Mosaic hotspot *PIK3CA* mutations cause non-cell-autonomous vascular overgrowth and pan-lineage dysregulation at disease onset

Hannah Brunsdon[1,2,3], Nuoya Wang[4], Micha Sam Brickman Raredon[4], Ralitsa R Madsen[5], Robert K Semple[1, 3*], E Elizabeth Patton[1, 2*]

1. MRC Human Genetics Unit, Institute of Genetics and Cancer, The University of Edinburgh, EH4 2XU, United Kingdom.
2. Edinburgh Cancer Research, CRUK Scotland Centre, Institute of Genetics and Cancer, The University of Edinburgh, EH4 2XR, United Kingdom.
3. Centre for Cardiovascular Science, The University of Edinburgh, Edinburgh, EH16 4TZ, United Kingdom.
4. Department of Anesthesiology, Vascular Biology and Therapeutics, Program in Translational Biomedicine, Yale School of Medicine, USA
5. MRC Protein Phosphorylation and Ubiquitylation Unit, School of Life Sciences, University of Dundee, Dundee, DD1 5EH, United Kingdom

* Corresponding authors

Author Contributions: Conceptualization: HB, RKS, EEP; Data curation: HB; Code checking: RRM, NW, MSBR; Formal analysis: HB; Funding acquisition: RRM, MSBR, RKS, EEP. Investigation: HB; Methodology: HB, RRM, MSBR, EEP; Project administration: RKS, EEP; Supervision: MSBR, RKS, EEP; Visualization: HB; Writing – original draft: HB, RRM, RKS, EEP; Writing – review & editing: All authors

Competing Interest Statement: RKS has received speaker fees and consulted for Novartis. RRM has received consulting fees from Nested Therapeutics (Cambridge, U.S.) and serves on the Scientific Advisory Board of CLOVES Syndrome Community. EEP is funded by QBiotics Group (Brisbane, Australia) for research not related to CLOVES/PROS.

Classification: Biological Sciences, Developmental Biology.

Keywords: PIK3CA, overgrowth, vascular malformations, PI3K, zebrafish, PROS, CLOVES

Abstract

Post-zygotic gain-of-function *PIK3CA* mutations arising during embryonic development cause disorders collectively known as the *PIK3CA*-related overgrowth spectrum (PROS). This ranges from minor, localized overgrowth to devastating multi-tissue overgrowth. Disease severity is widely attributed to a combination of PIK3CA genotype, affected cell type, and developmental timing of mutation acquisition. However, in PROS this explains neither the biased pattern of overgrowth - disproportionately affecting mesoderm and neuroectoderm-derived tissues - nor the typical low mutation burden within areas of extensive tissue overgrowth. Indeed, growing evidence suggests PROS mutations cause overgrowth non-cell-autonomously, although mechanisms of this are poorly understood. Here, we develop mosaic zebrafish models of PROS with overexpression of orthologous hotspot *pik3ca* mutations (*pik3ca*^{PROS}) to visualize and examine the effects of mutated cells on early development, in whole live animals. Reminiscent of PROS, we observe a spectrum of embryonic vasculature malformations (VMs), accompanied by larval muscle and bone overgrowth. Surprisingly, VMs only rarely expressed *pik3ca*^{PROS} in constituent endothelial cells, with *pik3ca*^{PROS}-expressing cells often closely abutting malformations instead. Single-cell transcriptomics of *pik3ca*^{PROS} mosaic zebrafish prior to VM onset revealed that most *pik3ca*^{PROS} cells are relatively immature and developmentally inert or constitute a small minority of mesodermal-derived cell types. Despite this constriction, global changes to cell fate were evident, alongside pervasive, pan-lineage abnormalities of gene expression, and rewiring of predicted ligand-receptor communication between lineages. We propose that targeting signals that indirectly propagate overgrowth through non-cell autonomous mechanisms - as well as PI3K activation itself - is worthy of therapeutic investigation.

Significance statement

Patchy, or mosaic, activating mutations in *PIK3CA* cause asymmetric human overgrowth due to aberrant hyperactivation of phosphoinositide 3-kinase (PI3K) signaling. Overgrowth prominently affects blood vessels and may be severely debilitating. While causation by *PIK3CA* mutations is clear, the explanation for the extent and pattern of associated overgrowth is not. Leveraging the power of zebrafish models for observation of early development, we now provide evidence for widely pervasive developmental effects of PIK3CA mutations extending beyond transcriptionally and phenotypically affected cells and lineages. This suggests potential therapeutic value of targeting secondary effects of PIK3CA activation as well as the activated PI3K itself.

Introduction

Coordination of lineage determination and growth rates within and among cell lineages is central to healthy human development. Class 1 phosphoinositide 3-kinases (PI3Ks) are one crucial group of signaling enzymes that mediate such coordination. PI3Ks transduce numerous upstream growth cues into widely ramifying downstream effects on cell growth, metabolism and other behaviors (1). They do this in a cell type, developmental stage and tissue specific manner. These growth cues include hormones and growth factors which activate receptor tyrosine kinases. The dominant class 1A PI3K subtypes regulating cell growth and metabolism in non-hematopoietic cell lineages include the p110α catalytic subunit encoded by *PIK3CA*, acting in concert with one of several regulatory subunits (2, 3). The heterodimeric enzyme is typically referred to as PI3Kα.

Hyperactivation of PI3Kα is common in solid human cancers, most often attributable to somatic activating *PIK3CA* mutations. The mutations with the greatest hyperactivity fall in 'hotspot' regions at the C terminal of the kinase domain (Histidine 1047) or in the helical domain (Glutamates 542 or 545) (4). These same mutations, if acquired during human development, also cause a wide range of asymmetric overgrowth disorders. Many different disease labels are used for these based on phenotypic descriptors, and the term *PIK3CA*-related overgrowth spectrum (PROS) is now used as an umbrella term for the whole range, reflecting its shared etiopathogenesis (5, 6). PROS ranges from very localized overgrowth, for example a circumscribed vascular malformation or single digit overgrowth, to severe and sometimes life-threatening multisystem disorders such as CLOVES syndrome (Congenital Lipomatous Overgrowth, Vascular Malformations, Epidermal Nevus, Spinal/Skeletal anomalies/Scoliosis) (5, 7). Vascular complications such as thromboembolism, hemorrhage, mass effects or high output heart failure are major causes of morbidity and mortality in PROS.

The wide phenotypic spectrum of PROS has most commonly been rationalized in terms of the timing and cell of origin of the founder *PIK3CA* mutation, which determines the anatomical site(s) of overgrowth, with the *PIK3CA* genotype (hotspot *vs* non hotspot) an important accessory factor determining the rate of overgrowth of affected tissues. This formulation fails to provide a complete account of the observed phenotypic heterogeneity, however. One unexplained observation is that not all tissues are equally affected in PROS. Mesoderm-derived veins, capillaries, lymphatics, adipose tissue, skeletal muscle, and bone are commonly and severely affected, for example, as are ectoderm-derived brain and peripheral nerves. In contrast, mesoderm-derived arteries and hematopoietic cells, and endoderm-derived organs are rarely affected (4, 7). Given the presumed stochastic nature of

35 spontaneous *PIK3CA* mutagenesis, this implies that mutations are excluded from, not
36 tolerated in, or not phenotypically penetrant in some lineages (4).

37 Another important observation is that the burden of *PIK3CA* mutation is variable, but often in
38 the range of <1-10% in tissue biopsies, even from macroscopically severely affected regions,
39 meaning that the bulk of sampled cells are genetically wild-type (WT) (4, 8-10). This will partly
40 reflect sampling of a genetically mosaic tissue, but the extent of this mismatch also suggests
41 that *PIK3CA* mutations may exert both cell-autonomous and non-cell-autonomous effects on
42 tissue growth. Work emerging using other model systems has shown that non-blood vessel
43 cells, such as PIK3CA-mutant lymphatic endothelial cells or Schwann cells, can increase PI3K
44 activation in neighboring endothelia to worsen vascular lesions and drive proliferation of WT
45 cells, through recruitment of macrophages or secretion of pro-inflammatory eicosanoids and
46 cytokines to drive proliferation of WT cells (11-14). Beyond its conceptual interest, identifying
47 mediators of such non-cell-autonomous mechanisms may be of translational value if these
48 secondary drivers of pathologic growth can be targeted. This is highly relevant in PROS as
49 available potent PI3K inhibitors have limited efficacy in only a subset of patients in studies so
50 far (15, 16), underscoring the need for better classification and understanding of the wider
51 signaling landscape within VMs to ensure optimal treatment strategies (17).

52 Dissecting how the presence of *PIK3CA*-mutated cells corrupts normal development is ideally
53 interrogated *in situ* in developing animals. Numerous animal models of PROS have indeed
54 been reported to date, using strategies based on endogenous (18-20) or transgenic
55 overexpression of *PIK3CA* hotspot variants (21-27), or a strongly activating but synthetic
56 *Pik3ca* allele (28, 29). However, although such strong genetic *Pik3ca* activation has been
57 shown to drive overgrowth in numerous lineages, the early events in overgrowth *in utero* are
58 not visible in mammals. We now present a zebrafish model of PROS circumventing this
59 limitation. The genetic tractability and transparency of zebrafish from their first cell divisions
60 into adulthood means that critical stages in disease onset and evolution hidden *in utero* can
61 be visualized and followed at cellular resolution from disease onset to adulthood in an intact
62 organism (30). The large clutch size and rapid development of zebrafish also offer high
63 experimental throughput to enable rapid testing of multiple PROS alleles in distinct cellular
64 lineages.

65 We use this new model to investigate the cellular and transcriptional landscape at the onset
66 of PROS vascular overgrowth. We show that overexpression of *pik3ca* hotspot alleles
67 (henceforth collectively "*pik3ca*^{PROS}") causes extensive vascular malformations, despite the
68 majority of *pik3ca*^{PROS} cells localizing adjacent to, rather than within lesions. Overexpression
69 limited to the mesoderm had widespread effects on subsequent development, manifesting as

aberrant partitioning of mutant cells among mesoderm-derived lineages. Single cell transcriptomics suggested more pervasive effects on gene expression and signaling in non-mesodermal lineages, with potential loosening of the normal pattern of developmental signal crosstalk among key embryonic structures. These findings consolidate evidence for developmental effects of *pik3ca*^{PROS} mosaicism that extend far beyond affected cells. This aligns with observations made in human PROS, presents a versatile animal model for further mechanistic interrogation, and opens new possibilities for accessory therapeutic targets beyond PI3K itself.

