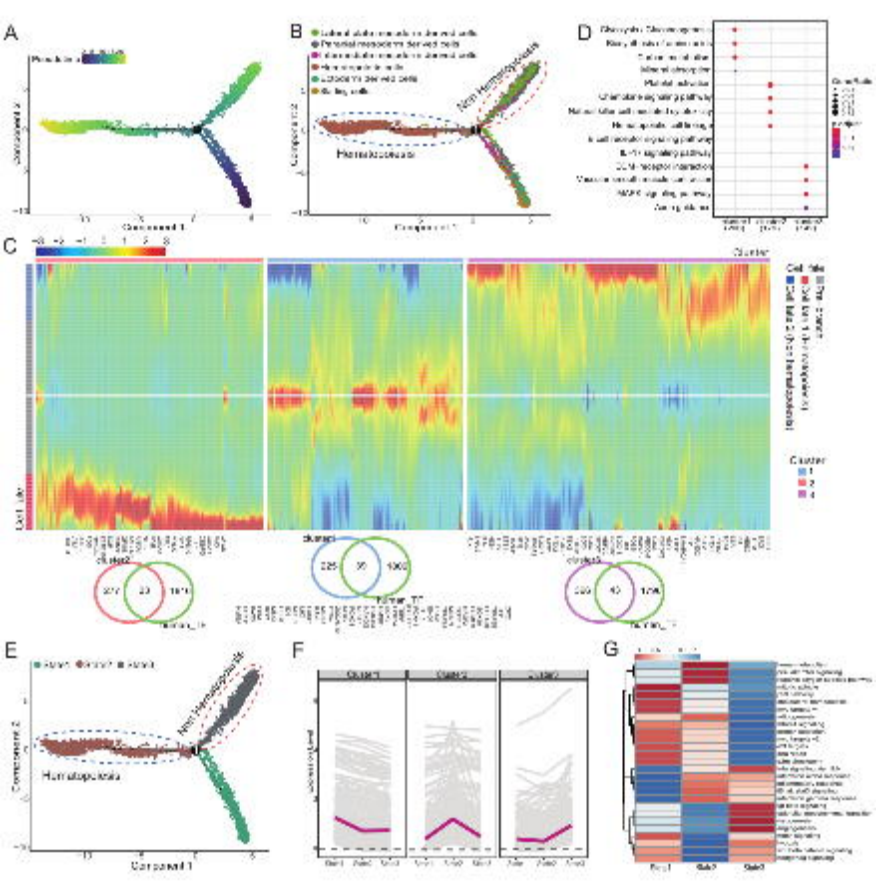
Table S2 qPCR primer for globin, cell adhesion and estradiol related genes

