RESOURCE ARTICLE

Prox1 dynamically regulates downstream targets and chromatin accessibility during venous to lymphatic endothelial cell transdifferentiation in the embryo

Lin Grimm^{1,2,3,*}, Elizabeth Mason^{1,2*}, Stefanie Dudczig^{1,2}, Jan Kazenwadel^{4,5}, Tyrone Chen^{1,2}, Oliver Yu^{1,2}, Neil I. Bower³, Scott Paterson^{1,2,3}, Kazuhide Okuda^{1,2}, Maria Rondon Galeano^{1,2}, Sakurako Kobayashi^{1,2}, Anne Senabouth^{3,4}, Anne K. Lagendijk³, Joseph Powell^{3,6,7,8}, Kelly A. Smith⁹, Natasha L. Harvey^{4,5}, Katarzyna Koltowska¹⁰ and Benjamin M. Hogan^{1,2,3,9 #}.

¹ Organogenesis and Cancer Program, Peter MacCallum Cancer Centre, Melbourne, VIC 3000, Australia.

² Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, VIC 3000, Australia.

³ Division of Genomics of Development and Disease, Institute for Molecular Bioscience, The University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia.

⁴ Centre for Cancer Biology, University of South Australia and SA Pathology, Adelaide, South Australia, 5001, Australia.

⁵ SA Pathology, Adelaide, South Australia, 5001, Australia.

⁶ Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, Sydney, 2010, Australia.

⁷ School of Medical Sciences, 18 High Street, University of New South Wales, Kensington, Sydney, Australia.

⁸ Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, Sydney, 2010, Australia.

⁹ Department of Anatomy and Physiology, University of Melbourne, Melbourne, VIC 3000, Australia.

¹⁰ Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden.

Author for Correspondence:

Professor Ben Hogan
Organogenesis and Cancer Program,
Peter MacCallum Cancer Centre,
Melbourne, VIC 3000, Australia
E-mail: ben.hogan@petermac.org

Keywords: Lymphatics, Lymphangiogenesis, Prox1, Single cell, RNA Seq, ATAC seq.

Summary: 149 words

Main Text: 4378 words

During development, lymphatic vasculature forms as a second, distinct vascular network derived from blood vessels. Transdifferentiation of venous into lymphatic endothelial cells (LECs) is the first step in this process. Specification, transdifferentiation and maintenance of LEC fate requires the transcription factor Prox1 but downstream mechanisms and markers remain to be described in full. We present a single cell transcriptomic atlas of lymphangiogenesis in zebrafish revealing new markers and hallmarks of LEC differentiation. We further profile single cell transcriptomic and chromatin accessibility changes in Prox1 mutants. This work reveals dynamic, multifaceted roles for Prox1 that include blocking blood vascular and hematopoietic fate and up-regulating essential factors for lymphatic development. We reveal that Prox1 controls chromatin state and evidence for Prox1-independent regulation of early LEC differentiation. This work thus provides new insights and a detailed resource for further understanding an enigmatic developmental process that is central in pathological lymphangiogenesis in disease.

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 ocular disease ¹. In the embryo, lymphangiogenesis begins when the first LEC progenitors depart the cardinal veins (CVs) from E9.5 in mice and 32 hours post fertilisation (hpf) in zebrafish ². Vegfc-Flt4 signalling drives LEC progenitor sprouting but also up-regulates Prox1 expression in both zebrafish and mice ³⁻⁷. The transcription factor Prox1 acts as the master regulator of LEC fate ⁸ 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 ^{4,9}. Following departure from the CV, LEC progenitors go on to colonise embryonic tissues and organs and remodel to form functional lymphatic vessels ¹⁰. While at later stages in mammals there are contributions of LECs from non-venous origins, early embryonic lymphangiogenesis occurs chiefly from the CVs ¹¹⁻¹⁷. 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 ^{9,18}.

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) markers¹⁸. Prox1 is known to autoregulate its own expression and to also regulate Flt4 expression in a positive regulatory loop during early development⁷. Yet how Prox1 controls the transcriptome during LEC specification, differentiation and maintenance has not been described. This is in part because of the technical challenge of accessing early mouse endothelial cells in wildtype 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^{4,6,19,20}, we here took advantage of the accessibility of the zebrafish embryo to examine developmental lymphangiogenesis using single cell transcriptomics. We further analysed zebrafish zygotic *prox1a* mutants with single cell RNA sequencing and single cell ATAC sequencing, as well as maternal-zygotic double *prox1a/prox1b* mutant (null Prox1 mutants) vasculature with single cell RNA sequencing. Overall, we reveal that embryonic lymphangiogenesis is a highly conserved process involving dynamic and discrete Prox1-dependent and independent mechanisms. We provide a resource of new markers of VEC-LEC transdifferentiation, differentiating and mature cell types and we demonstrate the utility of this resource to identify a new Prox1-dependent regulator of lymphangiogenesis, the transmembrane protein Tspan18.

Results

A single cell RNA-seq atlas of embryonic lymphangiogenesis

Zebrafish secondary angiogenesis occurs when Prox1-positive LECs and VECs both sprout from the cardinal vein (CV) in the trunk or the head in a progressive process between ~32 and ~48hpf. In the trunk, sprouting LECs migrate dorsally and invest the horizontal myoseptum, where they form a transient pool of parachordal LECs (PLs) from approximately 48hpf^{21,22}. In craniofacial regions of the embryo, LECs sprout from the CVs at several locations^{23,24}. After

this, LECs throughout the embryo proliferate and migrate extensively (between ~56hpf-80hpf) to colonise new regions and tissues²⁴⁻²⁷. In the trunk, LECs anastomose to form the first lymphatic vessels at around 4 days post fertilisation; forming the thoracic duct (TD), dorsal longitudinal lymphatic vessels (DLLVs) and intersegmental lymphatic vessels (ISLVs). In craniofacial regions, they assemble from disparate sources into lateral (LFL), medial (MFL), otolith 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 mural LEC population in the brain (muLECs, also known as FGPs or brain LECs)²⁸⁻³⁰. By 5dpf, major lymphatics in the craniofacial and trunk regions of the embryo are functional and drain dyes and fluids deposited in the peripheral tissues^{22,31}.

To profile stages of development spanning key steps in lymphatic differentiation, we selected: 40hpf, when specification and initial sprouting of LECs are occurring; 3dpf, when immature LECs are migrating through the embryo; 4dpf when LECs are assembling into vessels; and 5dpf when lymphatics are functional and maturing³² (**Fig. 1a**). We used transgenic zebrafish strains that specifically labelled embryonic vasculature to allow for fluorescence activated cell sorting (FACS, full details in methods). We dissociated whole transgenic embryos and FAC sorted endothelial cells, for profiling on the 10X Chromium scRNA-seq platform. We sequenced 35,634 cells across 3 runs, then merged and normalised the data^{33,34}, filtering low-quality libraries (**Extended data Fig 1a**). To define the cellular identity of each cluster, we systematically evaluated the expression of known markers summarised in **Extended data Table 1a**. We identified 9,771 lymphatic and venous endothelial cells that comprise our single cell atlas of lymphangiogenesis (**Fig 1b, Extended data Fig 1b-c**).

This dataset is displayed³⁵ 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 analysis³⁶. Comparison of cells at 40hpf with later stages revealed that the earlier populations are transcriptionally distinct from VECs and LECs at 3, 4 and 5dpf. Notably, the 40hpf “VEC preLEC” cluster contained both *prox1a+* and *prox1a-* cells forming a single population rather than discrete clusters. This indicates that LECs are not transcriptionally distinct from VECs at 40hpf, suggesting early *prox1a+* cells are specified but not yet differentiated LECs (**Fig 1c**)⁴. We identified 3 main classes of LECs at 3, 4 and 5dpf:

muLECs (1511 cells) marked by expression of *osr2*²⁸, canonical LECs marked by expression of *cdh6* (2669 cells), and a smaller sub-population of LECs expressing low levels of *prox1a* (LEC_low_prox1a, 310 cells) (**Fig 1c, Extended data Fig 1d-e**). To define markers of differentiating canonical LECs at each developmental stage, we applied differential expression (DE) analysis (**Extended data Table 1b**). 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 *grp156*, *hapln3*, *cdh6* and *tspan18a* (**Fig 1d, Extended data Fig 1f**). We extended this approach and evaluated global differences between all canonical LECs and all VECs (n=1240 genes, **Extended data table 1c**). GO analysis³⁷ confirmed the association of biological processes known to be associated with lymphatics with genes upregulated in LEC (n=752 LEC gene set), these terms included “lymphangiogenesis”, “glycolytic process”, “lymph vessel development” and “ameboidal-type cell migration” (**Fig 1e, Extended data Table 1d**).

To provide further confidence in the specificity of the new LEC markers defined by this scRNA-seq resource, we used DE analysis to defined cluster specific gene expression. We noted that this analysis suggested higher expression of the well-known marker *lyve1b* in muLECs than in 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 *slc7a7a* as uniquely expressed in muLEC cells and *fabp11a* as a canonical LEC and VEC marker excluded from muLECs. We generated both new *slc7a7a*-Citrine and a *fabp11a*-Citrine BAC transgenic strains, which confirmed the restricted muLEC and vascular expression patterns and further validated our scSeq atlas of lymphatic differentiation (**Fig 2b-c**). Overall, this atlas therefore identifies n=752 specific markers of lymphangiogenesis, the majority of which are new markers, representing new candidate regulators and spanning four developmental stages.