Results

Mosaic *pik3ca* hotspot transgenes cause vascular malformations in zebrafish embryos

Zebrafish Pik3ca protein has 91% homology with human PIK3CA, and AlphaFold predicts closely similar higher order structure to the experimentally determined structure of *PIK3CA*, evidenced by nearly identical positioning of PROS-related hotspot variants in both species **(Figure 1A)**. To develop a zebrafish *pik3ca-PROS* model, we generated Tol2-based vectors permitting *β-actin* promoter-driven (ubiquitous) expression of zebrafish *pik3ca*; either wild-type (WT) or harboring one of the hotspot mutations, namely *H1048R*, *H1048L* (orthologous to human *H1047R* and *H1047L* respectively), *E542K*, or *E545K* **(Figure 1B)**. A downstream IRES-*mScarlet* sequence enabled co-expression of the mScarlet reporter with mutant *pik3ca*. Additionally, the presence of an uncoupled secondary *cmcl2:GFP* cardiac reporter was included to provide a secondary readout of transgene integration. The *pik3ca*^{H1048R}*-mScarlet* construct was introduced into WT embryos to produce F0 mosaic transgenic embryos, which were raised and screened for transgene integration and overgrowth.

Many *pik3ca*^{H1048R}*-mScarlet*-injected embryos appeared morphologically normal, but 15% developed vascular lesions, observed as blood pooling and poor circulation most evident at 72 hours post-fertilization (hpf) **(Figure 1B)**. To investigate this phenotype further, we repeated *pik3ca*^{H1048R}*-mScarlet* transgene injections into the *Tg(fli1:eGFP)* endothelial reporter line, enabling co-visualization of blood vessels and transgene expression (31). *Fli1:eGFP* is first detectable during somitogenesis, during which angioblasts migrate from the lateral plate mesoderm (LPM) between the 10 and 18 somite stage (~16 hpf) towards the midline, with the main embryonic artery and vein being formed at 24 hpf, and trunk circulation being completed at 48 hpf (31, 32). In the trunk of zebrafish embryos at 72 hpf, arteries and veins are well established, with regularly-spaced intersegmental vessels (ISVs) connecting the ventral major aortic and venous vessels with the dorsal longitudinal anastomotic vessel (DLAV) (31, 33)

104 **(Figure 1C)**. Consistent with the variability observed in PROS, the severity of vascular
105 malformations resulting from transgene injection ranged from mild to severe, the latter obvious
106 from dramatic enlargement of the major blood vessels, causing blood pooling in the tail or
107 head, with some vascular overgrowths large enough to compromise swimming (**Figure 1D**).

108 In the same system, we next tested the propensity of other PROS hotspot variants to cause
109 vascular malformations, injecting WT, *H1048R, H1048L, E542K* or *E545K pik3ca*^{PROS}
110 constructs into *Tg*(*fli1:eGFP)* embryos. Transgenic expression of *pik3ca*^{WT}*-mScarlet* did not
111 induce overgrowth (**Figure S1A**). However, all four hotspot variants, whether affecting the
112 kinase (*H1048R/L*) or helical domains (*E542K/E545K*) of Pik3ca, reproducibly caused a
113 spectrum of vascular anomalies, from mild mis-patterning of *fli1:eGFP*+ ISVs only evident by
114 fluorescent transgene expression pattern, to severe balloon-like overgrowth, of the major
115 vessels in the tail **(Figure 1D,E; Figure S1A)**. Severity scoring of malformations for each of
116 the four PROS alleles and of WT or uninjected controls showed that transgenic expression of
117 mosaic *pik3ca*^{PROS} variants in zebrafish caused vascular malformations at similar levels of
118 severity (**Figure 1F**).

119 To test whether our model could recapitulate other overgrowth elements seen in PROS, we
120 selected 72 hpf embryos displaying a high level of *pik3ca*^{PROS} transgene integration but
121 appearing otherwise normal, and grew them to larvae stages (to 14 days). Transgene
122 integration was assessed either by high numbers of *mScarlet*+ cells or by the surrogate marker
123 of expression of the co-injected heart-targeted *cmcl2:GFP* reporter in most cardiac cells. We
124 imaged these fish as they developed into larvae. Despite no visible vascular abnormality in
125 the embryos, these fish still developed vascular malformations by larval stages, seen as
126 thickening of vessels, or as vascular protrusions from the skin **(Figure 1G)**. Aside from
127 vascular malformations, some *pik3ca*^{PROS} zebrafish larvae also showed appreciable
128 *mScarlet*+ expression and overgrowth in the musculature, as well as premature mineralization
129 in the notochord, fusing and thickening of vertebrae, and abnormal patterning of hypural
130 cartilage in the tail **(Figure S1B-D)**. These findings suggest that in the fish model, as in human
131 PROS, the trajectory of malformations is not exclusively determined in embryos, and
132 overgrowth can sometimes manifest or progress post-embryonically, notably in tissues of
133 mesodermal origin.

134

135

136

137 **Evidence for cell-autonomous and non-cell-autonomous mechanisms of *pik3ca^{PROS}***
138 **vascular overgrowth**

139 Surprisingly, we observed that only a minority of endothelial cells within vascular
140 malformations expressed both mScarlet and eGFP fluorophores, proxies for *pik3ca^{PROS}* and
141 *fli1* expression, respectively. We used confocal microscopy to build a 3D image of the vascular
142 lesions in higher resolution to investigate this further **(Figure 2A; Movie S1)**. In one example
143 of a mild VM, we observed thickening of the DLAV without mScarlet cells in the lesion itself,
144 but with *pik3ca^{PROS}-mScarlet*+ cells abutting the overgrown region **(Figure 2A; Movie S2)**. In
145 other moderate and severe vascular malformations, *pik3ca^{PROS}-mScarlet*+ cells were again
146 most often found adjacent to malformations rather than comprising abnormal vessel walls
147 (**Figure 2A; Movie S3**). Indeed, GFP and mScarlet double positive cells accounted for <5%
148 of total lesion volume on average **(Figure 2B).** This provides circumstantial evidence for the
149 hypothesized non-cell autonomous mechanisms in PROS. Specifically, it argues that
150 unidentified *pik3ca^{PROS}* cells in or close to the vascular niche may exert non-cell-autonomous
151 effects upon WT endothelial cells.

152

153 **Early mesodermal *pik3ca^{PROS}* transgene expression causes endothelial mis-patterning**

154 Many tissues prominently affected in PROS are mesoderm-derived, indicating biases in cell
155 lineage contribution to overgrowth, and the possibility of spatial and temporal windows of
156 especial vulnerability to PROS mutations during development (4). We therefore sought to
157 investigate the consequences of disease onset in the early mesoderm, which has hitherto not
158 been possible to observe *in vivo* in mammalian systems, by swapping our ubiquitous *β-actin*
159 promoter for that of the pan-mesodermal marker *tbxta* (**Figure 3A, B**). The activation of *tbxta*
160 prior to endothelial specification would also address the possibility that *mScarlet*+ endothelial
161 cells had been present during vasculogenesis but had either silenced transgene expression
162 or died prior to imaging and scoring at 72 hpf. As all tested *pik3ca^{PROS}* alleles gave rise to VMs
163 with the same frequency and severity (**Figure 1F**), we focused on one hotspot variant for
164 subsequent work, *pik3ca^{E545K},* for studying mesodermal PI3K gain-of-function.

165 *pik3ca^{E545K}-mScarlet*+ expressed from the *tbxta* promoter was as effective as the *β-actin*
166 construct in inducing the *pik3ca^{PROS}* vascular phenotype in *Tg(fli1:eGFP)* embryos (**Figure**
167 **S2**). *tbxta:pik3ca^{PROS}-mScarlet* expression was visible from early gastrulation, before the
168 earliest plausible stage of vascular malformation onset, as angioblasts migrate from the lateral
169 plate mesoderm (LPM) towards the midline during somitogenesis (10-22 hpf) (31). At 18 hpf,
170 *mScarlet* expression in injected controls was predominantly in the notochord and myotome,

171 but co-expressed with very few GFP+ endothelial cells, as expected from known expression
172 patterns of *tbxta* at this developmental stage (34) (**Figure 3C**). In controls, very occasional
173 ectopic *fli1:eGFP* cells were observed dorsal to the developing DA and unattached to the
174 nascent vasculature, thus seemingly independent of normal ISV sprouting, which occurs from
175 20 hpf (**Figure 3C, Movie S4**). In embryos overexpressing *pik3ca*^{PROS} however, the frequency
176 of such ectopic *fli1:eGFP* cells was markedly increased (**Figure 3D**). These ectopic cells
177 persisted through ISV sprouting and subsequent branching to form the DLAV at ~22 hpf.
178 Occasionally, at this time point, we observed rare examples of *eGFP* and mScarlet double
179 positive cells dying (**Figure 3C,** orange arrow). By 28 hpf, we observed DLAV cells forming
180 premature dilated vessels (**Movie S5**).

181 We thus demonstrate that *pik3ca*^{PROS} expression in mesoderm alone disrupts the earliest
182 stages of vascular development, causing dysregulation of endothelial cell sprouting and
183 branching. Occasional *mScarlet*-expressing endothelial cells were present, some of which
184 were observed dying during aberrant development, which may indicate heightened sensitivity
185 of endothelial cells to high levels of *pik3ca*^{PROS} transgene overexpression. However, as in later
186 vascular development, vascular defects in this model appear to be largely a non-cell
187 autonomous effect of *pik3ca*^{PROS} expression on *fli1:GFP+, mScarlet+* endothelial cells.

188

189 **Mosaic mesodermal *pik3ca*^{PROS} expression induces pan-lineage dysregulation**

190 Given this evidence for non-cell autonomous effects of mesodermal *pik3ca*^{PROS} expression on
191 early vascular development, we next sought to interrogate more agnostically the extent of the
192 developmental disruption. We used FACS to sort live cells from *tbxta:pik3ca*^{PROS}*-mScarlet* and
193 *tbxta:mScarlet* injected controls at 19 hpf and subjected these to single-cell RNA sequencing
194 (scRNA-seq) (**Figure 4A**). Using Seurat, we identified 40 distinct cell populations common to
195 'control' and 'PROS mosaic' embryos. These were named using a combination of marker gene
196 expression patterns and projection of two previously published scRNA-seq datasets onto the
197 data (**Figure 4B, Figure S3A, B**) (35, 36). This confirmed the presence of all cell lineages
198 expected at this developmental stage, including endothelial cells, CNS, epithelial, lateral plate
199 and presomitic mesoderm (LPM, PSM), and notochord (**Figure 4B, C**).

