

Direct specification of lymphatic endothelium from mesenchymal progenitors

Received: 3 April 2024

Accepted: 25 October 2024

Published online: 2 January 2025

 Check for updates

Irina-Elena Lupu ^{1,2,12}, David E. Grainger ^{1,2,12}, Nils Kirschnick ^{3,4,12}, Sarah Weischer ⁵, Erica Zhao ^{1,2}, Ines Martinez-Corral ⁶, Hans Schoofs ⁶, Marie Vanhollebeke ^{1,2}, Grace Jones ^{1,2}, Jonathan Godwin ^{1,2}, Aden Farrow ⁷, Ines Lahmann ^{8,9}, Paul R. Riley ^{1,2}, Thomas Zobel ⁵, Kari Alitalo ^{10,11}, Taija Mäkinen ^{6,10}, Friedemann Kiefer ^{3,4}  & Oliver A. Stone ^{1,2} 


During embryogenesis, endothelial cells (ECs) are generally described to arise from a common pool of progenitors termed angioblasts, which diversify through iterative steps of differentiation to form functionally distinct subtypes of ECs. A key example is the formation of lymphatic ECs (LECs), which are thought to arise largely through transdifferentiation from venous endothelium. Opposing this model, here we show that the initial expansion of mammalian LECs is primarily driven by the in situ differentiation of mesenchymal progenitors and does not require transition through an intermediate venous state. Single-cell genomics and lineage-tracing experiments revealed a population of paraxial mesoderm-derived *Etv2*⁺*Prox1*⁺ progenitors that directly give rise to LECs. Morphometric analyses of early LEC proliferation and migration, and mutants that disrupt lymphatic development supported these findings. Collectively, this work establishes a cellular blueprint for LEC specification and indicates that discrete pools of mesenchymal progenitors can give rise to specialized subtypes of ECs.

The lymphatic vasculature is a blunt-ended, unidirectional vessel network essential for the maintenance of tissue fluid homeostasis, immune cell trafficking and lipid absorption¹. Although defects in the formation and function of lymphatic vessels have been described in a broad range of diseases, including lymphedema, cancer, obesity and neurodegeneration¹, our understanding of the mechanisms governing their formation and function has historically trailed behind that of the cardiovascular system². The inner lining of lymphatic vessels is formed

by a single layer of lymphatic endothelial cells (LECs), which have been shown to play important roles in organ development³, homeostasis⁴ and regeneration³. Despite these key functions, our understanding of how LECs are specified during embryonic development remains incomplete, with the developmental origins of LECs debated for over a century^{5–11}.

During embryogenesis, formation of the vertebrate vasculature begins with the de novo specification of endothelial cells (ECs) from mesoderm-derived mesenchymal progenitors known as angioblasts^{12,13}.

¹Department of Physiology, Anatomy & Genetics, University of Oxford, Oxford, UK. ²Institute of Developmental & Regenerative Medicine, University of Oxford, Oxford, UK. ³University of Münster, European Institute for Molecular Imaging, Münster, Germany. ⁴Max Planck Institute for Molecular Biomedicine, Münster, Germany. ⁵University of Münster, Cells in Motion Interfaculty Centre, Imaging Network, Münster, Germany. ⁶Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden. ⁷Mathematical Institute, University of Oxford, Oxford, UK. ⁸Developmental Biology/Signal Transduction, Max-Delbrück-Center for Molecular Medicine, Berlin, Germany. ⁹Neurowissenschaftliches Forschungszentrum, NeuroCure Cluster of Excellence, Charité-Universitätsmedizin Berlin, Berlin, Germany. ¹⁰Wihuri Research Institute, Helsinki, Finland. ¹¹Translational Cancer Medicine Program, Biomedicum Helsinki, University of Helsinki, Helsinki, Finland. ¹²These authors contributed equally: Irina-Elena Lupu, David E. Grainger, Nils Kirschnick.

 e-mail: fkiefer@uni-muenster.de; oliver.stone@dpag.ox.ac.uk

In response to locally deposited growth factors, angioblasts further differentiate to form ECs that migrate and coalesce to form the first functional vessels. It is from these initial vessels that most blood and lymphatic vessel networks are thought to emerge¹⁴. Early histological assessment of lymphatic vessel formation in mammalian embryos described venous endothelium⁸ and mesenchymal cells⁵ as sources of lymphatics. However, loss-of-function¹⁵ and genetic lineage-tracing⁹ analyses in mouse later indicated that LECs mainly transdifferentiate from venous endothelium, with live imaging of lymphatic development in zebrafish showing that a similar process occurs in teleosts^{7,11}. In contrast, live imaging of facial lymphatic specification in zebrafish¹⁶, chimeric transplantation analyses in avian embryos¹⁰ and distinct lineage-tracing analyses in mouse^{6,17} revealed alternative non-venous sources of LECs in various tissues.

Using genetic lineage tracing in mouse, we previously showed that the overarching source of most mammalian LECs is the paraxial mesoderm¹⁸. These detailed imaging analyses revealed that paraxial mesoderm-derived cells contribute to venous endothelium and form LECs in most tissues. Here, using a combination of single-cell genomics, inducible lineage tracing and high-resolution imaging, we show that LECs can arise directly from a specialized pool of *Etv2*⁺*Prox1*⁺ angioblasts, in a spatiotemporal pattern that mimics migration from venous endothelium into the surrounding mesenchyme.

Results

Single-cell analysis of EC specification

Immunofluorescence analyses of *Pax3*^{Cre/+};*Rosa26*^{tdTomato} embryos, where paraxial mesoderm-derived LECs and their ancestors are labeled by tdTomato¹⁸, revealed that VEGFR2⁺ETV2⁺ angioblasts begin to emerge from somitic mesoderm at embryonic day (E) 8.25 (Fig. 1a–b'). Subsequently, paraxial mesoderm-derived cells contribute to the common cardinal vein (CV) and intersegmental vessels at E8.75 (Extended Data Fig. 1a), the sinus venosus, cardinal, umbilical and vitelline veins at E9.25 (Extended Data Fig. 1b,c) and E9.5 (Fig. 1c) and PROX1⁺ ECs sitting both inside and outside of the CV at E10.5 (Fig. 1d). Flow cytometry analysis of dissected *Pax3*^{Cre/+};*Rosa26*^{tdTomato} embryos at E13.5 showed that, although the proportion of paraxial mesoderm-derived blood ECs (BECs) is relatively limited (Extended Data Fig. 2a,b), most LECs are paraxial mesoderm derived at this developmental stage (Extended Data Fig. 2a,b).

To more precisely define molecular transitions during LEC differentiation, we performed single-cell RNA sequencing (scRNA-seq) at the onset of LEC specification (E9.5), during the emergence of LEC progenitors from venous endothelium (E10.5) and as primordial thoracic duct (pTD) formation begins (E11.5) (Fig. 1e)^{18,19}. Due to the restricted contribution of the *Pax3* lineage to LECs of the trunk²⁰, at each stage we dissected embryos at the level of the otic vesicle and first pharyngeal arch (Fig. 1e). To further enrich for cells of interest, at E9.5

VEGFR2⁺ cells were FACS-sorted with the goal of capturing ECs as well as any residual VEGFR2⁺ progenitors differentiating from mesoderm²¹ (Extended Data Fig. 2c). To maximize the diversity of EC subtypes collected at E10.5/E11.5, we captured cells that expressed VEGFR2 and/or PECAM1 (Extended Data Fig. 2d,e). A total of 19,699 high-quality cells were collected across three developmental stages (Extended Data Fig. 3a,b), with clustering of the merged dataset using Seurat version 4 revealing 30 distinct cellular states (Extended Data Fig. 3c,d). For visualization and to characterize putative transitions between cellular states, we used partition-based graph abstraction (PAGA) followed by ForceAtlas2 embedding (Fig. 1f). Each cellular state was comprised, to differing extents, of cells from each lineage (Fig. 1g and Extended Data Fig. 3d) and developmental stage (Fig. 1h and Extended Data Fig. 3d) and included paraxial mesoderm (*Pax3*⁺, *Lbx1*⁺, *Prrx2*⁺), angioblasts (*Etv2*⁺, *Tal1*⁺), venous ECs (*Cdh5*⁺, *Nr2f2*⁺, *Dab2*⁺), artery ECs (*Cdh5*⁺, *Gja5*⁺, *Bmx*⁺), LEC1 (*Prox1*⁺) and LEC2 (*Prox1*⁺, *Lyve1*⁺) (Fig. 1i and Extended Data Fig. 4a). Differential gene expression analyses between *Pax3* lineage-positive and lineage-negative ECs identified a number of genes with lineage-biased expression patterns (Extended Data Fig. 4b), including enhanced expression of genes linked to retinoic acid signaling (*Cyp26b1*, *Rbp1*, *Strat6*), a signaling pathway previously linked to lymphatic development^{22,23}. Cluster identities were experimentally validated in E10.5 embryos using immunofluorescence and fluorescence in situ hybridization (Extended Data Fig. 5a–d'). We found that VWF (Extended Data Fig. 5a–b'), enriched in venous and arterial clusters (Fig. 1f and Extended Data Fig. 4a), was expressed throughout the dorsal aorta and CV but not by *Prox1*⁺ ECs in the surrounding mesenchyme. Expression of *Lyve1*, which was enriched in the venous, sinus venosus and liver EC clusters (Fig. 1f and Extended Data Fig. 4a), was restricted to the dorsal CV and not expressed by *Prox1*⁺ ECs in the surrounding mesenchyme at this stage (Extended Data Fig. 5c–d').

LECs arise directly from specialized mesenchymal progenitors

To gain further insight into the cell state transitions that occur during LEC specification, we used PAGA-based trajectory inference. These analyses revealed known developmental transitions within our dataset, including a trajectory from sinus venosus ECs to liver ECs²⁴ (Extended Data Fig. 5e), and indicated that LECs can arise directly from angioblasts without transition through a venous EC state (Extended Data Fig. 5e). These findings were supported by Waddington Optimal Transport (Waddington-OT) (Extended Data Fig. 5f–k). In agreement with these findings, detailed examination of gene expression revealed an angioblast-like cellular state comprising cells that express mesenchymal markers, including *Prrx2* (ref. 25), the endothelial pioneer factor *Etv2* (ref. 26) and moderate expression of the master regulator of LEC fate *Prox1* (Fig. 1i). Furthermore, co-expression analyses identified single cells that express both *Etv2* and *Prox1* (Fig. 1j), which may represent bona fide lymphangioblasts.

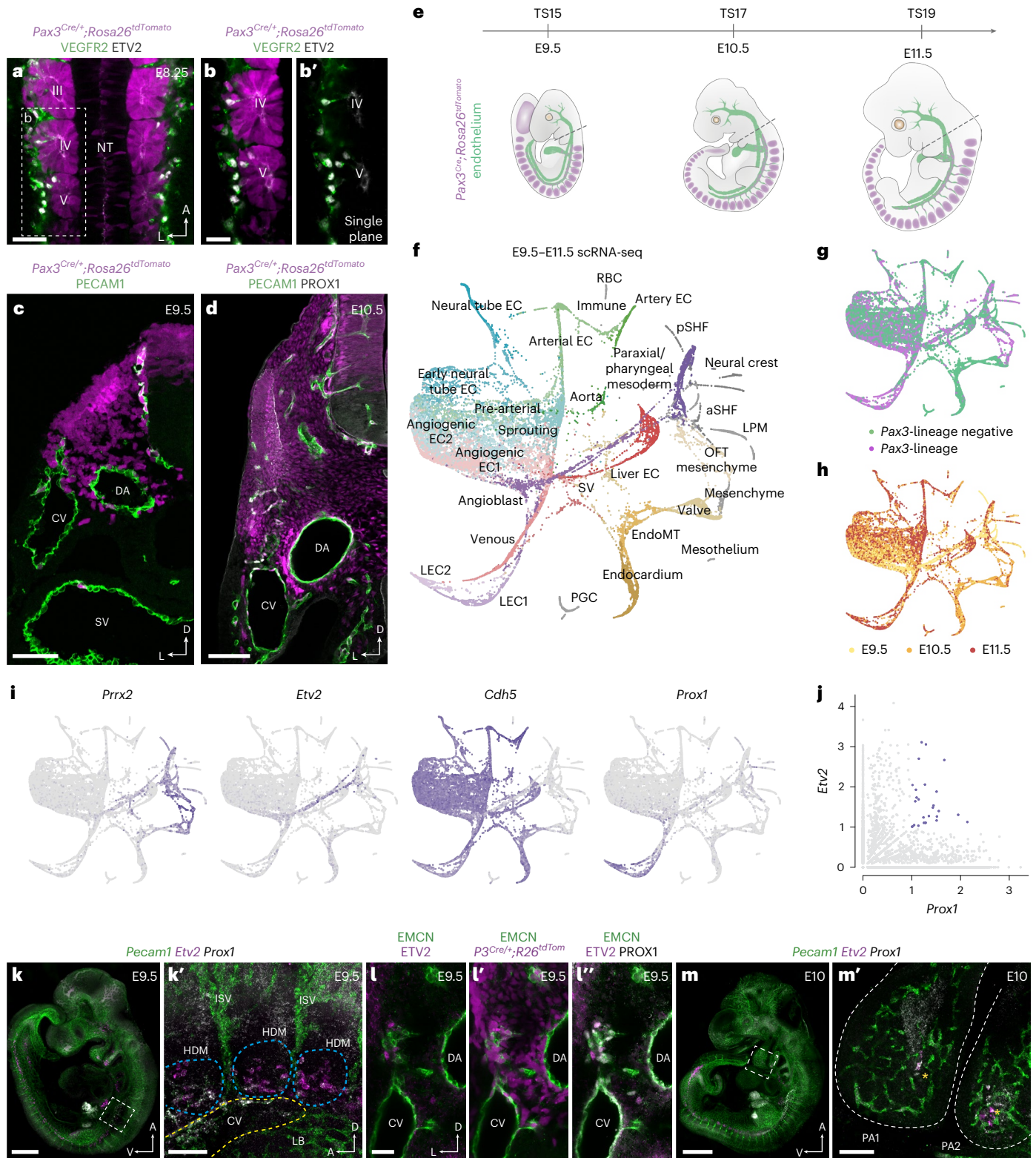
Fig. 1 | Spatiotemporal analyses of EC specification. **a**, Representative whole-mount immunofluorescence for tdTomato, VEGFR2 and ETV2 in an E8.25 *Pax3*^{Cre/+};*Rosa26*^{tdTomato} embryo. Dorsal view at the level of somites III–V ($n = 4$). **b, b'**, High-magnification single confocal plane of boxed area in **a**, highlighting a population of VEGFR2⁺ETV2⁺ angioblasts emerging from the lineage-traced somite. **c**, Representative immunofluorescence for tdTomato and PECAM1 on a transverse cryosection from a *Pax3*^{Cre/+};*Rosa26*^{tdTomato} embryo at E9.5 ($n = 6$). **d**, Representative immunofluorescence for tdTomato, PECAM1 and PROX1 on a transverse vibratome section from a *Pax3*^{Cre/+};*Rosa26*^{tdTomato} embryo at E10.5 ($n = 6$). **e**, Schematic highlighting embryonic stages and dissection strategy (dashed line) for scRNA-seq analyses. **f**, ForceAtlas2 (FA) embedding of 19,699 cells based on PAGA, with each dot representing a single cell. Cellular states were manually annotated based on known gene expression patterns. **g**, FA embedding showing the relationship between cell lineage and cell state. **h**, FA embedding showing the relationship between embryonic stage and cell state. **i**, FA embedding showing gene expression of mesenchymal (*Prrx2*),

angioblast (*Etv2*), pan-endothelial (*Cdh5*) and lymphatic endothelial (*Prox1*) markers. **j**, Scatterplot showing co-expression of *Etv2* and *Prox1* in single cells. **k, k'**, Representative whole-mount analysis of *Pecam1*, *Etv2* and *Prox1* expression at E9.5 using HCR shows *Etv2* and *Prox1* co-expressing cells in the hypaxial somite ($n = 6$). **l–l'**, Representative immunofluorescence for EMCN, ETV2, PROX1 and tdTomato on transverse vibratome sections from a *Pax3*^{Cre/+};*Rosa26*^{tdTomato} embryo at E9.5 ($n = 6$). **m, m'**, Representative whole-mount analysis of *Pecam1*, *Etv2* and *Prox1* expression at E10.0 using HCR shows *Etv2* and *Prox1* co-expressing cells in the pharyngeal arches ($n = 6$). TS, Thielier stage; aSHF, anterior second heart field; CCV, common cardinal vein; DA, dorsal aorta; HDM, hypaxial dermomyotome; ISV, intersegmental vessel; LB, limb bud; LPM, lateral plate mesoderm; NC, mesenchyme, neural crest-derived mesenchyme; NT, neural tube; OFT, outflow tract; PA, Pharyngeal arch; PGC, primordial germ cell; pSHF, posterior second heart field; RBC, red blood cell; SHF, second heart field; SV, sinus venosus. Scale bars, 50 μ m (**a, c** and **l–l'**), 25 μ m (**b**), 100 μ m (**d, k'** and **m'**), 500 μ m (**k**), 1 mm (**m**).

Collectively, these data suggest that LECs may arise directly from a dedicated pool of *Etv2*⁺*Prox1*⁺ angioblasts.

To characterize the spatiotemporal distribution of *Etv2*⁺*Prox1*⁺ cells during development, we performed hybridization chain reaction (HCR) and immunofluorescence on whole embryos and vibratome sections between E9.0 and E10.0 (Fig. 1k–m' and Extended Data Fig. S1–o'). Whole-mount imaging of *Etv2*, *Prox1* and *Pecam1* expression identified *Etv2*⁺*Prox1*⁺*Pecam1*⁺ angioblasts within the somitic paraxial mesoderm

at E9.0 (Extended Data Fig. S1–m'). By E9.5, *Etv2* expression was reduced in the anterior somites (Fig. 1k), where strong *Prox1* expression was observed anterior to the forelimb bud, as previously described¹⁵. High-resolution imaging of *Etv2* and *Prox1* co-expression revealed a population of *Etv2*⁺*Prox1*⁺*Pecam1*⁺ angioblasts emerging from the hypaxial dermatomyotome at E9.5 (Fig. 1k'). These findings were confirmed with immunofluorescence imaging in *Pax3*^{Cre/+};*Rosa26*^{tdTomato} embryos, showing that paraxial mesoderm-derived, ETV2⁺PROX1⁺



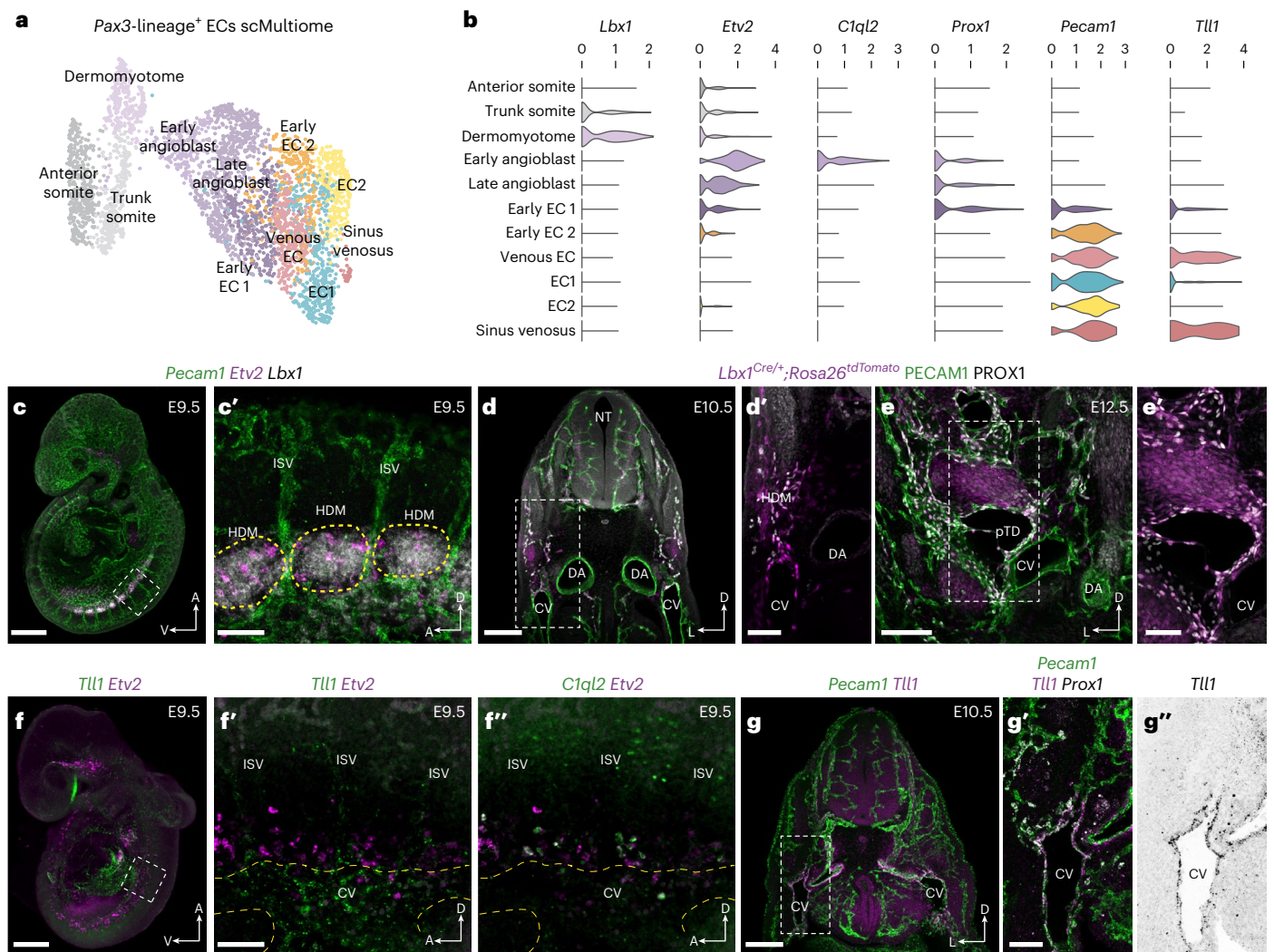


Fig. 2 | Multiomic analysis of paraxial mesoderm-derived endothelium. **a**, Weighted nearest neighbor UMAP (wnnUMAP) embedding of single-cell multiome analyses of 3,606 tdTomato⁺ VEGFR2⁺ PECAM1⁺ cells FACS-sorted from *Pax3*^{Cre/+};*Rosa26*^{tdTomato} embryos at E9.5. Cellular states were manually annotated based on known gene expression patterns and downstream analyses. **b**, Violin plots showing expression of selected genes across each cellular state. **c**, Representative whole-mount analysis of *Pecam1*, *Etv2* and *Lbx1* expression at E9.5 using HCR (*n* = 6). **c'**, High-magnification image of boxed area in **c** highlighting a population of *Etv2*⁺ angioblasts emerging from the *Lbx1*⁺ hypaxial dermomyotome. **d, d'**, Representative immunofluorescence for tdTomato, PECAM1 and PROX1 on a transverse vibratome section from an *Lbx1*^{Cre/+};*Rosa26*^{tdTomato} embryo at E10.5 (*n* = 4). **e, e'**, Representative

immunofluorescence for tdTomato, PECAM1 and PROX1 on a transverse vibratome section from an *Lbx1*^{Cre/+};*Rosa26*^{tdTomato} embryo at E12.5 (*n* = 4). **f**, Representative whole-mount analysis of *Tll1* and *Etv2* expression at E9.5 using HCR (*n* = 6). High-magnification images of boxed area in **f** highlighting *Tll1* expression in CV ECs (**f'**) and *C1ql2* expression in a subset of *Etv2*⁺ angioblasts (**f''**). **g**, Representative HCR analysis of *Pecam1* and *Tll1* on a vibratome section from an E10.5 embryo. **g'**, **g''**, High-magnification images of boxed area in **g** showing *Pecam1*, *Tll1* and *Prox1* expression (*n* = 4). CV, cardinal vein; DA, dorsal aorta; HDM, hypaxial dermomyotome; ISV, intersegmental vessel; pTD, primordial thoracic duct. Scale bars, 500 μm (**c** and **f**), 100 μm (**c'**, **d'**, **e'** and **g'**), 200 μm (**d** and **g**), 50 μm (**e'**).

angioblasts differentiate within mesenchyme surrounding the CV at E9.5 (Fig. 1l–l''). At E10, *Etv2*⁺ cells were largely absent from the lymphatic anlage anterior to the limb bud, highlighting the transient nature of this cellular state (Extended Data Fig. 5n–o'). Notably, analysis of *Etv2*, *Prox1* and *Pecam1* expression revealed the presence of *Etv2*⁺ *Prox1*⁺ *Pecam1*⁻ angioblasts in the pharyngeal mesoderm (Fig. 1m, m'), a known source of cardiac LECs^{27,28}. Collectively, these analyses identify specialized angioblasts for LECs that arise directly from paraxial and pharyngeal mesoderm.

