

1 **Systematic single-cell analysis reveals dynamic control of transposable element
2 activity orchestrating the endothelial-to-hematopoietic transition**

3 Cong Feng^{1,2†}, Ruxiu Tie^{3,4,5,6,7,8†}, Saige Xin^{1†}, Yuhao Chen¹, Sida Li¹, Xiaotian Hu¹,
4 Yincong Zhou¹, Yongjing Liu², Yueming Hu¹, **Yanshi Hu¹**, Hang Pan⁹, Zexu Wu¹, Haoyu
5 Chao¹, Shilong Zhang¹, Qingyang Ni¹, Jinyan Huang², Wenda Luo^{8*}, He Huang^{3,4,5,6*}, Ming
6 Chen^{1,2*}

7 ¹Department of Bioinformatics, Zhejiang University College of Life Sciences, Hangzhou
8 310058, China

9 ²Bioinformatics Center, The First Affiliated Hospital, Zhejiang University School of Medicine,
10 Hangzhou 310058, China

11 ³Bone Marrow Transplantation Center, the First Affiliated Hospital, Zhejiang University
12 School of Medicine, Hangzhou 310058, China

13 ⁴Liangzhu Laboratory, Zhejiang University Medical Center, Hangzhou 310058, China

14 ⁵Institute of Hematology, Zhejiang University, Hangzhou 310058, China

15 ⁶Zhejiang Province Engineering Laboratory for Stem Cell and Immunity Therapy,
16 Hangzhou 310058, China

17 ⁷Department of Hematology, The Second Clinical Medical College of Shanxi Medical
18 University, Shanxi Medical University, Taiyuan 030000, China

19 ⁸Department of Hematology-Oncology, Taizhou Hospital of Zhejiang Province, Linhai
20 317000, China

21 ⁹Department of Veterinary Medicine, Zhejiang University College of Animal Sciences,
22 Hangzhou 310058, China

23 *Correspondence: mchen@zju.edu.cn; hehuangyu@126.com; luowd@enzemed.com

24 †These authors have contributed equally to this work.

25 **Abstract**

26 **Background:** The endothelial-to-hematopoietic transition (EHT) process during definitive
27 hematopoiesis in vertebrate is highly conserved. Stage-specific expression of transposable
28 elements (TEs) has been detected during zebrafish EHT and may promote hematopoietic
29 stem cell formation by activating inflammatory signaling. However, little is known about how
30 TEs contribute to the EHT process in human and mouse.

31 **Results:** We reconstructed the single-cell EHT trajectories of human and mouse, and
32 resolved the dynamic expression patterns of TEs during EHT. Most TEs presented a
33 transient co-upregulation pattern along the conserved EHT trajectories. Enhanced TE
34 activation was tightly associated with the temporal relaxation of epigenetic silencing
35 systems. TE products can be sensed by multiple pattern recognition receptors, triggering
36 inflammatory signaling to facilitate the emergence of hematopoietic stem cells.
37 Furthermore, we observed that hypoxia-related signals were enriched in cells with higher
38 TE expression. Additionally, we constructed the hematopoietic cis-regulatory network of
39 accessible TEs and identified potential enhancers derived by TEs, which may boost the
40 expression of specific EHT marker genes.

41 **Conclusions:** Our study provides a systematic vision on how TEs are dynamically
42 controlled to promote the hematopoietic fate decision through transcriptional and cis-
43 regulatory networks, and pre-train the immunity of nascent hematopoietic stem cells.

44 **Keywords:** Endothelial-to-hematopoietic transition, Transposable element, Hematopoietic
45 stem cell, Inflammatory signaling, Cis-regulatory element, Hypoxia

46 **Introduction**

47 Hematopoietic stem cells (HSCs) are a pluripotent cell population in the blood system,
48 which possess the ability of self-renewal and lineage differentiation to maintain the function
49 of the hematopoietic system throughout the lifespan. In embryos, HSCs originate from a
50 subpopulation of endothelial cells (ECs) with hematopoietic potential in the aorta-gonad-
51 mesonephro (AGM) region, which have been directly tracked by using time-lapse imaging
52 methods [1-5]. The endothelial-to-hematopoietic transition (EHT) process is highly
53 conserved in vertebrate embryos including zebrafish, mouse, and human [6]. HSCs first
54 emerge during embryonic day (E) 10.5-11.5 in mouse AGM [7, 8] and Carnegie stage 13-
55 17 (CS13-17; 4-6 weeks) in human AGM [9, 10], and migrate to the fetal liver under the
56 promotion of blood flow, and finally colonize the fetal bone marrow. The identification and
57 functional analysis of heterogeneous and intermediate cell clusters during EHT *in vivo* has
58 been a challenging task in understanding and probing embryonic hematopoiesis [11]. It is
59 presented that a specific cluster of arterial endothelial cells (AECs) from primitive blood
60 vessels undergoes hemogenic fate decision to become HSC-primed hematopoietic
61 endothelial cells (HECs) [12-14]. In addition to HECs, at least two kinds of HSC precursors
62 (distinguished by CD45) are found in mouse intra-aortic hematopoietic clusters (IAHCs)
63 [15-18]. In the last years, the rapid development of single-cell sequencing technologies has
64 greatly broadened the insights into cellular heterogeneity and complex relationships during

65 developmental hematopoiesis [19, 20]. Using single-cell RNA sequencing (scRNA-seq)
66 and single-cell sequencing assay for transposase-accessible chromatin (scATAC-seq), the
67 continuous EHT trajectory has been constructed and an intermediate cell population
68 proximal to HECs is identified (termed pre-HECs) in mouse [21, 22]. Recently, a research
69 team discovered a signature gene set RUNX1+HOXA9+MLLT3+MECOM+HLF+SPINK2+
70 that distinguishes human HSCs from other hematopoietic progenitor cells, and for the first
71 time depicted the single cell landscape of HSCs from origination to maturation in the human
72 embryo [23]. However, whether the cell types are comparable in human and mouse EHT,
73 and how many marker genes are conserved, has not been systematically investigated.

74 It is recognized that the EHT process is strictly regulated by multiple regulatory factors at
75 the transcriptional and epigenetic levels [24, 25]. Various transcription factors such as
76 Runx1, Gfi1 and Gata2 have been shown to play vital roles in HSC development [26-28].
77 Signaling pathways like NOTCH, WNT, YAP and VEGF are also involved in HSC fate
78 decision [25]. Besides, it is worth noting that some inflammatory signals are highlighted to
79 regulate the HSC emergence [29, 30]. For example, interleukins like IL-1, IL-3, and IL-6
80 regulate HSC development endogenously in the AGM region [31, 32]. Tumor necrosis
81 factors (TNF) and interferon signals (IFN) can also promote the development of HSCs [33,
82 34]. In the innate immune system, pattern recognition receptors (PRRs) such as Toll-like
83 receptors (TLRs), RIG-I-like receptors (RLRs), NOD-like receptors (NLRs), and C-type
84 lectin receptors (CLRs) are considered to be key activators of these inflammatory
85 responses [35]. It has been pointed out that Toll-like receptor 4 (TLR4) can regulate the
86 formation and development of HSCs by promoting Notch activity through MyD88-mediated

87 NF-κB signaling [30]. A recent study in zebrafish unraveled that RLRs (including RIG-I,
88 MDA5 and LGP2) are also involved in HSC formation through activating downstream
89 inflammatory signaling such as TNF receptor-associated factors (TRAFs) [36]. Typically,
90 these PRRs induce antiviral immune responses by recognizing single-stranded RNAs
91 (ssRNAs) or double-stranded RNAs (dsRNAs) produced by exogenous pathogens [37, 38].
92 During the formation of HSCs, the ligands for PRRs are puzzling because the AGM region
93 is supposed to be in a sterile niche. Nonetheless, transposable elements (TEs) abundantly
94 distributed throughout the genome may provide endogenous nucleic acids for PRRs [39].
95 Strikingly, the expression TEs has been detected during zebrafish EHT, and is
96 demonstrated to affect HSC generation through the RLR pathway [36]. However, as yet,
97 little is known about the contributions of TEs to human and mouse EHT.

98 TEs are consisting of retrotransposons and DNA transposons (DNAs). Retrotransposons
99 transpose by a copy-and-paste mechanism, whereby an RNA intermediate is reverse-
100 transcribed and then inserted into a new genomic locus. Most of the TEs in human and
101 mouse are retrotransposons, whether long interspersed nuclear elements (LINEs), short
102 interspersed nuclear elements (SINEs) or hominid SVAs (SINE-VNTR-Alu), and long
103 terminal repeats (LTRs), most of which also known as endogenous retroviruses (ERVs).
104 The jumping mechanism of TEs may induce genome instability, and uncontrolled
105 transposition can lead to disease [40, 41]. Therefore, a variety of defense systems have
106 evolved to domesticate TEs, including chromatin modification, small RNA silencing, and
107 post-transcriptional repression [42, 43]. In vertebrates, Krüppel-associated box zinc finger
108 protein (KRAB-ZFP) is one of the prominent TE silencing systems, which inhibits TEs

109 through interacting with KAP1 to recruit DNA methyltransferases (DNMT), SETDB1, HP1
110 and the nucleosome remodeling deacetylase (NuRD) complex [42, 44]. The human
111 silencing hub (HUSH) complex coupled with the ATPase MORC2 to deposit H3K9me3 for
112 de novo silencing of TEs. Although the activity of TEs in the genome is often silenced, they
113 can be activated in a temporary or tightly fashion both at transcriptional and epigenetic
114 levels to shape embryonic development [45, 46]. Studies at the single cell level show that
115 TEs have cell type-specific expression during gastrulation and organogenesis, and
116 participate in dynamic regulation of pluripotency reprogramming and lineage differentiation
117 [47-49]. There are also related evidences show that TEs can contribute to the
118 hematopoietic regeneration and fate decision [50-52], which may provide new insights on
119 therapies of certain hematopoietic diseases.

120 Despite progress in understanding the expression and potential RLR pathway of TEs
121 during zebrafish EHT, there is still limited exploration of the single-cell expression and
122 regulatory landscape of TEs, as well as the underlying mechanisms of TE activation in
123 human and mouse. In this study, we first performed a comprehensive survey of the
124 genomic landscape and potential regulatory functions of TEs in human and mouse, laying
125 a foundation for further analysis of TE activity in embryonic HSC development. We
126 reconstructed the human and mouse EHT trajectories using scRNA-seq datasets,
127 demonstrated the conservation of EHT cell types between species, and resolved the
128 dynamic expression patterns of TEs during EHT. We observed that most TEs presented
129 low cell type-specificity on the conserved EHT trajectories, but form a co-upregulation
130 pattern during pre-HEC specification. The activation of TEs in the pre-HEC stage was

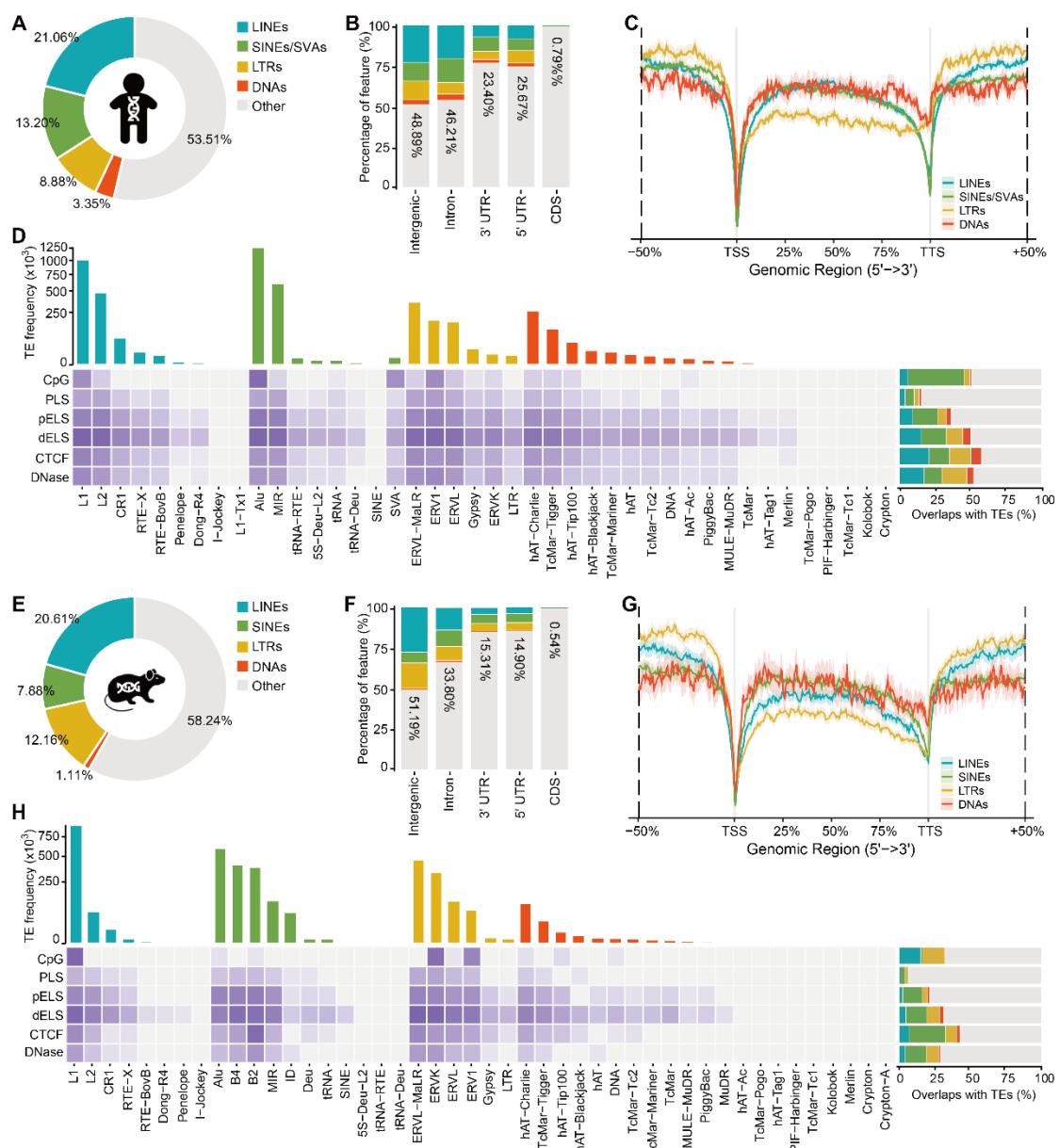
131 strongly associated with the transient relaxation of several TE silencing systems. The
132 delayed TE product sensing through various PRRs can trigger inflammatory signaling to
133 facilitate immune activation of HSCs. Importantly, these observations are highly conserved
134 in human and mouse. Additionally, by analyzing scATAC-seq data, we constructed the
135 hematopoietic cis-regulatory network of accessible TEs and identified two potential TE
136 derived enhancers, which may promote the expression of a specific pre-HEC marker
137 (Gja5). Furthermore, by combining the spatial transcriptome of human AGM and a bulk
138 RNA-seq of hypoxia response, we hypothesized that the hypoxic AGM niche may be
139 partially responsible for transient TE activation preceding hematopoietic fate commitment.
140 Taken together, our study provides a systematic single-cell investigation into the
141 contribution of TEs to the expression and regulatory landscape of EHT, which may shed
142 lights on studying TEs in the context of stem cell development and other cell type transition
143 systems.

144 **Results**

145 **Widespread TEs harbor great regulatory potential in human and mouse**

146 TEs in human and mouse genomes can be divided into 4 classes (LINEs, SINEs, LTRs
147 and DNAs), and the hominids SVAs (6 families) are included into SINEs for calculation.
148 These TEs can be further classified into 42 superfamilies (1176 families) and 41
149 superfamilies (1256 families) in human and mouse, respectively (Additional file 1: Table S1
150 and S2). Collectively, TEs account for about 46.38% and 41.76% of the human and mouse
151 genomes, respectively (Fig. 1A, E; Additional file 1: Table S3 and S4). The majority of TEs
152 are located in non-coding regions, including intergenic regions, introns and UTRs (Fig. 1B,

153 F; Additional file 1: Table S5 and S6). Interestingly, TEs are less distributed near the
154 transcription start sites (TSS) and transcription termination sites (TTS) (Fig. 1C, G), which
155 may be important for maintaining the specificity of gene transcription [53]. To explore the
156 regulatory potential of TEs, we calculated the copy number of each TE superfamily
157 overlapping with CpG islands and candidate cis-regulatory elements (cCREs, downloaded
158 from ENCODE-SCREEN [54]). In the human genome, SINEs/SVAs (including Alu and SVA)
159 contribute more than 39% of CpG islands (Fig. 1D; Additional file 2: Table S1 and S13), but
160 in mouse, SINEs only overlap with about 1.47% of CpG islands, while LINEs (especially
161 L1) and LTRs (ERV1 and ERVK) both contribute more than 15% of CpG islands (Fig. 1H;
162 Additional file 2: Table S2 and S14). In most cases, abundant CpG sites keep TEs
163 repressed in the methylated state. However, through demethylation processes such as
164 epigenetic reprogramming, it is possible for the TEs to be activated and play a role in
165 embryonic development [55, 56]. Among the cCREs, in addition to promoter-like sites
166 (PLS), a considerable proportion (36.39%-57.55% in human and 21.34%-42.95% in mouse)
167 of proximal enhancer-like signatures (pELS), distal enhancer-like signatures (dELS), CTCF
168 signatures and DNase-H3K4me3 signatures have intersections with TEs (Figure 1D, 1H;
169 Additional file 2: Table S3-S12). These evidences suggest that widely distributed TEs have
170 evolved enormous regulatory potential and may exert unique contributions in pluripotency
171 and early embryogenesis both at transcriptional and epigenetic levels [57-61]. In this study,
172 we will focus on analyzing the expression and chromatin accessibility of TEs during EHT,
173 expecting to reveal the potential regulatory mechanism of TE in the formation of HSCs.



