Novel markers for haemogenic endothelium and haematopoietic progenitors in the mouse yolk sac

Guillermo Diez-Pinel[1], Alessandro Muratore[1,2], Christiana Ruhrberg[1]*, Giovanni Canu[1]*

[1] UCL Institute of Ophthalmology, University College London, United Kingdom
[2] School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy

* Corresponding authors:
Christiana Ruhrberg: c.ruhrberg@ucl.ac.uk
Giovanni Canu: g.canu@ucl.ac.uk

KEYWORDS:
Haemogenic endothelium, erythro-myeloid progenitors, EMPs, haematopoietic development, haematopoietic differentiation, yolk sac, mouse embryo, single-cell transcriptomic, Fxyd5/Dysadherin, Scarf1, Fcer1g

ABSTRACT

Erythro-myeloid progenitors (EMPs) originate from the haemogenic endothelium in the yolk sac via an endothelial-to-haematopoietic transition (EHT) to generate blood and immune cells that support embryo development. Yet, the transitory nature of EHT and the limited availability of molecular markers has constrained our understanding of the origin, identity and differentiation dynamics of EMPs. Here, we have refined the annotation of yolk sac haemato-vascular populations in publicly available single-cell RNA sequencing (scRNAseq) datasets from mouse embryos to identify novel molecular markers of haemogenic endothelium and EMPs. By sub-clustering key cell populations followed by pseudotime analysis, we refined cluster annotations and then reconstructed differentiation trajectories. Subsequent differential gene expression analysis between clusters identified novel cell surface markers for haemogenic endothelial cells (*Fxyd5* and *Scarf1*) and EMPs (*Fcer1g*, *Tyrobp* and *Mctp1*). Further, we have identified candidate signalling and metabolic pathways that may regulate yolk sac haematopoietic emergence and differentiation. The specificity of FXYD5, SCARF1 and FCER1G for haemogenic endothelium and EMPs was validated by immunostaining of mouse yolk sac. These insights into the transcriptional dynamics in the yolk sac should support future investigation of EHT and haematopoietic differentiation during early mammalian development.

INTRODUCTION

Haematopoiesis in the developing embryo proceeds in three temporally and spatially overlapping waves in close proximity to vascular endothelial cells (Ivanovs *et al.*, 2017; Canu and Ruhrberg, 2021). The first wave, known as primitive haematopoiesis, arises in the yolk sac, when extra-embryonic mesoderm differentiates into vascular and haematopoietic progenitors that progressively organise into blood islands, and produces nucleated erythrocytes, megakaryocytes and macrophages (Ferkowicz and Yoder, 2005). This process occurs in the mouse from embryonic day (E) 7.0 onwards, equivalent to 2-3 post-conception weeks in human.

The second haematopoietic wave, known as pro-definitive or transient definitive haematopoiesis, also emerges in the yolk sac, when a change in cellular identity causes haematopoietic progenitors to bud off the endothelium as clusters of round cells, a process called endothelial-to-haematopoietic transition (EHT) (Frame *et al.*, 2016). In the mouse yolk sac, EHT generates mainly erythro-myeloid progenitors (EMPs) from E8.5, alongside lympho-myeloid progenitors (LMPs) from E9.5 (Böiers *et al.*, 2013; Frame, McGrath and Palis, 2013). Once the blood circulation is established by E9.5, these progenitors are transported out of the yolk sac and seed the foetal liver, from where they contribute blood and immune cells from mid-gestation until birth (Hoeffel *et al.*, 2015). Additionally, we recently showed that haematopoietic progenitors akin to EMPs can also arise from a *Pax3*-expressing lineage in the mouse and transiently seed the foetal liver by E12.5 to contribute erythroid and myeloid cells to foetal haematopoiesis (Canu *et al.*, 2024). Based on their origin from a *Pax3*-expressing progenitor, these haematopoietic cells may arise via EHT from a paraxial mesoderm-derived haemogenic endothelium, although further work is needed to address this possibility. Among the early haematopoietic populations, yolk sac-derived EMPs generate blood and immune cells important for embryogenesis (Frame, McGrath and Palis, 2013; Hoeffel *et al.*, 2015) and self-renewing tissue-resident macrophages that persist into adulthood (Moskalik, Niderla-Bielińska and Ratajska, 2021; Jiang *et al.*, 2024; Liu *et al.*, 2024).

The third haematopoietic wave, known as definitive haematopoiesis, originates intra-embryonically in the dorsal aorta, where EHT produces haematopoietic stem cells (HSCs) that enter the circulation to first colonise the foetal liver by mid-gestation and later seed the bone marrow to establish long-term, self-renewing adult haematopoiesis (Tavian *et al.*, 1996; Oberlin *et al.*, 2010; Ivanovs *et al.*, 2011; Canu and Ruhrberg, 2021; Sánchez-Lanzas, Jiménez-Pompa and Ganuza, 2024; Sugimura *et al.*, 2025). HSCs are first produced at E10.5 in the mouse, and 4-5 post-conception weeks in human.

71 Despite their key roles in embryonic haematopoiesis and the generation of tissue
72 macrophages with a role in adult homeostasis, the transitory nature of EMPs and their
73 overlapping markers with HSC lineages have made the study of EMP-derived cell lineages
74 challenging. For example, the Lin$^-$ KIT$^+$ CD41^{low} CD16/32$^+$ cell surface profile allows EMP
75 isolation by flow cytometry from the early embryo, but is insufficient to unambiguously
76 distinguish EMPs from HSC-derived myeloid progenitors at mid-gestation (McGrath *et al.*,
77 2015; Van Deren *et al.*, 2022). Temporally-controlled lineage tracing via Cre-mediated
78 reporter gene activation has therefore been used as a complementary approach to
79 distinguish EMPs from HSC-derived progenitors based on the earlier production of EMPs
80 (Gomez Perdiguero *et al.*, 2014; Hoeffel *et al.*, 2015; McGrath, Frame and Palis, 2016).
81 However, this approach relies on transgenic mice expressing suitable Cre drivers and
82 expression reporters. Therefore, specific molecular markers are needed to allow accurate
83 EMP isolation and to study their formation and behaviour in the mouse embryo. Moreover,
84 such markers might facilitate the identification of analogous markers suitable to study foetal
85 haematopoiesis in human embryos, in which genetic lineage tracing is not an option.
86 Here, we have leveraged publicly available mouse embryonic single-cell RNA
87 sequencing (scRNAseq) data to investigate cell populations in the developing mouse yolk
88 sac to identify novel candidate markers for haemogenic endothelium and EMPs.
89

90 **METHODS**

91 **Integration of publicly available E8.5 scRNAseq mouse embryo data**

92 The file 'embryo_sce.rds' containing a processed SingleCellExperiment object including
93 scRNAseq data from the ArrayExpress dataset E-MTAB-6967 (Pijuan-Sala *et al.*, 2019) and
94 ArrayExpress dataset E-MTAB-11763 (Goh *et al.*, 2023) was downloaded from
95 https://marionilab.github.io/ExtendedMouseAtlas/. The cells annotated as embryonic day
96 E8.5 were subset for analysis in R v4.5.1, with Rstudio v2025.05.1 using the Seurat v4.5.1
97 R package (Satija *et al.*, 2015). The function 'as.Seurat' was used to generate a Seurat
98 object from the SingleCellExperiment object. For sample integration, the E-MTAB-11763
99 cells were subset and split into layers containing individual samples, counts were log-
100 normalised with the 'NormalizeData' function, the top-2000 variable features were identified
101 with the 'FindVariableFeatures' and scaled and centred with the 'ScaleData' function. Then,
102 PCA dimensionality reduction was performed with the 'RunPCA' function and Harmony
103 Integration (Korsunsky *et al.*, 2019) was performed using the 'IntegrateLayers' function. A
104 Uniform Manifold Approximation and Projection (UMAP) reduction was obtained with the
105 'RunUMAP' function, for which the number of principal components was selected by

106 examination of an elbow plot. Lastly, the layers were joined with the 'JoinLayers' function.
107 This same integration pipeline was used for all other data integration in the study.
108

109 **Re-annotation of yolk sac cell types in E8.5 scRNAseq data**

110 Cells with anatomy 'Yolk sac' (YS) from haemato-vascular cell types were subset and
111 integrated by sample. For clustering, a k-nearest neighbour (KNN) graph was calculated (k
112 = 20), and the Louvain algorithm was used for graph-clustering. Then, the expression of
113 known cell type markers in the yolk sac clusters was used alongside the anatomical origin
114 (yolk sac or embryo) to annotate these haemato-vascular yolk sac cells as well as the
115 embryo haemato-vascular cells more accurately. Afterwards, both yolk sac and embryo cells
116 were merged and integrated by sample. These data were then used as the reference dataset
117 for reference-based annotation of the E-MTAB-6967 haemato-vascular cells using the
118 SingleR v2.10.0 R package (Aran *et al.*, 2019), and then merged and integrated by sample
119 with the E-MTAB-11763 data.
120

121 **Analysis of gene expression along the EHT differentiation trajectory**

122 To identify haemogenic endothelium cells, we sub-clustered all cells annotated as
123 endothelial cells and EMPs in the combined dataset, including Allantois endothelium,
124 Venous endothelium, Embryo proper endothelium, Endocardium, YS endothelium, Vitellin
125 vein endothelium, EMPs and EMP-like embryo, and then examined the expression of
126 haemato-vascular markers.To analyse the EHT differentiation trajectory, we sub-clustered
127 yolk sac haemato-vascular cells, including YS endothelium, haemogenic endothelium,
128 EMPs, BP1, BP2 and Erythroid, and the resulting UMAP reduction was used to carry out a
129 pseudotime analysis to predict differentiation trajectories with the Monocle3 v1.4.26 R
130 package (Cao *et al.*, 2019). Genes that changed as a function of pseudotime were identified
131 with the 'graph_test' function and grouped according to their expression patterns into
132 modules with the 'find_gene_modules', which were then grouped manually into four
133 supermodules (downwards trend, upwards trend, peaking at EMPs and peaking at BP1).
134

135 **Gene ontology and differential gene expression of haemato-vascular clusters**

136 Enrichment analyses against the GeneOntology: Biological Pathway and WikiPathways
137 databases were carried out using the 'enrichGO' and 'enrichWP' functions from the
138 clusterProfiler v4.16.0 R package (Wu *et al.*, 2021), respectively. All genes with at least 1
139 count were used as the background universe. Cytoscape 3.10.3 (Shannon *et al.*, 2003) and
140 the RCy3 v2.28.1 R package (Gustavsen *et al.*, 2019) were used to plot the TYROBP

141 WikiPathway pathway. For surface marker identification, differential expression analyses
142 were carried out with the 'FindMarkers' function, with an adjusted P-value threshold of 0.05.
143 Then, Uniprot accession IDs were retrieved from the org.Mm.eg.db v3.21.0 R package
144 (Carlson, 2025) using the 'select' function from the AnnotationDbi v1.70.0 R package (Pagès
145 *et al.*, 2025). A list of curated transmembrane proteins was obtained from Uniprot
146 (https://www.uniprot.org/) on the 30th of July 2025 using the query '(keyword:KW-0812)
147 AND (reviewed:true)', and filtering for 'Mouse' as the relevant organism. This list was then
148 used to filter the differential expression results for transmembrane proteins. Plotting of
149 UMAP reductions was carried out with Seurat's 'DimPlot' and the function
150 'DimPlot_scCustom' from the scCustomize v3.2.0 package (Marsh *et al.*, 2025). Plotting of
151 normalised gene expression onto UMAP reductions was carried out with Seurat's
152 'FeaturePlot' and scCustomize's 'FeaturePlot_scCustom' functions. Other plots were
153 generated using the ggplot2 v4.0.0 R package (Wickham *et al.*, 2025).
154