To better characterize the molecular transitions associated with LEC specification from paraxial mesoderm, we performed single-cell Multiome (scMultiome) analyses of *Pax3* lineage-positive VEGFR2⁺ nuclei isolated from *Pax3*^{Cre/+};*Rosa26*^{tdTomato} embryos at E9.5 (Extended Data Fig. 6a). A total of 3,801 high-quality single cells were collected

(Extended Data Fig. 6b–d), with clustering of the merged dataset using Seurat version 5 revealing 12 distinct cellular states (Extended Data Fig. 6d). To focus on cells of interest, we computationally removed the neural, neural crest and mixed mesoderm populations. Re-clustering of this subset of 3,606 cells revealed 11 cellular states (Fig. 2a, b and Extended Data Fig. 6e, f), including dermomyotome (*Lbx1*⁺, *Pax3*⁺), early angioblast (*Etv2*^{High}, *C1ql2*⁺, *Tal1*⁺, *Lmo2*⁺), late angioblast (*Etv2*⁺, *Prox1*⁺, *Pecam1*⁻), early EC1 (*Prox1*⁺, *Pecam1*⁺), early EC2 (*Prox1*^{low}, *Pecam1*^{High}), venous ECs (*Tll1*⁺) and sinus venosus ECs (*Gata4*⁺, *Tll1*⁺). Expression of the artery marker *Gja5* suggests that there is no clear contribution of paraxial mesoderm-derived cells to arterial endothelium (Extended Data Fig. 6f), as previously observed in our scRNA-seq analyses across multiple stages (Extended Data Fig. 3d). We experimentally validated cluster identity and differentiation potential

using HCR, immunofluorescence imaging and genetic lineage tracing (Fig. 2c–g’). Combined analysis of *Pecam1*, *Etv2* and *Lbx1* expression showed that *Etv2*⁺*Pecam1*⁺ angioblasts are specified within the *Lbx1*⁺ hypaxial domain of the dermomyotome²⁹ at E9.5 (Fig. 2c,c’). Immunofluorescence imaging of lineage-traced *Lbx1*^{Cre/+}; *Rosa26*^{tdTomato} embryos showed tdTomato labeling of PROX1⁺ ECs in the dorsal portion of the CV and surrounding mesenchyme at E10.5 (Fig. 2d,d’) and the pTD and jugular lymph sac (JLS) at E12.5 (Fig. 2e,e’). Whole-mount imaging of *Tll1* and *Etv2* showed restricted expression of *Tll1* in the anterior portion of the CV as well as substantial labeling of the sinus venosus at E9.5 (Fig. 2f,f’). Limited expression of *Tll1* was also observed in venous intersegmental vessels but did not overlap with expression of *Etv2* (Fig. 2f’). Notably, imaging of *C1ql2* and *Etv2* revealed co-expression in a subset of angioblasts (Fig. 2f’), supporting our scMultiome analyses (Fig. 2a,b). Imaging of transverse sections from E10.5 embryos revealed restricted expression of *Tll1* in the CV endothelium (Fig. 2g), which was absent from *Prox1*⁺ ECs in the surrounding mesenchyme (Fig. 2g’,g’). *Tll1* expression was also observed in mural cells surrounding the dorsal aorta (Fig. 2g–g’). Collectively, these analyses validate our scMultiome sequencing, identify molecular differences between *Prox1*⁺ ECs found in the CV and those found in the surrounding mesenchyme and suggest that a considerable proportion of *Prox1*⁺ ECs are specified directly from angioblasts.

To investigate transcription factor (TF) activity and chromatin dynamics during LEC specification, we performed gene regulatory network (GRN) analyses and assessed chromatin composition and putative TF binding at the *Prox1* locus (Fig. 3a–f). We used SCENIC+ to identify enhancer-driven GRNs (eGRNs) comprising *cis*-regulatory elements, upstream TFs and candidate target genes (Fig. 3b and Extended Data Fig. 6h,i). These analyses highlighted the roles of ETS TFs during EC specification (ETV2; Fig. 3b and Extended Data Fig. 6h) and differentiation (FLI1, ETV6, ETS2, ELK3, ETS1, ERG; Fig. 3b and Extended Data Fig. 6h) and identified known regulators of EC gene expression (MEF2A, MEF2C; Fig. 3b) and LEC specification (SOX18; Fig. 3b)³⁰. Additionally, eGRN analyses revealed putative regulators that may play roles in the differentiation of ECs and LECs (Fig. 3b), including regulatory nodes driven by the early B cell factor family members EBF1 and EBF2 (Fig. 3b and Extended Data Fig. 6g,i). EBF1 is a pioneer factor required for B cell differentiation³¹ and neuronal development³², whereas EBF2 regulates bone formation³³ and specification of brown adipose tissue³⁴. Network analyses revealed *cis*-regulatory elements and target genes that may be co-regulated by ETV2 and EBF1/EBF2, indicating that these TFs may also act during EC specification (Extended Data Fig. 6i).

In a more targeted approach, we investigated regulation of *Prox1* (Fig. 3c–f). We characterized the expression of known regulators of *Prox1*, which have been studied at later stages of embryonic development or in mature LECs^{30,35}, including *Sox18*, *Gata2*, *Nfatc1* and *Foxc2* (Fig. 3c). Notably, these TFs are expressed in angioblasts and, thus, may play similar roles in the initial induction of *Prox1* expression during LEC specification from angioblasts (Fig. 3c). To assess chromatin composition at the *Prox1* locus, we generated pseudobulk assay for transposase-accessible chromatin with sequencing (ATAC-seq) profiles by combining reads from individual cells within clusters and contrasted these with bulk ATAC-seq data from FACS-sorted LECs and BECs (Fig. 3d). These analyses identified a number of putative *cis*-regulatory elements that display increased accessibility during differentiation from dermomyotome to endothelium (Fig. 3d,e), including a previously described enhancer element that sits 11 kb upstream of the *Prox1* transcription start site (mm10:chr1:190,181,400–190,182,400 bp)³⁶. Additionally, we identified elements at +63 kb (mm10:chr1:190,106,700–190,107,700), –330 kb (mm10:chr1:190,500,600–190,501,600), –347 kb (mm10:chr1:190,517,800–190,518,800), –362 kb (mm10:chr1:190,532,650–190,533,650) and –371 kb (mm10:chr1:190,541,800–190,542,400), which are accessible at the onset of *Prox1* expression in angioblasts (Fig. 3c,e). To investigate regulation of *Prox1* by these

putative *cis*-regulatory elements, we performed ATAC-seq footprinting using TOBIAS (Fig. 3f). TOBIAS is a digital genomic footprinting tool that maps Tn5 insertion events to characterize TF occupancy at known binding motifs. As individual motifs may be bound by multiple TFs from the same family in a manner that is indistinguishable to TOBIAS, we highlighted motifs bound by TF family members (for example, ETS, SOXF, GATA, NFAT, FOX) rather than binding by individual TFs (for example, ETV2, ERG, SOX18, GATA2, NFATC1, FOXC2) (Fig. 3f). These analyses suggest that described regulators of *Prox1* may bind to proximal promoter elements and *cis*-regulatory regions to drive *Prox1* expression in angioblasts and early ECs, displaying TF co-occupancy at several of these regulatory elements (Fig. 3f). As extensive binding of ETS TFs is predicted in angioblasts, we probed the possibility of direct regulation of *Prox1* by ETV2 by comparing our pseudobulk and bulk ATAC profiles with published ETV2 chromatin immunoprecipitation followed by sequencing (ChIP-seq) analyses in mouse embryonic stem cells³⁷. These analyses identified an ETV2 ChIP-seq peak overlapping the accessible element at –371 kb, which was also predicted to be bound by ETS TFs in early angioblasts (Fig. 3f); it is possible that ETV2 also binds additional elements in the *Prox1* locus in angioblasts that are not bound in embryonic stem cells. Enhancer usage is known to be highly dynamic during embryonic development, with individual enhancers driving stage-specific biological functions³⁸. Our data suggest dynamic remodeling of chromatin composition and TF binding within the *Prox1* locus during development, as we observed that regions of highly accessible chromatin in E9.5 angioblasts and ECs are closed in E13.5 LECs, whereas other regions display increased accessibility at later stages (Fig. 3d,e). Collectively, our analyses suggest that LEC specification from paraxial mesoderm is accompanied by chromatin remodeling and TF binding in angioblasts and implicate a number of known regulators of *Prox1* in this process.

Temporal labeling of paraxial mesoderm derivatives

We next sought to assess the anatomical distribution of LECs derived from the paraxial mesoderm at distinct developmental stages using a tamoxifen-inducible *Pax3*^{CreERT2} driver³⁹. For validation, we performed immunofluorescence with an ESRI antibody to detect CreERT2. Strong CreERT2 expression was detected in the dorsal neural tube and dermomyotome at E9.5 (Extended Data Fig. 7a–b’) and E10.5 (Extended Data Fig. 7c–d’). Notably, in agreement with previous reports²¹, we found that *Pax3*-driven CreERT2 expression overlapped with VEGFR2 in the hypaxial dermomyotome at E9.5 (Extended Data Fig. 7a–b’) and was absent from PECAM1⁺ ECs at E9.5 and E10.5 (Extended Data Fig. 7a–d’), demonstrating the utility of this line for labeling paraxial mesoderm-derived ECs. To assess the timing of LEC specification from the paraxial mesoderm, we administered a single dose of tamoxifen to pregnant *Pax3*^{CreERT2/+}; *Rosa26*^{tdTomato} animals at E7.0, E8.0, E9.0 or E10.0 to label cells from –E7.25, –E8.25, –E9.25 or –E10.25, respectively. Embryos were collected at E13.5 and analyzed by flow cytometry to compare BEC and LEC labeling with constitutive labeling in *Pax3*^{Cre/+}; *Rosa26*^{tdTomato} animals (Extended Data Fig. 7e). These analyses revealed that, although labeling of BECs is very limited after tamoxifen administration beyond E8, significant labeling of LECs persists after induction at E9, indicating that the fate of paraxial mesoderm-derived angioblasts becomes restricted to LECs as development progresses.

To investigate the spatial contribution of paraxial mesoderm derivatives to blood and lymphatic endothelium, we administered tamoxifen to *Pax3*^{CreERT2/+}; *Rosa26*^{tdTomato} animals at multiple stages of development and performed immunofluorescence imaging (Fig. 4a–j’ and Extended Data Fig. 7i–j’). Imaging of transverse vibratome sections at E10.5 showed that most PROX1⁺ ECs in the mesenchyme surrounding the CV are labeled after tamoxifen administration at E8.0 (Fig. 4a,a’,c) or E9.0 (Fig. 4b,c). In contrast, labeling of PROX1⁺ venous ECs was significantly reduced after tamoxifen administration at E9.0 (Fig. 4b,c), with a limited proportion of PROX1⁺ ECs on the dorsal aspect of the CV

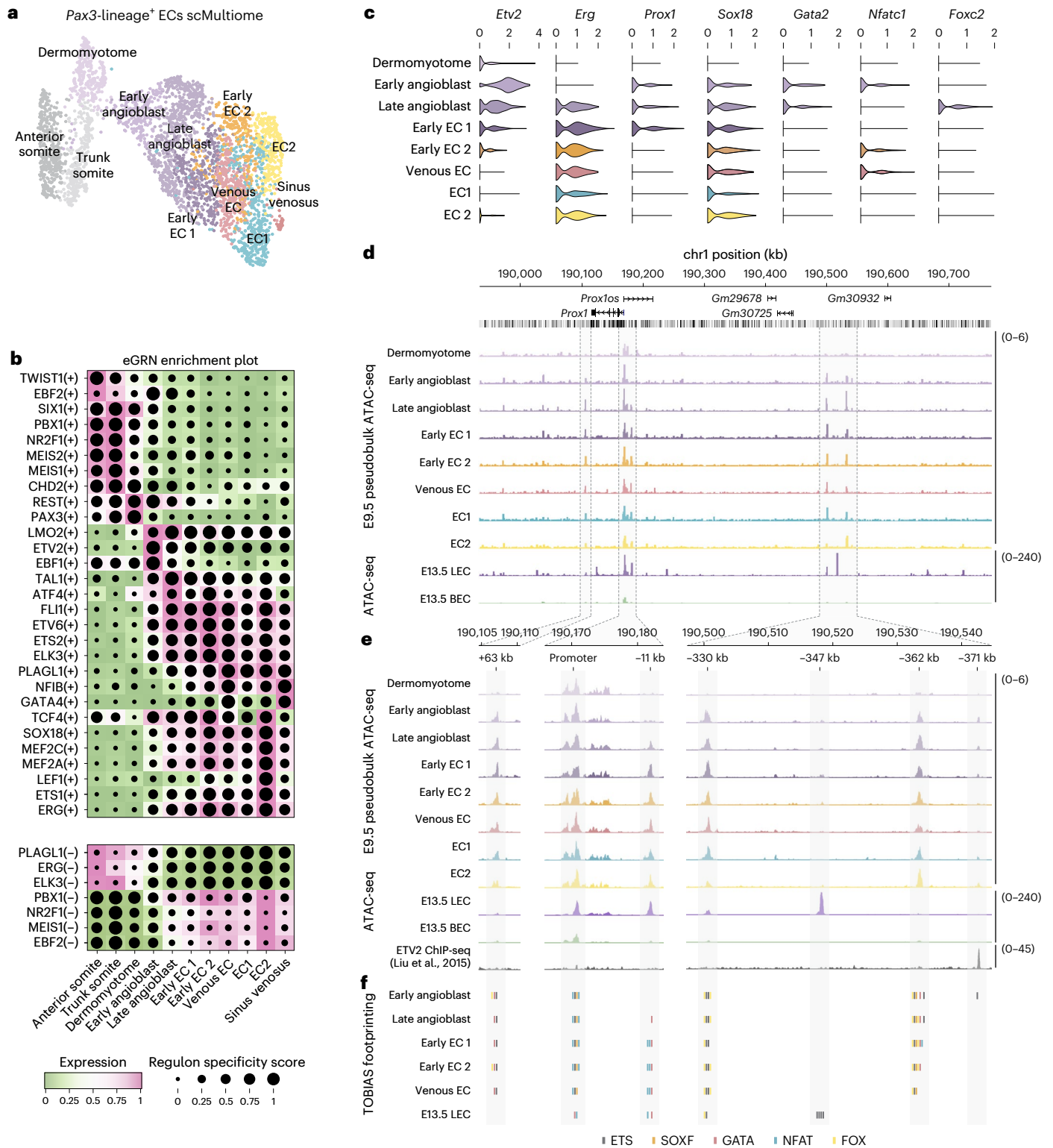


Fig. 3 | GRN analyses during LEC specification. **a**, Weighted nearest neighbor UMAP (wnnUMAP) embedding of single-cell multiome analyses of 3,606 tdTomato⁺ VEGFR2⁺ PECAM1⁺ cells FACS-sorted from *Pax3^{Cre/+}; Rosa26^{tdTomato}* embryos at E9.5. **b**, Combined heatmap and dotplot showing TF expression of individual eRegulons and cell type specificity (regulon specificity score), respectively. Activation or repression of gene expression is indicated by (+) or

(-) **c**, Violin plots showing expression of known transcriptional regulators of *Prox1* across each cellular state. **d**, Chromatin profiles at the *Prox1* genomic locus for E9.5 pseudobulk and E13.5 bulk ATAC-seq analyses. Pseudobulk profiles are derived from combining cells within each cellular state. **e**, Zoomed views of regions highlighted in **d**, including ETV2 ChIP-seq³⁷. **f**, Putative binding of ETS, SOXF, GATA, NFAT and FOX TFs identified with TOBIAS⁸¹.

labeled under these conditions (Fig. 4b,c). Of note, after tamoxifen administration at E9.0, we observed embryos in which a substantial proportion of PROX1⁺ ECs in the mesenchyme surrounding the CV were labeled in the absence of venous labeling, again suggesting that LEC specification can occur without transition through a venous intermediate. The observation that paraxial mesoderm-derived angioblasts continue to support venous expansion after E9.0 is in agreement with a recent study reporting labeled ECs in the CV after tamoxifen induction of *Etv2*^{CreERT2};*Rosa26*^{tdTomato} animals at E9.5 (ref. 40). Notably, using a Cre driver expressed from the *Myf5* locus (a myogenic transcription factor gene expressed from -E9.0), we again found a contribution of labeled cells to the dorsal aspect of the CV at E10.5 (Extended Data Fig. 7f-h).

In agreement with our flow cytometry analysis of BEC and LEC labeling at E13.5 (Extended Data Fig. 7e), imaging of transverse vibratome sections at E12.0 showed a substantial contribution of *Pax3*-lineage cells to BECs in the neural tube after tamoxifen at E8.0 (Fig. 4d,d'), which was lost after E9.0 tamoxifen (Fig. 4e,e'). In contrast, imaging of transverse sections at E12.0 showed that most PROX1⁺ LECs in the pTD and lymph sac are labeled after tamoxifen administration at E8.0 (Fig. 4f-f'') or E9.0 (Fig. 4g-g''), whereas labeling is largely absent after tamoxifen administration at E10.0 (Fig. 4h-h''). These analyses show that, between E8.0 and E9.0, the fate of paraxial mesoderm-derived angioblasts becomes restricted to lymphatic endothelium and indicate that tamoxifen-independent tracing of LECs does not occur in *Pax3*^{CreERT2/+};*Rosa26*^{tdTomato} animals. After tamoxifen administration to *Pax3*^{CreERT2/+};*Rosa26*^{tdTomato} animals at E9, almost complete labeling of PROX1⁺ LECs was observed in subcutaneous tissues at E13 (Fig. 4i,i') and in both lymphangiogenic sprouts (Fig. 4j,j') and isolated LEC clusters (Fig. 4j,j'') of the lumbar dermis at E16; lymphatic vessels that were previously proposed to arise from a non-venous source⁶. In contrast, labeling of LECs in the heart at E13.0 (Extended Data Fig. 7i-i'') and E16.0 (Extended Data Fig. 7j,j'') was more limited than that observed in *Pax3*^{Cre/+};*Rosa26*^{tdTomato} embryos (Extended Data Fig. 7k-l'). Collectively, these analyses provide evidence for a specialized paraxial mesoderm-derived angioblast that is a major source of lymphatic endothelium.

Reassessing the venous source of LECs

We next set out to re-evaluate the extent to which venous endothelium may serve as a source of mammalian LECs. Live imaging analyses in zebrafish embryos showed that LECs emerge from a subpopulation of venous ECs⁷, which undergo cell divisions that give rise to LECs and venous ECs⁴¹. The evidence for a predominantly venous source of LECs in mammals comes from lineage-tracing analyses of *Prox1*^{CreERT2} and *Tg(Tek-cre)* mice⁹. In contrast to our findings, which show concomitant initiation of PROX1 expression in the CV and surrounding mesenchyme at E9.5 (Fig. 1k-l''), tamoxifen administration to *Prox1*^{CreERT2};*Rosa26*^{LacZ} animals at E9.5 led to sparse labeling of the dorsal CV and no labeling of non-venous cells at E13.5 (ref. 9), suggesting that early labeling of *Prox1*-expressing cells using this approach is inefficient. Analysis of our scRNA-seq dataset showed that *Tek* is expressed not only in venous ECs but also in angioblasts during LEC specification

(Extended Data Fig. 8a). Furthermore, our previous analyses of dermal lymphatic vessel development in *Tg(Tek-cre)*^{12Flw/J};*Rosa26*^{mTmG} embryos identified double-reporter-positive LECs, indicative of recent recombination events induced by *Tg(Tek-cre)* in differentiated LECs⁶. Collectively, these observations bring into question the use of *Tg(Tek-cre)* mice as a tool for labeling only venous-derived LECs. Given these caveats, we used *Tg(Tek-cre)*^{S326Sato};*Rosa26*^{tdRFP} and *Tg(Tek-cre)*^{12Flw/J};*Rosa26*^{tdTomato} mice to revisit these analyses. Whole-mount immunofluorescence imaging of *Tg(Tek-cre)*^{S326Sato};*Rosa26*^{tdRFP} embryos revealed tdRFP⁺ PROX1⁺ ECs dorsal to the CV at E9.5 (Extended Data Fig. 8b). Unlabeled PROX1⁺ ECs were also observed in vibratome sections of *Tg(Tek-cre)*^{S326Sato};*Rosa26*^{tdRFP} embryos at E10.5 (Extended Data Fig. 8c,c') and E11.5 (Extended Data Fig. 8d) and in the pTD at E12.5 (Extended Data Fig. 8e,e'). Quantification of labeling at each stage revealed that approximately 30% of PROX1⁺ ECs outside of the venous endothelium are *Tg(Tek-cre)*^{S326Sato} lineage negative (Extended Data Fig. 8f), a figure that likely underestimates the true non-venous contribution given that *Tek* is expressed in angioblasts (Extended Data Fig. 8a). Similar observations were made in *Tg(Tek-cre)*^{12Flw/J};*Rosa26*^{tdTomato} embryos at E10.5 (Extended Data Fig. 8g,g'). Notably, the formation of lymph sac-like structures comprising LYVE1⁺PROX1⁺ LECs in *Tg(Tek-cre)*^{12Flw/J};*Prox1*^{fl/fl} embryos provides further evidence for a non-venous source of LECs (Extended Data Fig. 8h-h'). Collectively, these analyses provide evidence for a major non-venous contribution to developing lymphatics and raise questions about the use of *Tg(Tek-cre)* mice as a tool to specifically label vein-derived LECs.

