bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dissecting endothelial to haematopoietic stem cell transition by single-cell transcriptomic and functional analyses

Siyuan Hou[1,2,4,9], Zongcheng Li[2,9], Xiaona Zheng[3,9], Yun Gao[5,9], Ji Dong[5,9], Yanli Ni[2,9], Xiaobo Wang[3], Yunqiao Li[3], Xiaochen Ding[3], Zhilin Chang[3], Shuaili Li[3], Yuqiong Hu[5], Xiaoying Fan[5], Yu Hou[5], Lu Wen[5,6], Bing Liu[1,2]*, Fuchou Tang[5,6,7]*, Yu Lan[1,8]*

[1]Key Laboratory for Regenerative Medicine of Ministry of Education, Institute of Haematology, School of Medicine, Jinan University, Guangzhou 510632, China

[2]State Key Laboratory of Experimental Haematology, Fifth Medical Center of Chinese PLA General Hospital, Beijing 100071, China

[3]State Key Laboratory of Proteomics, Academy of Military Medical Sciences, Academy of Military Sciences, Beijing 100071, China

[4]Integrated Chinese and Western Medicine Postdoctoral research station, Jinan University, Guangzhou 510632, China

[5]Beijing Advanced Innovation Center for Genomics and Biomedical Institute for Pioneering Investigation via Convergence, College of Life Sciences, Peking University, Beijing 100871, China

[6]Ministry of Education Key Laboratory of Cell Proliferation and Differentiation, Beijing 100871, China

[7]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China

[8]Guangzhou Regenerative Medicine and Health-Guangdong Laboratory (GRMH-GDL),

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

23 Guangzhou 510530, China

24 [9]These authors contributed equally to this work.

25 *Correspondence: Bing Liu (bingliu17@yahoo.com); Fuchou Tang (tangfuchou@pku.edu.cn)

26 and Yu Lan (rainyblue_1999@126.com)

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Haematopoietic stem cells (HSCs) in adults are believed to be born from hemogenic endothelial cells (HECs) in mid-gestational mouse embryos. Due to rare and transient nature, the HSC-competent ECs have never been stringently identified and accurately captured, let alone their genuine vasculature precursors. Here, we firstly used high-precision single-cell transcriptomics to unbiasedly examine relevant EC populations at continuous developmental stages and transcriptomically identified putative HSC-primed HECs. Combining computational prediction and in vivo functional validation, we precisely captured HSC-competent HECs by newly constructed Neurl3-EGFP reporter mouse model, and realized enrichment further by surface marker combination. Surprisingly, endothelial-haematopoietic bi-potential was rarely but reliably witnessed in culture of single HECs. Noteworthy, primitive vascular ECs experienced two-step fate choices to become HSC-primed HECs, resolving several previously observed contradictions. Taken together, comprehensive understanding of endothelial evolutions and molecular programs underlying HSC-primed HEC specification in vivo will facilitate future investigations directing HSC production in vitro.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

The adult haematopoietic system, consisted mainly of haematopoietic stem cells (HSCs) and their multi-lineage progenies, is believed to be derived from hemogenic endothelial cells (HECs) in mid-gestational embryos [1,2]. It is generally accepted that while still embedded in the endothelial layer and presenting endothelial characteristics, HECs begin to express key hemogenic transcription factor Runx1 and have hemogenic potential [3,4]. Different from haematopoietic progenitors, HECs lack the expression of haematopoietic surface markers, such as CD41 and CD45, which mark the population capable of generating haematopoietic progenies when directly tested in colony-forming unit assays [3,5]. Haematopoietic stem and progenitor cells (HSPCs) are visualized to emerge from aortic endothelial cells (ECs) via a transient and dynamic process called endothelial-to-haematopoietic transition to form intra-aortic haematopoietic clusters (IAHCs) [6-10]. Being located within IAHCs or to the deeper sub-endothelial layers, pre-HSCs serve as the important cellular intermediates between HECs and HSCs, featured by their inducible repopulating capacity and priming with haematopoietic surface markers [11-15]. The specification of HSC-primed HECs is the initial and one of the most pivotal steps for vascular ECs to choose a HSC fate. However, as the precise identity of HSC-primed HECs is not clear, contradictory notions regarding whether primordial or arterial fated ECs are the direct origin of HSC-primed HECs are still on the debate. It is proposed that definitive HECs and arterial ECs represent distinct lineages [16,17]. Moreover, HSCs and arterial ECs are proposed to arise from distinct precursors, characterized by different Notch signaling strengths [18]. Most recently, HSC-primed HECs have been

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transcriptionally identified in human embryos, which present an unambiguous arterial property, indicative of their arterial EC origin [19].

In order to deeply investigate the cellular evolutions and molecular events underlying the specification of HSC-primed HECs and their subsequent commitment to HSPCs, it is necessary to efficiently isolate the HSC-primed HECs, which is proven to be difficult not only because the population is proposed to be small and transient, but also due to the technical challenging to determine their HSC competence [20]. Considering that not only HSCs but also the transient definitive haematopoiesis during embryogenesis are derived from HECs, repopulating capacity is required for the functional evaluation of the HSC-primed HECs. Previous studies have reported the HSC competence of $CD47^{+}$ but not $CD47^{-}$ ECs in embryonic day (E) 10.5 aorta-gonad-mesonephros (AGM) region and both $Dll4^{+}$ and $Dll4^{-}$ ECs in E9.5 para-aortic splanchnopleura (P-Sp) region [11,21]. Nevertheless, the enrichment of the above surface markers is far from efficient. Several transgenic reporter mouse models have been established by which HECs could be distinguished from non-HECs, including *Ly6a-GFP*, GFP transgenic reporter under the control of Runx1 +23 enhancer (*Runx1 +23GFP*) and *Gfi1-Tomato*, and the usage of these reporters largely helps to delineate the process of endothelial-to-haematopoietic transition [3,14,22-25]. Although expected to a certain extent, the HSC competence of the HECs labeled by these reporters has not been functionally validated. Up-to-date, efficient isolation of the HSC-primed HEC population has not yet been achieved.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

With the aim of delineating the molecular events underlying HSC emergence, several single-cell transcriptional profiling studies on HECs, IAHC cells, and HSPCs in the AGM region have been reported in recent years. Using either Runx1 +23GFP or Gfi1-Tomato as the marker of putative HECs, several defined cell populations are transcriptionally profiled by Fluidigm single-cell qPCR or single-cell RNA sequencing (scRNA-seq) [3,14,22]. Moreover, the cellular components of IAHCs are investigated at single-cell level by mechanically picking up single whole IAHCs in the aortas, showing cells with pre-HSC feature are predominantly involved [14]. Interestingly, contradiction still exists regarding whether HECs and non-HECs are molecularly similar and to what extent the two populations are distinguishable [14,22]. Since the enrichment efficiency or specificity of the above markers to define the HEC population might be not enough, an unsupervised screening of the embryonic endothelial pool within haematopoietic tissues is required for the precise recognition of HSC-primed HECs.

Here, we firstly used high-precision single-cell transcriptomics to unbiasedly examine all the EC populations spanning continuous developmental stages covering the presumed time points for the specification of HSC-primed HECs, transcriptomically identified them, and computationally screened for their candidate markers. Based on the consequently precise capture and isolation of the HSC-competent HECs using surface marker combination or newly constructed fluorescent reporter mice, we further decoded the cellular evolutions and molecular programs underlying the stepwise hemogenic fate settling from the initial primordial vascular ECs. A series of new findings, including the endothelial-haematopoietic bi-potential of HECs and the multi-step fate choice for the specification of HSC-primed HECs,

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

unprecedentedly enrich our understanding of HSC generation in vivo and should be extremely critical to inspire new approaches for stepwise HSC regeneration from pluripotent stem cells [26].

RESULTS

Transcriptomic identification of the HECs in AGM region

We first analyzed mouse embryos from E9.5, when the initial IAHC formation in the aorta occurs [7], to the stage of the appearance of HSCs at E11.0 [27] (Supplementary information, Fig. S1a). For each embryo, the embryo proper was isolated and the head, limb buds, heart, visceral bud, and vitelline and umbilical vessels outside the embryo proper were excluded (Fig. 1a). To specifically capture aortic luminal ECs of AGM region, we performed microinjection of fluorescent dye Oregon green into the dorsal aortas of E10.0-E11.0 embryos as reported [12] (Fig. 1a; Supplementary information, Fig. S1b). The sampled cells were purified by FACS as $CD45^{-}CD31^{+}CD144^{+}$, which contained predominantly vascular ECs and $CD41^{+}$ haematopoietic cells. Meanwhile, $CD45^{-}CD31^{-}CD144^{-}$ non-EC cells in the body were used as negative controls (Fig. 1a). We used unique molecular identifier (UMI)-based scRNA-seq method to accurately measure the gene expression profiles within individual cells. In totally 662 sequenced single cells, 597 single-cell transcriptomes passed rigorous quality control. On average we detected 7,035 genes (from 2,266 to 10,843) and 636,418 transcripts (from 103,793 to 2,959,573) expressed in each individual cell (Supplementary information, Fig. S1c).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

131 According to a graph-based clustering approach from Seurat software [28], all cells were
132 separated into six clusters, including one negative (Neg) cluster containing mainly non-EC
133 negative control cells, and five sample clusters comprising almost all FACS-isolated sample
134 cells (Supplementary information, Fig. S1d and Table S1). Featured by the obvious *Runx1*
135 and *Itga2b* (encoding CD41) expression, the haematopoietic cell (HC) cluster was distributed
136 away from the other four vascular EC clusters which presented apparent arterial or venous
137 characteristics (Supplementary information, Fig. S1d, e). One venous EC (vEC) cluster was
138 readily recognized by the exclusive expression of *Nr2f2* in all vascular EC populations
139 (Supplementary information, Fig. S1d, e). Two arterial EC clusters showed similar *Gja5*
140 expression but different level of *Ltbp4* expression [29]. Together with their different sampling
141 stages (mainly from E9.5-E10.0 and E10.5-E11.0, respectively), they were annotated as early
142 arterial EC (earlyAEC) and late arterial EC (lateAEC) cluster, respectively (Supplementary
143 information, Fig. S1d, e). The left one cluster basically met the criteria of the molecular
144 definition of HEC, showing apparent *Runx1* expression upon endothelial property, and was
145 consequently named as HEC cluster (Supplementary information, Fig. S1d, e). To more
146 strictly define the HEC population, cells within Neg cluster and those transcriptionally
147 expressing *Ptprc* (encoding CD45) or *Spn* (encoding CD43) were excluded for the
148 subsequent analysis (Supplementary information, Fig. S1f).
149
150 HEC and other two AEC clusters were further focused as they were either molecularly or
151 spatiotemporally near to each other (Fig. 1b, Supplementary information, Fig. S1d). To
152 exclude the possibility that we failed to identify important populations relevant to hemogenic

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

153 specification in earlyAEC cluster, which contributed evidently to the aortic inner layer of AGM
154 region at E10.0 (Supplementary information, Fig. S1f), we performed forced clustering within
155 the given cluster. *Runx1* (signature of hemogenic specification) was not significantly
156 differentially expressed between the two sub-clusters, suggesting that no population with sign
157 of hemogenic specification was missed by our clustering (Supplementary information, Fig.
158 S1g). Moreover, very few genes were significantly differentially expressed in the forced
159 sub-clusters of HEC, and none of them was related to hemogenic or haematopoietic features,
160 indicative of the largely homogeneous property of the HEC cluster (Supplementary
161 information, Fig. S1g). HEC was reduced promptly in number at E10.5, and became hardly
162 detectable by E11.0 (Fig. 1b; Supplementary information, Fig. S1f). The highly expressed
163 genes in HEC as compared to earlyAEC and lateAEC were mainly enriched in the terms
164 related to cell cycle and ribosome biogenesis (Fig. 1c; Supplementary information, Table S2).
165 Cell cycle analysis demonstrated a remarkably activated cycling in HEC, in sharp contrast to
166 the quiescent state by arterial EC maturation (Fig. 1d). On average, each cell in HEC cluster
167 expressed more mRNA molecules and ribosomal genes than either earlyAEC or lateAEC (Fig.
168 1e; Supplementary information, Fig. S1h), supportive of the globally up-regulated
169 transcriptional and translational activity during hemogenic specification, which was in line with
170 the finding in human embryo that translational initiation is overrepresented in HSC-primed
171 HECs than in arterial ECs [19]. We further evaluated the arteriovenous scores of the
172 populations we defined, and found similar results in mouse and human that HEC rather than
173 haematopoietic populations manifested certain arterial feature (Fig. 1f).
174

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Trajectory analysis by Monocle 2 suggested that along the arterial maturation path from earlyAEC towards lateAEC, HEC was segregated out from earlyAEC at E9.5-E10.0 (Fig. 1g). The gradual up-regulation of hemogenic genes, including *Runx1* and *Spi1*, was accompanied by the gradual down-regulation of both endothelial and arterial genes along the HEC specification pseudotime, with the endothelial-haematopoietic dual-feature of the HEC population presenting as a dynamic continuum (Fig. 1h). The finding was in line with previous report about the reciprocal expression of Runx1 and Sox17 in HECs [30]. To search for the genes that would be potentially meaningful to the distinct fate choices of earlyAEC, those differentially expressed between earlyAEC and its downstream population HEC or lateAEC were screened out, and eight major patterns were witnessed (Fig. 1i; Supplementary information, Fig. S1i and Table S3). Most of these genes showed altered expression along one but not both specification paths from earlyAEC (Pattern I, II IV, and V) (Fig. 1i; Supplementary information, Fig. S1i). Most transcription factors (TFs) within these patterns were those up-regulated along either HEC specification or arterial EC maturation (Fig. 1j; Supplementary information, Fig. S1j). Interestingly, both *Hoxa5* and *Hoxa9* belonged to the same pattern as *Runx1*, although their expression was not well-correlated with *Runx1* (Fig. 1j; Supplementary information, Fig. S1j, k). The data suggested that the gene expressions should be orchestrated and precisely regulated for the subsequent cell fate choice from earlyAEC.

Efficient capture of the HSC-competent and endothelial-haematopoietic bi-potent HECs in AGM region

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We next made an effort to identify surface marker combination to highly enrich the HECs for functional evaluation (Supplementary information, Table S4). *Cd44*, *Procr* (coding CD201) and *Kit* were screened out by differentially expressed genes and correlation analysis (Fig. 2a). We specifically focused on E10.0 in the following functional assays to keep consistent with the transcriptomic finding. Whole-mount immunostaining showed that in addition to the scattered blood cells throughout the tissue, the expression of CD44 was detected in the whole endothelial layer of dorsal aorta and very proximal part of its segmental branches (Fig. 2b). Using similar strategy as for pre-HSC identification [11], we found only the derivatives induced from $CD41^{-}CD43^{-}CD45^{-}CD31^{+}Kit^{+}CD201^{+}$ rather than $CD41^{-}CD43^{-}CD45^{-}CD31^{+}Kit^{+}CD201^{-}$ population at E9.5-E10.0 could long-term (16 weeks) and multi-lineage reconstitute lethally irradiated adult recipients, although both populations generated haematopoietic clusters with different frequencies upon 7 days culture on OP9-DL1 stromal cells (Fig. 2c-e; Supplementary information, Fig. S2a-c). Self-renewal capacity of the HSCs was further validated by secondary transplantation (Fig. 2d, e; Supplementary information, Fig. S2b). Within $CD41^{-}CD43^{-}CD45^{-}CD201^{+}$ population, induced HSC potential was exclusively detected in $CD44^{+}$ subpopulation (Fig. 2c-e; Supplementary information, Fig. S2a-c). Thus, our data identified $CD41^{-}CD43^{-}CD45^{-}CD31^{+}CD201^{+}Kit^{+}CD44^{+}$ (PK44) population in E10.0 caudal half as the HSC-competent HECs.

As compared to other endothelial surface markers, Flk1 (encoded by *Kdr*) is known to be specifically localized within the vessel lumen layer, and very few and only the basal-most localized IAHC cells express Flk1 [7]. Here, almost all PK44 cells expressed Flk1 by FACS

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

analysis, indicative of their endothelial layer localization (Fig. 2f). To determine the transcriptomic identity of the HSC-competent HECs we isolated, totally 96 PK44 single cells derived from E10.0 AGM were sequenced (Supplementary information, Table S1). The PK44 cells were clustered together with HEC by computational assignment (Fig. 2g), and showed a similar high expression of the HEC feature genes (Fig. 2h). The ubiquitous and obvious expression of several key haematopoietic TFs, including *Runx1*, *Spi1*, *Gfi1*, and *Myb* [3,22], in PK44 cells inferred the enrichment of hemogenic potential (Fig. 2h; Supplementary information, Fig. S2d). Therefore, immunophenotypically purified PK44 cells elegantly represented the transcriptomically defined HEC.

We next explored whether endothelial-haematopoietic bi-potential existed in these HSC-competent HECs, since that if a cell population is experiencing fate choice so that the transient intermediate state might be captured. Firstly, we found that the Kit^+CD201^+ ECs in the body part of embryo proper at E9.5-E10.0 had a relatively higher endothelial tube-forming capacity as compared to the Kit^- or Kit^+CD201^- endothelial populations (Supplementary information, Fig. S2e). Furthermore, $CD44^+$ and $CD44^-$ fractions within Kit^+CD201^+ ECs showed comparable endothelial tube-forming capacity whereas the generation of haematopoietic cells in the cultures was exclusively detected in the $CD44^+$ ones under the bi-potential induction system (Supplementary information, Fig. S2f). The data suggested the largely maintenance of endothelial potential in the HECs. By single-cell in vitro induction, 40.6% (106/261) PK44 cells gave rise to only haematopoietic progenies and 23.0% (60/261) only endothelial tubules. Remarkably, 2.7% (7/261) had both haematopoietic and endothelial

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

potential (Fig. 2i; Supplementary information, Fig. S2f). All three kinds of potential did not present an obviously biased distribution regarding Kit or CD201 expression level by index sorting analysis (Fig. 2j). Such rare bi-potential should properly represent the intermediate cellular state in HECs along their specification path (Fig. 1h), with both endothelial and haematopoietic competencies being reflected by the asymmetric cell division under in vitro culture condition, which further emphasized the efficiency of capturing such a dynamic functional population via unsupervised computational screening.

