Title: Multi-organ functions of yolk sac during human early development

Authors: Rachel A Botting[1]†, Issac Goh[1,2]† Antony Rose[1,2]‡, Simone Webb[1,2]‡, Justin Engelbert[1], Yorick Gitton[3], Emily Stephenson[1,2], Mariana Quiroga Londoño[4], Michael Mather[1], Nicole Mende[4], Ivan Imaz-Rosshandler[4,5], Dave Horsfall[1], Daniela Basurto-Lozada[1], Nana-Jane Chipampe[2], Victoria Rook[2], Pavel Mazin[2], MS Vijayabaskar[4], Rebecca Hannah[4], Laure Gambardella[2], Kile Green[6], Stephane Ballereau[2], Megumi Inoue[3], Liz Tuck[2], Valentina Lorenzi[2], Kwasi Kwakwa[2], Clara Alsinet[2], Bayanne Olabi[1], Mohi Miah[1], Chloe Admane[1,2], Dorin-Mirel Popescu[1], Meghan Acres[1], David Dixon[1], Rowen Coulthard[7], Steven Lisgo[1], Deborah J Henderson[1], Emma Dann[2], Chenqu Suo[2], Sarah J Kinston[4], Jong-eun Park[8], Krzysztof Polanski[2], Stijn Van Dongen[2], Kerstin B Meyer[2], Marella de Bruijn[9], James Palis[10], Sam Behjati[2,11], Elisa Laurenti[4], Nicola K Wilson[4], Roser Vento-Tormo[2], Alain Chédotal[3], Omer Bayraktar[2], Irene Roberts[12], Laura Jardine[1]*, Berthold Göttgens[4]*, Sarah A Teichmann[2]*, Muzlifah Haniffa[2,1]*

†/‡ Equal contribution

* Corresponding authors. Email: mh32@sanger.ac.uk

Affiliations:

[1]Biosciences Institute, Newcastle University, NE2 4HH, UK

[2]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK

[3]Sorbonne Université, INSERM, CNRS, Institut de la Vision, Paris, France

[4]Department of Haematology, Wellcome-MRC Cambridge Stem Cell Institute, CB2 0AW, UK

[5]MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, CD2 0QH, UK

[6]Translational and Clinical Research Institute, Newcastle University, NE2 4HH, UK

[7]NovoPath, Department of Pathology, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

[8]Korea Advanced Institute of Science and Technology, Daejeon, South Korea

[9]MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, OX3 9DS, UK

[10]Department of Pediatrics, University of Rochester Medical Center, Rochester, 14642, NY, USA

[11]Department of Paediatrics, University of Cambridge, Cambridge, UK

[12]Department of Paediatrics, University of Oxford, OX3 9DS, UK

Abstract: The yolk sac (YS) represents an evolutionarily-conserved extraembryonic structure that ensures timely delivery of nutritional support and oxygen to the developing embryo. However, the YS remains ill-defined in humans. We therefore assemble a complete single cell 3D map of human YS from 3-8 post conception weeks by integrating multiomic protein and gene expression data. We reveal the YS as a site of primitive and definitive haematopoiesis including a YS-specific accelerated route to macrophage production, a source of nutritional/metabolic support and a regulator of oxygen-carrying capacity. We reconstruct the emergence of primitive haematopoietic stem and progenitor cells from YS hemogenic endothelium and their decline upon stromal support modulation as intraembryonic organs specialise to assume these functions. The YS therefore functions as 'three organs in one' revealing a multifaceted relay of vital organismal functions as pregnancy proceeds.

One Sentence Summary: Human yolk sac is a key staging post in a relay of vital organismal functions during human pregnancy.

Main Text:

The primary human YS derives from the hypoblast at around the time of embryo implantation (Carnegie stage 4, CS4; ~1 post conception week (PCW)) (*1*, *2*). A secondary YS beneath the embryonic disc supersedes the primary structure at around CS6 (~2.5PCW) and persists until 8PCW (*1*, *2*). The secondary YS surrounds a vitelline fluid-filled cavity with three tissue compartments: mesothelium (a mesoderm-derived epithelial layer interfacing the amniotic fluid), mesoderm (which contains an array of cell types, including endothelial cells, blood cells and smooth muscle), and endoderm (interfacing the yolk sac cavity) (*1*). In phylogenetic terms, the YS is first seen in vertebrates with yolk-rich eggs e.g., birds, reptiles and amphibians, where its role is to extract macronutrients from yolk to sustain the embryo (*3*). The capacity to uptake, transport and metabolise nutrients is retained in mouse and human YS (*2*). Haematopoiesis originates in the YS in mammals, birds and some ray-finned fishes (*4*). The first wave of mouse YS haematopoiesis (primitive) yields primitive erythroid cells, macrophages and megakaryocytes from embryonic day 7.5 (E7.5) (*4*, *5*). Following the onset of circulation, a second wave of erytho-myeloid and lympho-myeloid progenitors arise in the YS and supply the embryo (*6*). Finally, definitive haematopoietic stem cells arise in the aorta-gonad-mesonephros (AGM) region of the dorsal aorta and seed the fetal liver.

Limited evidence suggests that YS also provides the first blood cells during development in humans. Primitive erythroblasts expressing embryonic globin genes, surrounded by endothelium, emerge at CS6 (~2.5PCW) (*7*, *8*). YS also produces megakaryocytes, mast cells

and myeloid cells (*9*), although this has not yet been evidenced directly in functional studies. Transplantation of human developmental tissues into immunodeficient mice has pinpointed the origin of definitive haematopoietic stem and progenitor cells (HSPCs) (defined as long-term multilineage repopulating cells) within the AGM region of embryo at CS14 (~5PCW) (*10*). Equivalent cells are subsequently found in the YS at CS16 and in the liver from CS17 (*10*). This sequence was also documented by following the transcriptional signature of definitive HSPCs across organs (*11*). While the process of definitive HSPC emergence from hemogenic endothelium (HE) has been reconstructed in human AGM (*11*, *12*), the process by which earlier progenitors arise in human YS has not been studied. Several key questions about human YS haematopoiesis remain unanswered: what is the full repertoire of human YS-derived blood cells, does the YS produce limited progenitors or HSPCs, do YS progenitors/HSPCs contribute to long-lived populations such as tissue-resident macrophages, do YS progenitors/HSPCs arise from HE, and what are the extrinsic regulators of this process.

In this study we report a time-resolved atlas of the human YS combining single cell multiomics with 3D light-sheet microscopy and multiplex RNA *in situ* hybridisation, providing the first comprehensive depiction of the metabolic and haematopoietic functions of the human YS, as well as a benchmark for *in vitro* culture systems aiming to recapitulate early human development.

Results

A single cell atlas of the human yolk sac

We performed droplet-based scRNA-seq profiling of human YS, including both membrane and contents, and integrated with external datasets to yield 169,798 high quality cells from 10 samples spanning 4-8PCW (CS10-CS23), which can be interrogated on our interactive web portal (https://developmental.cellatlas.io/yolk-sac; password: ys2022) (*13*) (**Fig. 1A-C, S1A-C, Table S1-5**). All datasets used for cell state validation, *in vitro* iPSC culture and cross-species comparisons are shown in **Fig. 1A (Table S6-7**).

With iterative graph-based leiden clustering and annotation, the integrated YS scRNA-seq dataset yielded 43 cell types, which we grouped into 15 broad categories including haematopoietic cells, endoderm, mesoderm and mesothelium (**Fig. 1B-C, S1B-C, Table S3-5**). We consistently apply the term HSPC for cells expressing a core HSPC signature e.g. *CD34*, *SPINK2*, *HLF*, without implying long-term repopulating capacity or multilineage potential. With comparison datasets, unless otherwise specified, we adopt published annotations (**Table S6-7**). We demonstrate the key marker genes for these broad cell categories, validated by plate-based scRNA-seq (**Fig. S1D-E, Table S8, S5**) and surface protein expression from CITE-seq analysis of n=2 YS cell suspension (**Fig. 1C, S1F-G, Table 9)**. We used surface protein expression data to construct a decision tree that identified combinatorial antibodies deployable for YS cell type purification and functional characterisation (**Fig. S2A**). FACS isolation of $CD45^-$ cells with high scatter and Smart-seq2 analysis enabled enrichment and validation of YS endoderm, mesothelium and erythroid cells (**Fig. S1D-E, Table S8, S5**). We generated matched embryonic liver scRNA-seq data for established and late stage YS at CS18 and CS22-

23 respectively (n=3; n=2 previously reported (*9*)) (**Fig. S2B-F**), validated by liver CITE-seq data (**Fig. S3A-C, Table S10**), which confirmed the presence of pre-macrophages only in YS and discrete B-cell progenitor stages only in the liver (**Fig. S1C, S3C)**. Around half of YS lymphoid cells were progenitors, which terminated in NK and ILC precursor states on force directed graph (FDG) visualisation (**Fig. S3D**). A small population of cells were termed 'B lineage' due to expression of *CD19, CD79B* and *IGLL1*. They did not express genes typical of B1 cells (*CCR10, CD27, CD5)*. Given the absence of distinct B cell progenitor stages and their later emergence (>5PCW), these may constitute migratory B cells of fetal liver origin (**Fig. S3D-E**).

To facilitate future use of our YS atlas, we employed a low-dimensional logistic regression (LR) framework (see Methods). The trained LR models and weights from trained scVI models are provided via our interactive web portal to enable mapping of scRNA-seq datasets using CellTypist (*14*) and transfer learning with single-cell architectural surgery (scArches) (*15*), respectively. Corresponding cross-tissue projection probability matrices are provided as Supplementary Tables (**Table S11-18**).

Using selected marker genes and proteins from our droplet-based datasets (**Fig. 1C**), we performed 3D visualisation of YS by light sheet microscopy, demonstrating the $CD34^{hi}PVLAP^{hi}LYVE1^{lo}$ vitelline artery and a $CD34^{lo}PVLAP^{lo}LYVE1^{hi}$ vitelline vein contiguous with a branching network of $CD34^{-}PVLAP^{lo}LYVE1^{hi}$ vessels within the YS (**Fig. 1D, Fig. S3F**). The $LYVE1^{hi}$ vessels localised within $HNF4A^{+}$ endoderm and adjacent to the $ECAD^{+}$ mesothelium (**Fig. 1D, S3G-I**). Rare *IL7R*$^{+}$ lymphoid cells, and *CD1C*$^{+}$*C1QA*$^{+/-}$ dendritic cells (DCs) were identified within the mesoderm, while *ACTA2*$^{+}$ smooth muscle cells

surrounded *IL33*$^{+}$ vessels, forming a sub-layer between mesoderm and *SPINK1*$^{+}$ endoderm (**Fig. 1E, S4A-B**).

To investigate the functional relevance of changes in YS cell composition during development, we next assessed proportional representation of cell states by gestation. In early YS (CS10; ~4PCW) HSPCs, erythroid cells, macrophages and megakaryocytes (MK) were the most prevalent cell types with both HSPCs and MKs proportionately diminished thereafter, while production of erythroid cells and macrophages was sustained. DC and *TREM2*$^{+}$ microglia-like cells did not emerge until >6PCW (**Fig. 1F, Table S4**). The ratio of haematopoietic to non-haematopoietic cells was around 3:1 in early YS (CS10; ~4PCW), but with expansion of fibroblasts particularly, the ratio in late YS (CS22-23; ~8PCW) approached 1:3 (**Fig. 1F**). We performed graph-based differential abundance testing with Milo (*16*) to assess heterogeneity within cell states by gestation. Both erythroid cells and macrophages had early and late-gestation specific molecular states but MKs were transcriptionally homogenous across gestation (**Fig. 1F, Table S19**). Early gestation molecular states were characterised by significantly upregulated ribosomal and glycolytic genes (e.g., *RPS29*, *ENO1*, *LDHB*), suggesting a common early burst of translation (**Fig. 1F, S4C, Table S4**). In contrast, cells at later gestation upregulated specific lineage-defining genes compared to their earlier counterparts (**Fig. 1F**).