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

Prox1 is both necessary and sufficient to drive VEC to LEC trans-differentiation in mammals and this is proposed to occur by Prox1 simultaneously initiating LEC fate and repressing the VEC fate program^{8,18}. Prox1 expression is also necessary for the maintenance of LEC identity during development^{18,38,39}. Despite the role for Prox1 being well-studied in mouse models, the transcriptomic program controlled by Prox1 *in vivo* has never been profiled. Zebrafish *prox1a* zygotic and maternal zygotic mutants have been previously described^{4,20}. 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 4dpf^{4,19}, however mutants have a reduction in total LEC numbers throughout the face and trunk⁴. When the maternal contribution of *prox1a* is removed, lymphatic development is more severely impaired⁴. We hypothesised that the zygotic mutants which form lymphatic vessels, likely have abnormal vessel identity in the absence of zygotic Prox1. Thus, we first addressed the question of how Prox1 regulates lymphatic differentiation by taking advantage of this unique model. We applied scRNA-seq to endothelial cells FAC sorted from *Zprox1a*^{-/-} mutants²⁰ and WT sibling zebrafish at 4dpf, during lymphatic vessel assembly (n=8075 cells, **Extended data Fig 2a, Extended data Table 2a**). Cluster analysis revealed 3 populations of LECs (n=2068) and a single population of VECs (n=1051) comprising both mutant and wild type cells, and perhaps most striking, a single population of mutant cells (n=484) transcriptionally similar to VECs, marked by the expression of *aqp1a.1* (**Fig 3a-b, Extended data Figs 2b-e**). All three LEC clusters (LEC, LEC_S1, 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 sub-types with different proliferative potential based on S-phase occupancy and cell cycle marker *mki67* (**Fig 3b, Extended data Fig 2e**). The expression of *prox1a* was lowest in the more proliferative LEC_S1 cells, and highest in the less proliferative LEC_S2 (**Fig 3c**). RNA-velocity analysis suggested a trajectory between the mutant cluster and the LEC cluster. 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 dedifferentiation.

To survey how Prox1 maintains normal LEC differentiation, we performed DE analysis comparing the mutant cluster with WT LEC (**Fig 3f, Extended data Table 2b**). Overall, a much

larger set of genes were downregulated than upregulated in the mutant cluster (n=1034 vs n=294) with almost half of the most downregulated genes (AvLogFC) highly enriched for LEC markers (eg. *tbx1*, *cdh6*, *cldn11b*). Of the upregulated genes, almost 75% were VEC markers (eg. *sox7*, *kdrl*, *cdh5*) and again this was enriched in the most highly upregulated genes (**Fig 3f, Extended data Fig 2f**). Comparing the change in gene expression between wildtype LEC cluster and VEC cluster, with the change between wildtype LEC and mutant cluster, we found striking concordance suggesting the mutant cluster is shifted along a LEC fate to VEC fate trajectory (**Fig 3g**). Overall, there is a simultaneous loss of lymphatic fate and re-acquisition of blood vascular gene expression, consistent with work in mouse and with Prox1 function being highly conserved between vertebrates. To validate these observations, we used confocal imaging of *Zprox1a*^{-/-} mutants and observed upregulation of blood vascular markers *cdh5* (**Fig 3h-i, Extended data Fig 3a,c**) and *kdrl* (**Fig 3j-k, Extended data Fig 3d-g**) in *lyve1b*-positive lymphatics vessels. We saw a coincident reduction in the expression of *lyve1b* in these vessels (**Extended data Fig3b**). Interestingly, the difference in *kdrl* levels (relative intensity) in lymphatics between mutants and wildtype was more significant at 5dpf than 4dpf, suggesting that progressive dedifferentiation may be occurring in the mutant (**Fig 3k**). This single cell profile thus demonstrates that Prox1 alone is sufficient to maintain LEC fate and identifies the gene expression maintained by Prox1 in the differentiating lymphatic vasculature.

Prox1 mutant cells that fail to differentiate, display a unique, immature chromatin signature

At the level of gene expression, Prox1 function is essential for VEC to LEC differentiation, however it is unknown if Prox1 controls chromatin accessibility during this process. Thus, we next profiled *Zprox1a*^{-/-} mutants at 4dpf during vessel assembly using single nuclei (sn) ATAC-seq (n=3,731 nuclei, **Extended data Fig 4a, Extended data Table 3a**). 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). Notably, we found that there was almost no contribution of mutant cells to the LEC clusters (**Fig 4a** lower bar plot), indicating a loss of fate when cell identity is determined at the level of chromatin state (**Extended data Fig 4b-e**). To

validate the utility of this dataset, we assessed concordance between gene expression and chromatin accessibility in WT LECs and WT VECs (**Fig 4c**). DA (differential accessibility) analysis revealed the more accessible regions in the LEC clusters were associated with LEC genes identified in our atlas, and the less accessible regions with VEC genes (**Fig 4c, Extended data table 3b**). These regions were enriched for motifs of lymphatic-specific transcription factors including Mafa, Mafb, and the Tbx family (**Extended data table 3c**). Several markers of LEC and VEC identity demonstrated overall concordance between chromatin accessibility (gene score) and gene expression (**Figure 4d**), including DA regions in *cdh6* (**Fig 4d and e**) and *cdh5* (**Fig 4d and f**). This was consistent with observations in the scRNA-seq data, providing further evidence that the Prox1a deficient cells assembling into fragmented lymphatic vessels are not differentiated LECs.

Given that cells in the mutant specific cluster lacking Prox1 demonstrate loss of lymphatic gene expression and acquisition of blood vascular gene expression, we expected chromatin accessibility to change in an equally coordinated manner. However, analysis of accessibility at individual genes (gene score) for the Mutant Cluster revealed changes inconsistent with a simple fate shift (**Fig 5a-b, Extended data Fig 4f**). Notably, at the level of individual peaks we found that 1726 peaks displayed a striking increase in accessibility compared to WT LEC and 1794 peaks an increase compared with WT VEC (**Fig 5b, Extended data tables 4a-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 5b**). This suggests that some regions of chromatin open up more than usual during LEC or VEC differentiation in the absence of Prox1. Most genes with increased accessibility did not show a concurrent increase in gene expression, and we observed no pattern or trend in expression overall (**Fig 5c**). These more open regions were enriched for motifs of early acting transcription factors involved in embryonic vasculogenesis and haematopoiesis, including Erg, Etv2, Ets1, Fli1 and Spi1/Pu.1 (**Fig 5d, Extended data table 4c**). We take this to suggest that in the absence of Prox1, transcription factors that normally drive earlier transitions during vasculogenesis and haematopoiesis activate a unique developmental chromatin signature. This demonstrates that Prox1 coordinates the correct accessibility of the chromatin while maintaining LEC fate in endothelial cells. Furthermore, 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 genes (**Extended data table 4c**).

Prox1 is primarily a negative regulator of blood and blood vascular fate during the specification of LEC fate and VEC-LEC transdifferentiation

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 LEC⁹. To ask how Prox1 controls the earliest stages of venous to lymphatic transdifferentiation, we first needed to generate a complete Prox1 loss of function model in zebrafish. We generated *prox1a*^{+/-}, *prox1b*^{+/-} double heterozygous animals (lacking both Prox1 homologues) and then used germline transplantation approaches to produce animals carrying a double mutant germline (**Extended data Fig 5a-b**). Genetic crossing of these animals generated embryos that were maternal zygotic (MZ) *MZprox1a*^{-/-}, *MZprox1b*^{-/-} mutant zebrafish completely lacking *prox1a/b* transcript expression or maternal deposition (**Fig 6a-b; Extended data Fig 5c**). 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 4dpf (**Fig 6d-e, Extended data Fig 5f-h**)⁴. Furthermore, *MZprox1a*^{-/-}, *MZprox1b*^{-/-} mutants initially show a delay in the formation of PLs as these cells emigrate the CV to invest the HM, but despite this delay, these cells eventually seed the HM before failing to undertake any further migration (**Fig 6c; Extended data Fig 5i-m**). We saw evidence for genetic interaction between *prox1a* and *prox1b*, suggesting that both can contribute in zebrafish lymphatic development (**Extended data Fig 5**).

To test whether LECs require Prox1a in a cell autonomous manner, we performed embryonic transplantation to generate chimeric embryos, transplanting *MZprox1a*^{-/-} mutant cells into wildtype hosts (**Extended data Fig 6a**). *MZprox1a*^{-/-} mutant cells failed to form lymphatic vessels in otherwise wildtype hosts and we noted mutant PLs stuck in the HM by 5dpf, confirming that Prox1a functions cell autonomously during zebrafish lymphangiogenesis (**Extended data Fig 6b-d**). Taken together, the *MZprox1a*^{-/-}, *MZprox1b*^{-/-} mutant phenotype is more severe than any previously described zebrafish Prox1 mutants and Prox1 is necessary cell autonomously for zebrafish lymphatic development.

Next, we profiled the endothelium of *MZprox1a*^{-/-}, *MZprox1b*^{-/-} mutant zebrafish at 40hpf (when cells are both being specified to LEC fate and sprouting from the CVs) using single cell RNA seq as described above (**Extended data table 5a**). Using cluster analysis and key marker expression, we identified populations of cells that included VECs and also included *prox1a*-positive cells. Based on expression of markers associated with EC sprouting (*mki67*, *pcna*), 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 representing the cardinal vein (PCV n=812 cells; **Fig 6f-i**). Consistent with our observations in **Fig 1**, cells expressing *prox1a* at 40hpf 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. To identify the earliest transcriptional changes controlled by Prox1 in lymphatic development, we used DE to evaluate the global differences between *MZprox1a*^{-/-}, *MZprox1b*^{-/-} and WT in the LEC_VEC_01 and LEC_VEC_02 clusters of sprouting cells (**Fig 6j, Extended data table 5b**). 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 6k**, n=1,027 genes, **Extended data table 5c**), but we saw little evidence of positively regulated genes downstream of Prox1 (just n=27 genes downregulated in *MZprox1a*^{-/-}, *MZprox1b*^{-/-}). This demonstrates that in the earliest stages of LEC specification Prox1 is not actively driving a fate program but rather is acting primarily to down-regulate blood vascular and hematopoietic fate. This profile identifies these earliest Prox1-dependent changes.