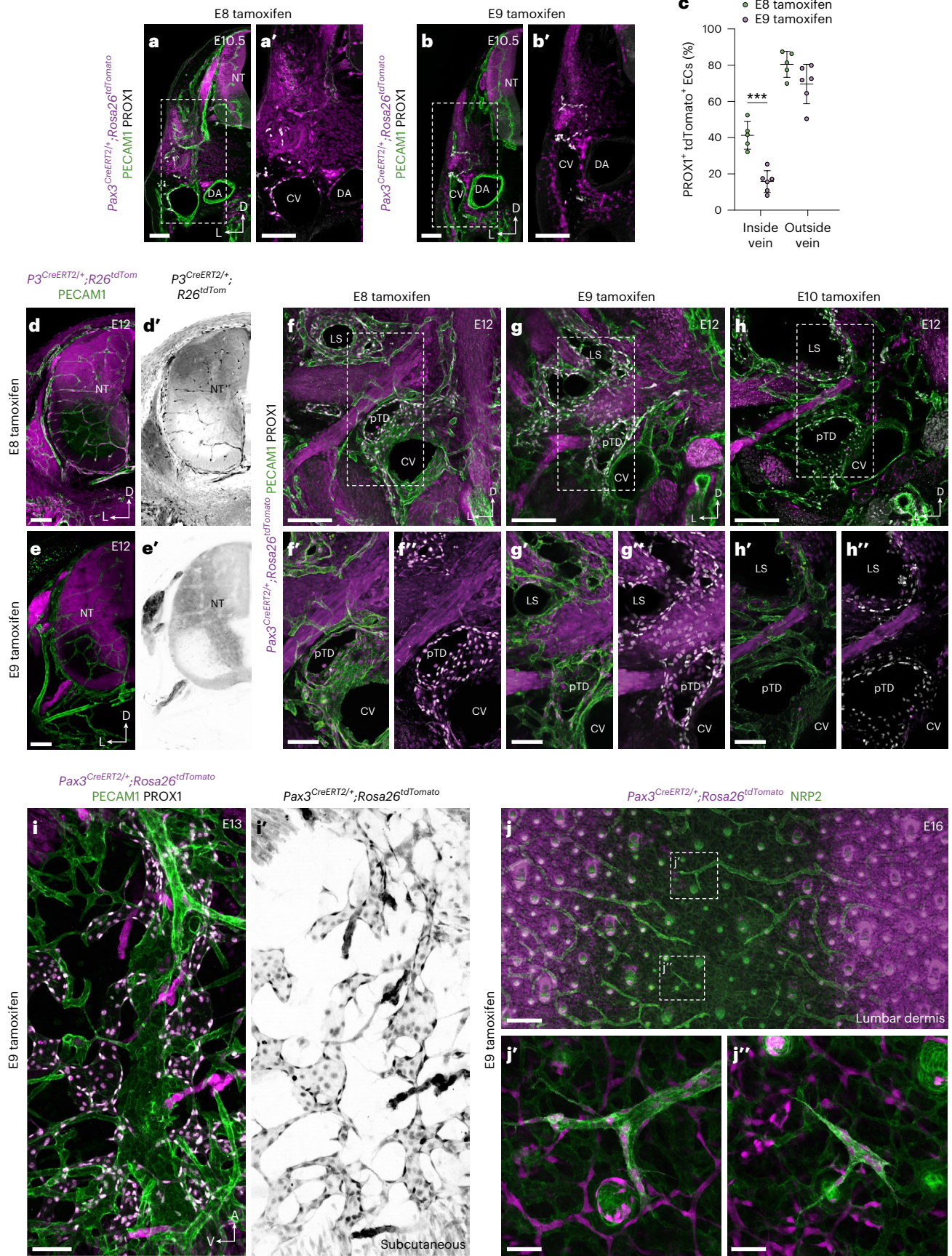
To further evaluate the developmental source of LECs, we assessed cell proliferation, reasoning that if migrating venous ECs are the major source of LECs, a substantial increase in EC numbers would be required to maintain venous integrity during the rapid expansion of lymphatics that occurs from E10. Characterization of cell cycle phase (Extended Data Fig. 9a,b) and gene expression (Extended Data Fig. 9c) in our scRNA-seq dataset revealed that most cells in the venous cluster are in G1 (Extended Data Fig. 9a,b) and express low levels of genes associated with cell cycle progression (Extended Data Fig. 9c). In contrast, most cells in the angioblast and LEC1 clusters are in S/G2/M (Extended Data Fig. 9a-c). These analyses suggest that angioblast-derived LECs and their ancestors, and not venous ECs, are endowed with the proliferative capacity to support rapid growth of LECs. To assess the expansion of venous ECs, and PROX1⁺ ECs inside and outside of the veins, we performed quantitative whole-mount light sheet imaging of embryos stained for ERG and PROX1 between E9.5 and E11.0 (Fig. 5a-e). These analyses revealed a rapid expansion of PROX1⁺ ECs outside the vein, increasing approximately 32-fold between E9.5 and E11, from 351 ± 85 to 11,510 ± 2,832 cells (±s.d.) (Fig. 5a-e). In contrast, more modest increases were observed in total venous ECs and PROX1⁺ venous ECs, increasing approximately 6.5-fold (1,250 ± 301 to 8,025 ± 557 cells (±s.d.)) and approximately nine-fold (140 ± 32 to 1,242 ± 328 cells (±s.d.)), respectively (Fig. 5a-e).

To empirically assess the proliferation of PROX1⁺ ECs inside and outside of the venous endothelium, we used a dual-pulse labeling strategy (Fig. 5f-i and Extended Data Fig. 9d). To maintain

Fig. 4 | Temporal analysis of LEC specification from paraxial mesoderm.

Representative immunofluorescence images of tdTomato, PECAMI and PROX1 on transverse vibratome sections from *Pax3*^{CreERT2/+};*Rosa26*^{tdTomato} embryos at E10.5 after tamoxifen administration at E8 (a,a') or E9 (b,b'). c, Quantification of percentage tdTomato labeling of PROX1⁺ ECs present inside or outside of the venous endothelium of *Pax3*^{CreERT2/+};*Rosa26*^{tdTomato} embryos at E10.5, after tamoxifen administration at E8.0 (n = 5, three pregnant dams) or E9.0 (n = 6, three pregnant dams). Each symbol represents data from an individual embryo (*** P < 0.001). Representative immunofluorescence images of tdTomato and PECAMI on transverse vibratome sections from *Pax3*^{CreERT2/+};*Rosa26*^{tdTomato} embryos at E12.0 after tamoxifen administration at E8.0 (d,d'; n = 6) or E9.0 (e,e'; n = 6). Representative immunofluorescence images of tdTomato, PECAMI

and PROX1 on transverse vibratome sections from *Pax3*^{CreERT2/+};*Rosa26*^{tdTomato} embryos at E12.0 after tamoxifen administration at E8.0 (f-f'; n = 6), E9.0 (g-g'; n = 6) or E10.0 (h-h'; n = 6). i,i', Representative immunofluorescence image of tdTomato, PECAMI and PROX1 on a sagittal vibratome section from a *Pax3*^{CreERT2/+};*Rosa26*^{tdTomato} embryo at E13 after tamoxifen administration at E9 (n = 4). j-j', Representative immunofluorescence images of tdTomato, NRP2 and PROX1 on whole-mount skin from *Pax3*^{CreERT2/+};*Rosa26*^{tdTomato} embryos after tamoxifen administration at E9 (n = 5). CV, cardinal vein; DA, dorsal aorta; LS, lymph sac; NT, neural tube; pTD, primordial thoracic duct. Scale bars, 100 μm (a,a', b,b', d,d', e,e', f,f', g,g', h,h' and i,i'), 200 μm (f, g and h), 250 μm (j), 50 μm (j', j''). Statistical analyses were performed using unpaired Student's *t*-test. Data are presented as mean ± s.d.



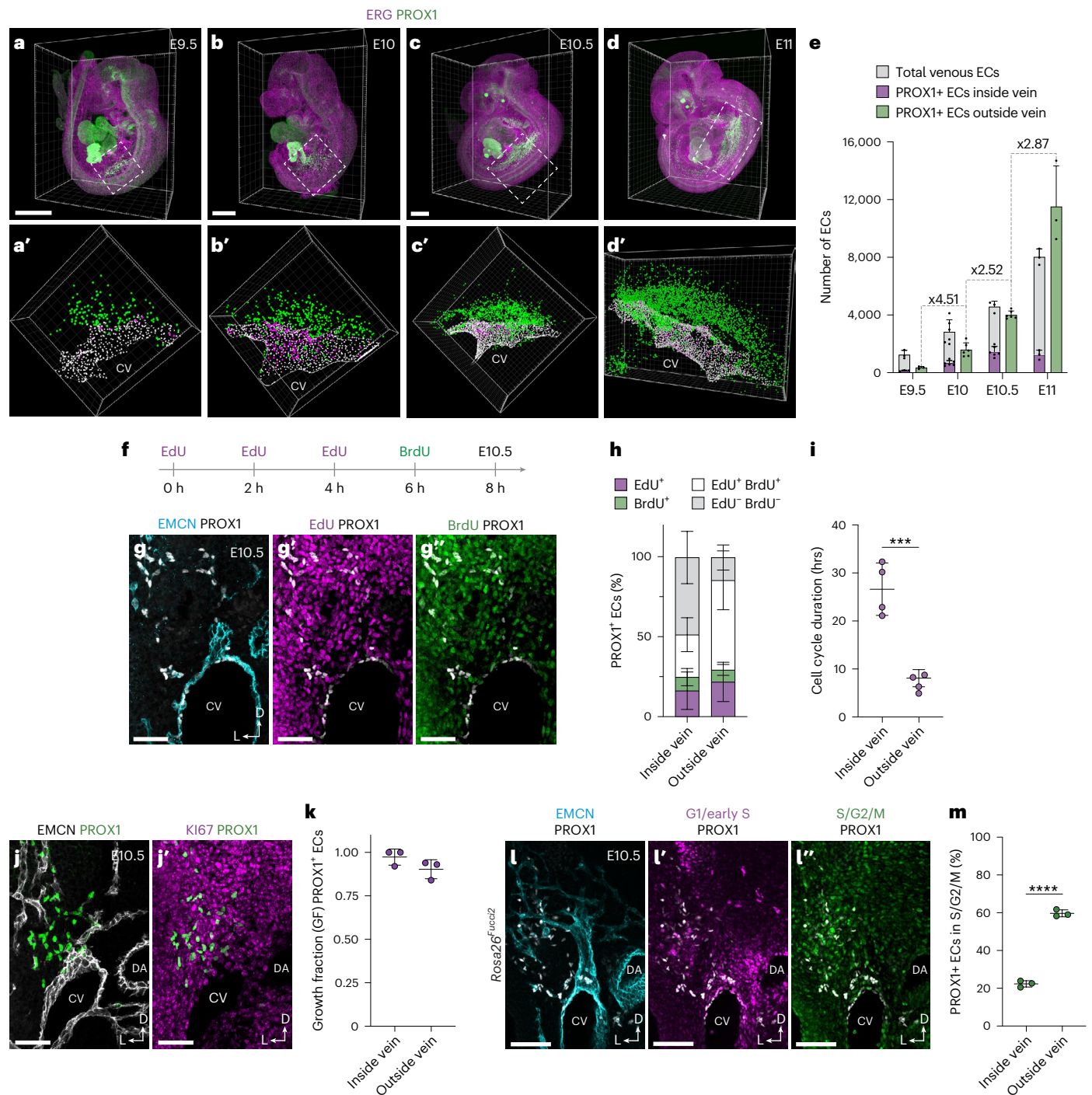


Fig. 5 | Quantitative analysis of LEC expansion. Representative whole-mount immunofluorescence for ERG and PROX1 in E9.5 (**a,a'**), E10 (**b,b'**), E10.5 (**c,c'**) and E11.0 (**d,d'**) embryos. 3D projections of regions of interest were segmented (**a'–d'**) for temporal quantification of ERG⁺PROX1⁺ ECs within the venous endothelium, ERG⁺PROX1⁺ ECs within the venous endothelium and ERG⁺PROX1⁺ ECs outside of the venous endothelium. **e**, Quantification of total ECs and PROX1⁺ ECs inside and outside of the venous endothelium between E9.5 and E11.0 (E9.5, *n* = 3; E10, *n* = 6; E10.5, *n* = 6; E11, *n* = 3). **f**, Schematic representation of the dual-pulse labeling strategy for analysis of cell cycle dynamics. **g–g''**, Representative immunofluorescence for EMCN, PROX1, EdU and BrdU on a transverse vibratome section from an E10.5 embryo. **h**, Quantification of labeling of PROX1⁺ ECs inside and outside of the venous endothelium with EdU

and/or BrdU (*n* = 4). **i**, Quantification of cell cycle duration inside and outside of the venous endothelium at E10.5 (*n* = 4, ****P* < 0.001). **j,j'**, Representative immunofluorescence for EMCN, PROX1 and Ki67 on a transverse vibratome section from an E10.5 embryo. **k**, Quantification of growth fraction for PROX1⁺ ECs present inside or outside of the venous endothelium (*n* = 3). **l–l''**, Representative immunofluorescence for EMCN, PROX1, mCherry (G1/early S) and mVenus (S/G2/M) on a transverse vibratome section from an E10.5 *Rosa26^{fucd2}* embryo. **m**, Quantification of PROX1⁺ ECs in S/G2/M phases of the cell cycle inside or outside of the venous endothelium (*n* = 3, *****P* < 0.001). DA, dorsal aorta. Scale bars, 500 μm (**a–d**), 50 μm (**g–g''**), 100 μm (**j,j'** and **l–l''**). Statistical analyses were performed using unpaired Student's *t*-test. Data are presented as mean ± s.d.

ethynyl-2'-deoxyuridine (EdU) bioavailability over the course of the experiment, we administered EdU three times at 2-h intervals, followed by a 2-h 5-bromo-2'-deoxyuridine (BrdU) pulse (Fig. 5f). The growth fraction of PROX1⁺ ECs was calculated using KI67 immunofluorescence to account for potential differences in growth state between cells inside and outside the vein (Fig. 5j,k and Extended Data Fig. 9d) and contrasted with growth fraction in the entire CV or dorsal aorta (Extended Data Fig. 9e,f). Analysis of EdU and BrdU labeling revealed higher incorporation in PROX1⁺ ECs outside of the vein at E10.5 (Fig. 5g,h), which translated to significantly shorter cell cycle duration in these cells (26.6 ± 5.5 h inside vein and 8.1 ± 1.8 h outside vein (±s.d.); Fig. 5i). In agreement with these findings, analysis of *Rosa26^{luciferase}* embryos⁴², where cells in G1/early S-phase are labeled with mCherry and cells in S/G2/M are labeled with mVenus, revealed that a higher proportion of PROX1⁺ ECs outside of the vein are in S/G2/M phase at E10.5 (Fig. 5l,m). Collectively, these findings show that rapid expansion of PROX1⁺ ECs outside of the vein (Fig. 5a–e) occurs despite only modest venous proliferation. Given the limited proliferation of venous endothelium at these stages, CV expansion may be supported by continued addition of cells from the paraxial mesoderm, as indicated by our lineage-tracing analyses (Fig. 4b,c). This model further challenges the prevailing view that, upon induction of PROX1 expression in venous ECs, LEC progenitors migrate dorsally from venous endothelium to form the first lymphatic structures.

The migration of LECs is highly dependent on vascular endothelial growth factor-C (VEGFC)^{19,43}, however, the role of VEGF signaling during LEC specification from paraxial mesoderm is unknown. Our scMultiome analyses indicate that *Etv2*⁺ angioblasts are initially specified from *Kdr*⁺ (*Vegfr2*) mesoderm and then sequentially upregulate *Flt4* (*Vegfr3*) and *Prox1* (Fig. 2a,b and Extended Data Fig. 10b). Interestingly, although we did not observe substantial expression of *Vegfa* (Extended Data Fig. 10b), *Vegfc* expression was observed in dermomyotome and angioblasts as well as in sinus venosus and neural tube ECs (Extended Data Fig. 10b). To validate these observations, we performed HCR analyses in whole-mount embryos at E9.5 (Fig. 6a–d^{''}) and Extended Data Fig. 10d–g^{''}). Combined analysis of *Kdr*, *Flt4* and *Etv2* expression revealed that all *Etv2*⁺ cells were *Kdr*⁺, and some *Etv2*⁺*Kdr*⁺ cells also expressed *Flt4* (Fig. 6a–b^{''}), indicating that *Etv2* and *Kdr* co-expression precedes *Flt4* expression. *Etv2* was shown to directly regulate *Flt4* expression in zebrafish embryos⁴⁴. To investigate regulation of *Flt4* by ETV2 in mice, we compared our pseudobulk and bulk ATAC profiles with published ETV2 ChIP-seq³⁷ and identified multiple overlapping peaks in the *Flt4* locus (Extended Data Fig. 10c), suggesting that ETV2 may bind these putative regulatory elements in angioblasts to drive chromatin remodeling or induce *Flt4* expression. FLT4 is required to maintain PROX1 expression in early LECs⁴⁵, and our analysis of *Flt4*, *Prox1* and *Etv2* expression revealed that *Etv2*⁺*Prox1*⁺ cells always expressed *Flt4* (Fig. 6c,d^{''}), suggesting that FLT4 may act upstream of PROX1 during angioblast differentiation. HCR analysis of *Vegf* ligands demonstrated broad, largely unrestricted *Vegfa* expression (Extended Data Fig. 10d–e^{''}). In contrast, *Vegfc* expression was enriched in the dermomyotome and dorsal aorta and in *Etv2*⁺ angioblasts surrounding the CV (Extended Data Fig. 10f–g^{''}), suggesting that it may act through both paracrine and autocrine signaling during LEC specification. Collectively, these findings suggest that VEGFC/VEGFR3 signaling acts upstream of PROX1 during LEC specification from paraxial mesoderm-derived angioblasts.

To address the functional importance of VEGFC signaling, we performed whole-mount immunofluorescence for PROX1 and ETV2 in *Vegfc^{+/+}* and *Vegfc^{LacZ/LacZ}* embryos (Fig. 6e–f^{''}). As previously reported⁴³, these analyses revealed a reduction in PROX1-expressing cells surrounding the CV (Fig. 6e–f^{''}). Notably, this reduction was coupled to an increase in ETV2-expressing cells (Fig. 6e–f^{''}), suggesting that, in the absence of VEGFC/VEGFR3 signaling, specification of LECs from paraxial mesoderm stalls at the level of the angioblast. These findings

indicate that, rather than stimulating migration of ECs from venous endothelium into the surrounding mesenchyme, the primary function of VEGFC may be to drive in situ differentiation of angioblasts into PROX1⁺ ECs. To further assess the plausibility of PROX1⁺ EC migration from the CV, we combined in situ hybridization and immunofluorescence for *Ccbe1*, *Vegfc*, PECAM1 and PROX1 at E10.5 (Fig. 6g,h). These analyses revealed that *Vegfc* is expressed along the lateral body wall and that co-expression with *Ccbe1*, which is essential for processing of VEGFC into its lymphangiogenic form⁴⁶, occurs in two domains (Fig. 6g,h). One domain sits dorsal to the CV, in the region of the hypaxial dermomyotome, and a second domain immediately flanks the ventrolateral CV. The absence of *Ccbe1* from the mesenchyme immediately dorsal to the CV, as well as the failure of venous PROX1⁺ ECs to enter the ventrolateral VEGFC/CCBE1 expression domain, suggests that paraxial mesoderm-derived angioblasts are more likely to be responsive to VEGFC/CCBE1 activity than venous PROX1⁺ ECs. To assess the direction of PROX1⁺ EC migration, we analyzed vibratome sections from E10.5 embryos stained for PROX1 and the Golgi marker GM130 (Fig. 6i–i^{''}). During cell migration, the Golgi apparatus is positioned ahead of the nucleus in the direction of movement and can, thus, be used to infer directionality of migration⁴⁷. Segmentation and vectorization of PROX1⁺ EC nuclei and their Golgi allowed quantification of Golgi orientation in the dorsal-ventral and lateral-medial axes, revealing that migration of PROX1⁺ ECs inside and outside of the vein is randomized (Fig. 6j). Collectively, our analyses challenge the view that venous endothelium is the primary source of LECs. Instead, we show that VEGFC induces in situ differentiation of LECs from a population of specialized angioblasts, which, in turn, give rise to a majority lymphatic endothelium.

Discussion

At the beginning of the 20th century, Florence Sabin's anatomical studies of pig⁸ and human⁴⁸ embryos suggested that lymphatic vessels form through budding from large veins. An alternative model of LEC specification, directly from mesenchymal precursors, was later proposed by Huntington and McClure⁵. Debate over these two opposing models of LEC specification has continued for over a century, with evidence from a range of model systems suggesting the existence of venous, non-venous and dual sources^{6,10,11,16,17,28,49}. In mammalian embryos, the consensus was that LECs arise predominantly through centrifugal sprouting from venous endothelium, with non-venous sources making limited contributions to organ-specific lymphatic networks^{1,6,17}. Here, using a range of single-cell genomics, lineage tracing and high-resolution imaging, we show that LECs are specified in situ from specialized mesenchymal progenitors (Fig. 6k,l). Collectively, these analyses provide evidence for a specialized paraxial mesoderm-derived angioblast that is a major source of lymphatic endothelium at E13.5.

The term 'specialized angioblast' was previously used to describe a population of ECs in the zebrafish CV that give rise to arterial, venous and lymphatic endothelium⁷. Here, we identify bona fide ETV2⁺ mesenchymal angioblasts that do not express markers of mature endothelium, are not exposed to blood flow and directly give rise to PROX1⁺ ECs. Due to the transient nature of ETV2 and PROX1 co-expression, we were unable to determine if all angioblast-derived LECs co-express these two genes during their differentiation. However, our analyses show that transition from ETV2⁺ angioblast to PROX1⁺ EC is rapid. Notably, our single-cell analyses revealed not only that specification of LECs from *Prox1*⁺ venous endothelium is limited but also that paraxial mesoderm-derived angioblasts continue to contribute to venous expansion, observations supported by lineage-tracing and morphometric analyses. Indeed, given the limited proliferative capacity of venous endothelium, if all PROX1⁺ ECs were to migrate from the CV and sinus venosus, these vessels would cease to exist. Instead, our analyses show that there is a steady increase in venous EC numbers at this stage, collectively supporting a model of limited PROX1⁺ EC migration from

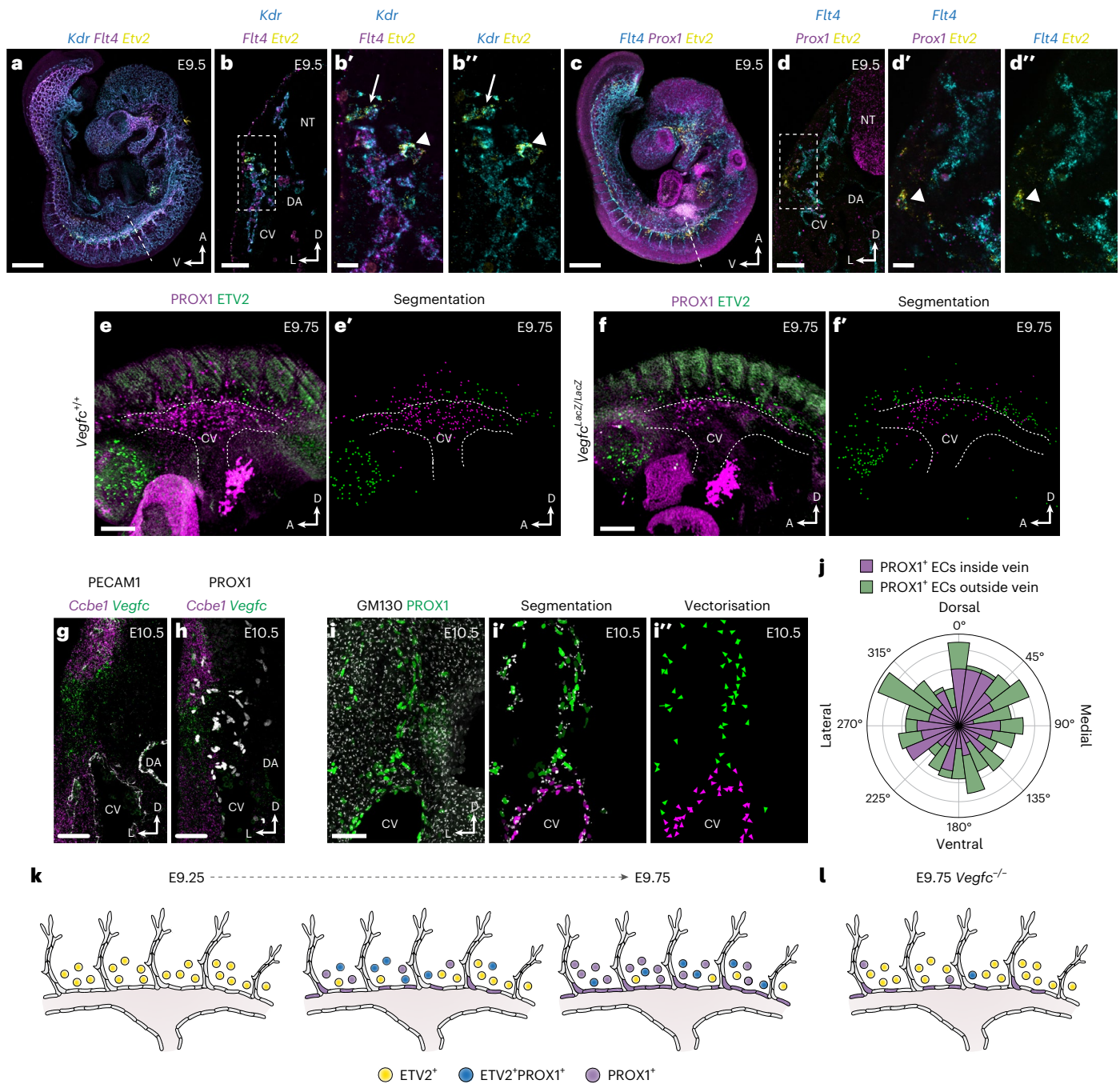


Fig. 6 | Temporal analyses of LEC morphogenesis. **a**, Representative confocal image of *Kdr*, *Flt4* and *Etv2* expression in a whole-mount E9.5 embryo using HCR ($n = 6$). **b**, Transverse vibratome section showing *Kdr*, *Flt4* and *Etv2* expression at the level of dashed line in **a** ($n = 4$). **b'**, **b''**, High-magnification images of boxed area in **b** showing *Kdr*, *Flt4* and *Etv2* expression. Arrow indicates a *Kdr*⁺ cell, arrowhead indicates a *Kdr*⁺*Etv2*⁺ cell. **c**, Representative confocal image of *Flt4*, *Prox1* and *Etv2* expression in a whole-mount E9.5 embryo using HCR ($n = 6$). **d**, Transverse vibratome section showing *Flt4*, *Prox1* and *Etv2* expression at the level of dashed line in **c** ($n = 4$). **d'**, **d''**, High-magnification images of boxed area in **d** showing *Flt4*, *Prox1* and *Etv2* expression. Arrowhead indicates a *Flt4*⁺*Prox1*⁺*Etv2*⁺ cell. Representative whole-mount immunofluorescence for PROX1 and ETV2 and image segmentation in *Vegfc*^{+/+} (**e, e'**) and *Vegfc*^{LacZ/LacZ} (**f, f'**)

embryos at E9.75 ($n = 3$). Combined in situ hybridization for *Ccbe1* and *Vegfc* and immunofluorescence for PECAM1 (**g**) or PROX1 (**h**). **i**, Immunofluorescence for GM130 and PROX1 on a transverse vibratome section from an E10.5 embryo. Segmentation of GM130⁺ Golgi and PROX1⁺ EC nuclei (**i'**) and vectorization of Golgi orientation for quantification ($n = 8$) (**i''**). **j**, Rose diagram illustrating Golgi orientation in PROX1⁺ ECs inside and outside of the venous endothelium. Proposed model of LEC differentiation through in situ specification from ETV2⁺PROX1⁺ paraxial mesoderm-derived lymphangioblasts in WT (**k**) and ETV2⁺PROX1⁺ cells represent individual cells expressing the indicated genes. DA, dorsal aorta; NT, neural tube. Scale bars, 500 μ m (**a** and **c**), 100 μ m (**b**, **d**, **g** and **h**), 25 μ m (**b'** and **d'**), 200 μ m (**e** and **f**), 50 μ m (**i**).

venous endothelium. Ultimately, technological advances that permit the live imaging of mammalian embryos, at currently infeasible stages of development, may provide definitive evidence for venous or non-venous sources of LECs.