Multi-organ functions of YS

To probe the nutritional/metabolic functions of YS, we next focused our scRNA-seq analysis on the non-haematopoietic cells. We identified an endoderm cell state co-expressing *APOA1/2*, *APOC3* and *TTR*, similar to embryonic/fetal hepatocytes, which was present from gastrulation

(**Fig. 2A, S4D, Table S3, S20, S7**). Compared with embryonic liver hepatocytes, YS endoderm expressed higher levels of transcripts for serine protease 3 (*PRSS3*), Glutathione S-Transferase Alpha 2 (*GSTA2*) and multi-functional protein Galectin 3 (*LGALS3*), while embryonic liver hepatocytes expressed a more extensive repertoire of detoxification enzymes, including alcohol and aldehyde dehydrogenases (*ADH1A*, *ALDH1A1*) and cytochrome P450 enzymes involved in metabolism of steroid hormones and vitamins (*CYP3A7*), fatty acids (*CYP4A11*) and the conversion of cholesterol to bile acids (*CYP27A1*) (**Fig. S4D, Table S20**). YS endoderm and embryonic liver hepatocytes shared gene modules implicated in coagulation and lipid and glucose metabolism (**Fig. 2B, Table S21**). These modules were also conserved between human and mouse extraembryonic endoderm (**Fig. S4E-F, Table S21**). We validated *in situ* the expression of lipid transport (alpha-fetoprotein, albumin and alpha-1-antitrypsin) and coagulation proteins (fibrin) in YS endoderm and embryonic liver hepatocytes (**Fig. 2C, D**).

We further explored the gene module 'coagulation regulation', conserved between YS and liver, (**Table S21**) by examining expression of pro- and anticoagulant proteins across samples, grouped by gestational age (**Fig. 2D**). From the earliest timepoints, YS endoderm expressed components of the *F3* (Tissue factor)-activated coagulation pathway, particularly *F2* (Thrombin) and *F10* (Factor X) and anticoagulant proteins *SERPINC1* (Antithrombin III) and *PROS1* (Protein S) (**Fig. 2D**). Tissue factor, Antithrombin III and Fibrinogen subunits were also expressed in mouse extraembryonic endoderm (**Fig. S4E**). Factors for the intrinsic pathway (triggered by exposed collagen) were minimally expressed in YS, but were expressed by embryonic liver hepatocytes (**Fig. 2D**). Given the importance of Tissue factor in YS angiogenesis (*17*, *18*), it is likely the coagulant and anticoagulant pathways develop in parallel as a means to balance haemostasis.

In addition to their metabolic and coagulation functions, YS endoderm cells expressed EPO and THPO that are critical for erythropoiesis and megakaryopoiesis (**Fig 2D, S4G**). EPO is known to be produced by mouse YS endoderm (*19*), and both growth factors are produced by human FL hepatocytes (*9*), but their combined presence has not previously been reported in human YS.

We next investigated how multi-organ functions of YS endoderm change over time. Milo-generated DEGs between early and late YS endoderm neighbourhoods revealed active retinoic acid and lipid metabolic processes until 7PCW, after which genes associated with cell stress and death were expressed (**Fig. 2E, Table S22**). Collectively, our findings describe a critical role of the human YS to support haematopoiesis, metabolism, coagulation and erythroid cell mass regulation, prior to these functions being taken over by fetal liver, and ultimately, by adult liver (metabolism and coagulation), bone marrow (haematopoiesis) and kidney (erythroid cell mass regulation) (**Fig. 2F**).

Primitive versus definitive haematopoiesis in YS and liver

Whether the YS is a site of definitive haematopoiesis in humans has not been fully resolved (*20*). Human YS progenitors spanned two groups- HSPC characterised by *SPINK2*, *CYTL1* and *HOXB9*, and cycling HSPC characterised by cell-cycle associated genes such as *MKI67* and *TOP2A* (**Fig. S5A**). HSPC and cycling HSPC had the highest probability of class prediction to fetal/embryonic liver HSC and MPP respectively (**Fig. S5B**). We further identified HSPC sub-states expressing markers characteristic of primitive (*DDIT4, SLC2A3, RGS16, LIN28A*) and definitive (*KIT, ITGA4, CD74, PROCR*) HSPCs in human YS/AGM (*11*), present in both HSPC and cycling HSPC fractions (**Fig. 3A-C**). Both the primitive and definitive HSPC sub-

states expressed canonical HSPC genes such as *SPINK2*, *HOPX* and *HLF* but diverged in expression of genes involved in multiple processes such as enzymes (*GAD1*), growth factors (*FGF23*), adhesion molecules (*SELL*) and patterning genes (*HOXA7*) (**Fig. 3A, S5C**). Differential protein expression in CITE-seq data indicated that CD194 (CCR4), CD357 and CD122 mark primitive HSPCs while CD197 (CCR7), CD193 and CD48 are preferentially expressed on definitive HSPCs (**Fig. S5D**). We confirmed that an iPSC-derived culture system reported to generate definitive HSPCs did express markers characteristic of definitive HSPCs (*11*), but an iPSC-derived culture system optimised for macrophage production (*21*) did not (**Fig. 3A**). To assess cross tissue HSPC heterogeneity, we integrated HSPCs across hematopoietic organs (see Methods) (**Fig. 3C, S5E)**. By kernel density estimation in integrated UMAP embeddings, YS definitive HPSCs co-localized with definitive HSPCs from age-matched liver (**Fig. 3C, S5E).** From exclusively primitive HSPCs at ~3PCW, we observed rapid accumulation of definitive HSPCs after AGM development CS14 (~5PCW), likely accounting for the increase in YS HSPC/progenitor fraction at 8PCW (**Fig. 1F, 3B, S5F**).

Next, we aimed to establish when the liver takes over from YS haematopoiesis and whether it initially uses primitive or definitive HSPCs. Prior to AGM (and to circulation), the human embryonic liver is macroscopically pale, suggesting that erythropoiesis predominantly occurs in the YS (**Fig. 3D**). We tracked the proportional representation of haemoglobin (Hb) subtypes over time as a proxy for YS versus embryonic liver contributions, as the zeta globin chain (*HBZ*) is restricted to primitive erythroblasts while definitive erythroblasts in fetal liver express gamma globins (*HBG1*) (*22*, *23*) (**Fig. 3E)**. Our observation of sustained *HBZ* production in YS, for several days prior to liver bud formation (4 PCW), is consistent with a scenario where YS supports sustained erythropoiesis. The minimal zeta globin expression and expression of *HBZ* repressors by embryonic liver erythroblasts suggests that embryonic liver erythropoiesis

is predominantly definitive, as we have previously shown (*9*), and that the transition to liver-dominated erythropoiesis occurs around 8PCW (**Fig. 3E**). This differs from the situation in mice, whereby immature definitive erythroid progenitors exit the YS and rapidly mature in other sites (*24*), as evidenced by the macroscopically red appearance of the mouse embryonic liver prior to AGM maturation (**Fig. 3D**). Tracking Hb subtype usage in the mouse, we noted two waves of erythropoiesis pre-AGM: primitive haematopoiesis in the YS (initially *Hbb BH1* and *Hba X-Hba1*) and pro-definitive haematopoiesis mirrored in both YS and torso (*Hbb BT1* and *Hbb BS*) (**Fig. S5G-I**).

Given the co-existence of primitive and definitive HSPCs in most of our YS samples, we examined an earlier reference (prior to AGM-HSPC derivation) to explore the differentiation potential of primitive HSPCs. At gastrulation (~3PCW), the YS haematopoietic landscape had a tripartite differentiation structure, with erythroid, MK and myeloid differentiation (**Fig. 3F)**. This structure was also observed in mouse YS **(Fig. S6A-B, Table S13, S5**). A differential-fate-prediction analysis demonstrated that primitive HSPCs pre-AGM at CS10-11 (~4PCW) were biased towards myeloid cell fates, however the abundance of differentiating erythroid and MK cells at this point suggests that an earlier wave of erythroid/MK production occurred (**Fig. 3G, S6C**). Post-AGM, the model predicted that remaining primitive HSPCs were erythroid/MK-biased while definitive HSPCs were biased towards lymphoid and MK fates (**Fig. 3G**). This is in keeping with the first appearance of YS lymphoid cells (ILC progenitors and NK cells, and B lineage cells) post CS14 (**Fig. 1F**). Differential-fate-prediction analyses revealed that iPSC-derived HSPCs were embryonic erythroid (i.e. erythroid cells expressing *HBZ*), myeloid and MK-biased whilst definitive iPSC-derived HSPCs were lymphoid, MK, non-embryonic erythroid and myeloid-primed, consistent with the predicted lineage potential of their *in vivo* primitive and definitive counterparts (**Fig. 3H, S6D**).

The lifespan of YS HSPCs

Following predictions that primitive HSPC differentiation potential changes over time, we sought to examine the extrinsic support received by HSPCs across their lifespan. HSPCs arise from hemogenic endothelium (HE) in the aorta, YS, bone marrow, placenta and embryonic head in mice (*25–29*). In human AGM, definitive HSPCs have recently been shown to emerge from *IL33*$^{+}$*ALDH1A1*$^{+}$ arterial endothelial cells (AEC) via *KCNK17*$^{+}$*ALDH1A1*$^{+}$ HE (*12*). Dissecting YS endothelial cell (EC) states in greater detail, the broad category of PVLAP^{+} ECs included AEC and HE, while LYVE1^{+} ECs encompassed sinusoidal, immature and VWF-expressing ECs (**Fig. 4A, S7A-B, Table S4-5**). HE was a transient feature of early YS (**Fig. 4A**). Along inferred trajectories, YS HSPCs appeared to arise from AEC via HE as in AGM (*11*), sequentially upregulating expected genes such as *ALDH1A1 (30)* (**Fig. 4B**). The same EC intermediate states and transition points could be identified in both iPSC culture systems (**Fig. 4B, S7C**). In keeping with their more recent endothelial origin, we found that YS HSPCs and AGM HSPCs, but not embryonic liver or FBM HSPCs retained an EC gene signature characterised by the expression of *KDR*, *CDH5*, *ESAM* and *PLVAP* (**Fig. 4C**)

Receptor-ligand interactions capable of supporting YS HSPC expansion and maintenance were predicted using CellphoneDB (*31*). We identified ECs, fibroblasts, smooth muscle cells and endoderm as likely interacting partners (**Fig. 4D-E, Table S23**). ECs were predicted to maintain and support the HSPC pool through production of stem cell factor (KITLG) and NOTCH 1/2 ligands DLL1 and JAG1 (*32*, *33*) (**Fig. 4D**). ECs potentially expand the HSPC pool via FN1 while fibroblasts and smooth muscle cells contribute via CSF1 and WNT5A (*34–36*) (**Fig. 4D**). WNT5A may also regulate lineage specification in YS HSPC as non-canonical WNT5A signalling promotes myelopoiesis at the expense of B lymphopoiesis in mouse (*37*).

Endoderm is predicted to interact with HSPCs *via* VTN and THPO, reported to promote haematopoiesis and long-term-HSC-like quiescence, respectively (*38*, *39*) (**Fig. 4D**).

Examining change in interactions over time, the key finding was diminishing interactions between CS17-CS23 (4-8PCW) including loss of cytokine and growth factor support, loss of TFGβ, WNT and NOTCH2 signals by all stromal fractions (**Fig. 4D-E, S7D, Table S24**). Many of these diminishing stromal ligands were still expressed in age-matched liver and AGM stromal cells (**Fig. S7E, Table S24**). Adhesive interactions in YS were also predicted to be significantly modulated (**Fig. S7F, Table S24**). While aged-matched liver provided opportunities for adhesion with stromal cells, AGM did not (**Fig. S7F, Table S24**). Interactions gained between CS17 and CS23 included endoderm-derived IL13 signalling to the *TMEM219*-encoded receptor implicated in apoptosis-induction (**Fig. 4D**). While limited conclusions can be made from studying cells that passed quality control for cell viability, we did observe upregulation in pro-apoptotic gene scores in late stage YS HSPCs, both primitive and definitive (**Fig. S7G**).

Despite marked change in the stromal environment of later stage YS, the proportion of HSPC to cycling HSPC remained stable (**Fig. S5F**). Differential lineage priming analysis revealed that very few HSPCs and mostly terminally differentiated cell states remained at CS22-23/8PCW YS (**Fig. S6C**) Together these observations are in keeping with an early burst of primitive HSPC production from transient YS HE, a later influx of definitive HSPCs derived from AGM, a loss of stromal support between 6-8 PCW, resulting in apoptosis and depletion of remaining HSPCs by terminal differentiation.

An accelerated route to macrophage production in YS and iPSC culture

While YS progenitors are restricted to a short time-window in early gestation, mouse models suggest that they contribute to long-lived macrophage populations in some tissues (*40*). In our YS data, non-progenitor myeloid cells (lacking *AZU1*, *PRTN3* and *MPO*) included *HMGB2*, *LYZ* and *LSP1*-expressing promonocytes, pre-macrophages (expression profile below) and *C1QA/B/C* and *MRC* macrophages (**Fig. S8A**). Monocytes were observed only after liver-bud formation and AGM-derived haematopoiesis at CS14 (~5PCW) (**Fig. 5A, Fig. S8B).** A high probability of class prediction between YS and embryonic liver monocytes was noted (**Fig. 5B, Table S12**). A sub-population of YS monocytes (Monocyte2) expressed *ICAM3*, *SELL*, and *PLAC8* (adhesion molecules expressed by embryonic liver but not YS HSPCs), in keeping with liver-derived monocytes migrating to YS (**Fig. 5C)**. However, sequential waves of monopoiesis occurring within the YS cannot be excluded. YS CITE-seq data confirmed differentiatial expression of CD62L (*SELL)* and CD14 (*CD14*) on Monocyte2 compared to Monocyte1 and identified additional discriminatory markers; e.g., CD15, CD43 for Monocyte1 and CD9, CD35 for Monocyte2 (**Fig. S8C, Table S25**).

