

Article

Deciphering human macrophage development at single-cell resolution

<https://doi.org/10.1038/s41586-020-2316-7>

Received: 15 June 2019

Accepted: 11 March 2020

Published online: 20 May 2020

 Check for updates

Zhilei Bian^{1,2,3,17}, Yandong Gong^{4,17}, Tao Huang^{4,17}, Christopher Z. W. Lee^{5,6,17}, Lihong Bian^{7,17}, Zhijie Bai^{4,17}, Hui Shi^{4,17}, Yang Zeng⁸, Chen Liu⁴, Jian He⁴, Jie Zhou⁸, Xianlong Li⁴, Zongcheng Li⁹, Yanli Ni⁸, Chunyu Ma⁷, Lei Cui⁹, Rui Zhang^{10,11}, Jerry K. Y. Chan^{12,13,14}, Lai Guan Ng⁵, Yu Lan^{1,2}✉, Florent Ginhoux^{5,6,15,16}✉ & Bing Liu^{1,4,8}✉

Macrophages are the first cells of the nascent immune system to emerge during embryonic development. In mice, embryonic macrophages infiltrate developing organs, where they differentiate symbiotically into tissue-resident macrophages (TRMs)¹. However, our understanding of the origins and specialization of macrophages in human embryos is limited. Here we isolated CD45⁺ haematopoietic cells from human embryos at Carnegie stages 11 to 23 and subjected them to transcriptomic profiling by single-cell RNA sequencing, followed by functional characterization of a population of CD45⁺CD34⁺CD44⁺ yolk sac-derived myeloid-biased progenitors (YSMPs) by single-cell culture. We also mapped macrophage heterogeneity across multiple anatomical sites and identified diverse subsets, including various types of embryonic TRM (in the head, liver, lung and skin). We further traced the specification trajectories of TRMs from either yolk sac-derived primitive macrophages or YSMP-derived embryonic liver monocytes using both transcriptomic and developmental staging information, with a focus on microglia. Finally, we evaluated the molecular similarities between embryonic TRMs and their adult counterparts. Our data represent a comprehensive characterization of the spatiotemporal dynamics of early macrophage development during human embryogenesis, providing a reference for future studies of the development and function of human TRMs.

Although macrophages are best known for their immune functions, there is accumulating evidence that they have additional roles, including regulating the haematopoietic microenvironment, influencing metabolism, mediating tissue repair, and overseeing the maturation of embryonic tissue¹. These diverse roles are mirrored by their varied ontogenetic pathways, with some TRMs arising from adult bone marrow-derived haematopoietic stem cells (HSCs) and some instead being specified early in development from embryonic haematopoietic precursors².

Mammalian embryonic haematopoiesis is a complex process, involving multiple temporally overlapping programs^{3–5}. Fate-mapping studies in mice have shown that macrophages develop much earlier during embryonic haematopoiesis than do functional HSCs^{6–8}, with the earliest progenitors emerging in the yolk sac in at least two distinct waves: an early c-Myb-independent primitive wave arising at embryonic day

(E) 7.5, and a later c-Myb-dependent erythromyeloid progenitor (EMP) wave arising at E8.25^{9–12}. The progenitors in the first wave differentiate *in situ* into macrophages that migrate into the brain rudiment, being the major source of microglia⁷. The second wave progenitors, initially defined as EMPs, differentiate *in situ* into macrophages and also migrate into the fetal liver, giving rise to several lineages, including monocytes that will migrate into diverse developing tissues and begin differentiating into TRMs before birth^{13,14}.

Although the characteristics and development of embryonic macrophages in mice have been characterized in detail, they are not well understood in humans. We know that the first functional HSCs appear in the aorta–gonad–mesonephros region at Carnegie stage (CS) 14, and slightly later in the yolk sac at CS16, whereas the first HSCs in the liver are detected at CS17¹⁵. However, further progress has been hampered by the difficulty of obtaining early human embryos for study, and because

¹Key Laboratory for Regenerative Medicine of Ministry of Education, Institute of Hematology, School of Medicine, Jinan University, Guangzhou, China. ²Guangzhou Regenerative Medicine and Health-Guangdong Laboratory (GRMH-GDL), Guangzhou, China. ³Integrated Chinese and Western Medicine Postdoctoral Research Station, Jinan University, Guangzhou, China. ⁴State Key Laboratory of Proteomics, Academy of Military Medical Sciences, Academy of Military Sciences, Beijing, China. ⁵Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), BIOPOLIS, Singapore, Singapore. ⁶School of Biological Sciences, Nanyang Technological University, Singapore, Singapore. ⁷State Department of Gynecology, Fifth Medical Center of Chinese PLA General Hospital, Beijing, China. ⁸State Key Laboratory of Experimental Hematology, Fifth Medical Center of Chinese PLA General Hospital, Beijing, China. ⁹Laboratory of Hematologic Diseases, Beijing Pediatric Research Institute, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China. ¹⁰Beijing Key Laboratory of Pediatric Hematology Oncology, Beijing Children's Hospital, Capital Medical University, Beijing, China. ¹¹Hematology Oncology Center, National Center for Children's Health, Beijing Children's Hospital, Capital Medical University, Beijing, China. ¹²Department of Reproductive Medicine, KK Women's and Children's Hospital, Singapore, Singapore. ¹³Department of Obstetrics & Gynaecology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ¹⁴OBGYN-Academic Clinical Program, Duke-NUS, Duke-NUS Medical School, Singapore, Singapore. ¹⁵Shanghai Institute of Immunology, Shanghai JiaoTong University School of Medicine, Shanghai, China. ¹⁶Translational Immunology Institute, SingHealth Duke-NUS Academic Medical Centre, Singapore, Singapore. ¹⁷These authors contributed equally: Zhilei Bian, Yandong Gong, Tao Huang, Christopher Z. W. Lee, Lihong Bian, Zhijie Bai, Hui Shi.

✉e-mail: rainyblue_1999@126.com; florent_ginhoux@immunol.a-star.edu.sg; bingliu17@yahoo.com

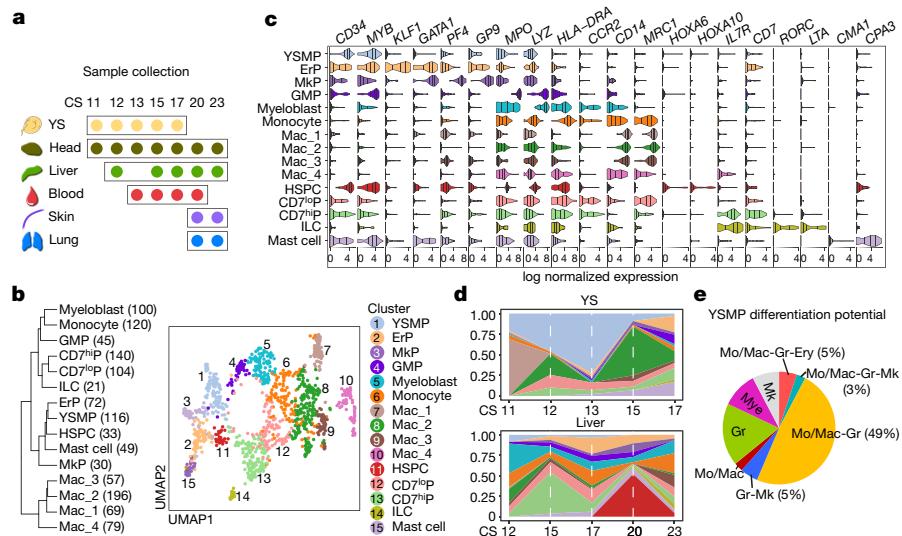


Fig. 1 | Transcriptomic landscape and functional characterization of CD45⁺ haematopoietic cells in tissues from human embryos. **a**, Location and Carnegie stage information for samples used for STRT-seq ($n=8$ biologically independent embryo samples). Details can be found in Supplementary Table 1. **b**, Uniform manifold approximation and projection (UMAP) visualization of all haematopoietic cell clusters identified (right), and hierarchical clustering using Euclidean distance with the number of cells in each cluster ($n=1,231$ cells total) shown in parentheses (left). Detailed cell information and DEGs can be found in Supplementary Table 2. **c**, Violin plot overview of expression of key lineage-associated genes by the haematopoietic clusters (number of cells in

each cluster shown in **b**), with twenty-fifth, fiftieth and seventy-fifth percentiles marked. **d**, Area charts showing changes in the proportions of the haematopoietic clusters (represented by the same colours as in **b**) between CS11 and CS17 for the yolk sac ($n=5$ biologically independent embryo samples and 238 cells), and CS12 and CS23 for the liver ($n=6$ biologically independent embryo samples and 354 cells). **e**, Pie chart showing lineage potential of YSMPs defined by single cell culture and subsequent flow cytometry analysis ($n=39$ clones generated from a CS13 yolk sac). Mo/mac, monocytes/macrophages (CD45⁺CD33⁺CD14⁺); Gr, granulocytes (CD45⁺CD33⁺CD66b⁺); Ery, erythrocytes (CD235a⁺); Mk, megakaryocytes (CD41a⁺).

conventional methods are unable to accurately analyse the extremely small number of early embryonic cells isolated.

In this study, we present findings from eight early human embryos obtained after abortion, from which we isolated a range of tissues before carrying out single-cell RNA sequencing (scRNA-seq) on the CD45⁺ haematopoietic cell populations. This enabled us to study the spatiotemporal distribution and dynamic process of human embryonic haematopoiesis and thereby to answer questions around the origins and specification of human macrophages.

scRNA-seq of human embryo haematopoiesis

To study human macrophage development, we used fluorescence-activated cell sorting (FACS) to isolate CD45⁺CD235a⁻ haematopoietic cells from various anatomical sites (yolk sac, head, liver, blood, skin and lung) of eight embryos that had been aborted at multiple Carnegie stages (from CS11 to CS23, equivalent to approximately days 24–56 of development) (Fig. 1a, Supplementary Fig. 1). We then used a modified scRNA-seq approach called single-cell tagged reverse transcription and sequencing (STRT-seq), as previously reported^{16–18}. Using high-precision quality-controlled transcriptomic data from 1,231 single cells (Fig. 1b, Extended Data Fig. 1, Supplementary Table 1), we annotated 15 clusters according to the expression of known feature genes (Fig. 1c, c, Supplementary Table 2).

A yolk sac-derived progenitor population and macrophages were among the first CD45⁺ haematopoietic cells to emerge, appearing in the yolk sac at CS11 (Fig. 1d, Extended Data Fig. 1f, g). Notably, these progenitors showed much weaker transcriptomic erythroid features than mouse EMPs¹⁹ (Extended Data Fig. 1k, l), suggestive of their myeloid-biased nature, and therefore were annotated as yolk sac-derived myeloid-biased progenitors (YSMPs). Both YSMPs and haematopoietic stem and progenitor cells (HSPCs) expressed high levels of CD34 and MYB, but HSPCs were seen only in the liver after CS17, and

specifically expressed HOX family transcription factors such as HOXA6 and HOXA10 (Fig. 1c, d, Extended Data Fig. 2, Supplementary Table 3). From CS12 onwards, we detected a population of granulocyte–monocyte progenitors (GMPs) in the liver that were characterized by high expression of MYB, MPO and LYZ. This contrasted with CD7^{lo} progenitors that arose at CS12, which expressed both lymphoid-related (CD7 and IL7R) and myeloid (MRC1 and CD14) genes, and with CD7^{hi} progenitors that were identified in the liver from CS15 onwards, which expressed higher levels of IL7R and CD7. Mast cells, which expressed high levels of CMA1 and CPA3, emerged in the yolk sac at CS13, and then in the liver, blood and skin from CS15 onwards. Innate lymphoid cells (ILCs), which are characterized by RORC expression, were detected in the yolk sac from as early as CS17 (Fig. 1c, d).

Notably, two populations (Mac_1 and Mac_4) clustered away from the other groups (Fig. 1b). Mac_1 cells were mainly found in the yolk sac at CS11 (Extended Data Fig. 1f, g) and expressed the yolk sac-related gene S100A1²⁰ as well as high levels of the endothelial marker CDH5²¹ (Extended Data Fig. 1j, Supplementary Table 2). By contrast, Mac_4 cells were predominantly located in the head, and this was the only population that expressed SALL1²² and gliomedin (GLDM)²³, indicating that these cells represent developing microglia (Extended Data Fig. 1j).

We next validated our STRT-seq results by high-throughput 10x Genomics analysis (Extended Data Fig. 3a–f, Supplementary Table 4). First, we re-analysed the yolk sac (YS) cells that had been sampled between CS11 and CS17 by STRT-seq, and found two macrophage populations: YS-Mac_1 cells were found mainly in the CS11 embryo, and YS-Mac_2 cells were found predominantly at CS15. 10x Genomics data from CS11 and CS15 yolk sacs similarly identified two macrophage populations separated stage-wise. We then projected the top ten differentially expressed genes (DEGs) identified by STRT-seq from both the YS-Mac_1 and YS-Mac_2 clusters on to the 10x Genomics data and found similar gene expression patterns. Finally, we identified CD34 and CD44 as putative markers of the YSMP population

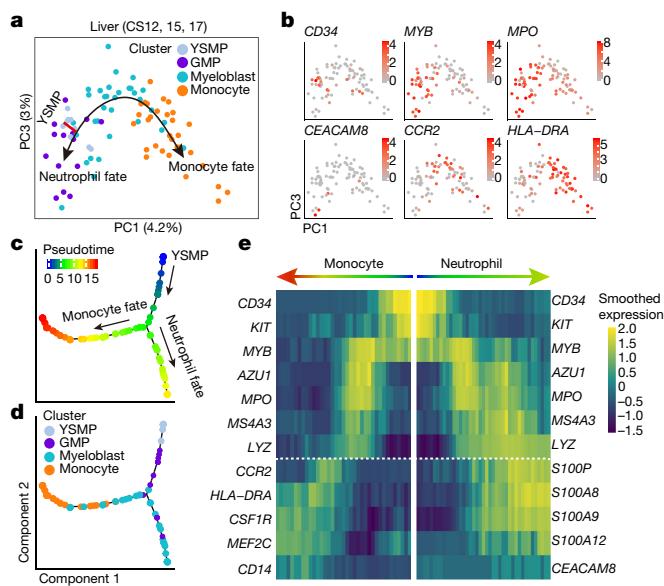


Fig. 2 | Developmental trajectory of YSMPs in human embryonic liver. **a**, PCA of YSMP, GMP, myeloblast and monocyte populations ($n=88$ cells) sampled from livers at CS12–CS17 ($n=4$ biologically independent embryo samples) suggests two distinct fates of YSMPs. **b**, Expression levels of the indicated genes projected onto PCA. **c, d**, Monocle prediction of YSMP developmental trajectory with pseudotime (**c**) and cluster information (**d**) mapped on. **e**, Heat map showing scaled expression of branching curated genes of monocyte and neutrophil fates ordered by pseudotime.

(Extended Data Fig. 3g, h) and sorted these cells to determine their functional identity.