Transcriptional and functional relationship between HECs and T1 pre-HSCs

Since the transcriptomically identified HECs and the immunophenotypically defined HSC-competent HECs (PK44) presented a largely similar molecular features (Fig. 2g, h; Supplementary information, Fig. S2d), we combined them as transcriptomic & immunophenotypic & functional HEC (tif-HEC) for the subsequent analysis. tif-HEC expressed a series of pre-HSC signature genes we previously identified [11], including *Hlf*, *Gfi1*, *Neurl3*, *Bcl11a*, *Adgrg1*, *Ikzf2*, *Angpt1*, *Mycn*, and *Procr*, suggestive of their HSC-related identity (Fig. 2h). We further performed scRNA-seq of 47 T1 pre-HSCs ($CD31^{+}CD45^{-}CD41^{low}Kit^{+}CD201^{high}$) from E11.0 AGM using the same sequencing strategy as other cells in the present study [11] (Fig. 3a; Supplementary information, Table S1). As compared to tif-HEC, T1 pre-HSC expressed similar level of *Runx1* and *Gfi1* but obviously higher level of *Spn* (encoding CD43), validating its haematopoietic cell identity (Fig. 3b). The distribution of most T1 pre-HSCs was adjacent to tif-HEC via t-SNE visualization (Fig. 3c). Of note, principal component (PC) 2 by PCA analysis largely captured the transcriptomic

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

differences between tif-HEC and T1 pre-HSC (Fig. 3d). The genes enriched in PC2 positive direction, where tif-HECs were mainly localized, were related to cell division, vascular development and cell spreading (Fig. 3e). Consistently, approximately 90% cells in HEC were proliferative (Fig. 1d), whereas the constitution is only about half in the T1 pre-HSCs [11]. Serving as the extracellular matrix component of blood vessels, *Col4a1* was expressed higher in tif-HEC than in T1 pre-HSC, further confirming the vascular endothelial property of the HECs we identified [31] (Fig. 3f). In comparison, the genes enriched in PC2 negative direction mainly related to RNA splicing and blood coagulation (Fig. 3e). Together with the overrepresented *Spi1* in T1 pre-HSC (Fig. 3f), the data suggested that haematopoietic activity has been activated in T1 pre-HSC as compared to HEC.

The developmental path from tif-HEC to T1 pre-HSC was inferred by Mpath trajectory analysis (Fig. 3g). Consistently, during the course of in vitro culture of the PK44 population from E10.0 AGM region on OP9-DL1 stromal cells to induce its HSC activity, we could witness the generation of immunophenotypic T1 pre-HSCs (Fig. 3h). We also evaluated the endothelial and haematopoietic potentials of the T1 pre-HSCs ($CD31^{+}CD45^{-}CD41^{low}Kit^{+}CD201^{high}$) in E11.0 AGM region at single-cell level. Surprisingly, we found that although displayed largely decreased endothelial potential as compared to E10.0 PK44 cells (Fig. 2i), T1 pre-HSCs still maintained comparable endothelial-haematopoietic bi-potential as that in PK44 population (Fig. 3i). This finding implied that the extremely rare and enriched T1 pre-HSC population has not completely fulfilled the endothelial-to-haematopoietic fate transition.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Enrichment of the HSC-competent HECs by newly established Neurl3-EGFP reporter

In an effort to search for single markers to distinguish HSC-primed HECs from non-HECs or those $CD45^{-}CD43^{-}$ haematopoietic cells sharing an endothelial immunophenotype, we computationally screened for the genes significantly overrepresented in HEC cluster as compared to each of the other four clusters, including one haematopoietic cluster (HC) and three vascular EC clusters (vEC, earlyAEC and lateAEC) (Supplementary information, Fig. S1f). Totally eleven genes were screened out, which were then designated as signature genes of HSC-primed HEC, including three TFs (*Mycn*, *Hlf* and *Gfi1*) but no cell surface markers (Fig. 4a; Supplementary information, Table S5). Most of them manifested similarly high expression in T1 pre-HSCs, with six of them, namely *Neurl3*, *Dnmt3b*, *Mycn*, *Hlf*, *Gfi1*, and *Gck*, belonged to pre-HSC signature genes [11] (Fig. 4a).

To further validate the bioinformatics findings and precisely determine the localization of these HSC-primed HECs, we specially chose *Neurl3* to establish a fluorescence reporter mouse line with the mind of possessing enough sensitivity, as the median expression of *Neurl3* was the highest among these signature genes (Fig. 4a; Supplementary information, Table S5). By CRISPR/Cas9-mediated gene knockin strategy, the *EGFP* was inserted into the translational initiation codon of mouse *Neurl3* gene to ensure that EGFP would be expressed in exactly the same way as Neurl3 (Fig. 4b). We firstly evaluated the Neurl3-EGFP expression by flow cytometric analysis (Fig. 4c). At E10.0 AGM region, about half of the Neurl3-$EGFP^{+}$ cells were haematopoietic (CD41/CD43/CD45-positive) cells, which

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

307 constituted about one fourth of haematopoietic population (Fig. 4c). All the Neurl3-EGFP$^+$
308 cells were CD31$^+$, and nearly all of them expressed CD44, indicative of the predominant
309 aortic localization of Neurl3-EGFP$^+$ ECs (Fig. 2b, 4c). Importantly, most PK44 cells were
310 Neurl3-EGFP$^+$, highly suggesting the enrichment of HSC-competence by the Neurl3-EGFP$^+$
311 ECs (Fig. 4c). To confirm the HSC-competence of the Neurl3-EGFP-labeled ECs, we
312 performed co-culture plus transplantation assay using E10.0 Neurl3-EGFP mouse embryos
313 (Fig. 4d). Although both could generate haematopoietic clusters under the in vitro cultures, all
314 the long-term (16 weeks) repopulations were detected exclusively in the recipients
315 transplanted with the derivatives from CD44$^+$Neurl3-EGFP$^+$ ECs but not from
316 CD44$^+$Neurl3-EGFP$^-$ ECs (Fig. 4e, f; Supplementary information, Fig. S3a).
317
318 We next investigated the transcriptomic identity of the Neurl3-EGFP$^+$ ECs and totally 48 ECs
319 with an immunophenotype of CD41$^-$CD43$^-$CD45$^-$CD31$^+$CD44$^+$Neurl3-EGFP$^+$ (NE$^+$) from
320 E10.0 AGM region were sequenced. All the NE$^+$ cells ubiquitously expressed EGFP as
321 expected and most of them expressed *Nerul3* and *Runx1* (Supplementary information, Fig.
322 S3b, c). They distributed close to tif-HEC and were predominantly located between tif-HEC
323 and earlyAEC by t-SNE visualization (Fig. 4g; Supplementary information, Table S1).
324 Accordingly, NE$^+$ cells demonstrated the increased cycling as compared to earlyAEC,
325 presenting an intermediate proliferative status between earlyAEC and tif-HEC
326 (Supplementary information, Fig. S3d). Similar to tif-HEC, NE$^+$ cells showed relatively high
327 expression of a set of HEC feature genes and pre-HSC signature genes (Fig. 2h, 4h).
328 Correlation analysis revealed that NE$^+$ cells showed the highest similarity with tif-HEC and

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

329 they were clustered together by hierarchical clustering, whereas earlyAEC and lateAEC were
330 much correlated (Fig. 4i). Therefore, from the immunophenotypic, functional and
331 transcriptomic evaluation, the performance of the Neurl3-EGFP-marked ECs was consistent
332 with the prediction of HSC-primed HEC by unsupervised computational screening.
333
334 **In situ localization and in vitro function of the HECs marked by Neurl3-EGFP reporter**
335 At the AGM region of E9.5-E11.0 embryos, CD44 expression marked the whole endothelial
336 layer of dorsal aorta in addition to IAHC cells, in line with the whole mount staining (Fig. 2b,
337 5a; Supplementary information, Fig. S3e). Of note, Neurl3-EGFP expression was restricted to
338 the IAHCs and partial aortic ECs, where Neurl3-EGFP and Runx1 presented a highly
339 co-expressed pattern (Fig. 5a; Supplementary information, Fig. S3e). Thus the Neurl3-EGFP$^+$
340 cells embedded in the endothelial layer largely enriched the putative HECs. By FACS
341 analysis, the average constitution of Neurl3-EGFP$^+$ cells in CD44$^+$ ECs were 37.7%, 50.2%
342 and 18.3% in E9.5 caudal half, E10.0 and E10.5 AGM region, respectively. Considering the
343 slightly over-estimation due to the much sensitivity of FACS, the data were basically in
344 accordance with the morphological finding (Fig. 5a) and the estimated HEC composition by
345 scRNA-seq (Supplementary information, Fig. S1f). The temporal dynamics of the HEC we
346 defined here was in line with that of Runx1 expression in aortic endothelial layer [32], and the
347 peaking of which at E10.0 was about 0.5 days earlier than the time point that the number of
348 IAHC cells reaches the peak and the first HSCs are detected in AGM region [7,27].
349

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Given the lacking of suitable antibodies to directly determine the anatomical distribution of PK44 cells, which have been proven as HSC-competent HECs (Fig. 2c-e), we compared the immunophenotype of PK44 and IAHC cells, known as $CD31^{+}Kit^{high}$ [7], regarding their relationship with Neurl3-EGFP expression whose localization was clearly defined (Fig. 5a). Both of them were mainly Neurl3-$EGFP^{+}$, with most $CD31^{+}Kit^{high}$ cells being CD41/CD43/CD45-positive haematopoietic cells as previously reported [7] (Fig. 5b). PK44 showed an expression pattern largely different from $CD31^{+}Kit^{high}$ cells, suggestive of their predominant non-IAHC localization (Fig. 5b). The expression of Neurl3-EGFP was completely absent from the sub-aortic mesenchyme, in contrast to the widespread distribution of Runx1 there (Fig. 5a; Supplementary information, Fig. S3e). Although scattered $Runx1^{+}CD44^{+}$ round blood cells were easily witnessed, much fewer Neurl3-EGFP-expressed cells outside dorsal aorta were detected, even at E11.0 (Fig. 5a; Supplementary information, Fig. S3e).

As less than half of Neurl3-$EGFP^{+}$ ECs were PK44 cells (Fig. 5b), we next explored the in vitro functional relationship of PK44 and non-PK44 fractions within $CD44^{+}$Neurl3-$EGFP^{+}$ ECs by index-sorting. From E9.5 to E10.5, all three kinds of potential, including endothelial-only, haematopoietic-only, and endothelial-haematopoietic bi-potential, could be detected in $CD44^{+}$Neurl3-$EGFP^{+}$ ECs, with different frequencies (Fig. 5c-e; Supplementary information, Fig. S3f). In E10.5, the potential was remarkably biased to endothelial as compared to E9.5 and E10.0 (Fig. 5d), which should be due to the prompt loss of Neurl3-EGFP-labeled HECs and the possible labeling of some lateAECs by Neurl3-EGFP (Fig. 4a, h). Of note, all three

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

kinds of potential were obviously higher in PK44 than non-PK44 fraction, with endothelial-haematopoietic bi-potential exclusively detected in PK44 cells (Fig. 5d). Therefore, PK44 represented the enriched functional sub-populations within Neurl3-EGFP^{+} HECs. All three kinds of potential did not show an evidently biased distribution regarding CD44 or Neurl3-EGFP expression level by index sorting analysis (Fig. 5e). Interestingly, cells with the haematopoietic rather than endothelial potential intended to have smaller side scatter density on FACS (Fig. 5e).

Stepwise fate choices of HSC-primed HECs from primitive vascular ECs

In an effort to decipher the stepwise specification of the HSC-primed HECs, we added the immunophenotypic EC samples, from the stage of initial aortic structure formation at E8.0 [33] to E9.0, to achieve seamless sampling with continuous developmental stages (Supplementary information, Fig. S4a). All the transcriptomically identified ECs were re-clustered into six clusters, with four of them basically consistent with those previously defined, namely vEC, earlyAEC, lateAEC, and HEC. The newly added samples were mainly distributed into three clusters, vEC, primitive EC (pEC) featured by *Etv2* expression and involving almost all E8.0 cells, and primitive arterial EC (pAEC) given the expression of arterial marker *Gja5* and serving as the earliest arterial EC population, with the latter two clusters as newly identified (Supplementary information, Fig. S4b, c).

Trajectory analysis by Mpath demonstrated two bifurcations along the path from pEC to HEC and revealed a predominant two-step fate choice (Fig. 6a). pEC firstly chose an arterial but

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

not venous fate to become pAEC, then upon maturing into earlyAEC and lateAEC, HEC chose to segregate from the intermediate arterial population earlyAEC (Fig. 6a), in line with the finding that the HEC displayed certain arterial characteristics but was completely devoid of venous feature (Fig. 1f). To decipher the underlying molecular programs for HEC specification, we specifically selected four clusters, excluding vEC and lateAEC branched out from the path from pEC to HEC, and added T1 pre-HSC as the end point for the subsequent analysis (Fig. 6b). Monocle 2 elegantly recapitulated the sequential sampling stages and the deduced cellular evolution upon stepwise hemogenic specification along the inferred pseudotime (Fig. 6c; Supplementary information, Fig. S4d).

We identified totally 2,851 genes whose expression levels were changed significantly among five clusters, which were further grouped into five principal expression patterns along the inferred pseudotime (Fig. 6d; Supplementary information, Fig. S4e and Table S6). In general, genes in Pattern 1 showed the highest expression in pEC, decreased apparently upon arterial specification, whereas slightly increased upon hemogenic specification, which were mainly related to rRNA processing and mitotic nuclear division (Fig. 6d; Supplementary information, Fig. S4e, f). Genes in Pattern 2 showed the highest expression in the initial arterial specification, and the lowest expression upon hemogenic specification, which were mainly related to organization of intra-cellular actin filament and inter-cellular junctions (Fig. 6d; Supplementary information, Fig. S4e, f). Genes in Pattern 3, which were related to endothelium development and cell migration, showed the highest level in earlyAEC, whereas relatively low expression in the upstream pEC and pAEC and downstream HEC and T1

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pre-HSC (Fig. 6d; Supplementary information, Fig. S4e, f). Genes in Pattern 4 and Pattern 5 both had the highest expression in the final T1 pre-HSC and both exhibited haematopoiesis-related terms, with those in Pattern 4 reaching the relatively high level from earlyAEC and those in Pattern 5 showing a gradual increase (Fig. 6d; Supplementary information, Fig. S4e,f).

Among the above 2,851 pattern genes, 75 TFs belonged to the core TFs of the regulons where the genes included significantly overlapped with the pattern genes (Fig. 6e; Supplementary information, Table S6). Given the simultaneous co-expression of the core TF and its predicted targets in a given regulon, these core TFs were considered to presumably play a role to drive or orchestrate the dynamic molecular program during HEC specification (Fig. 6e). Most of these TFs belonged to Pattern 4 and Pattern 5, indicating that most activated TFs along HEC specification from primitive vascular ECs were those overrepresented in the final hemogenic and haematopoietic populations (Fig. 6e). We also examined the expression patterns of totally 28 TFs previously reported to have a role in HSPC regeneration in vitro [34-37]. 19 of these presumed functional TF were dynamically changed and 15 of them were core TFs of regulons, with most of them belonging to Pattern 5 (Supplementary information, Fig. S4g).

We next evaluated the pathway enrichment for each cell to depict dynamic changes at pathway level. The pathways significantly changed among the five candidate clusters showed the dynamic patterns similar to gene expression patterns (Fig. 6f). Among them, cell cycle,

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ribosome and spliceosome were the pathways that were down-regulated with arterial specification whereas turned to be moderately up-regulated by hemogenic specification (Fig. 6f, g). In contract, several pathways experienced a completely opposite change, such as Rap1 signaling pathway (Fig. 6f). Artery development, together with its pivotal executor Notch signaling pathway [2,38], firstly rose to peak in earlyAEC and then modestly fell down upon hemogenic specification (Fig. 6f, g). Some inflammation related pathways, including NF κ B and TNF signaling, were activated from earlyAEC to the final T1 pre-HSC, in line with the notion about the requirement of inflammatory signaling during HSC generation [39] (Fig. 6f).

DISCCUSSION

Here via unbiasedly going through all the relevant EC populations, HSC-primed HECs were transcriptomically identified. More importantly, combining the computational prediction and in vivo functional evaluation, we precisely captured the HSC-competent HECs by a newly constructed fluorescent reporter mouse model, Neurl3-EGFP, and revealed further functionally enriched sub-population within Neurl3-EGFP-labeled ECs by a set of surface marker combination PK44. Serving as the putative marker of HSC-primed HECs [14,22], *Gfi1* was specifically expressed in HEC but not other EC-related populations (Fig. 2h, 4h), supportive of the cluster assignment. Belonging to the gene family of E3 ubiquitin ligases, the expression and role of Neurl3 in spermatogenesis and inflammation has been reported [40-42], whereas that relevant to vascular and haematopoietic development remains barely known. Neurl3 was screened out by unsupervised bioinformatics analysis, and fortunately, the expression of which in AGM region was restricted to aorta and largely consistent with that of

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

460 *Runx1* both transcriptomically (Supplementary information, Fig. S1k) and anatomically (Fig.
461 5a) regarding endothelial expression. Although highly expressed in tif-HEC, *Runx1* and
462 *Adgrg1* were also highly expressed in the $CD45^{-}CD43^{-}$ haematopoietic population (Fig. 4h),
463 which should be the derivatives of non-HSC haematopoiesis. This suggested that they may
464 not distinguish the precursors of HSCs and non-HSCs [32,43], thus *Runx1* and *Adgrg1* were not
465 included in the list of the signature genes of HSC-primed HEC (Fig. 4a). The specificity of
466 *Nerul3* expression related to HSC generation suggests that the Neurl3-EGFP would be a
467 good reporter for the studies of both HSC development and regeneration.
468
469 Based on the in vivo functional validation of the HSC-primed HECs and the sampling of
470 continuous developmental stages with intervals of 0.5 days, we had a good opportunity to
471 evaluate the dynamics and functional heterogeneity of these important transient populations.
472 Unexpectedly, the HSC-competent HECs demonstrated a previously unresolved
473 endothelial-haematopoietic bi-potential. The HECs we defined showed a higher enrichment
474 of the expression of key haematopoietic TFs (Supplementary information, Fig. S2d) and of
475 both haematopoietic and endothelial potential than using Runx1 $+23GFP^{+}$ as the maker of
476 HECs [3], which might partially explain why the rare endothelial-haematopoietic bi-potential is
477 hardly detected around the timing of HSC emergence in previous report [3]. Thus, our findings
478 well supplement the functional evaluation of putative HECs, which have a dynamic and
479 transient nature, that without catching the endothelial-haematopoietic bi-potential, it is hard to
480 define a given population to belonging to the ones being experiencing
481 endothelial-to-hemogenic fate determination. Both the constitution and the hemogenic

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

potential of the HSC-competent HECs reached the peak at the time point about 0.5 days before the first HSC emergence, and rapidly decreased thereafter (Fig. 5d; Supplementary information, Fig. S1f and S3f). Interestingly, the endothelial-haematopoietic bi-potential was still maintained until T1 pre-HSC stage at E11.0 (Fig. 3i), when cells have begun to express haematopoietic surface markers (Fig. 3b) and turned the shape into round [15]. The data suggest that the haematopoietic fate might not have been fixed in T1 pre-HSC, which needs further investigations.

