












# Single-cell analysis of lymphatic endothelial cell fate specification and differentiation during zebrafish development

Lin Grimm<sup>1,2,3,†</sup> , Elizabeth Mason<sup>1,2,\*,†</sup> , Hujun Yu<sup>1,2</sup>, Stefanie Dudczig<sup>1,2</sup> , Virginia Panara<sup>4</sup> , Tyrone Chen<sup>1,2</sup> , Neil I Bower<sup>3,‡</sup>, Scott Paterson<sup>1,2,3</sup>, Maria Rondon Galeano<sup>1,2</sup>, Sakurako Kobayashi<sup>1,2</sup> , Anne Senabouth<sup>3,5</sup> , Anne K Lagendijk<sup>3</sup> , Joseph Powell<sup>3,5,6,7</sup>, Kelly A Smith<sup>8</sup>, Kazuhide S Okuda<sup>1,2</sup> , Katarzyna Koltowska<sup>4</sup>  & Benjamin M Hogan<sup>1,2,3,8,\*\*</sup> 

## Abstract

During development, the lymphatic vasculature forms as a second network derived chiefly from blood vessels. The transdifferentiation of embryonic venous endothelial cells (VECs) into lymphatic endothelial cells (LECs) is a key step in this process. Specification, differentiation and maintenance of LEC fate are all driven by the transcription factor *Prox1*, yet the downstream mechanisms remain to be elucidated. We here present a single-cell transcriptomic atlas of lymphangiogenesis in zebrafish, revealing new markers and hallmarks of LEC differentiation over four developmental stages. We further profile single-cell transcriptomic and chromatin accessibility changes in zygotic *prox1a* mutants that are undergoing a LEC-VEC fate shift. Using maternal and zygotic *prox1a/prox1b* mutants, we determine the earliest transcriptomic changes directed by *Prox1* during LEC specification. This work altogether reveals new downstream targets and regulatory regions of the genome controlled by *Prox1* and presents evidence that *Prox1* specifies LEC fate primarily by limiting blood vascular and haematopoietic fate. This extensive single-cell resource provides new mechanistic insights into the enigmatic role of *Prox1* and the control of LEC differentiation in development.

**Keywords** lymphangiogenesis; lymphatics; Notch1; *Prox1*; *Vegfc* single-cell sequencing

**Subject Categories** Development; Methods & Resources; Vascular Biology & Angiogenesis

DOI 10.15252/embj.2022112590 | Received 13 September 2022 | Revised 24 January 2023 | Accepted 3 February 2023

The EMBO Journal (2023) e112590

## Introduction

Lymphatic vasculature plays crucial physiological roles that include the drainage of interstitial fluids, trafficking of immune cells and drainage of dietary lipids. The formation of new lymphatic vessels from pre-existing vessels (lymphangiogenesis) occurs in both development and disease. Signalling through *Vegfr3* (*Flt4*) can be triggered by *Vegfc* or *Vegfd* and drives lymphangiogenesis in settings as diverse as development, cancer metastasis, inflammation and cardiac repair (Oliver *et al*, 2020). In the embryo, lymphangiogenesis begins when the first LEC progenitors depart the cardinal veins (CVs) from E9.5 in mice and 32-h post-fertilisation (hpf) in zebrafish (Koltowska *et al*, 2013). *Vegfc-Flt4* signalling drives LEC progenitor sprouting but also upregulates *Prox1* expression in both zebrafish and mice (Deng *et al*, 2013; Srinivasan *et al*, 2014; Koltowska *et al*, 2015a; Shin *et al*, 2016; Baek *et al*, 2019). The transcription factor (TF) *Prox1* acts as the master regulator of LEC fate (Oliver & Srinivasan, 2010) and is exclusively expressed in developing LECs in early embryonic vasculature. Loss of *Prox1* (*Prox1a* and *Prox1b* in zebrafish) leads to a loss of developing lymphatic vessels (Wigle & Oliver, 1999; Koltowska *et al*, 2015a). Following departure from the CV, LEC progenitors go on to colonise embryonic tissues and organs and remodel to form functional lymphatic vessels (Hagerling *et al*,

1 Organogenesis and Cancer Program, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia

2 Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, VIC, Australia

3 Division of Genomics of Development and Disease, Institute for Molecular Bioscience, The University of Queensland, St Lucia, Brisbane, QLD, Australia

4 Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden

5 Garvan Institute of Medical Research, Sydney, NSW, Australia

6 School of Medical Sciences, University of New South Wales, Kensington, Sydney, NSW, Australia

7 Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia

8 Department of Anatomy and Physiology, University of Melbourne, Melbourne, VIC, Australia

\*Corresponding author. Tel: +61 3 8559 5000; E-mail: [elizabeth.mason@petermac.org](mailto:elizabeth.mason@petermac.org)

\*\*Corresponding author. Tel: +61 3 8559 5000; E-mail: [ben.hogan@petermac.org](mailto:ben.hogan@petermac.org)

†These authors contributed equally to this work

‡Deceased

2013). While at later stages in mammals there are contributions of LECs from non-venous origins, early embryonic lymphangiogenesis occurs chiefly from the CVs (Srinivasan *et al*, 2007, 2010; Yang *et al*, 2012; Klotz *et al*, 2015; Martinez-Corral *et al*, 2015; Stanczuk *et al*, 2015; Stone & Stainier, 2019). In mice, both the earliest stages of LEC progenitor sprouting from CVs and maintenance of LEC fate in stable lymphatics are dependent on the function of Prox1 (Wigle & Oliver, 1999; Johnson *et al*, 2008).

Despite over two decades of study of this enigmatic developmental process, the transcriptomic changes that occur as embryonic venous endothelial cells (VECs) transdifferentiate into LEC progenitors and further differentiate into mature lymphatics, have not been transcriptionally profiled *in vivo*. In the absence of Prox1 in conditional knockout mice, LECs have been shown to lose the expression of some LEC markers and to gain expression of some blood vascular endothelial cell (BEC) markers (Johnson *et al*, 2008). Prox1 is known to auto-regulate its own expression and to also regulate Flt4 expression in a positive regulatory loop during early development (Johnson *et al*, 2008; Srinivasan *et al*, 2014). Yet, how Prox1 controls the transcriptome during LEC specification, differentiation and maintenance has not been described in detail. This is in part because of the technical challenge of accessing early mouse endothelial cells (ECs) in wild-type and mutant embryos, a problem that is not limiting when using the zebrafish embryo.

As recent studies have demonstrated highly conserved expression and function of Prox1 homologues in zebrafish (Tao *et al*, 2011; van Impel *et al*, 2014; Koltowska *et al*, 2015a; Shin *et al*, 2016), we here took advantage of the accessibility of the zebrafish embryo to examine developmental lymphangiogenesis using single-cell transcriptomics. We provide a resource of new markers of VECs and LECs and reveal the timing of differentiation *in vivo*. We analysed zebrafish zygotic *prox1a* mutants with single-cell RNA sequencing and single-cell ATAC sequencing. This identified a LEC-VEC fate reversion in the absence of zygotic Prox1a, defined key Prox1-dependent genes in fate maintenance and discovered regulatory regions of chromatin (enhancers) controlled by Prox1. Profiling maternal-zygotic double *prox1a/prox1b* (null) mutant vasculature unexpectedly revealed that Prox1's earliest function is mostly to negatively regulate expression of a network of conserved haematopoietic and blood vascular fate regulators, rather than positively driving LEC fate. Finally, our single-cell resource provides new mechanistic insights, such as expression of Notch signalling

pathway components in LECs following specification and migration, and essential for LEC development in the embryo. Overall, this single-cell resource reveals new markers and mechanisms of embryonic lymphangiogenesis and Prox1 control of LEC fate.

## Results

### A single-cell RNA sequencing atlas of embryonic lymphangiogenesis

Zebrafish secondary angiogenesis occurs when Prox1-positive LECs and VECs both sprout from the cardinal vein (CV) in the trunk and the head in a progressive process between ~32 and ~48 hpf (Koltowska *et al*, 2015a; Shin *et al*, 2016). In the trunk, sprouting LECs migrate dorsally and invest the horizontal myoseptum, where they form a transient pool of parachordal LECs (PLs) from approximately 48 hpf (Yaniv *et al*, 2006; Hogan *et al*, 2009). In craniofacial regions of the embryo, LECs sprout from the CVs at several locations (Okuda *et al*, 2012; Eng *et al*, 2019). After this, LECs throughout the embryo proliferate and migrate extensively (between ~56 hpf and 80 hpf) to colonise new regions and tissues (Busmann *et al*, 2010; Cha *et al*, 2012; Okuda *et al*, 2012; Jung *et al*, 2017). In the trunk, LECs anastomose to form the first lymphatic vessels at around 4-day post-fertilisation (dpf), forming the thoracic duct (TD), dorsal longitudinal lymphatic vessels (DLLVs) and intersegmental lymphatic vessels (ISLVs). In cranial regions, they assemble from disparate sources into lateral (LFL), medial (MFL), otolithic lymphatic vessels (OLV) and lymphatic branchial arches (LBA), as well as forming a lymphatic loop (LL) in the head that will later give rise to a unique cranial mural LEC population (muLECs, also known as FGPs or brain LECs) (van Lessen *et al*, 2017; Venero Galanternik *et al*, 2017; Bower *et al*, 2017a). By 5 dpf, major lymphatics in the craniofacial and trunk regions of the embryo are functional and drain dyes and fluids deposited in the peripheral tissues (Kuchler *et al*, 2006; Yaniv *et al*, 2006).

To profile stages of development spanning key steps in lymphatic differentiation, we selected: 40 hpf, when specification and initial sprouting of LECs are ongoing; 3 dpf, when immature LECs are migrating through the embryo; 4 dpf, when LECs are assembling into vessels; and 5 dpf, when lymphatics are functional and maturing (Hogan & Schulte-Merker, 2017; Fig 1A). We used transgenic

#### Figure 1. Single-cell RNA-seq developmental atlas of lymphangiogenesis in zebrafish.

- Schematic representation of four key stages of zebrafish lymphatic development in head and trunk: 40 hpf encompasses both specification (Prox1-induction) and sprouting; 3 dpf migration of LECs; 4 dpf assembly of lymphatic vessels; 5 dpf maturation of functional lymphatics.
- UMAP visualisation of  $n = 9,771$  cells filtered for VEC and LEC populations ( $n = 6$  samples; see Dataset EV1A for number of cells per cluster, cluster identification and Fig EV1 for whole dataset) coloured according to developmental stage (top), predicted cell phenotype (middle) and RNA velocity (bottom).
- UMAP visualisation of key marker gene expression. Colour scale represents log-normalised expression. LEC markers: *prox1a*, *prox1b* (*prox3*), *cdh6*. BEC markers: *cdh5*. LEC and VEC markers: *mrc1a*, *flt4*. AEC marker: *flt1*. muLEC markers: *osr2*.
- Heatmap of top genes commonly up-regulated in LECs at 3, 4 and 5 dpf, with expression displayed in VECs (all stages), VEC\_preLEC (40 hpf) and LECs at 3, 4 and 5 dpf (see also Dataset EV1B, Fig EV1F). Colour scale represents average log-normalised expression, x-axis columns group cells by phenotype and developmental stage, and y-axis rows indicate the genes.
- Bar plot summarising GO BP analyses of genes commonly DE between 3, 4 and 5 dpf LEC and VEC populations (Dataset EV1C and D). Y-axis represents enriched BP term, x-axis represents the  $-\log_{10}$  (adjusted  $P$ -value), and bars are coloured and ordered according to fold enrichment of the GO term in LEC. GO terms enriched in LEC are coloured purple, and terms enriched in VEC are coloured red.

Data information: Hpf, hours post-fertilisation. Dpf, days post fertilisation. BEC, blood vascular endothelial cell. VEC, venous endothelial cell. LEC, lymphatic endothelial cell. muLEC, mural lymphatic endothelial cell (a.k.a FGP or brain LEC).

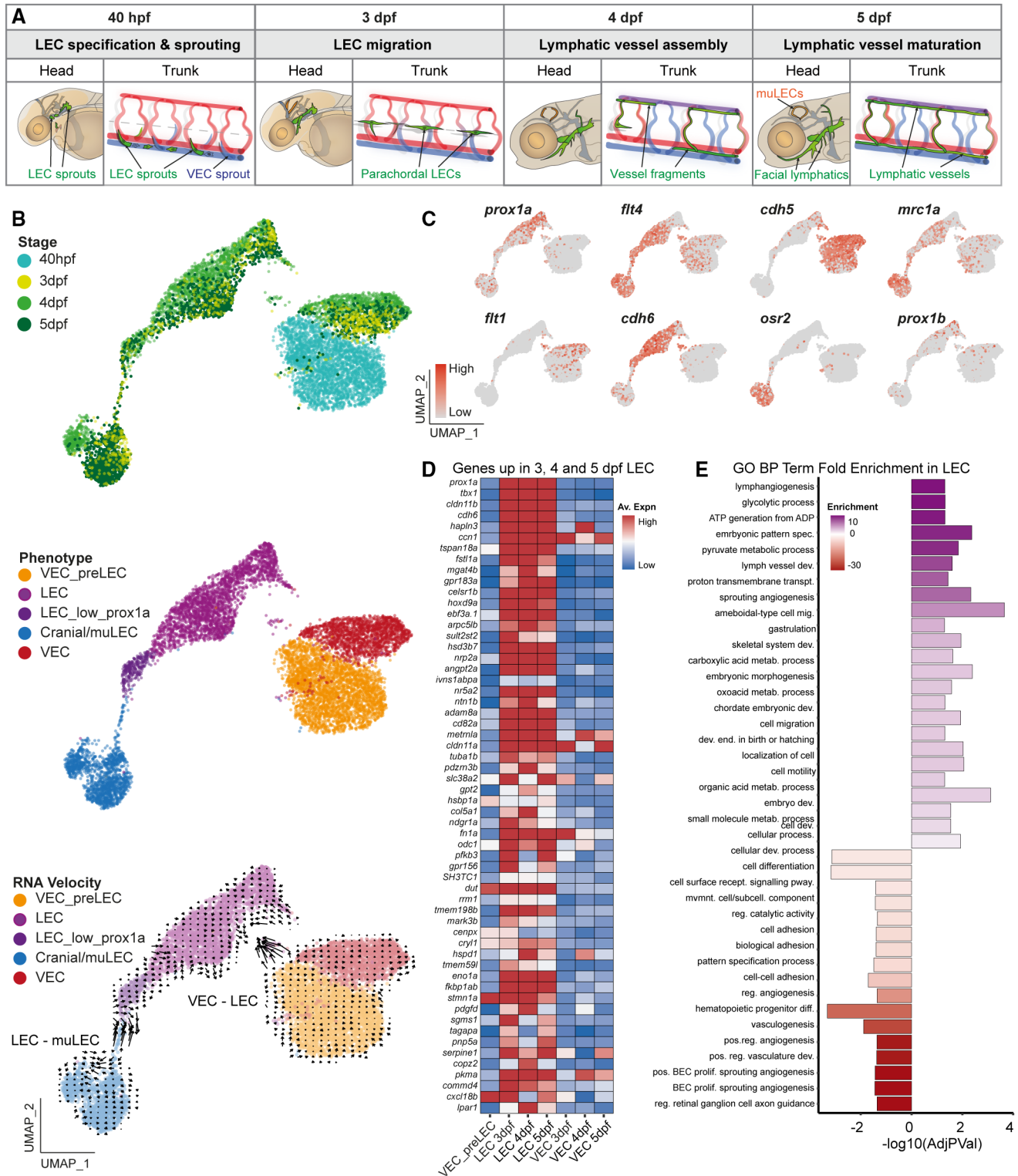


Figure 1.

zebrafish strains that label embryonic vasculature to allow for fluorescence-activated cell sorting (FACS, see methods) and single-cell RNA sequencing on the 10X Chromium platform. We sequenced 35,634 cells across three runs, then merged and normalised the data

(Stuart *et al*, 2019; He *et al*, 2020), filtering low-quality libraries (Fig EV1A–C, full details in methods). To define the cellular identity of each cluster, we systematically evaluated the expression of known markers summarised in Dataset EV1A. We identified 9,771 lymphatic

and venous ECs that comprise our single-cell atlas of lymphangiogenesis, which is accessible in an interactive CellXGene explorer web app at the link <https://115.146.95.206:5009/> (Figs 1B and EV1D–G).

This dataset is displayed (Becht *et al*, 2018) as a UMAP in Fig 1B coloured according to developmental stage and cell phenotype, with venous (VEC) to lymphatic (LEC) trajectory of differentiation confirmed by RNA velocity analysis (La Manno *et al*, 2018). Comparison of cells at 40 hpf with later stages revealed that the earlier populations are transcriptionally distinct from VECs and LECs at 3, 4 and 5 dpf, revealing that differentiation of LECs from VECs occurs between 40 hpf and 3 dpf. Notably, the 40 hpf “VEC\_preLEC” cluster contained both *prox1a*<sup>+</sup> and *prox1a*<sup>-</sup> cells. This indicates that LECs are not sufficiently transcriptionally distinct from VECs at 40 hpf to be assigned as a unique cluster, further suggesting that while early *prox1a*<sup>+</sup> cells are specified they have not yet differentiated (Fig 1C; Koltowska *et al*, 2015a). We identified three main classes of LECs at 3, 4 and 5 dpf: cranial (e.g. future muLECs; 1,511 cells) marked by expression of *osr2* (Bower *et al*, 2017a), canonical LECs marked by expression of *cdh6* (2,669 cells) and a smaller subpopulation of LECs expressing low levels of *prox1a* (LEC\_low\_*prox1a*, 310 cells) likely to be transitioning from canonical to cranial populations (Figs 1C and EV1D–G). To define markers of differentiating canonical LECs at each developmental stage, we applied differential expression (DE) analysis (Dataset EV1B). This analysis not only captured known LEC markers including *prox1a*, *angpt2a*, *tbx1* and *cldn11b* but also uncovered new genes commonly expressed in canonical LECs across all developmental stages including *gpr156*, *hapln3*, *cdh6* and *tspan18a* (Figs 1D and EV1H). We extended this approach and evaluated global differences between all canonical LECs and all VECs ( $n = 1,240$  genes, Dataset EV1C). GO analysis (Thomas *et al*, 2006) confirmed the association of biological processes known to be associated with lymphatics with genes upregulated in LECs ( $n = 752$  LEC gene set), and these terms included “lymphangiogenesis,” “glycolytic process,” “lymph vessel development” and “ameboidal-type cell migration” (Fig 1E, Dataset EV1D).

