

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

3

4 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
5 Wang³, Yunqiao Li³, Xiaochen Ding³, Zhilin Chang³, Shuaili Li³, Yuqiong Hu⁵, Xiaoying Fan⁵,
6 Yu Hou⁵, Lu Wen^{5,6}, Bing Liu^{1,2*}, Fuchou Tang^{5,6,7*}, Yu Lan^{1,8*}

7

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

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

12 ³State Key Laboratory of Proteomics, Academy of Military Medical Sciences, Academy of
13 Military Sciences, Beijing 100071, China

14 ⁴Integrated Chinese and Western Medicine Postdoctoral research station, Jinan University,
15 Guangzhou 510632, China

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

19 ⁶Ministry of Education Key Laboratory of Cell Proliferation and Differentiation, Beijing 100871,
20 China

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

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

23 Guangzhou 510530, China

24 ⁹These 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)

27 **ABSTRACT**

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

44 INTRODUCTION

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

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

67

68 In order to deeply investigate the cellular evolutions and molecular events underlying the
69 specification of HSC-primed HECs and their subsequent commitment to HSPCs, it is
70 necessary to efficiently isolate the HSC-primed HECs, which is proven to be difficult not only
71 because the population is proposed to be small and transient, but also due to the technical
72 challenging to determine their HSC competence ²⁰. Considering that not only HSCs but also
73 the transient definitive haematopoiesis during embryogenesis are derived from HECs,
74 repopulating capacity is required for the functional evaluation of the HSC-primed HECs.

75 Previous studies have reported the HSC competence of CD47⁺ but not CD47⁻ ECs in
76 embryonic day (E) 10.5 aorta-gonad-mesonephros (AGM) region and both Dll4⁺ and Dll4⁻
77 ECs in E9.5 para-aortic splanchnopleura (P-Sp) region ^{11,21}. Nevertheless, the enrichment of
78 the above surface markers is far from efficient. Several transgenic reporter mouse models
79 have been established by which HECs could be distinguished from non-HECs, including
80 *Ly6a-GFP*, GFP transgenic reporter under the control of Runx1 +23 enhancer (*Runx1*
81 +23*GFP*) and *Gfi1-Tomato*, and the usage of these reporters largely helps to delineate the
82 process of endothelial-to-haematopoietic transition ^{3,14,22-25}. Although expected to a certain
83 extent, the HSC competence of the HECs labeled by these reporters has not been
84 functionally validated. Up-to-date, efficient isolation of the HSC-primed HEC population has
85 not yet been achieved.

86

87 With the aim of delineating the molecular events underlying HSC emergence, several
88 single-cell transcriptional profiling studies on HECs, IAHC cells, and HSPCs in the AGM
89 region have been reported in recent years. Using either Runx1 +23GFP or Gfi1-Tomato as
90 the marker of putative HECs, several defined cell populations are transcriptionally profiled by
91 Fluidigm single-cell qPCR or single-cell RNA sequencing (scRNA-seq)^{3,14,22}. Moreover, the
92 cellular components of IAHCs are investigated at single-cell level by mechanically picking up
93 single whole IAHCs in the aortas, showing cells with pre-HSC feature are predominantly
94 involved¹⁴. Interestingly, contradiction still exists regarding whether HECs and non-HECs are
95 molecularly similar and to what extent the two populations are distinguishable^{14,22}. Since the
96 enrichment efficiency or specificity of the above markers to define the HEC population might
97 be not enough, an unsupervised screening of the embryonic endothelial pool within
98 haematopoietic tissues is required for the precise recognition of HSC-primed HECs.

99

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

109 unprecedently enrich our understanding of HSC generation *in vivo* and should be
110 extremely critical to inspire new approaches for stepwise HSC regeneration from pluripotent
111 stem cells²⁶.

112

113 **RESULTS**

114 **Transcriptomic identification of the HECs in AGM region**

115 We first analyzed mouse embryos from E9.5, when the initial IAHC formation in the aorta
116 occurs⁷, to the stage of the appearance of HSCs at E11.0²⁷ ([Supplementary information, Fig. S1a](#)). For each embryo, the embryo proper was isolated and the head, limb buds, heart,
117 visceral bud, and vitelline and umbilical vessels outside the embryo proper were excluded
118 ([Fig. 1a](#)). To specifically capture aortic luminal ECs of AGM region, we performed
119 microinjection of fluorescent dye Oregon green into the dorsal aortas of E10.0-E11.0 embryos
120 as reported¹² ([Fig. 1a; Supplementary information, Fig. S1b](#)). The sampled cells were
121 purified by FACS as CD45⁻CD31⁺CD144⁺, which contained predominantly vascular ECs and
122 CD41⁺ haematopoietic cells. Meanwhile, CD45⁻CD31⁻CD144⁻ non-EC cells in the body were
123 used as negative controls ([Fig. 1a](#)). We used unique molecular identifier (UMI)-based
124 scRNA-seq method to accurately measure the gene expression profiles within individual cells.
125 In totally 662 sequenced single cells, 597 single-cell transcriptomes passed rigorous quality
126 control. On average we detected 7,035 genes (from 2,266 to 10,843) and 636,418 transcripts
127 (from 103,793 to 2,959,573) expressed in each individual cell ([Supplementary information, Fig. S1c](#)).

130

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

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

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

194

195 **Efficient capture of the HSC-competent and endothelial-haematopoietic bi-potent**

196 **HECs in AGM region**

197 We next made an effort to identify surface marker combination to highly enrich the HECs for
198 functional evaluation ([Supplementary information, Table S4](#)). *Cd44*, *Procr* (coding CD201)
199 and *Kit* were screened out by differentially expressed genes and correlation analysis ([Fig. 2a](#)).
200 We specifically focused on E10.0 in the following functional assays to keep consistent with
201 the transcriptomic finding. Whole-mount immunostaining showed that in addition to the
202 scattered blood cells throughout the tissue, the expression of CD44 was detected in the
203 whole endothelial layer of dorsal aorta and very proximal part of its segmental branches ([Fig.](#)
204 [2b](#)). Using similar strategy as for pre-HSC identification ¹¹, we found only the derivatives
205 induced from CD41⁻CD43⁻CD45⁻CD31⁺Kit⁺CD201⁺ rather than
206 CD41⁻CD43⁻CD45⁻CD31⁺Kit⁺CD201⁻ population at E9.5-E10.0 could long-term (16 weeks)
207 and multi-lineage reconstitute lethally irradiated adult recipients, although both populations
208 generated haematopoietic clusters with different frequencies upon 7 days culture on
209 OP9-DL1 stromal cells ([Fig. 2c-e; Supplementary information, Fig. S2a-c](#)). Self-renewal
210 capacity of the HSCs was further validated by secondary transplantation ([Fig. 2d, e;](#)
211 [Supplementary information, Fig. S2b](#)). Within CD41⁻CD43⁻CD45⁻CD201⁺ population, induced
212 HSC potential was exclusively detected in CD44⁺ subpopulation ([Fig. 2c-e; Supplementary](#)
213 [information, Fig. S2a-c](#)). Thus, our data identified CD41⁻CD43⁻CD45⁻CD31⁺CD201⁺Kit⁺CD44⁺
214 (PK44) population in E10.0 caudal half as the HSC-competent HECs.

215
216 As compared to other endothelial surface markers, Flk1 (encoded by *Kdr*) is known to be
217 specifically localized within the vessel lumen layer, and very few and only the basal-most
218 localized IAHC cells express Flk1 ⁷. Here, almost all PK44 cells expressed Flk1 by FACS

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

228

229 We next explored whether endothelial-haematopoietic bi-potential existed in these
230 HSC-competent HECs, since that if a cell population is experiencing fate choice so that the
231 transient intermediate state might be captured. Firstly, we found that the Kit⁺CD201⁺ ECs in
232 the body part of embryo proper at E9.5-E10.0 had a relatively higher endothelial tube-forming
233 capacity as compared to the Kit⁻ or Kit⁺CD201⁻ endothelial populations (Supplementary
234 information, Fig. S2e). Furthermore, CD44⁺ and CD44⁻ fractions within Kit⁺CD201⁺ ECs
235 showed comparable endothelial tube-forming capacity whereas the generation of
236 haematopoietic cells in the cultures was exclusively detected in the CD44⁺ ones under the
237 bi-potential induction system (Supplementary information, Fig. S2f). The data suggested the
238 largely maintenance of endothelial potential in the HECs. By single-cell in vitro induction, 40.6%
239 (106/261) PK44 cells gave rise to only haematopoietic progenies and 23.0% (60/261) only
240 endothelial tubules. Remarkably, 2.7% (7/261) had both haematopoietic and endothelial

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

248

249 **Transcriptional and functional relationship between HECs and T1 pre-HSCs**

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

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

273
274 The developmental path from tif-HEC to T1 pre-HSC was inferred by Mpath trajectory
275 analysis (Fig. 3g). Consistently, during the course of in vitro culture of the PK44 population
276 from E10.0 AGM region on OP9-DL1 stromal cells to induce its HSC activity, we could
277 witness the generation of immunophenotypic T1 pre-HSCs (Fig. 3h). We also evaluated the
278 endothelial and haematopoietic potentials of the T1 pre-HSCs
279 ($CD31^+CD45^-CD41^{low}Kit^+CD201^{high}$) in E11.0 AGM region at single-cell level. Surprisingly, we
280 found that although displayed largely decreased endothelial potential as compared to E10.0
281 PK44 cells (Fig. 2i), T1 pre-HSCs still maintained comparable endothelial-haematopoietic
282 bi-potential as that in PK44 population (Fig. 3i). This finding implied that the extremely rare
283 and enriched T1 pre-HSC population has not completely fulfilled the
284 endothelial-to-haematopoietic fate transition.

285

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

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

297

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

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 *Neru13* 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

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

350 Given the lacking of suitable antibodies to directly determine the anatomical distribution of
351 PK44 cells, which have been proven as HSC-competent HECs (Fig. 2c-e), we compared the
352 immunophenotype of PK44 and IAHC cells, known as CD31⁺Kit^{high}⁷, regarding their
353 relationship with Neurl3-EGFP expression whose localization was clearly defined (Fig. 5a).
354 Both of them were mainly Neurl3-EGFP⁺, with most CD31⁺Kit^{high} cells being
355 CD41/CD43/CD45-positive haematopoietic cells as previously reported⁷ (Fig. 5b). PK44
356 showed an expression pattern largely different from CD31⁺Kit^{high} cells, suggestive of their
357 predominant non-IAHC localization (Fig. 5b). The expression of Neurl3-EGFP was
358 completely absent from the sub-aortic mesenchyme, in contrast to the widespread
359 distribution of Runx1 there (Fig. 5a; Supplementary information, Fig. S3e). Although scattered
360 Runx1⁺CD44⁺ round blood cells were easily witnessed, much fewer Neurl3-EGFP-expressed
361 cells outside dorsal aorta were detected, even at E11.0 (Fig. 5a; Supplementary information,
362 Fig. S3e).

