Tracking Early Mammalian Organogenesis – Prediction and Validation of Differentiation Trajectories at Whole Organism Scale

Ivan Imaz-Rosshandler[1,2,3,*], Christina Rode[4,*], Carolina Guibentif[5,*], Luke T. G. Harland[1,2,*], Mai-Linh N. Ton[1,2], Parashar Dhapola[10], Daniel Keitley[6], Ricard Argelaguet[11,12], Fernando J. Calero-Nieto[2], Jennifer Nichols[2], John C. Marioni[7,8,9,‡], Marella F. T. R. de Bruijn[4,‡], Berthold Göttgens[1,2,‡]

[1]Department of Haematology, University of Cambridge, Cambridge CB2 0RE, UK
[2]Wellcome-Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge CB2 0AW, UK
[3]MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK
[4]MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DS, UK
[5]Department of Microbiology and Immunology, University of Gothenburg, Gothenburg, Sweden
[6]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK
[7]Wellcome Sanger Institute, Wellcome Genome Campus, Saffron Walden CB10 1SA, UK
[8]European Molecular Biology Laboratory, European Bioinformatics Institute, Saffron Walden CB10 1SA, UK
[9]Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge CB2 0RE, UK
[10]Division of Molecular Hematology, Lund Stem Cell Center, Lund University, Sweden
[11]Epigenetics Programme, Babraham Institute, Cambridge CB22 3AT, UK
[12]Altos Labs Cambridge Institute, Granta Park, Cambridge, CB21 6GP, UK

*****Equal contribution authors
‡Co-corresponding authors

Summary Statement

Here we generate a transcriptional roadmap of mouse gastrulation to organogenesis (E6.5-9.5) and utilize it uncover complex waves of blood and endothelial development and interpret state-fate transplantation experiments.

Abstract

Early organogenesis represents a key step in animal development, where pluripotent cells diversify to initiate organ formation. Here, we sampled 300,000 single cell transcriptomes from mouse embryos between E8.5 and E9.5 in 6-hour intervals and combined this new dataset with our previous atlas (E6.5-E8.5) to produce a densely sampled time course of >400,000 cells from early gastrulation to organogenesis. Computational lineage reconstruction identified complex waves of blood and endothelial development, including a new programme for somite-derived endothelium. We also dissected the E7.5 primitive streak into four adjacent regions, performed scRNA-Seq and predicted cell fates computationally. Finally, we defined developmental state/fate relationships by combining orthotopic grafting, microscopic analysis and scRNA-Seq to transcriptionally determine cell fates of grafted primitive streak regions after 24h of *in vitro* embryo culture. Experimentally determined fate outcomes were in good agreement with computationally predicted fates, demonstrating how classical grafting experiments can be revisited to establish high-resolution cell state/fate relationships. Such interdisciplinary approaches will benefit future studies in developmental biology and guide the *in vitro* production of cells for organ regeneration and repair.

Introduction

Single cell transcriptomics has significantly contributed to our understanding of cell type diversity across species, organs and developmental processes. Efforts to build cell atlases from different model organisms include extensive transcriptomic profiling of embryonic development, with profiling of the mouse being particularly relevant given its broad use as a model for mammalian development. Independent efforts now provide coverage from embryonic days E3.5-E6.5, E6.5-E7.5, E4.5-E7.5, E6.5-E8.5, E6.5-E8.25, E9.5-13.5 and E10.5-E15.0 (Argelaguet et al., 2019; Cao et al., 2019; Chan et al., 2019; Grosswendt et al., 2020; Han et al., 2018; He et al., 2020; Ibarra-Soria et al., 2018; Lescroart et al., 2018; Mittnenzweig et al., 2021; Mohammed et al., 2017; Pijuan-Sala et al., 2019; Scialdone et al., 2016), complemented by detailed analysis of specific organs such as the brain (La Manno et al., 2021) and the heart (de Soysa et al., 2019) or with emphasis on specific germ layers (Nowotschin et al., 2019). Combining datasets to provide an integrated time course over a longer timespan has been achieved by overcoming the challenge of integrating different sequencing technologies (Qiu et al., 2022). Of note, single cell atlases of normal development have rapidly been utilised as a reference to interpret mutant phenotypes, providing new insights into the cellular and molecular processes controlled by key developmental regulators [Montague et al., 2018], (Barile et al., 2021; Clark et al., 2022;

Grosswendt et al., 2020; Guibentif et al., 2021; Mittnenzweig et al., 2021; Pijuan-Sala et al., 2019).

Integrating and annotating transcriptomic profiles in different contexts (e.g., across species, technologies, experiments, time points, etc.) represents a foundational step towards the construction and leveraging of cell atlases but poses substantial challenges due to the difficulty of distinguishing between highly similar or transitional cell populations arising along complex differentiation trajectories. Without lineage tracing, computational inference of differentiation trajectories may provide useful information for understanding the dynamics of cellular diversification but need to be interpreted with caution. The increasingly large number of methods for trajectory reconstruction furthermore highlights the need for benchmarking strategies (Saelens et al., 2019). Many trajectory inference methods are based on the concept of pseudotime, a latent dimension representing the transition between progenitors and differentiating cells as a function of transcriptional similarity. Though related, pseudotime and experimental time are different concepts. Thus, incorporating real time should improve reconstruction of developmental processes (Tritschler et al., 2019). Inspired by the Waddington landscape (Conrad Hal Waddington, 1957) and Optimal Transport theory (Monge, 1781), the probabilistic framework Waddington-OT takes advantage of experimental time and stochastic modelling to estimate the coupling probabilities of cells between consecutive time points and to reconstruct differentiation trajectories (Schiebinger et al., 2019).

Here we report a densely sampled scRNA-Seq atlas covering mouse development from E6.5 to E9.5 in 6h intervals. This new atlas includes our 116,000 previously published E6.5-E8.5 transcriptomes that cover gastrulation and the initial phase of early organogenesis, complemented by 314,000 new E8.5-E9.5 transcriptomes that bridge a critical gap in development not captured in existing datasets, i.e., the period of major morphological and organogenesis changes that occur between E8.5 and E9.5. This includes embryo turning, emergence of definitive-type Haematopoietic cells, and initiation of the heartbeat and circulation. This new combined E6.5 to E9.5 atlas delivers the most comprehensive transcriptome dataset for mammalian gastrulation and early organogenesis to date. To address the challenge of reconstructing and annotating cell lineages, we combined expert curation with a variety of computational methodologies to generate a resource of broad utility for the developmental biology community. We investigated further the intricate process of haemato-endothelial development, highlighting the multiple origins of endothelial cells and the formation of blood progenitors in asynchronous waves in spatially distinct sites. Moving beyond atlas generation, we also utilised precise embryo dissections to profile spatially-defined areas of the developing embryo. Finally, computational cell fate predictions were contrasted with cell fates observed in state-of-the-art cell grafting experiments, where individual E7.5 primitive streak segments were orthotopically grafted into recipient embryos and resulting cell fates analysed after 24 hours of culture by spatial and scRNA-Seq analysis. Our study provides a blueprint for cell state/cell fate analysis during key stages of mammalian development, paving the way for future interdisciplinary studies of cell fates and origins.

Results

A densely sampled scRNA-Seq atlas from E6.5 to E9.5 of mouse development

We previously reported a scRNA-Seq atlas covering mouse gastrulation and the early initiation of organogenesis between E6.5 to E8.5 (Pijuan-Sala et al., 2019). To capture the critical organogenesis period between E8.5 and E9.5 we undertook a new time-course experiment and integrated the new sampling time points (Fig.1a). Thus, 116,312 cells distributed across 9 time points from the original atlas were complemented with 314,027 new cells distributed across four new time points (E8.75-E9.5) as well as one overlapping time point (E8.5) to facilitate data integration (Fig.1b). Combined, the new extended atlas, ranging from E6.5 to E9.5, contains 430,339 cells across 13 time points spanning 3 days of mouse development (Fig.1a-c).

All embryos for the new dataset were dissected prior to droplet capture, to (i) profile yolk sac (YS) cells separately and (ii) dissect the embryo proper to provide single cell data anchored by anatomical location (Fig.1a). As illustrated below, sequencing the various tissue segments independently aids the disambiguation of transcriptionally similar cell states (Fig.1d-f). A combination of computational approaches and manual curation (see Methods, Fig.S1-S6 and Supplementary Table S1-S2) allowed us to define 88 major cell states, more than double the number identified in the previous atlas and reflecting the rapid diversification of cell states during the 24h time window from E8.5 to E9.5 (Fig.1g and Fig. 2). The new combined dataset has been made freely available through a user-friendly web portal to be explored and leveraged by the wider scientific community (see Data availability).

Inference of haemato-endothelial development reveals independent intra- and extra-embryonic trajectories

The computational reconstruction of developmental processes using single cell transcriptomics remains a major challenge despite the large number of available methods.

The time-course experimental design of this atlas provides the advantage of incorporating developmental stage information thus permitting the use of methods such as Waddington-Optimal Transport (W-OT) that anchor inferred developmental progression in real time rather than pseudotime (Schiebinger et al., 2019). In W-OT, developmental processes are modelled using a probabilistic framework that allows inferring ancestor and descendant probabilities between cells sampled at consecutive time points, thus allowing for time series analysis. We applied W-OT to the developing haematopoietic and endothelial systems, which are needed early during development to enable effective circulation once the heart starts beating around E8.25-E8.5. Blood and endothelium arise in waves, at least some of which are thought to entail shared haematoendothelial progenitors (reviewed in (de Bruijn and Dzierzak, 2017; Elsaid et al., 2020)). The first blood cells are so-called primitive erythrocytes, arising at E7.5 from mesodermal thickenings in the prospective blood islands of the YS. Next is a wave of definitive-type blood progenitors that originate from the YS vasculature from E8.5. The haematopoietic stem cell (HSC) lineage is the last to emerge from the endothelium of major arteries of the embryo, the dorsal aorta and vitelline and umbilical arteries, starting from E9.5. Both YS definitive-type blood progenitors and HSCs are derived from a specialised subset of endothelium, the haemogenic endothelium, through a so-called endothelial-to-haematopoietic transition. The extended mouse atlas covers the two YS-derived waves of blood emergence, as well as a variety of extra and intra-embryonic populations of endothelial cells, providing a unique opportunity to explore the highly complex emergence and diversification of the initial haemato-endothelial landscape during mammalian development.

To encompass all haemato-endothelial lineages, a W-OT fate matrix was computed for all blood and endothelial cell populations in the landscape (erythroid, megakaryocyte, megakaryocyte-erythroid progenitors (MEP), erythroid-myeloid progenitors (EMP), blood progenitors (BP), haemato-endothelial progenitors (HEP), endothelial populations from YS (YS EC), embryo proper (EP EC), and allantois (allantois EC), as well as venous endothelium and endocardium) leaving all other cells in the extended atlas grouped as “other fate” (Fig.3a). In brief, a W-OT fate matrix is a transition probability matrix from all cells to several target cell sets at a given time point, commonly the end point of the time course experiment (here E9.25 and E9.5). Following this strategy, three main W-OT studies were performed. As a first proof-of-concept, only YS blood and endothelial cells were considered with E9.25-E9.5 cells from these two populations defined as targets (Fig.3), which recovered the known divergence between the primitive and definitive-type YS waves over time. Specifically, the W-OT inferred primitive wave mainly generates nucleated primitive erythrocytes, as well as a small number of macrophage and megakaryocyte progenitors (red trajectory in Fig.3b) (Tober et al., 2007). The second, definitive-type YS wave starts with the emergence of EMPs followed by MEPs(McGrath et al., 2015) arising from YS HE as highlighted by a collection of known markers for all these populations (blue trajectory in Fig. 3b, and gene marker inspection in Fig.3c and Fig.S4-6). To further characterize gene expression changes along the inferred primitive and definitive hematopoietic trajectories in the WOT YS landscape, we employed Slingshot (Street et al., 2018) to infer lineage pseudotimes and utilized Tradeseq (Van den Berge et al., 2020) to identify gene changes associated with each trajectory (Fig.S7a-b and Table S3). It is worth noting that cells expressing lymphoid and microglial-like progenitor markers were also detected during the YS definitive wave, albeit at a low frequency (Fig. S8).

For the second W-OT analysis, we wanted to explore the potential origins of endothelial cells together with blood cells, and thus considered the complete haemato-endothelial progenitor population, which resulted in a considerably more complex hematoendothelial landscape as endothelial cells are found in multiple regions across the embryo (Fig.4a-b and Fig.S9). For our third WOT study, we merged blood cells with ‘other fates’ in the W-OT matrix, allowing us to focus explicitly on the complex endothelial landscape (third W-OT analysis). It is worth noting, that due to high levels of transcriptional convergence at endpoints during blood and endothelial differentiation, it is technically challenging to interrogate differentiation towards specific terminal cell types. Therefore, we opted to include multiple terminal blood and endothelial cell states in our WOT analyses, allowing us to explore differentiation trajectories that produce molecularly convergent cell states from transcriptionally distinct starting populations. Importantly, the comprehensive metadata associated with our dataset makes it straightforward for the wider research community to explore alternative data analysis options that may be better aligned with their research questions.

Altogether, our WOT landscapes revealed at least three putative endothelial differentiation trajectories that occur in different spatial locations and time windows during embryogenesis (Fig.S10a and Fig.S11a-c). The first wave of endothelial production initiates early in the YS (Fig.S7,c and Fig.S11c). By contrast, the second wave(s) occurs in posterior regions of the embryo at intermediate timepoints (Fig.S11b,e) and the final wave initiates at later stages and occurs in anterior/medial sections (Fig.S11a,d). Dynamic gene expression changes were explored with Tradeseq (Van den Berge et al., 2020) along these inferred trajectories (Fig.S7c, Fig.S11d-g) and detailed results of these analyses are presented in Supplementary Tables S3-S5. In addition to revealing multiple spatiotemporally separated trajectories with different gene expression patterns, these analyses revealed a strong connection between somitic tissues and EP ECs, highlighted by differences in the UMAP coloured by log odds between the first landscape and the second/third landscapes (Fig.3a versus Fig.4a and Fig.S10a). This connection led to the identification of the endotome, a cell population not previously reported in mammalian embryos, which will be explored in the following section.

Finally, to identify distinctions between downstream endothelial populations we performed differential expression analysis (Fig.S10c). As expected, YS endothelium expressed *Lyve1*, venous endothelium showed high expression of genes associated with vasculogenesis and angiogenesis such as *Clec1b and Cldn5* and the endocardium expressed markers indicative of Bmp (*Id1* and *Id3*) and Notch signalling (*Hey*). By contrast, embryo proper endothelium was relatively immature as indicated by lower levels of endothelial genes such as *Cdh5*, *Pecam1* and *Dlk1.* Furthermore, embryo proper endothelial cells that arise in different anatomical locations expressed distinct sets of Hox genes (Fig.S10d).

A previously unrecognised developmental trajectory involving novel intraembryonic endotome-like cells

A small but clear subset of endothelium appeared to be part of a continuous differentiation trajectory originating from cells transcriptionally similar to endotome, a somitic subset defined in zebrafish embryos (Nguyen et al., 2014) (Fig.4b and Fig.S11a,d). Added credence is given to this transcriptionally defined trajectory because both the contributing endotome and endothelial cells are derived from the same portions of the embryo (anterior/medial regions), demonstrating the utility of sub dissecting embryos before generating cell suspensions for sequencing. Furthermore, the mouse endotome cells are characterised by the expression of marker genes including *Cxcl12, Pax3, Meox1*, *Foxc2*, *Pdgfra* and *Alcam,* consistent with their recent discovery in zebrafish (Murayama et al., 2023; Nguyen et al., 2014; Tani et al., 2020) as well as by *Hlf expression* which has been associated with intra-aortic haematopoietic clusters (Yokomizo et al., 2019)(Fig.3c). Further characterization of the mouse endotome-like population (Fig. S12 and Supplementary Tables S6 and S7) reveals it also expresses genes related to skeletal system development (*Snai1, Sox9, Tbx1, Tcf15, Twist1, Irx5, Pkdcc, Foxp1*), vasculature development (*Apoe ,Nr2f2, Col3a1, Col4a1, Ednra*), muscle formation (*Fzd2, Six1, Tcf15, Twist1*) and VEGF signalling (*Vegfb, Vegfc*).

In the zebrafish embryo, endotome cells migrate towards the dorsal aorta and differentiate into endothelial cells that contribute to the niche for the emerging HSCs (Nguyen et al., 2014). Similarly, earlier studies on chick embryos showed a somite-derived cell population that contributes non-hemogenic endothelium, replenishing vascular cells that have undergone an endothelial-to-haematopoietic transition (EHT) (Pardanaud et al., 1996; Pouget et al., 2006; Sato et al., 2008). More recently, angioblast-like cells with similarities to somitic mesoderm-derived angioblast and endotome cells of chicken and zebrafish embryos were also reported in pluripotent stem cell-derived human axioloids (Yamanaka et al., 2023). In lower vertebrate models, the endotome was shown to originate at the ventral–posterior region of the sclerotome(Tani et al., 2020) which agrees with the transcriptional neighbours for the endotome population we discover here in mouse embryos. It took our densely-sampled and sub dissected scRNA-Seq approach to discover a likely connection between the newly-discovered mouse endotome and intraembryonic endothelium (Fig. 3a-c, Fig. S10a and Fig. S11a).

Intriguingly, in the Force Atlas representation of the hematoendothelial landscape a subset of endotome-like cells present in later stage embryos (E9.25-E9.5) are placed next to EMP blood progenitors (Fig. 4d, red box). Additionally, these cells cluster with EMPs (Fig. 4di, cluster 20) when performing louvian clustering using the top 50 principal components from the hematoendothelial landscape. However, this association is only visually highlighted when a force directed layout is generated on a subset of the atlas. By contrast, clustering, cell type annotation and UMAP embedding over the entire atlas, as well as the endothelial only landscape, demonstrates these cells have a transcriptional identity aligned with earlier stage endotome cells that are transcriptionally distinctive compared to EMPs (Fig. 4dii versus Fig.2 and Fig. S11a)."

Further exploration of these later stage endotome cells reveal they lack a clear haemogenic signature (Fig.4e and Fig.S13) and instead likely represent progenitors of endothelium and/or vascular mural and connective tissue cells, so-called vascular associated cells (VACs, Fig.4b-e, Fig. S11a and Fig.S13). Our densely sampled and regionally sub dissected atlas therefore suggests substantial transcriptional plasticity and developmental potential within mesodermal cells involved in the formation of intraembryonic blood vessels. Future studies will be required to investigate how this plasticity extends to the specification of intraembryonic HSCs within the vascular niche at E10.5.

A gradient of heterogeneous molecular states along the anterior-posterior axis of the primitive streak

The complexity of this atlas is both an opportunity and a challenge for computational inference of developmental processes at the whole organism scale. Importantly, the atlas also provides a framework for complementary experiments which can in turn expand the atlas' impact. Of particular interest to developmental biologists is the question of whether trajectory reconstruction analysis can unveil the timing at which cells start to differentiate towards either one or a set of specific cell fates. While two-dimensional representations of single cell expression data are often taken as a starting point, such representations lend themselves to overinterpretation, thus highlighting the need for experimental validation. We therefore complemented our extended atlas with (i) scRNA-Seq of carefully dissected subregions of the primitive streak at E7.5 and (ii) orthotopic transplantation of the same streak regions into recipient embryos followed by extended *in vitro* embryo culture and both microscopic as well as molecular fate analysis.

Previous primitive streak cell transplant and labelling experiments produced a fate map of cells along the anterior-posterior axis of the gastrulating mouse embryo revealing reproducible region-specific contribution to all the major cell lineages of the embryo(Kinder et al., 1999). To link these data to the extended mouse atlas, we revisited these experiments and dissected the primitive streak of E7.5 (early bud, EB) mouse embryos into four sequential regions labelled A to D (with A being most posterior and D being most anterior), generated single cell suspensions and profiled the regionally dissected cells using a modified Smart-seq2 scRNA-seq protocol providing deep coverage of each cell (Fig.5a, Fig.S9). We mapped these primitive streak cells onto the extended atlas and employed label transfer to assign cell types (Methods). Approximately 27% of the isolated cells mapped to pluripotent cell types such as epiblast or primitive streak, while 46% mapped to mesoderm, 24% to ectoderm and 3% to endoderm lineages (Fig 5b).

There were clear differences across primitive streak regions (Fig.5c,e). For instance, only cells from the posterior-most region A mapped to lateral plate mesoderm (LPM) and primordial germ cells (PGCs) as well as expressing known posterior markers such as *Msx2(Sun et al., 2016)* (expressed in the allantois and known to be involved in PGC migration), and *Bmp4* (Fig 5c,d,e, Fig S15). In contrast, cells from the most anterior region D mapped to the node and expressed several brain and cilia-associated genes, including *Riiad1* and *Pifo* (Fig 5e, Fig S15). Cells from regions B and C were similar to each other and mapped to paraxial, presomitic and somitic mesoderm as well as neural and gut cells. However, these cell types were not unique to regions B and C and were found across the entire posterior-anterior axis of the primitive streak (Fig 5c,d, Fig S15). Consistently, *Robo1*, involved in central nervous system and heart development, *Fzd10*, associated with neural induction, *Prickle1*, a limb development gene, and *Gas1*, a somitic gene also present in LPM, all showed no bias for a particular segment and were expressed throughout the entire length of the primitive streak (Fig 5e). Mesoderm-associated populations such as paraxial, presomitic and somitic mesoderm can be seen as transcriptionally-defined subsets during later development (Guibentif et al., 2021), but these were not yet observed in this EB stage-derived primitive streak dataset. In summary, transcriptional signatures of cells obtained from anterior to posterior primitive streak regions were notably heterogeneous and already showed a bias towards particular molecular states.

Cell fate analysis of orthotopic primitive streak grafts shows concordance between predicted and observed cells fates

Having identified molecular differences along the anterior to posterior axis of the primitive streak, we next used the extended cell atlas to predict the fates of regions A to D computationally and examined how these compare with experimentally determined cell fates. To this end, using computational fate inference, the closest neighbouring cells of the E7.5 primitive streak cells were identified in the atlas, and downstream fates were inferred based on the W-OT framework outlined above. Based on an initial survey of all predicted fates, a subset of major predictions was selected and visualised in so-called fate plots (Fig.6c and Fig S16e), which revealed associations between predicted cell fates and specific portions of the primitive streak. Cells from the posterior-most region A, for example, were primarily associated with mesodermal fates such as allantois, somites and the non-neural ectoderm, whereas the notochord and neural tube fates clearly favoured the most anterior region D (Fig 6c and Fig S16e).