Sprouting Prox1-expressing cells show changes in expression that are not controlled by Prox1

During their departure from the CV, zebrafish LECs show differences from sprouting VECs in their proliferation and morphology^{40,41}. While Prox1 specifies and maintains LEC fate, these early changes could potentially be regulated by Prox1-independent mechanisms that have yet to be described. To test this, we first examined the 27 genes that we identified as positively regulated in response to Prox1 function (down in mutants; **Fig 6l**). We found that none of these genes were known regulators of lymphangiogenesis. Next, we compared

wildtype *prox1a*-expressing and non-expressing cells in the 40hpf secondary sprout clusters using DE analysis (**Fig 6m,o; Extended data table 5d**). We found that n=71 genes were more highly expressed in the early specified Prox1+ LECs than in sprouting VECs, however there was essentially no overlap with the genes that were Prox1-dependent (**Fig 6n**). Despite this, the n=71 genes were clearly enriched for known regulators of early lymphatic vascular development. This included the Yap target *ccn2a*, as well as *s1pr1*, *nrp2b*, *rasa1b*, *mafbb* and *cldn11a*, all up-regulated in LECs at this early stage and previously linked to lymphatic development⁴²⁻⁴⁷ (**Fig 6o**). Together, we take this as evidence that Prox1-independent mechanisms exist in early specified LECs and that they control at least several essential regulators and markers of lymphangiogenesis.

Tspan18 homologues are Prox1 targets essential for embryonic lymphangiogenesis

While we find that Prox1 homologues function primarily to negatively regulate gene expression at the earliest stages of lymphangiogenesis, there were still 27 genes positively regulated by Prox1 at 40hpf (**Fig 7a**). One of these genes, *tspan18b* encodes a four-transmembrane domain Tetraspanin protein. Tetraspanins are a highly evolutionarily conserved family of proteins involved in the organisation of membrane microdomains and have been shown to be able to influence cell adhesion, signalling and behaviour⁴⁸. No Tetraspanins have been specifically linked to the process of developmental lymphangiogenesis. We noted that both *tspan18b* and *tspan18a* were down-regulated in *MZprox1a*^{-/-}, *MZprox1b*^{-/-} mutants, although *tspan18a* was expressed at much lower levels than *tspan18b* at this stage (**Fig 7a**). In our 4dpf *Zprox1a*^{-/-} mutant dataset, both paralogues were down-regulated in the mutant cluster compared with mature LECs (**Fig 7b**). Both genes displayed expression in LECs in our time course atlas that was enriched in lymphatic vessels compared with muLECs and *tspan18a* was highly enriched in LECs compared with VECs at 3, 4 and 5dpf (**Fig 7c**). In our sn-ATAC seq dataset at 4dpf, we identified LEC specific peaks of chromatin accessibility that were increased in the absence of Prox1 for *tspan18b* but Prox1-dependent for *tspan18a* (**Fig 7d**). This suggests that Prox1 regulation of Tspan18 paralogues is likely to be direct in the earliest stages of VEC-LEC transdifferentiation, during specification of the lymphatic lineage. To test if this may be correct and to probe evolutionary conservation of this relationship, we performed ChIP seq for PROX1 in human LECs. We found that PROX1 binds directly to intragenic sites and the TSS at the *TSPAN18* locus in human LECs (**Fig 7e**).

Finally, we tested if *tspan18a* and *tspan18b* are functionally relevant targets of Prox1 in zebrafish lymphatic development. We used a CRISPR F0 knockout approach⁴⁹ and confirmed CRISPR targeting by PCR. We carefully quantified LEC numbers in the trunk of *tspan18a/tspan18b* crispants at 5dpf and found a loss of lymphatic vessels and strong reduction in LEC numbers but no change in the number of VECs in venous ISVs, which form concomitant with LECs during secondary sprouting (**Fig 7f**). The reduction in LECs in this setting was not a phenocopy of the more severe phenotype in *MZprox1a^{-/-}*, *MZprox1b^{-/-}* mutants, suggesting that the *tspan18* paralogues likely perform only a part of the job necessary for normal lymphatic development downstream of Prox1 (**Fig 7f**). Overall, the example of *tspan18a* and *tspan18b* demonstrates the utility of this resource of single cell RNA and snATAC seq data for the discovery of highly conserved targets of Prox1 and of functionally relevant regulators of lymphatic development.

Discussion

In this study, we 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 profile the endothelial cells of *Zprox1a* mutants, which form lymphatic structures, but these “lymphatics” dedifferentiate. This identifies the transcriptional code maintained by Prox1 in order to maintain LEC identity and the highly conserved nature of Prox1 function comparing zebrafish with mice. We further probed the *Zprox1a* mutant model using snATAC seq and found that while in mutants the transcriptome reverts to a blood vascular fate in the absence of Prox1, the chromatin displays ectopically open regions and peaks that suggest a more immature state. In particular, in the absence of Prox1 at 4dpf, a large number of opened chromatin regions are enriched for transcription factor motifs for earlier acting vasculogenic and haematopoietic transcription factors. This suggests that regulators of earlier vascular fate become more active and increase chromatin accessibility in the absence of Prox1. This role of Prox1 has not been previously appreciated and this may suggest its combinatorial function with other developmental transcription factors that are dysfunctional in the absence of Prox1.

Additional biological insights from this study come from our analysis of double maternal zygotic *prox1a*, *prox1b* mutants. These mutants are presumed “null” mutants for Prox1 orthologues and they revealed (through scRNA seq) that during its earliest role in vascular development, in LEC fate specification and VEC-LEC transdifferentiation, Prox1 functions 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 manner⁵⁰⁻⁵⁴. We also found that there is significant change in early sprouting *prox1a*-positive LECs that is not regulated by Prox1-function. While this may seem counter-intuitive, many of the genes increased in these cells independent of Prox1 function are already implicated in lymphatic development and we take this to suggest that mechanisms regulated independently of Prox1 homologues exist in the earliest sprouting LECs. The regulators of these factors at this stage remain to be described. Finally, by further characterisation of Prox1 regulation of Tspan18 homologues (a and b), we identify that Prox1 controls lymphangiogenesis through a previously undescribed factor that is likely to act at the LEC membrane. The Tspan18 CRISPR knockout phenotype is more subtle than *MZprox1a*^{-/-}, *MZprox1b*^{-/-} mutants, suggesting that multiple downstream targets are likely at play and orchestrate lymphatic development in response to Prox1 function.

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 will help to understand lymphangiogenesis in contexts beyond the embryo, such as in pathological lymphangiogenesis in metastasis, inflammation and tissue repair. Finally, we suggest that combining zebrafish mutant models, with their inherent ease of cellular accessibility, and the resolution afforded by single cell profiling, will be an approach that yields powerful new insights in many areas of developmental biology in the future.

Acknowledgements

This project was supported in part by NHMRC Ideas grant 2004300 and ARC Discovery project DP180102846. B.M.H. was supported by NHMRC Fellowship 1155221. We thank Angelika

Christ and the IMB Sequence Facility (University of Queensland), Tim Semple and Peter Mac Genomics Facility, the Research Computing Facility at Peter Mac, FACS facilities at UQ and Petermac and Isaac Virshup for assistance.

Author contributions

L.G. and E.M. performed, analysed experiments and co-wrote manuscript. B.M.H, K.K and N.L.H conceptualised experiments, analysed data and co-wrote manuscript. S.D, T.C, O.Y, N.I.B, S.P, K.O, A.S and A.L. performed and analysed experiments. J.P, and K.A.S, provided key unpublished reagents and analysed data.

Conflicts of interest

The authors declare no conflict of interests.

Online Methods

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: *Tg(fli1a:nEGFP)^{y7 55}*; *Tg(-5.2lyve1b:DsRed)^{nz101 24}*; and *Tg(kdrl:Has.HRAS-mCherry)^{s916 21}*. Published mutant lines used were *prox1a^{i278 20}*, *prox1b^{4,19}*. *Tg(lyve1b:Venus)^{uq51bh}* was generated as previously described²⁸ but here using an independent genomic integration with the same construct.

Generation of maternal zygotic mutants

Germline replacement was performed using embryonic transplantation as previously described^{4,56}. Maternal zygotic (MZ) *prox1a* mutant embryos were made by crossing germline replaced *prox1a^{i278-/-}*; *prox1b^{sa0035+/-}* females with *prox1a^{i278+/-}*; *prox1b^{sa0035+/-}* males^{4,56}. Genotyping of individual embryos during transplantation and phenotypic analysis was performed as previously described⁴.

Transgenesis, genome-editing and genotyping

All microinjections were performed as previously described⁵⁷. CRISPR genome editing for *tspan18a/b* was performed as previously described⁵⁸, and all genotyping confirmed using

PCR⁴⁹. *slc7a7a*^{BAC}:*slc7a7a-Citrine*^{uom10} and *fabp11a*^{BAC}:*fabp11a-Citrine*^{uom10} recombineering was performed as previously described⁵⁹.

Guide RNA primers previously described⁴⁹:

tspan18a-T1-5'-CCCAGATTCCCACCCCTAGA-3' (product size 290bp)
tspan18a-T2-5'-GGCGGGATGCTGTTCCCTACT-3' (product size 270bp)
tspan18b-T1-5'-TGAAGCCTCCCCGAGCCCCT-3' (product size 285bp)
tspan18b-T2-5'-TGTGGACCCACAGGATTTA-3' (product size 110bp)

Primers for PCR genotyping:

tspan18a-T1-geno-forward:5'-AGTGCTGTGGGAAAATGAGC-3'
tspan18a-T1-geno-reverse:5'-AAATCCCATACAAACAATAGAAGG-3'
tspan18a-T2-geno-forward:5'-AGTGCTGTGGGAAAATGAGC-3'
tspan18a-T2-geno-reverse:5'-AAATCCCATACAAACAATAGAAGG-3'
tspan18b-T1-geno-forward:5'-AGTGCTTTAGTATGCATGTTTCTGT-3'
tspan18b-T1-geno-reverse:5'-TGCTAGATCAATCGTGGAATGG-3'
tspan18b-T2-geno-forward:5'-GTGGCTCGTTCCTCCTCG-3'
tspan18b-T2-geno-reverse:5'-CAGGATGATGTAGACGCCCG-3'

Primers for BAC recombineering:

slc7a7a-BAC-Citrine-forward:5'-
AACTGCTTTAGACAGTGTTTTTGGTACCATCCCATATATTTAAAAAACAGCCACCATGGTGA
GCAAGGGCGAGGAG-3'
slc7a7a-BAC-Citrine-reverse:5'-
TTCGACACCTCAGGGGATGCCTCTTCTGCAGGCGTAGGGCTGTAGGACGCTCAGAAGAACT
CGTCAAGAAGGCG-3'
fabp11a-BAC-Citrine-forward:5'-
TTACAGCTGTTGCGAGATTGAAAAGTAGAGGAGCATCATTATTCGGGAAAGCCACCATGGT
GAGCAAGGGCGAGGAG-3'
fabp11a-BAC-Citrine-reverse:5'-
TCAAAGTTGTCGCTGGTGGTCATTTCCACGTTCTACGAATTTGTCAACTCAGAAGAACTC
GTCAAGAAGGCG-3'

Imaging and quantification

Imaging was conducted at the Centre for Advanced Histology and Microscopy (Peter

MacCallum Cancer Centre). Imaging of live samples was performed using a Zeiss LSM 710 FCS confocal microscope, a Zeiss LSM 780 FCS confocal microscope, or an Olympus FV3000 confocal microscope. Mounting and imaging were performed as previously described⁶⁰.