174

175 **Figure 1. Genomic landscape of TEs in human and mouse. A, E Genome coverage of TEs**
 176 **in human and mouse. B, F Overlaps of TEs with gene structures in human and mouse. C,**
 177 **G Distribution of TEs along the gene body in human and mouse. D, H The frequency of TE**
 178 **superfamilies (upper bar plot) and overlaps of TEs with CpG islands and cis-regulatory**
 179 **elements (heatmap and right bar plot) in human and mouse.**

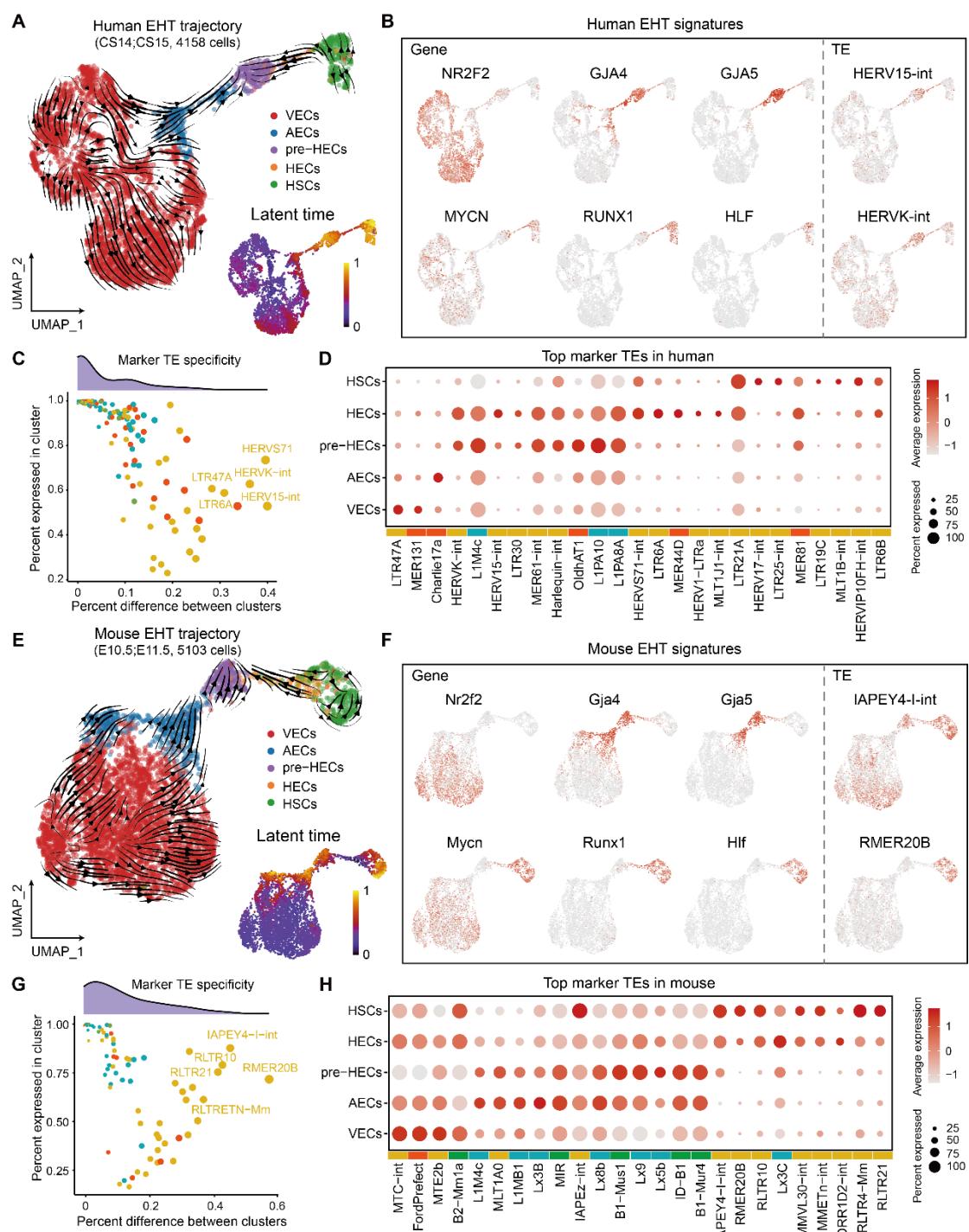
180 **TEs are detected with low cell type-specificity on the conserved EHT trajectories of**
 181 **human and mouse**

182 Accurate cell subtyping is the basis for studying various molecular regulatory activities
 183 during EHT. With the aid of published scRNA-seq data of the human and mouse AGM

184 (Additional file 3: Table S1) [22, 23], we reconstructed the EHT trajectories *de novo* by
185 analyzing RNA velocities (Fig. 2A, E). Interestingly, we observed aberrant orientation of
186 RNA velocity during the progression from pre-HECs to HSCs (which was also reflected in
187 the latent time) in both human and mouse, suggesting a differentiation bottleneck during
188 hematopoietic specification [22]. The cell type annotation of the AGM region and the
189 characterization of EHT cell subtypes were relied on the known marker genes and
190 signatures proposed in the human EHT study [23]. First, we isolated cell clusters of ECs
191 and HSCs from the AGM UMAP, and obtained EHT clusters through dimension reduction
192 and clustering. By using common marker genes of VECs (CDH5, NRP2, NR2F2), AECs
193 (GJA4, HEY1, DLL4), pre-HECs (TMEM100, GJA5, EDN1) and HSCs (RUNX1, MYB,
194 HLF), we accurately identified these four cell types within the EHT cell clusters (Fig. 2B, F;
195 Additional file 4: Fig. S1A-C, F and Fig. S2A-C, F). However, distinguishing HECs directly
196 from the cluster level can be challenging since they serve as transitional cells from pre-
197 HECs to HSCs. Therefore, HECs were selected based on the co-expression of CDH5,
198 RUNX1 and MYCN, along with the absence of PTPRC (Fig. 2B, F; Additional file 4: Fig.
199 S1D-F and Fig. S2D-F). After reconstructing the EHT trajectory, we integrated the human
200 and mouse EHT data based on the shared homologous genes. The results showed that
201 the EHT of both showed a highly conserved pattern, although a relatively larger number of
202 HECs were captured in mouse (Additional file 4: Fig. S3A, B). The majority of EHT marker
203 genes were found to be conserved in human and mouse. For instance, ACE is positive,
204 CD44 is low and KIT is negative in pre-HEC [21, 62]. However, there are also some
205 species-specific EHT marker genes, such as IL33 and SPINK2 only express in human pre-

206 HECs and HSCs, respectively, while in mouse, *Ikzf2* is more enriched HECs and HSCs
207 (Additional file 4: Fig. S3D).

208 To explore the dynamics of TEs on the EHT trajectories, we computed the family-level TE
209 expression to each EHT cell. Through differential expression analysis, we identified 214
210 and 96 marker TEs in human and mouse EHT cells (average $\log_{2}FC \geq 0.25$ and adjusted
211 P-value ≤ 0.05), respectively. Among these, 198 TEs (92.52%) in human were enriched in
212 pre-HECs, while 72 TEs (75%) in mouse were belonged to AECs and pre-HECs (Additional
213 file 5: Table S1 and Table S2). Notably, in human EHT, HERV15-int and HERVK-int
214 appeared to be enriched in pre-HECs and HECs, whereas in mouse EHT, IAPEY4-I-int
215 and RMER20B were highly expressed in HSCs (Fig. 2B, F). It should be noted that the
216 specificities of most marker TEs are not significant, that is, although TEs are enriched in
217 the target cell population, the differences in expression percentage compared with other
218 cell populations are mostly less than 0.2 (Fig. 2C, G). The top marker TEs for each cell
219 type of human and mouse EHT were displayed in Fig. 2D, H. Surprisingly, the marker TEs
220 that showed relatively higher specificities (percent difference > 0.25) are mostly ERVs,
221 which is consistent in both human and mouse. In particular, among the relatively highly
222 specific marker TEs in human EHT (Fig. 2C), primate-restricted HERVK transcripts have
223 been reported to be abundant in primordial germ cells, endodermal cells and blood
224 progenitors during human gastrulation, whereas HERV57 transcripts are frequently
225 detected in primitive streak, nascent mesoderm and definitive endoderm [48]. These
226 observations provide valuable insights into the regulatory roles of TEs to EHT and early
227 embryogenesis.



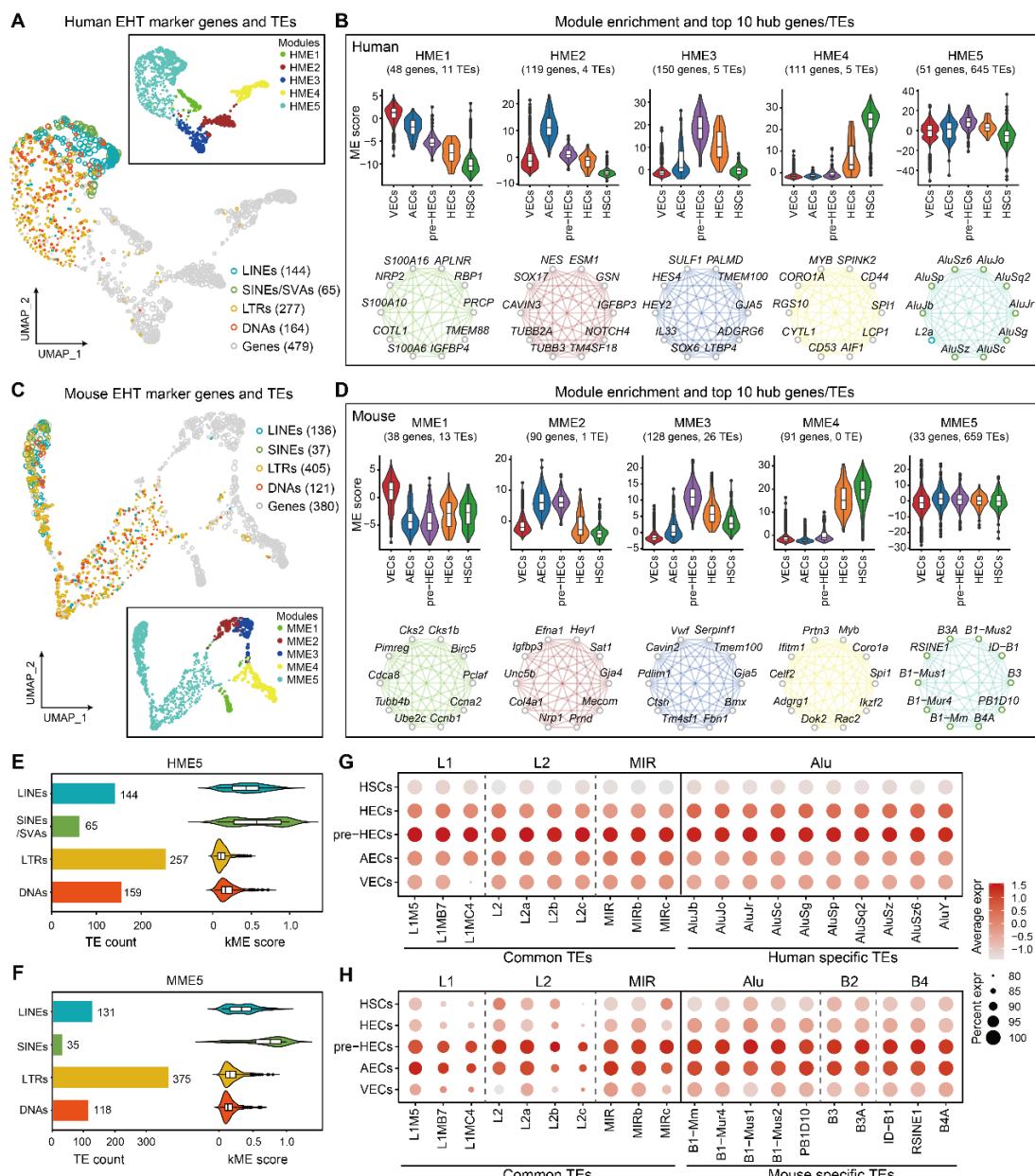
228

229 **Figure 2.** The EHT trajectories and TE expression overview in human and mouse. **A, E**
230 Human and mouse EHT trajectories and predicted latent time. **B, F** Human and mouse
231 EHT signatures. NR2F2 marks VECs, GJA4 marks AECs, GJA5 marks late AECs and pre-
232 HECs, MYCN and RUNX1 marks HECs and HSCs, HLF marks HSCs. These markers are
233 conserved between human and mouse. Only a few marker TEs were identified in the
234 analysis. **C, G** The specificities of marker TEs in human and mouse. The majority of TEs
235 with high cell type-specificities are LTRs. **D, H** Top markers of each cell type in human and
236 mouse EHT.

237 **TEs form a distinguished co-upregulation pattern during pre-HEC specification**

238 To identify modules that may participate in common regulatory processes during EHT, we
239 clustered the dynamic expression profiles of genes and TEs by co-expression network
240 analysis. A total of 988 filtered TEs, 528 marker genes (average log₂FC≥0.5 and adjusted
241 P-value≤0.05) in human and 864 filtered TEs, 421 marker genes in mouse were included
242 for co-expression analysis (Additional file 5: Table S3 and S4). Those selected genes and
243 TEs were clustered into 5 modules (HME1-5 and MME1-5) in both human and mouse, in
244 which HME1-4 (MME1-4) were enriched in VECs, AECs, pre-HECs and HECs/HSCs,
245 respectively (Fig. 3A-D; Additional file 6: Table S1 and S2). The top 10 hub genes/TEs of
246 each module were listed in Fig. 3B and Fig. 3D, characterizing the features that may play
247 a leading role in each module. Some conserved hub genes can be found in human and
248 mouse, such as GJA5 and TMEM100 are hub genes in pre-HECs, while MYB, SPI1 and
249 CORO1A are hub genes in HSCs. Interestingly, most TEs (645 of 670 in human and 659
250 of 699 in mouse) were clustered in HME5 and MME5 (Fig. 3B, D). The expression patterns
251 of these TEs were that, there was a consistent expression in almost the whole EHT process,
252 but specifically showed an upward trend in pre-HECs (appeared earlier in the AEC stage
253 in mouse). Among the TEs included in HME5 and MME5, although LTRs accounted for the
254 largest proportion (39.84% and 56.90% in human and mouse, respectively), their module
255 connectivity scores (kME scores) were relatively low (Fig. 3E, F). In contrast, SINEs and
256 LINEs exhibited higher kME scores in both HME5 and MME5. It is evidenced that
257 overexpression of a SINE copy (sine3-1a) can enhance HSC formation in zebrafish [36].
258 In particular, among TEs with kME scores greater than 0.3, L1 was the most abundant in

both human and mouse (94 in 190 and 68 in 179) (Additional file 6: Table S1 and S2). This observation could be attributed to the abundant CpG islands on the L1 elements, implying a potential reduction in methylation levels of TEs during the pre-HEC stage. In addition, 260 common TEs in human and mouse are found to be enriched in HME5 and MME5 (most with the top average kME scores are L1 elements) (Additional file 6: Table S3). Some top-ranked (according to kME scores) common TEs (L1, L2 and MIR) and species-specific TEs (Alu and mouse-specific B2 and B4) were displayed in Fig. 3G, H.