Bulk cultures of yolk sac cells revealed that haematopoietic progenies were readily detected in the $CD45^+CD34^+CD44^+$ population (YSMPs), with mostly $CD33^+$ myeloid cells, but not in the $CD45^+CD34^-CD44^-$ population (Extended Data Fig. 4a–c). Haematopoietic clusters were found in 67 (36%) of 184 YSMP single-cell cultures (Extended Data Fig. 4d). Thirty-nine of these culture wells were then selected randomly and analysed using flow cytometry. Sixty-two per cent (24 wells) had multi-lineage potential, with most of them (22 wells) containing both monocytes/macrophages ($CD14^+$) and granulocytes ($CD66b^+$). Of note, 92% (36 wells) exhibited myeloid potential ($CD33^+$), among which 23 wells showed monocyte/macrophage differentiation, in contrast to only 5% (2 wells) with erythroid potential (Fig. 1e, Extended Data Fig. 4d, e). These data confirmed the myeloid-biased multi-lineage potential of human YSMPs, in line with their transcriptional features (Extended Data Fig. 1k, l).

YSMP development in embryonic liver

To investigate whether human YSMPs could also give rise to HSPC-independent monocytes in vivo, we studied the myeloid-related clusters in liver up to CS17, a stage before transcriptomically defined HSPCs were first detected. Principal component analysis (PCA) suggested that the YSMPs, which expressed *CD34* and *MYB*, were located between the GMP and myeloblast populations (Fig. 2a, b, Extended Data Fig. 5a, b). These cells gradually expressed neutrophilic genes such as *CEACAM8*, while also increasing their expression of the monocyte markers *CCR2* and *HLA-DRA* (Fig. 2b), indicating that YSMPs can differentiate along two distinct paths. Trajectory analysis using Monocle similarly revealed two distinct cell fates arising from YSMPs (Fig. 2c, d, Extended Data Fig. 5c). Based on the gene expression profiles, it appeared likely (although we were unable to capture mature neutrophils in our analysis) that YSMPs could give rise to both monocytes and neutrophils in vivo,

similar to our observations in the in vitro functional assay of YSMPs. Comparing the DEGs between the two branches revealed that the cells shared a common signature including expression of *MYB*, *MPO* and *MS4A3*, while the monocyte branch exclusively expressed genes such as *CCR2* and *CD14*, and the neutrophil branch began to express canonical neutrophil markers such as *S100A9* and *CEACAM8* (Fig. 2e, Extended Data Fig. 5d–g, Supplementary Table 5).

Two waves of yolk sac-derived embryonic TRMs

An important question for us to answer was whether the origin and specification of macrophages is similar in humans to those in mice, in which there are two yolk sac-derived waves: a monocyte-independent primitive wave in early yolk sac, and a later fetal liver monocyte-derived wave¹⁴.

To study this, we first re-clustered the identified macrophages and macrophage-related populations alone (Fig. 3a, Extended Data Fig. 6a, b), and annotated them on the basis of their unique expression characteristics, staging information and localization, supporting our annotation by comparison with a curated list of mouse TRM-specific genes²⁴ (Extended Data Fig. 6c–e). A closer look at the unique DEGs between these populations suggested that these cells may have already initiated their tissue residency programs, with the Liver_Mac population expressing *SPIC*, the Blood_Mac population expressing *CCL13*, and the Lung_Mac and Skin_Mac populations expressing their tissue-related *BMX* and *MMP1* genes, respectively (Fig. 3b, Supplementary Table 6). Hierarchical clustering revealed that Head_Mac3, Head_Mac4 and Liver_Mac were clustered away from the rest of the macrophage groups (Fig. 3c): we hypothesized that this may be because they were more mature TRMs and therefore selected them for further analysis.

The YS_Mac1 group corresponded to the Mac_1 group that we previously identified (Fig. 1c), which uniquely expressed the endothelial gene *CDHS*²¹ among macrophages (Fig. 3d) and notably also expressed the red blood cell-related *HBE1* gene. High expression of *HBE1* is a hallmark of yolk sac-derived nucleated primitive erythrocytes²⁵, which led us to question whether YS_Mac1/Mac_1 might be a related lineage belonging to the early yolk sac-derived primitive macrophage wave. At CS11, this population was found mainly in the yolk sac, although some were also present in the head (Extended Data Fig. 6c–e). Together, our analyses suggest that these cells are yolk sac-derived primitive macrophages, with some of them migrating early to the head as microglial precursors, as in mouse development⁷.

After establishing the distinct identity of the YS_Mac1 population, we investigated the developmental trajectory of early human yolk sac-derived embryonic macrophages, considering their dual origins. We analysed macrophage-associated populations between CS11 and CS17, before the appearance of HSPCs, which are likely to arise from a separate definitive lineage. A temporal assessment of the Carnegie staging of these populations revealed that YS_Mac1 cells appear earliest, at CS11, and contribute to the main macrophage populations first, whereas YSMPs gave rise to monocyte-derived macrophages, but only after CS17 (Fig. 3e). The expression of key lineage-defining genes showed a similar pattern (Fig. 3d), with YSMPs expressing higher levels of *MYB* and giving rise to *MPO*-expressing progenitors, consistent with studies on fetal monocytes in zebrafish¹¹. Likewise, the YS_Mac1 population had a strong macrophage identity from the start, expressing *CD163* and *MRC1*. Both YSMPs and YS_Mac1 cells expressed *CDHS*, consistent with their yolk sac-derived endothelial origins. Thus, our data suggest that early embryonic macrophage development in humans closely mirrors the processes seen in mice and zebrafish.

Human microglia origin and specification

In mice, microglia primarily arise from the first wave of yolk sac-derived primitive macrophages⁷. To study the origin and specification of

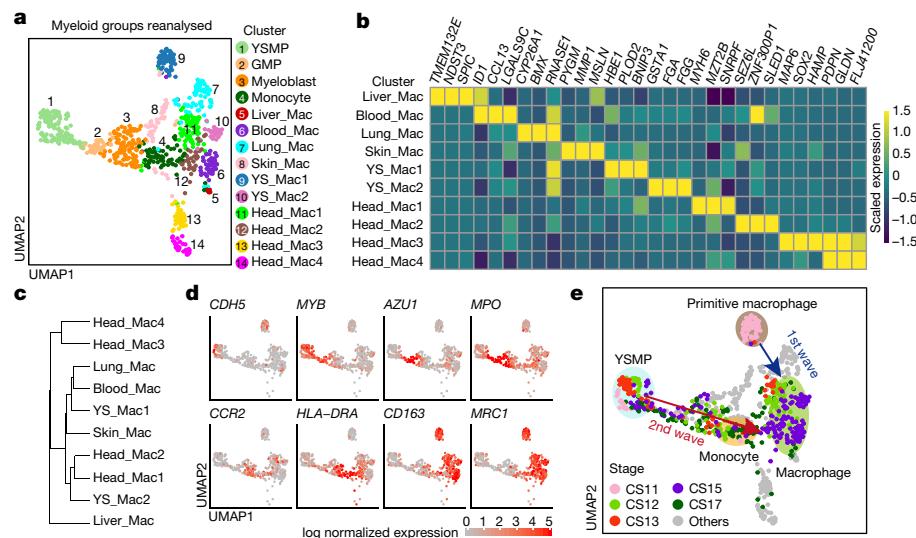


Fig. 3 | Two distinct waves of yolk sac-derived macrophages contribute to TRM populations in human embryos. **a**, UMAP visualization of myeloid-related cells ($n=8$ biologically independent embryo samples and 782 cells) with 14 re-clustered populations mapped on. **b**, Heat map showing scaled expression of the top three DEGs between macrophage populations ($n=450$ cells). DEGs were detected using FindAllMarkers function in Seurat (one-sided Wilcoxon rank-sum test, with P value adjusted for multiple testing using

Bonferroni correction), and the genes with fold change >1.5 and adjusted $P<0.05$ were selected (Supplementary Table 6). **c**, Hierarchical clustering based on Euclidean distance of macrophage populations ($n=450$ cells).

d, Expression of the indicated genes projected on to UMAP of myeloid-related clusters. **e**, UMAP visualization of myeloid-related clusters with cells from CS11 to CS17 ($n=6$ biologically independent embryo samples and 268 cells) mapped on, showing sequential appearance of macrophage populations.

microglia in human embryos, we focused on macrophages sampled from head tissues at CS11–CS23. Clustering analysis revealed that these macrophages could be divided into five groups (Head_Mac0–Head_Mac4), with features highly related to their developmental stage. The signature of Head_Mac0 overlapped completely with that of YS_Mac1; together with the observation that the yolk sac was the sole source of macrophages at this time-point, this suggests that YS_Mac1 gives rise to the macrophages found in the head at CS11 (Fig. 4a). A small population of monocytes derived from YSMPs was also seen at CS17 in the head. However, the disparity between their number and those of the other Head_Mac populations, as well as their late appearance, together suggested that they are likely to make only a minor contribution (Extended Data Fig. 7a, b). The sequential emergence of the five head macrophage populations indicated that these populations were part of a continuum of developing cells, and their expression of classical microglia-associated genes including *CX3CR1*, *SALL1* and *SPP1* further supported the idea that these cells were moving towards a microglial fate^{22,26} (Fig. 4b).

Next, we studied the dynamics of microglial specification by generating a profile of gene expression changes across the five Head_Mac clusters based on their sequential real-time emergence between CS11 and CS23 (Fig. 4c, Supplementary Table 7). The changes largely followed five patterns. Most genes in patterns 1 and 2 were downregulated and some of them, such as *CD163* and *ID1*, were associated with immune activity. We also witnessed increased expression of tissue development and neurodevelopmental genes such as *IGF1*²⁶ and *TMSB4X*²⁷. A survey of the transcription factor landscape revealed a similar trend (Fig. 4d), with head macrophages losing expression of inflammatory transcription factors such as *IRF7* and *STAT1*, and gaining microglial identity with expression of *BHLHE41*, *JUN*, *FOSB*, *NR4A1* and *SALL1*²²; these changes mirror the pattern observed during mouse microglial development.

Finally, we integrated our embryonic data with publicly available data for adult TRMs from the head, liver and lungs, along with our own data for skin sampled from children aged eight and ten. The results confirmed that specification had already occurred in the microglial population in embryos, with the embryonic and adult microglia cluster together (Extended Data Fig. 8, Supplementary Table 8). Likewise,

specification towards liver TRM fate had begun, although to a lesser extent than for head macrophages (Extended Data Fig. 7d–g, Supplementary Table 9). This is in contrast with skin TRMs, where there was only minor expression of *CD207* at CS23 (Extended Data Fig. 7h) and slight overlap between the embryonic and paediatric samples in the integrated dataset (Extended Data Fig. 8a, b). Similarly, no commitment had yet occurred in the embryonic lung (Extended Data Figs. 7c, 8a, b, Supplementary Table 8).

Discussion

Much of our current knowledge of embryonic haematopoiesis is based on findings in animal models such as mice or zebrafish. This study paves the way for a wide range of explorations and analyses that were previously difficult to approach owing to the ambiguity surrounding human yolk sac-derived macrophages and their progenitors. Although it is widely acknowledged that mammalian haematopoiesis is highly conserved³, the characterization of these cells in humans has so far been restricted to either microscopic observations^{3,28} or explant experiments²⁹. By leveraging the maturation of single-cell sequencing technology and bioinformatics, our analyses shed light on this issue in an unbiased and unsupervised manner, while maintaining tissue site and temporal information. Although it is still difficult to determine with certainty the ontogeny of the various human TRM subsets without the use of fate-mapping tools typically used in mouse models, we have identified two distinct HSC-independent waves of macrophages in humans that correspond to those seen in mice. This is especially important in the clinical context, as macrophages are essential regulators of tissue development and homeostasis³⁰, and understanding their functions and developmental pathways is key to the diagnosis and treatment of pathologies caused by their dysregulation. The contribution of these HSC-independent waves must be considered when characterizing the subsets of macrophages found in disease, as numerous studies have established that HSC-independent macrophages maintain a distinct transcriptomic and epigenetic identity from their HSC-derived counterparts³¹. Specifically, the human YSMPs that we have transcriptomically and functionally characterized here might correspond to mouse

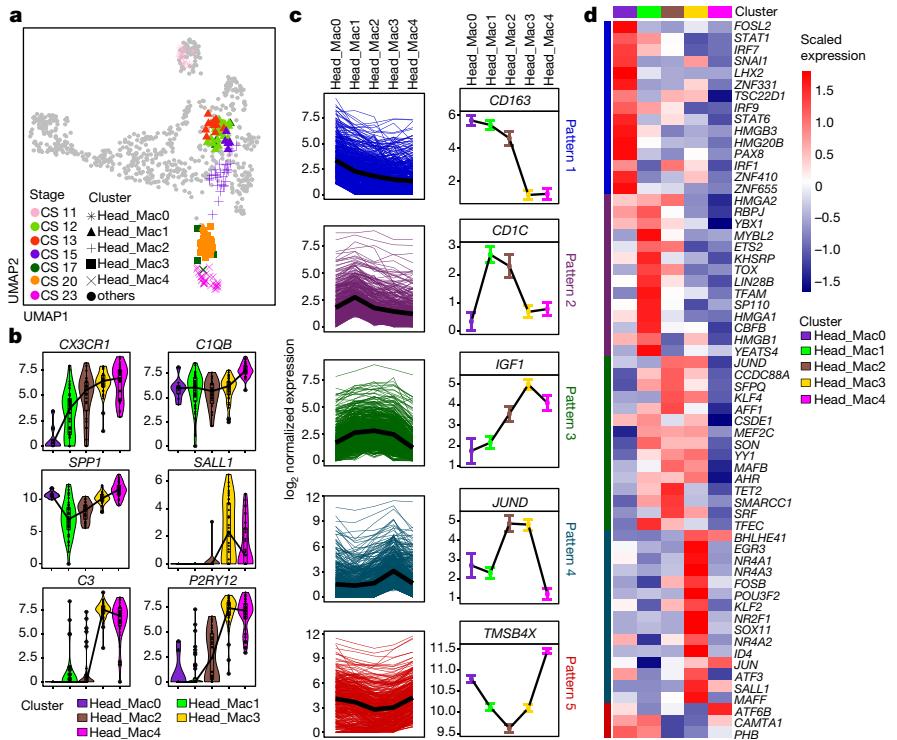


Fig. 4 | Origin and specification of microglia in human embryos. **a**, UMAP visualization of myeloid-related cells with macrophages in human embryonic head ($n=8$ biologically independent embryo samples and 155 cells) mapped on. Stage and cluster information indicated by colours and symbols, respectively. **b**, Violin plots showing changes in expression of microglia-related genes between head macrophage clusters. For box plot within each violin plot, centre black lines indicate median values, boxes range from the 25th to 75th percentiles, and whiskers correspond to $1.5 \times$ interquartile range (IQR). **c**, Five

main patterns revealed by DEGs between head macrophage clusters. Left, expression levels of all pattern genes (coloured lines) and the average expression of each pattern (black line). Complete gene list can be found in Supplementary Table 7. Right, kinetics of expression of a representative gene for each pattern (mean \pm s.e.m.). **d**, Heat map showing scaled expression of transcription factors within the pattern genes. Colours to the left indicate the patterns as in **c**.

EMPs, which exhibit similar spatiotemporal dynamics and haematopoietic progenitor characteristics¹³. However, erythroid potential is readily observed in cultures of mouse EMPs¹³, consistent with their prominent erythroid properties transcriptomically¹⁹ (Extended Data Fig. 1k, l). These observations warrant further investigation towards their in-depth comparison and the identification of related governing mechanisms.