The YS pre-macrophage uniquely expressed high levels of *PTGS2*, *MSL1* and *SPIA1*, as well as expressing progenitor genes (*SPINK2*, *CD34*, *SMIM24*), macrophage genes (*C1QA*, *MRC1*), and *CD52* which is typically associated with monocytes (**Fig. S8A**). This YS pre-macrophage rapidly declined by 5PCW and had no equivalent in embryonic liver (**Fig. 5A-B**), which led us to investigate putative macrophage differentiation trajectories in YS. FDG and partition-based graph abstraction (PAGA) visualisation revealed a direct, monocyte-independent trajectory to YS macrophages prior to CS14 (**Fig. 5D)**. In this pre-AGM trajectory, a transition from cycling HSPC to pre-macrophages, then macrophages (**Fig. 5D**) supports our predictions that primitive

pre-AGM HSPCs exhibit myeloid bias (**Fig. 3G**). After CS14, a clear differentiation trajectory from cycling HSPC to monocytes and monocyte-macrophages was seen, but there were few transitional cells connecting to macrophages (**Fig. 5D)**. 15.33% of this macrophage pool was proliferating and CellRank RNA state transition analysis was in keeping with active self-renewal (**Fig. S8D**).

To explore how the two YS macrophage differentiation pathways are regulated, we examined differential transcription factor (TF) usage with PySCENIC (**Fig. 5E, Table S26**). YS pre-macrophages were predicted to use a group of TFs, including FLI1 and MEF2C, that have been reported in the differentiation of multiple lineages (*41*, *42*), whereas the monocyte-dependent route (CMPs, MOP, promonocytes and monocytes) relied on recognised myeloid transcription factors such as SPI1, CEBPA and IRF8 (**Fig. 5E**).

TREM2$^+$ macrophages (*43*), which had a high probability of correlation with fetal brain microglia (*44*), fetal skin TREM2$^+$ (*45*) and fetal testes TREM2$^+$ macrophages (*46*), were observed in YS only after CS14 (**Fig. 5A, 5D, 5F, S8E, Table S18**). By FDG, PAGA and CellRank RNA state transition analysis, TREM2$^+$ macrophages were closely connected to the self-renewing macrophage population (**Fig. 5D, S8D**). *In situ*, YS TREM2$^+$ macrophages were adjacent to the mesothelium, in a region enriched by EC (**Fig. S8F**). CellphoneDB predicted interactions between TREM2$^+$ macrophages and VWF$^+$ EC, via CXCL8 and NRP1, both of which are involved in angiogenic pathways (*47*, *48*) (**Fig. 5G, Table S23**). TREM2$^+$ macrophages also expressed the purinergic receptor P2RY12 that supports trafficking towards ATP/ADP-expressing ECs, as reported in mouse CNS (*49*) (**Fig. 5F**).

In a recent dissection of mouse macrophage heterogeneity across tissues and time, TLF^+ macrophages (emerging from YS progenitors both directly and via a fetal monocyte intermediate) contributed to long-lived self-renewing tissue macrophage populations, while $CCR2^+$ and MHC-II^{hi} macrophage pools received continual input from monocytes (*50*). To explore the contribution of YS-derived macrophages in prenatal human tissues, we interrogated a human pan-fetal immune cell atlas (*45*) to explore the proportion of cells representing each macrophage pool over gestational time (**Fig. S8G**). TLF^+ cells first emerged in YS, and were subsequently found in both liver and skin. The proportion of TLF^+ macrophages decreased over gestational time in all tissues, while the proportion of MHC-II^{hi} macrophages increased, particularly after 10PCW. While our data cannot discern whether TLF^+ macrophages are upregulating an MHC-II^{hi} transcriptional programme or whether TLF^+ macrophages are being replaced by a second wave of monocyte-derived macrophages, we can conclude that the TLF^+ signature, temporally consistent with a YS origin, does not persist in humans to the extent observed in mouse (*50*). TLF^+ macrophages in fetal liver were found within proliferating Kupffer cells ('Kupffer 1' in Popescu *et al* (*9*)). Liver and YS CITE-seq data confirmed protein expression of CD206 and CD163 in TLF^+ macrophages and further identified CD28, CD49a, TSLPR and CD144 as positive identification markers that are minimally expressed in remaining Kupffer cells (**Fig. S8H, Table S27)**.

As macrophage differentiation from iPSCs permits high-resolution sampling over time, we sought to establish whether the macrophage subsets and developmental pathways we observed were recapitulated *in vitro*. To this end, we integrated our YS gene expression data with scRNA-seq data from iPSC-derived macrophage differentiation (n=19; k=50,512) by Alsinet et al (*21*) after refining the annotations of iPSC-derived cell-states (see Methods) (**Fig. S9A-D, Table S14, S5**). Non-adherent, CD14-expressing cells appearing after week 2 of differentiation

394 expressed *C1QA*, *C1QB* and *APOC1* in keeping with a macrophage identity, while *S100A8/9*,
395 *FCN1*, *CD52* and C*D14*-expressing monocytes only emerged after week 3 (**Fig. S9A-D**). Prior
396 to monocyte emergence, a monocyte-independent macrophage differentiation trajectory was
397 observed, consistent with the observations by Alsinet et al (*21*) (**Fig. 5H, Fig. S9A-D**). TF
398 regulatory profiles of iPSC-derived macrophage differentiation were consistent with both pre-
399 macrophage and monocyte-dependent *in vivo* TF profiles, including usage of *MEF2C, SPI1*,
400 *CEBPA* and *IRF8* in iPSC-derived pre-macrophages (**Fig. 5I, Table 26**). However, neither
401 system recapitulated TREM2^{+} macrophages suggesting that stromal cells, specifically ECs,
402 may be required to acquire the TREM2^{+} molecular profile.
403
404 **Discussion**
405
406 Using state-of-the-art single cell multiomic and imaging technologies we delineate the dynamic
407 composition and functions of human YS *in vivo* from the 3rd post conception week (PCW),
408 when the three embryonic germ layers form, to the 8th PCW when the majority of organ
409 structures are already established (*22*). We detail how the YS endoderm shares metabolic and
410 biosynthetic functions with liver and erythropoiesis-stimulating functions with liver and
411 kidney. In part, this shared functionality may relate to a common role in creating a niche for
412 haematopoiesis (*51*). Unlike in mice, where primitive erythroid progenitors mature in the YS
413 (prior to circulation being established) but erythromyeloid progenitors can exit the YS and
414 mature in the fetal liver, we show that active differentiation of erythroid and macrophage cells
415 occurs for several weeks in human YS, prior to liver handover. The multi-organ functions,
416 including extended haematopoiesis, of human YS may be an evolutionary adaptation to the
417 longer gestation in humans. While an earlier study in humans, based on colony assays,

418 suggested that the transition from YS to liver occurred at 5PCW, no YS samples were studied
419 after this point (*52*).
420
421 The developmental window investigated here encompasses haematopoiesis from HSPCs
422 arising both within the YS and within the embryo proper. We reconstruct YS HSPCs
423 emergence from a temporally-restricted HE, featuring similar transition states and molecular
424 regulation to AGM HSPCs. By gastrulation (CS7; ~3PCW), YS HSPCs already differentiate
425 into primitive erythroid, MK and myeloid lineages. Building on a recent compilation of gene
426 scorecards that characterise primitive and definitive HSPCs (*11*), we were able to parse the two
427 fractions and document transition to definitive HSPC-dominance after CS14 (~5PCW). This
428 separation also allowed us to identify a primitive HSPC bias towards myeloid, erythroid and
429 megakaryocyte lineages and a definitive HSPCs bias towards megakaryocyte and lymphoid
430 lineages. Both primitive and definitive HSPCs in the YS became more quiescent and
431 upregulated apoptosis-related genes between CS17 and CS23 (~6-8PCW). Stromal cell ligands
432 predicted to support HSPCs were markedly disrupted during this time, suggesting that the
433 barriers to the survival of YS HSPCS may be extrinsic.
434
435 Primitive HSPCs uniquely employ an accelerated route to macrophage production independent
436 of monocytes. The monocyte-dependent route may provide more tunable macrophage
437 production via circulating innate immune cells to facilitate macrophage regeneration in
438 response to tissue damage, inflammation or infection. While both primitive and definitive
439 HSPCs, 'accelerated' and monocyte-dependent macrophages were recapitulated during *in*
440 *vitro* differentiation of iPSCs, TREM2^{+} macrophages were not. TREM2^{+} macrophages, which
441 are transcriptionally aligned with brain microglia, fetal skin, testes and AGM TREM2^{+}
442 macrophages, were predicted to interact with endothelial cells, potentially supporting

443 angiogenesis as described in mouse brain (*44*). Benchmarking of *in vitro* cultures against *in*
444 *vivo* cell states and trajectories can facilitate more faithful replication of early blood and
445 immune cells.
446
447 There is a growing appreciation of the potentially life-long consequences of early
448 developmental processes. Our study illuminates a previously obscure phase of human
449 development, where vital organismal functions are delivered by a transient extraembryonic
450 organ employing non-canonical cellular differentiation paths. It will be fascinating to explore
451 how these processes may impact on tissue homeostasis and disease across the human lifespan.

References and Notes:

1. C. Ross, T. E. Boroviak, Origin and function of the yolk sac in primate embryogenesis. *Nat. Commun.* **11**, 3760 (2020).

2. T. Cindrova-Davies, E. Jauniaux, M. G. Elliot, S. Gong, G. J. Burton, D. S. Charnock-Jones, RNA-seq reveals conservation of function among the yolk sacs of human, mouse, and chicken. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E4753–E4761 (2017).

3. H. W. Mossman, *Vertebrate Fetal Membranes: Comparative Ontogeny and Morphology, Evolution, Phylogenetic Significance, Basic Functions, Research Opportunities* (1987).

4. T. Yamane, Mouse Yolk Sac Hematopoiesis. *Front. Cell Dev. Biol.* **0** (2018), doi:10.3389/fcell.2018.00080.

5. J. Palis, J. Malik, K. E. Mcgrath, P. D. Kingsley, Primitive erythropoiesis in the mammalian embryo. *The International Journal of Developmental Biology*. **54** (2010), pp. 1011–1018.

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

7. M. Tavian, M. F. Hallais, B. Péault, Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo. *Development*. **126**, 793–803 (1999).

8. C. Peschle, A. R. Migliaccio, G. Migliaccio, M. Petrini, M. Calandrini, G. Russo, G. Mastroberardino, M. Presta, A. M. Gianni, P. Comi, Embryonic----Fetal Hb switch in humans: studies on erythroid bursts generated by embryonic progenitors from yolk sac and liver. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 2416–2420 (1984).

9. D.-M. Popescu, R. A. Botting, E. Stephenson, K. Green, S. Webb, L. Jardine, E. F. Calderbank, K. Polanski, I. Goh, M. Efremova, M. Acres, D. Maunder, P. Vegh, Y. Gitton, J.-E. Park, R. Vento-Tormo, Z. Miao, D. Dixon, R. Rowell, D. McDonald, J. Fletcher, E. Poyner, G. Reynolds, M. Mather, C. Moldovan, L. Mamanova, F. Greig, M. D. Young, K. B. Meyer, S. Lisgo, J.

476 Bacardit, A. Fuller, B. Millar, B. Innes, S. Lindsay, M. J. T. Stubbington, M. S. Kowalczyk, B.
477 Li, O. Ashenberg, M. Tabaka, D. Dionne, T. L. Tickle, M. Slyper, O. Rozenblatt-Rosen, A.
478 Filby, P. Carey, A.-C. Villani, A. Roy, A. Regev, A. Chédotal, I. Roberts, B. Göttgens, S.
479 Behjati, E. Laurenti, S. A. Teichmann, M. Haniffa, Decoding human fetal liver haematopoiesis.
480 *Nature*. **574**, 365–371 (2019).

481 10. A. Ivanovs, S. Rybtsov, L. Welch, R. A. Anderson, M. L. Turner, A. Medvinsky, Highly potent
482 human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros
483 region. *J. Exp. Med.* **208**, 2417–2427 (2011).