200 One clear difference between control and *pik3ca*^{PROS} embryos lay in so-called cluster '0', which
201 could not be mapped to a uniquely annotated lineage, instead featuring weak signatures of
202 many different cell types, hinting at a "confused" lineage identity (**Figure S3C**). We confirmed
203 upregulation of the *pik3ca* transcript itself in this cluster, largely co-segregating with *mScarlet*
204 in PROS mosaic animals (**Figure 4D**). Velocity analysis showed a lack of clear developmental

205 directionality of transcription in PROS mosaic embryos compared to the small corresponding
206 cluster in control animals. In contrast, velocity in the PSM and neural lineages was increased,
207 suggesting non-cell autonomous perturbations of lineage transitions in both mesodermal
208 (*tbxta*-driven) and non-mesodermal (i.e. neural) lineages (**Figure S3D**).

209 Next, we quantified the changes in relative sizes of different cell clusters between experimental
210 conditions by expressing the number of cells within each cluster as a percentage of the total
211 cell number for control and PROS datasets, respectively. After calculating the differences in
212 these percentages, a dramatic expansion of cluster 0 was observed in PROS mosaic embryos
213 (**Figure 4E; Figure S3E**). Some clusters of mesodermal origin also showed expansion in
214 PROS mosaics, as may be expected from mesodermal activation of PI3K signaling, but,
215 surprisingly, this was not true of all mesoderm-derived clusters, and ectoderm-derived neural
216 clusters were also enriched in PROS mosaics (**Figure 4D, E; Figure S3E**). This suggests that
217 *pik3ca*^{PROS} cells have developmental effects extending beyond mesodermal lineages
218 themselves, unbalancing cell fate decisions and thereby cell type proportions in the embryo.

219 We next asked whether *pik3ca*^{PROS}*-mScarlet* cells exert transcriptional effects on admixed WT
220 cells. Although within PROS mosaic cells, 319 genes were significantly upregulated in
221 mScarlet+ PROS cells relative to control, including *pik3ca, tbxta* and *mScarlet*, surprisingly
222 few genes were significantly upregulated in PROS mosaic cluster 0 relative to control,
223 including *pik3ca* itself (**Figure S4A**), which precluded subsequent gene enrichment analysis.
224 We found widespread enrichment of Kegg or Reactome pathway components directly related
225 to PI3K signaling in clusters from the PROS zebrafish, including modules related to Pten,
226 Insulin, PIP3, mTor and Foxo1 (**Figure 4F**). Other gene sets were also frequently enriched,
227 including Notch, Wnt and Mapk pathways, as well as shared responses to cell metabolism
228 and cell division processes. Strikingly, these transcriptional changes extended to clusters in
229 the PROS zebrafish of non-mesoderm origin that expressed almost no mScarlet cells (**Figure**
230 **4F**). Moreover, these gene expression patterns persisted when we removed the PROS
231 mScarlet cells from the analysis entirely, indicating that otherwise wild type mesodermal and
232 non-mesodermal lineages are responding transcriptionally to PROS cells even without
233 expressing *pik3ca*^{PROS} themselves (**Figure S4B**).

234

235 ***pik3ca*^{PROS} elicits widespread transcriptional changes in ligand-receptor expression**
236 **patterns**

237 Finally, we sought evidence from single cell transcriptomic data implicating potential mediators
238 of the observed non-cell autonomous effects of genetic PI3K hyperactivation. We did this using

239 CellChat and NICHES as analytic tools to infer communication among cell lineages using the
240 proxy of expression patterns of ligand/receptor pairs (37, 38). Analyses of the top 10%
241 predicted cluster interactions by CellChat suggested major transcriptional changes in
242 signalling mediators in PROS embryos. In WT embryos the transcriptomic landscape
243 appeared dominated by reciprocal expression of ligand receptor pairs (loosely "signaling
244 traffic") among the notochord, and mixed epithelial/pharyngeal lineages, while in PROS
245 embryos a more distributed pattern was seen, with presomitic mesoderm and various neural
246 lineages much more prominent (**Figure 5A**).

247 To assess potential specific signaling pathways driving these predicted changes, we used the
248 NICHES algorithm, which analyses ligand-receptor mRNA expression in one-to-one pairs of
249 cells to infer communication (38). We subsetted our data to include the clusters found most
250 active by CellChat and ran NICHES to convert the single-cell gene expression dataset to a
251 single-cell signalling dataset based on LR signaling pattern genes, grouped by sending
252 clusters, receiving clusters, or individual pair-wise interactions (**Figure 5B)**. To identify
253 potential changes in cell signalling capacity between lineages in PROS mosaic animals
254 relative to control, we performed DE analysis on 'sending' and 'receiving' transcriptional
255 signatures for each cluster (**Figure S4C**). To gain a comprehensive overview of signaling
256 changes, we simplified the outputs by grouping and averaging ligand and receptor fold
257 changes respectively (**Figure 5C**). We observed a marked increase in Notch, Wnt and Ephrin
258 ligand-receptor pairs in both mesodermal and non-mesodermal lineages, but again cluster 0
259 exhibited general downregulation of sending and receiving signaling components. This is in
260 keeping with the non-cell autonomous responses to small numbers of *pik3ca*^{PROS} cells in the
261 PROS zebrafish, and the relatively inert state of C0 (**Figure 5C**; **Figure S3 C,D**).

262 We next calculated 'module scores' by averaging gene expression of all ligand-receptor pairs
263 within a pathway (e.g. all combinations of *jag*/*dll*/*dld-notch1a/1b/2/3* for notch signalling) and
264 compared the intensities of this score on cell-to-cell UMAPs, split by experimental condition
265 (**Figure 5D)**. Comparing the connectivity of PROS mosaics to control, we observed a reduction
266 in extent and intensity of mRNA expression of *egf* and *pdgf* ligand-receptor pairs in PROS,
267 further highlighting disruption of normal signaling through Pik3ca throughout the embryo
268 (**Figure 4F, 5D**). In contrast, we observed widespread upregulation of pairs of *notch*, *wnt* and
269 *ephrin* signaling genes (**Figure 5D**). Focusing on individual ligand-receptor pairs, we found
270 evidence for transcriptional "rewiring" of these pathways between lineages. For example, *dla-*
271 *notch3, wnt4-fzd5-lrp6* and *efnb3b-epha4a* ligand-receptor pair expression is increased for
272 neural, PSM, myotome and Epi/Ph nodes (**Figure 5D**).

273 We conceptualize these data to indicate that *pik3ca*^{PROS} has significant effects on cell fate,
274 behavior and likely signaling which ramify across many lineages, indicating pervasive and
275 disruptive non-cell-autonomous effects on early development. Overall, these analyses
276 highlight transcriptional changes in ligand and receptor gene expression that may contribute
277 to the non-cell-autonomous phenotypes observed and serve as potential drug targets worthy
278 of further investigation.

279

280 **Discussion**

281 Over ten years since mosaic, activating PIK3CA mutations were identified as the cause of a
282 group of asymmetric overgrowth disorders (39, 40) much has been learned. The wide human
283 clinical spectrum has been delineated, the natural history described, and nomenclature
284 harmonized (5, 6). Furthermore, case series and registry based open label studies have
285 established that the PI3Kα inhibitor alpelisib is of some clinical utility, leading to licensing in
286 the USA (16, 29) and establishment of an international randomized clinical trial still underway
287 (NCT04589650) (41). However, despite this major progress, important fundamental and
288 translational questions about PROS remain. Specifically, the explanation for the skewed
289 appearance of PROS-related pathology in some tissues but not others, and the involvement
290 and extent of non-cell-autonomous mechanisms of overgrowth in PROS pathogenesis have
291 eluded comprehensive analysis. The latter issue is of translational importance as it may yield
292 novel diagnostic and therapeutic targets; given current evidence that alpelisib has only partial
293 efficacy in only some patients with PROS (15), this is of great value.

294 Fundamental to translational progress is the fidelity of models to human pathology. A wealth
295 of murine models has advanced our understanding of PROS, by showing definitive causal
296 links between acquisition of PROS mutations and vascular malformations (22, 42), identifying
297 critical time windows during which endothelia are most sensitive to excessive PI3K signaling
298 (25), and demonstrating that PI3K and related pathway inhibition can mitigate or rescue
299 important disease manifestations (14, 21-23, 42, 43). More recently, mouse PROS or cancer
300 models have implicated signaling from non-endothelial cell types as drivers of vascular
301 overgrowth (11, 25, 27), but otherwise few animal studies have focused on non-cell-
302 autonomous contributors to PROS pathology to build a holistic view of multi-lineage
303 interactions during early development.

304 Overexpression of an activating PIK3CA allele has been shown to recapitulate several aspects
305 of PROS-related overgrowth in zebrafish (28), and the successful usage of zebrafish as a
306 preclinical model of lymphatic overgrowth or neurovascular malformations (44, 45)

307 encouraged us to develop patient-led zebrafish models to interrogate the complex interplay of
308 Pik3ca genotype and lineage context, and to examine overgrowth dynamics at early stages of
309 development. We now present a zebrafish model with mosaic overexpression of naturally
310 occurring hotspot *pik3ca*^{PROS} variants. This enables observation of the effects of PROS
311 mutations *in vivo*, from onset, at cellular resolution for the first time. Our model recapitulates
312 a spectrum of overgrowth severity affecting different cell lineages, as in PROS, with
313 vasculature most prominently affected, while scRNA-seq suggests that *pik3ca*^{PROS} expressing
314 cells appeared 'stuck' in a less differentiated cell state. This is reminiscent of iPSC models of
315 *PIK3CA*^{H1047R} activation, where homozygous mutant cells show transcriptional remodeling and
316 impaired differentiation (46). Our model also allows comparison of four different *pik3ca*
317 hotspot gain-of-function variants in parallel, which proved to cause vascular overgrowth with
318 similar frequency and severity. This argues against major differences in phenotypic expression
319 across different hotspot Pik3ca variants, whether affecting helical or kinase domains, although
320 the high level of expression in our model leaves open the possibility that differences may be
321 unmasked at endogenous expression levels.