The embryonic macrophage signature identified in our study also presents an interesting opportunity to re-evaluate our commonly used TRM models. Conventionally, for human studies, macrophages have been generated from blood monocytes via in vitro culture with CSF-1. Comparison of these in vitro monocyte-derived macrophages with the bona fide yolk sac-derived macrophages that initially seed the embryonic tissues might reveal crucial similarities and differences between these cells and allow better experimental design and interpretation of results. Our study also supports the case for a greater understanding of induced pluripotent stem (iPS) cell-derived macrophages and the various TRM systems that have been developed based on these cells³². Although a direct comparison between mouse primitive macrophages and iPS cell-derived macrophages has already proven that they are similar to each other and yet distinct from HSC-derived macrophages³³, the same comparison has not, to our knowledge, been carried out for their human counterparts. Our study will contribute to an understanding of the ontogeny of these iPS cell-derived macrophages.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information,

acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2316-7>.

- Wynn, T. A., Chawla, A. & Pollard, J. W. Macrophage biology in development, homeostasis and disease. *Nature* **496**, 445–455 (2013).
- Ginhoux, F. & Guilliams, M. Tissue-resident macrophage ontogeny and homeostasis. *Immunity* **44**, 439–449 (2016).
- Palis, J. et al. Spatial and temporal emergence of high proliferative potential hematopoietic precursors during murine embryogenesis. *Proc. Natl. Acad. Sci. USA* **98**, 4528–4533 (2001).
- Dzierzak, E. & Bigas, A. Blood development: hematopoietic stem cell dependence and independence. *Cell Stem Cell* **22**, 639–651 (2018).
- Zhou, F. et al. Tracing haematopoietic stem cell formation at single-cell resolution. *Nature* **533**, 487–492 (2016).
- Samokhvalov, I. M., Samokhvalova, N. I. & Nishikawa, S. Cell tracing shows the contribution of the yolk sac to adult haematopoiesis. *Nature* **446**, 1056–1061 (2007).
- Ginhoux, F. et al. Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science* **330**, 841–845 (2010).
- Yona, S. et al. Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis. *Immunity* **38**, 79–91 (2013).
- Hoeffel, G. & Ginhoux, F. Ontogeny of tissue-resident macrophages. *Front. Immunol.* **6**, 486 (2015).
- Frame, J. M., McGrath, K. E. & Palis, J. Erythro-myeloid progenitors: “definitive” hematopoiesis in the conceptus prior to the emergence of hematopoietic stem cells. *Blood Cells Mol. Dis.* **51**, 220–225 (2013).
- Bertrand, J. Y. et al. Definitive hematopoiesis initiates through a committed erythromyeloid progenitor in the zebrafish embryo. *Development* **134**, 4147–4156 (2007).
- Ivanovs, A. et al. Human haematopoietic stem cell development: from the embryo to the dish. *Development* **144**, 2323–2337 (2017).
- McGrath, K. E. et al. Distinct sources of hematopoietic progenitors emerge before HSCs and provide functional blood cells in the mammalian embryo. *Cell Rep.* **11**, 1892–1904 (2015).
- Hoeffel, G. et al. C-Myb⁺ erythro-myeloid progenitor-derived fetal monocytes give rise to adult tissue-resident macrophages. *Immunity* **42**, 665–678 (2015).
- Ivanovs, A. et al. Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. *J. Exp. Med.* **208**, 2417–2427 (2011).

Article

16. Ivanovs, A., Rybtsov, S., Anderson, R. A., Turner, M. L. & Medvinsky, A. Postmenstrual gestational age should be used with care in studies of early human hematopoietic development. *Blood* **121**, 3051–3052 (2013).
17. Zeng, Y. et al. Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing. *Cell Res.* **29**, 881–894 (2019).
18. Zeng, Y. et al. Single-cell RNA sequencing resolves spatiotemporal development of pre-thymic lymphoid progenitors and thymus organogenesis in human embryos. *Immunity* **51**, 930–948.e936 (2019).
19. Dong, J. et al. Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hybrid state during mouse organogenesis. *Genome Biol.* **19**, 31 (2018).
20. Kiewitz, R. et al. S100A1, a new marker for acute myocardial ischemia. *Biochem. Biophys. Res. Commun.* **274**, 865–871 (2000).
21. Costa, G. et al. SOX7 regulates the expression of VE-cadherin in the haemogenic endothelium at the onset of haematopoietic development. *Development* **139**, 1587–1598 (2012).
22. Buttigereit, A. et al. Sall1 is a transcriptional regulator defining microglia identity and function. *Nat. Immunol.* **17**, 1397–1406 (2016).
23. Gosselin, D. et al. An environment-dependent transcriptional network specifies human microglia identity. *Science* **356**, eaal3222 (2017).
24. Mass, E. et al. Specification of tissue-resident macrophages during organogenesis. *Science* **353**, aaf4238 (2016).
25. Peschle, C. et al. Haemoglobin switching in human embryos: asynchrony of $\zeta \rightarrow \alpha$ and $\epsilon \rightarrow \gamma$ -globin switches in primitive and definite erythropoietic lineage. *Nature* **313**, 235–238 (1985).
26. Włodarczyk, A. et al. A novel microglial subset plays a key role in myelinogenesis in developing brain. *EMBO J.* **36**, 3292–3308 (2017).
27. Wirsching, H. G. et al. Thymosin β 4 induces folding of the developing optic tectum in the chicken (*Gallus domesticus*). *J. Comp. Neurol.* **520**, 1650–1662 (2012).
28. Enzan, H. Electron microscopic studies of macrophages in early human yolk sacs. *Acta Pathol. Jpn.* **36**, 49–64 (1986).
29. Tavian, M., Robin, C., Coulombel, L. & Péault, B. The human embryo, but not its yolk sac, generates lympho-myeloid stem cells: mapping multipotent hematopoietic cell fate in intraembryonic mesoderm. *Immunity* **15**, 487–495 (2001).
30. Gordon, S. & Plüddemann, A. Tissue macrophages: heterogeneity and functions. *BMC Biol.* **15**, 53 (2017).
31. Lavin, Y. et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312–1326 (2014).
32. Lee, C. Z. W., Kozaki, T. & Ginhoux, F. Studying tissue macrophages in vitro: are iPSC-derived cells the answer? *Nat. Rev. Immunol.* **18**, 716–725 (2018).
33. Takata, K. et al. Induced-pluripotent-stem-cell-derived primitive macrophages provide a platform for modeling tissue-resident macrophage differentiation and function. *Immunity* **47**, 183–198.e186 (2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Human samples and quality control

The embryos were obtained from pregnant women undergoing medical abortions at The Fifth Medical Center of the PLA General Hospital (Beijing, China) the women had provided informed consent. The integrity and morphology of the embryos were evaluated, the somite pairs were counted, and the crown–rump lengths were measured under the microscope to define developmental stage³⁴. Paediatric skin samples from children aged eight and ten years were obtained from tissue discarded after elective circumcision at Beijing Children's Hospital, Capital Medical University with parental consent. All protocols were approved by the institutional review boards (approval number: ky-2017-3-5 and IEC-C-008-A08-2018-75) and in accordance with the regulations of the Declaration of Helsinki. All protocols were compliant with the Interim Measures for the Administration of Human Genetic Resources, administered by the Ministry of Science and Technology of China.

Preparation of cell suspensions and FACS

Embryonic tissues (yolk sac, head, lung, skin, and liver) were carefully dissected under the microscope, with the exception of blood, which was collected using a suction pipe directly from the heart. After being washed in PBS three times to remove any residual blood contamination, the yolk sac, head, lung and skin were transferred to pre-warmed digestion medium containing 0.1 g/ml collagenase I (Sigma, C2674, preheated to 37 °C) in RPMI1640 medium (Gibco, 11875093). The samples were enzymatically digested at 37 °C in a humidified incubator for 20–30 min, with the samples being shaken periodically every 5 min until they were digested into a single-cell suspension. The liver was first cut into small pieces using scalpels, and then mechanically dissociated using syringes into a single-cell suspension, before removal of erythrocytes with lysis buffer (BD). The cells were then filtered through a 70-µm cell strainer after the enzymes were neutralized using serum. Child skin samples were first incubated in RPMI 1640 medium with 1.2 U/ml dispase II (Roche) at 4 °C for 8 h before separation of the epidermal layer. After separation, the epidermal layer was cut into pieces and digested with 0.25% trypsin-EDTA (Gibco) containing 0.25 mg/ml DNase I (Sigma DN25) for 30 min at 37 °C in a humidified incubator. After that, the epidermal layer was passed through a 70-µm cell strainer by grinding, and then washed with PBS.

The following antibodies were used to label the embryonic cells: CD45 (BV421, 563879 BD, lot 9066960), CD235a (APC-Cy7, 349116 BioLegend, lot 7355682) and 7-amino-actinomycin D (7-AAD) (PerCP-Cy5.5, 00699350 eBioscience, lot 1910559). Langerhans cells were isolated using CD45 (BV421, 563879 BD, lot 9066960), 7-AAD (PerCP-Cy5.5, 00699350 eBioscience, lot 1910559), CD207 antibodies (PE, 564727 BD, lot 8135683) and CD1a antibodies (APC, 559775 BioLegend, lot 8164562). Cells were sorted using an Aria 2 Flow Cytometer (BD Bioscience). Data were analysed using BD FACSDIVA V8.0.1 and Flowjo (V10).

Haematopoietic progenitor culture in vitro

MS5 stromal cells²⁹ were seeded into 48-well or 96-well flat-bottom plates with α-MEM (Gibco, 12561-056) supplemented with 20% fetal bovine serum (HyClone, SH30070.03) at a density of 20,000 (48-well) or 10,000 cells (96-well) per well about 24 h before use. One hour before co-culture, stromal medium was replaced with serum-free StemPro 34 (Gibco, 1063901) supplemented with 50 ng/ml hSCF (PeproTech, 300-07-2), 50 ng/ml hFLT3 ligand (PeproTech, 300-19-2), 10 ng/ml hIL-3 (PeproTech, 200-03-2), 10 ng/ml hIL-6 (PeproTech, 200-06-5), 5 ng/ml hIL-11 (PeproTech, 200-11-10), 25 ng/ml hTPO (PeproTech, 300-18-10), 20 ng/ml hGM-CSF (PeproTech, 300-03-5), 3 U/ml EPO (Pepro Tech, 100-64-10), 1% 2-mercaptoethanol (Gibco, 21985-023), 1% L-glutamine (Gibco, 21051024) and 1% penicillin/streptomycin (Gibco, 15140-122).

Candidate cells were sorted into wells and cultured at 37 °C for 10 or 14 days. The morphologies of haematopoietic clusters were imaged using a Leica camera (A05C872000). At the end of the cultures, all cells including MS5 stromal cells in the selected positive wells were collected and FACS analysis was used for identification of myeloid cells (Mye, CD45⁺CD33⁺), monocytes/macrophages (Mo/Mac, CD45⁺CD33⁺CD14⁺), granulocytes (Gr, CD45⁺CD33⁺CD66b⁺), erythrocytes (Ery, CD235a⁺) and megakaryocytes (Mk, CD41a⁺).

For bulk cultures, YSMP (CD45⁺CD34⁺CD44⁺) and negative control (CD45⁺CD34⁻CD44⁻) populations were sorted and co-cultured with MS5 for 14 days (100 cells per well). The assays were performed in $n=3$ biologically independent experiments (using one CS11 and two CS12 yolk sacs, 21 wells in total). For the single cell cultures, 184 single YSMPs from a CS13 yolk sac were sorted and cultured in individual wells. Wells were observed under a light microscope and those with haematopoietic clusters (more than 50 round haematopoietic-like cells) at day 10 were counted as positive wells, which were stochastically chosen for further analyses with FACS to determine the lineage differentiation potential. For each well, 20,000 living cells were recorded, which took around one-third of the total number. The numbers of haematopoietic cells (including CD45⁺, CD235a⁺ and CD41⁺ cells) generated in each well thus ranged from 200 to 4,500 according to the calculation. Lineage potential was ascertained when the number of cells that expressed the given lineage marker (described above) was more than 20.

The following antibodies were used for staining and sorting: anti-CD45 (BV421, 563879 BD, lot 9066960), anti-CD34 (PE, 550761 BD, lot 7129824), anti-CD44 (BV605, 562991 BD, lot 7103609). The following antibodies were used for the identification of haematopoietic clusters and lineages: anti-CD235a (APC-Cy7, 349116 BioLegend, lot B289027), CD235a (Pacific Blue, 306611 BioLegend, lot B224563), anti-CD41a (APC, 17-0419-42 eBioscience, lot 2073742), anti-CD45 (FITC, 11-0459-42 eBioscience, lot 4310016), anti-CD33 (PE, 555450 BD, lot 8074660), anti-CD33 (APC-Cy7, 366614 BioLegend, lot B252646), anti-CD14 (BV786, 563698 BD, lot 8351911) and anti-CD66b (PE, 561650 BD, lot 7264511).

Single-cell RNA-seq library preparation and sequencing

Sequencing libraries were constructed following a modified single-cell tagged reverse transcription (STRT) protocol as previously reported^{35–37}. The cells were first judged by morphology under a microscope to assess their condition, with cells in good condition being picked by mouth pipette and directly placed into lysis buffer. The reverse transcription reaction was performed using a sample-specific 25-nucleotide (nt) oligo dT primer containing an 8-nt barcode (TCAGACGTGTGCTCTCCGATCT-XXXXXX-YYYYYYYY-T25, where X represents the sample-specific barcode, and D stands for the unique molecular identifier (UMI)). After reverse transcription and second-strand cDNA synthesis, the cDNAs were amplified by 16 cycles of PCR. The barcoded DNAs were then pooled together and purified using Agencount AMPure XP beads. Biotinylated pre-indexed primers were used to further amplify the PCR product by an additional four cycles of PCR to introduce biotin tags to the 3' ends of the amplified cDNAs. Approximately 300 ng cDNA was sonicated into 300-bp fragments using the Covaris S2 system and enriched with Dynabeads MyOneTM Streptavidin C1 beads. Libraries were constructed using a Kapa Hyper Prep Kit (Kapa Biosystems) and were then submitted to 150-bp paired-end sequencing on an Illumina Hiseq X Ten platform (Novogene).

For 10x Genomics single-cell RNA sequencing, we first used FACS to isolate living cells (7AAD⁻ for CS11 and CS15 yolk sac), and then implemented the Chromium Single Cell 3' v2 libraries, under the guidance of the official instruction manual (<https://support.10xgenomics.com/single-cell-gene-expression/library-prep/doc/technical-note-assay-scheme-and-configuration-of-chromium-single-cell-3-v2-libraries>).

Pre-processing scRNA-seq data

For scRNA-seq data, the raw reads from each cell were first split by the specific barcode sequence attached in Read 2. The template switch oligo (TSO) sequence and polyA tail sequence were trimmed for the corresponding Read 1 after UMI information was aligned to it. Subsequently, reads with adaptor contaminants or low-quality bases ($n > 10\%$) were discarded. Next, the stripped Read1 sequences were aligned to the hg19 human transcriptome (UCSC) using Hisat2 (version 2.1.0)³⁸. Uniquely mapped reads were counted using the HTSeq package³⁹ and grouped by the cell-specific barcodes. Duplicated transcripts were removed based on the UMI information for each gene. Finally, for each individual cell, the copy number of transcripts of a given gene was the number of distinct UMIs for that gene.

To filter low quality cells, count values for each cell were first grouped into an expression matrix; only cells with more than 2,000 genes and 10,000 transcripts detected and below 50% transcripts mapped on the External RNA Controls Consortium (ERCC) reference were retained. Also, cells with too many raw reads ($> 1,000,000$) and genes ($> 10,000$) were excluded because these cells might not be real single cells. We obtained a total of 1,461 cells after sequencing, of which only 119 cells (8.15% of total) were excluded for failing the quality control threshold (Supplementary Table 1), leaving us with a median UMI of 132,783 and gene number of 4,860. The distribution of the UMI can be seen in Extended Data Fig. 1a, with the majority of the cells that showed more than 50% of UMIs mapping to the ERCCs falling below the UMI threshold of 1×10^4 , while cells above 1×10^4 showed robust expression of housekeeping genes (Extended Data Fig. 1b).