Mechanistically, we provide insight into the acquisition of LEC identity through global GRN analyses and examination of *cis*-regulatory elements in the *Prox1* locus. In agreement with the well-described roles of ETS factors in directing^{26,50} and maintaining⁵¹

EC fate, our analyses identified multiple eGRNs regulated by ETS TFs, including ETV2 and ERG. Interestingly, we also identified a number of putative regulators of LEC specification, including EBF1 and EBF2. EBF1 is a pioneer factor essential for the acquisition of B cell identity⁵² and heart development⁵³, which can bind and remodel naive chromatin⁵². Although no obvious vascular defects have been described after loss of either *Ebf1* or *Ebf2*, EBF1 has been shown to functionally compensate for the loss of *Ebf2* in mouse adipocytes⁵⁴. Thus, it is tempting to speculate that these factors may also contribute to establishment of a chromatin landscape permissive to LEC specification. Although the importance of PROX1 activity for acquisition and maintenance of LEC fate is well established, knowledge of the factors controlling its transcription is relatively limited. Detailed characterization of the *Prox1* –11-kb enhancer, which is bound by GATA2, FOXC2, NFATC1 and PROX1, recently revealed a role for this element in the repression of hemogenic fate in LECs³⁶. Furthermore, elegant work in zebrafish has shown that multiple *cis*-regulatory elements are required to drive anatomically distinct patterns of *prox1a* expression during embryonic and larval development⁵⁵. Here, our scMultiome analyses uncovered multiple putative enhancers of *Prox1* open in angioblasts and predicted to be bound by known transcriptional regulators. In future work, it will be of interest to investigate the spatiotemporal control of LEC specification and *Prox1* expression by the TFs and *cis*-regulatory elements identified here.

Our analyses suggest that, in addition to important roles in control of LEC migration^{43,56} and maintenance of LEC identity⁵⁷, VEGFC/VEGFR3 is critical for the transition of differentiating angioblasts into PROX1⁺ ECs. During the process of LEC specification from paraxial mesoderm, VEGFR2 and ETV2 are expressed before VEGFR3. Given the key roles of VEGFR2 in angioblast specification at earlier stages of development, it is likely that initial specification of ETV2⁺ angioblasts in the hypaxial dermomyotome is VEGFR2 dependent, with subsequent differentiation of VEGFR2⁺ETV2⁺ angioblasts into PROX1⁺ ECs requiring VEGFC/VEGFR3. Our revised model of mammalian lymphatic vessel development suggests that, rather than sprouting from a contiguous venous-derived structure, expansion of the initial lymphatic network occurs through coalescence of extravascular PROX1⁺ ECs. This process resembles formation of lymphatic vessels from LEC clusters, which has been observed in the mesentery¹⁷, dermis⁶, heart⁵⁸ and kidney⁵⁹, and raises the possibility that the coalescence of clustered LECs is a more general mechanism for expansion of lymphatic networks throughout development.

The identification of angioblasts that directly give rise to a subset of mammalian endothelium has broad implications for our understanding of vascular development. Subtypes of angioblasts that form arterial, venous and intestinal ECs have been described in zebrafish embryos^{7,12,40}. Furthermore, chimeric transplantation in avian embryos demonstrated the restricted potential of dermomyotome-derived cells, which were found to contribute to endothelium of the body wall and kidney, but did not integrate into endothelium of the visceral organs or the ventral wall of the dorsal aorta⁶⁰. Based on our findings, it is tempting to speculate that further aspects of EC fate and function are underpinned by differences established at the level of the angioblast source from which they arise and before they acquire endothelial identity. An improved understanding of the determinants of BEC and LEC fate and function will shed further light on these processes, enhancing understanding of congenital and acquired vascular diseases and aiding attempts to engineer bona fide organ- and system-specific ECs in vitro.

Methods

Animal experiments

All procedures were carried out in accordance with local legislation: University of Oxford Animal Welfare and Ethical Review Boards in accordance with Animals (Scientific Procedures) Act 1986 under Home Office project licences PPL PC013B246 or PP6588077;

German animal protection legislation (Tierschutzgesetz und Tierschutz-versuchstierverordnung); and Uppsala Animal Experiment Ethics Board (permit number 130/15).

Mouse strains, husbandry and embryo collection

The following mice lines were used in this study: *Pax3*^{Cre} (*Pax3*^{tm1(cre)loj})⁶¹; *Pax3*^{CreERT2} (*Pax3*^{tm1.1(cre/ERT2)Lepr})³⁹; *Lbx1*^{Cre} (*Lbx1*^{tm3.1(cre)Cbm})⁶²; *Myf5*^{Cre} (*Myf5*^{tm3(cre)Sor})⁶³; Tie2-Cre (Tg(Tek-cre)^{12Flv/J})⁶⁴; Tie2-Cre (Tg(Tek-cre)^{5326Sato})⁶⁵; *Prox1*^{fl} (*Prox1*^{tm1a(EUCOMM)Wtsi})⁶; *Rosa26*^{tdTomato} (*Gt(ROSA)26Sor*^{tm9(CAG-tdTomato)Hze})⁶⁶; *Rosa26*^{tdTomato} (*Gt(ROSA)26Sor*^{tm14(CAG-tdTomato)Hze})⁶⁶; *Rosa26*^{drFP} (*Gt(ROSA)26Sor*^{tm1Hlf})⁶⁷; *Rosa26*^{Fucci2} (Tg(Gt(ROSA)26Sor-Fucci2)^{Sia})⁴²; and *Vegfc*^{LacZ} (*Vegfc*^{tm1Ali})⁴³. With the exception of the *Vegfc*^{tm1Ali} strain, which was maintained on a CD1 background, all animals were maintained on a C57Bl/6 background. Mice were maintained in individually ventilated cages (IVCs) and ventilated racks at 22 °C and 55% humidity. For embryo collection, mice were paired overnight, and females were checked the next morning for the presence of a vaginal plug. For inducible Cre induction, pregnant females were gavaged at the specified timepoints with 80 mg kg⁻¹ tamoxifen (Sigma-Aldrich, T5648) dissolved in peanut oil with 10% ethanol at a final concentration of 10 mg ml⁻¹.

Immunofluorescence staining of embryonic sections, whole-mount embryos and tissues

The following antibodies were used for immunofluorescence staining of cryosections, vibratome sections and/or whole-mount tissues: ETV2 (Abcam, ab181847, 1:100), VEGFR2 (BD Pharmingen, 550549, 1:200), PECAM1 (R&D Systems, AF3628, 1:250), PECAM1 (D. Vestweber, clones 5D2.6 and 1G5.1, 15 µg ml⁻¹), PROX1 (R&D Systems, AF2727, 1:200), PROX1 (Proteintech, 11067-2-AP, 1:100), PROX1 (Reliatech, 102-PA32, 1:100), PROX1 (Abcam, ab225414, 1:100), EMCN (Santa Cruz Biotechnology, sc-65495, 1:50), EMCN (D. Vestweber, VE44, 1:100), EMCN (Santa Cruz Biotechnology, sc-53941, 1:50), VWF (Abcam, 11713, 1:100), ESRI (Abcam, ab16660, 1:100), NRP2 (R&D Systems, AF567, 1:250), ERG (Abcam, ab92513, 1:200), GM130 (BD Pharmingen, 610823, 1:100), RFP (Rockland, 600-401-379, 1:500), BrdU (Abcam, ab6326, 1:100), KI67 (Thermo Fisher Scientific, 14-5698-82, 1:100) and GFP (Thermo Fisher Scientific, A-21311, 1:100). The following Alexa Fluor-conjugated secondary antibodies (Thermo Fisher Scientific) were used at 1:500–1:1,000: donkey anti-rat IgG Alexa Fluor Plus 405 (A48268), donkey anti-rabbit IgG Alexa Fluor 488 (A21206), donkey anti-goat IgG Alexa Fluor Plus 488 (A32814), donkey anti-sheep IgG Alexa Fluor 488 (A11015), donkey anti-rabbit IgG Alexa Fluor 555 (A32814), donkey anti-rat IgG Alexa Fluor 555 (A48270), donkey anti-goat IgG Alexa Fluor 647 (A32849), donkey anti-rabbit IgG Alexa Fluor Plus 647 (A32795) and donkey anti-sheep IgG Alexa Fluor 647 (A21448).

For immunofluorescence staining of cryosections, samples were fixed in 4% paraformaldehyde (PFA) overnight at 4 °C. Samples were washed in 1× PBS and then cryoprotected in sucrose and mounted in Optimal Cutting Temperature (OCT) compound. Next, 10-µm cryosections were blocked (PBS containing 0.1% Triton X-100 (PBX), 1% BSA and 2% donkey serum) for 1 h at room temperature, and then primary antibodies diluted in blocking buffer were incubated overnight. After three 10-min washes in PBX, secondary antibodies diluted in blocking buffer were incubated for 1 h at room temperature. Slides were then washed three times for 10 min in PBX before mounting with VECTASHIELD Antifade Mounting Medium.

For immunofluorescence staining of vibratome sections, fixed embryos were mounted in 6% low-melting-temperature agarose. Then, 150–200-µm vibratome sections were cut using a Leica VT1000S or a Leica VT1200S. Tissue slices were permeabilized in 0.5% PBX and then incubated in either PermBlock solution (PBS containing 3% BSA and 0.1% Tween 20) or blocking buffer (PBS containing 0.5% PBX, 0.5% Tween 20, 1% BSA and 3% donkey serum) for 2 h at room temperature and primary antibodies overnight at 4 °C. Subsequently, tissues were

washed three times for 20 min in PBX and incubated with Alexa Fluor–conjugated secondary antibodies overnight at 4 °C. Tissues were then washed in PBX before mounting.

For whole-mount staining of fixed embryos, samples were permeabilized in 0.5% PBX for 12 h at room temperature, blocked in PermBlock solution for 1–2 d at room temperature and stained with primary antibodies at 37 °C for at least 2 d. After three washing steps with PBST (1× PBS containing 0.1% Tween 20), tissues were incubated with Alexa Fluor–conjugated secondary antibodies for at least 1 d at 37 °C and then washed three times with PBST. Stained embryos were stored in PBST at 4 °C until further processing for optical clearing.

For whole-mount staining of the embryonic skin, tissues were fixed in 4% PFA for 2 h at room temperature, washed in 1× PBS and then incubated in blocking buffer (1× PBS containing 0.3% PBX, 1% BSA and 3% donkey serum) for 2 h at room temperature. Tissues were then incubated with primary antibodies diluted in blocking solution overnight at 4 °C. After primary antibody incubation, tissues were washed five times for 10 min in PBX and then incubated with Alexa Fluor–conjugated secondary antibodies for 3 h at room temperature. Tissues were then washed five times for 10 min in PBX and mounted in VECTASHIELD.

For whole-mount imaging of embryonic hearts, dissected hearts were fixed in 4% PFA overnight at 4 °C, washed in 1× PBS and then blocked (1× PBS containing 0.5% PBX, 0.5% PBST, 1% BSA and 3% donkey serum) overnight at 4 °C. Samples were incubated overnight at 4 °C with primary antibodies diluted in incubation buffer (1× PBS containing 0.25% PBX, 0.25% PBST, 0.5% BSA and 1.5% donkey serum). After primary antibody incubation, tissues were washed five times for 30 min in PBX and then incubated with Alexa Fluor–conjugated secondary antibodies overnight at 4 °C. Tissues were then washed five times for 30 min in PBX and mounted in 0.5% low-melting-temperature agarose for imaging.

Confocal laser scanning microscopy

Imaging of immunostained tissues was performed using Zeiss LSM780, LSM880 or LSM980 confocal microscopes, equipped with the following objectives: ×10 Plan-Apo, numerical aperture (NA) = 0.45; ×20 Plan-Apo, NA = 0.8; ×40 C-Apo water, NA = 1.2; ×63 Plan-Apo oil, NA = 1.4. Datasets were recorded and processed with ZEN Pro (Zeiss). All confocal images represent maximum intensity projections of z-stacks of either single tile or multiple tile scans. Mosaic tile scans with 10% overlap between neighboring z-stacks were stitched in ZEN software. Confocal single and multiple tile scans were processed in Fiji. If necessary, adjustments to brightness, contrast and intensity were made uniformly across individual channels and datasets.

Embryo dissociation for flow cytometry and FACS

All dissections were performed in PBS with 2% heat-inactivated FBS. For scRNA-seq experiments, embryos were dissected at the level of the otic vesicle and first pharyngeal arch to remove cranial tissues, with embryos from the same stage pooled as follows: E9.5: 18 embryos from four litters; E10.5: 10 embryos from three litters; E11.5: nine embryos from two litters. For scMultiome-seq experiments, 25 E9.5 *Pax3^{Cre/+};Rosa26^{tdTomato}* embryos were dissected at the level of the otic vesicle and first pharyngeal arch to remove cranial tissues. Single-cell suspensions were obtained and blocked as described above. For each stage, embryos were incubated in PBS with 10% FBS, 2 mg ml⁻¹ Collagenase IV (Gibco, 9001-12-1) and 0.2 mg ml⁻¹ DNase I (Roche, 10104159001) for 20–45 min at 37 °C until fully dissociated. The cell suspension was resuspended every 5 min. After digestion, PBS containing 0.5% FBS and 2 mM EDTA was added in a 1:1 ratio, with the resulting suspension passed through a 40-µm filter and centrifuged at 500g for 5 min at 4 °C. Cell pellets were resuspended in Cell Staining Buffer (BioLegend, 420201). Cell counting was performed using a Countess 3 Automated Cell Counter (Thermo Fisher Scientific).

Flow cytometry

Single-cell suspensions, generated as above, were incubated with Zombie Aqua Fixable Viability Kit (BioLegend, 423101, 1:1,000) for 15 min at room temperature. Cells were washed with Cell Staining Buffer (BioLegend, 420201) and then washed, centrifuged at 500g for 5 min and blocked with Fc block CD16/32 (BioLegend, 101302, 1:100) for 5 min on ice and stained with PECAMI-BV605 (BioLegend, 102427, 1:1,000), CD45-FITC (BioLegend, 157607, 1:200), CD41-BV421 (BioLegend, 133911, 1:200), PDPN-eF660 (eBioscience, 50-5381-82, 1:100) and LYVE1-PECy7 (eBioscience, 25-0443-82, 1:400) for 30 min on ice and then washed and resuspended in Cell Staining Buffer. The samples were either analyzed immediately on a BD LSRFortessa X20 cytometer or stored in IC Fixation Buffer (eBioscience, 00-8222-49), washed and analyzed the next day.

FACS

Single-cell suspensions were obtained as described above and subsequently blocked for 5 min on ice with Fc block CD16/32 (BioLegend, 101302, 1:100), followed by the addition of antibodies for 30 min on ice: VEGFR2-PECy7 (BioLegend, 136414, 1:100), CD45-APC (BioLegend, 103111, 1:200) and CD41-Alexa 647 (BioLegend, 133933, 1:200). For E10.5 and E11.5 suspensions, PECAMI-PECy7 (BioLegend, 102418, 1:100) was also included to capture as broad a pool of phenotypically diverse ECs as possible. The cell suspension was washed using Cell Staining Buffer (BioLegend, 420201) and then incubated with SYTOX Blue Dead Cell Stain (Invitrogen, S34857, 1:10,000) 10 min on ice and then washed. Single cells were sorted using a BD FACSAria III and collected into PBS containing 0.5% BSA for scRNA-seq experiments and PBS containing 0.04% BSA for scMultiome-seq experiments (Miltenyi Biotec, 130-091-376).

scRNA-seq

Library generation and pre-processing. scRNA-seq was performed using the 10x Genomics Chromium platform, and libraries were generated using the Next GEM Single Cell 3' GEM, Library & Gel Bead Kit version 3.1 (10x Genomics, PN-1000128). Libraries were sequenced with the standalone mode set to the manufacturer's protocol on the Illumina NextSeq 500 platform using the NextSeq 500/550 High Output Kit version 2.5, 150 cycles (Illumina, 20024907), to a depth of approximately 50,000 reads per cell. Raw base call files were demultiplexed using bcl2fastq version 2.20 software (Illumina) according to 10x Genomics instructions. Reads were aligned to the mm10 genome with the tdTomato-WPRE sequence added, and cells were called using Cell Ranger 5.0 (10x Genomics).

Quality control and normalization. The filtered barcode matrices were loaded into RStudio version 1.4 and further analyzed using Seurat 4.0 (refs. 68,69). Cells with fewer than 2,500 detected genes and more than 100,000 unique molecular identifiers (UMIs) and 7% of mitochondrial reads were removed. The data were normalized using the NormalizeData function, and the highly variable features were calculated using FindVariableFeatures. The data were further scaled using ScaleData, and principal component analysis (PCA) was performed using the variable features previously calculated. Cell cycle stage was predicted using G2M and S-phase genes⁷⁰.

Batch correction. To remove technical batch effects due to the different collection and processing times between the samples, the Seurat wrapper FastMNN, a faster version of MNN⁷¹, was used. The RunFastMNN function was used with auto.merge = T, and *k* was decreased to 10 to avoid overcorrection.

Clustering. A nearest neighbor graph was calculated using FindNeighbors using 10 corrected embeddings from FastMNN and 10 neighbors (*k.param* = 10). The clusters were found using FindClusters

with resolution = 2. The clusters were manually annotated according to marker genes determined with FindMarkers. The clusters clearly separated based on cell cycle stage alone were manually merged back together.

Visualization and trajectory analyses. The integrated Seurat object was converted to H5AD format for import into Python (Jupyter Notebook interface⁷²) using the packages SeuratDisk and SeuratData. The Scanpy 1.8 package⁷³ was used to run the PAGA⁷⁴ algorithm (scanpy.tl.paga) to produce an embedding that captures the connectivity of the data accurately using the corrected principal components (PCs) from FastMNN and the clusters previously calculated. To visualize the data, a force-directed graph⁷⁵ was calculated using scanpy.tl.draw_graph and pre-computed coordinates from PAGA.

Waddington-OT analyses. Trajectories from E9.5 to E11.5 were also assessed using Waddington-OT⁷⁶, a pipeline based on optimal transport. The results corroborate independent analyses with PAGA (Extended Data Fig. 5e). We used Waddington-OT with default parameters (entropic regularization $\epsilon = 0.05$, early timepoint balance regularization $\lambda_1 = 1$ and late timepoint balance regularization $\lambda_2 = 50$). Given a series of population snapshots at times t_1, \dots, t_T , Waddington-OT fits a coupling between consecutive pairs of populations. Each coupling, mathematically a joint distribution on the product space of early and late gene expression, connects cells at the earlier timepoint to their predicted descendants and cells at the later timepoint to their predicted ancestors. To compute the couplings, Waddington-OT solves an entropically regularized unbalanced optimal transport problem, minimizing the difference in gene expression between predicted ancestors and descendants subject to the constraint that the population at t_i maps to the population at t_{i+1} . Entropic regularization allows for stochasticity in cell fate decisions, whereas unbalanced transport accounts for uncertainty in the relative growth and death rates of different cells. In the absence of clear prior information about the relative growth rates of cells, we left the initial growth rate estimates uniform. By concatenating the coupling from E9.5 to E10.5 with the coupling from E10.5 to E11.5, we are able to predict ancestors and descendants across the timecourse.

scMultiome-seq

Nuclei isolation and transposition. FACS-sorted cells were gently resuspended in 50 μ l of chilled lysis buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl₂, 0.1% Tween 20, 0.1% NP-40 substitute, 0.01% digitonin, 1% BSA, 1 mM DTT and 1 U μ l⁻¹ RNase inhibitor) for 3 min on ice. Nuclei were washed with diluted nuclei buffer, visualized with trypan blue staining under a light microscope for quality control and transposed according to the manufacturer's instructions (Chromium Next GEM Single-Cell Multiome ATAC + Gene Expression User Guide, Rev F).

Library generation and pre-processing. Multiome sequencing of transposed nuclei was performed using the 10x Genomics Chromium platform, and sequencing libraries were generated using the Chromium Next GEM Single-Cell Multiome ATAC + Gene Expression Kit (10x Genomics, PN-1000285). Libraries were sequenced with the standalone mode set to the manufacturer's protocol on the Illumina NextSeq 500 platform using the NextSeq 500/550 High Output Kit version 2.5, 150 cycles (Illumina, 20024907), to a depth of approximately 27,300 reads per cell for the gene expression library and approximately 33,500 reads per cell for the ATAC library. FASTQ files were processed using the 10x Cell Ranger Arc pipeline (version 2.0.0) count function with default settings and mapped to the refdata-cellranger-arc-mm10-2020-A-2.0.0 reference.

Quality control and normalization. Cell Ranger Arc outputs were loaded into RStudio version 1.4 and further analyzed using Seurat version 5.0.1 (ref. 77) for gene expression and ATAC analyses and Signac

version 1.12.0 (ref. 78) for ATAC analyses. Before filtering, the ATAC data were processed to retain only peaks in standard chromosomes. Nuclei with fewer than 1,000 or more than 20,000 cDNA reads, fewer than 1,000 ATAC fragments or more than 15% mitochondrial reads were removed to retain 3,801 high-quality nuclei for analysis. RNA data were normalized using the NormalizeData function before calculation of variable features with FindVariableFeatures and scaling data with ScaleData. RunPCA was used to calculate the first 50 PCs. ATAC data were normalized with RunTFIDF, and top features were identified with FindTopFeatures before calculating dimensionality reduction with RunSVD.

Clustering. A weighted nearest neighbor (WNN) graph was constructed with FindMultimodalNeighbors using the 1–25 and 2–50 dimensions from the PCA and latent semantic indexing (LSI) reductions, respectively. This graph was subsequently used to calculate a uniform manifold approximation and projection (UMAP) for visualization and clustering with FindClusters (resolution = 0.8). Clusters were assigned biological annotation according to the top 10 enriched genes (FindAllMarkers) and visualization of known marker genes.

GRN analysis. To construct GRNs during LEC specification and differentiation, we applied the SCENIC+ pipeline⁷⁹ to our scMultiome dataset. High-quality singlets identified from the Seurat analysis described above were selected for SCENIC+ analysis.

Identification of enhancer candidates and topic modeling. pycisTopic⁷⁹ was used to identify enhancer candidates. ATAC fragment files were processed to generate pseudobulk BigWig and BAM files for each cell type and used to call peaks with MACS2 (ref. 80) and merge into consensus peaks with get_consensus_peaks(). Together with filtered fragments, these consensus regions were used to create a cisTopic object. We ran latent Dirichlet allocation (LDA) models on the cisTopic object with between two and 50 topics using run_cgs_models_mallet() and used evaluatemodels() to select 30 as the optimal number of co-accessible region sets. Topics were manually inspected for their enrichment in cell types to ensure that a range of cell-type-specific topics had been identified. Finally, differentially accessible regions between cell types were identified using pycisTopic's find_diff_features().