363

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

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

379

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

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

391

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

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

403

404 We identified totally 2,851 genes whose expression levels were changed significantly among
405 five clusters, which were further grouped into five principal expression patterns along the
406 inferred pseudotime (Fig. 6d; Supplementary information, Fig. S4e and Table S6). In
407 general, genes in Pattern 1 showed the highest expression in pEC, decreased apparently
408 upon arterial specification, whereas slightly increased upon hemogenic specification, which
409 were mainly related to rRNA processing and mitotic nuclear division (Fig. 6d; Supplementary
410 information, Fig. S4e, f). Genes in Pattern 2 showed the highest expression in the initial
411 arterial specification, and the lowest expression upon hemogenic specification, which were
412 mainly related to organization of intra-cellular actin filament and inter-cellular junctions (Fig.
413 6d; Supplementary information, Fig. S4e, f). Genes in Pattern 3, which were related to
414 endothelium development and cell migration, showed the highest level in earlyAEC, whereas
415 relatively low expression in the upstream pEC and pAEC and downstream HEC and T1

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

421

422 Among the above 2,851 pattern genes, 75 TFs belonged to the core TFs of the regulons
423 where the genes included significantly overlapped with the pattern genes (Fig. 6e;
424 [Supplementary information](#), Table S6). Given the simultaneous co-expression of the core TF
425 and its predicted targets in a given regulon, these core TFs were considered to presumably
426 play a role to drive or orchestrate the dynamic molecular program during HEC specification
427 (Fig. 6e). Most of these TFs belonged to Pattern 4 and Pattern 5, indicating that most
428 activated TFs along HEC specification from primitive vascular ECs were those
429 overrepresented in the final hemogenic and haematopoietic populations (Fig. 6e). We also
430 examined the expression patterns of totally 28 TFs previously reported to have a role in
431 HSPC regeneration in vitro³⁴⁻³⁷. 19 of these presumed functional TF were dynamically
432 changed and 15 of them were core TFs of regulons, with most of them belonging to Pattern 5
433 ([Supplementary information](#), Fig. S4g).

434

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

438 ribosome and spliceosome were the pathways that were down-regulated with arterial
439 specification whereas turned to be moderately up-regulated by hemogenic specification (Fig.
440 6f, g). In contrast, several pathways experienced a completely opposite change, such as
441 Rap1 signaling pathway (Fig. 6f). Artery development, together with its pivotal executor Notch
442 signaling pathway^{2,38}, firstly rose to peak in earlyAEC and then modestly fell down upon
443 hemogenic specification (Fig. 6f, g). Some inflammation related pathways, including NF κ B
444 and TNF signaling, were activated from earlyAEC to the final T1 pre-HSC, in line with the
445 notion about the requirement of inflammatory signaling during HSC generation³⁹ (Fig. 6f).

446

447 DISCUSSION

448 Here via unbiasedly going through all the relevant EC populations, HSC-primed HECs were
449 transcriptomically identified. More importantly, combining the computational prediction and in
450 vivo functional evaluation, we precisely captured the HSC-competent HECs by a newly
451 constructed fluorescent reporter mouse model, Neurl3-EGFP, and revealed further
452 functionally enriched sub-population within Neurl3-EGFP-labeled ECs by a set of surface
453 marker combination PK44. Serving as the putative marker of HSC-primed HECs^{14,22}, *Gfi1*
454 was specifically expressed in HEC but not other EC-related populations (Fig. 2h, 4h),
455 supportive of the cluster assignment. Belonging to the gene family of E3 ubiquitin ligases, the
456 expression and role of Neurl3 in spermatogenesis and inflammation has been reported⁴⁰⁻⁴²,
457 whereas that relevant to vascular and haematopoietic development remains barely known.
458 Neurl3 was screened out by unsupervised bioinformatics analysis, and fortunately, the
459 expression of which in AGM region was restricted to aorta and largely consistent with that of

460 *Runx1* both transcriptomically ([Supplementary information, Fig. S1k](#)) and anatomically ([Fig. 5a](#)) regarding endothelial expression. Although highly expressed in tif-HEC, *Runx1* and *Adgrg1* were also highly expressed in the CD45⁻CD43⁻ haematopoietic population ([Fig. 4h](#)), which should be the derivatives of non-HSC haematopoiesis. This suggested that they may not distinguish the precursors of HSCs and non-HSCs^{32,43}, thus *Runx1* and *Adgrg1* were not included in the list of the signature genes of HSC-primed HEC ([Fig. 4a](#)). The specificity of *Neru13* expression related to HSC generation suggests that the *Neru13*-EGFP would be a 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 HECs³, which might partially explain why the rare endothelial-haematopoietic bi-potential is
477 hardly detected around the timing of HSC emergence in previous report³. 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

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

489

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

504 precursors¹⁸. Our findings further emphasize that arterial specification and Notch signaling
505 should be precisely and stepwise controlled for HSC generation. Although both showing
506 obvious similarity regarding the arterial feature and anatomical distribution, the difference
507 between earlyAEC and lateAEC should also be paid attention to as the former but not the
508 latter is the direct origin of the HSC-primed HECs.

509

510 It is generally accepted that haematopoietic cells in the IAHCs are proliferative^{48,49}, within
511 which pre-HSCs are mainly involved^{14,49}. Supportively, enriched functional T1 pre-HSCs
512 manifested a relatively proliferative status¹¹. On the other hand, slow cycling is witnessed at
513 the base of IAHCs⁴⁹, and it is suggested that exit from cell cycle is necessary for HEC
514 development and endothelial-to-haematopoietic transition^{44,50}. Nevertheless, based on the
515 precise recognition of the HSC-primed HECs here, we showed that proliferation was
516 gradually decreased upon arterial specification and maturation, whereas re-activated once
517 the arterial ECs chose a hemogenic fate featured by the simultaneous *Runx1* expression (Fig.
518 6e-g). The functional requirement of cell cycle control for the specification of the HSC-primed
519 HECs needs to be investigated, which would depend on the initiating cell populations.

520

521 We also revealed several similarities regarding the molecular events underlying the
522 development of HSC-primed HECs between in mouse and human embryos we have reported
523 very recently¹⁹, including the arterial feature and the overrepresented ribosome and
524 translational activity in the HSC-primed HECs. Such conservation further assures the mouse
525 model as an adequate animal model for HSC development studies. The comprehensive

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.

530 **MATERIALS AND METHODS**

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

534

535 **Mice**

536 Mice were handled at the Laboratory Animal Center of Academy of Military Medical Sciences
537 in accordance with institutional guidelines. Mouse manipulations were approved by the
538 Animal Care and Use Committee of the Institute. The *Neurl3^{EGFP/+}* reporter mouse lines were
539 generated with the CRISPR/Cas9 technique by Beijing Biocytogen. All mice were maintained
540 on C57BL/6 background. Embryos were staged by somite pair (sp) counting: E8.0, 1-7 sp;
541 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,
542 41-45 sp. In some experiments, caudal half of E10.0 embryo was dissected under heart with
543 limbs removed. AGM region was dissected as reported ¹². The fluorescent dye Oregon green
544 488 was purchased from Invitrogen. Staining was performed as previously described ¹²
545 except that the concentration of staining solution was 5 μ mol/L and the time of staining was 3
546 minutes before washed. Primary embryonic single-cell suspension was acquired by type I
547 collagenase digestion.

548

549 **Flow cytometry**

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

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 cells⁵¹ 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

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 $\geq 5\%$ 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

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 μ m 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

618 methanol/PBS (50%, 100%; 10 minutes each). To block endogenous peroxidase, samples
619 were bleached in 5% H₂O₂ 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°C,
621 incubated with PBSMT containing anti-CD44 (1:25) overnight at 4°C, then washed 3 times in
622 PBSMT each for 1 hour at 4°C. 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°C. 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°C, the samples were soaked in
626 DendronFluor TSA (Histova, NEON 4-color IHC Kit for Wholomount/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°C 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 (TCAGACGTGTGCTCTCCGATCT-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

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

650

651 **Quantification of gene expression for scRNA-seq data**

652 We used unique molecular identifier (UMI)-based scRNA-seq method to measure the gene
653 expression profiles within individual cells. Raw reads were firstly split by specific barcode
654 attached in Read 2 for individual cells and UMI information was aligned to the corresponding
655 Read 1. Read 1 was trimmed to remove the template switch oligo (TSO) sequence and polyA
656 tail sequence. Subsequently, quality control was conducted to discard reads with adapter
657 contaminants or low-quality bases (N > 10%). Next, the mm10 mouse transcriptome (UCSC)
658 was used to align the clean reads using TopHat (version 2.0.12)⁵⁹. Uniquely mapped reads
659 were obtained using HTSeq package⁶⁰ and grouped by the cell-specific barcodes.
660 Transcripts of each gene were deduplicated based on the UMI information, while
661 mitochondrial genes were not included for quantification. Finally, for each gene in each

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

664

665 **Quality control and normalization of sequencing data**

666 For the 662 sequenced single cells from E9.5-E11.0 embryos of totally 29 embryos, we only
667 retained cells with more than 2,000 genes and 100,000 transcripts detected. Then, 597 cells
668 passed the filter standards. Gene expression levels in each cell were normalized by
669 $\log_2(\text{TPM}/10+1)$, where TPM (transcripts-per-million) was calculated as (the number of UMIs
670 of each gene / all UMIs of a given cell) $\times 1,000,000$. Since the UMI number of most of our
671 samples was less than the order of 1,000,000 transcripts, the TPM values were divided by 10
672 to avoid counting each transcript for several times. On average we detected 7,035 genes
673 (range from 2,266 to 10,843) and 636,418 transcripts (range from 103,793 to 2,959,573)
674 expressed in each individual cell.

675

676 Additionally, we also sequenced 96 single cells with a PK44 immunophenotype
677 (CD41⁻CD43⁻CD45⁻CD31⁺CD201⁺Kit⁺CD44⁺) from E10.0 AGM regions of totally 9 embryos,
678 47 T1 pre-HSCs (CD31⁺CD45⁻CD41^{low}Kit⁺CD201^{high}) from E11.0 AGM regions of totally 18
679 embryos, 48 single cells with an immunophenotype of
680 CD41⁻CD43⁻CD45⁻CD31⁺CD44⁺Neurl3-EGFP⁺ from Neurl3-EGFP reporter mouse embryos
681 and 579 single cells from E8.0-E9.0 body regions of totally 24 embryos. The same quality
682 control criteria and normalization method described above were applied to these additional
683 datasets. In total, 1,432 single cells were sequenced and 1,325 cells passed the filter

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,

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

728 populations were considered. Genes with fold-change ≥ 2 and adjusted P value ≤ 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 *Apnlr*)^{33,62-64}, were selected
735 to perform the arteriovenous feature scores. First, we scaled the $\log_2(\text{TPM}/10+1)$ 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

750 the average expression of each gene set as corresponding scores, and manually assigned
751 cells to approximate cell cycle phases based on the scores. Namely, cells with G1/S score <
752 2 and G2/M score < 2 were assigned as 'quiescent', otherwise 'proliferative'. Among
753 proliferative cells, those with G2/M score > G1/S score were assigned as 'G2/M', and those
754 with G1/S score > G2/M score were assigned as 'G1' when G2/M score < 2, or as 'S' when
755 G2/M score ≥ 2 .

756

757 **Constructing single cell trajectories**

758 Monocle 2⁶⁷ (version 2.6.4) and Mpath⁶⁸ (version 1.0) were adopted to infer the development
759 trajectory of selected cell populations. Monocle 2 can construct single-cell trajectories and
760 place each cell at its proper position in the trajectory, even a "branched" trajectory
761 corresponding to cellular "decisions". We followed the official vignette with recommended
762 parameters. Briefly, UMI count data of given cell populations was used as input and genes
763 with more than 1.5 times of fitted dispersion evaluated using dispersionTable function were
764 identified as HVGs. To reduce the influence of cell cycle effect, HVGs not included in the
765 direct cell cycle GO term (GO:0007049) were retained as ordering genes for the subsequent
766 ordering cells.