Regarding the droplet-based scRNA-seq data from 10x Genomics, the dataset was aligned and quantified using the CellRanger software package (version 2.1.0) with default parameters, giving a total of 11,944 cells from the CS11 and CS15 yolk sacs. For 10x Genomics, we adopted a more relaxed quality control standard because the sequencing depth is lower than for STRT-seq. Cells with $> 1,000$ genes expressed were retained, resulting in 25 cells excluded in total. The droplet-based scRNA-seq was expected to generate cell doublets at a low frequency, which could be incorrectly interpreted as novel cell types, and so to avoid the effects of doublets, 480 cells identified using Doubletdetection (<https://doi.org/10.5281/zenodo.2678042>) were removed from our data. We finally removed 1,874 cells with matured erythrocyte characteristics, leaving a total of 9,565 cells for our final analysis.

Cell-type detection and dimensionality reduction

Downstream analysis for well-based modified STRT-seq, such as data normalization, clustering, differential expression analysis, was implemented using the R package Seurat 2. After quality control, 1,342 cells were retained. To start with, preliminary UMAP and clustering analysis were performed, after which stromal cells (111 cells), including epithelial cells and mesenchymal cells, were identified on the basis of expression of *EPCAM* and *PDGFRA* and removed from our data. After quality control, 1,231 cells were retained. To start with, 888 highly variable genes (HVGs) were detected using the 'FindVariableGenes' function with the parameters of $y.\text{cutoff} = 1$ and $x.\text{low.cutoff} = 1$. These HVGs were used to perform PCA and the top 20 significant PCs were selected using the elbow of standard deviations of principal components (PCs). Next, selected PCs were used to perform UMAP analysis and cluster detection using the RunUMAP and FindClusters functions, respectively. Finally, eight progenitor clusters were annotated as YSMP, ErP, MkP, GMP, myeloblast, CD7^{lo}P, CD7^{hi}P and HSPC, characterized by high expression of *CD34* and *MYB* and expressing specific genes for ErP (*GATA1*, *KLF1*), MkP (*GATA1*, *PF4*), GMP (*MPO*), myeloblast (*LYZ*), CD7^{lo}P (*IL7R*, *CD7*), CD7^{hi}P (*IL7R*, *CD7*) and HSPC (*HOXA6*, *HOXA10*). In addition, seven mature cell types were identified and defined as monocytes (*CCR2*, *HLA-DRA*), macrophages (*CD14*, *MRC1*), ILCs (*RORC*, *LTA*) and mast cells (*CPA3*, *CMA1*) for further downstream analysis.

To decode the cell type heterogeneity of the myeloid group, cells of unrelated types (ErP, MkP, mast cells, CD7^{lo}P, CD7^{hi}P, ILC and HSPC) were excluded from downstream analysis. We detected 2,195 HVGs with setting parameters $x.\text{low.cutoff} = 0.3$ and $y.\text{cutoff} = 0.75$, then selected the top 30 PCs to find clusters and to implement UMAP analysis, which resulted in 14 discrete cell clusters. Among the 14 clusters, in addition to YSMP, GMP, myeloblasts and monocytes, 10 macrophage clusters were identified and recognized as Blood_Mac, Liver_Mac, Skin_Mac, Lung_Mac, YS_Mac1, YS_Mac2, Head_Mac1, Head_Mac2, Head_Mac3, Head_Mac4, based on sampling site information.

For droplet-based 10x Genomics data, the pooled raw data from CS11 and CS15 yolk sacs were objected to Seurat 3 for data integration based on identification of 'anchors' between pairs of datasets following the tutorial at <https://satijalab.org/seurat/v3.1/integration.html>. In brief, SCTtransform normalization was implemented separately for each dataset, after which the top 3,000 feature genes were selected and used for data integration. Next, the integrated data were used for dimensionality reduction and cluster detection. To start with, PCA was performed and the top 30 PCs were used for UMAP analysis using the umap package; then eight discrete clusters were detected using FindClusters with setting parameters $\text{dims} = 1:10$, $\text{resolution} = 0.1$ and annotated as YSMP (*CD34*, *MYB*), ErP/MkP (*GATA1*, *PF4*, *KLF1*), Mac (*CD14*, *CD163*), Endo (*CDH5*, *SOX7*), Epi (*EPCAM*) or Mes (*PDGFRA*) on the basis of feature genes. The same analysis procedures were applied for yolk sac dataset integration between modified STRT-seq and 10x data.

To identify sub-clusters in the Mac clusters, macrophage clusters were separately reanalysed. The subset data were normalized and the top 2,000 HVGs were recognized after ranking by residual variance using the 'vst' method through the SCTtransform function. All HVGs were imported into PCA analysis carried out by RunPCA next.

SCENIC and GSEA analysis

To further assess the transcriptional and regulatory characteristics of the different progenitors including YSMP, ErP, MkP, myeloblast, CD7^{lo}P, CD7^{hi}P and HSPC populations, gene regulatory network analysis was performed using SCENIC⁴⁰. First, regulatory modules (regulons) were inferred from co-expression and DNA motif analysis. These regulons were then evaluated in each cell to ascertain their activity before a binary matrix was obtained. To profile the gene regulatory module features of all progenitors, the Spearman correlation coefficient between regulons was calculated, and only regulons with a correlation coefficient larger than 0.3 with at least one other regulon and activated in at least 30% cells in any progenitor clusters were included for visualization.

To find statistically significantly different regulons between YSMP and HSPC, gene set enrichment analysis (GSEA)⁴¹ was performed on the regulon gene sets produced by SCENIC analysis, and the top two regulons ordered by *P* value were used for visualization for each population.

To compare the differential expression of erythroid and myeloid signature between YSMP in human and EMP in mouse, we performed GSEA analysis, in which the erythroid and myeloid gene sets refer to published data⁴².

Developmental pseudotime analysis

The Monocle 2 package (version 2.8.0)^{43,44} in R was used to determine the pseudotime of YSMP development in liver. First, cells sampled from liver (CS11 to CS17) belonging to the clusters related to myeloid lineage and excluding mature macrophages, namely YSMP, GMP, myeloblast, and monocyte, were selected and subjected to Monocle 2. Then, UMI data and HVGs obtained by FindVariableGenes ($x.\text{low.cutoff} = 0.5$, $y.\text{cutoff} = 0.5$) in Seurat were input for unsupervised ordering of the cells. The remaining parameters were default. To find genes that changed their expression during the process of monocyte specification, we calculated Spearman correlations between predicted pseudotime

and gene expression: only genes with Spearman correlations higher than 0.4 were selected for visualization.

To interpret macrophages specification in head, only head macrophages (Head_Mac0, Head_Mac1, Head_Mac2, Head_Mac3 and Head_Mac4) were included and projected onto the UMAP plot. Clearly, head macrophages are ordered by sampling stage. To detect genes with dynamic changes in expression during head macrophage specification, we performed analysis of variance (ANOVA, one-sided). A total of 2,438 DEGs were identified and subjected to partitioning around medoids (PAM) analysis, a more robust K-means version, resulting in five gene expression patterns.

To predict the developmental trajectory of liver macrophages, gene expression data from all macrophages sampled from the liver were extracted. First, PCA was implemented after data were normalized and log transformed. Then, the top ten PCs were selected to create a diffusion map of cells through the destiny package with default parameters⁴⁵, after which DPT analysis was used to align cells in pseudotime order with setting cells of CS12 as the start point.

Signature gene detection

To find the signature genes of YSMPs, we compared YSMPs with other haematopoietic cells in the yolk sac and obtained DEGs with adjusted $P < 0.05$ and fold change > 2 .

We also probed signature genes in 10x data when comparing YSMPs (YSMP1 and YSMP2) with other haematopoietic clusters (ErP/MkP and Mac), before selecting DEGs with an adjusted $P < 0.05$ and fold change > 1.5 .

DEG and cluster biomarker identification

DEGs were identified by running the ‘FindAllMarkers’ function in Seurat. All DEGs of specific clusters are listed in the Supplementary Tables.

The datasets of surface markers and transcription factors were downloaded from Cell Surface Protein Atlas (<http://wlab.ethz.ch/cspa/>) and HumanTFDB3.0 (<http://bioinfo.life.hust.edu.cn/HumanTFDB/>) respectively.

Statistics and reproducibility

For all human embryos, specific sample information can be found in Fig. 1a and Supplementary Table 1.

For the STRT-seq analysis of human embryos, a total of $n = 8$ biologically independent samples across 7 time points were included. The time points and anatomical locations are CS11 (head and yolk sac, $n = 1$), CS12 (head, liver and yolk sac, $n = 1$), CS13 (head, blood and yolk sac, $n = 1$), CS15 (head, blood and liver, $n = 2$; yolk sac, $n = 1$), CS17 (head, blood, liver and yolk sac, $n = 1$), CS20 (brain, blood, skin, liver and lung, $n = 1$), and CS23 (brain, liver, skin, lung, $n = 1$). The total number of cells used in the final analysis was 1,231 (Figs. 1b–d, 2a–e, 3a–e, 4a–d and Extended Data Figs. 1–3, 5–8).

For the STRT-seq analysis of paediatric skin, a total of $n = 2$ biologically independent samples were obtained after elective circumcision. In total, 37 cells were used in the final analysis (Extended Data Fig. 8).

For the 3'10x data from yolk sac, a total of $n = 2$ biologically independent samples were included, at CS11 ($n = 1$) and CS15 ($n = 1$). In total, 9,565 cells were used in the final analysis (Extended Data Fig. 3).

Detailed cell numbers for scRNA-seq in each figure are shown below, with those from 10x data specifically indicated. Figure 1b, c and Supplementary Table 2: 1,231 haematopoietic cells (116 YSMP cells, 72 ErP cells, 30 MkP cells, 45 GMP cells, 100 myeloblast cells, 120 monocyte cells, 69 Mac_1 cells, 196 Mac_2 cells, 57 Mac_3 cells, 79 Mac_4 cells, 33 HSPCs, 104 CD7^{hi}P cells, 140 CD7^{lo}P cells, 21 ILCs, and 49 mast cells); Fig. 1d: 238 yolk sac haematopoietic cells and 354 liver haematopoietic cells; Fig. 2a–e and Supplementary Table 5: 88 liver haematopoietic cells; Fig. 3a: 782 myeloid cells (128 YSMP cells, 37 GMP cells, 103 myeloblast cells, 64 monocyte cells, 7 Liver_Mac cells, 51 Blood_Mac cells, 71 Lung_Mac cells, 46 Skin_Mac cells, 61 YS_Mac1 cells, 29 YS_Mac2 cells, 73 Head_Mac1 cells, 38 Head_Mac2 cells, 41 Head_Mac3 cells, and 33 Head_Mac4 cells); Fig. 3b, c: 450 macrophages (7 Liver_Mac cells,

51 Blood_Mac cells, 71 Lung_Mac cells, 46 Skin_Mac cells, 61 YS_Mac1 cells, 29 YS_Mac2 cells, 73 Head_Mac1 cells, 38 Head_Mac2 cells, 41 Head_Mac3 cells, and 33 Head_Mac4 cells); Fig. 3b: DEGs were detected using FindAllMarkers function in Seurat (one-sided Wilcoxon rank-sum test, with P value adjusted for multiple testing using Bonferroni correction), genes with fold change > 1.5 and adjusted $P < 0.05$ were selected, and the top three DEGs for each cluster were used for visualization; Fig. 3d, e: 514 myeloid cells with CS20 and CS23 cells excluded; Fig. 4a–d and Supplementary Table 7: 155 macrophages in head (9 Head_Mac0 cells, 45 Head_Mac1 cells, 29 Head_Mac2 cells, 39 Head_Mac3 cells and 33 Head_Mac4 cells).

Extended Data Figure 1a–c: 1,461 cells from $n = 8$ biologically independent embryo samples; Extended Data Fig. 1d–j: 1,231 haematopoietic cells; Extended Data Fig. 1j: DEGs were detected using FindAllMarkers function in Seurat (one-sided Wilcoxon rank-sum test, with P value adjusted for multiple testing using Bonferroni correction), genes with fold change > 1.5 and adjusted $P < 0.05$ were selected, and the top five DEGs for each cluster were used for visualization; Extended Data Fig. 1k, l: 188 progenitor cells (118 mouse EMP cells and 70 human YSMP cells); Extended Data Fig. 2a, d, f and Supplementary Table 3: 640 progenitor cells (116 YSMP cells, 72 ErP cells, 30 MkP cells, 45 GMP cells, 100 myeloblast cells, 33 HSPCs, 104 CD7^{hi}P cells, and 140 CD7^{lo}P cells); Extended Data Fig. 2b, c: 116 YSMP cells and 33 HSPC cells; Extended Data Fig. 2c, f: DEGs were detected using FindAllMarkers function in Seurat (one-sided Wilcoxon rank-sum test, with P value adjusted for multiple testing using Bonferroni correction), genes with fold change > 1.5 (1.25 for Extended Data Fig. 2c) and adjusted $P < 0.05$ were selected, and the top ten DEGs (top five for Extended Data Fig. 2f) for each cluster were used for visualization; Extended Data Fig. 2e: 259 progenitor cells in liver and 131 progenitor cells in blood; Extended Data Fig. 3a: 11,944 cells from 10x data; Extended Data Fig. 3b, c: 238 yolk sac cells from STRT-seq; Extended Data Fig. 3d, e and sheet 1 of Supplementary Table 4: 9,565 yolk sac cells from 10x data; Extended Data Fig. 3f: 93 macrophages from STRT-seq and 1,259 macrophages from 10x; Extended Data Fig. 3g, h: 9,803 cells (6,449 cells in CS11 yolk sac from 10x data, 3,116 cells in CS15 yolk sac from 10x data, and 238 yolk sac cells from STRT-seq); Extended Data Fig. 3i: DEGs were detected using FindAllMarkers function in Seurat (one-sided Wilcoxon rank-sum test, with P value adjusted for multiple testing using Bonferroni correction), surface marker genes with fold change > 1.25 and adjusted $P < 0.05$ were selected, and the top five DEGs for each cluster were used for visualization; Extended Data Fig. 5a–d: 88 liver haematopoietic cells from 4 biologically independent embryo samples; Extended Data Fig. 5e–g: 66 liver haematopoietic cells; Extended Data Fig. 6a–d and Supplementary Table 6: 782 myeloid cells (128 YSMP cells, 37 GMP cells, 103 myeloblast cells, 64 monocyte cells, 7 Liver_Mac cells, 51 Blood_Mac cells, 71 Lung_Mac cells, 46 Skin_Mac cells, 61 YS_Mac1 cells, 29 YS_Mac2 cells, 73 Head_Mac1 cells, 38 Head_Mac2 cells, 41 Head_Mac3 cells, and 33 Head_Mac4 cells); Extended Data Fig. 6c: DEGs were detected using FindAllMarkers function in Seurat (one-sided Wilcoxon rank-sum test, with P value adjusted for multiple testing using Bonferroni correction), genes with fold change > 1.5 and adjusted $P < 0.05$ were selected, and the top five DEGs for each cluster were used for visualization; Extended Data Fig. 6e: 450 macrophages (7 Liver_Mac cells, 51 Blood_Mac cells, 71 Lung_Mac cells, 46 Skin_Mac cells, 61 YS_Mac1 cells, 29 YS_Mac2 cells, 73 Head_Mac1 cells, 38 Head_Mac2 cells, 41 Head_Mac3 cells, and 33 Head_Mac4 cells); Extended Data Fig. 7a: 64 monocytes; Extended Data Fig. 7b: 176 monocytes and macrophages in head; Extended Data Fig. 7c: 64 monocytes and macrophages in lung; Extended Data Fig. 7d–g and Supplementary Table 9: 41 macrophages in liver; Extended Data Fig. 7h: 49 macrophages in skin; Extended Data Fig. 8a–c, e and Supplementary Table 8: 464 macrophages (20 adult head macrophages, 39 embryonic head macrophages, 97 adult liver macrophages, 9 embryonic liver macrophages, 160 adult lung macrophages, 59 embryonic lung macrophages, 37 paediatric skin macrophages, 43 embryonic skin

Article

macrophages); Extended Data Fig. 8d: 464 macrophages (58 head macrophages, 104 liver macrophages, 156 lung macrophages, 45 skin macrophages, and 101 unspecified macrophages).