To obtain experimental cell fate data, E7.5 embryos were orthotopically grafted with primitive streak regions A to D, cultured for 24h and the fate contributions of the donor regions analysed at E8.25 by microscopic assessment and scRNA-seq (Fig.6a). To facilitate analysis of cell fates, transgenic embryos carrying a ubiquitous membrane-bound tdTomato (mTom, (Muzumdar et al., 2007)[were used as graft donors. Cultured whole embryos and sections were immunostained (Fig 6b, Fig S16b) and careful observation of the location of mTom+ cells showed clear differences in tissue distribution between the donor primitive streak regions (summarised in Fig 6d and detailed in Suppl.Table S8). Notably, region A of the primitive streak contributed predominantly to the most posterior vasculature of the embryo, including the paired dorsal aortae and so-called vessel-of-confluence (VoC, (Rodriguez et al., 2017)), and to the allantois and some YS cells. Regions B and C also contributed to the paired dorsal aortae, as well as to somites, and to the LPM. Region D contributed mostly to the neural tube and notochord (Fig 6d). To analyse the cellular contribution of the grafted primitive streak regions at the transcriptional level, we flow-sorted the mTom+ single cells from additional batches of cultured embryos and performed scRNA-seq (Smart-seq2). After low-level pre-processing, cells were mapped onto the extended atlas as before. Compared to the cells directly isolated from the E7.5 streak, donor primitive streak-derived cells in the E8.25 cultured embryos mapped to almost twice as many cell types (38 vs 63, respectively, compare Fig 5b and Fig 6e), highlighting their advanced differentiation and diversification.

Comparison between the microscopically and transcriptionally-observed fates of the grafted primitive streak regions showed good concordance with only subtle differences in the contribution to distinct lineages (Fig 6d and 6f; microscopically-observed fates are based on embryo counts, while transcriptionally-observed fates are based on individual cell counts). Importantly, comparison between observed and predicted fates showed largely concordant patterns across computationally inferred fates, molecularly (scRNA-Seq) mapped fates and microscopically assigned fates (Fig 6c,d,f). It is worth noting that scRNA-Seq based analysis of fates has substantially enhanced granularity over microscopic analysis, with fate assignment based on our extended atlas for example providing up to 88 different cell types/molecular states for fine-grained annotation. In summary, combining multi-disciplinary approaches, we established primitive streak cell ‘end fates’ at a single cell resolution. Placing these in the context of the extended mouse gastrulation and organogenesis atlas allowed establishing the proof-of-concept for a fate-predictive algorithm, able to forecast fate trajectories of individual primitive streak cells.

Discussion

To realise the full potential of single cell atlas efforts for developmental biology research, molecular profiling datasets need to (i) provide sufficient sampling density to enable a time-series capable of capturing rapid developmental processes, and (ii) contain enough single cells sequenced at reasonable depth, well annotated and provided to the broader community through user-friendly web portals. Here we report such a resource covering the critical stages of mouse gastrulation and early organogenesis, from E6.5 to E9.5 sampled every 6 hours in 13 individual time steps. This dataset transforms our previous effort by more than tripling the number of cells and more than doubling the number of defined cell states. Furthermore, we revisit classical embryo grafting experiments with transgenic and single cell analysis tools, providing a foundation for future efforts aiming to fully reconstruct cell lineage trees.

Previous single cell atlas efforts from us and others placed observed molecular states into the context of existing knowledge of mouse development (reviewed by (Tam and Ho, 2020) and further integrated by (Qiu et al., 2022; Qiu et al., 2023)), altogether contributing to a deeper understanding of the molecular heterogeneity accompanying lineage differentiation during mouse embryogenesis. Our present study advances beyond these efforts in several impactful ways. By incorporating WOT, spatial information and encompassing the developing yolk-sac region, our analysis achieves a comprehensive characterization of blood and endothelial formation spanning the critical developmental window of E8.5-E9.5. Our dense sampling, deep sequencing and regional sub dissection of embryos allowed us to identify cell states not previously observed in early mouse development, such as the endotome and VACs. Moreover, our trajectory analyses suggest that the endotome-like progenitor populations is related to downstream endothelial cells, suggesting a greater intricacy in endothelial differentiation than previously appreciated.

Similar cell types have previously been described in zebrafish and chicken embryos and recently in pluripotent stem cell-derived human axioloids (Nguyen et al., 2014; Pardanaud et al., 1996; Pouget et al., 2006; Sato et al., 2008; Yamanaka et al., 2023), highlighting the value and complementarity of using a variety of different model organisms. Moreover, the zebrafish studies suggested that endotome and VACs may contribute to a cellular niche that promotes intraembryonic formation of blood stem/progenitor cells (Nguyen et al., 2014). In the mouse, the best-understood site of HSC emergence is the haemogenic endothelium of the dorsal aorta where pro-HSCs are generated from E9.5 (Rybtsov et al., 2014). These haemogenic endothelial cells are lateral plate-/ splanchnopleuric mesoderm- derived and were shown to be replaced by somite-derived endothelial cells concomitant with the extinction of HSC generation (Pouget et al., 2006). Interestingly, another report suggests that 7 days later during mouse gestation, endothelial cells in fetal bone marrow undergo haemogenic transdifferentiation and produce blood progenitor and differentiated cells (Yvernogeau et al., 2019). Pax3-Cre lineage tracing further suggested a somitic origin of those haemogenic cells.

Intriguingly, more recent unpublished Tbx6+ (Yvernogeau et al., 2020) and Pax3+ lineage tracing studies (Lupu et al., 2022) have been performed at embryonic timepoints (E7.5-E11.5) exploring developing anatomical regions (anterior/medial embryonic sections) that overlap with the emergence of endotome, VACs and endotome derived endothelial cells in our extended atlas. These studies suggest somitic mesodermal precursors give rise to endothelial cells of the limb and trunk region in developing mouse embryos, as well as stromal cells juxtaposed to the dorsal aorta, which may be like the VAC population we identify in this present study. In these studies, somite derived endothelial cells make only a minimal contribution to dorsal aorta endothelium by E9.5-10.5. Future work will need to dissect the underlying connection we identify here between a putative endotome cell population, VACs and embryo proper endothelium, such studies will (i) enhance our understanding of early blood and endothelium development, (ii) likely reveal principles relevant to cell plasticity and potential in other developmental contexts, and (iii) provide new mechanistic insights that could be exploited to control differentiation, for example for directed differentiation of pluripotent cells for cell therapy applications.

By combining scRNA-Seq with classical embryo grafting experiments, we show how single cell atlases provide a powerful resource to revisit classical concepts of developmental biology. The results presented in this study are limited in scope but already illustrate how time course-guided trajectory reconstruction performed with computational methods such as W-OT represents a promising approach to dissect complex developmental processes. Our findings also support substantial fate bias present in the E7.5 primitive streak, based on deeply sequenced single cell profiles. Since we did not perform heterotopic transplants, our experiments could not assess fate plasticity nor full fate potential. Moreover, deeper and more comprehensive transplant experiments should be designed around new experimental tools not available when this work was performed, including extended *in vitro* embryo culture(Aguilera-Castrejon et al., 2021), as well as single cell barcoding to permit reconstruction of single cell phylogenies(Bowling et al., 2020).

A future thereby emerges where a confluence of complementary technologies will transform our understanding of early mouse development to a level previously only attained with non-mammalian organisms. Building on data resources such as the one reported here, mechanistic insights will continue to require perturbation experiments, with the important proviso that carefully chosen experimental perturbations have the added benefit to provide new insights into disease processes, in particular congenital defects associated with mutations in developmental regulator genes. Our extended mouse gastrulation atlas provides the developmental biology community with a significant new resource to probe novel hypotheses concerning cell fate acquisition and lineage commitment during embryogenesis.

Materials and methods

E8.5-E9.5 embryos collection for the extended atlas

Mouse embryos were collected under the project licence number PPL 70/8406. Animals used in this study were 6-10 week-old females, maintained on a lighting regime of 14h light and 10h darkness with food and water supplied *ad libitum*. Following wildtype C57BL/6 matings, females were killed by cervical dislocation at E8.5, E8.75, E9.0, E9.25 and E9.5. The uteri were collected into PBS with 2% heat-inactivated FCS on ice and the embryos were immediately dissected and processed for scRNA-seq. For each timepoint, four embryos were selected based on morphology and somite counts, in order to span the range expected for the given timepoint according to [Theiler, K., 1989] and processed individually. The exception is the E9.5 time-point, where embryos were smaller than expected and only two embryos were collected with lower somite numbers. The yolk sac was systematically separated from the rest of the embryo and processed as a separate sample. Of the four selected embryos at each time-point, two were partitioned in defined anterior-posterior sections dissociated as separate samples, and two were dissociated as bulk and the suspension then divided into two separate samples for 10X RNA-Seq analysis. For the E8.5 partitioned embryos, they were divided into two halves with the cut being made at the 4th somite level (i.e., right before the 4th somite level). The anterior portion (including headfolds, branchial arches and heart rudiment) and the posterior portion (incl. allantois, hindgut, primitive streak) were processed as individual samples. For E8.5 bulk-dissociated embryos, the single-cell suspension was divided into two separate samples for 10X RNA-Seq analysis.

E8.75-E9.5 embryos were divided into 3 segments, with cuts made below the otic pit and below the heart (at the 10th-12th somite level). The anterior most portion (incl. brain structures anterior to rhombomere 6 and branchial bars), mid-portion (incl. the heart and remains of vitelline vessels) and posterior portion (incl. allantoic structures, hindgut and posterior most somites) were then singularized and further processed as separate samples. Single cell suspensions were prepared by incubating the samples with TrypLE Express dissociation reagent (Life Technologies) at 37 °C for 7 min under agitation and quenching in PBS with 10% heat-inactivated serum. The resulting single-cell suspension was washed and resuspended in PBS with 0.4% BSA and filtered through a Flowmi Tip Strainer with 40 μm porosity (ThermoFisher Scientific, # 136800040). Cell counts were then assessed with a haemocytometer. Single-cell RNA-seq libraries were generated using the 10X Genomics Chromium system (version 3 chemistry), and samples were sequenced according to the manufacturer’s instructions on the Illumina NovaSeq 6000 platform.

Publicly available mouse gastrulation data

The mouse gastrulation atlas was processed exactly as described in [Pijuan Sala et al., 2019]. This time-course experiment contains 116,312 cells distributed across 9 time points sampled across E6.5-E8.5 at six hours intervals. Only two samples of this dataset (where a sample is a single lane of a 10x Chromium chip) contained pooled embryos staged across several time points. Cells from these samples are denoted as ‘Mixed gastrulation’ in the metadata.

10x Genomics data Low level analysis

Raw reads were processed with Cell Ranger 3.1.0 using the mouse reference 1.2.0, mm10 (Ensembl 92) and default mapping arguments. The following steps of pre-processing were performed with R using the same functions, parameters and software versions broadly described in [Pijuan Sala et al., 2019]: swapped molecule removal, cell calling, quality control, normalisation, selection of highly variable genes, doublet removal, batch correction and stripped nucleus removal. Therefore, singularity containers available in https://github.com/MarioniLab/EmbryoTimecourse2018 provide the necessary software for reproducing these steps.

Generating an integrated atlas

Log transformed normalised counts obtained from (Pijuan-Sala et al., 2019) and those generated here were integrated into an extended time course experiment across mouse developmental stages E6.5-E9.5 (13 time points). Highly variable genes (HVGs) were calculated using ‘trendVar’ and ‘decomposeVar’ from the scran R package, with loess span of 0.05. Genes that had significantly higher variance than the fitted trend (Benjamini–Hochberg-corrected $P < 0.05$) were retained. Genes with mean log2 normalized count <10−3; genes on the Y chromosome; the gene *Xist*; and the reads mapping to the tdTomato construct (where applicable) were excluded. Hence, this expression matrix contained 23,972 genes and 430,339 cells. Batch correction with fastMNN function from scran (Lun et al., 2016) was performed as described above, resulting in 5,665 genes and 75 batch-corrected principal components.

Mapping stage and cell type annotations within the extended gastrulation atlas.

Metadata annotations such as embryonic developmental stages and cell types were assigned to the mixed time points (annotated as Mixed gastrulation in the metadata) and newly generated E8.5 samples respectively using a strategy based on fastMNN. In this approach, UMI counts from both, the reference and the query datasets, are merged, normalised and log transformed together. Then, highly variable genes and top principal components are computed to subsequently use fastMNN for re-scaling the PCA space from both datasets. The annotations from the reference metadata are assigned to the query data as the mode among k nearest neighbours (KNN) between the query and the reference PCA subspaces using queryX function from Biocneighbours. The number of nearest neighbours is chosen depending on the resolution of transferred annotations. Mixed gastrulation time points were allocated to embryonic developmental stages using the 30 nearest neighbours queried from the top 50 batch-corrected principal components from the E6.5-E8.5 reference dataset. New E8.5 cell type annotations were assigned using 10 nearest neighbours with respect to E8.5 cells from the reference atlas in the corresponding E8.5 subspace of the integrated batch corrected PCA described above.

Constructing optimal transport maps

The W-OT approach was conceived to model time course experiments of developmental processes as a generalisation of a stochastic process using unbalanced optimal transport theory. Thus, it allows estimating the coupling probabilities between cells of consecutive time points, while taking into account cell growth and death(Schiebinger et al., 2019). The transport maps of consecutive time points were constructed over the entire set of cells with Waddington-OT (wot 1.0.8.post1) using default settings except for skipping the dimension-reduction step, and instead using the batch-corrected principal components as input as well as three iterations for learning the cell growth rate. Embryonic stages E9.25 and E9.5 were collapsed into a single time point for computing the transport maps because the somite count (a more accurate measure of developmental stage), overlapped between these embryos.

Estimating the descendants from cell populations at E8.5

The transport maps above explained were used to estimate the full trajectories of every cell population present at E8.5. That is, for each cell population (i.e., Erythroid3) the coupling probabilities were used to reconstruct the sequences of ancestors and descendants distributions by pushing the cell set through the transport matrix backwards and forwards respectively. Cells were allocated as descendants from a cell population at E8.5 by selecting those with maximum mass across all trajectories in time points E8.75-E9.5.

Integrating the brain and gut development atlases to support cell type annotations

Mapping publicly available data from the atlas of brain development(La Manno et al., 2021), and both atlases of gut development (Nowotschin et al., 2019) against the extended gastrulation atlas was performed following the strategy based on fastMNN above mentioned. However, in this case the matrices of principal components from the query dataset were randomly split into subsets smaller than 10,000 cells for fastMNN and merged back to generate a single mapping output.

Expansion and refinement of cell population annotations.

A combination of complementary strategies was used for defining final cell type annotations (Fig S1). First, cell type annotations from(Pijuan-Sala et al., 2019) were transferred within overlapping time points (E8.5) and cell descendants were estimated as described above (Supp. Figure 1a-b). The landscape was then split into two subsets, a mesodermal and an ectodermal-endodermal landscape (the latter including NMPs) (Fig S1c), and subsequently clustered using scanpy's implementation of Leiden's algorithm(Traag et al., 2019) with resolution 5 (Fig S1d). Highly variable genes and batch corrected principal components were recomputed on the subsets before clustering. Then, annotations from the original atlas expanded through mapping and estimation of cell descendants were manually refined by means of differential expression analysis between clusters (using findMarkers from scran's package version (Lun et al., 2016), as well as literature and visual inspection of gene markers resulting in major 88 cell type populations (Figure 2). Additionally, we utilized the FindAllMarkers (min_pct = 0.25, logfc_threshold = 0.25) function in Seurat (4.2.0)(Hao et al., 2021) to identify marker genes for all 88 cell types (Table S1).

Generating the landscapes of haemato-endothelial trajectories

The strategy for generating the haemato-endothelial landscapes was based on the so-called W-OT fate matrix. In brief, a W-OT fate matrix is a transition probability matrix (the rows of this matrix add up to 1) from all cells to a number of target cell sets at a given time point, commonly the ending point of the time course experiment. To fully cover the haemato-endothelium, three W-OT fates matrices were computed, one for each landscape presented in our results. One using E9.25-E9.5 YS-blood and YS endothelial cells as targets, another one with YS blood as well as both, YS and embryonic endothelial tissues, and a third one where only endothelial cell types were considered targets (Figures 2, 4 and S10 respectively). Then, to identify cells that form a differentiation trajectory towards any of the above mentioned cell fates, for every cell in the landscape the likelihood probabilities associated with these fates and the remaining cells at E9.25-E9.5 cell types (i.e., cells grouped as "Other fate") were compared using the log odds or ratio of probabilities. The log odds is defined as the logarithm of the ratio of probabilities of two different categorical and mutually exclusive outcomes $log(p/1 - p)$. These ratios provided a quantitative setting for more interpretable thresholds when selecting cells that potentially belong to a cell fate trajectory. These cells are then retained and used to generate new layouts including only potentially fated cells towards blood and endothelium. That is, selecting pluripotent and mesodermal cells above a reasonable log odds threshold, recomputing highly variable genes, correcting for batch effects and generating new force directed graphs from inferred trajectories. Cells with log odds > 0 were considered fate biased towards the set of selected population targets, such as in the proof of concept performed in Fig.3. Since the log odds is computed by summing over the cell probabilities for all population targets (here denoted as p), divided by the probability of all other fates grouped together (1-p); cells with negative values but close to 0, although with higher uncertainty might well be contributing to the population targets of interest. Thus, for more complex landscapes the threshold was lowered to log odds > -1 (Fig.4 and Fig.S10), in order to exclude all cells not associated with haematoendothelial fates with high confidence, while keeping a higher degree of uncertainty for those retained.

Visualisation

To generate the UMAP layout of the whole embryo, we used the top 50 batch corrected principal components to generate a BBKNN graph(Polanski et al., 2020) and then scanpy's implementation of UMAP, with parameter min_distance = 0.99. To generate the force directed layouts of the hematoendothelial landscapes, we recomputed highly variable genes for each subset as well as batch correction of PCA manifolds (We again retained the top 50 principal components). We then built a KNN graph (K=50) and used ForceAtlas2(Jacomy et al., 2014) implementation of forced directed layouts included in scanpy.

Slingshot trajectory inference

Slingshot (2.7.0)(Street et al., 2018) was used, in a semi-supervised fashion, to infer differentiation trajectories and produce average pseudotimes along various putative lineage trajectories that were highlighted during WOT when generating the YS landscape (primitive, YS definitive and YS endothelial trajectories) and the endothelial landscape (anterior/medial and posterior trajectories) using the getLineages and getCurves functions (default settings, starting clusters were provided). To getLineages the first 45 batch corrected principal components were provided as input data for the respective landscapes as well as cluster labels generated using the default Seurat (4.2.0,(Hao et al., 2021)) functions FindNeighbors (dims = 1:50, reduction = "PCA") and FindClusters (resolution = 3) for the respective landscapes. Cells that were members of relatively small clusters (< 100 cells) were excluded by setting their cluster value equal to -1 when running getLineages. Prior to running getLineages, supervised filtering of cells based on metadata (stage, anatomy and predicted anatomy) was performed. Predicted anatomy for cells in the hematoendothelial landscape was determined by performing label transfer from cells with specific anatomical labels (Anterior, Posterior, Anterior section, Medial section and Posterior section) to cells from EP sections using the embeddingKNN function from StabMAP(Ghazanfar et al., 2023) (type = "uniform_fixed", k_values = 5, cords = top 50 PCs). For the YS primitive blood trajectory inference, cells from stages E6.5-E8.25 were considered, for the YS definitive blood trajectory cells from stages E7.75-E9.5 were considered, for the YS endothelial trajectory inference cells from anatomy YS and pooled were considered and blood cells were excluded (EMP, MEP, Blood progenitors, Erythroid, Megakaryocyte progenitors). For the anterior/medial endothelial trajectory inference cells from stages E8.5-E9.5 were considered and cells from predicted anatomy Posterior section and Posterior were excluded. For the posterior endothelial trajectory inference cells from stages E7.5-E9.5 were considered and cells from predicted anatomy Anterior section, Medial section and Anterior were excluded. When multiple, similar lineages towards downstream blood or endothelial cells were predicted by getLineages, average pseudotimes were calculated using the slingAvgPseudotime function. These average pseudotimes were utilized in subsequent Tradseq analyses.

Tradeseq differential gene expression analyses

To identify genes with altered expression over the inferred trajectories the average pseudotimes were provided to Tradeseq (1.10.0, (Van den Berge et al., 2020)). Generalized additive models were fit (fitGAM, nknots = 6, cellWeights = 1) to genes that were highly variable (top 2000 highly variable genes were identifed using VariableFeatures function in Seurat (4.2.0)) amongst the cells that were part of the inferred trajectories. The associationTest function was then used to identify which of these genes had altered expression over the inferred trajectories. ComplexHeatmap (2.15.3, (Gu et al., 2016)) was utilized to visualize the expression patterns of the genes with the highest 300 waldStat scores (pValue < 0.01 and meanLogFC > 2). Metascape (Zhou et al., 2019) was used to identify GO terms that were enriched for clusters of genes that were associated with the anterior/medial and posterior endothelial differentiation trajectories (Table S4 and Table S5).

Cell communication analyses with CellChat

To identify predicted ligand-receptor interactions amongst cells that were present in the YS landscape, the hematoendothelial landscape and the anterior/medial sections of the cells in the hematoendothelial landscape, (Table S7) we utilized the computeCommunProb (default settings) and filterCommunication (min_cells = 10) in CellChat (1.6.1, (Jin et al., 2021))

Differential gene expression and Canonical Correlation Analysis of Endotome derived VACs

Subsequent to the construction of a complete haemato-endothelial landscape, louvain (Subelj and Bajec, 2011) clusters (as implemented in the R package igraph (Csardi Gabor, 2006)) were generated from the batch corrected principal components obtained as explained above (Fig.4c). Cluster 20 was of particular interest for investigating both shared and non-shared gene expression profiles between the anatomically distinct populations YS EMP and NYS EMP, and a subset of endotome cells that is placed next to them (Fig.4d-e). Differential gene expression analysis was performed using the function findMarkers from the scran package(Lun et al., 2016). The statistical test performed by findMarkers uses a general linear model and moderated t-statistics to perform differential expression, as implemented in the R package limma (Ritchie et al., 2015). The threshold established was FDR < 0.05 and LogFoldChange > 0.5. Furthermore, only genes among the top 20 rank genes are displayed in heatmaps (considering that some genes can have tied ranks, gene lists usually contain more than 20 genes). CCA as implemented in the R package CCA (Gittins, 1985) and (Kanti Mardia, 1979) was applied to mean metacell expression values of YS EMPs and the endotome derived VACs. Positively correlated genes were then identified by selecting only those with positive coefficients (canonical scores) over the two canonical variates and visualised using a scatter plot fitting a linear regression. Importantly, such analysis was performed using Metacells (Baran et al., 2019) to strengthen gene expression signals.

Metacells

We identified metacells (i.e., groups of cells that represent singular cell-states from single-cell data) with the goal of achieving a resolution that retains the continuous nature of differentiation trajectories while overcoming the sparsity issues of single-cell data. We adopted the SEACells implementation (Sitara Persad). Following the method guidelines, metacells were computed separately for each sample using approximately one metacell for every seventy-five cells. Following metacell identification, we regenerated the gene expression matrices summarised at the metacell level. Sample-specific count matrices were then concatenated and normalised together. Metacells were used exclusively to inspect the correlation between genes expressed in YS EMPs and endotome derived VACs, as well as the lack of a clear hemogenic signature. However, metacells were computed for all the haematoendothelial landscapes.