322 The non-cell autonomous influences of *pik3ca*^{PROS} cells found in our *pik3ca*^{PROS} model is in
323 keeping with mounting evidence that paracrine signals from *PI3KCA*^{PROS} cells corrupt
324 neighboring wildtype cells. For example, eicosanoids secreted from *PIK3CA* mutant breast
325 cancer cells induce proliferation in WT cells (13), paracrine VEGF and IL-6 signaling from
326 mammary duct culture lead to dysfunction in neighboring innervating vasculature *in vitro* (12).
327 Moreover, *in vivo* activating Akt mutations in murine mammary tissue cause hyperplasia of
328 nearby WT cells in a paracrine manner (47), and lymphatic capillary endothelial cells
329 expressing *Pik3ca*^{H1047R} recruit VEGFC-producing macrophages which cause immune
330 activation and worsening of vascular malformations in a mouse lymphoma model (11). Lastly,
331 driving *Pik3ca*^{E545K} in the murine neural crest muscle and nerves induced hypertrophy of
332 vasculature through changes to neural-crest derived pericytes (48). Paracrine signals also
333 have a role in normal endothelial and hematopoietic development, as a subtype of endothelial
334 cells derived from the paraxial mesoderm have been shown to support induction of
335 hematopoietic cells from neighboring hemogenic endothelial cells in zebrafish (49). It is
336 important to note that cell-autonomous and non-cell-autonomous contributions to VMs in our
337 model are not mutually exclusive, and we did consistently observe minorities of *fli1:eGFP*+
338 endothelia co-expressing *mScarlet* indicative of a mixture of signaling mechanisms occurring
339 in VMs. Likewise, gain-of-function *PIK3CA* mutations are found through genotyping
340 endothelial cells in PROS patients, and are thought to be a driver of overgrowth in their
341 wildtype counterparts. Therefore, both cell-autonomous and non-cell-autonomous signals
342 likely underpin PROS pathology.

343 A limitation of our model, as for many other published models, is that it relies on transgenic
344 *pik3ca* overexpression, although the variants we overexpressed were orthologues to human
345 disease-causing mutations. The extent of overactivation through transgenesis is likely higher
346 than seen in PROS where *PIK3CA* is regulated endogenously, meaning the lack of *mScarlet*
347 in cell types most sensitive to PI3K levels, such as the endothelia, may be due to an intolerably
348 high levels of PI3K signaling causing cell death, as we observed in some co-expressing cells.
349 Furthermore, the timing of p110α activation is important for phenotype severity, with prenatal
350 but not postnatal induction of $Pik3ca^{E545K}$ causing increased brain size, and timed induction of
351 $Pik3ca^{H1047R}$ in endothelial cells revealing critical windows of PI3K sensitivity during
352 development (25, 26). Further temporal resolution of $pik3ca^{PROS}$ during zebrafish
353 development, and to an extent the 'dosage' of PI3K activation would be possible using a drug-
354 inducible version of our transgene. Finally, further testing of our transgenic construct in older
355 fish, perhaps with more restricted spatial expression of PROS mutations, is required to see if
356 adipose tissue and other tissue defects observed in PROS occur, and to the same penetrance.

357 Our novel zebrafish $pik3ca^{PROS}$ model has enabled *in vivo* observation of early stages of
358 evolution of PROS-like vascular malformations at a cellular level. It has moreover offered
359 evidence for surprisingly pervasive developmental effects of mosaic *pik3ca* activation,
360 including extensive non-cell autonomous effects ramifying beyond the affected lineage. This
361 complements and extends existing models of PROS and provides a tractable platform for more
362 in-depth interrogation of specific candidate mediators of such remote effects of *pik3ca*
363 activation. Therapeutic targeting of these may in principle be a valuable adjunct or alternative
364 to PI3K inhibition in future.

365

366 **Materials and Methods**

367 **Structural comparison of zebrafish and human PIK3CA**

368 Protein structures for human PIK3CA (AF-P27986-F1-v4) and the predicted structure of
369 zebrafish Pik3ca (AF-F1QAD7-F1-v4) were visualized and annotated using PyMOL
370 (Schrodinger).

371

372 **Fish husbandry, fish lines**

373 Zebrafish were maintained in accordance with UK Home Office regulations, UK Animals
374 (Scientific Procedures) Act 1986, amended in 2013, and European Directive 2010/63/EU
375 under project licenses P8F7F7E52 and PP7317786. All experiments were approved by the

376 Home Office and AWERB (University of Edinburgh Ethics Committee). Fish stocks used were
377 wild-type AB, *Tg(fli1:eGFP)* (50), *casper* (51). Combined transgenic lines were generated by
378 crossing. Adult fish were maintained at ~28.5°C under 14:10 light-dark cycles. Embryos were
379 kept at 28.5°C and staged according to the reference table provided by (52) and (53).

380 **$Pik3ca^{PROS}$ transgene plasmid construction**

381 First, the *eGFP* sequence in the p3E Gateway entry vector IRES-eGFP-pA (54) was swapped
382 for the *mScarlet* CDS (Thermo Fisher Scientific) with the HiFi DNA assembly kit (NEB) to make
383 the p3E entry vector p3E-IRES-mScarlet-pA. The wild type zebrafish *pik3ca* CDS was
384 introduced into pDonr221 (Invitrogen) using Gateway™ BP Clonase™ II (ThermoFisher
385 Scientific) to make pME-$pik3ca^{WT}$. *pik3ca* hotspot mutation sites were introduced into pME-
386 $pik3ca^{WT}$ sequence by designing back-to-back mutagenesis primers incorporating hotspot
387 mutations *H1048R/L*, *E542K* or *E545K*, phosphorylating them using T4 PNK (NEB) according
388 to manufacturer's instructions, and amplifying using Q5 polymerase (NEB). T4 ligase (NEB)
389 was used to re-ligate the mutated amplicon, and the original template digested using Dpni
390 (NEB). LR Clonase II+ (ThermoFisher Scientific) was used to combine p5E entry vectors p5E-
391 *β-actin* (54), p5E-*fli1ep* (50), p5E- *tbxta* (this study, based off (55), with a *pik3ca-PROS* pME,
392 and p3E-IRES-mScarletpA (this study, modified from p3E-IRES-GFP) into pDestTol2CG2 (54)
393 which carries a *cmlc2:eGFP* reporter. Sanger sequencing was used to confirm successful
394 mutation of pME-$pik3ca^{WT}$ to pME- $pik3ca^{H1048/H1048L/E542K/E545K}$ and correct integration of all
395 components. Sequences of primers used for plasmid construction are found in **Table S1**.

396 **Generation of mosaic $pik3ca^{PROS}$ fish**

397 Approx. 2 nl of 30 ng/µl transgenesis construct was co-injected with 35 ng/µl Tol2 mRNA into
398 zebrafish embryos at the 1 cell stage. Embryos were *Tg(fli1:eGFP)* to enable visualization of
399 blood vessels, in either a wildtype or *casper* (unpigmented) genetic background – the latter to
400 more easily image embryos beyond stages of melanocyte development. Embryos were
401 screened for successful PROS transgene integration from 30 hpf onwards by assessment of
402 *cmlc2:GFP* cardiac fluorescence and/or *mScarlet* fluorescence. Morphologically abnormal
403 embryos at this stage were discarded, as vascular development is disrupted due to defects in
404 somite and heart development, for example (56). Scoring of vascular malformations was
405 performed at 72 hpf. For studies on older fish, embryos were raised to juvenile stages.

406 **Imaging**

407 Embryos and juvenile fish were anaesthetized with MS:222 and if necessary, placed on 3%
408 methylcellulose. Fluorescent mesoscope images were acquired using a Leica M165FCA

409 fluorescence stereomicroscope fitted with a 0.63x and 1x Plan-Apochromat objective, and a
410 Leica LMT260 XY scanning stage (Leica Microsystems, Germany). Image acquisition was
411 performed using a Leica K8 monochrome cMOS camera in Leica Application Suite software
412 (LASX; version 3.7). Color images were acquired as above but using a Leica KC3 color cMOS
413 camera (Leica, Microsystems, Germany). Images were acquired in Leica Application Suite
414 software (LASX; version 3.7). Movies were generated using Imaris (v9.6.0). Images were
415 processed using FIJI (v2.9.0).

416 Confocal images were acquired using a 20X/0.75 lens on the multimodal Imaging Platform
417 Dragonfly (Andor technologies, Belfast UK) equipped with 488 and 561 lasers built on a Nikon
418 Eclipse TI-E inverted microscope body with Perfect focus system (Nikon Instruments, Japan.
419 Data were collected in Spinning Disk 25 µm pinhole mode on the iXon 888 EMCCD camera
420 using a Bin of 1 and no frame averaging, and using Andor Fusion acquisition software. Z
421 stacks were collected using the Nikon TiE focus drive. Movies were generated using Imaris
422 (v9.6.0). Images were processed using FIJI (v2.9.0).

423 **Bone and Cartilage staining**

424 To stain bone and cartilage in juvenile fish, the Juvenile No Acid Bone & Cartilage Stain was
425 followed as described (zfin.atlassian.net/wiki/spaces/prot/pages/352157857) using a 0.02%
426 Alcian Blue 8GX (Merck)/10 mM $MgCl_2$ solution to stain cartilage and then a 0.01% Alizarin
427 red S (Merck) solution to stain mineralized bone.

428 **FACS and library preparation**

429 *Tg(fli1:eGFP)* embryos were injected with either *tbxta:pik3ca*^{E545K}*-mScarlet* or *tbxta:mScarlet*
430 and >300 embryos per condition raised to 19 hpf. Each batch was dissociated to single cell
431 suspensions using the methods described in (57). Cells from stage matched wild-type
432 embryos were used as negative controls to gate fluorescence. Live cells were sorted using a
433 FACSAria SORP instrument (BD Biosciences UK), with GFP (525/50, BP/488 nm laser), RFP
434 (582/15 BP, 561 nm laser), and DAPI (450/20BP, 405 nm laser) filters into 2% BSA in PBS
435 and immediately processed using the Chromium platform (10x Genomics). Libraries were
436 prepared using v3.1 Standard Throughput kits according to 10x protocols (10x Genomics),
437 and quality control and quantification assays performed using High Sensitivity DNA kits on a
438 Bioanalyzer (Agilent). Libraries were initially sequenced on an Illumina NovaSeq SP Flow cell
439 with 100 cycles with two lanes per sample. The imbalances in cell number/sequencing depth
440 between the two samples necessitated re-sequencing of the PROS sample as before. This
441 resulted in an average of 120,316 reads/cell after normalization.