We also precisely decoded the developmental path of HSC-primed HECs from the initially specified vascular ECs, the view of which has been generally neglected previously. We found that the genes and pathways involved in arterial development and Notch signaling were firstly increased and then decreased once upon HEC specification (Fig. 6e, g). Supportively, several seemingly contradictory findings have been reported regarding the role of Notch signaling in HEC specification. For example, activation of arterial program or Notch signaling is known to be required for HEC specification in mouse embryos or generation of HECs with lymphoid potential from human pluripotent stem cells [44-46]. On the other hand, repression of arterial genes in EC after arterial fate acquisition leads to augmented haematopoietic output [47]. Noteworthy, we revealed two bifurcates during HSC-primed HEC specification along the path from primitive vascular EC, suggesting two-step fate choice occurred for hemogenic fate settling (Fig. 6a). Serving as the two presumed final fates of earlyAEC, HEC and lateAEC displayed a series of differences (Fig. 1i), which better explains the presumably misinterpreted notion in previous report that arterial ECs and HSCs originate from distinct

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

precursors [18]. Our findings further emphasize that arterial specification and Notch signaling should be precisely and stepwise controlled for HSC generation. Although both showing obvious similarity regarding the arterial feature and anatomical distribution, the difference between earlyAEC and lateAEC should also be paid attention to as the former but not the latter is the direct origin of the HSC-primed HECs.

It is generally accepted that haematopoietic cells in the IAHCs are proliferative [48,49], within which pre-HSCs are mainly involved [14,49]. Supportively, enriched functional T1 pre-HSCs manifested a relatively proliferative status [11]. On the other hand, slow cycling is witnessed at the base of IAHCs [49], and it is suggested that exit from cell cycle is necessary for HEC development and endothelial-to-haematopoietic transition [44,50]. Nevertheless, based on the precise recognition of the HSC-primed HECs here, we showed that proliferation was gradually decreased upon arterial specification and maturation, whereas re-activated once the arterial ECs chose a hemogenic fate featured by the simultaneous *Runx1* expression (Fig. 6e-g). The functional requirement of cell cycle control for the specification of the HSC-primed HECs needs to be investigated, which would depend on the initiating cell populations.

We also revealed several similarities regarding the molecular events underlying the development of HSC-primed HECs between in mouse and human embryos we have reported very recently [19], including the arterial feature and the overrepresented ribosome and translational activity in the HSC-primed HECs. Such conservation further assures the mouse model as an adequate animal model for HSC development studies. The comprehensive

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

526 understanding of cellular evolutions and molecular programs underlying the specification of
527 HSC-primed HECs combined with the important spatiotemporal cues *in vivo* will facilitate
528 future investigations directing HSC formation *in vitro* and other related regeneration
529 strategies.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MATERIALS AND METHODS

No statistical methods were used to predetermine the sample size. The experiments were not randomized at any stage. The investigators were not blinded to allocation during the experiments and outcome assessment.

Mice

Mice were handled at the Laboratory Animal Center of Academy of Military Medical Sciences in accordance with institutional guidelines. Mouse manipulations were approved by the Animal Care and Use Committee of the Institute. The *Neurl3*$^{EGFP/+}$ reporter mouse lines were generated with the CRISPR/Cas9 technique by Beijing Biocytogen. All mice were maintained on C57BL/6 background. Embryos were staged by somite pair (sp) counting: E8.0, 1-7 sp; E8.5, 8-12 sp; E9.0, 13-20 sp; E9.5, 21-30 sp; E10.0, 31-35 sp; E10.5, 36-40 sp; and E11.0, 41-45 sp. In some experiments, caudal half of E10.0 embryo was dissected under heart with limbs removed. AGM region was dissected as reported [12]. The fluorescent dye Oregon green 488 was purchased from Invitrogen. Staining was performed as previously described [12] except that the concentration of staining solution was 5 μmol/L and the time of staining was 3 minutes before washed. Primary embryonic single-cell suspension was acquired by type I collagenase digestion.

Flow cytometry

Cells were sorted and analyzed by flow cytometers FACS Aria 2 and Calibur (BD Biosciences), and the data were analyzed using FlowJo software (Tree Star). Cells were

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

552 stained by the following antibodies: B220 (eBioscience, RA3-6B2), CD3 (eBioscience,
553 145-2C11), CD4 (eBioscience, GK1.5), CD8a (eBioscience, 53-6.7), CD31 (BD or BioLegend,
554 MEC13.3), CD41 (BD or eBioscience, MWReg30), CD43 (BD, S7), CD44 (eBioscience or
555 BioLegend, IM7), CD45.1 (eBioscience, A20), CD45.2 (eBioscience, 104), CD45
556 (eBioscience, 30-F11), CD144 (eBioscience, eBioBV13), CD201 (eBioscience, eBio1560),
557 Flk1 (eBioscience, Avas12a1), Kit (eBioscience, 2B8), Ly-6G (BioLegend, 1A8), and Mac-1
558 (eBioscience, M1/70). 7-amino-actinomycin D (7-AAD; eBioscience) was used to exclude
559 dead cells. For index sorting, the FACS Diva 8 "index sorting" function was activated and
560 sorting was performed in single-cell mode.
561
562 **OP9-based haematopoietic and endothelial potential assay**
563 Cells were sorted by flow cytometry in single-cell mode and were then plated on the OP9 or
564 OP9-DL1 stromal cells [51] in IMDM (Hyclone) containing 15% fetal bovine serum (Hyclone), 1%
565 bovine serum albumin (Sigma), 10 μg/mL insulin (Macgene), 200 μg/mL transferrin (Sigma),
566 and 5.5 x 10^{-5} mol/L 2-mercaptoethanol (Gibco). For the endothelial potential assay, 100
567 ng/mL rhVEGF-165 (PeproTech) was supplemented. For haematopoietic and endothelial
568 bi-potential assay with 10 cells or single cell plated per well, both 100 ng/mL rhVEGF-165 and
569 50 ng/mL SCF (PeproTech) were supplemented. After 7 days of co-culture, cells were fixed in
570 4% paraformaldehyde for 30 minutes and stained with PE-conjugated or purified CD45
571 antibody (eBioscience, 30-F11 or BD Biosciences) to ascertain the generation of
572 haematopoietic progeny. Subsequently, CD31 (BD Pharmingen, MEC13.3)
573 immunohistochemistry staining was performed using standard procedures, and the formation

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

574 of CD31-positive tubules in the wells was considered as having endothelial potential.

575

576 **OP9-DL1 co-culture and transplantation assay**

577 To investigate the HSC potential of the PK44 population in E10.0 caudal half, male CD45.1/1
578 and female CD45.2/2 mice were mated to obtain CD45.1/2 embryos. FACS purified cell
579 populations from E10.0 caudal half (CD45.1/2) were plated on the OP9-DL1 stromal cells in
580 α-MEM (Gibco) supplemented with 10% fetal bovine serum (Hyclone) and cytokines (100
581 ng/mL SCF, 100 ng/mL IL-3 and 100 ng/mL Flt3 ligand, all from PeproTech). After 7 days of
582 co-culture, cells were harvested and then injected into 8-12 weeks female recipients
583 (CD45.2/2) via tail vein, along with 2×10^4 nucleated fresh bone marrow carrier cells
584 (CD45.2/2) per recipient. Recipients were pre-treated by a split dose of 9 Gy γ-irradiation
585 (^{60}Co). Peripheral blood cells of recipients were analyzed by flow cytometry at the indicated
586 time points to determine the chimerism. The recipients demonstrating ⩾5% donor-derived
587 chimerism in $CD45^+$ cells of peripheral blood were considered as successfully reconstituted.
588 Multi-organ and multi-lineage reconstitution was evaluated as reported [52]. Totally 1×10^7 bone
589 marrow cells obtained from the reconstituted primary recipients at 16 weeks
590 post-transplantation were injected into the secondary recipients to investigate HSC
591 self-renewal potential.

592

593 To investigate the HSC potential of the $CD41^-CD43^-CD45^-CD31^+CD44^+$Neurl3-$EGFP^+$
594 population in E10.0 caudal half, male *Neurl3*$^{EGFP/+}$ reporter mice (CD45.2/2 background) were
595 crossed to female CD45.2/2 mice to generate *Neurl3*$^{EGFP/+}$ embryos. Then the co-culture and

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

596 transplantation strategy were same as mentioned above except that the recipients were
597 female 8-12 weeks CD45.1/2 mice and the carrier cells were obtained from CD45.1/1 mice.
598
599 **Immunofluorescence**
600 Embryos were isolated, fixed with 4% paraformaldehyde for 30 minutes to 2 hours at 4°C,
601 embedded in paraffin, and sectioned at 5-6 μm with Leica RM2235. Sections were
602 deparaffinized with ethanol of gradient concentration, then blocked in blocking solution
603 (Zhongshan golden bridge) for 30 minutes at room temperature, followed by incubation with
604 primary antibodies overnight at 4°C. After 3 washes (3 minutes each) in PBS, sections were
605 incubated with corresponding secondary antibodies (Zhongshan golden bridge) for 30
606 minutes at room temperature. After 3 washes in PBS, sections were stained with
607 DendronFluor TSA (Histova, NEON 4-color IHC Kit for FFPE, NEFP450, 1:100, 20–60-sec).
608 The primary and secondary antibodies were thoroughly eluted by heating the slides in citrate
609 buffer (pH 6.0) for 10 minutes at 95°C using microwave. In a serial fashion, each antigen was
610 labeled by distinct fluorophores. After all the antibodies were detected sequentially, the slices
611 were finally stained with DAPI. Images were collected by confocal microscope (Nikon Ti-E
612 A1/ ZEISS LSM 880). The primary antibodies were as follows: CD31 (BD Biosciences), CD44
613 (BD Biosciences), Endomucin (eBioscience), GFP (Cell Signaling), and Runx1 (Abcam).
614
615 **Whole-mount Immunofluorescence**
616 The body part between forelimb buds and hindlimb buds of E10.0 embryo was dissected,
617 fixed in 2% PFA/PBS for 20 minutes on ice and dehydrated in graded concentrations of

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

618 methanol/PBS (50%, 100%; 10 minutes each). To block endogenous peroxidase, samples
619 were bleached in 5% H_2O_2 for 1 hour on ice. For staining, the samples were blocked in
620 PBSMT (1% skim milk and 0.4%Triton X-100 in PBS) containing 0.2% BSA for 1 hour at 4℃,
621 incubated with PBSMT containing anti-CD44 (1:25) overnight at 4℃, then washed 3 times in
622 PBSMT each for 1 hour at 4℃. The primary antibody was developed by incubating
623 HRP-conjugated anti-rat Ig antibody (1:2000 in PBSMT; Zhongshan golden bridge) overnight
624 at 4℃. After extensive washing with more than 3 exchanges of PBSMT, including the final 20
625 minutes wash in PBST (0.1% Triton X-100 in PBS) at 4℃, the samples were soaked in
626 DendronFluor TSA (Histova, NEON 4-color IHC Kit for Wholemount/Cytometry, NEWM450)
627 for 10–30 minutes, and hydrogen peroxide was added to 0.03%. The enzymatic reaction was
628 allowed to proceed until the desired color intensity was reached, and the samples were rinsed
629 3 times in PBST. Finally, the samples were dehydrated in 100% methanol and soaked in
630 graded concentrations of BABB (phenylcarbinol and benzyl benzoate, 1:2)/methanol (50%,
631 100%; 1 minute each), stored at -20℃ until photographed.
632
633 **Single cell RNA-seq library construction**
634 Single cells in good condition were picked into lysis buffer by mouth pipetting. The single cell
635 RNA-seq preparation procedure was based on STRT with some modifications [53 54 55]. cDNAs
636 were synthesized using sample-specific 25 nt oligo dT primer containing 8 nt barcode
637 (TCAGACGTGTGCTCTTCCGATCT-XXXXXXXX-NNNNNNNN-T25, X representing
638 sample-specific barcode whereas N standing for unique molecular identifiers, UMI, see Table
639 S7) and TSO primer for template switching [56 57 58]. After reverse transcription and

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

second-strand cDNA synthesis, the cDNAs were amplified by 17 cycles of PCR using ISPCR primer and 3' Anchor primer (see Table S7). Up to 56 samples were pooled and purified using Agencourt AMPure XP beads (Beckman). 4 cycles of PCR were performed to introduce index sequence (see Table S7). After this step, 400 ng cDNAs were fragmented to around 300 bp by covaris S2. The cDNA was incubated with Dynabeads MyOne™ Streptavidin C1 beads (Thermo Fisher) for 1 hour at room temperature. Libraries were generated using KAPA Hyper Prep Kit (Kapa Biosystems). After adaptor ligation, the libraries were amplified by 7 cycles of PCR using QP2 primer and short universal primer (see Table S7). The libraries were sequenced on Illumina HiSeq 4000 platform in 150bp pair-ended manner (sequenced by Novogene).

Quantification of gene expression for scRNA-seq data

We used unique molecular identifier (UMI)-based scRNA-seq method to measure the gene expression profiles within individual cells. Raw reads were firstly split by specific barcode attached in Read 2 for individual cells and UMI information was aligned to the corresponding Read 1. Read 1 was trimmed to remove the template switch oligo (TSO) sequence and polyA tail sequence. Subsequently, quality control was conducted to discard reads with adapter contaminants or low-quality bases (N > 10%). Next, the mm10 mouse transcriptome (UCSC) was used to align the clean reads using TopHat (version 2.0.12) [59]. Uniquely mapped reads were obtained using HTSeq package [60] and grouped by the cell-specific barcodes. Transcripts of each gene were deduplicated based on the UMI information, while mitochondrial genes were not included for quantification. Finally, for each gene in each

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

individual cell, the number of the distinct UMIs derived from that gene was regarded as its copy number of transcripts.

Quality control and normalization of sequencing data

For the 662 sequenced single cells from E9.5-E11.0 embryos of totally 29 embryos, we only retained cells with more than 2,000 genes and 100,000 transcripts detected. Then, 597 cells passed the filter standards. Gene expression levels in each cell were normalized by $\log_2$(TPM/10+1), where TPM (transcripts-per-million) was calculated as (the number of UMIs of each gene / all UMIs of a given cell) ×1,000,000. Since the UMI number of most of our samples was less than the order of 1,000,000 transcripts, the TPM values were divided by 10 to avoid counting each transcript for several times. On average we detected 7,035 genes (range from 2,266 to 10,843) and 636,418 transcripts (range from 103,793 to 2,959,573) expressed in each individual cell.

Additionally, we also sequenced 96 single cells with a PK44 immunophenotype ($CD41^-CD43^-CD45^-CD31^+CD201^+Kit^+CD44^+$) from E10.0 AGM regions of totally 9 embryos, 47 T1 pre-HSCs ($CD31^+CD45^-CD41^{low}Kit^+CD201^{high}$) from E11.0 AGM regions of totally 18 embryos, 48 single cells with an immunophenotype of $CD41^-CD43^-CD45^-CD31^+CD44^+Neurl3\text{-}EGFP^+$ from Neurl3-EGFP reporter mouse embryos and 579 single cells from E8.0-E9.0 body regions of totally 24 embryos. The same quality control criteria and normalization method described above were applied to these additional datasets. In total, 1,432 single cells were sequenced and 1,325 cells passed the filter

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

684 standards and were used for downstream analyses (see Table S1).

685

686 **Dimensional reduction and clustering**

687 We used Seurat R package [61] (version 2.3.4) for further analyses and exploration of our
688 single cell RNA sequencing data, such as identification of highly variable genes (HVGs) and
689 differentially expressed genes (DEGs), dimension reduction using PCA or t-SNE,
690 unsupervised clustering and so on. A standard analysis process is briefly described below.
691 First, only genes expressed in at least 3 single cells were retained so as to exclude genes
692 that were hardly expressed. Then, FindVariableGenes function was used to select HVGs on
693 log2 (TPM/10+1) transformed expression values. Genes with average expression more than
694 1 and less than 8 and dispersion greater than 1 were identified as HVGs. To mitigate the
695 effect of cell cycle, HVGs not included in the direct cell cycle GO term (GO:0007049) (Table
696 S7) were used as inputs for PCA dimension reduction. Elbow method was employed to select
697 the top relevant PCs for subsequent t-SNE dimension reduction and graph-based clustering
698 [28].

699

700 For the initial dataset from E9.5-E11.0 body and DA locations, we select top 15 PCs for
701 clustering using FindClusters with default settings, to obtain 6 major clusters. Negative
702 control cells with a non-EC immunophenotype and cells grouped with these negative control
703 cells were reclassified specifically into the Neg cluster. The remaining cells were assigned as
704 vEC, earlyAEC, lateAEC, HEC and HC clusters based on the clustering results. Next, cells in
705 Neg cluster and cells with *Ptprc* or *Spn* expression level greater than 1 were removed. Then,

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

706 the filtered initial dataset was used for analyses of subdatasets, including subdividing of HEC
707 cluster, subdividing of eaAEC cluster and in-depth analyses of earlyAEC, lateAEC and HEC
708 clusters. The filtered initial dataset was also included in three combined datasets of
709 combining PK44 cell population, PK44 and T1 pre-HSC cell populations, and PK44 and
710 Neurl3-EGFP cell populations, respectively. Dimension reduction and clustering analyses for
711 subdatasets and combined datasets abovementioned also followed the same procedure as
712 described above. See Table S1 for detailed cell information.
713
714 For combined dataset from earlier dataset (E8.0-E9.0 body location) and initial dataset
715 (E9.5-E11.0 body and DA locations), we redid the dimension reduction and clustering
716 analyses over again. Same as the processing of initial dataset, negative control cells with a
717 non-EC immunophenotype and cells grouped with these negative control cells were
718 reclassified manually into Neg cluster. The remaining cells were assigned as vEC, pEC,
719 pAEC, earlyAEC, lateAEC, HEC and HC based on the clustering results. The new clustering
720 results are highly consistent with the previous ones within the common cell populations. Next,
721 cells in Neg cluster and cells with *Ptprc* or *Spn* expression level greater than 1 were removed.
722 Cells in pEC, pAEC, lateAEC and HEC and cells from T1 pre-HSC dataset were retained for
723 further analysis.
724
725 **Identification of DEGs**
726 DEGs were identified using FindMarkers or FindAllMarkers functions with default Wilcoxon
727 rank sum test and only genes detected in a minimum fraction of 0.25 cells in either of the two

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

728 populations were considered. Genes with fold-change $\geqslant$ 2 and adjusted *P* value $\leqslant$ 0.05
729 were selected as DEGs.
730
731 **Arterial and venous feature score**
732 Arteriovenous marker genes previously known or inferred from the artery development
733 pattern genes, including 10 arterial genes (*Dll4*, *Igfbp3*, *Unc5b*, *Gja4*, *Hey1*, *Mecom*, *Efnb2*,
734 *Epas1*, *Vegfc* and *Cxcr4*) and 3 venous genes (*Nr2f2*, *Nrp2*, and *Aplnr*) [33,62-64], were selected
735 to perform the arteriovenous feature scores. First, we scaled the $\log_2$(TPM/10+1) expression
736 values of each marker gene to 0-10 scale among all the sample cells after quality control.
737 Second, for each cell, we averaged the scaled values of arterial genes and venous genes,
738 respectively. Third, the averaged values were rescaled to 0-10 scale across all the sample
739 cells to finally achieve the arterial and venous scores. For each population, the arterial and
740 venous scores of all of the cells within the population were average. The 50% confidence
741 ellipses were also calculated to show the main distribution ranges. We chose score value = 5
742 as the threshold to infer the arterial or venous identity of vascular ECs, as the distribution of
743 individual cells was in line with the notion showing essentially no arterial/venous double
744 positive cells.
745
746 **Cell cycle analysis**
747 For cell cycle analysis, cell cycle-related genes consisting of a previously defined core set of
748 43 G1/S genes and 54 G2/M genes were used [58,65] (see Table S7 for detailed gene lists). We
749 used a way similar to Tirosh, et al. [66] to classify the cycling phases of the cells. We calculated

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the average expression of each gene set as corresponding scores, and manually assigned cells to approximate cell cycle phases based on the scores. Namely, cells with G1/S score < 2 and G2/M score < 2 were assigned as 'quiescent', otherwise 'proliferative'. Among proliferative cells, those with G2/M score > G1/S score were assigned as 'G2/M', and those with G1/S score > G2/M score were assigned as 'G1' when G2/M score < 2, or as 'S' when G2/M score $\geqslant$ 2.

