RESOURCE ARTICLE

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

Lin Grimm[1,2,3,*], Elizabeth Mason[1,2*,#], Oliver Yu[1,2], Stefanie Dudczig[1,2], Virginia Panara[4], Tyrone Chen[1,2], Neil I. Bower[3], Scott Paterson[1,2,3], Kazuhide Okuda[1,2], Maria Rondon Galeano[1,2], Sakurako Kobayashi[1,2], Anne Senabouth[3,5], Anne K. Lagendijk[3], Joseph Powell[3,5,6,7], Kelly A. Smith[8], Katarzyna Koltowska[4] and Benjamin M. Hogan[1,2,3,8 #].

[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] Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden.
[5] Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, Sydney, 2010, Australia.
[6] School of Medical Sciences, 18 High Street, University of New South Wales, Kensington, Sydney, Australia.
[7] Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, Sydney, 2010, Australia.
[8] Department of Anatomy and Physiology, University of Melbourne, Melbourne, VIC 3000, Australia.

[#] *Authors for Correspondence:*

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

Dr Elizabeth Mason (computational biology lead)
Organogenesis and Cancer Program,
Peter MacCallum Cancer Centre,
Melbourne, VIC 3000, Australia
E-mail: elizabeth.mason@petermac.org

Keywords: Lymphatics, Lymphangiogenesis, Prox1, Single cell, RNA Seq, ATAC seq.
Summary: 181 words
Main Text: 4,700 words

During development, the lymphatic vasculature forms as a second, new vascular network derived from blood vessels. The transdifferentiation of embryonic venous endothelial cells (VECs) into lymphatic endothelial cells (LECs) is the first step in this process. Specification, differentiation and maintenance of LEC fate are all driven by the transcription factor Prox1, yet downstream mechanisms remain to be elucidated. We 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 VEC-LEC fate reversion during differentiation. Using maternal and zygotic *prox1a/prox1b* mutants, we determine the earliest transcriptomic changes directed by Prox1 during LEC specification. This work altogether reveals new transcriptional targets and regulatory regions of the genome downstream of Prox1 in LEC maintenance, as well as showing that Prox1 specifies LEC fate primarily by limiting blood vascular and hematopoietic fate. This extensive single cell resource provides new mechanistic insights into the enigmatic role of Prox1 and the control of LEC differentiation in development.

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 (TF) 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,

77 early embryonic lymphanigogenesis occurs chiefly from the CVs [11-17]. In mice, both the earliest
78 stages of LEC progenitor sprouting from CVs and maintenance of LEC fate in stable lymphatics
79 are dependent on the function of Prox1 [9,18].
80
81 Despite over two decades of study of this enigmatic developmental process, the
82 transcriptomic changes that occur as embryonic venous endothelial cells (VECs)
83 transdifferentiate into LEC progenitors, and further differentiate into mature lymphatics,
84 have not been transcriptionally profiled *in vivo*. In the absence of Prox1 in conditional
85 knockout mice, LECs have been shown to lose the expression of some LEC markers and to gain
86 expression of some blood vascular endothelial cell (BEC) markers[18]. Prox1 is known to
87 autoregulate its own expression and to also regulate Flt4 expression in a positive regulatory
88 loop during early development[7]. Yet how Prox1 controls the transcriptome during LEC
89 specification, differentiation and maintenance has not been described in detail. This is in part
90 because of the technical challenge of accessing early mouse endothelial cells (ECs) in wildtype
91 and mutant embryos, a problem that is not limiting when using the zebrafish embryo.
92
93 As recent studies have demonstrated highly conserved expression and function of Prox1
94 homologues in zebrafish[4,6,19,20], we here took advantage of the accessibility of the zebrafish
95 embryo to examine developmental lymphangiogenesis using single cell transcriptomics. We
96 provide a resource of new markers of VEC-LEC transdifferentiation, differentiating and
97 mature cell types and validate several new markers with transgenic approaches. We analysed
98 zebrafish zygotic *prox1a* mutants with single cell RNA sequencing and single cell ATAC
99 sequencing. This identified a VEC-LEC fate reversion in the absence of zygotic Prox1a, defined
100 key Prox1-dependent genes in fate maintenance and discovered regulatory regions of
101 chromatin (enhancers) controlled by Prox1. Finally, profiling maternal-zygotic double
102 *prox1a/prox1b* (null) mutant vasculature with single cell RNA sequencing revealed that Prox1
103 reduces gene expression in early LEC progenitors, including for a core network of conserved
104 haematopoietic and blood vascular fate regulators. Overall, this single cell resource reveals
105 embryonic lymphangiogenesis with Prox1-dependent mechanisms central in specification,
106 differentiation and maintenance of LEC fate.
107
108

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 tissues[24-27]. 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), 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 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 maturing[32] (**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 ECs, 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**, full details in methods). 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 ECs that comprise our single cell atlas of lymphangiogenesis, which is accessible in an interactive CellXGene explorer web app at the link http://115.146.95.206:5006/ (**Fig 1b**, **Extended data Fig 1b-c**).

141 This dataset is displayed[35] as a UMAP in **Fig 1b** coloured according to developmental stage
142 and cell phenotype, with venous (VEC) to lymphatic (LEC) trajectory of differentiation
143 confirmed by RNA-velocity analysis[36]. Comparison of cells at 40hpf with later stages revealed
144 that the earlier populations are transcriptionally distinct from VECs and LECs at 3, 4 and 5dpf.
145 Notably, the 40hpf "VEC preLEC" cluster contained both *prox1a+* and *prox1a-* cells forming a
146 single population rather than discrete clusters. This indicates that LECs are not
147 transcriptionally distinct from VECs at 40hpf, suggesting that while early *prox1a+* cells are
148 specified they have not yet differentiated (**Fig 1c**)[4]. We identified 3 main classes of LECs at 3,
149 4 and 5dpf: muLECs (1,511 cells) marked by expression of *osr2*[28], canonical LECs marked by
150 expression of *cdh6* (2,669 cells), and a smaller sub-population of LECs expressing low levels
151 of *prox1a* (LEC_low_prox1a, 310 cells) (**Fig 1c**, **Extended data Fig 1d-e**). To define markers of
152 differentiating canonical LECs at each developmental stages, we applied differential
153 expression (DE) analysis (**Extended data Table 1b**,). This analysis not only captured known
154 LEC markers including *prox1a, angpt2a, tbx1* and *cldn11b*, but also uncovered new genes
155 commonly expressed in canonical LECs across all developmental stages including *grp156,*
156 *hapln3, cdh6* and *tspan18a* (**Fig 1d, Extended data Fig 1f**). We extended this approach and
157 evaluated global differences between all canonical LECs and all VECs (n=1,240 genes,
158 **Extended data Table 1c**). GO analysis[37] confirmed the association of biological processes
159 known to be associated with lymphatics with genes upregulated in LEC (n=752 LEC gene set),
160 these terms included "lymphangiogenesis", "glycolytic process", "lymph vessel development"
161 and "ameboidal-type cell migration" (**Fig 1e**, **Extended data Table 1d**).

162 To provide further confidence in the specificity of the new LEC markers defined by this scRNA-
163 seq resource, we used DE analysis to define cluster specific gene expression. We noted that
164 this analysis suggested higher expression of the well-known marker *lyve1b* in muLECs than in
165 LECs, which we validated by examining expression levels of *lyve1b* using the established
166 transgenic line *Tg(lyve1b:DsRed2)* (heat map in **Fig 2a**). We further identified expression of
167 the known kidney epithelial solute transporter *slc7a7* as uniquely expressed in muLEC cells
168 and *fabp11a* as a canonical LEC and VEC marker excluded from muLECs. We generated both
169 new *slc7a7a*–Citrine and a *fabp11a-Citrine* BAC transgenic strains, which confirmed the
170 muLEC and vascular expression patterns and further validated our scSeq atlas (**Fig 2b-c**).
171 Overall, this atlas therefore identifies n=752 specific markers of lymphangiogenesis, the

172 majority of which are new markers (**Extended data Table 1c**), representing new candidate
173 regulators and spanning four developmental stages.

174

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

177 Prox1 is both necessary and sufficient to drive VEC to LEC trans-differentiation in mammals
178 and this is proposed to occur by Prox1 simultaneously initiating LEC fate and repressing the
179 VEC fate program[8,18]. Prox1 expression is also necessary for the maintenance of LEC identity
180 during development[18,38,39]. Despite the role for Prox1 being well-studied in mouse models,
181 the transcriptomic program controlled by Prox1 *in vivo* has never been profiled. Zebrafish
182 *prox1a* zygotic and maternal zygotic mutants have been previously described[4,20]. The zygotic
183 mutants retain maternal deposition of *prox1a* in the oocyte, sufficient to drive normal PL
184 formation, LEC migration to the horizontal myoseptum (HM) and initial assembly of
185 lymphatics by 4dpf[4,19], however mutants have a reduction in total LEC numbers throughout
186 the face and trunk[4]. When the maternal contribution of *prox1a* is removed, lymphatic
187 development is more severely impaired[4]. We hypothesised that the zygotic mutants which
188 form lymphatic vessels, likely have abnormal vessel identity in the absence of zygotic Prox1.
189 Thus, we applied scRNA-seq to ECs FAC sorted from *Zprox1a*$^{-/-}$ mutants[20] and WT sibling
190 zebrafish at 4dpf, during lymphatic vessel assembly (n=8,075 cells, **Extended data Fig 2a**,
191 **Extended data Table 2a**). Cluster analysis revealed 3 populations of LECs (n=2,068) and a
192 single population of VECs (n=1,051) comprising both mutant and wild type cells, and perhaps
193 most striking, a single population of mutant cells (n=484) transcriptionally similar to VECs,
194 marked by the expression of *aqp1a.1* (**Fig 3a-d**, **Extended data Figs 2b-e**). All three LEC
195 clusters (LEC, LEC_S1, LEC_S2) showed graded expression of lymphatic markers *prox1a*,
196 *cldn11b*, *cdh6* and *angpt2a*, such that "LEC" most closely resembles canonical LECs from our
197 WT atlas. LEC_S1 and S2 represent alternative LEC sub-types with different proliferative
198 potential based on S-phase occupancy and cell cycle marker *mki67* (**Fig 3b,e**, **Extended data**
199 **Fig 2e**). The expression of *prox1a* was lowest in the more proliferative LEC_S1 cells, and
200 highest in the less proliferative LEC_S2 (**Fig 3b,e**). RNA-velocity analysis suggested a
201 trajectory between the mutant cluster and the LEC cluster (**Fig 3c** lower). Taken together,

202 these findings suggest that the mutant specific cluster sits on a trajectory between LEC and
203 VEC fate, representing cells either failing to fully differentiate from VEC to LEC or LECs
204 undergoing dedifferentiation and fate reversion.
205
206 To survey how Prox1 maintains normal LEC differentiation, we performed DE analysis
207 comparing the mutant cluster with WT LEC (**Fig 3f**, **Extended data Table 2b**). Overall, a much
208 larger set of genes were downregulated than upregulated in the mutant cluster (n=1,034 vs
209 n=294) with almost half of the most downregulated genes (AvLogFC) highly enriched for LEC
210 markers (eg. *tbx1*, *cdh6*, *cldn11b*). Of the upregulated genes, almost 75% were VEC markers
211 (eg. *sox7*, *kdrl*, *cdh5*) and again this was enriched in the most highly upregulated genes (**Fig**
212 **3f, Extended data Fig 2f**). Comparing the change in gene expression between the wildtype
213 LEC cluster and VEC cluster, with the change between wildtype LEC and mutant cluster, we
214 found striking concordance suggesting the mutant cluster is shifted along a LEC fate to VEC
215 fate trajectory (**Fig 3g**). Overall, there is a simultaneous loss of lymphatic fate and re-
216 acquisition of blood vascular gene expression, consistent with work in mouse and with Prox1
217 function being highly conserved between vertebrates. To validate these observations, we
218 used confocal imaging of Z*prox1a*$^{-/-}$ mutants and observed upregulation of blood vascular
219 markers *cdh5* (**Fig 3h-i, Extended data Fig 3a,c**) and *kdrl* (**Fig 3j-k, Extended data Fig 3d-g**) in
220 *lyve1b*-positive lymphatics vessels. We saw a coincident reduction in the expression of *lyve1b*
221 in these vessels (**Extended data Fig3b).** Interestingly, the difference in *kdrl* levels (relative
222 intensity) in lymphatics between mutants and wildtype was more significant at 5dpf than
223 4dpf, suggesting that progressive dedifferentiation may be occurring in the mutant (**Fig 3k**).
224 This single cell profile thus demonstrates that Prox1 alone is sufficient to maintain LEC fate
225 and identifies the gene expression maintained by Prox1 during lymphatic differentiation.
226
227 **Single cell ATAC sequencing reveals chromatin accessibility signatures in LECs and VECs,**
228 **identifying lymphatic specific enhancers and predicting key LEC TF families**
229
230 At the level of gene expression, Prox1 function is essential for VEC to LEC differentiation,
231 however it is unknown if Prox1 controls chromatin accessibility during this process. Thus, we
232 next profiled *Zprox1a*$^{-/-}$ mutants at 4dpf during vessel assembly using single nuclei (sn) ATAC-
233 seq (n=3,731 nuclei, **Extended data Fig 4a, Extended data Table 3a**). Cluster analysis (**Fig 4a**)

234 and overall accessibility of key markers (**Fig 4b**) identified similar populations to the scRNA-
235 seq profiling: canonical LEC (LEC_01 n=114 nuclei, LEC_02 n=213 nuclei) and VEC (VEC_01
236 n=157 nuclei) clusters, and a small but discrete population comprised almost entirely of
237 mutant cells (Mutant Cluster n=47 nuclei). Notably, we found that there was almost no
238 contribution of mutant cells to the LEC clusters (**Fig 4a** lower bar plot), indicating a loss of fate
239 when cell identity is determined at the level of chromatin state **(Extended data Fig 4b)**.
240 Differential accessibility (DA) analysis revealed unique sets of peaks that mark individual VEC,
241 LEC and AEC (arterial EC) clusters (**Fig 4c**).
242
243 To identify phenotype specific chromatin accessibility, we performed DA analysis between
244 WT LEC and WT VEC. This revealed the more accessible regions in the LEC clusters were
245 associated with LEC genes identified in our atlas, and the less accessible regions with VEC
246 genes (**Fig 4d, Extended data table 3b**). DA identified n=1,561 LEC specific peaks and n=2,624
247 VEC specific peaks representing putative lineage specific enhancers or regulatory elements
248 (**Extended data table 3b**). To test if these regions identified enhancers, we used the zebrafish
249 enhancer detector plasmid system (ZED vector[40]) and tested peaks that were uniquely open
250 5' of the newly identified LEC marker gene *cdh6.* To exclude non-specific reporter expression,
251 we also generated F1 embryos using a ZED vector only control. This did display mosaic
252 neuronal and dorsal root ganglion expression **Extended data Fig 4h**, suggesting that the
253 expression from the *cdh6* enhancer in neural structures using this vector should be taken with
254 some caution. No vascular expression was detected. Transgenesis identified one new
255 functional and LEC specific enhancer upstream of *cdh6*, validating the use of this dataset for
256 enhancer discovery (**Fig. 4e-f**).
257
258 We next aimed to use this new set of putative enhancer regions in an unbiased manner to
259 identify key TF families likely to regulate LEC development. We performed motif enrichment
260 analysis using HOMER[41] for all LEC enriched DA peaks identifying n=64 TF family motifs
261 (**Extended data table 3c**). Notably, we found motifs for TCF, ETS, SCL (TAL1), NFAT, TBX, MAF,
262 SLUG and RBPJ family TFs to be enriched in peaks more accessible in LECs (**Fig. 4g**). Analysis
263 of the human homologues in ENCODE data[42] revealed these TFs regulate a highly connected
264 network of genes expressed in our LEC atlas (**Fig. 4h**). Importantly, a number of these TFs are
265 already known to play important roles in lymphatics (e.g. NFAT[43,44], MAFB[45,46], TBX1[47],

266 TCF[48,49]), supporting the prediction that members of these TF families will play important
267 function roles in LEC development.
268
269 **Prox1 reduces the accessibility of chromatin peaks associated with blood vascular and**
270 **haematopoetic TF motifs**
271 Given that cells in the mutant specific cluster analysed by scRNA sequencing demonstrate loss
272 of LEC gene expression and acquisition of VEC gene expression, we expected chromatin
273 accessibility to change in an equally coordinated manner. However, analysis of accessibility
274 at individual genes (gene score) for the Mutant Cluster revealed changes inconsistent with a
275 simple fate shift (**Extended data tables 4a-b**). Some LEC specific genes (eg. *prox1a*, *cdh6* and
276 *tbx1*) showed loss of transcription in the mutant cluster but increased chromatin accessibility,
277 and some specific VEC genes with increased transcription in the mutant cluster (eg. *cdh5, flt1,*
278 *gata6*) also showed discordant chromatin changes (**Fig 5a-b, Extended data Fig 4g**). A DAP
279 analysis revealed little concordance between regions with increased accessibility, and
280 expression in LEC or VEC in the scSeq atlas (**Fig 5c**). We identified a subset of genes with more
281 accessible chromatin overall in the Mutant Cluster than either WT LEC or VEC settings (**Fig**
282 **5d**). Notably, at the level of individual peaks we found that n=1,726 peaks displayed a striking
283 increase in accessibility in the mutant cluster compared with WT LEC and n=1,794 peaks an
284 increase compared with WT VEC (**Fig 5e**, **Extended data tables 4a-b**). Of these, 431 were
285 common peaks identifying more accessible chromatin regions in the mutant cluster than in
286 either WT LEC or WT VEC (**Fig 5e,** examples in **Fig. 5f**). This suggests that some regions of
287 chromatin open up more than usual during LEC or VEC differentiation in the absence of Prox1.
288
289 To investigate the nature of the peaks that were opening in the absence of Prox1, we used an
290 unbiased assessment of TF motifs within these regions. The more open regions were highly
291 enriched for motifs of early acting TFs involved in embryonic vasculogenesis and
292 haematopoiesis, including Erg, Etv2, Etv4, Ets1, Fli1 and Spi1/Pu.1 (**Fig 5g**, **Extended data**
293 **table 4c**). We examined the human homologues of these TFs in ENCODE[42] data together with
294 homologues of genes expressed in our atlas and identified a highly connected putative gene
295 regulatory network (GRN) made up of early blood and blood vascular TFs (including ETV2,
296 TAL1, SOX7, SOX17 ERG and other key fate regulators) driving target genes including each
297 other, FOS, JUN, MYC and STAT3 (**Fig. 5h**). We take this to suggest that in the absence of