442 **Bioinformatics**

443 FastQ files were aligned using the CellRanger (v8.0.1, 10x Genomics) mkref, mkgtf and aggr
444 pipelines to custom zebrafish STAR genome index using annotations from Ensembl GRCz11
445 release 112 with manual addition of *eGFP* and *mScarlet*. The resultant h5 matrices were read
446 into R (v4.4.2) and subjected to quality control using Seurat (58) guidelines (v5.2.1, features
447 >200, <13000 and mitochondrial gene < 5%). This resulted in 10306 cells from PROS embryos
448 and 3505 from control. Data were scaled and regressed to cell cycle stage before performing
449 PCA, finding the dimensionality of the data by elbow plot (dims=50) to plot 40 clusters on
450 UMAP plots (SeuratExtend (v1.1.4 (59)). SingleR (60) (v2.8.0) was used with previously
451 published datasets GEO: GSE112294 (35) and NCBI SRA PRJNA940501 (36) to call clusters.
452 DE analyses were performed using zinbwave (v 1.28.0 (61)) and DESeq2 (v1.46.0(62)),
453 selecting genes with padj<0.05 before using clusterProfiler (v4.14.3 (63)) to perform gene
454 enrichment analyses searching Reactome, GO:BP and KEGG libraries, filtering by padj <0.05.
455 Proportions of cells within cluster for PROS and control data were found by dividing the
456 number of cells per cluster/total cell number per experimental condition, with Pearson's chi-
457 squared test (two-tailed) used to determine significance (raw data are in **Table S2**). RNA
458 velocity analysis was performed on loom files for control and PROS data generated with
459 velocyto (v. 0.17 (64)), and velocity computed using scVelo's (v.0.2.5 (65)) dynamical
460 modelling approach, and visualized using heatmaps and stream plots. CellChat (v. 2.1.2 (37))
461 was used to infer potential signaling events between clusters according to relevant package
462 guidelines. NICHES (v. 1.0.0 (38, 66)) was used to perform ligand-receptor expression
463 analysis at single cell resolution, with differential expression analysis within Seurat, using the
464 MAST test, performed to identify significant changes to LR gene expression. Additionally,
465 circuit diagrams were generated using code from
466 https://github.com/RaredonLab/NICHESMethods. Plots were generated using Seurat,
467 CellChat, NICHES, Complex Heatmap (v. 2.22.0 (67)) and ggplot2 (v. 4.0.0).

468

469 **Acknowledgments**

470 This work is supported by funding from the CLOVES Syndrome Community [c-13277479] and
471 the UK Medical Research Council [MR/Z506321/1]. We are grateful to Kristen Davies and
472 Seth Haddix from the CLOVES Syndrome Community for discussions about the experimental
473 aims and design of this work, and to Mihaly Badonyi and Joe Marsh (MRC Human Genetics
474 Unit) for assistance with Pymol. RKS is also supported by the BHF Centre for Research
475 Excellence Award III [RE/18/5/34216]. RRM is funded by a UKRI Future Leaders Fellowship
476 [MR/Y017439/1]. MSBR and NW are funded by T32GM086287 from the NIGMS, and
477 laboratory startup funds from the Yale School of Medicine and the Yale Department of

Anesthesiology. EEP is funded by the Medical Research Council (MC_UU_00035/13) and the Cancer Research UK Scotland Centre (CTRQQR-2021\100006).

Data and Code availability

scRNA-seq experiments generated in this study have been deposited at GEO and will become publicly available as of the date of publication. Computational code and plasmid constructs generated for this are available from the lead contact upon request.

References

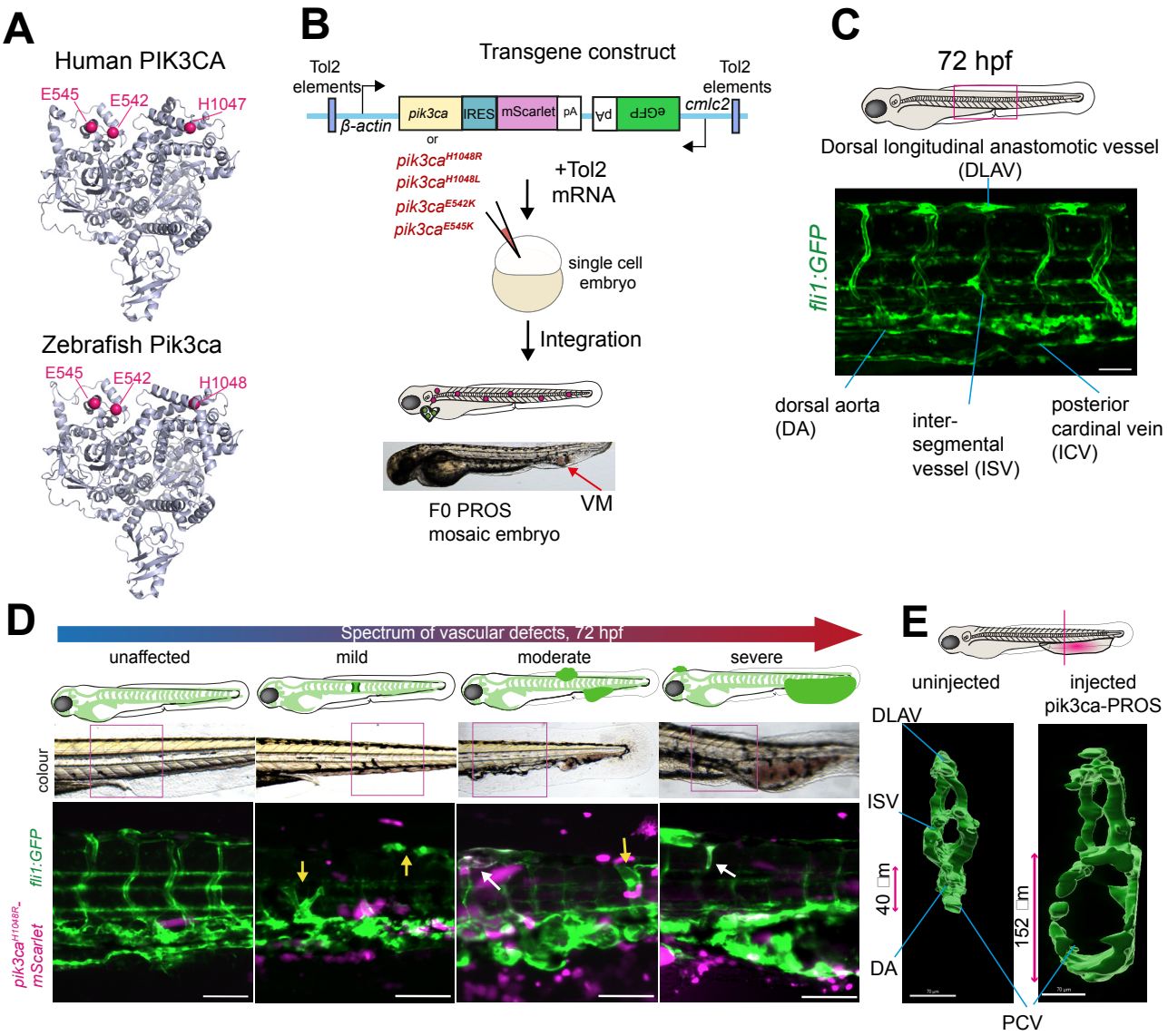
1. R. R. Madsen, A. Toker (2023) PI3K signaling through a biochemical systems lens. in *Journal of Biological Chemistry* (American Society for Biochemistry and Molecular Biology Inc.).
2. D. A. Fruman *et al.* (2017) The PI3K Pathway in Human Disease. in *Cell* (Cell Press), pp 605-635.
3. F. Sanchez-Vega *et al.*, Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **173**, 321-337.e310 (2018).
4. R. R. Madsen, B. Vanhaesebroeck, R. K. Semple (2018) Cancer-Associated PIK3CA Mutations in Overgrowth Disorders. in *Trends in Molecular Medicine* (Elsevier Ltd), pp 856-870.
5. K. M. Keppler-Noreuil *et al.*, Clinical delineation and natural history of the PIK3CA-related overgrowth spectrum. *American Journal of Medical Genetics, Part A* **164**, 1713-1733 (2014).
6. K. M. Keppler-Noreuil *et al.*, PIK3CA-related overgrowth spectrum (PROS): Diagnostic and testing eligibility criteria, differential diagnosis, and evaluation. *American Journal of Medical Genetics, Part A* **167**, 287-295 (2015).
7. A. Martinez-Lopez *et al.* (2017) CLOVES syndrome: review of a PIK3CA-related overgrowth spectrum (PROS). in *Clinical Genetics* (Blackwell Publishing Ltd), pp 14-21.
8. M. De Bortoli *et al.*, Somatic Loss-of-Function PIK3R1 and Activating Non-hotspot PIK3CA Mutations Associated with Capillary Malformation with Dilated Veins (CMDV). *J Invest Dermatol* **144**, 2066-2077 e2066 (2024).
9. B. N. Nriagu *et al.*, Microcystic lymphatic malformations in Turner syndrome are due to somatic mosaicism of PIK3CA. *Am J Med Genet A* **194**, 64-69 (2024).
10. D. Li *et al.*, Genomic profiling informs diagnoses and treatment in vascular anomalies. *Nat Med* **29**, 1530-1539 (2023).
11. M. Petkova *et al.*, Immune-interacting lymphatic endothelial subtype at capillary terminals drives lymphatic malformation. *J Exp Med* **220** (2023).
12. M. L. Kutys *et al.*, Uncovering mutation-specific morphogenic phenotypes and paracrine-mediated vessel dysfunction in a biomimetic vascularized mammary duct platform. *Nat Commun* **11**, 3377 (2020).
13. N. Koundouros *et al.*, Metabolic Fingerprinting Links Oncogenic PIK3CA with Enhanced Arachidonic Acid-Derived Eicosanoids. *Cell* **181**, 1596-1611 e1527 (2020).
14. Q. Venot *et al.*, PIK3CA gain-of-function mutation in Schwann cells leads to severe neuropathy and aerobic glycolysis through a non-cell autonomous effect. *Proc Natl Acad Sci U S A* **122**, e2424867122 (2025).

15. E. M. A. EMA/492147/2023 (2023) EMEA/H/C/005468 Withdrawal of application for the marketing authorisation of Vijoice (alpelisib).