484 11. V. Calvanese, S. Capellera-Garcia, F. Ma, I. Fares, S. Liebscher, E. S. Ng, S. Ekstrand, J.
485 Aguadé-Gorgorió, A. Vavilina, D. Lefaudeux, B. Nadel, J. Y. Li, Y. Wang, L. K. Lee, R.
486 Ardehali, M. L. Iruela-Arispe, M. Pellegrini, E. G. Stanley, A. G. Elefanty, K. Schenke-Layland,
487 H. K. A. Mikkola, Mapping human haematopoietic stem cells from haemogenic endothelium to
488 birth. *Nature*. **604**, 534–540 (2022).

489 12. Y. Zeng, J. He, Z. Bai, Z. Li, Y. Gong, C. Liu, Y. Ni, J. Du, C. Ma, L. Bian, Y. Lan, B. Liu,
490 Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA
491 sequencing. *Cell Res.* **29**, 881–894 (2019).

492 13. D. Horsfall, J. McGrath, Adifa software for Single Cell Insights (2022),
493 doi:10.5281/zenodo.5824896.

494 14. C. Domínguez Conde, C. Xu, L. B. Jarvis, D. B. Rainbow, S. B. Wells, T. Gomes, S. K. Howlett,
495 O. Suchanek, K. Polanski, H. W. King, L. Mamanova, N. Huang, P. A. Szabo, L. Richardson, L.
496 Bolt, E. S. Fasouli, K. T. Mahbubani, M. Prete, L. Tuck, N. Richoz, Z. K. Tuong, L. Campos, H.
497 S. Mousa, E. J. Needham, S. Pritchard, T. Li, R. Elmentaite, J. Park, E. Rahmani, D. Chen, D. K.
498 Menon, O. A. Bayraktar, L. K. James, K. B. Meyer, N. Yosef, M. R. Clatworthy, P. A. Sims, D.
499 L. Farber, K. Saeb-Parsy, J. L. Jones, S. A. Teichmann, Cross-tissue immune cell analysis
500 reveals tissue-specific features in humans. *Science*. **376**, eabl5197 (2022).

15. M. Lotfollahi, M. Naghipourfar, M. D. Luecken, M. Khajavi, M. Büttner, M. Wagenstetter, Ž. Avsec, A. Gayoso, N. Yosef, M. Interlandi, S. Rybakov, A. V. Misharin, F. J. Theis, Mapping single-cell data to reference atlases by transfer learning. *Nat. Biotechnol.* **40**, 121–130 (2022).

16. E. Dann, N. C. Henderson, S. A. Teichmann, M. D. Morgan, J. C. Marioni, Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat. Biotechnol.* (2021), doi:10.1038/s41587-021-01033-z.

17. J. Xue, Q. Wu, L. A. Westfield, E. A. Tuley, D. Lu, Q. Zhang, K. Shim, X. Zheng, J. E. Sadler, Incomplete embryonic lethality and fatal neonatal hemorrhage caused by prothrombin deficiency in mice. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 7603–7607 (1998).

18. W. Ruf, N. Yokota, F. Schaffner, Tissue factor in cancer progression and angiogenesis. *Thrombosis Research*. **125** (2010), pp. S36–S38.

19. Y. Yasuda, M. Okano, M. Nagao, S. Masuda, H. Konishi, K. Ueda, T. Matsuo, K. Tsujiguchi, S. Tajima, R. Sasaki, T. Tanimura, Erythropoietin in mouse avascular yolk sacs is increased by retinoic acid. *Dev. Dyn.* **207**, 184–194 (1996).

20. J. Palis, M. C. Yoder, Yolk-sac hematopoiesis: the first blood cells of mouse and man. *Exp. Hematol.* **29**, 927–936 (2001).

21. C. Alsinet, M. N. Primo, V. Lorenzi, E. Bello, I. Kelava, C. P. Jones, R. Vilarrasa-Blasi, C. Sancho-Serra, A. J. Knights, J.-E. Park, B. S. Wyspianska, G. Trynka, D. F. Tough, A. Bassett, D. J. Gaffney, D. Alvarez-Errico, R. Vento-Tormo, Robust temporal map of human in vitro myelopoiesis using single-cell genomics. *Nat. Commun.* **13**, 1–17 (2022).

22. C. Peschle, F. Mavilio, A. Carè, G. Migliaccio, A. R. Migliaccio, G. Salvo, P. Samoggia, S. Petti, R. Guerriero, M. Marinucci, Haemoglobin switching in human embryos: asynchrony of zeta----alpha and epsilon----gamma-globin switches in primitive and definite erythropoietic lineage. *Nature*. **313**, 235–238 (1985).

525 23. J. Palis, Primitive and definitive erythropoiesis in mammals. *Frontiers in Physiology*. **5** (2014), ,
526 doi:10.3389/fphys.2014.00003.

527 24. K. McGrath, J. Palis, Ontogeny of erythropoiesis in the mammalian embryo. *Curr. Top. Dev.*
528 *Biol.* **82**, 1–22 (2008).

529 25. T. Jaffredo, R. Gautier, A. Eichmann, F. Dieterlen-Lièvre, Intraaortic hemopoietic cells are
530 derived from endothelial cells during ontogeny. *Development*. **125**, 4575–4583 (1998).

531 26. L. Yvernogeau, R. Gautier, L. Petit, H. Khoury, F. Relaix, V. Ribes, H. Sang, P. Charbord, M.
532 Souyri, C. Robin, T. Jaffredo, In vivo generation of haematopoietic stem/progenitor cells from
533 bone marrow-derived haemogenic endothelium. *Nat. Cell Biol.* **21**, 1334–1345 (2019).

534 27. Z. Li, Y. Lan, W. He, D. Chen, J. Wang, F. Zhou, Y. Wang, H. Sun, X. Chen, C. Xu, S. Li, Y.
535 Pang, G. Zhang, L. Yang, L. Zhu, M. Fan, A. Shang, Z. Ju, L. Luo, Y. Ding, W. Guo, W. Yuan,
536 X. Yang, B. Liu, Mouse Embryonic Head as a Site for Hematopoietic Stem Cell Development.
537 *Cell Stem Cell*. **11** (2012), pp. 663–675.

538 28. J. M. Frame, K. H. Fegan, S. J. Conway, K. E. McGrath, J. Palis, Definitive Hematopoiesis in
539 the Yolk Sac Emerges from Wnt-Responsive Hemogenic Endothelium Independently of
540 Circulation and Arterial Identity. *Stem Cells*. **34**, 431–444 (2016).

541 29. K. E. Rhodes, C. Gekas, Y. Wang, C. T. Lux, C. S. Francis, D. N. Chan, S. Conway, S. H. Orkin,
542 M. C. Yoder, H. K. A. Mikkola, The emergence of hematopoietic stem cells is initiated in the
543 placental vasculature in the absence of circulation. *Cell Stem Cell*. **2**, 252–263 (2008).

544 30. R. Thambyrajah, M. Mazan, R. Patel, V. Moignard, M. Stefanska, E. Marinopoulou, Y. Li, C.
545 Lancrin, T. Clapes, T. Möröy, C. Robin, C. Miller, S. Cowley, B. Göttgens, V. Kouskoff, G.
546 Lacaud, GFI1 proteins orchestrate the emergence of haematopoietic stem cells through
547 recruitment of LSD1. *Nat. Cell Biol.* **18**, 21–32 (2016).

548 31. M. Efremova, M. Vento-Tormo, S. A. Teichmann, R. Vento-Tormo, CellPhoneDB v2.0:

549 Inferring cell-cell communication from combined expression of multi-subunit receptor-ligand
550 complexes, , doi:10.1101/680926.

551 32. L. Jardine, S. Webb, I. Goh, M. Quiroga Londoño, G. Reynolds, M. Mather, B. Olabi, E.
552 Stephenson, R. A. Botting, D. Horsfall, J. Engelbert, D. Maunder, N. Mende, C. Murnane, E.
553 Dann, J. McGrath, H. King, I. Kucinski, R. Queen, C. D. Carey, C. Shrubsole, E. Poyner, M.
554 Acres, C. Jones, T. Ness, R. Coulthard, N. Elliott, S. O'Byrne, M. L. R. Haltalli, J. E. Lawrence,
555 S. Lisgo, P. Balogh, K. B. Meyer, E. Prigmore, K. Ambridge, M. S. Jain, M. Efremova, K.
556 Pickard, T. Creasey, J. Bacardit, D. Henderson, J. Coxhead, A. Filby, R. Hussain, D. Dixon, D.
557 McDonald, D.-M. Popescu, M. S. Kowalczyk, B. Li, O. Ashenberg, M. Tabaka, D. Dionne, T. L.
558 Tickle, M. Slyper, O. Rozenblatt-Rosen, A. Regev, S. Behjati, E. Laurenti, N. K. Wilson, A.
559 Roy, B. Göttgens, I. Roberts, S. A. Teichmann, M. Haniffa, Blood and immune development in
560 human fetal bone marrow and Down syndrome. *Nature*. **598**, 327–331 (2021).

561 33. J. Fröbel, T. Landspersky, G. Percin, C. Schreck, S. Rahmig, A. Ori, D. Nowak, M. Essers, C.
562 Waskow, R. A. J. Oostendorp, The Hematopoietic Bone Marrow Niche Ecosystem. *Front Cell*
563 *Dev Biol*. **9**, 705410 (2021).

564 34. P. Zhang, C. Zhang, J. Li, J. Han, X. Liu, H. Yang, The physical microenvironment of
565 hematopoietic stem cells and its emerging roles in engineering applications. *Stem Cell Res. Ther*.
566 **10**, 327 (2019).

567 35. S. Kaur, A. Sehgal, A. C. Wu, S. M. Millard, L. Batoon, C. J. Sandrock, M. Ferrari-Cestari, J.-P.
568 Levesque, D. A. Hume, L. J. Raggatt, A. R. Pettit, Stable colony-stimulating factor 1 fusion
569 protein treatment increases hematopoietic stem cell pool and enhances their mobilisation in mice.
570 *J. Hematol. Oncol.* **14**, 3 (2021).

571 36. B. Murdoch, K. Chadwick, M. Martin, F. Shojaei, K. V. Shah, L. Gallacher, R. T. Moon, M.
572 Bhatia, Wnt-5A augments repopulating capacity and primitive hematopoietic development of
573 human blood stem cells in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3422–3427 (2003).

574 37. H. Liang, Q. Chen, A. H. Coles, S. J. Anderson, G. Pihan, A. Bradley, R. Gerstein, R. Jurecic, S.
575 N. Jones, Wnt5a inhibits B cell proliferation and functions as a tumor suppressor in
576 hematopoietic tissue. *Cancer Cell*. **4**, 349–360 (2003).

577 38. J. Shen, Y. Zhu, S. Zhang, S. Lyu, C. Lyu, Z. Feng, D. L. Hoyle, Z. Z. Wang, T. Cheng,
578 Vitronectin-activated αvβ3 and αvβ5 integrin signalling specifies haematopoietic fate in human
579 pluripotent stem cells. *Cell Prolif.* **54**, e13012 (2021).

580 39. H. Yoshihara, F. Arai, K. Hosokawa, T. Hagiwara, K. Takubo, Y. Nakamura, Y. Gomei, H.
581 Iwasaki, S. Matsuoka, K. Miyamoto, H. Miyazaki, T. Takahashi, T. Suda, Thrombopoietin/MPL
582 signaling regulates hematopoietic stem cell quiescence and interaction with the osteoblastic
583 niche. *Cell Stem Cell*. **1**, 685–697 (2007).

584 40. F. Ginhoux, M. Guilliams, Tissue-Resident Macrophage Ontogeny and Homeostasis. *Immunity*.
585 **44**, 439–449 (2016).

586 41. C. Gekas, K. E. Rhodes, L. M. Gereige, H. Helgadottir, R. Ferrari, S. K. Kurdistani, E.
587 Montecino-Rodriguez, R. Bassel-Duby, E. Olson, A. V. Krivtsov, S. Armstrong, S. H. Orkin, M.
588 Pellegrini, H. K. A. Mikkola, Mef2C is a lineage-restricted target of Scl/Tal1 and regulates
589 megakaryopoiesis and B-cell homeostasis. *Blood*. **113**, 3461–3471 (2009).

590 42. E. Suzuki, S. Williams, S. Sato, G. Gilkeson, D. K. Watson, X. K. Zhang, The transcription
591 factor Fli-1 regulates monocyte, macrophage and dendritic cell development in mice.
592 *Immunology*. **139**, 318–327 (2013).