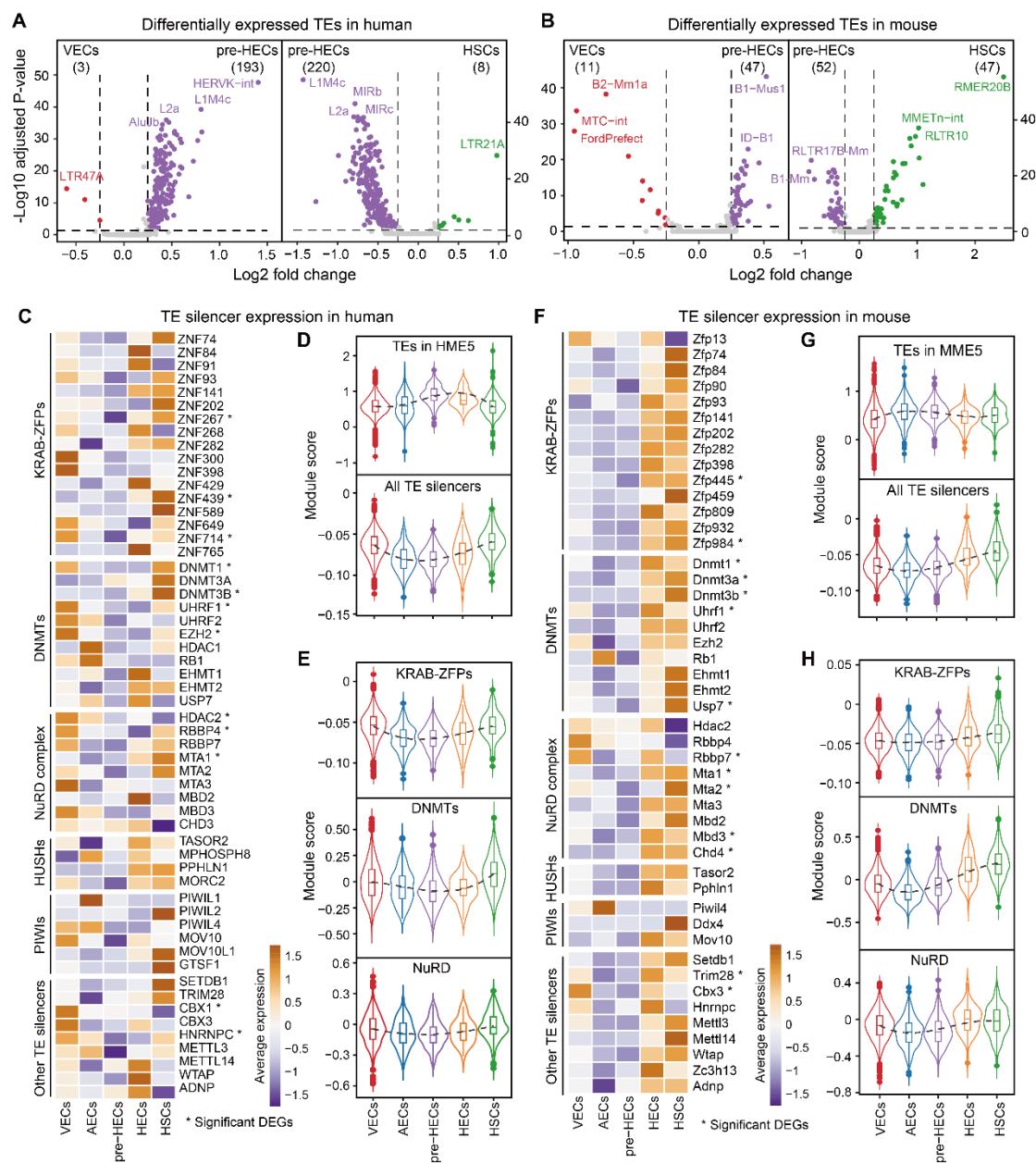
267 **Figure 3.** Co-expression network analysis of marker genes and expressed TEs. **A, C** Co-
268 expression modules of marker genes and TEs in human and mouse. Most TEs tend to
269 cluster together as distinct modules (HME5 and MME5). **B, D** The expression patterns and
270 top 10 hub genes/TEs of each module in human and mouse. TEs show a common
271 upregulation trend in pre-HECs. In the case of mouse, this upregulation appears to occur
272 even earlier, during the AEC stage. **E, F** TE composition (bar plot) and module connectivity
273 (kME, violin plot) of HME5 and MME5. **G, H** Dot plots show the expression levels of
274 selected common and species-specific TEs.

275 **TE silencers are transiently relaxed in pre-HECs**

276 We performed differential expression analysis on pre-HECs against VECs and HSCs, and
277 the results showed that TEs were pervasively upregulated in pre-HECs in human (Fig. 4A;
278 Additional file 7: Table S1 and S3), while this phenomenon was relatively insignificant in
279 mouse (Fig. 4B), but the upregulated TEs in pre-HECs were still more than other cell types
280 (Additional file 7: Table S5 and S7). This is consistent with our observation of TE-enriched
281 modules (HME5 and MME5) upregulated at the pre-HEC stage in co-expression network
282 analysis (Fig. 3B, D). However, in mouse, it seems that there is already a large-scale TE
283 activation in AECs, which may be related to the presence of more pre-HEC-primed cells in
284 AECs (Fig. 3D). It can also be seen that the fold changes of differentially expressed TEs
285 were generally low, which also explained why we identified few cell type-specific marker
286 TEs during EHT (Fig. 2C, G). Since TEs are normally repressed in most cases, it is
287 reasonable to assume that the transient activation of TEs in pre-HECs is due to the
288 downregulation of TE silencers. In fact, vertebrates have evolved a complex set of
289 epigenetic modules to control TEs, such as KRAB-ZFPs, DNA methylation, small RNAs
290 (piRNAs), and histone modifications [42]. We therefore systematically screened the
291 expression patterns of these TE silencing systems during EHT. Surprisingly, the majority
292 of TE silencers were downregulated in pre-HECs against VECs and HSCs (for example,

293 84.07% and 77.49% of KRAB-ZFPs in human and mouse, respectively) (Fig. 4C, F;
294 Additional file 7: Table S9 and S10; KRAB-ZFP genes are available from [63]). Among the
295 downregulated KRAB-ZPFs, ZNF84, ZNF382 and ZNF429 were found to bind significantly
296 to L1 superfamily in human [64]. In embryonic stem cells, ZNF91 and ZNF93 can
297 respectively repress SVAs and L1 in human [65], while Zfp932 regulates ERVK in mouse.
298 In human, there is also evidence that ZNF268, ZNF300 and ZNF589 are related to
299 hematopoietic differentiation [66]. It can be observed that some co-factors recruited by
300 KRAB-ZFPs, such as TRIM28 (KAP1), CBX3 (HP1) and SETDB1, also exhibited relatively
301 low expression levels in pre-HECs (Fig. 4C, F). TE silencers closely related to KRAB-ZFPs,
302 such as DNMTs and NuRD complex also showed low expression in pre-HECs (Fig. 4E, H).
303 In addition, although the overall expression levels were low, the HUSH complex (HUSHs),
304 P-element induced Wimpy testis-related genes (PIWIs) and some other TE silencers were
305 also expressed relatively lower in pre-HECs than in other cell types (Fig. 4C, F). Therefore,
306 it can be inferred that various TE silencers were relaxed by specific mechanisms, leading
307 to transient activation of TEs during pre-HEC specification (Fig. 4D, G). Interestingly, after
308 the pre-HEC stage, those TE silencers were upregulated to re-suppress the TE activity.
309 This also explains why some members of the DNMT complex (such as DNMT1 [67] and
310 EZH2 [68]) and NuRD complex (such as HDAC1 and HDAC2 [69]) are required for HSC
311 formation [70].

312



313

314 **Figure 4.** Expression of known TE silencers in human and mouse. **A, B** Differential
315 expression of TEs in pre-HECs versus VECs and HSCs in human and mouse. **C, F**
316 Expression heatmap of known TE silencing systems in human and mouse, including
317 KRAB-ZFPs, DNMTs, NuRD complex, HUSHs, PIWIs and other TE silencers. **D, G**
318 Expression trend of TEs (HME5 and MME5, $kME \geq 0.3$) and all TE silencers. **E, H**
319 Expression trend of specific TE silencing systems in human and mouse.

320

TE product sensing facilitates immune activation during HSC orientation

321

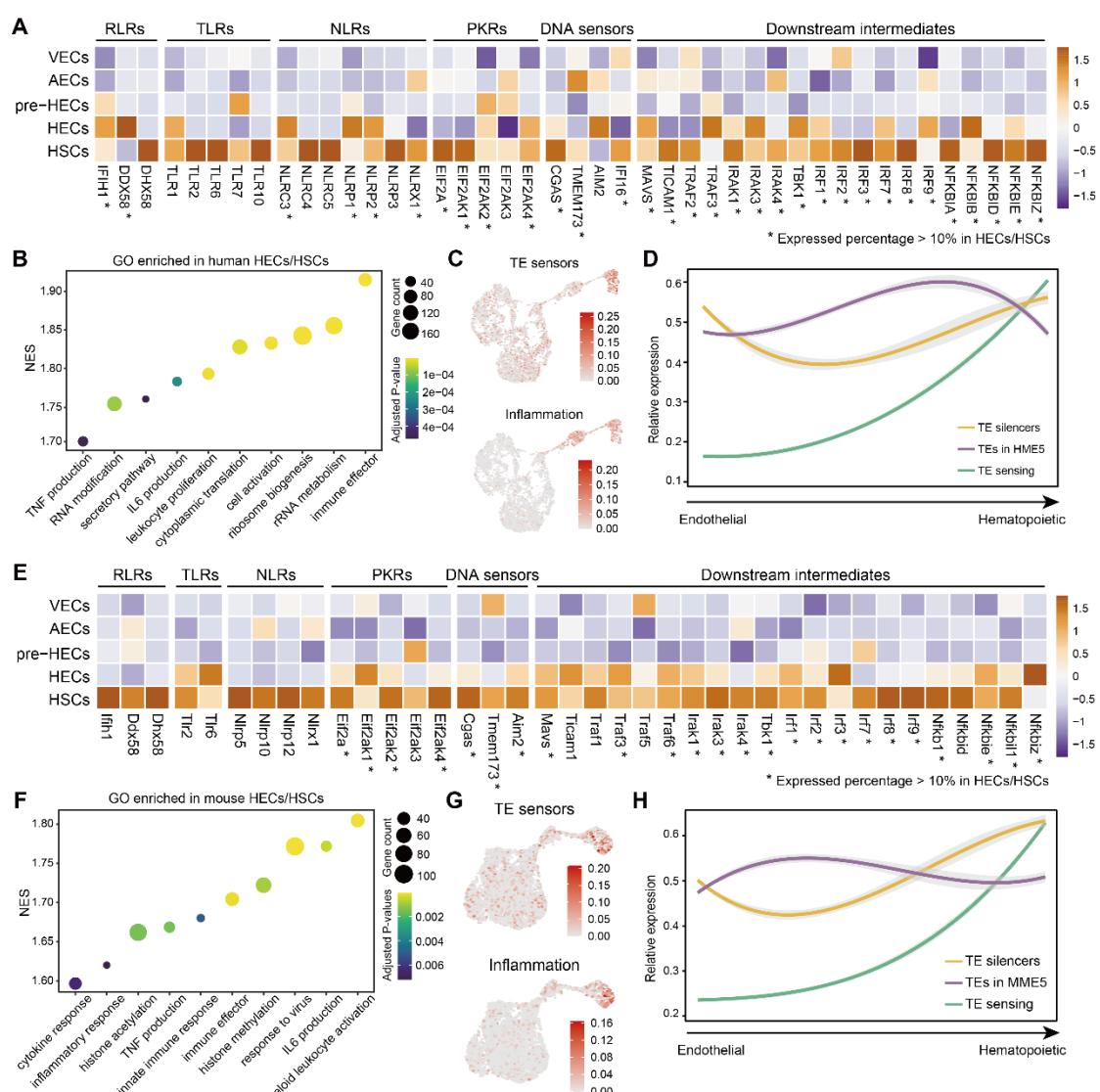
It was found that massive TEs were activated at the pre-HEC stage, but the contribution of

322

these TEs to the human and mouse EHT was unknown. Studies have shown that TEs are

323 the main source of endogenous RNAs or cDNAs [39, 42]. The products transcribed from
324 TEs are likely to activate downstream inflammatory and immune signaling pathways
325 through pattern recognition receptors. We investigated genes that were more expressed
326 in HECs/HSCs than in endothelial cells, and identified upregulation of a large number of
327 RNA and DNA sensors both in human and mouse (Fig. 5A, E; Additional file 8: Table S1
328 and S2). IFIH1 (MDA5), DDX58 (RIG-I) of RLRs and NLRP1, NLRP2, NLRC3 and NLRX1
329 of NLRs seemed to be significantly upregulated in human HECs/HSCs. Although a few
330 members of TLRs showed an upregulation trend in HECs/HSCs, the expression levels
331 were low in both human and mouse (expression percentage is less than 10%). In addition,
332 protein kinase R genes (PKRs) including EIF2A, EIF2AK1, EIF2AK2 and EIF2AK4 were
333 both significantly upregulated in human and mouse HECs/HSCs. DNA sensors like the
334 cGAS/STING (TMEM173) signals also appeared to be elevated in human and mouse
335 HECs/HSCs, which were possibly activated by cDNA intermediates of retrotransposons
336 (LINEs, SINEs and ERVs) [71, 72]. Typical downstream intermediates of RLRs and
337 cGAS/STING, such as MAVS, TRAF3, TBK1, IRF3, and NF- κ B, all showed highly
338 conserved upregulation patterns in human and mouse HECs/HSCs (Fig. 5A, E; Additional
339 file 8: Table S3 and S4). By functional enrichment analysis, both interferon alpha (IFN α)
340 and interferon gamma (IFN γ) response pathways were detected to be enriched in human
341 and mouse HECs/HSCs, confirming the activation of these PRRs (Additional file 8: Table
342 S5 and S6). Interestingly, IFNAR1 and IFNGR2 consistently appeared to function earlier in
343 human and mouse (immediately after TE expression upregulation), while IFNAR2 and
344 IFNGR1 showed complementary patterns (Additional file 8: Table S3 and S4). Further gene

345 set enrichment analysis (GSEA) on gene ontology (GO) showed that inflammatory signals
 346 (such as TNF and IL6) and immune response were enriched in human and mouse
 347 HECs/HSCs (Fig. 5B, F; Additional file 8: Table S5 and S6). Taking the above evidence
 348 together, we speculated that TE products (pervasively elevated in pre-HECs or earlier in
 349 partial AECs) could induce inflammatory signals through various PRRs during EHT, and
 350 trigger immune response pathways to activate HSC progression (Fig. 5C, G and D, H).



351
 352 **Figure 5. Expression of common TE sensors and functional enrichment for HECs/HSCs.**
 353 **A, E** Expression heatmap of common TE sensors in human and mouse, including RLRs,
 354 TLRs, NLRs, PKRs, DNA sensors and downstream intermediates. **B, F** Gene set
 355 enrichment analysis of GO terms in human and mouse HECs/HSCs. **C, G** Module scores
 356 of TE sensors and inflammatory genes in human and mouse. **D, H** Expression trends of

357 TE silencers, TEs (HME5 and MME5, kME \geqslant 0.3) and TE sensing genes during human and
358 mouse EHT. The expression pattern of TEs is opposite to that of TE silencers, whereas TE
359 sensors are less active until the HSC stage.

360 **TE accessibility is dynamically controlled during EHT**

361 TEs are known to play important roles in development and disease processes as cis-
362 regulatory elements [59]. To explore the potential cis-regulatory function of TEs on HSC
363 origination, we systematically analyzed the scATAC-seq data in the E10.5 mouse AGM.

364 Using cell types transferred from scRNA-seq data, a coherent EHT process can still be
365 achieved based on scATAC-seq data (Fig. 6A; Additional file 9: Fig. S1A-D). The gene
366 activities of specific EHT signatures obtained from scRNA-seq were well fitted to the
367 developmental stage of EHT cell clusters (Fig. 6B; Additional file 9: Fig. S1E). For example,
368 Gja5 showed pre-HEC-specific high activity on the UMAP embedding of scATAC-seq data.