Smart-seq2 data low level analysis

Sequencing reads were aligned against the *mm10* genome following the *GRCm38.95* genome annotation (*Ensembl 95*) using *STAR* version 2.5 (Dobin et al., 2013) with --intronMotif and 2-pass option mapping for each sample separately (i.e., --twopassMode basic) to improve detection of spliced reads mapping to novel junctions. Sequence alignment map (SAM) files were then processed with *samtools-1.6* in order to generate the corresponding count matrix with *htseq-count* from *HTSeq*, version 0.12.4. Cells with total counts lower than 10,000 reads were filtered out from downstream analysis based on observed library size distributions. The mitochondrial fraction of reads per cell and library complexity were computed as part of the QC. Only two cells showed a mitochondrial fraction larger than 4% and the lowest number of genes detected was 3,113. Thus, no cells were excluded during QC. Size factors were computed with *scran* and log transformed normalised counts were obtained with the *logNormCounts* function from *scater* using size factors centred at unity prior to calculation of normalised expression values. Then, highly variable enes were extracted using *ModelVar* function from *scran*, according to a significant deviation above the mean-variance fitted trend (BH-corrected p<0.05). These genes were selected for computing a PCA. The top 50 PCs were retained for batch correction using *reducedMNN*, which is an updated version of *fastMNN*. Figure 4.3 shows the two-dimensional PCA, before and after batch correction. The resulting batch-corrected PCs were used for clustering and visualisation in downstream analysis.

Predicting transcriptional identity and cell fate of primitive streak cells by mapping them against the atlas

Following the strategy to transfer annotations from the atlas to other datasets previously described above, cell type annotations were assigned to primitive streak cells according to the 15 nearest neighbours between the two manifolds (query data and atlas reference). The annotation transfer was performed separately for each batch to avoid introducing an extra confounding factor when applying MNN batch correction. Also, prior to mapping against the atlas, genes with total counts lower than 100 were filtered out from the Smart-seq2 datasets due to the large differences in dropouts. The resulting annotations were projected onto the UMAP extended atlas embryo landscape by highlighting the closest cells on it. By retaining the closest cell in the atlas and selecting the cell fate with highest probability allocated in a W-OT fates matrix, we aimed to not only map the observed transcriptome at E7.5 but make predictions about likely cell fates at future timepoints (eg. E8.5 and E9.5).

Primitive streak dissections and processing for Smart-Seq

Mouse embryos were collected under the UK Home Office project licence number PP9552402. Animals used in this study were 6-16 week-old wild type females of a mixed (CBAxC57BL/6)F2 background, maintained on a lighting regime of 14h light and 10h darkness with food and water supplied *ad libitum*. Following timed mating crosses between transgenic mTmG (B6.129(Cg)-*Gt(ROSA)26Sor*$^{tm4(ACTB\text{-}tdTomato,\text{-}EGFP)Luo}$/J (Muzumdar et al., 2007) males and wildtype females, females were killed 7 days after a vaginal plug was observed via a Schedule 1 method. The uterine tissue and the decidua were cut away to extract the E7 embryo. The Reichert's membrane was peeled off and embryos were staged according to (Downs and Davies, 1993). Early bud (EB) stage embryos were taken for further analysis. For primitive streak dissections, the extra-embryonic part of the conceptus was cut off, the anterior and posterior embryonic portions were separated, and the posterior portion flattened out, taking care not to mix up the orientation. The posterior portion containing the primitive streak was cut at the halfway point and each half was halved again creating 4 similar-sized primitive streak segments. All embryo and primitive streak dissections were performed in Dulbecco's Phosphate Buffered Saline (PBS; with CaCl2 and MgCl2, Gibco) supplemented with 10% FCS (Gibco), 50U/ml penicillin and 50U/ml streptomycin.

For scRNA-seq of primitive streak cells, individual primitive streak regions were collected in separate Eppendorf tubes in FACS buffer (PBS without CaCl2 and MgCl2 (Gibco) supplemented with 10% FCS (Gibco), 50U/ml penicillin and 50U/ml streptomycin). After centrifugation at 200G for 5 min, tissues were resuspended in 200-250ul of TrypLE Express (Life Technologies) and incubated at 37C for 7 min with regular agitation and dissociated by pipetting. Cells were washed with 1ml of FACS buffer, centrifuged at 200G for 5min and resuspended in FACS buffer with Hoechst 33358 for sorting. In order to reduce cell loss from these small cell populations, wild type adult mouse thymocytes were added to the samples and were gated out based on their typical FSC-SSC profiles and absence of tdTomato expression. Each primitive streak region yielded on average 200 cells. Cells were kept on ice throughout the procedure. Individual cells were sorted directly into 96 well plates (Starlab E1403-1200) into 2.3ul of lysis buffer containing SUPERase-In RNase Inhibitor 20U/ul (Ambion, AM2694), 10% Triton X-100 (Sigma, 93443) and RNAse-free water. Plates were processed using a combination of Smart-Seq2(Picelli et al., 2014) and mcSCRB-seq (Bagnoli et al., 2018) For detailed protocol see (Sturgess et al., 2021) Pooled libraries were run on the Illumina HiSeq4000 at Cancer Research UK Cambridge Institute Genomics Core.

Primitive streak grafts and static culture

mTmG transgenic embryos were dissected to harvest 4 equal primitive streak regions as described above. Each segment was deposited into a 50ul dissection buffer drop on the inside lid of a 5cm dish and labelled. One larger drop was deposited in the centre into which up to 5 dissected wild type embryos of the same stage (EB) were transferred. Primitive streak regions were grafted into orthotopic positions in the wild type embryos using a pulled glass capillary needle attached to a mouth pipette. Cells from each donor region were divided over 2-3 recipient embryos with each recipient embryo receiving 50-100 primitive streak cells, mostly as a chain of cells rather than in suspension. Embryos were gently washed in a drop of culture media and gently transferred into individual wells of a 96 well Ultra-Low adhesive plate (Corning) into 200ul culture media. Culture media was 100% rat serum (Envigo, custom collected). Embryos were cultured at 37C, 5% CO2 for 23-26h.

Fate contribution analysis

Embryos were taken out of the incubator and assessed for (the beginnings of) a heartbeat. Only embryos that lacked deformities and had a heartbeat were used for further analysis by microscopy or scRNA-Seq to assess the fates of the grafted primitive streak cells. For microscopically observed fates, reporter gene expression (mTom) was enhanced (anti-RFP antibody) and embryonic vasculature (CD31) visualised by immunostaining. Wholemount embryos were imaged to assess overall contribution of mTom in relation to embryo anatomy. Embryos were next embedded and cryosectioned, followed by imaging of all areas with mTom+ contribution and careful analysis and scoring for detailed contribution to a selection of most anatomically distinct cell lineages (Suppl Table 1). For scRNA-seq, grafted, post-culture embryos were staged by somite pair counts and individual donor embryo and single cell suspensions prepared for sorting as described above for E7.5 primitive streak region. Single mTom+ cells from the grafted embryos were sorted into 2.3ul lysis buffer in 96 well plates and processed for Smart-Seq2 as above.

Data availability

This data has been made available at https://marionilab.github.io/ExtendedMouseAtlas/ where links to both, raw and processed data are found. Memory usage to process and analyse this data was optimised by means of *scarf*, an efficient single cell analysis framework [Dhapola et al., 2022].

Acknowledgements/Disclosure statements

We would like to thank all the people and the different funding instances contributing to this project. Work at Cambridge was supported by Wellcome Collaborative award 220379/B/20/Z (Cambridge Gastrulation Consortium), Wellcome award 221052/C/20/Z (Human Cell Atlas - development atlas extension), and Wellcome award 206328/Z/17/Z (Defining the Haematopoietic System through Integrated Multi-Scale Analysis). Work in the Gottgens group was supported by Wellcome, Bloodwise, MRC, CRUK, by core support grants from Wellcome to the Wellcome-MRC Cambridge Stem Cell Institute. We thank Sarah Kinston for technical assistance. Work in the Marioni group was supported by core funding from CRUK (C9545/A29580) and by the European Molecular Biology Laboratory. J.C.M. has been an employee of Genentech since September 2022. Work at Oxford in de Bruijn group was supported by programmes MC_UU_00016/2 (MdB and CR) and MC_UU_00029/5 (MdB) in the MRC Molecular Haematology Unit Core award. We thank Biomedical Services staff at Oxford for animal care. We would like to acknowledge Kevin Clark in the flow cytometry facility at the MRC WIMM for providing cell sorting services. The facility is supported by the MRC HIU; MRC MHU (MC_UU_12009); NIHR Oxford BRC; Kay Kendall Leukaemia Fund

(KKL1057), John Fell Fund (131/030 and 101/517), the EPA fund (CF182 and CF170) and by the MRC WIMM Strategic Alliance awards G0902418 and MC_UU_12025. We thank Christoffer Lagerholm and Jana Koth of the Wolfson Imaging Centre Oxford for their assistance. The centre is supported by the MRC via the WIMM Strategic Alliance (G0902418), the MHU (MC_UU_12009), the HIU (MC_UU_12010), the Wolfson Foundation (18272) and the Wellcome Trust (Micron 107457/Z/15Z) grants. I.I. was funded by the Wellcome Mathematical Genomics and Medicine Programme at the University of Cambridge (203942/Z/16/Z RDAG/426 RG86191). C.G. was funded by the Swedish Research Council (2017-06278) and by a Swedish Childhood Cancer Fund position grant (TJ2021-0009). M.-L.N.T. was funded by a Herchel Smith PhD Fellowship in Science. D.K. was funded by the Wellcome Mathematical Genomics and Medicine Programme at the University of Cambridge (PFZH/158 RG92770).

Author Contributions

J.N. set up mouse timed mating. C.G., M.-L.N.T. and J.N. performed embryo dissection for the atlas. C.G. and F.J.C., performed sample and sequencing library preparations for 10X scRNA-Seq. I.I., performed the majority of computational analyses except for Metacells, CellChat and the inferred trajectory analyses with Slingshot and Tradeseq. L.T.G.H performed CellChat and the inferred trajectory analyses with Slingshot and Tradeseq. R.A. performed Metacells analysis. I.I, C.G., M.-L.N.T., and D.K. annotated atlas cell types. I.I. and P.D., developed the shiny app with scarf in the backend. C.R. performed Primitive Streak dissections and sample preparation for Smart-Seq2 experiments, Primitive Streak grafts and static cultures. I.I., C.R. and L.T.G.H prepared the manuscript figures. J.C.M, M.B. and B.G. supervised the study. I.I., C.R., C.G., L.T.G.H, J.C.M, M.B. and B.G wrote the manuscript. All authors read and approved the final manuscript.

References

Aguilera-Castrejon, A., Oldak, B., Shani, T., Ghanem, N., Itzkovich, C., Slomovich, S., Tarazi, S., Bayerl, J., Chugaeva, V., Ayyash, M., et al. (2021). Ex utero mouse embryogenesis from pre-gastrulation to late organogenesis. *Nature* **593**, 119-124.

Argelaguet, R., Clark, S. J., Mohammed, H., Stapel, L. C., Krueger, C., Kapourani, C. A., Imaz-Rosshandler, I., Lohoff, T., Xiang, Y., Hanna, C. W., et al. (2019). Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* **576**, 487-491.

Bagnoli, J. W., Ziegenhain, C., Janjic, A., Wange, L. E., Vieth, B., Parekh, S., Geuder, J., Hellmann, I. and Enard, W. (2018). Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. *Nat Commun* **9**, 2937.

Baran, Y., Bercovich, A., Sebe-Pedros, A., Lubling, Y., Giladi, A., Chomsky, E., Meir, Z., Hoichman, M., Lifshitz, A. and Tanay, A. (2019). MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biol* **20**, 206.

Barile, M., Imaz-Rosshandler, I., Inzani, I., Ghazanfar, S., Nichols, J., Marioni, J. C., Guibentif, C. and Göttgens, B. (2021). Coordinated changes in gene expression kinetics underlie both mouse and human erythroid maturation. *Genome Biol* **22**, 197.

Bowling, S., Sritharan, D., Osorio, F. G., Nguyen, M., Cheung, P., Rodriguez-Fraticelli, A., Patel, S., Yuan, W. C., Fujiwara, Y., Li, B. E., et al. (2020). An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells. *Cell* **181**, 1410-1422.e1427.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F. J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502.

Chan, M. M., Smith, Z. D., Grosswendt, S., Kretzmer, H., Norman, T. M., Adamson, B., Jost, M., Quinn, J. J., Yang, D., Jones, M. G., et al. (2019). Molecular recording of mammalian embryogenesis. *Nature* **570**, 77-82.

Clark, S. J., Argelaguet, R., Lohoff, T., Krueger, F., Drage, D., Göttgens, B., Marioni, J. C., Nichols, J. and Reik, W. (2022). Single-cell multi-omics profiling links dynamic DNA methylation to cell fate decisions during mouse early organogenesis. *Genome Biol* **23**, 202.

Conrad Hal Waddington, H. K. (1957). *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology*: Allen & Unwin.

Csardi Gabor, N. T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems* **1695**.

de Bruijn, M. and Dzierzak, E. (2017). Runx transcription factors in the development and function of the definitive hematopoietic system. *Blood* **129**, 2061-2069.

de Soysa, T. Y., Ranade, S. S., Okawa, S., Ravichandran, S., Huang, Y., Salunga, H. T., Schricker, A., Del Sol, A., Gifford, C. A. and Srivastava, D. (2019). Single-cell analysis of cardiogenesis reveals basis for organ-level developmental defects. *Nature* **572**, 120-124.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21.

Downs, K. M. and Davies, T. (1993). Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development* **118**, 1255-1266.

Elsaid, R., Soares-da-Silva, F., Peixoto, M., Amiri, D., Mackowski, N., Pereira, P., Bandeira, A. and Cumano, A. (2020). Hematopoiesis: A Layered Organization Across Chordate Species. *Front Cell Dev Biol* **8**, 606642.

Ghazanfar, S., Guibentif, C. and Marioni, J. C. (2023). Stabilized mosaic single-cell data integration using unshared features. *Nat Biotechnol*.

Gittins, R. (1985). *Canonical analysis; a review with applications in ecology*: Springer-Verlag.

Grosswendt, S., Kretzmer, H., Smith, Z. D., Kumar, A. S., Hetzel, S., Wittler, L., Klages, S., Timmermann, B., Mukherji, S. and Meissner, A. (2020). Epigenetic regulator function through mouse gastrulation. *Nature* **584**, 102-108.

Gu, Z., Eils, R. and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849.

Guibentif, C., Griffiths, J. A., Imaz-Rosshandler, I., Ghazanfar, S., Nichols, J., Wilson, V., Göttgens, B. and Marioni, J. C. (2021). Diverse Routes toward Early Somites in the Mouse Embryo. *Dev Cell* **56**, 141-153.e146.

Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., et al. (2018). Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**, 1091-1107.e1017.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., 3rd, Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529.

He, P., Williams, B. A., Trout, D., Marinov, G. K., Amrhein, H., Berghella, L., Goh, S. T., Plajzer-Frick, I., Afzal, V., Pennacchio, L. A., et al. (2020). The changing mouse embryo transcriptome at whole tissue and single-cell resolution. *Nature* **583**, 760-767.

Ibarra-Soria, X., Jawaid, W., Pijuan-Sala, B., Ladopoulos, V., Scialdone, A., Jörg, D. J., Tyser, R. C. V., Calero-Nieto, F. J., Mulas, C., Nichols, J., et al. (2018). Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation. *Nat Cell Biol* **20**, 127-134.

Jacomy, M., Venturini, T., Heymann, S. and Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One* **9**, e98679.

Jin, S., Guerrero-Juarez, C. F., Zhang, L., Chang, I., Ramos, R., Kuan, C. H., Myung, P., Plikus, M. V. and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088.

Kanti Mardia, J. K., J. Bibby (1979). *Multivariate Analysis*.

Kinder, S. J., Tsang, T. E., Quinlan, G. A., Hadjantonakis, A. K., Nagy, A. and Tam, P. P. (1999). The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo. *Development* **126**, 4691-4701.

La Manno, G., Siletti, K., Furlan, A., Gyllborg, D., Vinsland, E., Mossi Albiach, A., Mattsson Langseth, C., Khven, I., Lederer, A. R., Dratva, L. M., et al. (2021). Molecular architecture of the developing mouse brain. *Nature* **596**, 92-96.

Lescroart, F., Wang, X., Lin, X., Swedlund, B., Gargouri, S., Sànchez-Dànes, A., Moignard, V., Dubois, C., Paulissen, C., Kinston, S., et al. (2018). Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. *Science* **359**, 1177-1181.

Lun, A. T., McCarthy, D. J. and Marioni, J. C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**, 2122.

Lupu, I.-E., Kirschnick, N., Weischer, S., Martinez-Corral, I., Forrow, A., Lahmann, I., Riley, P. R., Zobel, T., Makinen, T., Kiefer, F., et al. (2022). Direct specification of lymphatic endothelium from non-venous angioblasts. *bioRxiv*, 2022.2005.2011.491403.

McGrath, K. E., Frame, J. M., Fegan, K. H., Bowen, J. R., Conway, S. J., Catherman, S. C., Kingsley, P. D., Koniski, A. D. and Palis, J. (2015). Distinct Sources of Hematopoietic Progenitors Emerge before HSCs and Provide Functional Blood Cells in the Mammalian Embryo. *Cell Rep* **11**, 1892-1904.

Mittnenzweig, M., Mayshar, Y., Cheng, S., Ben-Yair, R., Hadas, R., Rais, Y., Chomsky, E., Reines, N., Uzonyi, A., Lumerman, L., et al. (2021). A single-embryo, single-cell time-resolved model for mouse gastrulation. *Cell* **184**, 2825-2842.e2822.

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., et al. (2017). Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. *Cell Rep* **20**, 1215-1228.

Monge, G. (1781). *Mémoire sur la théorie des déblais et des remblais. Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même anné*.

Murayama, E., Vivier, C., Schmidt, A. and Herbomel, P. (2023). Alcam-a and Pdgfr-α are essential for the development of sclerotome-derived stromal cells that support hematopoiesis. *Nat Commun* **14**, 1171.

Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. and Luo, L. (2007). A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593-605.

Nguyen, P. D., Hollway, G. E., Sonntag, C., Miles, L. B., Hall, T. E., Berger, S., Fernandez, K. J., Gurevich, D. B., Cole, N. J., Alaei, S., et al. (2014). Haematopoietic stem cell induction by somite-derived endothelial cells controlled by meox1. *Nature* **512**, 314-318.

Nowotschin, S., Setty, M., Kuo, Y. Y., Liu, V., Garg, V., Sharma, R., Simon, C. S., Saiz, N., Gardner, R., Boutet, S. C., et al. (2019). The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361-367.

Pardanaud, L., Luton, D., Prigent, M., Bourcheix, L. M., Catala, M. and Dieterlen-Lievre, F. (1996). Two distinct endothelial lineages in ontogeny, one of them related to hemopoiesis. *Development* **122**, 1363-1371.

Picelli, S., Faridani, O. R., Björklund, A. K., Winberg, G., Sagasser, S. and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181.

Pijuan-Sala, B., Griffiths, J. A., Guibentif, C., Hiscock, T. W., Jawaid, W., Calero-Nieto, F. J., Mulas, C., Ibarra-Soria, X., Tyser, R. C. V., Ho, D. L. L., et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490-495.

Polanski, K., Young, M. D., Miao, Z., Meyer, K. B., Teichmann, S. A. and Park, J. E. (2020). BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964-965.

Pouget, C., Gautier, R., Teillet, M. A. and Jaffredo, T. (2006). Somite-derived cells replace ventral aortic hemangioblasts and provide aortic smooth muscle cells of the trunk. *Development* **133**, 1013-1022.

Qiu, C., Cao, J., Martin, B. K., Li, T., Welsh, I. C., Srivatsan, S., Huang, X., Calderon, D., Noble, W. S., Disteche, C. M., et al. (2022). Systematic reconstruction of cellular trajectories across mouse embryogenesis. *Nat Genet* **54**, 328-341.

Qiu, C., Martin, B. K., Welsh, I. C., Daza, R. M., Le, T.-M., Huang, X., Nichols, E. K., Taylor, M. L., Fulton, O., O'Day, D. R., et al. (2023). A single-cell transcriptional timelapse of mouse embryonic development, from gastrula to pup. *bioRxiv*, 2023.2004.2005.535726.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. and Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47.

Rodriguez, A. M., Jin, D. X., Wolfe, A. D., Mikedis, M. M., Wierenga, L., Hashmi, M. P., Viebahn, C. and Downs, K. M. (2017). Brachyury drives formation of a distinct vascular branchpoint critical for fetal-placental arterial union in the mouse gastrula. *Dev Biol* **425**, 208-222.

Rybtsov, S., Batsivari, A., Bilotkach, K., Paruzina, D., Senserrich, J., Nerushev, O. and Medvinsky, A. (2014). Tracing the origin of the HSC hierarchy reveals an SCF-dependent, IL-3-independent CD43(-) embryonic precursor. *Stem Cell Reports* **3**, 489-501.

Saelens, W., Cannoodt, R., Todorov, H. and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. *Nat Biotechnol* **37**, 547-554.

Sato, Y., Watanabe, T., Saito, D., Takahashi, T., Yoshida, S., Kohyama, J., Ohata, E., Okano, H. and Takahashi, Y. (2008). Notch mediates the segmental specification of angioblasts in somites and their directed migration toward the dorsal aorta in avian embryos. *Dev Cell* **14**, 890-901.

Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., et al. (2019). Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell* **176**, 928-943.e922.

Scialdone, A., Tanaka, Y., Jawaid, W., Moignard, V., Wilson, N. K., Macaulay, I. C., Marioni, J. C. and Göttgens, B. (2016). Resolving early mesoderm diversification through single-cell expression profiling. *Nature* **535**, 289-293.

Sitara Persad, Z.-N. C., Christine Dien, Ignas Masilionis, Ronan Chaligné, Tal Nawy, Chrysothemis C Brown, Itsik Pe'er, Manu Setty, Dana Pe'er SEACells: Inference of transcriptional and epigenomic cellular states from single-cell genomics data. *bioRxiv*.

Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E. and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477.

Sturgess, K. H. M., Calero-Nieto, F. J., Göttgens, B. and Wilson, N. K. (2021). Single-Cell Analysis of Hematopoietic Stem Cells. *Methods Mol Biol* **2308**, 301-337.

Subelj, L. and Bajec, M. (2011). Unfolding communities in large complex networks: combining defensive and offensive label propagation for core extraction. *Phys Rev E Stat Nonlin Soft Matter Phys* **83**, 036103.

Sun, J., Ting, M. C., Ishii, M. and Maxson, R. (2016). Msx1 and Msx2 function together in the regulation of primordial germ cell migration in the mouse. *Dev Biol* **417**, 11-24.

Tam, P. P. L. and Ho, J. W. K. (2020). Cellular diversity and lineage trajectory: insights from mouse single cell transcriptomes. *Development* **147**.

Tani, S., Chung, U. I., Ohba, S. and Hojo, H. (2020). Understanding paraxial mesoderm development and sclerotome specification for skeletal repair. *Exp Mol Med* **52**, 1166-1177.

Tober, J., Koniski, A., McGrath, K. E., Vemishetti, R., Emerson, R., de Mesy-Bentley, K. K., Waugh, R. and Palis, J. (2007). The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis. *Blood* **109**, 1433-1441.

Traag, V. A., Waltman, L. and van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* **9**, 5233.

Tritschler, S., Büttner, M., Fischer, D. S., Lange, M., Bergen, V., Lickert, H. and Theis, F. J. (2019). Concepts and limitations for learning developmental trajectories from single cell genomics. *Development* **146**.