Constructing single cell trajectories

Monocle 2 [67] (version 2.6.4) and Mpath [68] (version 1.0) were adopted to infer the development trajectory of selected cell populations. Monocle 2 can construct single-cell trajectories and place each cell at its proper position in the trajectory, even a "branched" trajectory corresponding to cellular "decisions". We followed the official vignette with recommended parameters. Briefly, UMI count data of given cell populations was used as input and genes with more than 1.5 times of fitted dispersion evaluated using dispersionTable function were identified as HVGs. To reduce the influence of cell cycle effect, HVGs not included in the direct cell cycle GO term (GO:0007049) were retained as ordering genes for the subsequent ordering cells.

For Mpath analysis, the $\log_2$(TPM/10+1) normalized data of HVGs identified by using Seurat method were used as inputs. The cluster labels defined by clustering procedures described above were used as landmark cluster assignment of individual cells. Based on the results of the Mpath analyses, we specified the starting point and developing directions according to

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the development time and visualized the results on t-SNE plot.

Patterns of DEGs among multiple clusters

In the case of identification of gene patterns in more than two clusters, analysis of variance followed by Tukey's HSD test for pairwise comparison was adopted to identify DEGs (genes with adjusted *P* value < 0.05 and fold change > 2 or < 0.5). For identification of patterns in earlyAEC, lateAEC and HEC clusters, only 1,005 DEGs resulted from the pairwise comparisons of earlyAEC and lateAEC and of earlyAEC and HEC were retained. According to the changed directions of HEC and lateAEC as compared to earlyAEC, we could assign these DEGs into 8 patterns as illustrated. Transcription factors network visualization was implemented as follows. First, the average expression values of genes included in each pattern were calculated as their representative expression levels. Then, the representative expression levels of 8 patterns and the expression profile data of transcription factors included in these patterns were combined as input for construction of "signed hybrid" weighted gene co-expression network analysis using WGCNA [69]. Next, we used 0.01 as adjacency threshold for including edges in the output to export network, which was then imported into Cytoscape[70] for visualization. We also calculated Pearson correlation coefficient between each transcription factor and the pattern it belongs to.

For identification of patterns in pEC, pAEC, earlyAEC and HEC clusters, all 2,851 DEGs among them were retained. We used ConsensusClusterPlus function with k-means algorithm on top 500 DEGs to achieve five stable clusters. Then all DEGs were reassigned into one of

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

794 the five patterns according to which pattern had maximum average Pearson correlation
795 coefficient with a given DEG. Note that we used a downsampled dataset in the visualization
796 related to the five patterns in order to show more detailed changes along the development
797 trajectory. Sixty cells were randomly sampled from pEC and pAEC clusters, respectively.
798
799 **Identification of HEC signature genes**
800 Firstly, we compared HEC to every cluster to get the overrepresented genes, which were
801 upregulated across each of the other 4 clusters (vEC, earlyAEC, lateAEC and HC) within
802 filtered initial dataset. To make sure the accuracy of HEC overrepresented genes, we used
803 both wilcox and roc method to perform the DEG analysis. Only the DEGs identified by both
804 methods were regarded as HEC overrepresented genes. Finally, 25 cluster HEC
805 overrepresented genes were retained. In order to not only identify the endothelium with
806 hemogenic potential, but also discriminate those HSC-primed hemogenic ECs from yolk
807 sac-derived early haematopoietic populations such as erythro-myeloid progenitors, genes
808 highly expressed in erythro-myeloid progenitors (*Gsta4*, *Spi1*, *Alox5ap* and *Myb*) as reported
809 [71 72] were excluded. In addition, genes not highly expressed ($\log_2$(TMP/10+1) < 2) in HEC or
810 highly expressed ($\log_2$(TPM/10+1) > 2) in every clusters were also excluded. Finally, eleven
811 HEC overrepresented genes were retained as HEC signature genes.
812
813 **SCENIC analysis**
814 SCENIC [73] could reconstruct gene regulatory networks from single-cell RNA-seq data based
815 on co-expression and DNA motif analysis. Here, we used SCENIC R package (version

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

816 1.1.1-9) to identify refined regulons, each of which represented a regulatory network that
817 connects a core TF with its target genes. We followed the “Running SCENIC” vignette in the
818 R package with default settings. We identified 507 unique regulons, among which 75
819 regulons significantly overlapped with the 2851 significantly changed genes were retained.
820 Fisher's exact test was employed for estimate the statistical significance of their overlaps.
821 The 75 core TFs were considered as putative driving force to orchestrate the dynamic
822 molecular program during HEC specification, given the simultaneous co-expression of the
823 core TF and its predicted targets in a given regulon.

824

825 **Gene set variation analysis**

826 Through gene set variation analysis, gene-level expression profiles could be transformed into
827 pathway-level enrichment score profiles using GSVA R package [74] coupled with KEGG
828 pathways [75]. We used ssgsea method [76] to estimate gene-set enrichment scores of each cell.
829 Two-sample Wilcoxon test was employed to find differentially enriched pathways between
830 involved clusters. Adjusted *P* value < 0.05 was considered statistically significant.

831

832 **TFs and cell surface molecules**

833 Genes were marked as TFs according to 1,485 TFs included in AnimalTFDB 2.0 [77], and
834 marked as surface molecules according to 871 high-confidence surfaceome proteins
835 identified in Cell Surface Protein Atlas [78]. See Table S7 for the detailed gene lists.

836

837 **Statistical analysis**

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

838 All statistical analyses were conducted in R version 3.4.3. Two-sample Wilcoxon Rank Sum
839 test was employed for comparisons of gene numbers, transcript counts, or gene expression
840 levels between two clusters of cells. We referred to statistically significant as $P < 0.05$ (if not
841 specified). Network enrichment analyses and gene ontology biological process enrichment
842 analyses were performed using Metascape [79] (http://metascape.org) and clusterProfiler [80],
843 respectively.
844
845 **Data and Code Availability**
846 The scRNA-seq data has been deposited in the NCBI Gene Expression Omnibus, the
847 accession number for the data is pending. Code is available on reasonable request.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

REFERENCES

1 Dzierzak, E. & Bigas, A. Blood Development: Hematopoietic Stem Cell Dependence and Independence. *Cell Stem Cell* **22**, 639-651, doi:10.1016/j.stem.2018.04.015 (2018).

2 Potente, M. & Makinen, T. Vascular heterogeneity and specialization in development and disease. *Nat Rev Mol Cell Biol* **18**, 477-494, doi:10.1038/nrm.2017.36 (2017).

3 Swiers, G. *et al.* Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. *Nat Commun* **4**, 2924, doi:10.1038/ncomms3924 (2013).

4 Chen, M. J. *et al.* Erythroid/myeloid progenitors and hematopoietic stem cells originate from distinct populations of endothelial cells. *Cell Stem Cell* **9**, 541-552, doi:10.1016/j.stem.2011.10.003 (2011).

5 Taoudi, S. *et al.* Progressive divergence of definitive haematopoietic stem cells from the endothelial compartment does not depend on contact with the foetal liver. *Development* **132**, 4179-4191, doi:10.1242/dev.01974 (2005).

6 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, doi:10.1038/nature07619 (2009).

7 Yokomizo, T. & Dzierzak, E. Three-dimensional cartography of hematopoietic clusters in the vasculature of whole mouse embryos. *Development* **137**, 3651-3661, doi:10.1242/dev.051094 (2010).

8 Kissa, K. & Herbomel, P. Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature* **464**, 112-115, doi:10.1038/nature08761 (2010).

9 Boisset, J. C. *et al.* In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature* **464**, 116-120, doi:10.1038/nature08764 (2010).

10 Bertrand, J. Y. *et al.* Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature* **464**, 108-111, doi:10.1038/nature08738 (2010).

11 Zhou, F. *et al.* Tracing haematopoietic stem cell formation at single-cell resolution. *Nature* **533**, 487-492, doi:10.1038/nature17997 (2016).

12 Rybtsov, S. *et al.* Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region. *J Exp Med* **208**, 1305-1315, doi:10.1084/jem.20102419 (2011).

13 Taoudi, S. *et al.* Extensive hematopoietic stem cell generation in the AGM region via maturation of VE-cadherin+CD45+ pre-definitive HSCs. *Cell Stem Cell* **3**, 99-108, doi:10.1016/j.stem.2008.06.004 (2008).

14 Baron, C. S. *et al.* Single-cell transcriptomics reveal the dynamic of haematopoietic stem cell production in the aorta. *Nat Commun* **9**, 2517, doi:10.1038/s41467-018-04893-3 (2018).

15 Boisset, J. C. *et al.* Progressive maturation toward hematopoietic stem cells in the mouse embryo aorta. *Blood* **125**, 465-469, doi:10.1182/blood-2014-07-588954 (2015).

16 Ditadi, A., Sturgeon, C. M. & Keller, G. A view of human haematopoietic development from the Petri dish. *Nat Rev Mol Cell Biol* **18**, 56-67, doi:10.1038/nrm.2016.127 (2017).

17 Ditadi, A. *et al.* Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages. *Nat Cell Biol* **17**, 580-591, doi:10.1038/ncb3161 (2015).

18 Gama-Norton, L. *et al.* Notch signal strength controls cell fate in the haemogenic endothelium. *Nat Commun* **6**, 8510, doi:10.1038/ncomms9510 (2015).

19 Zeng, Y. *et al.* Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing. *Cell Res*, doi:10.1038/s41422-019-0228-6 (2019).

20 Gritz, E. & Hirschi, K. K. Specification and function of hemogenic endothelium during embryogenesis.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cell Mol Life Sci **73**, 1547-1567, doi:10.1007/s00018-016-2134-0 (2016).

21 Hadland, B. K. *et al.* A Common Origin for B-1a and B-2 Lymphocytes in Clonal Pre- Hematopoietic Stem Cells. *Stem Cell Reports* **8**, 1563-1572, doi:10.1016/j.stemcr.2017.04.007 (2017).

22 Thambyrajah, R. *et al.* GFI1 proteins orchestrate the emergence of haematopoietic stem cells through recruitment of LSD1. *Nat Cell Biol* **18**, 21-32, doi:10.1038/ncb3276 (2016).

23 Solaimani Kartalaei, P. *et al.* Whole-transcriptome analysis of endothelial to hematopoietic stem cell transition reveals a requirement for Gpr56 in HSC generation. *J Exp Med* **212**, 93-106, doi:10.1084/jem.20140767 (2015).

24 Bee, T. *et al.* Nonredundant roles for Runx1 alternative promoters reflect their activity at discrete stages of developmental hematopoiesis. *Blood* **115**, 3042-3050, doi:10.1182/blood-2009-08-238626 (2010).

25 de Bruijn, M. F. *et al.* Hematopoietic stem cells localize to the endothelial cell layer in the midgestation mouse aorta. *Immunity* **16**, 673-683, doi:10.1016/s1074-7613(02)00313-8 (2002).

26 Guibentif, C. & Gottgens, B. Blood: Education for stem cells. *Nature* **545**, 415-417, doi:10.1038/nature22496 (2017).

27 Dzierzak, E. & Speck, N. A. Of lineage and legacy: the development of mammalian hematopoietic stem cells. *Nat Immunol* **9**, 129-136, doi:10.1038/ni1560 (2008).

28 Levine, J. H. *et al.* Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162**, 184-197, doi:10.1016/j.cell.2015.05.047 (2015).

29 Zilberberg, L. *et al.* Specificity of latent TGF-beta binding protein (LTBP) incorporation into matrix: role of fibrillins and fibronectin. *J Cell Physiol* **227**, 3828-3836, doi:10.1002/jcp.24094 (2012).

30 Bos, F. L., Hawkins, J. S. & Zovein, A. C. Single-cell resolution of morphological changes in hemogenic endothelium. *Development* **142**, 2719-2724, doi:10.1242/dev.121350 (2015).

31 Jeanne, M., Jorgensen, J. & Gould, D. B. Molecular and Genetic Analyses of Collagen Type IV Mutant Mouse Models of Spontaneous Intracerebral Hemorrhage Identify Mechanisms for Stroke Prevention. *Circulation* **131**, 1555-1565, doi:10.1161/CIRCULATIONAHA.114.013395 (2015).

32 Tober, J., Yzaguirre, A. D., Piwarzyk, E. & Speck, N. A. Distinct temporal requirements for Runx1 in hematopoietic progenitors and stem cells. *Development* **140**, 3765-3776, doi:10.1242/dev.094961 (2013).

33 Chong, D. C., Koo, Y., Xu, K., Fu, S. & Cleaver, O. Stepwise arteriovenous fate acquisition during mammalian vasculogenesis. *Dev Dyn* **240**, 2153-2165, doi:10.1002/dvdy.22706 (2011).

34 Sugimura, R. *et al.* Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature* **545**, 432-438, doi:10.1038/nature22370 (2017).

35 Lis, R. *et al.* Conversion of adult endothelium to immunocompetent haematopoietic stem cells. *Nature* **545**, 439-445, doi:10.1038/nature22326 (2017).

36 Wahlster, L. & Daley, G. Q. Progress towards generation of human haematopoietic stem cells. *Nat Cell Biol* **18**, 1111-1117, doi:10.1038/ncb3419 (2016).

37 Vo, L. T. & Daley, G. Q. De novo generation of HSCs from somatic and pluripotent stem cell sources. *Blood* **125**, 2641-2648, doi:10.1182/blood-2014-10-570234 (2015).

38 Gridley, T. Notch signaling in vascular development and physiology. *Development* **134**, 2709-2718, doi:10.1242/dev.004184 (2007).

39 Li, Y. *et al.* Inflammatory signaling regulates embryonic hematopoietic stem and progenitor cell production. *Genes Dev* **28**, 2597-2612, doi:10.1101/gad.253302.114 (2014).

40 Zhao, Y. *et al.* Neuralized E3 Ubiquitin Protein Ligase 3 Is an Inducible Antiviral Effector That Inhibits

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hepatitis C Virus Assembly by Targeting Viral E1 Glycoprotein. *J Virol* **92**, doi:10.1128/JVI.01123-18 (2018).

41 Xu, W. *et al.* Ubiquitin ligase gene neurl3 plays a role in spermatogenesis of half-smooth tongue sole (Cynoglossus semilaevis) by regulating testis protein ubiquitination. *Gene* **592**, 215-220, doi:10.1016/j.gene.2016.07.062 (2016).

42 Hu, Y., Nguyen, T. T., Bui, K. C., Demello, D. E. & Smith, J. B. A novel inflammation-induced ubiquitin E3 ligase in alveolar type II cells. *Biochem Biophys Res Commun* **333**, 253-263, doi:10.1016/j.bbrc.2005.05.102 (2005).

43 Frame, J. M., Fegan, K. H., Conway, S. J., McGrath, K. E. & Palis, J. Definitive Hematopoiesis in the Yolk Sac Emerges from Wnt-Responsive Hemogenic Endothelium Independently of Circulation and Arterial Identity. *Stem Cells* **34**, 431-444, doi:10.1002/stem.2213 (2016).

44 Marcelo, K. L. *et al.* Hemogenic endothelial cell specification requires c-Kit, Notch signaling, and p27-mediated cell-cycle control. *Dev Cell* **27**, 504-515, doi:10.1016/j.devcel.2013.11.004 (2013).

45 Uenishi, G. I. *et al.* NOTCH signaling specifies arterial-type definitive hemogenic endothelium from human pluripotent stem cells. *Nat Commun* **9**, 1828, doi:10.1038/s41467-018-04134-7 (2018).

46 Park, M. A. *et al.* Activation of the Arterial Program Drives Development of Definitive Hemogenic Endothelium with Lymphoid Potential. *Cell Rep* **23**, 2467-2481, doi:10.1016/j.celrep.2018.04.092 (2018).

47 Lizama, C. O. *et al.* Repression of arterial genes in hemogenic endothelium is sufficient for haematopoietic fate acquisition. *Nat Commun* **6**, 7739, doi:10.1038/ncomms8739 (2015).

48 Zape, J. P., Lizama, C. O., Cautivo, K. M. & Zovein, A. C. Cell cycle dynamics and complement expression distinguishes mature haematopoietic subsets arising from hemogenic endothelium. *Cell Cycle* **16**, 1835-1847, doi:10.1080/15384101.2017.1361569 (2017).

49 Batsivari, A. *et al.* Understanding Hematopoietic Stem Cell Development through Functional Correlation of Their Proliferative Status with the Intra-aortic Cluster Architecture. *Stem Cell Reports* **8**, 1549-1562, doi:10.1016/j.stemcr.2017.04.003 (2017).

50 Ottersbach, K. Endothelial-to-haematopoietic transition: an update on the process of making blood. *Biochem Soc Trans* **47**, 591-601, doi:10.1042/BST20180320 (2019).

51 Lan, Y. *et al.* Endothelial Smad4 restrains the transition to hematopoietic progenitors via suppression of ERK activation. *Blood* **123**, 2161-2171, doi:10.1182/blood-2013-09-526053 (2014).

52 Li, Z. *et al.* Mouse embryonic head as a site for hematopoietic stem cell development. *Cell Stem Cell* **11**, 663-675, doi:10.1016/j.stem.2012.07.004 (2012).

53 Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* **10**, 1096-1098, doi:10.1038/nmeth.2639 (2013).

54 Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181, doi:10.1038/nprot.2014.006 (2014).

55 Li, L. *et al.* Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions. *Cell Stem Cell* **20**, 858-873 e854, doi:10.1016/j.stem.2017.03.007 (2017).

56 Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep* **2**, 666-673, doi:10.1016/j.celrep.2012.08.003 (2012).

57 Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods* **11**, 163-166, doi:10.1038/nmeth.2772 (2014).

58 Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

59 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).

60 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).

61 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, doi:10.1038/nbt.3192 (2015).

62 Corada, M., Morini, M. F. & Dejana, E. Signaling pathways in the specification of arteries and veins. *Arterioscler Thromb Vasc Biol* **34**, 2372-2377, doi:10.1161/ATVBAHA.114.303218 (2014).

63 Simons, M. & Eichmann, A. Molecular controls of arterial morphogenesis. *Circ Res* **116**, 1712-1724, doi:10.1161/CIRCRESAHA.116.302953 (2015).