In Figs. 2b, 6a-c, and 7f, quantification of vascular phenotypes was performed as previously described^{60,61}. In Figs 3h-k, *cdh5:Venus* or *kdrl:GFP* intensity in the TD was measured using Imaris software (Bitplane) and normalised to fluorescence intensity of the DA for *cdh5* and to the fluorescence intensity PCV for *kdrl* in the same embryos.

Fluorescence activated cell sorting

FACS was performed at the Peter MacCallum Cancer Centre and the University of Queensland. For **Fig 1**, we isolated cells using the following transgenic lines: 40hpf, *Tg(fli1a:nEGFP)^{y7}*; 3,4 and 5dpf, *Tg(-5.2lyve1b:Venus)^{uq47bh}*, *Tg(kdrl:Has.HRAS-mCherry)^{s916}*. For **Fig 3 and Fig 4**, we isolated cells using the following transgenic lines: 4dpf, *Tg(fli1a:nEGFP)^{y7}*, *Tg(-5.2lyve1b:DsRed)^{nz101}*. For **Fig 6**, we isolated cells using the following transgenic lines: 40hpf, *Tg(fli1a:nEGFP)^{y7}*. To dissociate embryos and obtain single cell suspensions, we followed published protocols⁶². Briefly, at the desired developmental stage we deyolked embryos by pipetting up and down and rinsing in calcium free ringers solution, we centrifuged at 2000rpm for 5' at 4°C and remove supernatant and dissociated the cells by incubating in liberase [2.5mg/mL] (Cat #5401119001 Sigma-Aldrich) diluted at a 1:35 ratio in DPBS at 28.5°C for approximately 5', homogenizing the samples during and after the incubation. To stop the reaction, we added CaCl₂ to a final concentration of 1-2mM and FBS to a final concentration of 5-10%. We centrifuged at 2000rpm for 5' at 4°C. discarded the supernatant, in order to be able to asses live vs. dead cells, we re-suspend 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, for ATAC-seq experiments samples were resuspended in 2%BSA/PBS. Suspension was filtered through a strainer and taken to the FAC sorting facility. In the Flow Cytometry facility, we used the BD FACS Aria Fusion sorter (BD Biosciences), we based the selection for the desired population on FSC and SCC, alive 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 in 300uL 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 Mac Genomics Facility. Single cell suspensions were sorted by FACS, spun down to concentrate and a cell count was performed to determine post-sort viability and cell concentration. Single cell suspension was partitioned and barcoded using the 10X Genomics Chromium Controller (10X Genomics) and the Single Cell 3' Library and Gel Bead Kit (V2 10X Genomics PN-120237; V3.1; 10X Genomics; PN-1000123). The cells were loaded onto the Chromium Single Cell Chip A (10X Genomics; PN-120236), B (10X Genomics; PN-1000073 or PN-1000074) or G (10X Genomics; PN-1000120) to target 10,000 cells. GEM generation and barcoding, cDNA amplification, and library construction was performed according to the 10X 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 re-pooled according to estimated captured cells as determined using the Cell Ranger software (10X 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: 26bp (Read1), 8bp (i7 index), 98bp (Read2). Read1 supplies the cell barcode and UMI, i7 the sample index, and Read2 the 3' sequence of the transcript. At the Peter Mac 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 28bp (Read1), 8bp (i7 index), 91bp (Read2), ensuring compatibility with the 10X Genomics analysis software, Cell Ranger.

scRNA-seq data processing and analysis

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

https://atlassian.petermac.org.au/bitbucket/users/tyrone.chen/repos/hogan_lab/browse/2022_Grimm_Mason_et_al_PROX1/

Relevant functions are in italics for reference. Where necessary fastq files were made using *Cell Ranger*⁶³ (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*⁶⁴ in *Python* version 3.6 and filtered from subsequent analyses. For the *MZprox1*^{-/-} mutant and *Zprox1*^{-/-} mutant datasets filtered aggregated count files were processed, sc-transform normalised, filtered and clustered (louvain) using *Seurat* version 2.0⁶⁵ and 3.0³⁴ respectively for *R statistical software* version 4.0.2. QC was evaluated before and after normalisation using plot functions in *Seurat* and *scater* 1.20.1⁶⁶, and all thresholds and settings are described in scripting. Cluster solutions were evaluated using *ClusTree*⁶⁷. Datasets used in the atlas of lymphangiogenesis were processed, filtered, merged and log-normalised using *Seurat* version 3.0³⁴, with QC and settings as above. Merged data was clustered and normalised using *CSS simspect*³³ and clustering and cluster evaluation performed on this object only, as described above. For all scRNA-seq datasets cluster phenotype was determined using the expression of key markers (Supplementary Table), with the aid of *CellXGene* visualisation software⁶⁸. All downstream analysis and plotting were performed using *Seurat* version 3.0³⁴ with default settings. All gene ontology analyses were performed using Panther.db³⁷ (Biological Process Complete).

Preparation of single nuclei for snATAC-seq

Single cell suspensions were sorted by FACS and prepared for nuclei isolation as previously described by 10x Genomics Demonstrated Protocol for Single Cell ATAC Sequencing (CG000169 - Rev D). Cell suspensions were pelleted (300 x g for 5 minutes) and rinsed with PBS + 0.04% BSA. Cells were resuspended in 95uL of freshly prepared lysis buffer (10mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% Tween-20, 0.1% NP40 Substitute, 0.01%

Digitonin, and 1% BSA) and incubated on ice for 1 minute. 100uL of chilled wash buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% Tween-20, 1% BSA) was used to neutralise the reaction, before the nuclei were pelleted (500 x g for 5 minutes) and resuspended again in 7uL of 1x Nuclei Buffer (10X Genomics Cat# PN-2000153/2000207). 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 5000 nuclei per μ L before undergoing tagmentation for 60min at 37°C. After tagmentation nuclei were partitioned and barcoded using the 10X Genomics Chromium Controller (10X Genomics) and the Single Cell ATAC Reagent Kit (V1.1; 10X Genomics; PN-1000176). Tagmented nuclei were loaded onto the Chromium Single Cell Chip H (10X Genomics; PN-1000162), GEM generation, barcoding and library construction was performed according to the 10X 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: 50bp (Read1), 8bp (i7 index), 16bp (i5 index), 50bp (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 10X 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* for *R statistical software 4.0.5* 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 greater than 4 and greater than 1,000 unique nuclear fragments were retained. Doublets were predicted using *addDoubletScores* and filtered using *filterDoublets*. Data normalization and dimensionality reduction were performed using iterative Latent Semantic Indexing (LSI) and Uniform Manifold

Approximation and Projection (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 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. Differentially accessible genes were identified by differential testing using *getMarkerFeatures*. Local chromatin accessibility of the marker genes was visualised using the *plotBrowserTrack*.

ChIP-seq library preparation and analysis

For ChIP, antibodies used were goat anti-Prox1 (R&D Systems, AF2727), rabbit anti-H3K4Me1 (Cell Signaling, #5326), rabbit anti-H3K27Me3 (Cell Signaling, #9733), rabbit anti-H3K27Ac (Cell Signaling, #8173), rabbit anti-H3K4Me3 (Cell Signaling, #9751). hLEC were harvested and processed for ChIP as previously described⁶⁹ using a truChIP™ Low Cell Chromatin Shearing Kit with SDS Shearing Buffer (Covaris). Briefly, 10 million cells/ml were cross-linked using 1% formaldehyde for 5 minutes, neutralised with glycine, lysed and nuclei washed. For transcription factor ChIP, chromatin from 3 million cells was sheared using a 130µl microtube in a Covaris sonicator at the recommended settings for 8 minutes. 5µg antibody was used to immunoprecipitate sheared DNA. For histone mark ChIP, chromatin from 1 million cells was sheared and 1µg antibody was used. Following washing and reversal of crosslinks, DNA was purified using a Qiagen MinElute PCR Purification Kit. ChIP-Seq library preparation and sequencing was carried out at the ACRF Cancer Genomics Facility, Centre for Cancer Biology, Adelaide. ChIP-Seq libraries were prepared according to Illumina's TruSeq Sample Prep Guide Revision A with some minor modifications as previously described⁶⁹. Briefly, DNA was end-repaired, followed by adenylation of the 3' ends. Next, Illumina indexing adapters were ligated to the DNA. 18 cycles of PCR were performed to enrich for successfully adapter ligated molecules. Ligation products were purified with a 2% Pippin Prep gel (Sage Science), selecting a size range of 250 – 300 base pairs. The size distribution and yield of the purified libraries were determined using an Agilent Bioanalyzer High Sensitivity chip and Qubit dsDNA HS assay (Invitrogen), respectively. Libraries were pooled in equimolar ratios and sequenced in one lane of a HiSeq 2500 short read flowcell (1 x 50bp). Reads were mapped to the UCSC hg19 genome with BWA⁷⁰ version 0.7.9a-r786, allowing at most 3 alignments. Read depth was

646 calculated via *genomeCoverageBed* from BEDTools⁷¹, using the scale parameter to normalise
647 to reads-per-million. Read depth was plotted relative to annotated transcription start sites
648 via custom Python scripts utilising Matplotlib.