Van den Berge, K., Roux de Bezieux, H., Street, K., Saelens, W., Cannoodt, R., Saeys, Y., Dudoit, S. and Clement, L. (2020). Trajectory-based differential expression analysis for single-cell sequencing data. *Nat Commun* **11**, 1201.

Yamanaka, Y., Hamidi, S., Yoshioka-Kobayashi, K., Munira, S., Sunadome, K., Zhang, Y., Kurokawa, Y., Ericsson, R., Mieda, A., Thompson, J. L., et al. (2023). Reconstituting human somitogenesis in vitro. *Nature* **614**, 509-520.

Yokomizo, T., Watanabe, N., Umemoto, T., Matsuo, J., Harai, R., Kihara, Y., Nakamura, E., Tada, N., Sato, T., Takaku, T., et al. (2019). Hlf marks the developmental pathway for hematopoietic stem cells but not for erythro-myeloid progenitors. *J Exp Med* **216**, 1599-1614.

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

Yvernogeau, L., Klaus, A., Rooijen, C. v. and Robin, C. (2020). Generation of a new Tbx6-inducible reporter mouse line to trace presomitic mesoderm derivatives throughout development and in adults. *bioRxiv*, 2020.2012.2010.419275.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C. and Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523.

Figures

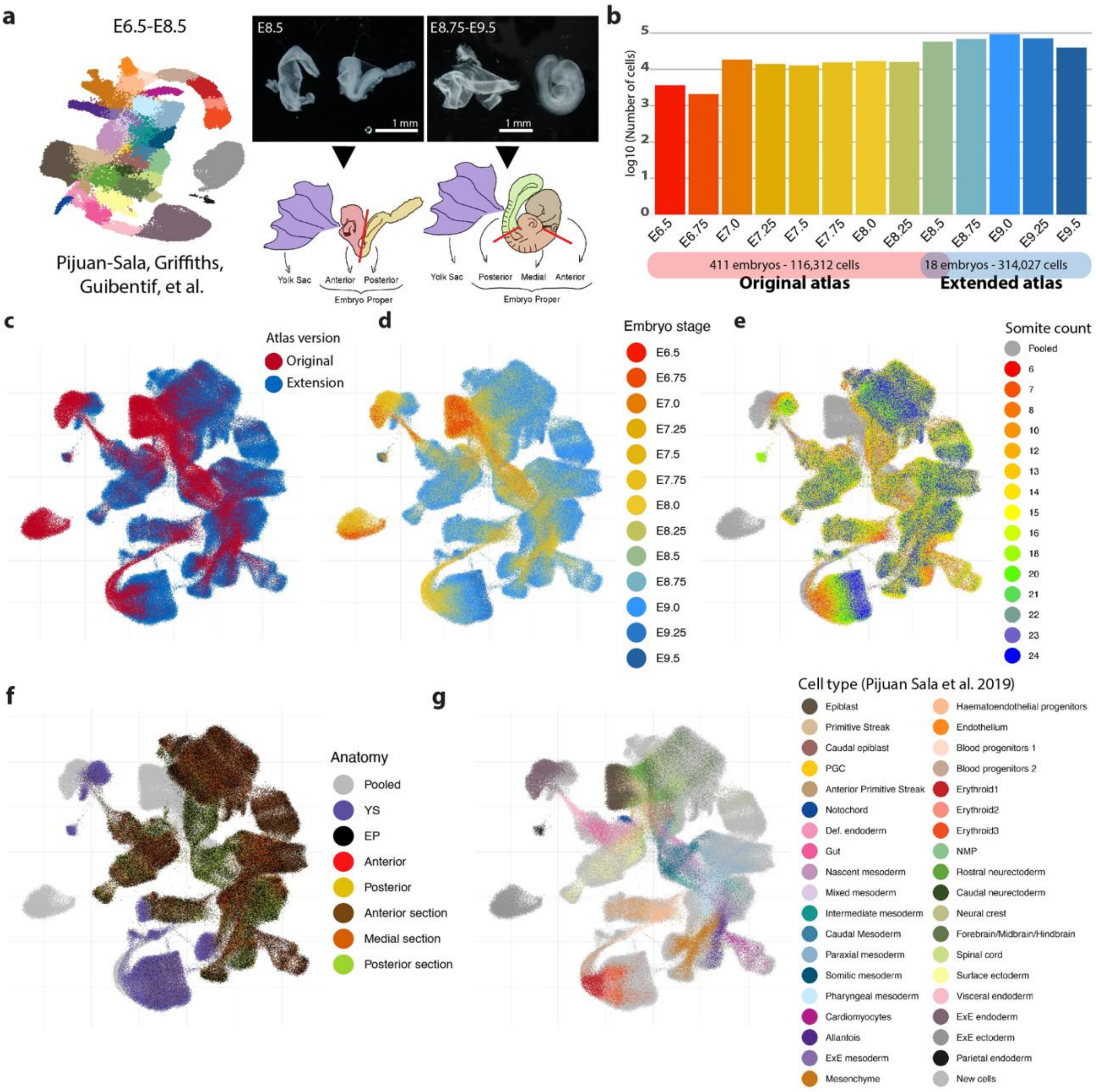


Fig. 1. Extending a single cell transcriptomic atlas of mouse gastrulation and early organogenesis.

a) Schematic representation of the experimental design. The publicly available E6.5-E8.5 time course experiment (Pijuan-Sala et al., 2019) was extended towards E9.5 across 24 hours at each six hour interval. An overlapping time point was generated (E8.5) to facilitate batch correction. Information of embryo dissections was recorded to support cell type annotations. **b)** Bar plot showing the number of cells per time point after data integration. In total, the number of transcriptomic profiles increased from 116,312 to 314,027. **c-h)** UMAP layout of the extended atlas. Cells are coloured by **c)** atlas version, the original atlas and the atlas extension, **d)** time-point, **e)** somite counts, as an alternative indication of developmental stage, **f)** anatomical dissection (pooled cells correspond to the original atlas), **g)** cell type annotations provided for the original atlas (newly generated profiles are highlighted in light grey)

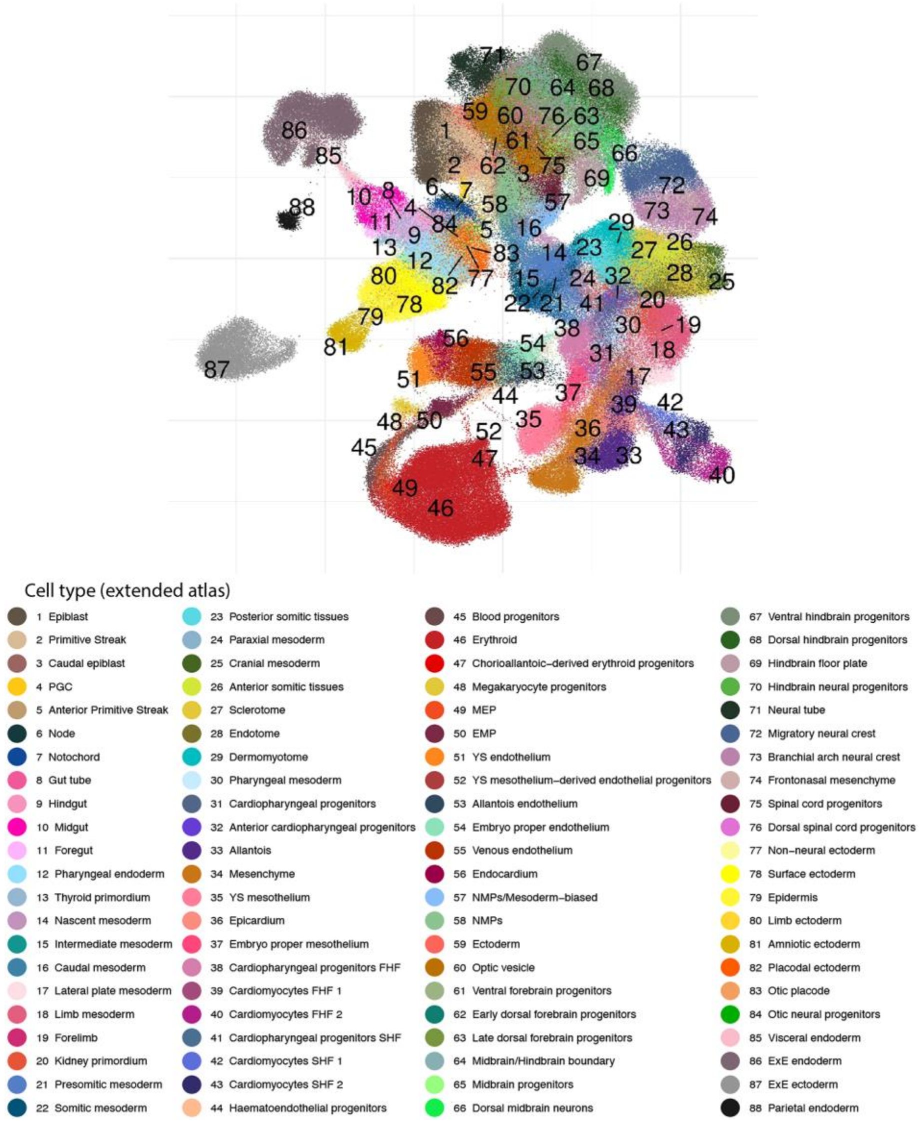


Fig. 2. An extended transcriptional atlas of mouse gastrulation and early organogenesis.

Cell type annotations resulting from the integration of both atlases and re-annotation process.

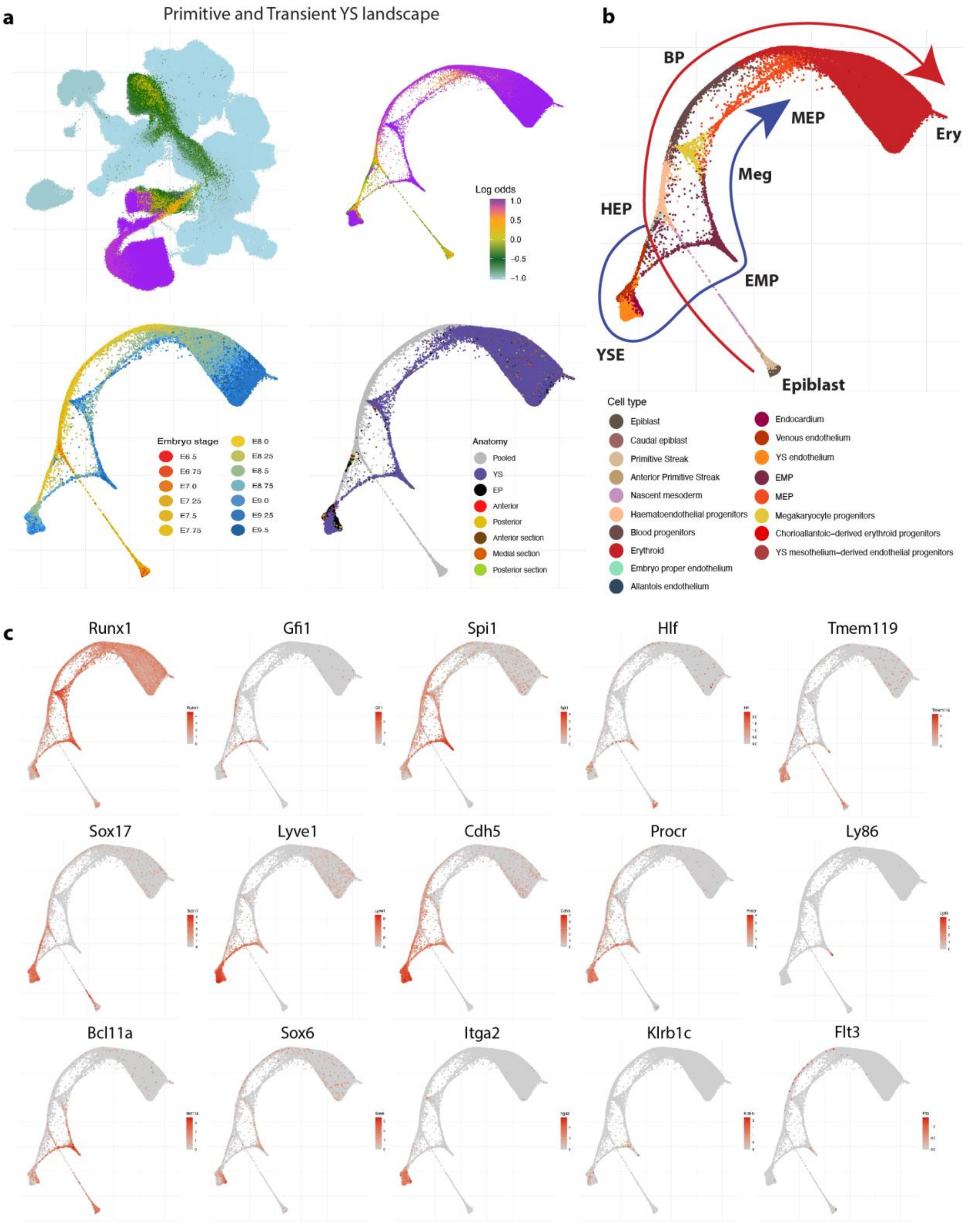


Fig. 3. Primitive and definitive YS waves of blood production.

a) UMAP layout of the mouse extended atlas displaying the log odds of fate probabilities associated with the primitive and YS definitive haemato-endothelial landscape (top left). Cells with log odds > - 0.5 were retained to generate a force directed layout. Cells are coloured by Log odds of fate probabilities of YS blood progenitors and YS endothelial cells, embryo stage and anatomical region. **b)** Force directed layout with cells coloured by cell type, displaying the trajectories of the two distinct waves, clearly distinguished by time. HEP: haemato-endothelial progenitors, BP: blood progenitors. ERY: Erythroid. YSE: Yolk Sack endothelium. EMP: Erythroid-Myeloid progenitors. MEP: Megakaryocyte-Erythroid progenitors. **c)** Force directed layout showing a collection of gene markers associated with these populations, including those represented in small fractions as lymphocytes and microglial progenitors.

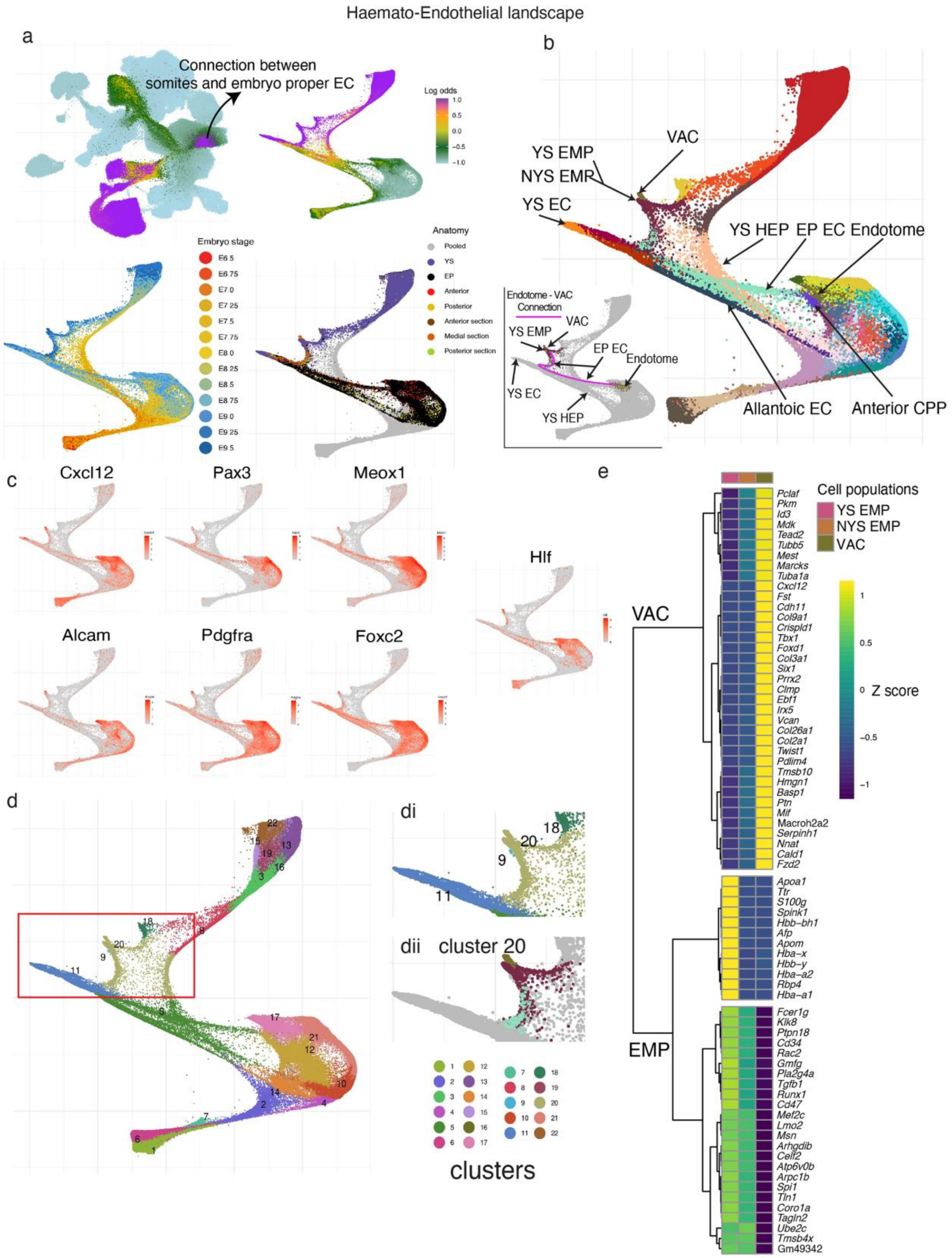


Fig. 4. The haemato-endothelial landscape.

a) UMAP layout of the mouse extended atlas displaying the log odds of fate probabilities associated with a complete haemato-endothelial landscape (top left). Cells with log odds > -1 were retained to generate a force directed layout. Cells are coloured by Log odds of fate probabilities of all hematoendothelial cells, embryo stage and anatomical regions. **b)** Force directed layout with cells coloured by cell type, highlighting multiple anatomical origins of haemato-endothelial cells and presence of blood cell types across YS and Embryo proper tissues. A subcluster of Endotome cells is splitted from its major cell type origin and cluster together with EMPs (here named as Endotome derived VACs). Cells are coloured by cell type (see cell type labels at Fig.1h). The bottom left plot further highlights two trajectories and intermediate populations, one for YS EMPs and one for VACs **c)** Collection of gene markers associated with the Endotome population as reported by [Dang Nguyen et al., 2014], [Tani et al., 2020] and *hlf,* which has been associated to HSCs induction [Yokomizo et al., 2019]. Following the landscape in b), the two circles highlight the location of endotome cells and VACs in the landscape, with the trajectory through embryonic endothelial cells indicated by an arrow. **d)** Newly calculated Louvain clusters identified in the landscape. The region of interest is highlighted in the red box. Cluster 20 is extracted and further splitted into anatomically distinct populations and used for differential expression and correlation analysis. Notice that cluster 9 are lymphocyte progenitors. **e)** Heat map displaying differentially expressed genes across different cell populations clustered together in the region highlighted in Dii. YS EMP: Yolk Sack EMP, NYS EMP: Non-Yolk Sack EMP, Endotome derived VAC: Endotome derived vascular associated cells. Mean gene expression values were computed and scaled by rows (Z-score).

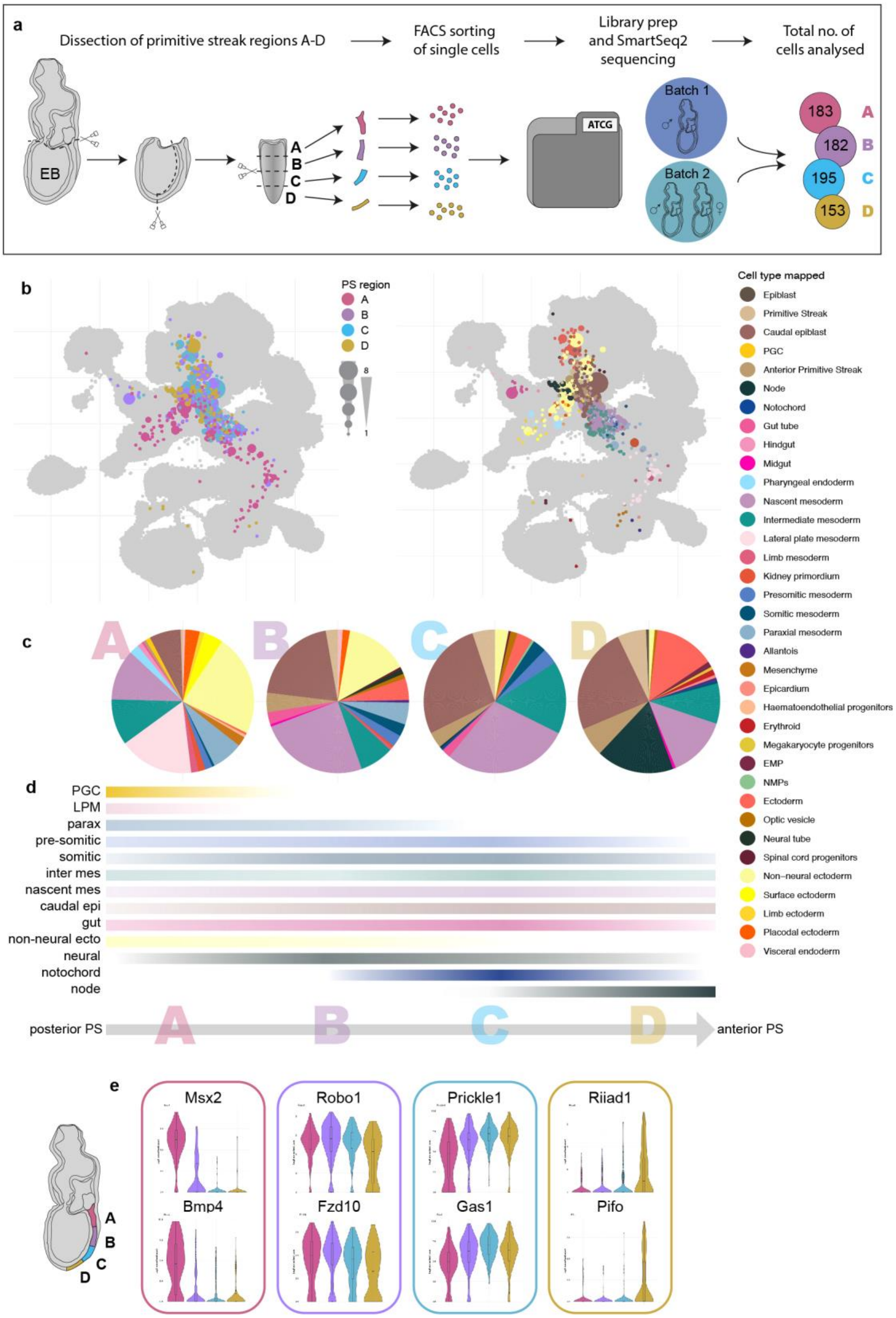


Fig. 5. A gradient of transcriptomic differences between four sequential regions of the primitive streak at E7.5.

a) Schematic of the experimental set up. Single cells isolated from four sequential early bud (EB)-stage primitive streak regions were analysed by scRNA-Seq. A total of 3 primitive streaks were analysed over two experiments. The final cell numbers analyzed for each primitive streak region are indicated. **b)** UMAPs of primitive streak cells mapped onto the extended mouse atlas. Cells are coloured by primitive streak region of origin (left) or transferred cell type label (right). The size of layout dots is proportional to the number of shared closest neighbours across primitive streak cells. **c)** Pie charts showing the relative proportion of the different cell types of individual primitive streak cells mapped to by region of origin (A to D). Five cells from the most distal region D were unexpectedly annotated as erythroid progenitors (Fig.4b, c). These may have inadvertently been included in the analyses during the removal of yolk sac tissues. **d)** Percentile representation of cell type mapping along the primitive streak axis (from A to D, respectively). The colour intensity indicates cell type abundance and resembles the gradient of cell type bias from specific PS portions. **e)** Examples of the most differentially expressed genes between the 4 primitive streak regions.