298 Prox1, TFs that are part of a GRN normally suppressed by Prox1 function are reactivated to
299 drive blood vascular and blood fates. This demonstrates that Prox1 coordinates the correct
300 accessibility of the chromatin and likely controls other TF functions while maintaining LEC fate
301 in ECs. Furthermore, we note that a large number of the genes with increased chromatin
302 accessibility are known regulators of lymphangiogenesis and so this increased accessibility
303 may be a sensitive way to identify key regulatory genes (**Extended data table 4c**).
304

305 **Prox1 is required cell autonomously for the normal sprouting of lymphatic progenitors and**
306 **their contribution to the lymphatic lineage**

307 While the above data identifies the role of Prox1 at stages when it is maintaining LEC identity
308 in assembling LECs, Prox1 is essential for the very earliest decision made when a VEC becomes
309 specified to develop into a LEC[9]. To ask how Prox1 controls the earliest stages of VEC to LEC
310 transdifferentiation, we first needed to generate a complete Prox1 loss of function model in
311 zebrafish. We generated *prox1a*$^{+/-}$, *prox1b*$^{+/-}$ double heterozygous animals (lacking both Prox1
312 homologues) and then used germline transplantation approaches to produce animals
313 carrying a double mutant germline (**Extended data Fig 5a**). Genetic crossing of these animals
314 generated embryos that were maternal zygotic (MZ) *MZprox1a*$^{-/-}$, *MZprox1b*$^{-/-}$ mutant
315 zebrafish completely lacking *prox1a/b* transcript expression or maternal deposition (**Fig 6a;**
316 **Extended data Fig 5b-c**). A quantitative phenotypic analysis revealed double MZ mutants
317 show a severe reduction of facial lymphatics and a near complete loss of lymphatic vessels in
318 the trunk by 4dpf (**Fig 6e**, **Extended data Fig 5d-g**) [4]. Furthermore, *MZprox1a*$^{-/-}$, *MZprox1b*$^{-/-}$
319 mutants initially show a delay in the formation of PLs as these cells emigrate the CV to invest
320 the HM, but despite this delay, these cells eventually seed the HM before failing to undertake
321 any further migration (**Fig 6d,f**). We saw evidence for genetic interaction between *prox1a* and
322 *prox1b* in PL formation and no change in the overall number of cells sprouting from the PCV
323 or number of cells contributing to the venous ISVs (blood vessels) suggesting both *prox1a* and
324 *prox1b* contribute specifically to early PL development (**Fig 6g-k)**. To test whether LECs require
325 Prox1a in a cell autonomous manner, we performed embryonic transplantation to generate
326 chimeric embryos. Due to challenges generating and maintaining large numbers of double MZ
327 mutant embryos, we performed transplantion of *MZprox1a*$^{-/-}$ mutant cells into wildtype hosts
328 only and assessed the contributions of vascular grafts to arteries, veins and lymphatics
329 (**Extended data Fig 6a**). MZ*prox1a*$^{-/-}$ mutant cells efficiently contributed AECs, VECs but not

330 LECs to developing vessels in otherwise wildtype hosts at 5dpf (**Extended data Fig 6b-d**). Thus,
331 we confirmed that Prox1 is necessary cell autonomously for zebrafish lymphatic development
332 and that the *MZprox1a*$^{-/-}$, *MZprox1b*$^{-/-}$ mutant phenotype is more severe than any previously
333 described zebrafish Prox1 mutants.
334
335 **Prox1 functions to suppress blood and blood vascular fate during the early specification of**
336 **LEC fate in the embryo**
337 To understand the very earliest role that Prox1 plays in LEC development, we profiled the
338 endothelium of *MZprox1a*$^{-/-}$, *MZprox1b*$^{-/-}$ mutant zebrafish at 40hpf (when cells are both
339 being specified to LEC fate and also actively sprouting from various regions of the PCV[3,4,6])
340 using single cell RNA seq as described above (**Extended data table 5a, Extended data Fig 7a-**
341 **c**). Using key marker expression and DE analysis, we identified populations of cells that
342 included VECs, endocardium, mixed populations of VECs and Prox1+ LEC progenitors (named
343 LEC_VEC) and clusters with expression of some VEC and AEC markers that were likely still
344 differentiating (**Fig 7a-c, Extended data Fig 7d-h**). Based on expression of known markers and
345 genes associated with EC sprouting (*mki67*, *pcna*), we defined two populations of secondary
346 sprouts of venous origin (LEC VEC 01 n=713 cells, LEC VEC 02 n=677 cells) and a single
347 population of cells representing the cardinal vein (PCV n=812 cells; **Fig 7d-e**). Consistent with
348 our observations in **Fig 1,** cells expressing *prox1a* at 40hpf were not transcriptionally distinct
349 from sprouting VECs and failed to form a "lymphatic progenitor" cluster, suggesting they are
350 specified and express *prox1a* but not yet differentiated. RNA velocity analysis suggested that
351 these three clusters remained closely related, consistent with little differentiation between
352 these populations at this stage (**Fig 7a** right).
353
354 To identify the earliest transcriptional changes controlled by Prox1 in lymphatic development,
355 we used DE to evaluate the global differences between *MZprox1a*$^{-/-}$, *MZprox1b*$^{-/-}$ and WT in
356 the LEC VEC 01, LEC VEC 02 and PCV clusters of sprouting cells (**Fig 7f, Extended data table**
357 **5b**). We found a significant upregulation of a large set of genes in the absence of Prox1 that
358 were identified as enriched for blood vascular and haematopoietic genes by GO-term and
359 marker analysis (**Fig 6g-h,** n=1,137 genes, **Extended data table 5c**). We also saw increased
360 expression of mitochondrial metabolism genes and of pre-mRNA splicing genes in the
361 absence of Prox1 (**Fig. 7f, Extended data table 5c**). At this stage of development, in contrast

362 to the 4 dpf stage, we saw little evidence of positively regulated genes downstream of Prox1
363 (just n=49 genes downregulated in *MZprox1a*$^{-/-}$, *MZprox1b*$^{-/-}$). We next examined the genes
364 up-regulated in Prox1 mutants that we identified as key early blood and vascular
365 developmental regulators. This includes *esm1, flt4, mef2c, hdac1, lmo2, lmo4, rasip1, egfl7,*
366 *dusp5, clec14a, fli1b* and others (**Fig. 7h-i**). STRING[50] analysis suggested the presence of inter-
367 related gene networks and so we further examined up-regulated TFs in this network by
368 leveraging human homologues in ENCODE data and identifying known target genes expressed
369 in our dataset. This identified a GRN containing SOX7, FLI1, LMO2, ETV2, TCF12, LMO4 and
370 MAFB as well as known targets of these TFs, that is repressed in response to Prox1-function
371 during LEC specification (**Fig. 7j**). This is highly concordant with observations at later stages of
372 up-regulated TF activity in 4-day old mutants (**Fig. 5**). Overall, we believe this suggests that in
373 the earliest stages of LEC specification Prox1 is not actively driving a fate program but rather
374 is acting primarily to down-regulate blood vascular and hematopoietic fate. It seems likely
375 that the negative regulation of core blood and blood vascular fate genes (those in **Fig. 7j**),
376 would be sufficient to drive the differentiation in these earliest specified LECs.
377
378
379 **Discussion**
380 In this study, we first present a single cell RNA seq analysis of four key stages of embryonic
381 lymphangiogenesis in a vertebrate embryo, revealing new markers and potential regulators
382 of LEC differentiation. We further profile the ECs of *Zprox1a* mutants, which form lymphatic
383 structures, but these "lymphatics" dedifferentiate or revert their fate to blood vascular in a
384 mutant specific, fate-shifted, single cell cluster. This analysis identified the transcriptional
385 code maintained by Prox1 in order to maintain LEC identity and also shows the highly
386 conserved nature of Prox1 function comparing zebrafish with mice. Overall, the striking fate
387 shift identified across the whole transcriptome confirms that the function of just one
388 transcription factor (Prox1) is sufficient to maintain cellular identity, validating its status as
389 the master regulator of LEC identity.
390
391 As well as single cell profiling of the developing LEC transcriptome, we also performed snATAC
392 seq on 4 dpf zygotic *prox1a* mutants and wildtype, again identifying a mutant specific, fate
393 shifted cluster in this analysis. Analysis of the wildtype LECs and VECs revealed strong

394 concordance between chromatin accessibility at LEC enhancers and the transcriptional profile
395 for LEC and VEC specific genes. Interestingly however, analysis of the mutant cluster
396 chromatin identified ectopically open regions and peaks with a distinct discordance between
397 chromatin accessibility at enhancers and transcriptional profile. In the mutant cluster at 4dpf,
398 a large number of opened chromatin regions were enriched for TF motifs for earlier acting
399 vasculogenic and haematopoietic TFs. This suggests that regulators of earlier vascular fate
400 become more active and increase chromatin accessibility at specific targets in the absence of
401 Prox1. Overall, this revealed that at the level of chromatin accessibility the fate-shifted ECs
402 display a more immature state, perhaps a consequence of regulatory "confusion" due to a
403 failed fate transition. It seems likely that Prox1 has a combinatorial function as part of a larger
404 GRN of developmental TFs, which has yet to be studied in detail.
405
406 Additional biological insights from this study come from our analysis of double maternal
407 zygotic *prox1a, prox1b* mutants. These mutants are presumed "null" mutants for Prox1
408 orthologues and they revealed (through scRNA seq) that during its earliest role in vascular
409 development, in LEC fate specification and VEC-LEC transdifferentiation, Prox1 functions
410 primarily to negatively regulate blood and blood vascular fate. While this may represent a
411 downstream program rather than direct repression of gene expression by Prox1 orthologues,
412 it is notable that Prox1 and *Drosophila* Prospero have been reported to be able to function as
413 repressors in a context dependent manner [51-55]. Interestingly, the analysis of upregulated
414 genes identified a set of TFs that are known early regulators of blood and blood vascular fates
415 during embryonic haematopoiesis and vasculogenesis. These included Sox7, Etv2, Lmo2 and
416 Lmo4 and we take this to suggest that Prox1 functions during VEC-LEC transdifferentiation at
417 least in part by blocking expression of early acting blood vascular fate driving TFs. This is in
418 line with maintaining a negative regulatory relationship with these and other fate driving TFs
419 at 4 dpf as described above. It will be interesting in the future to understand at a mechanistic
420 level if Prox1 is actively repressing gene expression at bound targets to block alternative cell
421 fates and if so, if the activity of Prox1 switches between repressor or activator at different
422 stages or target genes.
423
424 Altogether this study describes the process of developmental venous to lymphatic
425 transdifferentiation, early LEC differentiation and LEC maintenance in detail, *in vivo*. We

426 describe the role of Prox1 during this process, revealing a conserved and dynamic regulatory
427 process with unprecedented resolution. This resource will help to understand
428 lymphangiogenesis in contexts beyond the embryo, such as in pathological
429 lymphangiogenesis in metastasis, inflammation and tissue repair. Finally, we suggest that
430 combining zebrafish mutant models, with their inherent ease of cellular accessibility, and the
431 resolution afforded by single cell profiling, will be an approach that yields powerful new
432 insights in many areas of developmental biology in the future.
433

434 **Acknowledgements**
435 This project was supported in part by NHMRC Ideas grant 2004300 and ARC Discovery project
436 DP180102846. B.M.H. was supported by NHMRC Fellowship 1155221. We thank Angelika
437 Christ and the IMB Sequence Facility (University of Queensland), Tim Semple and Peter Mac
438 Genomics Facility, the Research Computing Facility at Peter Mac, FACS facilities at UQ and
439 Petermac and Isaac Virshup for assistance.
440

441 **Author contributions**
442 L.G. and E.M. performed, analysed experiments and co-wrote manuscript. B.M.H, K.K and
443 N.L.H conceptualised experiments, analysed data and co-wrote manuscript. S.D, T.C, O.Y,
444 N.I.B, S.P, K.O, A.S and A.L. performed and analysed experiments. J.P, and K.A.S, provided key
445 unpublished reagents and analysed data.
446

447 **Conflicts of interest**
448 The authors declare no conflict of interests.
449

450 **Online Methods**
451 **Zebrafish husbandry**
452 Zebrafish work was conducted in compliance with animal ethics committees at the Peter
453 MacCallum Cancer Centre, the University of Melbourne and the University of Queensland.
454 Published transgenic lines used were: *Tg(fli1a:nEGFP)*^{y7} [56]*; Tg(-5.2lyve1b:DsRed)*^{nz101} [24]*;* and
455 *Tg(kdrl:Has.HRAS-mCherry)*^{s916} [21]*.* Published mutant lines used were *prox1a*^{i278} [20], *prox1b* [4,19].
456 *Tg(lyve1b:Venus)*^{uq51bh} was generated as previously described[28] but here using an
457 independent genomic integration with the same construct.

458

459 **Generation of maternal zygotic mutants**

460 Germline replacement was performed using embryonic transplantation as previously
461 described [4,57]. Maternal zygotic (MZ) *prox1a* mutant embryos were made by crossing germline
462 replaced *prox1a*$^{i278-/-}$;*prox1b*$^{sa0035+/-}$ females with *prox1a*$^{i278+/-}$;*prox1b*$^{sa0035+/-}$ males [4,57].
463 Genotyping of individual embryos during transplantation and phenotypic analysis was
464 performed as previously described [4].

465

466 **Transgenesis, genome-editing and genotyping**

467 All microinjections were performed as previously described [58]. *slc7a7a*^{BAC}*:slc7a7a-Citrine*^{uom10}
468 and *fabp11a*^{BAC}*:fabp11a-Citrine*^{uom10} recombineering was performed as previously described
469 [59].

470 Primers for BAC recombineering:

471 *slc7a7a*-BAC-Citrine-forward:5'-
472 AACTGCTTTAGACAGTGTTTTTTGGTACCATCCCATATATTTAAAAAACAGCCACCATGGTGA
473 GCAAGGGCGAGGAG-3'

474 *slc7a7a*-BAC-Citrine-reverse:5'-
475 TTCGACACCTCAGGGGATGCCTCTTCTGCAGGCGTAGGGCTGTAGGACGCTCAGAAGAACT
476 CGTCAAGAAGGCG-3'

477 *fabp11a*-BAC-Citrine-forward:5'-
478 TTACAGCTGTTGCGAGATTGAAAAGTAGAGGAGCATCATTATTCGGGAAAGCCACCATGGT
479 GAGCAAGGGCGAGGAG-3'

480 *fabp11a*-BAC-Citrine-reverse:5'-
481 TCAAAGTTGTCGCTGGTGGTCATTTTCCACGTTCCTACGAATTTGTCAACTCAGAAGAACTC
482 GTCAAGAAGGCG-3'

483 **Enhancer detection transgenesis and analysis**

484 A 501bp PCR fragment of *cdh6* enhancer (*chr2:28150709-28151209*) was cloned into the
485 zebrafish enhancer detection (ZED) vector[40]. Empty ZED vector was injected as previously
486 described[40]. Briefly, 1nL of construct at 40 ng/µl or 45ng/µL and tol2 transposase mRNA at
487 100 ng/µl or 55ng/µl was injected into the one-cell stage wild type zebrafish embryos. All F0
488 embryos were screened for skeletal muscle DsRed2 expression. Stable F1 embryos were
489 imaged on a Zeiss LSM 780 or Leica TCS SP8 DLS microscope. Images were processed using

490 ImageJ 2.0.0. or 2.3.0.

491

492 **Imaging and quantification**

493 Imaging was conducted at the Centre for Advanced Histology and Microscopy (Peter
494 MacCallum Cancer Centre). Imaging of live samples was performed using a Zeiss LSM 710 FCS
495 confocal microscope, a Zeiss LSM 780 FCS confocal microscope, or an Olympus FV3000
496 confocal microscope. Mounting and imaging were performed as previously described [60]. In
497 **Figs 6e-f,i-k**, and **Extended data Fig 5f-h, k-m** quantification of vascular phenotypes was
498 performed as previously described [60,61]. In Figs 3h-k, *cdh5:Venus* or *kdrl:GFP* intensity in the
499 TD was measured using Imaris software (Bitplane) and normalised to fluorescence intensity
500 of the DA for *cdh5* and to the fluorescence intensity PCV for *kdrl* in the same embryos.