155 **Animal procedures and tissue staining**

156 Animal procedures were performed with Animal Welfare Ethical Review Body (AWERB) and
157 UK Home Office approval. C57BL/6J mice were timed-mated for embryo collection at E8.5
158 after cervical dislocation culling of the pregnant dam. Yolk sacs were fixed in 4%
159 formaldehyde, washed in phosphate-buffered saline (PBS) and then blocked in PBS
160 containing 2% serum-free protein block (DAKO), 2% bovine serum albumin and 0.4% Triton
161 X-100 before staining with the following antibodies: goat anti-mouse KIT (R&D Systems
162 #AF1356-SP, 1:50), rat anti-mouse PECAM1 (BD Pharmingen #553370, 1:50), rabbit anti-
163 mouse FXYD5/Dysadherin (Proteintech #12166-1-AP, 1:50), rabbit anti-mouse SCARF1
164 (Proteintech #13702-1-AP, 1:100) or rabbit anti-mouse FCER1G (GeneTex #GTX108487s,
165 1:50), followed by the appropriate secondary antibodies, which were Alexa Fluor 488-
166 conjugated donkey anti-goat Fab fragments (Stratech #705-547-003, 1:200), Cy3-
167 conjugated donkey anti-rat Fab fragments (Stratech #712-166-150, 1:200), and Alexa Fluor
168 647-conjugated donkey anti-rabbit Fab fragments (Stratech #711-607-003, 1:200).
169 Secondary-only controls were included but are not shown. DAPI-counterstained sections
170 were imaged on a LSM710 confocal microscope (Zeiss).
171
172
173
174

175 **RESULTS**

176

177 **Refined annotation of E8.5 yolk sac and embryo cell clusters**

178 To identify the transcriptomic signature of yolk sac haemogenic endothelium and that
179 of its haemopoietic progeny, we re-analysed published scRNAseq data from E8.5 mouse
180 embryos (**Fig. 1A**), which were extracted from the ArrayExpress datasets E-MTAB-
181 6967 (Pijuan-Sala *et al.*, 2019) and E-MTAB-11763 (Goh *et al.*, 2023). E-MTAB-
182 6967 contains data from E6.5 to E8.5 mouse embryos, including yolk sac and embryo
183 proper, which were sequenced and analysed together to construct a molecular atlas
184 describing early mouse gastrulation (Pijuan-Sala *et al.*, 2019). E-MTAB-11763 instead
185 contains data that were generated by collecting yolk sac and embryo proper from E8.5 to
186 E9.5 mouse embryos and sequencing them separately. These two published datasets
187 overall comprise 400,000 cells from E6.5 to E9.5. A subsequent *in silico* merge of E-MTAB-
188 11763 data and E-MTAB-6967 embryo and yolk sac data was previously analysed to
189 compare the mouse versus human yolk sac contribution to metabolic and nutritional support
190 as well as to early erythroid and myeloid lineages (Goh *et al.*, 2023). As the goal of our
191 analysis was to define the transcriptome of yolk sac endothelium and its haematopoietic
192 progeny at the time when EMPs emerge from yolk sac haemogenic endothelium, we subset
193 only the E8.5 data from E-MTAB-6967 and E-MTAB-11763 (**Fig. 1A**).

194 First, we analysed E8.5 data from E-MTAB-11763, which contained yolk sac and
195 embryo proper data annotated separately, and generated a UMAP including 75 clusters,
196 using the cluster annotations provided by the authors of the original publication (**Fig. S1**).
197 Next, we selectively sub-clustered the 8,453 haemato-vascular cells that had been
198 sequenced from the yolk sac (**Fig. 1B,C**). To confirm the identity of the resulting clusters,
199 we analysed the expression of known markers of endothelial and haematopoietic progenitor
200 populations (**Fig. 2A-D**). The yolk sac endothelial cell (EC) cluster was identified via
201 expression of the pan-EC markers *Kdr* and *Cdh5* together with *Lyve1* and *Stab2* (**Fig. 2A**)
202 as two markers typical for yolk sac ECs (Gordon, Gale and Harvey, 2008; Shvartsman *et*
203 *al.*, 2019). The EMP cluster was identified via detection of *Csf1r* and *Ptprc* as well as the
204 myeloid markers *Spi1 and Mpo* (**Fig. 2B**). Notably, *Kdr*, *Cdh5* and *Lyve1* were also detected
205 in the EMP cluster (**Fig. 2A**), agreeing with their known endothelial origin (Gomez
206 Perdiguero *et al.*, 2014). Clusters expressing haematopoietic progenitor markers (*Vav1*,
207 *Myb*, *Gata2*) (**Fig. 2C**) together with erythroid (*Gata1*, *Klf1*) or myeloid (*Spi1, Mpo)*
208 differentiation markers (**Fig. 2D**) represented intermediate progenitors at different states of

209 commitment and were annotated as blood progenitors (BP) 1 and BP2 (**Fig. 1C**). Finally,
210 high expression levels of differentiation markers (*Alas2*, *Gypa*) (**Fig. 2D**), in the absence of
211 endothelial (*Cdh5*, *Lyve1*), EMP (*Csf1r*, *Ptprc*) or progenitor markers (*Spi1*, *Gata2*),
212 identified a cluster of differentiated erythroid cells.

213 As expected, the detection levels of endothelial and haematopoietic genes displayed
214 opposite trends between the endothelial and haematopoietic clusters. Thus, *Kdr* and *Cdh5*
215 appeared highly expressed in ECs and EMPs but their detection declined in BP1, whilst they
216 were barely detectable in BP2 and absent from erythroid cells (**Fig. 2A**). Conversely,
217 erythroid genes such as *Klf1* and *Gypa* were not detected in ECs and EMPs, detected at
218 low levels in BP1, at higher levels in BP2 and at very high levels in erythroid cells (**Fig. 2D**).
219 Known haematopoietic progenitor markers (*e.g.*, *Myb*, *Gata2, Runx1*) displayed a gradual
220 decrease from EMPs to BP1 and BP2, supporting the notion that the three clusters
221 represented successive maturation stages (**Fig. 2C**). Myeloid markers (*e.g.*, *Csf1r, Spi1,*
222 *Mpo*) did not identify a separate cluster but were present in the EMP cluster and, at lower
223 levels, also in BP1 (**Fig. 2B**), agreeing with prior studies suggesting that macrophages are
224 only present in the yolk sac from E9.0 onwards (Hoeffel and Ginhoux, 2018).

225 Unexpectedly, we observed that E-MTAB-11763 contained cells that had been
226 isolated and sequenced from the embryo proper but nevertheless clustered with yolk sac
227 EMPs and BPs; accordingly, we annotated these cells as EMP-like embryo and blood
228 progenitors embryo, respectively (**Fig. 3A,B**). Also unexpectedly, some cells sequenced
229 from the embryo proper had been annotated in the original publication describing E-MTAB-
230 11763 as representing yolk sac-derived ECs. As the endothelium in the yolk sac vasculature
231 forms a continuum with the endothelium of the vitelline vein, cells within the embryo proper
232 that resemble yolk sac ECs likely represent vitelline vein ECs, and we re-annotated them
233 accordingly (**Fig. 3A,B**).

234 Next, we integrated E-MTAB-11763 with E-MTAB-6967 to increase the number of
235 cells available for downstream analysis in the haemato-vascular clusters to identify yolk sac
236 haemogenic endothelium. As cells derived from the yolk sac and embryo proper had not
237 been separately sequenced to generate E-MTAB-6967, we used a validated, reference-
238 based annotation method (Aran *et al.*, 2019) to harmonise cell type annotations and
239 distinguish cells from these sources before integrating the two datasets (**Fig. 3A**).
240 Specifically, this method uses gene expression information to assign the most likely cluster
241 from a reference dataset to a query dataset, which in our case involved applying our revised
242 annotation for E8.5 data in E-MTAB-11763 to the E8.5 data in E-MTAB-6967. This approach

243 increased the overall number of haemato-vascular cells from 8,453 to 12,351 for further
244 analysis (**Fig. 3B**). We also surveyed the expression of key known marker genes of the cell
245 populations of interest to confirm the appropriate annotation of the integrated dataset (**Fig.**
246 **3C-F**).
247

248 **Identifying yolk sac haemogenic ECs via their transcriptomic signature**

249 The haemogenic endothelium is a rare and transient population, and thus difficult to
250 study (Gritz and Hirschi, 2016; Scarfò *et al.*, 2024). To identify haemogenic ECs within the
251 yolk sac endothelial cluster, we subset EMPs and ECs from both yolk sac and embryo proper
252 and re-clustered them to generate a new UMAP including only these cells, which was
253 annotated according to our refined nomenclature. Using this approach, yolk sac ECs
254 clustered in an intermediate position between embryo-derived ECs and EMPs and also split
255 into two sub-clusters, one positioned closer to EMPs and the other closer to endocardial
256 ECs (**Fig. 4A**). The analysis of known marker genes showed that both sub-clusters
257 expressed markers of yolk sac endothelium at similar levels (*Lyve1*, *Stab2*), whereas only
258 the EMP-proximal sub-cluster co-expressed haematopoietic progenitor genes (*Runx1*, *Myb*,
259 *Spi1*) (**Fig. 4B**). Together, these findings indicate that a subset of yolk sac ECs have a dual
260 endothelial and haematopoietic cell identity, consistent with a presumed identify of
261 haemogenic ECs that give rise to EMPs.
262

263 **Pseudotime analysis identifies transcriptomic changes accompanying EHT**

264 Having identified a cell cluster with haemogenic endothelial identity, we next used our
265 refined annotation of the yolk sac populations to study signalling pathways and markers
266 defining the transition from haemogenic ECs into EMPs and BPs. When analysing samples
267 that contain cells along a continuous differentiation process, single-cell transcriptomic data
268 can be used to arrange cells along a pseudotime trajectory (Cannoodt, Saelens and Saeys,
269 2016). In essence, pseudotime analysis uses gene expression patterns to reorder cells and
270 arrange them based on their presumed progressive gene expression differences, thus
271 reconstructing trajectories along which cells may be differentiating. Thus, we used
272 pseudotime analysis of yolk sac cell populations to study the transition of ECs into
273 haematopoietic populations. In agreement with prior knowledge of EHT, pseudotime
274 analysis of our integrated E8.5 dataset suggested a continuum of cell differentiation from
275 endothelial to haemogenic endothelium and then to haematopoietic cells (**Fig. 5A**), and also
276 highlighted progressive erythroid maturation (**Fig. 5B**). A refined pseudotime analysis after

277 exclusion of erythroid cells as the predominant cell type corroborated the expected
278 progressive transition of yolk sac ECs to haemogenic ECs, EMPs, BP1 and BP2 (**Fig. 5C**).

279 Next, we took advantage of a method that analyses genes that change as a function
280 of pseudotime (Sharon *et al.*, 2019) to identify genes that showed an upward or downward
281 trend during EHT (**Fig. 5D**, **Table S1**). We found that 2,924 genes were progressively
282 downregulated, whereas 1,776 genes were progressively upregulated, during the predicted
283 transition from endothelial to haematopoietic cells. We also found that 231 genes were
284 upregulated during the predicted transition from ECs to EMPs, but downregulated during
285 subsequent differentiation towards BP1 and then BP2 (peak transcript levels in EMPs).
286 Moreover, 848 genes were progressively upregulated from ECs to BP1, but downregulated
287 in BP2 (peak transcript levels in BP1) (**Fig. 5D**).

288 To identify potential biological processes underpinning EHT, we used gene ontology
289 (GO) analysis on these 4 groups of co-regulated genes (**Fig. 5E-H**, **Table S2**). The GO
290 analysis of genes with a downward expression trend during the predicted progression from
291 endothelial to haematopoietic cells revealed enrichment in biological processes related to
292 angiogenesis, cytoskeleton re-organisation, cell adhesion, endothelial cell migration and
293 endothelial development (**Fig. 5E**). Genes with an upward expression trend instead were
294 involved in mRNA processing, splicing and ribonucleoprotein biogenesis (**Fig. 5F**). Genes
295 that peaked in EMPs showed enrichment in terms involved in leukocyte migration, myeloid
296 proliferation and immune response (**Fig. 5G**), in agreement with EMPs being responsible
297 for the production of foetal blood and immune cells. Finally, genes that peaked in BP1 are
298 involved in oxidative phosphorylation, cellular respiration and ATP synthesis (**Fig. 5H**),
299 raising the possibility that these cells have increased metabolism.