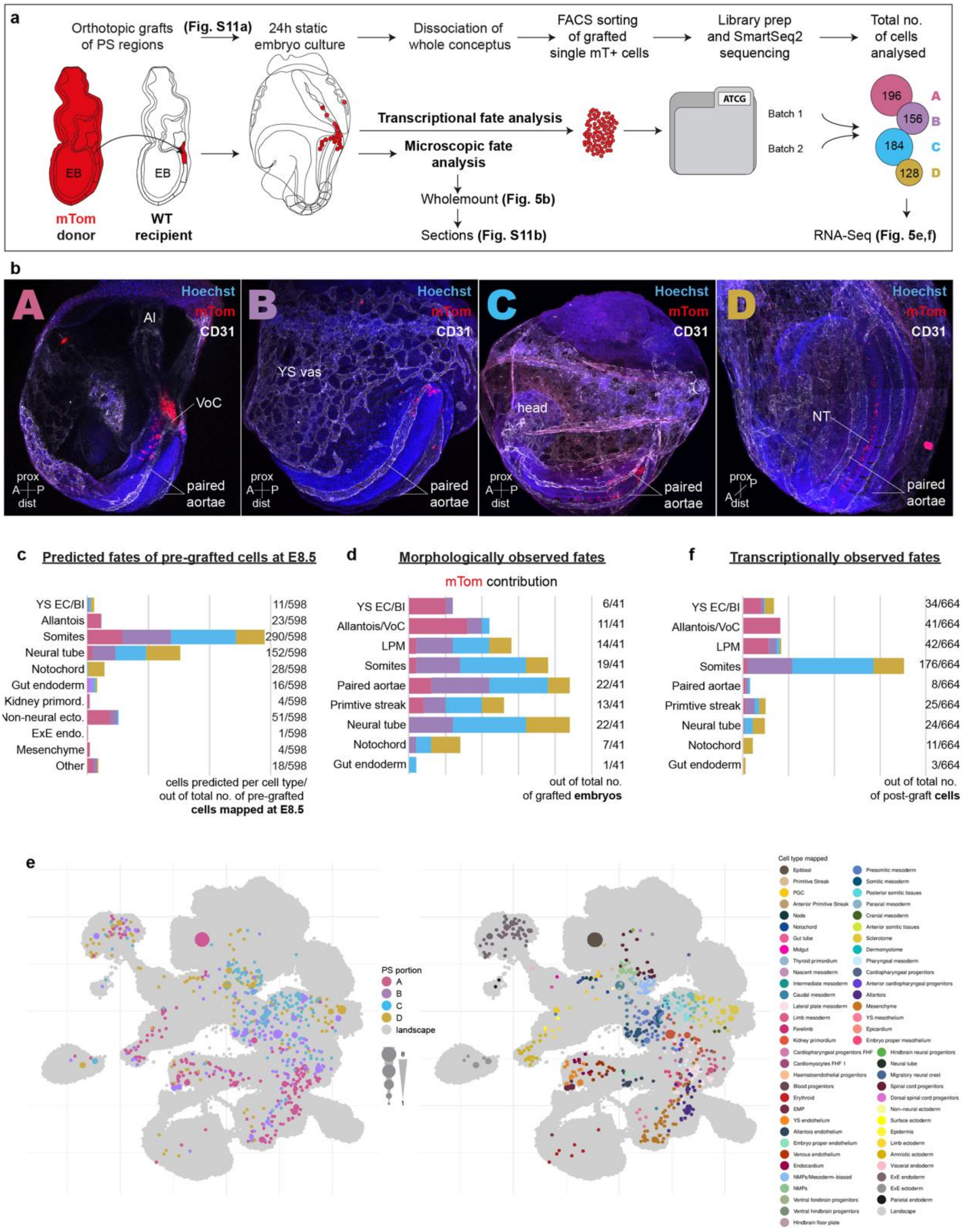


Fig. 6. Experimental validation of cell fates in primitive streak grafts.

a) Schematic of the orthotopic graft experiments with subsequent analysis of cell progeny by imaging and RNA-Seq. Primitive streak regions A to D isolated from E7.5 (EB) mouse embryos carrying a ubiquitous membrane bound tdTomato (mTom) reporter were orthotopically grafted as small cell clusters into E7.5 (EB) wild type recipient embryos. After

24h of culture grafted embryos were either fixed, stained for wholemount analysis and cryo-sectioned for detailed analysis of mTom+ cell contribution; or sorted for mTom+ single cells and processed for SmartSeq2 scRNA-Seq. The total number of mTom+ cells sequenced is given. **b)** Representative wholemount images of grafted embryos after culture, showing mTom (red) contribution to different tissues. Vasculature was stained with CD31 (white). **c)** Fate map of predicted fates of cells freshly isolated from E7.5 (EB) primitive streak regions A to D (documented in Fig.4), determined with the W-OT algorithm. Prediction was capped at E8.5 to allow direct comparison to microscopical and transcriptional fates of the post-grafted donor primitive streak-derived cells. **d)** Fate map of experimentally observed fates of the primitive streak regions A to D based on microscopic analysis (wholemount and sections) of mTom contribution in different tissues of the grafted embryos after culture (Suppl.Table 1). The number of grafted embryos with mTom contribution to particular tissues out of the total number of grafted embryos is shown for each fate. **e)** UMAP of the mouse extended atlas highlighting the closest neighbouring cells to the transcriptomic profiles generated from mTom+ cells isolated from grafted embryos. Cells are coloured by primitive streak region (left) and transferred cell type label (right). The size of layout dots is proportional to the number of shared closest neighbours across primitive streak cells. **f)** Fate map of transcriptionally observed fates based on the mapping of the post-grafted cells on the extended atlas. Only selected cell types are shown here to match the microscopically observed fates. The complete data is provided in Fig. S11d. prox - proximal, dist - distal, A - anterior, P - posterior, Al - Allantois, VoC - Vessel of confluence, YS vas - yolk sac vasculature, NT - neural tube.

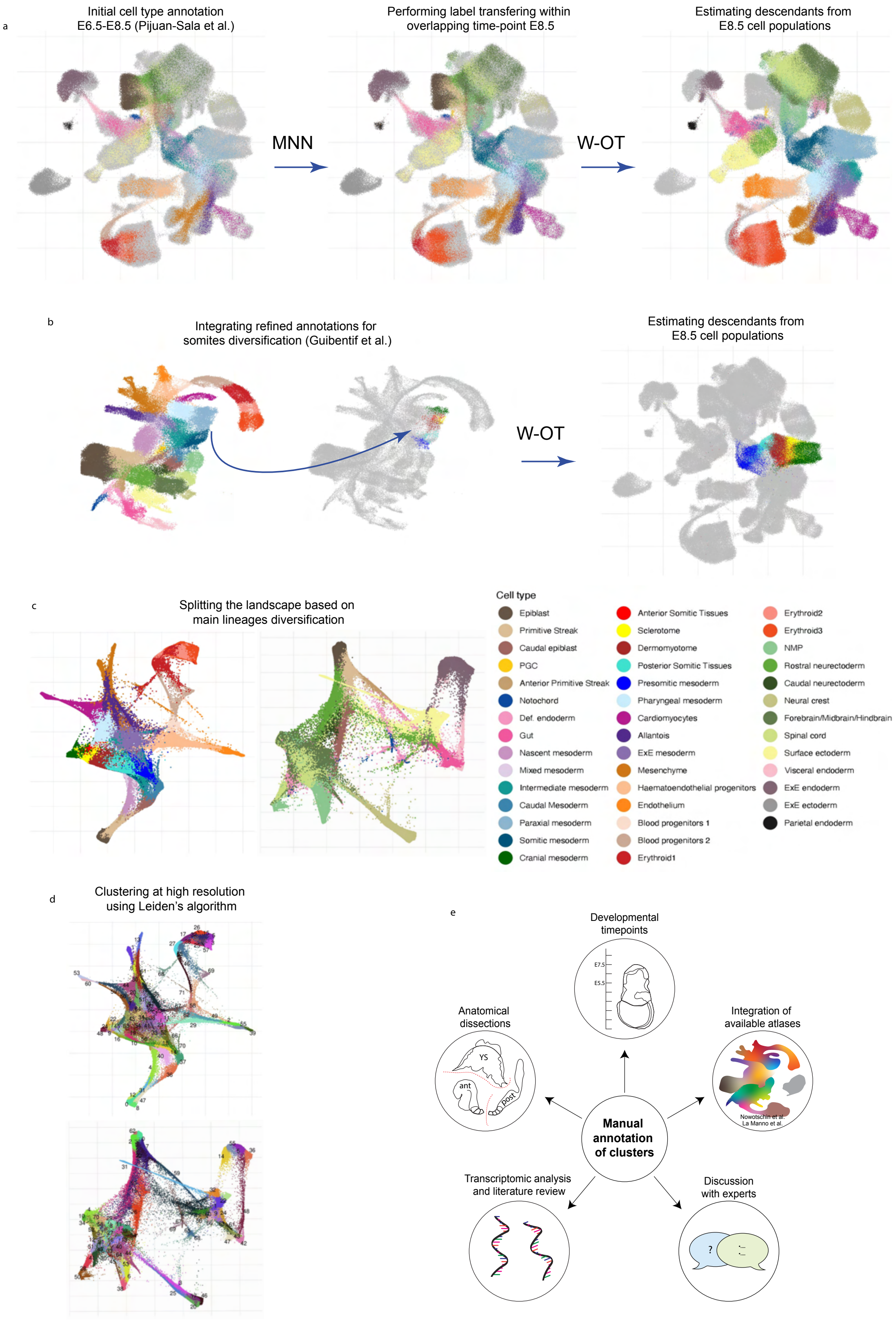

a
Initial cell type annotation
E6.5-E8.5 (Pijuan-Sala et al.)
MNN
Performing label transfering within
overlapping time-point E8.5
W-OT
Estimating descendants from
E8.5 cell populations
b
Integrating refined annotations for
somites diversification (Guibentif et al.)
W-OT
Estimating descendants from
E8.5 cell populations
c
Splitting the landscape based on
main lineages diversification
Cell type
Epiblast
Primitive Streak
Caudal epiblast
PGC
Anterior Primitive Streak
Notochord
Def. endoderm
Gut
Nascent mesoderm
Mixed mesoderm
Intermediate mesoderm
Caudal Mesoderm
Paraxial mesoderm
Somitic mesoderm
Cranial mesoderm
Anterior Somitic Tissues
Sclerotome
Dermomyotome
Posterior Somitic Tissues
Presomitic mesoderm
Pharyngeal mesoderm
Cardiomyocytes
Allantois
ExE mesoderm
Mesenchyme
Haematoendothelial progenitors
Endothelium
Blood progenitors 1
Blood progenitors 2
Erythroid1
Erythroid2
Erythroid3
NMP
Rostral neurectoderm
Caudal neurectoderm
Neural crest
Forebrain/Midbrain/Hindbrain
Spinal cord
Surface ectoderm
Visceral endoderm
ExE endoderm
ExE ectoderm
Parietal endoderm
d
Clustering at high resolution
using Leiden's algorithm
e
Developmental
timepoints
E7.5
E5.5
Anatomical
dissections
YS
ant
post
Integration of
available atlases
Nowotschin et al.
La Manno et al.
Manual
annotation
of clusters
Transcriptomic analysis
and literature review
Discussion
with experts
?

Fig. S1. Extending cell type annotations step by step.

a) UMAP layouts are used to visualise the first stage of cell type annotation, cells are coloured by cell type [Pijuan-Sala et al. 2019] but light grey cells represent those new that have not yet been annotated. Initially, mutual nearest neighbours (MNN) were used to transfer E8.5 cell type annotations from [Pijuan-Sala et al. 2019] to newly profiled cells within the overlapping time-point. Next, cell descendants are estimated for all existing E8.5 cell populations using W-OT by pushing forward the mass over the transport maps in the following time-points (E8.75-E9.5). Here, only cell descendant populations are coloured by cell type and light grey cells refer instead, to those from the original atlas (E6.5-E8.5). **b)** Refined cell type annotations for Paraxial and Somitic mesoderm from Guibentif et al. [Pijuan-Sala et al. 2019] are integrated through the same process described in a). **c)** The atlas is splitted into two landscapes based on the annotations deriving from a) and b). The mesodermal landscape at the left and the ectodermal/ endodermal one at the right. Cells are coloured by original Atlas cell types and descendants. **d)** Cells are coloured by high resolution louvain clustering as the starting point of new cell type annotations. **e)** Highly variable genes, batch correction and leiden clustering was performed independently for the two landscapes obtained from C. These clusters are manually annotated using different sources of information as illustrated in the right side scheme. For instance, cell type annotations from other atlases such as [Nowotschin et al. 2019 and La Manno et. al, 2021] were used as guidance, differential expression analysis and literature marker inspection and trajectory analysis.

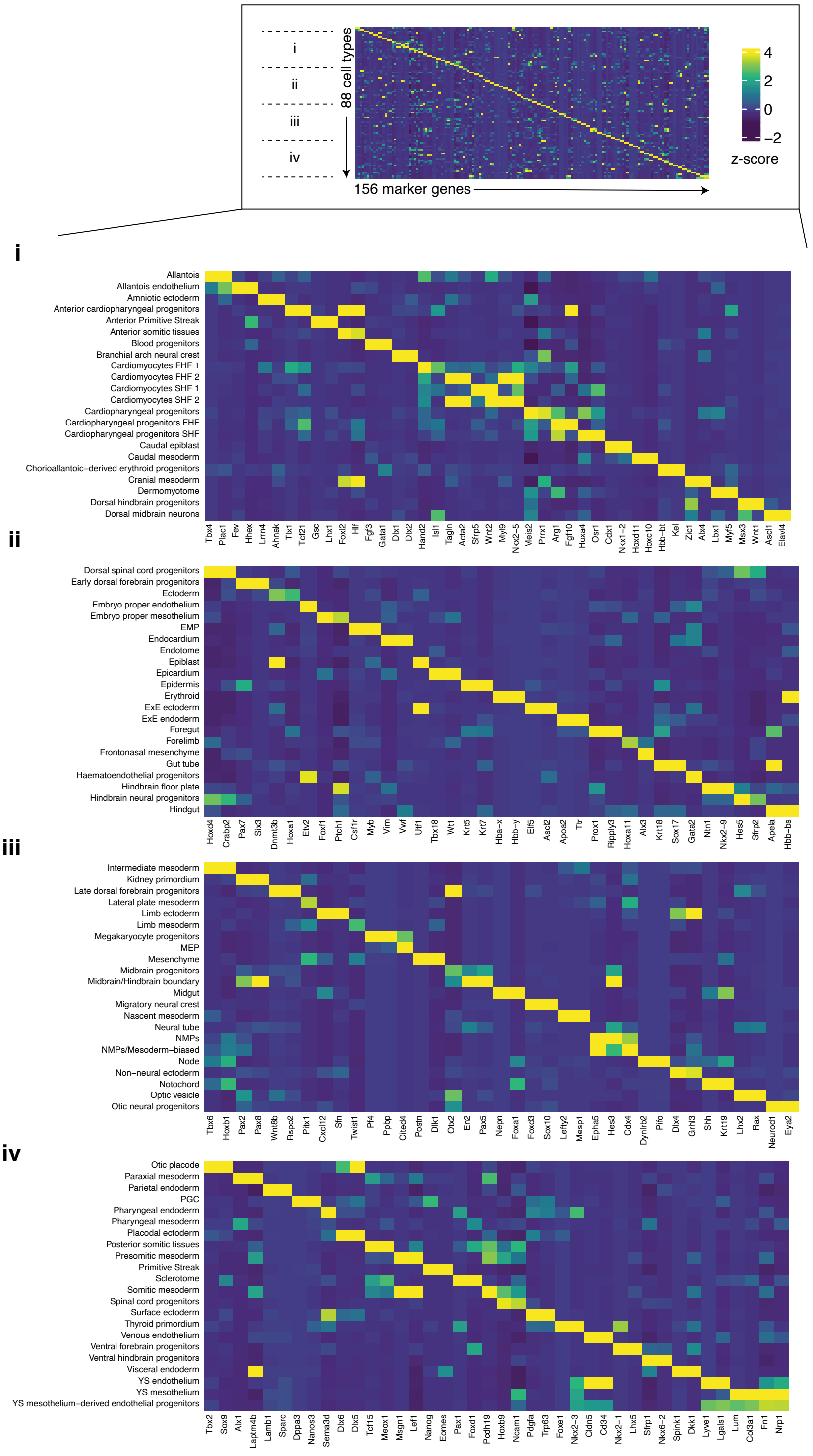


Fig. S2. Cell type marker genes.

Collection of gene markers (~ 2 genes/cell type) for the 88 cell types shown as a complement for Figure 1. Mean gene expression values were computed and scaled by rows (Z-score), cell types are listed alphabetically.

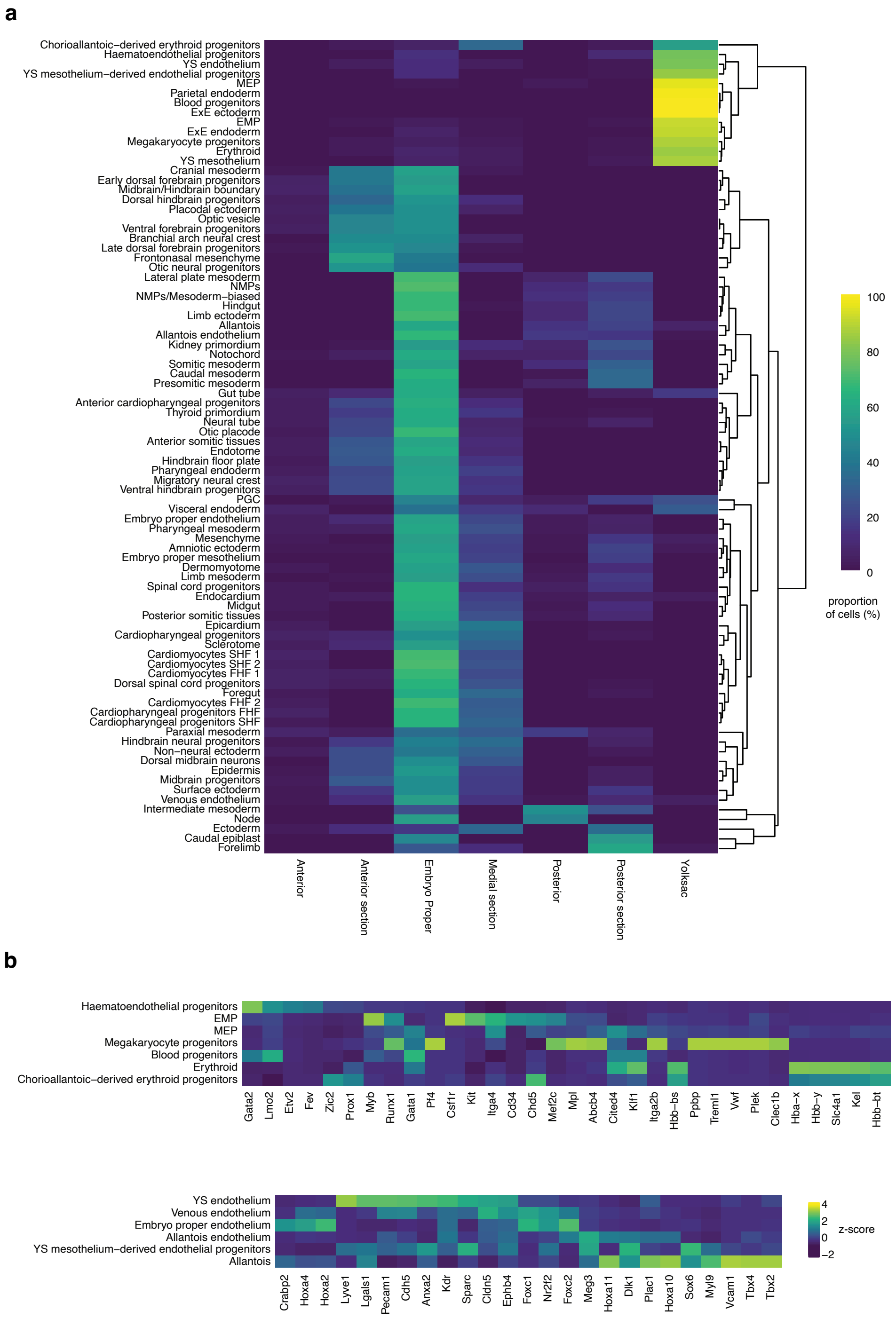


Fig. S3. Cell type distribution in different anatomical regions.

(a) Heatmap showing the proportion of cells (%) in different anatomical regions, normalized per row (cell type). Rows are hierarchically clustered and only cells from E8.5-E9.5 sub-dissected embryos are included. **(b)** Collection of gene markers for endothelium, haemato-endothelial and blood progenitors shown as a complement for Figure 2. Mean gene expression values were computed and scaled by rows (Z-score).