16. G. Canaud *et al.*, Alpelisib for treatment of patients with PIK3CA-related overgrowth spectrum (PROS). *Genet Med* **25**, 100969 (2023).
17. E. Seront, A. Queisser, L. M. Boon, M. Vikkula, Molecular landscape and classification of vascular anomalies. *Hematology Am Soc Hematol Educ Program* **2024**, 700-708 (2024).
18. S. D. Castillo *et al.*, Somatic activating mutations in Pik3ca cause sporadic venous malformations in mice and humans. *Science Translational Medicine* **8** (2016).
19. K. M. Kinross *et al.*, Ubiquitous expression of the Pik3caH1047R mutation promotes hypoglycemia, hypoinsulinemia, and organomegaly. *FASEB J* **29**, 1426-1434 (2015).
20. I. M. Berenjeno *et al.*, Oncogenic PIK3CA induces centrosome amplification and tolerance to genome doubling. *Nat Commun* **8**, 1773 (2017).
21. A. Roy *et al.*, Mouse models of human PIK3CA-related brain overgrowth have acutely treatable epilepsy. *Elife* **4** (2015).
22. P. Castel *et al.*, Somatic PIK3CA mutations as a driver of sporadic venous malformations. *Sci Transl Med* **8**, 332ra342 (2016).
23. L. di Blasio *et al.*, PI3K/mTOR inhibition promotes the regression of experimental vascular malformations driven by PIK3CA-activating mutations. *Cell Death Dis* **9**, 45 (2018).
24. H. Sabata *et al.* (2025) Context-dependent response of endothelial cells to PIK3CA mutation.
25. I. Martinez-Corral *et al.*, Blockade of VEGF-C signaling inhibits lymphatic malformations driven by oncogenic PIK3CA mutation. *Nat Commun* **11**, 2869 (2020).
26. S. Torii *et al.*, Embryological cellular origins and hypoxia-mediated mechanisms in PIK3CA-driven refractory vascular malformations. *EMBO Mol Med* **17**, 1289-1324 (2025).
27. M. Kraft *et al.*, Angiopoietin-TIE2 feedforward circuit promotes PIK3CA-driven venous malformations. *Nat Cardiovasc Res* **4**, 801-820 (2025).
28. N. Blum, M. P. Harris, Localized heterochrony integrates overgrowth potential of oncogenic clones. *DMM Disease Models and Mechanisms* **16** (2023).
29. Q. Venot *et al.*, Targeted therapy in patients with PIK3CA-related overgrowth syndrome. *Nature* **558**, 540-546 (2018).
30. E. E. Patton, L. I. Zon, D. M. Langenau, Zebrafish disease models in drug discovery: from preclinical modelling to clinical trials. *Nat Rev Drug Discov* **20**, 611-628 (2021).
31. N. D. Lawson, B. M. Weinstein, In vivo imaging of embryonic vascular development using transgenic zebrafish. *Dev Biol* **248**, 307-318 (2002).
32. W. Zhou *et al.*, Akt is a mediator of artery specification during zebrafish development. *Development (Cambridge, England)* **151** (2024).
33. L. J. Greenspan, B. M. Weinstein, To be or not to be: endothelial cell plasticity in development, repair, and disease. *Angiogenesis* **24**, 251-269 (2021).
34. B. P. Thisse, S.; Fürthauer, M.; Loppin, B.; Heyer, V.; Degrave, A.; Woehl, R.; Lux, A.;Steffan, T.; Charbonnier, X.Q; and Thisse, C. (2001) Expression of the zebrafish genome during embryogenesis (NIH R01 RR15402). ZFIN Direct Data Submission. in *ZFIN Direct Data Submission* (NIH R01 RR15402, http://zfin.org).
35. D. E. Wagner *et al.*, Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981-987 (2018).
36. M. Lange *et al.*, A multimodal zebrafish developmental atlas reveals the state-transition dynamics of late-vertebrate pluripotent axial progenitors. *Cell* **187**, 6742-6759 e6717 (2024).
37. S. Jin *et al.*, Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088 (2021).
38. M. S. B. Raredon *et al.*, Comprehensive visualization of cell-cell interactions in single-cell and spatial transcriptomics with NICHES. *Bioinformatics* **39** (2023).

39. M. J. Lindhurst *et al.*, Mosaic overgrowth with fibroadipose hyperplasia is caused by somatic activating mutations in PIK3CA. *Nat Genet* **44**, 928-933 (2012).
40. J. B. Riviere *et al.*, De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. *Nat Genet* **44**, 934-940 (2012).
41. R. R. Madsen, R. K. Semple, PIK3CA-related overgrowth: silver bullets from the cancer arsenal? *Trends Mol Med* **28**, 255-257 (2022).
42. S. D. Castillo *et al.*, Somatic activating mutations in Pik3ca cause sporadic venous malformations in mice and humans. *Sci Transl Med* **8**, 332ra343 (2016).
43. L. Rodriguez-Laguna *et al.*, Somatic activating mutations in PIK3CA cause generalized lymphatic anomaly. *J Exp Med* **216**, 407-418 (2019).
44. D. Li *et al.*, ARAF recurrent mutation causes central conducting lymphatic anomaly treatable with a MEK inhibitor. *Nat Med* **25**, 1116-1122 (2019).
45. E. Martin-Valiente *et al.*, Reversal of cerebrovascular anomalies in a zebrafish model of vein of Galen aneurysm. *Nat Cardiovasc Res* **4**, 773-789 (2025).
46. R. R. Madsen *et al.*, Oncogenic PIK3CA promotes cellular stemness in an allele dose-dependent manner. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 8380-8389 (2019).
47. M. J. Lindhurst *et al.*, A mouse model of Proteus syndrome. *Hum Mol Genet* **28**, 2920-2936 (2019).
48. E. Marechal *et al.*, Multiple congenital malformations arise from somatic mosaicism for constitutively active Pik3ca signaling. *Frontiers in Cell and Developmental Biology* **10** (2022).
49. P. Sahai-Hernandez *et al.*, Dermomyotome-derived endothelial cells migrate to the dorsal aorta to support hematopoietic stem cell emergence. *Elife* **12** (2023).
50. J. A. Villefranc, J. Amigo, N. D. Lawson, Gateway compatible vectors for analysis of gene function in the zebrafish. *Dev Dyn* **236**, 3077-3087 (2007).
51. R. M. White *et al.*, Transparent adult zebrafish as a tool for in vivo transplantation analysis. *Cell Stem Cell* **2**, 183-189 (2008).
52. C. B. Kimmel, W. W. Ballard, S. R. Kimmel, B. Ullmann, T. F. Schilling, Stages of embryonic development of the zebrafish. *Dev Dyn* **203**, 253-310 (1995).
53. D. M. Parichy, M. R. Elizondo, M. G. Mills, T. N. Gordon, R. E. Engeszer, Normal table of postembryonic zebrafish development: staging by externally visible anatomy of the living fish. *Dev Dyn* **238**, 2975-3015 (2009).
54. K. M. Kwan *et al.*, The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. *Dev Dyn* **236**, 3088-3099 (2007).
55. S. A. Harvey, S. Tumpel, J. Dubrulle, A. F. Schier, J. C. Smith, no tail integrates two modes of mesoderm induction. *Development* **137**, 1127-1135 (2010).
56. A. N. Stratman, B. M. Weinstein, Assessment of Vascular Patterning in the Zebrafish. *Methods Mol Biol* **2206**, 205-222 (2021).
57. M. Manoli, W. Driever, Fluorescence-activated cell sorting (FACS) of fluorescently tagged cells from zebrafish larvae for RNA isolation. *Cold Spring Harb Protoc* **2012** (2012).
58. Y. Hao *et al.*, Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* **42**, 293-304 (2024).
59. Y. Hua, L. Weng, F. Zhao, F. Rambow, SeuratExtend: Streamlining Single-Cell RNA-Seq Analysis Through an Integrated and Intuitive Framework. *bioRxiv* 10.1101/2024.08.01.606144, 2024.2008.2001.606144 (2024).
60. D. Aran *et al.*, Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* **20**, 163-172 (2019).
61. D. Risso, F. Perraudeau, S. Gribkova, S. Dudoit, J. P. Vert, Publisher Correction: A general and flexible method for signal extraction from single-cell RNA-seq data. *Nat Commun* **10**, 646 (2019).
62. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

63. T. Wu *et al.*, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021).
64. G. La Manno *et al.*, RNA velocity of single cells. *Nature* **560**, 494-498 (2018).
65. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* **38**, 1408-1414 (2020).
66. A. M. Greaney *et al.*, Engineered Whole Lungs for Tissue Biology. *bioRxiv* 10.1101/2024.10.02.616240, 2024.2010.2002.616240 (2024).
67. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016).


A
Human PIK3CA
E545
E542
H1047
Zebrafish Pik3ca
E545
E542
H1048
B
Transgene construct
Tol2 elements
β-actin
pik3ca
IRES
mScarlet
pA
pA
eGFP
cmlc2
Tol2 elements
or
pik3ca^H1048R
pik3ca^H1048L
pik3ca^E542K
pik3ca^E545K
+Tol2 mRNA
single cell embryo
Integration
F0 PROS mosaic embryo
VM
C
72 hpf
Dorsal longitudinal anastomotic vessel (DLAV)
fli1:GFP
dorsal aorta (DA)
inter-segmental vessel (ISV)
posterior cardinal vein (ICV)
D
Spectrum of vascular defects, 72 hpf
unaffected
mild
moderate
severe
colour
fli1:GFP
pik3ca^H1048R-mScarlet
E
uninjected
injected pik3ca-PROS
DLAV
ISV
40 μm
DA
152 μm
PCV

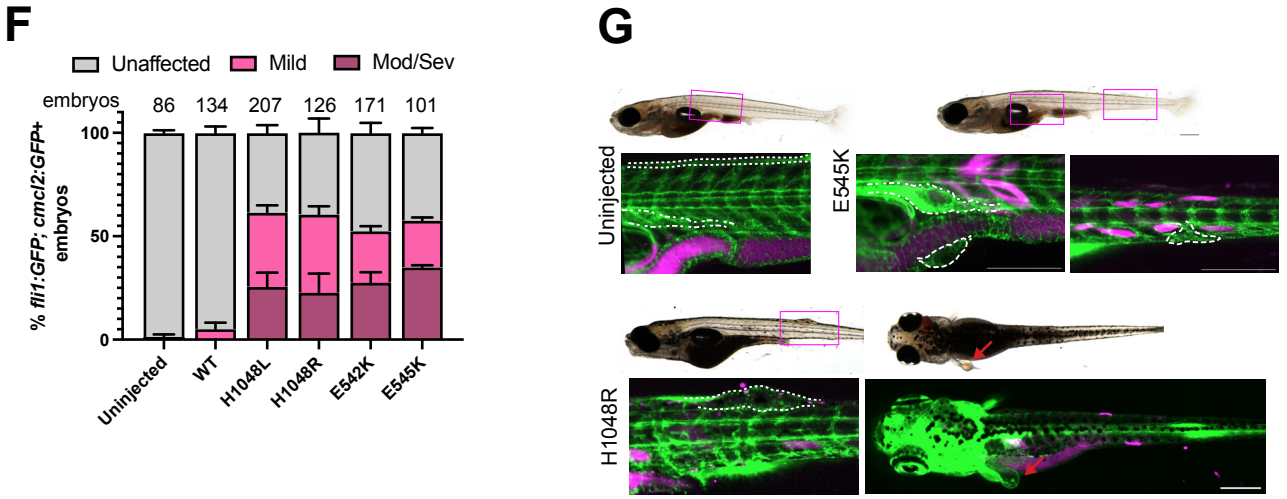

F
Unaffected
Mild
Mod/Sev
embryos
86
134
207
126
171
101
% fli1:GFP; cmcl2:GFP+ embryos
100
50
0
Uninjected
WT
H1048L
H1048R
E542K
E545K
G
Uninjected
E545K
H1048R

Figure 1

Figure Legends

Figure 1: Mosaic *pik3ca* hotspot transgenes cause vascular malformations in zebrafish embryos

A) The structure of human PIK3CA and the predicted structure of the zebrafish Pik3ca orthologue is remarkably conserved, with one additional amino acid in zebrafish Pik3ca shifting the kinase domain amino acids by 1. The position of E542, E545 and H1047 residues found mutated in PROS (and the orthologous E542, E545 and H1048) are labelled in magenta and are predicted to occur in very similar structural regions of the protein in both human and fish.