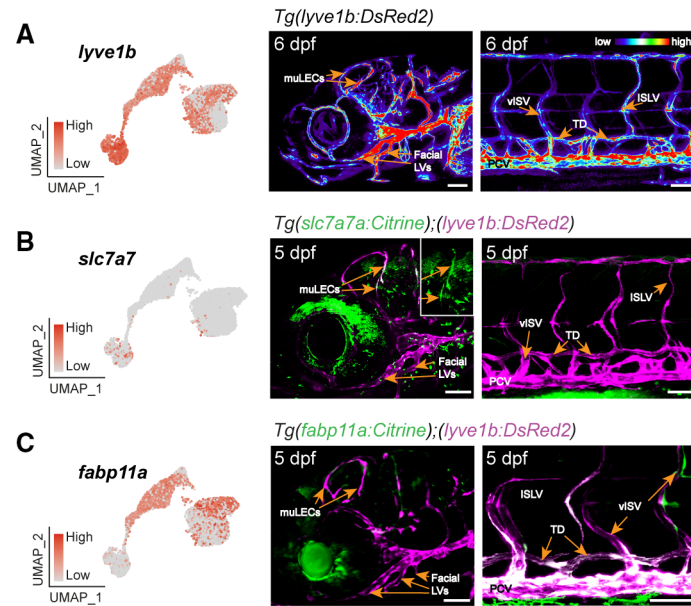
To confirm the specificity of the new LEC markers defined by this scRNA-seq resource, we used DE analysis to define cluster-specific gene expression. We noted that this analysis suggested higher expression of the well-known marker *lyve1b* in cranial than in canonical LECs, which we validated by examining expression levels of *lyve1b* using the established transgenic line *Tg (lyve1b:DsRed2;* heat map in Fig 2A). We further identified expression of the known kidney epithelial solute transporter *slc7a7* as uniquely expressed in the cranial populations and *fabp11a* as a LEC and VEC marker excluded from cranial populations. We generated both new *slc7a7a*-Citrine and *fabp11a*-Citrine BAC transgenic strains, which confirmed these expression patterns *in vivo* and further validated our single cell (sc) RNASeq atlas (Fig 2B and C). Overall, this atlas identifies  $n = 752$  specific markers of lymphangiogenesis, the majority of which are new markers (Dataset EV1C), representing new candidate regulators and spanning four developmental stages.

### Prox1 maintains LEC identity by repressing blood vascular fate and promoting lymphatic vascular fate at the level of the transcriptome

Prox1 is necessary to drive VEC to LEC transdifferentiation in mammals and loss of Prox1 leads to reduced expression of select LEC

markers and increased expression of some BEC markers (Johnson *et al*, 2008; Oliver & Srinivasan, 2010). Prox1 expression is also necessary for the maintenance of LEC identity during development (Wigle *et al*, 2002; Johnson *et al*, 2008; Kang *et al*, 2010). However, the transcriptomic program controlled by Prox1 *in vivo* has never been profiled and whether Prox1 controls all transcriptional differences found between LECs and VECs remains unclear. Zebrafish *prox1a* zygotic and maternal zygotic mutants have been previously described (van Impel *et al*, 2014; Koltowska *et al*, 2015a). The zygotic mutants retain maternal deposition of *prox1a* in the oocyte, sufficient to drive normal PL formation, LEC migration to the horizontal myoseptum (HM) and initial assembly of lymphatics by 4 dpf (van Impel *et al*, 2014; Koltowska *et al*, 2015a). However, mutants do have a reduction in total LEC numbers throughout the face and trunk (Koltowska *et al*, 2015a). When the maternal contribution of *prox1a* is removed, lymphatic development is more severely impaired with a loss of lymphatics (Koltowska *et al*, 2015a). We hypothesised that the zygotic mutants which form lymphatic vessels likely have abnormal vessel identity in the absence of zygotic Prox1. Thus, we applied scRNA-seq to ECs FAC sorted from *Zprox1a*<sup>-/-</sup> mutants (van Impel *et al*, 2014) and WT sibling zebrafish at 4 dpf ( $n = 8,075$  cells, Fig EV2A–C, Dataset EV2A). Cluster analysis revealed 3 populations of LECs ( $n = 2,068$ ) and a single population of VECs ( $n = 1,051$ ) comprising both mutant and wild-type (WT) cells, and perhaps most striking, a single population of mutant cells ( $n = 484$ ), marked by the expression of *aqp1a.1* (Figs 3A–D and EV2D–F). All three LEC clusters (LEC, LEC\_S1 and LEC\_S2) showed graded expression of lymphatic markers *prox1a*, *cldn11b*, *cdh6* and *angpt2a*, such that “LEC” most closely resembles canonical LECs from our WT atlas. LEC\_S1 and S2 represent alternative LEC subtypes with different proliferative potential based on S-phase occupancy and cell cycle marker *mki67* (Figs 3B and E, and EV2E middle). The expression of *prox1a* was lowest in the more proliferative LEC\_S1 cells and highest in the less proliferative LEC\_S2 (Fig 3B and E). RNA velocity analysis suggested a trajectory between the mutant cluster and the LEC cluster, while other trajectory analysis methods were of limited utility (Figs 3C lower, and EV2G). Taken together, these findings suggest that the mutant-specific cluster sits on a trajectory between LEC and VEC fate, representing cells either failing to fully differentiate from VEC to LEC or LECs undergoing de-differentiation and fate reversion.

To survey how Prox1 maintains normal LEC differentiation, we performed DE analysis comparing the mutant cluster with WT LECs (Fig 3F, Dataset EV2B). Overall, a larger set of genes were downregulated than upregulated in the mutant cluster ( $n = 1,107$  vs.  $n = 326$ ) with almost half of the most downregulated genes (AvLogFC) highly enriched for LEC markers (e.g. *tbx1*, *cdh6* and *cldn11b*). Of the upregulated genes, almost 60% ( $n = 192/326$ ) were VEC markers (e.g. *sox7*, *kdrl* and *cdh5*) and again this was enriched in the most highly upregulated genes (Figs 3F and EV2H). Comparing the change in gene expression between the WT LEC cluster and VEC cluster, with the change between WT LEC and mutant cluster, we found striking concordance, with no VEC or LEC marker genes unchanged in the absence of Prox1 (Fig 3G). This suggests (i) that the mutant cluster is wholly shifted along a LEC to VEC fate trajectory, and (ii) that all of the change that occurs between LEC and VEC is explained by the ongoing function of Prox1. Overall, there is a simultaneous loss of lymphatic fate and reacquisition of blood vascular gene expression, consistent with work in mouse



**Figure 2. Transgenic marker strains confirm cluster identity and identify vessel-specific markers.**

- A UMAP of *lyve1b* expression predicts higher expression in muLECs than LEC populations (left), confirmed by a heat map of a *Tg(lyve1b:DsRed2)* zebrafish larvae at 6 dpf showing high *lyve1b* expression in craniofacial (middle) and moderate expression in trunk lymphatic vessels (right). Full larvae image in Fig EV11.
- B UMAP of *slc7a7* expression predicts restricted expression in muLEC populations (left), confirmed in lateral confocal projections showing co-expression of *lyve1b* (magenta) and a new BAC transgenic strain for *slc7a7a* (green) in the zebrafish head at 5 dpf (middle), which is not expressed in veins and lymphatic vessels of the trunk (right).
- C UMAP of *fabp11a* expression predicts expression in LEC and VEC but not muLEC populations (left), confirmed in lateral confocal projections showing co-expression of *lyve1b* and a new BAC transgenic strain for *fabp11a* marking venous and lymphatic vessels in the trunk (right) without showing expression in the vasculature of the head (middle) at 5 dpf.

Data information: Expression in UMAPs is log normalised. Lateral confocal images, anterior to the left. muLECs, mural lymphatic endothelial cells; facial LV, facial lymphatic vessels; vISV, venous intersegmental vessel; PCV, posterior cardinal vein; TD, thoracic duct; ISLV, intersegmental lymphatic vessel. Scale bars, 80  $\mu$ m for head (middle) in (A) and 50  $\mu$ m for other images.

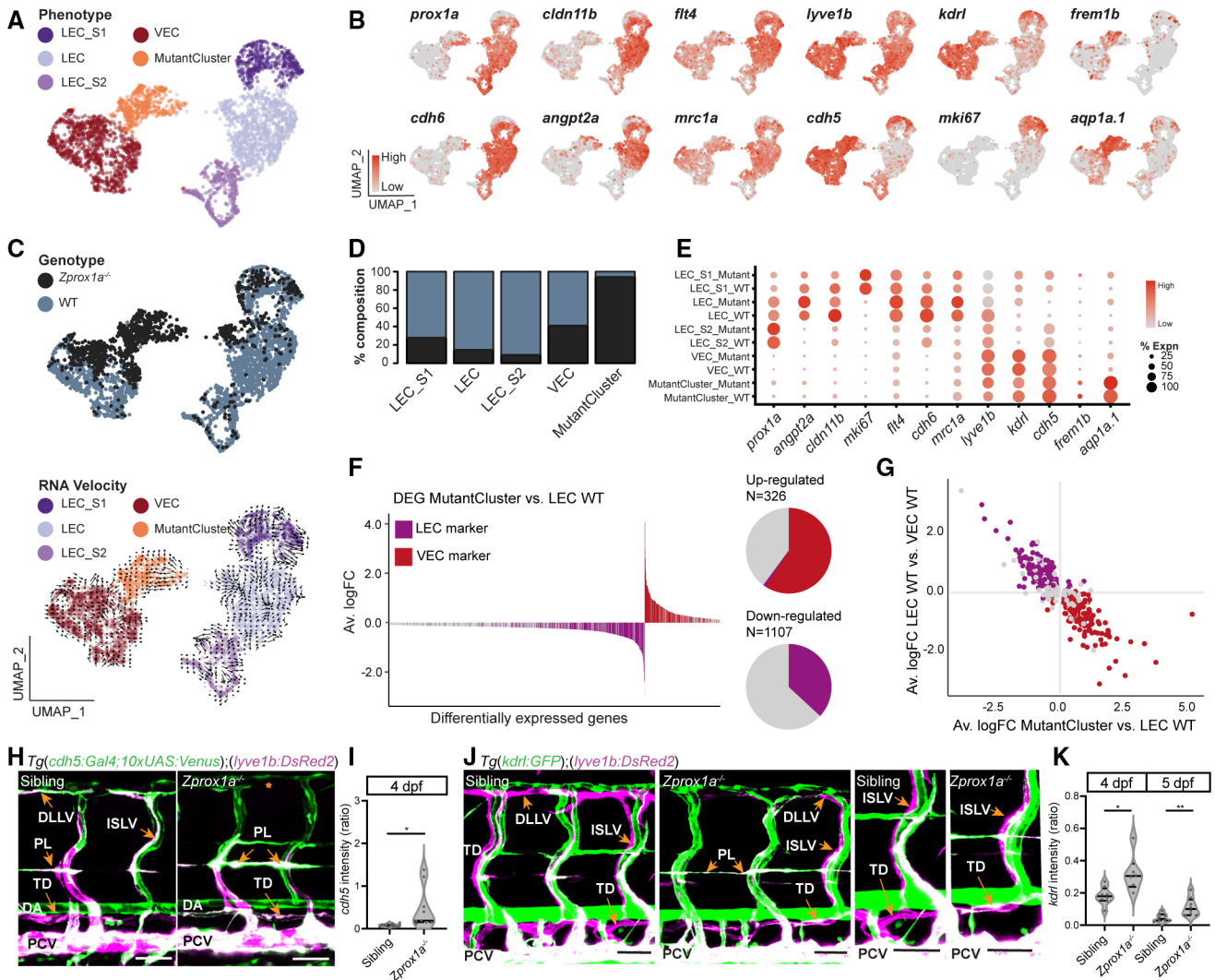
that relied on fewer marker genes. To validate these observations, we used confocal imaging of *Zprox1a*<sup>-/-</sup> mutants and observed upregulation (relative intensity) of blood vascular markers *cdh5* (Figs 3H and I, and EV3A and C) and *kdrl*, the later progressively increasing between 4 and 5 dpf (Figs 3J and K, and EV3D–G) in *lyve1b*-positive lymphatic vessels. This suggests that progressive dedifferentiation is likely occurring in mutant lymphatics. We also saw a coincident reduction in the expression of *lyve1b* in these vessels (Fig EV3B). This demonstrates that *Prox1* alone is sufficient to maintain LEC fate at the whole transcriptomic level and identifies the gene expression maintained by *Prox1* during lymphatic differentiation.

### Single-cell ATAC sequencing reveals chromatin accessibility signatures in LECs and VECs, identifying lymphatic-specific enhancers and predicting key LEC TF families

*Prox1* function is essential for VEC to LEC differentiation; however, whether *Prox1* controls chromatin accessibility during this process is unknown. Thus, we next profiled *Zprox1a*<sup>-/-</sup> mutants at 4 dpf using single-nucleus (sn) ATAC-seq ( $n = 3,731$  nuclei, Fig EV4A, Dataset EV3A). Cluster analysis (Fig 4A) and overall accessibility of key markers (Fig 4B) identified similar populations to the scRNA-seq profiling: canonical LEC (LEC\_01  $n = 114$  nuclei, LEC\_02  $n =$

213 nuclei) and VEC (VEC\_01  $n = 157$  nuclei) clusters, and a small but discrete population comprised almost entirely of mutant cells (mutant cluster  $n = 47$  nuclei) as was seen in our scRNA-seq dataset (and considered the equivalent cluster of mutant cells because all other cell populations were accounted for). Notably, we found that there was almost no contribution of mutant cells to the LEC clusters (Fig 4A lower bar plot), indicating a more profound fate shift when cell identity is determined at the level of chromatin accessibility (Fig EV4B). Differential accessibility (DA) analysis revealed unique sets of peaks that mark individual VEC, LEC and AEC (arterial EC) clusters (Fig 4C).

To identify phenotype-specific chromatin accessibility, we performed DA analysis between WT LEC and WT VEC. This revealed the more accessible regions in the LEC clusters were associated with LEC genes (based on our scRNA-seq atlas) and the less accessible regions with VEC genes (Fig 4D, Dataset EV3B). DA identified  $n = 1,561$  LEC-specific peaks (Fig EV4F) and  $n = 2,624$  VEC-specific peaks representing putative lineage-specific regulatory elements, such as enhancers (Fig EV4E, Dataset EV3B). To test whether these regions identified enhancers, we used the zebrafish enhancer detector plasmid system (ZED vector; Bessa et al, 2009) and tested peaks that were uniquely open 5' of the newly identified LEC marker gene *cdh6*. To exclude non-specific reporter expression, we examined injected F0 embryos using a ZED vector-only control.



**Figure 3. Single-cell RNA-seq analysis reveals a fate shift from LEC to VEC in the absence of Prox1 in zygotic *prox1a* mutants.**

- A UMAP visualisation of  $n = 8,075$  cells filtered for VEC and LEC subpopulations ( $n = 2$  samples; see Dataset EV2A for number of cells per cluster and cluster annotation, and see Fig EV3 for full dataset) coloured according to predicted cell phenotype.
- B UMAP visualisation of marker gene expression. Colour scale represents SCT-normalised expression. LEC markers: *prox1a*, *cldn11b*, *cdh6*, *angpt2a*. LEC and VEC markers: *fit4*, *mrc1a*, *lyve1b*. BEC markers: *cdh5*, *kdrl*. Mutant cluster markers: *frem1b*, *aqp1a.1*.
- C UMAP visualisation with cells coloured according to genotype (*Zprox1a*<sup>-/-</sup> or WT; top) identifies a mutant-specific cluster and RNA velocity analysis (bottom) suggests a trajectory between the mutant cluster and the LEC cluster, likely indicative of a fate shift.
- D Stacked bar plot representing the genotype composition of cells in a given phenotypic cluster.
- E Dot plot of marker expression across defined clusters (indicated on Y-axis). Colour scale represents average SCT-normalised expression, and point size represents percentage of cells expressing gene.
- F Bar plot of average log fold change in gene expression comparing WT LECs with the mutant cluster. Y-axis represents average log fold change, x-axis represents differentially expressed genes (Wilcoxon rank-sum adjusted  $P$ -value  $< 0.05$ ), and bars are coloured according to status as a LEC (purple) or VEC (red) marker in Fig 1. Pie charts (right) indicate the LEC and VEC marker composition of genes upregulated ( $n = 326$ ) and downregulated ( $n = 1,107$ ) in the mutant cluster, demonstrating a fate shift.
- G Concordance in the fate shift between mutant cluster and WT LEC with the WT LEC and VEC trajectory. Each point represents significant DE genes ( $n = 2,287$ ) between mutant cluster and LEC WT, coloured according to LEC or VEC marker status. X-axis represents average log fold change relative to the mutant cluster, and Y-axis represents average log fold change relative to LEC WT.
- H Lateral confocal images of *cdh5* (green) and *lyve1b* (magenta) expression in the developing trunk in WT and *Zprox1a* mutants at 4 dpf.
- I Quantification of *cdh5* intensity in the thoracic duct in WT ( $n = 13$ ) and mutants ( $n = 9$ ) (relative to expression in the DA).
- J Lateral confocal images of *kdrl* (green) and *lyve1b* (magenta) expression at 4 dpf (left) and 5 dpf (right).
- K Quantification of *kdrl* intensity in the thoracic duct in WT and mutants (relative to expression in the PCV) at 4 (WT  $n = 8$ , mutant  $n = 6$ ) and 5 dpf (WT  $n = 8$ , mutant  $n = 8$ ).

