



ARTICLE OPEN

Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing

Yang Zeng¹, Jian He², Zhijie Bai², Zongcheng Li¹, Yandong Gong², Chen Liu², Yanli Ni¹, Junjie Du², Chunyu Ma³, Lihong Bian³, Yu Lan⁴ and Bing Liu^{1,2,4,5}

Tracing the emergence of the first hematopoietic stem cells (HSCs) in human embryos, particularly the scarce and transient precursors thereof, is so far challenging, largely due to the technical limitations and the material rarity. Here, using single-cell RNA sequencing, we constructed the first genome-scale gene expression landscape covering the entire course of endothelial-to-HSC transition during human embryogenesis. The transcriptomically defined HSC-primed hemogenic endothelial cells (HECs) were captured at Carnegie stage (CS) 12–14 in an unbiased way, showing an unambiguous feature of arterial endothelial cells (ECs) with the up-regulation of *RUNX1*, *MYB* and *ANGPT1*. Importantly, subcategorizing CD34⁺CD45⁻ ECs into a CD44⁺ population strikingly enriched HECs by over 10-fold. We further mapped the developmental path from arterial ECs via HSC-primed HECs to hematopoietic stem progenitor cells, and revealed a distinct expression pattern of genes that were transiently over-represented upon the hemogenic fate choice of arterial ECs, including *EMCN*, *PROCR* and *RUNX1T1*. We also uncovered another temporally and molecularly distinct intra-embryonic HEC population, which was detected mainly at earlier CS 10 and lacked the arterial feature. Finally, we revealed the cellular components of the putative aortic niche and potential cellular interactions acting on the HSC-primed HECs. The cellular and molecular programs that underlie the generation of the first HSCs from HECs in human embryos, together with the ability to distinguish the HSC-primed HECs from others, will shed light on the strategies for the production of clinically useful HSCs from pluripotent stem cells.

Cell Research (2019) 29:881–894; <https://doi.org/10.1038/s41422-019-0228-6>

INTRODUCTION

Hematopoietic stem cells (HSCs) give rise to all blood lineages and permanently maintain the adult hematopoietic system throughout the entire lifespan of an individual via self-renewal and differentiation. Although the ontogeny of HSCs during development has been extensively investigated in animal models including zebrafish and mice, it is largely unclear in human embryos, given the limited accessibility of human embryonic tissues. By functional assessment with xenotransplantation, human HSCs are reported to be detected sequentially in multiple embryonic sites. Long-term multi-lineage repopulating HSCs are detected firstly in the aorta–gonad–mesonephros (AGM) region at Carnegie stage (CS) 14 (32 days post coitus, dpc), with a frequency of less than one per embryo equivalent, and then in the yolk sac several days later (CS 16, 35–38 dpc), showing an even lower frequency than that in the AGM region.^{1,2} The first human HSCs manifest a phenotype of CD34⁺CD144⁺CD45⁺KIT⁺CD90⁺Endoglin⁺RUNX1⁺CD38⁻CD45RA⁻, similar to those in the embryonic liver or cord blood,^{3–5} although the enrichment is yet far from being efficient. An evident presence of HSCs in the embryonic liver, the major organ for HSC expansion during embryogenesis, is

witnessed only after CS 17 (39–42 dpc), generally from 7 to 8 weeks of gestation.^{1,2,6}

Histologically, the first HSCs are supposed to be located within the intra-aortic hematopoietic clusters (IAHCs) predominantly on the ventral wall of the dorsal aorta in human embryos.³ Given the expression of multiple endothelial markers in emerging HSCs, and the spatiotemporally intimate relationship between them and vascular endothelial cells (ECs), it has been widely proposed that human HSCs are derived from hemogenic ECs (HECs), a specified endothelial lineage with blood-forming potential, similar to what has been comprehensively validated in mice.^{7,8} Up to date, little is known about the developmental dynamics and molecular identity of the HSC-primed HECs in human embryos, except for the simultaneous expression of endothelial feature genes and *RUNX1*, a transcription factor (TF) known to be expressed in hematopoietic-fated ECs and critically required for endothelial-to-HSC transition,^{9,10} and the lack of known hematopoietic markers, such as CD43 (encoded by *SPN*) and CD45 (encoded by *PTPRC*).

Regarding the developmental path of HECs that give rise to hematopoietic stem progenitor cells (HSPCs), most proposed

¹State Key Laboratory of Experimental Hematology, Fifth Medical Center of Chinese PLA General Hospital, Beijing 100071, China; ²State Key Laboratory of Proteomics, Academy of Military Medical Sciences, Academy of Military Sciences, Beijing 100071, China; ³Department of Gynecology, Fifth Medical Center of Chinese PLA General Hospital, Beijing 100071, China; ⁴Guangzhou Regenerative Medicine and Health-Guangdong Laboratory (GRMH-GDL); Key Laboratory for Regenerative Medicine of Ministry of Education, Institute of Hematology, Department of Pathophysiology, School of Medicine, Jinan University, Guangzhou 510632, China and ⁵State Key Laboratory of Experimental Hematology, Institute of Hematology and Blood Diseases Hospital, Chinese Academy of Medical Sciences, Tianjin 300020, China

Correspondence: Yu Lan (rainyblue_1999@126.com) or Bing Liu (bingliu17@yahoo.com)

These authors contributed equally: Yang Zeng, Jian He, Zhijie Bai, Zongcheng Li

Received: 15 July 2019 Accepted: 13 August 2019

Published online: 9 September 2019

views are established on the in vitro differentiation system of human pluripotent stem cells (PSCs), which still fails to efficiently and reliably generate functional HSCs.^{11–14} It is suggested that the HECs and vascular ECs are specified early as distinct lineages from KDR^+ mesoderm, and HECs subsequently integrate with vascular ECs of the dorsal aorta, undergoing endothelial-to-hematopoietic transition to form IAHCs.^{15,16} Most recently, activation of the arterial program in HECs, which is regulated by MAPK/ERK and Notch signaling, has been reported to be required for the lymphopoiesis in PSC differentiation system, whereas the HECs without arterialization demonstrate only primitive and myeloid-restricted hematopoietic potential.^{14,17} It remains elusive whether the HSC-primed HECs are derived from arterial ECs and what are the differences between the putative distinct HEC populations, if they exist, during human embryogenesis.

HSC development in AGM region needs the signals from surrounding niche cells, including at least sub-aortic mesenchymal cells. The cellular and molecular components of AGM niche have mainly been investigated using animal models, showing the presumed involvement of BMP and SCF signaling, both with a polarized distribution predominantly to the ventral part of sub-aortic mesenchyme.^{18,19} Interestingly, previous study has described a successful reprogramming from adult mouse ECs to HSCs through transient expression of four TFs combined with an endothelial feeder-dependent induction, emphasizing the role of vascular niche-derived angiocrine factors in the generation of HSCs.²⁰ Deciphering the putative interactions that potentially support the specification of the HSC-primed HECs, the initial step for hematopoietic fate choice, in human embryos is of great importance but has not been achieved.

Single-cell transcriptional profiling has been widely used to decipher the developmental events.^{21,22} It shows unique advantage in the capture of the transient and rare cell populations, such as the emerging HSCs and HSC-primed HECs, from invaluable human embryonic samples in the present study. Here, we performed both well-based single-cell RNA-sequencing (scRNA-seq) of 509 cells and droplet-based scRNA-seq of 11,440 cells to firstly construct a molecular landscape of HSC generation in human embryos (Supplementary information, Fig. S1a). Precisely understanding the cellular and molecular programs and interactions that underlie the generation of the first HSCs from HECs, together with distinguishing the HSC-primed HECs from the ones related to transient hematopoiesis, will unambiguously shed light on the strategies for the generation of clinically useful HSCs from human PSCs.

RESULTS

Transcriptional capture of HECs in human embryonic dorsal aorta around HSC emergence

We proposed that HSC-primed HECs existed at stages around HSC emergence at CS 14,¹ thus should be transcriptionally captured within the aortic structure at CS 13. Therefore, we firstly performed droplet-based scRNA-seq (10X Chromium) with 7-AAD⁻CD235a⁻ cells from the anatomically dissected dorsal aorta in human AGM region at CS 13 (30 dpc). An average of 2968 genes were detected in each individual cell after quality control (Supplementary information, Fig. S1b). Eight cell clusters featured by the expression of known marker genes were readily recognized and annotated, including EC, arterial EC (AEC), hematopoietic-related cell (Hem), Epithelial cell (Epi) and four distinct mesenchymal cell clusters (Fig. 1a–c; Supplementary information, Fig. S1c, d, f). As compared to AEC that showed a typical arterial feature such as expression of *GJA5*, *GJA4*, *HEY2*, and *CXCR4*,^{23–27} EC cluster exhibited a relatively venous feature such as over-representation of *APLN*, *NRP2* and *NT5E*^{28,29} (Fig. 1c; Supplementary information, Fig. S1c). The Hem cluster was characterized by the expression of *SP11* and *SPINK2*^{30,31} (Fig. 1b, c). Taken together, we successfully

captured endothelial and hematopoietic cell populations in the aortic structure, constituting about half of the total cells (48%), together with putative AGM niche cells, mainly involving different types of mesenchymal cells (46%) (Fig. 1d).

Pearson correlation analysis revealed that AEC showed the highest similarity with Hem cluster, whereas mesenchymal and Epi clusters correlated closely (Supplementary information, Fig. S1e). This finding suggested that the hemogenic activity was presumably located within AEC and Hem clusters. To precisely capture the putative HECs, AEC and Hem clusters were taken out separately and further divided into three sub-clusters (Fig. 1e). The largest one was basically corresponding to the previous AEC cluster, thus was still annotated as AEC. Of note, one of the other two sub-clusters showed high expression of *RUNX1* together with the endothelial feature, thus was annotated as HEC (Fig. 1e, f; Supplementary information, Fig. S1g). The other one was named as hematopoietic cell (HC) given the expression of hematopoietic genes *SPN* and *PTPRC* but the lack of endothelial property (Fig. 1e; Supplementary information, Fig. S1g). Compared among these three sub-clusters, the major biological processes enriched in AEC were related to extracellular matrix organization and vasculature/endothelium development, in accord with that the dorsal aorta at this stage is undergoing a maturation process³² (Fig. 1g; Supplementary information, Fig. S1d). In addition to *RUNX1*, a novel long non-coding RNA (lncRNA)—*SNHG16* was found as the most significant differentially expressed genes (DEGs) in HEC (Fig. 1f). Genes related to RNA catabolic process were enriched in HEC sub-cluster, also evidenced by the relatively high expression of *RPL5*, *RPL12* and *RPL6*³³ (Fig. 1f, g).

Next, we computationally screened for the surface markers that might help to enrich and prospectively isolate the HEC population. Among the three sub-clusters, no surface markers specific for the HEC population were screened out, which was conceivable as the feature of HEC was similar with both AEC and HC clusters. Among the top 10 significantly differentially expressed surface markers between AEC and HC, *CD44* showed relatively more abundant expression in HEC than in AEC, serving as a potential candidate for the enrichment of HEC population (Fig. 1h).

