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

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 and EMPs. Further, we have identified candidate signalling and metabolic pathways that may regulate yolk sac haematopoietic emergence and differentiation. Novel markers for haemogenic endothelium and EMPs were 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 (Proteintech #12166-1-AP, 1:50), rabbit anti-mouse SCARF1 (Proteintech
164 #13702-1-AP, 1:100) or rabbit anti-mouse FCER1G (GeneTex #GTX108487s, 1:50),
165 followed by the appropriate secondary antibodies, which were Alexa Fluor 488-conjugated
166 donkey anti-goat Fab fragments (Stratech #705-547-003, 1:200), Cy3-conjugated donkey
167 anti-rat Fab fragments (Stratech #712-166-150, 1:200), and Alexa Fluor 647-conjugated
168 donkey anti-rabbit Fab fragments (Stratech #711-607-003, 1:200). DAPI-counterstained
169 sections were imaged on a LSM710 confocal microscope (Zeiss).
170
171
172
173

174 **RESULTS**

175

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

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

193 First, we analysed E8.5 data from E-MTAB-11763, which contained yolk sac and
194 embryo proper data annotated separately, and generated a UMAP including 75 clusters,
195 using the cluster annotations provided by the authors of the original publication (**Fig. S1**).
196 Next, we selectively sub-clustered the 8,453 haemato-vascular cells that had been
197 sequenced from the yolk sac (**Fig. 1B,C**) and analysed them for known markers of
198 endothelial and haematopoietic progenitor populations (**Fig. 2A-D**). The yolk sac endothelial
199 cell (EC) cluster was identified via expression of the pan-EC markers *Kdr* and *Cdh5* together
200 with *Lyve1* and *Stab2* (**Fig. 2A**) as two markers typical for yolk sac ECs (Gordon, Gale and
201 Harvey, 2008; Shvartsman *et al.*, 2019). The EMP cluster was identified via detection of
202 *Csf1r* and *Ptprc* as well as the myeloid markers *Spi1 and Mpo* (**Fig. 2B**). Notably, *Kdr*, *Cdh5*
203 and *Lyve1* were also detected in the EMP cluster (**Fig. 2A**), agreeing with their known
204 endothelial origin (Gomez Perdiguero *et al.*, 2014). Clusters expressing haematopoietic
205 progenitor markers (*Vav1*, *Myb*, *Gata2*) (**Fig. 2C**) together with haematopoietic
206 differentiation markers (*Gata1*, *Klf1*) (**Fig. 2D**) represented intermediate progenitors at
207 different states of commitment and were annotated as blood progenitors (BP) 1 and BP2

208 (**Fig. 1C**). Finally, high expression levels of differentiation markers (*Alas2*, *Gypa*) (**Fig. 2D**),
209 in the absence of endothelial (*Cdh5*, *Lyve1*), EMP (*Csf1r*, *Ptprc*) or progenitor markers
210 (*Spi1*, *Gata2*), identified a cluster of differentiated erythroid cells.

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

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

232 Next, we integrated E-MTAB-11763 with E-MTAB-6967 to increase the number of
233 cells available for downstream analysis in the haemato-vascular clusters to identify yolk sac
234 haemogenic endothelium. As cells derived from the yolk sac and embryo proper had not
235 been separately sequenced to generate E-MTAB-6967, we used a validated, reference-
236 based annotation method (Aran *et al.*, 2019) to harmonise cell type annotations and
237 distinguish cells from these sources before integrating the two datasets (**Fig. 3A**).
238 Specifically, this method uses gene expression information to assign the most likely cluster
239 from a reference dataset to a query dataset, which in our case involved applying our revised
240 annotation for E8.5 data in E-MTAB-11763 to the E8.5 data in E-MTAB-6967. This approach
241 increased the overall number of haemato-vascular cells from 8,453 to 12,351 for further

242 analysis (**Fig. 3B**). We also surveyed the expression of key known marker genes of the cell
243 populations of interest to confirm the appropriate annotation of the integrated dataset (**Fig.**
244 **3C-F**).
245

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

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

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

262 Having identified a cell cluster with haemogenic endothelial identity, we next used our
263 refined annotation of the yolk sac populations to study signalling pathways and markers
264 defining the transition from haemogenic ECs into EMPs and BPs. When analysing samples
265 that contain cells along a continuous differentiation process, single-cell transcriptomic data
266 can be used to arrange cells along a pseudotime trajectory (Cannoodt, Saelens and Saeys,
267 2016). In essence, pseudotime analysis uses gene expression patterns to reorder cells and
268 arrange them based on their presumed progressive gene expression differences, thus
269 reconstructing trajectories along which cells may be differentiating. Thus, we used
270 pseudotime analysis of yolk sac cell populations to study the transition of ECs into
271 haematopoietic populations. In agreement with prior knowledge of EHT, pseudotime
272 analysis of our integrated E8.5 dataset suggested a continuum of cell differentiation from
273 endothelial to haemogenic endothelium and then to haematopoietic cells (**Fig. 5A**), and also
274 highlighted progressive erythroid maturation (**Fig. 5B**). A refined pseudotime analysis after
275 exclusion of erythroid cells as the predominant cell type corroborated the expected
276 progressive transition of yolk sac ECs to haemogenic ECs, EMPs, BP1 and BP2 (**Fig. 5C**).

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

286 To identify potential biological processes underpinning EHT, we used gene ontology
287 (GO) analysis on these 4 groups of co-regulated genes (**Fig. 5E-H**, **Table 2**). The GO
288 analysis of genes with a downward expression trend during the predicted progression from
289 endothelial to haematopoietic cells revealed enrichment in biological processes related to
290 angiogenesis, cytoskeleton re-organisation, cell adhesion, endothelial cell migration and
291 endothelial development (**Fig. 5E**). Genes with an upward expression trend instead were
292 involved in mRNA processing, splicing and ribonucleoprotein biogenesis (**Fig. 5F**). Genes
293 that peaked in EMPs showed enrichment in terms involved in leukocyte migration, myeloid
294 proliferation and immune response (**Fig. 5G**), in agreement with EMPs being responsible
295 for the production of foetal blood and immune cells. Finally, genes that peaked in BP1 are
296 involved in oxidative phosphorylation, cellular respiration and ATP synthesis (**Fig. 5H**),
297 raising the possibility that these cells have increased metabolism.