Data information: WT, wild-type. Z, zygotic. Vessel abbreviations as in earlier figures and DA, dorsal aorta. PCV posterior cardinal vein. DLLV, dorsal longitudinal lymphatic vessel. PL, parachordal LEC. Scale bars, 50  $\mu$ m. Violin plots show the median and upper/lower quartiles. \* $P < 0.05$  and \*\* $P < 0.01$  from an unpaired, two-sided  $t$ -test.

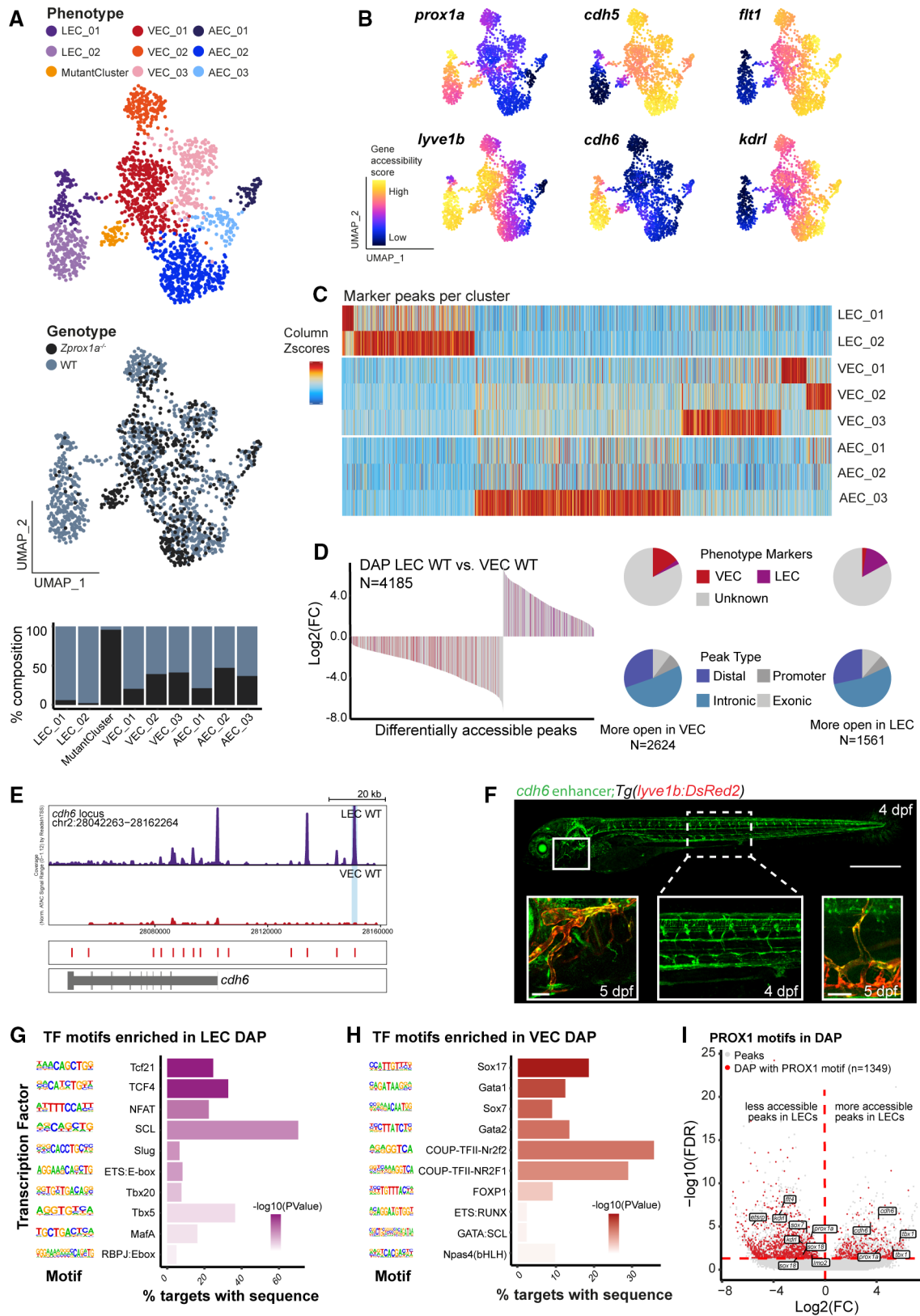


Figure 4.

**Figure 4. Single-nuclei ATAC-seq identifies lineage-specific regulatory regions in VECs and LECs.**

- A UMAP visualisation of snATAC-seq from *Zprox1a*<sup>-/-</sup> mutant and WT endothelial cells at 4 dpf ( $n = 3,731$  nuclei), coloured by cell phenotype (top) and genotype (middle; see Fig EV4 for full dataset). Stacked bar plot (bottom) summarising the composition of wild-type and mutants in each cluster. A mutant-specific cluster is identified, similar to Fig 3.
- B UMAP visualisation indicating accessibility of key marker genes (gene accessibility score with imputation) that confirm predicted cluster phenotypes. LEC markers: *prox1a*, *cdh6*. VEC markers: *cdh5*, *flt1*, *kdrl*.
- C Heatmap of cluster-specific accessible peaks defined using DAP analysis, for all endothelial cells. Colour reflects a column-wise Z-score, rows represent clusters defined in Fig EV4A, and columns are peaks.
- D Bar plot of log<sub>2</sub> fold change for the differentially accessible peaks ( $n = 4,185$  peaks, Wilcoxon rank-sum,  $FDR < 0.05$ ) between WT LECs and WT VECs. LEC/VEC markers identified in scRNA-seq are coloured in red/purple, respectively, and demonstrate a strong correlation between chromatin state and the VEC to LEC fate shift. 228/1,561 more accessible peaks in the LEC cluster were associated with LEC markers and 398/2,624 more accessible peaks in VEC cluster associated with VEC markers. Pie charts (right) summarising the proportion of LEC/VEC markers in the differentially accessible peaks (top) and the type of peak (bottom).
- E Genome accessibility track of LEC marker *cdh6*. Red bars represent peaks in the reproducible peak set from snATAC-seq. The peak highlighted blue indicates a potential enhancer of *cdh6* (*chr2:28150709-28151209*) with significantly more accessible chromatin in WT LECs compared with WT VECs (Wilcoxon rank-sum,  $FDR < 0.05$ ).
- F Overall GFP expression of *cdh6* enhancer (*chr2:28150709-28151209*) reporter at 4 dpf (top). Lateral confocal image of *cdh6* enhancer reporter in the trunk at 4 dpf (bottom middle). Co-expression of *cdh6* enhancer reporter (green) and *lyve1b* (red) in the facial lymphatics (bottom left) and trunk lymphatics (bottom right) at 5 dpf. Scale bars, 500  $\mu\text{m}$  for (top) and 50  $\mu\text{m}$  for (bottom left and right).
- G Vascular TF motifs ( $n = 10$  from top  $n = 50$ ) enriched in peaks that are more permissive in WT LEC compared with WT VEC. The depth of bar colour represents the  $-\log_{10}(\text{RawPVal})$ , y-axis displays individual motifs and schematic (left) and x-axis represents the percentage of target regions enriched for the motif in the  $n = 1,561$  peak set more open in WT LEC. See Dataset EV3C for complete list.
- H Vascular TF motifs ( $n = 10$  from top  $n = 100$ ) enriched in peaks that are less permissive in WT LEC compared with WT VEC. The depth of bar colour represents the  $-\log_{10}(\text{RawPVal})$ , y-axis displays individual motifs and schematic (left), and x-axis represents the percentage of target regions enriched for the motif in the  $n = 2,624$  peak set less open in WT LEC. See Dataset EV3D for complete list.
- I Volcano plot of differentially accessible peaks (DAPs;  $n = 66,568$  peaks, Wilcoxon rank-sum) between WT LECs and WT VECs. Y-axis represents the  $-\log_{10}$  of the false discovery rate ( $FDR$ ), and the x-axis represents the log<sub>2</sub> fold change of each peak. Each point above the horizontal red line represents significant DAPs between WT LECs and WT VECs ( $n = 4,185$ ,  $FDR < 0.05$ ). Points coloured in red represent DAPs with predicted PROX1 binding motif ( $P$ -value for motif detection  $< 7 \times 10^{-4}$ ). Hypergeometric tests for over-representation revealed that the PROX1 motif is more frequently observed in the closed DAP set ( $P = 1.28 \times 10^{-13}$ ) but not the open DAP set ( $P > 0.05$ ), relative to the total peak set.

Data information: DAP, differentially accessible peaks. FDR, false discovery rate. snATAC-seq, single-nuclei ATAC-seq. TF, transcription factor.

This demonstrated mosaic neuronal and dorsal root ganglion expression Fig EV4I, but no lymphatic vascular expression was detected. Transgenesis identified a new functional LEC-specific enhancer upstream of *cdh6*, validating the use of this dataset for lymphatic enhancer discovery (Fig 4E and F).

We next aimed to use this new set of putative enhancer regions in an unbiased manner to identify key TF families likely to regulate LEC development. We performed motif enrichment analysis using HOMER (Heinz *et al.*, 2010) for LEC- and VEC-enriched DA peaks, respectively (Dataset EV3C and D). Notably, we found motifs for TCF, ETS, SCL (TAL1), NFAT, TBX, MAF, SLUG and RBPJ family TFs to be enriched in peaks more accessible in LECs (Fig 4G). Analysis of the human homologues in ENCODE data (Rouillard *et al.*, 2016) revealed these TFs regulate a highly connected network of genes expressed in our LEC atlas (Fig EV4G). Importantly, a number of these TFs are already known to play important roles in lymphatics (e.g. NFAT Kulkarni *et al.*, 2009; Norrmen *et al.*, 2009, MAFB Dieterich *et al.*, 2015; Koltowska *et al.*, 2015b, TBX1 Chen *et al.*, 2010, and TCF Nicenboim *et al.*, 2015; Cha *et al.*, 2018), supporting the prediction that members of these TF families will play important functional roles in LEC development. TF motifs enriched in DAPs that were more open in VECs were associated with blood vascular and venous identity as would be expected (e.g. SOX17, SOX7, COUPTFII and GATA2; Fig 4H). Interestingly, we did not find enrichment for the PROX1 motif in either LEC or VEC gene-associated DAPs in our dataset using this method. This prompted us to look for the presence of the consensus PROX1 motif across the two DAP sets (Fig 4I) using the *Fimo* (Grant *et al.*, 2011) tool in MEME suite. We found PROX1 motifs in some DAPs that were more accessible in LECs compared with VECs (including at LEC marker genes *cdh6*, *tbx1* and *prox1a*), but there was no

over-representation of the PROX1 motif in this dataset. Unexpectedly, we observed strong statistical enrichment for the PROX1 motif at DAPs that were closed in LECs but open in VECs (*Hypergeometric test* for over-representation  $P$ -value  $1.28 \times 10^{-13}$ ). This included DAPs associated with key early blood vascular fate regulatory genes such as *etsrp/etv2*, *sox7*, *sox18* and *lmo2* (Fig 4I). This may suggest a role for Prox1 in shutting down regulatory regions at key fate regulating genes. Overall, motifs for TFs of the NFAT, MAF, TCF, TBX and ETS families are associated with regions of open chromatin open in LECs and motifs for PROX1 are more frequently associated with regions of closed chromatin in LECs compared with VECs.

#### Prox1 reduces the accessibility of chromatin peaks associated with blood vascular and haematopoietic transcription factor motifs

Given that cells in the mutant-specific cluster analysed by scRNA-Seq demonstrate loss of LEC gene expression and acquisition of VEC gene expression, we expected chromatin accessibility to change in an equally coordinated manner in these cells. However, analysis of overall accessibility at individual genes (gene score) for the mutant cluster revealed changes inconsistent with a simple fate shift (Dataset EV4A and B). Some LEC-specific genes (e.g. *prox1a*, *cdh6* and *tbx1*) showed loss of transcription in the mutant cluster but increased chromatin accessibility and some specific VEC genes with increased transcription in the mutant cluster (e.g. *cdh5*, *flt1* and *gata6*) also showed discordant chromatin changes (Figs 5A and B, and EV4H). A DAP analysis revealed little concordance between regions with increased accessibility and expression in LEC or VEC in the scRNA-Seq atlas (Fig 5C). We identified a subset of genes with



more accessible chromatin overall in the mutant cluster than either WT LEC or VEC settings (Fig 5D). Notably, at the level of individual peaks we found that  $n = 1,726$  peaks displayed a striking increase in accessibility in the mutant cluster compared with WT LEC and

$n = 1,794$  peaks an increase compared with WT VEC (Fig 5E, Dataset EV4A and B). Of these, 431 were common peaks identifying more accessible chromatin regions in the mutant cluster than in either WT LEC or WT VEC (Fig 5E, examples in Fig 5F).

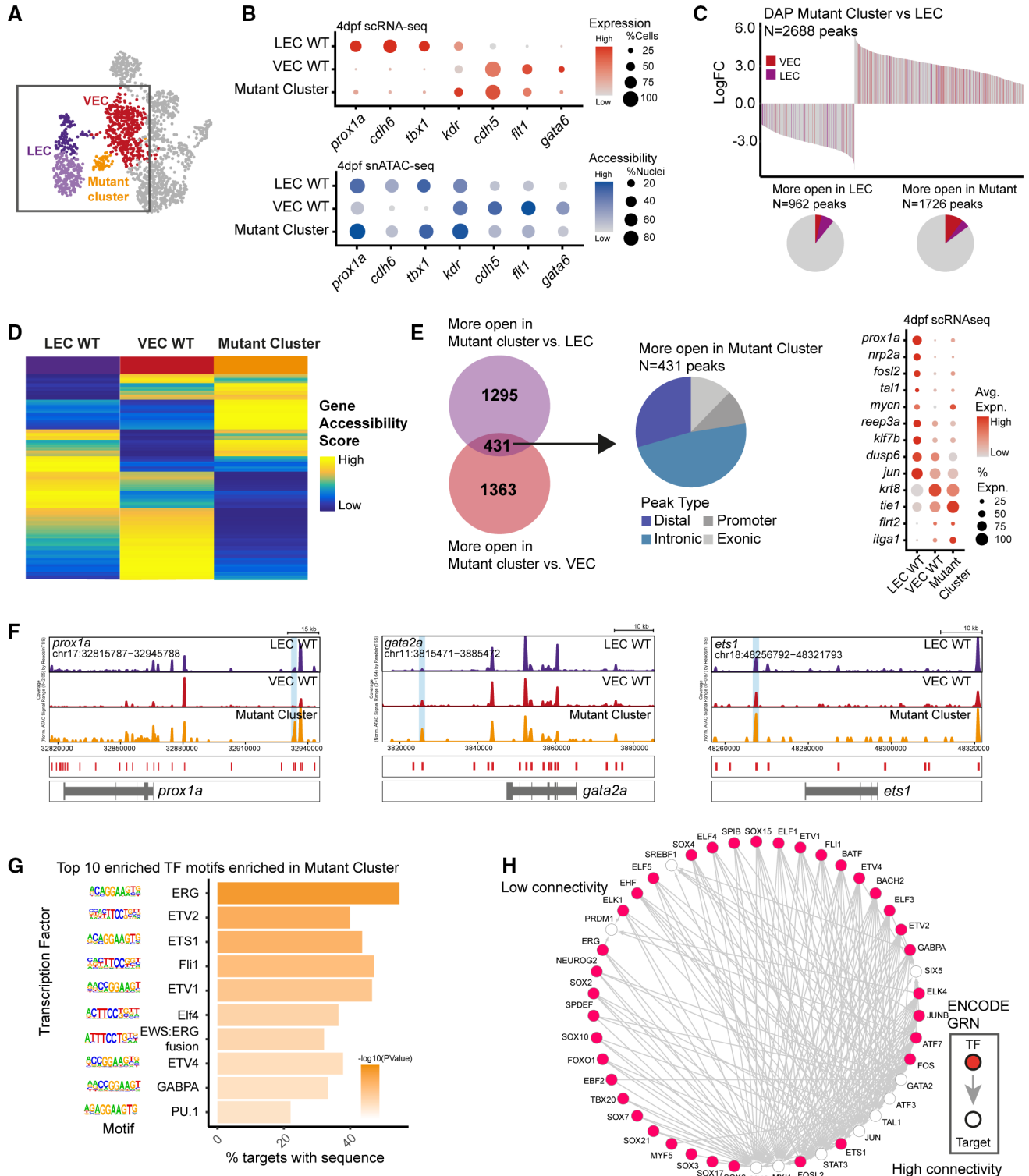


Figure 5.