HECs in human AGM region exhibited unambiguous arterial feature and were efficiently enriched in phenotypic CD44⁺ ECs. Due to the limited resolution of droplet-based scRNA-seq strategy including 10X Chromium, we subsequently performed well-based scRNA-seq (modified STRT-seq) to more precisely decode the HECs in human AGM region at stages shortly before or upon the generation of HSCs (Supplementary information, Fig. S1a). The appearance of intra-aortic IAHCs on the ventral wall of human dorsal aorta represents the morphological manifestations of endothelial-to-hematopoietic transition, via which HSPCs are generated. IAHCs firstly emerge at CS 12 (27 dpc),³⁴ and the first HSCs are detected at CS 14.¹ Therefore, CS 12 to CS 14 should be the time window for detecting HSC-primed HECs in human embryos. Immunophenotypic CD235a⁻CD45⁻CD34⁺CD44⁻ cells (CD44⁻ ECs) and CD235a⁻CD45⁻CD34⁺CD44⁺ cells (CD44⁺ ECs) were simultaneously sampled with similar cell numbers, although the ratio the latter population took in ECs was at least 10-fold less than the former (Fig. 2a). Cells were collected from CS 12 (27 dpc) caudal half (CH), CS 13 (29 dpc) and CS 14 (32 dpc) AGM regions of human embryos (Supplementary information, Fig. S1a). An average of 6011 genes were detected in each individual cell and the transcriptional expression of sorting markers basically matched the immunophenotypes (Supplementary information, Fig. S2a–c). By unsupervised clustering, the ECs were mainly divided into two populations, largely in line with the immunophenotypes regarding the expression of CD44 (Fig. 2b; Supplementary information, Fig. S2d). The cluster composed mainly of CD44⁺ ECs was of arterial feature, with ubiquitous expression of *GJA5*, *GJA4* and *DLL4*, and thus was defined as arterial EC (aEC)

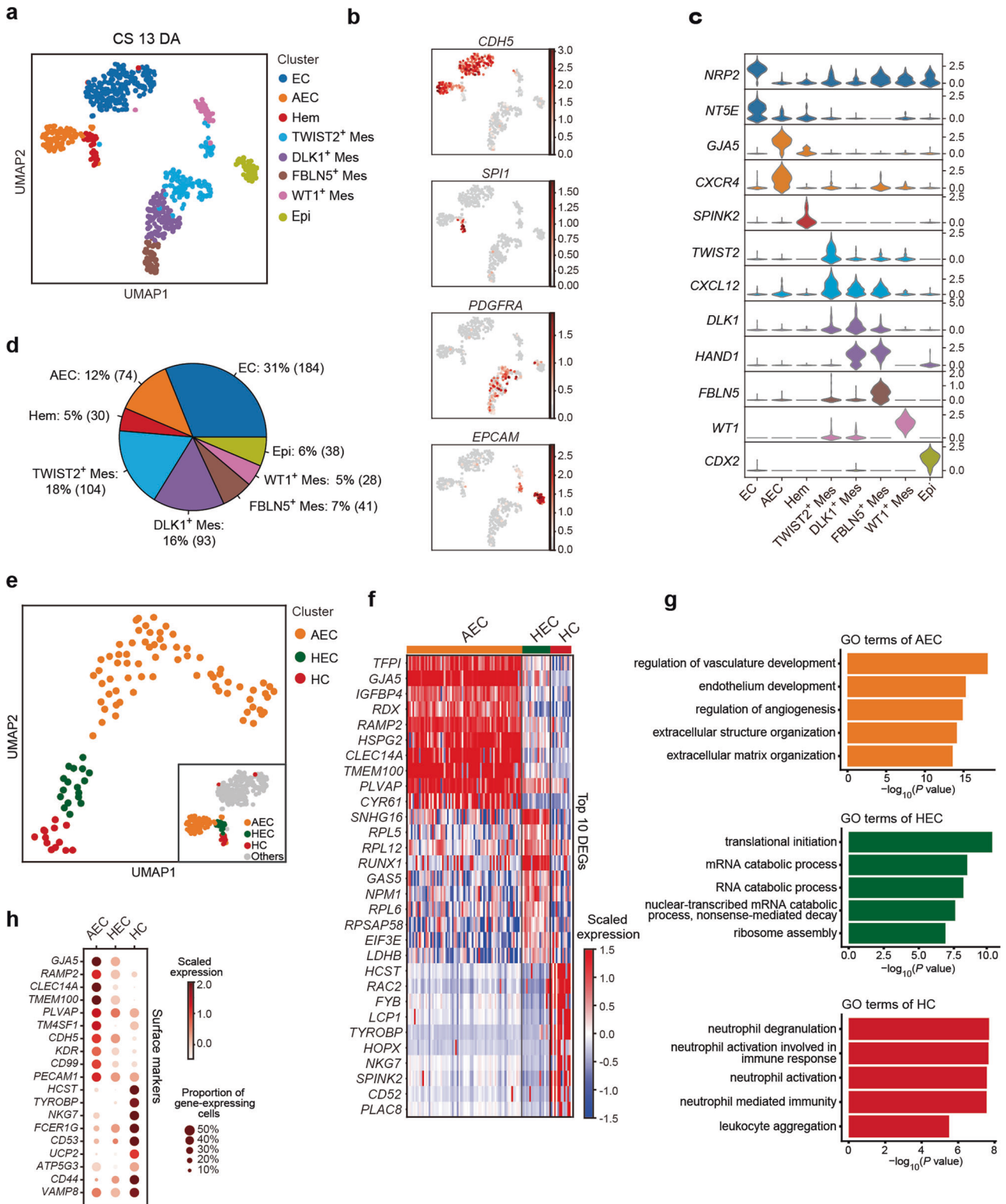


Fig. 1 Transcriptomic identification of different cell populations in human CS 13 dorsal aorta (DA) and the capture of HECs. **a** Identification of cell populations in CS 13 DA visualized by UMAP. Each dot represents one cell and colors represent cell clusters as indicated. **b** UMAP visualization of the expression of curated feature genes for the identification of cell clusters (*CDH5*, *SPI1*, *PDGFRA* and *EPCAM*). **c** Violin plots showing the expression of feature genes in each cell cluster. Colors represent cell clusters indicated in **a**. **d** Pie chart showing the percentages and absolute numbers of each cell cluster involved in CS 13 DA. **e** UMAP visualization of AEC, HEC and HC clusters, resulted from sub-dividing the cells in AEC and Hem clusters described in **a** as indicated in the lower right frame. **f** Heatmap showing the scaled expression of top 10 differentially expressed genes (DEGs) in AEC, HEC and HC clusters. **g** The major Gene Ontology biological process (GO:BP) terms in which DEGs are enriched for each cluster. **h** Dot plots showing the scaled expression level of top 10 significantly differentially expressed surface marker genes in AEC, HEC and HC. Note no specific marker for HEC and only nine genes for HC cluster met the criteria. Colors represent the scaled expression and size encodes the proportion of gene-expressing cells

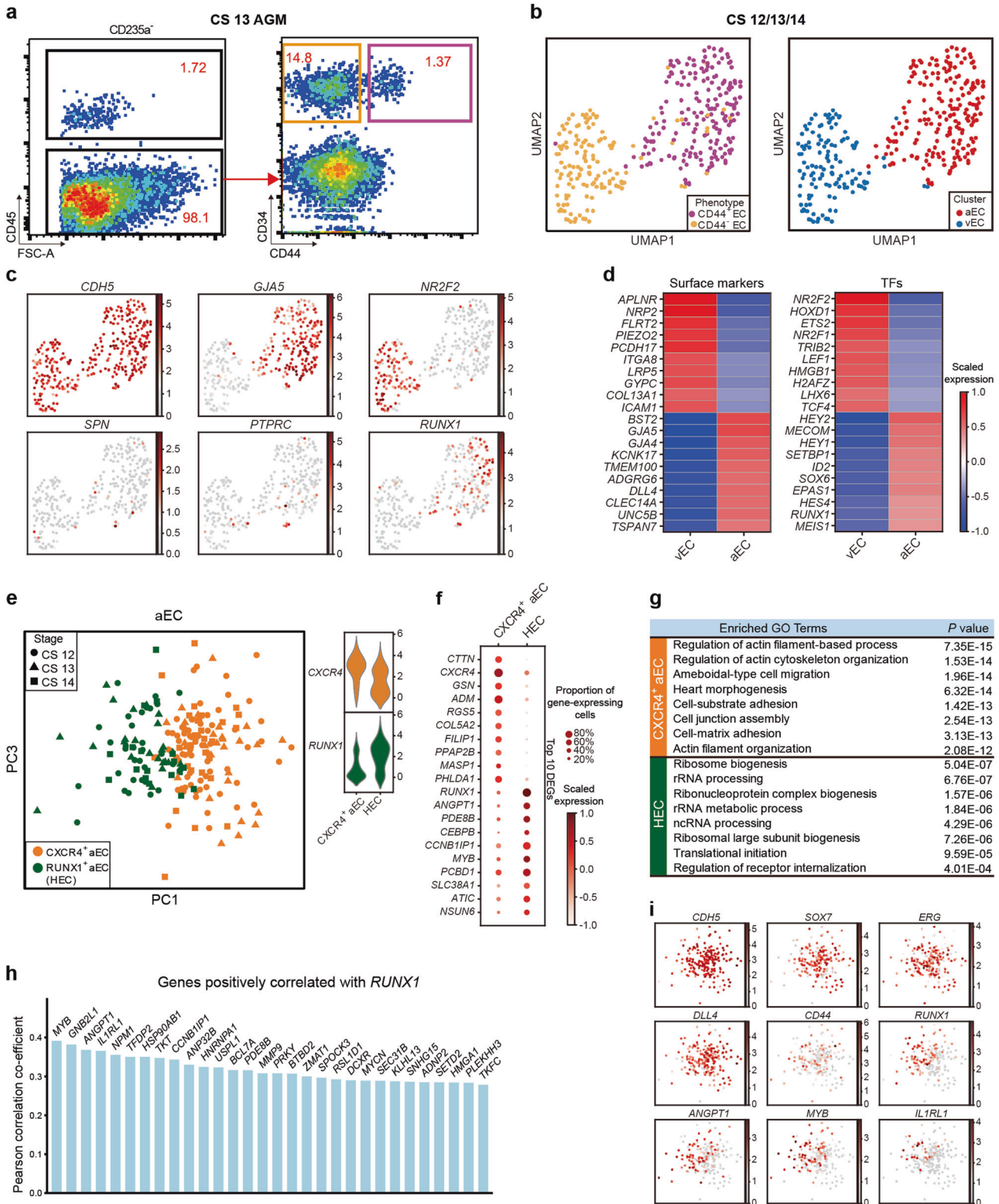


Fig. 2 Capture and further analysis of HECs in CD44⁺ ECs from CS 12/13/14 embryos. **a** Sorting strategy of CD44⁺ EC and CD44⁻ EC. **b** UMAP with phenotypically different populations (left panel) and two transcriptionally distinct clusters (right panel) mapped on it. **c** UMAP plots displaying the expression of hematopoietic and endothelial genes. **d** Heatmaps showing the average expressions of top 10 differentially expressed surface markers and TFs between aEC and vEC clusters. **e** The aEC cluster is divided into two sub-clusters. The expression of their feature genes is shown on the violin plots to the right. **f** Dot plots showing top 10 DEGs in the two sub-clusters. **g** Enriched GO terms in CXCR4⁺ aEC and HEC, respectively. **h** Bar plots displaying top genes positively correlated with RUNX1 in aEC cluster. Genes related to ribosome biogenesis were removed from the gene list. **i** PCA plot showing expression of endothelial and arterial genes and representative genes from **h** in aEC cluster. HEC shares endothelial and arterial features with CXCR4⁺ aEC. Hematopoietic genes correlated with RUNX1 are enriched in HEC

(Fig. 2c, d; Supplementary information, Fig. S2e). The other cluster showed higher expression of venous genes such as *NR2F2* and *NRP2* and was consequently annotated as venous EC (vEC) (Fig. 2c, d). Importantly, hematopoietic gene *RUNX1* was also exhibited in the top 10 over-represented TF genes of aEC population (Fig. 2d). Of note, immunophenotypic $CD45^-CD34^+CD44^+$ cells ($CD44^+$ ECs) enriched most, if not all, *RUNX1*-expressing cells in the AGM region. In contrast, very few, if any, *RUNX1*-expressing cells were detected in $CD44^-$ ECs (Fig. 2b, c).