593 43. F. Wang, P. Ding, X. Liang, X. Ding, C. B. Brandt, E. Sjöstedt, J. Zhu, S. Bolund, L. Zhang, L.
594 P. M. H. de Rooij, L. Luo, Y. Wei, W. Zhao, Z. Lv, J. Haskó, R. Li, Q. Qin, Y. Jia, W. Wu, Y.
595 Yuan, M. Pu, H. Wang, A. Wu, L. Xie, P. Liu, F. Chen, J. Herold, J. Kalucka, M. Karlsson, X.
596 Zhang, R. B. Helmig, L. Fagerberg, C. Lindskog, F. Pontén, M. Uhlen, L. Bolund, N. Jessen, H.
597 Jiang, X. Xu, H. Yang, P. Carmeliet, J. Mulder, D. Chen, L. Lin, Y. Luo, Endothelial cell
598 heterogeneity and microglia regulons revealed by a pig cell landscape at single-cell level. *Nat.*

Commun. **13**, 1–18 (2022).

44. U. C. Eze, A. Bhaduri, M. Haeussler, T. J. Nowakowski, A. R. Kriegstein, Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. *Nat. Neurosci.* **24**, 584–594 (2021).

45. C. Suo, E. Dann, I. Goh, L. Jardine, V. Kleshchevnikov, J.-E. Park, R. A. Botting, E. Stephenson, J. Engelbert, Z. K. Tuong, K. Polanski, N. Yayon, C. Xu, O. Suchanek, R. Elmentaite, C. Domínguez Conde, P. He, S. Pritchard, M. Miah, C. Moldovan, A. S. Steemers, P. Mazin, M. Prete, D. Horsfall, J. C. Marioni, M. R. Clatworthy, M. Haniffa, S. A. Teichmann, Mapping the developing human immune system across organs. *Science*. **376**, eabo0510 (2022).

46. R. Vento-Tormo, L. Garcia-Alonso, V. Lorenzi, C. Mazzeo, C. Sancho-Serra, K. Roberts, J. Engelbert, J. Alves-Lopes, M. Marečková, R. Botting, T. Li, B. Crespo, S. van Dongen, V. Kiselev, E. Prigmore, A. Moffett, M. Herbert, O. A. Bayraktar, A. Surani, M. Haniffa, Single-cell roadmap of human gonadal development (2021), doi:10.21203/rs.3.rs-496470/v1.

47. M. L. Petreaca, M. Yao, Y. Liu, K. DeFea, M. Martins-Green, Transactivation of Vascular Endothelial Growth Factor Receptor-2 by Interleukin-8 (IL-8/CXCL8) Is Required for IL-8/CXCL8-induced Endothelial Permeability. *Mol. Biol. Cell* (2007), doi:10.1091/mbc.e07-01-0004.

48. Z. Lyu, H. Jin, Z. Yan, K. Hu, H. Jiang, H. Peng, H. Zhuo, Effects of NRP1 on angiogenesis and vascular maturity in endothelial cells are dependent on the expression of SEMA4D. *Int. J. Mol. Med.* **46**, 1321–1334 (2020).

49. K. Bisht, K. A. Okojie, K. Sharma, D. H. Lentferink, Y.-Y. Sun, H.-R. Chen, J. O. Uweru, S. Amancherla, Z. Calcuttawala, A. B. Campos-Salazar, B. Corliss, L. Jabbour, J. Benderoth, B. Friestad, W. A. Mills 3rd, B. E. Isakson, M.-È. Tremblay, C.-Y. Kuan, U. B. Eyo, Capillary-associated microglia regulate vascular structure and function through PANX1-P2RY12 coupling in mice. *Nat. Commun.* **12**, 5289 (2021).

50. S. A. Dick, A. Wong, H. Hamidzada, S. Nejat, R. Nechanitzky, S. Vohra, B. Mueller, R. Zaman, C. Kantores, L. Aronoff, A. Momen, D. Nechanitzky, W. Y. Li, P. Ramachandran, S. Q. Crome, B. Becher, M. I. Cybulsky, F. Billia, S. Keshavjee, S. Mital, C. S. Robbins, T. W. Mak, S. Epelman, Three tissue resident macrophage subsets coexist across organs with conserved origins and life cycles. *Sci Immunol*. **7**, eabf7777 (2022).

51. S. Chou, H. F. Lodish, Fetal liver hepatic progenitors are supportive stromal cells for hematopoietic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **107** (2010), doi:10.1073/pnas.1003586107.

52. G. Migliaccio, A. R. Migliaccio, S. Petti, F. Mavilio, G. Russo, D. Lazzaro, U. Testa, M. Marinucci, C. Peschle, Human embryonic hemopoiesis. Kinetics of progenitors and precursors underlying the yolk sac----liver transition. *Journal of Clinical Investigation*. **78** (1986), pp. 51–60.

53. H. Wang, J. He, C. Xu, X. Chen, H. Yang, S. Shi, C. Liu, Y. Zeng, D. Wu, Z. Bai, M. Wang, Y. Wen, P. Su, M. Xia, B. Huang, C. Ma, L. Bian, Y. Lan, T. Cheng, L. Shi, B. Liu, J. Zhou, Decoding Human Megakaryocyte Development. *Cell Stem Cell*. **28**, 535–549.e8 (2021).

54. L. Jardine, S. Webb, I. Goh, M. Quiroga Londoño, G. Reynolds, M. Mather, B. Olabi, E. Stephenson, R. A. Botting, D. Horsfall, J. Engelbert, D. Maunder, N. Mende, C. Murnane, E. Dann, J. McGrath, H. King, I. Kucinski, R. Queen, C. D. Carey, C. Shrubsole, E. Poyner, M. Acres, C. Jones, T. Ness, R. Coulthard, N. Elliott, S. O'Byrne, M. L. R. Haltalli, J. E. Lawrence, S. Lisgo, P. Balogh, K. B. Meyer, E. Prigmore, K. Ambridge, M. S. Jain, M. Efremova, K. Pickard, T. Creasey, J. Bacardit, D. Henderson, J. Coxhead, A. Filby, R. Hussain, D. Dixon, D. McDonald, D.-M. Popescu, M. S. Kowalczyk, B. Li, O. Ashenberg, M. Tabaka, D. Dionne, T. L. Tickle, M. Slyper, O. Rozenblatt-Rosen, A. Regev, S. Behjati, E. Laurenti, N. K. Wilson, A. Roy, B. Göttgens, I. Roberts, S. A. Teichmann, M. Haniffa, Blood and immune development in human fetal bone marrow and Down syndrome. *Nature*. **598**, 327–331 (2021).

649 55. B. J. Stewart, J. R. Ferdinand, M. D. Young, T. J. Mitchell, K. W. Loudon, A. M. Riding, N.
650 Richoz, G. L. Frazer, J. U. L. Staniforth, F. A. Vieira Braga, R. A. Botting, D.-M. Popescu, R.
651 Vento-Tormo, E. Stephenson, A. Cagan, S. J. Farndon, K. Polanski, M. Efremova, K. Green, M.
652 Del Castillo Velasco-Herrera, C. Guzzo, G. Collord, L. Mamanova, T. Aho, J. N. Armitage, A.
653 C. P. Riddick, I. Mushtaq, S. Farrell, D. Rampling, J. Nicholson, A. Filby, J. Burge, S. Lisgo, S.
654 Lindsay, M. Bajenoff, A. Y. Warren, G. D. Stewart, N. Sebire, N. Coleman, M. Haniffa, S. A.
655 Teichmann, S. Behjati, M. R. Clatworthy, Spatiotemporal immune zonation of the human
656 kidney. *Science*. **365**, 1461–1466 (2019).

657 56. G. L. Website, Tracking Early Mammalian Organogenesis – Prediction and Validation of
658 Differentiation Trajectories at Whole Organism Scale. *ExtendedMouseAtlas*, (available at
659 https://marionilab.github.io/ExtendedMouseAtlas/).

660 57. R. C. V. Tyser, E. Mahammadov, S. Nakanoh, L. Vallier, A. Scialdone, S. Srinivas, Single-cell
661 transcriptomic characterization of a gastrulating human embryo. *Nature*. **600**, 285–289 (2021).

662 58. E. I. Crosse, S. Gordon-Keylock, S. Rybtsov, A. Binagui-Casas, H. Felchle, N. C. Nnadi, K.
663 Kirschner, T. Chandra, S. Tamagno, D. J. Webb, F. Rossi, R. A. Anderson, A. Medvinsky,
664 Multi-layered Spatial Transcriptomics Identify Secretory Factors Promoting Human
665 Hematopoietic Stem Cell Development. *Cell Stem Cell*. **27**, 822–839.e8 (2020).

666 59. T. Strachan, S. Lindsay, D. I. Wilson, *Molecular Genetics of Early Human Development* (Taylor
667 & Francis, 1997).

668 60. M. Belle, D. Godefroy, G. Couly, S. A. Malone, F. Collier, P. Giacobini, A. Chédotal,
669 Tridimensional Visualization and Analysis of Early Human Development. *Cell*. **169**, 161–
670 173.e12 (2017).

671 61. A.-C. Villani, R. Satija, G. Reynolds, S. Sarkizova, K. Shekhar, J. Fletcher, M. Griesbeck, A.
672 Butler, S. Zheng, S. Lazo, L. Jardine, D. Dixon, E. Stephenson, E. Nilsson, I. Grundberg, D.
673 McDonald, A. Filby, W. Li, P. L. De Jager, O. Rozenblatt-Rosen, A. A. Lane, M. Haniffa, A.

Regev, N. Hacohen, Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*. **356** (2017), doi:10.1126/science.aah4573.

62. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* **8**, 281–291.e9 (2019).

63. S. J. Fleming, J. C. Marioni, M. Babadi, CellBender remove-background: a deep generative model for unsupervised removal of background noise from scRNA-seq datasets, , doi:10.1101/791699.

64. F. A. Wolf, F. Alexander Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*. **19** (2018), , doi:10.1186/s13059-017-1382-0.

65. M. P. Mulè, A. J. Martins, J. S. Tsang, Normalizing and denoising protein expression data from droplet-based single cell profiling, , doi:10.1101/2020.02.24.963603.

66. R. Lopez, J. Regier, M. B. Cole, M. I. Jordan, N. Yosef, Deep generative modeling for single-cell transcriptomics. *Nat. Methods*. **15**, 1053–1058 (2018).

67. A. Gayoso, R. Lopez, G. Xing, P. Boyeau, V. Valiollah Pour Amiri, J. Hong, K. Wu, M. Jayasuriya, E. Mehlman, M. Langevin, Y. Liu, J. Samaran, G. Misrachi, A. Nazaret, O. Clivio, C. Xu, T. Ashuach, M. Gabitto, M. Lotfollahi, V. Svensson, E. da Veiga Beltrame, V. Kleshchevnikov, C. Talavera-López, L. Pachter, F. J. Theis, A. Streets, M. I. Jordan, J. Regier, N. Yosef, A Python library for probabilistic analysis of single-cell omics data. *Nat. Biotechnol.* **40**, 163–166 (2022).

68. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P.-R. Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods*. **16**, 1289–1296 (2019).

69. M. Büttner, Z. Miao, F. A. Wolf, S. A. Teichmann, F. J. Theis, A test metric for assessing single-cell RNA-seq batch correction. *Nat. Methods*. **16**, 43–49 (2019).

70. V. A. Traag, L. Waltman, N. J. van Eck, From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 1–12 (2019).

71. C. D. Conde, C. Domínguez Conde, C. Xu, L. B. Jarvis, T. Gomes, S. K. Howlett, D. B. Rainbow, O. Suchanek, H. W. King, L. Mamanova, K. Polanski, N. Huang, E. S. Fasouli, K. T. Mahbubani, M. Prete, L. Tuck, N. Richoz, Z. K. Tuong, L. Campos, H. S. Mousa, E. J. Needham, S. Pritchard, T. Li, R. Elmentaite, J. Park, D. K. Menon, O. A. Bayraktar, L. K. James, K. B. Meyer, M. R. Clatworthy, K. Saeb-Parsy, J. L. Jones, S. A. Teichmann, Cross-tissue immune cell analysis reveals tissue-specific adaptations and clonal architecture in humans, , doi:10.1101/2021.04.28.441762.

72. A. J. Tarashansky, J. M. Musser, M. Khariton, P. Li, D. Arendt, S. R. Quake, B. Wang, Mapping single-cell atlases throughout Metazoa unravels cell type evolution. *Elife*. **10** (2021), doi:10.7554/eLife.66747.

73. A. Edelstein, N. Amodaj, K. Hoover, R. Vale, N. Stuurman, Computer Control of Microscopes Using μManager. *Current Protocols in Molecular Biology*. **92** (2010), , doi:10.1002/0471142727.mb1420s92.