**Figure 5. Zygotic *prox1a* mutants display a unique chromatin accessibility state consistent with increased activity of early blood and blood vascular fate transcription factors.**

- A Schematic illustrates DAP analyses between mutant cluster vs. LEC WT and mutant cluster vs. VEC WT.
- B Dot plots of scRNA-seq (top) and snATAC-seq accessibility (bottom) data summarising the behaviour of key LEC (*prox1a*, *cdh6*, *tbx1* and *kdr*) and VEC (*cdh5*, *flt1* and *gata6*) markers in WT LECs, WT VECs and the mutant cluster at 4 dpf. LEC genes are upregulated (scRNA-seq) and chromatin is more permissive (snATAC-seq) in WT LEC, and VEC genes are upregulated and more permissive in WT VEC. This concordance between gene expression and chromatin accessibility is lost in the mutant cluster. The size of the dots represents the proportion of cells or nuclei, and colour represents either SCT-normalised expression or gene score of accessibility.
- C Bar plot of log<sub>2</sub> fold change for the differentially accessible peaks ( $n = 2,688$  peaks, Wilcoxon rank-sum,  $FDR < 0.05$ ) between mutant cluster and WT LECs. LEC/VEC markers identified in scRNA-seq are coloured in purple/red, respectively, and demonstrate regions more open in the Mutant are associated with more vascular than lymphatic genes. Pie charts (bottom) summarising the proportion of LEC/VEC markers in the differentially accessible peaks.
- D Heatmaps of accessibility (gene accessibility score) for all genes ( $n = 32,020$ ) in WT LECs, WT VECs and mutant cluster, showing that mutant cluster cells display a unique chromatin state at many genes. Colour scale indicates level of accessibility.
- E Venn diagram (left) indicates all individual peaks with increased accessibility in the mutant cluster vs LEC or VEC, with  $n = 431$  DAP more open in both LEC WT and VEC WT ( $RawPVal < 0.05$ , log<sub>2</sub> fold change  $> 1.5$ ). Pie chart (middle) indicates the proportion of peak types for  $n = 431$  DAP classified as distal, intronic, promoter and exonic, respectively. Dot plot (right) summarising the scRNA-seq expression level of  $n = 13$  genes with DAP more open in the mutant cluster at 4 dpf, demonstrating that accessibility changes for these genes did not correlate with changes in transcription. The size of the dot represents the proportion of cells that express the markers in the cluster, and colour represents SCT-normalised expression.
- F Genome accessibility tracks for key markers with DAP more permissive in the mutant cluster: *prox1a*, *gata2a* and *ets1* (left to right). Red bars represent peaks in the reproducible peak set from snATAC-seq. Blue bars highlight DAP (Wilcoxon rank-sum,  $FDR < 0.05$ ).
- G Top 10 enriched motifs (HOMER analysis, adjusted  $P$ -value  $< 0.05$ ) in the  $n = 431$  peaks that are more open in the mutant cluster than WT LEC and VEC. The depth of colour represents the  $-\log_{10}$  ( $RawPVal$ ), y-axis displays individual motifs and schematics (left), and x-axis represents the percentage of peaks enriched for the motif in the  $n = 431$  peak set.
- H Degree-sorted gene regulatory network displaying known TF binding at genes with more permissive chromatin in mutant cluster compared with WT LECs. TFs are represented by red circles (nodes), target genes by white circles (nodes), and known binding of TF to target in ENCODE by a grey arrow (edges). Nodes with a larger number of edges are more highly connected (bottom right), and nodes with fewer edges are less connected (top left).
- Data information: DAP, differentially accessible peaks. FDR, false discovery rate. snATAC-seq, single-nuclei ATAC-seq. TF, transcription factor.

To investigate the nature of the peaks that were opening in the absence of Prox1, we used an unbiased assessment of TF motifs within these regions. The more open regions were highly enriched for motifs of early acting TFs involved in embryonic vasculogenesis and haematopoiesis, including Erg, Etv2, Etv4, Ets1, Fli1 and Spi1/Pu.1 (Fig 5G, Dataset EV4C). We examined the human homologues of these TFs in ENCODE (Rouillard *et al*, 2016) data together with homologues of genes expressed in our atlas and identified a highly connected putative gene regulatory network (GRN) made up of early blood and blood vascular TFs (including ETV2, TAL1, SOX7, SOX17 ERG and other key fate regulators) driving target genes, including each other, FOS, JUN, MYC and STAT3 (Fig 5H). We take this to suggest that in the absence of Prox1, TFs that are part of a GRN normally suppressed by Prox1 function are reactivated to drive blood vascular and blood fates. This further suggests that Prox1 coordinates the correct accessibility of the chromatin and likely controls other TF functions while maintaining LEC fate. Finally, we note that a large number of the genes with increased chromatin accessibility are known regulators of lymphangiogenesis and so this increased accessibility may be a sensitive way to identify key regulatory factors (Dataset EV4C).

**Prox1 is required cell autonomously for lymphatic development and double maternal zygotic *prox1a*<sup>-/-</sup>, *prox1b*<sup>-/-</sup> mutants show delayed sprouting of LEC progenitors**

While the above data identifies the role of Prox1 at stages when it is maintaining LEC identity in assembling LECs, Prox1 is essential for the very earliest decision made when a VEC becomes specified to develop into a LEC (Wigle & Oliver, 1999). To ask how Prox1 controls the earliest stages of VEC to LEC transdifferentiation, we first needed to generate a complete Prox1 loss of function model in zebrafish. We crossed *prox1a*<sup>+/-</sup>, *prox1b*<sup>+/-</sup> double heterozygous

animals (lacking both Prox1 homologues) and then used germline transplantation approaches to produce animals carrying a double mutant germline (Fig EV5A). Crossing of these animals generated embryos that were maternal zygotic (MZ) *MZprox1a*<sup>-/-</sup>, *MZprox1b*<sup>-/-</sup> double mutant zebrafish completely lacking *prox1a/b* transcript expression and maternal deposition (Figs 6A and EV5B and C). A quantitative phenotypic analysis revealed double MZ mutants show a severe reduction of facial lymphatics and a near complete loss of lymphatic vessels in the trunk by 4 dpf (Figs 6B–C,E and EV5D–G). Furthermore, *MZprox1a*<sup>-/-</sup>, *MZprox1b*<sup>-/-</sup> mutants initially show a delay in the formation of PLs as these cells emigrate the CV to invest the HM, but despite this initial delay, these cells eventually seed the HM before failing to undertake any further migration (Fig 6D and F). This observation is different to observations from mouse studies that Prox1 function is needed for LEC sprouting (Yang *et al*, 2012) and suggests that sprouting and cell fate acquisition can be genetically separable events. We saw evidence for genetic interaction between *prox1a* and *prox1b* in PL formation and no change in the overall number of cells sprouting from the PCV or number of cells contributing to the venous ISVs (blood vessels) suggesting both *prox1a* and *prox1b* contribute specifically to lymphatic development (Fig 6G–K). To test whether LECs require Prox1a in a cell-autonomous manner, we performed embryonic transplantation to generate chimeric embryos. Due to challenges maintaining germline mutant females to generate double MZ mutants, we performed transplantation only of *MZprox1a*<sup>-/-</sup> mutant cells into wild-type hosts and assessed the contributions of vascular grafts to arteries, veins and lymphatics (Appendix Fig S1A). *MZprox1a*<sup>-/-</sup> mutant cells efficiently contributed AECs, VECs but not LECs to developing vessels in otherwise wild-type hosts at 5 dpf (Appendix Fig S1B–D). Thus, we confirmed that Prox1 is necessary cell autonomously for zebrafish lymphatic development and that the *MZprox1a*<sup>-/-</sup>, *MZprox1b*<sup>-/-</sup> mutant phenotype is more severe than any previously described zebrafish Prox1 mutants.

**Prox1 functions to suppress blood and blood vascular fate during the early specification of LEC fate in the embryo**

To understand the very earliest role that Prox1 plays in LEC development, we profiled the endothelium of *MZprox1a*<sup>-/-</sup>, *MZprox1b*<sup>-/-</sup>

mutant zebrafish at 40 hpf (when cells at various regions of the PCV are both being specified to LEC fate and also actively sprouting; Koltowska et al, 2015a; Shin et al, 2016; Baek et al, 2019) using scRNA-seq as described previously (Dataset EV5A, Appendix Fig S2A–C). Using key marker expression and DE analysis, we identified

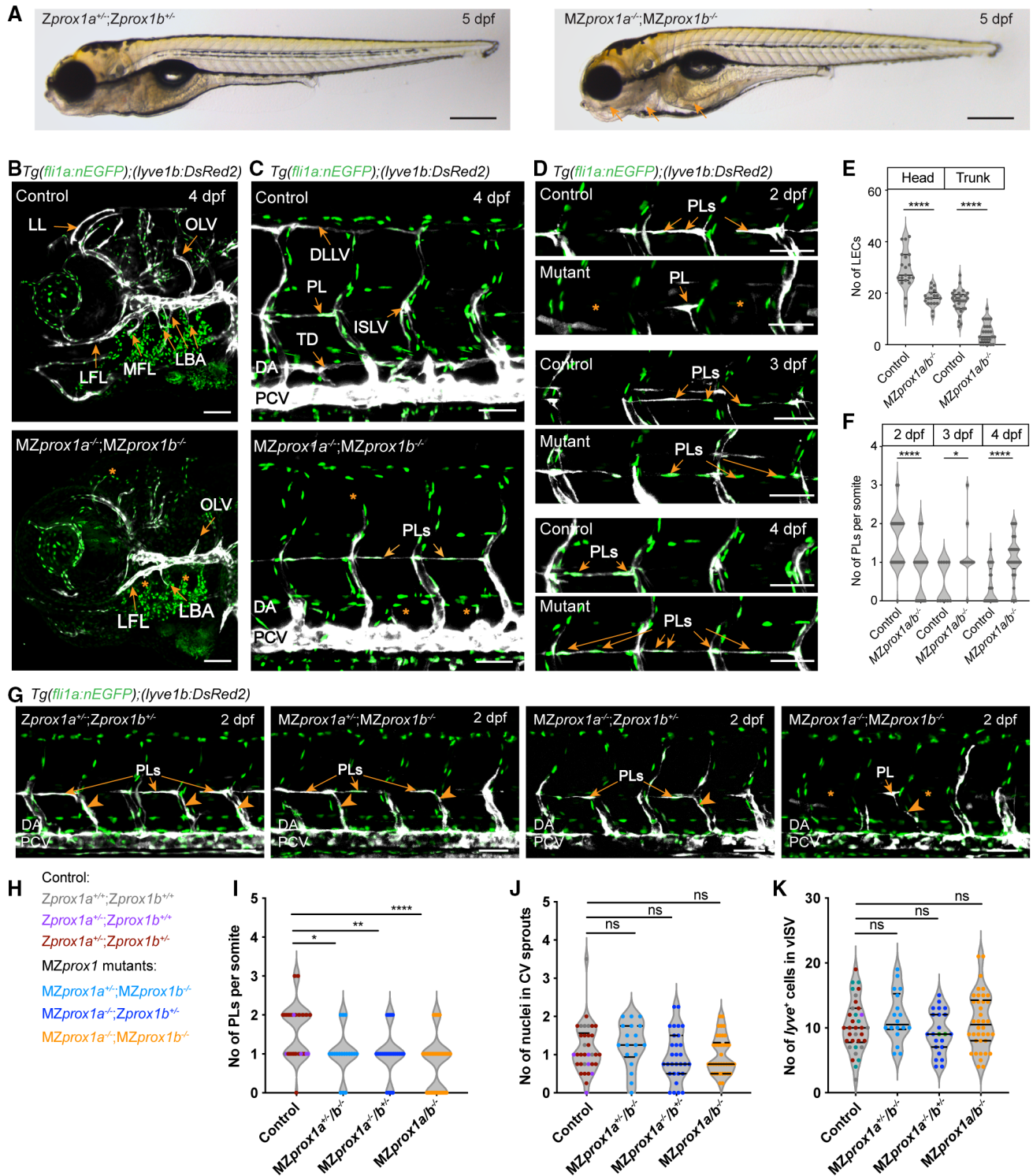


Figure 6.

**Figure 6. Maternal zygotic *prox1a*, *prox1b* double mutants display a specific loss of lymphatic vessels throughout the developing embryo.**

- A Overall morphology of control and *MZprox1a*<sup>-/-</sup>; *MZprox1b*<sup>-/-</sup> mutants at 5 dpf. Arrows indicate oedema around eyes, heart and intestine. Scale bars 500  $\mu$ m.
- B Lateral confocal images of zebrafish heads at 4 dpf showing endothelial cell nuclei (green) and venous and lymphatic vessels (white) in control (upper) and *MZprox1a*<sup>-/-</sup>; *MZprox1b*<sup>-/-</sup> embryos (lower). Facial lymphatics are absent or shorter (asterisk) in the mutants (asterisk). Scale bar, 80  $\mu$ m.
- C Lateral confocal images of zebrafish trunks at 4 dpf in control (upper) and *MZprox1a*<sup>-/-</sup>; *MZprox1b*<sup>-/-</sup> mutants (lower), showing absent lymphatic vessels (asterisk) but retained PLs (arrows in lower). Scale bars, 50  $\mu$ m.
- D Lateral confocal images of PLs in the horizontal myoseptum at 2, 3 and 4 dpf. PLs form later in *MZprox1a*<sup>-/-</sup>; *MZprox1b*<sup>-/-</sup> mutants and accumulate in the horizontal myoseptum while the PLs of control embryos emigrate the HM by 4 dpf. Scale bars, 50  $\mu$ m.
- E Quantification of lymphatic endothelial cell (LEC) number from 4 dpf heads (control: *n* = 19, *MZprox1a*<sup>-/-</sup>; *MZprox1b*<sup>-/-</sup> mutants: *n* = 18) and trunks (control: *n* = 29, *MZprox1a*<sup>-/-</sup>; *MZprox1b*<sup>-/-</sup> mutants: *n* = 25).
- F Quantification of number of PLs per somite in trunks at 2 dpf in controls (*n* = 41) and *MZprox1a*<sup>-/-</sup>; *MZprox1b*<sup>-/-</sup> mutants (*n* = 38), at 3 dpf in controls (*n* = 11) and *MZprox1a*<sup>-/-</sup>; *MZprox1b*<sup>-/-</sup> mutants (*n* = 10) and at 4 dpf in controls (*n* = 29) and *MZprox1a*<sup>-/-</sup>; *MZprox1b*<sup>-/-</sup> mutants (*n* = 25).
- G Lateral confocal images showing defects in the formation of PLs in mutants upon loss of *prox1a* compared with controls in the trunk. Genotypes are indicated.
- H Colour-coded list of analysed genotypes abbreviated in (I–K). Each embryo has a defined genotype for *prox1b*, which is represented in colour code as displayed in (J).
- I Quantification of the number of PLs formed (per somite) at 2 dpf in control (*n* = 41), *MZprox1a*<sup>+/-</sup>; *MZprox1b*<sup>-/-</sup> (*n* = 18), *MZprox1a*<sup>-/-</sup>; *MZprox1b*<sup>+/-</sup> (*n* = 31), *MZprox1a*<sup>-/-</sup>; *MZprox1b*<sup>-/-</sup> (*n* = 38). Decreasing gene dosage for *prox1a* and *prox1b* progressively reduced the initial seeding of the HM by PLs.
- J The number of cells in sprouts departing the PCV at 2 dpf in control (*n* = 41), *MZprox1a*<sup>+/-</sup>; *MZprox1b*<sup>-/-</sup> (*n* = 18), *MZprox1a*<sup>-/-</sup>; *MZprox1b*<sup>+/-</sup> (*n* = 31), *MZprox1a*<sup>-/-</sup>; *MZprox1b*<sup>-/-</sup> (*n* = 38) is unchanged in mutants showing that sprouting occurs normally.
- K The number of cells in vISVs at 2 dpf in control (*n* = 41), *MZprox1a*<sup>+/-</sup>; *MZprox1b*<sup>-/-</sup> (*n* = 18), *MZprox1a*<sup>-/-</sup>; *MZprox1b*<sup>+/-</sup> (*n* = 31), *MZprox1a*<sup>-/-</sup>; *MZprox1b*<sup>-/-</sup> (*n* = 38) is unchanged in mutants indicating no effect on the venous endothelium.

Data information: Vessel abbreviations as in earlier figures and LL, lymphatic loop. OLV, otolithic lymphatic vessel. LFL, lateral facial lymphatic vessel. MFL, medial facial lymphatic vessel. LBA, lymphatic branchial arches. Violin plots show the median and upper/lower quartiles. \**P* < 0.05, \*\**P* < 0.01, \*\*\*\**P* < 0.0001. Unpaired, two-sided *t*-test (E, F), ordinary one-way ANOVA test with Dunnett's multiple comparison test (K) and Kruskal–Wallis test with Dunn's multiple comparison test (I, J).

populations of cells that included AECs, endocardium, VEC, mixed populations of VECs and Prox1<sup>+</sup> LEC progenitors (named LEC\_VEC) and clusters with expression of some VEC and AEC markers that were likely still differentiating (Fig 7A–C, Appendix Fig S2D–H). Based on expression of known markers and genes associated with EC sprouting, we defined two populations of secondary sprouts of venous origin (LEC\_VEC 01 *n* = 713 cells, LEC\_VEC 02 *n* = 677 cells) and a single population of cells of the cardinal vein (PCV *n* = 812 cells; Fig 7C–F, Appendix Fig S3C). Consistent with our observations in Fig 1, cells expressing *prox1a* at 40 hpf were not transcriptionally distinct from sprouting VECs and failed to form a “lymphatic progenitor” cluster, suggesting they are specified and express *prox1a* but not yet differentiated. RNA velocity analysis suggested that these three clusters remained closely related, consistent with little

differentiation between these populations at this stage (Fig 7A right).