Since *RUNX1* was exclusively expressed in a small part of cells in aEC cluster (Fig. 2c), aEC was further sub-divided in an unsupervised way into two subsets, featured by the expression of *CXCR4* (*CXCR4*⁺ aEC) and *RUNX1* (HEC), respectively (Fig. 2e, f; Supplementary information, Fig. S2f). The cellular contributions to each subset were similar among three stages (Supplementary information, Fig. S2c, f). Enrichment of pathways involved in the regulation of ribosome and translation initiation within HEC was in accord with the role of *RUNX1* in regulating ribosome biogenesis³⁵ (Fig. 2g; Supplementary information, Fig. S2g). Myb is expressed by HSCs and required for definitive hematopoiesis in mice.^{36,37} *Angpt1* is highly expressed by HSCs and may be involved in regulating the regeneration of their niche in murine bone marrow.³⁸ The respective homologs of these two genes, *MYB* and *ANGPT1*,^{38,39} were included in the genes whose expression patterns were positively correlated with that of *RUNX1* (Fig. 2h; Supplementary information, Table S1), and were also enriched in HEC (Fig. 2i). The expression of *IL1RL1* (the gene encoding receptor for IL33), which was reported co-expressed with *RUNX1* in mouse and human leukemia cells,⁴⁰ was also positively correlated with that of *RUNX1* (Fig. 2h, i). Taken together, the HEC cluster, exhibiting a feature of expressing *RUNX1*, *MYB*, *ANGPT1* as well as endothelial genes *CDH5*, *SOX7* and *ERG* (Fig. 2i), without apparent expression of hematopoietic surface markers *SPN* and *PTPRC* (Fig. 2c), was transcriptionally identified as HEC. These HECs were characterized with clear arterial feature represented by the expression of *GJA5*, *GJA4* and *DLL4*^{27,41,42} (Fig. 2c, d, i). More importantly, considering the physiological cellular constitution, the scarce HECs were efficiently enriched in the phenotypic $CD44^+$ EC population, by at least 10-fold, in the human AGM region.

Developmental path from arterial ECs via HSC-primed HECs to HSPCs in human AGM region

Functionally, the most efficient marker combination reported to enrich the emerging HSCs in human AGM region is $CD34^+CD45^+$.³ Although other markers have been described, the enrichment is not further improved.³ In order to cover the rare HSCs, here we performed well-based scRNA-seq with sorted $CD235a^-CD45^+CD34^+$ cells, which were proven to contain most, if not all, functional human AGM HSCs by xenotransplantations,³ from the dorsal aorta in human AGM region at CS 15 (Fig. 3a; Supplementary information, Fig. S3a). By unsupervised clustering, the immunophenotypic $CD235a^-CD45^+CD34^+$ population was divided into five clusters (Fig. 3b). Among them, two clusters exhibited feature of hematopoietic differentiation, with one towards myeloid lineage represented by the obvious expression of *GATA1* and *ITGA2B* and the other having the sign of lymphoid lineage potential evidenced by the expression of *IL7R* and *LEF1*^{43,44} (Fig. 3c; Supplementary information, Fig. S3b). The remaining three clusters with stemness signatures (*FGD5* and *HLF*) were thus recognized as HSPCs, constituting about 75.4% of the immunophenotypic $CD235a^-CD45^+CD34^+$ population^{45,46} (Fig. 3b, c). Notably, as a marker previously recognized for identifying functional long-term HSCs in mouse bone marrow, *FGD5* was among the top 10 DEGs highly expressed in HSPCs when compared to the two differentiated populations⁴⁵ (Fig. 3c, d). *HLF*, as reported to be expressed in HSCs in multiple mouse

embryonic hematopoietic sites and might maintain their quiescence,^{46,47} was included in top 10 TFs enriched in HSPC clusters (Fig. 3c; Supplementary information, Fig. S3b). The results further validated our identification of HSPCs in the dorsal aorta.

Three HSPC clusters showed distinct features. The one farthest from the lineage-primed clusters, which should be developmentally the most upstream population within $CD45^+CD34^+$ cells (Fig. 3b), manifested higher endothelial (*KDR* and *FLT1*) and arterial (*HEY1*, *GJA5* and *GJA4*) features as compared to the other two HSPC populations, thus was annotated as $GJA5^+$ HSPC (Fig. 3e). The finding suggested arterial-featured HECs as the source of the first HSCs in human AGM region. The most significant difference between the other two HSPC clusters was the cell cycle status, and the one with highly proliferative property was consequently annotated as Cycling HSPC (Supplementary information, Fig. S3d). The other one was annotated as $GFI1B^+$ HSPC given its highest expression of *GFI1B* among the three HSPC clusters (Fig. 3e). As a direct target of *Runx1*, *Gfi1b* is a marker whose expression can be used in enriching mouse AGM HSCs with long-term repopulating capacity.⁴⁸

To investigate the molecular events during the emergence of HSCs, we pooled together the datasets of relevant cell populations from modified STRT-seq data, including vEC, *CXCR4*⁺ aEC and HEC in CS 12 (CH), CS 13 and CS 14 of the AGM region (Fig. 2b, e), and the three HSPC populations in CS 15 for further analysis. Dimension reduction was performed by principal component analysis (PCA) (Fig. 3f). Of note, principal component (PC) 1 captured the differences between endothelial and hematopoietic properties, with HEC and $GJA5^+$ HSPC located adjacent to the middle part of the axis (Fig. 3f; Supplementary information, Fig. S3c). PC2 positive direction represented the arterial feature, exhibiting the gradual down-regulation along with hematopoietic specification of HSPCs from HEC (Fig. 3f; Supplementary information, Fig. S3c). The result was in line with previous reports in mouse embryos showing reciprocal relationship of arterial gene expression and hematopoietic fate acquisition.^{49,50} In contrast to the lack of the expression of venous feature genes, including *NR2F2* and *APLN*, HSPCs expressed different levels of canonical arterial genes, such as *SOX17*, *GJA5*, *HEY2*, and *MECOM*,⁵¹ indicative of the arterial EC origin of HSCs (Fig. 3g). The expression patterns of *CD44*, *RUNX1*, *MYB* and *ANGPT1* were similar, being seldom expressed in vEC, gradually increased from *CXCR4*⁺ aEC via HEC to HSPCs, with obvious expression in all HSPC clusters (Fig. 3g). Of note, *HLF*, *SPN* and *GFI1B* were exclusively expressed in three HSPC clusters⁴⁶ (Fig. 3g). In accord with their involvement in arterial specification,⁴⁹ several genes of NOTCH signaling pathway, including *DLL4*, *JAG1*, *NOTCH1* and *NOTCH4*,¹⁴ were most highly expressed in *CXCR4*⁺ aEC, and gradually down-regulated along the specification via HEC towards HSPCs (Fig. 3g). We also witnessed the difference in cell cycle status among distinct clusters. *CXCR4*⁺ aEC and HEC showed a relatively quiescent status as compared to vEC, with a dramatically increased proliferation in HSPCs (Supplementary information, Fig. S3d). These findings suggested that once the hematopoietic fate is acquired, HSPCs may rapidly expand in situ before differentiation.

Monocle analysis revealed the developmental trajectory of endothelial-to-hematopoietic transition in human AGM region, clearly showing the path from *CXCR4*⁺ aEC via HEC, $GJA5^+$ HSPC, to Cycling and $GFI1B^+$ HSPC clusters (Fig. 3h; Supplementary information, Fig. S3e). Genes that were significantly differentially expressed along the pseudotime axis were subjected to k-means clustering analysis, resulting in four distinct gene expression patterns (Fig. 3i, j; Supplementary information, Fig. S3f, g and Table S2). Genes grouped in pattern 1 highly expressed only in *CXCR4*⁺ aEC while rapidly decreased in HEC and HSPCs, represented by *GJA1*, *NES*, *S100A10* and *PREX1*, were related to the regulation of cytoskeleton organization. Genes in both pattern 2 and pattern 3 were related to angiogenesis, EC migration and cell