Generating a custom cisTarget database. pycisTarget⁸⁰ was used to create pre-computed enrichment scores for all SCENIC+ motifs across our consensus peaks. FASTA files were generated from consensus regions identified above with an additional kilobase of padding as a background sequence for cluster-buster. Finally, we used pycisTarget's create_cistarget_motif_databases.py to generate our custom database.

Inferring eGRN. Using the outputs from pycisTopic and pycisTarget, in combination with snRNA-seq data exported from Seurat to Scanpy, we ran the SCENIC+ snakemake pipeline using default settings to predict the enhancer-driven regulons constituting the GRN. The direct eRegulon metadata were used to examine the activity of gene-based and region-based eRegulons (AUCell scores) in heatmaps using heatmap_dotplot(). To export eRegulon network graphs, the mudata object output from the snakemake pipeline was converted to a sceniplus object for backwards compatibility. For network graphs of known EC differentiation (Extended Data Fig. 6h), the top 350 most variable genes and regions were selected as nodes. Conversely, when investigating putative regulators (Extended Data Fig. 6i), the genes and regions with the highest triplet score were selected.

In silico prediction of TF binding. To predict the binding of TFs in silico on a per-cell-type basis, we first generated ATAC pseudobulks using sinto (<https://github.com/timoast/sinto>). Peaks were called for each

pseudobulk using Genrich (<https://github.com/jsh58/Genrich>). The resulting .bam and .bed files were used to run the TOBIAS footprinting pipeline⁸¹. First, pseudobulks were corrected for Tn5 cutting bias with ATACorrect, followed by calculation of footprinting scores with FootprintingScores. Scores were calculated for all motifs in JASPAR 2024 core motif collection with motifs of interest manually inspected in Integrative Genomics Viewer (IGV).

Analysis of publicly available ETV2 ChIP-seq data. Publicly available ETV2 ChIP-seq data were retrieved from GSE59402 (ref. 37). BigWig files were converted from the mm9 to the mm10 genome for display with our snATAC-deq data using CrossMap⁸² and the mm9 to mm10 chain file. Converted BigWig files were displayed using IGV.

Bulk ATAC-seq analysis

Nuclei isolation and transposition. Bulk ATAC-seq libraries were generated following the Omni-ATAC protocol⁸³. In brief, BECs (tdTomato⁺CD41⁺CD45⁺PECAM1⁺PDPN⁻) and LECs (tdTomato⁺CD41⁺CD45⁻PECAM1⁺PDPN⁺) were FACS-sorted from E13.5 *Pax3^{Cre/+}; Rosa26^{tdTomato}* embryos. In total, 10,000 cells were gently resuspended in 50 μ l of chilled lysis buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl₂, 0.1% Tween 20, 0.1% NP-40 substitute, 0.01% digitonin, 1% BSA, 1 mM DTT and 1 U μ l⁻¹ RNase inhibitor) for 3 min on ice. Nuclei were washed with 990 μ l of chilled resuspension buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl₂ and 0.1% Tween 20), incubated in transposition buffer (1 \times TD buffer, 100 nM Tn5 Transposase (FC-121-1030, Illumina), 0.01% digitonin, 0.1% Tween 20 and 0.33 \times PBS) for 30 min at 37 °C and 1,000 rpm. Reactions were subsequently cleaned up with the MinElute Reaction Cleanup Kit (Qiagen, 28204).

Library generation and pre-processing. Sequencing libraries were generated following standard protocols⁸⁴ and amplified with a target concentration of 4 nM. Tagmentation efficiency was assessed with an Agilent TapeStation, and libraries were sequenced using paired-end 40-bp reads on an Illumina NextSeq 500. Reads were mapped to the mm10 genome assembly using Bowtie 2, and duplicates were removed using Picard's MarkDuplicates. To visualize peaks at regions of interest, BigWig files were generated using bamCoverage with a binSize of 10 for loading into IGV.

In situ HCR

Whole-mount in situ HCR is commercially available from Molecular Technologies⁸⁵. The HCR version 2.0 protocol for whole-mount mouse embryos was performed according to the manufacturer's instructions. The following probes were used: Etv2-B4 (NM_007959.2), Prox1-B2 (NM_008937.3), Pecam1-B3 (NM_008816.3), Lbx1-B1 (NC_000085.7), C1ql2-B3 (NM_207233.1), Tll1-B1 (NM_009390.3), Kdr-B3 (NM_010612.3), Flt4-B5 (NM_008029.3), Vegfa-B2 (NM_009505) and Vegfc-B1 (NM_009506.2). Embryos were mounted in SlowFade Gold Antifade Mountant (Invitrogen, S36936) and analyzed using a Zeiss inverted LSM 880 or LSM 980 or an Olympus FV1000 laser scanning confocal microscope.

RNAscope fluorescence in situ hybridization

RNA in situ hybridization on embryonic tissue was performed using the commercially available RNAscope Multiplex Fluorescent version 2 assays (Advanced Cell Diagnostics (ACD), 323100). Frozen tissue sections of 20- μ m thickness were processed according to the manufacturer's protocols for fresh-frozen samples. For thick vibratome sections of 300 μ m, a modified protocol was applied. In brief, tissue sections were dehydrated in 50%, 75% and 100% methanol for 10 min and then washed with PBS-T (PBS with 0.1% Tween 20). Sections were then pre-treated with hydrogen peroxide for 10 min at room temperature and Protease III reagent for 20 min at room temperature, followed by washing with PBS supplemented with protease inhibitor (Sigma-Aldrich, I1873580001).

Sections were post fixed using 4% PFA for 30 min at room temperature and washed with 0.2 \times saline sodium citrate (SSC) buffer. For RNA detection, sections were incubated with the following probes for 3 h at 40 °C: Mm-Prox1-C2 (ACD, 488591), Mm-Pecam1-C1 (ACD, 316721), Mm-Cdh5-C3 (ACD, 31253), Mm-Lyve1-C1 (ACD, 428451), Mm-Vegfc-C2 (ACD, 492701-C2) and Mm-Cbe1-C1 (ACD, 485651). Subsequent amplification steps were performed at 40 °C (AMP1-FL and AMP2-FL: 50 min each; AMP3-FL: 20 min), and each amplifier was removed by washing using 0.2 \times SSC buffer. For signal detection, sections were incubated with the channel-specific HRP for 20 min at 40 °C and incubated with the respective fluorophores (PerkinElmer: Fluorescein, 1:500; Cy3, 1:1,000; Cy5, 1:1,500; Opal 520, 1:750; Opal 620, 1:750) for 40 min at 40 °C, followed HRP blocker incubation for 20 min at 40 °C.

Analysis of cell cycle kinetics in mid-gestation embryos

The mean total cell cycle (T_c) and S-phase duration (T_s) of initial LECs at E10.5 were determined using an adapted dual-pulse labeling protocol^{86,87}. The experimental setup and calculation of cell cycle kinetics are outlined in Fig. 5f and Extended Data Fig. 9d, respectively. EdU (50 mg kg⁻¹, intraperitoneal) was administered to pregnant females three times, with a time interval of 2 h, to maintain bioavailability during the experiment. After 6 h, BrdU (50 mg kg⁻¹, intraperitoneal) was administered, and, 2 h later, embryos were dissected and fixed overnight in 4% PFA at 4 °C. Vibratome sections (150 μ m) were permeabilized in 0.5% Triton X-100 in PBS for 30 min at room temperature. Heat-induced antigen retrieval (HIER) was used to facilitate BrdU antibody labeling. Sections were submerged in sodium citrate buffer (10 mM sodium citrate, 0.05% Tween 20, pH 6.0) and heated to 98 °C for 30 min. Sections were cooled to room temperature with fresh sodium citrate buffer and washed for 5 min with PBS. After HIER, sections were incubated in blocking solution (3% BSA, 0.1% Tween 20 in PBS) for 2 h at room temperature and then incubated with antibodies to BrdU (Abcam, ab6326, 1:100), EMCN (D. Vestweber, VE44, 1:100) and PROX1 (Reliatech, 102-PA32, 1:100). Subsequently, slides were washed in PBS-T (PBS with 0.1% Tween 20) 3 times for 10 min each and incubated with Alexa Fluor-conjugated secondary antibodies for 2 h at room temperature or overnight at 4 °C. For EdU detection, the Click-iT Alexa Fluor 555 reaction cocktail (Thermo Fisher Scientific) was freshly prepared and incubated for 30 min at room temperature. Sections were rinsed in PBS and then mounted with Mowiol. Sections were imaged on a Zeiss LSM 880 confocal microscope using a \times 40 water immersion objective (NA = 1.2). Cell counts were performed on at least three 150- μ m sections, and individual data points were calculated as the mean of all sections analyzed per embryo. LEC progenitors at E10.5 are represented by PROX1⁺ nuclei (hereafter, P_{cells}). EdU⁺ and BrdU⁺ cells were scored as any nuclei showing immunoreactivity for these markers regardless of staining pattern.

The T_s of P_{cells} is given by equation (1):

$$T_s = \frac{T_i}{L_{\text{cells}}/S_{\text{cells}}} \quad (1)$$

T_i is the injection interval during which cells can incorporate at least one of the thymidine analogues. EdU⁺/BrdU⁺ double-positive cells reflect all cells in the S-phase at the end of the experiment (S_{cells}). Cells of the initial EdU-labeled S-phase population leave the S-phase during T_i , with this fraction labeled with EdU but not BrdU (L_{cells}).

The T_c of P_{cells} cells is given by equation (2):

$$T_c = \frac{T_s}{S_{\text{cells}}/(P_{\text{cells}} \times \text{GF})} \quad (2)$$

To avoid overestimation of T_c , which could occur if not all cells are actively progressing through the cell cycle, a growth fraction (GF) that represents the proportion of P_{cells} that are in the cell cycle was used for

calculation of T_c . GF can be determined by examining the proportion of P_{cells} labeled with the cell cycle marker KI67 using equation (3):

$$GF = \frac{P_{\text{cells}}^{\text{KI67}^+}}{P_{\text{cells}}} \quad (3)$$

Optical tissue clearing

The methods used for optical tissue clearing were described previously⁸⁸. In brief, before tissue clearing, whole-mount stained samples were embedded in 1% low-melting agarose to facilitate sample handling. Samples were dehydrated in increasing concentrations of methanol (50%, 70%, >99.5% and >99.5% (v/v)) for at least 1 h each. After dehydration, tissues were incubated in a 1:1 mixture of >99.5% methanol and BABB (1:2 benzyl alcohol:benzyl benzoate) for 1 h and, finally, in BABB for at least 4 h.

Light sheet microscopy

Optically cleared samples were imaged using an UltraMicroscope II Super Plan (Miltenyi Biotec) equipped with $\times 4$ MI Plan Objective (NA = 0.35, working distance (WD) ≥ 15 mm). An NKT SuperK supercontinuum white light laser served as excitation light source. For excitation and emission detection of specific fluorophores, custom band-pass filters (excitation 470/40, 577/25 or 640/30 nm; emission 525/50, 632/60 or 690/50 nm) were used in combination with a pco.edge 4.2 sCMOS camera. Images were acquired with 2- μm steps in the z axis.

Quantification of cell numbers

Confocal and light sheet image stacks were rendered into three-dimensional (3D) volumes and analyzed using Imaris version 9.5 (Bitplane; RRID: SCR_007370). Quantification of absolute cell numbers is based on staining of specific TFs to visualize the nuclei of cells of interest. Thus, the number of nuclei reflect the number of cells. Nuclei were automatically annotated using the 'Spots' function, which automatically detects point-like structures with a pre-defined diameter. Accurate quantification required an appropriate estimate of cell nuclei diameter and filtration of selected nuclei by tuning the quality parameters. The accuracy of this automatic counting procedure was verified by visual inspection, which herein served as ground truth. Using the 'manual Surface creation' function, vascular structures were segmented based on specific EC marker expression. Thereby, cell populations inside and outside of segmented vascular structures were defined by filtering the shortest distance between 'Spots' and 'Surface'.

Directional migration of LEC progenitors

To assess directional migration of PROX1⁺ LEC progenitors, we established an image analysis pipeline to automatically define and quantify cell polarity in large tissue sections. Vibratome sections (200 μm) of E10.5 embryos were stained for Golgi (GM130) and LEC nuclei (PROX1), and sections were imaged on a Zeiss LSM 880 confocal microscope using a $\times 40$ water immersion objective (NA = 1.2). In a first post-processing step, Golgi and nuclei in each image stack were segmented using the 'Surface' function in Imaris version 9.5 (Bitplane; RRID: SCR_007370). Cell populations inside and outside of defined vascular structures were defined as described above ('Quantification of cell numbers' subsection). Surface masks were exported and processed in Fiji version 1.53. Subsequently, nearest neighbor analyses were used to pair individual nuclei with their corresponding Golgi (closest border–border distance), and the centroid of each object was computed using 3D ImageJ Suite version 4.0.36 (ref. 89). Two-dimensional vectorization images were obtained by drawing arrows from the nuclei centroid toward the Golgi centroid using Fiji. Nuclei–Golgi pairs with a border-to-border distance larger than 5 μm were excluded from further analysis. Centroid vectors were produced using the x–y coordinates of nuclei and Golgi centroids and

transformed to unit vectors (A). The dorsal body axis served as reference vector (B). The angle was obtained by calculating the inverse cosine of the dot product of centroid unit vectors (A) and reference vector (B) ($\theta = \arccos(A \cdot B)$). All calculations were performed using Python version 3.8. Angles were transformed to represent the body axes (0°, dorsal; 90°, lateral; 180°, ventral; 270°, medial), and a histogram on a polar axis was used to display the angular distribution of individual LECs representing their migration direction.

Statistical analysis

Statistical analyses were performed by unpaired, two-tailed Student's *t*-test or non-parametric one-way ANOVA followed by Tukey's honestly significant difference (HSD) using GraphPad Prism software. Data are presented as mean \pm s.d. (error bars). $P < 0.05$ was considered statistically significant (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.001$).

Reporting summary

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

Data availability

scRNA-seq (GSE276281) and scMultiome-seq and bulk ATAC-seq (GSE276282) data were deposited in the Gene Expression Omnibus (GEO). ETV2 ChIP-seq data were downloaded from the GEO (GSE59402). Correspondence and requests for materials should be addressed to F.K. and O.A.S.

Code availability

All scripts required to reproduce the data presented are available on GitHub at https://github.com/StoneLabGH/LEC_Specification.

References

1. Oliver, G., Kipnis, J., Randolph, G. J. & Harvey, N. L. The lymphatic vasculature in the 21st century: novel functional roles in homeostasis and disease. *Cell* **182**, 270–296 (2020).
2. Mortimer, P. S. & Rockson, S. G. New developments in clinical aspects of lymphatic disease. *J. Clin. Invest.* **124**, 915–921 (2014).
3. Liu, X. et al. Lymphoangiocrine signals promote cardiac growth and repair. *Nature* **588**, 705–711 (2020).
4. Zhang, F. et al. Lacteal junction zippering protects against diet-induced obesity. *Science* **361**, 599–603 (2018).
5. Huntington, G. S. M. & McClure, C. F. W. The anatomy and development of the jugular lymph sacs in the domestic cat (*Felis domestica*). *Am. J. Anat.* **10**, 177–312 (1910).
6. Martinez-Corral, I. et al. Nonvenous origin of dermal lymphatic vasculature. *Circ. Res.* **116**, 1649–1654 (2015).
7. Nicenboim, J. et al. Lymphatic vessels arise from specialized angioblasts within a venous niche. *Nature* **522**, 56–61 (2015).
8. Sabin, F. R. On the origin of the lymphatic system from the veins and the development of the lymph hearts and thoracic duct in the pig. *Am. J. Anat.* **1**, 367–389 (1902).
9. Srinivasan, R. S. et al. Lineage tracing demonstrates the venous origin of the mammalian lymphatic vasculature. *Genes Dev.* **21**, 2422–2432 (2007).
10. Wilting, J. et al. Dual origin of avian lymphatics. *Dev. Biol.* **292**, 165–173 (2006).
11. Yaniv, K. et al. Live imaging of lymphatic development in the zebrafish. *Nat. Med.* **12**, 711–716 (2006).
12. Kohli, V., Schumacher, J. A., Desai, S. P., Rehn, K. & Sumanas, S. Arterial and venous progenitors of the major axial vessels originate at distinct locations. *Dev. Cell* **25**, 196–206 (2013).
13. Reischauer, S. et al. Cloche is a bHLH-PAS transcription factor that drives haemato-vascular specification. *Nature* **535**, 294–298 (2016).

14. Potente, M. & Makinen, T. Vascular heterogeneity and specialization in development and disease. *Nat. Rev. Mol. Cell Biol.* **18**, 477–494 (2017).
15. Wigle, J. T. & Oliver, G. *Prox1* function is required for the development of the murine lymphatic system. *Cell* **98**, 769–778 (1999).
16. Eng, T. C. et al. Zebrafish facial lymphatics develop through sequential addition of venous and non-venous progenitors. *EMBO Rep.* **20**, e47079 (2019).
17. Stanczuk, L. et al. *cKit* lineage hemogenic endothelium-derived cells contribute to mesenteric lymphatic vessels. *Cell Rep.* **10**, 1708–1721 (2015).
18. Stone, O. A. & Stainier, D. Y. R. Paraxial mesoderm is the major source of lymphatic endothelium. *Dev. Cell* **50**, 247–255 e243 (2019).
19. Hagerling, R. et al. A novel multistep mechanism for initial lymphangiogenesis in mouse embryos based on ultramicroscopy. *EMBO J.* **32**, 629–644 (2013).
20. Stone, O. A., Zhou, B., Red-Horse, K. & Stainier, D. Y. R. Endothelial ontogeny and the establishment of vascular heterogeneity. *Bioessays* **43**, e2100036 (2021).
21. Mayeuf-Louchart, A. et al. Notch regulation of myogenic versus endothelial fates of cells that migrate from the somite to the limb. *Proc. Natl Acad. Sci. USA* **111**, 8844–8849 (2014).
22. Bowles, J. et al. Control of retinoid levels by CYP26B1 is important for lymphatic vascular development in the mouse embryo. *Dev. Biol.* **386**, 25–33 (2014).
23. Marino, D., Dabouras, V., Brandli, A. W. & Detmar, M. A role for all-trans-retinoic acid in the early steps of lymphatic vasculature development. *J. Vasc. Res.* **48**, 236–251 (2011).
24. Zhang, H. et al. Genetic lineage tracing identifies endocardial origin of liver vasculature. *Nat. Genet.* **48**, 537–543 (2016).
25. Soldatov, R. et al. Spatiotemporal structure of cell fate decisions in murine neural crest. *Science* **364**, eaas9536 (2019).
26. Gong, W. et al. ETV2 functions as a pioneer factor to regulate and reprogram the endothelial lineage. *Nat. Cell Biol.* **24**, 672–684 (2022).
27. Maruyama, K. et al. The cardiopharyngeal mesoderm contributes to lymphatic vessel development in mouse. *eLife* **11**, e81515 (2022).
28. Maruyama, K., Miyagawa-Tomita, S., Mizukami, K., Matsuzaki, F. & Kurihara, H. *Isl1*-expressing non-venous cell lineage contributes to cardiac lymphatic vessel development. *Dev. Biol.* **452**, 134–143 (2019).
29. Jagla, K. et al. Mouse *Lbx1* and human *LBX1* define a novel mammalian homeobox gene family related to the *Drosophila* lady bird genes. *Mech. Dev.* **53**, 345–356 (1995).
30. Francois, M. et al. Sox18 induces development of the lymphatic vasculature in mice. *Nature* **456**, 643–647 (2008).
31. Lin, H. & Grosschedl, R. Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature* **376**, 263–267 (1995).
32. Garel, S., Yun, K., Grosschedl, R. & Rubenstein, J. L. The early topography of thalamocortical projections is shifted in *Ebf1* and *Dlx1/2* mutant mice. *Development* **129**, 5621–5634 (2002).
33. Kieslinger, M. et al. EBF2 regulates osteoblast-dependent differentiation of osteoclasts. *Dev. Cell* **9**, 757–767 (2005).
34. Rajakumari, S. et al. EBF2 determines and maintains brown adipocyte identity. *Cell Metab.* **17**, 562–574 (2013).
35. Kazenwadel, J. et al. GATA2 is required for lymphatic vessel valve development and maintenance. *J. Clin. Invest.* **125**, 2979–2994 (2015).
36. Kazenwadel, J. et al. A *Prox1* enhancer represses haematopoiesis in the lymphatic vasculature. *Nature* **614**, 343–348 (2023).
37. Liu, F. et al. Induction of hematopoietic and endothelial cell program orchestrated by ETS transcription factor ER71/ETV2. *EMBO Rep.* **16**, 654–669 (2015).
38. Nord, A. S. et al. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* **155**, 1521–1531 (2013).
39. Southard, S. et al. A series of Cre-ER^{T2} drivers for manipulation of the skeletal muscle lineage. *Genesis* **52**, 759–770 (2014).
40. Metikala, S. et al. Integration of vascular progenitors into functional blood vessels represents a distinct mechanism of vascular growth. *Dev. Cell* **57**, 767–782 (2022).
41. Koltowska, K. et al. Vegfc regulates bipotential precursor division and *Prox1* expression to promote lymphatic identity in zebrafish. *Cell Rep.* **13**, 1828–1841 (2015).
42. Abe, T. et al. Visualization of cell cycle in mouse embryos with Fucci2 reporter directed by *Rosa26* promoter. *Development* **140**, 237–246 (2013).
43. Karkkainen, M. J. et al. Vascular endothelial growth factor C is required for sprouting of the first lymphatic vessels from embryonic veins. *Nat. Immunol.* **5**, 74–80 (2004).
44. Davis, J. A. et al. ETS transcription factor *Etsrp/Etv2* is required for lymphangiogenesis and directly regulates *vegfr3/flt4* expression. *Dev. Biol.* **440**, 40–52 (2018).
45. Srinivasan, R. S. et al. The *Prox1*-*Vegfr3* feedback loop maintains the identity and the number of lymphatic endothelial cell progenitors. *Genes Dev.* **28**, 2175–2187 (2014).
46. Jeltsch, M. et al. *CCBE1* enhances lymphangiogenesis via A disintegrin and metalloprotease with thrombospondin motifs-3-mediated vascular endothelial growth factor-C activation. *Circulation* **129**, 1962–1971 (2014).
47. Franco, C. A. et al. Dynamic endothelial cell rearrangements drive developmental vessel regression. *PLoS Biol.* **13**, e1002125 (2015).
48. Sabin, F. R. The lymphatic system in human embryos, with a consideration of the system as a whole. *Am. J. Anat.* **9**, 43–91 (1909).
49. Klotz, L. et al. Cardiac lymphatics are heterogeneous in origin and respond to injury. *Nature* **522**, 62–67 (2015).
50. Ginsberg, M. et al. Efficient direct reprogramming of mature amniotic cells into endothelial cells by ETS factors and TGFβ suppression. *Cell* **151**, 559–575 (2012).
51. Gomez-Salainero, J. M. et al. Cooperative ETS transcription factors enforce adult endothelial cell fate and cardiovascular homeostasis. *Nat. Cardiovasc. Res.* **1**, 882–899 (2022).
52. Bossen, C. et al. The chromatin remodeler Brg1 activates enhancer repertoires to establish B cell identity and modulate cell growth. *Nat. Immunol.* **16**, 775–784 (2015).
53. Kim, E. E. et al. The transcription factor EBF1 non-cell-autonomously regulates cardiac growth and differentiation. *Development* **150**, dev202054 (2023).
54. Angueira, A. R. et al. Early B cell factor activity controls developmental and adaptive thermogenic gene programming in adipocytes. *Cell Rep.* **30**, 2869–2878 (2020).
55. Panara, V. et al. Multiple cis-regulatory elements control *prox1a* expression in distinct lymphatic vascular beds. *Development* **151**, dev202525 (2024).
56. Makinen, T. et al. Inhibition of lymphangiogenesis with resulting lymphedema in transgenic mice expressing soluble VEGF receptor-3. *Nat. Med.* **7**, 199–205 (2001).
57. Nurmi, H. et al. VEGF-C is required for intestinal lymphatic vessel maintenance and lipid absorption. *EMBO Mol. Med.* **7**, 1418–1425 (2015).
58. Gancz, D. et al. Distinct origins and molecular mechanisms contribute to lymphatic formation during cardiac growth and regeneration. *eLife* **8**, e44153 (2019).
59. Jafree, D. J. et al. Spatiotemporal dynamics and heterogeneity of renal lymphatics in mammalian development and cystic kidney disease. *eLife* **8**, e48183 (2019).