767

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

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

773

774 **Patterns of DEGs among multiple clusters**

775 In the case of identification of gene patterns in more than two clusters, analysis of variance

776 followed by Tukey's HSD test for pairwise comparison was adopted to identify DEGs (genes

777 with adjusted *P* value < 0.05 and fold change > 2 or < 0.5). For identification of patterns in

778 earlyAEC, lateAEC and HEC clusters, only 1,005 DEGs resulted from the pairwise

779 comparisons of earlyAEC and lateAEC and of earlyAEC and HEC were retained. According

780 to the changed directions of HEC and lateAEC as compared to earlyAEC, we could assign

781 these DEGs into 8 patterns as illustrated. Transcription factors network visualization was

782 implemented as follows. First, the average expression values of genes included in each

783 pattern were calculated as their representative expression levels. Then, the representative

784 expression levels of 8 patterns and the expression profile data of transcription factors

785 included in these patterns were combined as input for construction of "signed hybrid"

786 weighted gene co-expression network analysis using WGCNA⁶⁹. Next, we used 0.01 as

787 adjacency threshold for including edges in the output to export network, which was then

788 imported into Cytoscape⁷⁰ for visualization. We also calculated Pearson correlation coefficient

789 between each transcription factor and the pattern it belongs to.

790

791 For identification of patterns in pEC, pAEC, earlyAEC and HEC clusters, all 2,851 DEGs

792 among them were retained. We used ConsensusClusterPlus function with k-means algorithm

793 on top 500 DEGs to achieve five stable clusters. Then all DEGs were reassigned into one of

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 ($\log_2(\text{TPM}/10+1) < 2$) in HEC or
810 highly expressed ($\log_2(\text{TPM}/10+1) > 2$) in every clusters were also excluded. Finally, eleven
811 HEC overrepresented genes were retained as HEC signature genes.

812

813 **SCENIC analysis**

814 SCENIC⁷³ 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

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**

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.

848 **REFERENCES**

849 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).

850 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).

851 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).

852 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).

853 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).

854 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).

855 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).

856 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).

857 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).

858 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).

859 11 Zhou, F. *et al.* Tracing haematopoietic stem cell formation at single-cell resolution. *Nature* **533**, 487-492, doi:10.1038/nature17997 (2016).

860 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).

861 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).

862 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).

863 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).

864 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).

865 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).

866 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).

867 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).

868 20 Gritz, E. & Hirschi, K. K. Specification and function of hemogenic endothelium during embryogenesis.

891 1547-1567, doi:10.1007/s00018-016-2134-0 (2016).

892 21 Hadland, B. K. *et al.* A Common Origin for B-1a and B-2 Lymphocytes in Clonal Pre- Hematopoietic
893 Stem Cells. *Stem Cell Reports* **8**, 1563-1572, doi:10.1016/j.stemcr.2017.04.007 (2017).

894 22 Thambyrajah, R. *et al.* GFI1 proteins orchestrate the emergence of haematopoietic stem cells through
895 recruitment of LSD1. *Nat Cell Biol* **18**, 21-32, doi:10.1038/ncb3276 (2016).

896 23 Solaimani Kartalaei, P. *et al.* Whole-transcriptome analysis of endothelial to hematopoietic stem cell
897 transition reveals a requirement for Gpr56 in HSC generation. *J Exp Med* **212**, 93-106,
898 doi:10.1084/jem.20140767 (2015).

899 24 Bee, T. *et al.* Nonredundant roles for Runx1 alternative promoters reflect their activity at discrete
900 stages of developmental hematopoiesis. *Blood* **115**, 3042-3050, doi:10.1182/blood-2009-08-238626
901 (2010).

902 25 de Bruijn, M. F. *et al.* Hematopoietic stem cells localize to the endothelial cell layer in the midgestation
903 mouse aorta. *Immunity* **16**, 673-683, doi:10.1016/s1074-7613(02)00313-8 (2002).

904 26 Guibentif, C. & Gottgens, B. Blood: Education for stem cells. *Nature* **545**, 415-417,
905 doi:10.1038/nature22496 (2017).

906 27 Dzierzak, E. & Speck, N. A. Of lineage and legacy: the development of mammalian hematopoietic stem
907 cells. *Nat Immunol* **9**, 129-136, doi:10.1038/ni1560 (2008).

908 28 Levine, J. H. *et al.* Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that
909 Correlate with Prognosis. *Cell* **162**, 184-197, doi:10.1016/j.cell.2015.05.047 (2015).

910 29 Zilberberg, L. *et al.* Specificity of latent TGF-beta binding protein (LTBP) incorporation into matrix: role
911 of fibrillins and fibronectin. *J Cell Physiol* **227**, 3828-3836, doi:10.1002/jcp.24094 (2012).

912 30 Bos, F. L., Hawkins, J. S. & Zovein, A. C. Single-cell resolution of morphological changes in hemogenic
913 endothelium. *Development* **142**, 2719-2724, doi:10.1242/dev.121350 (2015).

914 31 Jeanne, M., Jorgensen, J. & Gould, D. B. Molecular and Genetic Analyses of Collagen Type IV Mutant
915 Mouse Models of Spontaneous Intracerebral Hemorrhage Identify Mechanisms for Stroke Prevention.
916 *Circulation* **131**, 1555-1565, doi:10.1161/CIRCULATIONAHA.114.013395 (2015).

917 32 Tober, J., Yzaguirre, A. D., Piwarzyk, E. & Speck, N. A. Distinct temporal requirements for Runx1 in
918 hematopoietic progenitors and stem cells. *Development* **140**, 3765-3776, doi:10.1242/dev.094961
919 (2013).

920 33 Chong, D. C., Koo, Y., Xu, K., Fu, S. & Cleaver, O. Stepwise arteriovenous fate acquisition during
921 mammalian vasculogenesis. *Dev Dyn* **240**, 2153-2165, doi:10.1002/dvdy.22706 (2011).

922 34 Sugimura, R. *et al.* Haematopoietic stem and progenitor cells from human pluripotent stem cells.
923 *Nature* **545**, 432-438, doi:10.1038/nature22370 (2017).

924 35 Lis, R. *et al.* Conversion of adult endothelium to immunocompetent haematopoietic stem cells. *Nature*
925 **545**, 439-445, doi:10.1038/nature22326 (2017).

926 36 Wahlster, L. & Daley, G. Q. Progress towards generation of human haematopoietic stem cells. *Nat Cell
927 Biol* **18**, 1111-1117, doi:10.1038/ncb3419 (2016).

928 37 Vo, L. T. & Daley, G. Q. De novo generation of HSCs from somatic and pluripotent stem cell sources.
929 *Blood* **125**, 2641-2648, doi:10.1182/blood-2014-10-570234 (2015).

930 38 Gridley, T. Notch signaling in vascular development and physiology. *Development* **134**, 2709-2718,
931 doi:10.1242/dev.004184 (2007).

932 39 Li, Y. *et al.* Inflammatory signaling regulates embryonic hematopoietic stem and progenitor cell
933 production. *Genes Dev* **28**, 2597-2612, doi:10.1101/gad.253302.114 (2014).

934 40 Zhao, Y. *et al.* Neuralized E3 Ubiquitin Protein Ligase 3 Is an Inducible Antiviral Effector That Inhibits

935 Hepatitis C Virus Assembly by Targeting Viral E1 Glycoprotein. *J Virol* **92**, doi:10.1128/JVI.01123-18
936 (2018).

937 41 Xu, W. *et al.* Ubiquitin ligase gene *neurl3* plays a role in spermatogenesis of half-smooth tongue sole
938 (Cynoglossus semilaevis) by regulating testis protein ubiquitination. *Gene* **592**, 215-220,
939 doi:10.1016/j.gene.2016.07.062 (2016).

940 42 Hu, Y., Nguyen, T. T., Bui, K. C., Demello, D. E. & Smith, J. B. A novel inflammation-induced ubiquitin E3
941 ligase in alveolar type II cells. *Biochem Biophys Res Commun* **333**, 253-263,
942 doi:10.1016/j.bbrc.2005.05.102 (2005).

943 43 Frame, J. M., Fegan, K. H., Conway, S. J., McGrath, K. E. & Palis, J. Definitive Hematopoiesis in the Yolk
944 Sac Emerges from Wnt-Responsive Hemogenic Endothelium Independently of Circulation and Arterial
945 Identity. *Stem Cells* **34**, 431-444, doi:10.1002/stem.2213 (2016).

946 44 Marcelo, K. L. *et al.* Hemogenic endothelial cell specification requires c-Kit, Notch signaling, and
947 p27-mediated cell-cycle control. *Dev Cell* **27**, 504-515, doi:10.1016/j.devcel.2013.11.004 (2013).

948 45 Uenishi, G. I. *et al.* NOTCH signaling specifies arterial-type definitive hemogenic endothelium from
949 human pluripotent stem cells. *Nat Commun* **9**, 1828, doi:10.1038/s41467-018-04134-7 (2018).

950 46 Park, M. A. *et al.* Activation of the Arterial Program Drives Development of Definitive Hemogenic
951 Endothelium with Lymphoid Potential. *Cell Rep* **23**, 2467-2481, doi:10.1016/j.celrep.2018.04.092
952 (2018).

953 47 Lizama, C. O. *et al.* Repression of arterial genes in hemogenic endothelium is sufficient for
954 haematopoietic fate acquisition. *Nat Commun* **6**, 7739, doi:10.1038/ncomms8739 (2015).

955 48 Zape, J. P., Lizama, C. O., Cautivo, K. M. & Zovein, A. C. Cell cycle dynamics and complement expression
956 distinguishes mature haematopoietic subsets arising from hemogenic endothelium. *Cell Cycle* **16**,
957 1835-1847, doi:10.1080/15384101.2017.1361569 (2017).

958 49 Batsivari, A. *et al.* Understanding Hematopoietic Stem Cell Development through Functional
959 Correlation of Their Proliferative Status with the Intra-aortic Cluster Architecture. *Stem Cell Reports* **8**,
960 1549-1562, doi:10.1016/j.stemcr.2017.04.003 (2017).

961 50 Ottersbach, K. Endothelial-to-haematopoietic transition: an update on the process of making blood.
962 *Biochem Soc Trans* **47**, 591-601, doi:10.1042/BST20180320 (2019).

963 51 Lan, Y. *et al.* Endothelial Smad4 restrains the transition to hematopoietic progenitors via suppression
964 of ERK activation. *Blood* **123**, 2161-2171, doi:10.1182/blood-2013-09-526053 (2014).

965 52 Li, Z. *et al.* Mouse embryonic head as a site for hematopoietic stem cell development. *Cell Stem Cell* **11**,
966 663-675, doi:10.1016/j.stem.2012.07.004 (2012).

967 53 Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods*
968 **10**, 1096-1098, doi:10.1038/nmeth.2639 (2013).

969 54 Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181,
970 doi:10.1038/nprot.2014.006 (2014).

971 55 Li, L. *et al.* Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal
972 Niche Interactions. *Cell Stem Cell* **20**, 858-873 e854, doi:10.1016/j.stem.2017.03.007 (2017).

973 56 Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear
974 amplification. *Cell Rep* **2**, 666-673, doi:10.1016/j.celrep.2012.08.003 (2012).

975 57 Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods* **11**,
976 163-166, doi:10.1038/nmeth.2772 (2014).

977 58 Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using
978 Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).

979 59 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq.
980 *Bioinformatics* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).

981 60 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput
982 sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).

983 61 Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene
984 expression data. *Nat Biotechnol* **33**, 495-502, doi:10.1038/nbt.3192 (2015).

985 62 Corada, M., Morini, M. F. & Dejana, E. Signaling pathways in the specification of arteries and veins.
986 *Arterioscler Thromb Vasc Biol* **34**, 2372-2377, doi:10.1161/ATVBAHA.114.303218 (2014).

987 63 Simons, M. & Eichmann, A. Molecular controls of arterial morphogenesis. *Circ Res* **116**, 1712-1724,
988 doi:10.1161/CIRCRESAHA.116.302953 (2015).

989 64 Su, T. *et al.* Single-cell analysis of early progenitor cells that build coronary arteries. *Nature* **559**,
990 356-362, doi:10.1038/s41586-018-0288-7 (2018).

991 65 Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq.
992 *Science* **352**, 189-196, doi:10.1126/science.aad0501 (2016).