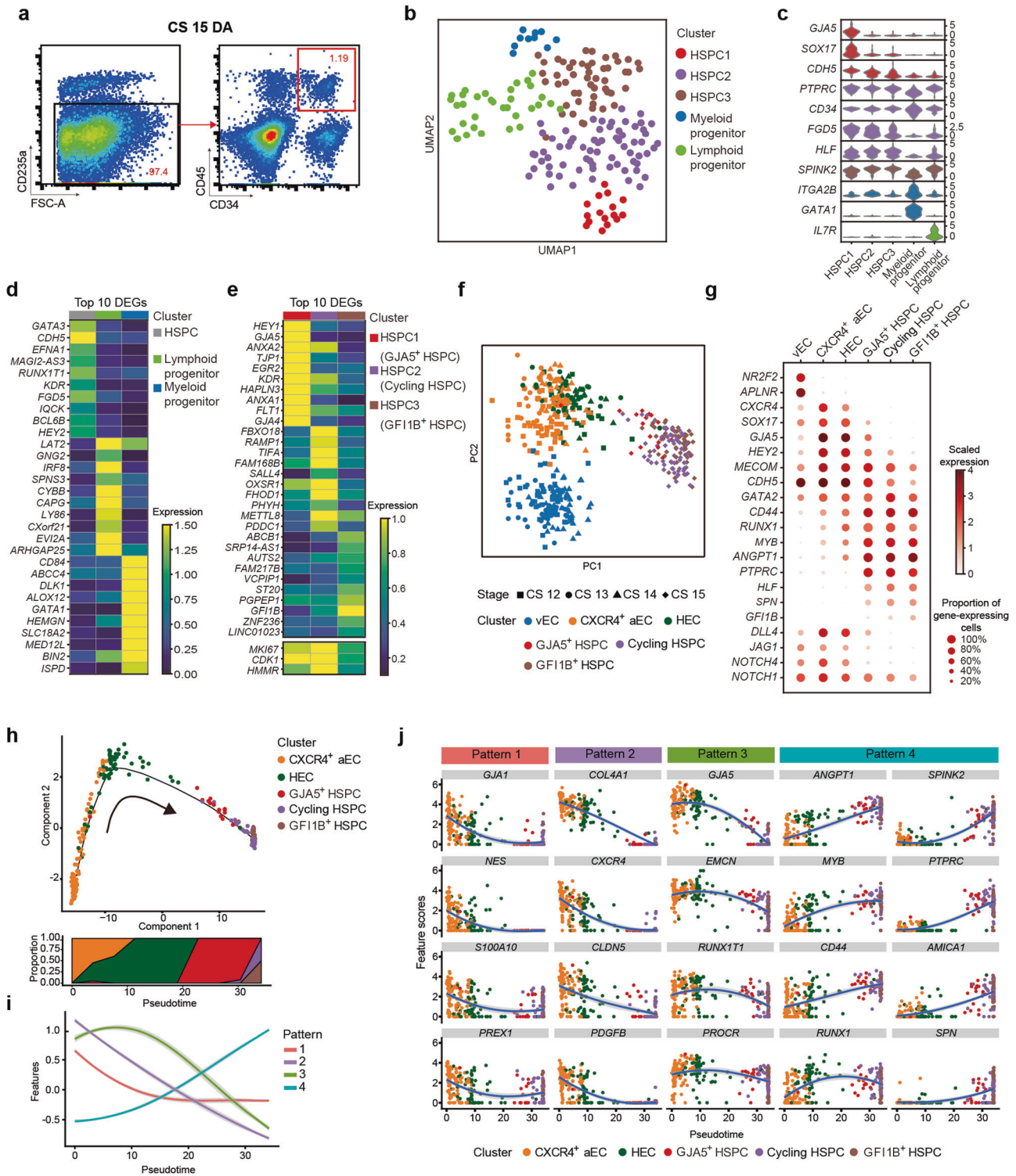


Fig. 3 The developmental path from arterial ECs via HSC-primed HECs to HSPCs in the AGM region. **a** Sorting strategy of CD235a⁻CD45⁺CD34⁺ hematopoietic progenitors in CS 15 dorsal aorta. **b** Identities of five cell populations in CS 15 dorsal aorta visualized by UMAP. **c** Violin plots showing the expression of feature genes in each cell cluster. **d** Heatmap showing the average expressions of top 10 DEGs expressed in HSPC (HSPC1/2/3), Myeloid progenitor and Lymphoid progenitor clusters. **e** Heatmap showing the average expressions of top 10 DEGs in HSPC1 (GJA5⁺ HSPC), HSPC2 (Cycling HSPC) and HSPC3 (GFI1B⁺ HSPC) clusters. **f** PCA plot of vEC, CXCR4⁺ aEC, HEC, GJA5⁺ HSPC, Cycling HSPC and GFI1B⁺ HSPC. **g** Dot plots showing the scaled expression level of feature genes in the indicated clusters. Expression of four Notch signaling pathway genes is shown at the bottom. **h** Trajectory analysis by Monocle 2 combining two aEC sub-clusters (CXCR4⁺ aEC and HEC) from CS 12/13/14 with three HSPC clusters (GJA5⁺ HSPC, Cycling HSPC and GFI1B⁺ HSPC) from CS 15 indicates the developmental path from arterial ECs to HSPCs. Dynamic changes of proportion of clusters are shown on the bottom. **i** Four distinct gene expression patterns along the pseudotime axis inferred by Monocle 2. **j** Expression of representative genes of each pattern along pseudotime axis inferred by Monocle 2.

junction organization, with the former simultaneously sharing the biological terms with pattern 1 (Supplementary information, Fig. S3g). Genes in pattern 2 showed gradual decrease in expression accompanied with hematopoietic specification, represented by *COL4A1*, *CXCR4*, *CLDN5* and *PDGFB* (Fig. 3j; Supplementary information, Fig. S3f). Genes in pattern 3 maintained or even slightly increased expression level upon the initial step of hemogenic fate choice of arterial ECs, while being down-regulated in HSPCs, represented by *GJA5*, *EMCN*, *RUNX1T1* and *PROCR* (Fig. 3j; Supplementary information, Fig. S3f). Some of these genes were validly studied in HSCs. In mice, functional HSCs in E11.5 AGM region are exclusively within *Emcn*⁺ cells based on CD45⁺CD41⁺ phenotype.⁵² *PROCR* is reported to efficiently enrich HSCs in human fetal liver⁵³ and its homolog has been proven to be a marker for type I and type II pre-HSCs in mice.⁵⁴ *RUNX1T1*, whose homolog in mice, with other five TFs, could impart multi-lineage potential to lineage-primed hematopoietic progenitors in reprogramming research, implying its function in initiating the molecular program towards multi-potential HSPCs.⁵⁵ Genes in pattern 4 were those up-regulated in HSPCs, including the ones already highly expressed in HEC as compared to *CXCR4*⁺ aEC, such as *ANGPT1*, *MYB*, *CD44* and *RUNX1* (Fig. 3j; Supplementary information, Fig. S3f), and also those specifically expressed in HSPCs, such as *SPINK2*, *PTPRC*, *AMICA1* and *SPN* (Fig. 3j; Supplementary information, Fig. S3f). Of note, *SPINK2* has been reported to be expressed in HSCs in human umbilical cord blood.³¹

A distinct HEC population lacking arterial feature existed prior to HSC-primed HECs in human embryos

As the HSC-primed HECs in human embryos could be detected at CS 12, to explore whether HECs existed at stages much earlier, cells from CS 10 body part and CS 11 CH of embryo proper were collected for droplet-based scRNA-seq and analyzed together (Supplementary information, Fig. S1a, b). The strategy applied to define different populations was the same as that used in analysis for the data from CS 13 DA (Fig. 1a–c). For clearer presentation, populations from CS 10/11 were prefixed with “early” while those from CS 13 were prefixed with “late” (Supplementary information, Table S3). Endothelial population (Endo), featured by *CDH5* and *SOX7* expression, and hematopoietic population (Hema) with *RUNX1* expression much adjacent to Endo were picked out from CS 10/11 dataset and re-analyzed in searching of HECs (Supplementary information, Fig. S4a, b). Another hematopoietic population showing clear macrophage identity evidenced by *C1QC* and *CD68* expression was not included (Supplementary information, Fig. S4a, b).

Three EC populations and two hematopoietic populations were readily recognized by their distinct feature genes. They were vascular ECs with primitive and venous features (early EC), cardiac ECs (early cEC), arterial ECs (early AEC), megakaryocytic progenitors (early Mega) and erythroid progenitors (early Ery), respectively (Fig. 4a–c; Supplementary information, Fig. S4c). A putative HEC population, a part of which met the strict definition of HEC by expressing endothelial genes (*CDH5* and *ERG*) and *RUNX1* but lacking expression of hematopoietic marker genes (*ITGA2B*, *SPN* and *PTPRC*), was identified and annotated as early HEC (Fig. 4a, b). The early HEC was located nearer to EC populations than to hematopoietic populations in Uniform Manifold Approximation and Projection (UMAP) (Fig. 4a), implying their endothelial identity. Specifically, they highly expressed *CD44* and *SPI1* as compared to other endothelial and hematopoietic populations (Fig. 4c). To strictly comply with the definition of HEC, a part of cells within early HEC cluster expressing specific hematopoietic markers (*ITGA2B*, *SPN* or *PTPRC*), which was located further away from EC populations, were excluded from early HEC cluster and separately named as early HC in the subsequent analysis.

Different molecular features were found in early and late HEC. *ORAI1*, *PPP1R14B* and *S100A10* were among the top of DEGs

expressed in early HEC (Fig. 4d; Supplementary information, Table S4). Expression of genes representing an arterial feature including *HEY1* and *SOX17* was enriched in late HEC (Fig. 4d). *IL11RA* was also up-regulated in late HEC, accordant with the multiple biological function of its ligand IL11 in lymphohematopoietic cells⁵⁶ (Fig. 4d). Gene ontology (GO) terms related to oxygen transport, leukocyte cell-cell adhesion and differentiation were enriched in early HEC (Fig. 4e). In comparison, those related to endothelial cell differentiation, development and vasculogenesis were enriched in late HEC, indicating a more mature endothelial feature of them (Fig. 4e). Pathways related to NOTCH1 intracellular domain, positive epigenetic regulation of rRNA expression and RNA polymerase I transcription were enriched in late HEC, revealing a Notch-dependent arterial property together with an activated transcriptional and translational feature of late HEC (Fig. 4e).

Since HSC emergence occurs almost 10 days later than CS 10 (22–23 dpc),¹ it would be very likely that HECs existing at CS 10/11 (early HEC) and CS 13 (late HEC) had different derivations and represented endothelial precursors of different hematopoietic populations. Thus, endothelial, hemogenic and hematopoietic populations collected from CS 10/11 (Fig. 4a) were pooled with their counterparts from CS 13 embryo (Fig. 1a) for further analysis (Fig. 4f). Dimension reduction showed that a small population expressing the primitive EC feature gene *ETV2*,⁵⁷ which was predominantly sampled from CS 10/11, was localized in the center of EC populations, with one direction towards venous feature and the other towards arterial feature specification (Fig. 4f, g). In contrast to late HEC that was specified from late AEC with the strongest arterial feature, early HEC was segregated from the ECs showing ambiguous arteriovenous feature, indicating their generation did not transit through an arterial stage (Fig. 4f, g). Therefore, two distinct HEC populations formed two bridges connecting EC populations with hematopoietic populations (Fig. 4f, upper right panel). Notably, few cells in early HEC were from CS 11 (3 cells) while plenty of them (26 cells) from CS 10, albeit the numbers of endothelial populations sampled from both stages were comparable (Fig. 4a; Supplementary information, Fig. S4d). The temporal gap between the detection of two HEC populations excluded the possibility that late HEC might be developed from early HEC, and more likely, they developed independently. Spatiotemporal analyses have revealed that IAHCs firstly emerge at CS 12 (27 dpc) and disappear by CS 16 (35–48 dpc).³⁴ Early HECs mainly existing at CS 10 were thus considered not to be related to arterial vasculatures, in line with their lacking of arterial property (Fig. 4d). It is believed that the development of definitive multi-lineage hematopoietic progenitors that can generate lymphoid cells are restricted to arterial vessels during embryogenesis,⁵⁸ the early HEC detected in this study thus should not be related to multi-potential HSPC generation.

Computational analysis of the heterologous cellular interactions for the development of HSC-primed HECs in human embryos The ventral wall of dorsal aorta is the site where the first HSCs in human embryos are detected.³ However, the cellular constitution and molecular basis of the dorsal aortic niche remains unknown. Using the scRNA-seq dataset sampled from the anatomically dissected dorsal aorta in the AGM region at CS 13, we investigated the previously unknown cellular components and potential cellular interactions acting on the HECs, the development of which serves as the initiation of the cell fate specification towards HSCs. In addition to endothelial and hematopoietic populations, five cell populations, including epithelial cell (Epi) and four molecularly different mesenchymal cell populations (Mes), were identified (Fig. 1a). Combined with feature genes and GO terms, these populations were readily recognized (Fig. 1c; Supplementary information, Fig. S1c, d). Epithelial cells, featured by *EPCAM* and