For Supplementary Tables 2–4, 6 and 8, DEGs were detected using FindAllMarkers function in Seurat (one-sided Wilcoxon rank-sum test, with *P* value adjusted for multiple testing using Bonferroni correction). For Supplementary Tables 7 and 9, one-sided ANOVA was used to find DEGs between five head macrophage sub-clusters (Head_Mac0, Head_Mac1, Head_Mac2, Head_Mac3, and Head_Mac4) and four liver macrophage sub-clusters at different stages (CS12, CS15, CS17, and CS23), with *P* value adjusted using FDR using p.adjust function in R.

For cell culture of yolk sac populations, a total of $n = 4$ biologically independent embryo samples were used. For bulk culture of YSMP populations, 3 biologically independent samples (CS11, $n = 1$ and CS12, $n = 2$) were used, of which one CS12 sample was used for bulk co-culture of the CD45⁺CD34⁺CD44⁺ population. For the single-cell functional assay of YSMPs, one CS13 yolk sac was used, and 39 single-cell wells with haematopoietic clusters were analysed by FACS to detect haematopoietic lineages generated (Fig. 1e, Extended Data Fig. 4a–e).

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Raw data from scRNA-seq analysis have been deposited in the NCBI Gene Expression Omnibus (GEO) under accession numbers GSE133345 and GSE137010. Source Data for four Figures and eight Extended Data Figures are provided within the online content of this paper.

Code availability

All data were analysed with standard programs and packages, as detailed above. Scripts can be found at https://github.com/yand-gong307/human_macrophage_project.

34. Moore, K. L., Persaud, T. V. N. & Torchia, M. G. *The Developing Human E-Book: Clinically Oriented Embryology With STUDENT CONSULT Online Access* (Elsevier Health Sciences, 2011).
35. Li, J. et al. Single-cell RNA-seq analysis maps development of human germline cells and gonadal niche interactions. *Cell Stem Cell* **20**, 858–873.e854 (2017).
36. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).

37. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protocols* **9**, 171–181 (2014).
38. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
39. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
40. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
41. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
42. Chambers, S. M. et al. Hematopoietic fingerprints: an expression database of stem cells and their progeny. *Cell Stem Cell* **1**, 578–591 (2007).
43. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
44. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
45. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).
46. Darmanis, S. et al. A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. USA* **112**, 7285–7290 (2015).
47. Cohen, M. et al. Lung single-cell signalling interaction map reveals basophil role in macrophage imprinting. *Cell* **175**, 1031–1044.e1018 (2018).
48. MacParland, S. A. et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* **9**, 4383 (2018).

Acknowledgements We thank F. Tang, Q. Li and Y. Hu from Peking University for technical assistance during single cell library construction, and L. Robinson for language editing of the manuscript. This study was supported by grants from the National Key Research and Development Program of China, Stem Cell and Translational Research (2017YFA0103401, 2016YFA0100601 and 2019YFA0110201), the National Natural Science Foundation of China (31425012, 31930054, 81890991, 31871173, 81800102, 81600077, 81900115 and 31800978), the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (2017ZT07S347), the Key Research and Development Program of Guangdong Province (2019B020234002), the Beijing Municipal Science & Technology Commission (Z171100000417009 and Z171100001117159), the National Key Research and Development Plan Young Scientists Program (2017YFA0106000), the State Key Laboratory of Proteomics (SKLP-K201502) and the China Postdoctoral Science Foundation (2018M643373). F.G. is supported by Singapore Immunology Network (SIgN) core funding. F.G. is a European Molecular Biology Organization (EMBO) YIP awardee and is supported by the Singapore National Research Foundation Senior Investigatorship (NRFI) NRF2016NRF-NRFI001-02. C.Z.W.L. is supported by an A*STAR Graduate Scholarship.

Author contributions B.L., F.G. and Y.L. designed the study. Z. Bian and H.S. performed sample preparation and FACS with help from Y.Z., Z. Bai and Y.N. T.H., Z. Bian, H.S., C.L. and J.H. performed scRNA-seq with help from J.Z. and X.L. Z. Bai performed cell culture with help from L.B. and Y.Z. L.B., C.M., R.Z. and L.C. collected and prepared the samples. Y.G. and T.H. performed bioinformatics analysis with help from Z.L., B.L., Y.L., Z. Bian, C.Z.W.L., J.K.Y.C. and L.G.N. Z. Bian, C.Z.W.L., T.H., Y.G., H.S., Z. Bai, Y.L., F.G. and B.L. wrote the manuscript, with contributions from all authors.

Competing interests The authors declare no competing interests.

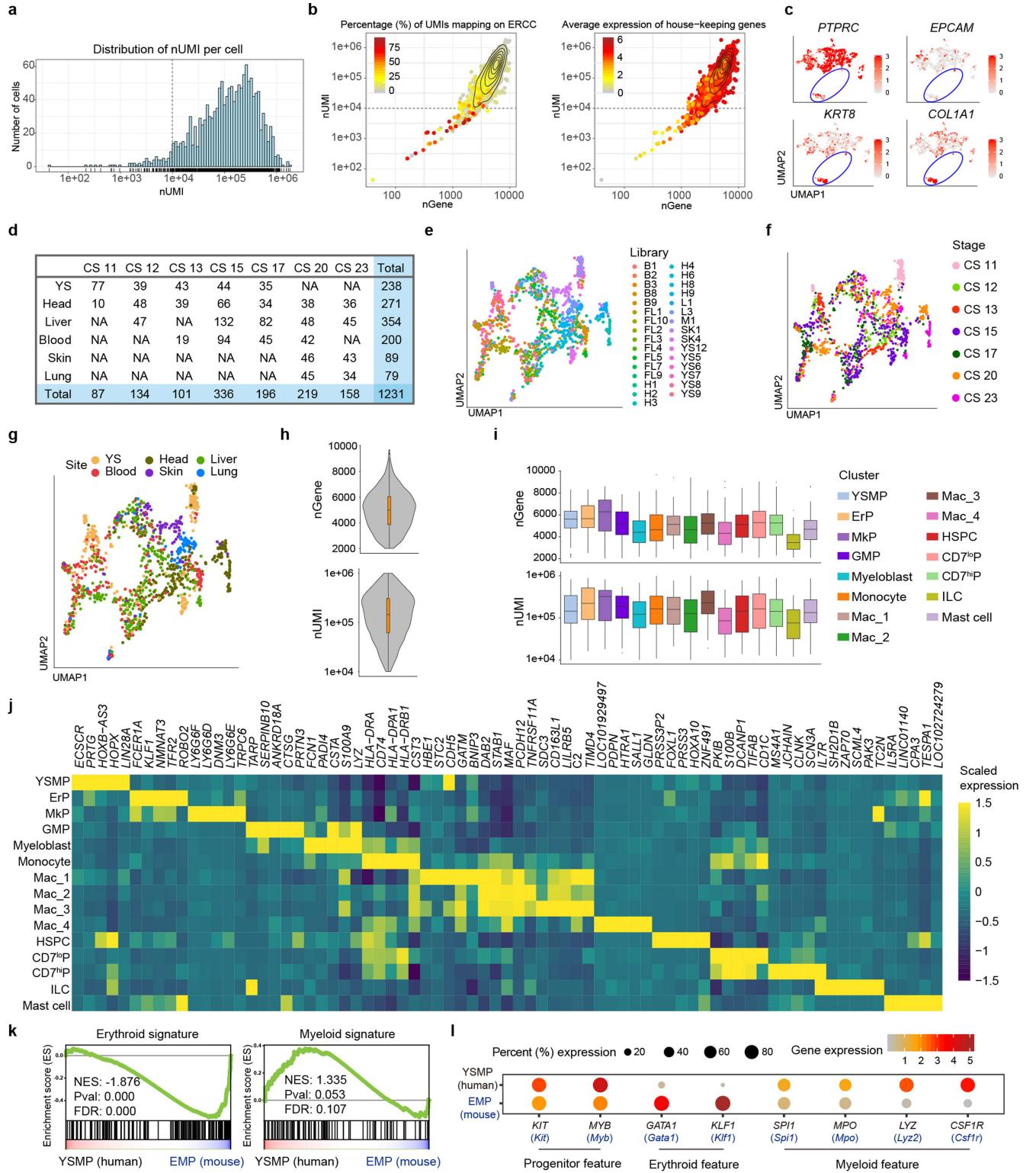
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2316-7>.

Correspondence and requests for materials should be addressed to Y.L., F.G. or B.L.

Peer review information *Nature* thanks Hans-Reimer Rodewald and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

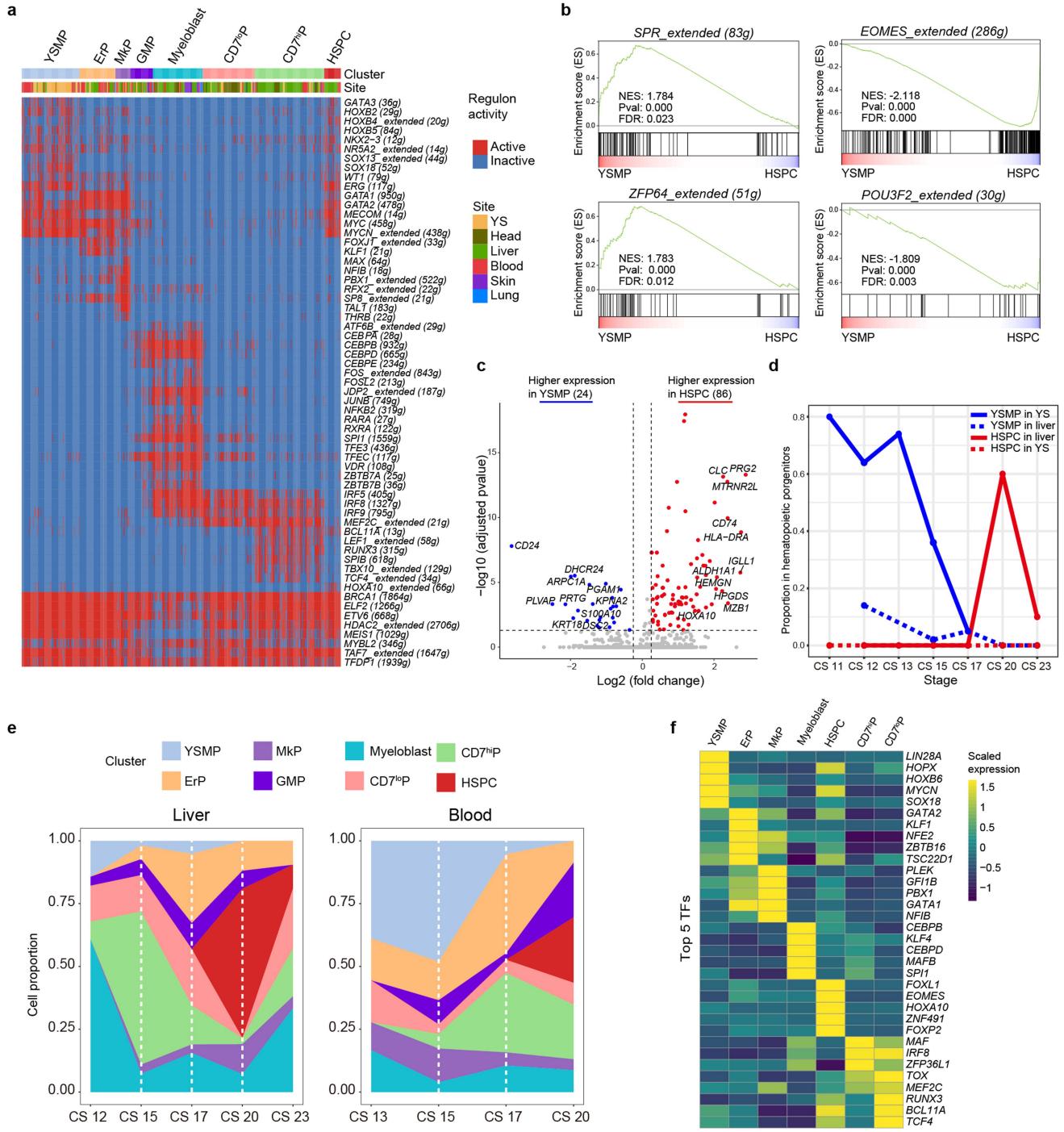
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | See next page for caption.

Article

Extended Data Fig. 1 | Technical information about scRNA-seq library and overview of haematopoietic clusters. **a**, Graph showing distribution of UMI per cell for STRT-seq data ($n = 8$ biologically independent embryo samples and 1,461 cells). Threshold for final analysis was set at $>10,000$. **b**, Percentage of UMIs mapped on to ERCC and average expression of house-keeping genes. nGene: number of genes expressed, nUMI: number of UMIs expressed. **c**, UMAP visualization showing exclusion of non-haematopoietic cells from final analysis based on expression of *PTPRC*, *EPCAM*, *KRT8* and *COL1A1*. **d**, Numbers, location and Carnegie stage information for cells used in final analysis. **e**, UMAP showing minimal batch effect among single cell libraries ($n = 1,231$ cells). **f, g**, UMAP visualization of all haematopoietic cells with Carnegie stage (**f**) and site (**g**) information mapped on. **h**, Violin plots of average gene and UMI numbers of scRNA-seq in STRT-seq data. For box plot within each violin plot, centre black lines indicate median values, boxes range from 25th to 75th percentiles, and whiskers correspond to $1.5 \times \text{IQR}$. **i**, Gene and UMI numbers of identified haematopoietic clusters. Centre black lines indicate median values, boxes range from 25th to 75th percentiles, and whiskers correspond to $1.5 \times \text{IQR}$. **j**, Heat map of the top five DEGs between haematopoietic clusters. DEGs were detected using *FindAllMarkers* function in Seurat (one-sided Wilcoxon rank-sum test, with P value adjusted for multiple testing using Bonferroni correction), and genes with fold change >1.5 and adjusted $P < 0.05$ were selected. **k**, GSEA plots of erythroid and myeloid signatures indicate significantly lower expression of erythroid signature in human YSMPs than in mouse EMPs (extracted from a published scRNA-seq dataset)¹⁹ ($n = 70$ human YSMP cells and 118 mouse EMP cells). P value was calculated using permutation test (one-sided) based on phenotype by GSEA 3.0 software, representing statistical significance of normalized enrichment score (NES). FDR, false discovery rate. **l**, Gene expression of haematopoietic progenitor feature (*CD34*, *MYB*), erythroid feature (*GATA1*, *KLF1*) and myeloid feature (*SPI1*, *MPO*, *LYZ*, *CSF1R*) in human YSMPs and mouse EMPs^{13,19}.