993 66 Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendrogloma.
994 *Nature* **539**, 309-313, doi:10.1038/nature20123 (2016).

995 67 Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**,
996 979-982, doi:10.1038/nmeth.4402 (2017).

997 68 Chen, J., Schlitzer, A., Chakarov, S., Ginhoux, F. & Poidinger, M. Mpath maps multi-branching single-cell
998 trajectories revealing progenitor cell progression during development. *Nat Commun* **7**, 11988,
999 doi:10.1038/ncomms11988 (2016).

1000 69 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC
1001 Bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).

1002 70 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular
1003 interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).

1004 71 Ibarra-Soria, X. *et al.* Defining murine organogenesis at single-cell resolution reveals a role for the
1005 leukotriene pathway in regulating blood progenitor formation. *Nat Cell Biol* **20**, 127-134,
1006 doi:10.1038/s41556-017-0013-z (2018).

1007 72 Hoeffel, G. *et al.* C-Myb(+) erythro-myeloid progenitor-derived fetal monocytes give rise to adult
1008 tissue-resident macrophages. *Immunity* **42**, 665-678, doi:10.1016/j.jimmuni.2015.03.011 (2015).

1009 73 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**,
1010 1083-1086, doi:10.1038/nmeth.4463 (2017).

1011 74 Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq
1012 data. *BMC Bioinformatics* **14**, 7, doi:10.1186/1471-2105-14-7 (2013).

1013 75 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on
1014 genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353-D361, doi:10.1093/nar/gkw1092
1015 (2017).

1016 76 Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require
1017 TBK1. *Nature* **462**, 108-112, doi:10.1038/nature08460 (2009).

1018 77 Zhang, H. M. *et al.* AnimalTFDB 2.0: a resource for expression, prediction and functional study of
1019 animal transcription factors. *Nucleic Acids Res* **43**, D76-81, doi:10.1093/nar/gku887 (2015).

1020 78 Bausch-Fluck, D. *et al.* A mass spectrometric-derived cell surface protein atlas. *PLoS One* **10**, e0121314,
1021 doi:10.1371/journal.pone.0121314 (2015).

1022 79 Tripathi, S. *et al.* Meta- and Orthogonal Integration of Influenza "OMICs" Data Defines a Role for UBR4

1023 in Virus Budding. *Cell Host Microbe* **18**, 723-735, doi:10.1016/j.chom.2015.11.002 (2015).

1024 80 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes
1025 among gene clusters. *OMICS* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).

1026

1027

1028 **FIGURE LEGENDS**

1029 **Fig. 1. Transcriptomic identification and molecular characteristics of the HECs in AGM**

1030 **Region.**

1031 (a) Schematic illustration of the strategies used for embryo dissection and cell preparation for
1032 the subsequent scRNA-seq. The involved body part and AGM region is indicated as blue and
1033 green, respectively, with head, limb buds, heart, visceral bud, and umbilical and vitelline
1034 vessels outside the embryo proper excluded.

1035 (b) PCA plots with three clusters (earlyAEC, lateAEC and HEC) (left), sampling locations
1036 (middle) and embryonic stages (right) mapped onto it.

1037 (c) Metascape network enrichment analysis with top 10 enriched terms exhibited to the right.
1038 Each cluster is represented by different colors and each enriched term is represented by a
1039 circle node. Number in the bracket indicates the *P* value based on -log10.

1040 (d) Classification of the indicated cells into quiescent phase and other cycling phases (G1, S
1041 and G2M) based on the average expression of G1/S and G2/M gene sets (left). Stacked bar
1042 chart showing the constitution of different cell cycle phases in the corresponding three
1043 clusters shown to the left (right).

1044 (e) Violin plot showing the number of transcripts for ribosomal related genes detected in each
1045 single cell of the indicated clusters. Wilcoxon Rank Sum test is employed to test the
1046 significance of difference and *P* values are indicated for the comparison. *P* < 0.05 is
1047 considered statistically significant.

1048 (f) Scatterplot showing the average arteriovenous scores of the cells in each cluster for
1049 mouse dataset in this paper (left) and human dataset from published articles (right),

1050 respectively. Main distribution ranges of arteriovenous scores in each cluster are also
1051 indicated as an oval shape.

1052 (g) Pseudotemporal ordering of the cells in select three clusters inferred by monocle 2, with
1053 pseudotime (left), clusters (middle) and sampling stages (right) mapped to it. HEC
1054 specification and AEC maturation directions are indicated as orange and deep red arrows,
1055 respectively.

1056 (h) Heatmap showing the expression of the indicated genes (smoothed over 15 adjacent cells)
1057 with cells ordered along the pseudotime axis of HEC specification branch inferred by monocle
1058 2.

1059 (i) Eight major expression patterns identified from the differentially expressed genes in HEC
1060 or lateAEC as compared to earlyAEC. Arrows showing the changes in HEC or lateAEC as
1061 compared to earlyAEC. The numbers of pattern genes are indicated to the right.

1062 (j) Heatmaps showing the relative expressions (smoothed over 20 adjacent cells) of the TFs
1063 belonging to the pattern genes with cells ordered along the pseudotime axis and genes
1064 ordered by patterns.

1065

1066 **Fig. 2. Efficiently isolating the HSC-competent and endothelial-hematopoietic
1067 bi-potent HECs before HSC emergence.**

1068 (a) Gene lists of the top ten cell surface molecules significantly overrepresented in HEC as
1069 compared to the indicated cell populations (first 3 lines) and those positively correlated with
1070 Runx1 within 4 EC clusters (vEC, earlyAEC, lateAEC and HEC, last line). Non-HEC, cells
1071 except for HEC within 4 EC clusters. Highlights in red font indicate the candidates used for
1072 further functional analysis.

1073 (b) Representative whole-mount staining of CD44 at E10.0 AGM region, showing CD44 is
1074 expressed in the whole endothelial layer of the dorsal aorta and roots of its proximal branches.
1075 DA, dorsal aorta; Scale bar, 100 μ m.

1076 (c) Representative FACS plots for cell sorting of the E9.5-E10.0 caudal half for
1077 co-culture/transplantation assay and the donor chimerism at 16 weeks after transplantation of
1078 the derivatives of the indicated cell populations.

1079 (d) Blood chimerism of the primary (I^o) and corresponding secondary (II^o) recipients at 16
1080 weeks post-transplantation. The primary recipients were transplanted with the derivatives of
1081 the indicated cells from the caudal half of E9.5-E10.0 embryos. The paired primary and
1082 corresponding secondary repopulated mice are shown as the same symbol and color.

1083 (e) Bars represent the percent donor contribution to the granulocytes/monocytes (GM, red), B
1084 lymphocytes (green), and T lymphocytes (purple) in the peripheral blood of the primary (I^o)
1085 and secondary (II^o) recipients at 16 weeks post-transplantation. The paired primary and
1086 corresponding secondary repopulated mice are shown as the same colors below.

1087 (f) FACS plot of Flk1 expression in the indicated population of E10.0 AGM region, with PK44
1088 (CD41⁺CD43⁺CD45⁺CD31⁺CD201⁺Kit⁺CD44⁺) cells (red) mapped onto it. Red box indicates
1089 the gate of Flk1⁺ cells.

1090 (g) t-SNE plot of the cells included in the filtered initial dataset and PK44 dataset, with
1091 clusters mapped on it. PK44, CD41⁺CD43⁺CD45⁺CD31⁺CD201⁺Kit⁺CD44⁺ population from
1092 E10.0 AGM region.

1093 (h) Heatmap showing the relative expressions of HEC feature genes, which are defined as
1094 those significantly highly expressed as compared to others including HC, vEC, earlyAEC and

1095 lateAEC, in the indicated cell populations. Selected HEC feature genes are shown to the right
1096 with pre-HSC signature genes marked as aquamarine.
1097 (i) Representative CD31 and CD45 immunostaining on the cultures of single PK44 cells from
1098 E10.0 AGM region, showing typical morphologies regarding distinct differentiation potentials.
1099 Cell frequencies of each kind of potential are also shown. Data are from 5 independent
1100 experiments with totally 15 embryos used. Scale bars, 400 μ m.
1101 (j) Expression of Kit and CD201 in the index-sorted single PK44 cells with differentiation
1102 potential based on in vitro functional evaluation. Cells with different kinds of potential are
1103 mapped onto the reference FACS plots (grey dots). Box in the middle plot indicates the gate
1104 for FACS sorting of PK44 cells in E10.0 AGM region and its enlarged view is shown to the
1105 right.

1106

1107 **Fig. 3. Relationship between HSC-primed HECs and T1 pre-HSCs.**

1108 (a) Representative FACS plots for sorting of the T1 pre-HSCs
1109 ($CD31^+CD45^-CD41^{low}Kit^+CD201^{high}$) from E11.0 AGM region of mouse embryos. Red box
1110 indicates the sampling cells for scRNA-seq.
1111 (b) Violin plots showing the expression levels of indicated genes in tif-HEC (including cluster
1112 HEC and PK44), T1 pre-HSC and lateAEC.
1113 (c) t-SNE plot of the cells included in the filtered initial dataset, PK44 dataset and T1 pre-HSC
1114 dataset, with clusters mapped on it. Cluster HEC and PK44 are combined as tif-HEC.
1115 (d) PCA plot of tif-HEC and T1 pre-HSC populations.
1116 (e) Enriched terms of PC2 positive and negative genes are shown, corresponding to the

1117 properties distinguishing tif-HEC and T1 pre-HSC, respectively.

1118 (f) Heatmap showing top 20 positive and negative genes of PC2. Genes were ordered by

1119 their contributions to PC2.

1120 (g) Trajectory of AEC clusters, tif-HEC and T1 pre-HSC inferred by Mpath. Arrows indicate

1121 the development directions predicted by sampling stages.

1122 (h) Representative FACS plots for sorting of the PK44 cells from E10.0 AGM region (left) and

1123 analysis of the immunophenotypic T1 pre-HSCs (right) after cultured in vitro for 4 days.

1124 (i) Representative CD31 immunostaining on the cultures of single T1 pre-HSCs from E11.0

1125 AGM region, showing typical morphologies regarding distinct differentiation capacities. Cell

1126 frequencies of each type are also shown. Data are from 7 independent experiments with

1127 totally 89 embryos used. Scale bars, 400 μ m.

1128

1129 **Fig. 4. Identifying *Neurl3* as a signature gene of HSC-primed HECs validated by**

1130 **functional and transcriptomic evaluation.**

1131 (a) Dot plot showing the average and percentage expression of HEC signature genes in the

1132 indicated clusters. Genes are ordered by their median expression level in tif-HEC. Pre-HSC

1133 signature genes are marked as aquamarine.

1134 (b) Schematic model of the gene targeting strategy for generating *Neurl3*^{EGFP/+} reporter

1135 mouse line via CRISPR/Cas9 system.

1136 (c) Representative FACS analysis of the E10.0 AGM region in *Neurl3*^{EGFP/+} embryos, FACS

1137 plot to the right showing PK44 cells (red dots) mapped on it.

1138 (d) Representative FACS plot for sorting of the indicated cell populations from E10.0 caudal

1139 half of *Neurl3*^{EGFP/+} embryos.

1140 (e) Graph showing the donor chimerism at 16 weeks after transplantation of the derivatives of
1141 the indicated populations from the caudal half of E10.0 *Neurl3*^{EGFP/+} embryos.

1142 (f) Graph showing the donor chimerism at 4-16 weeks post-transplantation. The recipients
1143 were transplanted with the derivatives of CD41-CD43-CD45-CD31⁺CD44⁺*Neurl3*-EGFP⁺
1144 population from the caudal half of E10.0 *Neurl3*^{EGFP/+} embryos. Number of repopulated/total
1145 recipients is shown in the brackets.