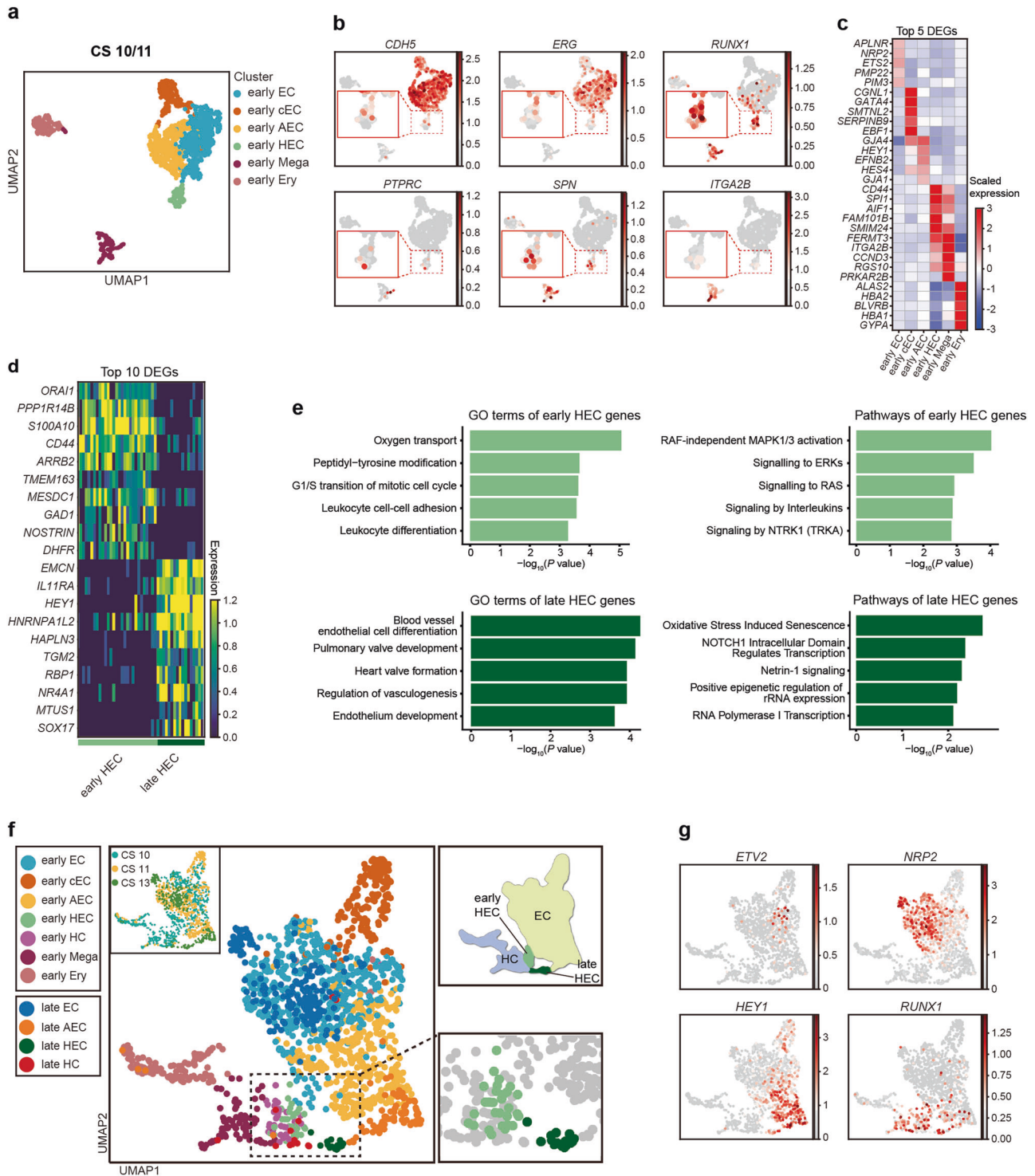


Fig. 4 Different features and origins of the early and late HECs. **a** UMAP showing early endothelial and hematopoietic populations in CS 10–11 of human embryos. Early HEC lies more closely to endothelial cell populations. **b** Expression patterns of typical endothelial and hematopoietic genes in early populations (CS 10 and CS 11). **c** Heatmap showing the average expressions of top 5 DEGs enriched in distinct populations. **d** The expression of top 10 DEGs in early and late HEC. **e** GO:BP terms (left) and pathways (right) enriched in early HEC (upper) and late HEC (lower). **f** UMAP of endothelial and hematopoietic cells from early (CS 10 and CS 11) and late (CS 13) stages with different populations mapped on it. Two distinct HEC clusters at high magnification are shown to the lower right. The distribution of two HEC clusters and endothelial (EC) and hematopoietic (HC) populations on UMAP is shown as schematic to the upper right. **g** UMAP with the expressions of indicated genes mapped on it

CDX2 expression (Fig. 1b, c), were likely the cells from developing gastrointestinal tract adjacent to the ventral wall of dorsal aorta.^{59,60} *WT1*⁺ Mes was probably from the mesonephros tissues nearby the dorsal aorta.⁶¹ The remaining three mesenchymal

populations extensively expressed *PDGFRA* (Fig. 1b), which marks the mesenchyme around the dorsal aorta.⁶² Of note, both *Dlk1* and *Fbln5* are expressed in aortic smooth muscle cells in mice.^{63,64} The *TWIST2*⁺ Mes seemed to have multiple differentiation

capacity, and might be the undifferentiated mesenchymal progenitors.⁶⁵

The cellular interactions were represented by the simultaneous expression of heterologous ligand–receptor gene pairs. Regarding mesenchymal and epithelial populations as potential niche components, only ligands expressed by these populations and receptors expressed by HEC were considered (Fig. 5a). Among a total of 128 heterologous ligand–receptor pairs detected, pathways regulating cytokine–cytokine receptor interaction, axon guidance and endocytosis, Hedgehog signaling pathway, and TGF- β signaling pathway were the pathways predominantly involved (Fig. 5b). Those related to NOTCH signaling pathway, focal adhesion, MAPK signaling pathway, dorso-ventral axis formation and mTOR signaling pathway were also recurrent (Fig. 5b). The ligand–receptor gene pairs exhibited different expression patterns among distinct stromal populations when coupled with HEC (Fig. 5c; Supplementary information, Fig. S5). Among three sub-aortic mesenchymal populations, DLK1⁺ Mes and FBLN5⁺ Mes, paired with HEC, showed relatively similar expression patterns of heterologous ligand–receptor gene pairs, while TWIST2⁺ Mes showed only a few interactions with HEC (Fig. 5c). DLK1–NOTCH1 were presented in DLK1⁺ Mes and FBLN5⁺ Mes, when coupled with HEC (Fig. 5d; Supplementary information, Fig. S5). More specifically, SPP1–CD44, WNT2B–FZD4 and IL11–IL11RA pairs were presented in TWIST2⁺ Mes, DLK1⁺ Mes and FBLN5⁺ Mes, when coupled with HEC, respectively (Fig. 5d). BMP4-involved pairs were shared by WT1⁺ Mes, DLK1⁺ Mes and FBLN5⁺ Mes, including BMP4–BMP2–BMPR1A, BMP4–AVR2A–BMPR1A and BMP4–AVR2B–BMPR1A (Fig. 5d; Supplementary information, Fig. S5), accordant with previous report that BMP4 is expressed in the ventral sub-aortic mesenchyme of human embryos⁶⁶ and the suggested role of Bmp4 in HSC generation in mice.^{67,68}

Considering the special architecture of endothelial layer in the dorsal aorta, cytokines and chemokines together with Notch signaling were taken into consideration in cell–cell interaction analysis. The interactions between either EC or AEC and HEC included NOTCH-related (JAG–NOTCH and DLL–NOTCH), TGF β -related (TGFB1–TGFB1, TGFB1–TGFB2 and BMP2–SMO) and VEGF-related (VEGFA–FLT1, VEGFA–KDR and VEGFA–NRP1) ligand–receptor interactions (Fig. 5e). These interactions have been widely studied in vertebrate embryonic hematopoiesis and for in vitro generation of HSCs from human PSC.^{69,70} However, their functional involvement in human HSC generation needs further investigations.

DISCUSSION

In the present study with the precious human embryonic tissues, we constructed for the first time a genome-scale gene expression landscape for HSC generation in the AGM region of human embryos, in which we focused specifically on the cell populations and molecular events involved in endothelial-to-hematopoietic transition (Fig. 6). We transcriptomically identified the HSC-primed HECs, which meet the widely acknowledged criteria for HECs that simultaneously express core feature genes of ECs (such as *CDH5* and *SOX7*) and critical hematopoietic TFs (such as *RUNX1* and *MYB*) but yet not specific surface markers of hematopoietic cells (including *PTPRC* and *SPN*)^{9,71} (Fig. 2c, i). The HECs exhibited a much more intimate relationship with vascular ECs than with the functionally validated HSPCs showing an immunophenotype of CD45⁺CD34⁺ (Fig. 3f), confirming their endothelial identity and origin. The finding is in line with the notion in the mouse embryos that HECs and non-HECs in the AGM region show great similarity in transcriptome⁷² (Fig. 3f, g). Of note, *RUNX1* was ranked as the most significant DEG in HEC when the EC population with an arterial feature in the AGM region was further sub-divided in an unsupervised way (Fig. 2f). Together with the specific GO terms and enriched pathways (Fig. 2g; Supplementary information,

Fig. S2g), the transcriptomically identified HECs here should be recognized as the earliest cell population that choose the fate specification towards HSCs, which has not been uncovered in human embryos.

We revealed that the HSPCs in CS 15 AGM region showed certain levels of arterial features but were nearly absent of venous feature (Fig. 3g), in line with the reports in the mouse embryos that functional type I and type II pre-HSCs in the AGM region express evident arterial markers but much lower levels of venous markers.⁵⁴ The data suggested an arterial EC origin of emerging HSCs in human, in accord with the apparent arterial feature of the HECs we transcriptomically identified in the CS 12–14 AGM region (Fig. 3g), which is previously unknown in human embryos. Concerning the in vitro differentiation system of human PSCs, the arterial-specific markers *DLL4* and *CXCR4* can be used to identify HECs with lympho-myeloid potential.¹⁴ Interestingly, we found that the HEC sub-population showed much lower *CXCR4* expression than the other one with comparable arterial feature in the AGM (Fig. 2f). Thus, albeit *CXCR4* is regarded as an arterial gene,^{17,73} *CXCR4*⁺ ECs represented a proportion but not all of arterial ECs, showing much less hemogenic characteristics. Therefore, it can be explained that *CXCR4*⁺ cells generated from human PSCs in vitro lack direct hematopoietic potential.⁷³ Our finding emphasizes the importance of discriminating cell identity transcriptomically at single-cell resolution and further suggests that *CXCR4* expression might be rapidly down-regulated upon the hemogenic specification in the arterial ECs.

Transcriptionally, the homing receptor *CD44* was expressed in a subset of ECs with arterial features but seldom in venous ECs of the AGM region in human embryos (Fig. 2b, c). Importantly, almost all the cells with hemogenic characteristics were enriched in the immunophenotypic CD44⁺ EC population, and the expression of *CD44* was gradually increased along the hematopoietic specification (Fig. 3g). The results are in line with the histological finding that *CD44* is expressed in the IAHC cells, as well as the nearby ECs lining the inner layer of the dorsal aorta in human embryos at the similar stages.⁷⁴ Therefore, together with *CD44* labeling, which constituted an average 7.4% of CD34⁺CD45[−] ECs in the AGM region, the HECs in human embryos are enriched, for the first time, more than 10-fold as compared to that by pan-EC markers. Combining flow cytometry with computational identification, the transcriptomically defined HECs constituted about 1/40–1/30 of the ECs in the AGM region. Previous study has reported the frequency of blood-forming ECs in human AGM at about 1/150 of CD34⁺CD45[−] cells at 30 dpc (CS 13), evaluated by an in vitro co-culture system.⁷⁵ It remains to be determined whether all the transcriptomically defined HECs in the present study are functional, especially with improvement of the culture system and even ex vivo functional evaluation.