74. D. Hörl, F. Rojas Rusak, F. Preusser, P. Tillberg, N. Randel, R. K. Chhetri, A. Cardona, P. J. Keller, H. Harz, H. Leonhardt, M. Treier, S. Preibisch, BigStitcher: reconstructing high-resolution image datasets of cleared and expanded samples. *Nat. Methods*. **16**, 870–874 (2019).

75. M. Gataric, J. S. Park, T. Li, V. Vaskivskyi, J. Svedlund, C. Strell, K. Roberts, M. Nilsson, L. R. Yates, O. Bayraktar, M. Gerstung, PoSTcode: Probabilistic image-based spatial transcriptomics decoder. *bioRxiv* (2021), p. 2021.10.12.464086.

76. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. *Nat. Methods*. **9**, 676–682 (2012).

Acknowledgements:

For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. We thank the Newcastle University Flow Cytometry Core Facility; Sanger Institute Cellular Genetics IT; Newcastle upon Tyne NHS Trust NovoPath; Tamilvendhan Dhanaseelan of HDBR for assistance with human fetal tissue processing and cell freezing; Natalina Elliott for contributions toward CITE-seq panel design.

Funding:

We acknowledge funding from the Wellcome Human Cell Atlas Strategic Science Support (WT211276/Z/18/Z), MRC Human Cell Atlas award, Wellcome Human Developmental Biology Initiative, and HDBR (MRC/Wellcome MR/R006237/1).

MH is funded by Wellcome (WT107931/Z/15/Z), The Lister Institute for Preventive Medicine and NIHR and Newcastle Biomedical Research Centre.

SAT is funded by Wellcome (WT206194) and the ERC Consolidator Grant ThDEFINE.

Relevant research in the BG group was funded by Wellcome (206328/Z/17/Z) and the MRC (MR/M008975/1 and MR/S036113/1).

IR is funded by Blood Cancer UK and by the NIHR Oxford Biomedical Centre Research Fund.

EL is funded by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (107630/Z/15/A).

LJ is funded by an NIHR Academic Clinical Lectureship.

MMa is funded by an Action Medical Research Fellowship (GN2779).

NM was funded by a DFG Research Fellowship (ME 5209/1-1).

KG is funded by NIHR (MIC-2016-014).

SBe is funded by a Wellcome Senior Research Fellowship (10104/Z/15/Z).

CAl is funded by the Open Targets consortium (OTAR026 project) and Wellcome Sanger core funding (WT206194).

MI is supported by the Wellcome Trust (215116/Z/18/Z) and thanks the PhD program FIRE and the Graduate School EURIP of Université Paris Cité for their financial support.

JPal is funded by NIH NHLBI R01 (HL151777).

Author contributions:

Conceived and directed the study: MH, SAT, BG

Acquired HDBR fetal samples: SL, DH

Generated scRNA-seq datasets: JE, ES, IG

Generated CITE-seq datasets: ES, NKW, NM, RH, MSV

Performed light-sheet microscopy: YG, MI

Performed immunofluorescence: DD, MA

Performed immunohistochemistry: RC

Performed RNAscope: KK, LT

Performed mouse and human embryo imaging: SJK

Generated and interpreted iPSC data: CAl, RVT, VL

Performed CITE-seq data analysis and interpretation: AR, MQL, NKW

Performed scRNA-seq data analysis and interpretation: IG, SW, AR, KG, IIR, DMP, KP, JPar, SvD

Interpreted the single cell data: OB, LJ, SBa, LG, MMa, KM, JPal, SBe, EL, AC, IR, MdB, ED, CS

Wrangled the single cell data: SW, MMa, MQL, AR, NKW, IG

Led web portal development: DH, DBL

Prepared the manuscript: BO, MMi

Drafted the manuscript: LJ, SW, RB, IG, MH

Designed the manuscript figures: JE, RB, CAd

Competing interests:

All authors declare no competing interests.

Data and materials availability:

All novel raw sequencing data from this study are made publicly available at ArrayExpress as FASTQs and count matrices as follows:

- i) Human embryonic liver and yolk sac 10x scRNA-seq (E-MTAB-10552)
- ii) Human embryonic yolk sac 10x scRNA-seq (E-MTAB-11673)
- iii) Human embryonic yolk sac Smart-seq2 scRNA-seq (E-MTAB-10888)
- iv) Human embryonic yolk sac CITE-seq (E-MTAB-11549)
- v) Human embryonic liver CITE-seq (E-MTAB-11618)
- vi) Human fetal liver CITE-seq (E-MTAB-11613)

Accessions for published data reused in this study are detailed comprehensively in **Table S6**.

798 Processed single cell datasets and supplementary tables are available for interactive
799 exploration and download as well as corresponding trained scVI and logistic regression
800 models via our interactive web portal (https://developmental.cellatlas.io/yolk-sac; password:
801 ys2022). Of note, data on portals are best used for rapid visualisation - for formal analysis it
802 is recommended to follow our GitHub code.
803
804 All code for reproducibility and trained logistic regression models of analysis is available at
805 https://github.com/haniffalab/FCA_yolkSac
806
807 **Supplementary Materials:**
808 Materials and Methods
809 Figs S1 to S9
810 Refs 59 to 76
811 Tables S1 to S32
812 Movies S1 to S3

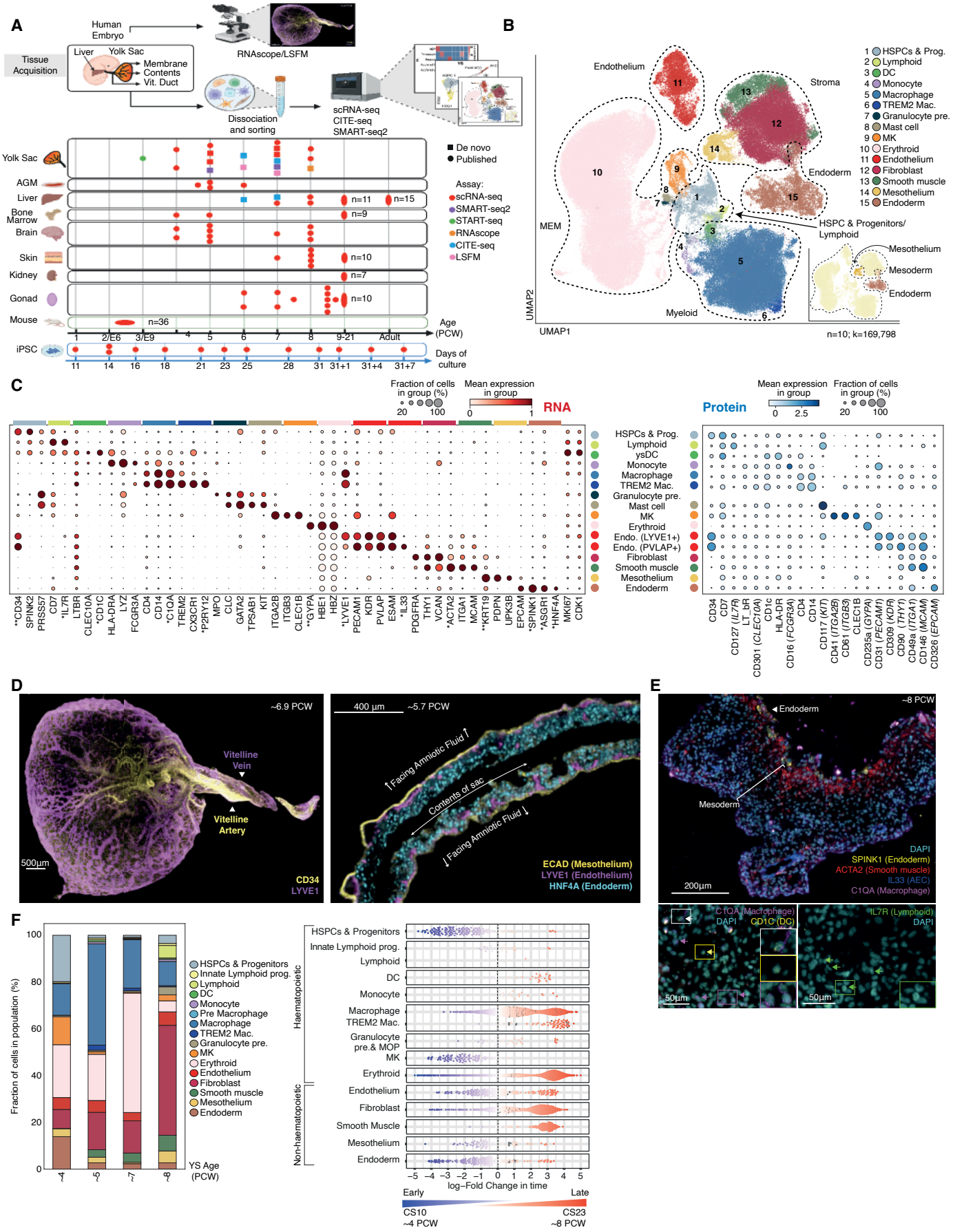

Fig. 1: A single cell atlas of the human yolk sac

(A) Summary of data included in analyses. Squares represent data generated for this study and circles represent published datasets: YS (*9*, *11*, *45*, *53*), AGM (*11*), liver (*9*), fetal BM (*54*), fetal brain (*44*), fetal skin (*45*), fetal kidney (*55*), fetal gonads (*46*), mouse (*56*), iPSC (*11*, *21*). Shape colour indicates the assay used to generate data (**Table S6**).

(**B)** UMAP visualisation of YS scRNA-seq data (n=10 independent biological replicates; k=169,798), with colours representing broad cell states: DC= dendritic cell, Mac= macrophage, MEM= megakaryocyte-erythroid-mast cell lineage, MK= megakaryocyte, pre.= precursor. Insert shows the same UMAP coloured by YS tissue layer (**Table S5)**.

(**C**) Left: Dot plot showing the expression level (by colour) and percent expression (by dot size) of broad cell state-defining genes in scRNA-seq data as shown in **b,** with data scaled to a maximum value=1. Right: Dot plot showing equivalent protein expression (by colour) and percent expression (by dot size) of broad cell states from n=2 biologically independent YS CITE-seq samples, with data scaled zero_centre=False. * indicates genes validated by RNAscope and ** indicates equivalent proteins validated by IHC/IF (**Table S4**).

(**D**) Light-sheet fluorescence microscopy images of YS. Left: $CD34^+$ and $LYVE1^+$ vascular structures (representative ~6.9PCW sample; scale bar=500µm; **Movie S1-2**). Right: Location of $LYVE1^+$ vascular structures within $HNF4A^+$ endoderm and adjacent to $ECAD^+$ mesothelium (representative ~5.7PCW sample; scale bar=400µm).

(**E**) RNAscope images of YS, showing a representative 8PCW sample. Top: demonstrating endoderm (*SPINK1*), smooth muscle (*ACTA2*), AEC (*IL33*) and macrophages (*C1QA*) (scale bar=200µm). Bottom: demonstrating DCs (*CD1C*; yellow arrow), macrophages (*C1QA*; magenta arrow), mac-DCs ($CD1C^+C1QA^+$; white arrow) and lymphoid cells (*IL7R*; green arrow) (scale bar=50µm).

(**F**) Left**:** Bar graph showing the proportional representation of broad cell states to YS scRNA-

seq data grouped by gestational age in PCW. Right: Milo beeswarm plot showing differential abundance of YS scRNA-seq neighbourhoods across time, where blue neighbourhoods are significantly enriched (SpatialFDR<0.1,logFC<0) early in gestation (CS10-11), red neighbourhoods are enriched later (CS22-23) (SpatialFDR<0.1,logFC>0) and colour intensity denotes degree of significance. Abbreviations: as per **(b)** and MOP= monocyte progenitor and pre= precursor (**Table S19**).