298 To identify genes belonging to known signalling pathways amongst each one of the
299 4 groups of co-regulated genes, we performed pathway analysis using the WikiPathways
300 resource (Agrawal *et al.*, 2024) (**Table 3**). Genes that were downregulated when endothelial
301 cells were predicted to transition to haematopoietic cells belonged to the cholesterol
302 biosynthesis, IL3, IL17A, PI3K/AKT/mTOR and EFGR1 signalling pathways. Genes that
303 were upregulated during this transition are involved in RNA processing and cell cycle
304 activation. Genes peaking in BP1 are involved in the regulation of cell metabolism, whereas
305 genes peaking in EMPs belonged to the TYROBP and IL5 signalling pathways (**Table 3**).
306 Notably, by comparing EMPs to other yolk sac clusters, we identified multiple genes within
307 the TYROBP signalling pathway that, in addition to peaking in EMPs, were significantly
308 enriched in EMPs compared to the other cells in the haemato-vascular cluster (**Fig. S2**).

309 Overall, this analysis provided insights into molecular changes predicted to
310 accompany EHT and haematopoietic differentiation in the yolk sac, and identified candidate

311 signalling pathways that may be studied further to understand the molecular control of these
312 complex cellular differentiation pathways.
313

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

315 To identify markers for haemogenic ECs and EMPs, we used our integrated E8.5
316 mouse dataset to perform differential expression (DE) analysis of endothelial and
317 haematopoietic cell populations in both the yolk sac and embryo. First, we subset the
318 clusters for both cell populations and then performed DE analysis by comparing clusters in
319 different combinations to identify the top differentially regulated genes. To identify genes
320 that may encode markers useful for cell identification via immunostaining or flow cytometric
321 cell sorting, we filtered DE genes for those that encoded transmembrane proteins (**Table 4**).

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

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

340 To confirm that *Fxyd5*, *Scarf1* and *Fcer1g* encode EMP markers, we performed
341 immunofluorescence staining of E8.5 yolk sac. Although we had found *Fxyd5* transcripts in
342 yolk sac ECs and EMPs, FXYD5 staining was restricted to round KIT^{+} PECAM1^{+} cells
343 consistent with newly emerging EMPs, and there was no obvious staining of ECs (**Fig. 6E**).
344 Additionally, FXYD5 immunostaining identified round KIT^{-} PECAM1^{low} cells, agreeing with
345 our transcriptomic analysis detecting BPs with no *Kit* transcripts and low levels of *Fxyd5*

346 (**Fig. 6B**). Consistent with the presence of *Scarf1* transcripts in both yolk sac ECs and EMPs,
347 immunofluorescence staining identified SCARF1 protein in flat KIT$^-$ PECAM1$^+$ ECs and
348 round KIT$^+$ PECAM1$^+$ emerging EMPs. Furthermore, SCARF1 immunostaining identified flat
349 KIT$^+$ PECAM1$^+$ cells that may be haemogenic ECs (**Fig. 6F**). Although *Fcer1g* transcripts
350 were detected in both EMPs and BP1, immunofluorescence identified FCER1G protein
351 within a subset of round KIT$^+$ cells. Additionally, FCER1G immunostaining identified flat KIT$^-$
352 PECAM1$^+$ ECs, consistent with our transcriptomic analysis detecting low levels of *Fcer1g*
353 transcripts in yolk sac ECs (**Fig. 6G**).

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

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

363

364 **DISCUSSION**

365 Our re-analysis of transcriptomic data from E8.5 mouse embryos provided a refined
366 annotation of endothelial and haematopoietic populations in the mouse yolk sac (**Fig. 3A,B**)
367 and identified a subset of endothelial cells co-expressing endothelial and haematopoietic
368 markers, consistent with a haemogenic endothelial phenotype (**Fig. 4A,B**). This refined
369 annotation also defined E8.5 yolk sac EMPs and demonstrated co-expression of myeloid,
370 erythroid and progenitor markers in this cluster, consistent with the known dual erythro-
371 myeloid differentiation potential of EMPs. Moreover, we identified EMP-like cells within the
372 E8.5 embryo proper, which have not been highlighted in prior transcriptomic studies of early
373 haematopoiesis (**Fig. 3B**). Knowing that haematopoietic cells enter the embryo only from
374 E9.5, when the heart begins to beat to initiate blood circulation (Frame *et al.*, 2016; Kasaai
375 *et al.*, 2017), we considered two possibilities to explain the existence of EMP-like cells in the
376 embryo. On the one hand, these cells may represent yolk sac-derived haematopoietic
377 progenitors that had contaminated the embryonic samples during dissection. On the other
378 hand, it is conceivable that these cells represent EMP-like cells originating in the embryo
379 rather than in the yolk sac. The latter possibility is consistent with our recent discovery of a

380 *Pax3* lineage-derived EMP-like progenitor in the mouse embryo (Canu *et al.*, 2024).
381 However, further work is required to accurately describe these EMP-like cells.

382 Others have used scRNAseq to map potential differentiation trajectories for EMPs in
383 the yolk sac (Zhao *et al.*, 2023). Additionally, our analysis identified distinct BP1 and BP2
384 clusters as consecutive stages of EMP maturation (**Fig. 5A-C**) and provided new insight into
385 candidate signalling pathways controlling this transition (**Fig. 5D-H**). As primitive
386 erythrocytes are produced from E7.0 onwards (Palis *et al.*, 1999; Ferkowicz and Yoder,
387 2005), but EMPs only emerge from E8.25 onwards (Frame, McGrath and Palis, 2013;
388 Gomez Perdiguero *et al.*, 2014; Hoeffel *et al.*, 2015; McGrath *et al.*, 2015), the erythroid
389 cluster in the E8.5 dataset is expected to contain mostly primitive erythrocytes that were
390 produced prior to or alongside EMP emergence, with only a potential minor contribution from
391 EMP-derived erythrocytes. No markers of fully differentiated macrophages or other myeloid
392 cell types were observed, confirming previous studies showing that macrophages form in
393 the yolk sac from E9.0 onwards (Hoeffel and Ginhoux, 2018). We also did not detect a
394 lymphoid cluster at E8.5, in agreement with prior reports of LMP presence in the yolk sac
395 only from E9.5 onwards (Böiers *et al.*, 2013).