300 To identify genes belonging to known signalling pathways amongst each one of the
301 4 groups of co-regulated genes, we performed pathway analysis using the WikiPathways
302 resource (Agrawal *et al.*, 2024) (**Table S3**). Genes that were downregulated when ECs were
303 predicted to transition to haematopoietic cells belonged to the cholesterol biosynthesis, IL3,
304 IL17A, PI3K/AKT/mTOR and EFGR1 signalling pathways. Genes that were upregulated
305 during this transition are involved in RNA processing and cell cycle activation. Genes
306 peaking in BP1 are involved in the regulation of cell metabolism, whereas genes peaking in
307 EMPs belonged to the TYROBP and IL5 signalling pathways (**Table S3**). Notably, by
308 comparing EMPs to other yolk sac clusters, we identified multiple genes within the TYROBP
309 signalling pathway that, in addition to peaking in EMPs, were significantly enriched in EMPs
310 compared to the other cells in the haemato-vascular cluster (**Fig. S2**).

311 Overall, this analysis provided insights into molecular changes predicted to
312 accompany EHT and haematopoietic differentiation in the yolk sac, and identified candidate
313 signalling pathways that may be studied further to understand the molecular control of these
314 complex cellular differentiation pathways.
315

316 **Differential expression analysis identifies EMP and haemogenic endothelium markers**

317 To identify markers for haemogenic ECs and EMPs, we used our integrated E8.5
318 mouse dataset to perform differential expression (DE) analysis of endothelial and
319 haematopoietic cell populations in both the yolk sac and embryo. First, we subset the
320 clusters for both cell populations and then performed DE analyses by comparing clusters in
321 different combinations to identify the top differentially regulated genes. To identify genes
322 that may encode markers useful for cell identification via immunostaining or flow cytometric
323 cell sorting, we filtered DE genes for those that encoded transmembrane proteins (**Table**
324 **S4**).

325 By comparing EMPs to haematopoietic and endothelial cells, we identified *Csf1r* and
326 *Alox5ap* as enriched in EMPs (**Fig. 6A**), agreeing with previous studies that identified them
327 as EMP markers (Hoeffel *et al.*, 2015; Sharon *et al.*, 2019). In addition, we identified *Fcer1g*,
328 *Tyrobp* and *Mctp1* as potential new markers for EMPs (**Fig. 6A,B**), whereas *Fxyd5* and
329 *Scarf1* were identified as markers of yolk sac ECs and EMPs, and to a lower extent of BP1
330 (**Fig. 6B**). None of these seven genes were expressed in more differentiated BP2 and
331 erythroid cells. A comparison of genes expressed in haemogenic ECs versus EMPs showed
332 that acquisition of an EMP identity was associated with a loss of *Stab2* (**Fig. 6C**), as
333 previously described (Shvartsman *et al.*, 2019). Furthermore, we identified *Icam2*, *Cldn5*
334 and *Gpr182* as additional genes that were expressed in yolk sac ECs and haemogenic ECs
335 but downregulated in EMPs (**Fig. 6C**).

336 By comparing EMPs to BP1, and BP1 to BP2, we identified transcriptional changes
337 associated with the EMP transition to more committed progenitor states. This transition was
338 associated with the downregulation of the endothelial genes *Ecscr* and *Cdh5*, the EMP
339 marker *Alox5ap*, as well as the downregulation of our predicted novel EMP marker *Fcer1g*
340 and yolk sac EC and EMP markers *Fxyd5* and *Scarf1* (**Fig. 6A-D**). By contrast, the EMP
341 transition to more committed progenitor states was associated with the upregulation of the
342 megakaryocytic gene *F2r* and the erythroid gene *Smim1* (**Fig. 6D**).

343 To confirm that *Fxyd5*, *Scarf1* and *Fcer1g* encode EMP and haemogenic endothelial
344 markers, we performed immunofluorescence staining of E8.5 yolk sac. Although we had
345 found *Fxyd5* transcripts in yolk sac ECs and EMPs, FXYD5 staining was restricted to round

346 KIT$^+$ PECAM1$^+$ cells consistent with newly emerging EMPs, and there was no obvious
347 staining of ECs (**Figs. 6E, S3A**). Additionally, FXYD5 immunostaining identified round KIT$^-$
348 PECAM1^{low} cells, agreeing with our transcriptomic analysis detecting BPs with no *Kit*
349 transcripts and low levels of *Fxyd5* (**Fig. 6B**). Consistent with the presence of *Scarf1*
350 transcripts in both yolk sac ECs and EMPs, immunofluorescence staining identified SCARF1
351 protein in flat KIT$^-$ PECAM1$^+$ ECs and round KIT$^+$ PECAM1$^+$ emerging EMPs. Furthermore,
352 SCARF1 immunostaining identified flat KIT$^+$ PECAM1$^+$ cells that may be haemogenic ECs
353 (**Figs. 6F, S3B**). Although *Fcer1g* transcripts were detected in both EMPs and BP1,
354 immunofluorescence identified FCER1G protein within a subset of round KIT$^+$ cells.
355 Additionally, FCER1G immunostaining identified flat KIT$^-$ PECAM1$^+$ ECs, consistent with our
356 transcriptomic analysis detecting low levels of *Fcer1g* transcripts in yolk sac ECs (**Figs. 6G,**
357 **S3C**).

358 A comparison with clusters from the embryo showed that the markers identified in our
359 analysis as expressed in yolk sac ECs (*Fxyd5*, *Scarf1*, *Icam2*, *Cldn5*, *Gpr182*) were also
360 detected in the endocardium and in venous ECs (**Fig. 7A-C**). In contrast, the EMP markers
361 we identified (*Csf1r*, *Alox5ap*, *Fcer1g*, *Tyrobp*, *Mctp1*) were not expressed in ECs from the
362 embryo, confirming their specificity. Nevertheless, these markers were also expressed in
363 EMP-like cells from the embryo, thus corroborating their appropriate annotation.

364 Taken together, these findings show that novel EMP markers can be identified by
365 scRNAseq analysis and that they could be used to investigate both *bona fide* EMPs in the
366 yolk sac as well as poorly characterised EMP-like cells found in the embryo.

367

368 **DISCUSSION**

369 To capture the time when EMPs emerge from the yolk sac haemogenic endothelium
370 and to define the transcriptome of the cell populations involved in EHT, we re-analysed
371 transcriptomic data from the E8.5 mouse embryo (**Fig. 1**). Our analysis provided a refined
372 annotation of endothelial and haematopoietic populations in the mouse yolk sac (**Fig. 3A,B**)
373 and identified a subset of ECs co-expressing endothelial and haematopoietic markers,
374 consistent with a haemogenic endothelial phenotype (**Fig. 4A,B**). This refined annotation
375 also defined E8.5 yolk sac EMPs and demonstrated co-expression of myeloid, erythroid and
376 progenitor markers in this cluster, consistent with the known dual erythro-myeloid
377 differentiation potential of EMPs. Moreover, we identified EMP-like cells within the E8.5
378 embryo proper, which have not been highlighted in prior transcriptomic studies of early
379 haematopoiesis (**Fig. 3B**). Knowing that haematopoietic cells enter the embryo only from
380 E9.5, when the heart begins to beat to initiate blood circulation (Frame *et al.*, 2016; Kasaai

381 *et al.*, 2017), we considered two possibilities to explain the existence of EMP-like cells in the
382 embryo. On the one hand, these cells may represent yolk sac-derived haematopoietic
383 progenitors that had contaminated the embryonic samples during dissection. On the other
384 hand, it is conceivable that these cells represent EMP-like cells originating in the embryo
385 rather than in the yolk sac. The latter possibility is consistent with our recent discovery of a
386 *Pax3* lineage-derived EMP-like progenitor in the mouse embryo (Canu *et al.*, 2024).
387 However, further work is required to accurately describe these EMP-like cells.

388 Others have used scRNAseq to map potential differentiation trajectories for EMPs in
389 the yolk sac (Zhao *et al.*, 2023). Additionally, our analysis identified distinct BP1 and BP2
390 clusters as consecutive stages of EMP maturation (**Fig. 5A-C**) and provided new insight into
391 candidate signalling pathways controlling this transition (**Fig. 5D-H**). As primitive
392 erythrocytes are produced from E7.0 onwards (Palis *et al.*, 1999; Ferkowicz and Yoder,
393 2005), but EMPs only emerge from E8.5 onwards (Frame, McGrath and Palis, 2013; Gomez
394 Perdiguero *et al.*, 2014; Hoeffel *et al.*, 2015; McGrath *et al.*, 2015), the erythroid cluster in
395 the E8.5 dataset is expected to contain mostly primitive erythrocytes that were produced
396 prior to or alongside EMP emergence, with only a potential minor contribution from EMP-
397 derived erythrocytes. No markers of fully differentiated macrophages or other myeloid cell
398 types were observed, confirming previous studies showing that macrophages form in the
399 yolk sac from E9.0 onwards (Hoeffel and Ginhoux, 2018). We also did not detect a lymphoid
400 cluster at E8.5, in agreement with prior reports of LMP presence in the yolk sac only from
401 E9.5 onwards (Böiers *et al.*, 2013).

402 Previous studies established that EHT is associated with downregulation of
403 endothelial genes and concomitant upregulation of haematopoietic genes (Palis *et al.*, 1999;
404 Frame, McGrath and Palis, 2013; Gritz and Hirschi, 2016; Lilly *et al.*, 2016; Canu *et al.*,
405 2020). Here, pseudotime reconstruction highlighted dynamic gene expression trends during
406 EHT and haematopoietic maturation in the yolk sac (**Fig. 5A-D**), whereby haemogenic ECs
407 form a transcriptional continuum bridging the transition from endothelial to haematopoietic
408 cell identities, thus implying a gradual and continuous process rather than a discrete lineage
409 switch. Altogether, these findings align with established models of yolk sac haematopoiesis
410 derived from genetic lineage tracing, flow cytometry experiments and prior transcriptomic
411 analysis (Frame, McGrath and Palis, 2013; Hoeffel and Ginhoux, 2018; Canu and Ruhrberg,
412 2021; Sánchez-Lanzas, Jiménez-Pompa and Ganuza, 2024; Sugimura *et al.*, 2025).

413 The progressive downregulation of genes associated with angiogenesis, cytoskeletal
414 organisation and endothelial adhesion during the predicted EHT (**Fig. 5E**) may indicate
415 structural and functional changes necessary for cell detachment and transition to a

416 haematopoietic phenotype. Conversely, the upregulation of genes associated with RNA
417 processing (**Fig. 5F**) suggests that haematopoietic specification is coupled to enhanced
418 transcriptional activity. Furthermore, upregulation of genes associated to cell cycle activation
419 and transition from the G1 to S phase (**Table S3**) agrees with previous findings that EHT
420 requires cell cycle progression and is controlled by the cell cycle regulators CDK4/6 and
421 CDK1 (Marcelo *et al.*, 2013; Canu *et al.*, 2020).

422 DE analysis identified components of the TYROBP signalling pathway as significantly
423 enriched in EMPs (**Fig. S2**). In the adult, the activation of the TYROBP signalling pathway
424 has been associated with the microglial response to stress (Haure-Mirande *et al.*, 2022;
425 Zhang *et al.*, 2024). However, to the best of our knowledge, this signalling pathway has not
426 been previously associated with EMP emergence. Further investigation should determine
427 whether this pathway may be functionally involved in the production of EMPs or other
428 haematopoietic progenitors (*e.g.*, HSCs from the dorsal aorta) and whether manipulating
429 TYROBP *in vitro* during differentiation of induced pluripotent stem cells may allow to devise
430 more efficient haematopoietic differentiation methods.

431 DE analysis also identified upregulation of genes associated with oxidative
432 phosphorylation upon differentiation of EMPs to BP1 (**Fig. 5H**). Previous studies of
433 pluripotent stem cells and neuronal progenitors showed that their differentiation is
434 associated with metabolic reprogramming, whereby differentiating progenitors switch from
435 glycolysis to oxidative phosphorylation (Zheng *et al.*, 2016; Dahan *et al.*, 2019; Mahmoud,
436 2023). Our analysis raises the possibility that an analogous process may occur when
437 haematopoietic progenitors in the yolk sac commit to differentiation.