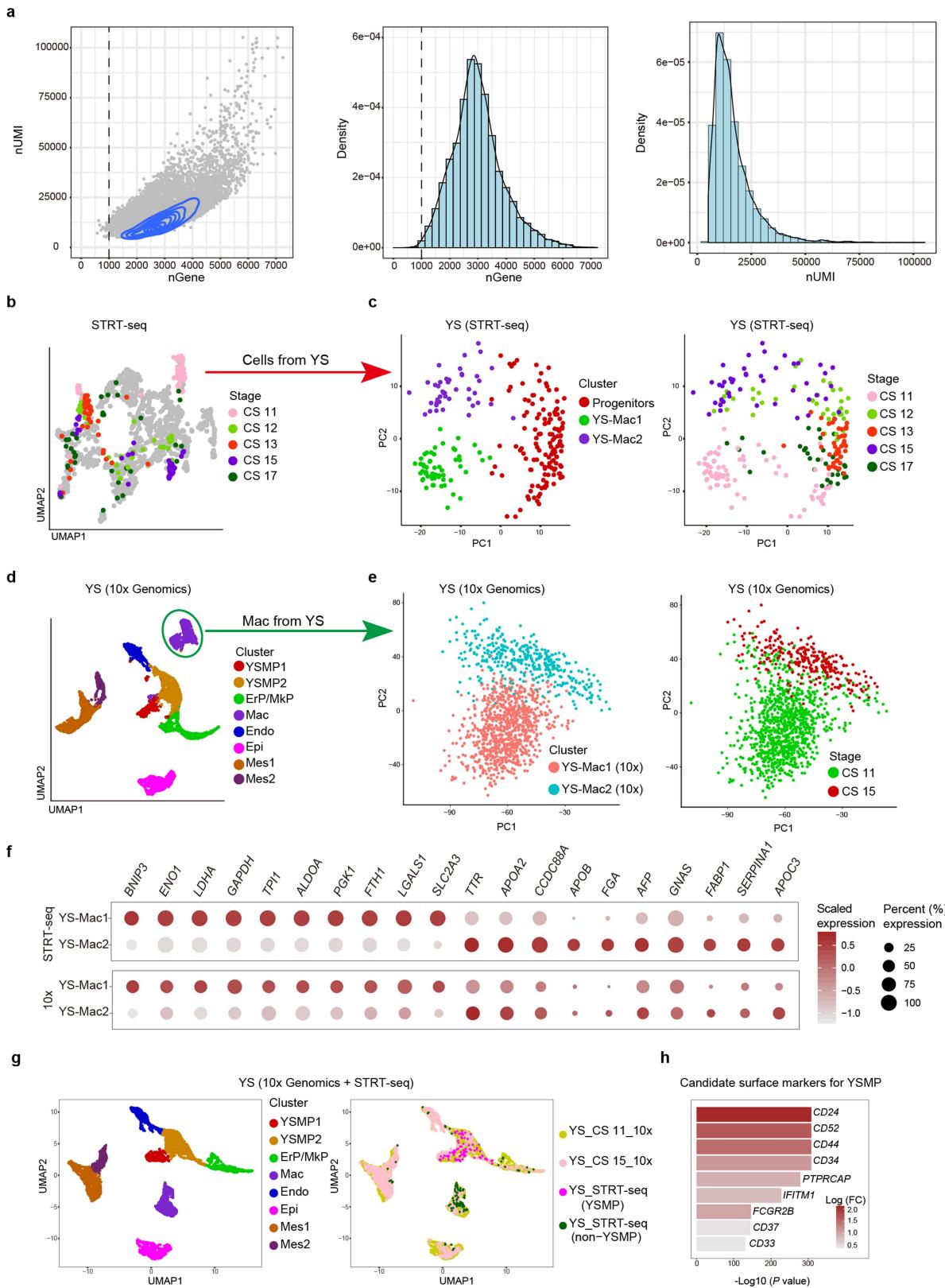


Extended Data Fig. 2 | See next page for caption.

Article

Extended Data Fig. 2 | Characteristics of haematopoietic progenitors in human embryos. **a**, Heat map showing differential regulon expression between haematopoietic progenitor clusters (YSMP, $n=116$; ErP, $n=72$; MkP, $n=30$; GMP, $n=45$; myeloblast, $n=100$; CD7^{lo}P, $n=140$; CD7^{hi}P, $n=104$; HSPC, $n=33$) generated by SCENIC and clear sets of cell-type specific regulons that may play critical roles in the development of each progenitor population. The number of genes associated with each regulon is listed in parentheses. ErP and MkP signatures were very similar, although MkP appeared to have downregulated expression of *KLF1* and up-regulated expression of other platelet-related transcription factors such as *TAL1* and *NFIB*. CD7^{lo}P had overlapping modules with myeloblast and GMP, sharing the myeloid-restricted *TFEC* pathways, but lacked expression of more myeloid-committed *CEBPs*. CD7^{hi}P showed many signatures typical of lymphoid potential, such as the activation of *LEF1* and *TCF4* signals. HSPC were characterized by activation of the *HOXA10* module, as well as higher levels of lymphoid-associated *BCL11A* when compared to YSMP. **b**, GSEA plots of the top two differentially expressed regulons between YSMP ($n=116$ cells) and HSPC ($n=33$ cells). GSEA analysis revealed that YSMPs highly enriched the *SPR* and *ZFP64* regulons, while HSPCs had higher expression of the *EOMES* and *POU3F2* modules. P value was calculated using permutation test (one-sided) based on phenotype by GSEA 3.0 software, representing the statistical significance of enrichment score. **c**, Volcano plot of DEGs between YSMP ($n=116$ cells) and

HSPC ($n=33$ cells), with the top 10 genes for each cluster indicated. Although the regulon landscape was similar between these two groups, we identified 110 DEGs (Supplementary Table 3). There were more upregulated genes in HSPC (86, red) than in YSMP (24, blue), with HSPC expressing genes related to antigen presentation including *CD74* and *HLA-DRA* as well as lymphoid-related genes including *IGLL1*. DEGs were detected using FindAllMarkers function in Seurat (one-sided Wilcoxon rank-sum test, with P value adjusted for multiple testing using Bonferroni correction), and genes with fold change >1.25 and adjusted $P<0.05$ were selected. **d**, Proportion changes of YSMPs and HSPCs in the haematopoietic progenitor populations of yolk sac and liver between CS11 and CS23 ($n=8$ biologically independent embryo samples). The proportion of the YSMP population peaked at CS11 before steadily decreasing, while that of the HSPC population expanded between CS17 and CS20 before reducing to about 10% at CS23. **e**, Proportion changes of different haematopoietic progenitor clusters from CS12 to CS23 in the liver ($n=6$ biologically independent embryo samples and 259 cells), and CS13 to CS20 in the blood ($n=5$ biologically independent embryo samples and 131 cells). **f**, Heat map showing expression levels of the top five differentially expressed transcription factors between YSMP, ErP, MkP, myeloblast, HSPC, CD7^{lo}P and CD7^{hi}P cells. DEGs were detected using FindAllMarkers function in Seurat (one-sided Wilcoxon rank-sum test, with P value adjusted for multiple testing using Bonferroni correction), and genes with fold change >1.5 and adjusted $P<0.05$ were selected.

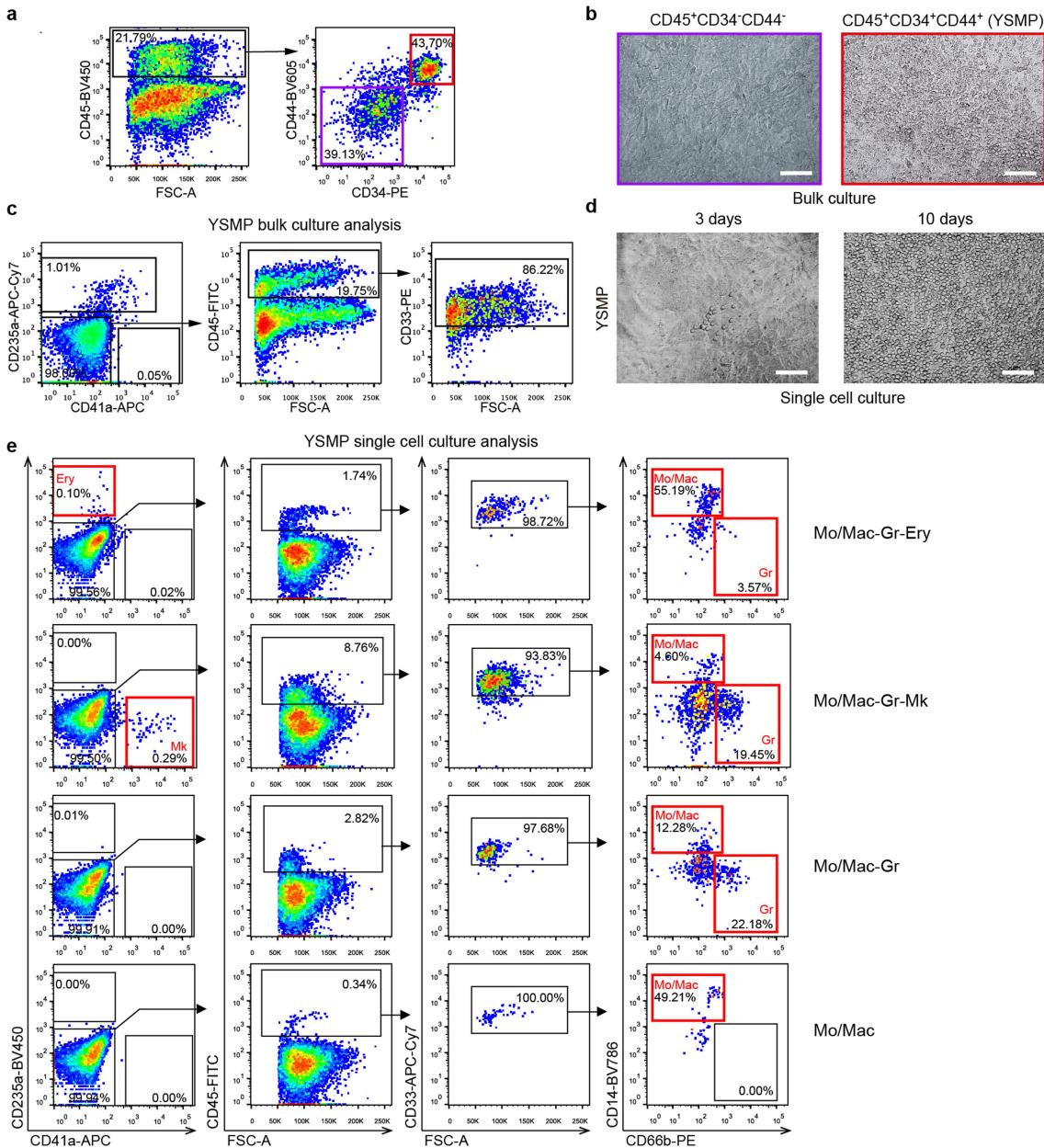


Extended Data Fig. 3 | See next page for caption.

Article

Extended Data Fig. 3 | Validation of STRT-seq macrophage clustering in yolk sac by 10x Genomics. **a**, Quality control for 10x Genomics data by UMI and gene numbers ($n = 2$ biologically independent embryo samples and 11,944 cells). The threshold for final analysis was set as gene number $>1,000$ per cell. **b**, UMAP visualization of total haematopoietic clusters generated via STRT-seq with cells from yolk sac ($n = 5$ biologically independent embryo samples and 238 cells) mapped on and coloured by stage information, which were extracted for further analysis. **c**, PCA of cells from yolk sac in STRT-seq with re-clustering (left) and stage (right) information mapped on ($n = 5$ biologically independent embryo samples and 238 cells). These cells were re-clustered into three clusters and annotated by gene expression profiles. The YS-Mac1 cluster mainly consisted of cells from CS11, while the YS-Mac2 cluster mainly consisted of cells from CS15. Based on these findings, we selected the 10x Genomics data from the CS11 and CS15 yolk sacs to validate our clustering. **d**, UMAP visualization of 10x Genomics data from CS11 and CS15 yolk sacs ($n = 2$ biologically independent embryo samples and 9,565 cells). The Mac cluster was extracted for further analysis. **e**, PCA of Mac cluster in 10x Genomics data with re-clustering (left)

and stage (right) information mapped on. **f**, Expression profile of top ten DEGs between YS-Mac1 and YS-Mac2 identified by STRT-seq ($n = 238$ cells) and projected onto the 10x Genomics data ($n = 9,565$ cells). Even though 10x Genomics data have lower depth compared to STRT-seq, similar expression profiles can be seen for the majority of genes. **g**, UMAP visualization of integrated yolk sac data from STRT-seq and 10x Genomics analysis ($n = 9,803$ cells). There is less overlap in the mesenchymal (Mes1 and Mes2) and epithelial (Epi1) clusters because the STRT-seq data were only from CD45⁺ haematopoietic cells, whereas the 10x Genomics data were from all yolk sac cells. Note that YSMPs from both datasets are well merged. **h**, Bar plot showing putative surface markers of YSMPs. On the basis of these data, CD34 and CD44 were selected for functional assays. DEGs were identified from combined STRT-seq and 10x Genomics data using FindAllMarkers function in Seurat (one-sided Wilcoxon rank-sum test with P value adjusted for multiple testing using Bonferroni correction), and surface marker genes with fold change >1.25 and adjusted $P < 0.05$ were selected.