To identify the earliest transcriptional changes controlled by Prox1 in lymphatic development, we used DE to evaluate the global differences between *MZprox1a*<sup>-/-</sup>, *MZprox1b*<sup>-/-</sup> and WT in the LEC\_VEC 01, LEC\_VEC 02 and PCV clusters of sprouting cells (Fig 7G, Dataset EV5B). We found a significant upregulation of a large set of genes in the absence of Prox1 that were identified as enriched for blood vascular and haematopoietic genes by GO term and marker analysis (Fig 7H and I, *n* = 1,137 genes, Dataset EV5C). We also saw increased expression of mitochondrial metabolism genes and of pre-mRNA splicing genes in the absence of Prox1 (Fig 7G, Dataset EV5C). At this stage of development, in contrast to the 4 dpf stage, we saw little evidence of positively regulated genes

**Figure 7. Single-cell profiling of maternal zygotic *prox1a*, *prox1b* double mutants reveals Prox1-negative regulation of gene expression during a fate transition.**

- A UMAP visualisation of endocardial, venous and arterial endothelial cells (Level 02 *n* = 5,347) coloured according to predicted cell phenotype (left), genotype (middle) and RNA velocity (right; *n* = 4 samples; see Dataset EV5A for number of cells per cluster and cluster annotation, and see Appendix Fig S2 for full dataset)
- B Heatmap displaying SCT-normalised expression of phenotype-specific genes defined using differential expression analysis. Columns are cells grouped by phenotype assignment, rows are genes, and colour indicates the average level of expression.
- C Dot plot indicating expression of key markers used to define cell phenotype. Colour scale represents average SCT-normalised expression, and point size represents percentage of cells expressing gene.
- D UMAP visualisation of endocardial, venous and arterial endothelial cells (Level 02 *n* = 5,347) with the LEC\_VEC and VEC populations selected for further analysis indicated by a red dashed box.
- E UMAP visualisation of the reclustered venous and lymphatic endothelial cell populations (Level 03 *n* = 2,747) coloured according to predicted cell phenotype (top) and genotype (bottom).
- F UMAP visualisation of key marker gene expression. Colour scale represents SCT-normalised expression.
- G Bar plot indicating average log fold change of *n* = 1,186 significantly different genes between all *MZprox1a/b*<sup>-/-</sup> and WT venous endothelial cells (LEC\_VEC 01, LEC\_VEC 02 and PCV combined) at 40 hpf (Wilcoxon rank-sum adjusted *P*-value < 0.05; Dataset EV5B). Colour indicates genes associated with one or more mitochondrial, mRNA processing and blood vascular GO terms.
- H Bar plot summarising GO BP analysis of the *n* = 1,137 genes upregulated in the *MZprox1a/b*<sup>-/-</sup> venous endothelial cells (Dataset EV5C). Y-axis represents enriched GO BP term, x-axis represents the  $-\log_{10}$  (adjusted *P*-value), and bars are coloured according to fold enrichment.
- I Dot plot of *n* = 36 key blood vascular markers upregulated in the *MZprox1a/b*<sup>-/-</sup> venous endothelial cells, indicating genotype-specific expression in LEC\_VEC and PCV cell phenotypes. Colour scale represents average SCT-normalised expression, and point size represents percentage of cells expressing gene.
- J STRING analysis of *n* = 36 key blood vascular markers upregulated in the *MZprox1a/b*<sup>-/-</sup> venous endothelial cells.
- K Degree-sorted gene regulatory network displaying known TF binding at *n* = 1,137 genes upregulated in the *MZprox1a/b*<sup>-/-</sup> venous endothelial cells. TFs are represented by red circles (nodes), target genes by white circles (nodes), and known binding of TF to target in ENCODE by a grey arrow (edges). Nodes with a larger number of edges are more highly connected (bottom right), and nodes with fewer edges are less connected (top left).

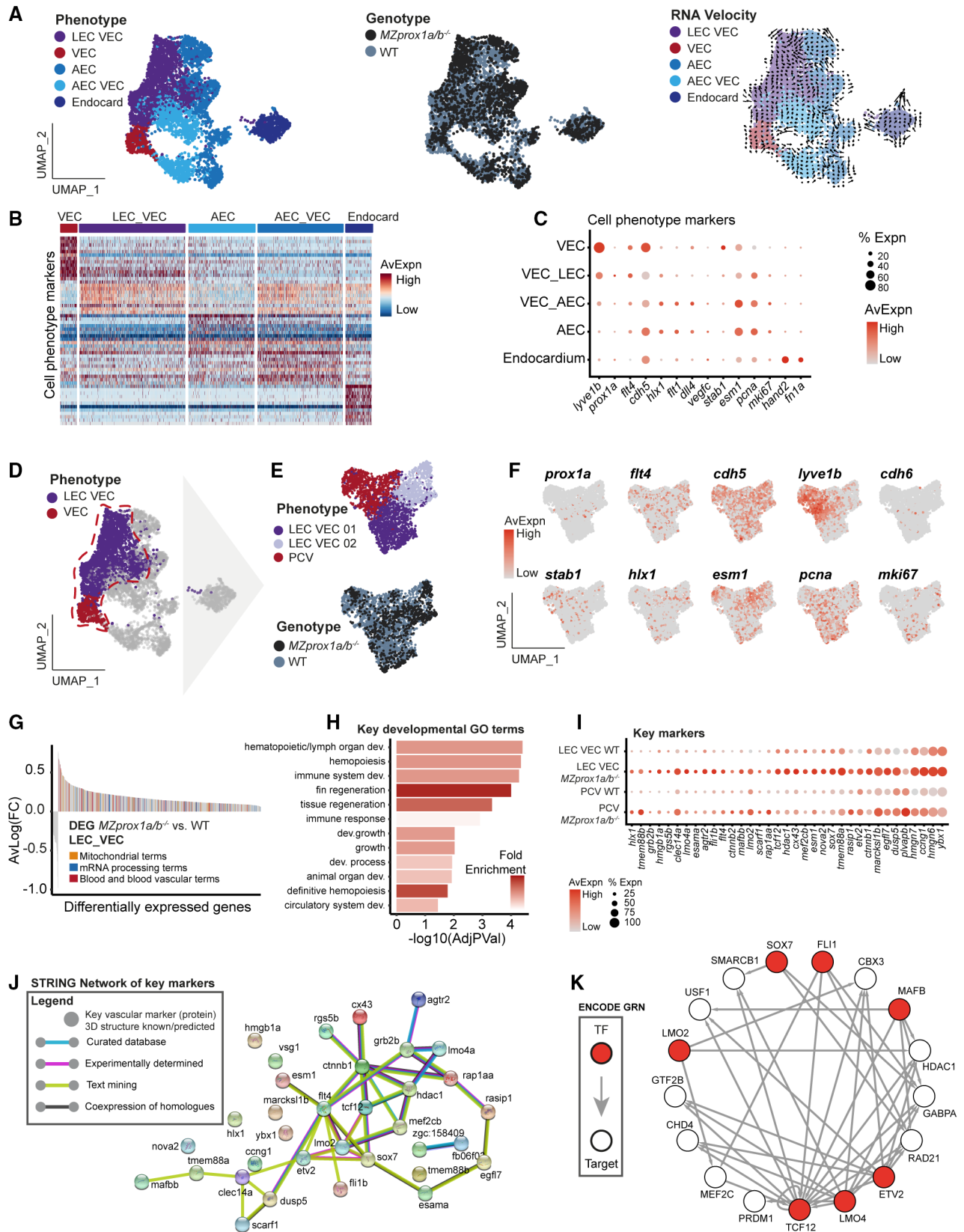


Figure 7.

downstream of Prox1 (just  $n = 27$  genes downregulated in *MZprox1a*<sup>-/-</sup>, *MZprox1b*<sup>-/-</sup>; Dataset EV5B). We next examined the genes upregulated in Prox1 mutants that we identified as key early blood and vascular developmental regulators. This includes *esm1*, *etv2*, *flt4*, *mef2c*, *hdac1*, *lmo2*, *lmo4*, *rasip1*, *egfl7*, *dup5*, *clec14a*, *fli1b* and others (Fig 7I, Dataset EV5C). STRING (Szklarczyk et al, 2017) analysis suggested the presence of inter-related gene networks (Fig 7J) and so we further examined upregulated TFs in this network by leveraging human homologues in ENCODE data and identifying known target genes expressed in our dataset. This identified a GRN containing SOX7, FLI1, LMO2, ETV2, TCF12, LMO4 and MAFB as well as known targets of these TFs, which is repressed in response to Prox1 function during LEC specification (Fig 7K). This is highly concordant with observations at later stages of upregulated TF expression and activity in 4 dpf mutants (Fig 5) and the observation that the Prox1 motif is over-represented in DAPs less accessible in LECs than in VECs at 4 dpf (Fig 4I). Overall, this suggests that in the earliest stages of LEC specification Prox1 is probably not actively driving a fate program but plays a major role downregulating expression and function of blood vascular and haematopoietic fate genes. It seems likely that the negative regulation of core blood and blood vascular fate genes (those in Fig 7K) could be sufficient to allow differentiation to then occur down a new route in these earliest specified LECs.

### Lymphatic Notch signalling is essential for lymphangiogenesis with Notch components and targets expressed at stages following LEC specification

To further demonstrate the utility of the single-cell data above, we next turned our attention to an area that has remained murky: the precise timing and role of Notch signalling in lymphatic

development. Work in mouse reported that Notch signalling negatively regulates VEC to LEC transition via suppression of Prox1 expression (Murtoomaki et al, 2013). This work contradicts the observation that endothelial deletion of the core Notch effector Rbpj (using *Tie2:Cre; Rbpj*<sup>f/f</sup>) has no effect on the expression of Prox1 (Srinivasan et al, 2010). In zebrafish, it was found that lymphatics do not form in the absence of Notch signalling, with PLs replaced by ectopic venous ISVs (vISVs), suggesting a fate shift (Geudens et al, 2010). However, this was recently explained with the demonstration that hard-wired arterial (aISV) fate through Notch signalling was responsible for the abnormal wiring of vISVs and subsequent, secondary loss of PLs (Geudens et al, 2019). These studies led to a current model whereby Notch signalling in zebrafish is not autonomous to the developing veins and lymphatics but rather acts upon arteries.

We examined our wild-type atlas for the expression of major Notch receptors, ligands and target genes. We found that *notch1a* and *notch1b*, but not *notch2* or *notch3*, were expressed in LECs at 3, 4 and 5 dpf. *dll4* showed strong expression in 4 and 5 dpf LECs and there were detectable levels of *jag2b* and *jag1b* in 4 and 5 dpf LECs (Fig 8A and B). Known Notch targets *hey1* and *hey2* were not expressed, but *her6* was a strong LEC-specific marker and *her9* was also detectable at 3, 4 and 5 dpf in LECs. Importantly, at 40 hpf there was little expression of any of these Notch pathway components (Fig 8B). In zygotic *prox1a* mutants when we compared differentiated mutant LECs with WT LECs (cells in the LEC cluster only), we noted that *notch1a* expression appeared to be reduced in the absence of zygotic *prox1a*, as did expression of *her6* (Fig 8C). Finally, we found evidence of differential regulation between LECs and VECs at DAPs at the loci for *notch1a*, *notch1b* and *jag2b* in our snATAC-seq data and *prox1a*-dependent peaks for *notch1b* and *jag2b* (Fig 8D).

#### Figure 8. Notch signalling regulates lymphatic development following LEC specification.

- A UMAP visualisation indicating expression of *her6*, *dll4*, *notch1a* and *notch1b* in the developmental time course ( $n = 9,771$  cells filtered for VEC and LEC populations). Colour scale represents log-normalised expression.
- B DotPlot indicating expression of major Notch pathway genes in LEC populations in the developmental time course. Colour scale represents average SCT-normalised expression and point size the percentage of cells expressing each gene.
- C UMAP visualisation (left) of the *Zprox1a*<sup>-/-</sup> dataset indicating key cell phenotypes ( $n = 8,075$  cells filtered for VEC and LEC subpopulations at 4 dpf). LEC populations depicted in violin plots are indicated by the circle. Violin plots (right) indicating SCT-normalised expression (y-axis) of Notch pathway components *her6*, *dll4*, *notch1a* and *notch1b* in WT LEC and *Zprox1a*<sup>-/-</sup> LEC cells (x-axis).
- D Genome accessibility track of key Notch pathway components *notch1a*, *notch1b* and *jag2b*. Red bars represent peaks in the reproducible peak set from snATAC-seq. DAPs between the mutant cluster and WT LECs are highlighted blue, and those between WT LEC and WT VEC are highlighted green (Wilcoxon rank-sum,  $FDR < 0.05$ ).
- E Lateral spinning disc confocal images of zebrafish larvae at 5 dpf showing blood vessels (black) in control (upper) and *notch1b*<sup>-/-</sup> larvae (lower). ISV hypersprouting (arrowheads) is observed in *notch1b*<sup>-/-</sup> larvae. Scale bar, 200  $\mu\text{m}$ .
- F Lateral spinning disc confocal images of zebrafish trunk at 5 dpf showing blood vessels (black) in control (upper) and *notch1b*<sup>-/-</sup> larvae (lower). Right bottom images are magnification of orange dotted region. ISV hypersprouting (arrowheads) is observed in *notch1b*<sup>-/-</sup> larvae ( $n = 35$ ). Scale bar, 100  $\mu\text{m}$ .
- G Quantification of venous/arterial ISV ratio at 5 dpf control ( $n = 11$ ) and *notch1b*<sup>-/-</sup> larvae ( $n = 10$ ). Venous/arterial ISV ratio is reduced in *notch1b*<sup>-/-</sup> larvae.
- H Lateral spinning disc confocal images of zebrafish trunk at 2.5 dpf showing endothelial cell nuclei (green) and venous and lymphatic vessels (white) in control (upper) and *notch1b*<sup>-/-</sup> embryos (lower). The number of PLs (arrowheads) is increased in *notch1b*<sup>-/-</sup> embryos. Scale bar, 100  $\mu\text{m}$ .
- I Lateral spinning disc confocal images of zebrafish trunk at 5 dpf showing endothelial cell nuclei (green) and venous and lymphatic vessels (white) in control (upper) and *notch1b*<sup>-/-</sup> larvae (lower). The *notch1b*<sup>-/-</sup> larvae (bottom) has wild-type venous/arterial ISV ratio. Lymphatic vessel formation is impaired in *notch1b*<sup>-/-</sup> larvae. Scale bar, 100  $\mu\text{m}$ .
- J Quantification of LEC number in 2.5, 3, 4 and 5 dpf control (2.5 ( $n = 36$ ), 3 ( $n = 31$ ), 4 ( $n = 31$ ) and 5 dpf ( $n = 25$ )) and *notch1b*<sup>-/-</sup> embryos/larvae (2.5 ( $n = 14$ ), 3 ( $n = 11$ ), 4 ( $n = 11$ ) and 5 dpf ( $n = 11$ )). LEC number is increased in 2.5 dpf *notch1b*<sup>-/-</sup> embryos but reduced in 3, 4 and 5 dpf *notch1b*<sup>-/-</sup> larvae.
- K Correlation plot showing LEC number (y-axis) and venous/arterial ISV ratio (x-axis) in 5 dpf *notch1b*<sup>-/-</sup> larvae. No correlation between LEC number and venous/arterial ISV ratio in *notch1b*<sup>-/-</sup> larvae (Pearson's  $r$  value =  $-0.031$ ,  $P$ -value =  $0.8979$ , Pearson's test of correlation).
- L Lateral confocal images of zebrafish trunk at 2 dpf showing endothelial cell nuclei (green) and *prox1a*-positive cells (magenta) in control (upper) and *notch1b*<sup>-/-</sup> embryos (lower). *notch1b* loss does not impair *prox1a* expression in endothelial cells. Scale bar, 50  $\mu\text{m}$ .
- Data information: Vessel abbreviations as in earlier figures. Violin plots show the median and upper/lower quartiles. \*\* $P < 0.01$ , \*\*\*\* $P < 0.0001$ . Unpaired, two-sided t-test (G and J (2.5, 3 and 5 dpf)) and two-tailed Mann-Whitney test (J (4 dpf)).

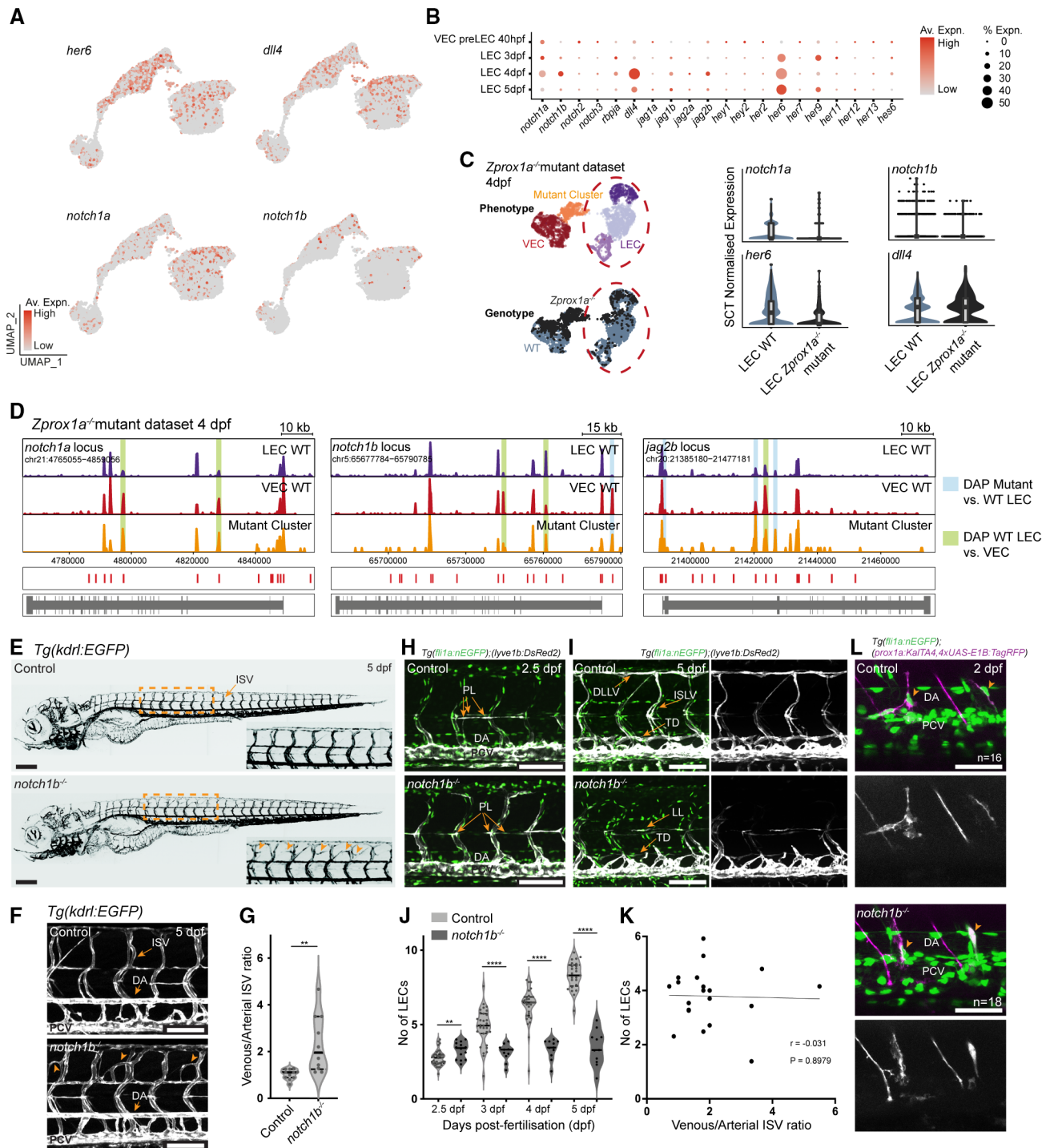


Figure 8.