438 Our DE analysis identified both known and previously undescribed candidate
439 markers associated with the formation of haemogenic endothelium and EMPs (**Fig. 6A-D**).
440 Established EMP markers such as *Csf1r* and *Alox5ap* were previously described by others
441 (Hoeffel *et al.*, 2015; Sharon *et al.*, 2019), thus validating the robustness of our
442 transcriptomic analysis. In addition, we identified *Fcer1g*, *Tyrobp*, and *Mctp1* as potential
443 novel EMP markers, and *Fxyd5* and *Scarf1* as shared markers between yolk sac endothelial
444 and haematopoietic cells. The expression of *Fxyd5* and *Scarf1* in endothelial and EMP
445 clusters, and to a lower extent in BP1, but not in more differentiated cells, suggests that
446 these genes may serve as new markers for yolk sac ECs and emerging EMPs.
447 Immunofluorescence staining of E8.5 yolk sac corroborated the predictions of our DE
448 analysis: FXYD5 was present on round KIT^+ $PECAM1^+$ cells consistent with emerging EMPs
449 (**Figs. 6E, S3A**); SCARF1 labelled flat KIT^- $PECAM1^+$ ECs, flat KIT^+ $PECAM1^+$ cells
450 consistent with haemogenic ECs and round KIT^+ $PECAM1^+$ cells consistent with emerging

451 EMPs (**Figs. 6F, S3B**); FCER1G marked flat KIT$^-$ PECAM1$^+$ ECs and a few round KIT$^+$
452 PECAM1$^+$ cells that may represent a subset of EMPs (**Figs. 6G, S3C**). Nevertheless, these
453 experiments also highlighted differences in RNA and protein expression. Thus, although
454 *Fxyd5* RNA was detected in yolk sac ECs, haemogenic ECs and EMPs, and at lower levels
455 in BP1, the FXYD5 protein appeared to be prevalent in round KIT$^+$ PECAM1$^+$ cells consistent
456 with EMPs and in round KIT$^-$ PECAM1^{low} cells presumed to be myeloid-committed BP1
457 (**Figs. 6E, S3A**). While *Fcer1g* RNA clearly identified the EMP cluster, the FCER1G protein
458 was detected in only a subset of round KIT$^+$ PECAM1$^+$ cells, possibly marking more mature
459 EMPs (**Figs. 6G, S3C**). This is reminiscent of observations for *Csf1r*, a gene expressed in
460 EMPs at the RNA level, whereas CSF1R protein only becomes translated upon
461 differentiation to myeloid cells (Hoeffel *et al.*, 2015). These discrepancies between RNA and
462 protein expression may be due differences in expression dynamics and half-lives for mRNAs
463 and proteins (Vogel and Marcotte, 2012). Overall, these immunostaining results validated
464 our transcriptomic predictions and highlight the utility of single-cell datasets in discovering
465 functional cell surface markers. Previously used markers for haemogenic ECs include the
466 surface markers CDH5, KIT and CD41, and haematopoietic transcription factors such as
467 RUNX1 (Mikkola *et al.*, 2003; Nadin, Goodell and Hirschi, 2003; Lilly *et al.*, 2016), whereas
468 previously used markers for EMPs include KIT, CD41 and CD16/32, as well as the gene
469 expression reporter *Csf1r*-eGFP (McGrath *et al.*, 2011; Azzoni *et al.*, 2018; Plein *et al.*,
470 2018). The addition of the novel markers identified in our analysis will provide
471 complementary tools to define haemogenic ECs and EMPs in the mouse yolk sac.
472 In summary, this study refines the cellular and molecular characterisation of yolk sac
473 haematopoiesis, provides a detailed transcriptomic roadmap of EHT in the yolk sac and
474 identifies novel candidate markers for haemogenic endothelium and EMPs. The
475 identification of FXYD5, SCARF1 and FCER1G as markers for EMPs and haemogenic ECs
476 may offers valuable tools for future *in vivo* and *in vitro* studies. Finally, future studies may
477 wish to determine whether the novel EMP markers are present also in foetal HSCs, and
478 whether they are conserved in human haemopoiesis. Defining the specificity of such novel
479 markers may allow researchers to better define EMP identity and lineage potential as a
480 prerequisite for deciphering the ontogeny of the immune system and, ultimately, to harness
481 early haematopoietic pathways for regenerative medicine.
482

483 **ACKNOWLEDGEMENTS**

484 We thank the staff of the Biological Resources, FACS and Imaging Facilities at the UCL

Institute of Ophthalmology for their technical support. This research was supported by grants from Wellcome (205099/Z/16/Z) and the British Heart Foundation (PG23/11301).

AUTHORSHIP CONTRIBUTIONS

GC, GDP and CR conceived and designed the study. GC and CR co-wrote the manuscript. GC and AM performed immunostaining experiments. GDP performed bioinformatic analyses. GC and CR supervised the project. All authors read and approved the submitted manuscript.

DISCLOSURE OF CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

498 **REFERENCES**

499 Agrawal, A. *et al.* (2024) 'WikiPathways 2024: next generation pathway database', *Nucleic*
500 *Acids Research*, 52(D1), pp. D679–D689.

501 Aran, D. *et al.* (2019) 'Reference-based analysis of lung single-cell sequencing reveals a
502 transitional profibrotic macrophage', *Nature Immunology 2019 20:2*, 20(2), pp. 163–172.

503 Azzoni, E. *et al.* (2018) 'Kit ligand has a critical role in mouse yolk sac and aorta–gonad–
504 mesonephros hematopoiesis', *EMBO reports*, 19(10).

505 Böiers, C. *et al.* (2013) 'Lymphomyeloid Contribution of an Immune-Restricted Progenitor
506 Emerging Prior to Definitive Hematopoietic Stem Cells', *Cell Stem Cell*, 13(5), pp. 535–
507 548.

508 Cannoodt, R., Saelens, W. and Saeys, Y. (2016) 'Computational methods for trajectory
509 inference from single-cell transcriptomics', *European Journal of Immunology*, 46(11), pp.
510 2496–2506.

511 Canu, G. *et al.* (2020) 'Analysis of endothelial-to-haematopoietic transition at the single cell
512 level identifies cell cycle regulation as a driver of differentiation', *Genome Biology*, 21(1),
513 pp. 1–28.

514 Canu, G. *et al.* (2024) 'A Pax3 lineage gives rise to transient haematopoietic progenitors',
515 *Development (Cambridge, England)*, 151(23).

516 Canu, G. and Ruhrberg, C. (2021) 'First blood: the endothelial origins of hematopoietic
517 progenitors', *Angiogenesis 2021 24:2*, 24(2), pp. 199–211.

518 Cao, J. *et al.* (2019) 'The single-cell transcriptional landscape of mammalian
519 organogenesis', *Nature 2019 566:7745*, 566(7745), pp. 496–502.

520 Carlson, M. (2025) *org.Mm.eg.db - Genome wide annotation for Mouse*, *Bioconductor*.

521 Dahan, P. *et al.* (2019) 'Metabolism in pluripotency: Both driver and passenger?', *Journal*
522 *of Biological Chemistry*, 294(14), pp. 5420–5429.

523 Van Deren, D.A. *et al.* (2022) 'Defining the Hoxb8 cell lineage during murine definitive
524 hematopoiesis', *Development (Cambridge, England)*, 149(8), p. dev200200.

525 Ferkowicz, M.J. and Yoder, M.C. (2005) 'Blood island formation: Longstanding
526 observations and modern interpretations', *Experimental Hematology*, 33(9 SPEC. ISS.),
527 pp. 1041–1047.

528 Frame, J.M. *et al.* (2016) 'Definitive Hematopoiesis in the Yolk Sac Emerges from Wnt-
529 Responsive Hemogenic Endothelium Independently of Circulation and Arterial Identity',
530 *Stem Cells*, 34(2), pp. 431–444.

531 Frame, J.M., McGrath, K.E. and Palis, J. (2013) 'Erythro-myeloid progenitors: "definitive"
532 hematopoiesis in the conceptus prior to the emergence of hematopoietic stem cells', *Blood*
533 *cells, molecules & diseases*, 51(4), pp. 220–225.

534 Goh, I. *et al.* (2023) 'Yolk sac cell atlas reveals multiorgan functions during human early
535 development', *Science*, 381(6659).

536 Gomez Perdiguero, E. *et al.* (2014) 'Tissue-resident macrophages originate from yolk-sac-
537 derived erythro-myeloid progenitors', *Nature 2014 518:7540*, 518(7540), pp. 547–551.

538 Gordon, E.J., Gale, N.W. and Harvey, N.L. (2008) 'Expression of the hyaluronan receptor
539 LYVE-1 is not restricted to the lymphatic vasculature; LYVE-1 is also expressed on
540 embryonic blood vessels', *Developmental Dynamics*, 237(7), pp. 1901–1909.

541 Gritz, E. and Hirschi, K.K. (2016) 'Specification and function of hemogenic endothelium
542 during embryogenesis', *Cellular and molecular life sciences : CMLS*, 73(8), pp. 1547–
543 1567.

544 Gustavsen, J.A. *et al.* (2019) 'RCy3: Network biology using Cytoscape from within R',
545 *F1000Research 2019 8:1774*, 8, p. 1774.

546 Haure-Mirande, J.V. *et al.* (2022) 'Microglial TYROBP/DAP12 in Alzheimer's disease:
547 Transduction of physiological and pathological signals across TREM2', *Molecular*
548 *Neurodegeneration 2022 17:1*, 17(1), pp. 1–15.

549 Hoeffel, G. *et al.* (2015) 'C-Myb+ Erythro-Myeloid Progenitor-Derived Fetal Monocytes
550 Give Rise to Adult Tissue-Resident Macrophages', *Immunity*, 42(4), pp. 665–678.

551 Hoeffel, G. and Ginhoux, F. (2018) 'Fetal monocytes and the origins of tissue-resident
552 macrophages', *Cellular Immunology*, 330, pp. 5–15.

553 Ivanovs, A. *et al.* (2011) 'Highly potent human hematopoietic stem cells first emerge in the
554 intraembryonic aorta-gonad-mesonephros region', *Journal of Experimental Medicine*,
555 208(12), pp. 2417–2427.

556 Ivanovs, A. *et al.* (2017) 'Human haematopoietic stem cell development: from the embryo
557 to the dish', *Development*, 144(13), pp. 2323–2337.

558 Jiang, Y. *et al.* (2024) 'Macrophages in organ fibrosis: from pathogenesis to therapeutic

559 targets', *Cell Death Discovery 2024 10:1*, 10(1), pp. 1–11.

560 Kasaai, B. *et al.* (2017) 'Erythro-myeloid progenitors can differentiate from endothelial cells
561 and modulate embryonic vascular remodeling', *Scientific Reports*, 7(1), pp. 1–12.

562 Korsunsky, I. *et al.* (2019) 'Fast, sensitive and accurate integration of single-cell data with
563 Harmony', *Nature Methods 2019 16:12*, 16(12), pp. 1289–1296.

564 Lilly, A.J. *et al.* (2016) 'Interplay between SOX7 and RUNX1 regulates hemogenic
565 endothelial fate in the yolk sac', *Development (Cambridge, England)*, 143(23), pp. 4341–
566 4351.

567 Liu, X. *et al.* (2024) 'Inducible CCR2+ nonclassical monocytes mediate the regression of
568 cancer metastasis', *The Journal of clinical investigation*, 134(22).

569 Mahmoud, A.I. (2023) 'Metabolic switches during development and regeneration',
570 *Development (Cambridge)*, 150(20).

571 Marcelo, K.L. *et al.* (2013) 'Hemogenic endothelial cell specification requires c-Kit, notch
572 signaling, and p27-mediated cell-cycle control', *Developmental Cell*, 27(5), pp. 504–515.

573 Marsh, S. *et al.* (2025) *scCustomize: Custom Visualizations & Functions for Streamlined*
574 *Analyses of Single Cell Sequencing*, *CRAN: Contributed Packages*.

575 McGrath, K.E. *et al.* (2011) 'Definitive Erythro-Myeloid Progenitors (EMP) Emerge in the
576 Yolk Sac From Hemogenic Endothelium and Share Transcriptional Regulators with Adult
577 Hematopoiesis', *Blood*, 118(21), p. 910.