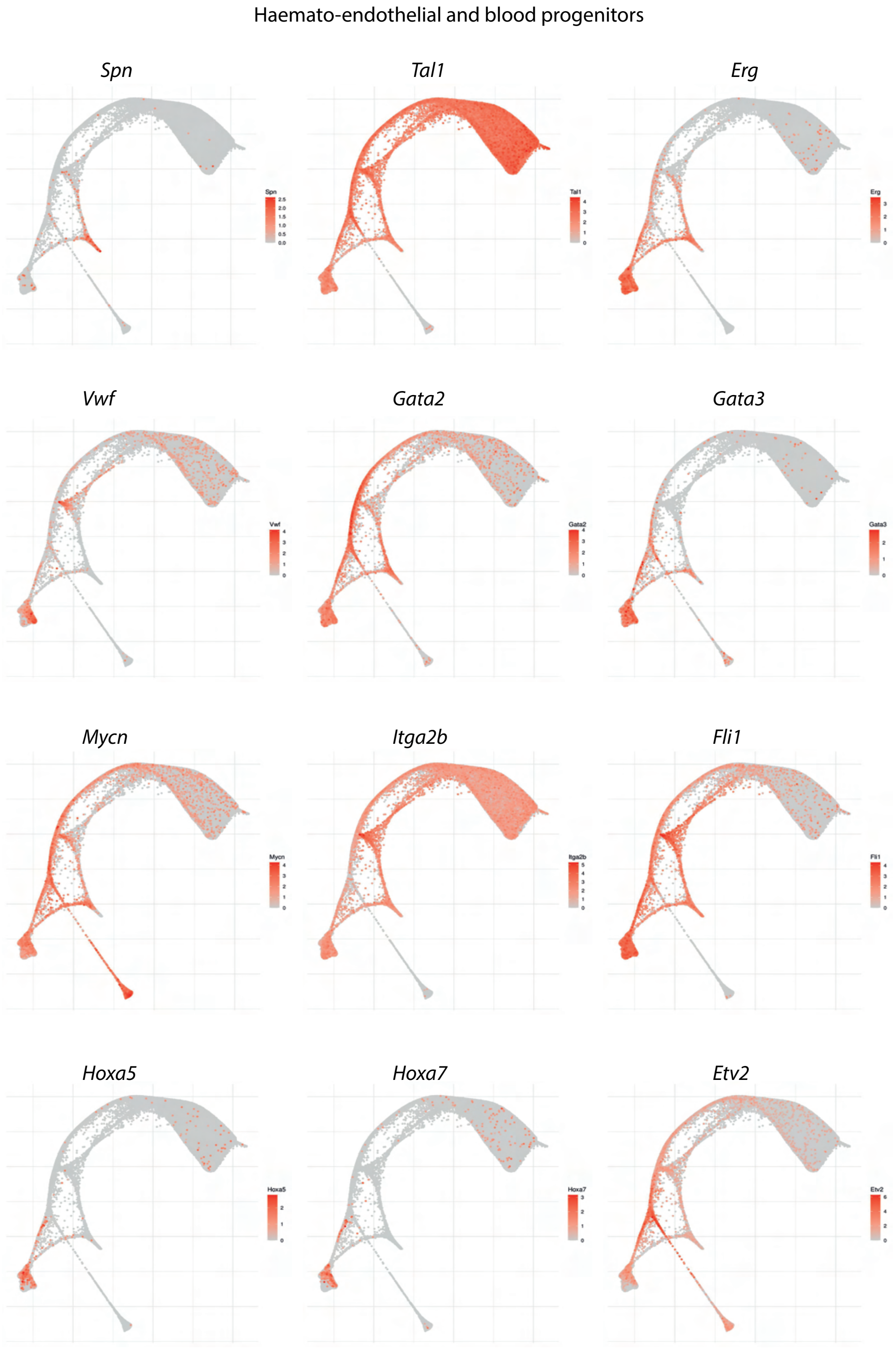


Fig. S4. Haemato-endothelial and blood progenitors.

Collection of gene markers for haemato-endothelial and blood progenitors shown as a complement for Figure 2. Force directed layouts displaying gene markers expression levels.

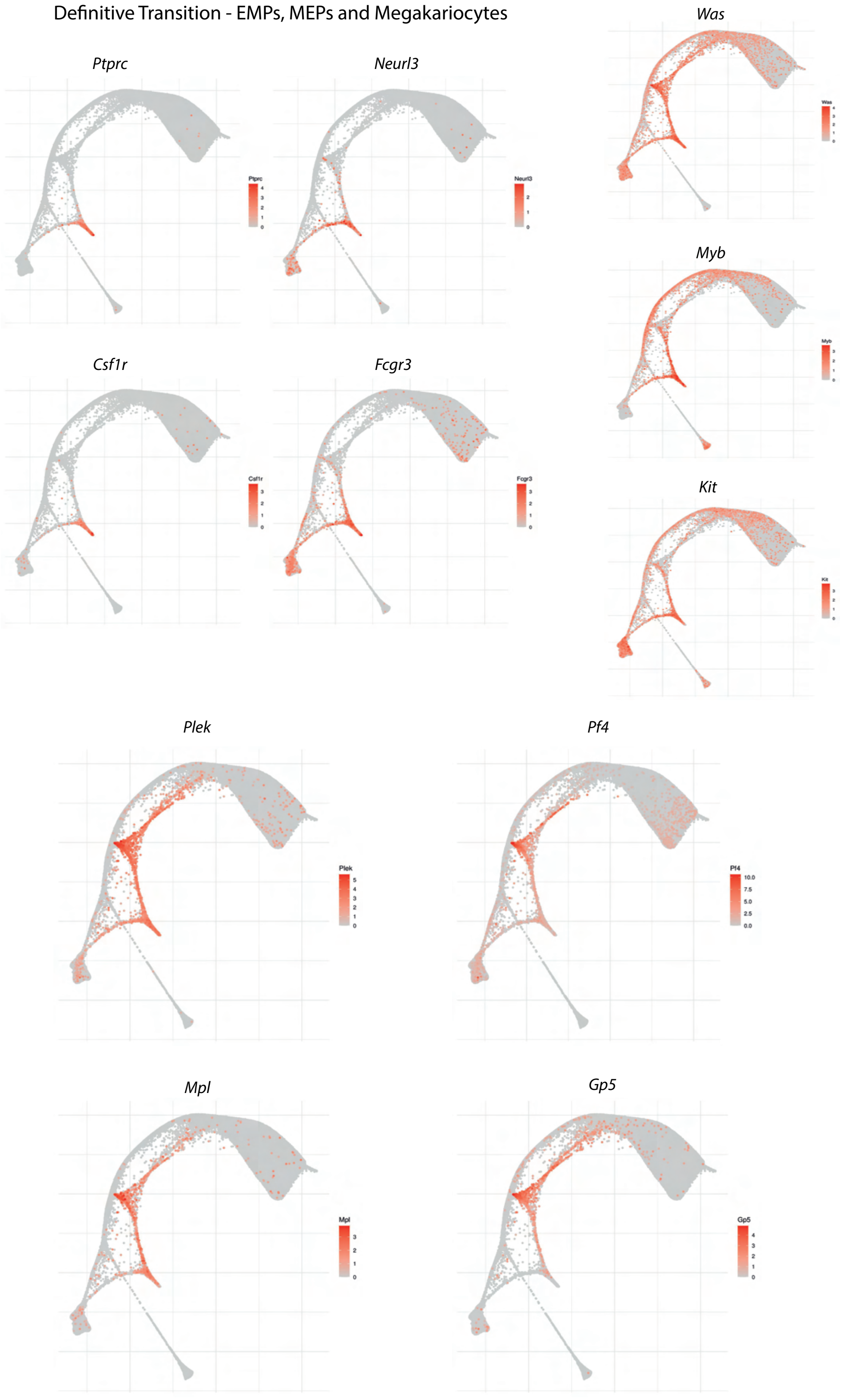


Fig. S5. Definitive transition - EMPs, MEPs and megakaryocytes.
Collection of gene markers for definitive blood populations shown as a complement for Figure 2. Force directed layouts displaying gene markers expression levels.

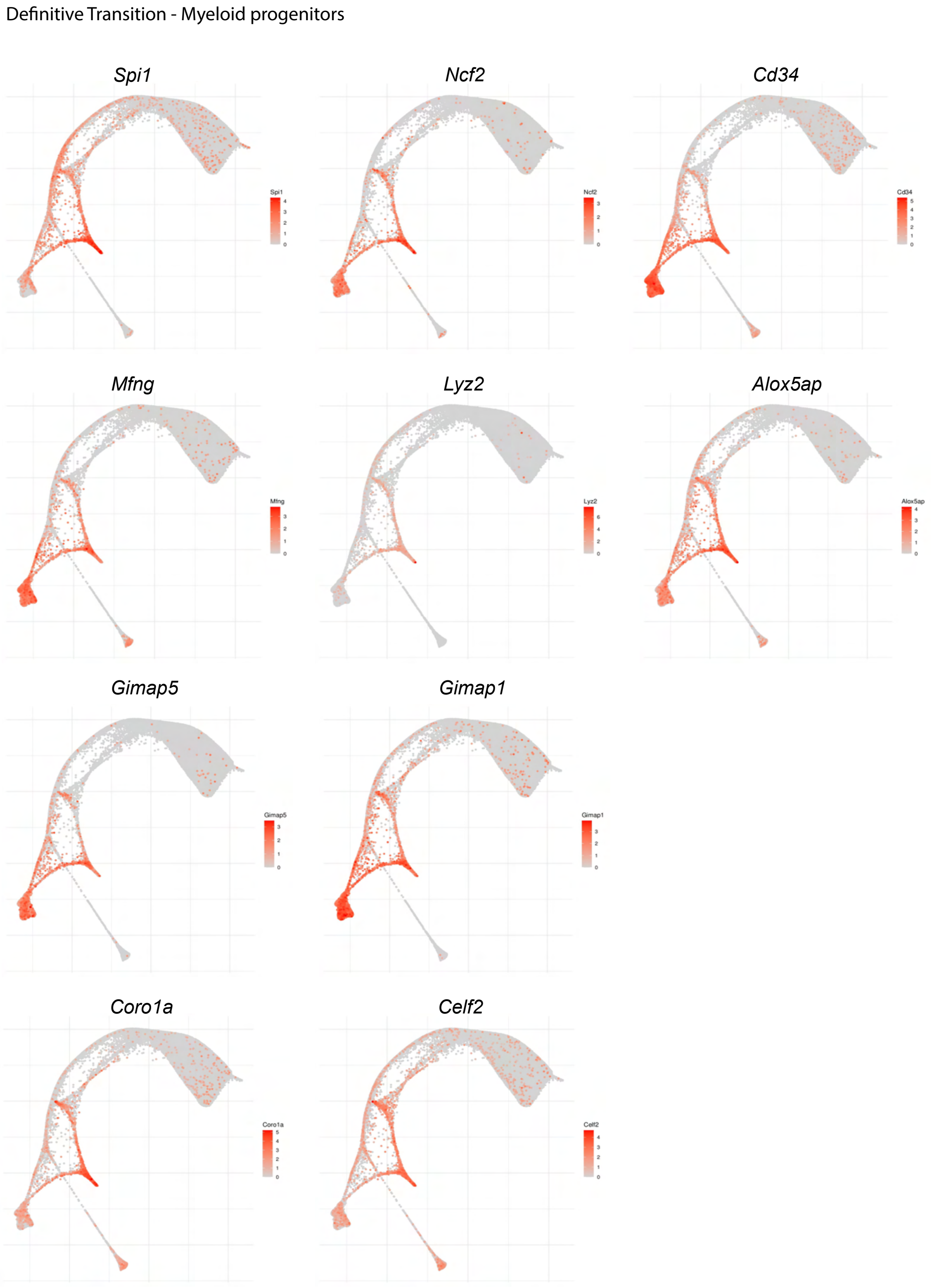


Fig. S6. Definitive transition - Myeloid progenitors.

Collection of gene markers for definitive blood populations shown as a complement for Figure 2. Force directed layouts displaying gene markers expression levels.

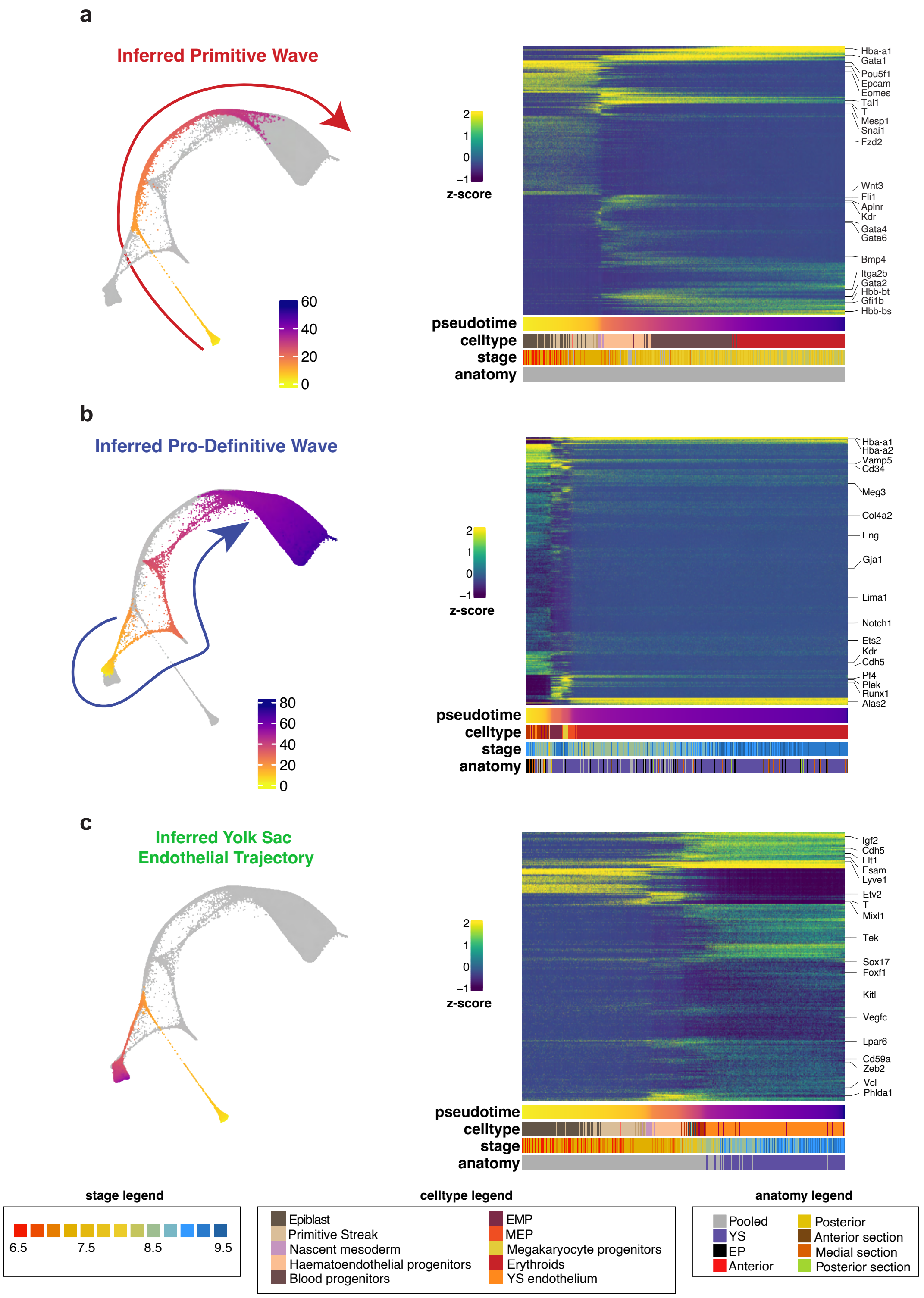

a
Inferred Primitive Wave
60
40
20
0
2
1
0
−1
z-score
Hba-a1
Gata1
Pou5f1
Epcam
Eomes
Tal1
T
Mesp1
Snai1
Fzd2
Wnt3
Fli1
Aplnr
Kdr
Gata4
Gata6
Bmp4
Itga2b
Gata2
Hbb-bt
Gfi1b
Hbb-bs
pseudotime
celltype
stage
anatomy
b
Inferred Pro-Definitive Wave
80
60
40
20
0
2
1
0
−1
z-score
Hba-a1
Hba-a2
Vamp5
Cd34
Meg3
Col4a2
Eng
Gja1
Lima1
Notch1
Ets2
Kdr
Cdh5
Pf4
Plek
Runx1
Alas2
pseudotime
celltype
stage
anatomy
c
Inferred Yolk Sac
Endothelial Trajectory
2
1
0
−1
z-score
Igf2
Cdh5
Flt1
Esam
Lyve1
Etv2
T
Mixl1
Tek
Sox17
Foxf1
Kitl
Vegfc
Lpar6
Cd59a
Zeb2
Vcl
Phlda1
pseudotime
celltype
stage
anatomy
stage legend
6.5
7.5
8.5
9.5
celltype legend
Epiblast
Primitive Streak
Nascent mesoderm
Haematoendothelial progenitors
Blood progenitors
EMP
MEP
Megakaryocyte progenitors
Erythroids
YS endothelium
anatomy legend
Pooled
YS
EP
Anterior
Posterior
Anterior section
Medial section
Posterior section

Fig. S7. Gene expression changes along the inferred primitive and definitive YS waves of blood and YS endothelial formation.

Force directed layout showing the subset of cells contributing to the primitive **(a)** definitive YS **(b)** and YS endothelial inferred differentiation trajectories **(c)** coloured by pseudotime. Heat maps display the top 300 genes associated with the pseudotimes for the various inferred trajectories (Tradeseq associationTest, p<0.01 and meanLogFC > 2). Rows (genes) are hierarchically clustered together. Accompanying metadata including cell type annotation, stage and anatomy is displayed below the heatmap of gene expression. Mean gene expression values were computed and scaled by rows (Z-score).

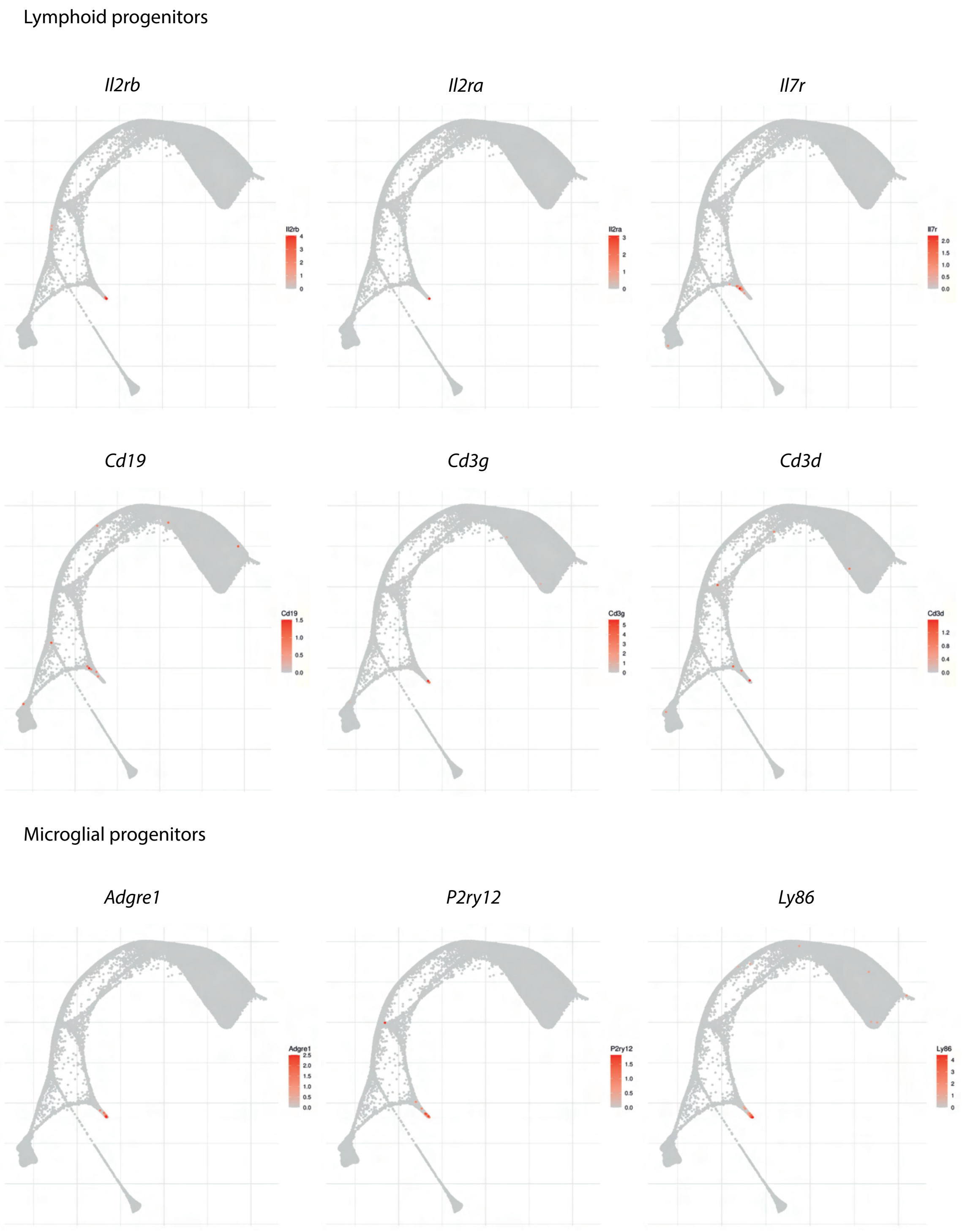


Fig. S8. Definitive transition - Lymphoid and microglial progenitors.
Collection of gene markers for definitive blood populations shown as a complement for Figure 2. Force directed layouts displaying gene markers expression levels.

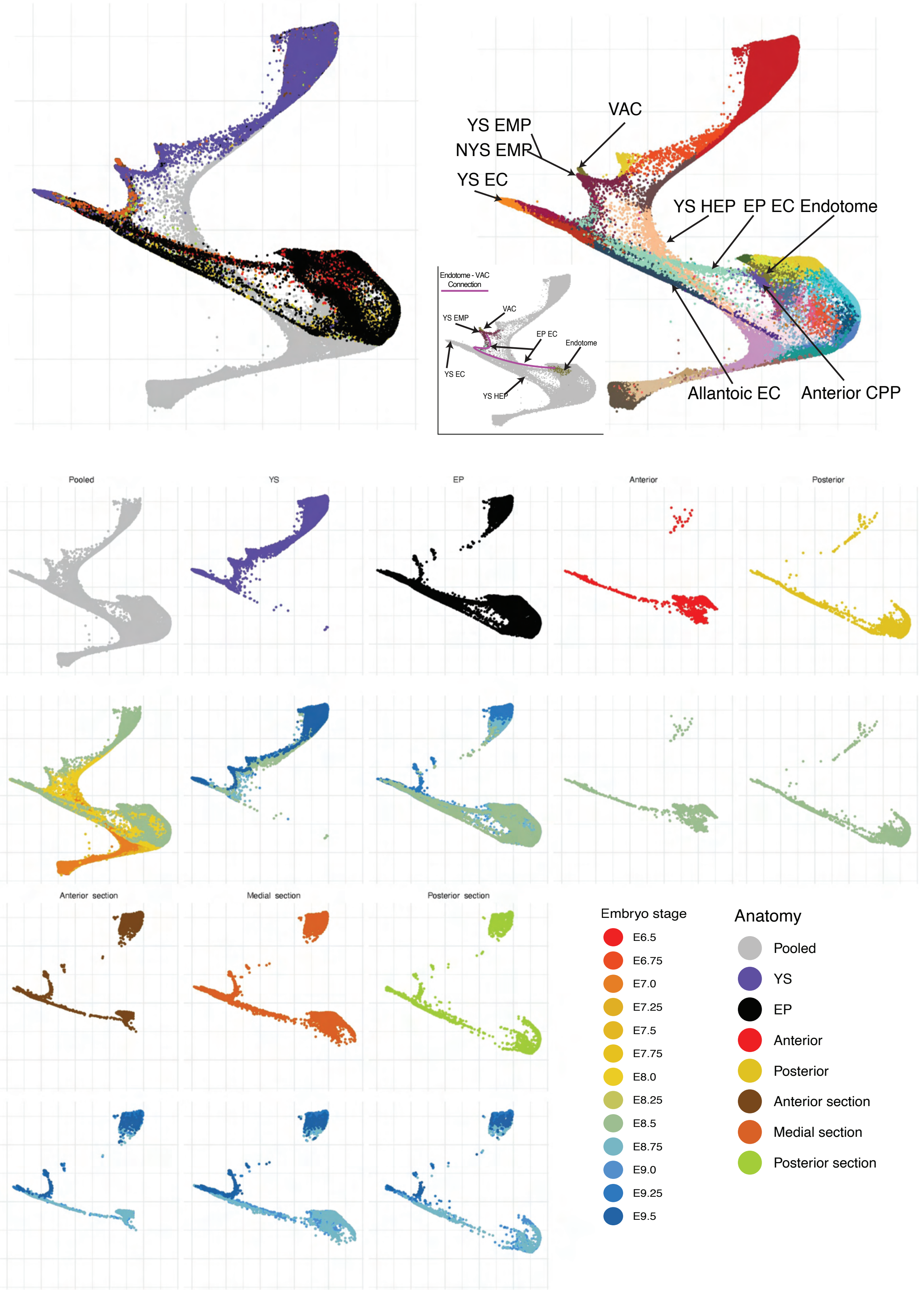

VAC
YS EMP
NYS EMP
YS EC
YS HEP
EP EC
Endotome
Allantoic EC
Anterior CPP
Endotome - VAC Connection
VAC
YS EMP
EP EC
Endotome
YS EC
YS HEP
Pooled
YS
EP
Anterior
Posterior
Anterior section
Medial section
Posterior section
Embryo stage
E6.5
E6.75
E7.0
E7.25
E7.5
E7.75
E8.0
E8.25
E8.5
E8.75
E9.0
E9.25
E9.5
Anatomy
Pooled
YS
EP
Anterior
Posterior
Anterior section
Medial section
Posterior section

Fig. S9. Diverse anatomical origins of the haemato-endothelial landscape. Force directed layout of the haemato-endothelial landscape highlighting anatomical locations and relevant cell type populations (top); each anatomical region is shown in a different panel (bottom). Here, cells are coloured by anatomical locations as well as by embryo stage. Note that embryo proper sub dissections where no distinction between anterior, medial and posterior was made before sequencing are labelled as EP.