1146 (g) t-SNE plot of the cells included in the filtered initial dataset and additional PK44 and NE+
1147 datasets, with clusters mapped on it. Cluster HEC and PK44 are combined as tif-HEC. NE+,
1148 CD41-CD43-CD45-CD31⁺CD44⁺*Neurl3*-EGFP⁺ population from E10.0 AGM region.

1149 (h) Dot plot showing the average and percentage expression of selected HEC feature genes
1150 in the indicated clusters. Pre-HSC signature genes are marked as aquamarine.

1151 (i) Heatmap showing the correlation coefficient between each two clusters with hierarchical
1152 clustering using average method. Pearson correlation coefficient is calculated using average
1153 expression of highly variable genes in each cluster.

1154

1155 **Fig. 5. In situ localization and in vitro function of the dynamic HECs marked by**
1156 ***Neurl3*-EGFP reporter.**

1157 (a) Representative immunostaining on cross sections at AGM region of E9.5 (upper), E10.0
1158 (middle) and E10.5 (lower) *Neurl3*^{EGFP/+} embryos. Arrows indicate *Neurl3*⁺ aortic ECs. Yellow
1159 arrowheads indicate *Neurl3*⁺ bulging and bulged cells and IAHCs. Aquamarine arrowheads
1160 indicate CD44⁺Runx1⁺*Neurl3*⁻ haematopoietic cells distributed outside the aorta. nt, neural
1161 tube; DA, dorsal aorta. Scale bars, 100 μ m.

1162 (b) Representative FACS analysis of the E10.0 AGM region of *Neurl3^{EGFP/+}* embryos. FACS
1163 plots to the right showing PK44 cells (red dots, upper) and CD31⁺Kit^{high} cells (blue dots, lower)
1164 mapped on, respectively, with their contributions to each gated population indicated.

1165 (c) Representative CD31 and CD45 immunostaining on the cultures of single
1166 CD41⁻CD43⁻CD45⁻CD31⁺CD44⁺*Neurl3-EGFP⁺* cells from E10.0 AGM region of *Neurl3^{EGFP/+}*
1167 embryos, showing typical morphologies regarding distinct differentiation potentials. Cell
1168 frequencies of each kind of potential are also shown. Data are from 5 independent
1169 experiments with totally 37 embryos used. Scale bars, 400 μ m.

1170 (d) Column charts showing the proportions of positive wells in the indicated populations
1171 (lower) for each kind of potential. The experiments were performed with
1172 CD41⁻CD43⁻CD45⁻CD31⁺CD44⁺*Neurl3-EGFP⁺* single cells from E9.5 caudal half or
1173 E10.0-E10.5 AGM region of *Neurl3^{EGFP/+}* embryos with PK44 indexed. Progenies from PK44
1174 and non-PK44 cells are represented by distinct fill patterns.

1175 (e) Expression of CD44 and *Neurl3-EGFP* and values of FSC-A and SSC-A in the
1176 index-sorted single CD41⁻CD43⁻CD45⁻CD31⁺CD44⁺*Neurl3-EGFP⁺* cells with differentiation
1177 potential based on in vitro functional evaluation. Cells with different kinds of potential are
1178 mapped onto the reference FACS plots (grey dots). Solid boxes (left of each stage) indicate
1179 the gates of the populations for FACS sorting. The enlarged views of solid boxes are shown
1180 below.

1181

1182 **Fig. 6. Molecular evolution underlying the specification of HSC-primed HECs from**
1183 **primitive vascular ECs.**

1184 (a) Trajectory of pEC, vEC, pAEC, earlyAEC, lateAEC and HEC inferred by Mpath. Arrows

1185 indicate the development directions predicted by sampling stages.

1186 (b) t-SNE plot showing the distribution of the four clusters involved in hemogenic specification.

1187 Other cells are in grey.

1188 (c) Pseudotemporal ordering of the cells included in the indicated five clusters inferred by

1189 monocle 2 (left), with clusters (upper left) and sampling stages (lower left) mapped to it. HEC

1190 specification directions are indicated as red arrows. Smooth distribution of clusters (upper

1191 right) and sampling stages (lower right) along pseudotime by using Gaussian kernel density

1192 estimate are shown to the right.

1193 (d) Dynamic changes of five gene expression patterns along the trajectory ordered by

1194 pseudotime inferred by monocle 2. For each pattern, principal curves are fitted on expression

1195 levels of the genes in that pattern along pseudotemporal order, using local polynomial

1196 regression fitting method. Randomly down-sampling is performed in pEC and pAEC clusters

1197 for better visualization.

1198 (e) Heatmap showing the relative expression of the core TFs which belong to the regulons

1199 that the genes within exhibit significant overlapping with the pattern genes. Cells are ordered

1200 by pseudotime and TFs are ordered by Patterns.

1201 (f) Heatmap showing smoothed (along adjacent 25 cells) and scaled enrichment scores of top

1202 50 KEGG pathways along the order by pseudotime. Pathways are ordered by hierarchical

1203 clustering using ward.D method.

1204 (g) Scatter plots showing the relative activity levels of pathways or GO terms with loess

1205 smoothed fit curves and 95% confidence interval indicated. Relative activity levels are

1206 represented by the PC1 scores of expression levels of the genes in a given set. The sign or

1207 direction of PC1 is corrected according to positive correlation with averaged expression
1208 levels.

1209 **SUPPLEMENTARY FIGURE LEGENDS**

1210 **Fig. S1. Information, clustering of initial dataset and molecular characteristics**
1211 **of major clusters.**

1212 (a) Embryo, independent experiment, and cell number information for scRNA-seq. DA,
1213 dorsal aortic luminal layer of AGM region.

1214 (b) Whole-mount image of the E11.0 AGM region labeled with Oregon Green 488.

1215 (c) Boxplots showing the number of genes (left) and transcripts (right) in each single
1216 cell of different locations.

1217 (d) t-SNE plots with clusters (left), sampling locations (right) and embryonic stages
1218 (right) mapped onto it.

1219 (e) Violin plots showing the expression levels of indicated genes in six clusters
1220 identified in the initial dataset.

1221 (f) Cell number information of the spatiotemporal distribution of distinct clusters.

1222 (g) Volcano plots showing differentially expressed genes (marked as blue or red)
1223 between two sub-clusters by forced clustering in earlyAEC and HEC, respectively.

1224 Top 10 (earlyAEC) or all (HEC) differentially expressed genes are indicated. *Runx1*
1225 and *Kit* are also indicated.

1226 (h) Violin plots showing the number of genes (left) and transcripts (right) in each
1227 single cell of the indicated clusters. Wilcoxon Rank Sum test is employed to test the
1228 significance of difference and *P* values are indicated for the comparison. *P* < 0.05 is
1229 considered statistically significant.

1230 (i) Heatmap showing the relative expression levels of genes in eight patterns among
1231 earlyAEC, lateAEC and HEC.

1232 (j) Network view of TFs positively correlated with the gene expression patterns. A
1233 deeper background color of the gene name indicates a higher positive correlation of
1234 the TF to that expression pattern.

1235 (k) Bar chart showing the top 50 genes positively correlated with Runx1 within cell
1236 population including earlyAEC, lateAEC and HEC. Genes included in the patterns
1237 identified above are marked as indicated.

1238

1239 **Fig. S2. Identification of the HSC-competent and endothelial-haematopoietic
1240 bi-potent HECs.**

1241 (a) Detailed information of the co-culture/transplantation assays performed with
1242 E9.5-E10.0 caudal half cells.

1243 (b) Blood chimerism of the primary and secondary recipients at 4-16 weeks
1244 post-transplantation. The primary recipients were transplanted with the derivatives of
1245 the indicated cell populations from the caudal half of E9.5-E10.0 embryos. The paired
1246 primary and corresponding secondary repopulated mice are show as the same
1247 symbol and color. Numbers of repopulated/total recipients are shown in the brackets.

1248 Only the recipients survived to 16 weeks post-transplantation are shown.

1249 (c) FACS plots showing representative primary recipients with long-term (16 weeks),
1250 multi-organ and multi-lineage repopulations transplanted with the derivatives of the
1251 indicated cell populations from the caudal half of E9.5-E10.0 embryos. Donor-derived

1252 (CD45.1+CD45.2+) myeloid (Gr-1+/Mac-1+), B lymphoid (B220+), and T lymphoid
1253 (CD3+) cells in multiple haematopoietic organs are shown.

1254 (d) Heatmap showing the expression of selected genes in earlyAEC, lateAEC, HEC
1255 and PK44 populations. Note the similarity of expression patterns between HEC and
1256 PK44.

1257 (e) Graph showing the endothelial potential of different cell populations in E9.5-E10.0
1258 body part of embryo proper. Cells with indicated immunophenotype were isolated by
1259 FACS, co-cultured with OP9 stromal cells for 7 days, and stained with CD31 to
1260 identify the endothelial tubes. Data are means \pm s.d.. For E9.5 embryos, data are
1261 from 3 independent experiments with 6-9 embryo equivalents pooled for each
1262 experiment. For E10.0 embryos, data are from 3 independent experiments with 8-9
1263 embryo equivalents pooled for each experiment.

1264 (f) Detailed information of endothelial-haematopoietic bi-potential induction assays
1265 performed with cells from E9.5-E10.0 caudal half or AGM region.

1266

1267 **Fig. S3. Identification of the HSC-competent HECs marked by Neurl3-EGFP
1268 reporter.**

1269 (a) Detailed information of the co-culture/transplantation assays performed with the
1270 caudal half cells from E10.0 *Neurl3^{EGFP/+}* embryos.

1271 (b) Boxplot showing the transcriptional expression level of *EGFP* in NE+ cell
1272 population, NE+, CD41⁺CD43⁺CD45⁺CD31⁺CD44⁺*Neurl3-EGFP*⁺ population from
1273 AGM region of E10.0 *Neurl3^{EGFP/+}* embryos.

1274 (c) Scatter plots showing correlation of the expression of *EGFP* with that of *Neurl3*
1275 and *Runx1*, respectively. Fitted line and 95% confidence interval are shown in red.
1276 Pearson correlation coefficients and *P* values are also shown in blue text.
1277 (d) Stacked bar chart showing the constitution of different cell cycle phases in the
1278 indicated clusters.
1279 (e) Representative immunostaining on cross sections at AGM region of E11.0
1280 *Neurl3*^{EGFP/+} embryos. Arrow indicates *Neurl3*⁺ aortic ECs; Yellow arrowheads
1281 indicate *Neurl3*⁺ bulging and bulged cells and IAHCs. Aquamarine arrowheads
1282 indicate CD44⁺*Runx1*⁺*Neurl3*⁻ haematopoietic cells distributed outside the aorta. nt,
1283 neural tube; DA, dorsal aorta. Scale bars, 100 μ m.
1284 (f) Detailed information of endothelial-haematopoietic bi-potential induction assays
1285 performed with cells from E9.5 caudal half and E10.0-E10.5 AGM region of
1286 *Neurl3*^{EGFP/+} embryos.

1287

1288 **Fig. S4. Molecular programs from primitive vascular ECs to HSC-primed HECs.**
1289 (a) Embryo, independent experiment, and cell number information for additional
1290 scRNA-seq. sp, somite pairs.
1291 (b) t-SNE plots with clusters (upper left), sampling locations (lower left), embryonic
1292 stages (upper right) and clusters previously defined (lower right) mapped onto it.
1293 (c) Dot plot showing the average and percentage expression of selected marker
1294 genes in the indicated clusters.
1295 (d) Pseudotemporal ordering of the cells involved in HEC specification, including
1296 those in pEC, pAEC, earlyAEC, HEC, and T1 pre-HSC, inferred by monocle 2, with

1297 pseudotime mapped to it.

1298 (e) Heatmap showing smoothed and scaled expression levels of 2,851 pattern genes.