649

650

651

FIGURES AND FIGURE LEGENDS

Figure 1: A single cell RNAseq developmental atlas of lymphangiogenesis in zebrafish

a. Schematic representation of four key stages of zebrafish lymphatic development in head and trunk: 40hpf encompasses both specification (Prox1-induction) and sprouting; 3dpf migration of LECs; 4dpf assembly of lymphatic vessels; 5dpf maturation functional lymphatics.

- b. UMAP visualization of 9771 cells filtered for VEC and LEC populations (n=6 samples; see **Extended data table 1a** for cluster identification and **Extended data fig 1** for whole dataset) coloured according to developmental stage (top), predicted cell phenotype (middle), and RNA-velocity (bottom).

c. UMAP visualization of key marker gene expression. Colour scale represents log-normalised expression. LEC markers: *prox1a*, *cdh6*. BEC marker: *cdh5*. LEC and VEC markers: *mrc1a*, *flt4*. AEC marker: *flt1*. muLEC markers: *osr2*, *hexa*.

d. Dot plot of top genes commonly marking LECs at 3, 4 and 5dpf, with expression displayed in VECs (all stages), VEC_preLEC (40hpf) and LECs at 3, 4 and 5dpf (see also **Extended data table 1b**, **Extended data figure 1f**). Colour scale represents average log-normalised expression and point size represents percentage of cells expressing gene.

e. Bar plot summarizing GO BP analyses of genes DE between 3, 4 and 5dpf LEC and VEC populations (**Extended data table 1c-d**). Y-axis represents enriched BP term, x-axis represents the $-\log_{10}(\text{adjusted } p \text{ value})$, bars are coloured and ordered according to fold enrichment of the GO term in LEC. GO terms enriched in LEC are coloured purple, terms enriched in VEC are coloured red.

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 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 6dpf showing high *lyve1b* expression in craniofacial (middle) and moderate expression in trunk lymphatic vessels (right).
- b. UMAP of *slc7a7a* 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 *slc7a7* (green) in the zebrafish head at 5dpf (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 5dpf.

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, intersomitic lymphatic vessel. Scale bars, 80 μ m for head (middle) in (a) and 50 μ m for trunk images (right).

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

- UMAP visualization of 8075 cells filtered for VEC and LEC subpopulations (n=2 samples; see **Extended data table 2a** for cluster annotation, **Extended data Fig3** for full dataset) coloured according to predicted cell phenotype.
- UMAP visualization of marker gene expression. Colour scale represents SCT-normalised expression. LEC markers: *prox1a*, *cldn11b*, *cdh6*, *angpt2a*. LEC and VEC markers: *flt4*, *mrc1a*, *lyve1b*. BEC markers: *cdh5*, *kdrl*. Mutant cluster markers: *frem1b*, *aqp1a.1*.
- UMAP visualization with cells coloured according to genotype (*Zprox1a*^{-/-} or WT) identifies a mutant specific cluster and RNA velocity analysis suggests a trajectory between the mutant cluster and the LEC cluster, likely indicative of a fate shift.
- 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 (Wilcoxin 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 up-regulated (n=365) and down regulated (n=1922) 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=2287) 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, 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 4dpf.
- i. Quantification of *cdh5* intensity in the thoracic duct in WT and mutants (relative to expression in the DA).
- j. Lateral confocal images of *kdrl* (green) and *lyve1b* (magenta) expression at 4dpf (left) and 5dpf (right).
- k. Quantification of *kdrl* intensity in the thoracic duct in WT and mutants (relative to expression in the PCV) at 4 and 5dpf.

WT, wildtype. Z, zygotic. TD, thoracic duct, DA, Dorsal Aorta. PCV posterior cardinal vein. ISLV, intersomitic lymphatic vessel. DLLV, Dorsal longitudinal lymphatic vessel. PL, parachordal LEC. Scale bars, 50 μ m. *, The error bars represent mean $\pm$ s.e.m.; *p=0.01 and **p=0.0094 from an unpaired, two-sided t-test.

Figure 4: Single nuclei ATAC seq reveals chromatin accessibility changes between venous and lymphatic endothelial cells

a. UMAP visualization of snATAC-seq from $Zprox1a^{-/-}$ mutant and WT endothelial cells at 4dpf (n=3731 nuclei), coloured by cell phenotype (top) and genotype (middle) (see

Extended data Fig4 for full dataset). Bar plot summarising the composition of wildtype and mutants in each cluster. A mutant specific cluster is identified, similar to **Fig 3**.

b. UMAPs visualization indicating accessibility of key marker genes (gene score) that confirm predicted cluster phenotypes.

c. Bar plot of log2 fold change for the differentially accessible peaks (n=4185 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. 228/1561 more accessible peaks in the LEC cluster were associated with LEC markers and 398/2624 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.

d. Dot plots of scRNA-seq (top) and snATAC-seq accessibility (bottom) data summarising the behaviour of key LEC (*prox1a*, *cdh6*, *tbx1*) and VEC (*cdh5*, *flt1*, *gata6*) markers in WT LECs, WT VECs and the Mutant Cluster at 4dpf. 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. The size of the dots represents the proportion of cells or nuclei, and colour represents either SCT-normalised expression or Gene Score of Accessibility.

e. Genome accessibility track of LEC marker *cdh6* (left) and VEC marker *cdh5* (right) indicating these genes are more accessible in WT LEC and WT VEC cell phenotypes respectively. Red bars represent peaks in the reproducible peak set from snATAC-seq. DAP between WT LECs and WT VECs (Wilcoxon Rank Sum, $FDR < 0.05$) were highlighted in blue.

DAP, differentially accessible peaks. FDR, false discovery rate. snATAC-seq, single nuclei ATAC-seq.

Figure 5: Zygotic *prox1a* mutants display a unique chromatin accessibility state in fate-shifted endothelial cells

a. Heatmaps of accessibility (gene score) for all genes (n=32 020 genes) in WT LECs, WT VECs and Mutant Cluster showing that mutant cluster cells display a unique chromatin state at many genes.

- b.** Schematic illustrates DAP analyses between Mutant Cluster vs. LEC WT and Mutant Cluster vs. VEC WT. Venn diagram indicates all individual peaks that were found to be increased in accessibility in the Mutant Cluster vs LEC or VEC with n=431 DAP commonly more open than in both LEC WT and VEC WT (FDR < 0.05, log2 fold change > 1.5).
- c.** Dot plot summarizing the scRNA-seq expression level of n=13 genes with DAP more open in the Mutant Cluster at 4dpf, 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.
- d.** 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. Percentages on the right-hand side represent the proportion of regions enriched for the motif in the n=431 peak set, and the whole genome (background).
- e.** Genome accessibility track of *prox1a*. Red bars represent peaks in the reproducible peak set from snATAC-seq.
- f.** Genome accessibility track of *tbx1*. Red bars represent peaks in the reproducible peak set from snATAC-seq
- g.** Genome accessibility track of *kdr*. Red bars represent peaks in the reproducible peak set from snATAC-seq

DAP, differentially accessible peaks. FDR, false discovery rate. snATAC-seq, single nuclei ATAC-seq.

Figure 6: The role of Prox1 in LEC specification and sprouting revealed by single cell profiling of maternal zygotic *prox1a*;*prox1b* double mutants

- a. Lateral confocal images of zebrafish heads at 4dpf showing endothelial cell nuclei (green) and venous and lymphatic vessels (white) in control (upper) and *MZprox1a*^{-/-}; *MZprox1b*^{-/-} embryos (lower). Facial lymphatics are absent or shorter or missing (asterisk) in the mutants (asterisk). Scale bar, 80 μm.
- b. Lateral confocal images of zebrafish trunks at 4dpf in control (upper) and *MZprox1a*^{-/-}; *MZprox1b*^{-/-} mutants (lower), showing absent lymphatic vessels (asterisk) but retained PLs (arrows in lower). Scale bars, 50 μm.
- c. Lateral confocal images of PLs in the horizontal myoseptum at 2, 3 and 4dpf. PLs form later in *MZprox1a*^{-/-}; *MZprox1b*^{-/-} mutants and accumulate in the horizontal myoseptum while the PLs of control embryos emigrate the HM by 4dpf. Scale bars, 50 μm.
- d. Quantification of lymphatic endothelial cell (LEC) number from 4dpf heads control: n=19, *MZprox1a*^{-/-}; *MZprox1b*^{-/-} mutants: n=18) and trunks (control: n=29, *MZprox1a*^{-/-}; *MZprox1b*^{-/-} mutants: n=25). Error bars represent mean ± s.e.m.; **** p < 0.0001, from an unpaired, two-sided t-test.
- e. Quantification of number of PLs per somite in trunks at 2dpf in controls (n=41) and *MZprox1a*^{-/-}; *MZprox1b*^{-/-} mutants (n=38), at 3dpf in controls (n=11) and *MZprox1a*^{-/-}; *MZprox1b*^{-/-} mutants (n=10) and at 4dpf in controls (n=29) and *MZprox1a*^{-/-}; *MZprox1b*^{-/-} mutants (n=25). Error bars represent mean ± s.e.m.; **** p < 0.0001, * p < 0.05, from an unpaired, two-sided t-test.
- f. UMAP visualization of filtered scRNA-seq data showing 2202 endothelial cells (n=4 samples; **Extended data table 6a**) coloured according to predicted cell phenotype (top), genotype (middle), and RNA-velocity (bottom).
- g. UMAP visualization of key marker gene expression. Colour scale represents SCT-normalised expression.
- h. Stacked bar plot representing the genotype composition of cells in each cluster.
- i. Dot plot of key marker expression. Colour scale represents average SCT-normalised expression and point size represents percentage of cells expressing gene.
- j. Bar plot indicating average log fold change of n=1054 significantly different genes between the *MZprox1a/b*^{-/-} and WT cells in LEC_VEC_01 and LEC_VEC_02 clusters at 40hpf (Wilcoxin Rank Sum *adjusted p value* < 0.05).