369 However, HECs were not able to be distinguished by scATAC-seq data, possibly due to
370 the small number of captured cells. Next, we calculated the TE activities in each cell
371 (reflecting the degree of TE accessibility) at the locus level. By applying differential
372 accessibility analysis, we surprisingly found that TEs were more accessible in pre-HECs

373 compared to endothelial cells (Fig. 6C; Additional file 10: Table S1), which aligned with our
374 previous finding that TE expression is generally elevated in pre-HECs (Fig. 4A, B). While

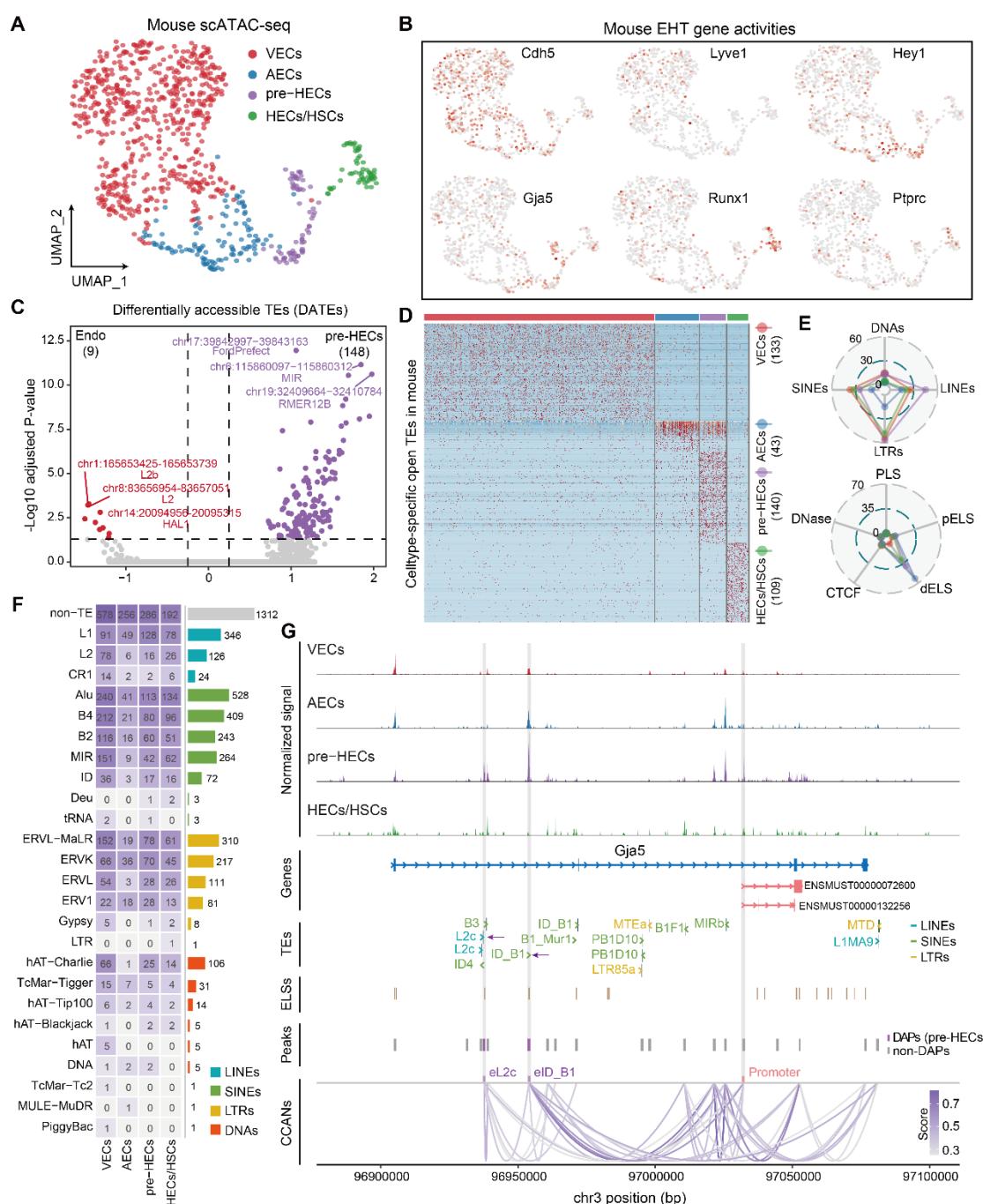
375 a total of 148 differentially accessible TEs (DATEs, average log2FC \geqslant 0.25 & adjusted P-
376 value \leqslant 0.05) were enriched in pre-HECs (Fig. 6C), only a few DATEs were identified
377 between pre-HECs and HECs/HSCs (Additional file 10: Table S2). Notably, when
378 differential accessibility analysis was performed on all peaks, we also found more
379 differentially accessible peaks (DAPs) in pre-HECs compared to endothelial cells. This

380 indicated that chromatin may undergo general regulation of accessibility, i.e., chromatin
381 reprogramming, during the pre-HEC stage. In addition, TE accessibility did not change
382 significantly when pre-HECs entered the HSC-committed stage, suggesting that the
383 observed downregulation of TE expression in HECs/HSCs may not be influenced by
384 chromatin accessibility. Unlike the low specificity of TE expression in scRNA-seq data, we
385 identified a considerable number of cell type-specific open TEs in scATAC-seq data (Fig.
386 6D; Additional file 10: Table S3). Among them, AECs had the fewest specific open TEs (43),
387 16 of which overlapped with pre-HEC. DNAs accounted for the least amount of these cell
388 type-specific open TEs, and LINEs accounted for more in pre-HECs than in other cell types
389 (Fig. 6E). As expected, the majority of these open TEs played roles as distal enhancers.

390 Considering that TEs may function together with other accessible regions, we re-performed
391 differential accessibility analysis for all peaks and identified a total of 4,230 cell type-
392 specific DAPs, of which 2,918 overlapped with TEs, named TEPs (Fig. 6F; Additional file
393 10: Table S4). It can be seen that SINEs (especially Alu superfamily) exhibit relatively
394 higher regulatory potential in each cell type. The gene regions closest to these TEPs
395 contain many EHT-associated signatures, such as Gja4 (AECs), Gja5, Edn1 (pre-HECs),
396 and Gata2, Cd44, Runx1 (HECs/HSCs) (Additional file 9: Fig. S2; Additional file 10: Table
397 S4). Gja5 is a member of the connexin gene family, which had elevated expression in late
398 AECs and pre-HECs (Fig. 2B, F). It can be noticed that the promoter of two transcripts
399 (ENSMUST00000072600 and ENSMUST00000132256, annotated in EPD [73]) of Gja5
400 was specifically more accessible in pre-HECs (Fig. 6G), while two upstream promoter-like
401 peaks showed high accessibility in both AECs and pre-HECs, which may permit the earlier

402 expression of Gja5 observed in AECs. In addition, we found that two TEs
403 (chr3:96937220–96937315-L2c and chr3:96954448–96954510: ID_B1) inside the gene
404 body of Gja5 may function as enhancers (termed eL2c and eID_B1) to promote Gja5
405 expression in pre-HECs. The regions where these two TEs located are also annotated as
406 ELSs in ENCODE. We applied Cicero [74] to predict the cis-co-accessibility networks
407 (CCANs) among peaks detected near or inside Gja5. Although the potential enhancer
408 eID_B1 had the greatest increase in accessibility in pre-HECs, it was also open in AECs
409 and may interact with the two upstream promoter-like regions (Fig. 6G). Interestingly, the
410 potential enhancer eL2c was only opened in pre-HECs, consistent with the accessibility
411 pattern of the proximal promoter, and thus could be more likely to cooperatively increase
412 the expression of Gja5. However, why Gja5 is upregulated in the pre-HEC stage of EHT
413 remains to be further explored, although a recent study has pointed out its importance for
414 HSCs to dampen oxidative stress [75].

415



416

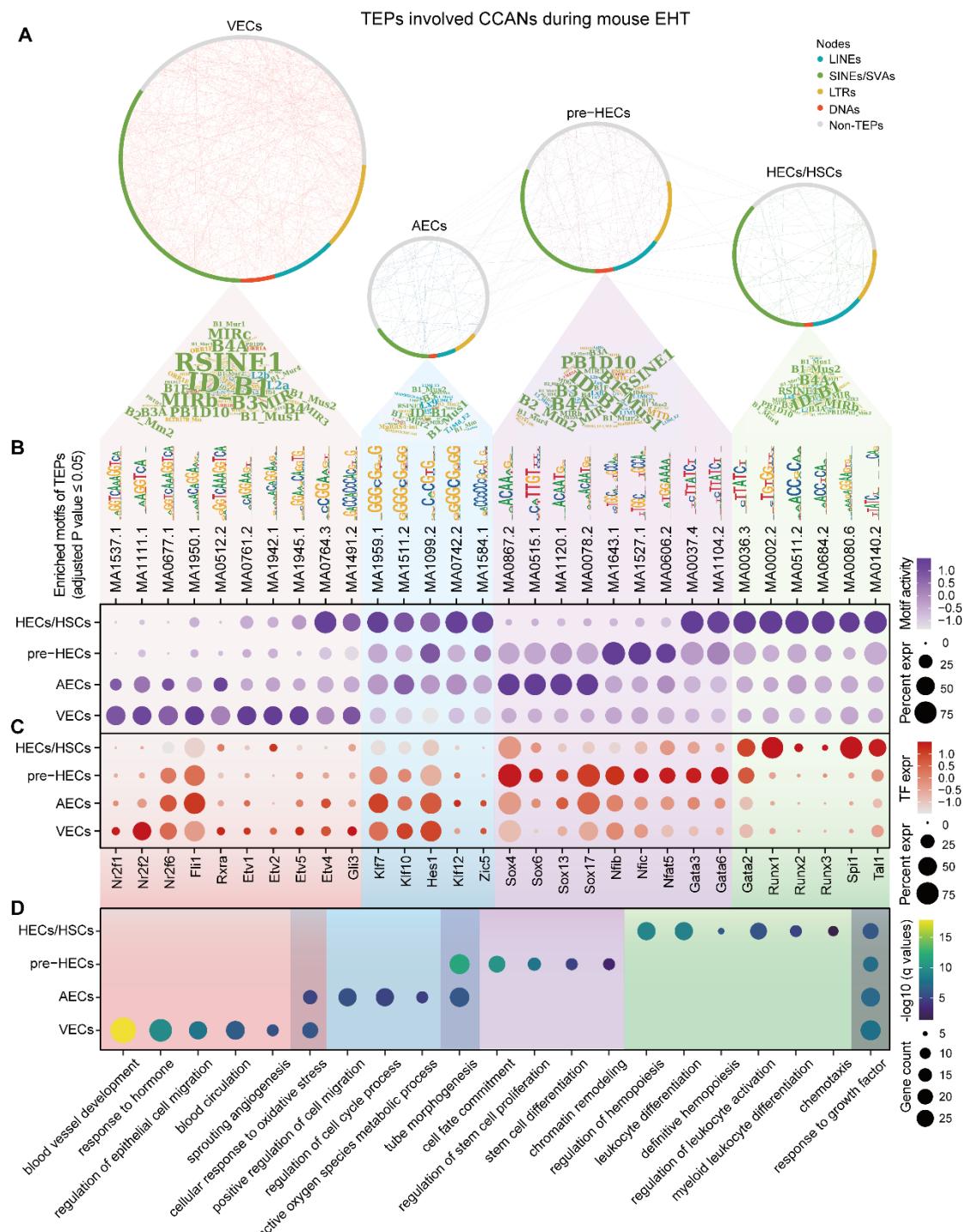
417 **Figure 6.** Mouse AGM scATAC-seq analysis and accessible TE identification. **A** Cell types
418 recovered from mouse scATAC-seq data. **B** UMAP of EHT marker gene activities. Cdh5
419 marks endothelial cells, Lyve1 marks VECs, Hey1 marks AECs, Gja5 marks pre-HECs and
420 partial AECs, Runx1 marks HECs/HSCs and Ptprc marks HSCs. **C** Differential accessible
421 analysis of TEs in pre-HECs versus endothelial clusters (VECs and AECs). **D** Cell type-
422 specific open TEs identified in mouse EHT. **E** TE composition and potential cis-regulatory
423 functions of cell type-specific open TEs. **F** Overlaps of all cell type-specific accessible
424 peaks with different TE superfamilies. **G** Tracks of normalized signals in each cell type,
425 genes, TEs, ELSSs, peaks and CCANs. Two potential enhancers derived by TEs and the
426 promoter of Gja5 transcripts are highlighted with light grey bars.

427 **Cell type-specific accessible TEs shape hematopoietic cis-regulatory networks**

428 It has been found that some TEs can act as enhancers to drive the expression of some
429 hematopoietic-related genes, but the co-regulation mode of these TEs and the
430 corresponding biological functions in the EHT process remain unclear. We used Cicero to
431 construct the CCANs of all cell type-specific DAPs (including TEPs and non-TEPs) by
432 filtering out links with co-access score less than 0.4 (Fig. 7A; Additional file 10: Table S5).
433 Analysis of TE compositions of cell type-specific TEPs revealed that ID_B1 (Alu
434 superfamily) was abundant in all cell types. The top-ranked TEs seemed to have high
435 consistency across cell clusters, but they were enriched to different motifs in different EHT
436 stages (Fig. 7B; Additional file 10: Table S6), which may be related to the variation
437 accumulated on different copies of TEs during evolution [76, 77]. Surprisingly, TEs almost
438 participate in shaping all cis-regulatory networks closely related to EHT process. For
439 example, SOX and GATA binding sites were mostly open in VECs and RUNX binding sites
440 gained increased accessibility in HECs/HSCs. We conducted a joint analysis of the
441 enriched motifs and the corresponding transcription factors (TFs) (Fig. 7B, C), and found
442 that although the motifs such as KLF (Klf7, Klf10 and Klf12) were active in AECs and later
443 stages, the expression of these TFs were downregulated to control the developmental fate
444 of AECs. In pre-HECs, the SOX motifs significantly increased the activity in AECs in
445 advance, but the expression of TFs (Sox4, Sox6, Sox13 and Sox17) peaked after entering
446 the pre-HEC stage. Similar to KLF, Gata3 and Gata6 had higher motif activities in both pre-
447 HECs and HECs/HSCs, but were only highly expressed in pre-HECs. This dual regulation

448 via motif binding activity and TF expression precisely shapes the lineage determination
449 and functional specification during EHT.

450 To explore the function of the TFs bound by TEPs, we predicted the cell type-specific TF-
451 target network based on interactions from TRRUST [78]. Surprisingly, some EHT
452 signatures were involved downstream of those TE-bound TFs (Additional file 9: Fig. S3;
453 Additional file 10: Table S7), such as Kdr, Flt1 (VECs), Smad6, Vegfc (pre-HECs) and Kit,
454 Ikzf1 (HECs/HSCs). The results of GO enrichment analysis also showed that these cis-
455 regulatory networks shaped by cell type-specific TEs were enriched in various important
456 functional modules during EHT (Fig. 7D; Additional file 10: Table S8), such as blood vessel
457 development and tube morphogenesis in VECs and AECs, stem cell differentiation and
458 chromatin reprogramming in pre-HECs, and regulation of hematopoiesis in HECs and
459 HSCs.



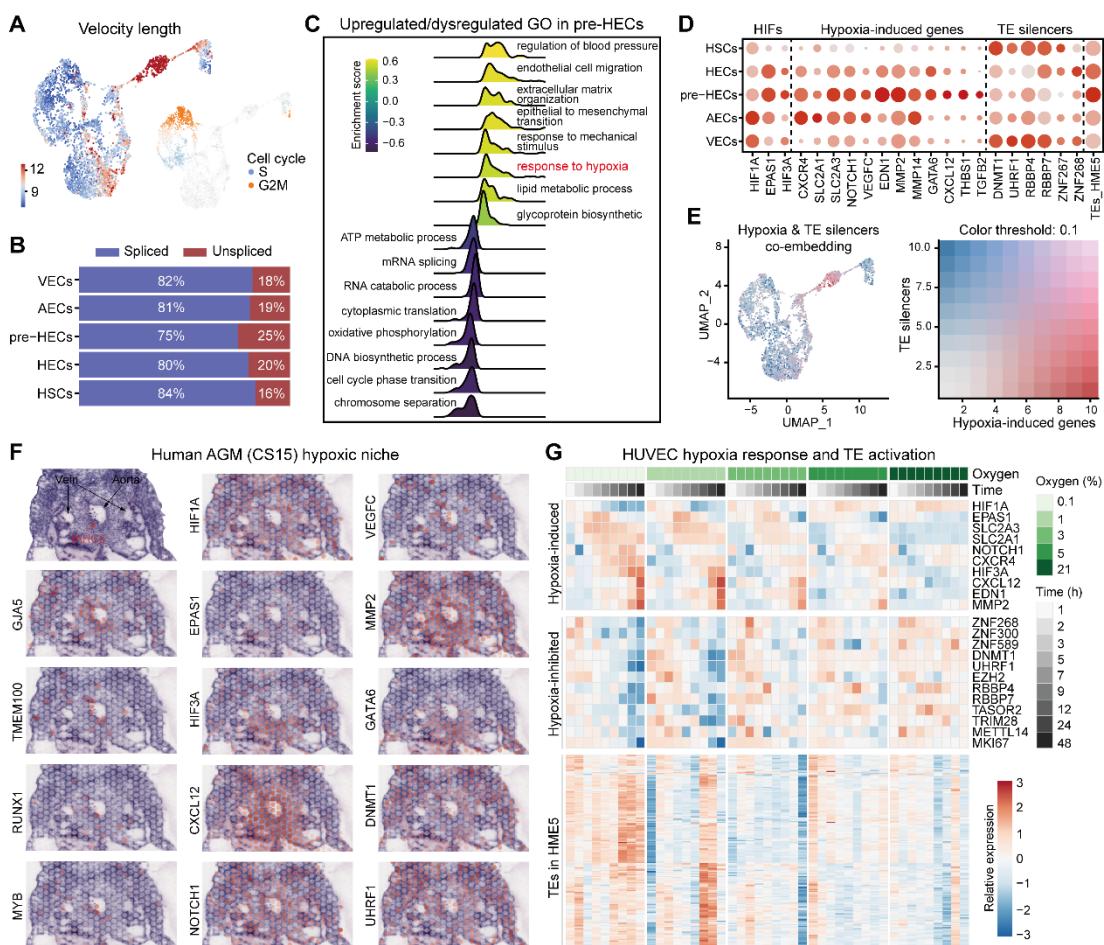
460

461 **Figure 7. Cis-regulatory network analysis of TE-associated accessible peaks. A** Cell type
462 CCANs predicted by Cicero. TE families in each cell type are displayed as word clouds. **B** Enriched
463 motifs of TE-associated accessible peaks in each cell type. **C** TF expression
464 corresponding to motifs in (B). The expression data is extracted from scRNA-seq data. **D** GO
465 enrichment of TF and target genes in each cell type. Overlaps are GO terms common
466 to two or more cell types.