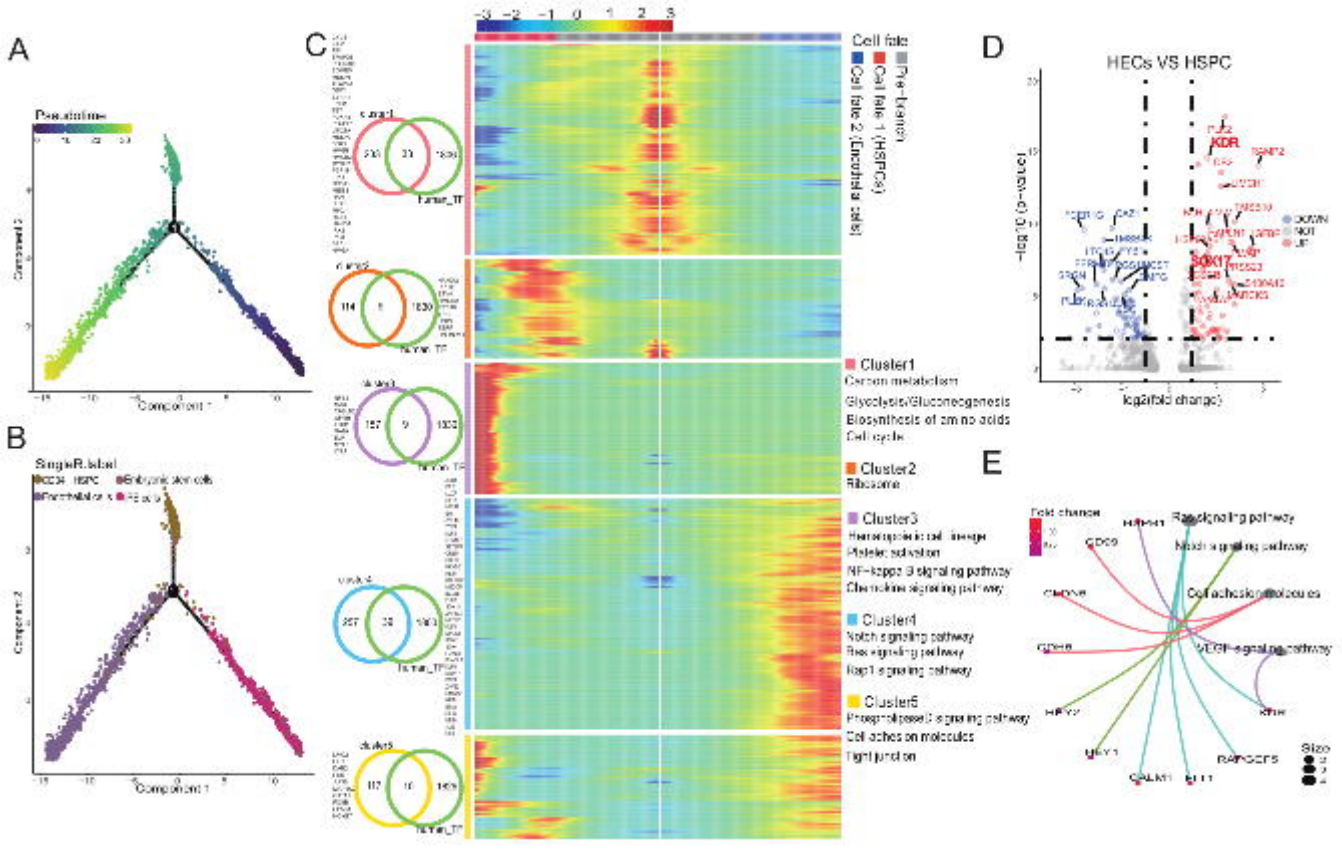

A
Day 0
Day 2
Day 11
Day 11
Day 23 or longer
B
C
CD34-FITC
D
E
F
Day 14
Day 17
Day 23
Day 28
G
H