578 McGrath, K.E. *et al.* (2015) 'Distinct sources of hematopoietic progenitors emerge before
579 HSCs and provide functional blood cells in the mammalian embryo', *Cell reports*, 11(12),
580 p. 1892.

581 McGrath, K.E., Frame, J.M. and Palis, J. (2016) 'Early hematopoiesis and macrophage
582 development', *Seminars in immunology*, 27(6), p. 379.

583 Mikkola, H.K.A. *et al.* (2003) 'Expression of CD41 marks the initiation of definitive
584 hematopoiesis in the mouse embryo', *Blood*, 101(2), pp. 508–516.

585 Moskalik, A., Niderla-Bielińska, J. and Ratajska, A. (2021) 'Multiple roles of cardiac
586 macrophages in heart homeostasis and failure', *Heart Failure Reviews*, 27(4), p. 1413.

587 Nadin, B.M., Goodell, M.A. and Hirschi, K.K. (2003) 'Phenotype and hematopoietic
588 potential of side population cells throughout embryonic development', *Blood*, 102(7), pp.
589 2436–2443.

590 Oberlin, E. *et al.* (2010) 'Definitive human and mouse hematopoiesis originates from the
591 embryonic endothelium: a new class of HSCs based on VE-cadherin expression', *The*
592 *International Journal of Developmental Biology*, 54(6–7), pp. 1165–1173.

593 Pagès, H. *et al.* (2025) *AnnotationDbi - Manipulation of SQLite-based annotations in*
594 *Bioconductor*, *Bioconductor*.

595 Palis, J. *et al.* (1999) 'Development of erythroid and myeloid progenitors in the yolk sac
596 and embryo proper of the mouse', *Development*, 126(22), pp. 5073–5084.

597 Pijuan-Sala, B. *et al.* (2019) 'A single-cell molecular map of mouse gastrulation and early
598 organogenesis', *Nature 2019 566:7745*, 566(7745), pp. 490–495.

599 Plein, A. *et al.* (2018) 'Erythro-myeloid progenitors contribute endothelial cells to blood
600 vessels', *Nature 2018 562:7726*, 562(7726), pp. 223–228.

601 Sánchez-Lanzas, R., Jiménez-Pompa, A. and Ganuza, M. (2024) 'The evolving
602 hematopoietic niche during development', *Frontiers in Molecular Biosciences*, 11, p.
603 1488199.

604 Satija, R. *et al.* (2015) 'Spatial reconstruction of single-cell gene expression data', *Nature*
605 *Biotechnology 2015 33:5*, 33(5), pp. 495–502.

606 Scarfò, R. *et al.* (2024) 'CD32 captures committed haemogenic endothelial cells during
607 human embryonic development', *Nature Cell Biology 2024 26:5*, 26(5), pp. 719–730.

608 Shannon, P. *et al.* (2003) 'Cytoscape: A Software Environment for Integrated Models of
609 Biomolecular Interaction Networks', *Genome Research*, 13(11), p. 2498.

610 Sharon, N. *et al.* (2019) 'A Peninsular Structure Coordinates Asynchronous Differentiation
611 with Morphogenesis to Generate Pancreatic Islets', *Cell*, 176(4), pp. 790-804.e13.

612 Shvartsman, M. *et al.* (2019) 'Single-cell atlas of major haematopoietic tissues sheds light
613 on blood cell formation from embryonic endothelium', *bioRxiv*, p. 774547.

614 Sugimura, R. *et al.* (2025) 'Editorial: Plasticity of the haematopoietic niche: from embryonic
615 development to aging and disease', *Frontiers in Molecular Biosciences*, 12, p. 1683902.

616 Tavian, M. *et al.* (1996) 'Aorta-Associated CD34+ Hematopoietic Cells in the Early Human
617 Embryo', *Blood*, 87(1), pp. 67–72.

618 Vogel, C. and Marcotte, E.M. (2012) 'Insights into the regulation of protein abundance from
619 proteomic and transcriptomic analyses', *Nature Reviews Genetics 2012 13:4*, 13(4), pp.
620 227–232.

621 Wickham, H. *et al.* (2025) ‘ggplot2: Create Elegant Data Visualisations Using the Grammar
622 of Graphics’, *CRAN: Contributed Packages* [Preprint].

623 Wu, T. *et al.* (2021) ‘clusterProfiler 4.0: A universal enrichment tool for interpreting omics
624 data’, *Innovation*, 2(3), p. 100141.

625 Zhang, M. *et al.* (2024) ‘TYROBP serve as potential immune-related signature genes in
626 the acute phase of intracerebral hemorrhage’, *Scientific Reports 2024 14:1*, 14(1), pp. 1–
627 12.

628 Zhao, Y.X. *et al.* (2023) ‘Single-cell RNA sequencing-guided fate-mapping toolkit
629 delineates the contribution of yolk sac erythro-myeloid progenitors’, *Cell Reports*, 42(11),
630 p. 113364.

631 Zheng, X. *et al.* (2016) ‘Metabolic reprogramming during neuronal differentiation from
632 aerobic glycolysis to neuronal oxidative phosphorylation’, *eLife*, 5(JUN2016), p. e13374.

633

FIGURE LEGENDS

Fig 1. Re-analysis of E8.5 mouse embryonic scRNAseq data to study yolk sac EHT.

A Workflow for the integrated re-analysis of mouse embryonic scRNAseq datasets E-MTAB-11763 and E-MTAB-6967.

B UMAP of dataset E-MTAB-11763 showing cells sequenced from the E8.5 mouse embryo and labelled according to their anatomical origin as embryo and yolk sac.

C UMAP obtained by subsetting haemato-vascular cells from E8.5 yolk sac from E-MTAB-11763 and labelled based on our refined cluster annotation.

EMP: erythro-myeloid progenitor; EC: endothelial cell; HE: haemogenic endothelium; YS: yolk sac.

Fig 2. Marker analysis identifies haemato-vascular cell populations in E8.5 yolk sac data in E-MTAB-11763.

UMAP representation of transcript levels for the indicated marker genes for (**A**) endothelial cells, (**B**) EMPs, (**C**) haematopoietic progenitors and (**D**) erythroid cells. The scale bar represents log-normalised transcript levels.

Fig 3. Refined cluster annotation of E8.5 embryo and yolk sac haemato-vascular cell clusters after integration of E-MTAB-11763 and E-MTAB-6967.

A Reference-based annotation based on gene expression was used to apply our revised annotation for E8.5 data in E-MTAB-11763 to the E8.5 data in E-MTAB-6967 before integrating the two datasets.

B UMAPs showing the E8.5 haemato-vascular clusters from both embryo and yolk sac for dataset E-MTAB-11763 before integration to dataset E-MTAB-6967 and after reference-mapping and integration of the two datasets.

C-F UMAP representation of transcript levels for the indicated marker genes for (**C**) endothelial cells, (**D**) EMPs, (**E**) haematopoietic progenitors and (**F**) erythroid cells. The scale bar represents log-normalised transcript levels.

Fig 4. Endothelial cell and EMP sub-clustering identifies haemogenic endothelium.

A UMAP obtained after subsetting and re-clustering all cells annotated as endothelial cells and EMPs after integration of E8.5 data in E-MTAB-11763 and E-MTAB-6967. The dotted box indicates an area that is zoomed on the cells previously annotated as yolk sac endothelial cells, vitellin vein endothelial cells and EMPs; note the formation of a separate cluster positioned between EMPs and the main yolk sac endothelial cluster consistent with a haemogenic endothelial identify.

B UMAP representation of key endothelial and haematopoietic markers corroborates the identification of the haemogenic endothelial cluster. The scale bars represent log-normalised transcript levels.

YS: yolk sac; EMPs: erythro-myeloid progenitors.

Fig 5. Pseudotime analysis identifies co-regulated genes and enriched GO pathways during EHT.

A UMAP obtained after subsetting and re-clustering yolk sac haemato-vascular cells after integration of E8.5 data in E-MTAB-11763 and E-MTAB-6967, and showing our refined cluster annotation.

B,C Pseudotime analysis highlights erythroid generation and progressive maturation (**B**) and, after excluding the erythroid cluster (**C**), the progressive differentiation from yolk sac endothelial cells to haemogenic endothelial cells and then EMPs followed by blood progenitors 1 and 2. The scale bars represent pseudotime.

D Genes that changed as a function of pseudotime were identified and grouped according to their expression patterns. Heatmap shows 4 groups of co-regulated genes during differentiation from yolk sac endothelium to blood progenitors 2.

E-H Gene ontology analysis for each group of co-regulated genes.

YS: yolk sac; EMP: erythro-myeloid progenitor; BP1: blood progenitors 1.

Fig 6. Differential gene (DE) expression analysis and immunostaining validation identifies new markers for haemogenic endothelial cells and EMPs.

A-D UMAP representation of transcript levels for the indicated marker genes identified by DE analysis between the different yolk sac haemato-vascular clusters. The scale bars represent log-normalised transcript levels.

E-G Immunofluorescence staining of E8.5 yolk sac for the EMP marker KIT together with PECAM1 and FXYD5, SCARF1 or FCER1G. Emerging EMPs are indicated with arrows (round KIT+ PECAM1+ cells), presumed blood (myeloid) progenitors with arrowheads (round KIT- PECAM1low cells), endothelial cells with empty arrows (flat KIT- PECAM1+ cells), haemogenic endothelial cells with empty arrowheads (flat KIT+ PECAM1+ cells) and mature EMPs with asterisk (round KIT+ PECAM1+ FCER1G+ cells). Scale bars: 50 µm; n = 3 embryos.

Fig 7. Expression analysis in haemato-vascular cells from both yolk sac and embryo shows specificity of the new markers for haemogenic ECs and EMPs.

A,B Dot plot representation of markers identified by DE analysis in (**A**) yolk sac and (**B**) embryo haemato-vascular clusters. The dot size reflects the percentage of cells in each cluster expressing the gene. The scale bar indicates log-normalised gene expression.

C UMAP of the E8.5 haemato-vascular clusters from both embryo and yolk sac showing transcript levels for the indicated markers identified by DE analysis. The scale bars represent log-normalised transcript levels.

Fig S1. Cluster annotation of E8.5 mouse scRNAseq data in E-MTAB-11763.

UMAP of the E8.5 cells in E-MTAB-11763 dataset, labelled with the cluster annotations provided by the authors of the original publication (Goh et al., 2023).

Fig S2. The TYROBP signalling pathway is enriched in emerging EMPs.

TYROBP pathway genes enriched in EMPs compared to haemato-vascular clusters from both yolk sac and embryo. The box colour indicates the average log2 fold change. A green border indicates significant enrichment in EMPs.

Fig S3. Immunostaining validation of new markers for haemogenic endothelial cells and EMPs.

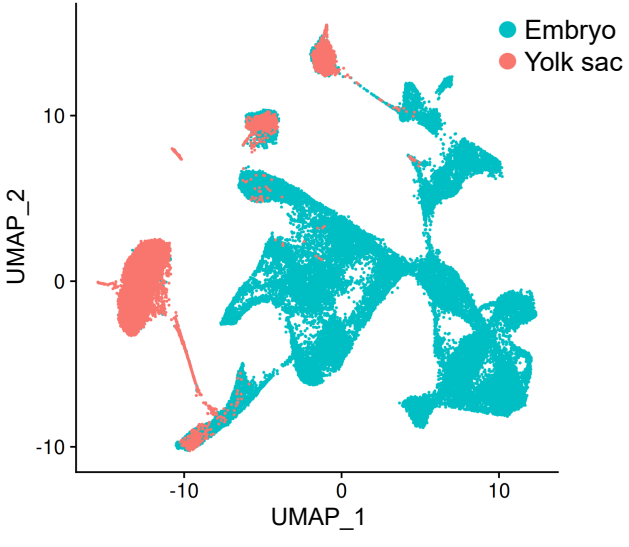
A-C Immunofluorescence staining of E8.5 yolk sac for the EMP marker KIT together with PECAM1 and FXYD5, SCARF1 or FCER1G. Emerging EMPs are indicated with arrows (round KIT+ PECAM1+ cells), presumed blood (myeloid) progenitors with arrowheads (round KIT- PECAM1low cells), endothelial cells with empty arrows (flat KIT- PECAM1+ cells), haemogenic endothelial cells with empty arrowheads (flat KIT+ PECAM1+ cells) and mature EMPs with asterisk (round KIT+ PECAM1+ FCER1G+ cells). Scale bars: 50 µm; n = 3 embryos.