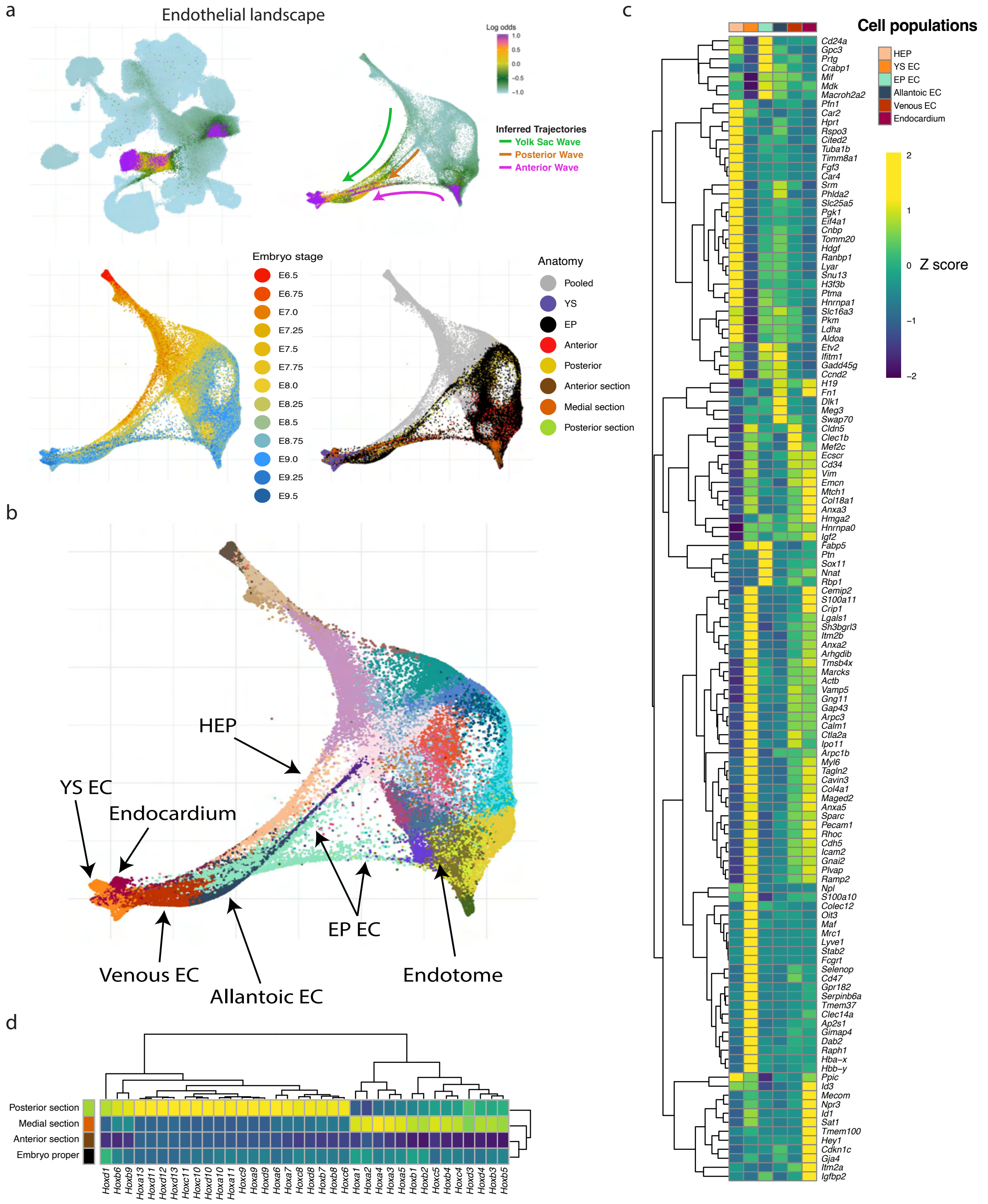
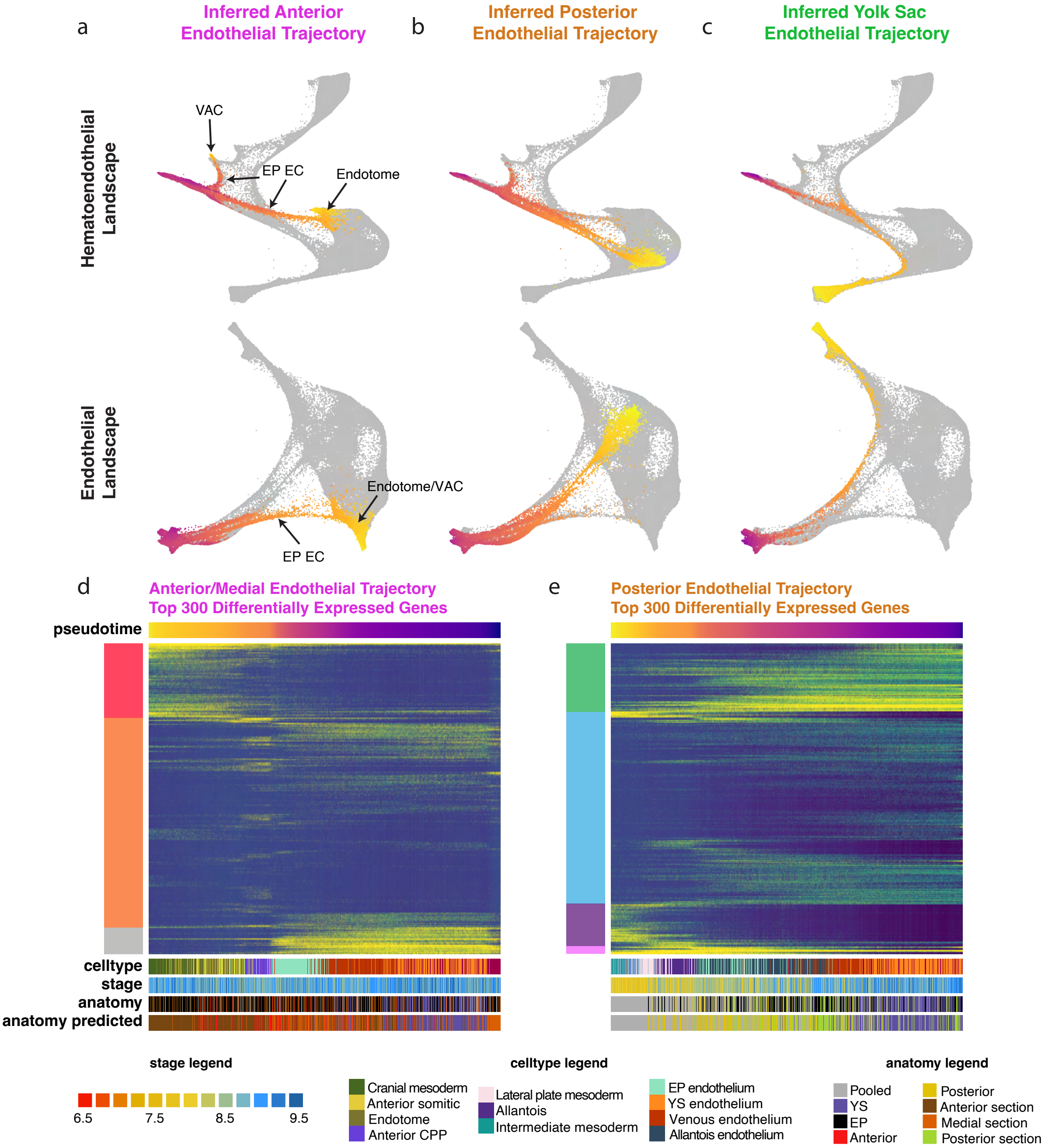

a
Endothelial landscape
Log odds
1.0
0.5
0.0
−0.5
−1.0
Inferred Trajectories
Yolk Sac Wave
Posterior Wave
Anterior Wave
Embryo stage
E6.5
E6.75
E7.0
E7.25
E7.5
E7.75
E8.0
E8.25
E8.5
E8.75
E9.0
E9.25
E9.5
Anatomy
Pooled
YS
EP
Anterior
Posterior
Anterior section
Medial section
Posterior section
b
HEP
YS EC
Endocardium
EP EC
Venous EC
Allantoic EC
Endotome
c
Cell populations
HEP
YS EC
EP EC
Allantoic EC
Venous EC
Endocardium
2
1
0
Z score
−1
−2
Cd24a
Gpc3
Prtg
Crabp1
Mif
Mdk
Macroh2a2
Pfn1
Car2
Hprt
Rspo3
Cited2
Tuba1b
Timm8a1
Fgf3
Car4
Srm
Phlda2
Slc25a5
Pgk1
Eif4a1
Cnbp
Tomm20
Hdgf
Ranbp1
Lyar
Snu13
H3f3b
Ptma
Hnrnpa1
Slc16a3
Pkm
Ldha
Aldoa
Etv2
Ifitm1
Gadd45g
Ccnd2
H19
Fn1
Dlk1
Meg3
Swap70
Cldn5
Clec1b
Mef2c
Ecscr
Cd34
Vim
Emcn
Mtch1
Col18a1
Anxa3
Hmga2
Hnrnpa0
Igf2
Fabp5
Ptn
Sox11
Nnat
Rbp1
Cemip2
S100a11
Crip1
Lgals1
Sh3bgrl3
Itm2b
Anxa2
Arhgdib
Tmsb4x
Marcks
Actb
Vamp5
Gng11
Gap43
Arpc3
Calm1
Ctla2a
Ipo11
Arpc1b
Myl6
Tagln2
Cavin3
Col4a1
Maged2
Anxa5
Sparc
Pecam1
Rhoc
Cdh5
Icam2
Gnai2
Plvap
Ramp2
Npl
S100a10
Colec12
Oit3
Maf
Mrc1
Lyve1
Stab2
Fcgrt
Selenop
Cd47
Gpr182
Serpinb6a
Tmem37
Clec14a
Ap2s1
Gimap4
Dab2
Raph1
Hba−x
Hbb−y
Ppic
Id3
Mecom
Npr3
Id1
Sat1
Tmem100
Hey1
Cdkn1c
Gja4
Itm2a
Igfbp2
d
Posterior section
Medial section
Anterior section
Embryo proper
Hoxd1
Hoxb6
Hoxb9
Hoxa13
Hoxd11
Hoxd12
Hoxd13
Hoxc11
Hoxc10
Hoxd10
Hoxa10
Hoxa11
Hoxc9
Hoxa9
Hoxd9
Hoxa6
Hoxa7
Hoxc8
Hoxd8
Hoxb7
Hoxb8
Hoxc6
Hoxa1
Hoxa2
Hoxa4
Hoxa3
Hoxa5
Hoxb1
Hoxb2
Hoxc5
Hoxb4
Hoxc4
Hoxd3
Hoxd4
Hoxb3
Hoxb5

Fig. S10. Different origins of endothelial cells.

UMAP layout of the mouse extended atlas displaying the log odds of fate probabilities exclusively associated with endothelial populations (top left). Cells with log odds > - 1 were retained to generate a force directed layout. Cells are coloured by Log odds of fate probabilities of endothelial cells, embryo stage and anatomical region. Arrows highlight three putative endothelial differentiation trajectories. **b)** Force directed layout of the endothelial landscape highlighting relevant cell populations. HEP: Haemato-endothelial progenitors. EC: Endothelial cell. **c)** Heat map of differentially expressed genes across distinct endothelial populations. Mean gene expression values were computed and scaled by rows (Z-score). **(d)** Heat map showing Hox gene expression of embryo proper endothelial cells from distinct anatomical locations. Mean gene expression values were computed and scaled by columns (Z-score).



f

	Gene Enrichment: GO Terms	-log10(P)
🟥	skeletal system development	10.93
	pattern specification process	9.38
	muscle structure development	5.74
🟧	vasculature development	15.50
	angiogenesis	8.05
	tissue morphogenesis	11.61
	epithelial tube morphogenesis	8.39
	mesenchyme development	6.57
⬜	vasculogenesis	9.56
	positive regulation of cell adhesion	6.34
	semaphorin-plexin signaling pathway	4.17

g

	Gene Enrichment: GO Terms	-log10(P)
🟩	angiogenesis	15.64
	endothelial cell migration	9.95
	bicellular tight junction assembly	8.90
🟦	blood vessel development	24.23
	angiogenesis	18.74
	regulation of biomineral tissue development	9.86
	muscle organ development	9.51
	artery morphogenesis	6.49
🟪	anterior/posterior pattern specification	15.61
	mesoderm development	8.65
	blood vessel development	7.07

Fig. S11. Gene expression changes along the inferred anterior and posterior endothelial differentiation trajectories.
(a) Force directed layouts of the hematoendothelial (top row) and endothelial landscapes (bottow row) showing the predicted subset of cells contributing to the anterior/medial (left column), posterior (middle column) and YS endothelial inferred differentiation trajectories coloured by pseudotime. **(b,c)** Heat maps displaying the top 300 genes associated with the anterior/medial **(d)** and posterior **(e)** inferred endothelial differentiation trajectories (Tradeseq associationTest, p<0.01 and meanLogFC > 2). Rows (genes) are hierarchically clustered together. Accompanying metadata including cell type annotations, stage, anatomy and predicted anatomy is displayed below the heatmap of gene expression. Mean gene expression values were computed and scaled by rows (Z-score). **(f,g)** Gene Ontology Biological Process term enrichment analysis was performed using Metascape on clusters of genes **(f: red, orange, grey; g: green, blue, purple, pink)** highlighted by the different colour blocks shown on the left of the heatmaps.

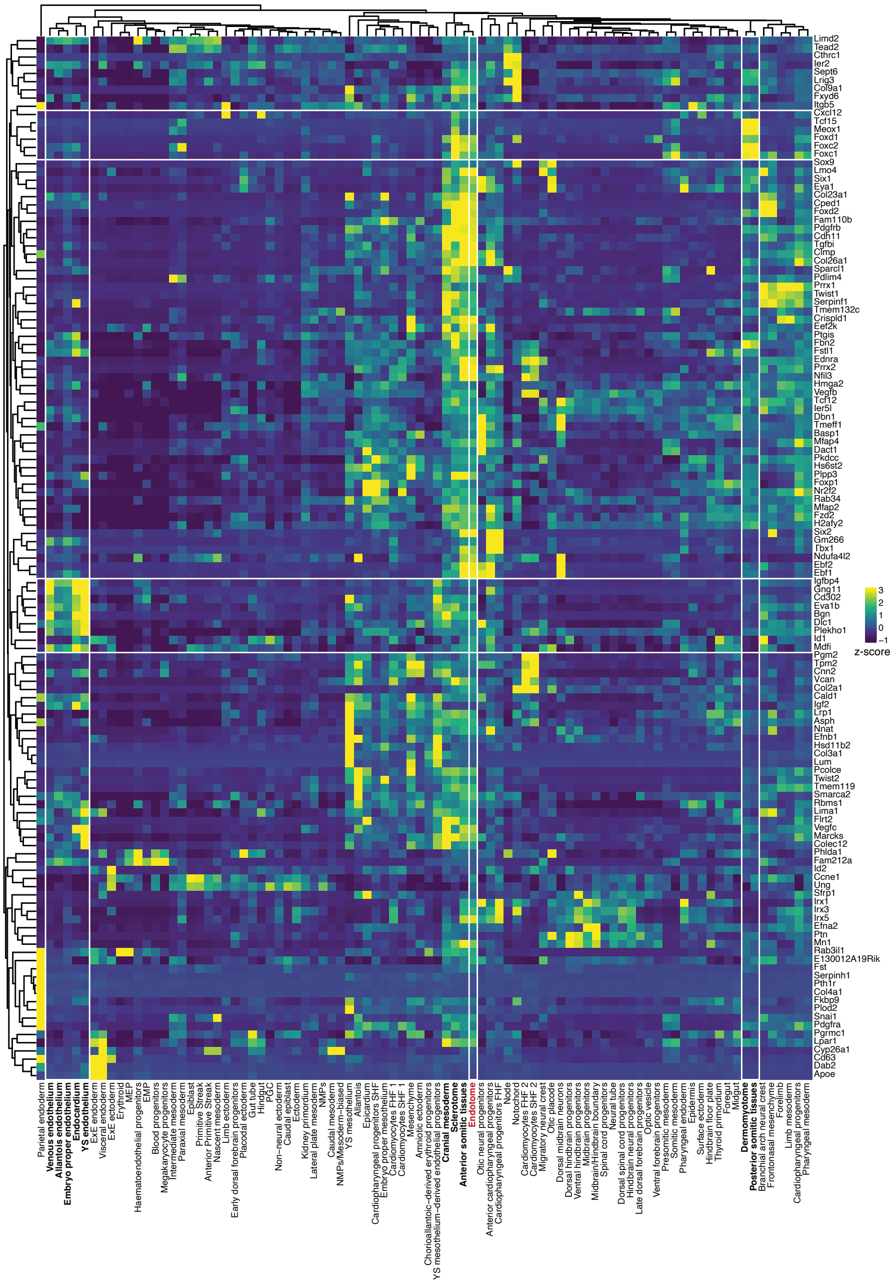


Fig. S12. Characterizing the gene expression signature of endotome cells. Heat map displaying positive genes markers that were identified for endotome cells (test = wilcox, logfc_threshold = 0.25, min_pct = 0.25), cell types (columns) and marker genes (rows) are clustered hierarchically. Mean gene expression values were computed and scaled by rows (Z-score). White boxes highlight endotome marker genes that are also expressed by endothelial cells or dermomyotome cell populations.