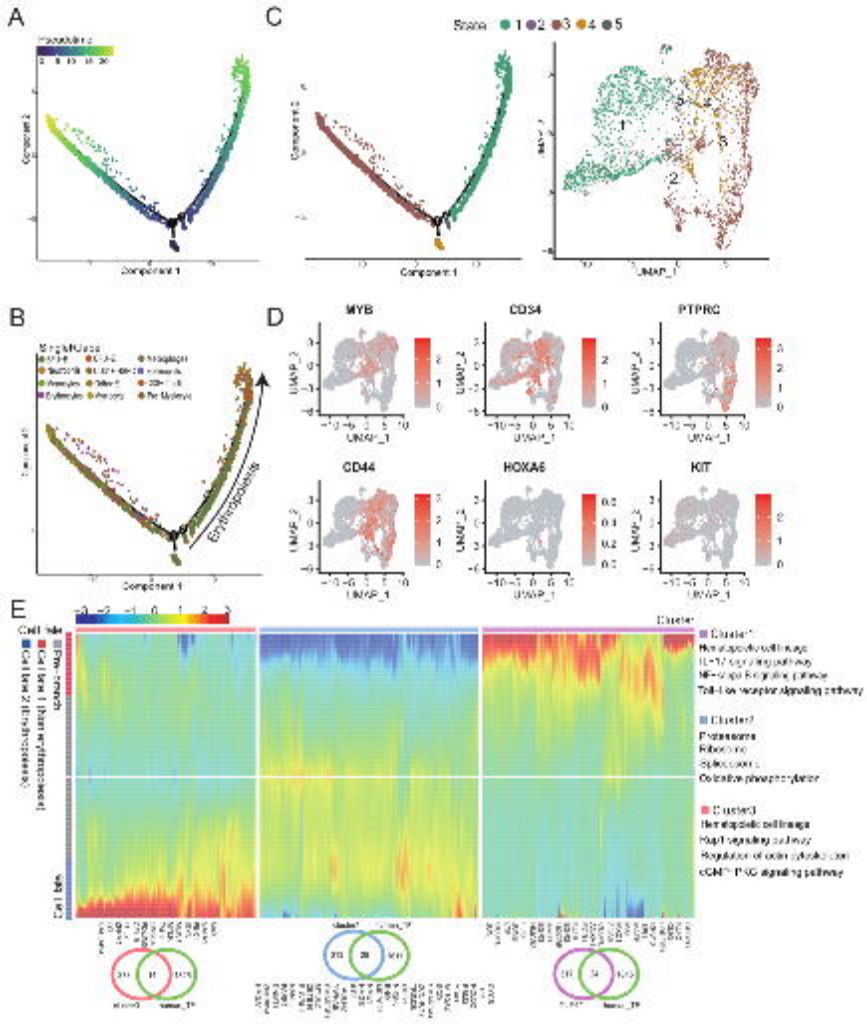

A
B
C
D
E
F
UMAP_1
UMAP_2
Average Expression
Percent Expressed
Features

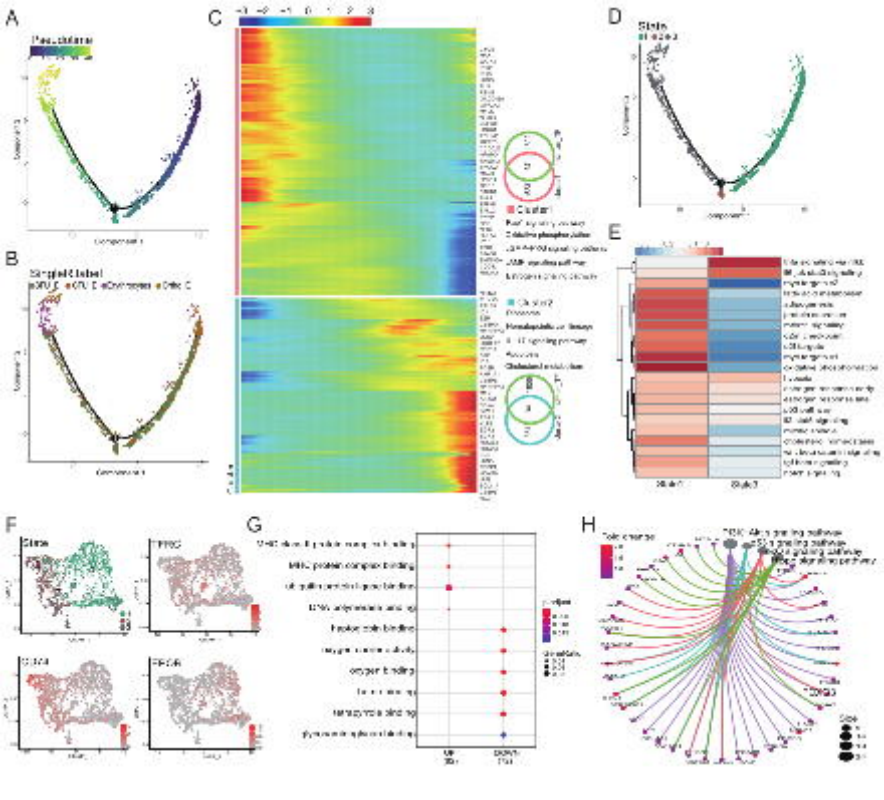

A
B
C
D
ESAM


A
Starting cells
intermediate mesoderm derived cells
Ectoderm derived cells
Hematopoietic cells
Paraxial mesoderm derived cells
Lateral plate mesoderm derived cells
UMAP_2
UMAP_1
B
C
CNECUT2
TEAD1
FOXA3