In an effort to determine the earliest time point at which the intra-embryonic HECs can be detected in human embryos, we unexpectedly revealed two temporally and molecularly distinct HEC populations, with the earlier one lacking obvious arterial features (Fig. 4c). It has been proven that distinct waves of hematopoiesis occur sequentially in model organisms including mouse and zebrafish embryos, with both transient definitive non-HSC hematopoiesis and definitive HSCs generated from HECs.^{8,75} We proposed that the previously undefined two distinct intra-embryonic HEC populations might correspond to the putative two definitive waves of hematopoiesis in human embryos (Fig. 4f). As a potential system for regenerating HSPCs in vitro, human PSC-derived lineages in most studies, in which the induction of arterial characteristics has long been ignored, resemble the transient definitive hematopoiesis from yolk sac, which lacks lymphopoiesis.⁷⁶ Recent researches have proved that activation of the arterial program, either through *ETS1* overexpression, by modulating MAPK/ERK signaling pathways, or activating Notch signaling, drives development of HECs with lymphoid potential in human PSC

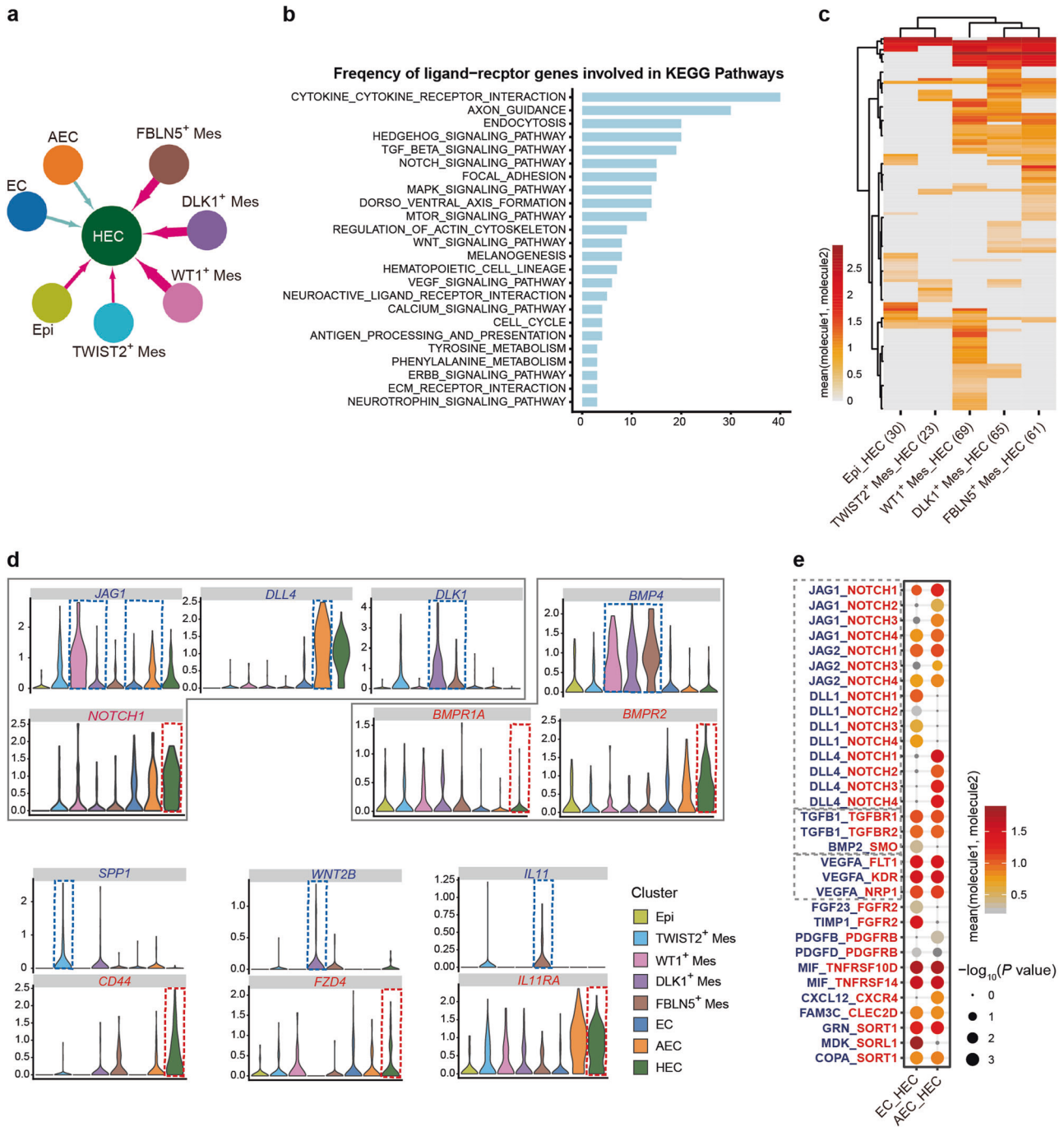


Fig. 5 Computational analysis of the heterologous cellular interactions for the development of HSC-primed HECs. **a** The schematic diagram indicates the heterologous cellular interactions of ligand–receptor pairs between HEC and other clusters including EC, AEC, epithelial and four mesenchymal cell clusters. The direction of arrows indicates from the ligands to the corresponding receptors. Different colors of arrows indicate different databases for ligand–receptor analysis. The thickness of arrows indicates the relative number of ligand–receptor pairs detected. **b** Frequency of ligand–receptor genes involved in KEGG pathways. **c** Heatmap showing the average expression of molecules for each ligand–receptor pair in distinct stromal clusters when coupled with HEC. The hierarchical clustering result of cell populations (columns) and ligand–receptor pairs (rows) is shown. **d** The expression of ligand (blue) and its receptor (red) genes of indicated pairs. Dashed line frames indicate the populations in which the expression of the given gene met the criteria. **e** Dot plots showing the ligand–receptor pairs between EC/AEC as ligand-expressing cell and HEC as receptor-expressing cell

differentiation system.^{14,17} Therefore, the early and late HECs we defined here, with the arterial feature as one of the most prominent differences between them, should be functionally distinct, at least, in the potential of generating lymphoid cells. The precise functional difference between these two transcriptomically identified intra-embryonic HEC populations in human embryos needs further

investigations. It awaits to be determined whether two HEC populations we transcriptomically identified have the common ancestors, such as the immature HECs proposed in human PSC differentiation system.^{14,17} Much likely, along the arterial specification path from primordial ECs, two cohorts of the intra-embryonic HECs are sequentially segregated out, with one before and the

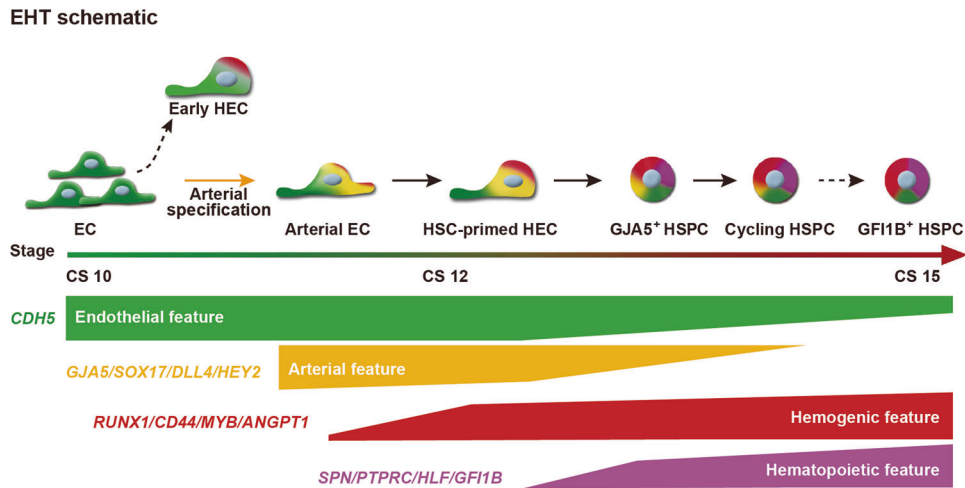


Fig. 6 The schematic showing the endothelial-to-hematopoietic transition (EHT) process in human early embryos. Different cell populations including two types of HECs, developmental stages, and cell type-specific marker genes along the endothelial-to-hematopoietic transition process identified in this study are shown. EC endothelial cells, HEC hemogenic endothelial cells, HSPC hematopoietic stem progenitor cells

other after arterial fate settling (Fig. 6). It is extremely pivotal to discriminate the initial steps for different types of hematopoiesis during human embryogenesis, which are presumably related to quite distinct self-renewal and differentiation capacities and physiological functions. More importantly, this will provide crucial clues for the *in vitro* blood regeneration studies.

MATERIALS AND METHODS

Ethics statement and sample collection

Healthy human embryonic samples were obtained with elective medical termination of pregnancy in Affiliated Hospital of Academy of Military Medical Sciences (the Fifth Medical Center of the PLA General Hospital). All experiments were performed in accordance with protocols approved by the Ethics Committee of the Affiliated Hospital of Academy of Military Medical Sciences (ky-2017-3-5), and local and state ethical guidelines and principles. The written informed consent was obtained before sample collection. Carnegie stages (CS) were used^{77,78} to determine the stages of embryos according to crown-rump length (CRL) measurement and number of somite pairs. Samples used in this study were from CS 10 (23 dpc), CS 11 (24 dpc), CS 12 (27 dpc), CS 13 (29 and 30 dpc), CS 14 (32 dpc) and CS 15 (36 dpc) embryos.

Preparation of single-cell suspensions

Human embryonic body, caudal half and AGM region were isolated and transferred to IMDM medium (Gibco, 12440053) containing 10% Fetal Bovine Serum (HyClone, SH30070.03) on ice. AGM region was washed by PBS and transferred to pre-warmed digestion medium containing 0.1 g/mL Collagenase I (Sigma, C2674), which was shaken vigorously for 30 s and further incubated at 37 °C for about 30 min in incubator with gentle shaking every 5 min to release cells. IMDM medium containing 10% fetal bovine serum was added to terminate digestion. Cells were then collected by centrifuging at 350 × *g* for 6 min, and resuspended in FACS sorting buffer (1 × PBS with 1% BSA) for subsequent staining.

Fluorescence activated cell sorting (FACS) and scRNA-seq

Cells were stained in FACS sorting buffer with specific antibodies for 30 min at 4 °C. The following antibodies were used: BV421-conjugated anti-human CD45 (BD Biosciences, 563879), FITC-conjugated anti-human CD34 (BD Biosciences, 555821), APC-Cy7-conjugated anti-human CD235a (BioLegend, 349116), BV605-conjugated anti-human CD44 (BD Biosciences, 562991) and

PerCP-Cy5.5-conjugated 7-AAD (eBioscience, 00-6993-50). After staining, cells were washed once and resuspended in FACS sorting buffer. Cells were sorted on BD FACS Aria II. Pre-gating was first done for live cells based on a 7-AAD staining. Data analysis was performed using FlowJo V10 software (<https://www.flowjo.com>).