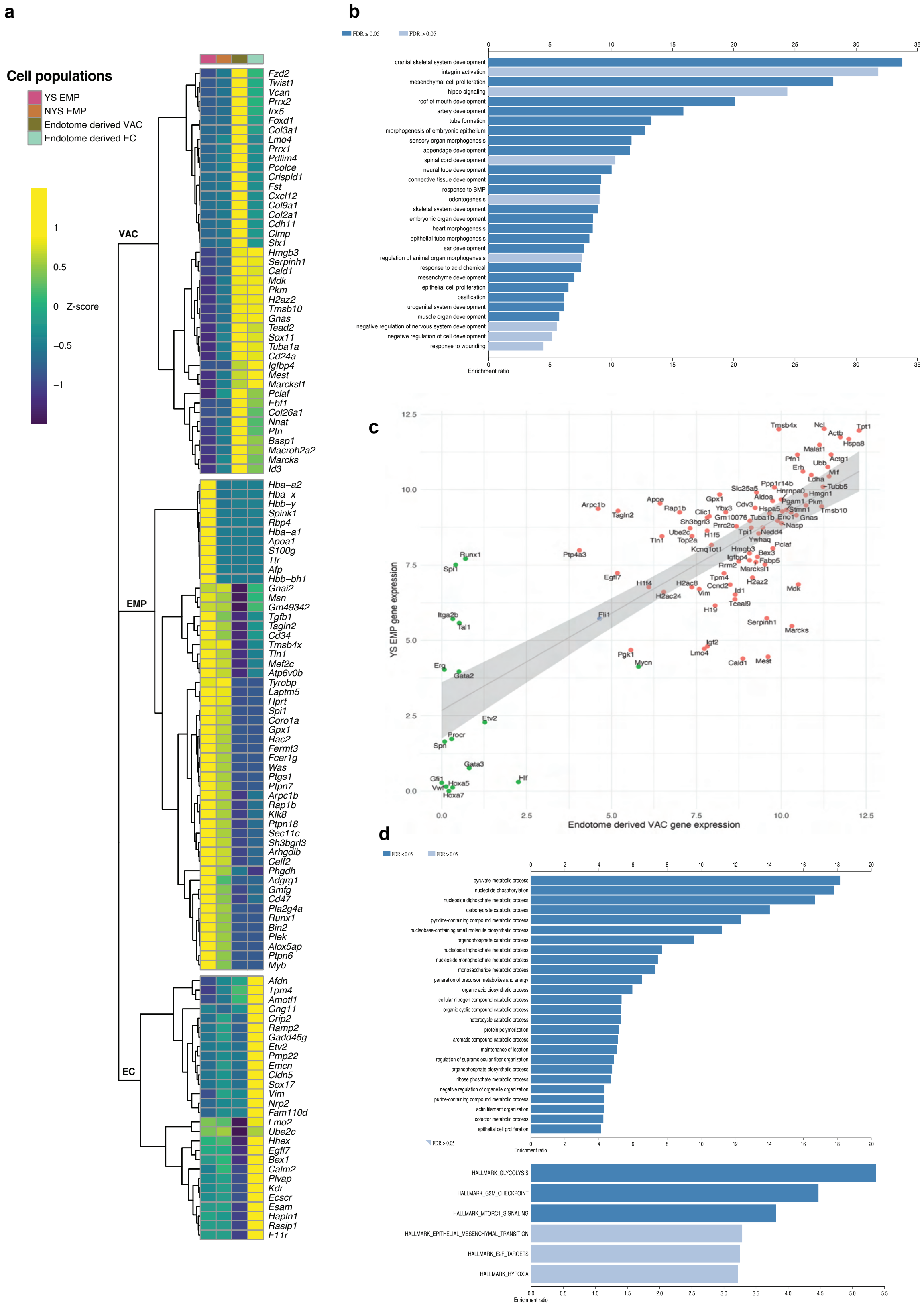

a
Cell populations
YS EMP
NYS EMP
Endotome derived VAC
Endotome derived EC
1
0.5
0 Z-score
−0.5
−1
VAC
EMP
EC
Fzd2
Twist1
Vcan
Prrx2
Irx5
Foxd1
Col3a1
Lmo4
Prrx1
Pdlim4
Pcolce
Crispld1
Fst
Cxcl12
Col9a1
Col2a1
Cdh11
Clmp
Six1
Hmgb3
Serpinh1
Cald1
Mdk
Pkm
H2az2
Tmsb10
Gnas
Tead2
Sox11
Tuba1a
Cd24a
Igfbp4
Mest
Marcksl1
Pclaf
Ebf1
Col26a1
Nnat
Ptn
Basp1
Macroh2a2
Marcks
Id3
Hba−a2
Hba−x
Hbb−y
Spink1
Rbp4
Hba−a1
Apoa1
S100g
Ttr
Afp
Hbb−bh1
Gnai2
Msn
Gm49342
Tgfb1
Tagln2
Cd34
Tmsb4x
Tln1
Mef2c
Atp6v0b
Tyrobp
Laptm5
Hprt
Spi1
Coro1a
Gpx1
Rac2
Fermt3
Fcer1g
Was
Ptgs1
Ptpn7
Arpc1b
Rap1b
Klk8
Ptpn18
Sec11c
Sh3bgrl3
Arhgdib
Celf2
Phgdh
Adgrg1
Gmfg
Cd47
Pla2g4a
Runx1
Bin2
Plek
Alox5ap
Ptpn6
Myb
Afdn
Tpm4
Amotl1
Gng11
Crip2
Ramp2
Gadd45g
Etv2
Pmp22
Emcn
Cldn5
Sox17
Vim
Nrp2
Fam110d
Lmo2
Ube2c
Hhex
Egfl7
Bex1
Calm2
Plvap
Kdr
Ecscr
Esam
Hapln1
Rasip1
F11r
b
FDR ≤ 0.05
FDR > 0.05
0
5
10
15
20
25
30
35
cranial skeletal system development
integrin activation
mesenchymal cell proliferation
hippo signaling
roof of mouth development
artery development
tube formation
morphogenesis of embryonic epithelium
sensory organ morphogenesis
appendage development
spinal cord development
neural tube development
connective tissue development
response to BMP
odontogenesis
skeletal system development
embryonic organ development
heart morphogenesis
epithelial tube morphogenesis
ear development
regulation of animal organ morphogenesis
response to acid chemical
mesenchyme development
epithelial cell proliferation
ossification
urogenital system development
muscle organ development
negative regulation of nervous system development
negative regulation of cell development
response to wounding
Enrichment ratio
c
12.5
10.0
7.5
5.0
2.5
0.0
YS EMP gene expression
0.0
2.5
5.0
7.5
10.0
12.5
Endotome derived VAC gene expression
Tmsb4x
Ncl
Actb
Tpt1
Hspa8
Malat1
Pfn1
Actg1
Ubb
Erh
Mif
Ldha
Tubb5
Ppp1r14b
Hnrnpa0
Hmgn1
Slc25a5
Gpx1
Aldoa
Pgam1
Pkm
Cdv3
Ybx3
Hspa5
Stmn1
Tmsb10
Apoe
Arpc1b
Rap1b
Tagln2
Clic1
Gm10076
Tuba1b
Eno1
Gnas
Sh3bgrl3
Prrc2c
Nasp
Ube2c
Tpi1
Nedd4
H1f5
Tln1
Top2a
Ywhaq
Pclaf
Kcnq1ot1
Hmgb3
Bex3
Igfbp4
Fabp5
Rrm2
Marcksl1
Runx1
Spi1
Ptp4a3
Egfl7
H1f4
H2ac8
Tpm4
H2az2
Ccnd2
Mdk
Vim
Id1
H2ac24
H19
Tceal9
Fli1
Itga2b
Tal1
Serpinh1
Marcks
Igf2
Lmo4
Pgk1
Mycn
Cald1
Mest
Erg
Gata2
Etv2
Procr
Spn
Gata3
Gfi1
Hoxa5
Vwf
Hoxa7
Hlf
d
FDR ≤ 0.05
FDR > 0.05
0
2
4
6
8
10
12
14
16
18
20
pyruvate metabolic process
nucleotide phosphorylation
nucleoside diphosphate metabolic process
carbohydrate catabolic process
pyridine-containing compound metabolic process
nucleobase-containing small molecule biosynthetic process
organophosphate catabolic process
nucleoside triphosphate metabolic process
nucleoside monophosphate metabolic process
monosaccharide metabolic process
generation of precursor metabolites and energy
organic acid biosynthetic process
cellular nitrogen compound catabolic process
organic cyclic compound catabolic process
heterocycle catabolic process
protein polymerization
aromatic compound catabolic process
maintenance of location
regulation of supramolecular fiber organization
organophosphate biosynthetic process
ribose phosphate metabolic process
negative regulation of organelle organization
purine-containing compound metabolic process
actin filament organization
cofactor metabolic process
epithelial cell proliferation
FDR > 0.05
Enrichment ratio
HALLMARK_GLYCOLYSIS
HALLMARK_G2M_CHECKPOINT
HALLMARK_MTORC1_SIGNALING
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
HALLMARK_E2F_TARGETS
HALLMARK_HYPOXIA
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
Enrichment ratio

Fig. S13. Gene expression signature of endotome and VAC populations.

a) Heat map displaying differentially expressed genes across different cell populations clustered together in the region highlighted in Fig.3. Dii. YS EMP: Yolk Sack EMP, NYS EMP: Non-Yolk Sack EMP, Endotome derived VAC: Endotome derived vascular associated cells (This heat map is an extended version of the analysis performed for Fig.3.E). Mean gene expression values were computed and scaled by rows (Z-score). **b)** Gene set over representation analysis (ORA) performed with *webgestalt*[Liao et al., 2019] against Gene Ontology Biological Processes of VAC genes highlighted in panel A. Significant terms (FDR < 0.1) are generally associated with connective tissue development. **c)** Gene correlation analysis between YS EMPs and endotome derived VACs (red dots: positively correlated genes identified by CCA, green dots: manually selected genes associated with HSC progenitors, blue dots: Intersecting genes between both). Metacells were used to strengthen the gene expression signals (see methods). **d)** ORA of positively correlated genes against hallmark gene sets (FDR < 0.1) and Gene Ontology Biological Processes (FDR < 0.05). The resulting hallmarks reflect cell growth (evidenced by cell cycle genes, mTORC1 signalling and E2f2 targets), glycolytic metabolism and *Myc* activation (the latter above the significance threshold) as well as epithelial-mesenchymal transition while no evidence of haematopoietic identity, but leaving the possibility of niche function akin to the situation observed in Zebrafish.

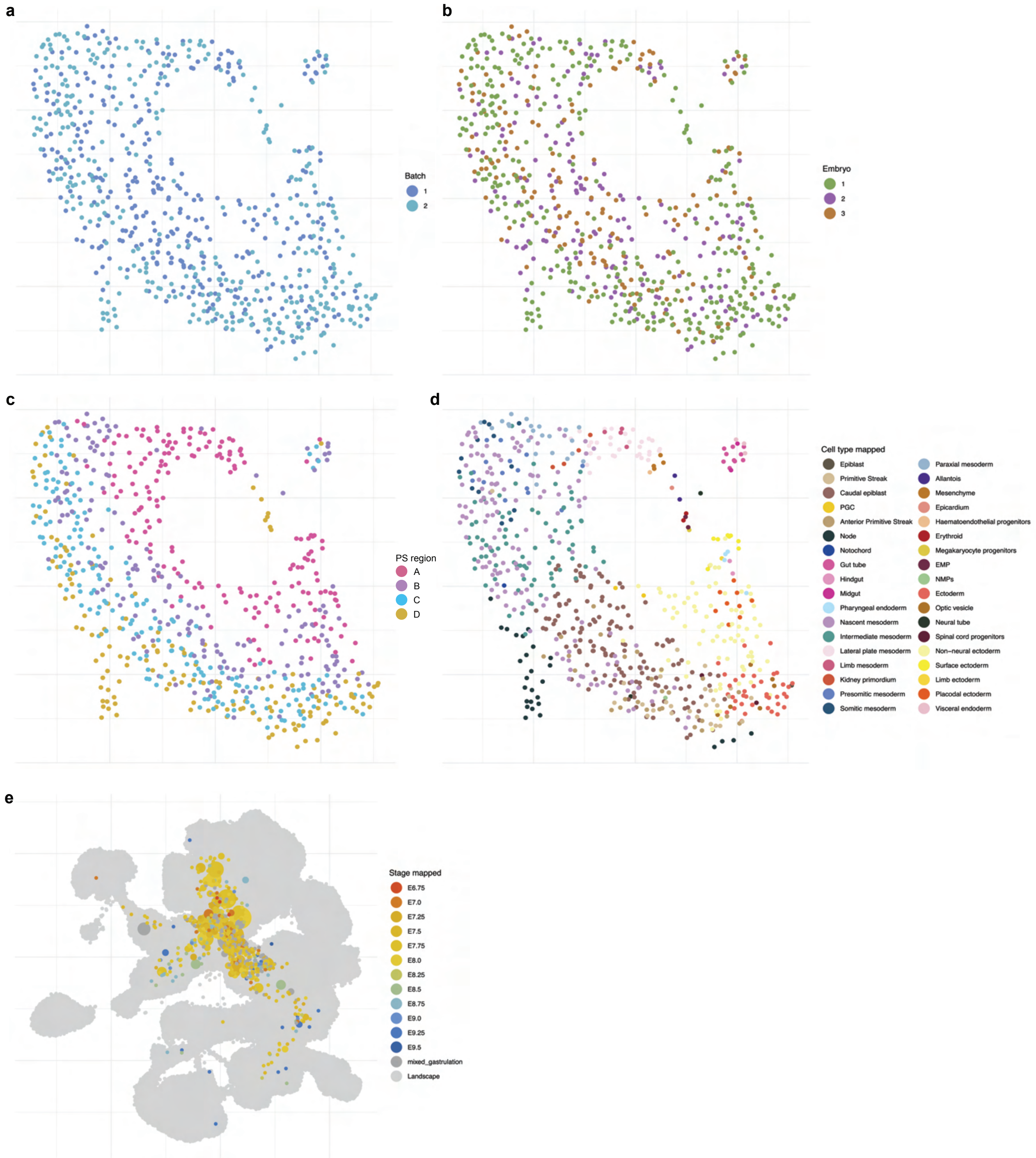


Fig. S14. Primitive streak dissections profiled with Smart-seq2.

UMAP layout of single cell transcriptomes from primitive streak dissections after batch correction (cells are coloured by batch). **b)** Three different embryos were profiled (cells are coloured by dissected embryos). **c)** Dissected portions of the Primitive Streak from Anterior to Posterior (cells are coloured by portions). **d)** Cell type annotations assigned to cells by transferring annotation from the whole atlas (cells are coloured by cell types). **e)** UMAP of primitive streak cells mapped onto the extended mouse atlas. Cells are coloured by the embryo stage corresponding to its closest neighbour in the atlas.

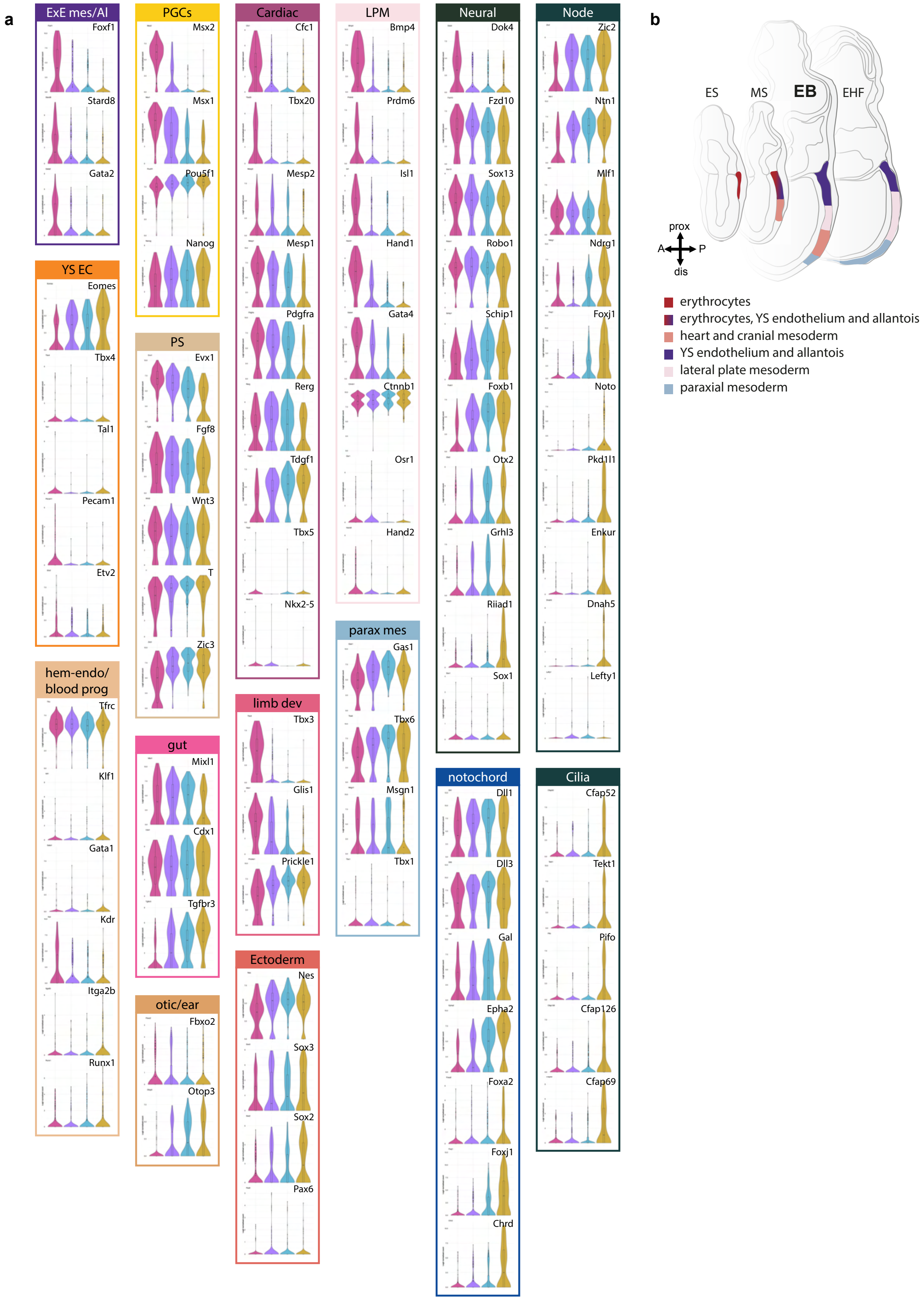
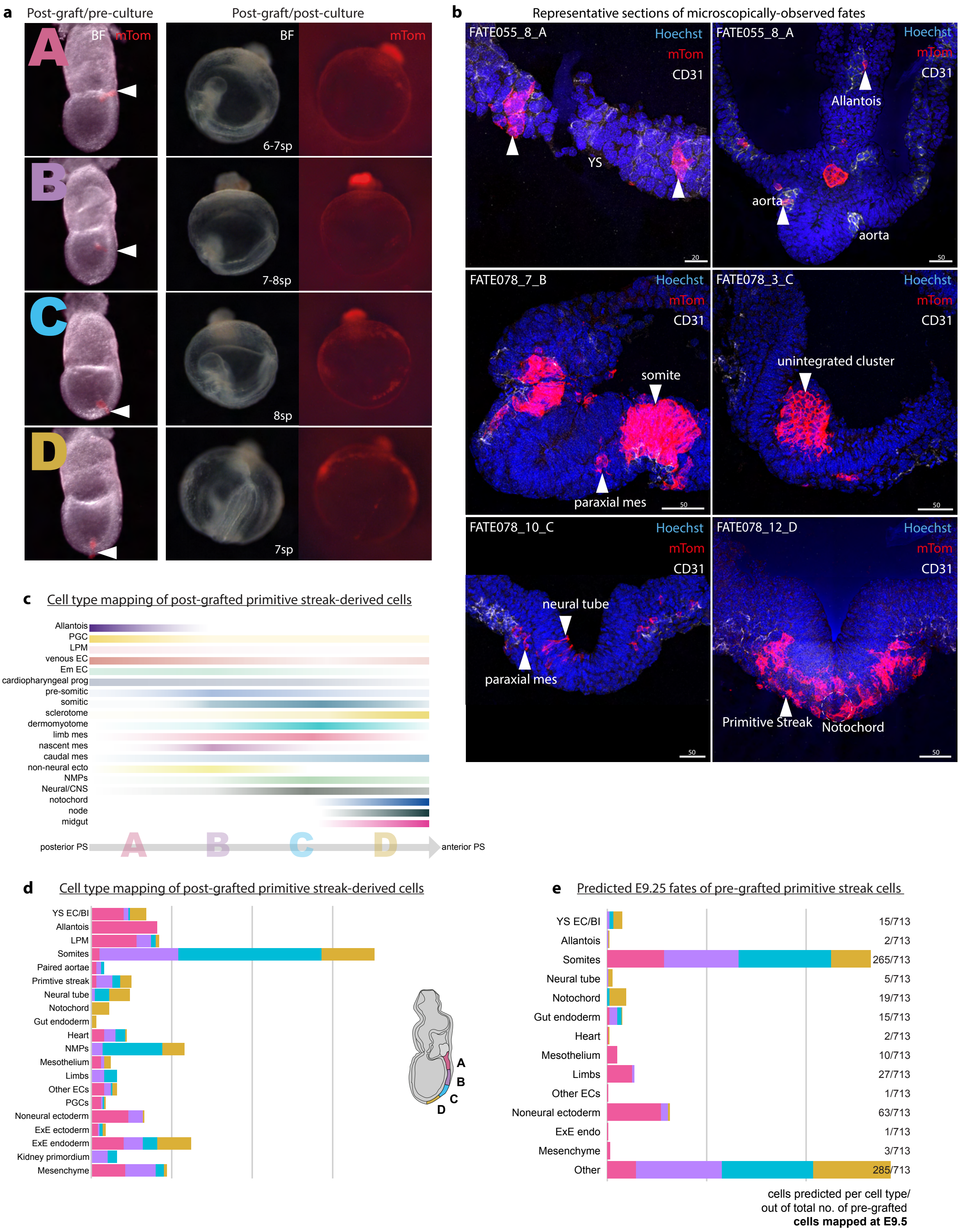

a
ExE mes/Al
Foxf1
Stard8
Gata2
YS EC
Eomes
Tbx4
Tal1
Pecam1
Etv2
hem-endo/ blood prog
Tfrc
Klf1
Gata1
Kdr
Itga2b
Runx1
PGCs
Msx2
Msx1
Pou5f1
Nanog
PS
Evx1
Fgf8
Wnt3
T
Zic3
gut
Mixl1
Cdx1
Tgfbr3
otic/ear
Fbxo2
Otop3
Cardiac
Cfc1
Tbx20
Mesp2
Mesp1
Pdgfra
Rerg
Tdgf1
Tbx5
Nkx2-5
limb dev
Tbx3
Glis1
Prickle1
Ectoderm
Nes
Sox3
Sox2
Pax6
LPM
Bmp4
Prdm6
Isl1
Hand1
Gata4
Ctnnb1
Osr1
Hand2
parax mes
Gas1
Tbx6
Msgn1
Tbx1
Neural
Dok4
Fzd10
Sox13
Robo1
Schip1
Foxb1
Otx2
Grhl3
Riiad1
Sox1
notochord
Dll1
Dll3
Gal
Epha2
Foxa2
Foxj1
Chrd
Node
Zic2
Ntn1
Mlf1
Ndrg1
Foxj1
Noto
Pkd1l1
Enkur
Dnah5
Lefty1
Cilia
Cfap52
Tekt1
Pifo
Cfap126
Cfap69
b
ES
MS
EB
EHF
prox
A
P
dis
erythrocytes
erythrocytes, YS endothelium and allantois
heart and cranial mesoderm
YS endothelium and allantois
lateral plate mesoderm
paraxial mesoderm

Fig. S15. Examples of cell type-specific gene signatures within primitive streak region A to D.

a) Violin plots of cell type-specific gene signatures within the 4 primitive streak regions are shown. Cell types are colour coded according to the UMAP legend (Fig.4b). **b)** Schematic representation of previously established primitive streak fates at different embryonic stages during gastrulation, adapted from Kinder et al., 2001.



Development • Supplementary information

Fig. S16. Grafting controls and full fate maps.

a) Wholemount images showing EB-stage wild type embryos orthotopically grafted with primitive streak regions A-D of mTom transgenic mouse embryos. Arrowheads point to the grafting site along the primitive streak axis with visible red cluster cells lodged in the primitive streak immediately after grafting and before embryo culture. After 24h culture, embryos developed around 7 somite pairs (Suppl.Table 1), had a beating heart and showed mTom contribution in their tissues. **b)** Representative immunofluorescent images of post-grafted/post-cultured embryo sections, where mTom contribution was assessed based on microscopic observations. **c)** Percentile representation of cell type mapping along the posterior-anterior axis of the primitive streak (from A to D, respectively) of post-grafted cells. The colour intensity indicates cell type abundance and resembles the gradient of cell type bias from specific primitive streak regions. **d)** Full fate map of transcriptionally observed fates based on the mapping of the post-grafted cells on the extended atlas. **e)** Full fate map of predicted fates of the pre-graft PS portion cells, determined with the Waddington-OT algorithm. Prediction was made including the full extended atlas, thus up to E9.25.

Table S1. Cell type markers across entire extended gastrulation atlas

Available for download at
https://journals.biologists.com/dev/article-lookup/doi/10.1242/dev.201867#supplementary-data

Table S2. Referencing genes markers for cell type annotation

Available for download at
https://journals.biologists.com/dev/article-lookup/doi/10.1242/dev.201867#supplementary-data

Table S3. Tradeseq Output: Genes that change along the YS Primitive, YS Definitive Blood and YS Endothelial Inferred Trajectories

Available for download at
https://journals.biologists.com/dev/article-lookup/doi/10.1242/dev.201867#supplementary-data

Table S4. Tradeseq Output: Genes that change along the Anterior/Medial Endothelial Inferred Trajectories with GO Terms for Clustered Genes

Available for download at
https://journals.biologists.com/dev/article-lookup/doi/10.1242/dev.201867#supplementary-data

Table S5. Tradeseq Output: Genes that change along the Posterior Endothelial Inferred Trajectories with GO Terms for Clustered Genes

Available for download at
https://journals.biologists.com/dev/article-lookup/doi/10.1242/dev.201867#supplementary-data

Table S6. Deeper characterization of endotome marker gene expression

Available for download at
https://journals.biologists.com/dev/article-lookup/doi/10.1242/dev.201867#supplementary-data

Table S7. CellComm Output: Ligand-Receptor Interactions in the YS Landscape and the Hematoendothelial Landscape

Available for download at
https://journals.biologists.com/dev/article-lookup/doi/10.1242/dev.201867#supplementary-data