1299 Genes are ordered by patterns. Cells are ordered by pseudotime.

1300 (f) Dot plot showing the top six enriched Gene Ontology biological process (GO:BP)

1301 terms for each pattern. Dot color indicates statistical significance of the enrichment

1302 and dot size represents the fraction of genes annotated to each term.

1303 (g) Scatter plots showing the expression levels of the TF genes previously reported to

1304 be functional in HSPC regeneration along the pseudotemporal order with loess

1305 smoothed fit curves and 95% confidence interval indicated. The patterns to which the

1306 genes belong are indicated by different fill colors. The core TFs of the significantly

1307 overlapped regulons are underlined.

1308

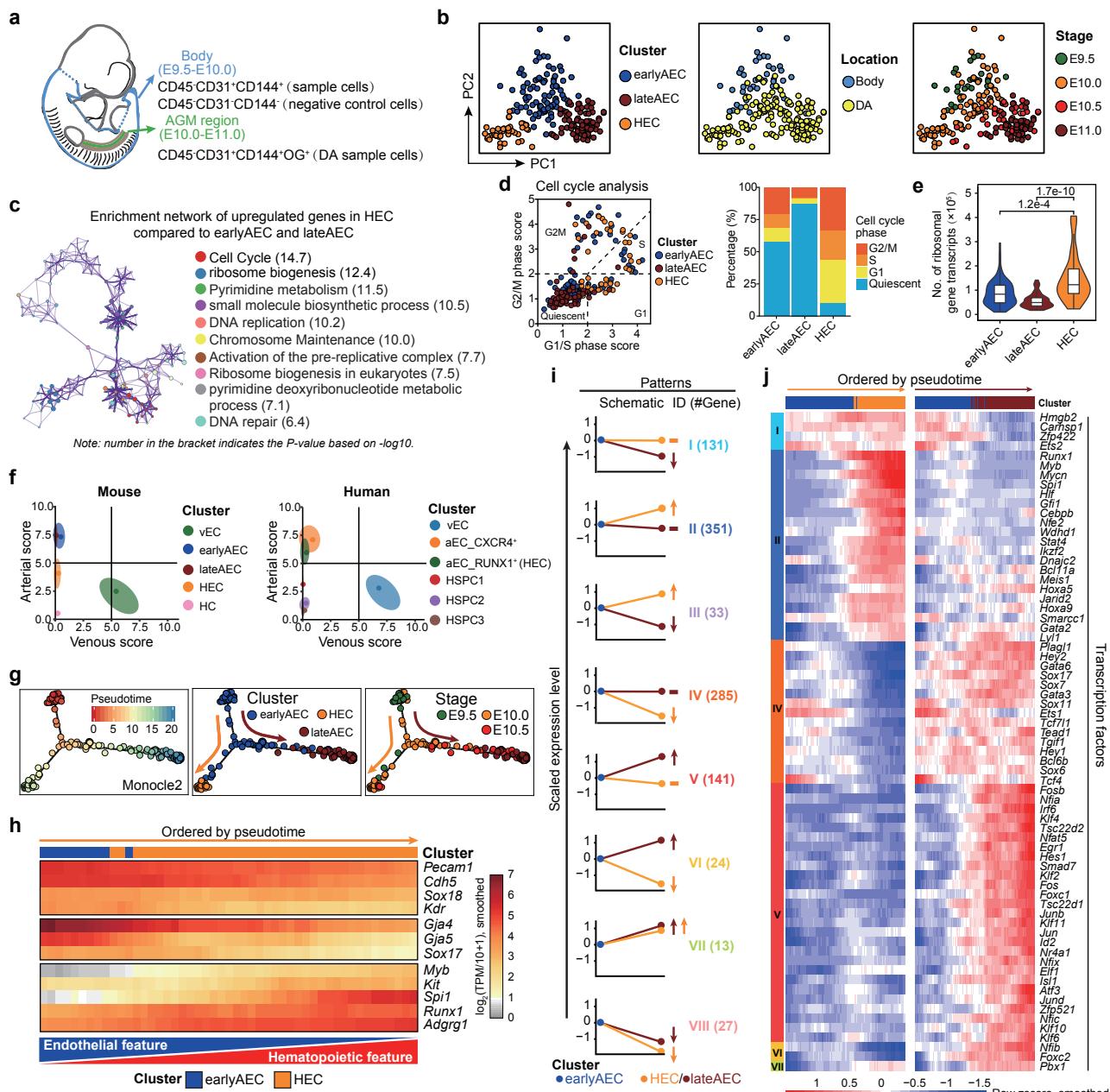
Fig. 1

Fig. 2

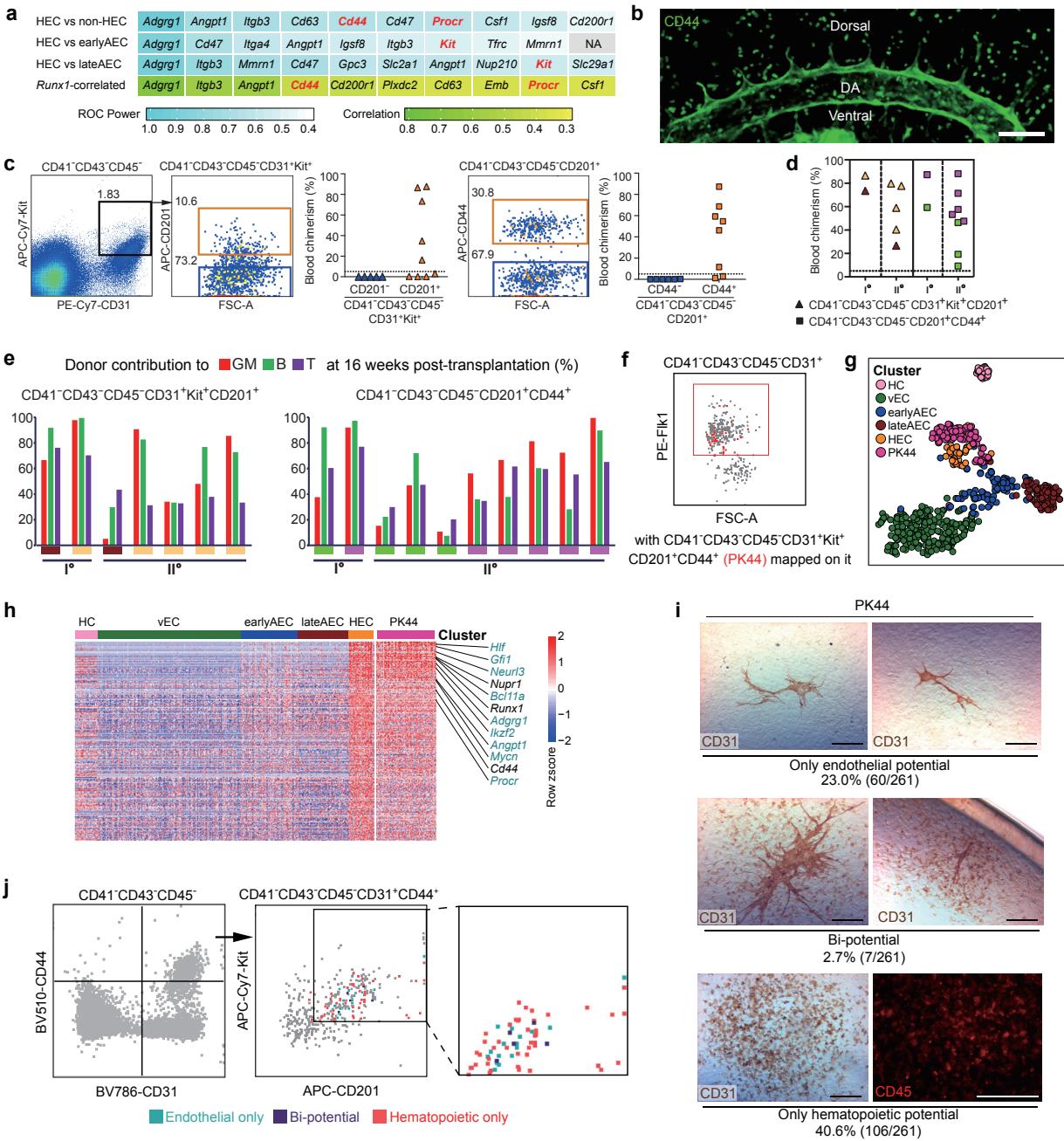


Fig. 3