64 Su, T. *et al.* Single-cell analysis of early progenitor cells that build coronary arteries. *Nature* **559**, 356-362, doi:10.1038/s41586-018-0288-7 (2018).

65 Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196, doi:10.1126/science.aad0501 (2016).

66 Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309-313, doi:10.1038/nature20123 (2016).

67 Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**, 979-982, doi:10.1038/nmeth.4402 (2017).

68 Chen, J., Schlitzer, A., Chakarov, S., Ginhoux, F. & Poidinger, M. Mpath maps multi-branching single-cell trajectories revealing progenitor cell progression during development. *Nat Commun* **7**, 11988, doi:10.1038/ncomms11988 (2016).

69 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).

70 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).

71 Ibarra-Soria, X. *et al.* Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation. *Nat Cell Biol* **20**, 127-134, doi:10.1038/s41556-017-0013-z (2018).

72 Hoeffel, G. *et al.* C-Myb(+) erythro-myeloid progenitor-derived fetal monocytes give rise to adult tissue-resident macrophages. *Immunity* **42**, 665-678, doi:10.1016/j.immuni.2015.03.011 (2015).

73 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-1086, doi:10.1038/nmeth.4463 (2017).

74 Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7, doi:10.1186/1471-2105-14-7 (2013).

75 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353-D361, doi:10.1093/nar/gkw1092 (2017).

76 Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108-112, doi:10.1038/nature08460 (2009).

77 Zhang, H. M. *et al.* AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res* **43**, D76-81, doi:10.1093/nar/gku887 (2015).

78 Bausch-Fluck, D. *et al.* A mass spectrometric-derived cell surface protein atlas. *PLoS One* **10**, e0121314, doi:10.1371/journal.pone.0121314 (2015).

79 Tripathi, S. *et al.* Meta- and Orthogonal Integration of Influenza "OMICs" Data Defines a Role for UBR4

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in Virus Budding. *Cell Host Microbe* **18**, 723-735, doi:10.1016/j.chom.2015.11.002 (2015).

80 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, doi:10.1089/omi.2011.0118 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE LEGENDS

Fig. 1. Transcriptomic identification and molecular characteristics of the HECs in AGM Region.

(a) Schematic illustration of the strategies used for embryo dissection and cell preparation for the subsequent scRNA-seq. The involved body part and AGM region is indicated as blue and green, respectively, with head, limb buds, heart, visceral bud, and umbilical and vitelline vessels outside the embryo proper excluded.

(b) PCA plots with three clusters (earlyAEC, lateAEC and HEC) (left), sampling locations (middle) and embryonic stages (right) mapped onto it.

(c) Metascape network enrichment analysis with top 10 enriched terms exhibited to the right. Each cluster is represented by different colors and each enriched term is represented by a circle node. Number in the bracket indicates the *P* value based on -log10.

(d) Classification of the indicated cells into quiescent phase and other cycling phases (G1, S and G2M) based on the average expression of G1/S and G2/M gene sets (left). Stacked bar chart showing the constitution of different cell cycle phases in the corresponding three clusters shown to the left (right).

(e) Violin plot showing the number of transcripts for ribosomal related genes detected in each single cell of the indicated clusters. Wilcoxon Rank Sum test is employed to test the significance of difference and *P* values are indicated for the comparison. $P < 0.05$ is considered statistically significant.

(f) Scatterplot showing the average arteriovenous scores of the cells in each cluster for mouse dataset in this paper (left) and human dataset from published articles (right),

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

respectively. Main distribution ranges of arteriovenous scores in each cluster are also indicated as an oval shape.

(g) Pseudotemporal ordering of the cells in select three clusters inferred by monocle 2, with pseudotime (left), clusters (middle) and sampling stages (right) mapped to it. HEC specification and AEC maturation directions are indicated as orange and deep red arrows, respectively.

(h) Heatmap showing the expression of the indicated genes (smoothed over 15 adjacent cells) with cells ordered along the pseudotime axis of HEC specification branch inferred by monocle 2.

(i) Eight major expression patterns identified from the differentially expressed genes in HEC or lateAEC as compared to earlyAEC. Arrows showing the changes in HEC or lateAEC as compared to earlyAEC. The numbers of pattern genes are indicated to the right.

(j) Heatmaps showing the relative expressions (smoothed over 20 adjacent cells) of the TFs belonging to the pattern genes with cells ordered along the pseudotime axis and genes ordered by patterns.

Fig. 2. Efficiently isolating the HSC-competent and endothelial-haematopoietic bi-potent HECs before HSC emergence.

(a) Gene lists of the top ten cell surface molecules significantly overrepresented in HEC as compared to the indicated cell populations (first 3 lines) and those positively correlated with Runx1 within 4 EC clusters (vEC, earlyAEC, lateAEC and HEC, last line). Non-HEC, cells except for HEC within 4 EC clusters. Highlights in red font indicate the candidates used for further functional analysis.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(b) Representative whole-mount staining of CD44 at E10.0 AGM region, showing CD44 is expressed in the whole endothelial layer of the dorsal aorta and roots of its proximal branches. DA, dorsal aorta; Scale bar, 100 μm.

(c) Representative FACS plots for cell sorting of the E9.5-E10.0 caudal half for co-culture/transplantation assay and the donor chimerism at 16 weeks after transplantation of the derivatives of the indicated cell populations.

(d) Blood chimerism of the primary (I°) and corresponding secondary (II°) recipients at 16 weeks post-transplantation. The primary recipients were transplanted with the derivatives of the indicated cells from the caudal half of E9.5-E10.0 embryos. The paired primary and corresponding secondary repopulated mice are shown as the same symbol and color.

(e) Bars represent the percent donor contribution to the granulocytes/monocytes (GM, red), B lymphocytes (green), and T lymphocytes (purple) in the peripheral blood of the primary (I°) and secondary (II°) recipients at 16 weeks post-transplantation. The paired primary and corresponding secondary repopulated mice are shown as the same colors below.

(f) FACS plot of Flk1 expression in the indicated population of E10.0 AGM region, with PK44 ($CD41^-CD43^-CD45^-CD31^+CD201^+Kit^+CD44^+$) cells (red) mapped onto it. Red box indicates the gate of $Flk1^+$ cells.

(g) t-SNE plot of the cells included in the filtered initial dataset and PK44 dataset, with clusters mapped on it. PK44, $CD41^-CD43^-CD45^-CD31^+CD201^+Kit^+CD44^+$ population from E10.0 AGM region.

(h) Heatmap showing the relative expressions of HEC feature genes, which are defined as those significantly highly expressed as compared to others including HC, vEC, earlyAEC and

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

lateAEC, in the indicated cell populations. Selected HEC feature genes are shown to the right with pre-HSC signature genes marked as aquamarine.

(i) Representative CD31 and CD45 immunostaining on the cultures of single PK44 cells from E10.0 AGM region, showing typical morphologies regarding distinct differentiation potentials. Cell frequencies of each kind of potential are also shown. Data are from 5 independent experiments with totally 15 embryos used. Scale bars, 400 μm.

(j) Expression of Kit and CD201 in the index-sorted single PK44 cells with differentiation potential based on in vitro functional evaluation. Cells with different kinds of potential are mapped onto the reference FACS plots (grey dots). Box in the middle plot indicates the gate for FACS sorting of PK44 cells in E10.0 AGM region and its enlarged view is shown to the right.

Fig. 3. Relationship between HSC-primed HECs and T1 pre-HSCs.

(a) Representative FACS plots for sorting of the T1 pre-HSCs ($CD31^{+}CD45^{-}CD41^{low}Kit^{+}CD201^{high}$) from E11.0 AGM region of mouse embryos. Red box indicates the sampling cells for scRNA-seq.

(b) Violin plots showing the expression levels of indicated genes in tif-HEC (including cluster HEC and PK44), T1 pre-HSC and lateAEC.

(c) t-SNE plot of the cells included in the filtered initial dataset, PK44 dataset and T1 pre-HSC dataset, with clusters mapped on it. Cluster HEC and PK44 are combined as tif-HEC.

(d) PCA plot of tif-HEC and T1 pre-HSC populations.

(e) Enriched terms of PC2 positive and negative genes are shown, corresponding to the

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

properties distinguishing tif-HEC and T1 pre-HSC, respectively.

(f) Heatmap showing top 20 positive and negative genes of PC2. Genes were ordered by their contributions to PC2.

(g) Trajectory of AEC clusters, tif-HEC and T1 pre-HSC inferred by Mpath. Arrows indicate the development directions predicted by sampling stages.

(h) Representative FACS plots for sorting of the PK44 cells from E10.0 AGM region (left) and analysis of the immunophenotypic T1 pre-HSCs (right) after cultured in vitro for 4 days.

(i) Representative CD31 immunostaining on the cultures of single T1 pre-HSCs from E11.0 AGM region, showing typical morphologies regarding distinct differentiation capacities. Cell frequencies of each type are also shown. Data are from 7 independent experiments with totally 89 embryos used. Scale bars, 400 μm.

Fig. 4. Identifying Neurl3 as a signature gene of HSC-primed HECs validated by functional and transcriptomic evaluation.

(a) Dot plot showing the average and percentage expression of HEC signature genes in the indicated clusters. Genes are ordered by their median expression level in tif-HEC. Pre-HSC signature genes are marked as aquamarine.

(b) Schematic model of the gene targeting strategy for generating *Neurl3*$^{EGFP/+}$ reporter mouse line via CRISPR/Cas9 system.

(c) Representative FACS analysis of the E10.0 AGM region in *Neurl3*$^{EGFP/+}$ embryos, FACS plot to the right showing PK44 cells (red dots) mapped on it.

(d) Representative FACS plot for sorting of the indicated cell populations from E10.0 caudal half of *Neurl3*$^{EGFP/+}$ embryos.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(e) Graph showing the donor chimerism at 16 weeks after transplantation of the derivatives of the indicated populations from the caudal half of E10.0 *Neurl3*$^{EGFP/+}$ embryos.

(f) Graph showing the donor chimerism at 4-16 weeks post-transplantation. The recipients were transplanted with the derivatives of CD41^{-}CD43^{-}CD45^{-}CD31^{+}CD44^{+}Neurl3-EGFP^{+} population from the caudal half of E10.0 *Neurl3*$^{EGFP/+}$ embryos. Number of repopulated/total recipients is shown in the brackets.

(g) t-SNE plot of the cells included in the filtered initial dataset and additional PK44 and NE+ datasets, with clusters mapped on it. Cluster HEC and PK44 are combined as tif-HEC. NE+, CD41^{-}CD43^{-}CD45^{-}CD31^{+}CD44^{+}Neurl3-EGFP^{+} population from E10.0 AGM region.

(h) Dot plot showing the average and percentage expression of selected HEC feature genes in the indicated clusters. Pre-HSC signature genes are marked as aquamarine.

(i) Heatmap showing the correlation coefficient between each two clusters with hierarchical clustering using average method. Pearson correlation coefficient is calculated using average expression of highly variable genes in each cluster.

Fig. 5. In situ localization and in vitro function of the dynamic HECs marked by Neurl3-EGFP reporter.

(a) Representative immunostaining on cross sections at AGM region of E9.5 (upper), E10.0 (middle) and E10.5 (lower) *Neurl3*$^{EGFP/+}$ embryos. Arrows indicate Neurl3^{+} aortic ECs. Yellow arrowheads indicate Neurl3^{+} bulging and bulged cells and IAHCs. Aquamarine arrowheads indicate CD44^{+}Runx1^{+}Neurl3^{-} haematopoietic cells distributed outside the aorta. nt, neural tube; DA, dorsal aorta. Scale bars, 100 μm.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(b) Representative FACS analysis of the E10.0 AGM region of *Neurl3*$^{EGFP/+}$ embryos. FACS plots to the right showing PK44 cells (red dots，upper) and CD31^{+}Kit^{high} cells (blue dots, lower) mapped on, respectively, with their contributions to each gated population indicated.

(c) Representative CD31 and CD45 immunostaining on the cultures of single CD41^{-}CD43^{-}CD45^{-}CD31^{+}CD44^{+}Neurl3-EGFP^{+} cells from E10.0 AGM region of *Neurl3*$^{EGFP/+}$ embryos, showing typical morphologies regarding distinct differentiation potentials. Cell frequencies of each kind of potential are also shown. Data are from 5 independent experiments with totally 37 embryos used. Scale bars, 400 μm.

(d) Column charts showing the proportions of positive wells in the indicated populations (lower) for each kind of potential. The experiments were performed with CD41^{-}CD43^{-}CD45^{-}CD31^{+}CD44^{-}Neurl3-EGFP^{+} single cells from E9.5 caudal half or E10.0-E10.5 AGM region of *Neurl3*$^{EGFP/+}$ embryos with PK44 indexed. Progenies from PK44 and non-PK44 cells are represented by distinct fill patterns.

(e) Expression of CD44 and Neurl3-EGFP and values of FSC-A and SSC-A in the index-sorted single CD41^{-}CD43^{-}CD45^{-}CD31^{+}CD44^{+}Neurl3-EGFP^{+} cells with differentiation potential based on in vitro functional evaluation. Cells with different kinds of potential are mapped onto the reference FACS plots (grey dots). Solid boxes (left of each stage) indicate the gates of the populations for FACS sorting. The enlarged views of solid boxes are shown below.

Fig. 6. Molecular evolution underlying the specification of HSC-primed HECs from primitive vascular ECs.

(a) Trajectory of pEC, vEC, pAEC, earlyAEC, lateAEC and HEC inferred by Mpath. Arrows

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

indicate the development directions predicted by sampling stages.

(b) t-SNE plot showing the distribution of the four clusters involved in hemogenic specification. Other cells are in grey.

(c) Pseudotemporal ordering of the cells included in the indicated five clusters inferred by monocle 2 (left), with clusters (upper left) and sampling stages (lower left) mapped to it. HEC specification directions are indicated as red arrows. Smooth distribution of clusters (upper right) and sampling stages (lower right) along pseudotime by using Gaussian kernel density estimate are shown to the right.

(d) Dynamic changes of five gene expression patterns along the trajectory ordered by pseudotime inferred by monocle 2. For each pattern, principal curves are fitted on expression levels of the genes in that pattern along pseudotemporal order, using local polynomial regression fitting method. Randomly down-sampling is performed in pEC and pAEC clusters for better visualization.

(e) Heatmap showing the relative expression of the core TFs which belong to the regulons that the genes within exhibit significant overlapping with the pattern genes. Cells are ordered by pseudotime and TFs are ordered by Patterns.

(f) Heatmap showing smoothed (along adjacent 25 cells) and scaled enrichment scores of top 50 KEGG pathways along the order by pseudotime. Pathways are ordered by hierarchical clustering using ward.D method.

(g) Scatter plots showing the relative activity levels of pathways or GO terms with loess smoothed fit curves and 95% confidence interval indicated. Relative activity levels are represented by the PC1 scores of expression levels of the genes in a given set. The sign or

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1207 direction of PC1 is corrected according to positive correlation with averaged expression

1208 levels.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SUPPLEMENTARY FIGURE LEGENDS

Fig. S1. Information, clustering of initial dataset and molecular characteristics of major clusters.

(a) Embryo, independent experiment, and cell number information for scRNA-seq. DA, dorsal aortic luminal layer of AGM region.

(b) Whole-mount image of the E11.0 AGM region labeled with Oregon Green 488.

(c) Boxplots showing the number of genes (left) and transcripts (right) in each single cell of different locations.

(d) t-SNE plots with clusters (left), sampling locations (right) and embryonic stages (right) mapped onto it.

(e) Violin plots showing the expression levels of indicated genes in six clusters identified in the initial dataset.

(f) Cell number information of the spatiotemporal distribution of distinct clusters.

(g) Volcano plots showing differentially expressed genes (marked as blue or red) between two sub-clusters by forced clustering in earlyAEC and HEC, respectively. Top 10 (earlyAEC) or all (HEC) differentially expressed genes are indicated. *Runx1* and *Kit* are also indicated.

(h) Violin plots showing the number of genes (left) and transcripts (right) in each single cell of the indicated clusters. Wilcoxon Rank Sum test is employed to test the significance of difference and *P* values are indicated for the comparison. $P < 0.05$ is considered statistically significant.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(i) Heatmap showing the relative expression levels of genes in eight patterns among earlyAEC, lateAEC and HEC.

(j) Network view of TFs positively correlated with the gene expression patterns. A deeper background color of the gene name indicates a higher positive correlation of the TF to that expression pattern.

(k) Bar chart showing the top 50 genes positively correlated with Runx1 within cell population including earlyAEC, lateAEC and HEC. Genes included in the patterns identified above are marked as indicated.

Fig. S2. Identification of the HSC-competent and endothelial-haematopoietic bi-potent HECs.

(a) Detailed information of the co-culture/transplantation assays performed with E9.5-E10.0 caudal half cells.

(b) Blood chimerism of the primary and secondary recipients at 4-16 weeks post-transplantation. The primary recipients were transplanted with the derivatives of the indicated cell populations from the caudal half of E9.5-E10.0 embryos. The paired primary and corresponding secondary repopulated mice are show as the same symbol and color. Numbers of repopulated/total recipients are shown in the brackets. Only the recipients survived to 16 weeks post-transplantation are shown.

(c) FACS plots showing representative primary recipients with long-term (16 weeks), multi-organ and multi-lineage repopulations transplanted with the derivatives of the indicated cell populations from the caudal half of E9.5-E10.0 embryos. Donor-derived

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(CD45.1+CD45.2+) myeloid (Gr-1+/Mac-1+), B lymphoid (B220+), and T lymphoid (CD3+) cells in multiple haematopoietic organs are shown.

(d) Heatmap showing the expression of selected genes in earlyAEC, lateAEC, HEC and PK44 populations. Note the similarity of expression patterns between HEC and PK44.

(e) Graph showing the endothelial potential of different cell populations in E9.5-E10.0 body part of embryo proper. Cells with indicated immunophenotype were isolated by FACS, co-cultured with OP9 stromal cells for 7 days, and stained with CD31 to identify the endothelial tubes. Data are means ± s.d.. For E9.5 embryos, data are from 3 independent experiments with 6-9 embryo equivalents pooled for each experiment. For E10.0 embryos, data are from 3 independent experiments with 8-9 embryo equivalents pooled for each experiment.

(f) Detailed information of endothelial-haematopoietic bi-potential induction assays performed with cells from E9.5-E10.0 caudal half or AGM region.

Fig. S3. Identification of the HSC-competent HECs marked by Neurl3-EGFP reporter.

(a) Detailed information of the co-culture/transplantation assays performed with the caudal half cells from E10.0 *Neurl3*$^{EGFP/+}$ embryos.

(b) Boxplot showing the transcriptional expression level of *EGFP* in NE+ cell population, NE+, CD41^{-}CD43^{-}CD45^{-}CD31^{+}CD44^{+}Neurl3-EGFP^{+} population from AGM region of E10.0 *Neurl3*$^{EGFP/+}$ embryos.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(c) Scatter plots showing correlation of the expression of *EGFP* with that of *Neurl3* and *Runx1*, respectively. Fitted line and 95% confidence interval are shown in red. Pearson correlation coefficients and *P* values are also shown in blue text.