467 **The hypoxic AGM niche may be partially responsible for the transient TE activation
468 preceding hematopoietic fate commitment**

469 The downregulation of TE silencing systems in pre-HECs may be the main reason for the
470 enhanced TE activity, however the underlying mechanisms regulating these TE silencers
471 remain unclear. Therefore, we focused on analyzing the specific gene expression and
472 functional status of pre-HECs. The RNA velocity map showed that pre-HECs had a
473 prominently high differentiation rate and are not in the cell cycle (Fig. 8A). In addition, more
474 unspliced RNAs were found in pre-HECs (Fig. 8B), explaining the disordered direction of
475 RNA velocity around pre-HECs (Fig. 2A). GSEA analysis showed that mRNA splicing, RNA
476 catabolic process and cell cycle were downregulated in the pre-HEC stage (Fig. 8C;
477 Additional file 11: Table S1). The downregulation of oxidative phosphorylation and
478 upregulation of lipid metabolic process suggested that pre-HECs may undergo metabolic
479 reprogramming [62]. Strikingly, genes related to epithelial to mesenchymal transition (EMT)
480 and response to hypoxia were upregulated in pre-HECs. In cancers, hypoxia has been
481 widely introduced to induce the EMT process and promote tumor cell metastasis [79, 80].
482 In zebrafish, hypoxia has been shown to strongly promote HSC formation through hypoxia-
483 inducible factors (HIFs, hif-1a and hif-2a) and Notch signaling [81], which coincides with a
484 study that induced human embryonic stem cells (hESCs) towards HSPC-like cells through
485 a hypoxia differentiation system *in vitro* [82]. By staining the mouse embryo (E10) with a
486 hypoxia indictor Pimonidazole (hypoxyprobe), it was also directly observed that the IAHC
487 cluster region was hypoxic [83, 84]. We analyzed hypoxia-related genes on human EHT
488 and found that EPAS1 (also known as HIF2A) and HIF3A [85] were highly expressed on

489 pre-HECs, while HIF1A was more expressed on VECs and AECs (Fig. 8D; Additional file
490 11: Table S2). Many hypoxia-induced downstream genes were also found to be enriched
491 in pre-HECs, such as SLC2A3 [86], CXCL12/CXCR4 [87], NOTCH1, VEGFC, EDN1,
492 MMP2/MMP14, GATA6 TGFB2 and THBS1, etc. Spatial transcriptome analysis of human
493 embryo (CS15) demonstrated that the expression of the above hypoxia-induced genes
494 was also enriched in the adjacent region of IAHCs (Fig. 8F). Furthermore, we observed
495 that the expression patterns of hypoxia-induced genes were exactly opposite to those of
496 TE silencers (Fig. 8D, E), especially DNMT1 and UHRF1 (Fig. 8D, F). However, to the best
497 of our knowledge, it has not been explored whether hypoxia can induce TE activation,
498 although it is widely recognized that TEs play an important role in stress response [50, 88,
499 89]. Therefore, we first recalculated the TE expression landscape in the human AGM
500 dataset. The results indicated that there seemed to be some other local hypoxic areas
501 around AGM besides the pre-HECs, including the stromal cells, which also exhibited higher
502 TE expression levels (Additional file 12: Fig. S1A-C). Few cell type-specific TEs were
503 identified for each cell type in the AGM region (Additional file 12: Fig. SD), which is
504 consistent with findings during the EHT trajectory (Fig. 2C, G). Besides, it was noticed that
505 the expression pattern of HIF3A was closer to that of TE in various cell types in the AGM
506 region (Additional file 12: Fig. S1B). HIF1A expressed in various cell types, whereas EPAS1
507 was enriched in endothelial cells (Additional file 12: Fig. S1E). Both stromal cells and
508 epithelial cells had a certain degree of HIF3A expression, which could be part of the reason
509 for their relatively higher TE expression levels (Additional file 12: Fig. S1A, B).



510
511 **Figure 8.** Hypoxia state analysis of pre-HECs and the AGM niche. **A** Velocity length and
512 cell cycle scores on the human EHT UMAP. **B** The proportions of spliced and unspliced
513 RNAs in each cell type. **C** Gene set enrichment analysis of GO terms in human pre-HECs
514 **D** Expression of hypoxia-related genes and TE silencers in human EHT. **E** Co-embedding
515 of expression of hypoxia-related genes and TE silencers. The expression patterns of the
516 two seem to be opposite. **F** Spatial expression of hypoxia-related genes in the human AGM.
517 **G** Expression heatmap of hypoxia-related genes, potential hypoxia-inhibited genes (TE
518 silencers) and TEs in HME5 ($kME \geq 0.3$).

519 Furthermore, here we included a comprehensive time-series RNA-seq study [90]
520 examining the hypoxia response of human umbilical vein endothelial cells (HUVECs) to
521 confirm the relationship between hypoxia and TE activation. Excitingly, it is observed that
522 TEs began to be broadly activated after 12 hours under extremely low oxygen
523 concentrations (0.1% and 1%), whereas no significant upregulation of TEs was observed
524 in groups with oxygen concentrations greater than 3% (Fig. 8G; Additional file 11: Table

525 S3). Different TE classes showed similar upregulation patterns under hypoxia condition
526 (Additional file 12: Fig. S2). Coincidentally, many TE silencers (such as KRAB-ZFP
527 members ZNF268, ZNF300, ZNF589) were greatly downregulated after 12 hours of
528 hypoxic culture. Interestingly, HIF3A still appears to be more correlated with TE expression
529 patterns than HIF1A and EPAS1 (Fig. 8G). Collectively, we hypothesized that the hypoxic
530 AGM niche might induce transient TE activation in pre-HECs by inhibiting the expression
531 of TE silencers, which is postulated to be critical for the EHT process (Fig. 5E, J). By
532 analyzing the expression of pre-HEC-specific markers in HUVEC data, it is observed that
533 SOX17, HEY1 and HEY2 were not upregulated under hypoxia induction (Additional file 12:
534 Fig. S2E, Fig. S3), suggesting that these genes may play distinct roles during pre-HEC
535 specification.

536 **Discussion**

537 TEs are abundant in the eukaryotic genomes and mounting evidence suggests that they
538 have evolved essential roles in transcriptional and epigenetic regulation [42, 43, 57, 59,
539 76]. Recent single cell sequencing technologies to characterize the transcriptomes and
540 epigenomes have revealed the broad expression and crucial roles of TEs in the developing
541 embryos [47-49]. At present, although TEs have been revealed to be specifically expressed
542 during definitive hematopoiesis and HSC regeneration [36, 50, 51], the underlying
543 mechanisms of TE activation have not yet been elucidated. In addition, TEs are known to
544 harbor a wide range of binding sites with regulatory potential (Fig. 1D, H), but their cis-
545 regulatory roles during the EHT process remains to be investigated. In this work, we
546 conducted a comprehensive analysis to understand the potential functionality of TEs during

547 the definitive hematopoiesis in human and mouse at the single cell resolution. We
548 demonstrated how cells conservatively program the EHT process and drive HSC formation
549 by dynamically regulating the expression and chromatin accessibility of TEs. Finally, we
550 deduced that the local hypoxic niche in AGM might be one of the important factors for the
551 unique developmental state and TE activation in pre-HECs.

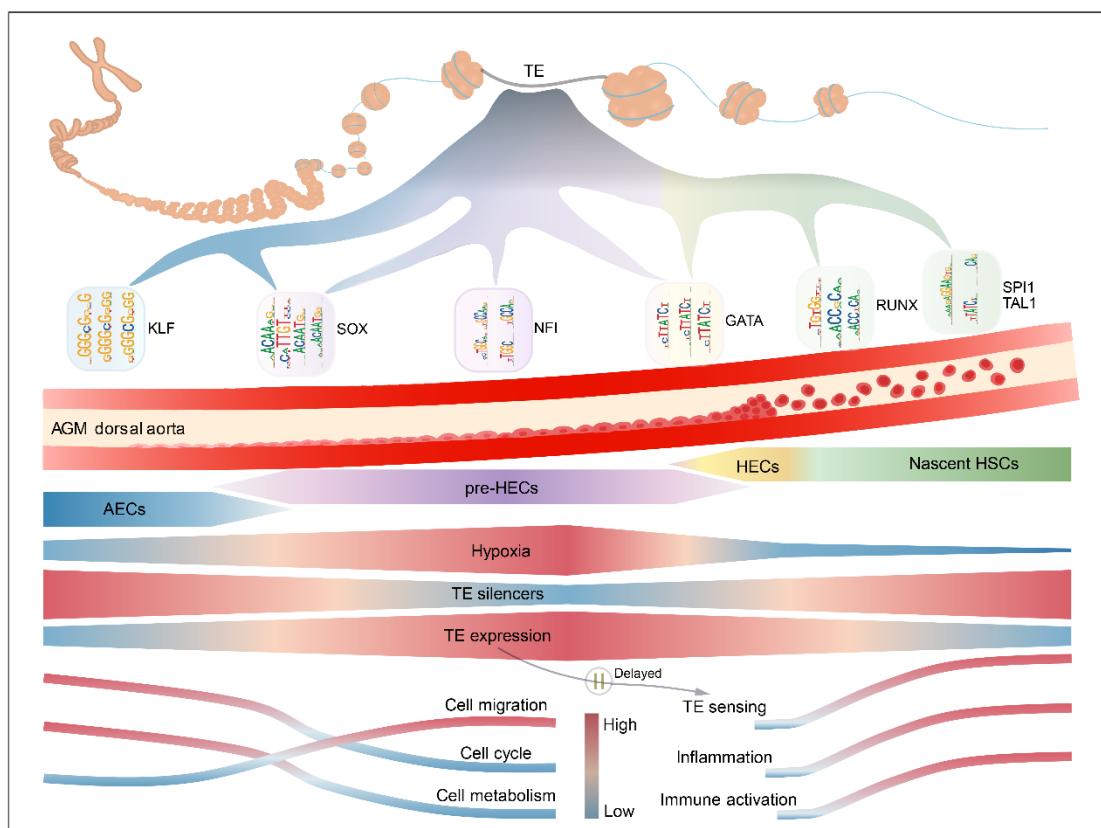
552 Leveraging the public single cell datasets of human and mouse AGM [22, 23], we
553 reconstructed the EHT trajectories and presented the landscape of dynamic TE expression.

554 Similar to the study in zebrafish [36], while only a few cell type-specific TEs (mainly LTRs)
555 were identified during EHT (Fig. 2C, G), we unexpectedly observed that clusters of TEs
556 were consistently upregulated during pre-HEC specification (Fig. 3B; Fig. 4A, B). The
557 upregulation of TEs in mouse seems to occur already in the late stage of AECs (Fig. 3D).

558 Coincidentally, TE silencing systems (including KRAB-ZFPs, DNMTs, NuRD complex, and
559 HUSHs, etc.) [42, 44, 64, 66] were at relatively low levels from AECs to pre-HECs (Fig. 4C-
560 H), which could at least partially account for the principle of TE activation in this period.
561 Interestingly, two RNA transferases METTL3 and METTL14, which can form nuclear
562 complex and control TE activity through m6A modification [91, 92], were also
563 downregulated in pre-HECs.

564 Previous studies have shown that TE products can activate inflammatory signals through
565 RLRs and promote HSC formation [36] and regeneration [50]. In contrast, we screened the
566 common PRRs exhaustedly and found that except RLRs, many other PRRs were also
567 upregulated in HECs/HSCs (Fig. 5A, E). For example, NLR family member X1 (NLRX1)

568 [93] is conservatively enriched in both human and mouse HSCs, but its role in
569 developmental hematopoiesis has not been described. DNA receptors (such as
570 cGAS/STING) [72] also appeared to be activated after the pre-HEC stage, suggesting the
571 possible presence of dsDNA from TEs or mitochondria. Many factors mediating these TE
572 sensors and downstream inflammatory signaling were also specifically turned on in
573 HECs/HSCs. It should be pointed out that dsRNAs from other sources, or some non-coding
574 RNA, can also activate these PRRs [36, 39]. Notably, TE activation and TE sensing are not
575 synchronized, that is, TEs that are largely transcribed in pre-HECs do not activate PRRs
576 and inflammatory signals until HECs/HSCs. A possible explanation for this might be that
577 the RNA catabolic process and cell metabolism are quiescent in pre-HECs (Fig. 8B, C),
578 which delays the TE sensing (Fig. 9). Such TE activation and sensing is conserved in
579 human and mouse, which resembles a rehearsal mechanism during the EHT process. After
580 hematopoietic differentiation, nascent HSCs learn antigen-like properties from
581 endogenous nucleic acid repertoire (e.g. TEs) and complete a pluripotent immune
582 activation, which lays the foundation for the anti-pathogen ability of various immune cells
583 differentiated from mature HSC.



584
585 **Figure 9.** Schematic diagram of dynamic regulations of TEs during EHT. The EHT process
586 in the AGM dorsal aorta is shown in the middle vessel. The process of TEs driving EHT by
587 providing different cis-regulatory elements through dynamic accessibility is presented
588 above. Below the blood vessel, the AGM hypoxic niche induces TE activation in pre-HECs
589 and triggers delayed TE sensing and inflammatory signaling through pattern recognition
590 receptors, thereby promoting the formation of nascent HSCs.

591 Cis-regulatory elements, especially enhancers, are another important way in which TEs
592 exert genome regulatory functions [57, 59, 61, 76, 77]. In this study, we observed that TEs
593 show cell type-specific dynamic accessible patterns (Fig. 6D). The distribution of TE-
594 related open regions was detected near many EHT marker genes (Additional file 9: Fig.
595 S2). We identified two TE-associated enhancers, eL2c and eID_B1 (upstream of the Gja5
596 transcript), whose accessibilities were specifically increased in pre-HECs and were
597 predicted to interact with the Gja5 promoter (Fig. 6G). However, the transcription factors
598 that bind them have not been determined. Through motif prediction and TF expression
599 analysis of cell type-specific TEPs, we recovered the complete TE-involved cis-regulatory

600 networks during EHT (Fig. 7). Interestingly, we found that the activity of many motifs and
601 the expression of corresponding TFs are not fully synchronized. The open range of motifs
602 seems to be wider, while the TF expression is more specific. For example, some key TF
603 binding motifs (such as SOX, GATA) appear to open earlier, while the TF expression is
604 tightly timed (Fig. 7B, C; Fig. 9).

605 What are the conditions for the activation of TEs during pre-HEC specification, and whether
606 the AGM niche has an effect on the TE activation? These are the objects of much
607 speculation. Many evidences show that environmental stress (including heat shock,
608 oxidative and chemotherapy) [50, 94] is one of the driving forces to induce TE activation.

609 In our study, we inferred that the AGM hypoxic niche may be at least partially responsible
610 for the activation of TEs in pre-HECs. Especially after the elevated expression of hypoxia-
611 inducible factors EPAS1 and HIF3A, the TE silencers were downregulated, accompanied
612 by the increase of TE expression (Fig. 8D, E). Indeed, hypoxia in the AGM region (IAHC
613 cluster) has already been observed and shown to promote HSC formation [81-84]. Our
614 study supports these findings and attributes the role of hypoxia on HSC development in
615 part to TE activation. Interestingly, the stromal cells and epithelial cells in AGM were also
616 hypoxic to some extent, and TE expression was also relatively high (Fig. 8F; Additional file

617 12: Fig. S1A), indicating that the activation of TEs by hypoxia appears to be cell type
618 insensitive. We recognize that HIF3A seems to be more related to the TE expression, as
619 evidenced by the upregulation of TEs in stromal and epithelial cells that do not express
620 EPAS1 but express HIF3A (Additional file 12: Fig. S1B, E). Furthermore, by analyzing a
621 time-series RNA-seq study examining the hypoxia response of HUVECs, we observed that

622 TE silencers were downregulated and the TE expression was broadly upregulated after 12
623 h of hypoxic culture, further confirming our hypothesis (Fig. 8G). In fact, the effects of
624 hypoxia on HSC development are multifaceted, for example, we found that cell cycle and
625 metabolism were inhibited in hypoxic pre-HECs (Fig. 8C; Fig. 9). Conversely, genes related
626 to cell migration were upregulated, which matches the findings in tumor EMT [95, 96].

627 **Conclusions**

628 TEs are domesticated during the evolution of eukaryotic genomes and mediate the
629 emergence of novel regulatory elements [42]. TEs have been found to be specifically
630 expressed to promote definitive hematopoiesis and HSC regeneration through RLRs and
631 inflammatory signaling [31, 36, 50, 51]. Our study not only extends the potential upstream
632 and downstream of TE transcription during EHT at the single-cell level, but also fills in the
633 gap of TEs as cis-regulatory elements driving HSC development. We found that TEs were
634 massively upregulated during pre-HEC specification, coinciding with the downregulation of
635 TE silencers at this stage. PRRs-mediated TE product sensing and activation of
636 inflammatory signaling are delayed until the HSC stage. These observations are highly
637 conserved between human and mouse. Analysis of scATAC-seq data reveals that
638 dynamically accessible TEs shape the hematopoietic cis-regulatory network to coordinate
639 the EHT process. We additionally reported that the hypoxic AGM niche may be partially
640 responsible for the transient TE activation before hematopoietic fate commitment. Further
641 investigations are required to confirm such a hypothesis. In summary, this study provides
642 a systematical single-cell analysis to uncover how TEs, through dynamic expression and
643 chromatin accessibility, orchestrate the EHT process and drive HSC formation.