To determine the functional relevance of the above observations, we took advantage of a *notch1b*<sup>luq53bh</sup> mutant strain that we had identified in an earlier forward genetic screen (Koltowska et al, 2015b). This mutant phenocopied earlier Notch1 morphant phenotypes (Geudens et al, 2010), mapped to the *notch1b* locus on Chromosome 5 by whole genome sequencing analysis and sequencing confirmed a nonsense mutation at amino acid position 128 (Fig 8D,

Appendix Fig S2I) predicted to be a loss of function mutation. As previous studies have shown an arterial–venous fate shift (aISV–vISV patterning defect) in the absence of Notch signalling (Geudens et al, 2010, 2019), we scored arterial and venous patterning of trunk ISVs but found that only a portion of the mutant embryos show increased vISV numbers (Fig 8E–G). Thus, this loss of Notch1b represents partial loss of Notch activity and serendipitously allowed us

to examine lymphatic development in embryos with normal aISV-vISV patterning. Different to earlier studies where Notch loss-of-function reduced PLs, we found that there was a mild increase in the number of PLs at 2.5 dpf, but this was coupled with a strong loss of lymphatic vessels in the trunk of these *notch1b*<sup>uq53bh</sup> mutants by 5 dpf (Fig 8H–J). Importantly, there was no correlation between the embryos with aISV-vISV changes and this strong loss of lymphatics (Fig 8K). In addition, we found that *prox1a* was expressed normally in these mutants (Fig 8L). Altogether, this demonstrates an essential role for Notch signalling in lymphangiogenesis that occurs after the stages of establishment of arterial, venous and lymphatic fates.

## Discussion

In this study, we first present a single-cell RNA-seq analysis of four key stages of embryonic lymphangiogenesis in a vertebrate embryo, revealing new markers and potential regulators of LEC differentiation. We further profile the ECs of *Zprox1a*<sup>-/-</sup> mutants, which form lymphatic structures, but these “lymphatics” dedifferentiate or revert their fate to blood vascular in a mutant-specific, fate-shifted, single-cell cluster. This analysis identified the transcriptional code maintained by Prox1 in order to maintain LEC identity and also shows the highly conserved nature of Prox1 function comparing zebrafish with published work from mice. Overall, the striking fate shift identified across the whole transcriptome confirms that the function of just one transcription factor (Prox1) is sufficient to maintain cellular identity, validating its status as the master regulator of LEC identity.

As well as single-cell profiling of the developing LEC transcriptome, we also performed snATAC-seq on 4 dpf zygotic *prox1a* mutants and wild-type, again identifying a mutant-specific, fate-shifted cluster in this analysis. Comparison of wild-type LECs with VECs revealed strong concordance between chromatin accessibility at LEC enhancers and the transcriptional profile for LEC- and VEC-specific genes. TFs motifs enriched at these putative enhancers were associated with known, key, regulatory TFs. Interestingly, however, analysis of the mutant cluster chromatin identified ectopically open regions and peaks with a distinct discordance between chromatin accessibility at enhancers and transcriptional profile. In the mutant cluster at 4 dpf, a large number of opened chromatin regions were enriched for TF motifs for early acting vasculogenic and haematopoietic TFs. This suggests that regulators of early blood vascular fate become more active and increase chromatin accessibility at specific targets in the absence of Prox1. Overall, this revealed that at the

level of chromatin accessibility the fate-shifted ECs display a more immature state, perhaps a consequence of regulatory “confusion” due to a failed fate transition and the presence of normally repressed TFs. It seems likely that Prox1 has a combinatorial function as part of a larger GRN of developmental TFs, which has yet to be studied in detail.

Additional biological insights from this study come from our analysis of double maternal zygotic *prox1a*, *prox1b* mutants. These mutants are presumed “null” for Prox1 orthologues and they revealed that during its earliest function in LEC fate specification and VEC-LEC transdifferentiation, Prox1 acts primarily to negatively regulate blood and blood vascular fate. While this may represent a downstream program rather than direct repression of gene expression by Prox1 orthologues, it is notable that Prox1 and *Drosophila* Prospero have been reported to be able to function as repressors in a context-dependent manner (Qin *et al*, 2004; Song *et al*, 2006; Takeda & Jetten, 2013; Armour *et al*, 2017; Liu *et al*, 2020). Interestingly, the analysis of upregulated genes identified a set of TFs that are known early regulators of blood and blood vascular fates during embryonic haematopoiesis and vasculogenesis. These included Sox7, Etv2/Etsrp, Lmo2 and Lmo4, and we take this to suggest that Prox1 probably functions primarily during VEC-LEC transdifferentiation by blocking expression of early acting blood vascular fate driving TFs. This is in line with motif enrichment for the PROX1 motif at DAPs that are closed in LECs and aligns well with the enrichment of blood and blood vascular fate regulating TF motifs at DAPs that open up in fate-shifted mutant cells at 4 dpf. It will be interesting in the future to understand at a mechanistic level if Prox1 is actively repressing gene expression at bound targets to block alternative cell fates and if so, if this activity of Prox1 is sufficient to determine LEC fate. Finally, as an example of the utility of the datasets generated here, we explored the data for evidence of expression and regulation of Notch signalling components in developing LECs. This led us to clarify a role for Notch signalling in developmental lymphangiogenesis. Analysis of a *notch1b* mutant revealing an essential function following the specification of LECs and not influenced by abnormal arterial–venous fates.

Altogether this study describes the process of developmental venous to lymphatic transdifferentiation, early LEC differentiation and LEC maintenance in detail, *in vivo*. We describe the role of Prox1 during this process, revealing a conserved and dynamic regulatory process with unprecedented resolution. This resource contributes a base of new knowledge that will help to understand lymphangiogenesis in contexts beyond the embryo, such as in pathological lymphangiogenesis in metastasis, inflammation and tissue repair.

## Materials and Methods

### Reagents and Tools table

Reagent/resource	Reference or source	Identifier or catalog number
<b>Experimental Models</b>		
<i>Tg(fli1a:nEGFP)<sup>y7</sup></i>	Lawson & Weinstein (2002) Dev. Biol.	y1
<i>Tg(-5.2lyve1b:DsRed)<sup>nz101</sup></i>	Okuda <i>et al</i> (2012) Development.	nz101



Reagents and Tools table (continued)

Reagent/resource	Reference or source	Identifier or catalog number
<i>Tg(prox1a:KalTA4,4xUAS-E1B:TagRFP)<sup>nim5</sup></i>	van Impel et al (2014) Development. & Dunworth et al (2014) Circ. Res.	nim5
<i>Tg(kdrl:EGFP)<sup>s843</sup></i>	Beis et al (2005) Development.	s843
<i>Tg(kdrl:Has.HRAS-mCherry)<sup>s916</sup></i>	Hogan et al (2009) Nat. Genet.	s916
<i>Tg(10xUAS:Venus)<sup>uq8bh</sup></i>	Lagendijk et al (2017) Nat. Commun.	uq8bh
<i>TgBAC(ve-cad:GALFF)mu101</i>	Bussmann & Schulte-Merker (2011) Development.	mu101
<i>prox1a<sup>i278</sup></i>	van Impel et al (2014)	i278
<i>prox1b<sup>sa35</sup></i>	Koltowska et al (2015a) Cell Rep. & Tao et al (2011) PLoS One.	sa35
Mutant: <i>notch1b<sup>uq53bh</sup></i>	This study	
<i>Tg(lyve1b:Venus)<sup>uq51bh</sup></i>	This study	
Maternal and zygotic mutant: <i>prox1a/b<sup>-/-</sup></i>	This study	
<i>TgBAC(fabp11a:Citrine)<sup>uom103</sup></i>	This study	
<i>TgBAC(slc7a7:Citrine)<sup>uom104</sup></i>	This study	
<i>Tg(cdh6 enhancer:EGFP)<sup>uom105</sup></i>	This study	
<b>Oligonucleotides and sequence-based reagents</b>		
Primers for BAC recombineering ( <i>slc7a7a</i> -BAC-Citrine, <i>fabp11a</i> -BAC-Citrine)		Methods: Transgenesis, genome-editing and genotyping
Primers for genotyping of <i>notch1b<sup>uq53bh</sup></i>		Methods: Transgenesis, genome-editing and genotyping
<b>Chemicals, enzymes and other reagents</b>		
Liberase [2.5 mg/ml]	Sigma-Aldrich	Cat # 5401119001
Zombie Violet TM Viability Dye	BioLegend	Cat # 423113
10× Genomics 1× Nuclei Buffer	10× Genomics	Cat # PN-2000153/2000207
Trypan Blue	Thermo Fisher	Cat # T10282
Countess Cell Counting Chamber Slides		Cat # C10228
<b>Software</b>		
Imaris	Bitplane	Version 9.6 or earlier
ImageJ (Fiji)	Schindelin et al (2012) Nat. Methods.	National Institutes of Health, (Version 2 or earlier)
10× Genomics Cell Ranger software	Zheng et al (2017) Nat Commun.	version 3.1.0 or 3.0.2
FASTQC	<a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc/">https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>	0.11.6
R statistical software		4.0.2
Python		3.6
Scrublet (python)	Wolock et al (2019) Cell Syst	
Velocyto (python)	La Manno et al (2018) Nature	
Seurat (R)	Stuart et al (2019) Cell	3
SC-Transform (R)	Hafemeister & Satija (2019) Genome Biol.	
Scater (R)	McCarthy et al (2017) Bioinformatics.	1.20.1
ClusTree (R)	Zappia & Oshlack (2018) Gigascience	
CSS-simspec (R)	He et al (2020) Genome Biol.	
CellXGene visualisation software	preprint: Li et al (2020) bioRxiv.	Chan-Zuckerberg Initiative
Cell Ranger ATAC	10× Genomics	
ArchR	Granja et al (2021) Nat. Genet.	1.0.1
HOMER	<a href="http://homer.ucsd.edu/homer/motif/">http://homer.ucsd.edu/homer/motif/</a>	4.11
FIMO (MEME suite)	Grant et al (2011) Bioinformatics	5.4.1

Reagents and Tools table (continued)

Reagent/resource	Reference or source	Identifier or catalog number
ENSEMBL Biomart	<a href="https://asia.ensembl.org/info/data/biomart/index.html">https://asia.ensembl.org/info/data/biomart/index.html</a>	
Cytoscape	Shannon <i>et al</i> (2003) Genome Res.	
String.db	Szklarczyk <i>et al</i> (2017) Nucleic Acids Res.	
<b>Other</b>		
Zeiss LSM 710 FCS confocal microscope	Zeiss	
Zeiss LSM 780 FCS confocal microscope	Zeiss	
Olympus FV3000 confocal microscope	Olympus	
Nikon Yokogawa CSU-W1 spinning disc confocal microscope	Nikon	
BD FACS Aria Fusion sorter	BD Biosciences	
10× Genomics Chromium Controller	10× Genomics	
10× Genomics Single Cell 3' Library and Gel Bead Kit	10× Genomics	V2 PN-120237; V3.1; 10× Genomics; PN-1000123
10× Genomics Chromium Single Cell Chip	10× Genomics	A PN-120236; B PN-1000073 or PN-1000074; G PN-1000120; H PN-1000162
Agilent BioAnalyzer 2100	Agilent	
Agilent High Sensitivity DNA Kit	Agilent	Cat # 5067-4626
KAPA Library Quantification Kit - Illumina/Universal	KAPA Biosystems	Cat # KK4824
Life Technologies Viia 7 real time PCR instrument	Life Technologies	
Illumina NextSeq-500	Illumina	
Illumina NovaSeq 6000	Illumina	
10× Genomics Single Cell ATAC Reagent Kit	10× Genomics	V1.1 PN-1000176

## Methods and Protocols

### Zebrafish husbandry

Zebrafish work was conducted in compliance with animal ethics committees at the Peter MacCallum Cancer Centre, the University of Melbourne and the University of Queensland. Published transgenic lines used were as follows: *Tg(fli1a:nEGFP)<sup>v7</sup>* (Lawson & Weinstein, 2002); *Tg(-5.2lyve1b:DsRed)<sup>nz101</sup>* (Okuda *et al*, 2012); *Tg(prox1a:KalTA4,4xUAS-E1B:TagRFP)<sup>nim5</sup>* (Dunworth *et al*, 2014; van Impel *et al*, 2014), *Tg(kdr:EGFP)<sup>s843</sup>* (Beis *et al*, 2005), *Tg(kdr:Has.HRAS-mCherry)<sup>s916</sup>* (Hogan *et al*, 2009), *Tg(10xUAS:Venus)<sup>uq8bh</sup>* (Legendijk *et al*, 2017) and *TgBAC(cd5:GALFF)<sup>mu10164</sup>*. Published mutant lines used were *prox1a<sup>i278</sup>* (van Impel *et al*, 2014) and *prox1b<sup>sa35</sup>* (Tao *et al*, 2011; Koltowska *et al*, 2015a). The *uq53bh* mutant was isolated in a previously described genetic screen (Koltowska *et al*, 2015b). The genetic mapping was performed as previously described (Koltowska *et al*, 2015b) and identified a single region of linkage on Chromosome 5 containing the *notch1b* locus (data not shown). The mutation identified was a nucleotide substitution from C to T early in the coding sequence of *notch1b* resulting in glutamine to stop codon conversion (encoding Q128\*). *Tg(lyve1b:Venus)<sup>uq51bh</sup>* was generated as previously described (Bower *et al*, 2017a) but here using an independent genomic integration with the same construct.

### Generation of maternal zygotic mutants

Germline replacement was performed using embryonic transplantation as described previously (Ciruna *et al*, 2002; Koltowska *et al*, 2015a). Maternal zygotic (MZ) *prox1a* mutant embryos were made by crossing

germline replaced (germline genotype = *prox1a<sup>i278-/-</sup>*; *prox1b<sup>sa003-/-</sup>*) females with *prox1a<sup>i278+/-</sup>*; *prox1b<sup>sa0035+/-</sup>* males (Ciruna *et al*, 2002; Koltowska *et al*, 2015a). Genotyping of individual embryos during transplantation and phenotypic analysis was performed as described previously (Koltowska *et al*, 2015a). Precise genotypes are indicated in Figs 6 and EV5.

### Transgenesis, genome editing and genotyping

All microinjections were performed as described previously (Westerfield & The Zebrafish Book, 2000). *TgBAC(fabp11a:Citrine)<sup>uom103</sup>* and *TgBAC(slc7a7:Citrine)<sup>uom104</sup>* recombineering was performed as described previously (Bussmann & Schulte-Merker, 2011). BAC clones used in recombineering were as follows: *fabp11a<sup>BAC</sup>* CH1073-188H4 and *slc7a7a<sup>BAC</sup>* CH73-370E22.

Primers for BAC recombineering:

*slc7a7a*-BAC-Citrine-forward: 5'-AACTGCTTAGACAGTGTTTTT GGTACCATCCCATATATTTAAAAACAGCCACCATGGTGAGCAA GGGCGAGGAG-3'.

*slc7a7a*-BAC-Citrine-reverse: 5'-TTCGACACCTCAGGGATGCCTCTT CTGCAGGCGTAGGGCTGTAGGACGCTCAGAAGAACTCGTCAAGAA GGCG-3'.

*fabp11a*-BAC-Citrine-forward: 5'-TTACAGCTGTTGCGAGATTGAAA AGTAGAGGAGCATCATTATTCGGGAAAGCCACCATGGTGAGCAAG GCGGAGGAG-3'.

*fabp11a*-BAC-Citrine-reverse: 5'-TCAAAGTTGTCGCTGGTGTCATT TTCCACGTTCTACGAATTTGTCAACTCAGAAGAACTCGTCAAGA AGGCG-3'.

For *Tg(cdh6 enhancer:EGFP)<sup>uom105</sup>*, a 501-bp PCR fragment of *cdh6* enhancer (*chr2:28150709-28151209*) was cloned into the zebrafish enhancer detection (ZED) vector (Bessa *et al*, 2009) using In-Fusion cloning. The ZED vector plasmid was digested with restriction enzymes BspEI and BmgBI. Empty ZED vector was injected as described previously (Bessa *et al*, 2009). Briefly, 1 nl of construct at 40 ng/μl or 45 ng/μl and tol2 transposase mRNA at 100 ng/μl or 55 ng/μl was injected into the one-cell stage wild-type zebrafish embryos. All F0 embryos were screened for skeletal muscle DsRed2 expression. Stable F1 embryos were imaged.