396 Pseudotime reconstruction highlighted dynamic gene expression trends during EHT
397 and haematopoietic maturation in the yolk sac (**Fig. 5A-D**), whereby haemogenic endothelial
398 cells form a transcriptional continuum bridging the transition from endothelial to
399 haematopoietic cell identities, thus implying a gradual and continuous process rather than a
400 discrete lineage switch. Altogether, these findings align with established models of yolk sac
401 haematopoiesis derived from genetic lineage tracing, flow cytometry experiments and prior
402 transcriptomic analysis (Frame, McGrath and Palis, 2013; Hoeffel and Ginhoux, 2018; Canu
403 and Ruhrberg, 2021; Sánchez-Lanzas, Jiménez-Pompa and Ganuza, 2024; Sugimura *et*
404 *al.*, 2025).

405 The progressive downregulation of genes associated with angiogenesis, cytoskeletal
406 organisation and endothelial adhesion during the predicted EHT (**Fig. 5E**) may indicate
407 structural and functional changes necessary for cell detachment and transition to a
408 haematopoietic phenotype. Conversely, the upregulation of genes associated with RNA
409 processing (**Fig. 5F**) suggests that haematopoietic specification is coupled to enhanced
410 transcriptional activity. Furthermore, upregulation of genes associated to cell cycle activation
411 and transition from the G1 to S phase (**Table 3**) agrees with previous findings that EHT
412 requires cell cycle progression and is controlled by the cell cycle regulators CDK4/6 and
413 CDK1 (Marcelo *et al.*, 2013; Canu *et al.*, 2020).

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

423 DE analysis also identified upregulation of genes associated with oxidative
424 phosphorylation upon differentiation of EMPs to BP1 (**Fig. 5H**). Previous studies of
425 pluripotent stem cells and neuronal progenitors showed that their differentiation is
426 associated with metabolic reprogramming, whereby differentiating progenitors switch from
427 glycolysis to oxidative phosphorylation (Zheng *et al.*, 2016; Dahan *et al.*, 2019; Mahmoud,
428 2023). Our analysis raises the possibility that an analogous process may occur when
429 haematopoietic progenitors in the yolk sac commit to differentiation.

430 Our DE analysis identified both known and previously undescribed candidate
431 markers associated with the formation of haemogenic endothelium and EMPs (**Fig. 6A-D**).
432 Established EMP markers such as *Csf1r* and *Alox5ap* were previously described by others
433 (Hoeffel *et al.*, 2015; Sharon *et al.*, 2019), thus validating the robustness of our
434 transcriptomic analysis. In addition, we identified *Fcer1g*, *Tyrobp*, and *Mctp1* as potential
435 novel EMP markers, and *Fxyd5* and *Scarf1* as shared markers between yolk sac endothelial
436 and haematopoietic cells. The expression of *Fxyd5* and *Scarf1* in both endothelial and EMP
437 clusters, but not in more differentiated progenitors, suggests that these genes may serve as
438 new markers for yolk sac ECs and emerging EMPs. Immunofluorescence staining of E8.5
439 yolk sac corroborated the predictions of our DE analysis: FXYD5 was present on round KIT$^+$
440 PECAM1$^+$ cells consistent with emerging EMPs (**Fig. 6E**); SCARF1 labelled flat KIT$^-$
441 PECAM1$^+$ ECs, flat KIT$^+$ PECAM1$^+$ cells consistent with haemogenic ECs and round KIT$^+$
442 PECAM1$^+$ cells consistent with emerging EMPs (**Fig. 6F**); FCER1G marked flat KIT$^-$
443 PECAM1$^+$ ECs and a few round KIT$^+$ PECAM1$^+$ cells that may represent a subset of EMPs
444 (**Fig. 6G**). Nevertheless, these experiments also highlighted differences in RNA and protein
445 expression. Although *Fxyd5* RNA was detected in yolk sac ECs, haemogenic ECs and
446 EMPs, the FXYD5 protein appeared to be predominantly expressed in round KIT$^+$ PECAM1$^+$
447 cells consistent with EMPs (**Fig. 6E**). While *Fcer1g* RNA clearly identified the EMP cluster,
448 the FCER1G protein was detected in only a subset of round KIT$^+$ PECAM1$^+$ cells, possibly

449 marking more mature EMPs (**Fig. 6G**). This is reminiscent of observations for *Csf1r*, a gene
450 expressed in EMPs at the RNA level, whereas CSF1R protein only becomes translated upon
451 differentiation to myeloid cells (Hoeffel *et al.*, 2015). Overall, these immunostaining results
452 validated our transcriptomic predictions and highlight the utility of single-cell datasets in
453 discovering functional cell surface markers.

454 In summary, this study refines the cellular and molecular characterisation of yolk sac
455 haematopoiesis, provides a detailed transcriptomic roadmap of EHT in the yolk sac and
456 identifies novel candidate markers for haemogenic endothelium and EMPs. The
457 identification of FXYD5, SCARF1 and FCER1G as markers for EMPs and haemogenic ECs
458 may offers valuable tools for future *in vivo* and *in vitro* studies. Finally, future studies may
459 wish to determine whether the novel EMP markers are present also in foetal HSCs, and
460 whether they are conserved in human haemopoiesis. Defining the specificity of such novel
461 markers may allow researchers to better define EMP identity and lineage potential as a
462 prerequisite for deciphering the ontogeny of the immune system and, ultimately, to harness
463 early haematopoietic pathways for regenerative medicine.

464

465 **ACKNOWLEDGEMENTS**

466 We thank the staff of the Biological Resources, FACS and Imaging Facilities at the UCL
467 Institute of Ophthalmology for their technical support. This research was supported by grants
468 from Wellcome (205099/Z/16/Z) and the British Heart Foundation (PG23/11301).

469

470 **AUTHORSHIP CONTRIBUTIONS**

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

475

476 **DISCLOSURE OF CONFLICTS OF INTEREST**

477 The authors declare that they have no competing interests.

478

479

480 **REFERENCES**

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

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

485 Böiers, C. *et al.* (2013) 'Lymphomyeloid Contribution of an Immune-Restricted Progenitor
486 Emerging Prior to Definitive Hematopoietic Stem Cells', *Cell Stem Cell*, 13(5), pp. 535–
487 548.

488 Cannoodt, R., Saelens, W. and Saeys, Y. (2016) 'Computational methods for trajectory
489 inference from single-cell transcriptomics', *European Journal of Immunology*, 46(11), pp.
490 2496–2506.