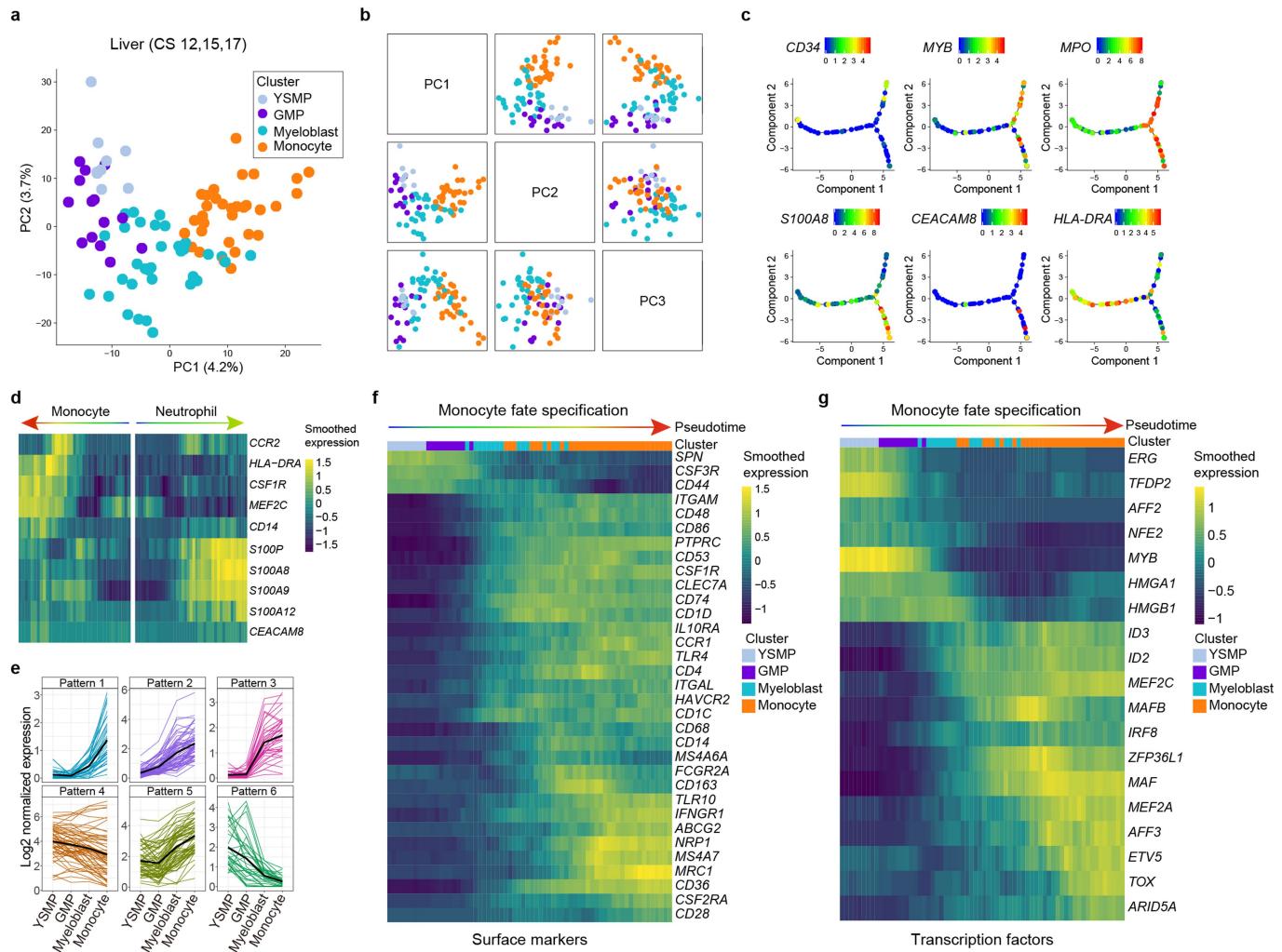
Extended Data Fig. 4 | In vitro functional assay of YSMPs. **a**, Gating strategy for sorting of the YSMPs ($CD45^+CD34^+CD44^+$) from a CS12 yolk sac.

b, Representative morphologies of bulk cultures (100 cells per well) of negative control cells ($CD45^+CD34^+CD44^-$, $n=5$ replication wells) and YSMPs ($CD45^+CD34^+CD44^+$, $n=16$ replication wells) after 14 days of culture on MS5 feeder layer. $n=3$ independent experiments from one sample of CS11 yolk sac and two samples of CS12 yolk sac for YSMPs. Scale bars, 100 μ m.

c, Representative FACS analysis of cells collected from bulk cultures of YSMPs. Note that the myeloid cells ($CD33^+$) are predominant, in contrast to a small number of erythroid cells ($CD235a^+$) detected ($n=3$ independent experiments).

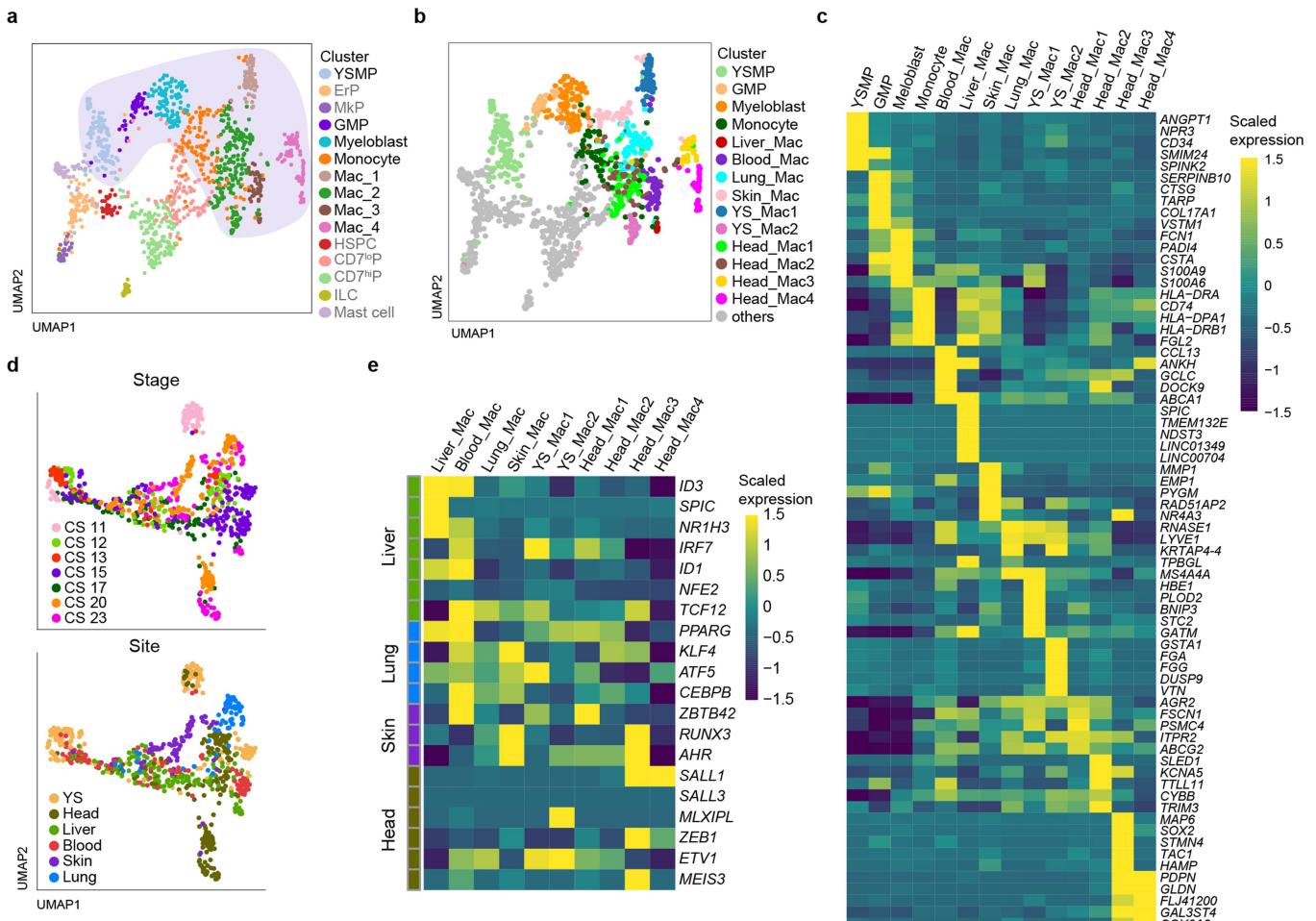
d, Representative morphologies of haematopoietic cells generated by a single YSMP from a CS13 yolk sac after 3 and 10 days of culture on MS5 feeder layer. In total, 184 YSMPs were individually cultured and 67 of them generated morphologically typical haematopoietic clusters. Scale bars, 100 μ m. **e**, Representative FACS analysis of four kinds of distinct differentiation potential of single YSMPs. Cells were collected from the single-cell YSMP cultures and in total 39 wells were individually analysed. Lineage differentiation potentials are indicated in red for each clone. Mo/Mac, monocytes/macrophages ($CD45^+CD33^+CD14^+$); Gr, granulocytes ($CD45^+CD33^+CD66b^+$); Ery, erythrocytes ($CD235a^+$); Mk, megakaryocytes ($CD41a^+$).

Article



Extended Data Fig. 5 | Developmental trajectory of YSMP in human embryonic liver. **a**, PCA of YSMP, GMP, myeloblast and monocyte populations sampled from CS12 to CS17 livers using PC1 and PC2 ($n = 4$ biologically independent embryo samples and 88 cells). **b**, PCA matrix of YSMP, GMP, myeloblast and monocyte populations. **c**, Monocle visualization of YSMP, GMP, myeloblast and monocyte populations sampled from CS12 to CS17 livers with the expression of the indicated genes mapped on. **d**, Heat map showing scaled expression of branching curated genes of monocyte and neutrophil fates

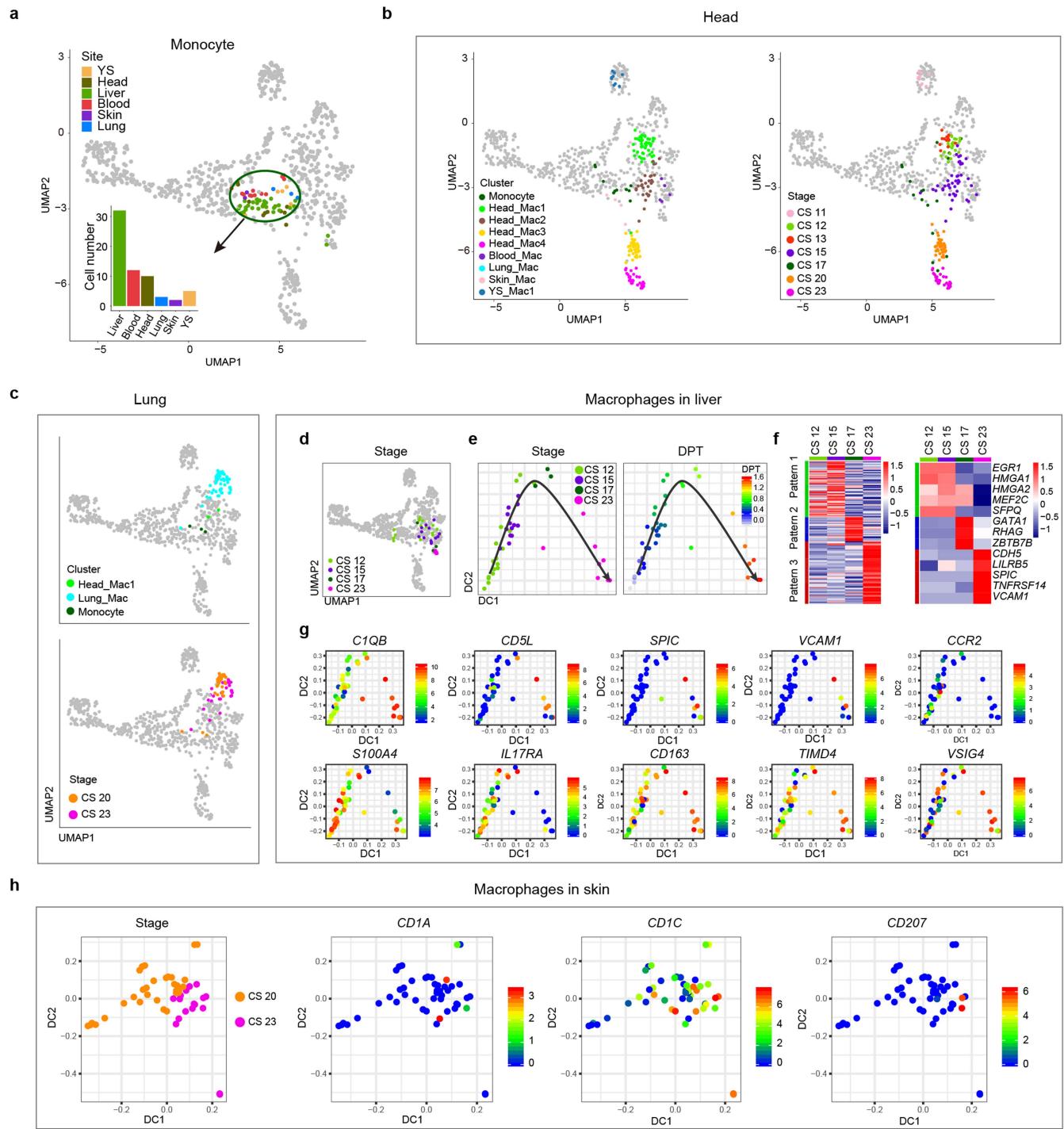
ordered by pseudotime. **e**, Six main patterns of gene expression compared between YSMP, GMP, myeloblast and monocyte clusters. The expression levels of all pattern genes (coloured lines) and the average expression of each pattern (black line) are shown. The complete list of genes can be found in Supplementary Table 5. **f,g**, Heat maps displaying expression of monocyte specification-related surface markers (**f**) and transcription factors (**g**) along pseudotime.



Extended Data Fig. 6 | Two distinct waves of yolk sac-derived macrophages contribute to TRM populations in human embryos. **a**, The area coloured lavender on the UMAP visualization highlights the myeloid groups selected for re-analysis ($n=8$ biologically independent embryo samples and 1,231 cells). **b**, UMAP visualization of all the haematopoietic cells with re-clustered myeloid and macrophage clusters ($n=782$ myeloid cells) mapped on. **c**, Heat map showing scaled expression of the top five DEGs for each re-clustered myeloid and macrophage population. DEGs were detected using FindAllMarkers function in Seurat (one-sided Wilcoxon rank-sum test, with P value adjusted for

multiple testing using Bonferroni correction), and genes with fold change >1.5 and adjusted $P<0.05$ were selected. **d**, UMAP visualization of myeloid cells with stage (top) and site (bottom) information mapped on. **e**, Heat map showing scaled expression of curated TRM signature genes from a previous mouse study²⁴ in the re-clustered macrophage populations ($n=450$ cells: 7 Liver_Mac cells, 51 Blood_Mac cells, 71 Lung_Mac cells, 46 Skin_Mac cells, 61 YS_Mac1 cells, 29 YS_Mac2 cells, 73 Head_Mac1 cells, 38 Head_Mac2 cells, 41 Head_Mac3 cells, and 33 Head_Mac4 cells).