501

502 **Fluorescence activated cell sorting**

503 FACS was performed at the Peter MacCallum Cancer Centre and the University of
504 Queensland. For **Fig 1**, we isolated cells using the following transgenic lines: 40hpf,
505 *Tg(fli1a:nEGFP)^{y7}*; 3,4 and 5dpf, *Tg(-5.2lyve1b:Venus)^{uq47bh}, Tg(kdrl:Has.HRAS-mCherry)^{s916}*
506 For **Fig 3 and Fig 4**, we isolated cells using the following transgenic lines: 4dpf,
507 *Tg(fli1a:nEGFP)^{y7}, Tg(-5.2lyve1b:DsRed)^{nz101}*. For **Fig 6**, we isolated cells using the following
508 transgenic lines: 40hpf, *Tg(fli1a:nEGFP)^{y7}* .To dissociate embryos and obtain single cell
509 suspensions, we followed published protocols [62]. Briefly, at the desired developmental stage
510 we deyolked embryos by pipetting up and down and rinsing in calcium free ringers solution,
511 we centrifuged at 2000rpm for 5' at 4°C and remove supernatant and dissociated the cells by
512 incubating in liberase [2.5mg/mL] (Cat #5401119001 Sigma-Aldrich) diluted at a 1:35 ratio in
513 DPBS at 28.5°C for approximately 5', homogenizing the samples during and after the
514 incubation. To stop the reaction, we added CaCl2 to a final concentration of 1-2mM and FBS
515 to a final concentration of 5-10%. We centrifuged at 2000rpm for 5' at 4°C. discarded the
516 supernatant, in order to be able to asses live vs. dead cells, we re-suspend the cells solution
517 in Zombie Violet TM Viability Dye (Cat# 423113, BioLegend) and incubated for 20' at RT softly
518 rocking, we rinsed the cells by centrifuging and resuspending in DPBS/EDTA, for ATAC-seq
519 experiments samples were resuspended in 2%BSA/PBS. Suspension was filtered through a
520 strainer and taken to the FAC sorting facility. In the Flow Cytometry facility, we used the BD
521 FACS Aria Fusion sorter (BD Biosciences), we based the selection for the desired population

522 on FSC and SCC, alive cells were selected based on the Zombie Violet profile and double
523 positive cells for the desired transgenics were targeted according to the expression profiles
524 of single cells. Double positive cells were sorted in 300uL 100% FBS in a cold block and taken
525 immediately to the sequencing facility.
526
527 **scRNA-seq library preparation**
528 Library preparation and sequencing was performed at the Institute for Molecular Bioscience
529 Sequencing Facility (University of Queensland) or Peter Mac Genomics Facility. Single cell
530 suspensions were sorted by FACS, spun down to concentrate and a cell count was performed
531 to determine post-sort viability and cell concentration. Single cell suspension was partitioned
532 and barcoded using the 10X Genomics Chromium Controller (10X Genomics) and the Single
533 Cell 3' Library and Gel Bead Kit (V2 10X Genomics PN-120237; V3.1; 10X Genomics; PN-
534 1000123). The cells were loaded onto the Chromium Single Cell Chip A (10X Genomics; PN-
535 120236), B (10X Genomics; PN-1000073 or PN-1000074) or G (10X Genomics; PN-1000120)
536 to target 10,000 cells. GEM generation and barcoding, cDNA amplification, and library
537 construction was performed according to the 10X Genomics Chromium User Guide. The
538 resulting single cell transcriptome libraries contained unique sample indices for each sample.
539 The libraries were quantified on the Agilent BioAnalyzer 2100 using the High Sensitivity DNA
540 Kit (Agilent, 5067-4626). Libraries were pooled in equimolar ratios, and the pool was
541 quantified by qPCR using the KAPA Library Quantification Kit - Illumina/Universal (KAPA
542 Biosystems, KK4824) in combination with the Life Technologies Viia 7 real time PCR
543 instrument. After the initial sequencing run, libraries were re-pooled according to estimated
544 captured cells as determined using the Cell Ranger software (10X Genomics).
545
546 **Sequencing of scRNA-seq libraries**
547 At the IMB(UQ) genomics facility, denatured libraries were loaded onto an Illumina NextSeq-
548 500 and sequenced using a 150-cycle High-Output Kit as follows: 26bp (Read1), 8bp (i7 index),
549 98bp (Read2). Read1 supplies the cell barcode and UMI, i7 the sample index, and Read2 the
550 3’ sequence of the transcript. At the Peter Mac Molecular Genomics facility, single cell
551 transcriptome libraries were sequenced on an Illumina NovaSeq 6000 using S4 300-cycle
552 chemistry. Read1 supplies the cell barcode and UMI, i7 the sample index, and Read2 the 3’
553 sequence of the transcript. Sequencing read lengths were trimmed to 28bp (Read1), 8bp (i7

554 index), 91bp (Read2), ensuring compatibility with the 10X Genomics analysis software, Cell
555 Ranger.
556
557 **scRNA-seq data processing and analysis**
558 All code and documentation associated with this analysis is publicly available under an open
559 source software license at:
560 https://atlassian.petermac.org.au/bitbucket/users/tyrone.chen/repos/hogan_lab/browse/2
561 022_Grimm_Mason_et_al_PROX1/
562 Relevant functions are in italics for reference. Where necessary fastq files were made using
563 *Cell Ranger* [63] (version 3.1.0 or 3.0.2) *mkfastq*. Sequencing QC was assessed using *FastQC*
564 0.11.6 and *MultiQC* viewer for aggregated reports. *Cell Ranger count* and *aggr* were used to
565 generate aggregated count files mapped to GRCz11 (Ensembl 101), without depth
566 normalisation. Doublets were identified from the filtered aggregated count files using
567 *Scrublet* [64] in *Python* version 3.6 and filtered from subsequent analyses. For the *MZprox1*$^{-/-}$
568 mutant and *Zprox1*$^{-/-}$ mutant datasets filtered aggregated count files were processed, sc-
569 transform normalised, filtered and clustered (louvain) using *Seurat* version 2.0 [65] and 3.0 [34]
570 respectively for *R statistical software* version 4.0.2. QC was evaluated before and after
571 normalisation using plot functions in *Seurat* and *scater* 1.20.1 [66], and all thresholds and
572 settings are described in scripting. Cluster solutions were evaluated using *ClusTree* [67].
573 Datasets used in the atlas of lymphangiogenesis were processed, filtered, merged and log-
574 normalised using *Seurat* version 3.0 [34], with QC and settings as above. Merged data was
575 clustered and normalised using *CSS simspec* [33] and clustering and cluster evaluation
576 performed on this object only, as described above. For all scRNA-seq datasets cluster
577 phenotype was determined using the expression of key markers (Supplementary Table), with
578 the aid of *CellXGene* visualisation software [68]. All downstream analysis and plotting were
579 performed using *Seurat* version 3.0 [34] with default settings. All gene ontology analyses were
580 performed using Panther.db [37] (Biological Process Complete).
581
582 **Preparation of single nuclei for snATAC-seq**
583 Single cell suspensions were sorted by FACS and prepared for nuclei isolation as previously
584 described by 10x Genomics Demonstrated Protocol for Single Cell ATAC Sequencing
585 (CG000169 - Rev D). Cell suspensions were pelleted (300 x g for 5 minutes) and rinsed with

586 PBS + 0.04% BSA. Cells were resuspended in 95uL of freshly prepared lysis buffer (10mM Tris-
587 HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% Tween-20, 0.1% NP40 Substitute, 0.01%
588 Digitonin, and 1% BSA) and incubated on ice for 1 minute. 100uL of chilled wash buffer
589 (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% Tween-20, 1% BSA) was used to
590 neutralise the reaction, before the nuclei were pelleted (500 x g for 5 minutes) and
591 resuspended again in 7uL of 1x Nuclei Buffer (10X Genomics Cat# PN-2000153/2000207).
592 Presence of healthy and intact nuclei was assessed by visual inspection on a brightfield
593 microscope using Trypan Blue staining (Thermo Fisher Cat# T10282) and Countess Cell
594 Counting Chamber Slides (Thermo Fisher Cat# C10228).
595
596 **snATAC-seq library preparation and sequencing**
597 Single nuclei suspensions were resuspended at approximately 5000 nuclei per μL before
598 undergoing tagmentation for 60min at 37°C. After tagmentation nuclei were partitioned and
599 barcoded using the 10X Genomics Chromium Controller (10X Genomics) and the Single Cell
600 ATAC Reagent Kit (V1.1; 10X Genomics; PN-1000176). Tagmented nuclei were loaded onto
601 the Chromium Single Cell Chip H (10X Genomics; PN-1000162), GEM generation, barcoding
602 and library construction was performed according to the 10X Genomics Chromium User
603 Guide. The resulting single cell ATAC libraries contained unique sample indices for each
604 sample. The libraries were quantified on the Agilent BioAnalyzer 2100 using the High
605 Sensitivity DNA Kit (Agilent, 5067-4626) and pooled in equimolar ratios. Sequencing was
606 performed on an Illumina NextSeq 500 using a 150-cycle High-Output Kit as follows: 50bp
607 (Read1), 8bp (i7 index), 16bp (i5 index), 50bp (Read2) achieving a read depth of 25,000 read
608 pairs per nucleus.
609
610 **snATAC-seq processing and analysis**
611 FASTQ files generated from sequencing were used as inputs to 10X Genomics *Cell Ranger*
612 *ATAC 2.0.0*. *cellranger-atac count* was used to generate count files mapped to GRCz11
613 (ENSEMBL 101), without depth normalisation. Resulting fragment files were read into *ArchR*
614 *1.0.1* for *R statistical software 4.0.5* as a tile matrix with 500-bp bins. All remaining steps in
615 the ATAC-Seq analysis were performed within *ArchR 1.0.1*. QC filtering was performed, and
616 only high-quality cells with a TSS enrichment score greater than 4 and greater than 1,000
617 unique nuclear fragments were retained. Doublets were predicted using *addDoubletScores*

618 and filtered using *filterDoublets.* Data normalization and dimensionality reduction were
619 performed using iterative Latent Semantic Indexing (LSI) and Uniform Manifold
620 Approximation and Projection (UMAP) embeddings were used for visualisation in reduced
621 dimension space. Separate from the *ArchR 1.0.1* package, cluster solutions were
622 independently evaluated using *clustree 0.4.3*. A Gene Score Matrix that stores predicted gene
623 expression was then generated based on the accessibility of regulatory elements in the
624 vicinity of the gene. We used gene scores of endothelial markers for cluster annotation and
625 subsetting. Differentially accessible genes were identified by differential testing using
626 *getMarkerFeatures*. Local chromatin accessibility of the marker genes was visualised using
627 the *plotBrowserTrack*.
628
629 **Gene regulatory network analyses**
630 For gene regulatory network construction we leveraged the ENCODE transcription factor
631 targets gene-attribute edge list from the Harmonize database
632 https://maayanlab.cloud/Harmonizome/dataset/ENCODE+Transcription+Factor+Targets
633 that includes information for n=181 transcription factors from ChIP-seq analyses[42]. Gene lists
634 generated from scRNA-seq or snATAC-seq analysis were mapped from Zebrafish genome
635 version GRCz11 to Human genome version GRCh38.p13 using ENSEMBL Biomart [69]. These
636 mapped gene sets were used to select relevant edges, that were visualised as a degree-sorted
637 circular network in Cytoscape[70].
638
639
640
641

FIGURES AND FIGURE LEGENDS

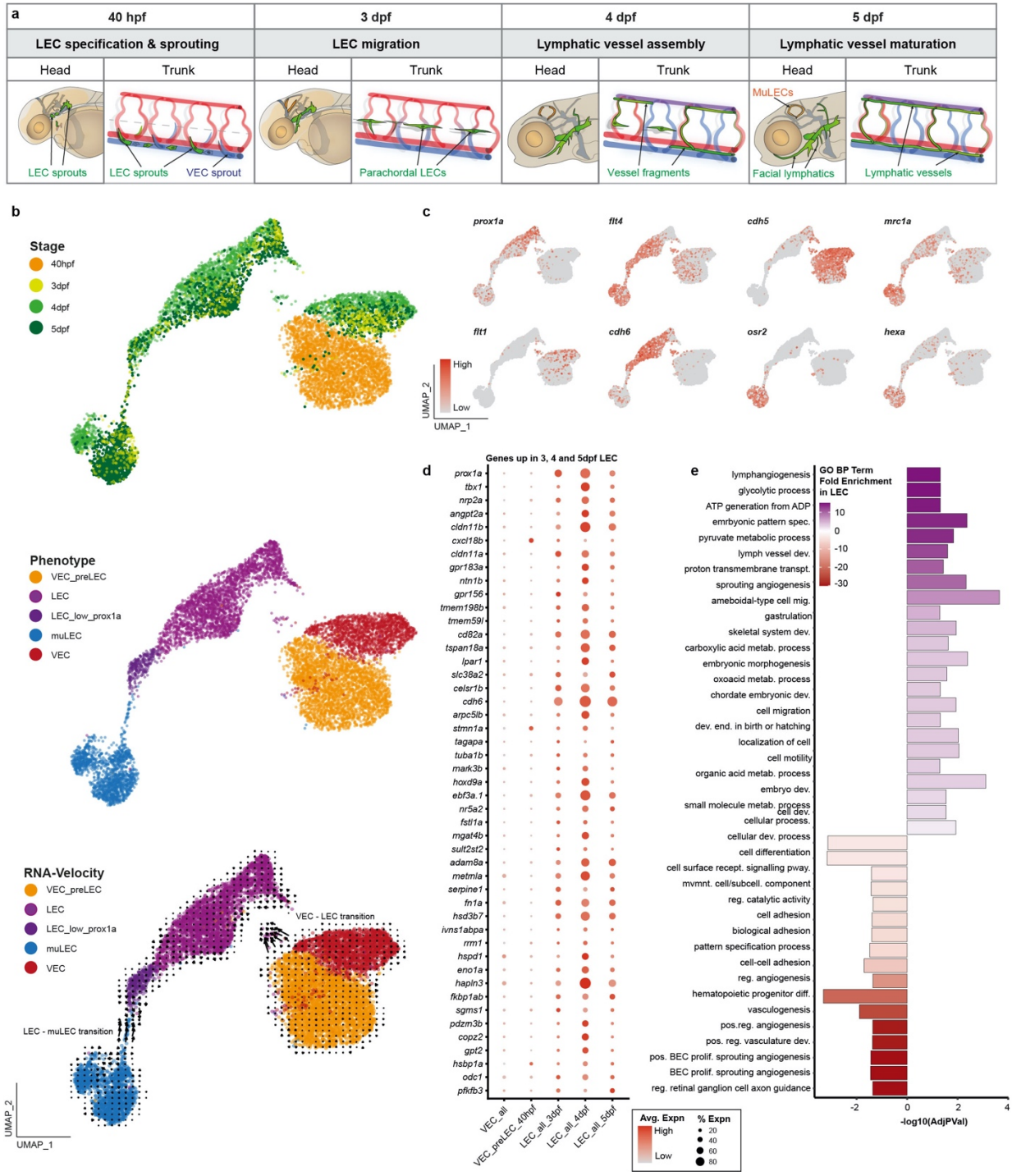


Figure 1: A single cell RNAseq developmental atlas of lymphangiogenesis in zebrafish

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.

649 **b.** UMAP visualization of n=9,771 cells filtered for VEC and LEC populations (n=6 samples;
650 see **Extended data table 1a** for cluster identification and **Extended data Fig 1** for
651 whole dataset) coloured according to developmental stage (top), predicted cell
652 phenotype (middle), and RNA-velocity (bottom).
653 **c.** UMAP visualization of key marker gene expression. Colour scale represents log-
654 normalised expression. LEC markers: *prox1a, cdh6*. BEC markers: *cdh5*. LEC and VEC
655 markers: *mrc1a, flt4*. AEC marker: *flt1*. muLEC markers: *osr2, hexa*.
656 **d.** Dot plot of top genes commonly marking LECs at 3, 4 and 5dpf, with expression
657 displayed in VECs (all stages), VEC_preLEC (40hpf) and LECs at 3, 4 and 5dpf (see also
658 **Extended data table 1b**, **Extended data Fig 1f**). Colour scale represents average log-
659 normalised expression and point size represents percentage of cells expressing gene.
660 **e.** Bar plot summarizing GO BP analyses of genes DE between 3, 4 and 5dpf LEC and VEC
661 populations (**Extended data table 1c-d**). Y-axis represents enriched BP term, x-axis
662 represents the -log10(*adjusted p value*), bars are coloured and ordered according to
663 fold enrichment of the GO term in LEC. GO terms enriched in LEC are coloured purple,
664 terms enriched in VEC are coloured red.
665 Hpf, Hours post-fertilisation.dpf, Days post fertilisation. BEC, Blood vascular endothelial cell.
666 VEC, Venous endothelial cell. LEC, Lymphatic endothelial cell. muLEC, mural Lymphatic
667 endothelial cell (a.k.a FGP or brain LEC).
668
669
670

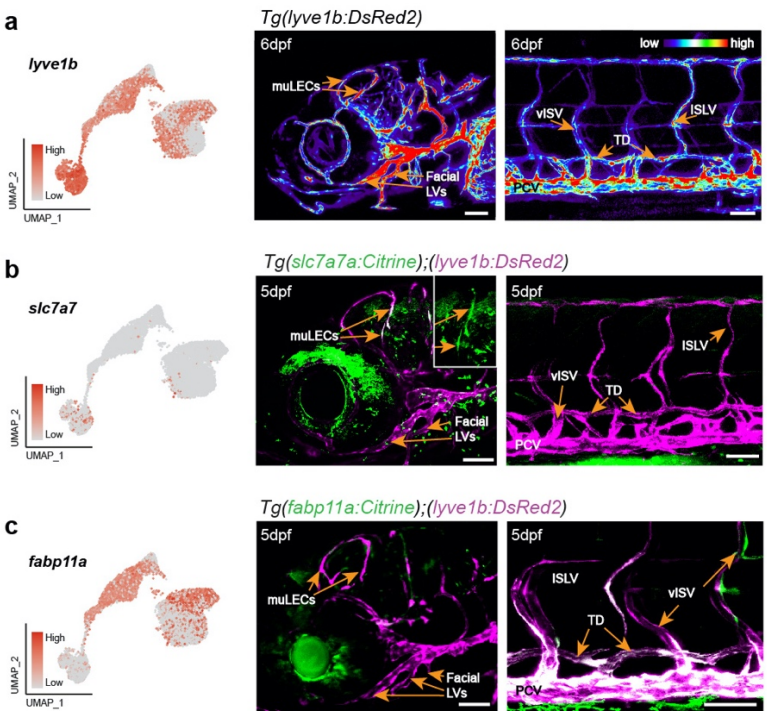


Figure 2: Transgenic marker strains confirm cluster identity and identify vessel specific markers

- **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 *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 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 µm for head (middle) in (a) and 50 µm for trunk images (right).