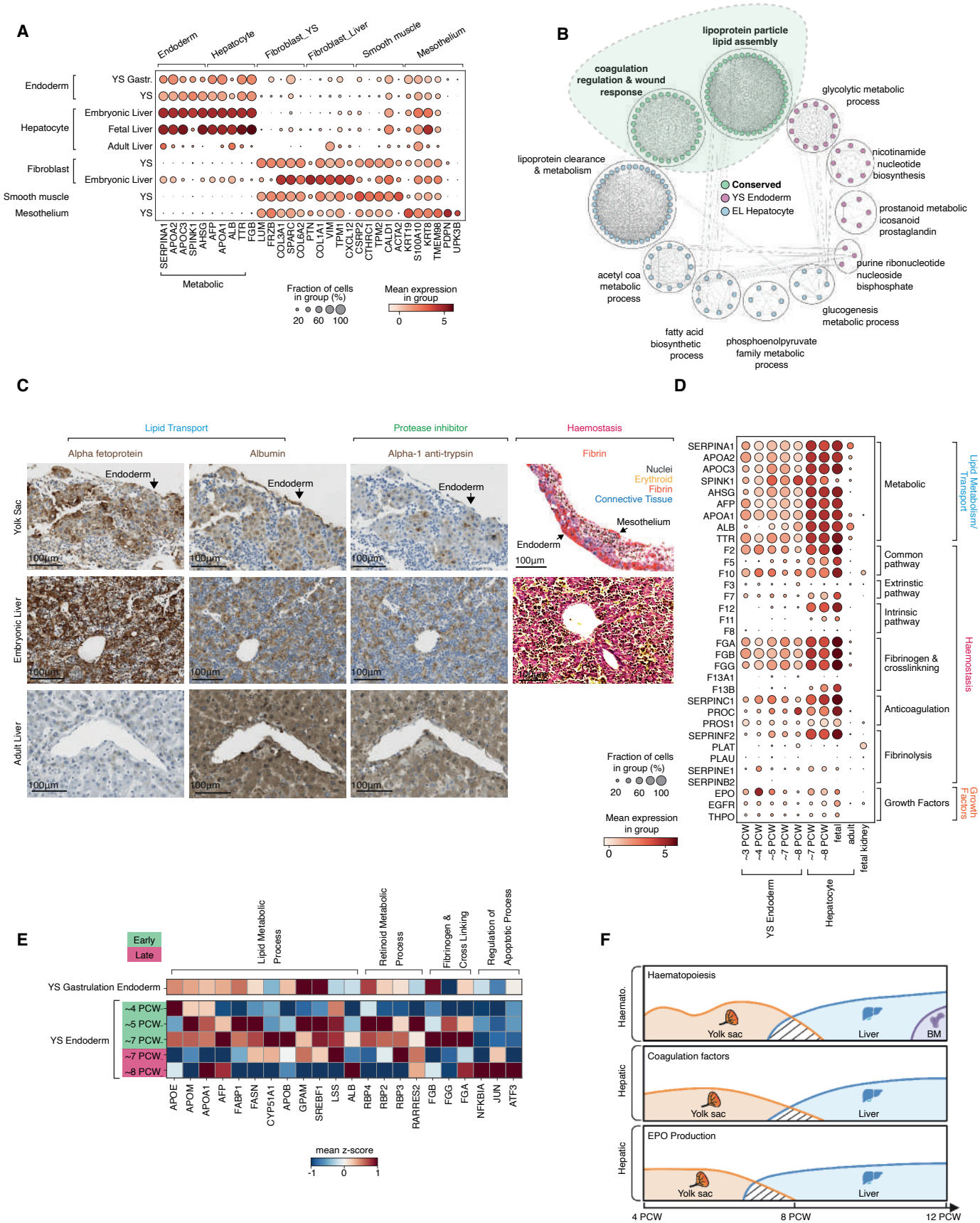

Fig. 2: Multi-organ functions of YS

(A) Dot plot showing the expression level (colour scale) and percent expression (dot size) of DEGs in YS (main and gastrulation (gastr.) data) stromal cell states, matched embryonic, fetal and adult liver stromal cell states and iPSC stromal cell states (*21*) (**Table S3, 20, 7**). All datasets independently scaled to max value=10 and then combined, except YS and matched EL scRNA-seq data which were scaled together. Genes involved in endoderm metabolic function are grouped.

(B) Flower plot of the significant pathways upregulated in YS endoderm (pink), embryonic liver (EL) hepatocytes (blue) and conserved across both tissues (green). Lines indicate connected nodes of expression (**Table S21**).

(C) Left and middle: IHC staining of alpha fetoprotein (*AFP*), albumin (*ALB*) and alpha-1 antitrypsin (*SERPINA1*) in 8PCW YS, 8PCW EL and healthy adult liver. Representative images from 1 of n=5 biological independent YS (4-8PCW), 1 of n=3 biologically independent ELs (7-8PCW) and 1 of n=3 biologically independent adult livers. Scale bar=100µm. Right: MSB-stained 8PCW EL (representative of n=3 biologically independent samples) and 4PCW YS (representative of n=3 biologically independent samples). Nuclei (grey), erythroid (yellow), fibrin (red), and connective tissue (blue). See 'Immunohistochemistry' section in Methods for details regarding pseudo-colouring shown. Scale bars=100um.

(D) Dot plot showing the expression level (colour scale) and percent expression (dot size) of haemostasis factors expressed by YS endoderm (main and gastrula data), embryonic, fetal and adult liver hepatocytes, and endoderm from fetal kidney (*55*). Grouped by pathway and role. All datasets independently scaled to max value=10 and then combined, except YS and matched EL scRNA-seq data which were scaled together.

876 **(E)** Matrix heatmap of Milo-generated DEGs in endoderm from YS scRNA-seq including
877 gastrulation datasets (**Table S19**). DEGs are grouped by function. Each dataset scaled zero
878 centre=False and YS endoderm standard_scale='var'.
879 **(F)** Schematic of the relative contribution of YS (orange), EL (blue), BM (purple) to
880 haematopoiesis, coagulation factors and EPO synthesis in the first trimester of human
881 development.
882
883
884
885

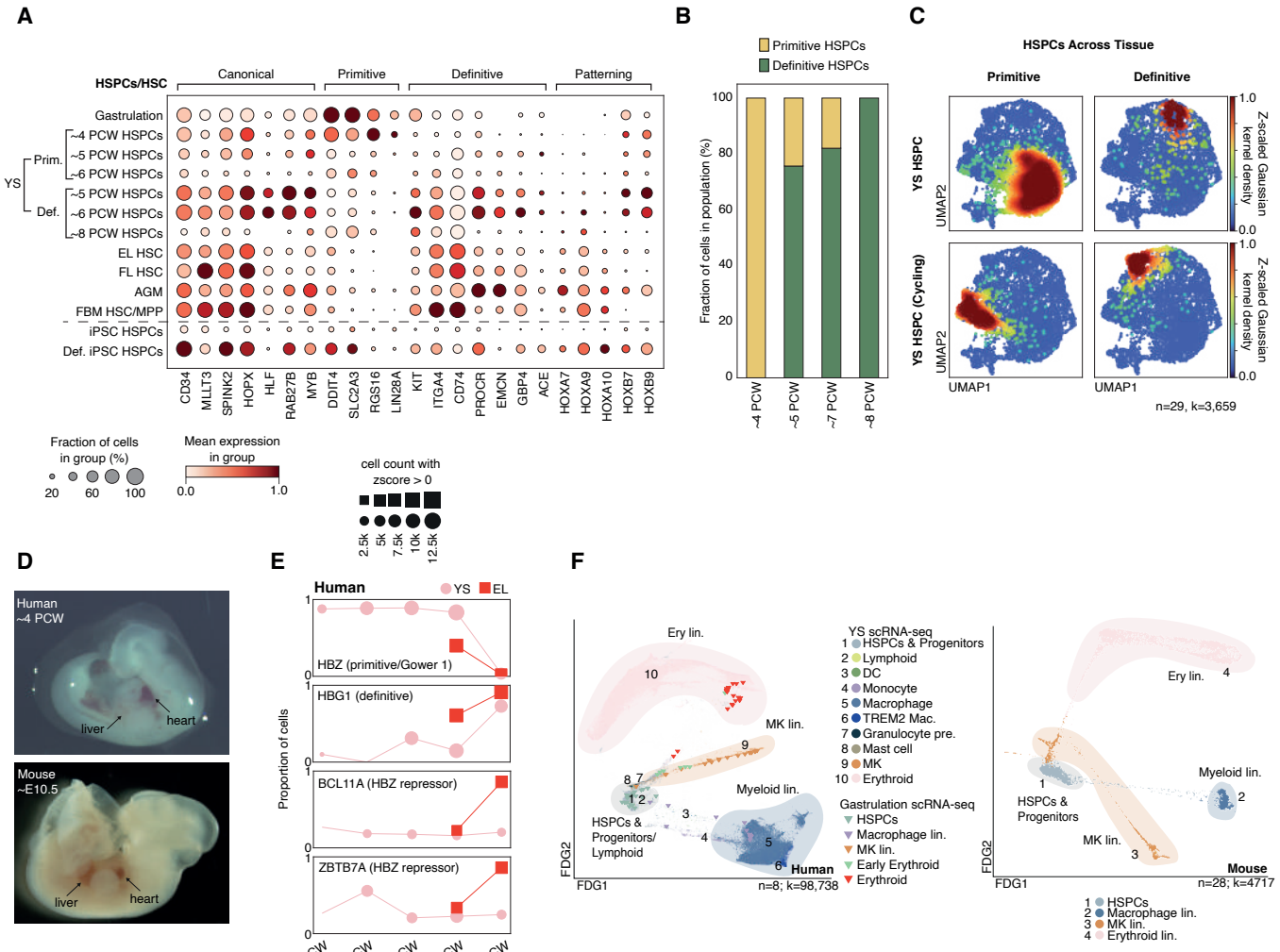

Fig. 3: Primitive versus definitive haematopoiesis in YS and liver

(A) Dot plot showing the mean variance scaled expression level (colour scale) and percent expression (dot size) of canonical, primitive and definitive, and patterning HSC markers expressed between YS HSPCs, including gastrulation (*57*), AGM HSPC (*58*), matched EL HSC, FL HSC (*9*), fetal BM HSC/MPP (*54*), iPSC-derived HSPC (*21*) and definitive iPSC-derived HSPC (*11*).

(B) Bar graph showing the proportional representation of primitive YS HSPCs to definitive YS HSPCs in the main YS scRNA-seq data (grouped by gestational age in PCW).

(C) Density plots showing the distribution of YS HSPC (top), cycling HSPC (bottom) with primitive (left) and definitive signatures (right) in the integrated UMAP landscape of HSPC/HSCs from YS including gastrulation (n=10, k=2,597), AGM (*58*) (n=1, k=28), matched embryonic liver (EL) (n=3, k=412), fetal liver (FL) (*9*), fetal bone marrow (*54*) (FBM) (n=9, k=92), iPSC-derived HSPC (*21*) and definitive iPSC-derived HSPC (*11*) scRNA-seq datasets. Colour of HSC/HSPC cells represents the z-scored kernel density estimation (KDE) score for each population (**Table S5**).

(D) Image of a ~4PCW/CS12 human embryo (top; representative from n=4 biologically independent samples) and ~E10.5/CS12 mouse embryo (bottom; n=1) with the heart and liver labelled.

(E) Line graphs showing the relative change in proportion of erythroid lineage cells (y-axis) enriched in expression of globins over gestational age, including *HBZ*, *HBG1* and HBZ-repressors *BCL11A* and *ZBTB7A*. Pink lines= human YS main and gastrulation scRNA-seq data. Red lines= matched scRNA-seq data. Globins are grouped by their role in primitive haematopoiesis, definitive haematopoiesis and repression.

(F) Force directed graph (FDG) visualisation of haematopoietic cell states in the YS scRNA-seq dataset (n=8, k=98,738; dots) integrated with human gastrulation (*57*) scRNA-seq dataset

(n=1, k=91; triangles) (left), and equivalent cell states found in the mouse gastrulation scRNA-seq dataset (*56*) (n=28, k=4,717; dots) (right). Colours represent cell states and clouds mark lineages

(G) Circular plots showing relative absorption probabilities of lineage-state transition between primitive HSPCs in the YS pre-AGM (CS10-11) (left) and primitive and definitive HSPCs in the YS post-AGM (>CS14) (right). Colour indicates the HSPC population density as a z-scored kernel density estimation (KDE) score and the position of HSPC population densities indicate respective lineage priming probability between Macrophage, lymphoid (NK and B lineage), erythroid and MK terminal states.

(H) Circular plots showing relative absorption probabilities of lineage-state transition between iPSC-derived HSPCs (left) and definitive iPSC-derived HSPCs (right). Colour indicates the HSPC population density as a z-scored kernel density estimation (KDE) score and the position of HSPC population densities indicate respective lineage priming probability between Macrophage, lymphoid (NK and B lineage), erythroid (any erythroid cell with individual Z-score of HBA1,HBA2,HBG1,HBG2,HBD > 0), embryonic erythroid (any erythroid cell with HBZ Z-score > 0) and MK terminal states.