Single-cell library construction

Droplet-based scRNA-seq datasets were produced with a Chromium system (10X Genomics, PN120263) following manufacturer's instruction. For well-based scRNA-seq, sorted single cells in good condition were picked into lysis by mouth pipetting, and the scRNA-seq libraries were constructed based on STRT-seq with some modifications.^{79–81} cDNAs were synthesized using sample-specific 25-nt oligo-dT primer containing 8-nt barcode (TCA-GACGTGTGCTCTCCGATCT-XXXXXXXX-DDDDDDDD-T25, X representing sample-specific barcode while D standing for unique molecular identifiers (UMI), shown in Supplementary information, Table S5) and template switch oligo (TSO) primer for template switching.^{82–84} After reverse transcription and second-strand cDNA synthesis, the cDNAs were amplified by 16 cycles of PCR using ISPCR primer and 3' Anchor primer (see Supplementary information, Table S5). Samples were pooled and purified using Agencourt AMPure XP beads (Beckman, A63882). 4 cycles of PCR were performed to introduce index sequence (shown in Supplementary information, Table S5) and subsequently, 400 ng cDNAs were fragmented to around 300 bp by Covaris S2. After being incubated with Dynabeads MyOne™ Streptavidin C1 beads (Thermo Fisher, 65002) for 1 h at room temperature, cDNA libraries were generated using KAPA Hyper Prep Kit (Kapa Biosystems, kk8505). After adaptor ligation, the libraries were amplified by 8 cycles of PCR using QP2 primer and short universal primer (shown in Supplementary information, Table S5). The libraries were sequenced on Illumina HiSeq X Ten platform in 150 bp paired-end manner (sequenced by Novogene).

Processing scRNA-seq data

Sequencing data from 10X genomics was processed with Cell Ranger software (version 2.1.0) with default mapping arguments. For modified STRT-seq data, raw reads were first split for each cell by specific barcode sequence attached in Read 2. The TSO sequence and polyA tail sequence were trimmed for the corresponding Read 1 after UMI information was aligned to it. Reads with adapter contaminants or low-quality bases ($N > 10\%$) were discarded. Subsequently, we aligned the stripped Read 1 sequences to hg19 human transcriptome (UCSC) using Hisat2

(version 2.10).⁸⁵ Uniquely mapped reads were counted by HTSeq package⁸⁶ and grouped by the cell-specific barcodes. Duplicated transcripts were removed based on the UMI information for each gene. Finally, for each individual cell, the copy number of transcripts of a given gene was the number of the distinct UMIs of that gene.

Quality control

Then quality control was performed to filter low-quality cells. For 10X-derived datasets, we only retained cells that had (1) more than 1000 genes, (2) less than 1,000,000 UMIs, and (3) less than 20% of reads mapped to mitochondrial genes. For modified STRT-seq datasets, UMI values for each cell were grouped in an expression matrix, cells were retained when more than 2000 genes, 10,000 UMIs and less than 20% of mitochondrial gene percentages were detected.

Dimension reduction and clustering

The Seurat (version 2.3.4)⁸⁷ implemented in R (version 3.4) was applied to reduce the dimension of CS 13 10X-derived datasets. UMI count matrix was applied a logarithmic transformation with the scale factor 10,000. High variable genes (HVGs) were calculated using FindVariableGenes function with default parameters except for “x.low.cutoff” = 0.0125. PCA was performed using HVGs, and significant PCs were selected using elbow method to perform dimension reduction and clustering. Cells were projected in 2D space using UMAP with default parameters. Using graph-based clustering function Findcluster with default parameters except for “resolution” = 0.6, we divided the 592 cells into 8 transcriptionally similar clusters. For CS 12/13/14 modified STRT-seq data, the normalization scale factor was set to 100,000. The PCA was performed using HVGs with parameter “x.low.cutoff” = 0.2, and then UMAP analysis with parameters “n_neighbors” = 10, “min_dist” = 0.3 and clustering with parameter “resolution” = 0.8, resulting in 2 endothelial clusters. For HEC analysis, we recalculated HVGs and performed PCA on the aEC subset, and PC1 and PC3 captured most of the variation between populations. CS 15 modified STRT-seq data was subjected to Scanpy (version 1.4.2),⁸⁸ 5 clusters were identified using Louvain clustering with 5 nearest neighbors and the first 20 PCs, and UMAP was used for dimension reduction and visualization of gene expression.

Integrated analysis of CS 10/11/13 10X-derived data

To account for batch differences among CS 10, CS 11 and CS 13 10X-derived datasets, we used the batch balanced KNN (BBKNN), a batch effect removal tool in Scanpy package. BBKNN actively combats technical artifacts by splitting data into batches and finding a smaller number of neighbors for each cell within each of the groups, which helps create connections between analogous cells in different batches without altering the counts or PCA space. We took the union of the top 2000 genes with the highest expression and dispersion from both datasets, which were used for PCA. We then aligned the subspaces on the basis of the first 20 PCs and selected top 10 neighbors to report for each batch, which generated new PCs that were used for subsequent analysis.

Differential expression analysis

To find DEGs among different clusters, we performed non-parametric Wilcoxon rank sum tests, as implemented in Seurat. DEGs with adjusted *P* value less than 0.01 were thought to be significant. We applied tl.rank_genes_groups function with default parameters to identify DEGs, which were then filtered with “min_fold_change” = 0.5, “min_in_group_fraction” = 0.2, as implemented in Scanpy.

Surface markers and TFs

Surface markers and TF lists (Supplementary information, Table S6) were downloaded from Cell Surface Protein Atlas (<http://wlab.ethz.ch/cspa>) and HumanTFDB3.0 (<http://bioinfo.life.hust.edu.cn/HumanTFDB>), respectively.

Correlation analysis

Pearson correlation analysis was performed using the top 50 PCs calculated in CS 13 10X-derived dataset. For genes positively correlated with *RUNX1*, we performed Pearson correlation using corr.test function in psych R package and *P* values were adjusted using “BH” method.

Monocle 2 analysis

Pseudotime trajectory was constructed with the Monocle 2 package (v2.10.1)⁸⁹ according to the documentation (<http://cole-trapnell-lab.github.io/monocle-release>). Ordering genes were selected based on PCA loadings. The Discriminative Dimensionality Reduction with Trees (DDRTree) method was used to reduce data to two dimensions. To investigate the different patterns of gene expression during this developmental path, significant DEGs along pseudotime were identified by Monocle 2's differentialGeneTest function. For identification of major patterns, we filtered out those genes with average normalized expression less than 1 in all involved clusters and then clustered the retained genes into four distinct patterns using k-means clustering.

Gene functional annotation analysis

GO term and KEGG pathway enrichment were done on DEGs using clusterProfiler R package (version 3.10.1)⁹⁰ with default parameters.

Cell cycle regression

To reduce the variation in cell cycle status which contribute to the heterogeneity in scRNA-seq datasets, we performed CellCycleScoring function in Seurat to evaluate the cell cycle status using the previously reported G1/S and G2/M phase-specific genes.^{84,91} Then using regressout in ScaleData function to remove cell cycle effects.

Cellular interaction analysis

Cellular interaction analysis was performed using CellphoneDB software (version 2.0)⁹² with default parameters. Significant ligand–receptor pairs were filtered with *P* value less than 0.01. Pairs of which ligand genes expressed in Mes, Epi, EC clusters and receptor genes expressed in HEC were retained. Visualization of directed network was done using Cytoscape (version 3.6.0).

Data availability

The scRNA-seq data reported in this study have been deposited in NCBI's Gene Expression Omnibus (GEO) with the accession number GSE135202. All other relevant data in this study are available from the corresponding authors upon reasonable request.

ACKNOWLEDGEMENTS

This study was supported by grants from the National Key R&D Program of China (2017YFA0103401, 2016YFA0100601, and 2019YFA0110201), the National Natural Science Foundation of China (31425012, 81890991, 31930054, 31871173, 81800102 and 81900115), the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (2017ZT075347), the Beijing Municipal Science & Technology Commission (Z171100000417009), the State Key Laboratory of Proteomics (SKLP-K201502) and the Key Research and Development Program of Guangdong Province (2019B020234002).

AUTHOR CONTRIBUTIONS

B.L. and Y.L. designed the study. Y.Z. and Z.B. performed the sample preparation and the FACS with help from Y.N.; Z.B. and C.L. performed the single-cell RNA-sequencing; L.B., C.M., Y.Z., and Z.B. collected and prepared the samples; J.H. and Z.L. performed

the bioinformatics analysis with help from Y.G. and J.D.; B.L., Y.L., Y.Z., J.H., Z.B., and Z.L. wrote the manuscript.

ADDITIONAL INFORMATION

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41422-019-0228-6>.

Competing interests: The authors declare no competing interests.