(d) Stacked bar chart showing the constitution of different cell cycle phases in the indicated clusters.

(e) Representative immunostaining on cross sections at AGM region of E11.0 *Neurl3*$^{EGFP/+}$ embryos. Arrow indicates Neurl3^{+} aortic ECs; Yellow arrowheads indicate Neurl3^{+} bulging and bulged cells and IAHCs. Aquamarine arrowheads indicate CD44^{+}Runx1^{+}Neurl3^{-} haematopoietic cells distributed outside the aorta. nt, neural tube; DA, dorsal aorta. Scale bars, 100 μm.

(f) Detailed information of endothelial-haematopoietic bi-potential induction assays performed with cells from E9.5 caudal half and E10.0-E10.5 AGM region of *Neurl3*$^{EGFP/+}$ embryos.

Fig. S4. Molecular programs from primitive vascular ECs to HSC-primed HECs.

(a) Embryo, independent experiment, and cell number information for additional scRNA-seq. sp, somite pairs.

(b) t-SNE plots with clusters (upper left), sampling locations (lower left), embryonic stages (upper right) and clusters previously defined (lower right) mapped onto it.

(c) Dot plot showing the average and percentage expression of selected marker genes in the indicated clusters.

(d) Pseudotemporal ordering of the cells involved in HEC specification, including those in pEC, pAEC, earlyAEC, HEC, and T1 pre-HSC, inferred by monocle 2, with

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.910356; this version posted January 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pseudotime mapped to it.

(e) Heatmap showing smoothed and scaled expression levels of 2,851 pattern genes. Genes are ordered by patterns. Cells are ordered by pseudotime.

(f) Dot plot showing the top six enriched Gene Ontology biological process (GO:BP) terms for each pattern. Dot color indicates statistical significance of the enrichment and dot size represents the fraction of genes annotated to each term.

(g) Scatter plots showing the expression levels of the TF genes previously reported to be functional in HSPC regeneration along the pseudotemporal order with loess smoothed fit curves and 95% confidence interval indicated. The patterns to which the genes belong are indicated by different fill colors. The core TFs of the significantly overlapped regulons are underlined.

Fig. 1

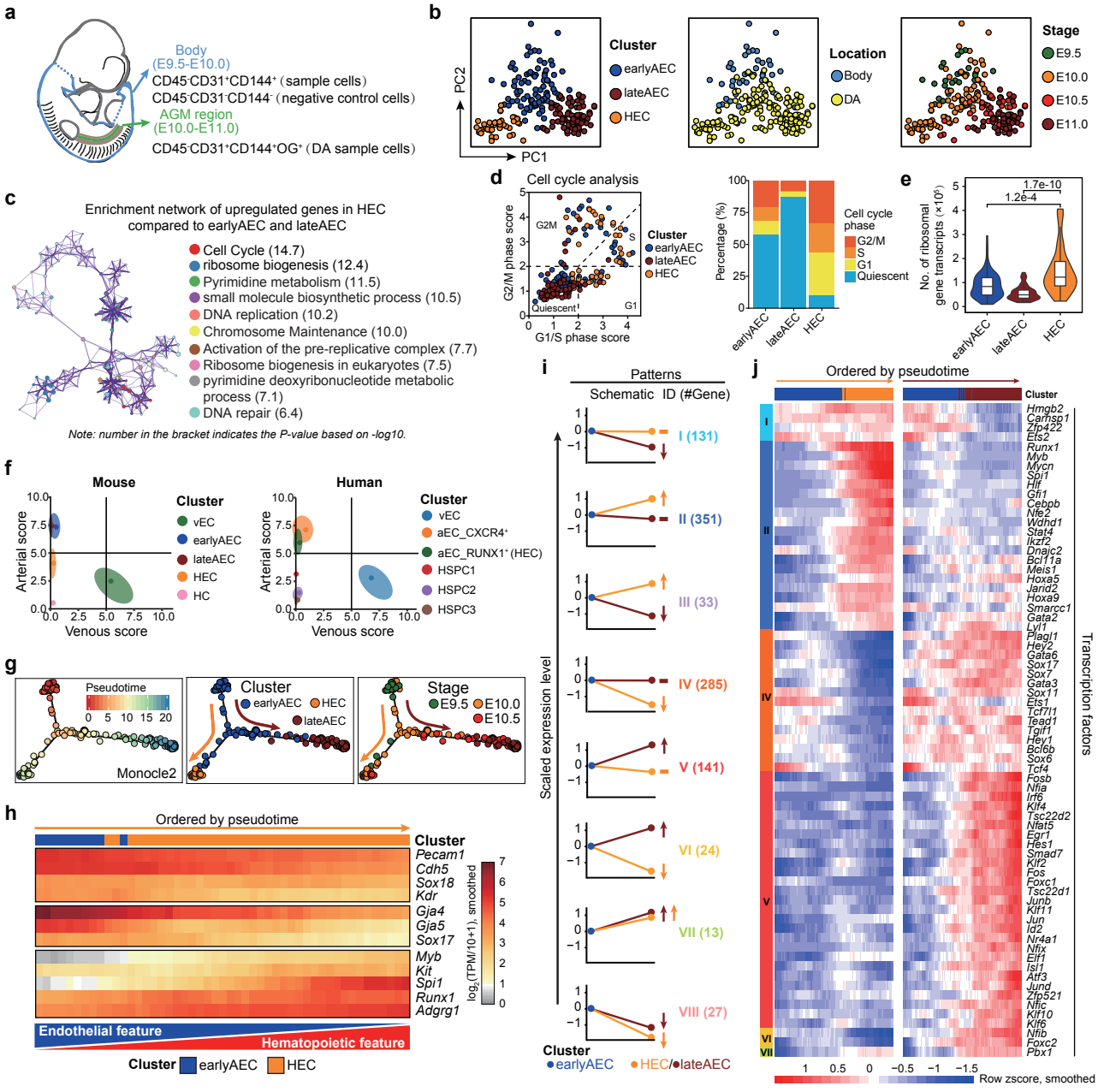

a
Body
(E9.5-E10.0)
CD45-CD31+CD144+ (sample cells)
CD45-CD31-CD144- (negative control cells)
AGM region
(E10.0-E11.0)
CD45-CD31+CD144+OG+ (DA sample cells)
b
PC2
PC1
Cluster
earlyAEC
lateAEC
HEC
Location
Body
DA
Stage
E9.5
E10.0
E10.5
E11.0
c
Enrichment network of upregulated genes in HEC compared to earlyAEC and lateAEC
Cell Cycle (14.7)
ribosome biogenesis (12.4)
Pyrimidine metabolism (11.5)
small molecule biosynthetic process (10.5)
DNA replication (10.2)
Chromosome Maintenance (10.0)
Activation of the pre-replicative complex (7.7)
Ribosome biogenesis in eukaryotes (7.5)
pyrimidine deoxyribonucleotide metabolic process (7.1)
DNA repair (6.4)
Note: number in the bracket indicates the P-value based on -log10.
d
Cell cycle analysis
G2/M phase score
G1/S phase score
G2M
S
G1
Quiescent
Cluster
earlyAEC
lateAEC
HEC
Percentage (%)
Cell cycle phase
G2/M
S
G1
Quiescent
e
No. of ribosomal gene transcripts (×10^5)
1.7e-10
1.2e-4
earlyAEC
lateAEC
HEC
f
Mouse
Human
Arterial score
Venous score
Cluster
vEC
earlyAEC
lateAEC
HEC
HC
aEC_CXCR4+
aEC_RUNX1+ (HEC)
HSPC1
HSPC2
HSPC3
g
Pseudotime
Monocle2
Cluster
earlyAEC
HEC
lateAEC
Stage
E9.5
E10.0
E10.5
h
Ordered by pseudotime
Cluster
Pecam1
Cdh5
Sox18
Kdr
Gja4
Gja5
Sox17
Myb
Kit
Spi1
Runx1
Adgrg1
log2(TPM/10+1), smoothed
Endothelial feature
Hematopoietic feature
Cluster
earlyAEC
HEC
i
Patterns
Schematic
ID (#Gene)
I (131)
II (351)
III (33)
IV (285)
V (141)
VI (24)
VII (13)
VIII (27)
Scaled expression level
Cluster
earlyAEC
HEC/lateAEC
j
Ordered by pseudotime
Cluster
Hmgb2
Camsp1
Zfp422
Ets2
Runx1
Myb
Mycn
Spi1
Hlf
Gfi1
Cebpb
Nfe2
Wdhd1
Stat4
Ikzf2
Dnajc2
Bcl11a
Meis1
Hoxa5
Jarid2
Hoxa9
Smarcc1
Gata2
Lyl1
Plagl1
Hey2
Gata6
Sox17
Sox7
Gata3
Sox11
Ets1
Tcf7l1
Tead1
Tgif1
Hey1
Bcl6b
Sox6
Tcf4
Fosb
Nfia
Irf6
Klf4
Tsc22d2
Nfat5
Egr1
Hes1
Smad7
Klf2
Fos
Foxc1
Tsc22d1
Junb
Klf11
Jun
Id2
Nr4a1
Nfix
Elf1
Isl1
Atf3
Jund
Zfp521
Nfic
Klf10
Klf6
Nfib
Foxc2
Pbx1
Transcription factors
Row zscore, smoothed

Fig. 2

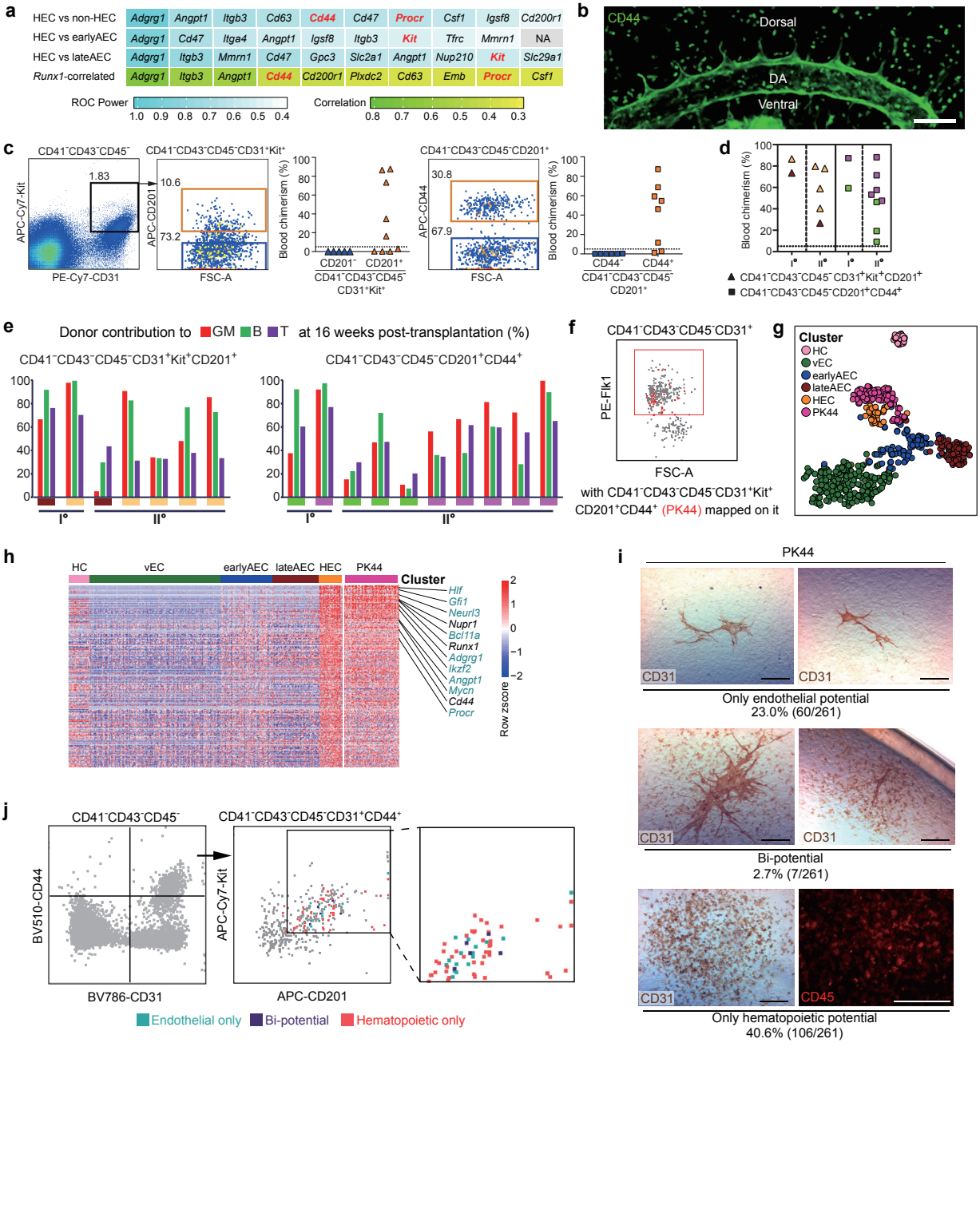

a
HEC vs non-HEC
Adgrg1 Angpt1 Itgb3 Cd63 Cd44 Cd47 Procr Csf1 Igsf8 Cd200r1
HEC vs earlyAEC
Adgrg1 Cd47 Itga4 Angpt1 Igsf8 Itgb3 Kit Tfrc Mmrn1 NA
HEC vs lateAEC
Adgrg1 Itgb3 Mmrn1 Cd47 Gpc3 Slc2a1 Angpt1 Nup210 Kit Slc29a1
Runx1-correlated
Adgrg1 Itgb3 Angpt1 Cd44 Cd200r1 Plxdc2 Cd63 Emb Procr Csf1
ROC Power
1.0 0.9 0.8 0.7 0.6 0.5 0.4
Correlation
0.8 0.7 0.6 0.5 0.4 0.3
b
CD44
Dorsal
DA
Ventral
c
CD41⁻CD43⁻CD45⁻
1.83
APC-Cy7-Kit
PE-Cy7-CD31
CD41⁻CD43⁻CD45⁻CD31⁺Kit⁺
10.6
73.2
APC-CD201
FSC-A
Blood chimerism (%)
CD201⁻
CD201⁺
CD41⁻CD43⁻CD45⁻ CD31⁺Kit⁺
CD41⁻CD43⁻CD45⁻CD201⁺
30.8
67.9
APC-CD44
CD44⁻
CD44⁺
CD41⁻CD43⁻CD45⁻ CD201⁺
d
Blood chimerism (%)
I°
II°
I°
II°
▲ CD41⁻CD43⁻CD45⁻CD31⁺Kit⁺CD201⁺
■ CD41⁻CD43⁻CD45⁻CD201⁺CD44⁺
e
Donor contribution to GM B T at 16 weeks post-transplantation (%)
CD41⁻CD43⁻CD45⁻CD31⁺Kit⁺CD201⁺
CD41⁻CD43⁻CD45⁻CD201⁺CD44⁺
I°
II°
I°
II°
f
CD41⁻CD43⁻CD45⁻CD31⁺
PE-Flk1
FSC-A
with CD41⁻CD43⁻CD45⁻CD31⁺Kit⁺
CD201⁺CD44⁺ (PK44) mapped on it
g
Cluster
HC
vEC
earlyAEC
lateAEC
HEC
PK44
h
HC
vEC
earlyAEC
lateAEC
HEC
PK44
Cluster
Hlf
Gfi1
Neurl3
Nupr1
Bcl11a
Runx1
Adgrg1
Ikzf2
Angpt1
Mycn
Cd44
Procr
Row zscore
2
1
0
−1
−2
i
PK44
CD31
CD31
Only endothelial potential
23.0% (60/261)
CD31
CD31
Bi-potential
2.7% (7/261)
CD31
CD45
Only hematopoietic potential
40.6% (106/261)
j
CD41⁻CD43⁻CD45⁻
CD41⁻CD43⁻CD45⁻CD31⁺CD44⁺
BV510-CD44
BV786-CD31
APC-Cy7-Kit
APC-CD201
Endothelial only
Bi-potential
Hematopoietic only