Article

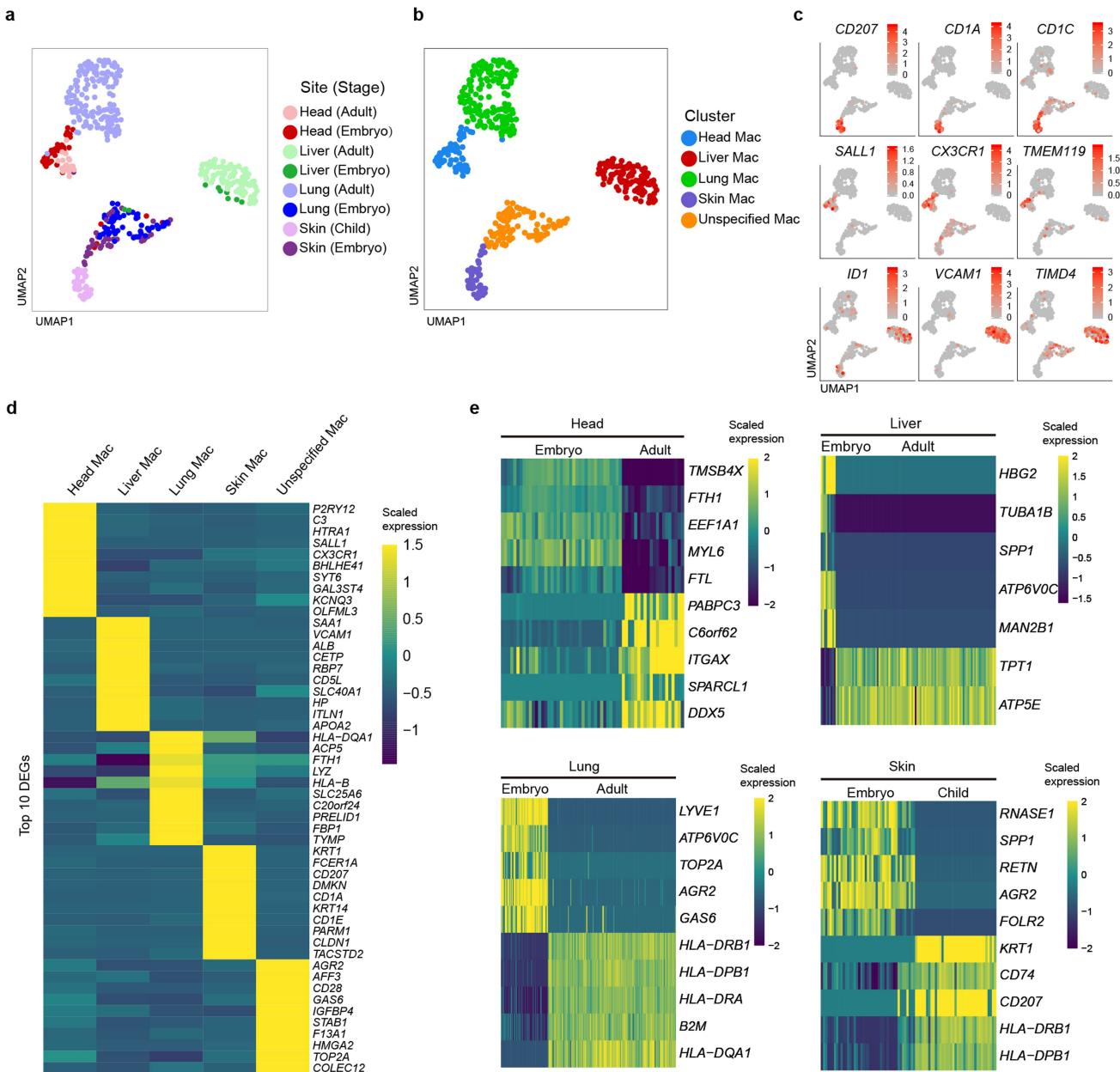


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Monocyte distribution and macrophage specification in human embryonic head, lung, liver and skin. **a**, UMAP visualization of myeloid cells with monocytes ($n=64$ cells) coloured by site information mapped on. Bar plot shows cell numbers at different sites. **b**, UMAP visualization of myeloid cells with monocytes and macrophages from human embryonic head ($n=176$ cells) mapped on. Cluster (left) and stage information (right) are indicated by colours. **c**, UMAP visualization of myeloid cells with monocytes and macrophages from human embryonic lung ($n=64$ cells) mapped on. Cluster (top) and stage information (bottom) is indicated by colours. **d**, UMAP visualization of the myeloid cells with macrophages from human embryonic liver ($n=41$ cells) coloured by stage information mapped on. These cells were used to study Kupffer cell specification *in situ*. **e**, DiffusionMap visualizing differentiation trajectory of embryonic Kupffer cells with stage information (left) and pseudo-order (right) mapped on. Note that the cells also lined up in a continuum from CS12 to CS23, suggesting the gradual and sequential acquisition of TRM identity. **f**, Heat maps showing scaled expression of DEGs (left) and transcription factors within DEGs (right) in embryonic Kupffer cells across stages with three main gene expression patterns identified. DEGs were detected using FindAllMarkers function in

Seurat (one-sided Wilcoxon rank-sum test, with P value adjusted for multiple testing using Bonferroni correction), and genes with fold change >1.5 and adjusted $P < 0.05$ were selected. Complete gene list can be found in Supplementary Table 9. **g**, DiffusionMap visualizing differentiation trajectory of embryonic Kupffer cells with expression levels of the indicated genes mapped on. Expression of *C1QB*, a gene associated with macrophage tissue residency, was gradually upregulated, while genes related to Kupffer cell function such as *CDSL*, *SPIC* and *VCAM1* were expressed only at the end of the developmental pathway, suggesting that specialized Kupffer cells began to appear after CS17. Many of the downregulated genes are inflammation- or migration-related, such as *CCR2*, *S100A4* and *IL17RA*, while the expression of residency and Kupffer cell identity genes such as *CD163*, *TIMD4* and *VSIG4* was increased. Many of the signature genes, such as *SPIC* and *VCAM1*, have been previously reported in TRMs using animal models, which confirms that these cells were moving towards a more differentiated tissue-resident state. **h**, DiffusionMap visualization of macrophages from human embryonic skin ($n=49$ cells) with stage information (left) and the expression levels of the indicated genes (right) mapped on.

Article



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Characteristics of human embryonic TRMs versus conventional TRMs. **a, b**, UMAP visualization of embryonic TRMs and their conventional TRM counterparts with site and stage (**a**) and cluster (**b**) information mapped on ($n=464$ cells: 20 adult head macrophages, 39 embryonic head macrophages, 97 adult liver macrophages, 9 embryonic liver macrophages, 160 adult lung macrophages, 59 embryonic lung macrophages, 37 paediatric skin macrophages, and 43 embryonic skin macrophages). We performed combined analysis including the four embryonic TRM populations in the present study (Head_Mac4, Liver_Mac, Lung_Mac, and Skin_Mac) and the corresponding conventional TRMs in adults (head, liver and lung from public scRNA-seq data)^{46–48} and children (skin). In total, five main macrophage clusters (head, liver, lung, skin and unspecified) were identified by unsupervised clustering. Two of the embryonic TRM populations (head and liver) clustered with their corresponding adult counterparts. The embryonic skin TRMs distributed into both specified and unspecified clusters, with the former cluster together with those from paediatric skin. The embryonic lung macrophages did not cluster at all with those in the adult lung, which indicated that the differentiation and specification of these TRMs had not yet occurred. **c**, UMAP visualizations of all TRM clusters with the expression levels of Langerhans (*CD207*, *CD1A* and *CD1C*), microglial (*SALL1*, *CX3CR1* and *TMEM119*) and Kupffer (*IDL*, *VCAM1* and *TIMD4*) cell-related genes mapped on. **d**, Heat map showing scaled expression of the top ten DEGs between the five identified human macrophage clusters ($n=464$ cells: 58 head macrophages, 104 liver

macrophages, 156 lung macrophages, 45 skin macrophages, and 101 unspecified macrophages). DEGs were detected using *FindAllMarkers* function in Seurat (one-sided Wilcoxon rank-sum test, with P value adjusted for multiple testing using Bonferroni correction), and genes with fold change >1.5 and adjusted $P < 0.05$ were selected. Note that they were distinguished by the expression of TRM genes that have been well described in previous animal and human studies, such as *CD207* for the skin, *VCAM1* for the liver and *P2RY12* for the head. **e**, Heat map showing scaled expression of DEGs between the embryonic TRMs and their conventional TRM counterparts in each tissue (head, liver, lung and skin). DEGs were detected using *FindAllMarkers* function in Seurat (one-sided Wilcoxon rank-sum test, with P value adjusted for multiple testing using Bonferroni correction), and genes with fold change >1.5 and adjusted $P < 0.05$ were selected. The complete list of genes can be found in Supplementary Table 8. Many of the upregulated genes in the embryonic TRMs are related to cell cycle or tissue development, whereas the upregulated genes in the conventional TRMs are more related to immune function. For example, in the head, the embryonic TRMs expressed the neurodevelopmental gene *TMSB4X* as well as the cell cycle-related gene *EEF1A1*, whereas conventional TRMs expressed the immune-related gene *ITGAX*. Embryonic skin macrophages expressed the chemokine *SPPI*, indicating that they are either cells in transition or have newly arrived in the niche, further supporting our prediction that skin TRM specification has just begun at this time-point.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

RS image expression (version 4.5.1.3)

Data analysis

Software used include: BD FACSDIVA (v8.01), Flowjo (v10), Cell Ranger (v2.10), Hisat2 (v2.10), HTSeq (v0.11.2), R (v3.50), Rstudio (v1.1), Seurat (v2.31 & v3.10), SCENIC (v0.96), Monocle 2 (v2.80), destiny (v2.10.0), GSEA (v3.0), Adobe Illustrator CS6 (16.0.0). Detailed parameters of each of the methods are mentioned in relevant sections in Methods.

All data were analyzed with standard programs and packages, as detailed above. Scripts can be available at https://github.com/yandgong307/human_macrophage_project.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw data from scRNA-seq analysis have been deposited in the NCBI Gene Expression Omnibus under accession number GSE133345 and GSE137010. Source data for 4 figures and 8 extended data figures are provided within the online content of this paper.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size of scRNA-seq were determined by availability of human tissues. Final dataset scale was determined according to the quality control criteria as described in the methods.
Data exclusions	Quality control standard is established in advance according to the previous literature, which was accord with the general standard of the single cell library construction. We have used the following cell exclusion criteria: low quality, doublets, non-CD45+CD235a- cells, and the details placed in the section of methods.
Replication	In this study, we have 2 biologically independent CS 15 embryos and 2 biologically independent foreskins for STRT-seq. There were no other replications for STRT-seq and 10x. For cell culture of yolk sac populations, a total of n = 4 biologically independent embryo samples were used. For the bulk culture of YSMP population, n = 3 (one CS 11 and two CS 12). For the single cell functional assay of YSMPs, n = 1 CS 13 yolk sac was used and n = 39 single-cell wells were analyzed by FACS to detect hematopoietic lineages generated. We used 10x Genomics and colony assay to verify the results found by STRT-seq, and found the cluster identity were consistent across different samples and sequencing techniques.
Randomization	We did not randomise the samples, as we were studying the trajectory of healthy embryonic hematopoietic development in general.
Blinding	Blinding is not relevant, as we are not studying pathology or disease.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

7-amino-actinomycin D (7-AAD) (PerCP-Cy5.5, 00699350 eBioscience, Lot 1910559)
anti-CD1a (APC, 559775 Biolegend, Lot 8164562)
anti-CD14 (BV786, 563698 BD, Lot 8351911)
anti-CD207 (PE, 564727 BD, Lot 8135683)
anti-CD235a (APC-Cy7, 349116 Biolegend, Lot 7355682 and B289027)
anti-CD235a (Pacific Blue, 306611 BioLegend, Lot B224563)
anti-CD33 (PE, 555450 BD, Lot 8074660)
anti-CD33 (APC-Cy7, 366614 BioLegend, Lot B252646)
anti-CD34 (PE, 550761 BD, Lot 7129824)
anti-CD41 (APC, 17-0419-42 eBioscience, Lot 2073742)
anti-CD44 (BV605, 562991 BD, Lot 7103609)
anti-CD45 (BV421, 563879 BD, Lot 9066960)
anti-CD45 (FITC, 11-0459-42 eBioscience, Lot 4310016)
anti-CD66b (PE, 561650 BD, Lot 7264511)

Validation

Antibodies used in the study are all commercial and verified by the manufacturer with catalog number and lot number provided.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	All human embryos included were between Carnegie stage 11 and 23. We collected 3 male, 6 female and 2 gender non-available embryos. The integrity of the samples was confirmed and examined at multiple steps. From the beginning, the Obstetrics and Gynecology clinicians confirmed the collected embryos and fetuses free of any known genetic or developmental abnormality. Before sample processing, the morphological examinations were performed to exclude samples with any potential development defects. The pediatric skin samples were obtained from foreskin of two children (8 and 10 years old).
Recruitment	The embryos were obtained from pregnant women undergoing drug abortion at The Fifth Medical Center of the PLA General Hospital (Beijing, China) after obtaining informed consent. Inclusion criteria: (1) Females older than 18 years of age and pregnant for 20-80 days (about 3-11 weeks) (2) Healthy pregnant women who intend to voluntarily terminate pregnancy through drug abortion and have no history of chronic diseases, genetic diseases and infectious diseases. Based on the principle of voluntary participation, no compensation had been paid in this study. Given the samples were extremely rare and precious, the sample selection was entirely determined by the availability of clinical specimens. Enrolled embryos were all quality controlled, which had been performed by two persons respectively. There was no self-selection bias in the progress of sample enrolled. Pediatric skin samples were obtained from tissue discarded after elective circumcision at Beijing Children's Hospital, Capital Medical University with parental consent.
Ethics oversight	All protocols were approved by the institutional review boards (The Affiliated Hospital of Academy of Military Medical Sciences Ethics Committee and Beijing Children's Hospital, Capital Medical University Ethics Committee; approval number: ky-2017-3-5 & IEC-C-008-A08-2018-75) and in accordance with the regulations of the Declaration of Helsinki. All the protocols were compliant with the Interim Measures for the Administration of Human Genetic Resources, administered by the Ministry of Science and Technology of China.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

The embryos were obtained from pregnant women undergoing drug abortion at The Fifth Medical Center of the PLA General Hospital (Beijing, China) after obtaining informed consent. The integrity and morphology of the embryos were evaluated, the somite pairs were counted, and the crown-rump length of the embryos were measured under the microscope to define the developmental stage1. The child skin was obtained from tissue discarded after elective circumcision at Beijing Children's Hospital, Capital Medical University with parental consent. Samples were immediately transported in RPMI1640 medium (Gibco,11875093) containing 10% Fetal Bovine Serum (HyClone, SH30070.03) on ice and processed within two hours. The various tissues (yolk sac, head, lung, skin, and liver) were carefully dissected under the microscope, with the exception of the blood which was collected by suction pipe from the heart directly. After washing in PBS 3 times to remove any residual blood contamination, the yolk sac, head, lung and skin were transferred to pre-warmed digestion medium containing 0.1 g/mL Collagenase I (Sigma, C2674, preheated to 37 °C) in RPMI 1640 medium. The samples were enzymatically digested at 37 °C in a humidified incubator for 20-30 minutes, with the samples being shaken periodically every five minutes until digested to single cell suspension. The liver was first cut into small pieces using scalpels, and then mechanically dissociated using syringes into single cell suspension, before removing the erythrocytes with lysis buffer (BD). The cells were then filtered through a 70 µm cell strainer after neutralizing enzymes. For the pediatric skin, the tissue was first incubated in RPMI 1640 with 1.2 U/ml dispase II (Roche, Indianapolis, IN) at 4 °C for 8 h prior to the separation of the epidermal layer. After separation, the epidermal layer was cut into pieces and digested with 0.25% trypsin-EDTA (Gibco) containing 0.25 mg/ml DNase I (Sigma DN25) for 30 min at 37°C in a humified incubator. After that, the epidermal layer was passed through a 70 µm cell strainer by grinding, and then washed with PBS.

For bulk cultures, cells were sorted and co-cultured with MS5 for 14 days (100 cells per well). For the single cell cultures, single YSMPs were sorted and cultured in individual wells. Wells were observed under a light microscope and those with hematopoietic

clusters (more than 50 round hematopoietic-like cells) at day 10 were counted as positive wells. At the end of the cultures, all cells including MS5 stromal cells in the selected positive wells were collected for further analyses with FACS to determine the lineage differentiation potential.

Instrument

BD ARIA 2

Software

BD FACSDIVA v8.0 and FlowJo X 10.0.7r2

Cell population abundance

The proportion of CD45+CD235a- cells in YS ranges from 10% to 30%. The cells, sorted by BD ARIA 2, were judged by morphology under microscope to assess their conditions, with the ones in good condition being picked by mouth pipette and directly placed into lysis buffer for single cell RNA library construction. The CD34+CD44+ population accounts for about 40% of CD45+ cells in YS.

Gating strategy

7AAD-CD45+CD235a-for STRT-seq; 7AAD- for 10x Genomics; 7AAD-CD45+CD34+CD44+ or 7AAD-CD45+CD34-CD44- for functional assay.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.