644 **Methods**

645 **TE coverage, distribution, and regulatory potential analysis**

646 The TE annotation data of human (hg38) and mouse (mm10) were obtained from the
647 UCSC Genome Browser database (<https://genome.ucsc.edu/>) [97]. The genomic
648 annotations (intergenic, intron, 3' UTR, 5' UTR, CDS) and unmasked CpG islands data
649 were also downloaded from UCSC. The intersection of TEs and gene structures was
650 measured using BEDTools (v2.30.0) [98]. ChIPseeker (v1.34.1) [99] was used to visualize
651 the TE distributions with respect to protein-coding genes. The cCREs annotations
652 (including PLS, pELS, dELS, CTCF-only and DNase-H3K4me3) were downloaded from
653 ENCODE [54]. To improve the annotation accuracy of TE regulatory potential, the
654 overlapping of TE and cCRE was required to be more than 50% of the TE length. The
655 heatmap representations were generated using the R package pheatmap (v1.0.12).

656 **Single cell RNA-seq data processing**

657 The single-cell raw sequencing data of human and mouse AGM were downloaded from
658 GEO (<https://www.ncbi.nlm.nih.gov/geo/>) with accession numbers GSE162950 [23] and
659 GSE137117 [22]. The detailed information of samples used in this study can be found in
660 Additional file 3: Table S1. Reads were mapped to the human (refdata-gex-GRCh38-2020-
661 A) and mouse (refdata-gex-mm10-2020-A) reference genomes using CellRanger (v7.1.0).
662 The R package Seurat (v4.3.0) [100] was used to perform downstream analysis. Cells with
663 less than 200 unique molecular identifiers (UMIs) or greater than 15% mitochondrial
664 expression were removed and clusters with unusual low RNA features or counts were also
665 filtered in further analysis. Batch effects were corrected by Harmony (v0.1.1) [101].

666 SCTransform (v0.3.5) [102] was used to normalize the clean data followed by dimension
667 reduction and clustering through Seurat. Marker genes were identified using
668 FindAllMarkers with MAST [103]. Cell types were annotated according to the marker genes
669 provided in [23] (Additional file 4). Integration of the human and mouse EHT data were
670 achieved by Seurat CCA based on the shared homologous genes. The R package biomaRt
671 (v2.54.1) [104] was used to map gene symbols of mouse to human.

672 **EHT trajectory reconstruction**

673 To reconstruct the EHT trajectory, Velocityo (v0.17.17) [105] was applied to estimate the
674 RNA velocity from the bam files generated by Cell Ranger. The velocity maps were
675 visualized by scVelo (v0.2.5) [106]. The latent time was estimated using dynamical
676 modeling that models the full splicing kinetics in scVelo. Cell cycle scores along EHT
677 trajectory were calculated based on phase marker genes [107].

678 **Single cell TE quantification and differential expression analysis**

679 We applied scTE (v1.0) [49] to quantify the TE expression at the family level in human and
680 mouse scRNA-seq data. To keep the consistency of the read counting results, we
681 incorporated the same gene annotations as Cell Ranger and TE annotations from UCSC
682 to build the genome indices. Count matrix of only LINEs, SINEs/SVAs, LTRs and DNAs
683 were kept and merged into the Seurat object for further analysis. Cell type-specific marker
684 TEs were identified using FindAllMarkers with default parameters. Differential expression
685 analysis of TEs was performed using FindMarkers with default parameters. The specificity

686 of a TE was measured by the percent difference in expression of the TE between the target
687 cell type and other cell types.

688 **Co-expression gene and TE module analysis**

689 Cell type-specific marker genes (average $\log_{2}FC \geq 0.5$ and adjusted P-value ≤ 0.05) and
690 TEs counting more than 50 were extracted for weighted gene co-expression network
691 analysis (WGCNA) using hdWGCNA (v0.2.16) [108], which extends the standard WGCNA
692 [109] pipeline into scRNA-seq analysis. The single cells were first aggregated into
693 pseudobulk (meta) cells to reduce the drop out effect. The co-expression networks were
694 visualized with UMAP. The module connectivity score (kME) was computed based on
695 module eigengene. The module scores of TEs with $kME \geq 0.3$ in HME5 and MME5 were
696 calculated using AddModuleScore in Seurat.

697 **TE silencing and sensing analysis**

698 Genes related to TE silencing were collected from the literature [42, 44]. Potential KRAB-
699 ZFP genes in human and mouse were obtained from [63]. The whole list of TE silencers
700 analyzed in this study can be found in Additional file 7: Table S9 and S10. TE silencers with
701 lower mean expression in pre-HECs than in other cell types were selected and displayed
702 as heatmaps. Genes associated with TE sensing (including PRRs and downstream
703 intermediates) were also extracted from publications [35, 37, 39, 71, 72]. TE sensing genes
704 and inflammatory factors are listed in Additional file 8: Table S3 and S4. Differential
705 expression analysis of genes between hematopoietic cells (HECs/HSCs) and endothelial

706 cells (VECs/AECs) were performed using FindMarkers with default parameters. The
707 module scores of different gene sets were calculated using AddModuleScore.

708 **Functional enrichment analysis**

709 The gene set enrichment analysis (GSEA) was performed using clusterProfiler (v4.6.2)
710 [110]. Both Gene Ontology (biological process) and Molecular Signatures Database
711 (hallmark gene sets) are included. Genes were ranked according to fold changes
712 calculated in Seurat.

713 **Single cell ATAC-seq data processing**

714 The raw sequencing data of mouse AGM (E10.5) was downloaded from GEO with
715 accession GSE137115 [22]. Reads were mapped to the mouse reference genome
716 (refdata-cellranger-arc-mm10-2020-A) using cellranger-atac (v2.1.0). The R package
717 Signac (v1.9.0) [111] was used to perform downstream analysis, including quality control,
718 normalization, dimension reduction and clustering. After estimating the gene activities, the
719 cell types of scATAC-seq data were annotated through cross-modality integration and label
720 transfer from scRNA-seq data using CCA [102]. The final cell types were corrected
721 according to the gene activities of known EHT markers.

722 **Single cell TE accessibility estimation and differential accessible analysis**

723 The count matrix of TEs was estimated using FeatureMatrix in Signac. Cell type-specific
724 open TEs were identified by FindAllMarkers. Differentially accessible peaks and TEs
725 between cell types were identified using FindMarkers. Each of the open TEs was assigned
726 to the closest gene using ClosestFeature. TE-related differentially accessible peaks

727 (Additional file 9: Fig. S2) were plotted on the mouse genome using karyoloteR (v1.24.0)
728 [112].

729 **Cis-co-accessible network (CCAN) construction**

730 The cis-co-accessible peaks were identified using Cicero (v1.3.9) [74]. The links with
731 coaccess score more than 0.4 were extracted to construct the CCANs. The CCAN network
732 of all differentially accessible peaks was visualized in Cytoscape (v3.9.0) [113].

733 **Motif enrichment and TF expression analysis**

734 The motif enrichment analysis was performed in Signac. The motif position frequency
735 matrices were from JASPAR [114]. Motifs enriched in TE-related differentially accessible
736 peaks were found by FindMotifs. The motif activity was computed by chromVAR (v1.20.2)
737 [115]. Active motifs were selected by combining with the expression of corresponding TFs
738 from scRNA-seq data. The cell type-specific TF-target network (Additional file 9: Fig. S3)
739 was constructed based on interactions from TRRUST [78]. The average expression of the
740 target genes in the target cell type were required to be more than 0.25.

741 **Spatial transcriptome data processing**

742 The raw spatial transcriptome sequencing data of human AGM (CS15, sample 7) was
743 downloaded from GEO with accession GSE162950 [23]. Reads were mapped to the
744 human reference genome (refdata-gex-GRCh38-2020-A) using Space Ranger (v2.0.1).
745 The R package Seurat was used to perform downstream analysis. The expression data
746 were normalized using SCTransform.

747 **HUVEC bulk RNA-seq data processing**

748 The raw sequencing data of HUVECs against hypoxia stress was downloaded from SRA
749 (<https://www.ncbi.nlm.nih.gov/Traces/study/>) with accession PRJNA561635 [90]. The
750 detailed information of samples used in this study can be found in Additional file 3: Table
751 S2. We treated each sample as a single cell and thus can still use scTE to quantify TE and
752 gene expression. Only wild type samples were included for further analysis. The gene
753 modules of pre-HEC markers (Additional file 12: Fig. S3) in HUVEC data were predicted
754 using WGCNA [109].

755 **Abbreviations**

756 EHT: endothelial-to-hematopoietic transition; TEs: transposable elements; HSCs:
757 hematopoietic stem cells; ECs: endothelial cells; AGM: aorta-gonad-mesonephro; AECs:
758 arterial endothelial cells; HECs: hematopoietic endothelial cells; IAHCs: intra-aortic
759 hematopoietic clusters; scRNA-seq: single-cell RNA sequencing; scATAC-seq: single-cell
760 sequencing assay for transposase-accessible chromatin; TNF: tumor necrosis factor; IFN:
761 interferon; PRRs: pattern recognition receptors; TLRs: Toll-like receptors; RLRs: RIG-I-like
762 receptors; NLRs: NOD-like receptors; CLRs: C-type lectin receptors; TRAFs: TNF
763 receptor-associated factors; ssRNAs: single-stranded RNAs; dsRNAs: double-stranded
764 RNAs; DNAs: DNA transposons; LINEs: long interspersed nuclear elements; SINEs: short
765 interspersed nuclear elements; SVA: SINE-VNTR-Alu; LTRs: long terminal repeats; ERVs:
766 endogenous retroviruses; KRAB-ZFP: Krüppel-associated box zinc finger protein; DNMT:
767 DNA methyltransferases; NuRD: nucleosome remodeling deacetylase; HUSH: human
768 silencing hub; TSS: transcription start sites; TTS: transcription termination sites; cCREs:

769 candidate cis-regulatory elements; PLS: promoter-like sites; pELS: proximal enhancer-like
770 signatures; dELS: distal enhancer-like signatures; kME: module connectivity; PIWIs: P-
771 element induced Wimpy testis-related genes; PKRs: protein kinase R genes; GSEA: gene
772 set enrichment analysis; GO: gene ontology; DATEs: differentially accessible TEs; DAPs:
773 differentially accessible peaks; TEPs: TE overlapped DAPs; CCANs: cis-co-accessibility
774 networks; EMT: epithelial to mesenchymal transition; HIFs: hypoxia-inducible factors;
775 hESCs: human embryonic stem cells; HUVECs: human umbilical vein endothelial cells;
776 UMIs: unique molecular identifiers; WGCNA: weighted gene co-expression network
777 analysis.

778 **Declarations**

779 **Ethics approval and consent to participate**

780 Not applicable.

781 **Consent for publication**

782 Not applicable.

783 **Availability of data and materials**

784 The scRNA-seq and spatial transcriptome data for human AGM that were analyzed in this
785 study are available from GEO (GSE162950) [23]. The scRNA-seq and scATAC-seq data
786 for mouse AGM are available from GEO (GSE137117) [22]. The time series RNA-seq data
787 for HUVEC are available from SRA (PRJNA561635) [90]. The detailed information of
788 samples used in this study can be found in Additional file 3. All analysis pipelines, in-house
789 scripts and files for reproducing the results in this study can be accessed at
790 <https://github.com/ventson/hscTE>. We also provide a web interface

791 (<https://bis.zju.edu.cn/hscTE>, implemented using UCSC Cell Browser [116]) to visualize TE
792 and gene expression during human and mouse EHT. The multi-faceted display (including
793 TEs, CpG, cCREs, peaks and genome coverages) of mouse EHT scATAC-seq data is
794 available from <https://bis.zju.edu.cn/hscTE/jbrowse/?data=mouse>, which is implemented
795 by JBrowse [117].

796 **Competing interests**

797 The authors declare that they have no competing interests.

798 **Funding**

799 This work was supported by the National Key Research and Development Program of
800 China (No. 2018YFC0310602; 2016YFA0501704), National Natural Sciences Foundation
801 of China (No. 31771477; 32070677), the 151 Talent Project of Zhejiang Province (first level),
802 the Fundamental Research Funds for the Central Universities, Jiangsu Collaborative
803 Innovation Center for Modern Crop Production and Collaborative Innovation Center for
804 Modern Crop Production co-sponsored by province and ministry, Doctor Foundation of The
805 Second Hospital of Shanxi Medical University (No. 201701-4), Natural Science Foundation
806 of Shanxi Province (No. 201901D111381).

807 **Authors' contributions**

808 MC, HH, and WDL conceived and supervised the study. CF, RXT, and SGX processed the
809 data and prepared the manuscript. YHC helped RNA velocity and trajectory analysis. SDL
810 and XTH helped the spatial transcriptome analysis. YCZ, YJL, YMH, YSH, HP, and ZXW
811 contributed to the analysis pipeline construction and modification. HYC, SLZ, and QYN
812 helped the scATAC-data analysis. JYH provided constructive discussion and suggestions.

813 CF, RXT, SGX, MC, HH, and WDL wrote and revised the manuscript with input from all the
814 other authors. All authors read and approved the final manuscript.

815 **Acknowledgements**

816 The authors would like to thank members in Ming Chen's lab for discussion and valuable
817 suggestions.

818 **Authors' information**

819 Department of Bioinformatics, Zhejiang University College of Life Sciences, Hangzhou
820 310058, China

821 Cong Feng, Saige Xin, Yuhao Chen, Sida Li, Xiaotian Hu, Yincong Zhou, Yueming Hu,
822 Yanshi Hu, Zexu Wu, Haoyu Chao, Shilong Zhang, Qingyang Ni, Ming Chen

823 Bioinformatics Center, The First Affiliated Hospital, Zhejiang University School of Medicine,
824 Hangzhou 310058, China

825 Cong Feng, Yongjing Liu, Jinyan Huang, Ming Chen

826 Bone Marrow Transplantation Center, the First Affiliated Hospital, Zhejiang University
827 School of Medicine, Hangzhou 310058, China

828 Ruxiu Tie, He Huang

829 Liangzhu Laboratory, Zhejiang University Medical Center, Hangzhou 310058, China

830 Ruxiu Tie, He Huang

831 Institute of Hematology, Zhejiang University, Hangzhou 310058, China

832 Ruxiu Tie, He Huang

833 Zhejiang Province Engineering Laboratory for Stem Cell and Immunity Therapy, Hangzhou

834 310058, China

835 Ruxiu Tie, He Huang

836 Department of Hematology, The Second Clinical Medical College of Shanxi Medical
837 University, Shanxi Medical University, Taiyuan 030000, China

838 Ruxiu Tie

839 Department of Hematology-Oncology, Taizhou Hospital of Zhejiang Province, Linhai
840 317000, China

841 Ruxiu Tie, Wenda Luo

842 Department of Veterinary Medicine, Zhejiang University College of Animal Sciences,
843 Hangzhou 310058, China

844 Hang Pan

845 References

- 846 1. Bertrand JY, Chi NC, Santoso B, Teng S, Stainier DY, Traver D: **Haematopoietic stem**
847 **cells derive directly from aortic endothelium during development.** *Nature* 2010,
848 **464:**108-111.
- 849 2. Boisset JC, van Cappellen W, Andrieu-Soler C, Galjart N, Dzierzak E, Robin C: **In vivo**
850 **imaging of haematopoietic cells emerging from the mouse aortic endothelium.**
851 *Nature* 2010, **464:**116-120.
- 852 3. Eilken HM, Nishikawa S, Schroeder T: **Continuous single-cell imaging of blood**
853 **generation from haemogenic endothelium.** *Nature* 2009, **457:**896-900.
- 854 4. Kiss K, Herbomel P: **Blood stem cells emerge from aortic endothelium by a novel**
855 **type of cell transition.** *Nature* 2010, **464:**112-115.
- 856 5. Lancrin C, Sroczynska P, Stephenson C, Allen T, Kouskoff V, Lacaud G: **The**
857 **haemangioblast generates haematopoietic cells through a haemogenic**
858 **endothelium stage.** *Nature* 2009, **457:**892-895.
- 859 6. Jagannathan-Bogdan M, Zon LI: **Hematopoiesis.** *Development* 2013, **140:**2463-2467.
- 860 7. de Bruijn MF, Speck NA, Peeters MC, Dzierzak E: **Definitive hematopoietic stem cells**
861 **first develop within the major arterial regions of the mouse embryo.** *EMBO J* 2000,
862 **19:**2465-2474.
- 863 8. Medvinsky A, Dzierzak E: **Definitive hematopoiesis is autonomously initiated by the**
864 **AGM region.** *Cell* 1996, **86:**897-906.
- 865 9. Ivanovs A, Rybtsov S, Ng ES, Stanley EG, Elefanti AG, Medvinsky A: **Human**
866 **haematopoietic stem cell development: from the embryo to the dish.** *Development*