Fig. 3

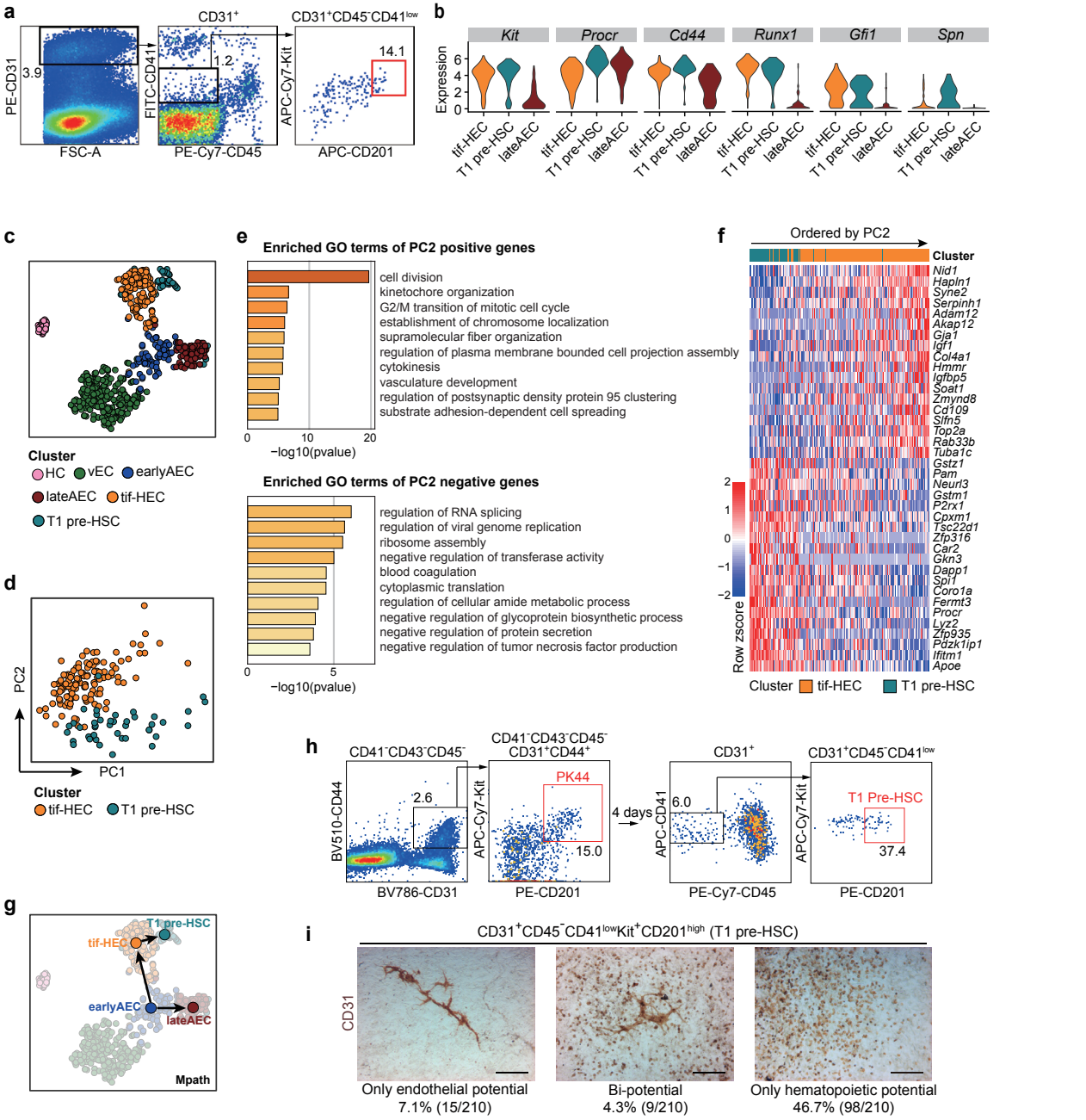

a
CD31+
CD31+CD45-CD41low
PE-CD31
FITC-CD41
APC-Cy7-Kit
3.9
1.2
14.1
FSC-A
PE-Cy7-CD45
APC-CD201
b
Kit
Procr
Cd44
Runx1
Gfi1
Spn
Expression
tif-HEC
T1 pre-HSC
lateAEC
c
Cluster
HC
vEC
earlyAEC
lateAEC
tif-HEC
T1 pre-HSC
d
PC2
PC1
Cluster
tif-HEC
T1 pre-HSC
e
Enriched GO terms of PC2 positive genes
cell division
kinetochore organization
G2/M transition of mitotic cell cycle
establishment of chromosome localization
supramolecular fiber organization
regulation of plasma membrane bounded cell projection assembly
cytokinesis
vasculature development
regulation of postsynaptic density protein 95 clustering
substrate adhesion-dependent cell spreading
−log10(pvalue)
Enriched GO terms of PC2 negative genes
regulation of RNA splicing
regulation of viral genome replication
ribosome assembly
negative regulation of transferase activity
blood coagulation
cytoplasmic translation
regulation of cellular amide metabolic process
negative regulation of glycoprotein biosynthetic process
negative regulation of protein secretion
negative regulation of tumor necrosis factor production
−log10(pvalue)
f
Ordered by PC2
Cluster
Nid1
Hapln1
Syne2
Serpinh1
Adam12
Akap12
Gja1
Igf1
Col4a1
Hmmr
Igfbp5
Soat1
Zmynd8
Cd109
Slfn5
Top2a
Rab33b
Tuba1c
Gstz1
Pam
Neurl3
Gstm1
P2rx1
Cpxm1
Tsc22d1
Zfp316
Car2
Gkn3
Dapp1
Spi1
Coro1a
Fermt3
Procr
Lyz2
Zfp935
Pdzk1ip1
Ifitm1
Apoe
Row zscore
Cluster
tif-HEC
T1 pre-HSC
g
T1 pre-HSC
tif-HEC
earlyAEC
lateAEC
Mpath
h
CD41-CD43-CD45-
CD41-CD43-CD45-
CD31+CD44+
PK44
2.6
15.0
BV510-CD44
APC-Cy7-Kit
BV786-CD31
PE-CD201
4 days
CD31+
CD31+CD45-CD41low
6.0
T1 Pre-HSC
37.4
APC-CD41
APC-Cy7-Kit
PE-Cy7-CD45
PE-CD201
i
CD31+CD45-CD41lowKit+CD201high (T1 pre-HSC)
CD31
Only endothelial potential
7.1% (15/210)
Bi-potential
4.3% (9/210)
Only hematopoietic potential
46.7% (98/210)

Fig. 4

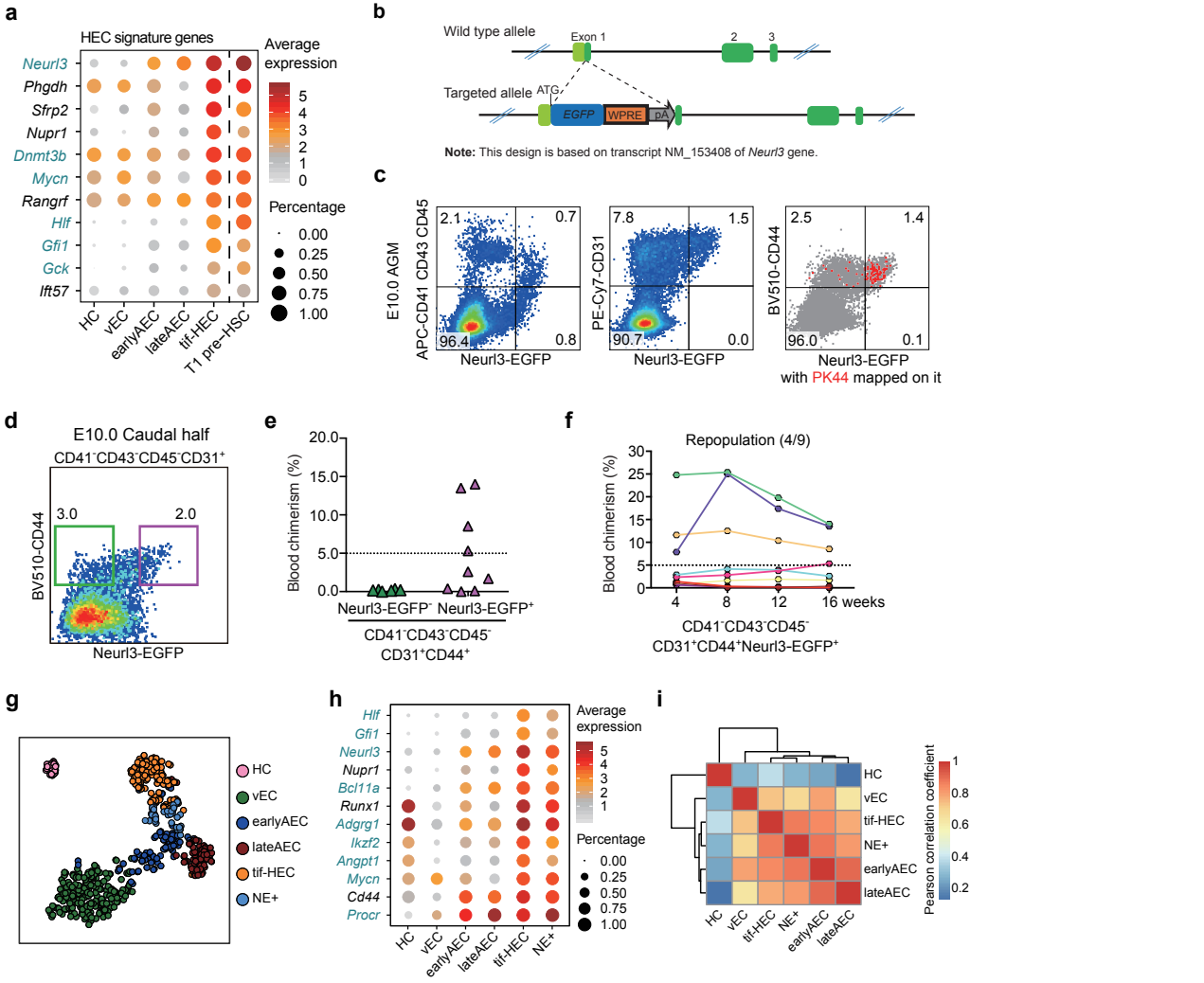

a
HEC signature genes
Neurl3
Phgdh
Sfrp2
Nupr1
Dnmt3b
Mycn
Rangrf
Hlf
Gfi1
Gck
Ift57
HC
vEC
earlyAEC
lateAEC
tif-HEC
T1 pre−HSC
Average expression
5
4
3
2
1
0
Percentage
0.00
0.25
0.50
0.75
1.00
b
Wild type allele
Exon 1
2
3
Targeted allele
ATG
EGFP
WPRE
pA
Note: This design is based on transcript NM_153408 of Neurl3 gene.
c
E10.0 AGM
APC-CD41 CD43 CD45
2.1
0.7
96.4
0.8
Neurl3-EGFP
PE-Cy7-CD31
7.8
1.5
90.7
0.0
Neurl3-EGFP
BV510-CD44
2.5
1.4
96.0
0.1
Neurl3-EGFP
with PK44 mapped on it
d
E10.0 Caudal half
CD41⁻CD43⁻CD45⁻CD31⁺
3.0
2.0
BV510-CD44
Neurl3-EGFP
e
Blood chimerism (%)
20.0
15.0
10.0
5.0
0.0
Neurl3-EGFP⁻
Neurl3-EGFP⁺
CD41⁻CD43⁻CD45⁻
CD31⁺CD44⁺
f
Repopulation (4/9)
Blood chimerism (%)
30
25
20
15
10
5
0
4
8
12
16 weeks
CD41⁻CD43⁻CD45⁻
CD31⁺CD44⁺Neurl3-EGFP⁺
g
HC
vEC
earlyAEC
lateAEC
tif-HEC
NE+
h
Hlf
Gfi1
Neurl3
Nupr1
Bcl11a
Runx1
Adgrg1
Ikzf2
Angpt1
Mycn
Cd44
Procr
HC
vEC
earlyAEC
lateAEC
tif-HEC
NE+
Average expression
5
4
3
2
1
Percentage
0.00
0.25
0.50
0.75
1.00
i
HC
vEC
tif-HEC
NE+
earlyAEC
lateAEC
Pearson correlation coefficient
1
0.8
0.6
0.4
0.2

Fig. 5

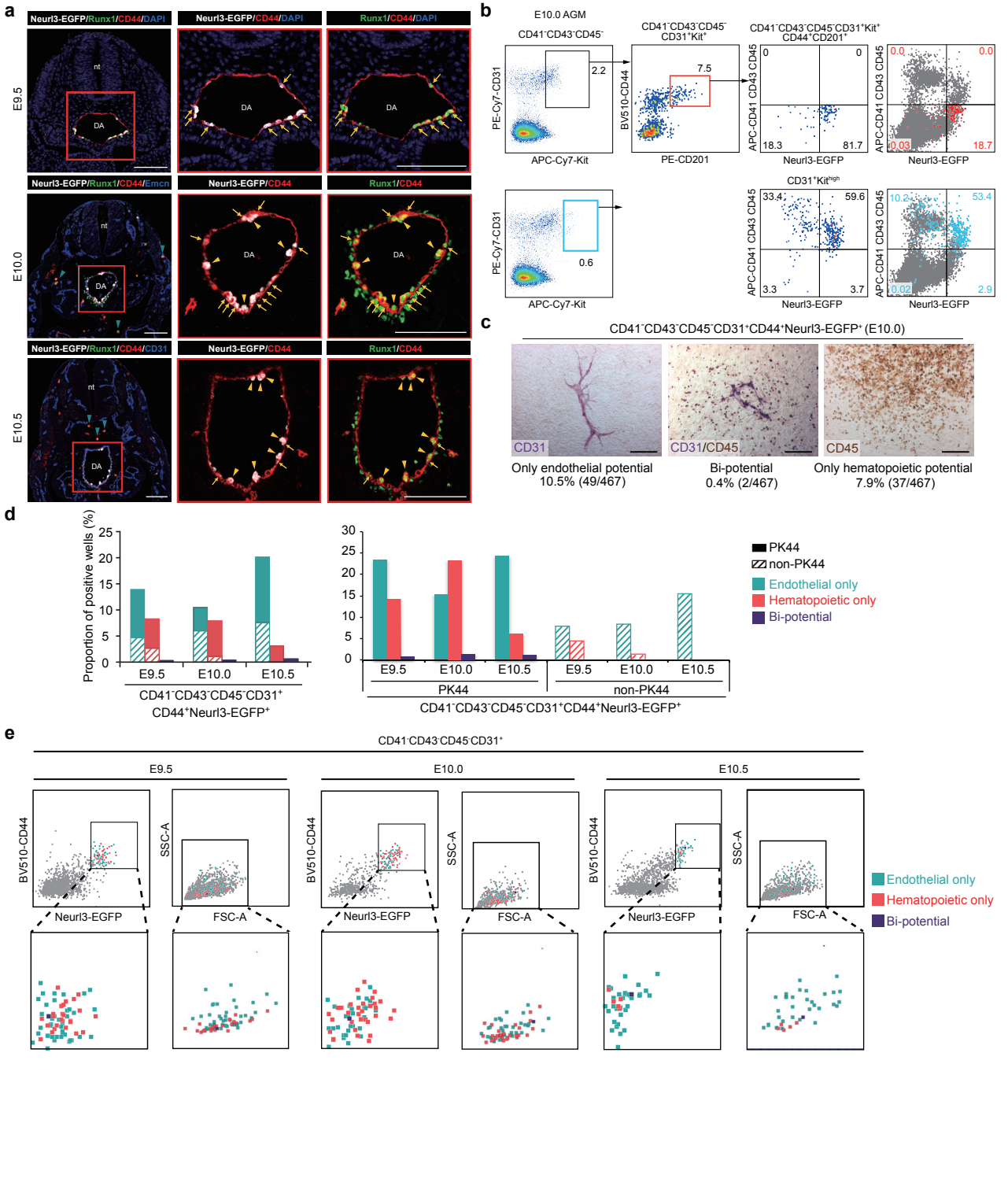

a
Neurl3-EGFP/Runx1/CD44/DAPI
Neurl3-EGFP/CD44/DAPI
Runx1/CD44/DAPI
E9.5
E10.0
E10.5
nt
DA
Neurl3-EGFP/Runx1/CD44/Emcn
Neurl3-EGFP/CD44
Runx1/CD44
Neurl3-EGFP/Runx1/CD44/CD31
b
E10.0 AGM
CD41-CD43-CD45-
PE-Cy7-CD31
APC-Cy7-Kit
2.2
CD41-CD43-CD45-CD31+Kit+
BV510-CD44
PE-CD201
7.5
CD41-CD43-CD45-CD31+Kit+CD44+CD201+
APC-CD41 CD43 CD45
Neurl3-EGFP
0
18.3
81.7
0.0
0.03
18.7
CD31+Kithigh
33.4
59.6
3.3
3.7
10.2
53.4
0.02
2.9
0.6
c
CD41-CD43-CD45-CD31+CD44+Neurl3-EGFP+ (E10.0)
CD31
CD31/CD45
CD45
Only endothelial potential
10.5% (49/467)
Bi-potential
0.4% (2/467)
Only hematopoietic potential
7.9% (37/467)
d
Proportion of positive wells (%)
E9.5
E10.0
E10.5
CD41-CD43-CD45-CD31+
CD44+Neurl3-EGFP+
PK44
non-PK44
CD41-CD43-CD45-CD31+CD44+Neurl3-EGFP+
Endothelial only
Hematopoietic only
Bi-potential
e
CD41-CD43-CD45-CD31+
E9.5
E10.0
E10.5
BV510-CD44
Neurl3-EGFP
SSC-A
FSC-A
Endothelial only
Hematopoietic only
Bi-potential