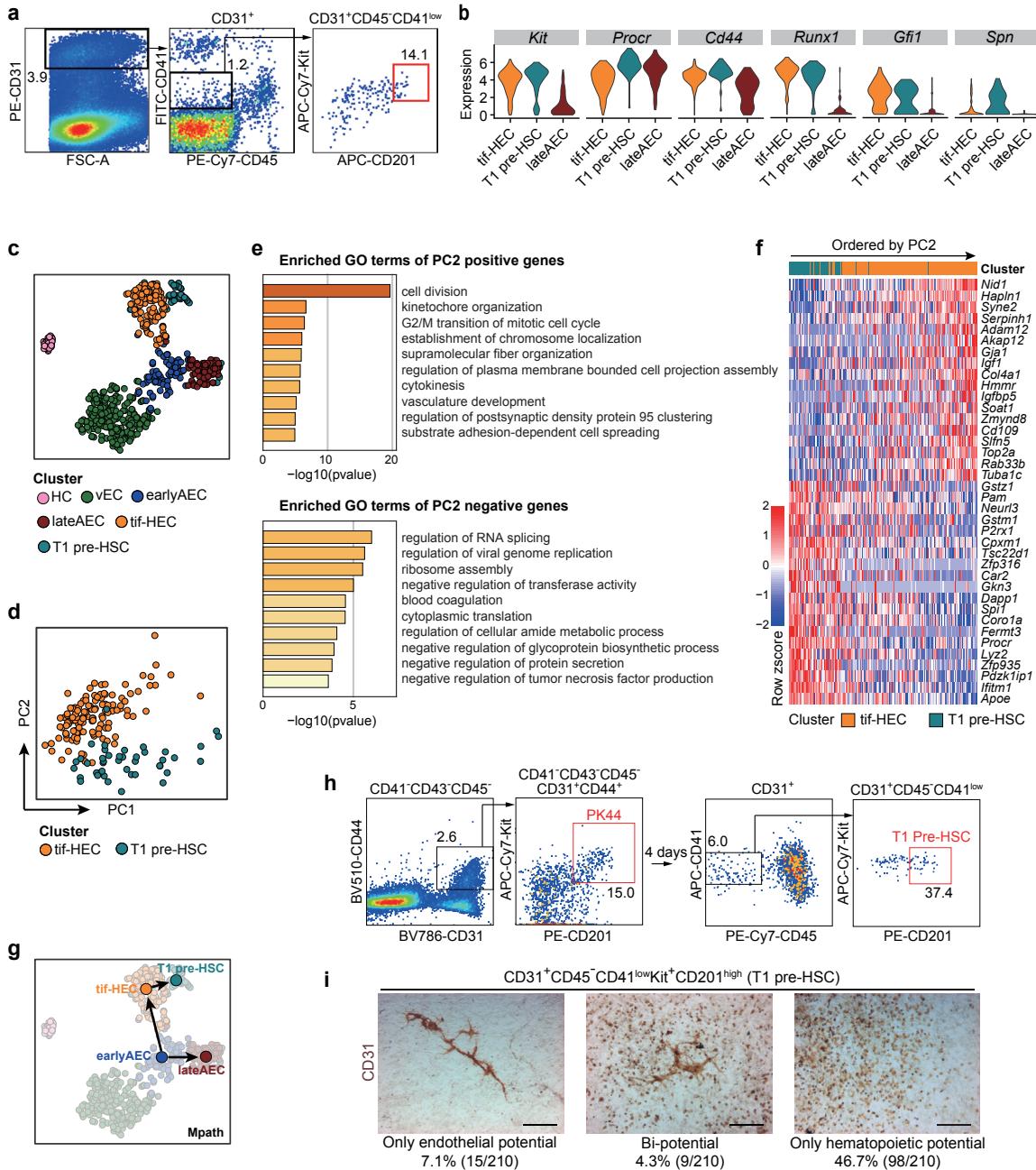


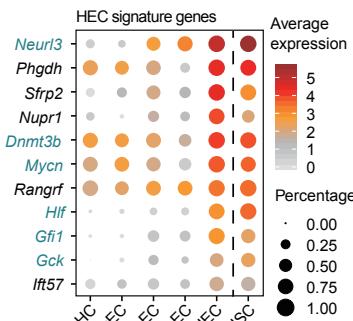
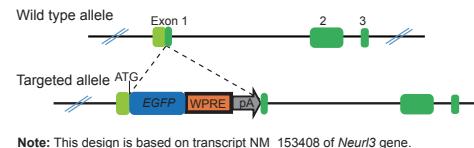
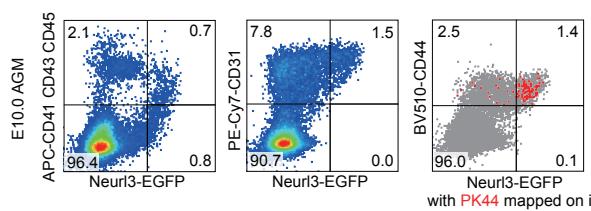
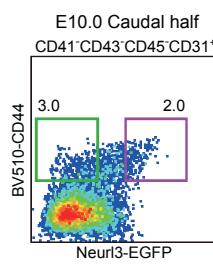
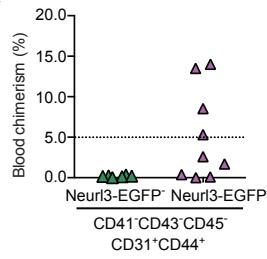
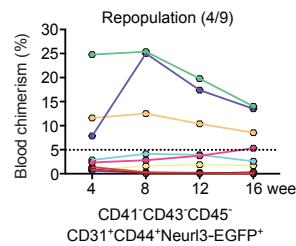
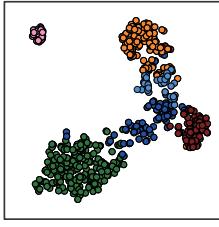
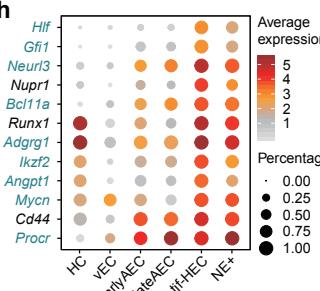
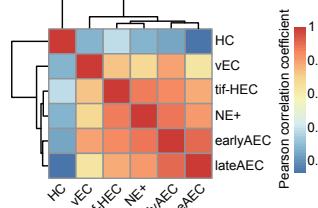
Fig. 4**a****b****c****d****e****f****g****h****i**

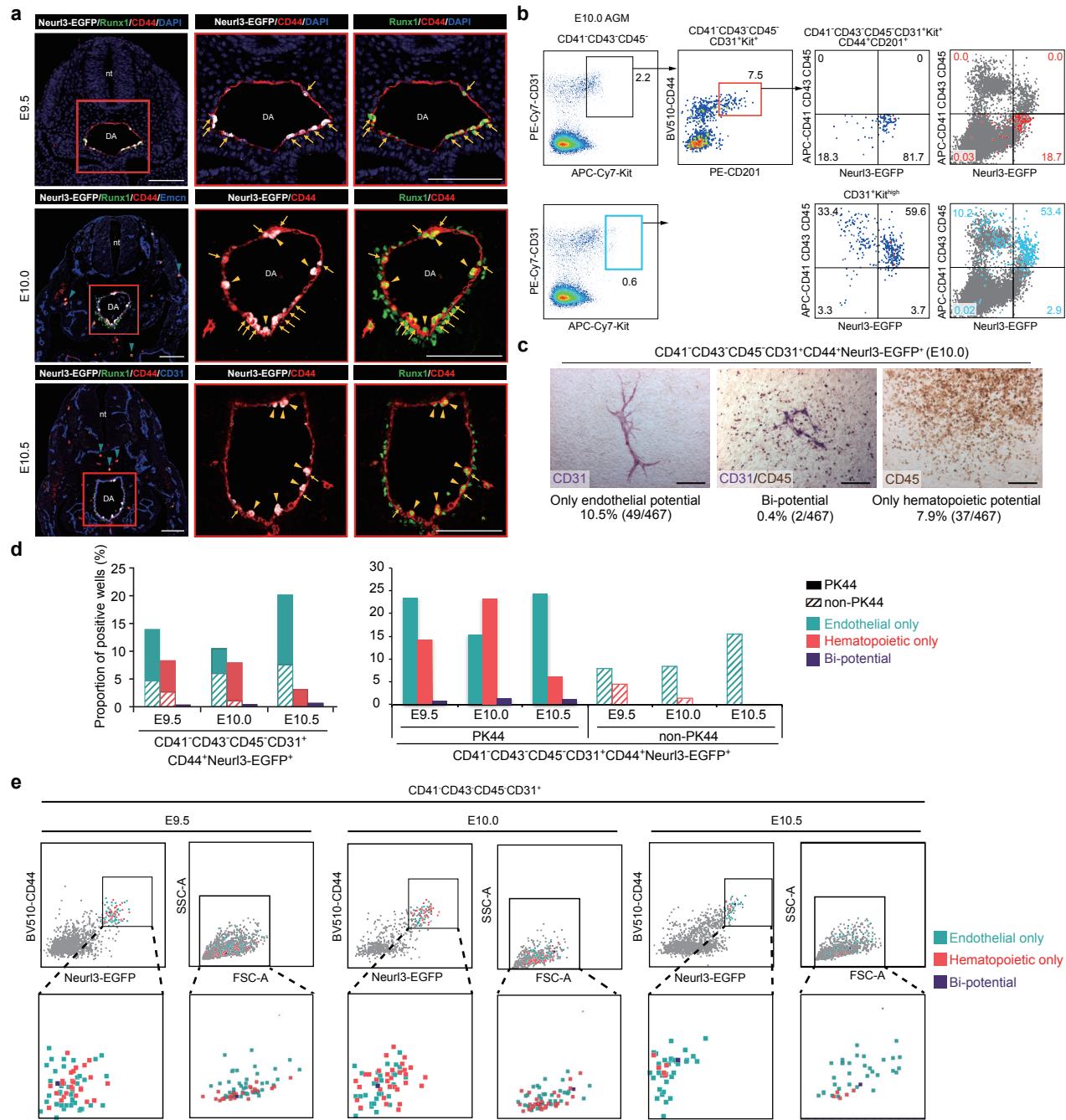
Fig. 5

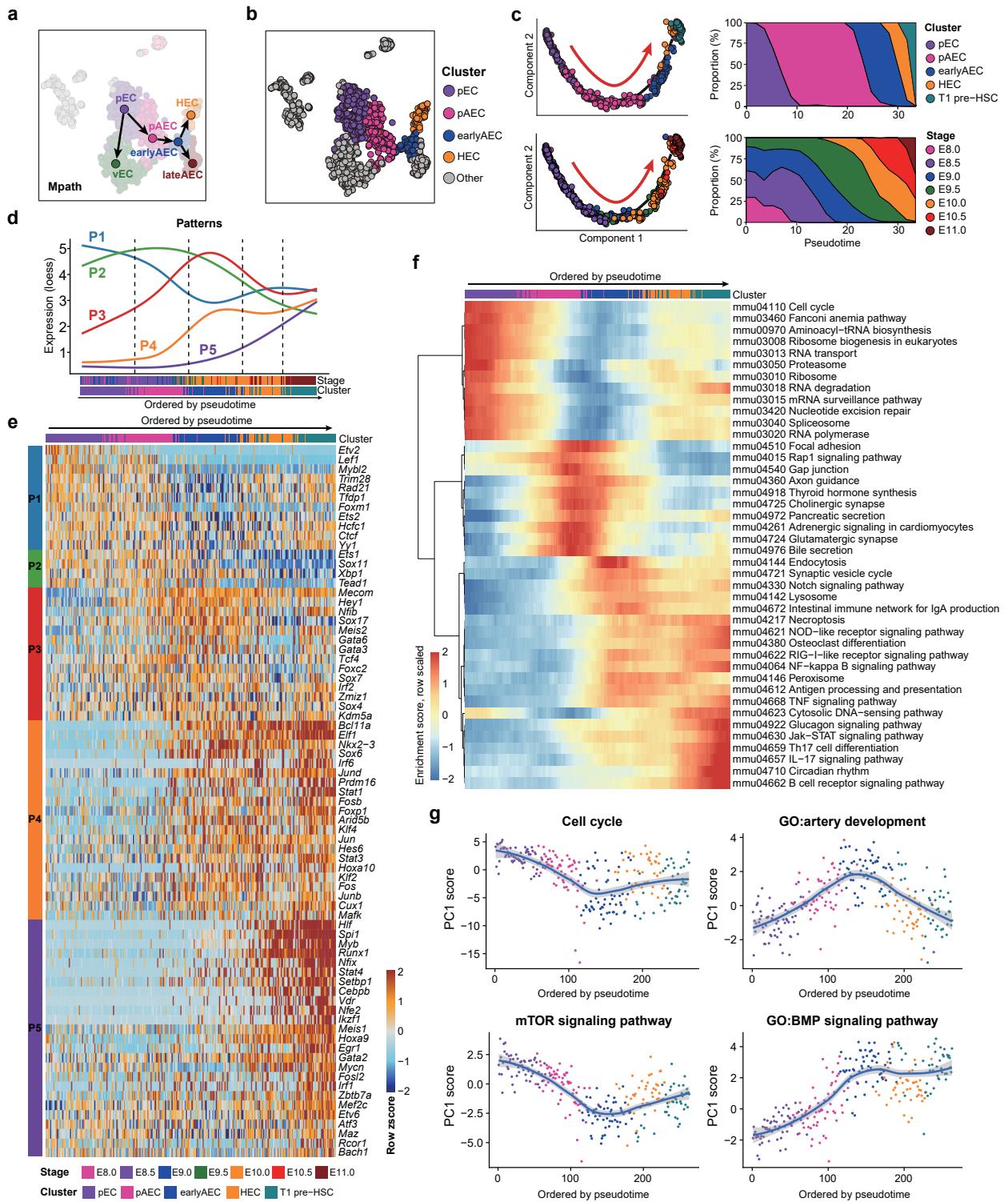
Fig. 6

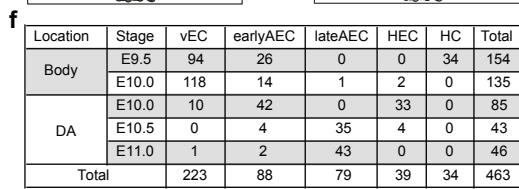
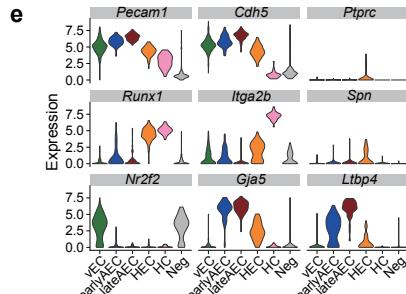
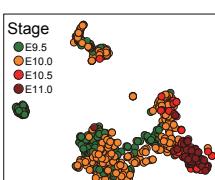
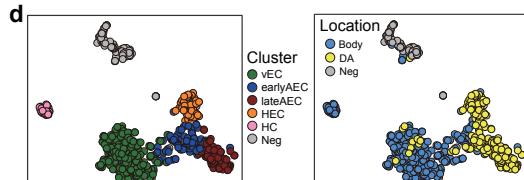
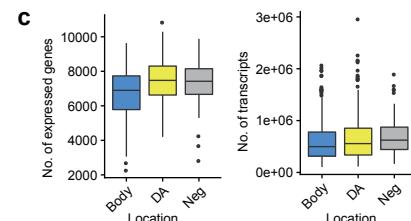
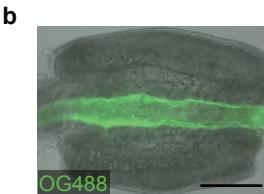
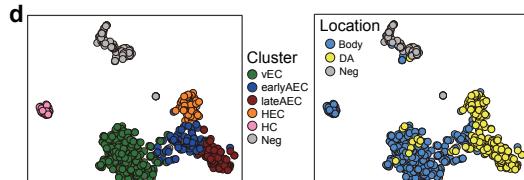
Fig. S1