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

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

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

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

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

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

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

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

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

511 Frame, J.M., McGrath, K.E. and Palis, J. (2013) ‘Erythro-myeloid progenitors: “definitive”
512 hematopoiesis in the conceptus prior to the emergence of hematopoietic stem cells’, *Blood*
513 *cells, molecules & diseases*, 51(4), pp. 220–225.

514 Goh, I. *et al.* (2023) ‘Yolk sac cell atlas reveals multiorgan functions during human early
515 development’, *Science*, 381(6659).

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

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

521 Gritz, E. and Hirschi, K.K. (2016) ‘Specification and function of hemogenic endothelium
522 during embryogenesis’, *Cellular and molecular life sciences : CMLS*, 73(8), pp. 1547–
523 1567.

524 Gustavsen, J.A. *et al.* (2019) ‘RCy3: Network biology using Cytoscape from within R’,
525 *F1000Research 2019 8:1774*, 8, p. 1774.

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

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

531 Hoeffel, G. and Ginhoux, F. (2018) ‘Fetal monocytes and the origins of tissue-resident
532 macrophages’, *Cellular Immunology*, 330, pp. 5–15.

533 Ivanovs, A. *et al.* (2011) ‘Highly potent human hematopoietic stem cells first emerge in the
534 intraembryonic aorta-gonad-mesonephros region’, *Journal of Experimental Medicine*,
535 208(12), pp. 2417–2427.

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

538 Jiang, Y. *et al.* (2024) ‘Macrophages in organ fibrosis: from pathogenesis to therapeutic
539 targets’, *Cell Death Discovery 2024 10:1*, 10(1), pp. 1–11.

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

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

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

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

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

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

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

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

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

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

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

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

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

568 Sánchez-Lanzas, R., Jiménez-Pompa, A. and Ganuza, M. (2024) 'The evolving
569 hematopoietic niche during development', *Frontiers in Molecular Biosciences*, 11, p.
570 1488199.

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

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

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

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

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

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

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

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

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

589 Zhang, M. *et al.* (2024) ‘TYROBP serve as potential immune-related signature genes in
590 the acute phase of intracerebral hemorrhage’, *Scientific Reports 2024 14:1*, 14(1), pp. 1–
591 12.

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

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

597

598

Figure 1

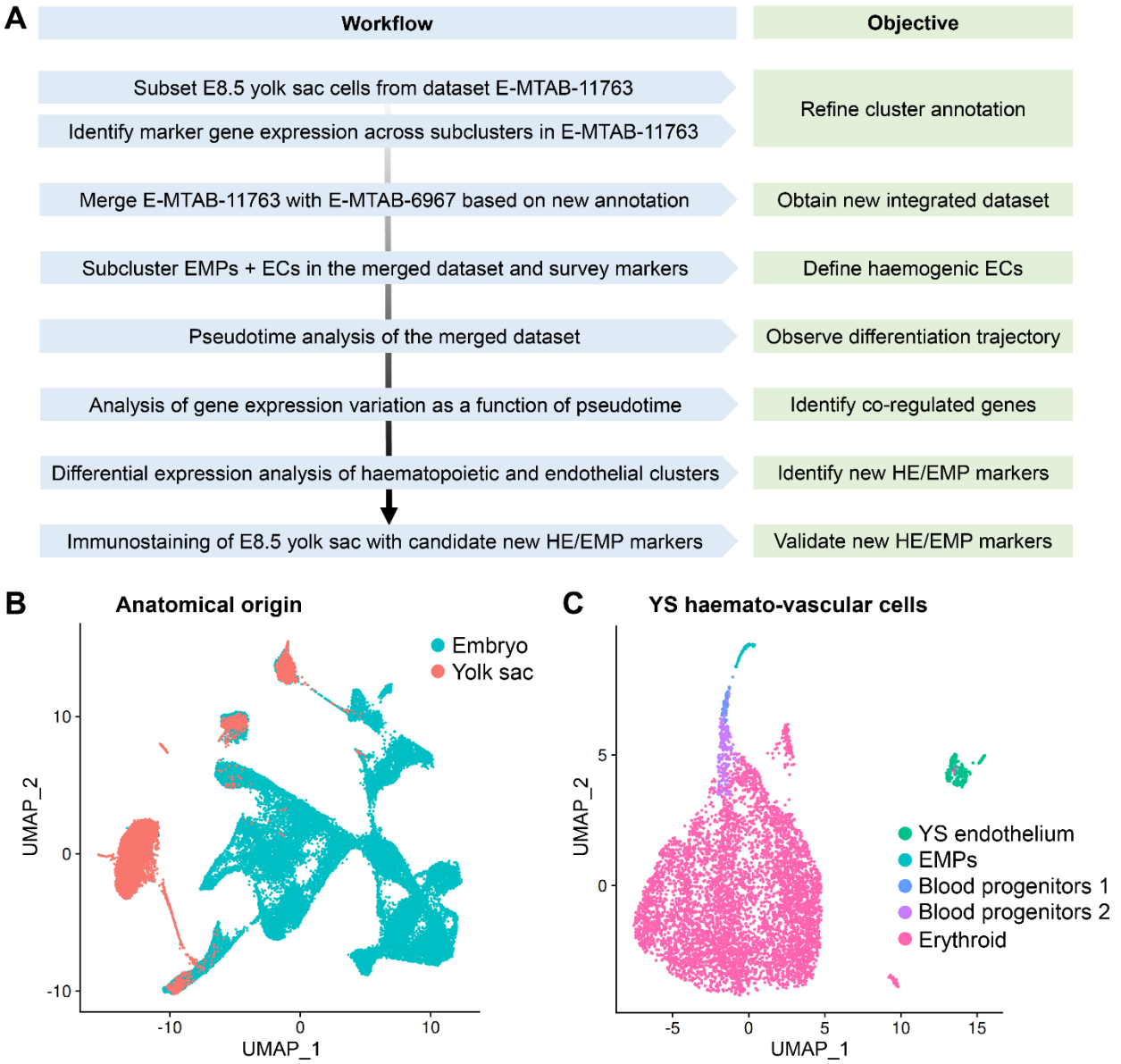

Fig 1. Re-analysis of E8.5 mouse embryonic scRNAseq data to study yolk sac EHT.

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.