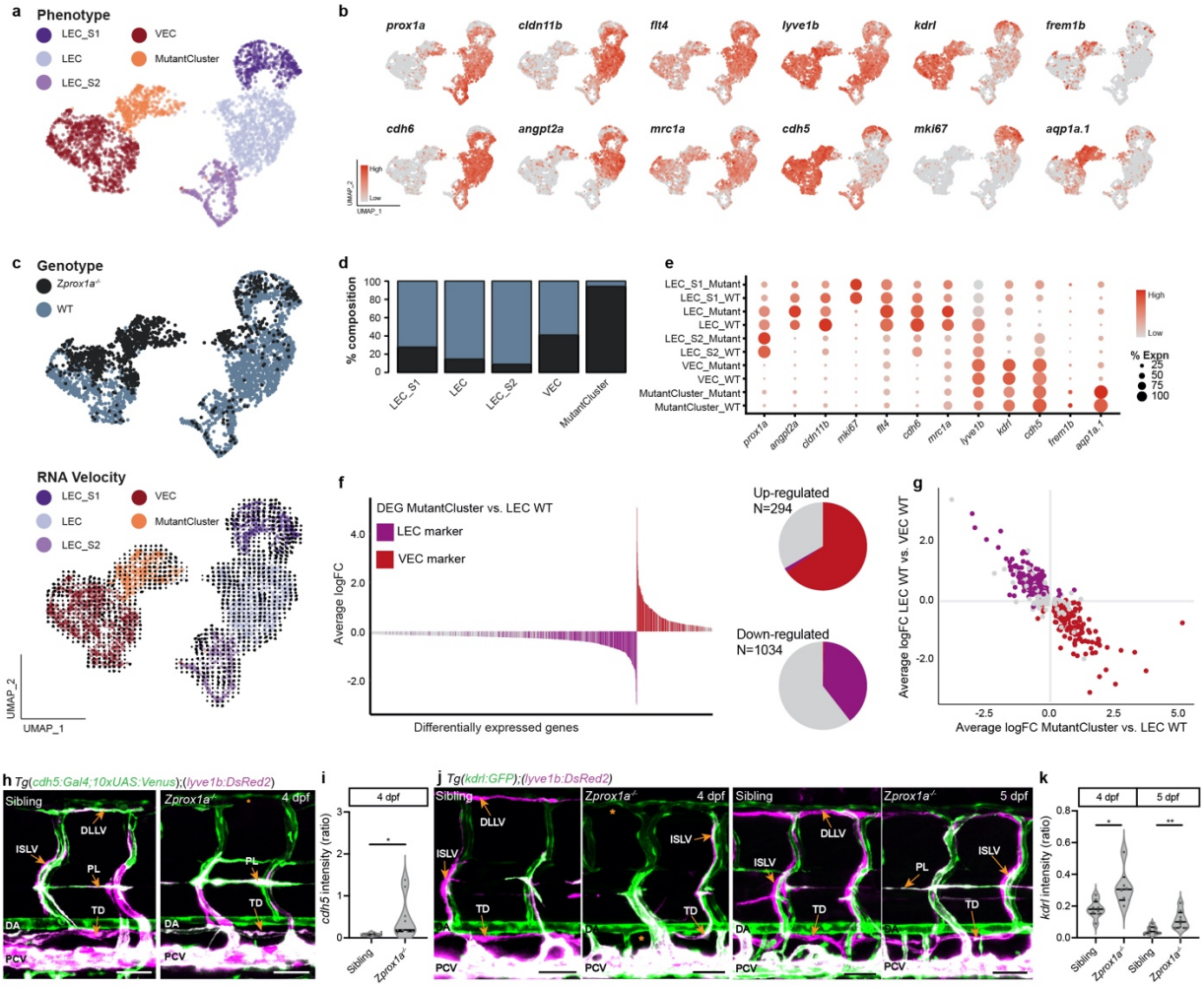


Figure 3: Single cell RNA seq analysis reveals a fate shift from LEC to VEC in the absence of Prox1 in zygotic *prox1a* mutants

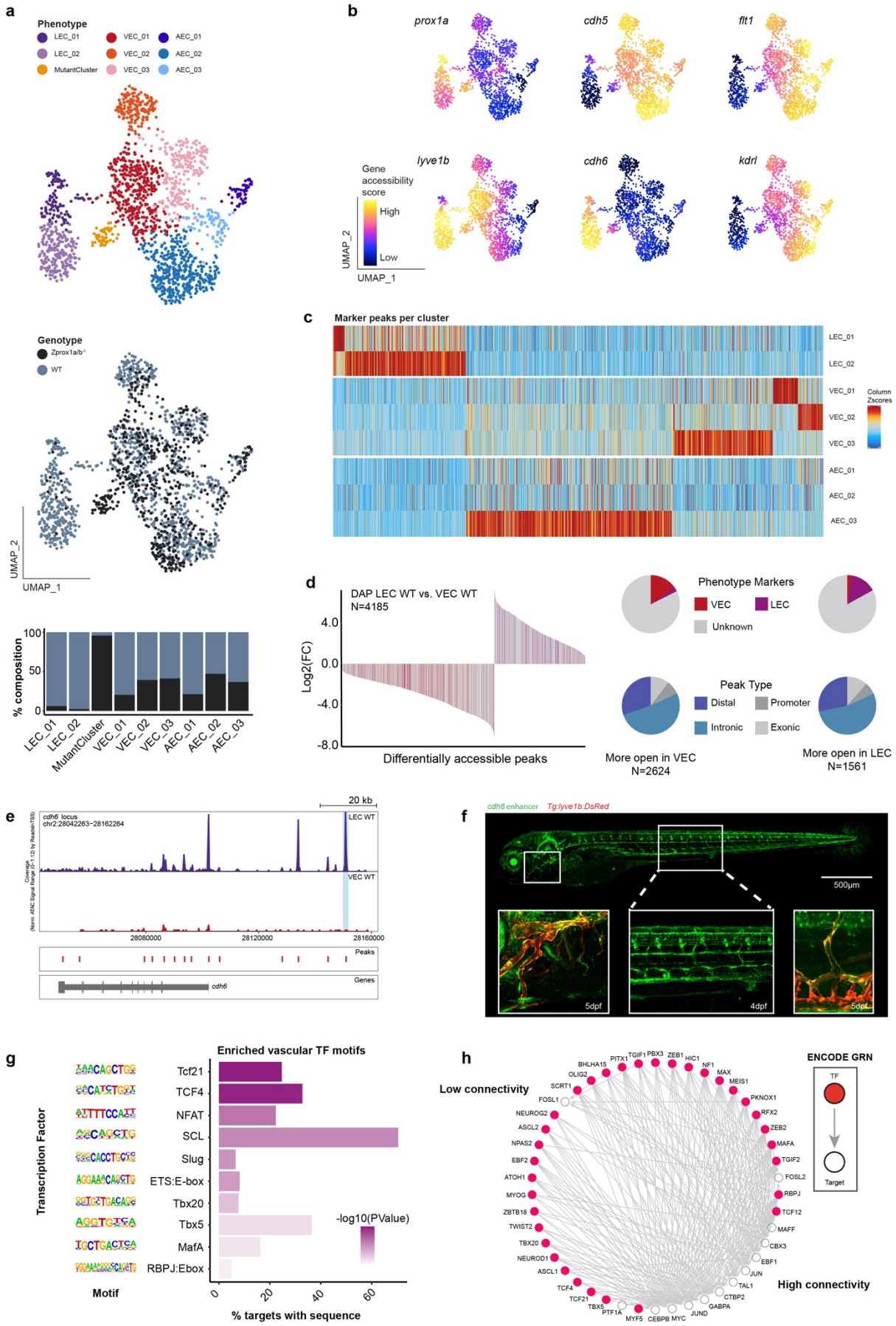
a. UMAP visualization of n=8,075 cells filtered for VEC and LEC subpopulations (n=2 samples; see **Extended data table 2a** for cluster annotation, **Extended data Fig3** for full dataset) coloured according to predicted cell phenotype.

b. 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.*

c. 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.

d. Stacked bar plot representing the genotype composition of cells in a given phenotypic cluster.

706 **e.** Dot plot of marker expression across defined clusters (indicated on Y-axis). Colour
707 scale represents average SCT-normalised expression and point size represents
708 percentage of cells expressing gene.
709 **f.** Bar plot of average log fold change in gene expression comparing WT LECs with the
710 Mutant Cluster. Y-axis represents average log fold change, x-axis represents
711 differentially expressed genes (Wilcoxin Rank Sum *adjusted p value* < 0.05) and bars
712 are coloured according to status as a LEC (purple) or VEC (red) marker in **Fig. 1**. Pie
713 charts (right) indicate the LEC and VEC marker composition of genes up-regulated
714 (n=365) and down regulated (n=1,922) in the Mutant Cluster, demonstrating a fate
715 shift.
716 **g.** Concordance in the fate shift between Mutant Cluster and WT LEC with the WT LEC
717 and VEC trajectory. Each point represents significant DE genes (n=2,287) between
718 Mutant Cluster and LEC WT, coloured according to LEC or VEC marker status. X-axis
719 represents average log fold change relative to the Mutant Cluster, Y-axis represents
720 average log fold change relative to LEC WT.
721 **h.** Lateral confocal images of *cdh5* (green) and *lyve1b* (magenta) expression in the
722 developing trunk in WT and *Zprox1a* mutants at 4dpf.
723 **i.** Quantification of *cdh5* intensity in the thoracic duct in WT and mutants (relative to
724 expression in the DA).
725 **j.** Lateral confocal images of *kdrl* (green) and *lyve1b* (magenta) expression at 4dpf (left)
726 and 5dpf (right).
727 **k.** Quantification of *kdrl* intensity in the thoracic duct in WT and mutants (relative to
728 expression in the PCV) at 4 and 5dpf.
729 WT, wildtype. Z, zygotic. TD, thoracic duct, DA, Dorsal Aorta. PCV posterior cardinal vein. ISLV,
730 intersomitic lymphatic vessel. DLLV, Dorsal longitudinal lymphatic vessel. PL, parachordal LEC.
731 Scale bars, 50 µm. *, The error bars represent mean ± s.e.m.; *p=0.01 and **p=0.0094 from
732 an unpaired, two-sided t-test.
733


a
Phenotype
LEC_01
LEC_02
MutantCluster
VEC_01
VEC_02
VEC_03
AEC_01
AEC_02
AEC_03
Genotype
Zprox1a/b-
WT
UMAP_2
UMAP_1
% composition
b
prox1a
cdh5
flt1
lyve1b
cdh6
kdrl
Gene accessibility score
High
Low
c
Marker peaks per cluster
Column Zscores
d
DAP LEC WT vs. VEC WT
N=4185
Log2(FC)
Differentially accessible peaks
Phenotype Markers
VEC
LEC
Unknown
Peak Type
Distal
Promoter
Intronic
Exonic
More open in VEC
N=2624
More open in LEC
N=1561
e
20 kb
cdh6 locus
chr2:28042263-28162264
LEC WT
VEC WT
Peaks
Genes
cdh6
f
cdh6 enhancer
Tg:lyve1b:DsRed
500μm
5dpf
4dpf
5dpf
g
Enriched vascular TF motifs
Transcription Factor
Tcf21
TCF4
NFAT
SCL
Slug
ETS:E-box
Tbx20
Tbx5
MafA
RBPJ:Ebox
Motif
-log10(PValue)
% targets with sequence
h
Low connectivity
High connectivity
ENCODE GRN
TF
Target

735 **Figure 4: Single nuclei ATAC seq identifies lineage specific regulatory regions in VECs and**
736 **LECs**

737 **a.** UMAP visualization of snATAC-seq from *Zprox1a*$^{-/-}$ mutant and WT endothelial cells at
738 4dpf (n=3,731 nuclei), coloured by cell phenotype (top) and genotype (middle) (see
739 **Extended data Fig 4** for full dataset). Stacked bar plot (bottom) summarising the
740 composition of wildtype and mutants in each cluster. A mutant specific cluster is
741 identified, similar to **Fig 3**.

742 **b.** UMAP visualization indicating accessibility of key marker genes (gene accessibility
743 score with imputation) that confirm predicted cluster phenotypes. LEC markers:
744 *prox1a, cdh6*. VEC markers: *cdh5*, *flt1*, *kdrl.*

745 **c.** Heatmap of cluster specific accessible peaks defined using DAP analysis, for all
746 endothelial cells. Colour reflects a column-wise Z-score, rows represent clusters
747 defined in **Extended data Fig 4a** and columns are peaks.

748 **d.** Bar plot of log2 fold change for the differentially accessible peaks (n=4,185 peaks,
749 Wilcoxon Rank Sum, *FDR* < 0.05) between WT LECs and WT VECs. LEC/VEC markers
750 identified in scRNA-seq are coloured in red/purple respectively and demonstrate a
751 strong correlation between chromatin state and the VEC to LEC fate shift. 228/1,561
752 more accessible peaks in the LEC cluster were associated with LEC markers and
753 398/2,624 more accessible peaks in VEC cluster associated with VEC markers. Pie
754 charts (right) summarising the proportion of LEC/VEC markers in the differentially
755 accessible peaks (top) and the type of peak (bottom).

756 **e.** Genome accessibility track of LEC marker *cdh6*. Red bars represent peaks in the
757 reproducible peak set from snATAC-seq. The peak highlighted blue indicates a
758 potential enhancer of *cdh6* (*chr2:28150709-28151209*) with significantly more
759 accessible chromatin in WT LECs compared to WT VECs (Wilcoxon Rank Sum, *FDR* <
760 0.05).

761 **f.** Overall GFP expression of *cdh6* enhancer (*chr2:28150709-28151209*) reporter at 4dpf
762 (top). Lateral confocal image of *cdh6* enhancer reporter in the trunk at 4dpf (bottom
763 middle). Co-expression of *cdh6* enhancer reporter (green) and *lyve1b* (red) in the
764 facial lymphatics (bottom left) and trunk lymphatics (bottom right) at 5dpf.

765 **g.** Vascular TF motifs (n=10 from top n=50) enriched in peaks that are more permissive
766 in WT LEC compared to WT VEC. The depth of bar colour represents the -

767 log10(*RawPVal*), y-axis displays individual motifs and schematic (left) and x-axis
768 represents the percentage of target regions enriched for the motif in the n=1561 peak
769 set more open in WT LEC.
770 **h.** Degree-sorted gene regulatory network displaying known TF binding at genes with
771 more permissive chromatin in WT LECs compared to WT VECs. TFs are represented
772 by red circles (nodes), target genes by white circles (nodes), and known binding of TF
773 to target by a grey arrow (edges). Nodes with a larger number of edges are more
774 highly connected (bottom right), and nodes with fewer edges are less connected (top
775 left).
776
777 DAP, differentially accessible peaks. FDR, false discovery rate. snATAC-seq, single nuclei
778 ATAC-seq. TF, transcription factor.
779

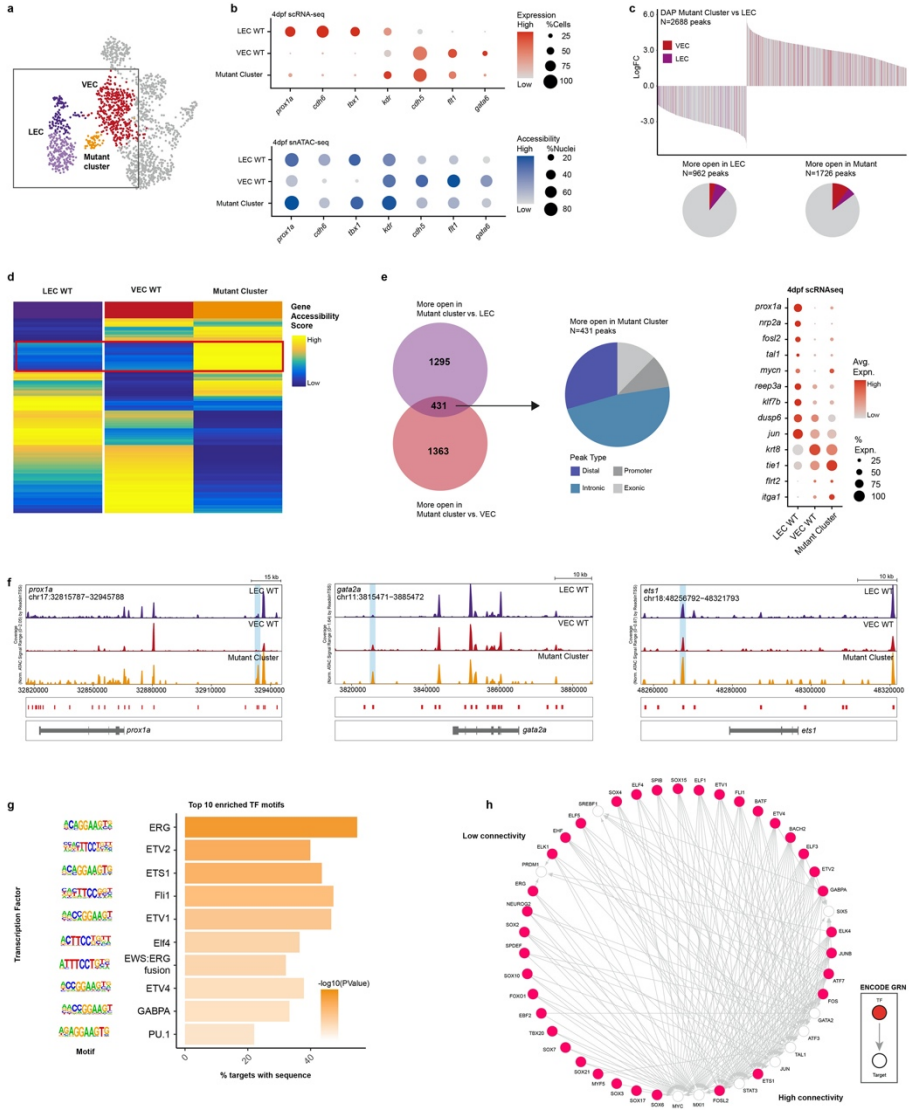


Figure 5: Zygotic *prox1a* mutants display a unique chromatin accessibility state consistent with increased activity of early blood and blood vascular fate transcription factors

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, kdr*) 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. 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.