Primers for cloning of *cdh6* enhancer:

*cdh6-forward*: 5' TGAATGCTCATCCGGAGTTTTGTCTGTAAATT CAATTGTAACAATAGTTTACTGT 3'.

*cdh6-reverse*: 5' TATCTGACAGCAGACGTGGAGCACATAAACAGCT CCCC 3'.

For *notch1b<sup>uq53bh</sup>* genotyping, primers for genotyping of *notch1b<sup>uq53bh</sup>* were:

*notch1b-geno-forward*: 5'-AATCCTGTGCC-3'

*notch1b-geno-reverse*: 5'-GTGCCACTCCGT-3'

*notch1b<sup>uq53bh</sup>* genotyping using KASP system (LGC group)

*notch1b-geno-KASP*: 5'-GGCAAACCTGCAAA[C/T]AAGATGTCAAC GAGT-3'

### Imaging and quantification

Imaging of live samples was performed using a Zeiss LSM 710 FCS confocal microscope, a Zeiss LSM 780 FCS confocal microscope, Olympus FV3000 confocal microscope, or a Nikon Yokogawa CSU-W1 spinning disc confocal microscope. Mounting and imaging were performed as described previously (Okuda *et al*, 2018). In Figs 6E and F, I–K, 8J and K, and EV5F–H, K–M quantification of vascular phenotypes was performed as described previously (Bower *et al*, 2017b; Okuda *et al*, 2018). Venous/arterial ISV ratios in Fig 8G and K were manually quantified with Fiji, ImageJ (National Institutes of Health) (Schindelin *et al*, 2012). In Fig 3H–K, *TgBAC(cdh5:GALFF; 10xUAS:Venus)* or *Tg(kdrl:GFP)<sup>s843</sup>* intensity in the TD was measured using Imaris software (Bitplane) and normalised to fluorescence intensity of the DA for *TgBAC(cdh5:GALFF; 10xUAS:Venus)* and to the fluorescence intensity PCV for *Tg(kdrl:GFP)<sup>s843</sup>* in the same embryos.

### Fluorescence-activated cell sorting

For Fig 1, we isolated cells using the following transgenic lines: 40 hpf, *Tg(fli1a:nEGFP)<sup>y7</sup>*; 3,4 and 5 dpf, *Tg(-5.2lyve1b:Venus)<sup>uq47bh</sup>*, *Tg(kdrl:Has.HRAS-mCherry)<sup>s916</sup>*.

For Figs 3 and 4, we isolated cells using the following transgenic lines: 4 dpf, *Tg(fli1a:nEGFP)<sup>y7</sup>*, *Tg(-5.2lyve1b:DsRed)<sup>nz101</sup>*. For Fig 6, we isolated cells using the following transgenic lines: 40 hpf, *Tg(fli1a:nEGFP)<sup>y7</sup>*. To dissociate embryos and obtain single-cell suspensions, we followed published protocols (Kartopawiro *et al*, 2014). Briefly, at the desired developmental stage we devolged embryos by pipetting up and down and rinsing in calcium-free ringers solution, we centrifuged at 376 g for 5' at 4°C and removed supernatant and dissociated the cells by incubating in liberase [2.5 mg/ml] (Cat #5401119001 Sigma-Aldrich) diluted at a 1:35 ratio in DPBS at 28.5°C for approximately 5', homogenising the samples during and

after the incubation. To stop the reaction, we added CaCl<sub>2</sub> to a final concentration of 1–2 mM and FBS to a final concentration of 5–10%. We centrifuged at 376 g for 5' at 4°C and discarded the supernatant, in order to be able to assess live vs. dead cells, we resuspend the cells solution in Zombie Violet TM Viability Dye (Cat# 423113, BioLegend) and incubated for 20' at RT softly rocking, we rinsed the cells by centrifuging and resuspending in DPBS/EDTA, and for snATAC-seq experiments, samples were resuspended in 2% BSA/PBS. Suspension was filtered through a strainer. We used the BD FACS Aria Fusion sorter (BD Biosciences), we based the selection of the desired population on FSC and SCC, live cells were selected based on the Zombie Violet profile and double-positive cells for the desired transgenics were targeted according to the expression profiles of single cells. Double-positive cells were sorted into 300 μl 100% FBS in a cold block and taken immediately to the sequencing facility.

### scRNA-seq library preparation

Library preparation and sequencing was performed at the Institute for Molecular Bioscience Sequencing Facility (University of Queensland) or Peter MacCallum Cancer Centre Genomics Facility. Single-cell suspensions were sorted by FACS and spun down to concentrate, and a cell count was performed to determine postsort viability and cell concentration. Single-cell suspension was partitioned and barcoded using the 10× Genomics Chromium Controller (10× Genomics) and the Single Cell 3' Library and Gel Bead Kit (V2 10× Genomics PN-120237; V3.1; 10× Genomics; PN-1000123). The cells were loaded onto the Chromium Single Cell Chip A (10× Genomics; PN-120236), B (10× Genomics; PN-1000073 or PN-1000074) or G (10× Genomics; PN-1000120) to target 10,000 cells. GEM generation and barcoding, cDNA amplification and library construction were performed according to the 10× Genomics Chromium User Guide. The resulting single-cell transcriptome libraries contained unique sample indices for each sample. The libraries were quantified on the Agilent BioAnalyzer 2100 using the High Sensitivity DNA Kit (Agilent, 5067-4626). Libraries were pooled in equimolar ratios, and the pool was quantified by qPCR using the KAPA Library Quantification Kit—Illumina/Universal (KAPA Biosystems, KK4824) in combination with the Life Technologies Viia 7 real-time PCR instrument. After the initial sequencing run, libraries were repooled according to estimated captured cells as determined using the Cell Ranger software (10× Genomics).

### Sequencing of scRNA-seq libraries

At the IMB(UQ) genomics facility, denatured libraries were loaded onto an Illumina NextSeq-500 and sequenced using a 150-cycle High-Output Kit as follows: 26 bp (Read1), 8 bp (i7 index), 98 bp (Read2). Read1 supplies the cell barcode and UMI, i7 the sample index, and Read2 the 3' sequence of the transcript. At the Peter MacCallum Cancer Centre Molecular Genomics facility, single-cell transcriptome libraries were sequenced on an Illumina NovaSeq 6000 using S4 300-cycle chemistry. Read1 supplies the cell barcode and UMI, i7 the sample index and Read2 the 3' sequence of the transcript. Sequencing read lengths were trimmed to 28 bp (Read1), 8 bp (i7 index), 91 bp (Read2), ensuring compatibility with the 10× Genomics analysis software, Cell Ranger.

### scRNA-seq data processing and analysis

Relevant functions are in italics for reference with default settings applied unless otherwise described, where necessary fastq files were

made using *Cell Ranger* (Zheng et al, 2017; version 3.1.0 or 3.0.2) *mkfastq*. Sequencing QC was assessed using *FastQC* 0.11.6 and *MultiQC* viewer for aggregated reports. *Cell Ranger count* and *aggr* were used to generate aggregated count files mapped to GRCz11 (Ensembl 101), without depth normalisation. Doublets were identified from the filtered aggregated count files using *Scrublet* (Wolock et al, 2019) in *Python* version 3.6 and filtered from subsequent analyses. RNA velocity analyses were run using *Velocyto* (La Manno et al, 2018) for *Python* version 3.6 and trajectory analysis performed using *Slingshot* (Street et al, 2018), with default settings.

*Zprox1a*<sup>-/-</sup> mutant and *MZprox1a/b*<sup>-/-</sup> mutant datasets were analysed using the same workflow: *Cell Ranger count* and *aggr* were used to generate aggregated count files mapped to GRCz11 (Ensembl 101), without depth normalisation, ribosomal and globin genes were filtered, filtered according to library size and mitochondrial content, sc-transform normalised (Hafemeister & Satija, 2019), followed by Uniform Manifold Approximation and Projection (UMAP) dimension reduction, clustering (louvain) and cell cycle analysis using *Seurat* version 3.0 (Stuart et al, 2019) for *R statistical software* version 4.0.2. All thresholds and settings are described in scripting. QC was evaluated before and after normalisation using both custom plots and built-in functions in *Seurat* and *scater* 1.20.1 (McCarthy et al, 2017). Cluster solutions were evaluated and most appropriate resolution selected using *ClusTree* (Zappia & Oshlack, 2018), and cluster identity was defined using key markers of phenotype described in Datasets EV2A and EV5A, respectively.

The atlas of lymphangiogenesis was generated using the following workflow: filtered aggregated count files were processed, ribosomal and globin genes were filtered, filtered according to library size and mitochondrial content, log-normalised, scaled and dimension reduction (PCA and UMAP) calculated using *Seurat* version 3.0 (Stuart et al, 2019). Merged data were clustered and normalised using *CSS simspect* (He et al, 2020) with cluster evaluation and all downstream analyses performed on this object only, as described previously.

All data were visualised at each level of subsetting with *CellXGene* visualisation software (preprint: Li et al, 2020). All DEG analysis and plotting were performed using *Seurat* version 3.0 (Stuart et al, 2019) with default settings unless otherwise described in scripting. All gene ontology analyses were performed using Panther.db (Thomas et al, 2006; Biological Process Complete, *FDR* < 0.05).

### Preparation of single nuclei for snATAC-seq

Single-cell suspensions were sorted by FACS and prepared for nuclei isolation as previously described by 10× Genomics Demonstrated Protocol for Single Cell ATAC Sequencing (CG000169—Rev D). Cell suspensions were pelleted (300 g for 5 min) and rinsed with PBS + 0.04% BSA. Cells were resuspended in 95 µl of freshly prepared lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% Tween-20, 0.1% NP40 Substitute, 0.01% Digitonin and 1% BSA) and incubated on ice for 1 min. 100 µl of chilled wash buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% Tween-20, 1% BSA) was used to neutralise the reaction, before the nuclei were pelleted (500 g for 5 min) and resuspended again in 7 µl of 1× Nuclei Buffer (10× Genomics Cat# PN-2000153/2000207). The presence of healthy and intact nuclei was assessed by visual inspection on a brightfield microscope using Trypan Blue staining (Thermo Fisher Cat# T10282) and Countess Cell Counting Chamber Slides (Thermo Fisher Cat# C10228).

### snATAC-seq library preparation and sequencing

Single-nuclei suspensions were resuspended at approximately 5,000 nuclei per µl before undergoing tagmentation for 60 min at 37°C. After tagmentation, nuclei were partitioned and barcoded using the 10× Genomics Chromium Controller (10× Genomics) and the Single Cell ATAC Reagent Kit (V1.1; 10× Genomics; PN-1000176). Tagmented nuclei were loaded onto the Chromium Single Cell Chip H (10× Genomics; PN-1000162), GEM generation, barcoding and library construction was performed according to the 10× Genomics Chromium User Guide. The resulting single-cell ATAC libraries contained unique sample indices for each sample. The libraries were quantified on the Agilent BioAnalyzer 2100 using the High Sensitivity DNA Kit (Agilent, 5067-4626) and pooled in equimolar ratios. Sequencing was performed on an Illumina NextSeq 500 using a 150-cycle High-Output Kit as follows: 50 bp (Read1), 8 bp (i7 index), 16 bp (i5 index) and 50 bp (Read2) achieving a read depth of 25,000 read pairs per nucleus.

### snATAC-seq processing and analysis

FASTQ files generated from sequencing were used as inputs to 10× Genomics Cell Ranger ATAC 2.0.0. *cellranger-atac count* was used to generate count files mapped to GRCz11 (ENSEMBL 101), without depth normalisation. Resulting fragment files were read into *ArchR* 1.0.1 (Granja et al, 2021) for *R statistical software* 4.0.5 or 4.1.0 as a tile matrix with 500-bp bins. All remaining steps in the ATAC-Seq analysis were performed within *ArchR* 1.0.1. QC filtering was performed, and only high-quality cells with a TSS enrichment score > 4 and > 1,000 unique nuclear fragments were retained. Doublets were predicted using *addDoubletScores* and filtered using *filterDoublets*. Data normalisation and dimensionality reduction were performed using iterative Latent Semantic Indexing (LSI), and UMAP embeddings were used for visualisation in reduced dimension space. Separate from the *ArchR* 1.0.1 package, cluster solutions were independently evaluated using *clustree* (Zappia & Oshlack, 2018) 0.4.3. A gene score matrix that stores predicted gene expression was then generated based on the accessibility of regulatory elements in the vicinity of the gene. We used gene scores of endothelial markers for cluster annotation and subsetting. Local chromatin accessibility of the marker genes was visualised using the *plotBrowserTrack*. Differentially accessible peaks were identified by differential testing using *getMarkerFeatures*. Peaks with false discovery rate (*FDR*) < 0.05 in Fig 4D and peaks with raw *P*-value < 0.05 and log<sub>2</sub> fold change > 1.5 or < -1.5 in Fig 5C were considered as significant DAPs.

Motif enrichment analysis was performed with *Homer* 4.11 using *findMotifsGenome.pl* with default parameters and the GRCz11 genome. Less accessible and more accessible DAPs between LEC WT and VEC WT were used as background in Fig 4G and H, respectively. Default background is used in Fig 5G. Motif scanning was performed to identify putative PROX1 binding sites in DAPs between LEC WT and VEC WT. *FIMO* (Grant et al, 2011) in *MEME Suite* 5.4.1 was used to scan human PROX1 binding motif (MA0794.1, JASPAR 2022) in the DAPs with a *P*-value threshold of 7e-4. Hypergeometric test for over-representation was performed to test the enrichment of PROX1 motif in each of the open and closed DA peak sets. The hypergeometric *P*-value is calculated as the probability of randomly selecting peaks from the entire peak set (*n* = 70,996) and having the same or larger

proportion of peaks with PROX1 motif as in the open ( $n = 1,561$ ) or closed ( $n = 2,624$ ) DAPs.

### Gene regulatory network analyses

For gene regulatory network construction, we used the ENCODE transcription factor targets gene-attribute edge list from the Harmonize database <https://maayanlab.cloud/Harmonizome/dataset/ENCODE+Transcription+Factor+Targets> that includes information for  $n = 181$  transcription factors from ChIP-seq analyses (Rouillard *et al*, 2016). Gene lists generated from scRNA-seq or snATAC-seq analysis were mapped from Zebrafish genome version GRCz11 to Human genome version GRCh38.p13 using ENSEMBL Biomart (Smedley *et al*, 2015). These mapped gene sets were used to select relevant edges that were visualised as a degree-sorted circular network in Cytoscape (Shannon *et al*, 2003).

### Statistics

Statistical analysis for imaging data was performed using GraphPad Prism 9. When comparing two groups, either a two-tailed Student's  $t$ -test (normal distribution) or a two-tailed Mann–Whitney test (non-normally distributed) was used. For comparing multiple groups, one-way analysis of variance (ANOVA) test (normal distribution) or Kruskal–Wallis test was used (non-normally distributed) was used. Normal distribution was assessed by Shapiro–Wilk test.

## Data availability

All code and documentation associated with this analysis is publicly available under an open-source software licence at:

[https://atlassian.petermac.org.au/bitbucket/users/tyrone.chen/repos/hogan\\_lab/browse/2022\\_Grimm\\_Mason\\_et\\_al\\_PROX1/](https://atlassian.petermac.org.au/bitbucket/users/tyrone.chen/repos/hogan_lab/browse/2022_Grimm_Mason_et_al_PROX1/)

All raw and processed data are available under GEO accession ID GSE188342 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE188342>).

All single-cell datasets appearing in this paper can be explored using *CellXGene* (preprint: Li *et al*, 2020) visualisation software:

Developmental atlas of lymphangiogenesis (Fig 1) <http://115.146.95.206:5009/>

Zygotic *prox1a*<sup>-/-</sup> mutant dataset 40 hpf (Fig 3) <http://115.146.95.206:5008/>

Maternal and zygotic *prox1a/b*<sup>-/-</sup> dataset 4 dpf (Fig 7) <http://115.146.95.206:5007/>

**Expanded View** for this article is available [online](#).

## Acknowledgements

This project was supported in part by NHMRC Ideas grant 2004300 and ARC Discovery project DP180102846. BMH was supported by NHMRC Fellowship 1155221. We thank Angelika Christ and the IMB Sequence Facility (University of Queensland, UQ), Tim Semple and Peter MacCallum Cancer Centre (Peter Mac) Genomics Facility, the Research Computing Facility at Peter Mac, FACS facilities at UQ and Peter Mac, Isaac Virshup for assistance and Nathan Lawson for sharing of laboratory protocols. Imaging was conducted at the Centre for Advanced Histology and Microscopy (Peter Mac) and the Australian Cancer Research Foundation Dynamic Imaging Facility (UQ).

## Author contributions

**Lin Grimm:** Formal analysis; investigation; methodology; writing – original draft. **Elizabeth Mason:** Formal analysis; supervision; investigation; methodology; writing – original draft; writing – review and editing. **Hujun Yu:** Investigation; methodology. **Stefanie Dudczig:** Investigation; methodology. **Virginia Panara:** Investigation. **Tyrone Chen:** Data curation; formal analysis. **Neil I Bower:** Formal analysis; investigation; methodology. **Scott Paterson:** Investigation; methodology. **Maria Rondon Galeano:** Investigation; methodology. **Sakurako Kobayashi:** Methodology. **Anne Senabouth:** Formal analysis. **Anne K Lagendijk:** Supervision. **Joseph Powell:** Resources; supervision. **Kelly A Smith:** Supervision. **Kazuhide S Okuda:** Formal analysis; investigation; methodology. **Katarzyna Koltowska:** Supervision; investigation. **Benjamin M Hogan:** Conceptualization; resources; formal analysis; supervision; funding acquisition; project administration.