- 867 2017, **144**:2323-2337.
- 868 10. Ivanovs A, Rybtsov S, Welch L, Anderson RA, Turner ML, Medvinsky A: **Highly potent**
869 **human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-**
870 **mesonephros region.** *J Exp Med* 2011, **208**:2417-2427.
- 871 11. Taoudi S, Medvinsky A: **Functional identification of the hematopoietic stem cell**
872 **niche in the ventral domain of the embryonic dorsal aorta.** *Proc Natl Acad Sci U S A*
873 2007, **104**:9399-9403.
- 874 12. Hou S, Li Z, Zheng X, Gao Y, Dong J, Ni Y, Wang X, Li Y, Ding X, Chang Z, et al:
875 **Embryonic endothelial evolution towards first hematopoietic stem cells revealed**
876 **by single-cell transcriptomic and functional analyses.** *Cell Res* 2020, **30**:376-392.
- 877 13. Zhao S, Feng S, Tian Y, Wen Z: **Hemogenic and aortic endothelium arise from a**
878 **common hemogenic angioblast precursor and are specified by the Etv2 dosage.**
879 *Proc Natl Acad Sci U S A* 2022, **119**:e2119051119.
- 880 14. Zeng Y, He J, Bai Z, Li Z, Gong Y, Liu C, Ni Y, Du J, Ma C, Bian L, et al: **Tracing the first**
881 **hematopoietic stem cell generation in human embryo by single-cell RNA**
882 **sequencing.** *Cell Res* 2019, **29**:881-894.
- 883 15. Taoudi S, Gonneau C, Moore K, Sheridan JM, Blackburn CC, Taylor E, Medvinsky A:
884 **Extensive hematopoietic stem cell generation in the AGM region via maturation of**
885 **VE-cadherin+CD45+ pre-definitive HSCs.** *Cell Stem Cell* 2008, **3**:99-108.
- 886 16. Zhou F, Li X, Wang W, Zhu P, Zhou J, He W, Ding M, Xiong F, Zheng X, Li Z, et al:
887 **Tracing haematopoietic stem cell formation at single-cell resolution.** *Nature* 2016,
888 **533**:487-492.
- 889 17. Baron CS, Kester L, Klaus A, Boisset JC, Thambyrajah R, Yvernogeau L, Kouskoff V,
890 Lacaud G, van Oudenaarden A, Robin C: **Single-cell transcriptomics reveal the**
891 **dynamic of haematopoietic stem cell production in the aorta.** *Nat Commun* 2018,
892 **9**:2517.
- 893 18. Vink CS, Calero-Nieto FJ, Wang X, Maglitto A, Mariani SA, Jawaid W, Gottgens B,
894 Dzierzak E: **Iterative single-cell analyses define the transcriptome of the first**
895 **functional hematopoietic stem cells.** *Cell Rep* 2020, **31**:107627.
- 896 19. Ye F, Huang W, Guo G: **Studying hematopoiesis using single-cell technologies.** *J*
897 *Hematol Oncol* 2017, **10**:27.
- 898 20. Zhang P, Li X, Pan C, Zheng X, Hu B, Xie R, Hu J, Shang X, Yang H: **Single-cell RNA**
899 **sequencing to track novel perspectives in HSC heterogeneity.** *Stem Cell Res Ther*
900 2022, **13**:39.
- 901 21. Fadlullah MZH, Neo WH, Lie ALM, Thambyrajah R, Patel R, Mevel R, Aksoy I, Do Khoa
902 N, Savatier P, Fontenille L, et al: **Murine AGM single-cell profiling identifies a**
903 **continuum of hemogenic endothelium differentiation marked by ACE.** *Blood* 2022,
904 **139**:343-356.
- 905 22. Zhu Q, Gao P, Tober J, Bennett L, Chen C, Uzun Y, Li Y, Howell ED, Mumau M, Yu W, et
906 al: **Developmental trajectory of prehematopoietic stem cell formation from**
907 **endothelium.** *Blood* 2020, **136**:845-856.
- 908 23. Calvanese V, Capellera-Garcia S, Ma F, Fares I, Liebscher S, Ng ES, Ekstrand S,
909 Aguade-Gorgorio J, Vavilina A, Lefaudeux D, et al: **Mapping human haematopoietic**
910 **stem cells from haemogenic endothelium to birth.** *Nature* 2022, **604**:534-540.

- 911 24. Orkin SH, Zon LI: **Hematopoiesis: an evolving paradigm for stem cell biology.** *Cell*
912 2008, **132**:631-644.
- 913 25. Ding Y, Liu Z, Liu F: **Transcriptional and epigenetic control of hematopoietic stem**
914 **cell fate decisions in vertebrates.** *Dev Biol* 2021, **475**:156-164.
- 915 26. de Bruijn M, Dzierzak E: **Runx transcription factors in the development and function**
916 **of the definitive hematopoietic system.** *Blood* 2017, **129**:2061-2069.
- 917 27. Katsumura KR, Bresnick EH, Group GFM: **The GATA factor revolution in hematology.**
918 *Blood* 2017, **129**:2092-2102.
- 919 28. Thambyrajah R, Mazan M, Patel R, Moignard V, Stefanska M, Marinopoulou E, Li Y,
920 Lancrin C, Clapes T, Moroy T, et al: **GFI1 proteins orchestrate the emergence of**
921 **haematopoietic stem cells through recruitment of LSD1.** *Nat Cell Biol* 2016, **18**:21-
922 32.
- 923 29. Clapes T, Lefkopoulos S, Trompouki E: **Stress and non-stress roles of inflammatory**
924 **signals during HSC emergence and maintenance.** *Front Immunol* 2016, **7**:487.
- 925 30. He Q, Zhang C, Wang L, Zhang P, Ma D, Lv J, Liu F: **Inflammatory signaling regulates**
926 **hematopoietic stem and progenitor cell emergence in vertebrates.** *Blood* 2015,
927 **125**:1098-1106.
- 928 31. Collins A, Mitchell CA, Passegue E: **Inflammatory signaling regulates hematopoietic**
929 **stem and progenitor cell development and homeostasis.** *J Exp Med* 2021, **218**.
- 930 32. Tie R, Li H, Cai S, Liang Z, Shan W, Wang B, Tan Y, Zheng W, Huang H: **Interleukin-6**
931 **signaling regulates hematopoietic stem cell emergence.** *Exp Mol Med* 2019, **51**:1-
932 12.
- 933 33. Espin-Palazon R, Stachura DL, Campbell CA, Garcia-Moreno D, Del Cid N, Kim AD,
934 Candel S, Meseguer J, Mulero V, Traver D: **Proinflammatory signaling regulates**
935 **hematopoietic stem cell emergence.** *Cell* 2014, **159**:1070-1085.
- 936 34. Li Y, Esain V, Teng L, Xu J, Kwan W, Frost IM, Yzaguirre AD, Cai X, Cortes M,
937 Maijenburg MW, et al: **Inflammatory signaling regulates embryonic hematopoietic**
938 **stem and progenitor cell production.** *Genes Dev* 2014, **28**:2597-2612.
- 939 35. Takeuchi O, Akira S: **Pattern recognition receptors and inflammation.** *Cell* 2010,
940 **140**:805-820.
- 941 36. Lefkopoulos S, Polyzou A, Derecka M, Bergo V, Clapes T, Cauchy P, Jerez-Longres C,
942 Onishi-Seebacher M, Yin N, Martagon-Calderon NA, et al: **Repetitive elements trigger**
943 **RIG-I-like receptor signaling that regulates the emergence of hematopoietic stem**
944 **and progenitor cells.** *Immunity* 2020, **53**:934-951 e939.
- 945 37. Bowie AG, Unterholzner L: **Viral evasion and subversion of pattern-recognition**
946 **receptor signalling.** *Nat Rev Immunol* 2008, **8**:911-922.
- 947 38. Kowalinski E, Lunardi T, McCarthy AA, Louber J, Brunel J, Grigorov B, Gerlier D, Cusack
948 **S: Structural basis for the activation of innate immune pattern-recognition receptor**
949 **RIG-I by viral RNA.** *Cell* 2011, **147**:423-435.
- 950 39. Chen YG, Hur S: **Cellular origins of dsRNA, their recognition and consequences.**
951 *Nat Rev Mol Cell Biol* 2022, **23**:286-301.
- 952 40. Burns KH: **Repetitive DNA in disease.** *Science* 2022, **376**:353-354.
- 953 41. Payer LM, Burns KH: **Transposable elements in human genetic disease.** *Nat Rev*
954 *Genet* 2019, **20**:760-772.

- 955 42. Almeida MV, Vernaz G, Putman ALK, Miska EA: **Taming transposable elements in**
956 **vertebrates: from epigenetic silencing to domestication.** *Trends Genet* 2022,
957 **38:**529-553.
- 958 43. Senft AD, Macfarlan TS: **Transposable elements shape the evolution of mammalian**
959 **development.** *Nat Rev Genet* 2021, **22:**691-711.
- 960 44. Ecco G, Imbeault M, Trono D: **KRAB zinc finger proteins.** *Development* 2017,
961 **144:**2719-2729.
- 962 45. Hutchins AP, Pei D: **Transposable elements at the center of the crossroads between**
963 **embryogenesis, embryonic stem cells, reprogramming, and long non-coding**
964 **RNAs.** *Sci Bull (Beijing)* 2015, **60:**1722-1733.
- 965 46. Low Y, Tan DEK, Hu Z, Tan SYX, Tee WW: **Transposable element dynamics and**
966 **regulation during zygotic genome activation in mammalian embryos and**
967 **embryonic stem cell model systems.** *Stem Cells Int* 2021, **2021:**1624669.
- 968 47. Chang NC, Rovira Q, Wells J, Feschotte C, Vaquerizas JM: **Zebrafish transposable**
969 **elements show extensive diversification in age, genomic distribution, and**
970 **developmental expression.** *Genome Res* 2022, **32:**1408-1423.
- 971 48. Pontis J, Pulver C, Playfoot CJ, Planet E, Grun D, Offner S, Duc J, Manfrin A, Lutolf MP,
972 Trono D: **Primate-specific transposable elements shape transcriptional networks**
973 **during human development.** *Nat Commun* 2022, **13:**7178.
- 974 49. He J, Babarinde IA, Sun L, Xu S, Chen R, Shi J, Wei Y, Li Y, Ma G, Zhuang Q, et al:
975 **Identifying transposable element expression dynamics and heterogeneity during**
976 **development at the single-cell level with a processing pipeline scTE.** *Nat Commun*
977 **2021, 12:**1456.
- 978 50. Clapes T, Polyzou A, Prater P, Sagar, Morales-Hernandez A, Ferrarini MG, Kehrer N,
979 Lefkopoulos S, Bergo V, Hummel B, et al: **Chemotherapy-induced transposable**
980 **elements activate MDA5 to enhance haematopoietic regeneration.** *Nat Cell Biol*
981 **2021, 23:**704-717.
- 982 51. Clapes T, Trompouki E: **Hematopoietic regeneration under the spell of epigenetic-**
983 **epitranscriptomic factors and transposable elements.** *Curr Opin Hematol* 2020,
984 **27:**264-272.
- 985 52. Ye M, Goudot C, Hoyler T, Lemoine B, Amigorena S, Zueva E: **Specific subfamilies of**
986 **transposable elements contribute to different domains of T lymphocyte enhancers.**
987 *Proc Natl Acad Sci U S A* 2020, **117:**7905-7916.
- 988 53. Feng C, Dai M, Liu Y, Chen M: **Sequence repetitiveness quantification and de novo**
989 **repeat detection by weighted k-mer coverage.** *Brief Bioinform* 2021, **22.**
- 990 54. Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shores N, Adrian J, Kawli T, Davis CA,
991 Dobin A, Kaul R, et al: **Expanded encyclopaedias of DNA elements in the human and**
992 **mouse genomes.** *Nature* 2020, **583:**699-710.
- 993 55. Walter M, Teissandier A, Perez-Palacios R, Bourc'his D: **An epigenetic switch ensures**
994 **transposon repression upon dynamic loss of DNA methylation in embryonic stem**
995 **cells.** *Elife* 2016, **5.**
- 996 56. Zhou W, Liang G, Molloy PL, Jones PA: **DNA methylation enables transposable**
997 **element-driven genome expansion.** *Proc Natl Acad Sci U S A* 2020, **117:**19359-19366.
- 998 57. de Souza FS, Franchini LF, Rubinstein M: **Exaptation of transposable elements into**

- 999 **novel cis-regulatory elements: is the evidence always strong?** *Mol Biol Evol* 2013,
1000 **30:**1239-1251.
- 1001 58. Lee HJ, Hou Y, Maeng JH, Shah NM, Chen Y, Lawson HA, Yang H, Yue F, Wang T:
1002 **Epigenomic analysis reveals prevalent contribution of transposable elements to**
1003 **cis-regulatory elements, tissue-specific expression, and alternative promoters in**
1004 **zebrafish.** *Genome Res* 2022, **32:**1424-1436.
- 1005 59. Sundaram V, Wysocka J: **Transposable elements as a potent source of diverse cis-**
1006 **regulatory sequences in mammalian genomes.** *Philos Trans R Soc Lond B Biol Sci*
1007 2020, **375:**20190347.
- 1008 60. Todd CD, Deniz O, Taylor D, Branco MR: **Functional evaluation of transposable**
1009 **elements as enhancers in mouse embryonic and trophoblast stem cells.** *Elife* 2019,
1010 **8.**
- 1011 61. Andrews G, Fan K, Pratt HE, Phalke N, Zoonomia Consortium section s, Karlsson EK,
1012 Lindblad-Toh K, Gazal S, Moore JE, Weng Z, et al: **Mammalian evolution of human**
1013 **cis-regulatory elements and transcription factor binding sites.** *Science* 2023,
1014 **380:**eabn7930.
- 1015 62. Oatley M, Bolukbasi OV, Svensson V, Shvartsman M, Ganter K, Zirngibl K, Pavlovich PV,
1016 Milchevskaya V, Foteva V, Natarajan KN, et al: **Single-cell transcriptomics identifies**
1017 **CD44 as a marker and regulator of endothelial to haematopoietic transition.** *Nat*
1018 *Commun* 2020, **11:**586.
- 1019 63. Zhang Y, He F, Zhang Y, Dai Q, Li Q, Nan J, Miao R, Cheng B: **Exploration of the**
1020 **regulatory relationship between KRAB-Zfp clusters and their target transposable**
1021 **elements via a gene editing strategy at the cluster specific linker-associated**
1022 **sequences by CRISPR-Cas9.** *Mob DNA* 2022, **13:**25.
- 1023 64. Imbeault M, Helleloid PY, Trono D: **KRAB zinc-finger proteins contribute to the**
1024 **evolution of gene regulatory networks.** *Nature* 2017, **543:**550-554.
- 1025 65. Jacobs FM, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B,
1026 Salama SR, Haussler D: **An evolutionary arms race between KRAB zinc-finger**
1027 **genes ZNF91/93 and SVA/L1 retrotransposons.** *Nature* 2014, **516:**242-245.
- 1028 66. Ecco G, Cassano M, Kauzlaric A, Duc J, Coluccio A, Offner S, Imbeault M, Rowe HM,
1029 Turelli P, Trono D: **Transposable elements and their KRAB-ZFP controllers regulate**
1030 **gene expression in adult tissues.** *Dev Cell* 2016, **36:**611-623.
- 1031 67. Li Y, Tang C, Liu F, Zhu C, Liu F, Zhu P, Wang L: **DNA methylation safeguards the**
1032 **generation of hematopoietic stem and progenitor cells by repression of Notch**
1033 **signaling.** *Development* 2022, **149.**
- 1034 68. Soto RA, Najia MAT, Hachimi M, Frame JM, Yette GA, Lummetz da Rocha E, Stankunas
1035 K, Daley GQ, North TE: **Sequential regulation of hemogenic fate and hematopoietic**
1036 **stem and progenitor cell formation from arterial endothelium by Ezh1/2.** *Stem Cell*
1037 *Reports* 2021, **16:**1718-1734.
- 1038 69. Thambyrajah R, Fadlullah MZH, Proffitt M, Patel R, Cowley SM, Kouskoff V, Lacaud G:
1039 **HDAC1 and HDAC2 modulate TGF-beta signaling during endothelial-to-**
1040 **hematopoietic transition.** *Stem Cell Reports* 2018, **10:**1369-1383.
- 1041 70. Kasper DM, Nicoli S: **Epigenetic and Epitranscriptomic Factors Make a Mark on**
1042 **Hematopoietic Stem Cell Development.** *Curr Stem Cell Rep* 2018, **4:**22-32.