793 **c.** Bar plot of log2 fold change for the differentially accessible peaks (n=2,688 peaks,
794 Wilcoxon Rank Sum, *FDR* < 0.05) between Mutant Cluster and WT LECs. LEC/VEC
795 markers identified in scRNA-seq are coloured in purple/red respectively and
796 demonstrate regions more open in the Mutant are associated with more vascular than
797 lymphatic genes. Pie charts (bottom) summarising the proportion of LEC/VEC markers
798 in the differentially accessible peaks.
799 **d.** Heatmaps of accessibility (gene accessibility score) for all genes (n=32,020) in WT LECs,
800 WT VECs and Mutant Cluster showing that mutant cluster cells display a unique
801 chromatin state at many genes. Red box indicates genes that are more permissive in
802 the Mutant Cluster than either WT LEC or WT VEC. Colour indicates level of
803 accessibility.
804 **e.** Venn diagram (left) indicates all individual peaks with increased in accessibility in the
805 Mutant Cluster vs LEC or VEC, with n=431 DAP commonly more open than in both LEC
806 WT and VEC WT (*FDR* < 0.05, log2 fold change > 1.5). Pie chart (middle) indicates the
807 proportion of peak types for n=431 DAP classified as distal, intronic, promoter and
808 exonic respectively. Dot plot (right) summarizing the scRNA-seq expression level of
809 n=13 genes with DAP more open in the Mutant Cluster at 4dpf, demonstrating that
810 accessibility changes for these genes did not correlate with changes in transcription.
811 The size of the dot represents the proportion of cells that express the markers in the
812 cluster, and colour represents SCT-normalised expression.
813 **f.** Genome accessibility tracks for key markers with DAP more permissive in the Mutant
814 Cluster: *prox1a, gata2a* and *ets1* (left to right). Red bars represent peaks in the
815 reproducible peak set from snATAC-seq. Blue bars highlight DAP (Wilcoxon Rank Sum,
816 *FDR* < 0.05).
817 **g.** Top 10 enriched motifs (HOMER analysis, *adjusted p value* < 0.05) in the n=431 peaks
818 that are more open in the Mutant Cluster than WT LEC and VEC. The depth of colour
819 represents the -log10(*RawPVal*), y-axis displays individual motifs and schematics (left),
820 and x-axis represents the percentage of peaks enriched for the motif in the n=431
821 peak set.
822 **h.** Degree-sorted gene regulatory network displaying known TF binding at genes with
823 more permissive chromatin in Mutant Cluster compared to WT LECs. TFs are
824 represented by red circles (nodes), target genes by white circles (nodes), and known

825 binding of TF to target by a grey arrow (edges). Nodes with a larger number of edges
826 are more highly connected (bottom right), and nodes with fewer edges are less
827 connected (top left).
828
829 DAP, differentially accessible peaks. FDR, false discovery rate. snATAC-seq, single nuclei
830 ATAC-seq. TF, transcription factor.
831

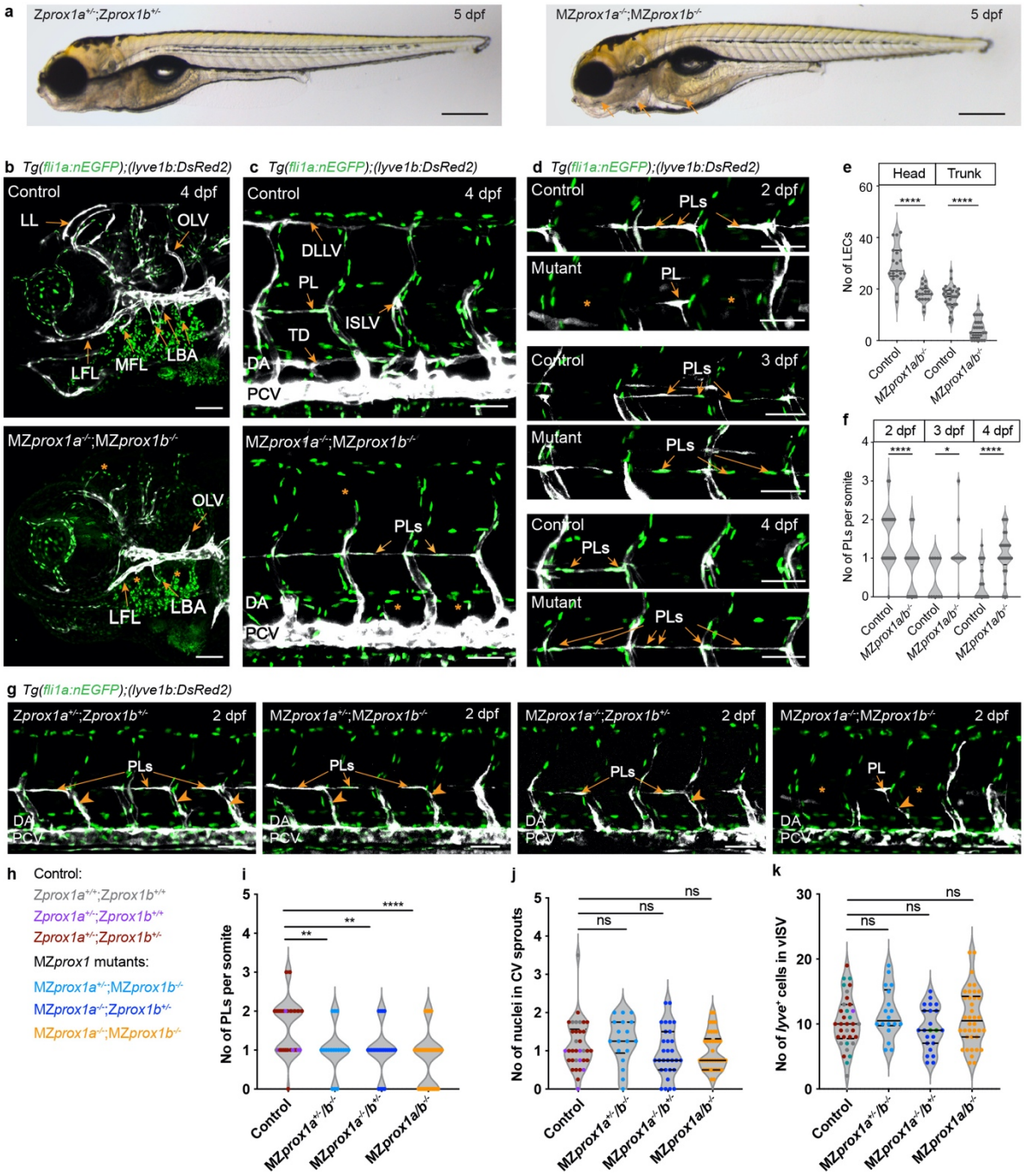
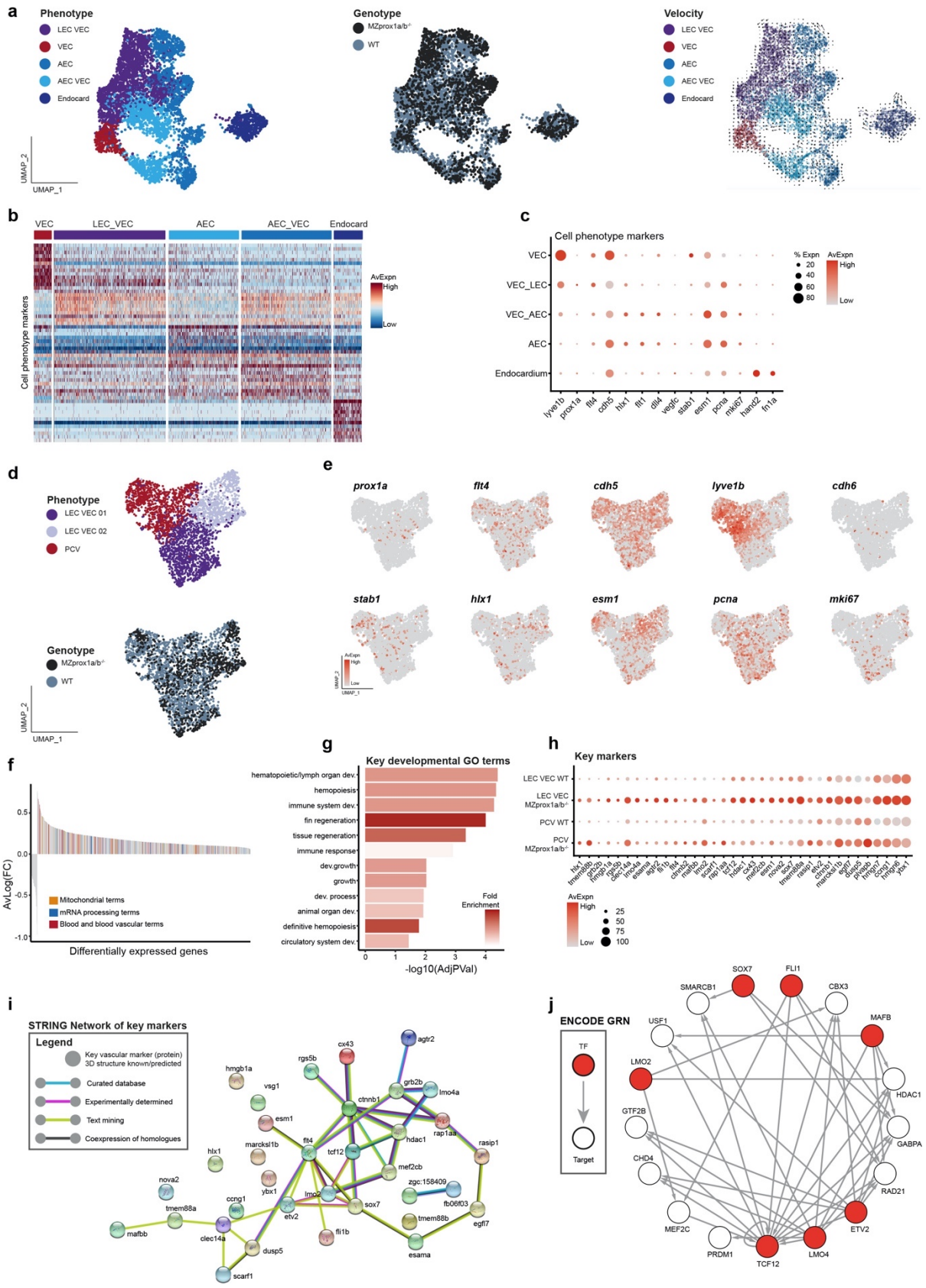


Figure 6: maternal zygotic *prox1a;prox1b* double mutants display a specific loss of lymphatic vessels throughout the developing embryo

a. Overall morphology of control and *MZprox1a*$^{-/-}$*;MZprox1b*$^{-/-}$ mutants at 5dpf. Arrows indicate oedema around eyes, heart and intestine. Scale bars 500 µm.

b. 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.

841 **c.** Lateral confocal images of zebrafish trunks at 4dpf in control (upper) and *MZprox1a*$^{-/-}$
842 *;MZprox1b*$^{-/-}$ mutants (lower), showing absent lymphatic vessels (asterisk) but
843 retained PLs (arrows in lower). Scale bars, 50 µm.
844 **d.** Lateral confocal images of PLs in the horizontal myoseptum at 2, 3 and 4dpf. PLs form
845 later in *MZprox1a*$^{-/-}$*;MZprox1b*$^{-/-}$ mutants and accumulate in the horizontal
846 myoseptum while the PLs of control embryos emigrate the HM by 4dpf. Scale bars,
847 50 µm.
848 **e.** Quantification of lymphatic endothelial cell (LEC) number from 4dpf heads control:
849 n=19, *MZprox1a*$^{-/-}$*;MZprox1b*$^{-/-}$ mutants: n=18) and trunks (control: n=29, *MZprox1a*$^{-/-}$
850 *;MZprox1b*$^{-/-}$ mutants:n=25). Error bars represent mean ± s.e.m.; **** $p < 0.0001$,
851 from an unpaired, two-sided t-test.
852 **f.** Quantification of number of PLs per somite in trunks at 2dpf in controls (n=41) and
853 *MZprox1a*$^{-/-}$*;MZprox1b*$^{-/-}$ mutants (n=38), at 3dpf in controls (n=11) and *MZprox1a*$^{-/-}$
854 *;MZprox1b*$^{-/-}$ mutants (n=10) and at 4dpf in controls (n=29) and *MZprox1a*$^{-/-}$
855 *;MZprox1b*$^{-/-}$ mutants (n=25). Error bars represent mean ± s.e.m.; **** $p < 0.0001$, *
856 $p < 0.05$, from an unpaired, two-sided t-test.
857 **g.** Lateral confocal images showing defects in the formation of PLs in mutants upon loss
858 of *prox1a* compared with controls in the trunk. Genotypes are indicated.
859 **h.** Colour coded list of analysed genotypes abbreviated in (**i-k**). Each embryo has a
860 defined genotype for *prox1b* which is represented in colour code as displayed in j.
861 **i.** The number of PLs formed (per somite) at 2dpf in genotypes indicated. Decreasing
862 gene dosage for *prox1a* and *prox1b* progressively reduced the initial seeding of the
863 HM by PLs.
864 **j.** The number of cells in sprouts departing the PCV is unchanged in mutants showing
865 that sprouting occurs normally.
866 **k.** The number of cells in vISVs is unchanged in mutants indicating no effect on the
867 venous endothelium.
868
869 LL, lymphatic loop. OLV, otholitic lymphatic vessel. LFL, lateral facial lymphatic vessel. MFL,
870 medial facial lymphatic vessel. LBA, lymphatic branchial arches. DLLV, dorsal longitudinal
871 lymphatic vessel. PL, parachordal LEC. ISLV, intersomitic lymphatic vessel. TD, thoracic duct.
872 DA, dorsal aorta. PCV, posterior cardinal vein.