## Disclosure and competing interests statement

The authors declare that they have no conflict of interest.

## References

- Armour SM, Remsberg JR, Damle M, Sidoli S, Ho WY, Li Z, Garcia BA, Lazar MA (2017) An HDAC3-PROX1 corepressor module acts on HNF4alpha to control hepatic triglycerides. *Nat Commun* 8: 549
- Baek S, Oh TG, Secker G, Sutton DL, Okuda KS, Paterson S, Bower NI, Toubia J, Koltowska K, Capon SJ *et al* (2019) The alternative splicing regulator Nova2 constrains vascular Erk signaling to limit specification of the lymphatic lineage. *Dev Cell* 49: 279–292
- Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, Ginhoux F, Newell EW (2018) Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* 37: 38–44
- Beis D, Bartman T, Jin SW, Scott IC, D'Amico LA, Ober EA, Verkade H, Frantsve J, Field HA, Wehman A *et al* (2005) Genetic and cellular analyses of zebrafish atrioventricular cushion and valve development. *Development* 132: 4193–4204
- Bessa J, Tena JJ, de la Calle-Mustienes E, Fernández-Miñán A, Naranjo S, Fernández A, Montoliu L, Akalin A, Lenhard B, Casares F *et al* (2009) Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev Dyn* 238: 2409–2417
- Bower NI, Koltowska K, Pichol-Thievend C, Virshup I, Paterson S, Lagendijk AK, Wang W, Lindsey BW, Bent SJ, Baek S *et al* (2017a) Mural lymphatic endothelial cells regulate meningeal angiogenesis in the zebrafish. *Nat Neurosci* 20: 774–783
- Bower NI, Vogrin AJ, le Guen L, Chen H, Stacker SA, Achen MG, Hogan BM (2017b) Vegfd modulates both angiogenesis and lymphangiogenesis during zebrafish embryonic development. *Development* 144: 507–518
- Bussmann J, Schulte-Merker S (2011) Rapid BAC selection for tol2-mediated transgenesis in zebrafish. *Development* 138: 4327–4332
- Bussmann J, Bos FL, Urasaki A, Kawakami K, Duckers HJ, Schulte-Merker S (2010) Arteries provide essential guidance cues for lymphatic endothelial cells in the zebrafish trunk. *Development* 137: 2653–2657
- Cha YR, Fujita M, Butler M, Isogai S, Kochhan E, Siekmann AF, Weinstein BM (2012) Chemokine signaling directs trunk lymphatic network formation along the preexisting blood vasculature. *Dev Cell* 22: 824–836
- Cha B, Geng X, Mahamud MR, Zhang JY, Chen L, Kim W, Jho EH, Kim Y, Choi D, Dixon JB *et al* (2018) Complementary Wnt sources regulate lymphatic

- vascular development via PROX1-dependent Wnt/beta-catenin signaling. *Cell Rep* 25: 571–584
- Chen L, Mupo A, Huynh T, Cioffi S, Woods M, Jin C, McKeenan W, Thompson-Snipes LA, Baldini A, Illingworth E (2010) Tbx1 regulates Vegfr3 and is required for lymphatic vessel development. *J Cell Biol* 189: 417–424
- Ciruna B, Weidinger G, Knaut H, Thisse B, Thisse C, Raz E, Schier AF (2002) Production of maternal-zygotic mutant zebrafish by germ-line replacement. *Proc Natl Acad Sci U S A* 99: 14919–14924
- Deng Y, Atri D, Eichmann A, Simons M (2013) Endothelial ERK signaling controls lymphatic fate specification. *J Clin Invest* 123: 1202–1215
- Dieterich LC, Klein S, Mathelier A, Sliwa-Primorac A, Ma Q, Hong YK, Shin JW, Hamada M, Lizio M, Itoh M et al (2015) DeepCAGE Transcriptomics reveal an important role of the transcription factor MAFB in the lymphatic endothelium. *Cell Rep* 13: 1493–1504
- Dunworth WP, Cardona-Costa J, Bozkulak EC, Kim JD, Meadows S, Fischer JC, Wang Y, Cleaver O, Qyang Y, Ober EA et al (2014) Bone morphogenetic protein 2 signaling negatively modulates lymphatic development in vertebrate embryos. *Circ Res* 114: 56–66
- Eng TC, Chen W, Okuda KS, Misa JP, Padberg Y, Crosier KE, Crosier PS, Hall CJ, Schulte-Merker S, Hogan BM et al (2019) Zebrafish facial lymphatics develop through sequential addition of venous and non-venous progenitors. *EMBO Rep* 20: e47079
- Geudens I, Hershers R, Hermans K, Segura I, Ruiz de Almodovar C, Bussmann J, de Smet F, Vandevelde W, Hogan BM, Siekmann A et al (2010) Role of delta-like-4/notch in the formation and wiring of the lymphatic network in zebrafish. *Arterioscler Thromb Vasc Biol* 30: 1695–1702
- Geudens I, Coxam B, Alt S, Gebala V, Vion AC, Meier K, Rosa A, Gerhardt H (2019) Artery-vein specification in the zebrafish trunk is pre-patterned by heterogeneous notch activity and balanced by flow-mediated fine-tuning. *Development* 146: dev181024
- Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY, Greenleaf WJ (2021) ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* 53: 403–411
- Grant CE, Bailey TL, Noble WS (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27: 1017–1018
- Hafemeister S, Satija R (2019) Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* 20: 296
- Hägerling R, Pollmann C, Andreas M, Schmidt C, Nurmi H, Adams RH, Alitalo K, Andresen V, Schulte-Merker S, Kiefer F (2013) A novel multistep mechanism for initial lymphangiogenesis in mouse embryos based on ultramicroscopy. *EMBO J* 32: 629–644
- He Z, Brazovskaja A, Ebert S, Camp JG, Treutlein B (2020) CSS: cluster similarity spectrum integration of single-cell genomics data. *Genome Biol* 21: 224
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38: 576–589
- Hogan BM, Schulte-Merker S (2017) How to plumb a Pisces: understanding vascular development and disease using zebrafish embryos. *Dev Cell* 42: 567–583
- Hogan BM, Bos FL, Bussmann J, Witte M, Chi NC, Duckers HJ, Schulte-Merker S (2009) Ccbe1 is required for embryonic lymphangiogenesis and venous sprouting. *Nat Genet* 41: 396–398
- van Impel A, Zhao Z, Hermkens DMA, Roukens MG, Fischer JC, Peterson-Maduro J, Duckers H, Ober EA, Ingham PW, Schulte-Merker S (2014) Divergence of zebrafish and mouse lymphatic cell fate specification pathways. *Development* 141: 1228–1238
- Johnson NC, Dillard ME, Baluk P, McDonald DM, Harvey NL, Frase SL, Oliver G (2008) Lymphatic endothelial cell identity is reversible and its maintenance requires Prox1 activity. *Genes Dev* 22: 3282–3291
- Jung HM, Castranova D, Swift MR, Pham VN, Venero Galanternik M, Isogai S, Butler MG, Mulligan TS, Weinstein BM (2017) Development of the larval lymphatic system in zebrafish. *Development* 144: 2070–2081
- Kang J, Yoo J, Lee S, Tang W, Aguilar B, Ramu S, Choi I, Otu HH, Shin JW, Dotto GP et al (2010) An exquisite cross-control mechanism among endothelial cell fate regulators directs the plasticity and heterogeneity of lymphatic endothelial cells. *Blood* 116: 140–150
- Kartopawiro J, Bower NI, Karnezis T, Kazenwadel J, Betterman KL, Lesieur E, Koltowska K, Astin J, Crosier P, Vermeren S et al (2014) Arap3 is dysregulated in a mouse model of hypotrichosis-lymphedema-telangiectasia and regulates lymphatic vascular development. *Hum Mol Genet* 23: 1286–1297
- Klotz L, Norman S, Vieira JM, Masters M, Rohling M, Dubé KN, Bollini S, Matsuzaki F, Carr CA, Riley PR (2015) Cardiac lymphatics are heterogeneous in origin and respond to injury. *Nature* 522: 62–67
- Koltowska K, Betterman KL, Harvey NL, Hogan BM (2013) Getting out and about: the emergence and morphogenesis of the vertebrate lymphatic vasculature. *Development* 140: 1857–1870
- Koltowska K, Lagendijk AK, Pichol-Thievend C, Fischer JC, Francois M, Ober EA, Yap AS, Hogan BM (2015a) Vegf regulates Bipotential precursor division and Prox1 expression to promote lymphatic identity in zebrafish. *Cell Rep* 13: 1828–1841
- Koltowska K, Paterson S, Bower NI, Baillie GJ, Lagendijk AK, Astin JW, Chen H, Francois M, Crosier PS, Taft RJ et al (2015b) Mafba is a downstream transcriptional effector of Vegfc signaling essential for embryonic lymphangiogenesis in zebrafish. *Genes Dev* 29: 1618–1630
- Küchler AM, Gjini E, Peterson-Maduro J, Cancilla B, Wolburg H, Schulte-Merker S (2006) Development of the zebrafish lymphatic system requires VEGFC signaling. *Curr Biol* 16: 1244–1248
- Kulkarni RM, Greenberg JM, Akeson AL (2009) NFATc1 regulates lymphatic endothelial development. *Mech Dev* 126: 350–365
- la Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastri ME, Lönnerberg P, Furlan A et al (2018) RNA velocity of single cells. *Nature* 560: 494–498
- Lagendijk AK, Gomez GA, Baek S, Hesselson D, Hughes WE, Paterson S, Conway DE, Belting HG, Affolter M, Smith KA et al (2017) Live imaging molecular changes in junctional tension upon VE-cadherin in zebrafish. *Nat Commun* 8: 1402
- Lawson ND, Weinstein BM (2002) In vivo imaging of embryonic vascular development using transgenic zebrafish. *Dev Biol* 248: 307–318
- van Lessen M, Shibata-Germanos S, van Impel A, Hawkins TA, Rihel J, Schulte-Merker S (2017) Intracellular uptake of macromolecules by brain lymphatic endothelial cells during zebrafish embryonic development. *Elife* 6: e25932
- Li K, Zhengyu O, Chen Y, Gagnon J, Lin D, Mingueneau M, Chen W, Sexton D, Zhang B (2020) Cellxgene VIP unleashes full power of interactive visualization, plotting and analysis of scRNA-seq data in the scale of millions of cells. *bioRxiv* <https://doi.org/10.1101/2020.08.28.270652> [PREPRINT]
- Liu X, Shen J, Xie L, Wei Z, Wong C, Li Y, Zheng X, Li P, Song Y (2020) Mitotic implantation of the transcription factor Prospero via phase separation drives terminal neuronal differentiation. *Dev Cell* 52: 277–293
- Martinez-Corral I, Ulvmar MH, Stanczuk L, Tatin F, Kizhatil K, John SWM, Alitalo K, Ortega S, Makinen T (2015) Nonvenous origin of dermal lymphatic vasculature. *Circ Res* 116: 1649–1654

- McCarthy DJ, Campbell KR, Lun AT, Wills QF (2017) Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* 33: 1179–1186
- Murtomaki A, Uh MK, Choi YK, Kitajewski C, Borisenko V, Kitajewski J, Shawber CJ (2013) Notch1 functions as a negative regulator of lymphatic endothelial cell differentiation in the venous endothelium. *Development* 140: 2365–2376
- Nicenboim J, Malkinson G, Lupo T, Asaf L, Sela Y, Maysel O, Gibbs-Bar L, Senderovich N, Hashimshony T, Shin M et al (2015) Lymphatic vessels arise from specialized angioblasts within a venous niche. *Nature* 522: 56–61
- Norrmen C, Ivanov KI, Cheng J, Zangger N, Delorenzi M, Jaquet M, Miura N, Puolakkainen P, Horsley V, Hu J et al (2009) FOXC2 controls formation and maturation of lymphatic collecting vessels through cooperation with NFATc1. *J Cell Biol* 185: 439–457
- Okuda KS, Astin JW, Misa JP, Flores MV, Crosier KE, Crosier PS (2012) Iyve1 expression reveals novel lymphatic vessels and new mechanisms for lymphatic vessel development in zebrafish. *Development* 139: 2381–2391
- Okuda KS, Baek S, Hogan BM (2018) Visualization and tools for analysis of zebrafish lymphatic development. *Methods Mol Biol* 1846: 55–70
- Oliver G, Srinivasan RS (2010) Endothelial cell plasticity: how to become and remain a lymphatic endothelial cell. *Development* 137: 363–372
- Oliver G, Kipnis J, Randolph GJ, Harvey NL (2020) The lymphatic vasculature in the 21(st) century: novel functional roles in homeostasis and disease. *Cell* 182: 270–296
- Qin J, Gao DM, Jiang QF, Zhou Q, Kong YY, Wang Y, Xie YH (2004) Prospero-related homeobox (Prox1) is a corepressor of human liver receptor homolog-1 and suppresses the transcription of the cholesterol 7 $\alpha$ -hydroxylase gene. *Mol Endocrinol* 18: 2424–2439
- Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, Ma'ayan A (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database (Oxford)* 2016: baw100
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B et al (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9: 676–682
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498–2504
- Shin M, Male I, Beane TJ, Villefranc JA, Kok FO, Zhu LJ, Lawson ND (2016) Vegfc acts through ERK to induce sprouting and differentiation of trunk lymphatic progenitors. *Development* 143: 3785–3795
- Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, Arnaiz O, Awedh MH, Baldock R, Barbiera G et al (2015) The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res* 43: W589–W598
- Song KH, Li T, Chiang JY (2006) A Prospero-related homeodomain protein is a novel co-regulator of hepatocyte nuclear factor 4 $\alpha$  that regulates the cholesterol 7 $\alpha$ -hydroxylase gene. *J Biol Chem* 281: 10081–10088
- Srinivasan RS, Dillard ME, Lagutin OV, Lin FJ, Tsai S, Tsai MJ, Samokhvalov IM, Oliver G (2007) Lineage tracing demonstrates the venous origin of the mammalian lymphatic vasculature. *Genes Dev* 21: 2422–2432
- Srinivasan RS, Geng X, Yang Y, Wang Y, Mukatira S, Studer M, Porto MPR, Lagutin O, Oliver G (2010) The nuclear hormone receptor coup-TFII is required for the initiation and early maintenance of Prox1 expression in lymphatic endothelial cells. *Genes Dev* 24: 696–707
- Srinivasan RS, Escobedo N, Yang Y, Interiano A, Dillard ME, Finkelstein D, Mukatira S, Gil HJ, Nurmi H, Alitalo K et al (2014) The Prox1-Vegfr3 feedback loop maintains the identity and the number of lymphatic endothelial cell progenitors. *Genes Dev* 28: 2175–2187
- Stanczuk L, Martinez-Corral I, Ulvmar MH, Zhang Y, Laviña B, Fruttiger M, Adams RH, Saur D, Betscholtz C, Ortega S et al (2015) cKit lineage Hemogenic endothelium-derived cells contribute to mesenteric lymphatic vessels. *Cell Rep* 10: 1708–1721
- Stone OA, Stainier DYR (2019) Paraxial mesoderm is the major source of lymphatic endothelium. *Dev Cell* 50: 247–255
- Street K, Rizzo D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S (2018) Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* 19: 477
- Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM III, Hao Y, Stoeckius M, Smibert P, Satija R (2019) Comprehensive integration of single-cell data. *Cell* 177: 1888–1902
- Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P et al (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 45: D362–D368
- Takeda Y, Jetten AM (2013) Prospero-related homeobox 1 (Prox1) functions as a novel modulator of retinoic acid-related orphan receptors  $\alpha$ - and  $\gamma$ -mediated transactivation. *Nucleic Acids Res* 41: 6992–7008
- Tao S, Witte M, Bryson-Richardson RJ, Currie PD, Hogan BM, Schulte-Merker S (2011) Zebrafish prox1b mutants develop a lymphatic vasculature, and prox1b does not specifically mark lymphatic endothelial cells. *PLoS One* 6: e28934
- Thomas PD, Kejariwal A, Guo N, Mi H, Campbell MJ, Muruganujan A, Lazareva-Ulitsky B (2006) Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res* 34: W645–W650
- Venero Galanternik M, Castranova D, Gore AV, Blewett NH, Jung HM, Stratman AN, Kirby MR, Iben J, Miller MF, Kawakami K et al (2017) A novel perivascular cell population in the zebrafish brain. *Elife* 6: e24369
- Westerfield M (2000) *The Zebrafish Book. A guide for the laboratory use of zebrafish (Danio rerio)*, 4th edn. Eugene, OR: University of Oregon Press
- Wigle JT, Oliver G (1999) Prox1 function is required for the development of the murine lymphatic system. *Cell* 98: 769–778
- Wigle JT, Harvey N, Detmar M, Lagutina I, Grosveld G, Gunn MD, Jackson DG, Oliver G (2002) An essential role for Prox1 in the induction of the lymphatic endothelial cell phenotype. *EMBO J* 21: 1505–1513
- Wolock SL, Lopez R, Klein AM (2019) Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst* 8: 281–291
- Yang Y, García-Verdugo JM, Soriano-Navarro M, Srinivasan RS, Scallan JP, Singh MK, Epstein JA, Oliver G (2012) Lymphatic endothelial progenitors bud from the cardinal vein and intersomitic vessels in mammalian embryos. *Blood* 120: 2340–2348
- Yaniv K, Isogai S, Castranova D, Dye L, Hitomi J, Weinstein BM (2006) Live imaging of lymphatic development in the zebrafish. *Nat Med* 12: 711–716
- Zappia L, Oshlack A (2018) Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience* 7: giy083
- Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott G, Zhu J et al (2017) Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 8: 14049