Figure 2

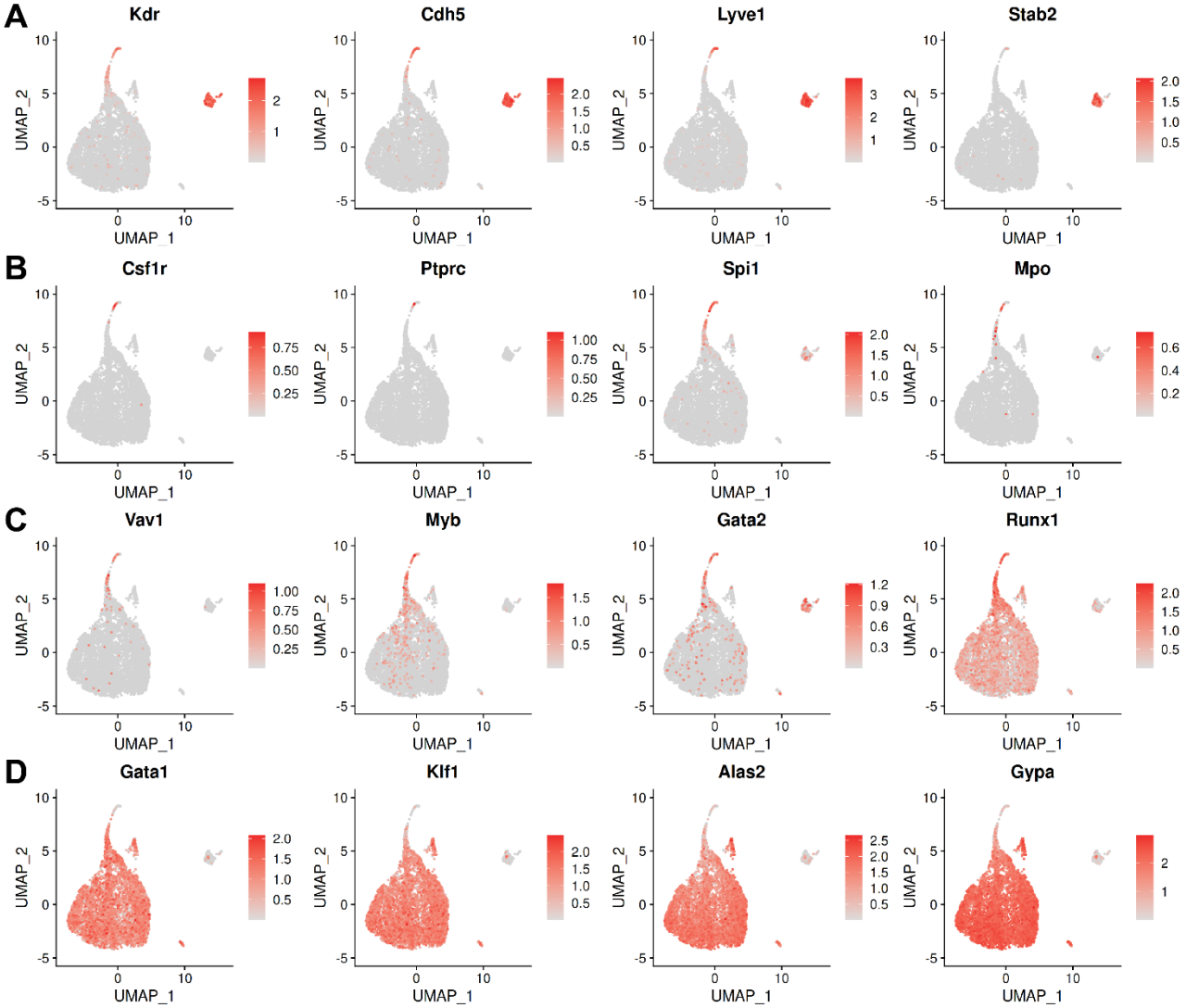

Fig 2. Marker analysis identifies haemato-vascular cell populations in E8.5 yolk sac data in E-MTAB-11763.

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.

Figure 3

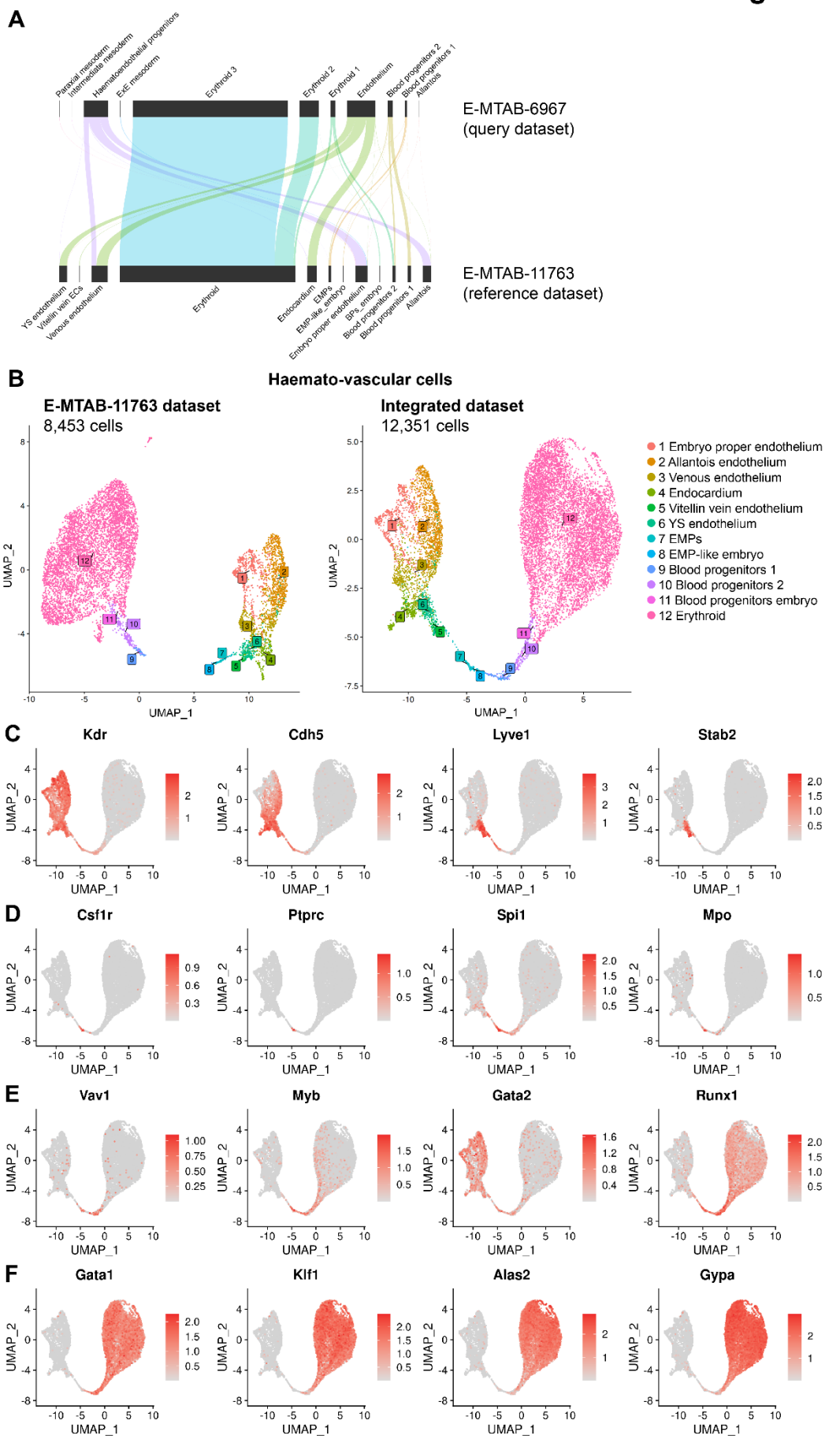
A
Paraxial mesoderm
Intermediate mesoderm
Haematoendothelial progenitors
ExE mesoderm
Erythroid 3
Erythroid 2
Erythroid 1
Endothelium
Blood progenitors 2
Blood progenitors 1
Allantois
E-MTAB-6967
(query dataset)
E-MTAB-11763
(reference dataset)
YS endothelium
Vitellin vein ECs
Venous endothelium
Erythroid
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
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
UMAP_1
UMAP_2
C
Kdr
Cdh5
Lyve1
Stab2
D
Csf1r
Ptprc
Spi1
Mpo
E
Vav1
Myb
Gata2
Runx1
F
Gata1
Klf1
Alas2
Gypa