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	
E11.0 (43-44 sp)	7	1 ^c	/	56	/	56	
Total	29	8	360	222	80		662

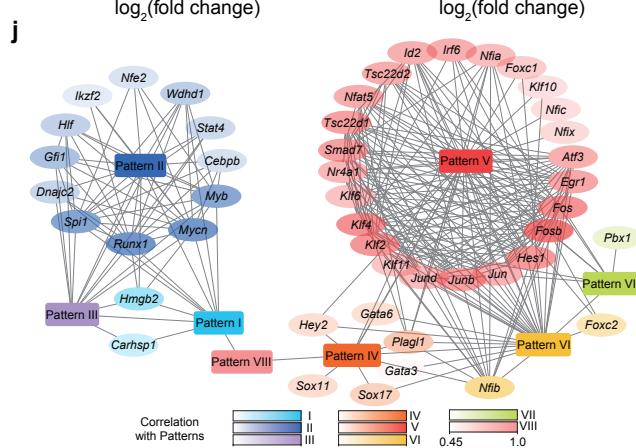
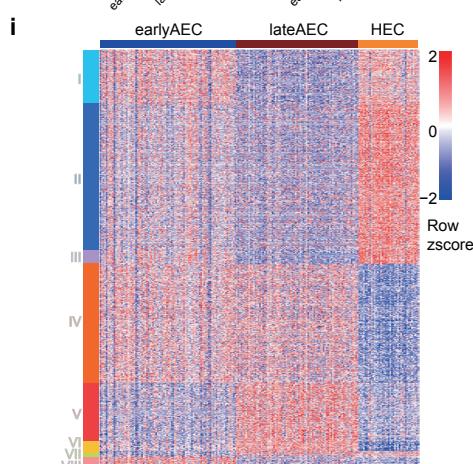
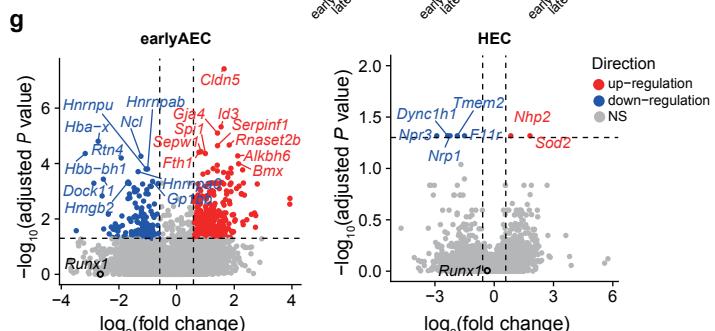
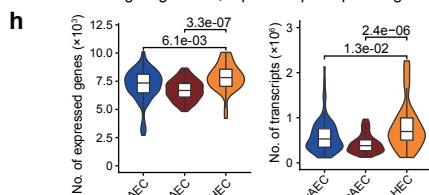
^a CD45⁻CD31⁻CD144⁻ negative control cells from body:

^b Independent experiments for collecting CD45⁻CD31⁺CD144⁺ sample cells from body;

Independent experiments for collecting CD45⁺CD31⁺CD144⁺ Oregon green⁺ sample cells from AGM region.



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



Genes positively correlated with *Runx1*

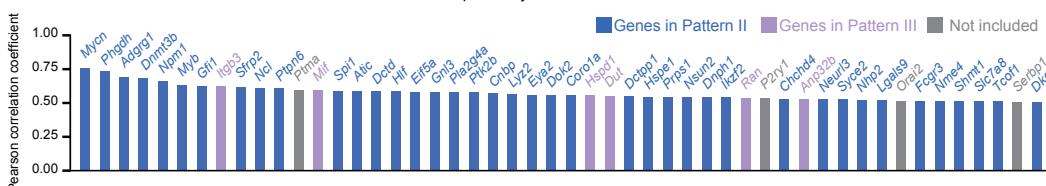


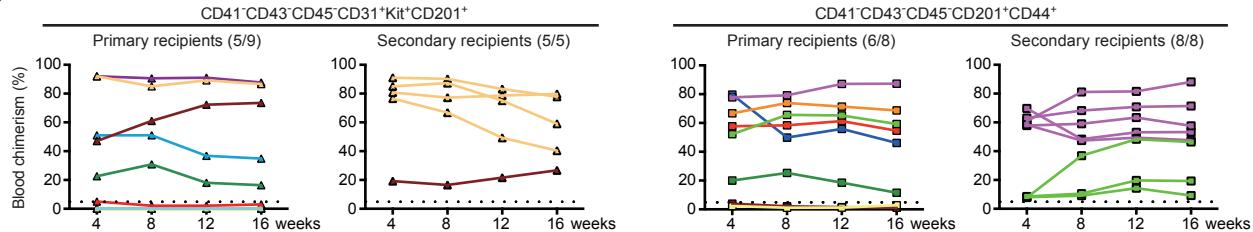
Fig. S2

a

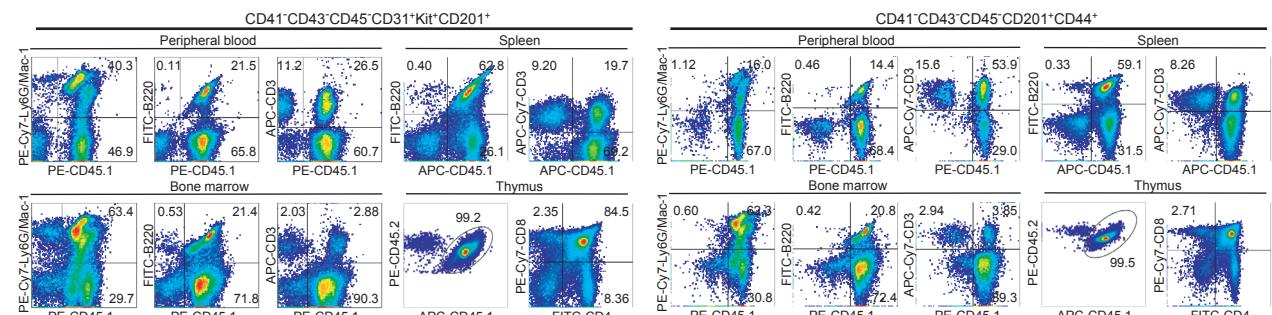
Stage	Organ	Exp (n)	Population	Constitution in the region (%)	Cell dose ^a	Clusters ^b
E9.5-E10.0 (30-32 sp ^c)	Caudal half	4	CD41 ⁺ CD43 ⁺ CD45 ⁺ CD31 ⁺ Kit ⁺ CD201 ⁺	0.20±0.07	3.0-4.0 ee ^d	13/13 (100%)
			CD41 ⁺ CD43 ⁺ CD45 ⁺ CD31 ⁺ Kit ⁺ CD201 ⁻	1.25±0.45	3.0-4.0 ee	2/13 (15.4%)
E9.5-E10.0 (30-34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

^acell dose per well and per recipient. ^bhematopoietic cluster positive wells/total wells. ^csomite pairs. ^dembryo equivalents.

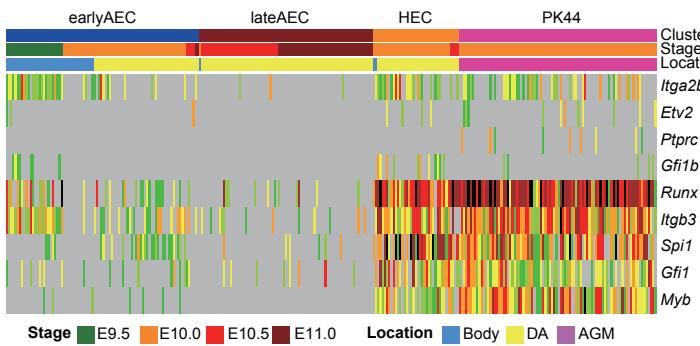
b



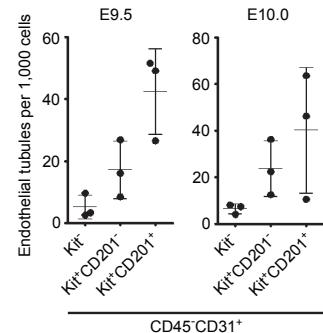
c



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%)

^acells per well. ^bhematopoietic and endothelial bi-potential positive wells/total wells. ^chematopoietic progeny positive wells/total wells.

^dendothelial tube positive wells/total wells. ^esomite pairs. ^f2 independent experiments, totally 12 embryos were used.

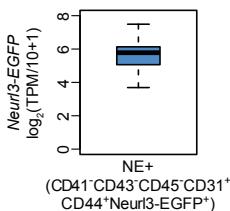
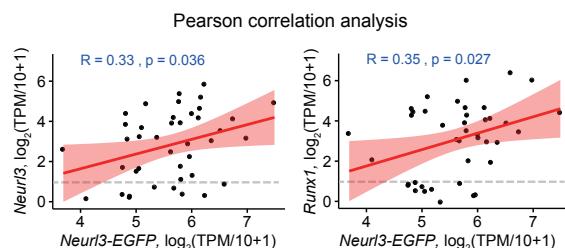
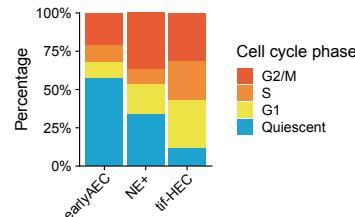
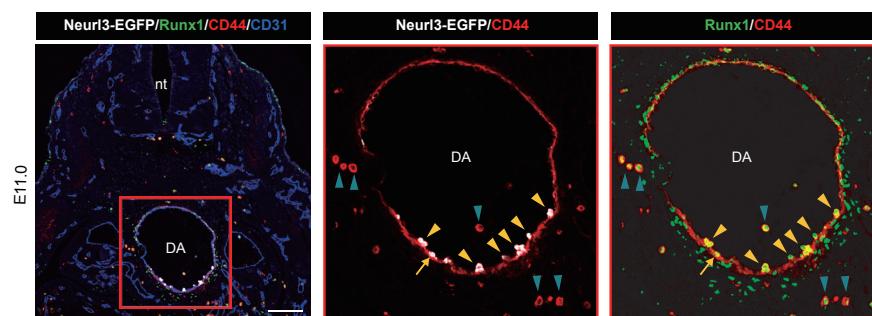
^g5 independent experiments, totally 15 embryos were used.

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%)

^aCell dose per recipient. ^bhematopoietic cluster positive wells/total wells. ^csomite pairs.

^dembryo 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%)

^ahematopoietic and endothelial bi-potential positive wells/total wells. ^bhematopoietic progeny positive wells/total wells. ^cendothelial tube positive wells/total wells.

^dsomite pairs. ^e8 independent experiments, totally 32 embryos were used. ^f5 independent experiments, totally 37 embryos were used.

^g6 independent experiments, totally 23 embryos were used.

Fig. S4