Fig. 6

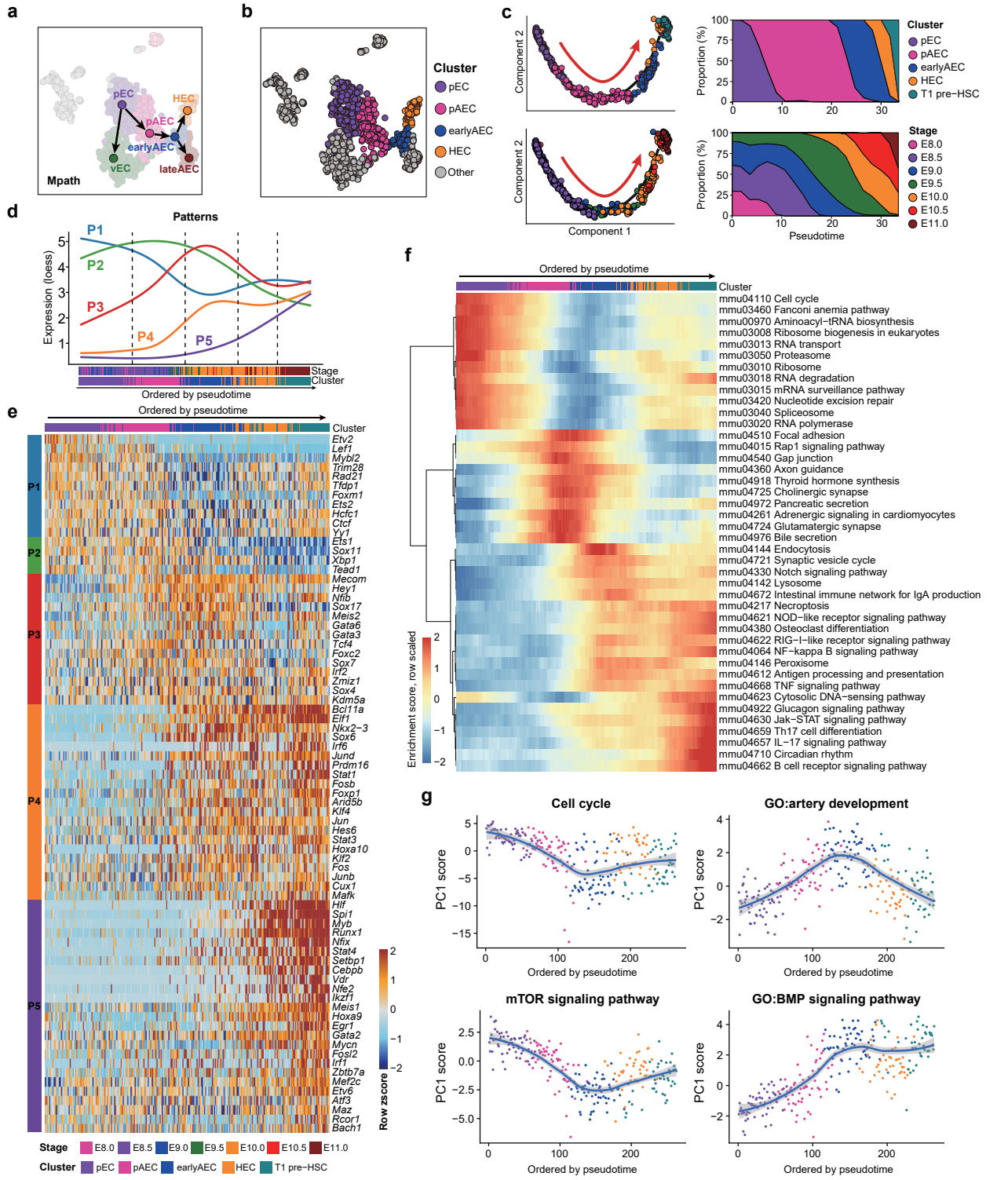

a
pEC
HEC
pAEC
earlyAEC
vEC
lateAEC
Mpath
b
Cluster
pEC
pAEC
earlyAEC
HEC
Other
c
Component 2
Component 1
Proportion (%)
Cluster
pEC
pAEC
earlyAEC
HEC
T1 pre−HSC
Stage
E8.0
E8.5
E9.0
E9.5
E10.0
E10.5
E11.0
Pseudotime
d
Patterns
P1
P2
P3
P4
P5
Expression (loess)
Stage
Cluster
Ordered by pseudotime
e
Ordered by pseudotime
Cluster
P1
P2
P3
P4
P5
Etv2
Lef1
Mybl2
Trim28
Rad21
Tfdp1
Foxm1
Ets2
Hcfc1
Ctcf
Yy1
Ets1
Sox11
Xbp1
Tead1
Mecom
Hey1
Nfib
Sox17
Meis2
Gata6
Gata3
Tcf4
Foxc2
Sox7
Irf2
Zmiz1
Sox4
Kdm5a
Bcl11a
Elf1
Nkx2−3
Sox6
Irf6
Jund
Prdm16
Stat1
Fosb
Foxp1
Arid5b
Klf4
Jun
Hes6
Stat3
Hoxa10
Klf2
Fos
Junb
Cux1
Mafk
Hlf
Spi1
Myb
Runx1
Nfix
Stat4
Setbp1
Cebpb
Vdr
Nfe2
Ikzf1
Meis1
Hoxa9
Egr1
Gata2
Mycn
Fosl2
Irf1
Zbtb7a
Mef2c
Etv6
Atf3
Maz
Rcor1
Bach1
Row zscore
Stage E8.0 E8.5 E9.0 E9.5 E10.0 E10.5 E11.0
Cluster pEC pAEC earlyAEC HEC T1 pre−HSC
f
Ordered by pseudotime
Cluster
mmu04110 Cell cycle
mmu03460 Fanconi anemia pathway
mmu00970 Aminoacyl−tRNA biosynthesis
mmu03008 Ribosome biogenesis in eukaryotes
mmu03013 RNA transport
mmu03050 Proteasome
mmu03010 Ribosome
mmu03018 RNA degradation
mmu03015 mRNA surveillance pathway
mmu03420 Nucleotide excision repair
mmu03040 Spliceosome
mmu03020 RNA polymerase
mmu04510 Focal adhesion
mmu04015 Rap1 signaling pathway
mmu04540 Gap junction
mmu04360 Axon guidance
mmu04918 Thyroid hormone synthesis
mmu04725 Cholinergic synapse
mmu04972 Pancreatic secretion
mmu04261 Adrenergic signaling in cardiomyocytes
mmu04724 Glutamatergic synapse
mmu04976 Bile secretion
mmu04144 Endocytosis
mmu04721 Synaptic vesicle cycle
mmu04330 Notch signaling pathway
mmu04142 Lysosome
mmu04672 Intestinal immune network for IgA production
mmu04217 Necroptosis
mmu04621 NOD−like receptor signaling pathway
mmu04380 Osteoclast differentiation
mmu04622 RIG−I−like receptor signaling pathway
mmu04064 NF−kappa B signaling pathway
mmu04146 Peroxisome
mmu04612 Antigen processing and presentation
mmu04668 TNF signaling pathway
mmu04623 Cytosolic DNA−sensing pathway
mmu04922 Glucagon signaling pathway
mmu04630 Jak−STAT signaling pathway
mmu04659 Th17 cell differentiation
mmu04657 IL−17 signaling pathway
mmu04710 Circadian rhythm
mmu04662 B cell receptor signaling pathway
Enrichment score, row scaled
g
Cell cycle
GO:artery development
mTOR signaling pathway
GO:BMP signaling pathway
PC1 score
Ordered by pseudotime

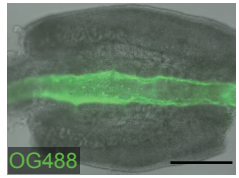
Fig. S1

a

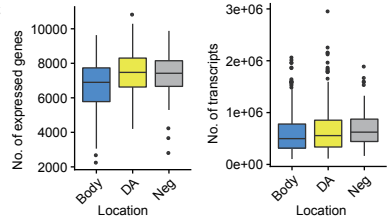
Stage	Embryos	Exp. (n)	Body	DA	Neg[a]	Total
E9.5 (25 sp)	4	2[b]	180	/	40	220
E10.0 (31-35 sp)	10	2[b]+2[c]	180	110	40	330
E10.5 (37-39 sp)	8	1[c]	/	56	/	56
E 11.0 (43-44 sp)	7	1[c]	/	56	/	56
Total	29	8	360	222	80	662

[a] $CD45^-CD31^-CD144^-$ negative control cells from body;
[b] Independent experiments for collecting $CD45^-CD31^+CD144^+$ sample cells from body;
[c] Independent experiments for collecting $CD45^-CD31^+CD144^+$ Oregon green$^+$ sample cells from AGM region.

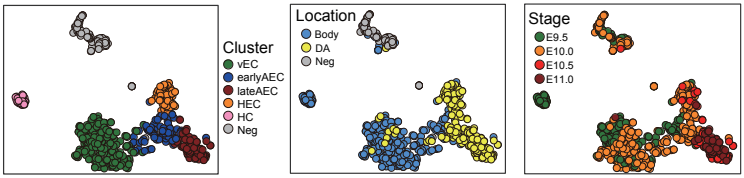
b



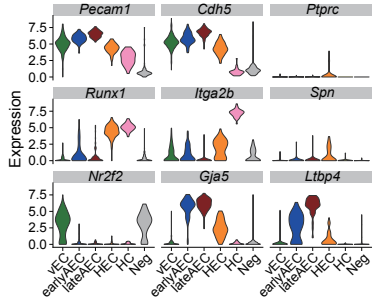
c



d



e

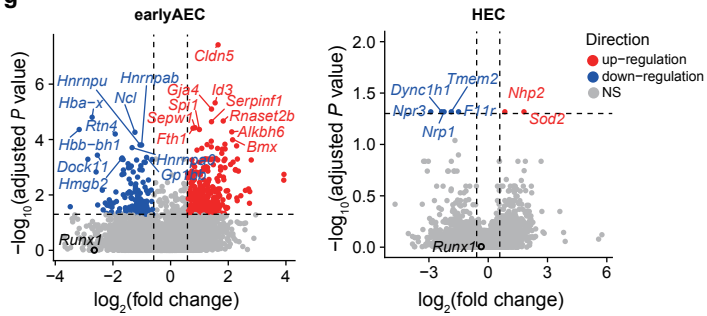


f

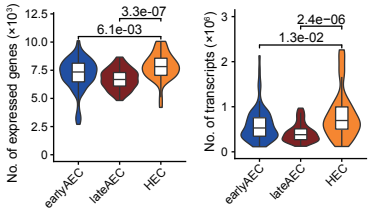
Location	Stage	vEC	earlyAEC	lateAEC	HEC	HC	Total
Body	E9.5	94	26	0	0	34	154
	E10.0	118	14	1	2	0	135
DA	E10.0	10	42	0	33	0	85
	E10.5	0	4	35	4	0	43
	E11.0	1	2	43	0	0	46
Total		223	88	79	39	34	463

Cells excluding Neg cluster, *Ptprc*- or *Spn*-expressing cells

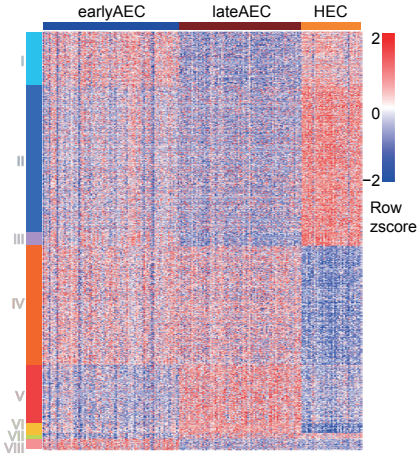
g



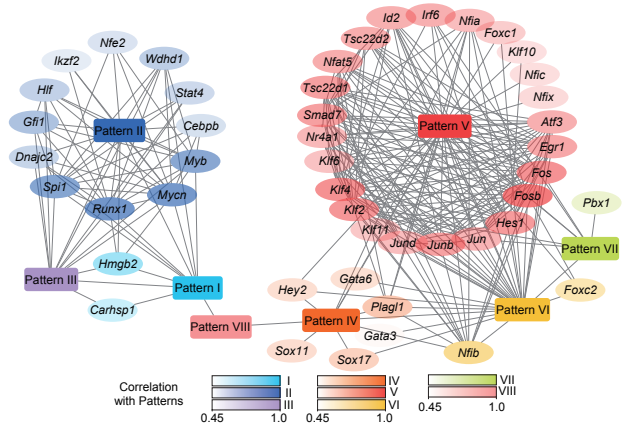
h



i



j



k

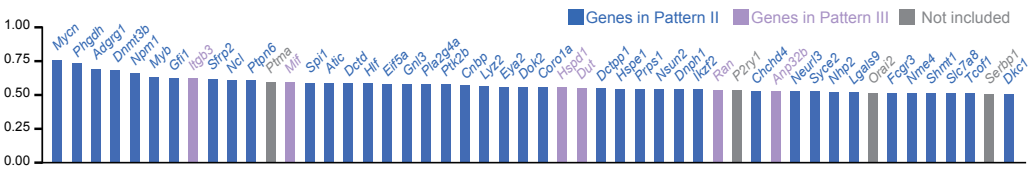


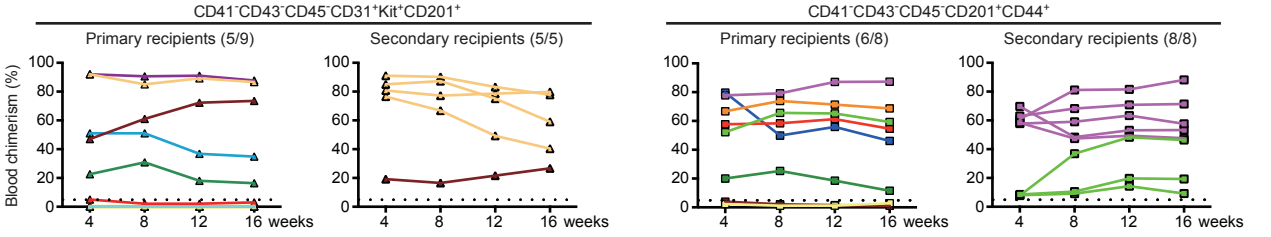
Fig. S2

a

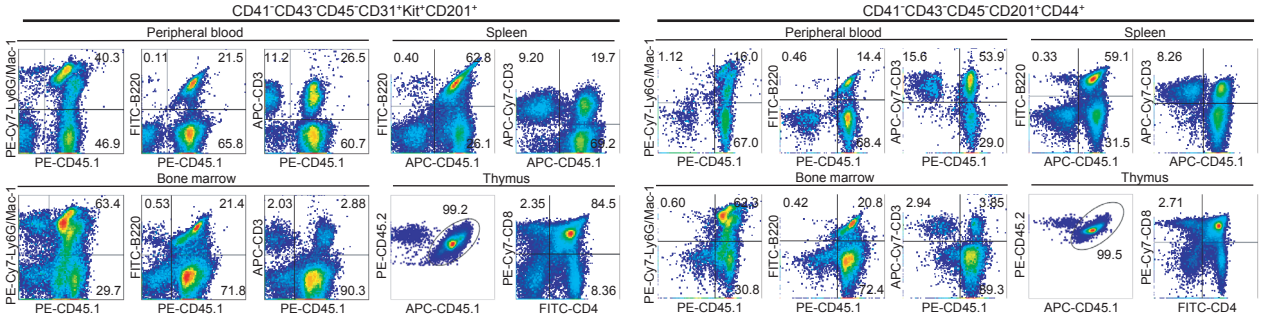
Stage	Organ	Exp (n)	Population	Constitution in the region (%)	Cell dose[a]	Clusters[b]
E9.5-E10.0 (30-32 sp[c])	Caudal half	4	$CD41^-CD43^-CD45^-CD31^+Kit^+CD201^+$	0.20±0.07	3.0-4.0 ee[d]	13/13 (100%)
			$CD41^-CD43^-CD45^-CD31^+Kit^+CD201^-$	1.25±0.45	3.0-4.0 ee	2/13 (15.4%)
E9.5-E10.0 (30-34 sp)	Caudal half	4	$CD41^-CD43^-CD45^-CD201^+CD44^+$	0.09±0.03	3.0-4.0 ee	11/11 (100%)
			$CD41^-CD43^-CD45^-CD201^+CD44^-$	0.33±0.21	3.0-4.0 ee	0/11

[a]cell dose per well and per recipient. [b]hematopoietic cluster positive wells/total wells. [c]somite pairs. [d]embryo equivalents.

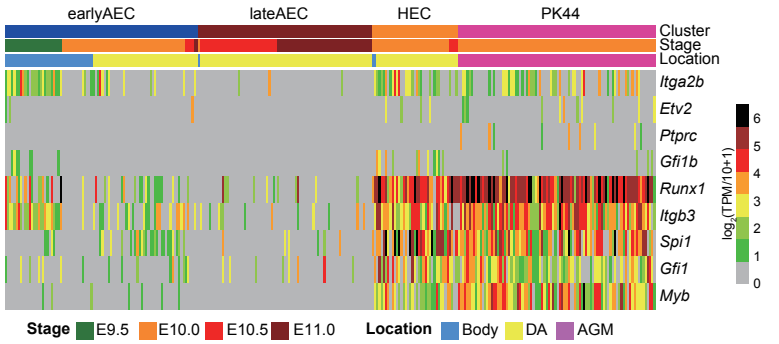
b



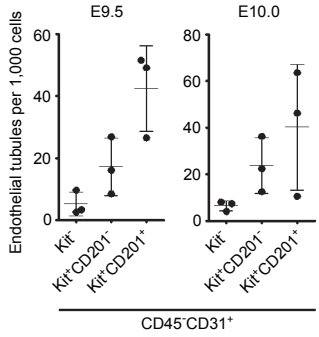
c



d



e



f

Stage	Region	Exp (n)	Population in EC ($CD41^-CD43^-CD45^-CD31^+$)	Constitution in the region (%)	Cell dose[a]	Bi-potential[b]	Hematopoietic[c]	Endothelial[d]
E9.5-E10.0 (29-31 sp[e])	Caudal half	2[f]	$Kit^+CD201^+CD44^+$	0.11±0.03	10 cells	/	44/44 (100%)	35/44 (79.5%)
			$Kit^+CD201^+CD44^-$	0.10±0.02	10 cells	/	0/42 (0%)	33/42 (78.5%)
E9.5-E10.0 (30-32 sp)	AGM	5[g]	$Kit^+CD201^+CD44^+$	0.31±0.09	Single cell	7/261 (2.7%)	113/261 (43.3%)	67/261 (25.7%)

[a]cells per well. [b]hematopoietic and endothelial bi-potential positive wells/total wells. [c]hematopoietic progeny positive wells/total wells.
[d]endothelial tube positive wells/total wells. [e]somite pairs. [f]2 independent experiments, totally 12 embryos were used.
[g]5 independent experiments, totally 15 embryos were used.

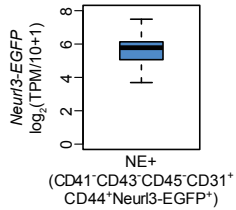
Fig. S3

a

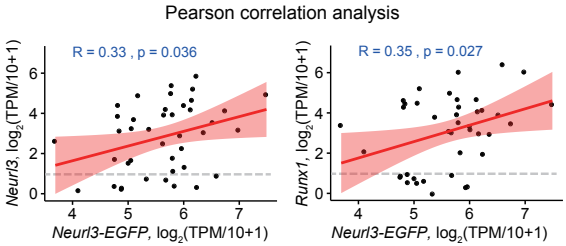
Stage	Region	Exp (n)	Population in EC ($CD41^-CD43^-CD45^-CD31^+$)	Cell dose[a]	Clusters[b]
E10.0 (30-34 sp[c])	Caudal half	5	$CD44^+$Neurl3-EGFP$^+$	1.5-2.5 ee[d]	6/6 (100%)
			$CD44^+$Neurl3-EGFP$^-$	1.5-5.0 ee	3/6 (50%)

[a]Cell dose per recipient. [b]hematopoietic cluster positive wells/total wells. [c]somite pairs. [d]embryo equivalents.

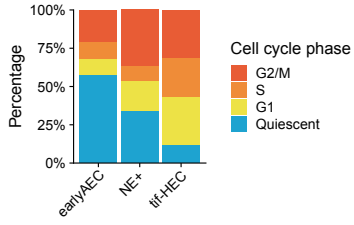
b



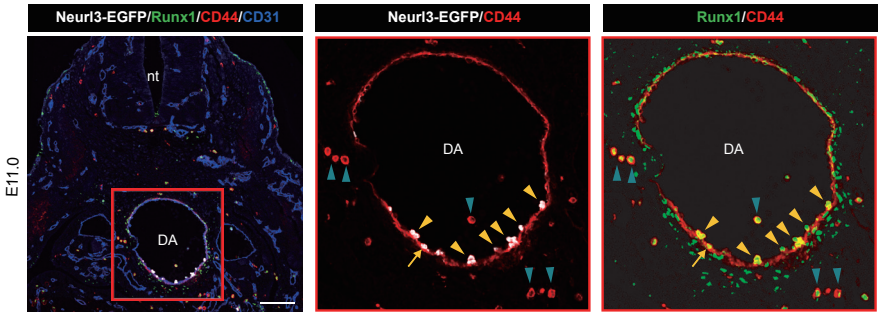
c



d



e



f

Single cell co-culture

Stage	Region	Exp (n)	Population in EC ($CD41^-CD43^-CD45^-CD31^+$)	Constitution in the region (%)	Constitution in EC (%)	Bi-potential[a]	Hematopoietic[b]	Endothelial[c]
E9.5 (26-30 sp[d])	Caudal half	8[e]	$CD44^+$Neurl3-EGFP$^+$	0.20±0.07	2.70±0.66	1/301 (0.3%)	26/301 (8.6%)	43/301 (14.3%)
E10.0 (31-35 sp)	AGM	5[f]	$CD44^+$Neurl3-EGFP$^+$	1.20±0.57	11.3±3.51	2/467 (0.4%)	39/467 (8.4%)	51/467 (10.9%)
E10.5 (36-39 sp)	AGM	6[g]	$CD44^+$Neurl3-EGFP$^+$	0.29±0.10	2.74±0.52	1/159 (0.6%)	6/159 (3.8%)	33/159 (20.8%)

[a]hematopoietic and endothelial bi-potential positive wells/total wells. [b]hematopoietic progeny positive wells/total wells. [c]endothelial tube positive wells/total wells. [d]somite pairs. [e]8 independent experiments, totally 32 embryos were used. [f]5 independent experiments, totally 37 embryos were used. [g]6 independent experiments, totally 23 embryos were used.

Fig. S4

a

Stage	Embryos	Exp. (n)	Sample[a]	Neg[b]	Total
E8.0 (3-5 sp)	4	1	40	8	48
E8.5 (8-9 sp)	14	3	236	50	286
E9.0 (17-18 sp)	6	2	205	40	245
Total	24	6	481	98	579

[a]CD45$^-$CD31$^+$CD144$^+$ sample cells from body
[b]CD45$^-$CD31$^-$CD144$^-$ negative control cells from body