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.

Figure 4

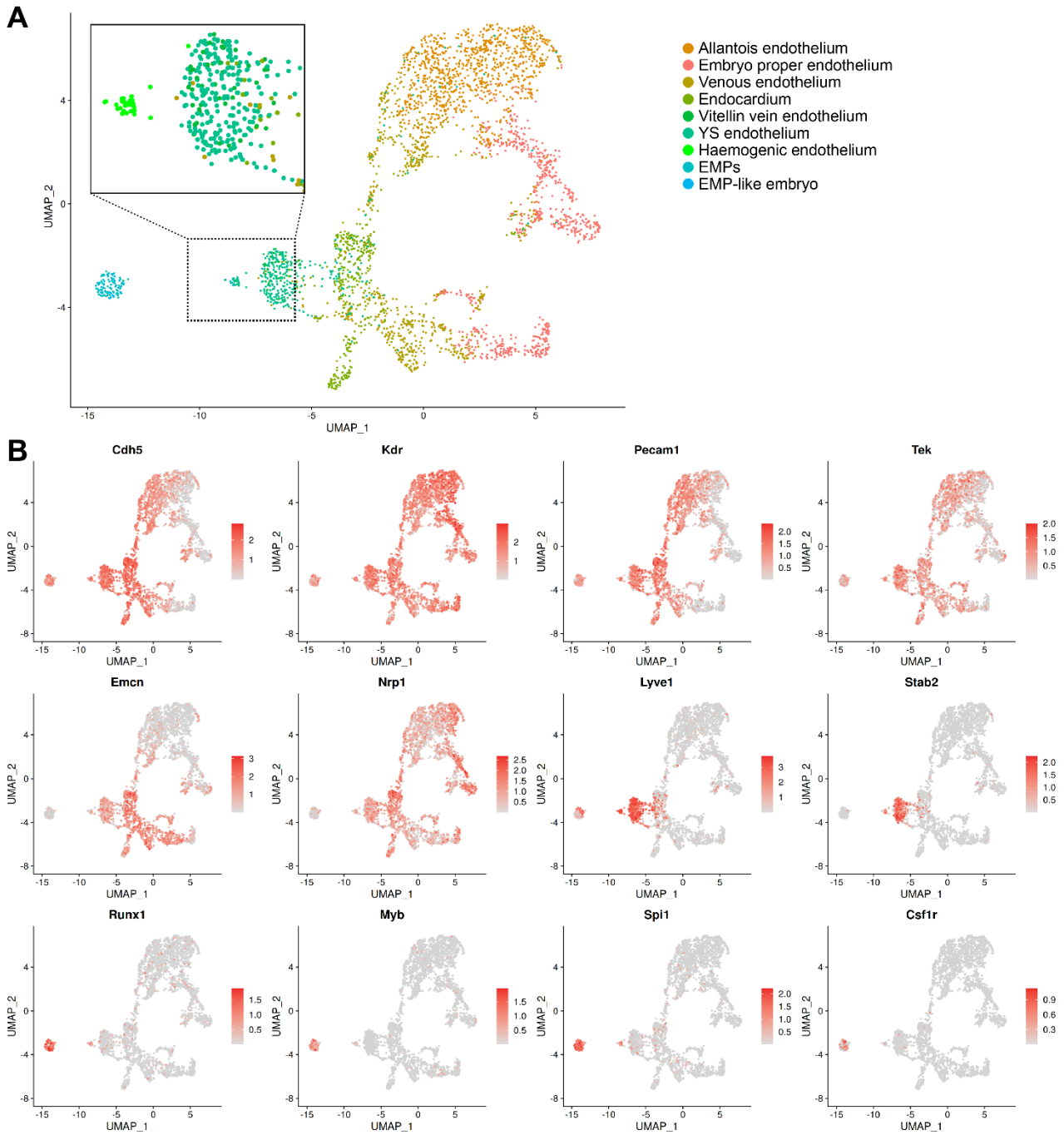

Fig 4. Endothelial cell and EMP sub-clustering identifies haemogenic endothelium.

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.