REFERENCES

- Ivanovs, A. et al. Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. *J. Exp. Med.* **208**, 2417–2427 (2011).
- Ivanovs, A. et al. Human haematopoietic stem cell development: from the embryo to the dish. *Development* **144**, 2323–2337 (2017).
- Ivanovs, A., Rytsov, S., Anderson, R. A., Turner, M. L. & Medvinsky, A. Identification of the niche and phenotype of the first human hematopoietic stem cells. *Stem Cell Rep.* **2**, 449–456 (2014).
- Notta, F. et al. Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science* **333**, 218–221 (2011).
- Zhang, Y. et al. VE-cadherin and ACE co-expression marks highly proliferative hematopoietic stem cells in human embryonic liver. *Stem Cells Dev.* **28**, 165–185 (2019).
- Oberlin, E. et al. VE-cadherin expression allows identification of a new class of hematopoietic stem cells within human embryonic liver. *Blood* **116**, 4444–4455 (2010).
- Tavian, M. et al. Aorta-associated CD34+ hematopoietic cells in the early human embryo. *Blood* **87**, 67–72 (1996).
- Zovein, A. C. et al. Fate tracing reveals the endothelial origin of hematopoietic stem cells. *Cell Stem Cell* **3**, 625–636 (2008).
- Chen, M. J., Yokomizo, T., Zeigler, B. M., Dzierzak, E. & Speck, N. A. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature* **457**, 887–891 (2009).
- Swiers, G. et al. Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. *Nat. Commun.* **4**, 2924 (2013).
- Garcia-Alegria, E. et al. Early human hemogenic endothelium generates primitive and definitive hematopoiesis in vitro. *Stem Cell Rep.* **11**, 1061–1074 (2018).
- Teichweyde, N. et al. HOXB4 promotes hemogenic endothelium formation without perturbing endothelial cell development. *Stem Cell Rep.* **10**, 875–889 (2018).
- Kang, H. et al. GATA2 is dispensable for specification of hemogenic endothelium but promotes endothelial-to-hematopoietic transition. *Stem Cell Rep.* **11**, 197–211 (2018).
- Uenishi, G. I. et al. NOTCH signaling specifies arterial-type definitive hemogenic endothelium from human pluripotent stem cells. *Nat. Commun.* **9**, 1828 (2018).
- Cortes, F., Debacker, C., Peault, B. & Labastie, M. C. Differential expression of KDR/VEGFR-2 and CD34 during mesoderm development of the early human embryo. *Mech. Dev.* **83**, 161–164 (1999).
- Boisset, J. C. et al. In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature* **464**, 116–120 (2010).
- Park, M. A. et al. Activation of the arterial program drives development of definitive hemogenic endothelium with lymphoid potential. *Cell Rep.* **23**, 2467–2481 (2018).
- Crisan, M. et al. BMP signalling differentially regulates distinct haematopoietic stem cell types. *Nat. Commun.* **6**, 8040 (2015).
- Souilhol, C. et al. Inductive interactions mediated by interplay of asymmetric signalling underlie development of adult haematopoietic stem cells. *Nat. Commun.* **7**, 10784 (2016).
- Lis, R. et al. Conversion of adult endothelium to immunocompetent haematopoietic stem cells. *Nature* **545**, 439–445 (2017).
- Pereira, C. F. et al. Hematopoietic reprogramming in vitro informs in vivo identification of hemogenic precursors to definitive hematopoietic stem cells. *Dev. Cell* **36**, 525–539 (2016).
- Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
- Buschmann, I. et al. Pulsatile shear and Gja5 modulate arterial identity and remodeling events during flow-driven arteriogenesis. *Development* **137**, 2187–2196 (2010).
- Aquila, G. et al. Distinct gene expression profiles associated with Notch ligands Delta-like 4 and Jagged1 in plaque material from peripheral artery disease patients: a pilot study. *J. Transl. Med.* **15**, 98 (2017).
- Red-Horse, K. & Siekmann, A. F. Veins and arteries build hierarchical branching patterns differently: bottom-up versus top-down. *Bioessays* **41**, e1800198 (2019).
- Aranguren, X. L. et al. Unraveling a novel transcription factor code determining the human arterial-specific endothelial cell signature. *Blood* **122**, 3982–3992 (2013).
- Somekawa, S. et al. Tmem100, an ALK1 receptor signaling-dependent gene essential for arterial endothelium differentiation and vascular morphogenesis. *Proc. Natl Acad. Sci. USA* **109**, 12064–12069 (2012).
- Lathen, C. et al. ERG-APLN axis controls pulmonary venule endothelial proliferation in pulmonary veno-occlusive disease. *Circulation* **130**, 1179–1191 (2014).
- Yuan, L. et al. Abnormal lymphatic vessel development in neuropilin 2 mutant mice. *Development* **129**, 4797–4806 (2002).
- Wilkinson, A. C. et al. Single-cell analyses of regulatory network perturbations using enhancer-targeting TALEs suggest novel roles for PU.1 during haematopoietic specification. *Development* **141**, 4018–4030 (2014).
- He, X. et al. Differential gene expression profiling of CD34+ CD133+ umbilical cord blood hematopoietic stem progenitor cells. *Stem Cells Dev.* **14**, 188–198 (2005).
- Potente, M. & Makinen, T. Vascular heterogeneity and specialization in development and disease. *Nat. Rev. Mol. Cell Biol.* **18**, 477–494 (2017).
- Zhou, X., Liao, W. J., Liao, J. M., Liao, P. & Lu, H. Ribosomal proteins: functions beyond the ribosome. *J. Mol. Cell Biol.* **7**, 92–104 (2015).
- Tavian, M., Hallais, M. F. & Peault, B. Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo. *Development* **126**, 793–803 (1999).
- Cai, X. et al. Runx1 deficiency decreases ribosome biogenesis and confers stress resistance to hematopoietic stem and progenitor cells. *Cell Stem Cell* **17**, 165–177 (2015).
- Palis, J. Primitive and definitive erythropoiesis in mammals. *Front. Physiol.* **5**, 3 (2014).
- Baker, S. J. et al. B-myb is an essential regulator of hematopoietic stem cell and myeloid progenitor cell development. *Proc. Natl Acad. Sci. USA* **111**, 3122–3127 (2014).
- Zhou, B. O., Ding, L. & Morrison, S. J. Hematopoietic stem and progenitor cells regulate the regeneration of their niche by secreting Angiopoietin-1. *Elife* **4**, e05521 (2015).
- Mucenski, M. L. et al. A functional c-myc gene is required for normal murine fetal hepatic hematopoiesis. *Cell* **65**, 677–689 (1991).
- Hyde, R. K. et al. Cbfb/Runx1 repression-independent blockage of differentiation and accumulation of Csf2rb-expressing cells by Cbfb-MYH11. *Blood* **115**, 1433–1443 (2010).
- Fang, J. S. et al. Shear-induced Notch-Cx37-p27 axis arrests endothelial cell cycle to enable arterial specification. *Nat. Commun.* **8**, 2149 (2017).
- Suchting, S. et al. The Notch ligand Delta-like 4 negatively regulates endothelial tip cell formation and vessel branching. *Proc. Natl Acad. Sci. USA* **104**, 3225–3230 (2007).
- Kueh, H. Y. & Rothenberg, E. V. Regulatory gene network circuits underlying T cell development from multipotent progenitors. *Wiley Int. Rev. Syst. Biol. Med.* **4**, 79–102 (2012).
- Hovanes, K., Li, T. W. & Waterman, M. L. The human LEF-1 gene contains a promoter preferentially active in lymphocytes and encodes multiple isoforms derived from alternative splicing. *Nucleic Acids Res.* **28**, 1994–2003 (2000).
- Gazit, R. et al. Fgd5 identifies hematopoietic stem cells in the murine bone marrow. *J. Exp. Med.* **211**, 1315–1331 (2014).
- Komorowska, K. et al. Hepatic leukemia factor maintains quiescence of hematopoietic stem cells and protects the stem cell pool during regeneration. *Cell Rep.* **21**, 3514–3523 (2017).
- Yokomizo, T. et al. Hlf marks the developmental pathway for hematopoietic stem cells but not for erythro-myeloid progenitors. *J. Exp. Med.* **216**, 1599–1614 (2019).
- Thambyrajah, R. et al. Gfi1 proteins orchestrate the emergence of haematopoietic stem cells through recruitment of LSD1. *Nat. Cell Biol.* **18**, 21–32 (2016).
- Lizama, C. O. et al. Repression of arterial genes in hemogenic endothelium is sufficient for haematopoietic fate acquisition. *Nat. Commun.* **6**, 7739 (2015).
- Bos, F. L., Hawkins, J. S. & Zovein, A. C. Single-cell resolution of morphological changes in hemogenic endothelium. *Development* **142**, 2719–2724 (2015).
- Garside, V. C. et al. SOX9 modulates the expression of key transcription factors required for heart valve development. *Development* **142**, 4340–4350 (2015).
- Matsubara, A. et al. Endomucin, a CD34-like sialomucin, marks hematopoietic stem cells throughout development. *J. Exp. Med.* **202**, 1483–1492 (2005).
- Subramaniam, A., Talkhoncheg, M. S., Magnusson, M. & Larsson, J. Endothelial protein C receptor (EPCR) expression marks human fetal liver hematopoietic stem cells. *Haematologica* **104**, e47–e50 (2019).
- Zhou, F. et al. Tracing haematopoietic stem cell formation at single-cell resolution. *Nature* **533**, 487–492 (2016).

55. Riddell, J. et al. Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors. *Cell* **157**, 549–564 (2014).
56. Cherel, M. et al. Molecular cloning of two isoforms of a receptor for the human hematopoietic cytokine interleukin-11. *Blood* **86**, 2534–2540 (1995).
57. Lammerts van Bueren, K. & Black, B. L. Regulation of endothelial and hematopoietic development by the ETS transcription factor Etv2. *Curr. Opin. Hematol.* **19**, 199–205 (2012).
58. Slukvin, I. I. & Uenishi, G. I. Arterial identity of hemogenic endothelium: a key to unlock definitive hematopoietic commitment in human pluripotent stem cell cultures. *Exp. Hematol.* **71**, 3–12 (2019).
59. Lei, Z. et al. EpCAM contributes to formation of functional tight junction in the intestinal epithelium by recruiting claudin proteins. *Dev. Biol.* **371**, 136–145 (2012).
60. Larsen, S., Davidsen, J., Dahlggaard, K., Pedersen, O. B. & Troelsen, J. T. HNF4alpha and CDX2 regulate intestinal YAP1 promoter activity. *Int. J. Mol. Sci.* **20**, <https://doi.org/10.3390/ijms20122981> (2019).
61. Haber, D. A. et al. An internal deletion within an 11p13 zinc finger gene contributes to the development of Wilms' tumor. *Cell* **61**, 1257–1269 (1990).
62. Ding, G., Tanaka, Y., Hayashi, M., Nishikawa, S. & Kataoka, H. PDGF receptor alpha + mesoderm contributes to endothelial and hematopoietic cells in mice. *Dev. Dyn.* **242**, 254–268 (2013).
63. Mirshekar-Syahkal, B. et al. Dlk1 is a negative regulator of emerging hematopoietic stem and progenitor cells. *Haematologica* **98**, 163–171 (2013).
64. Kobayashi, N. et al. A comparative analysis of the fibulin protein family. Biochemical characterization, binding interactions, and tissue localization. *J. Biol. Chem.* **282**, 11805–11816 (2007).
65. Lai, W. T., Krishnappa, V. & Phinney, D. G. Fibroblast growth factor 2 (Fgf2) inhibits differentiation of mesenchymal stem cells by inducing Twist2 and Spry4, blocking extracellular regulated kinase activation, and altering Fgf receptor expression levels. *Stem Cells* **29**, 1102–1111 (2011).
66. Marshall, C. J., Kinnon, C. & Thrasher, A. J. Polarized expression of bone morphogenetic protein-4 in the human aorta-gonad-mesonephros region. *Blood* **96**, 1591–1593 (2000).
67. McGarvey, A. C. et al. A molecular roadmap of the AGM region reveals BMPER as a novel regulator of HSC maturation. *J. Exp. Med.* **214**, 3731–3751 (2017).
68. Durand, C. et al. Embryonic stromal clones reveal developmental regulators of definitive hematopoietic stem cells. *Proc. Natl Acad. Sci. USA* **104**, 20838–20843 (2007).
69. Sturgeon, C. M., Ditadi, A., Awong, G., Kennedy, M. & Keller, G. Wnt signaling controls the specification of definitive and primitive hematopoiesis from human pluripotent stem cells. *Nat. Biotechnol.* **32**, 554–561 (2014).
70. Corada, M. et al. The Wnt/beta-catenin pathway modulates vascular remodeling and specification by upregulating Dll4/Notch signaling. *Dev. Cell* **18**, 938–949 (2010).
71. Vodyanik, M. A., Thomson, J. A. & Slukvin, I. I. Leukosialin (CD43) defines hematopoietic progenitors in human embryonic stem cell differentiation cultures. *Blood* **108**, 2095–2105 (2006).
72. Baron, C. S. et al. Single-cell transcriptomics reveal the dynamic of hematopoietic stem cell production in the aorta. *Nat. Commun.* **9**, <https://doi.org/10.1038/s41467-018-04893-3> (2018).
73. Ditadi, A. et al. Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages. *Nat. Cell Biol.* **17**, 580–591 (2015).
74. Watt, S. M. et al. Functionally defined CD164 epitopes are expressed on CD34(+) cells throughout ontogeny but display distinct distribution patterns in adult hematopoietic and nonhematopoietic tissues. *Blood* **95**, 3113–3124 (2000).
75. Oberlin, E., Tavian, M., Blazsek, I. & Peault, B. Blood-forming potential of vascular endothelium in the human embryo. *Development* **129**, 4147–4157 (2002).
76. Greenfest-Allen, E., Malik, J., Palis, J. & Stoeckert, C. J. Jr. Stat and interferon genes identified by network analysis differentially regulate primitive and definitive erythropoiesis. *BMC Syst. Biol.* **7**, 38 (2013).
77. Gasser, R. F. Atlas of human embryos. *J. Anat.* **120**, 607 (1975).
78. O'Rahilly, R., Muller, F. & Streeter, G. L. *Developmental Stages in Human Embryos: Including a Revision of Streeter's Horizons and a Survey of the Carnegie Collection* (Carnegie Institution of Washington, Washington, DC, 1987).
79. Li, L. et al. Single-cell RNA-Seq analysis maps development of human germline cells and gonadal niche interactions. *Cell Stem Cell* **20**, 891–892 (2017).
80. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
81. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
82. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).
83. Islam, S. et al. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11**, 163–166 (2014).
84. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
85. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
86. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
87. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
88. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
89. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
90. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
91. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
92. Vento-Tormo, R. et al. Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* **563**, 347–353 (2018).



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019