Figure 1

A

Workflow	Objective
Subset E8.5 yolk sac cells from dataset E-MTAB-11763	Refine cluster annotation
Identify marker gene expression across subclusters in E-MTAB-11763	
Merge E-MTAB-11763 with E-MTAB-6967 based on new annotation	Obtain new integrated dataset
Subcluster EMPs + ECs in the merged dataset and survey markers	Define haemogenic ECs
Pseudotime analysis of the merged dataset	Observe differentiation trajectory
Analysis of gene expression variation as a function of pseudotime	Identify co-regulated genes
Differential expression analysis of haematopoietic and endothelial clusters	Identify new HE/EMP markers
Immunostaining of E8.5 yolk sac with candidate new HE/EMP markers	Validate new HE/EMP markers

B



C

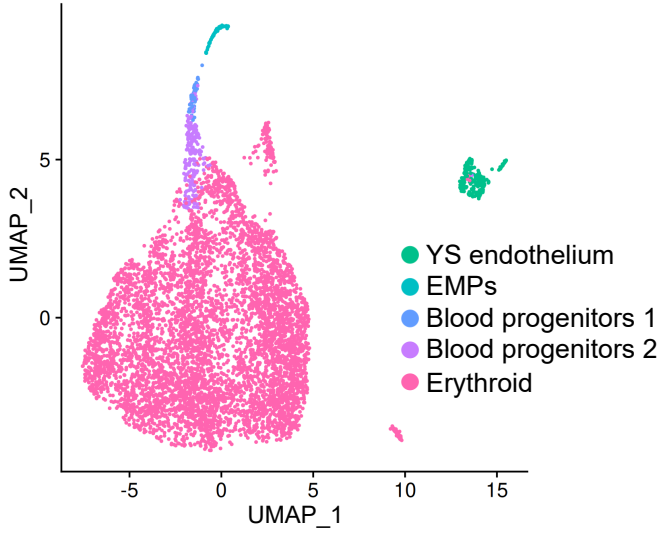


Figure 2

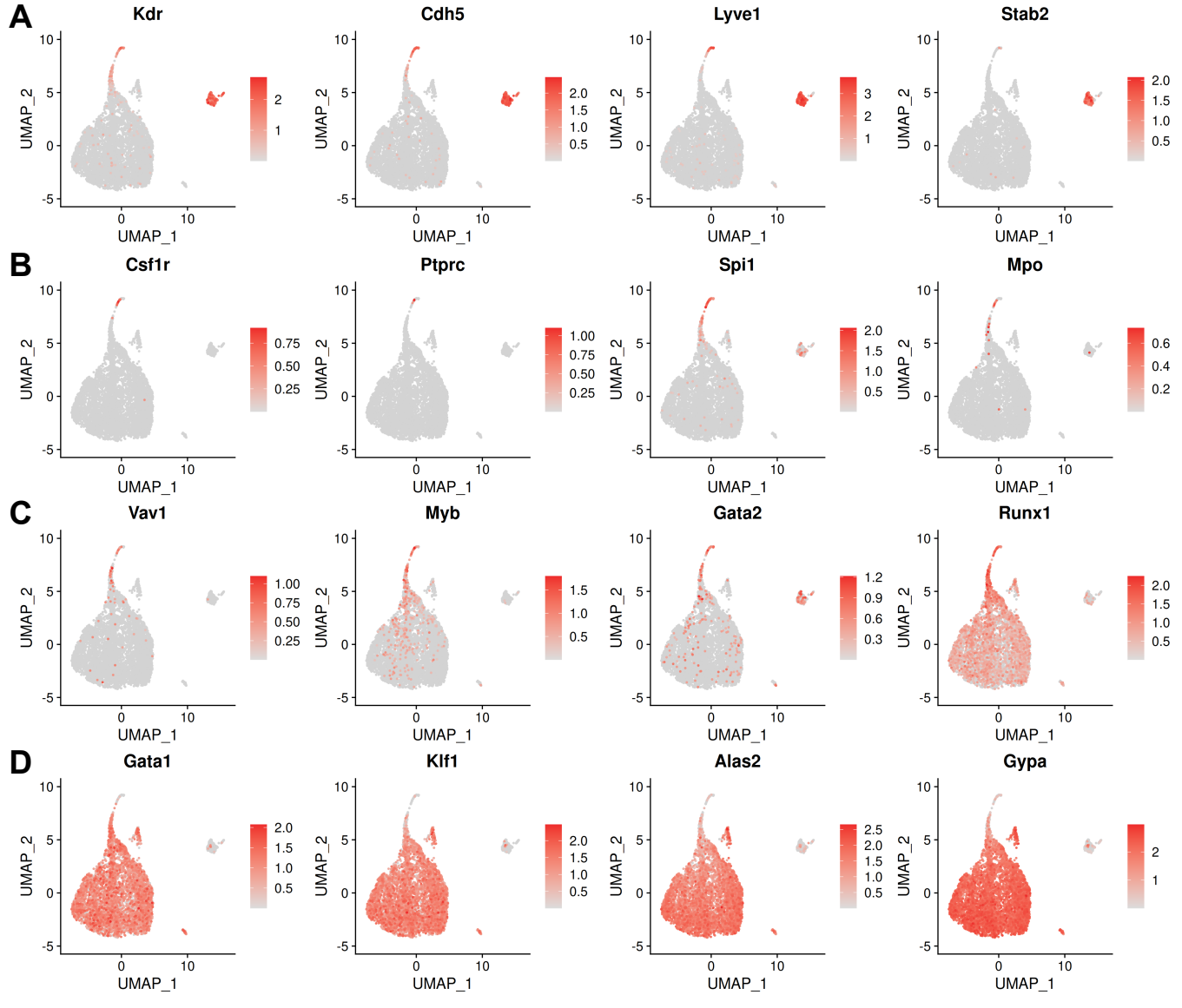

A
Kdr
Cdh5
Lyve1
Stab2
B
Csf1r
Ptprc
Spi1
Mpo
C
Vav1
Myb
Gata2
Runx1
D
Gata1
Klf1
Alas2
Gypa
UMAP_1
UMAP_2

Figure 3

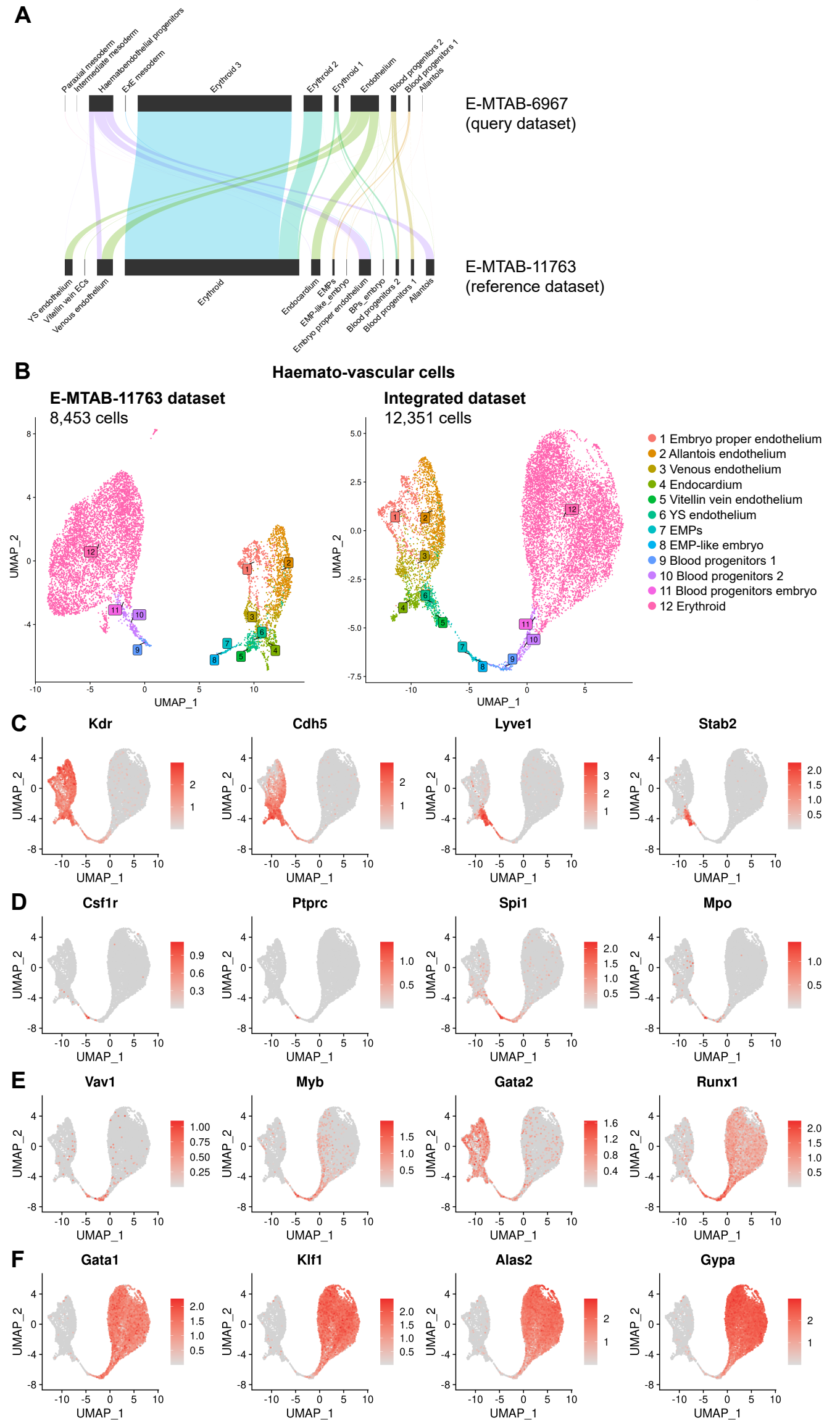
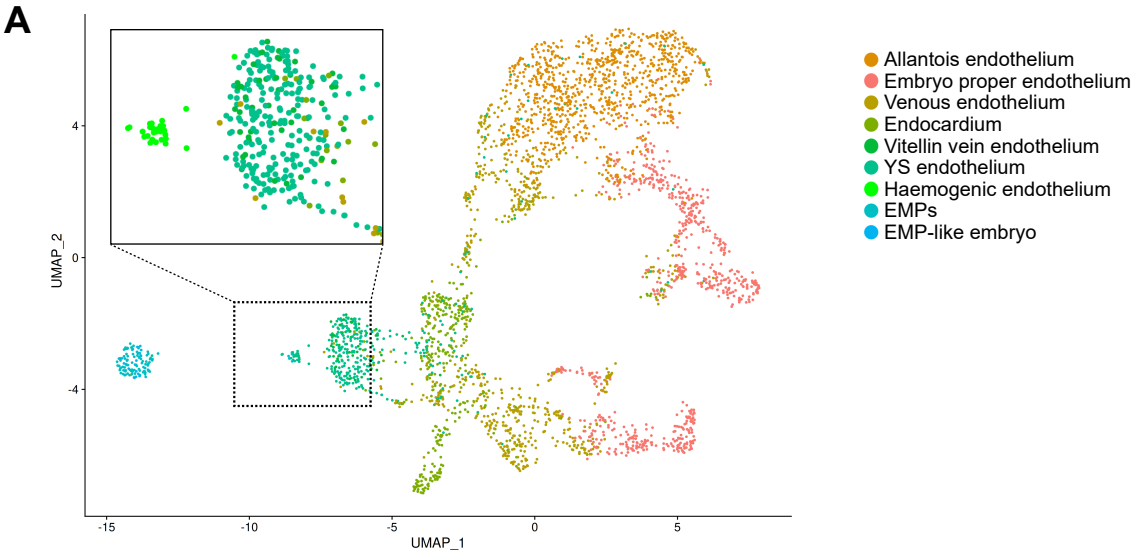