Figure 5

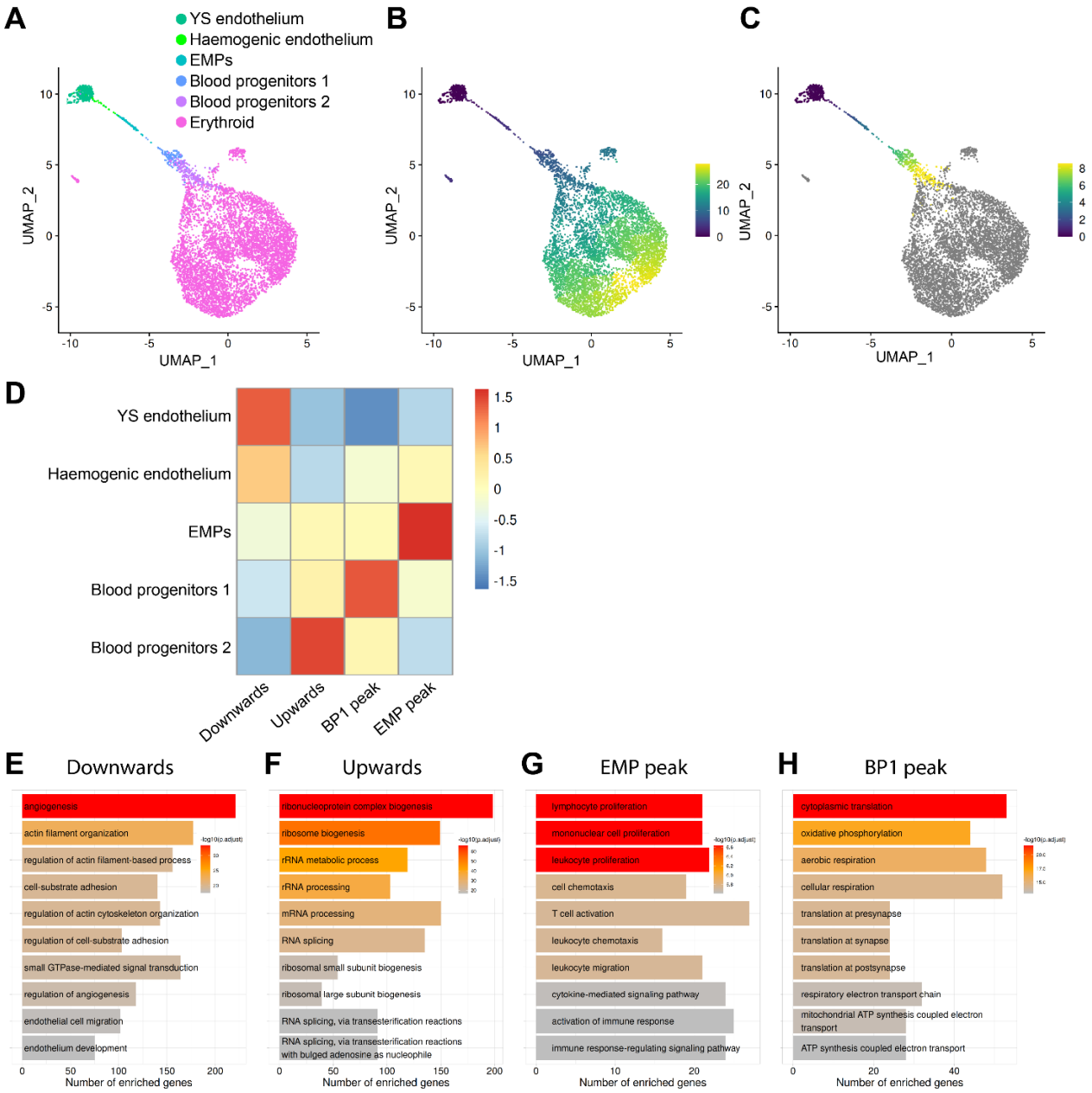

Fig 5. Pseudotime analysis identifies co-regulated genes and enriched GO pathways during EHT.

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.

Figure 6

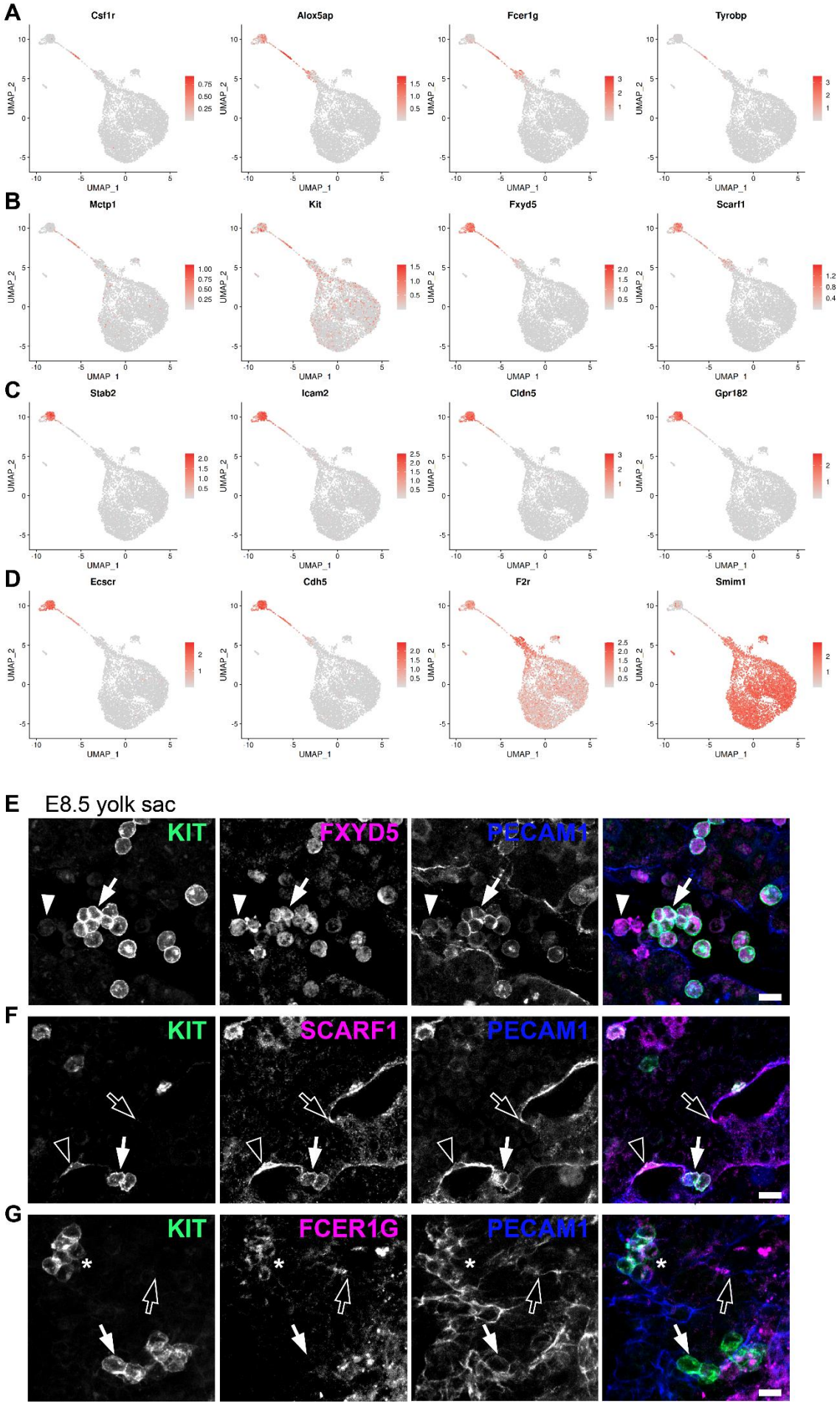
A
Csf1r
Alox5ap
Fcer1g
Tyrobp
B
Mctp1
Kit
Fxyd5
Scarf1
C
Stab2
Icam2
Cldn5
Gpr182
D
Ecscr
Cdh5
F2r
Smim1
UMAP_1
UMAP_2
E
E8.5 yolk sac
KIT
FXYD5
PECAM1
F
KIT
SCARF1
PECAM1
G
KIT
FCER1G
PECAM1