- k. Bar plot summarizing GO BP analyses of n=1027 genes upregulated in *MZprox1a/b^{-/-}*. Y-axis represents enriched BP term, x-axis represents the $-\log_{10}(\text{adjusted } p \text{ value})$, bars are coloured according to fold enrichment.
- l. Dot plot of n=27 genes downregulated in 40hpf *MZprox1a/b^{-/-}* and WT cells in LEC_VEC_01 and LEC_VEC_02 clusters.
- m. Schematic outlining the selection of n=109 *prox1a* expressing cells from n=915 *prox1a* negative counterparts in the 40hpf WT sample (whole dataset in **Fig1**). DEG analysis revealed n=71 genes upregulated in *prox1a* positive cells (Wilcoxin Rank Sum *adjusted p value* < 0.05).
- n. Venn diagram illustrating genes upregulated in *prox1a* positive (vs negative) WT cells at 40hpf, and downregulated in *MZprox1a/b^{-/-}* mutant cells at 40hpf.
- o. Dot plot of n=71 genes upregulated in WT *prox1a* positive (vs negative) cells at 40hpf. Colour scale represents average SCT-normalised expression and point size represents percentage of cells expressing gene. Bold labels indicate known lymphatic genes.

LL, lymphatic loop. OLV, otholitic lymphatic vessel. LFL, lateral facial lymphatic vessel. MFL, medial facial lymphatic vessel. LBA, lymphatic branchial arches. DLLV, dorsal longitudinal lymphatic vessel. PL, parachordal LEC. ISLV, intersomitic lymphatic vessel. TD, thoracic duct. DA, dorsal aorta. PCV, posterior cardinal vein.

Figure 7: Identification of Tspan18 paralogues as functionally essential downstream targets of Prox1 in lymphangiogenesis

- a. Dot plot of expression of *tspan18b* and *tspan18a* in WT and *MZprox1a/b*^{-/-} mutant cells at 40hpf. The size of the dot represents the proportion of cells that express the markers in the cluster, and colour represents SCT-normalised expression.
- b. Dot plot of expression of *tspan18b* and *tspan18a* in Wt and *Zprox1a*^{-/-} mutant cells at 4dpf. The size of the dot represents the proportion of cells that express the markers in the cluster, and colour represents SCT-normalised expression.
- c. UMAP visualization of *tspan18b* and *tspan18a* expression in the single cell atlas of lymphangiogenesis. Colour scale represents log-normalised expression.
- d. Genome accessibility track of *tspan18b* (left) and *tspan18a* (right). Red bars represent peaks in the reproducible peak set from snATAC-seq. Differentially accessible peaks between Mutant Cluster and WT LECs (*p* value < 0.05) were highlighted in blue
- e. ChIP-seq profile of Prox1 binding together with H3K27Me3, H3K4Me1, H3K27AC, and H3K4Me3 profiles at the TSPAN18 locus (hg19) in human LECs.
- f. Lateral confocal images of zebrafish trunks depicting mature lymphatic vessels in the uninjected control which are missing in *tspan18a;b* CRISPR injected embryos at 5dpf (asterisks) identifying a functional requirement for *tspan18* paralogues in lymphangiogenesis. Note that the formation of intersegmental vessels is also severely impaired in *tspan18a;b* crispants (arrowheads).. Scale bars, 50 µm. Quantification of LEC numbers across 4 somites and VEC numbers per vISV at 5dpf. Error bars represent mean ± s.e.m.; ** *p* = 0.0012, ns = not significant from an unpaired, two-sided t-test.

DLLV, dorsal longitudinal lymphatic vessel. ISLV, intersomitic lymphatic vessel. TD, thoracic duct. PCV, posterior cardinal vein. vISV, venous intersegmental vessel.

EXTENDED DATA FIGURES AND FIGURE LEGENDS

Extended data Fig 1: Full dataset and analysis for data in Fig 1.

- UMAP visualization of n=35634 endothelial cells sequenced for the single cell atlas of lymphangiogenesis (n=6 samples; see also **Extended data table 1a** for markers used to identify clusters) coloured according to developmental stage (top left), cell cycle phase (top right), predicted cell phenotype (bottom left), and cluster solution (RNA-snn-res.1).
- Clustree analysis demonstrating the relationship between different cluster resolutions for the n=9771 venous and lymphatic endothelial cells in Figure 1a. Resolution 0.4 (boxed) was used for all downstream analyses.
- UMAP visualization of cluster resolution 0.4 (as appears in **Fig 1**).
- Stacked bar plot of resolution 0.4 cluster composition cell cycle phase.
- Stacked bar plot of resolution 0.4 cluster composition developmental stage.

902 **f.** Venn diagram of significant genes commonly upregulated in LEC compared to VEC at
903 3, 4 and 5dpf.
904

Extended data Fig 2: Full dataset and analysis for data in Fig 3.

a. UMAP visualization of n=8075 endothelial cells sequenced comparing lymphangiogenesis in WT and *Zprox1a* mutants at 4dpf (n=2 samples; see also **Extended data table 2a** for markers used to identify clusters) coloured according to developmental stage (top left), cell cycle phase (top right), predicted cell phenotype (bottom left), and cluster solution (SCT-snn-res.8)

- b.** Clustree analysis demonstrating the relationship between different cluster resolutions for the n=3063 venous and lymphatic endothelial cells in **Figure 3a**. Resolution 0.2 (boxed) was used for all downstream analyses.
- c.** UMAP visualization of cluster resolution 0.2
- d.** Stacked bar plot of resolution 0.2 cluster composition genotype
- e.** Stacked bar plot of resolution 0.2 cluster composition cell cycle phase
- f.** Venn diagram indicates the overlap between genes significantly upregulated in VEC compared to LEC at each developmental stage of the single cell atlas, and those upregulated in the *Zprox1a*^{-/-} Mutant Cluster at 4dpf compared to WT LECs. N=85 genes (coloured red) are uniquely associated with the Mutant Cluster and WT VEC at 4dpf, compared to N=4 genes in WT VEC at 3 and 5dpf.

Extended data Fig 3: Additional phenotypic analysis of blood vascular markers in *Zprox1a*^{-/-} mutants associated with Fig 3

- Lateral confocal images of zebrafish trunks at 4dpf showing *cdh5* expression (green) and *lyve1b* expression in lymphatic vessels (magenta) in control (upper) and *Zprox1a* mutant embryos (lower). Transgenic lines are indicated. Scale bar, 50 μ m.
- Quantification of *lyve1b* expression intensity in the TD relative to the PCV in sibling and *Zprox1a* mutant embryos at 4dpf.

- c. Quantification of *cdh5* expression intensity in the TD relative to the DA in sibling and *Zprox1a* mutant embryos at 4dpf.
 - d. Lateral confocal images of zebrafish trunks at 4dpf showing *kdrl* expression (green) and *lyve1b* expression in lymphatic vessels (magenta) in control (upper) and *Zprox1a* mutants (lower). Transgenic lines are indicated. Scale bars, 50 μ m.
 - e. Quantification of *kdrl* expression intensity in the TD relative to the DA in sibling and *Zprox1a* mutant embryos at 4dpf and 5dpf (upper).
 - f. Lateral confocal images of zebrafish trunks at 5dpf showing *kdrl* expression (green) and *lyve1b* expression in lymphatic vessels (magenta) in control (upper) and *Zprox1a* mutants (lower). Transgenic lines are indicated. Scale bars, 50 μ m.
 - g. Quantification of the number of TD segments displaying *kdrl* expression in *Zprox1a* mutant embryos at 4dpf and 5dpf (lower).
- DLLV, dorsal longitudinal lymphatic vessel. ISLV, intersomitic lymphatic vessel. TD, thoracic duct. DA, Dorsal aorta. PCV, posterior cardinal vein. vISV, venous intersegmental vessel. PL, Parachoral lymphatic endothelial cells.

Extended data Fig 4: Additional data and analysis for data in Fig 4.

a. UMAP visualization of n=3731 single nuclei ATAC-seq data comparing lymphangiogenesis in WT and *Zprox1a* mutants at 4dpf. Nuclei coloured according to cluster (res-0.8).

- b.** UMAP visualisation with nuclei coloured according to predicted phenotype based on accessibility changes at the genes defined in **Extended data table 1c**.
- c.** UMAP visualisation with nuclei coloured according to genotype.
- d.** Stacked bar plot indicating the genotype composition of nuclei captured per cell type.
- e.** Density plot indicating the majority of snATAC-seq peaks fall around the TSS. Y-axis represents distance from TSS (bp) and y-axis represents normalised Tn5 insertion profile.
- f.** Heatmaps of accessibility (gene score) for DEG (**Fig 3f**) upregulated (left, n=300 mapped genes) and downregulated (right, n=1052 mapped genes) in the Mutant Cluster compared to WT LECs. Colour indicates row-wise Z-score, each row represents a differentially expressed gene in scRNA-seq data, columns are snATAC-seq clusters. Boxed genes downregulated in Mutant display more permissive chromatin in Mutant than WT LEC, showing that mutant cluster cells display a unique chromatin state at many genes.

Extended data Fig 5: Additional phenotypic analysis of *MZprox1a*^{-/-};*MZprox1b*^{-/-} mutants associated with Fig 6

- a. Schematic explanation of the germline replacement transplantation method used to generate *MZprox1a*^{-/-};*MZprox1b*^{-/-} mutants.
- b. Overall morphology of control and *MZprox1a*^{-/-};*MZprox1b*^{-/-} mutants at 5dpf. Arrows indicate oedema around eyes, heart and intestine. Scale bars 500 µm.
- c-d. Lateral confocal images showing vascular defects in control and mutant in the craniofacial (c) and trunk (d) regions of the larvae at 4dpf. Genotypes and transgenic labels are indicated. *MZprox1a*^{-/-} and *MZprox1a*^{+/-}, *MZprox1b*^{-/-} animals both showed a strong loss of lymphatic vessels in craniofacial and trunk regions. The *MZprox1a*^{-/-}, *MZprox1a*^{+/-} mutants showed a robust loss of craniofacial vessels equivalent to the other mutant phenotypes but a more severe loss of trunk lymphatics and an accumulation of PLs in the HM.
- e. Genotypes that are shown in the quantification. Colour codes are used to identify individual embryonic genotypes in the dot plots showing quantification in f-h.
- f. Quantification of the number of craniofacial (head) LECs within facial lymphatic vessels showing evidence for contributions from *prox1a* and *prox1b*.
- g. Quantification of the number of mural LECs (muLECs) showing a major role for *prox1a* and a role for *prox1b* in muLEC development.
- h. Quantification of the number of trunk LECs within vessels showing a dominant role for *prox1a* in trunk lymphangiogenesis.
- i. 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.
- j. Colour coded list of analysed genotypes abbreviated in (k-m). Each embryo has a defined genotype for *prox1b* which is represented in colour code as displayed in j.
- k. The number of PLs formed (per somite) at 2dpf in genotypes indicated. Decreasing gene dosage for *prox1a* and *prox1b* progressively reduced the initial seeding of the HM by PLs.
- l. The number of cells in sprouts departing the PCV is unchanged in mutants showing that sprouting occurs normally.
- m. The number of cells in vISVs is unchanged in mutants indicating no effect on the venous endothelium.