B) Schematic representation of the transgenic *pik3ca*^{PROS} constructs generated for this study, containing *β-actin:pik3ca*^{PROS}*-mScarlet* and the cardiac reporter *cmlc2:GFP* flanked by Tol2 response elements, and is co-injected with *Tol2* mRNA into 1-cell zebrafish embryos. Embryos were screened for successful integration of the transgene from 48 hpf onwards, indicated by mosaic cardiac *cmlc2:GFP*, and *mScarlet* expression. At 72 hpf, some PROS mosaic embryos had tail vascular malformations.

C) Maximum Z-projection of a 72 hpf *Tg(fli1:eGFP);casper* tail region, showing typical endothelial vessel structures marked by *fli1:eGFP* fluorescence (green). Scale bar = 50 µm

D) Representative diagrams and fluorescent microscopy images of the spectrum of vascular phenotypes seen in *Tg(fli1:eGFP)* embryos injected with *pik3ca*^{H1048R}*-mScarlet*. These ranged from unaffected and mild, to moderate and severe. *Fli1:eGFP* expression in unaffected embryos is as uninjected. Embryos scored 'mild' displayed expansion and mispatterning of isolated inter somitic vessels that was only visible when viewing eGFP fluorescence. Moderate and severe scores were applied to embryos where blood pooling was evident under brightfield, particularly obvious in the PCV. Yellow arrows indicate mis-patterned vessels in 'mild' embryos, and white arrows indicate overlap of *fli1:eGFP* (green) and *mScarlet* (magenta) expression in moderate and severe examples. Quantification of this and numbers screened can be seen in Figure 3. Scale bars = 100 µm.

E) Cross sections of trunk blood vessels at 72 hpf, showing the dramatic expansion of the PCV.

F) Quantification of these vascular malformation phenotypes for each construct, scored based on severity, as a percentage of the total numbers injected. Numbers indicate *fli1:eGFP; cmcl2:GFP*+ embryos screened for vessel malformations at 72 hpf, representative of three biological repeats per construct. Error bars represent S.E.M.

G) Representative images of juvenile *Tg(β-actin:pik3ca*^{E545K}*-mScarlet; fli1:eGFP); casper* (top) and *Tg(β-actin:pik3ca*^{H1048R}*-mScarlet; fli1:eGFP* fish that harboured no detectable vascular anomalies at 72 hpf, but were found to have vascular anomalies when imaged two weeks later (*E545K*-injected – 6/16, *H1048R*-injected – 3/9). Dotted lines encompass vessel malformations in PROS animals, and equivalent regions in control

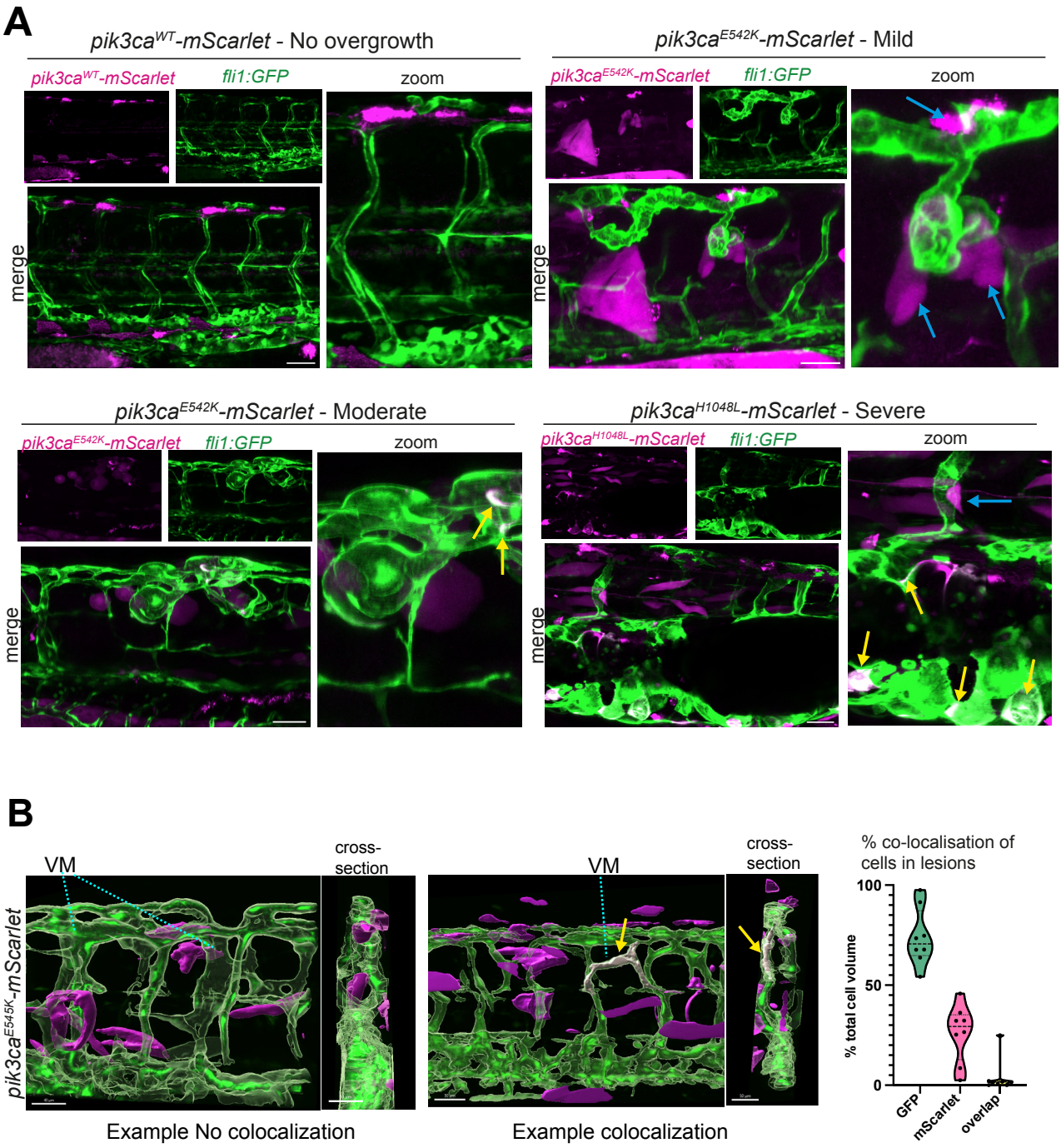

A
pik3ca^WT-mScarlet - No overgrowth
pik3ca^WT-mScarlet
fli1:GFP
zoom
merge
pik3ca^E542K-mScarlet - Mild
pik3ca^E542K-mScarlet
fli1:GFP
zoom
merge
pik3ca^E542K-mScarlet - Moderate
pik3ca^E542K-mScarlet
fli1:GFP
zoom
merge
pik3ca^H1048L-mScarlet - Severe
pik3ca^H1048L-mScarlet
fli1:GFP
zoom
merge
B
VM
cross-section
pik3ca^E545K-mScarlet
Example No colocalization
VM
cross-section
Example colocalization
% co-localisation of cells in lesions
% total cell volume
100
50
0
GFP
mScarlet
overlap

Figure 2

Figure 2: Evidence for cell-autonomous and non-cell-autonomous mechanisms of *pik3ca^{PROS}* vascular overgrowth

A) Representative maximum projection confocal images of 72 hpf *Tg(fli1:eGFP);casper* embryos expressing either wildtype *pik3ca-mScarlet* (no overgrowth) or a *pik3ca^{PROS}* hotspot variant and exhibiting a mild, moderate or severe vascular phenotype. For each, separated *mScarlet* fluorescence (magenta), *fli1:eGFP* channels (green) and merged channels are shown, with a zoomed area to the right. Blue arrows indicate *mScarlet* cells in close proximity to deformed vessels. Yellow arrows indicate white cells co-expressing *fli1:eGFP* and *mScarlet*. Scale bars = 50 µm

B) Quantification of *mScarlet* cells by volume within GFP+ endothelial cells. Imaris was used to visualize confocal Z stacks of vessel overgrowth at 72 hpf. 3D structures of each channel, plus a co-localization channel were converted into surfaces, enabling volume calculations for vessels (green), mScarlet+ cells outside vessels (purple) or areas of colocalization (white). Volumes of each were expressed as percentage of total cell volume in 20x magnification view, dots indicate individual images (8 lesions were analysed in total).