60. Pardanaud, L. et al. Two distinct endothelial lineages in ontogeny, one of them related to hemopoiesis. *Development* **122**, 1363–1371 (1996).
61. Lang, D. et al. Pax3 functions at a nodal point in melanocyte stem cell differentiation. *Nature* **433**, 884–887 (2005).
62. Sieber, M. A. et al. Lbx1 acts as a selector gene in the fate determination of somatosensory and viscerosensory relay neurons in the hindbrain. *J. Neurosci.* **27**, 4902–4909 (2007).
63. Tallquist, M. D., Weismann, K. E., Hellstrom, M. & Soriano, P. Early myotome specification regulates PDGFA expression and axial skeleton development. *Development* **127**, 5059–5070 (2000).
64. Koni, P. A. et al. Conditional vascular cell adhesion molecule 1 deletion in mice: impaired lymphocyte migration to bone marrow. *J. Exp. Med.* **193**, 741–754 (2001).
65. Theis, M. et al. Endothelium-specific replacement of the connexin43 coding region by a lacZ reporter gene. *Genesis* **29**, 1–13 (2001).
66. Madisen, L. et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
67. Luche, H., Weber, O., Nageswara Rao, T., Blum, C. & Fehling, H. J. Faithful activation of an extra-bright red fluorescent protein in 'knock-in' Cre-reporter mice ideally suited for lineage tracing studies. *Eur. J. Immunol.* **37**, 43–53 (2007).
68. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
69. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
70. Kowalczyk, M. S. et al. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* **25**, 1860–1872 (2015).
71. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).
72. Kluyver, T. et al. Positioning and power in academic publishing: players, agents and agendas. In *Proc. 20th International Conference on Electronic Publishing* (eds Loizides, F. & Schmidt, B.) 87–90 (IOS Press, 2016).
73. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15 (2018).
74. Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019).
75. Jacomy, M., Venturini, T., Heymann, S. & Bastian, M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE* **9**, e98679 (2014).
76. Schiebinger, G. et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176**, 928–943 (2019).
77. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
78. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
79. Bravo Gonzalez-Blas, C. et al. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat. Methods* **20**, 1355–1367 (2023).
80. Zhang, Y. et al. Model-based analysis of CHIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
81. Bentsen, M. et al. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat. Commun.* **11**, 4267 (2020).
82. Zhao, H. et al. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006–1007 (2014).
83. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
84. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.21–21.29.29 (2015).
85. Choi, H. M. T. et al. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145**, dev165753 (2018).
86. Harris, L., Zalucki, O. & Piper, M. BrdU/EdU dual labeling to determine the cell-cycle dynamics of defined cellular subpopulations. *J. Mol. Histol.* **49**, 229–234 (2018).
87. Martynoga, B., Morrison, H., Price, D. J. & Mason, J. O. Foxg1 is required for specification of ventral telencephalon and region-specific regulation of dorsal telencephalic precursor proliferation and apoptosis. *Dev. Biol.* **283**, 113–127 (2005).
88. Orlich, M. & Kiefer, F. A qualitative comparison of ten tissue clearing techniques. *Histol. Histopathol.* **33**, 181–199 (2018).
89. Ollion, J., Cochenec, J., Loll, F., Escude, C. & Boudier, T. TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization. *Bioinformatics* **29**, 1840–1841 (2013).

Acknowledgements

We thank H. Gerhardt (Max Delbrück Centre for Molecular Medicine) and C. Birchmeier (Max Delbrück Centre for Molecular Medicine) for facilitating this work; C. Franco (Católica Biomedical Research Centre) for providing reagents; S. Gauvrit (University of Saskatchewan) and T. Sauka-Spengler (University of Oxford) for discussions; and D. Stainer (Max Planck Institute for Heart and Lung Research), H. Gerhardt, S. De Val (University of Oxford), B. Vanhollebeke (Université libre de Bruxelles) and S. Herbert (University of Manchester) for reading the paper. We gratefully acknowledge the Oxford-ZEISS Centre of Excellence in Biomedical Imaging for their support and assistance in this work. The Oxford-ZEISS Centre of Excellence is supported by the Kennedy Trust for Rheumatology Research, the Institute of Developmental and Regenerative Medicine and Carl Zeiss Gmbh. This work was also supported by a Sir Henry Dale Fellowship, jointly funded by the Wellcome Trust and the Royal Society (218561/Z/19/Z, O.A.S.); the British Heart Foundation (RE/13/1/30181, O.A.S.); the John Fell Fund (O.A.S. and D.E.G.); the German Research Foundation (DFG) (386797833 (CRC1348/1, F.K.) and 431460824 (CRC1450/1, F.K.)); the IZKF Münster (Kief/019/20, F.K.); the Max Planck Society (F.K.); CiM-IMPRS (EXC 1003 – CiM, F.K.); the Knut and Alice Wallenberg Foundation (2018.0218, T.M.); the Göran Gustafsson Foundation (T.M.); and the Swedish Research Council (2020-0269, T.M.).

Author contributions

I.-E.L., N.K., F.K. and O.A.S. conceived the study. I.-E.L., D.E.G., N.K., I.M.-C., T.M., F.K. and O.A.S. designed experiments. I.-E.L., D.E.G. and O.A.S. performed scRNA-seq and scMultiome-seq experiments. I.-E.L., D.E.G., N.K., E.Z., I.M.-C. H.S., M.V., G.J., J.G., F.K. and O.A.S. performed experimental work. I.-E.L., D.E.G. and A.F. performed bioinformatic analyses. S.W. and T.Z. provided code for image analyses. I.L., P.R.R. and K.A. provided essential reagents and approvals. T.M., F.K. and O.A.S. supervised the work. I.-E.L., D.E.G., N.K., F.K. and O.A.S. wrote the paper, with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s44161-024-00570-5>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44161-024-00570-5>.

Correspondence and requests for materials should be addressed to Friedemann Kiefer or Oliver A. Stone.

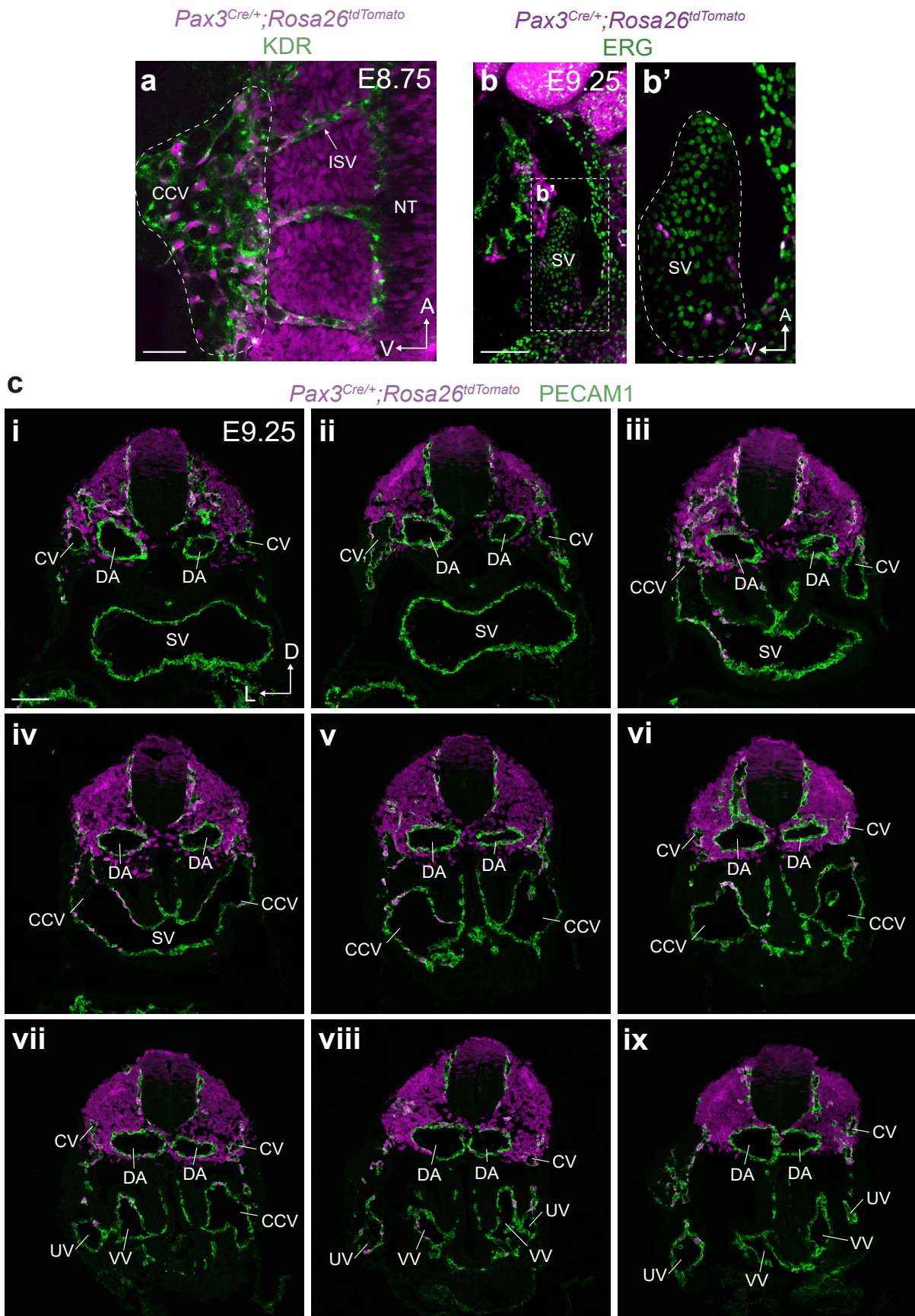
Peer review information *Nature Cardiovascular Research* thanks Gou Young Koh and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

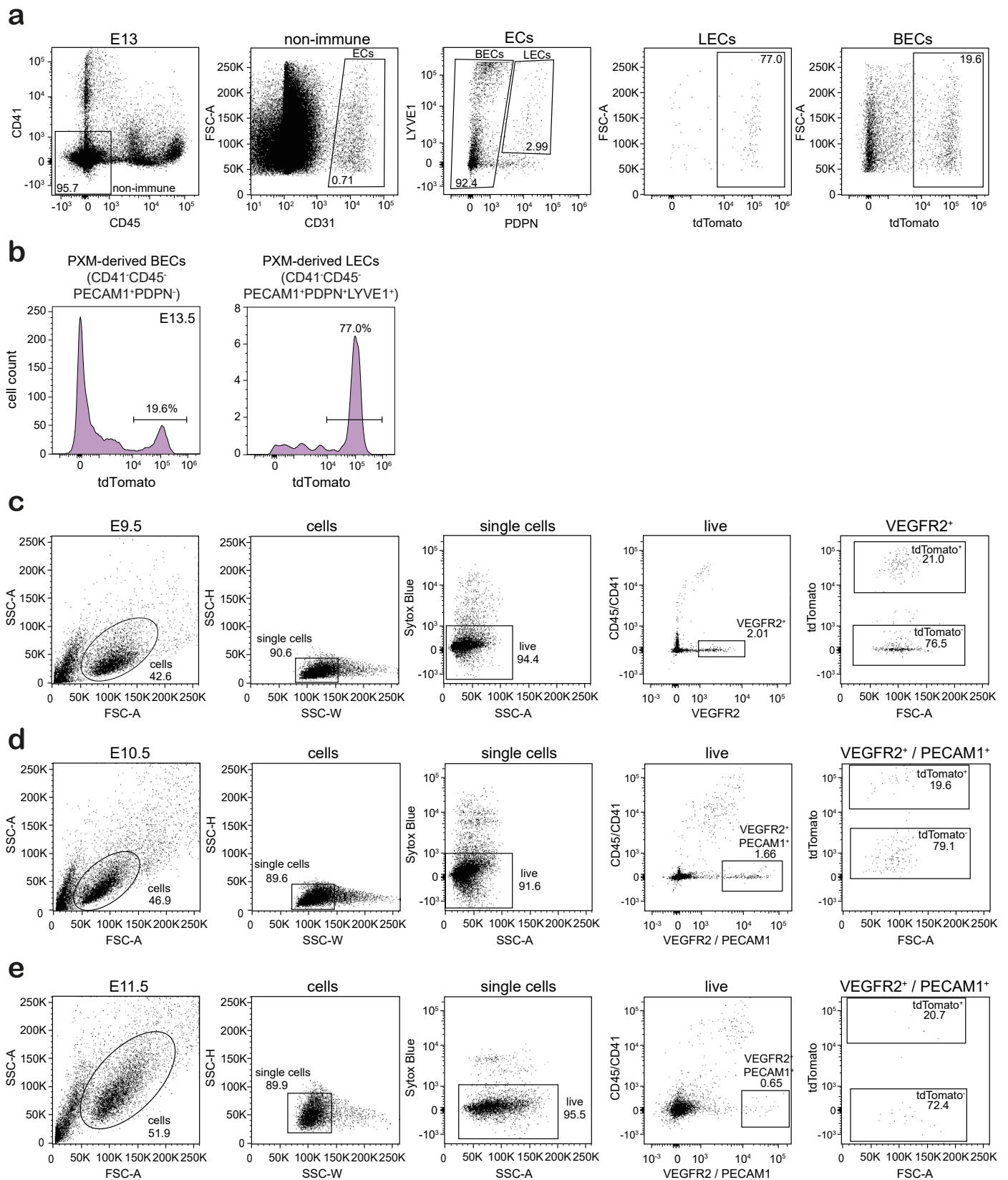
© The Author(s) 2025



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

Extended Data Fig. 1 | Contribution of *Pax3*-lineage cells to the developing vasculature. (a) Representative whole mount immunofluorescence for tdTomato and VEGFR2 in a *Pax3*^{Cre/+};*Rosa26*^{tdTomato} embryo at E8.75. Lateral view at the level of somites II-V (n = 4). (b) Representative whole mount immunofluorescence for ERG and tdTomato at the level of the sinus venosus in a *Pax3*^{Cre/+};*Rosa26*^{tdTomato} embryo at E9.25 (n = 4). (b') High magnification view

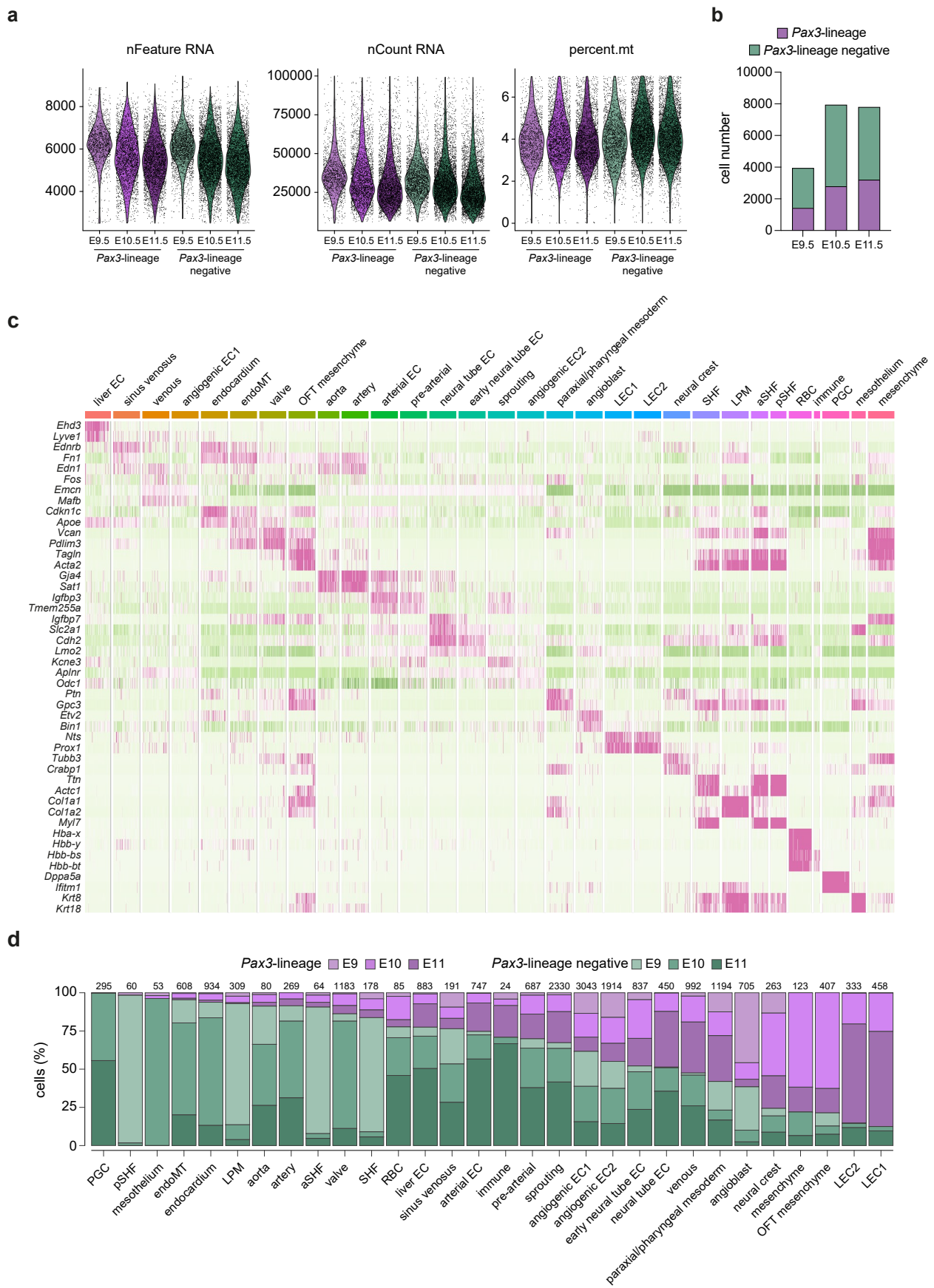
of boxed area in b. (c) Representative immunofluorescence for PECAM1 and tdTomato on transverse cryosections spanning anterior (i) to posterior (ix) regions of an individual *Pax3*^{Cre/+};*Rosa26*^{tdTomato} embryo at E9.25 (n = 3). (ISV, intersegmental vessel; CCV, common cardinal vein; SV, sinus venosus; CV, cardinal vein; DA, dorsal aorta; UV, umbilical vein; VV, vitelline vein. Scale bars - 50 μ m (a,c), 100 μ m (b)).



Extended Data Fig. 2 | Flow cytometry and FACS sorting strategies.

(a) Flow cytometry analysis of tdTomato labelling of LECs and BECs in *Pax3^{Cre/+}; Rosa26^{tdTomato}* embryos at E13.5. (b) Flow cytometry analysis of tdTomato labelling of blood endothelial cell (BEC) and lymphatic endothelial cell (LEC)

populations in *Pax3^{Cre/+}; Rosa26^{tdTomato}* embryos at E13.5. Gating strategy to sort individual cells from *Pax3^{Cre/+}; Rosa26^{tdTomato}* embryos for scRNA-seq at (c) E9.5, (d) E10.5 and (e) E11.5.

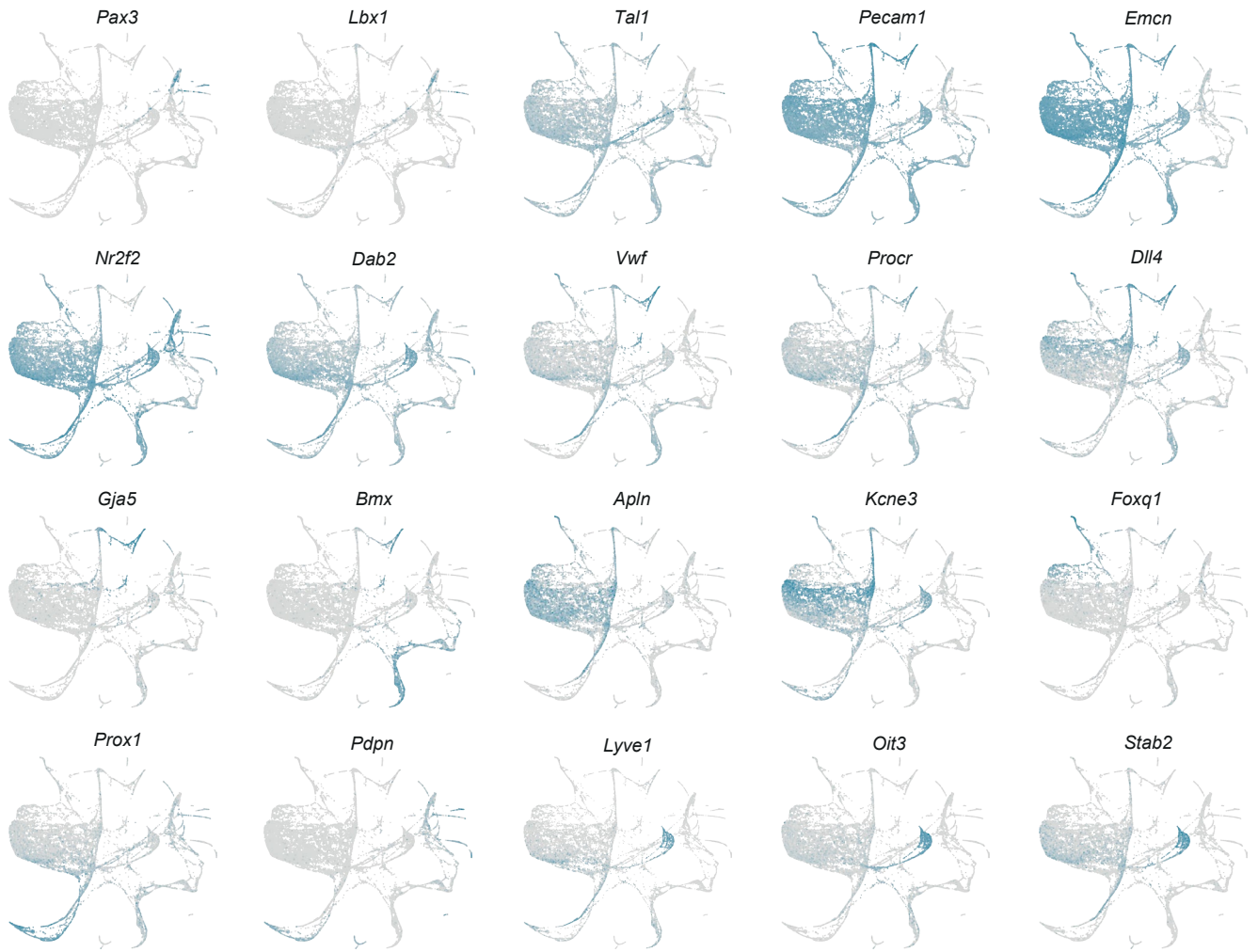


Extended Data Fig. 3 | See next page for caption.