874

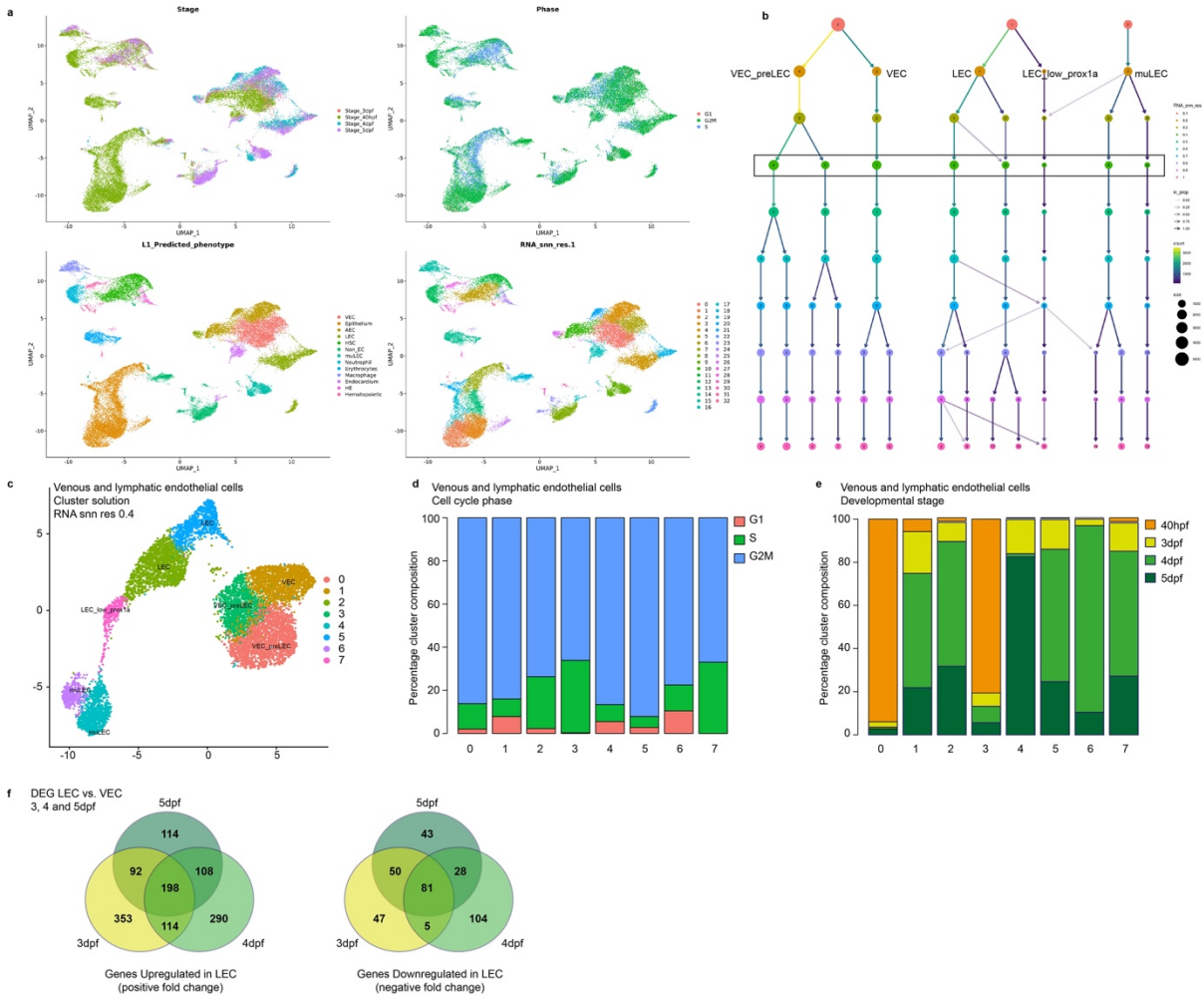
875 **Figure 7: Single cell profiling of maternal zygotic *prox1a;prox1b* double mutants reveals that**

876 **Prox1 homologues initially suppress gene expression to drive a fate transition**

877 **a.** UMAP visualisation of endocardial, venous and arterial endothelial cells (Level 02
878 n=5,347) coloured according to predicted cell phenotype (left), genotype (middle) and
879 RNA-velocity (right).
880 **b.** Heatmap displaying expression of phenotype specific genes defined using differential
881 expression analysis. Columns are cells grouped by phenotype assignment, rows are
882 genes, and colour indicates the average level of expression.
883 **c.** Dot plot indicating expression of key markers used to define cell phenotype. Colour
884 scale represents average SCT-normalised expression and point size represents
885 percentage of cells expressing gene.
886 **d.** UMAP visualisation of venous endothelial cells (Level 03 n=2,747) coloured according
887 to predicted cell phenotype (top) and genotype (bottom).
888 **e.** UMAP visualisation of key marker gene expression. Colour scale represents SCT-
889 normalised expression.
890 **f.** Bar plot indicating average log fold change of n=1,186 significantly different genes
891 between all *MZprox1a/b*$^{-/-}$ and WT venous endothelial cells (LEC VEC 01, LEC VEC 02
892 and PCV combined) at 40hpf (Wilcoxin Rank Sum *adjusted p value* < 0.05). Colour
893 indicates genes associated with one or more mitochondrial, mRNA processing and
894 blood and blood vascular GO terms.
895 **g.** Bar plot summarizing GO BP analysis of the n=1,137 genes upregulated in the
896 *MZprox1a/b*$^{-/-}$ venous endothelial cells. Y-axis represents enriched GO BP term, x-axis
897 represents the -log10(*adjusted p value*) and bars are coloured according to fold
898 enrichment.
899 **h.** Dot plot of n=36 key blood vascular markers upregulated in the *MZprox1a/b*$^{-/-}$ venous
900 endothelial cells, indicating genotype specific expression in LEC VEC and PCV cell
901 phenotypes. Colour scale represents average SCT-normalised expression and point
902 size represents percentage of cells expressing gene.
903 **i.** STRING analysis of n=36 key blood vascular markers upregulated in the *MZprox1a/b*$^{-/-}$
904 venous endothelial cells.
905 **j.** Degree-sorted gene regulatory network displaying known TF binding at n=1,137 genes
906 upregulated in the *MZprox1a/b*$^{-/-}$ venous endothelial cells. TFs are represented by red
907 circles (nodes), target genes by white circles (nodes), and known binding of TF to

target 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).

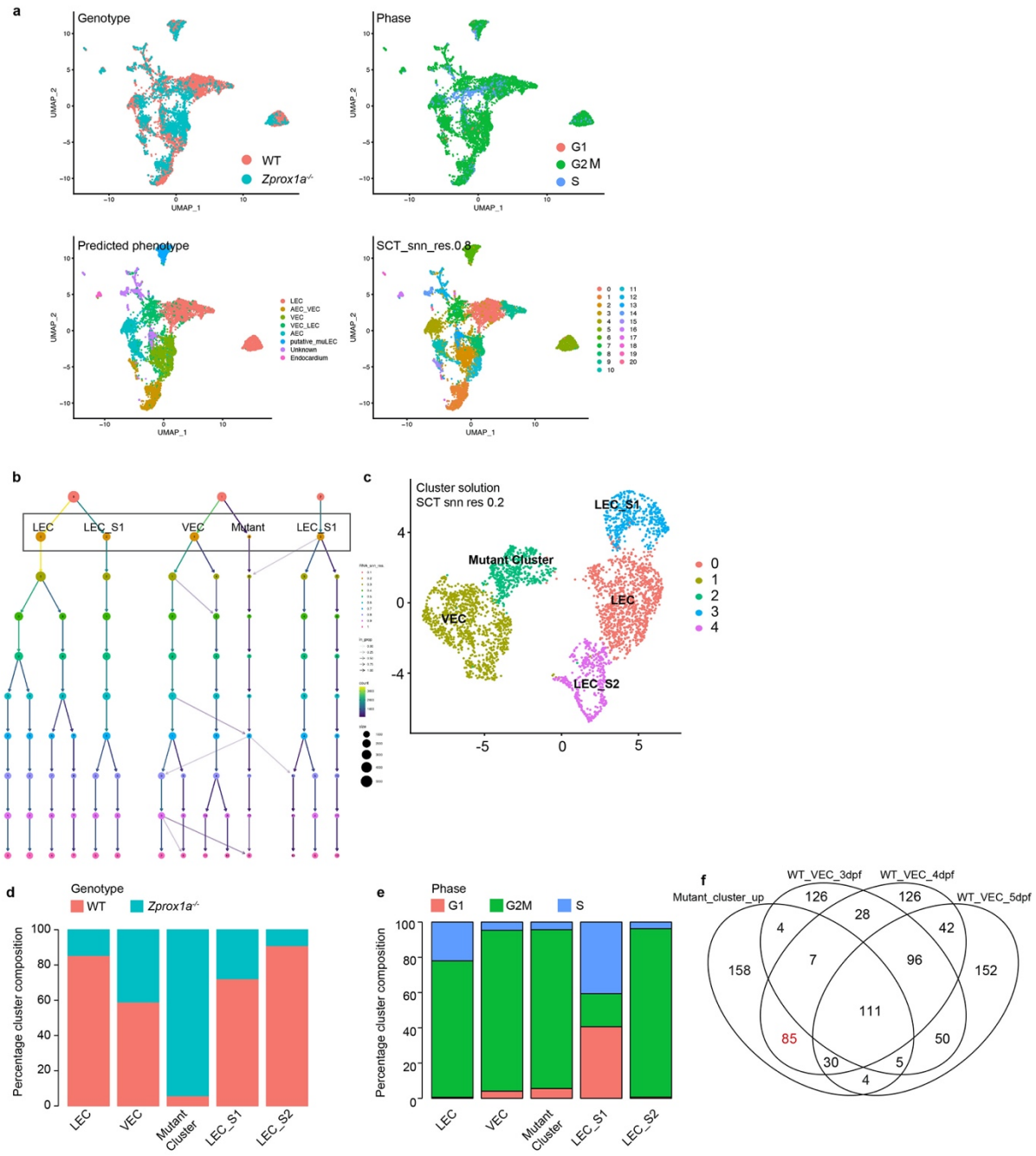
EXTENDED DATA FIGURES AND FIGURE LEGENDS



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

- **a.** 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).
- **b.** 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.
- **c.** UMAP visualization of cluster resolution 0.4 (as appears in **Fig 1**).
- **d.** Stacked bar plot of resolution 0.4 cluster composition cell cycle phase.
- **e.** Stacked bar plot of resolution 0.4 cluster composition developmental stage.

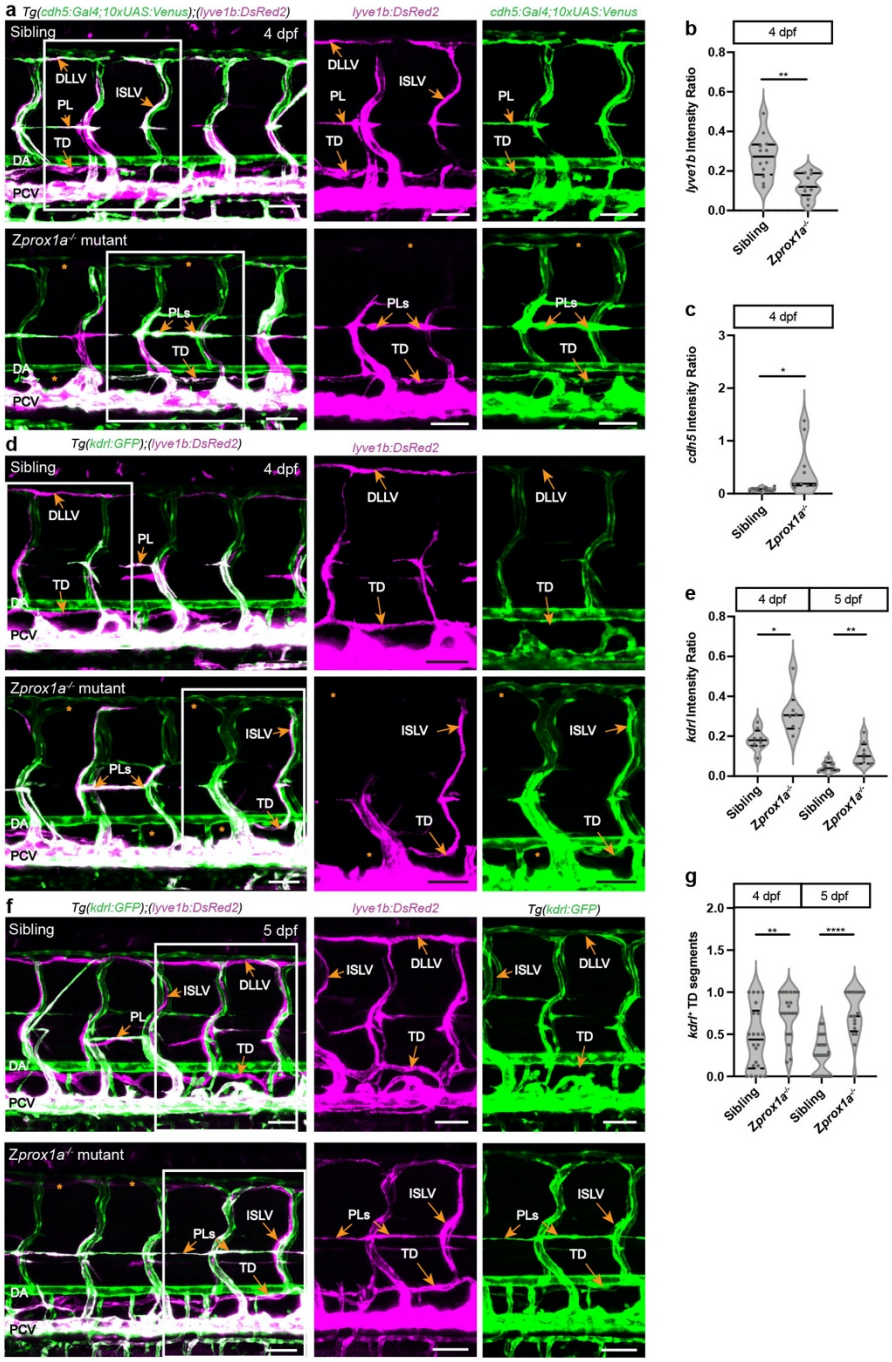
925 **f.** Venn diagram of significant genes commonly upregulated in LEC compared to VEC at
926 3, 4 and 5dpf.
927



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)

935 **b.** Clustree analysis demonstrating the relationship between different cluster resolutions
936 for the n=3063 venous and lymphatic endothelial cells in **Figure 3a**. Resolution 0.2
937 (boxed) was used for all downstream analyses.
938 **c.** UMAP visualization of cluster resolution 0.2
939 **d.** Stacked bar plot of resolution 0.2 cluster composition genotype
940 **e.** Stacked bar plot of resolution 0.2 cluster composition cell cycle phase
941 **f.** Venn diagram indicates the overlap between genes significantly upregulated in VEC
942 compared to LEC at each developmental stage of the single cell atlas, and those
943 upregulated in the *Zprox1a*$^{-/-}$ Mutant Cluster at 4dpf compared to WT LECs. N=85
944 genes (coloured red) are uniquely associated with the Mutant Cluster and WT VEC at
945 4dpf, compared to N=4 genes in WT VEC at 3 and 5dpf.
946

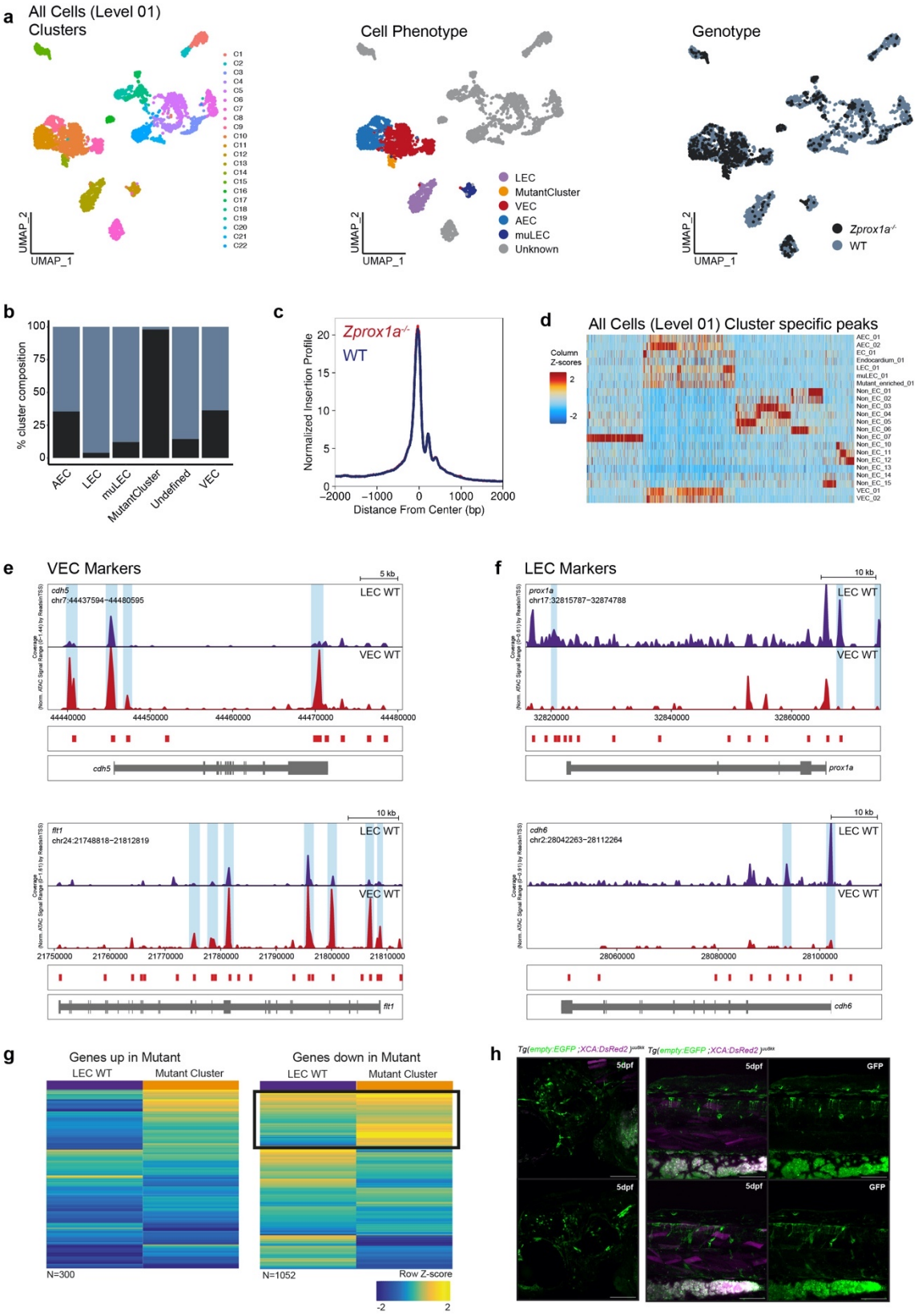


Extended data Fig 3: Additional phenotypic analysis of blood vascular markers in *Zprox1a*$^{-/-}$ mutants associated with Fig 3

a. 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 µm.

b. Quantification of *lyve1b* expression intensity in the TD relative to the PCV in sibling and *Zprox1a* mutant embryos at 4dpf.