1000 Z, zygotic; MZ, maternal and zygotic; HM, horizontal myoseptum; LL, lymphatic loop; OLV,
1001 otholitic lymphatic vessel; LFL, lateral facial lymphatic vessel; MFL, medial facial lymphatic
1002 vessel; LBA, lymphatic branchial arches; DLLV, dorsal longitudinal lymphatic vessel; PL,
1003 parachordal LEC; ISLV, intersomitic lymphatic vessel; TD, thoracic duct; DA, dorsal aorta; PCV,
1004 posterior cardinal vein.
1005

Extended data Fig 6: Embryonic transplantation data demonstrating endothelial cell autonomy of *prox1a* maternal zygotic mutant phenotype associated with Fig 6.

a. Schematic showing the endothelial cell transplantation approach used to test cell autonomy for the vascular phenotypes seen in *MZprox1a* mutants. Cell identities were traced following transplantation in unlabelled host embryos using expression of the pan-endothelial marker *fli1a* and venous/lymphatic marker *lyve1b* (indicated).

b. Examples of 5dpf wildtype hosts engrafted with wildtype donor cells (upper) that contribute to arteries (aISVs, DA), veins (PCV, vISVs) and lymphatic vessels (TD, ISLVs) and examples of wildtype embryos engrafted with *MZprox1a* mutant cells contributing arteries (aISVs, DA) and veins (vISV, PCV) but not lymphatic vessels. Note the PL accumulation in the horizontal myoseptum..

c. The number of embryos imaged and graft sizes scored post-tranplantation. As the contribution to the different EC types is affected by the graft size, we analysed grafts of comparable sizes and position from wildtype and mutant embryos. Grafts were categorised as large, medium or small. Small (spanning 1-2 somites), medium (spanning 2-3 somites) and large grafts (spanning 4-5 somites). The number of ECs contributing to each EC type was scored at single cell resolution and the percentage of grafts with a contribution to each lineage calculated accordingly. The numbers within the different graft groups indicate the number of analysed embryos within that group.

d. The percentage of embryos with vascular grafts that display contributions to arteries, veins and lymphatics. *MZprox1a*^{-/-} mutant cells contribute to arteries and veins but showed a near complete loss of contribution to lymphatics. The size of the grafts did not influence the outcome. *Z1b** indicates a mixed prox1b background. $p < 0.0001$ from an unpaired, two-sided t-test.

DA, dorsal aorta. PCV, posterior cardinal vein. PL, parachordal LEC. TD, thoracic duct. aISV, arterial intersegmental vessel. vISV, venous intersegmental vessel. ISLV, intersegmental lymphatic vessel.

References

- 1 Oliver, G., Kipnis, J., Randolph, G. J. & Harvey, N. L. The Lymphatic Vasculature in the 21(st) Century: Novel Functional Roles in Homeostasis and Disease. *Cell* **182**, 270-296, doi:10.1016/j.cell.2020.06.039 (2020).
- 2 Koltowska, K., Betterman, K. L., Harvey, N. L. & Hogan, B. M. Getting out and about: the emergence and morphogenesis of the vertebrate lymphatic vasculature. *Development* **140**, 1857-1870, doi:10.1242/dev.089565 (2013).
- 3 Baek, S. *et al.* The Alternative Splicing Regulator Nova2 Constrains Vascular Erk Signaling to Limit Specification of the Lymphatic Lineage. *Dev Cell* **49**, 279-292 e275, doi:10.1016/j.devcel.2019.03.017 (2019).
- 4 Koltowska, K. *et al.* Vegfc Regulates Bipotential Precursor Division and Prox1 Expression to Promote Lymphatic Identity in Zebrafish. *Cell Rep* **13**, 1828-1841, doi:10.1016/j.celrep.2015.10.055 (2015).
- 5 Deng, Y., Atri, D., Eichmann, A. & Simons, M. Endothelial ERK signaling controls lymphatic fate specification. *The Journal of clinical investigation* **123**, 1202-1215, doi:10.1172/JCI63034 (2013).
- 6 Shin, M. *et al.* Vegfc acts through ERK to induce sprouting and differentiation of trunk lymphatic progenitors. *Development* **143**, 3785-3795, doi:10.1242/dev.137901 (2016).
- 7 Srinivasan, R. S. *et al.* The Prox1-Vegfr3 feedback loop maintains the identity and the number of lymphatic endothelial cell progenitors. *Genes Dev* **28**, 2175-2187, doi:10.1101/gad.216226.113 (2014).
- 8 Oliver, G. & Srinivasan, R. S. Endothelial cell plasticity: how to become and remain a lymphatic endothelial cell. *Development* **137**, 363-372, doi:10.1242/dev.035360 (2010).
- 9 Wigle, J. T. & Oliver, G. Prox1 function is required for the development of the murine lymphatic system. *Cell* **98**, 769-778, doi:10.1016/s0092-8674(00)81511-1 (1999).
- 10 Hagerling, R. *et al.* A novel multistep mechanism for initial lymphangiogenesis in mouse embryos based on ultramicroscopy. *The EMBO journal* **32**, 629-644, doi:10.1038/emboj.2012.340 (2013).
- 11 Klotz, L. *et al.* Cardiac lymphatics are heterogeneous in origin and respond to injury. *Nature* **522**, 62-67, doi:10.1038/nature14483 (2015).
- 12 Martinez-Corral, I. *et al.* Nonvenous origin of dermal lymphatic vasculature. *Circ Res* **116**, 1649-1654, doi:10.1161/CIRCRESAHA.116.306170 (2015).
- 13 Stanczuk, L. *et al.* cKit Lineage Hemogenic Endothelium-Derived Cells Contribute to Mesenteric Lymphatic Vessels. *Cell Rep* **10**, 1708-1721, doi:10.1016/j.celrep.2015.02.026 (2015).
- 14 Srinivasan, R. S. *et al.* Lineage tracing demonstrates the venous origin of the mammalian lymphatic vasculature. *Genes Dev* **21**, 2422-2432, doi:10.1101/gad.1588407 (2007).
- 15 Srinivasan, R. S. *et al.* 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, doi:10.1101/gad.1859310 (2010).
- 16 Yang, Y. *et al.* Lymphatic endothelial progenitors bud from the cardinal vein and intersomitic vessels in mammalian embryos. *Blood* **120**, 2340-2348, doi:10.1182/blood-2012-05-428607 (2012).

1083 17 Stone, O. A. & Stainier, D. Y. R. Paraxial Mesoderm Is the Major Source of
1084 Lymphatic Endothelium. *Dev Cell* **50**, 247-255 e243,
1085 doi:10.1016/j.devcel.2019.04.034 (2019).
1086 18 Johnson, N. C. *et al.* Lymphatic endothelial cell identity is reversible and its
1087 maintenance requires Prox1 activity. *Genes Dev* **22**, 3282-3291,
1088 doi:10.1101/gad.1727208 (2008).
1089 19 Tao, S. *et al.* Zebrafish prox1b mutants develop a lymphatic vasculature, and prox1b
1090 does not specifically mark lymphatic endothelial cells. *PLoS One* **6**, e28934,
1091 doi:10.1371/journal.pone.0028934 (2011).
1092 20 van Impel, A. *et al.* Divergence of zebrafish and mouse lymphatic cell fate
1093 specification pathways. *Development* **141**, 1228-1238, doi:10.1242/dev.105031
1094 (2014).
1095 21 Hogan, B. M. *et al.* Ccbe1 is required for embryonic lymphangiogenesis and venous
1096 sprouting. *Nat Genet* **41**, 396-398, doi:10.1038/ng.321 (2009).
1097 22 Yaniv, K. *et al.* Live imaging of lymphatic development in the zebrafish. *Nat Med* **12**,
1098 711-716, doi:10.1038/nm1427 (2006).
1099 23 Eng, T. C. *et al.* Zebrafish facial lymphatics develop through sequential addition of
1100 venous and non-venous progenitors. *EMBO Rep* **20**, doi:10.15252/embr.201847079
1101 (2019).
1102 24 Okuda, K. S. *et al.* lyve1 expression reveals novel lymphatic vessels and new
1103 mechanisms for lymphatic vessel development in zebrafish. *Development* **139**, 2381-
1104 2391, doi:10.1242/dev.077701 (2012).
1105 25 Bussmann, J. *et al.* Arteries provide essential guidance cues for lymphatic endothelial
1106 cells in the zebrafish trunk. *Development* **137**, 2653-2657, doi:10.1242/dev.048207
1107 (2010).
1108 26 Cha, Y. R. *et al.* Chemokine signaling directs trunk lymphatic network formation
1109 along the preexisting blood vasculature. *Dev Cell* **22**, 824-836,
1110 doi:10.1016/j.devcel.2012.01.011 (2012).
1111 27 Jung, H. M. *et al.* Development of the larval lymphatic system in zebrafish.
1112 *Development* **144**, 2070-2081, doi:10.1242/dev.145755 (2017).
1113 28 Bower, N. I. *et al.* Mural lymphatic endothelial cells regulate meningeal angiogenesis
1114 in the zebrafish. *Nat Neurosci* **20**, 774-783, doi:10.1038/nn.4558 (2017).
1115 29 van Lessen, M. *et al.* Intracellular uptake of macromolecules by brain lymphatic
1116 endothelial cells during zebrafish embryonic development. *Elife* **6**,
1117 doi:10.7554/eLife.25932 (2017).
1118 30 Venero Galanter, M. *et al.* A novel perivascular cell population in the zebrafish
1119 brain. *Elife* **6**, doi:10.7554/eLife.24369 (2017).
1120 31 Kuchler, A. M. *et al.* Development of the zebrafish lymphatic system requires
1121 VEGFC signaling. *Curr Biol* **16**, 1244-1248, doi:10.1016/j.cub.2006.05.026 (2006).
1122 32 Hogan, B. M. & Schulte-Merker, S. How to Plumb a Pisces: Understanding Vascular
1123 Development and Disease Using Zebrafish Embryos. *Dev Cell* **42**, 567-583,
1124 doi:10.1016/j.devcel.2017.08.015 (2017).
1125 33 He, Z., Brazovskaja, A., Ebert, S., Camp, J. G. & Treutlein, B. CSS: cluster similarity
1126 spectrum integration of single-cell genomics data. *Genome Biol* **21**, 224,
1127 doi:10.1186/s13059-020-02147-4 (2020).
1128 34 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902
1129 e1821, doi:10.1016/j.cell.2019.05.031 (2019).
1130 35 Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP.
1131 *Nat Biotechnol*, doi:10.1038/nbt.4314 (2018).