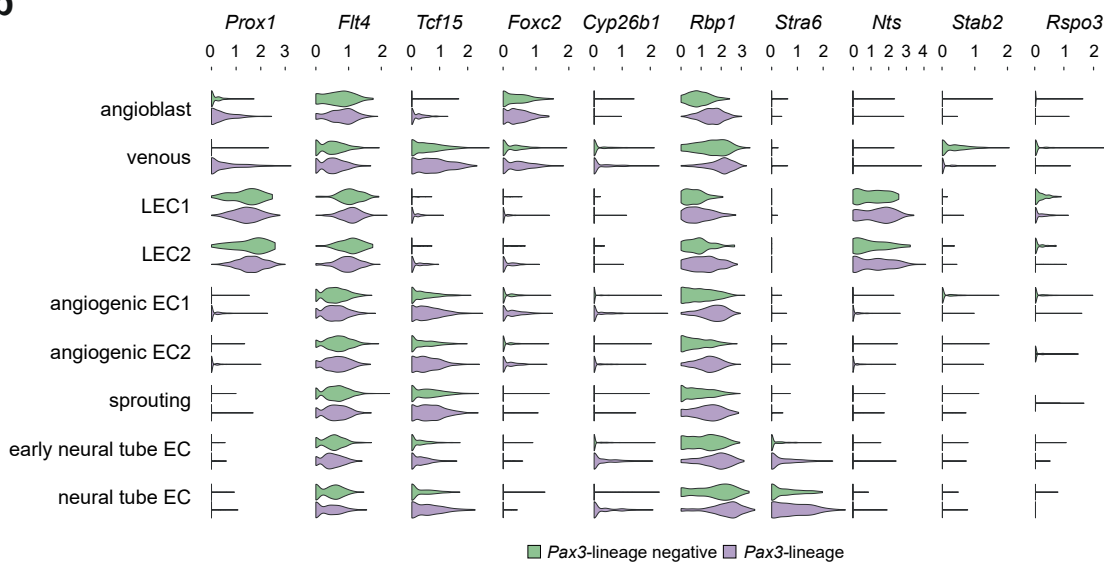
Extended Data Fig. 3 | scRNA-seq Quality Control. (a) Quality control (QC) plots showing RNA features, counts and percentage of mitochondrial reads per cell. Cells with less than 2500 detected features (genes), more than 100,000 UMIs (counts) and 7% of mitochondrial reads were excluded from downstream analyses. (b) Histogram showing the number of single cells from each lineage and stage that passed QC. (c) Heatmap showing normalized expression of two

diagnostic markers for each cell state. (d) Histogram showing the number and percentage of single cells from each lineage and stage assigned to each cell state. (EC, endothelial cell; OFT, outflow tract; LEC, lymphatic endothelial cell; SHF, second heart field; LPM, lateral plate mesoderm; NC mesenchyme, neural crest derived mesenchyme; aSHF, anterior second heart field; pSHF, posterior second heart field; RBC, red blood cell; PGC, primordial germ cell).

a



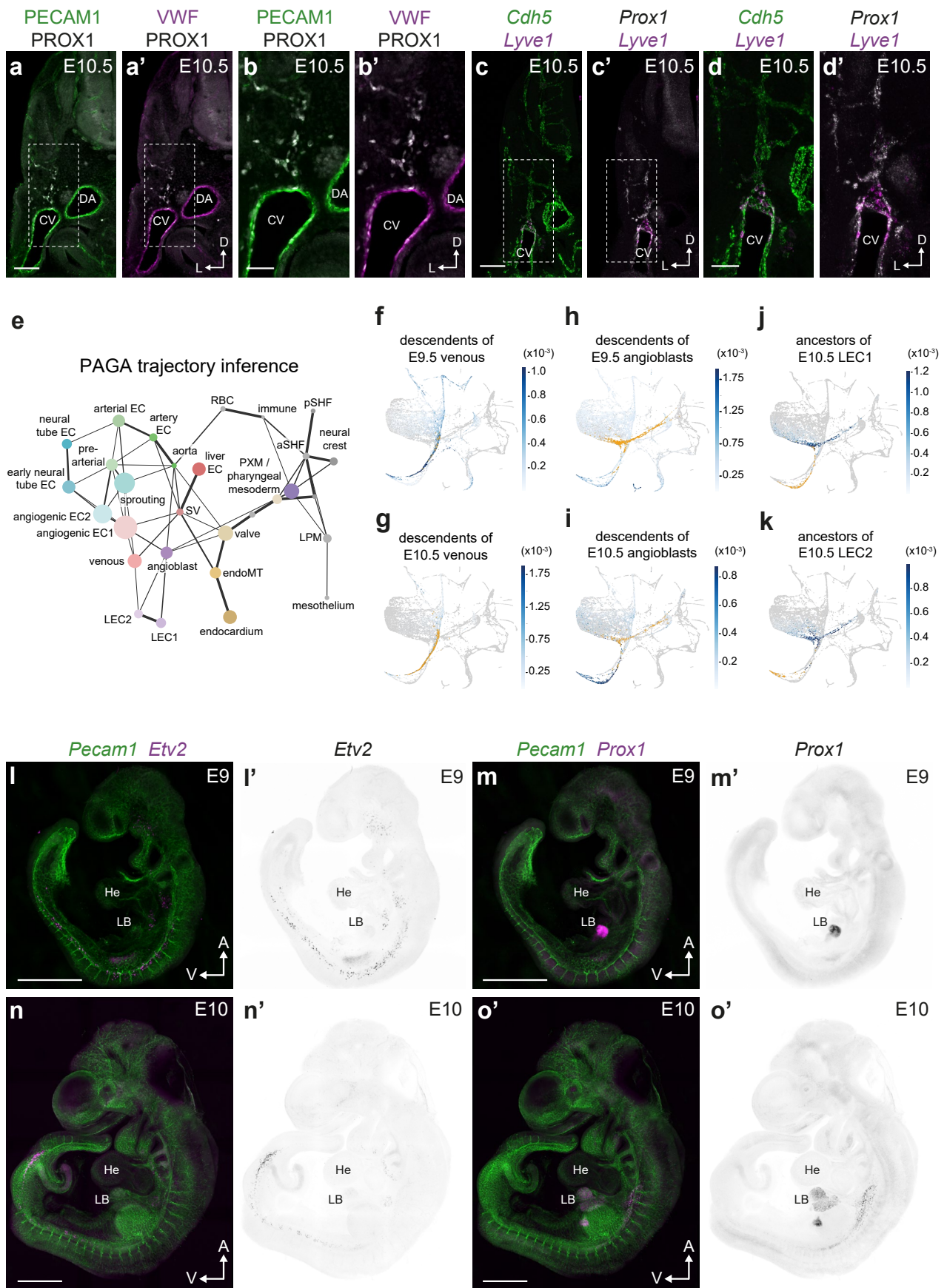
b



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | scRNA-seq gene expression analyses. (a) ForceAtlas2 (FA) embedding showing expression of somitic paraxial mesoderm (*Pax3*, *Lbx1*), angioblast/early EC (*Tal1*), endothelial (*Pecam1*, *Emcn*), venous (*Nr2f2*, *Dab2*), large vessel (*Vwf*, *Procr*), arterial (*Dll4*, *Gja5*, *Bmx*), angiogenic EC (*Apln*, *Kcne3*),

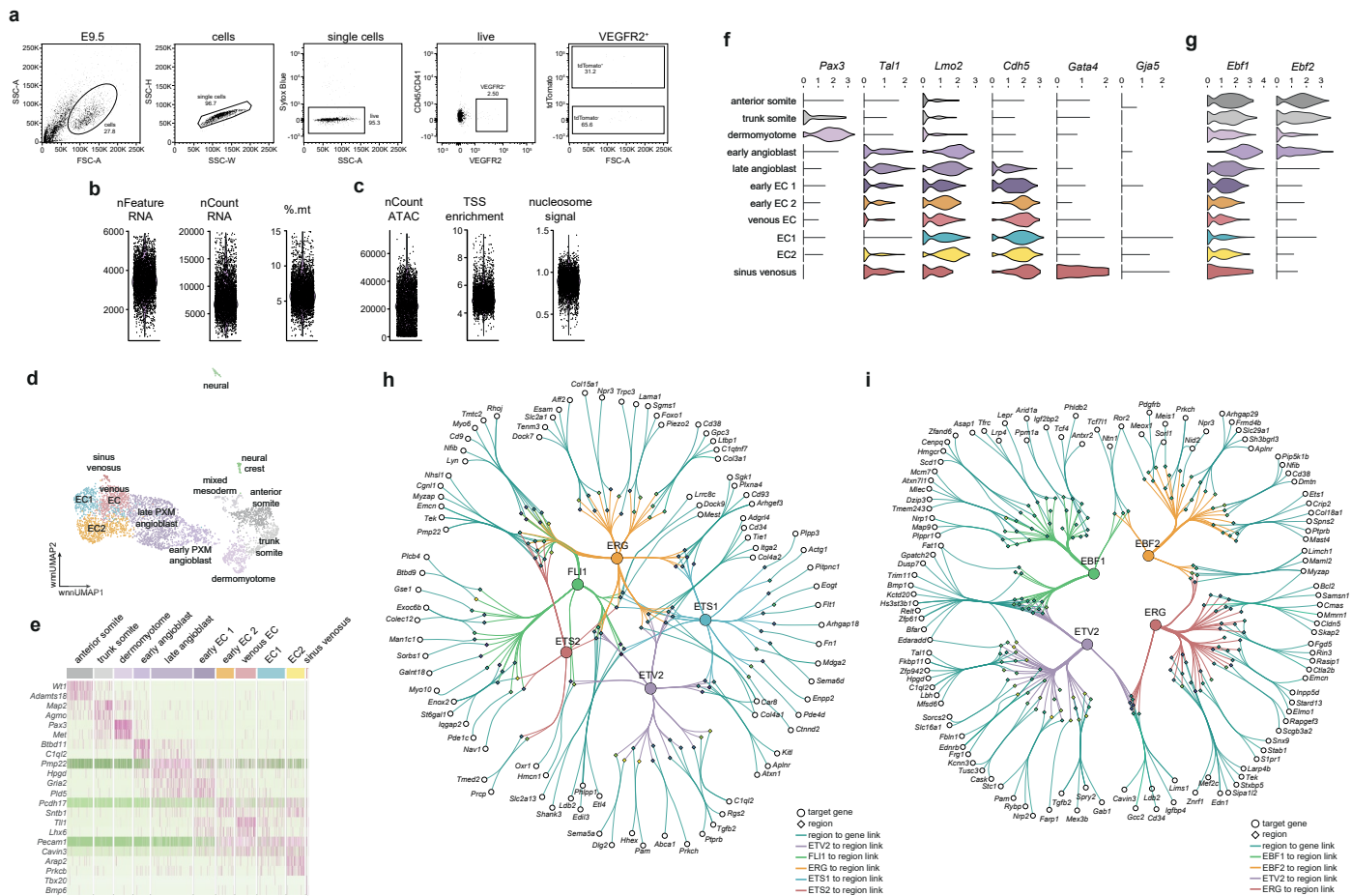
neural tube EC (*Foxq1*), lymphatic EC (*Prox1*, *Pdpr*, *Lyve1*) and sinus venosus/liver EC (*Oit3*, *Stab2*) markers. (b) Violin plots showing expression of selected genes in *Pax3*-lineage negative and *Pax3*-lineage positive cells across indicated cellular states.



Extended Data Fig. 5 | See next page for caption.

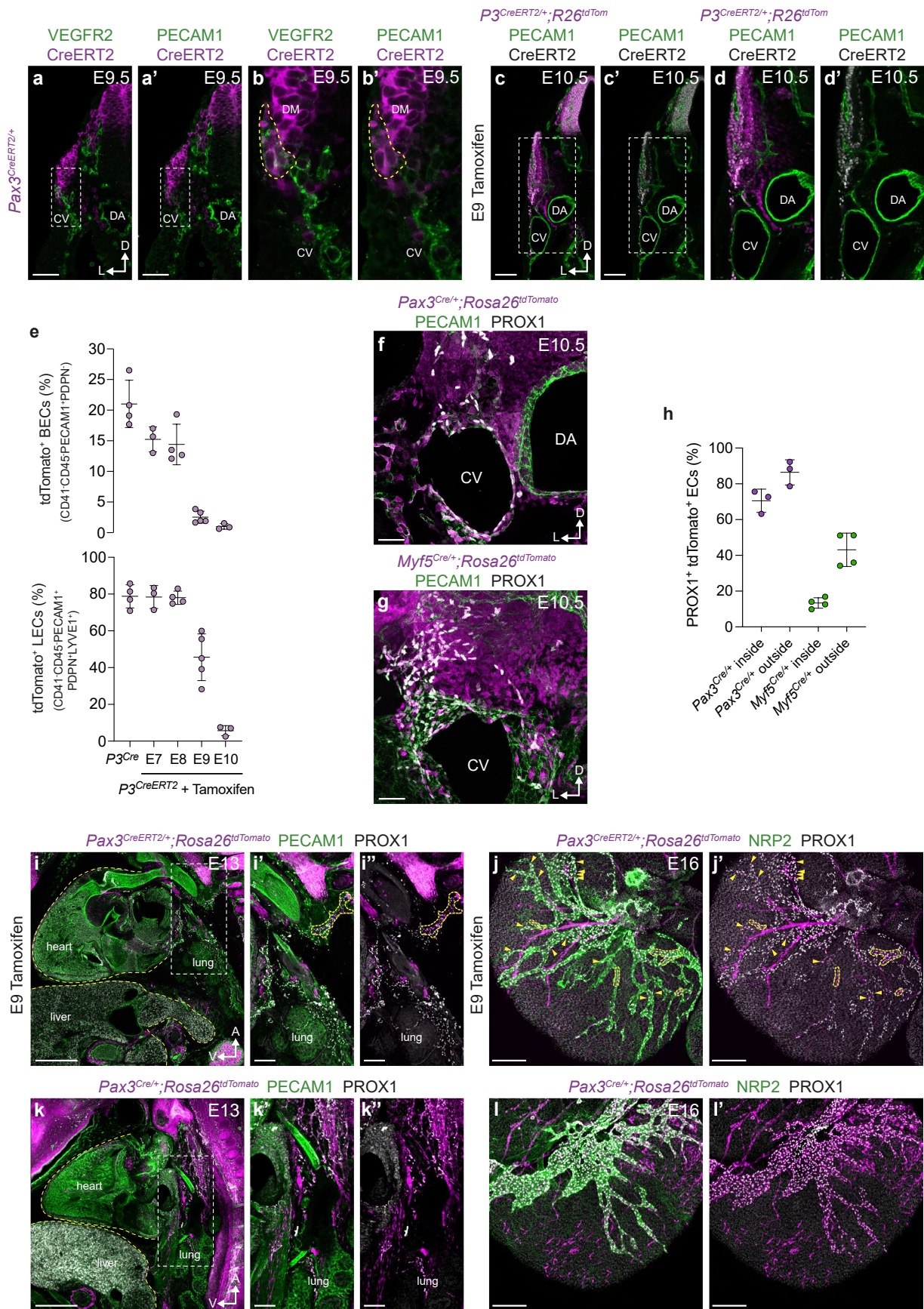
Extended Data Fig. 5 | Gene expression and trajectory analyses. Representative immunofluorescence images of PECAM1 and PROX1 (**a,b**) or VWF and PROX1 (**a',b'**) on a transverse vibratome section from an E10.5 embryo (**b-b'** - high magnification images of boxed areas in **a-a'**) (n = 4). *In situ* hybridization for *Cdh5* and *Lyve1* (**c,d**) or *Prox1* and *Lyve1* (**c'd'**) on a transverse vibratome section from an E10.5 embryo (**d-d'** - high magnification images of boxed areas in **c-c'**) (n = 4). (**e**) Partition-based graph abstraction (PAGA) inference of developmental trajectories on 19,699 cells. Cellular states were manually annotated based on known gene expression patterns. FA embedding showing Waddington-OT-based optimal transport analysis of (**f**) descendants of E9.5 venous cells, (**g**) descendants of E10.5 venous cells, (**h**) descendants of E9.5 angioblasts, (**i**)

descendants of E10.5 angioblasts, (**j**) ancestors of E10.5 LEC progenitors and (**k**) ancestors of E10.5 LECs. (**l-m'**) Representative whole mount imaging of *Pecam1*, *Etv2* and *Prox1* expression at E9.0 using hybridization chain reaction (n = 6). (**n-o'**) Representative whole mount imaging of *Pecam1*, *Etv2* and *Prox1* expression at E10.0 using hybridization chain reaction (n = 6). (CV, cardinal vein; DA, dorsal aorta; He, heart; LB, liver bud; EC, endothelial cell; LEC, lymphatic endothelial cell; SV, sinus venosus; SHF, second heart field; PXM, paraxial mesoderm; LPM, lateral plate mesoderm; aSHF, anterior second heart field; pSHF, posterior second heart field; RBC, red blood cell; PGC, primordial germ cell; Scale bars - 100 μ m (**a-a'**, **c-c'**), 50 μ m (**b-b'**, **d-d'**), 1 mm (**l-o'**)).



Extended Data Fig. 6 | scMultiome Quality Control and gene regulatory network analysis. (a) Gating strategy to sort individual cells from *Pax3^{Cre/+};Rosa26^{tdTomato}* embryos for scRNA-seq at E9.5. (b) Quality control (QC) plots showing RNA features, counts and percentage of mitochondrial reads per cell. (c) Quality control (QC) plots showing ATAC counts, transcription start site (TSS) and percentage nucleosome signal per cell. (d) Weighted nearest neighbour Uniform Manifold Approximation and Projection (wnnUMAP) embedding of single cell multiome sequencing of 3,801 *tdTomato⁺ VEGFR2⁺ PECAM1⁺* cells FACS-sorted from *Pax3^{Cre/+};Rosa26^{tdTomato}* embryos at E9.5. (e)

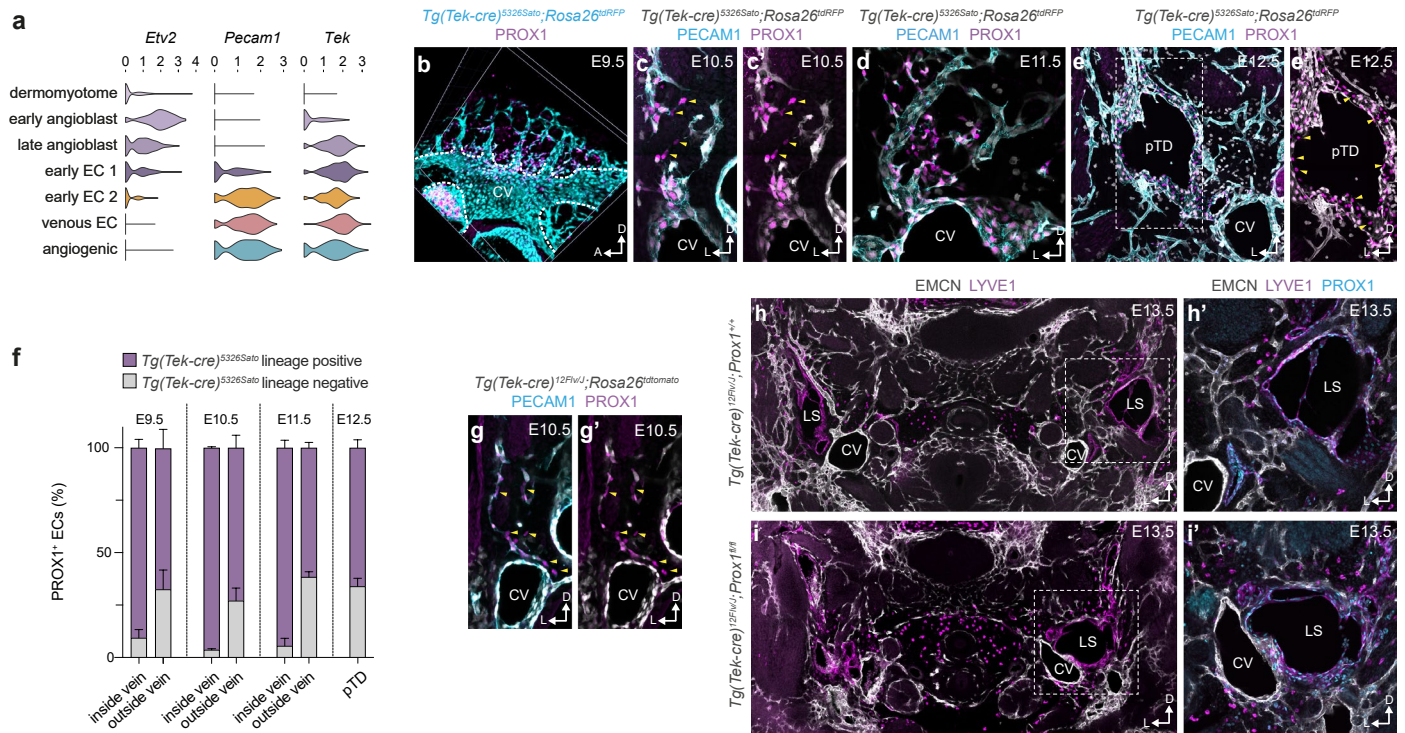
Heatmap showing normalized expression of two diagnostic markers for each cell state on the subset of 3,606 cells in Fig. 2a. (f) Violin plots showing expression of *Pax3*, *Tal1*, *Lmo2*, *Cdh5*, *Gata4* and *Gja5* across each cellular state. (g) Violin plots showing expression of *Ebf1* and *Ebf2* across each cellular state. (h) Visualisation of the eGRN formed by ETV2, FLI1, ERG, ETS1 and ETS2. The network consists of selected transcription factors and the top 300 most variable regions and genes. (i) Visualisation of the eGRN formed by ETV2, ERG, EBF1 and EBF2. The network consists of selected transcription factors and the regions and genes with the 30 highest triplet scores for each of the chosen eRegulons.



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

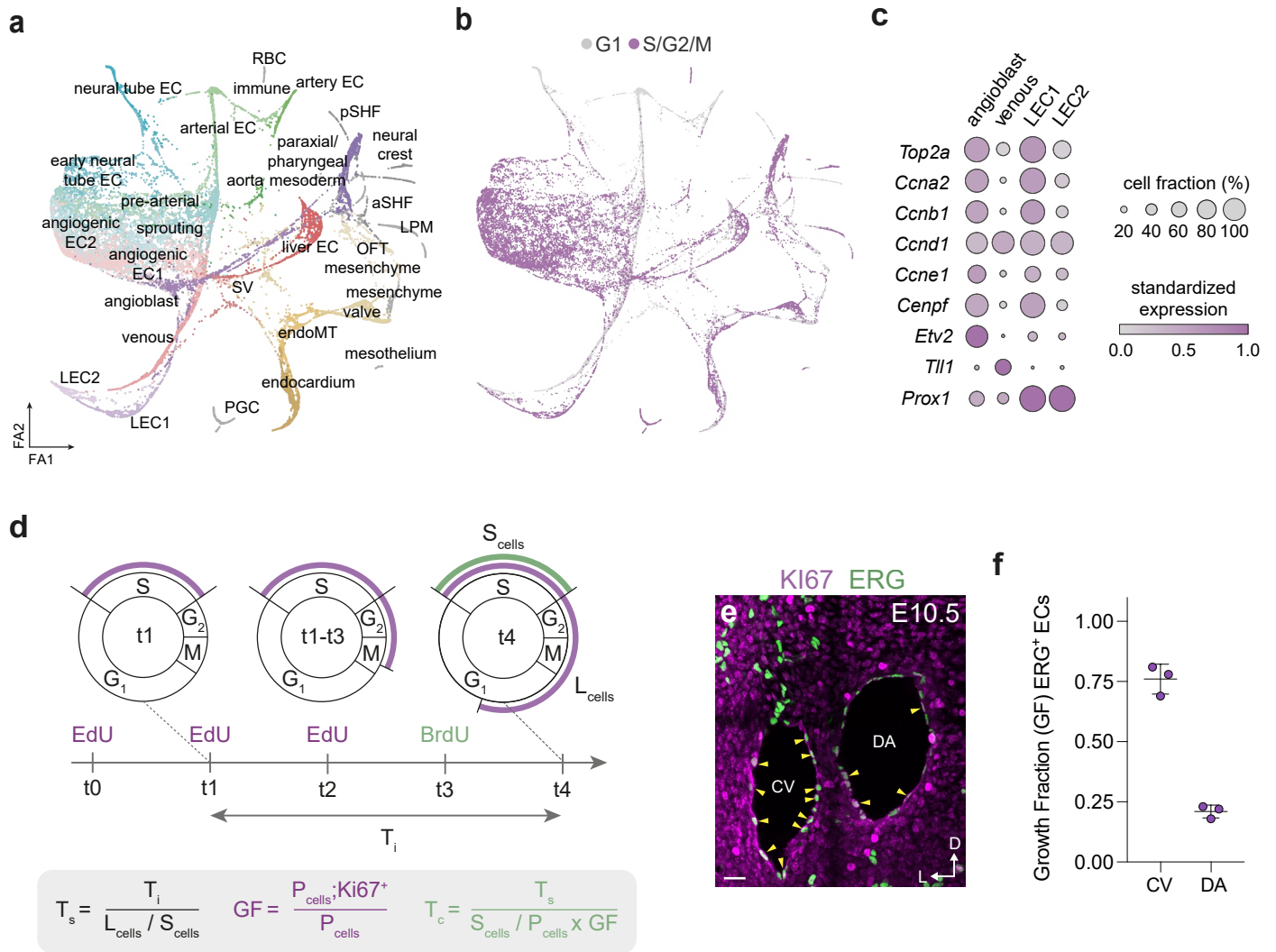
Extended Data Fig. 7 | Spatiotemporal analyses of LEC lineage. Representative immunofluorescence image of VEGFR2 and CreERT2 (detected with an ESRI antibody) (**a, b**) or PECAM1 and CreERT2 (**a', b'**) on a transverse vibratome section from an E9.5 *Pax3^{CreERT2/+}* embryo (**b-b'** - high magnification image of boxed area in **a-a'**) (n = 4). (**c-c'**) Representative immunofluorescence image of tdTomato, PECAM1 and CreERT2 on a transverse vibratome section from a *Pax3^{CreERT2/+};Rosa26^{tdTomato}* embryo at E10.5 following tamoxifen administration at E9 (n = 4). (**d-d'**) High magnification image of boxed area in **c-c'**. (**e**) Quantification of tdTomato labelling of BECs and LECs by flow cytometry in *Pax3^{Cre/+};Rosa26^{tdTomato}* and *Pax3^{CreERT2/+};Rosa26^{tdTomato}* embryos at E13.5. Tamoxifen was administered to *Pax3^{CreERT2/+};Rosa26^{tdTomato}* animals at E7 (n = 3, 1 pregnant dam), E8.0 (n = 4, 2 pregnant dams), E9.0 (n = 5, 3 pregnant dams) or E10.0 (n = 3, 2 pregnant dams). Representative immunofluorescence images of tdTomato, PECAM1 and PROX1 on transverse vibratome sections from (**f**) *Pax3^{Cre/+};Rosa26^{tdTomato}* (n = 3) or (**g**) *Myf5^{Cre/+};Rosa26^{tdTomato}* (n = 4) embryos

at E10.5. (**h**) Quantification of percentage tdTomato labelling of PROX1⁺ ECs present inside or outside of the venous endothelium in *Pax3^{Cre/+};Rosa26^{tdTomato}* or *Myf5^{Cre/+};Rosa26^{tdTomato}* embryos at E10.5. (**i-i''**) Representative immunofluorescence image of tdTomato, PECAM1 and PROX1 on a sagittal vibratome section from a *Pax3^{CreERT2/+};Rosa26^{tdTomato}* embryo at E13 following tamoxifen administration at E9 (n = 4). (**j-j'**) Representative immunofluorescence image of tdTomato, NRP2 and PROX1 on a whole mount heart from a *Pax3^{CreERT2/+};Rosa26^{tdTomato}* embryo at E16.0 following tamoxifen administration at E9 (n = 4). (**k-k''**) Representative immunofluorescence image of tdTomato, PECAM1 and PROX1 on a sagittal vibratome section from a *Pax3^{Cre/+};Rosa26^{tdTomato}* embryo (n = 4). (**l-l'**) Representative immunofluorescence image of tdTomato, NRP2 and PROX1 on a whole mount heart from a *Pax3^{Cre/+};Rosa26^{tdTomato}* embryo at E16.0 (n = 4). (DM, dermomyotome; CV, cardinal vein; DA, dorsal aorta; pTD, primordial thoracic duct; LS, lymph sac; Scale bars - 50 μ m (**a-a'**, **f, g**), 100 μ m (**c-c'**, **i-i''**), 200 μ m (**j-k'**), 500 μ m (**i**); Data are presented as mean \pm s.d. (error bars)).