955 **c.** Quantification of *cdh5* expression intensity in the TD relative to the DA in sibling and
956 *Zprox1a* mutant embryos at 4dpf.
957 **d.** Lateral confocal images of zebrafish trunks at 4dpf showing *kdrl* expression (green)
958 and *lyve1b* expression in lymphatic vessels (magenta) in control (upper) and *Zprox1a*
959 mutants (lower). Transgenic lines are indicated. Scale bars, 50 µm.
960 **e.** Quantification of *kdrl* expression intensity in the TD relative to the DA in sibling and
961 *Zprox1a* mutant embryos at 4dpf and 5dpf (upper).
962 **f.** Lateral confocal images of zebrafish trunks at 5dpf showing *kdrl* expression (green)
963 and *lyve1b* expression in lymphatic vessels (magenta) in control (upper) and *Zprox1a*
964 mutants (lower). Transgenic lines are indicated. Scale bars, 50 µm.
965 **g.** Quantification of the number of TD segments displaying *kdrl* expression in *Zprox1a*
966 mutant embryos at 4dpf and 5dpf (lower).
967 DLLV, dorsal longitudinal lymphatic vessel. ISLV, intersomitic lymphatic vessel. TD, thoracic
968 duct. DA, Dorsal aorta. PCV, posterior cardinal vein. vISV, venous intersegmental vessel. PL,
969 Parachoral lymphatic endothelial cells.
970



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

a. UMAP visualization of n=3,731 single nuclei ATAC-seq data comparing lymphangiogenesis in WT and *Zprox1a* mutants at 4dpf. Nuclei coloured according to

975 cluster (Res-0.8, left), predicted cell phenotype (middle) and genotype (right). Cell
976 phenotype was predicted based on accessibility changes at the genes defined in
977 **Extended data table 3a.**

978 **b.** Stacked bar plot indicating the genotype composition of nuclei captured per cell type.

979 **c.** Density plot indicating the majority of snATAC-seq peaks fall around the TSS. Y-axis
980 represents distance from TSS (bp) and y-axis represents normalised Tn5 insertion
981 profile.

982 **d.** Heatmap of cluster specific accessible peaks for all cell clusters. Colour reflects a
983 column-wise Z-score, each row represents a cluster defined in **Extended Data Fig. 4a**,
984 columns are peaks.

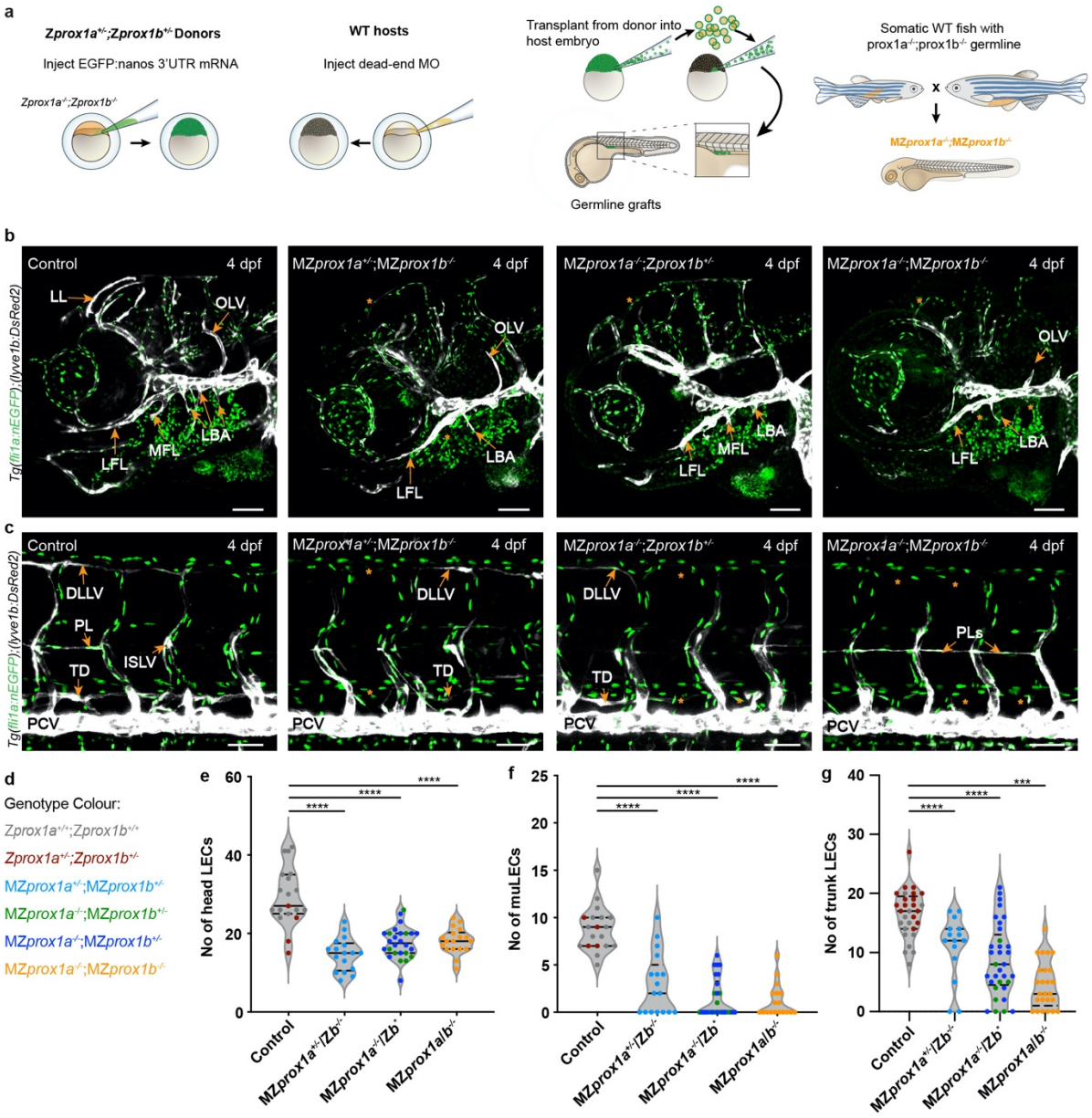
985 **e.** Genome accessibility track of key VEC markers *cdh5* (top) and *flt1* (bottom) indicating
986 these genes are more accessible in WT VEC than WT LEC cell phenotypes. Red bars
987 represent peaks in the reproducible peak set from snATAC-seq. DAP between WT VECs
988 and WT LECs (Wilcoxon Rank Sum, *FDR* < 0.05) are highlighted blue.

989 **f.** Genome accessibility track of key LEC markers *prox1a* (top) and *cdh6* (bottom)
990 indicating these genes are more accessible in WT LEC than WT VEC cell phenotypes.
991 Red bars represent peaks in the reproducible peak set from snATAC-seq. DAP between
992 WT LECs and WT VECs (Wilcoxon Rank Sum, *FDR* < 0.05) are highlighted blue.

993 **g.** Heatmaps of accessibility (gene score) for DEG (**Fig 3f**) upregulated (left, n=300
994 mapped genes) and downregulated (right, n=1,052 mapped genes) in the Mutant
995 Cluster compared to WT LECs. Colour indicates row-wise Z-score, each row represents
996 a differentially expressed gene in scRNA-seq data, columns are snATAC-seq clusters.
997 Boxed genes downregulated in Mutant display more permissive chromatin in Mutant
998 than WT LEC, showing that mutant cluster cells display a unique chromatin state at
999 many genes.

1000 **h.** Confocal projections of zebrafish head (left) labelled with *Tg(empty:EGFP;*
1001 *XCA:DsRed2)* at 5 dpf, showing GFP expression in neuronal tissues. Confocal
1002 projections of zebrafish trunk (right) labelled with *Tg(empty:EGFP; XCA:DsRed2)* at 5
1003 dpf, showing GFP expression in neuronal tissues. All scale bars = 100µm.

1004



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-c. Lateral confocal images showing vascular defects in control and mutant in the craniofacial (**b**) and trunk (**c**) 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-/-, MZprox1b-/-* 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.

1016

1017 **d.** Genotypes that are shown in the quantification. Colour codes are used to identify
1018 individual embryonic genotypes in the dot plots showing quantification in **e-g**.

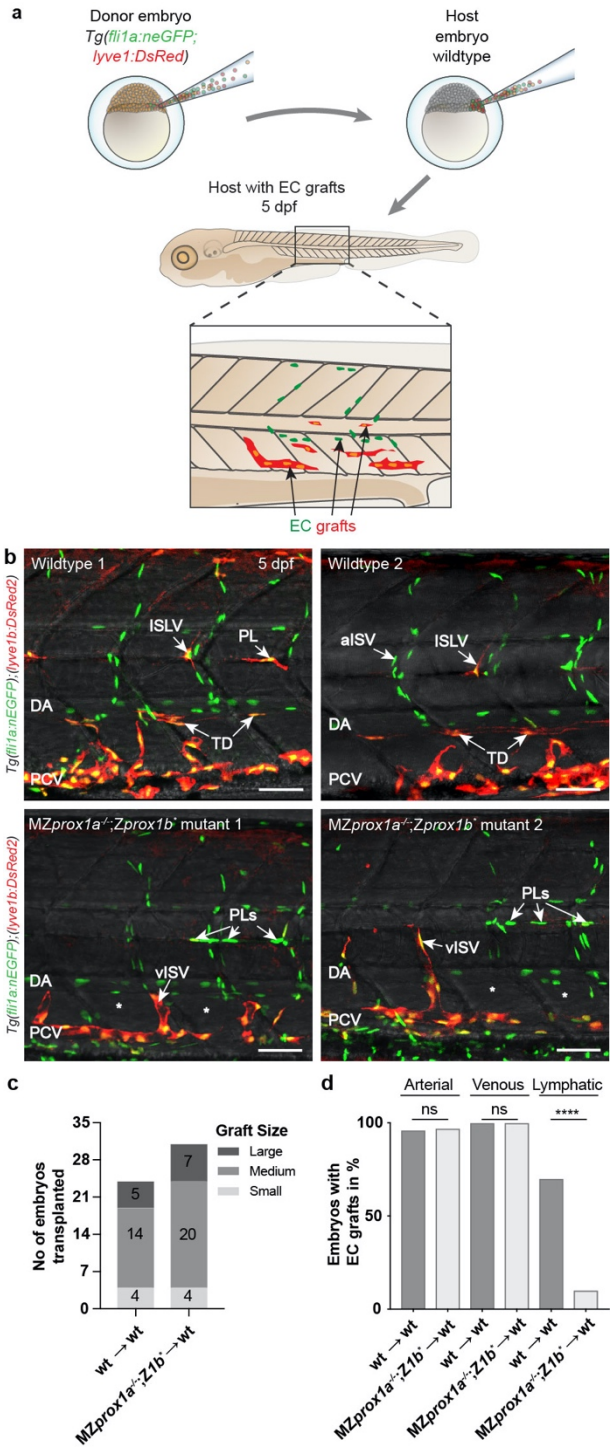
1019 **e.** Quantification of the number of craniofacial (head) LECs within facial lymphatic
1020 vessels showing evidence for contributions from *prox1a* and *prox1b*.

1021 **f.** Quantification of the number of mural LECs (muLECs) showing a major role for *prox1a*
1022 and a role for *prox1b* in muLEC development.

1023 **g.** Quantification of the number of trunk LECs within vessels showing a dominant role for
1024 *prox1a* in trunk lymphangiogenesis.

1025 Z, zygotic; MZ, maternal and zygotic; HM, horizontal myoseptum; LL, lymphatic loop; OLV,
1026 otholitic lymphatic vessel; LFL, lateral facial lymphatic vessel; MFL, medial facial lymphatic
1027 vessel; LBA, lymphatic branchial arches; DLLV, dorsal longitudinal lymphatic vessel; PL,
1028 parachordal LEC; ISLV, intersomitic lymphatic vessel; TD, thoracic duct; DA, dorsal aorta; PCV,
1029 posterior cardinal vein.

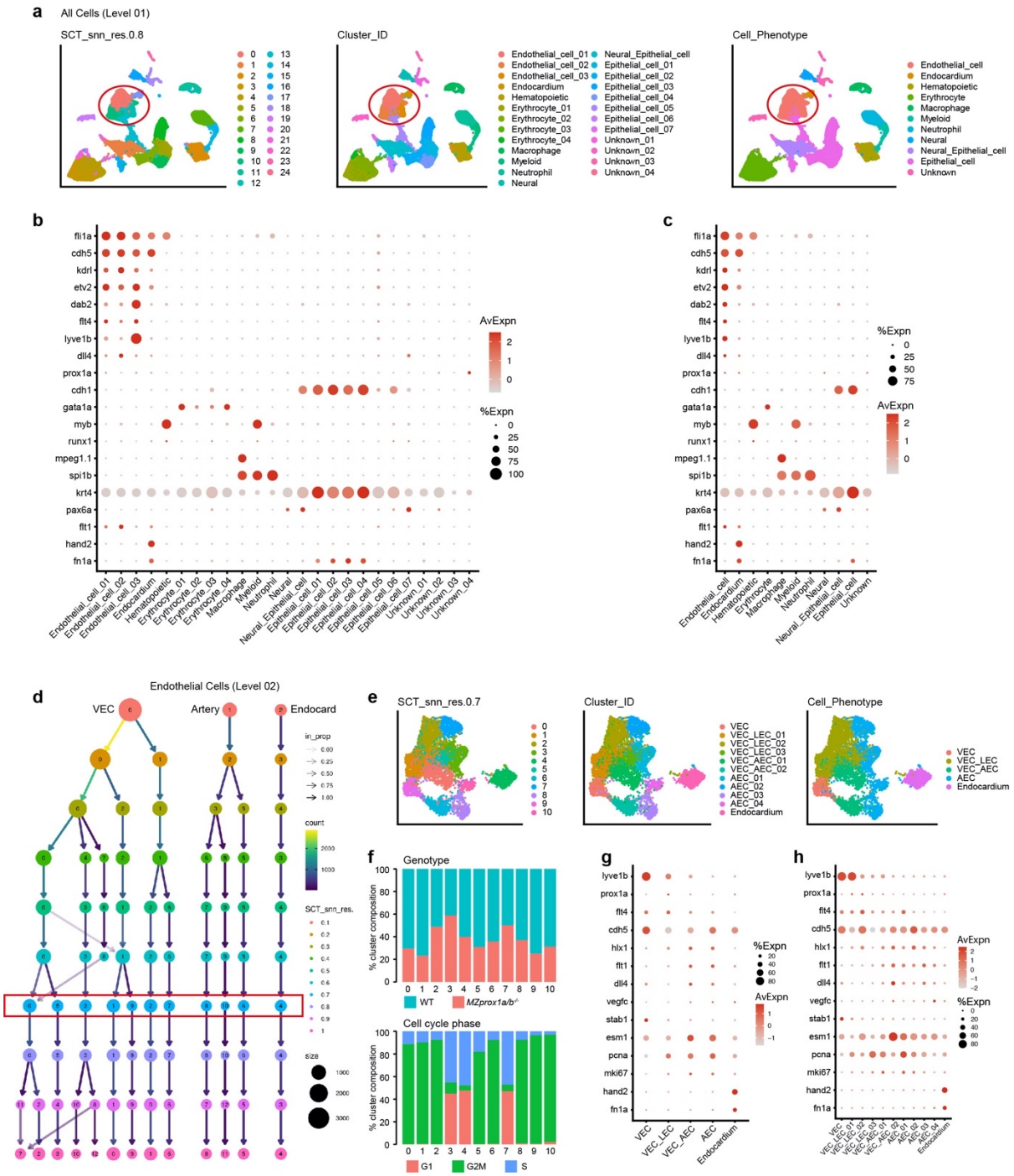
1030



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).

1038 **b.** Examples of 5dpf wildtype hosts engrafted with wildtype donor cells (upper) that
1039 contribute to arteries (aISVs, DA), veins (PCV, vISVs) and lymphatic vessels (TD, ISLVs)
1040 and examples of wildtype embryos engrafted with *MZprox1a* mutant cells
1041 contributing arteries (aISVs, DA) and veins (vISV, PCV) but not lymphatic vessels. Note
1042 the PL accumulation in the horizontal myoseptum..
1043 **c.** The number of embryos imaged and graft sizes scored post-tranplantation. As the
1044 contribution to the different EC types is affected by the graft size, we analysed grafts
1045 of comparable sizes and position from wildtype and mutant embryos. Grafts were
1046 categorised as large, medium or small. Small (spanning 1-2 somites), medium
1047 (spanning 2-3 somites) and large grafts (spanning 4-5 somites). The number of ECs
1048 contributing to each EC type was scored at single cell resolution and the percentage
1049 of grafts with a contribution to each lineage calculated accordingly. The numbers
1050 within the different graft groups indicate the number of analysed embryos within that
1051 group.
1052 **d.** The percentage of embryos with vascular grafts that display contributions to arteries,
1053 veins and lymphatics. *MZprox1a*$^{-/-}$ mutant cells contribute to arteries and veins but
1054 showed a near complete loss of contribution to lymphatics. The size of the grafts did
1055 not influence the outcome. *Z1b** indicates a mixed prox1b background. $p < 0.0001$
1056 from an unpaired, two-sided t-test.
1057 DA, dorsal aorta. PCV, posterior cardinal vein. PL, parachordal LEC. TD, thoracic duct. aISV,
1058 arterial intersegmental vessel. vISV, venous intersegmental vessel. ISLV, intersegmental
1059 lymphatic vessel.
1060



Extended data Fig 7: Additional data and analysis for data in Fig 7.

a. UMAP visualization of n=26,289 endothelial cells sequenced comparing lymphangiogenesis in WT and *MZprox1a/b* mutants at 40hpf (n=4 samples; see also **Extended data table 5a** for markers used to identify clusters) coloured according to cluster solution (SCT-snn-res.8, left), predicted phenotype per cluster (middle), and predicted cell phenotype (right). Endothelial cells selected for further analysis are indicated by a red circle.