A
B
C
D
E
F
G
Pseudotime
Component 1
Component 2
Hematopoietic
Non Hematopoietic
Cluster
Cell Type
Gene Module
Gene Ontology
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Cluster 10
Cluster 11
Cell Cycle
Chromosome Segregation
Mitotic Cell Cycle
Cell Division
Mitotic Sister Chromatid Segregation
Mitotic Nuclear Division
Cell Cycle Phase
Cell Cycle Process
Cell Cycle Checkpoint
Cell Cycle Phase Transition
Cell Cycle G2/M Phase Transition
Cell Cycle Mitotic
Cell Cycle G1/S Phase Transition
Cell Cycle DNA Replication
Cell Cycle Checkpoint Signaling
Cell Cycle Checkpoint Mitotic
Cell Cycle Checkpoint Cell Cycle
Cell Cycle Checkpoint Cell Cycle Phase
Cell Cycle Checkpoint Cell Cycle Process
Cell Cycle Checkpoint Cell Cycle Phase Transition
Cell Cycle Checkpoint Cell Cycle G2/M Phase Transition
Cell Cycle Checkpoint Mitotic
Cell Cycle Checkpoint Cell Cycle G1/S Phase Transition
Cell Cycle Checkpoint Cell Cycle DNA Replication
Cell Cycle Checkpoint Signaling
Cell Cycle Checkpoint Mitotic Cell Cycle
Cell Cycle Checkpoint Cell Division
Cell Cycle Checkpoint Mitotic Sister Chromatid Segregation
Cell Cycle Checkpoint Mitotic Nuclear Division
Cell Cycle Checkpoint Cell Cycle Phase
Cell Cycle Checkpoint Cell Cycle Process
Cell Cycle Checkpoint Cell Cycle Checkpoint


A
Pseudotime
Component 1
Component 2
B
SingleR.label
Component 1
Component 2
C
Cell fate
Cell fate 2 (Endothelial cells)
Cell fate 1 (HSPC)
Pre-branch
Cluster1
Carbon metabolism
Glycolysis/Gluconeogenesis
Biosynthesis of amino acids
Cell cycle
Cluster2
Ribosome
Cluster3
Hematopoietic cell lineage
Platelet activation
NF-kappa B signaling pathway
Chemokine signaling pathway
Cluster4
Notch signaling pathway
Ras signaling pathway
Rap1 signaling pathway
Cluster5
Phosphatidylinositol signaling pathway
Cell adhesion molecules
Tight junction
D
HECs VS HSPC
log2(fold change)
KDR
SOX17
E
Fold change
Ras signaling pathway
Notch signaling pathway
Cell adhesion molecules
VEGF signaling pathway
Size


A
Pseudotime
Component 1
Component 2
C
State 1 2 3 4 5
Component 1
Component 2
UMAP_1
UMAP_2
B
Erythroblasts
Component 1
Component 2
D
MYB
CD34
PTPRC
CD44
HOXA6
KIT
UMAP_1
UMAP_2
E
Cell fate
Cluster
Cluster1
Hematopoietic cell lineage
IL-17 signaling pathway
NF-κB signaling pathway
Toll-like receptor signaling pathway
Cluster2
Proteasome
Ribosome
Spliceosome
Oxidative phosphorylation
Cluster3
Hematopoietic cell lineage
Rap1 signaling pathway
Regulation of actin cytoskeleton
cGMP-PKG signaling pathway


A
Pseudotime
B
Singlet label
C
Cluster1
Cluster2
D
State
E
F
State
G
H