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 with arrows (round KIT+ PECAM1+ cells); presumed blood 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); mature EMPs with asterisk (round KIT+ PECAM1+ FCER1G+ cells). Scale bars: 50 µm.

Figure 7

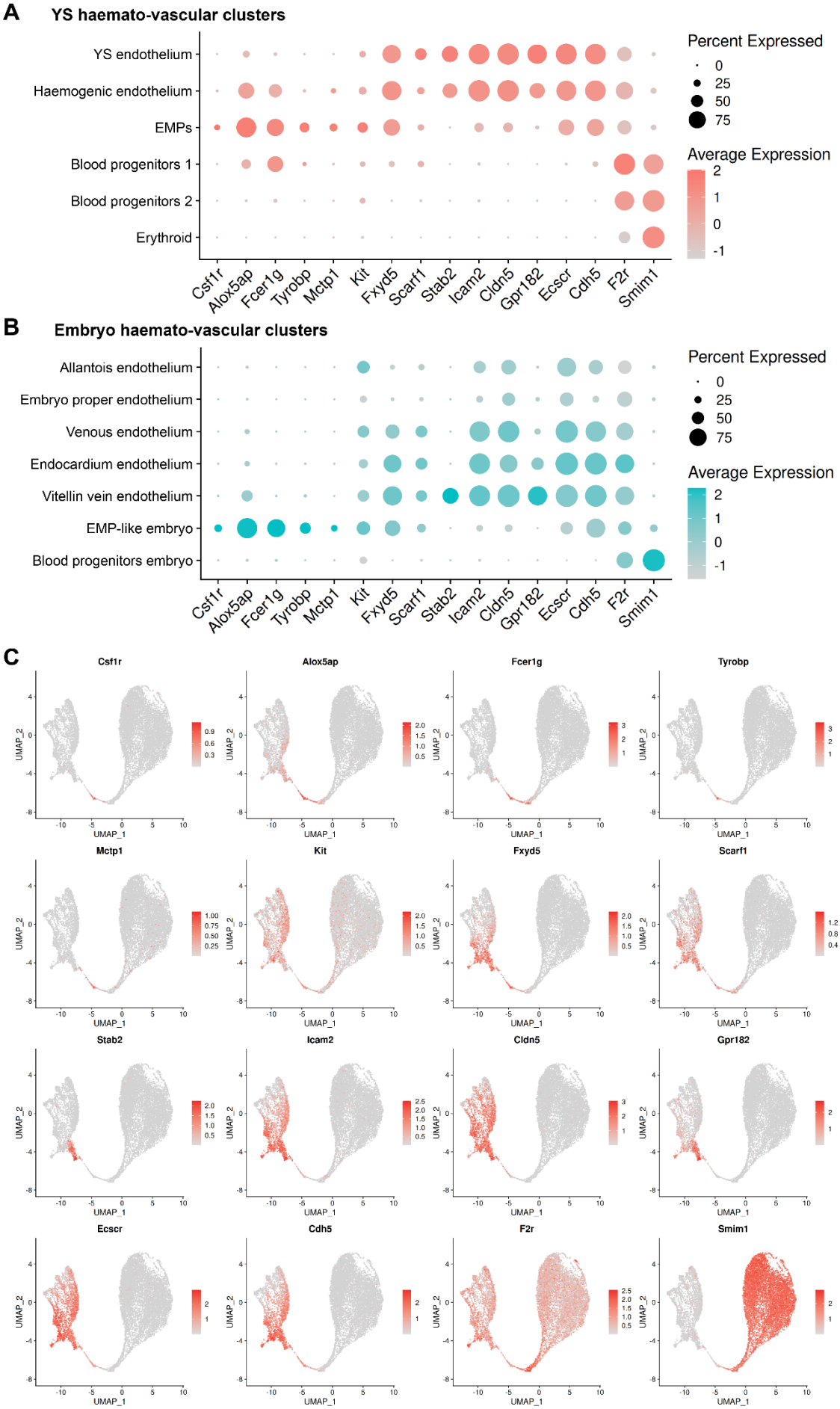
A
YS haemato-vascular clusters
YS endothelium
Haemogenic endothelium
EMPs
Blood progenitors 1
Blood progenitors 2
Erythroid
Csf1r
Alox5ap
Fcer1g
Tyrobp
Mctp1
Kit
Fxyd5
Scarf1
Stab2
Icam2
Cldn5
Gpr182
Ecscr
Cdh5
F2r
Smim1
Percent Expressed
0
25
50
75
Average Expression
2
1
0
-1
B
Embryo haemato-vascular 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

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.