1069 **b.** Dot plot summarizing the scRNA-seq expression level of n=20 key genes used to
1070 predict cell phenotypes of each cluster in the complete dataset (Level 01), for each
1071 cluster defined in **Extended Data Fig. 7a** (SCT-snn-res.8). The size of the dot
1072 represents the proportion of cells that express the markers in the cluster, and colour
1073 represents SCT-normalised expression.
1074 **c.** Dot plot summarizing the scRNA-seq expression level of n=20 key genes used to
1075 predict cell phenotype in the complete dataset. The size of the dot represents the
1076 proportion of cells that express the markers in the cluster, and colour represents SCT-
1077 normalised expression.
1078 **d.** Clustree analysis demonstrating the relationship between different cluster resolutions
1079 for the n=5,347 endocardial, venous and arterial endothelial cells in **Fig. 7a** and
1080 **Extended Data Fig. 7a** (Level 02). Cluster resolution SCT_snn_0.7 (boxed) was selected
1081 for all downstream analyses.
1082 **e.** UMAP visualization of cluster resolution SCT_snn_0.7 (left), predicted phenotype of
1083 each cluster (middle) and phenotypic group (right).
1084 **f.** Stacked bar plot of cluster resolution SCT_snn_0.7 cluster composition genotype (top)
1085 and cell cycle phase (bottom).
1086 **g.** Dot plot summarizing the scRNA-seq expression level of n=14 key genes used to
1087 predict cell phenotype in the endothelial cell populations (Level 02, n=5,347 cells).
1088 The size of the dot represents the proportion of cells that express the markers in the
1089 cluster, and colour represents SCT-normalised expression.
1090
1091
1092
1093
1094
1095 **References**

1096 1 Oliver, G., Kipnis, J., Randolph, G. J. & Harvey, N. L. The Lymphatic Vasculature in
1097 the 21(st) Century: Novel Functional Roles in Homeostasis and Disease. *Cell* **182**,
1098 270-296, doi:10.1016/j.cell.2020.06.039 (2020).
1099 2 Koltowska, K., Betterman, K. L., Harvey, N. L. & Hogan, B. M. Getting out and
1100 about: the emergence and morphogenesis of the vertebrate lymphatic vasculature.
1101 *Development* **140**, 1857-1870, doi:10.1242/dev.089565 (2013).
1102 3 Baek, S. *et al.* The Alternative Splicing Regulator Nova2 Constrains Vascular Erk
1103 Signaling to Limit Specification of the Lymphatic Lineage. *Dev Cell* **49**, 279-292
1104 e275, doi:10.1016/j.devcel.2019.03.017 (2019).

1105 4 Koltowska, K. *et al.* Vegfc Regulates Bipotential Precursor Division and Prox1
1106 Expression to Promote Lymphatic Identity in Zebrafish. *Cell Rep* **13**, 1828-1841,
1107 doi:10.1016/j.celrep.2015.10.055 (2015).
1108 5 Deng, Y., Atri, D., Eichmann, A. & Simons, M. Endothelial ERK signaling controls
1109 lymphatic fate specification. *The Journal of clinical investigation* **123**, 1202-1215,
1110 doi:10.1172/JCI63034 (2013).
1111 6 Shin, M. *et al.* Vegfc acts through ERK to induce sprouting and differentiation of
1112 trunk lymphatic progenitors. *Development* **143**, 3785-3795, doi:10.1242/dev.137901
1113 (2016).
1114 7 Srinivasan, R. S. *et al.* The Prox1-Vegfr3 feedback loop maintains the identity and the
1115 number of lymphatic endothelial cell progenitors. *Genes Dev* **28**, 2175-2187,
1116 doi:10.1101/gad.216226.113 (2014).
1117 8 Oliver, G. & Srinivasan, R. S. Endothelial cell plasticity: how to become and remain a
1118 lymphatic endothelial cell. *Development* **137**, 363-372, doi:10.1242/dev.035360
1119 (2010).
1120 9 Wigle, J. T. & Oliver, G. Prox1 function is required for the development of the
1121 murine lymphatic system. *Cell* **98**, 769-778, doi:10.1016/s0092-8674(00)81511-1
1122 (1999).
1123 10 Hagerling, R. *et al.* A novel multistep mechanism for initial lymphangiogenesis in
1124 mouse embryos based on ultramicroscopy. *EMBO J* **32**, 629-644,
1125 doi:10.1038/emboj.2012.340 (2013).
1126 11 Klotz, L. *et al.* Cardiac lymphatics are heterogeneous in origin and respond to injury.
1127 *Nature* **522**, 62-67, doi:10.1038/nature14483 (2015).
1128 12 Martinez-Corral, I. *et al.* Nonvenous origin of dermal lymphatic vasculature. *Circ Res*
1129 **116**, 1649-1654, doi:10.1161/CIRCRESAHA.116.306170 (2015).
1130 13 Stanczuk, L. *et al.* cKit Lineage Hemogenic Endothelium-Derived Cells Contribute to
1131 Mesenteric Lymphatic Vessels. *Cell Rep* **10**, 1708-1721,
1132 doi:10.1016/j.celrep.2015.02.026 (2015).
1133 14 Srinivasan, R. S. *et al.* Lineage tracing demonstrates the venous origin of the
1134 mammalian lymphatic vasculature. *Genes Dev* **21**, 2422-2432,
1135 doi:10.1101/gad.1588407 (2007).
1136 15 Srinivasan, R. S. *et al.* The nuclear hormone receptor Coup-TFII is required for the
1137 initiation and early maintenance of Prox1 expression in lymphatic endothelial cells.
1138 *Genes Dev* **24**, 696-707, doi:10.1101/gad.1859310 (2010).
1139 16 Yang, Y. *et al.* Lymphatic endothelial progenitors bud from the cardinal vein and
1140 intersomitic vessels in mammalian embryos. *Blood* **120**, 2340-2348,
1141 doi:10.1182/blood-2012-05-428607 (2012).
1142 17 Stone, O. A. & Stainier, D. Y. R. Paraxial Mesoderm Is the Major Source of
1143 Lymphatic Endothelium. *Dev Cell* **50**, 247-255 e243,
1144 doi:10.1016/j.devcel.2019.04.034 (2019).
1145 18 Johnson, N. C. *et al.* Lymphatic endothelial cell identity is reversible and its
1146 maintenance requires Prox1 activity. *Genes Dev* **22**, 3282-3291,
1147 doi:10.1101/gad.1727208 (2008).
1148 19 Tao, S. *et al.* Zebrafish prox1b mutants develop a lymphatic vasculature, and prox1b
1149 does not specifically mark lymphatic endothelial cells. *PLoS One* **6**, e28934,
1150 doi:10.1371/journal.pone.0028934 (2011).
1151 20 van Impel, A. *et al.* Divergence of zebrafish and mouse lymphatic cell fate
1152 specification pathways. *Development* **141**, 1228-1238, doi:10.1242/dev.105031
1153 (2014).

21 Hogan, B. M. *et al.* Ccbe1 is required for embryonic lymphangiogenesis and venous sprouting. *Nat Genet* **41**, 396-398, doi:10.1038/ng.321 (2009).
22 Yaniv, K. *et al.* Live imaging of lymphatic development in the zebrafish. *Nat Med* **12**, 711-716, doi:10.1038/nm1427 (2006).
23 Eng, T. C. *et al.* Zebrafish facial lymphatics develop through sequential addition of venous and non-venous progenitors. *EMBO Rep* **20**, doi:10.15252/embr.201847079 (2019).
24 Okuda, K. S. *et al.* lyve1 expression reveals novel lymphatic vessels and new mechanisms for lymphatic vessel development in zebrafish. *Development* **139**, 2381-2391, doi:10.1242/dev.077701 (2012).
25 Bussmann, J. *et al.* Arteries provide essential guidance cues for lymphatic endothelial cells in the zebrafish trunk. *Development* **137**, 2653-2657, doi:10.1242/dev.048207 (2010).
26 Cha, Y. R. *et al.* Chemokine signaling directs trunk lymphatic network formation along the preexisting blood vasculature. *Dev Cell* **22**, 824-836, doi:10.1016/j.devcel.2012.01.011 (2012).
27 Jung, H. M. *et al.* Development of the larval lymphatic system in zebrafish. *Development* **144**, 2070-2081, doi:10.1242/dev.145755 (2017).
28 Bower, N. I. *et al.* Mural lymphatic endothelial cells regulate meningeal angiogenesis in the zebrafish. *Nat Neurosci* **20**, 774-783, doi:10.1038/nn.4558 (2017).
29 van Lessen, M. *et al.* Intracellular uptake of macromolecules by brain lymphatic endothelial cells during zebrafish embryonic development. *Elife* **6**, doi:10.7554/eLife.25932 (2017).
30 Venero Galanternik, M. *et al.* A novel perivascular cell population in the zebrafish brain. *Elife* **6**, doi:10.7554/eLife.24369 (2017).
31 Kuchler, A. M. *et al.* Development of the zebrafish lymphatic system requires VEGFC signaling. *Curr Biol* **16**, 1244-1248, doi:10.1016/j.cub.2006.05.026 (2006).
32 Hogan, B. M. & Schulte-Merker, S. How to Plumb a Pisces: Understanding Vascular Development and Disease Using Zebrafish Embryos. *Dev Cell* **42**, 567-583, doi:10.1016/j.devcel.2017.08.015 (2017).
33 He, Z., Brazovskaja, A., Ebert, S., Camp, J. G. & Treutlein, B. CSS: cluster similarity spectrum integration of single-cell genomics data. *Genome Biol* **21**, 224, doi:10.1186/s13059-020-02147-4 (2020).
34 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821, doi:10.1016/j.cell.2019.05.031 (2019).
35 Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*, doi:10.1038/nbt.4314 (2018).
36 La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498, doi:10.1038/s41586-018-0414-6 (2018).
37 Thomas, P. D. *et al.* Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res* **34**, W645-650, doi:10.1093/nar/gkl229 (2006).
38 Wigle, J. T. *et al.* An essential role for Prox1 in the induction of the lymphatic endothelial cell phenotype. *EMBO J* **21**, 1505-1513, doi:10.1093/emboj/21.7.1505 (2002).
39 Kang, J. *et al.* An exquisite cross-control mechanism among endothelial cell fate regulators directs the plasticity and heterogeneity of lymphatic endothelial cells. *Blood* **116**, 140-150, doi:10.1182/blood-2009-11-252270 (2010).

1202 40 Bessa, J. *et al.* Zebrafish enhancer detection (ZED) vector: a new tool to facilitate
1203 transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev*
1204 *Dyn* **238**, 2409-2417, doi:10.1002/dvdy.22051 (2009).
1205 41 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors
1206 prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*
1207 **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
1208 42 Rouillard, A. D. *et al.* The harmonizome: a collection of processed datasets gathered
1209 to serve and mine knowledge about genes and proteins. *Database (Oxford)* **2016**,
1210 doi:10.1093/database/baw100 (2016).
1211 43 Kulkarni, R. M., Greenberg, J. M. & Akeson, A. L. NFATc1 regulates lymphatic
1212 endothelial development. *Mech Dev* **126**, 350-365, doi:10.1016/j.mod.2009.02.003
1213 (2009).
1214 44 Norrmen, C. *et al.* FOXC2 controls formation and maturation of lymphatic collecting
1215 vessels through cooperation with NFATc1. *J Cell Biol* **185**, 439-457,
1216 doi:10.1083/jcb.200901104 (2009).
1217 45 Dieterich, L. C. *et al.* DeepCAGE Transcriptomics Reveal an Important Role of the
1218 Transcription Factor MAFB in the Lymphatic Endothelium. *Cell Rep* **13**, 1493-1504,
1219 doi:10.1016/j.celrep.2015.10.002 (2015).
1220 46 Koltowska, K. *et al.* mafba is a downstream transcriptional effector of Vegfc
1221 signaling essential for embryonic lymphangiogenesis in zebrafish. *Genes Dev* **29**,
1222 1618-1630, doi:10.1101/gad.263210.115 (2015).
1223 47 Chen, L. *et al.* Tbx1 regulates Vegfr3 and is required for lymphatic vessel
1224 development. *J Cell Biol* **189**, 417-424, doi:10.1083/jcb.200912037 (2010).
1225 48 Nicenboim, J. *et al.* Lymphatic vessels arise from specialized angioblasts within a
1226 venous niche. *Nature* **522**, 56-61, doi:10.1038/nature14425 (2015).
1227 49 Cha, B. *et al.* Complementary Wnt Sources Regulate Lymphatic Vascular
1228 Development via PROX1-Dependent Wnt/beta-Catenin Signaling. *Cell Rep* **25**, 571-
1229 584 e575, doi:10.1016/j.celrep.2018.09.049 (2018).
1230 50 Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein-
1231 protein association networks, made broadly accessible. *Nucleic Acids Res* **45**, D362-
1232 D368, doi:10.1093/nar/gkw937 (2017).
1233 51 Armour, S. M. *et al.* An HDAC3-PROX1 corepressor module acts on HNF4alpha to
1234 control hepatic triglycerides. *Nat Commun* **8**, 549, doi:10.1038/s41467-017-00772-5
1235 (2017).
1236 52 Liu, X. *et al.* Mitotic Implantation of the Transcription Factor Prospero via Phase
1237 Separation Drives Terminal Neuronal Differentiation. *Dev Cell* **52**, 277-293 e278,
1238 doi:10.1016/j.devcel.2019.11.019 (2020).
1239 53 Qin, J. *et al.* Prospero-related homeobox (Prox1) is a corepressor of human liver
1240 receptor homolog-1 and suppresses the transcription of the cholesterol 7-alpha-
1241 hydroxylase gene. *Mol Endocrinol* **18**, 2424-2439, doi:10.1210/me.2004-0009 (2004).
1242 54 Song, K. H., Li, T. & Chiang, J. Y. A Prospero-related homeodomain protein is a
1243 novel co-regulator of hepatocyte nuclear factor 4alpha that regulates the cholesterol
1244 7alpha-hydroxylase gene. *J Biol Chem* **281**, 10081-10088,
1245 doi:10.1074/jbc.M513420200 (2006).
1246 55 Takeda, Y. & Jetten, A. M. Prospero-related homeobox 1 (Prox1) functions as a novel
1247 modulator of retinoic acid-related orphan receptors alpha- and gamma-mediated
1248 transactivation. *Nucleic Acids Res* **41**, 6992-7008, doi:10.1093/nar/gkt447 (2013).
1249 56 Lawson, N. D. & Weinstein, B. M. In vivo imaging of embryonic vascular
1250 development using transgenic zebrafish. *Dev Biol* **248**, 307-318,
1251 doi:10.1006/dbio.2002.0711 (2002).

57 Ciruna, B. *et al.* Production of maternal-zygotic mutant zebrafish by germ-line replacement. *Proc Natl Acad Sci U S A* **99**, 14919-14924, doi:10.1073/pnas.222459999 (2002).
58 Westerfield, M. *The Zebrafish Book. A Guide for the Laboratory use of Zebrafish (Danio rerio)*. 4 edn, (University of Oregorn Press, 2000).
59 Bussmann, J. & Schulte-Merker, S. Rapid BAC selection for tol2-mediated transgenesis in zebrafish. *Development* **138**, 4327-4332, doi:10.1242/dev.068080 (2011).
60 Okuda, K. S., Baek, S. & Hogan, B. M. Visualization and Tools for Analysis of Zebrafish Lymphatic Development. *Methods Mol Biol* **1846**, 55-70, doi:10.1007/978-1-4939-8712-2_4 (2018).
61 Bower, N. I. *et al.* Vegfd modulates both angiogenesis and lymphangiogenesis during zebrafish embryonic development. *Development* **144**, 507-518, doi:10.1242/dev.146969 (2017).
62 Kartopawiro, J. *et al.* Arap3 is dysregulated in a mouse model of hypotrichosis-lymphedema-telangiectasia and regulates lymphatic vascular development. *Hum Mol Genet* **23**, 1286-1297, doi:10.1093/hmg/ddt518 (2014).
63 Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049, doi:10.1038/ncomms14049 (2017).
64 Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281-291 e289, doi:10.1016/j.cels.2018.11.005 (2019).
65 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420, doi:10.1038/nbt.4096 (2018).
66 McCarthy, D. J., Campbell, K. R., Lun, A. T. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179-1186, doi:10.1093/bioinformatics/btw777 (2017).
67 Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience* **7**, doi:10.1093/gigascience/giy083 (2018).
68 Kejie Li, Z. O., Dongdong Lin, Michael Mingueneau, Will Chen, David Sexton, Baohong Zhang. cellxgene VIP unleashes full power of interactive visualization, plotting and analysis of scRNA-seq data in the scale of millions of cells. *bioRxiv* **2020.08.28.270652**, doi:https://doi.org/10.1101/2020.08.28.270652 (2020).
69 Smedley, D. *et al.* The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res* **43**, W589-598, doi:10.1093/nar/gkv350 (2015).
70 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).