- 1132 36 La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498,
1133 doi:10.1038/s41586-018-0414-6 (2018).
- 1134 37 Thomas, P. D. *et al.* Applications for protein sequence-function evolution data:
1135 mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res*
1136 **34**, W645-650, doi:10.1093/nar/gkl229 (2006).
- 1137 38 Wigle, J. T. *et al.* An essential role for Prox1 in the induction of the lymphatic
1138 endothelial cell phenotype. *EMBO J* **21**, 1505-1513, doi:10.1093/emboj/21.7.1505
1139 (2002).
- 1140 39 Kang, J. *et al.* An exquisite cross-control mechanism among endothelial cell fate
1141 regulators directs the plasticity and heterogeneity of lymphatic endothelial cells.
1142 *Blood* **116**, 140-150, doi:10.1182/blood-2009-11-252270 (2010).
- 1143 40 Jerafi-Vider, A. *et al.* VEGFC/FLT4-induced cell-cycle arrest mediates sprouting and
1144 differentiation of venous and lymphatic endothelial cells. *Cell Rep* **35**, 109255,
1145 doi:10.1016/j.celrep.2021.109255 (2021).
- 1146 41 Koltowska, K. *et al.* The RNA helicase Ddx21 controls Vegfc-driven developmental
1147 lymphangiogenesis by balancing endothelial cell ribosome biogenesis and p53
1148 function. *Nat Cell Biol* **23**, 1136-1147, doi:10.1038/s41556-021-00784-w (2021).
- 1149 42 Burrows, P. E. *et al.* Lymphatic abnormalities are associated with RASA1 gene
1150 mutations in mouse and man. *Proc Natl Acad Sci U S A* **110**, 8621-8626,
1151 doi:10.1073/pnas.1222722110 (2013).
- 1152 43 Frye, M. *et al.* EphrinB2-EphB4 signalling provides Rho-mediated homeostatic
1153 control of lymphatic endothelial cell junction integrity. *Elife* **9**,
1154 doi:10.7554/eLife.57732 (2020).
- 1155 44 Geng, X. *et al.* S1PR1 regulates the quiescence of lymphatic vessels by inhibiting
1156 laminar shear stress-dependent VEGF-C signaling. *JCI Insight* **5**,
1157 doi:10.1172/jci.insight.137652 (2020).
- 1158 45 Grimm, L. *et al.* Yap1 promotes sprouting and proliferation of lymphatic progenitors
1159 downstream of Vegfc in the zebrafish trunk. *Elife* **8**, doi:10.7554/eLife.42881 (2019).
- 1160 46 Koltowska, K. *et al.* mafba is a downstream transcriptional effector of Vegfc
1161 signaling essential for embryonic lymphangiogenesis in zebrafish. *Genes Dev* **29**,
1162 1618-1630, doi:10.1101/gad.263210.115 (2015).
- 1163 47 Yuan, L. *et al.* Abnormal lymphatic vessel development in neuropilin 2 mutant mice.
1164 *Development* **129**, 4797-4806 (2002).
- 1165 48 Termini, C. M. & Gillette, J. M. Tetraspanins Function as Regulators of Cellular
1166 Signaling. *Front Cell Dev Biol* **5**, 34, doi:10.3389/fcell.2017.00034 (2017).
- 1167 49 Wu, R. S. *et al.* A Rapid Method for Directed Gene Knockout for Screening in G0
1168 Zebrafish. *Dev Cell* **46**, 112-125 e114, doi:10.1016/j.devcel.2018.06.003 (2018).
- 1169 50 Armour, S. M. *et al.* An HDAC3-PROX1 corepressor module acts on HNF4alpha to
1170 control hepatic triglycerides. *Nat Commun* **8**, 549, doi:10.1038/s41467-017-00772-5
1171 (2017).
- 1172 51 Liu, X. *et al.* Mitotic Implantation of the Transcription Factor Prospero via Phase
1173 Separation Drives Terminal Neuronal Differentiation. *Dev Cell* **52**, 277-293 e278,
1174 doi:10.1016/j.devcel.2019.11.019 (2020).
- 1175 52 Qin, J. *et al.* Prospero-related homeobox (Prox1) is a corepressor of human liver
1176 receptor homolog-1 and suppresses the transcription of the cholesterol 7-alpha-
1177 hydroxylase gene. *Mol Endocrinol* **18**, 2424-2439, doi:10.1210/me.2004-0009 (2004).
- 1178 53 Song, K. H., Li, T. & Chiang, J. Y. A Prospero-related homeodomain protein is a
1179 novel co-regulator of hepatocyte nuclear factor 4alpha that regulates the cholesterol
1180 7alpha-hydroxylase gene. *J Biol Chem* **281**, 10081-10088,
1181 doi:10.1074/jbc.M513420200 (2006).

1182 54 Takeda, Y. & Jetten, A. M. Prospero-related homeobox 1 (Prox1) functions as a novel
1183 modulator of retinoic acid-related orphan receptors alpha- and gamma-mediated
1184 transactivation. *Nucleic Acids Res* **41**, 6992-7008, doi:10.1093/nar/gkt447 (2013).
1185 55 Lawson, N. D. & Weinstein, B. M. In vivo imaging of embryonic vascular
1186 development using transgenic zebrafish. *Dev Biol* **248**, 307-318,
1187 doi:10.1006/dbio.2002.0711 (2002).
1188 56 Ciruna, B. *et al.* Production of maternal-zygotic mutant zebrafish by germ-line
1189 replacement. *Proc Natl Acad Sci U S A* **99**, 14919-14924,
1190 doi:10.1073/pnas.2224599999 (2002).
1191 57 Westerfield, M. *The Zebrafish Book. A Guide for the Laboratory use of Zebrafish*
1192 *(Danio rerio)*. 4 edn, (University of Oregon Press, 2000).
1193 58 Gagnon, J. A. *et al.* Efficient mutagenesis by Cas9 protein-mediated oligonucleotide
1194 insertion and large-scale assessment of single-guide RNAs. *PLoS One* **9**, e98186,
1195 doi:10.1371/journal.pone.0098186 (2014).
1196 59 Bussmann, J. & Schulte-Merker, S. Rapid BAC selection for tol2-mediated
1197 transgenesis in zebrafish. *Development* **138**, 4327-4332, doi:10.1242/dev.068080
1198 (2011).
1199 60 Okuda, K. S., Baek, S. & Hogan, B. M. Visualization and Tools for Analysis of
1200 Zebrafish Lymphatic Development. *Methods Mol Biol* **1846**, 55-70, doi:10.1007/978-
1201 1-4939-8712-2_4 (2018).
1202 61 Bower, N. I. *et al.* Vegfd modulates both angiogenesis and lymphangiogenesis during
1203 zebrafish embryonic development. *Development* **144**, 507-518,
1204 doi:10.1242/dev.146969 (2017).
1205 62 Kartopawiro, J. *et al.* Arap3 is dysregulated in a mouse model of hypotrichosis-
1206 lymphedema-telangiectasia and regulates lymphatic vascular development. *Hum Mol*
1207 *Genet* **23**, 1286-1297, doi:10.1093/hmg/ddt518 (2014).
1208 63 Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells.
1209 *Nat Commun* **8**, 14049, doi:10.1038/ncomms14049 (2017).
1210 64 Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of
1211 Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281-291 e289,
1212 doi:10.1016/j.cels.2018.11.005 (2019).
1213 65 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell
1214 transcriptomic data across different conditions, technologies, and species. *Nat*
1215 *Biotechnol* **36**, 411-420, doi:10.1038/nbt.4096 (2018).
1216 66 McCarthy, D. J., Campbell, K. R., Lun, A. T. & Wills, Q. F. Scater: pre-processing,
1217 quality control, normalization and visualization of single-cell RNA-seq data in R.
1218 *Bioinformatics* **33**, 1179-1186, doi:10.1093/bioinformatics/btw777 (2017).
1219 67 Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at
1220 multiple resolutions. *Gigascience* **7**, doi:10.1093/gigascience/giy083 (2018).
1221 68 Kejie Li, Z. O., Dongdong Lin, Michael Mingueneau, Will Chen, David Sexton,
1222 Baohong Zhang. cellxgene VIP unleashes full power of interactive visualization,
1223 plotting and analysis of scRNA-seq data in the scale of millions of cells. *bioRxiv*
1224 **2020.08.28.270652**, doi:<https://doi.org/10.1101/2020.08.28.270652> (2020).
1225 69 Kazenwadel, J. *et al.* GATA2 is required for lymphatic vessel valve development and
1226 maintenance. *J Clin Invest* **125**, 2979-2994, doi:10.1172/JCI78888 (2015).
1227 70 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler
1228 transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
1229 71 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
1230 genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033
1231 (2010).

1232

1233

1234