A
Paraxial mesoderm
Intermediate mesoderm
Haematoendothelial progenitors
ExE mesoderm
Erythroid 3
Erythroid 2
Erythroid 1
Endothelium
Blood progenitors 2
Blood progenitors 1
Allantois
E-MTAB-6967
(query dataset)
E-MTAB-11763
(reference dataset)
YS endothelium
Vitellin vein ECs
Venous endothelium
Erythroid
Endocardium
EMPs
EMP-like_embryo
Embryo proper endothelium
BPs_embryo
Blood progenitors 2
Blood progenitors 1
Allantois
B
Haemato-vascular cells
E-MTAB-11763 dataset
8,453 cells
Integrated dataset
12,351 cells
UMAP_1
UMAP_2
1 Embryo proper endothelium
2 Allantois endothelium
3 Venous endothelium
4 Endocardium
5 Vitellin vein endothelium
6 YS endothelium
7 EMPs
8 EMP-like embryo
9 Blood progenitors 1
10 Blood progenitors 2
11 Blood progenitors embryo
12 Erythroid
C
Kdr
Cdh5
Lyve1
Stab2
D
Csf1r
Ptprc
Spi1
Mpo
E
Vav1
Myb
Gata2
Runx1
F
Gata1
Klf1
Alas2
Gypa

Figure 4


A
Allantois endothelium
Embryo proper endothelium
Venous endothelium
Endocardium
Vitellin vein endothelium
YS endothelium
Haemogenic endothelium
EMPs
EMP-like embryo
UMAP_1
UMAP_2

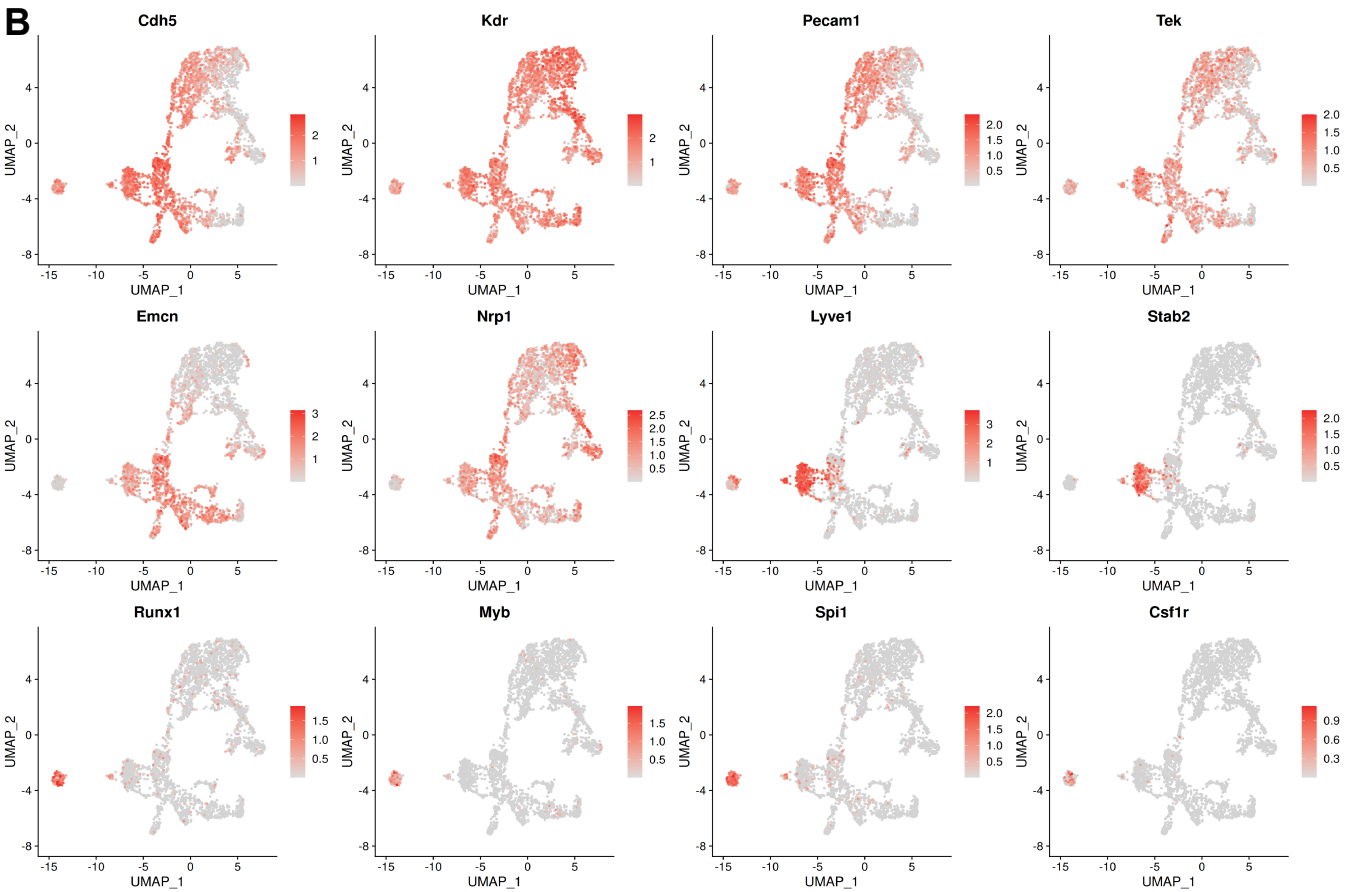

B
Cdh5
Kdr
Pecam1
Tek
Emcn
Nrp1
Lyve1
Stab2
Runx1
Myb
Spi1
Csf1r
UMAP_1
UMAP_2

Figure 5

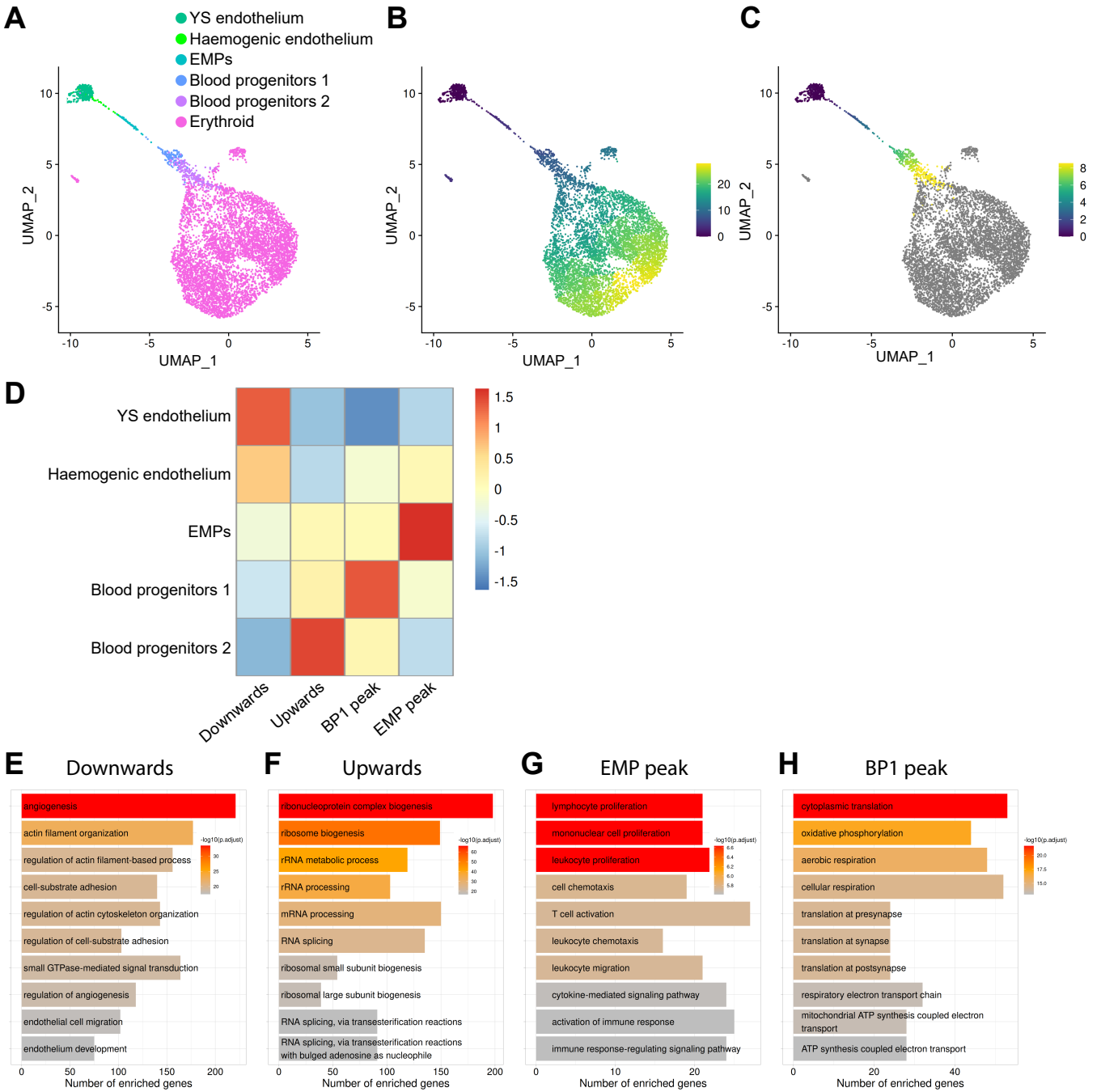
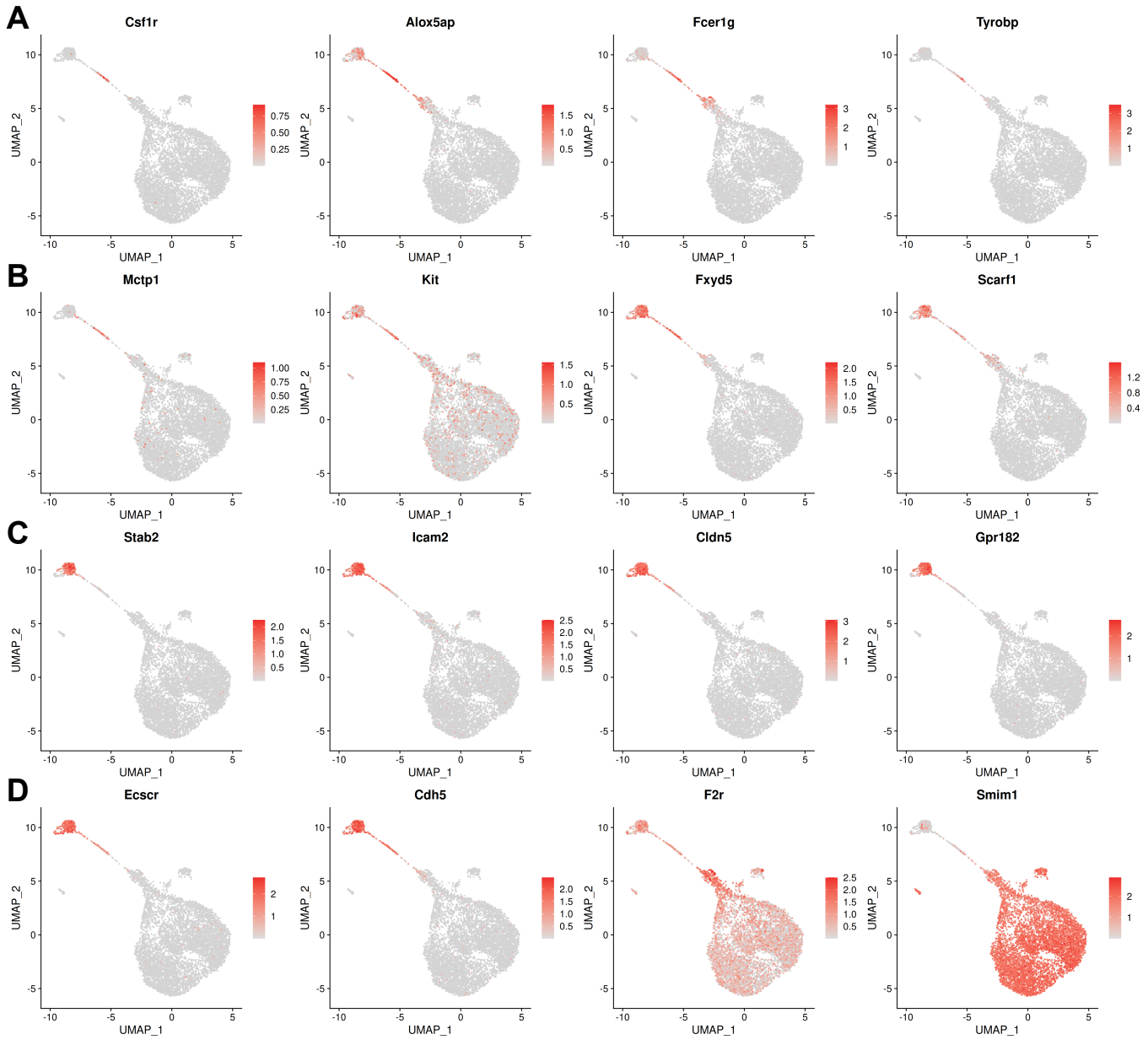