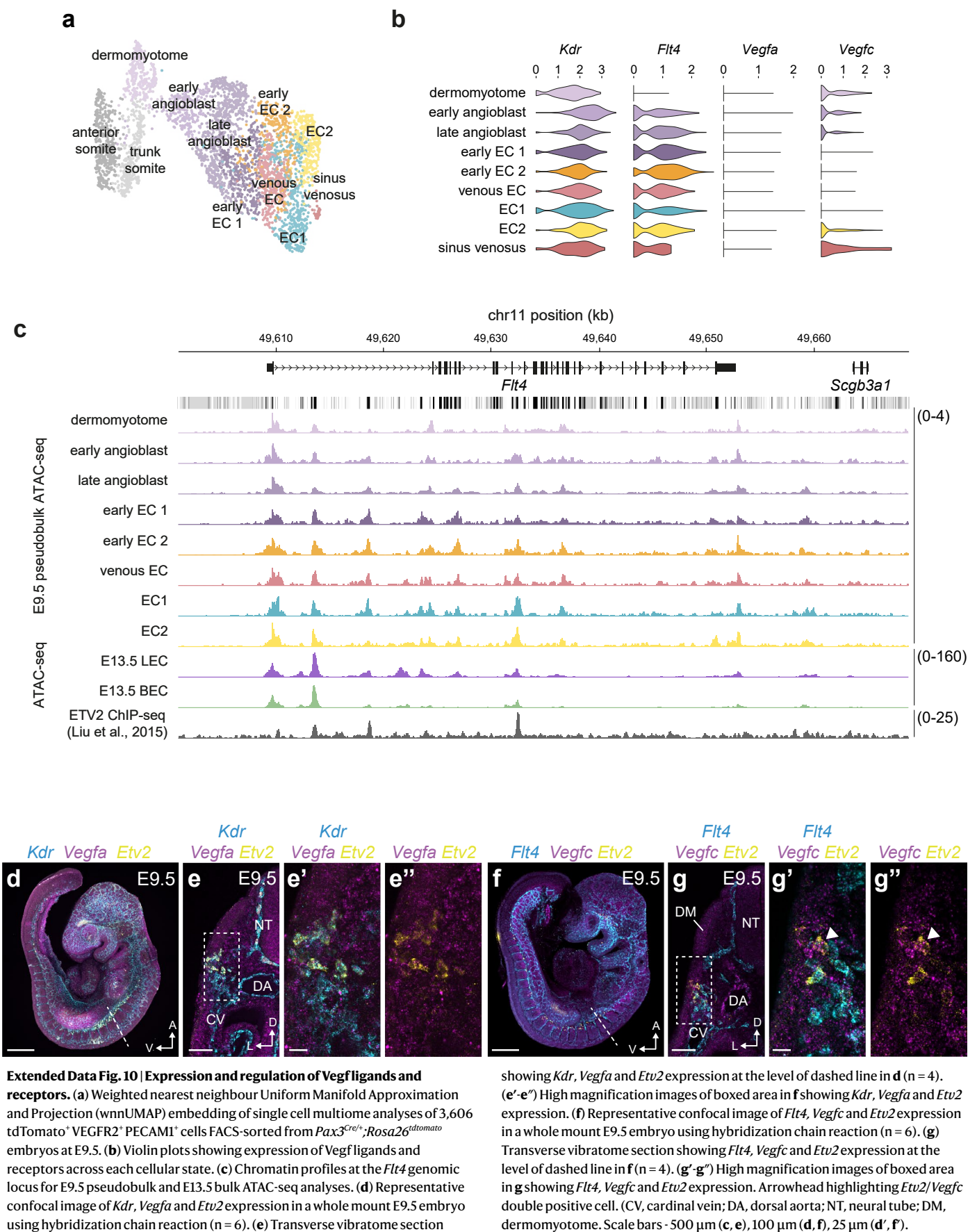
Extended Data Fig. 8 | Reassessment of the venous origin of lymphatic endothelium. (a) Violin plots showing expression of *Tek* in angioblasts at E9.5. (b) Immunofluorescence for tdRFP and PROX1 on a whole mount *Tg(Tek-Cre)^{5326Sato};Rosa26^{tdRFP}* embryo at E9.5 (n = 3). Representative immunofluorescence images of tdRFP, PECAM1 and PROX1 on transverse vibratome sections from *Tg(Tek-Cre)^{5326Sato};Rosa26^{tdRFP}* embryos at (c-c') E10.5, (d) E11.5 and (e-e') E12.5 (n = 3). (f) Quantification of the percentage of lineage traced PROX1⁺ ECs in *Tg(Tek-Cre)^{5326Sato};Rosa26^{tdRFP}* embryos at E9.5-E12.5

(n = 3 per stage). (g-g') Representative immunofluorescence image of tdTomato, PECAM1 and PROX1 from a transverse vibratome section of a *Tg(Tek-Cre)^{12FlvJ};Rosa26^{tdRFP}* embryo at E10.5 (n = 3). Immunofluorescence for EMCN, LYVE1 and PROX1 on transverse vibratome sections from (h-h') *Tg(Tek-Cre)^{12FlvJ};Prox1^{+/+}* and (i-i') *Tg(Tek-Cre)^{12FlvJ};Prox1^{fl/fl}* embryos at E13.5. (CV, cardinal vein; DA, dorsal aorta; pTD, primordial thoracic duct; LS, lymph sac; Scale bars - 25 μm (c-c'), 50 μm (d, g-g'), 200 μm (e, h, i), 100 μm (h', i'); Data are presented as mean ± s.d. (error bars)).



Extended Data Fig. 9 | Analysis of LEC proliferation. (a) ForceAtlas2 (FA) embedding of 19,699 cells based on partition-based graph abstraction (PAGA), with each dot representing a single cell. **(b)** FA embedding showing cell cycle phase in each cellular state. **(c)** Expression of selected cell cycle-related genes in single cells in angioblast, venous, LEC progenitor and lymphatic states. **(d)**

Dual-pulse labelling strategy for analysis of cell proliferation. **(e)** Representative immunofluorescence image of ERG and Ki67 expression in a transverse vibratome section from an E10.5 embryo (n = 3). **(f)** Quantification of growth fraction for ECs present inside the CV or DA. (CV, cardinal vein; DA, dorsal aorta; Scale bars - 25 μm (d); Data are presented as mean ± s.d. (error bars)).



Reporting Summary

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

Statistics

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

n/a | Confirmed

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

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

Software and code

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

Data collection

Confocal laser scanning microscopy:

Imaging of immunostained tissues was performed using Zeiss LSM780, LSM880 or LSM980 confocal microscopes. Datasets were recorded and processed with ZEN Pro (Carl Zeiss). All confocal images represent maximum intensity projections of z-stacks of either single tile or multiple tile scans. Mosaic tile-scans with 10% overlap between neighbouring z-stacks were stitched in ZEN software. Confocal single and multi-tile-scans were processed in Fiji. If necessary, adjustments to brightness, contrast and intensity were made uniformly across individual channels and datasets.

Directional migration of LEC progenitors:

To assess directional migration of PROX1+ LEC progenitors, we established an image analysis pipeline to automatically define and quantify cell polarity in large tissue sections. Vibratome sections were imaged on a Zeiss LSM 880 confocal microscope and in a first post-processing step, Golgi and nuclei in each image stack were segmented using the "Surface" function in Imaris v9.5 (Bitplane; RRID:SCR_007370).

Flow Cytometry and FACS: Flow cytometry was performed on a BD LSRFortessa X20 cytometer and FACS on a BD Aria III. Datasets were recorded and processed using BD FACS Diva software v8.0.

Data analysis

Single cell genomics data analysis:

All scripts required to reproduce the data presented are available on GitHub at https://github.com/StoneLabGH/LEC_Specification.

Data analysis softwares:

GraphPad Prism 10 was used for statistical analyses and presenting data. FlowJo v9 was used to analysis flow cytometry data. Confocal images were processed in Fiji v2.1.0 and Imaris v9.5 (Bitplane). The following softwares were used for single cell RNA sequencing analyses: Cell

Ranger 5.0 (10x Genomics), RStudio v1.4, Seurat 4.0, Python v3.11 (Jupyter Notebook interface), Scanpy 1.8, scanpy.tl.paga. The following softwares were used for single cell multiome analyses: 10x Cell Ranger Arc pipeline (v2.0.0), RStudio v2022.02.3, R v4.2.0, Seurat v5.0.1, Signac v1.9.0, SCENIC+ v1.0a1, pycisTopic v2.0a0, MACS2 (v.2.2.9.1), IGV(v2.13.1), TOBIAS (v.0.17.0), sinto (<https://github.com/timoast/sinto>), Genrich (<https://github.com/jsh58/Genrich>). For bulk ATAC-seq analyses, the following softwares were used: bowtie2 (v.2.4.4), bamCoverage (v2.1.0)

Quantification of cell numbers:

Confocal and light sheet image stacks were rendered into 3D volumes and analysed using Imaris v9.5 (Bitplane; RRID:SCR_007370). Quantification of absolute cell numbers is based on staining of specific transcription factors to visualize the nuclei of cells of interest. Thus, the number of nuclei reflect the number of cells. Nuclei were automatically annotated using the “Spots” function, which automatically detects point-like structures with a predefined diameter. Accurate quantification required an appropriate estimate of cell nuclei diameter and filtration of selected nuclei by tuning the quality parameters. The accuracy of this automatic counting procedure was verified by visual inspection, which herein served as ground truth. Using the “manual Surface creation” function, vascular structures were segmented based on specific EC marker expression. Thereby cell populations inside and outside of segmented vascular structures were defined by filtering the shortest distance between “Spots” and “Surface”.

Directional migration of LEC progenitors:

Cell populations inside and outside of defined vascular structures were defined as described above (Quantification of cell numbers). Surface masks were exported and processed in Fiji v1.53. Subsequently, nearest neighbour analyses were used to pair individual nuclei with their corresponding Golgi (closest border – border distance), and the centroid of each object was computed using the 3D ImageJ Suite v4.0.36 74. 2D vectorization images were obtained by drawing arrows from the nuclei centroid towards the Golgi centroid using Fiji. Nuclei – Golgi pairs with a border-to-border distance larger than 5 μm were excluded from further analysis. Centroid vectors were produced using the XY-coordinates of nuclei and Golgi centroids and transformed to unit vectors (A). The dorsal body axis served as reference vector (B). The angle was obtained by calculating the inverse cosine of the dot product of centroid unit vectors (A) and reference vector (B) ($\theta = \arccos(A \cdot B)$). All calculations were performed using python v3.8. Angles were transformed to represent the body axes (0°, dorsal; 90° lateral; 180°, ventral; 270°, medial) and a histogram on a polar axis was used to display the angular distribution of individual LECs representing their migration direction.

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

Data

Policy information about [availability of data](#)

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Data availability

scRNA-seq (GSE276281) and scMultiome-seq (GSE276282) data were deposited in the Gene Expression Omnibus (GEO). Correspondence and requests for materials should be addressed to F.K. and O.A.S.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	<input type="text" value="n/a"/>
Reporting on race, ethnicity, or other socially relevant groupings	<input type="text" value="n/a"/>
Population characteristics	<input type="text" value="n/a"/>
Recruitment	<input type="text" value="n/a"/>
Ethics oversight	<input type="text" value="n/a"/>

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

Field-specific reporting

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

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

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

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

Sample size	We used the NC3Rs experimental design assistant (EDA) to form an experimental plan that integrates statistics-based power calculations informed by our knowledge of estimated variability.
Data exclusions	No data were excluded
Replication	Experiments were performed following careful staging of embryos, and all presented analyses used the offspring of multiple pregnant dams. Analyses were reproduced through multiple independent experiments to confirm the obtained results. Specific numbers of replication for each experiments are included in the figures and figure legends.
Randomization	Due to the nature of most of the experiments performed here, allocation of embryos to distinct treatment groups is randomized. For example, Vegfc +/- and Vegfc -/- embryos were identified from a pool of littermates though genotyping following imaging analyses.
Blinding	For imaging and flow cytometry analyses, investigators were blinded to genotype and treatments until completion of data. collection/ quantification.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Included in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

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

Antibodies

Antibodies used

The following antibodies were used for immunofluorescence staining of cryosections, vibratome sections and/or whole mount tissues: ETV2 (Abcam, #ab181847, 1:100), VEGFR2 (BD Pharmingen, #550549, 1:200), PECAM1 (R&D Systems, #AF3628, 1:250), PECAM1 (D. Vestweber, Clone 5D2.6 and 1G5.1, 15 µg / mL), PROX1 (R&D Systems, #AF2727, 1:200), PROX1 (Proteintech, #11067-2-AP, 1:100), PROX1 (Reliatech, #102-PA32, 1:100), PROX1 (Abcam, #ab225414, 1:100), EMCN (Santa Cruz Biotechnology, SC-65495, 1:50), EMCN (D. Westweber, #VE44, 1:100), EMCN (Santa Cruz Biotechnology, SC-53941, 1:50), VWF (Abcam, #11713, 1:100), ESR1 (Abcam, #ab16660, 1:100), NRP2 (R&D Systems, #AF567, 1:250), ERG (Abcam, #ab92513, 1:200), GM130 (BD Pharmingen, #610823, 1:100), RFP (Rockland, #600-401-379, 1:500), BrdU (Abcam, #ab6326, 1:100), KI67 (Thermo Fisher, #14-5698-82, 1:100) and GFP (Thermo Fisher, #A-21311, 1:100), Donkey anti-Rat IgG Alexa Fluor Plus 405 (#A48268), Donkey anti-Rabbit IgG Alexa Fluor 488 (#A21206), Donkey anti-Goat IgG Alexa Fluor Plus 488 (#A32814), Donkey anti-Sheep IgG Alexa Fluor 488 (#A11015), Donkey anti-Rabbit IgG Alexa Fluor 555 (#A32814), Donkey anti-Rat IgG Alexa Fluor 555 (#A48270), Donkey anti-Goat IgG Alexa Fluor 647 (#A32849), Donkey anti-Rabbit IgG Alexa Fluor Plus 647 (#A32795), and Donkey anti-Sheep IgG Alexa Fluor 647 (#A21448).

Validation

Antibodies were purchased from commercial vendors and validated by the manufacturer, and/or by previous publications:

- ETV2 (Abcam, #ab181847) : Immunoblot and immunocytochemistry (manufacturer), PMID: 32908310, 29669933, 33522307.
- VEGFR2 (BD Pharmingen, #550549) : detection by immunohistochemistry, immunoblot and immunocytochemistry (manufacturer), PMID: 31943281, 31665628, 31118412.
- PECAM1 (R&D Systems, #AF3628) : detection by immunohistochemistry, immunoblot and immunocytochemistry (manufacturer), PMID: 34259830, 34450028, 37587341.
- PECAM1 (D. Vestweber, Clone 5D2.6) : PMID : 16818677
- PROX1 (R&D Systems, #AF2727) : detection by immunohistochemistry, immunoblot and immunocytochemistry (manufacturer), PMID: 35794479, 25992544, 32251437.
- PROX1 (Proteintech, #11067-2-AP) : detection by immunohistochemistry, immunoblot and immunocytochemistry (manufacturer), PMID: 31341278, 28024299, 34038712
- PROX1 (Reliatech, #102-PA32) : detection by immunoblot and immunofluorescence (manufacturer), PMID:31980640, 38177500, 25061877
- PROX1 (Abcam, #ab199359) : detection by immunohistochemistry, immunoblot and immunocytochemistry (manufacturer), PMID: 37576598, 37699906, 36994549
- EMCN (Santa Cruz Biotechnology, SC-65495) : detection by immunohistochemistry and immunoblot (manufacturer), PMID: 18924607, 24647000, 21909098

- EMCN (D. Westweber, #VE44) PMID: 11554756
 - EMCN (Santa Cruz Biotechnology, SC-53941) : detection by immunocytochemistry (manufacturer), PMID: 29773646, 25992544, 31701448.
 - VWF (Abcam, #ab11713) : detection by immunocytochemistry (manufacturer), PMID: 37834029, 35325071, 35001873.
 - ESR1 (Abcam, #ab16660) : detection by immunocytochemistry (manufacturer), PMID: 33999917, 36729672, 33623049
 - NRP2 (R&D Systems, #AF567) : detection by immunoblot, immunohistochemistry, flow cytometry, and immunocytochemistry (manufacturer), PMID: 1130354, 33931446, 30006544
 - ERG (Abcam, #ab92513) : detection by immunohistochemistry, flow cytometry, and immunocytochemistry (manufacturer), PMID: 30301887, 30178747, 29400648
 - GM130 (BD Pharmingen, #610823) : detection by immunocytochemistry and immunofluorescence (manufacturer), PMID:39198435, 38969763, 38874642.
 - RFP (Rockland, #600-401-379) : detection by immunohistochemistry and immunofluorescence (manufacturer), PMID: 38238288, 38579720, 38326622.
 - BrdU (Abcam, #ab6326) : detection by immunocytochemistry and immunofluorescence (manufacturer), PMID: 38123565, 37874652, 38036565.
 - Ki67 (Invitrogen 14-5698-82 rat monoclonal clone SolA15) - detection by immunohistochemistry and immunocytochemistry (manufacturer), PMID: 32750316, 30389919, 27699223
 - GFP (Thermo Fisher, #A-21311): detection by immunohistochemistry and immunocytochemistry (manufacturer), PMID: 31797533, 28442548, 28233777

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals

This study was performed using mouse embryos and the following lines were used: Pax3Cre (Pax3tm1(cre)Joe); Pax3CreERT2 (Pax3tm1.1(cre/ERT2)Lepr); Lbx1Cre (Lbx1tm3.1(cre)Cbm); Myf5Cre (Myf5tm3(cre)Sor); Tie2-Cre (Tg(Tek-cre)12Flv/J); Tie2-Cre (Tg(Tek-cre)5326Sato); Prox1fl (Prox1tm1a(EUCOMM)Wtsi); Rosa26tdTomato (Gt(ROSA)26Sortm9(CAG-tdTomato)Hze); Rosa26tdTomato (Gt(ROSA)26Sortm14(CAG-tdTomato)Hze); Rosa26tdRFP (Gt(ROSA)26Sortm1Hjf); Rosa26Fucci2 (Tg(Gt(ROSA)26Sor-Fucci2)Sia), VegfcLacZ (Vegfctm1Ali). With the exception of the VegfcLacZ (Vegfctm1Ali) line, which was maintained on a CD1 background, all lines were maintained on a C57Bl6/J background. Breeding pairs between 2 and 8 months were used to generate animals of the indicated genotypes. Both male and female embryos were used for analyses. Mice were maintained in IVC-cages and ventilated racks at 22°C and 55% humidity with a 12/12 hour light/dark cycle. For embryo collection, mice were paired overnight and females were checked the next morning for the presence of a vaginal plug. For inducible Cre induction, pregnant females were gavaged at the specified time points with 80mg/kg tamoxifen (Sigma, #T5648) dissolved in peanut oil with 10% ethanol at a final concentration of 10mg/ml.

Wild animals

No wild animals were used in this study.

Reporting on sex

As this study involved the analysis of blood and lymphatic vessel development in mouse embryos, we did not record or compare sex dependent differences.

Field-collected samples

No field collected samples were used in this study.

Ethics oversight

All procedures were carried out in accordance with local legislation: University of Oxford Animal Welfare and Ethical Review Boards in accordance with Animals (Scientific Procedures) Act 1986 under Home Office project licences PPL PC013B246 or PP6588077; German animal protection legislation (Tierschutzgesetz und Tierschutz-versuchstierverordnung); Uppsala Animal Experiment Ethics Board (permit number 130/15).

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

Plants

Seed stocks

n/a

Novel plant genotypes

n/a

Authentication

n/a

Plots

Confirm that:

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

Methodology

Sample preparation

Single cell suspensions, generated as above, were incubated with Zombie Aqua™ Fixable Viability Kit (Biolegend, #423101, 1:1000) for 15 min at room temperature. Cells were washed with Cell Staining Buffer (Biolegend, #420201), then washed, centrifuged and 500xg for 5 min and blocked with Fc block CD16/32 (Biolegend, #101302, 1:100), for 5-minutes on ice and stained with PECAM1-BV605 (Biolegend, #102427, 1:1000, 100), CD45-FITC (Biolegend, #157607, 1:200), CD41-BV421 (Biolegend, #133911, 1:200), PDPN-eF660 (eBioscience, #50-5381-82, 1:100), LYVE1-PECy7 (eBioscience, #25-0443-82, 1:400) for 30-minutes on ice, then washed and resuspended in Cell Staining Buffer.

Instrument

Samples were either analysed immediately on a BD LSRFortessa X20 cytometer or stored in IC Fixation Buffer (eBioscience, #00-8222-49), washed and analysed the next day.

Software

Data were collected using FACSDiva (v8.0) software and analysed with FlowJo (v9) software.

Cell population abundance

FACS sorted cells were subjected to scRNA-seq or scMultiome-seq, demonstrating the degree of purity.

Gating strategy

Gating strategies are outlined in extended data figure 1. The initial SSC/FSC gating was broad and fluorescence minus one (FMO) controls were used for all antibodies/fluorophores.

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