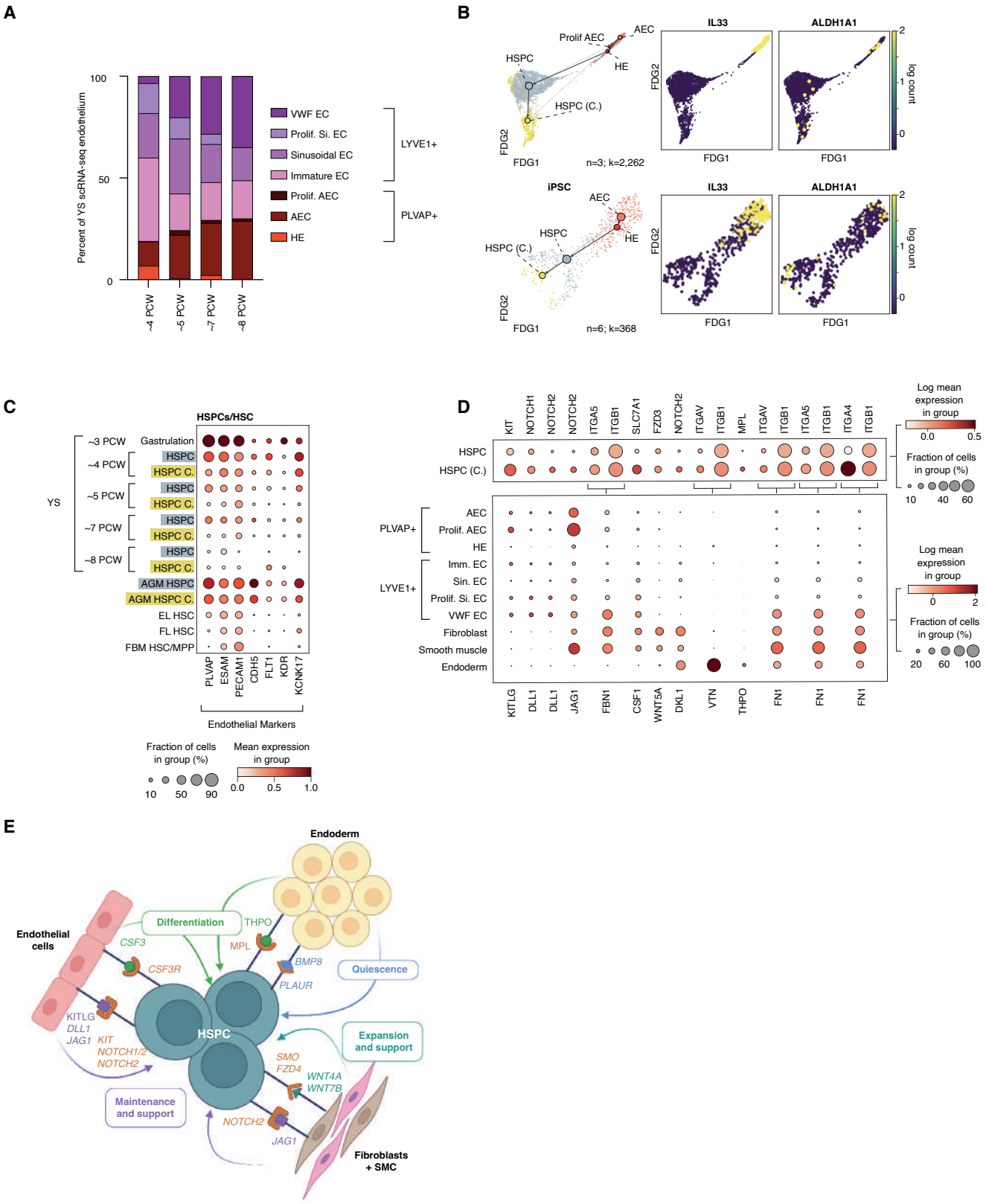

Fig. 4: The lifespan of YS HSPCs

(A) Barplot showing the relative proportion of YS endothelial cell subsets grouped by PCW.

(B) FDG overlaid with PAGA showing trajectory of HE transition to HSPC in YS scRNA-seq data (n=3; CS10, 11 and 14; k=2,262) (top) and iPSC-derived HSPC scRNA-seq data (n=7, k=437) (*21*) (bottom), with feature plots of key genes (*IL33*, *ALDH1A1*) involved in endothelial to hemogenic transition (**Table S5**).

(C) Dot plot showing the expression level (colour scale) and percent expression (dot size) of genes associated with endothelial cells in HSPCs from YS (including gastrulation) scRNA-seq, AGM (*11*), matched EL (embryonic liver), FL (fetal liver) (*9*) and fetal bone marrow BM (*32*).

(D) Heatmap showing relative mean expression z-scores of curated and statistically significant ($p<0.05$) CellphoneDB putative receptor ligand interactions between YS endothelial subsets vs HSPC across gestation. Growth factor, TGF beta and NOTCH ligand receptor -related gene interactions have been highlighted.

(E) Schematic of selected and statistically significant ($p<0.05$) CellphoneDB putative receptor ligand interactions between YS HSPC vs endoderm, fibroblasts (Fib), smooth muscle cells (SMC) and endothelial cells (EC) in scRNA-seq data. Receptors and ligands in italics significantly decrease CS17-12 (6-8PCW). See CellPhoneDB methods and **Table S23-24**.

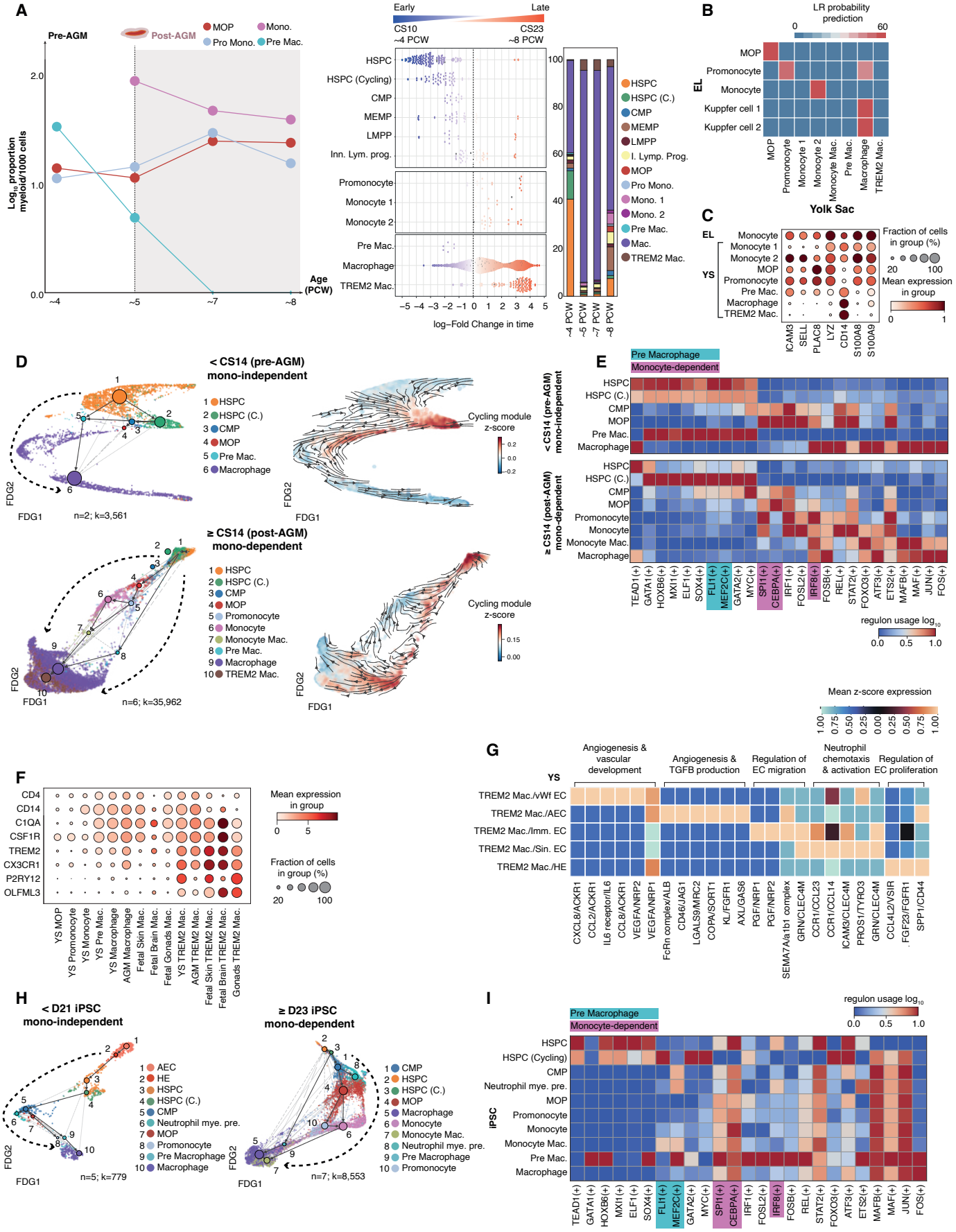

Fig. 5: Accelerated macrophage production in YS and iPSC culture

(A) Left: Line graph showing the proportion of monocyte progenitors (MOP), promonocytes (pro mono), monocytes (mono) and pre macrophage (pre mac) in yolk sac (YS) over gestational age. Dashed line indicates the stage before (left) and after (right) AGM involvement. Middle: Milo beeswarm plot showing differential abundance of YS scRNA-seq myeloid neighbourhoods across time, where blue neighbourhoods are significantly enriched (SpatialFDR<0.1,logFC<0) early in gestation (CS10), red neighbourhoods are enriched later (CS23) (SpatialFDR<0.1,logFC>0) and colour intensity denotes degree of significance. HSPC= haematopoietic stem and progenitor cells, CMP= common myeloid progenitors, MEMP= mast, erythroid and megakaryocyte progenitors, LMPP= lymphoid-primed multipotent progenitors, inn. lymph. prog.= innate lymphoid progenitor, MOP= monocyte progenitor, mono= monocyte, mac=macrophage Prog.= progenitor, pre=, precursor, MK= megakaryocyte (**Table S19, 4**). Right: Bar graph showing the proportional representation of myeloid cell states in YS scRNA-seq data grouped by gestational age in PCW.

(B) Median logistic regression class prediction probabilities for a model trained on YS scRNA-seq myeloid cell states (x-axis) projected onto equivalent cell states in matched embryonic liver (EL) scRNA-seq (y-axis)(**Table S12**).

(C) Dot plot showing the expression level (colour scale) and percent expression (dot size) of DEGs and known monocyte markers in YS myeloid cell states shown in **b** compared to matched EL monocytes (standard_scale='var') (**Table S28**).

(D) Left: FDG overlaid with PAGA showing monocyte-independent trajectory from YS scRNA-seq HSPC to macrophage prior to CS14 (pre-AGM; n=2; k=3,561; top) and monocyte-dependent trajectory to macrophage after CS14 (post-AGM; n=6; k=35,962; bottom). Left: Coloured by cell state. HSPC, haematopoietic stem cell; CMP, common myeloid progenitor; MOP, monocyte progenitor (**Table S5**). Right: CellRank state transition matrix inferred arrows

projected onto FDG indicate the trend of trajectory, and colour shows z-score enrichment in cycling module (GO:0007049) genes.

(E) Matrix heatmap showing the pySCENIC-derived regulons associated with the YS macrophage trajectories as shown in **d** for monocyte-independent, and monocyte-independent routes of differentiation.

(F) Dot plot showing the level (colour scale) and percent expression (dot size) of macrophage and microglia markers within the YS scRNA-seq monocyte and macrophage lineage in comparison to microglia, microglia-like and macrophage cell states in AGM (*11*), fetal skin (*45*), fetal gonads (*46*) and fetal brain (*44*). Fetal brain, skin and gonad cell states annotated in-house using LR output provided in **Fig. S8E.** Each dataset was scaled independently to a max_value = 10 then combined for plotting (**Table S18**).

nocyte-dependent trajectories in the YS scRNA-seq dataset (including gastrulation).

(G) Heatmap visualisation of CellphoneDB predicted interactions between TREM2^{+} microglia-like and endothelial cell states in the YS scRNA-seq (**Table S23**). Colour scale represents z-scored mean expression values of each gene pair.

(H) FDG overlaid with PAGA showing monocyte-independent trajectory from iPSC scRNA-seq (*21*) AEC to macrophage prior to D21 (n=5; k=779; left) and monocyte-dependent trajectory to macrophage after D21 (n=7; k=8,553; right). Arrows indicate the trend of trajectory. Coloured by cell state (**Table S7, S5).**

(I) Matrix heatmap showing the pySCENIC-derived regulons associated with the iPSC macrophage trajectories as shown in **i** for the iPSC scRNA-seq dataset (*21*).