A
YS endothelium
Haemogenic endothelium
EMPs
Blood progenitors 1
Blood progenitors 2
Erythroid
UMAP_1
UMAP_2
B
C
D
YS endothelium
Haemogenic endothelium
EMPs
Blood progenitors 1
Blood progenitors 2
Downwards
Upwards
BP1 peak
EMP peak
E Downwards
angiogenesis
actin filament organization
regulation of actin filament-based process
cell-substrate adhesion
regulation of actin cytoskeleton organization
regulation of cell-substrate adhesion
small GTPase-mediated signal transduction
regulation of angiogenesis
endothelial cell migration
endothelium development
-log10(p.adjust)
Number of enriched genes
F Upwards
ribonucleoprotein complex biogenesis
ribosome biogenesis
rRNA metabolic process
rRNA processing
mRNA processing
RNA splicing
ribosomal small subunit biogenesis
ribosomal large subunit biogenesis
RNA splicing, via transesterification reactions
RNA splicing, via transesterification reactions with bulged adenosine as nucleophile
-log10(p.adjust)
Number of enriched genes
G EMP peak
lymphocyte proliferation
mononuclear cell proliferation
leukocyte proliferation
cell chemotaxis
T cell activation
leukocyte chemotaxis
leukocyte migration
cytokine-mediated signaling pathway
activation of immune response
immune response-regulating signaling pathway
-log10(p.adjust)
Number of enriched genes
H BP1 peak
cytoplasmic translation
oxidative phosphorylation
aerobic respiration
cellular respiration
translation at presynapse
translation at synapse
translation at postsynapse
respiratory electron transport chain
mitochondrial ATP synthesis coupled electron transport
ATP synthesis coupled electron transport
-log10(p.adjust)
Number of enriched genes

Figure 6


A
Csf1r
Alox5ap
Fcer1g
Tyrobp
B
Mctp1
Kit
Fxyd5
Scarf1
C
Stab2
Icam2
Cldn5
Gpr182
D
Ecscr
Cdh5
F2r
Smim1
UMAP_1
UMAP_2

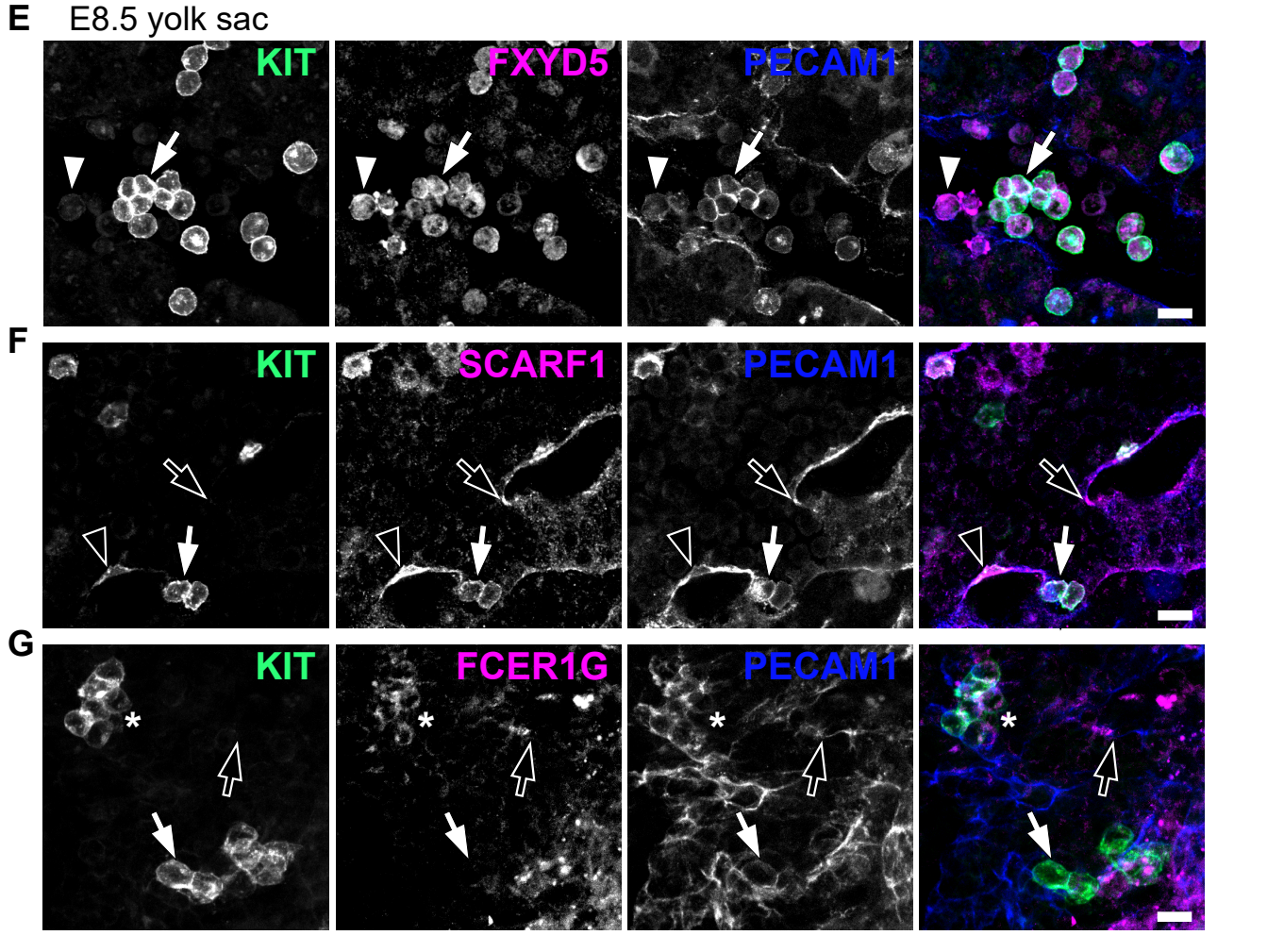
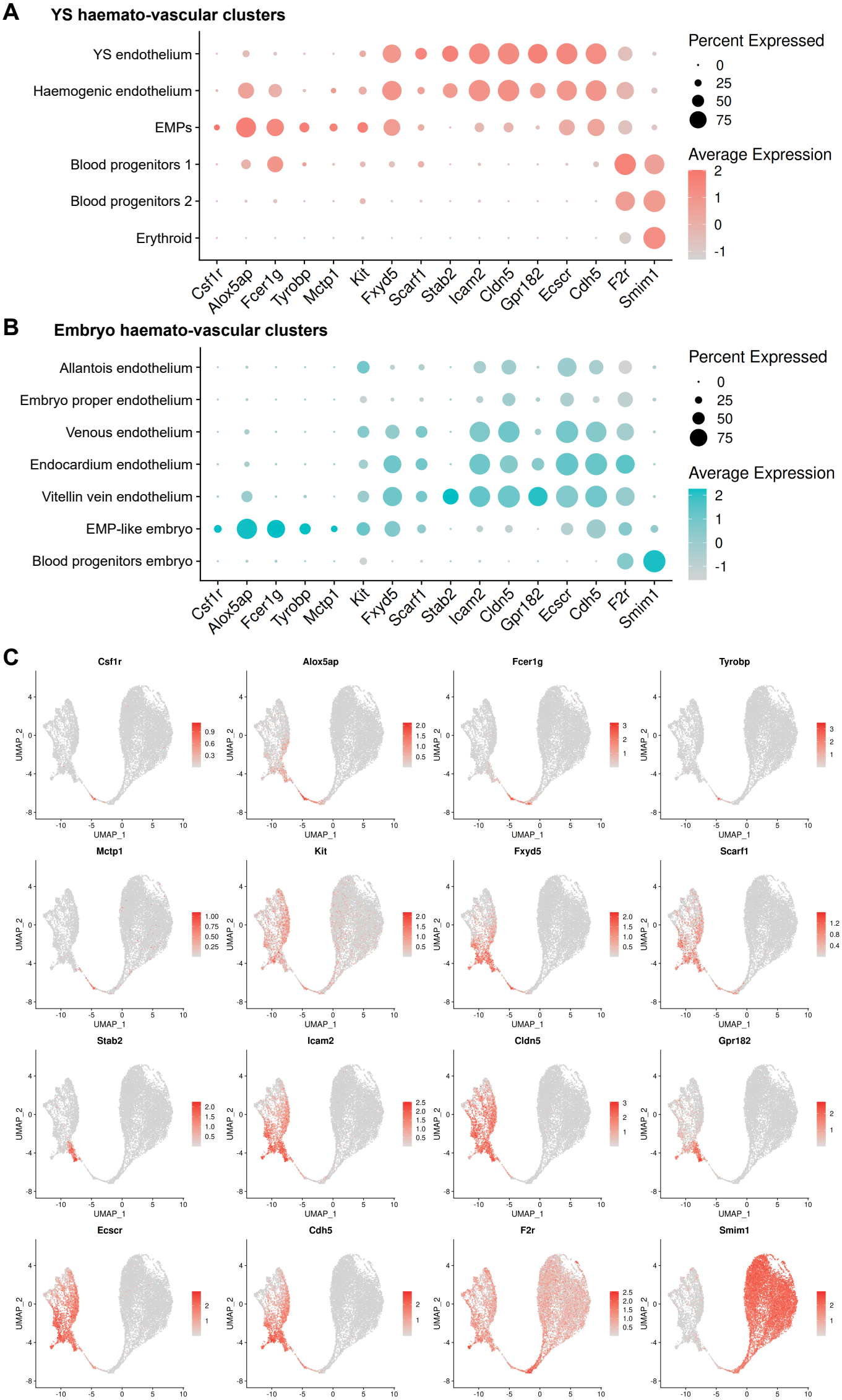

E
E8.5 yolk sac
KIT
FXYD5
PECAM1
F
KIT
SCARF1
PECAM1
G
KIT
FCER1G
PECAM1

Figure 7


A
YS haemato-vascular clusters
YS endothelium
Haemogenic endothelium
EMPs
Blood progenitors 1
Blood progenitors 2
Erythroid
Csf1r
Alox5ap
Fcer1g
Tyrobp
Mctp1
Kit
Fxyd5
Scarf1
Stab2
Icam2
Cldn5
Gpr182
Ecscr
Cdh5
F2r
Smim1
Percent Expressed
0
25
50
75
Average Expression
2
1
0
-1
B
Embryo haemato-vascular clusters
Allantois endothelium
Embryo proper endothelium
Venous endothelium
Endocardium endothelium
Vitellin vein endothelium
EMP-like embryo
Blood progenitors embryo
Percent Expressed
0
25
50
75
Average Expression
2
1
0
-1
C
Csf1r
Alox5ap
Fcer1g
Tyrobp
Mctp1
Kit
Fxyd5
Scarf1
Stab2
Icam2
Cldn5
Gpr182
Ecscr
Cdh5
F2r
Smim1
UMAP_1
UMAP_2