- 1043 71. Gazquez-Gutierrez A, Witteveldt J, S RH, Macias S: **Sensing of transposable**
1044 **elements by the antiviral innate immune system.** *RNA* 2021, **27**:735-752.
- 1045 72. Yang Y, Huang Y, Zeng Z: **Advances in cGAS-STING signaling pathway and**
1046 **diseases.** *Front Cell Dev Biol* 2022, **10**:800393.
- 1047 73. Meylan P, Dreos R, Ambrosini G, Groux R, Bucher P: **EPD in 2020: enhanced data**
1048 **visualization and extension to ncRNA promoters.** *Nucleic Acids Res* 2020, **48**:D65-
1049 D69.
- 1050 74. Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D,
1051 Srivatsan S, Qiu X, Jackson D, Minkina A, et al: **Cicero predicts cis-regulatory DNA**
1052 **interactions from single-cell chromatin accessibility data.** *Mol Cell* 2018, **71**:858-871
1053 e858.
- 1054 75. Cacialli P, Mahony CB, Petzold T, Bordignon P, Rougemont AL, Bertrand JY: **A**
1055 **connexin/ifi30 pathway bridges HSCs with their niche to dampen oxidative stress.**
1056 *Nat Commun* 2021, **12**:4484.
- 1057 76. Feschotte C: **Transposable elements and the evolution of regulatory networks.** *Nat*
1058 *Rev Genet* 2008, **9**:397-405.
- 1059 77. Floc'hlay S, Wong E, Zhao B, Viales RR, Thomas-Chollier M, Thieffry D, Garfield DA,
1060 Furlong EEM: **Cis-acting variation is common across regulatory layers but is often**
1061 **buffered during embryonic development.** *Genome Res* 2020, **31**:211-224.
- 1062 78. Han H, Cho JW, Lee S, Yun A, Kim H, Bae D, Yang S, Kim CY, Lee M, Kim E, et al:
1063 **TRRUST v2: an expanded reference database of human and mouse transcriptional**
1064 **regulatory interactions.** *Nucleic Acids Res* 2018, **46**:D380-D386.
- 1065 79. Sahlgren C, Gustafsson MV, Jin S, Poellinger L, Lendahl U: **Notch signaling mediates**
1066 **hypoxia-induced tumor cell migration and invasion.** *Proc Natl Acad Sci U S A* 2008,
1067 **105**:6392-6397.
- 1068 80. Saxena K, Jolly MK, Balamurugan K: **Hypoxia, partial EMT and collective migration:**
1069 **Emerging culprits in metastasis.** *Transl Oncol* 2020, **13**:100845.
- 1070 81. Gerri C, Marass M, Rossi A, Stainier DYR: **Hif-1alpha and Hif-2alpha regulate**
1071 **hemogenic endothelium and hematopoietic stem cell formation in zebrafish.** *Blood*
1072 2018, **131**:963-973.
- 1073 82. Wang N, Chen C, Cheng Y, Fu Y, Zhong Z, Yang Y, Lv L, Chen H, Huang J, Duan Y:
1074 **Hypoxia drives hematopoiesis with the enhancement of T lineage through eliciting**
1075 **arterial specification of hematopoietic endothelial progenitors from hESC.** *Stem*
1076 *Cell Res Ther* 2022, **13**:282.
- 1077 83. Imanirad P, Dzierzak E: **Hypoxia and HIFs in regulating the development of the**
1078 **hematopoietic system.** *Blood Cells Mol Dis* 2013, **51**:256-263.
- 1079 84. Imanirad P, Solaimani Kartalaei P, Crisan M, Vink C, Yamada-Inagawa T, de Pater E,
1080 Kurek D, Kaimakis P, van der Linden R, Speck N, Dzierzak E: **HIF1alpha is a regulator**
1081 **of hematopoietic progenitor and stem cell development in hypoxic sites of the**
1082 **mouse embryo.** *Stem Cell Res* 2014, **12**:24-35.
- 1083 85. Jaskiewicz M, Moszynska A, Serocki M, Kroliczewski J, Bartoszewska S, Collawn JF,
1084 Bartoszewska R: **Hypoxia-inducible factor (HIF)-3a2 serves as an endothelial cell**
1085 **fate executor during chronic hypoxia.** *EXCLI J* 2022, **21**:454-469.
- 1086 86. Mimura I, Nangaku M, Kanki Y, Tsutsumi S, Inoue T, Kohro T, Yamamoto S, Fujita T,

- 1087 Shimamura T, Suehiro J, et al: **Dynamic change of chromatin conformation in**
1088 **response to hypoxia enhances the expression of GLUT3 (SLC2A3) by cooperative**
1089 **interaction of hypoxia-inducible factor 1 and KDM3A.** *Mol Cell Biol* 2012, **32**:3018-
1090 3032.
- 1091 87. Sugiyama T, Kohara H, Noda M, Nagasawa T: **Maintenance of the hematopoietic stem**
1092 **cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell**
1093 **niches.** *Immunity* 2006, **25**:977-988.
- 1094 88. Casacuberta E, Gonzalez J: **The impact of transposable elements in environmental**
1095 **adaptation.** *Mol Ecol* 2013, **22**:1503-1517.
- 1096 89. Horvath V, Merenciano M, Gonzalez J: **Revisiting the relationship between**
1097 **transposable elements and the eukaryotic stress response.** *Trends Genet* 2017,
1098 **33**:832-841.
- 1099 90. Klomp J, Hyun J, Klomp JE, Pajcini K, Rehman J, Malik AB: **Comprehensive**
1100 **transcriptomic profiling reveals SOX7 as an early regulator of angiogenesis in**
1101 **hypoxic human endothelial cells.** *J Biol Chem* 2020, **295**:4796-4808.
- 1102 91. Chelmicki T, Roger E, Teissandier A, Dura M, Bonneville L, Rucli S, Dossin F, Fouassier
1103 C, Lameiras S, Bourc'his D: **m(6)A RNA methylation regulates the fate of**
1104 **endogenous retroviruses.** *Nature* 2021, **591**:312-316.
- 1105 92. Xiong F, Wang R, Lee JH, Li S, Chen SF, Liao Z, Hasani LA, Nguyen PT, Zhu X,
1106 Krakowiak J, et al: **RNA m(6)A modification orchestrates a LINE-1-host interaction**
1107 **that facilitates retrotransposition and contributes to long gene vulnerability.** *Cell*
1108 **Res** 2021, **31**:861-885.
- 1109 93. Pickering RJ, Booty LM: **NLR in eXile: Emerging roles of NLRX1 in immunity and**
1110 **human disease.** *Immunology* 2021, **162**:268-280.
- 1111 94. Villanueva-Canas JL, Horvath V, Aguilera L, Gonzalez J: **Diverse families of**
1112 **transposable elements affect the transcriptional regulation of stress-response**
1113 **genes in Drosophila melanogaster.** *Nucleic Acids Res* 2019, **47**:6842-6857.
- 1114 95. Hapke RY, Haake SM: **Hypoxia-induced epithelial to mesenchymal transition in**
1115 **cancer.** *Cancer Lett* 2020, **487**:10-20.
- 1116 96. Zhang L, Huang G, Li X, Zhang Y, Jiang Y, Shen J, Liu J, Wang Q, Zhu J, Feng X, et al:
1117 **Hypoxia induces epithelial-mesenchymal transition via activation of SNAI1 by**
1118 **hypoxia-inducible factor -1alpha in hepatocellular carcinoma.** *BMC Cancer* 2013,
1119 **13**:108.
- 1120 97. Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenblom KR, Raney BJ, Lee CM, Lee
1121 BT, Hinrichs AS, Gonzalez JN, et al: **The UCSC Genome Browser database: 2019**
1122 **update.** *Nucleic Acids Res* 2019, **47**:D853-D858.
- 1123 98. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic**
1124 **features.** *Bioinformatics* 2010, **26**:841-842.
- 1125 99. Yu G, Wang LG, He QY: **ChIPseeker: an R/Bioconductor package for ChIP peak**
1126 **annotation, comparison and visualization.** *Bioinformatics* 2015, **31**:2382-2383.
- 1127 100. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ,
1128 Darby C, Zager M, et al: **Integrated analysis of multimodal single-cell data.** *Cell* 2021,
1129 **184**:3573-3587 e3529.
- 1130 101. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M,

- 1131 Loh PR, Raychaudhuri S: **Fast, sensitive and accurate integration of single-cell data**
1132 **with Harmony.** *Nat Methods* 2019, **16**:1289-1296.
- 1133 102. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y,
1134 Stoeckius M, Smibert P, Satija R: **Comprehensive integration of single-cell data.** *Cell*
1135 2019, **177**:1888-1902 e1821.
- 1136 103. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW,
1137 McElrath MJ, Prlic M, et al: **MAST: a flexible statistical framework for assessing**
1138 **transcriptional changes and characterizing heterogeneity in single-cell RNA**
1139 **sequencing data.** *Genome Biol* 2015, **16**:278.
- 1140 104. Durinck S, Spellman PT, Birney E, Huber W: **Mapping identifiers for the integration of**
1141 **genomic datasets with the R/Bioconductor package biomaRt.** *Nat Protoc* 2009,
1142 **4**:1184-1191.
- 1143 105. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K,
1144 Kastriti ME, Lonnerberg P, Furlan A, et al: **RNA velocity of single cells.** *Nature* 2018,
1145 **560**:494-498.
- 1146 106. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ: **Generalizing RNA velocity to**
1147 **transient cell states through dynamical modeling.** *Nat Biotechnol* 2020, **38**:1408-
1148 1414.
- 1149 107. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, 2nd, Treacy D, Trombetta JJ, Rotem A,
1150 Rodman C, Lian C, Murphy G, et al: **Dissecting the multicellular ecosystem of**
1151 **metastatic melanoma by single-cell RNA-seq.** *Science* 2016, **352**:189-196.
- 1152 108. Morabito S, Miyoshi E, Michael N, Shahin S, Martini AC, Head E, Silva J, Leavy K,
1153 Perez-Rosendahl M, Swarup V: **Single-nucleus chromatin accessibility and**
1154 **transcriptomic characterization of Alzheimer's disease.** *Nat Genet* 2021, **53**:1143-
1155 1155.
- 1156 109. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network**
1157 **analysis.** *BMC Bioinformatics* 2008, **9**:559.
- 1158 110. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing**
1159 **biological themes among gene clusters.** *OMICS* 2012, **16**:284-287.
- 1160 111. Stuart T, Srivastava A, Madad S, Lareau CA, Satija R: **Single-cell chromatin state**
1161 **analysis with Signac.** *Nat Methods* 2021, **18**:1333-1341.
- 1162 112. Gel B, Serra E: **karyoploteR: an R/Bioconductor package to plot customizable**
1163 **genomes displaying arbitrary data.** *Bioinformatics* 2017, **33**:3088-3090.
- 1164 113. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski
1165 B, Ideker T: **Cytoscape: a software environment for integrated models of**
1166 **biomolecular interaction networks.** *Genome Res* 2003, **13**:2498-2504.
- 1167 114. Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Lemma RB, Turchi L, Blanc-
1168 Mathieu R, Lucas J, Boddie P, Khan A, Manosalva Perez N, et al: **JASPAR 2022: the**
1169 **9th release of the open-access database of transcription factor binding profiles.**
1170 *Nucleic Acids Res* 2022, **50**:D165-D173.
- 1171 115. Schep AN, Wu B, Buenrostro JD, Greenleaf WJ: **chromVAR: inferring transcription-**
1172 **factor-associated accessibility from single-cell epigenomic data.** *Nat Methods* 2017,
1173 **14**:975-978.
- 1174 116. Speir ML, Bhaduri A, Markov NS, Moreno P, Nowakowski TJ, Papatheodorou I, Pollen

1175 AA, Raney BJ, Seninge L, Kent WJ, Haeussler M: **UCSC Cell Browser: visualize your**
1176 **single-cell data.** *Bioinformatics* 2021, **37**:4578-4580.
1177 117. Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, Goodstein DM, Elsik
1178 CG, Lewis SE, Stein L, Holmes IH: **JBrowse: a dynamic web platform for genome**
1179 **visualization and analysis.** *Genome Biol* 2016, **17**:66.
1180

1181 **Supplementary information**

1182 **Additional file 1: Table S1.** TE families, superfamilies and classes in human. **Table S2.**

1183 TE families, superfamilies and classes in mouse. **Table S3.** Genome coverage of TEs in
1184 human. **Table S4.** Genome coverage of TEs in mouse. **Table S5.** Genomic distribution of
1185 TEs in human. **Table S6.** Genomic distribution of TEs in mouse.

1186 **Additional file 2: Table S1.** Overlaps among TEs and CpG islands in human. **Table S2.**

1187 Overlaps among TEs and CpG islands in mouse. **Table S3.** Overlaps among TEs and
1188 promoter-like sites (PLS) in human. **Table S4.** Overlaps among TEs and proximal
1189 enhancer-like sites (pELS) in human. **Table S5.** Overlaps among TEs and distal enhancer-

1190 like sites (dELS) in human. **Table S6.** Overlaps among TEs and CTCF-only sites in human.

1191 **Table S7.** Overlaps among TEs and DNase-H3K4me3 sites in human. **Table S8.** Overlaps

1192 among TEs and promoter-like sites (PLS) in mouse. **Table S9.** Overlaps among TEs and
1193 proximal enhancer-like sites (pELS) in mouse. **Table S10.** Overlaps among TEs and distal
1194 enhancer-like sites (dELS) in mouse. **Table S11.** Overlaps among TEs and CTCF-only
1195 sites in mouse. **Table S12.** Overlaps among TEs and DNase-H3K4me3 sites in mouse.

1196 **Table S13.** Copy number of cCREs overlapped with TEs in human. **Table S14.** Copy
1197 number of cCREs overlapped with TEs in mouse.

1198 **Additional file 3: Table S1.** Human and mouse AGM datasets included in this study. **Table**

1199 **S2.** HUVEC bulk RNA-seq dataset (PRJNA561635) included in this study.

1200 **Additional file 4: Figure S1.** Steps to reconstruct the human EHT trajectory. **Figure S2.**

1201 Steps to reconstruct the mouse EHT trajectory. **Figure S3.** Integration of the human and

1202 mouse EHT scRNA-seq data.

1203 **Additional file 5: Table S1.** Marker TEs in human EHT. **Table S2.** Marker TEs in mouse

1204 EHT. **Table S3.** Marker genes in human EHT. **Table S4.** Marker genes in mouse EHT.

1205 **Additional file 6: Table S1.** Co-expressed gene and TE modules in human. **Table S2.** Co-

1206 expressed gene and TE modules in mouse. **Table S3.** Common TEs in HME5 and MME5.

1207 **Additional file 7: Table S1.** Differentially expressed TEs between pre-HECs and VECs in

1208 human. **Table S2.** Differentially expressed genes between pre-HECs and VECs in human.

1209 **Table S3.** Differentially expressed TEs between pre-HECs and HSCs in human. **Table S4.**

1210 Differentially expressed genes between pre-HECs and HSCs in human. **Table S5.**

1211 Differentially expressed TEs between pre-HECs and VECs in mouse. **Table S6.**

1212 Differentially expressed genes between pre-HECs and VECs in mouse. **Table S7.**

1213 Differentially expressed TEs between pre-HECs and HSCs in mouse. **Table S8.**

1214 Differentially expressed genes between pre-HECs and HSCs in mouse. **Table S9.** Average

1215 expression of TE silencers in human. **Table S10.** Average expression of TE silencers in

1216 mouse.

1217 **Additional file 8: Table S1.** Differentially expressed genes in human hematopoietic cells

1218 (HECs/HSCs) vs endothelial cells (VECs/AECs). **Table S2.** Differentially expressed genes

1219 in mouse hematopoietic cells (HECs/HSCs) vs endothelial cells (VECs/AECs). **Table S3.**

1220 Average expression of pattern recognition receptors and downstream signals in human.

1221 **Table S4.** Average expression of pattern recognition receptors and downstream signals in

1222 mouse. **Table S5.** GSEA analysis of genes in human hematopoietic cells (HECs/HSCs) vs
1223 endothelial cells (VECs/AECs). **Table S6.** GSEA analysis of genes in mouse hematopoietic
1224 cells (HECs/HSCs) vs endothelial cells (VECs/AECs).

1225 **Additional file 9: Figure S1.** Steps to reconstruct the mouse EHT trajectory from scATAC-
1226 seq data. **Figure S2.** Genome landscape of differentially accessible peaks (DAPs). **Figure**
1227 **S3.** The TF-target network in mouse EHT.

1228 **Additional file 10: Table S1.** Differentially accessible TEs between pre-HECs and
1229 VECs/AECs. **Table S2.** Differentially accessible TEs between pre-HECs and HECs/HSCs.
1230 **Table S3.** Cell type-specific open TEs. **Table S4.** Closest genes and annotations of
1231 differentially accessible peaks in mouse EHT. **Table S5.** Co-access scores of cell type-
1232 specific peak pairs. **Table S6.** Enriched motifs of TE associated cell type-specific peaks
1233 (TEPs). **Table S7.** Activated TF-target pairs during EHT (TFs are corresponding to TE-
1234 related motifs). **Table S8.** GO enrichment results of cell type-specific network modules of
1235 TE-bound TFs and downstream targets.

1236 **Additional file 11: Table S1.** GSEA analysis of pre-HECs vs other cell types. **Table S2.**
1237 Average expression of hypoxia-related genes in human EHT cell types. **Table S3.**
1238 Normalized counts of genes and TEs in the HUVEC dataset (PRJNA561635).

1239 **Additional file 12: Figure S1.** The hypoxic niche and TE expression in human AGM
1240 scRNA-seq data. **Figure S2.** Expression heatmap of each TE family (grouped into four TE
1241 classes) in the HUVEC bulk RNA-seq data. **Figure S3.** Five gene modules of pre-HEC
1242 markers on the HUVEC data.

1243