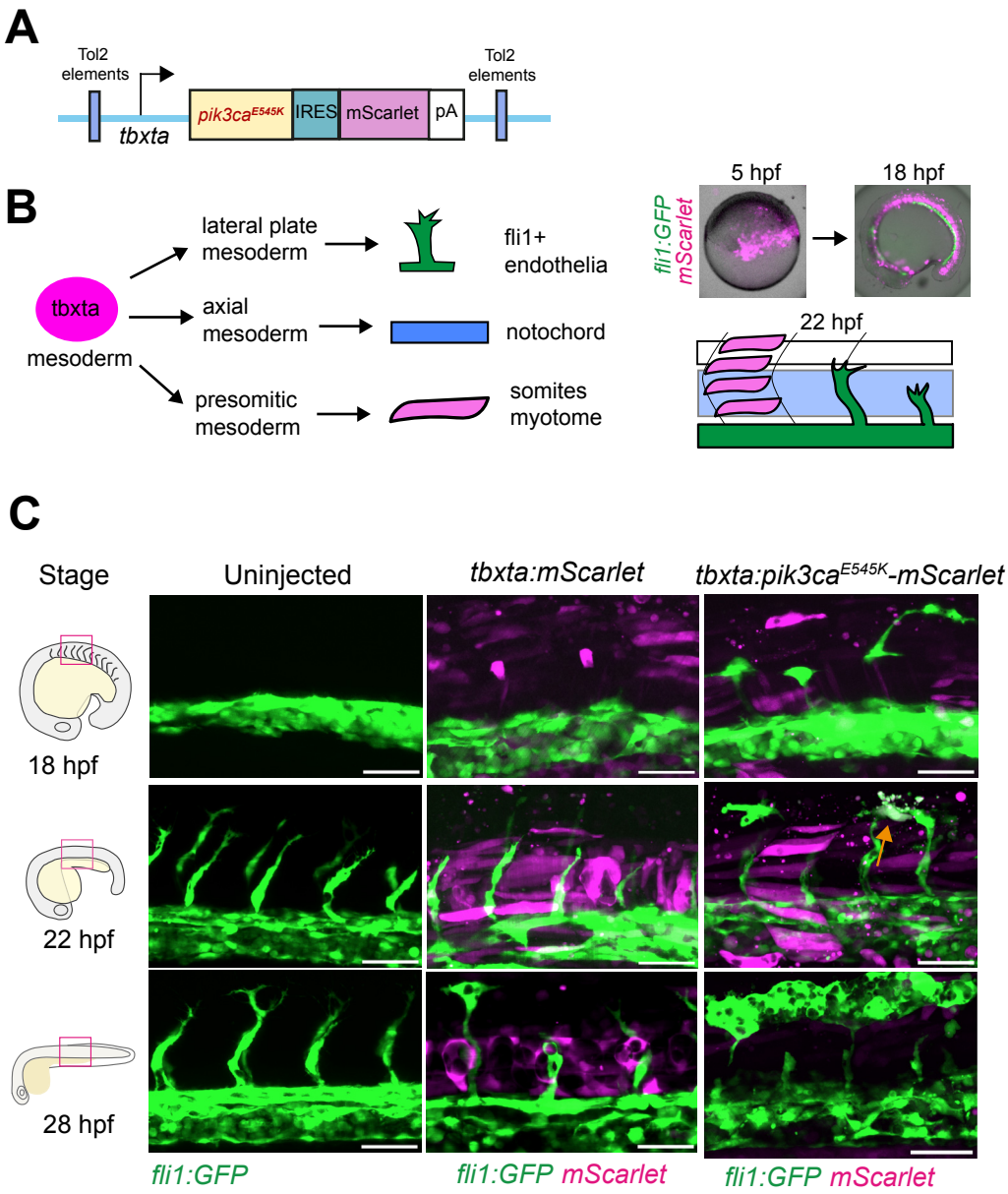

A
Tol2 elements
Tol2 elements
tbxta
pik3ca^{E545K}
IRES
mScarlet
pA
B
tbxta
mesoderm
lateral plate mesoderm
axial mesoderm
presomitic mesoderm
fli1+ endothelia
notochord
somites myotome
5 hpf
18 hpf
fli1:GFP mScarlet
22 hpf
C
Stage
Uninjected
tbxta:mScarlet
tbxta:pik3ca^{E545K}-mScarlet
18 hpf
22 hpf
28 hpf
fli1:GFP
fli1:GFP mScarlet
fli1:GFP mScarlet
D
18 hpf
****
****
ns
number ectopic eGFP+ cells
20
15
10
5
0
tbxta:mSc
tbxta:WT-mSc
tbxta:E545K-mSc

Figure 3

Figure 3: Early mesodermal *pik3ca*^{PROS} transgene expression causes endothelial mis-patterning

A) Plasmid design of the transgenic construct *tbxta:pik3ca*^{E545K}*-mScarlet*.

B) Simplified diagram of a *tbxta* progenitor cell which may differentiate into the lateral plate mesoderm, the notochord, or presomitic mesoderm which gives rise to endothelia, the notochord or muscle respectively. Mosaic *tbxta:mScarlet* is visible at 5 hpf (representative images) before onset of *fli1:eGFP* reporter expression. By 18 hpf, *tbxta* is most highly expressed in the notochord.

C) Representative maximum projection confocal images of embryos from a range of stages between somitogenesis (~18 hpf) and tailbud (~28 hpf) to visualise the start of *fli1:eGFP* expression and endothelial development. In uninjected embryos, lateral stripes of *fli1:eGFP* are present from 16hpf, with sprouting of ISVs occurring from anterior to posterior from around 19 hpf onwards. In *tbxta:pik3ca*^{E545K}*-mScarlet* injected embryos, ectopic *fli1:eGFP* cells can be seen dorsal to the initial lateral plate mesoderm stripe. Scale bars = 50 µm

D) Quantification of ectopic *fli1:eGFP*+ cells in injected *tbxta:pik3ca*^{E545K}*-mScarlet,* or injected with a similar construct lacking the *pik3ca-ires* element (unattached to the lateral mesoderm stripes of *fli1:eGFP*). 3 biological replicates shown with each dot representing 1 embryo. Significance was calculated with a two-tailed Mann Whitney test. ****=$p<0.001$.

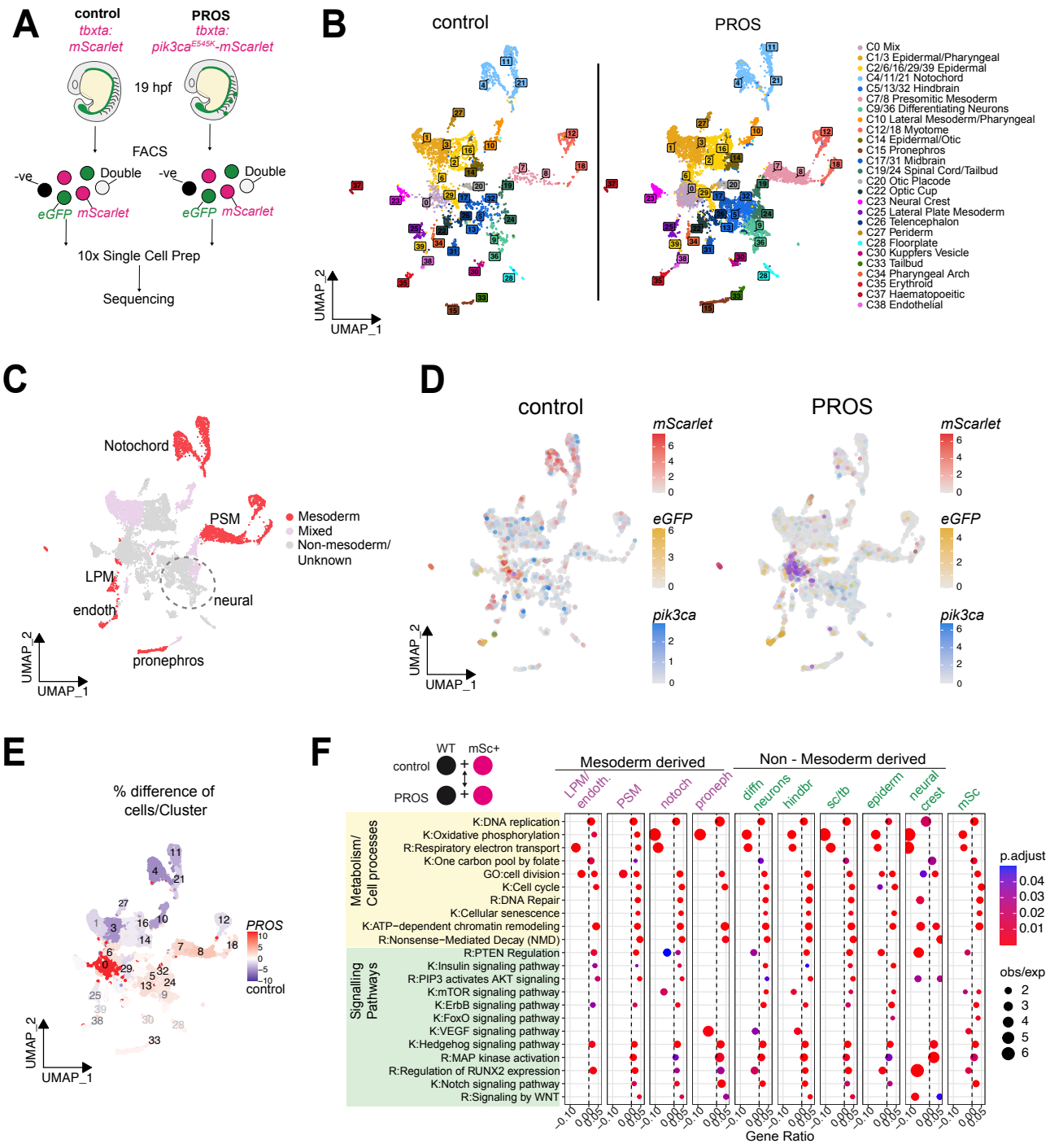

A
control
tbxta:
mScarlet
PROS
tbxta:
pik3ca E545K-mScarlet
19 hpf
FACS
-ve
Double
eGFP
mScarlet
10x Single Cell Prep
Sequencing
B
control
PROS
UMAP_2
UMAP_1
C0 Mix
C1/3 Epidermal/Pharyngeal
C2/6/16/29/39 Epidermal
C4/11/21 Notochord
C5/13/32 Hindbrain
C7/8 Presomitic Mesoderm
C9/36 Differentiating Neurons
C10 Lateral Mesoderm/Pharyngeal
C12/18 Myotome
C14 Epidermal/Otic
C15 Pronephros
C17/31 Midbrain
C19/24 Spinal Cord/Tailbud
C20 Otic Placode
C22 Optic Cup
C23 Neural Crest
C25 Lateral Plate Mesoderm
C26 Telencephalon
C27 Periderm
C28 Floorplate
C30 Kuppfers Vesicle
C33 Tailbud
C34 Pharyngeal Arch
C35 Erythroid
C37 Haematopoeitic
C38 Endothelial
C
Notochord
PSM
LPM
endoth
neural
pronephros
Mesoderm
Mixed
Non-mesoderm/
Unknown
D
control
PROS
mScarlet
eGFP
pik3ca
E
% difference of
cells/Cluster
PROS
control
F
WT
mSc+
Mesoderm derived
Non - Mesoderm derived
LPM/ endoth.
PSM
notoch
proneph
diffn
neurons
hindbr
sc/tb
epiderm
neural crest
mSc
Metabolism/ Cell processes
K:DNA replication
K:Oxidative phosphorylation
R:Respiratory electron transport
K:One carbon pool by folate
GO:cell division
K:Cell cycle
R:DNA Repair
K:Cellular senescence
K:ATP−dependent chromatin remodeling
R:Nonsense−Mediated Decay (NMD)
Signalling Pathways
R:PTEN Regulation
K:Insulin signaling pathway
R:PIP3 activates AKT signaling
K:mTOR signaling pathway
K:ErbB signaling pathway
K:FoxO signaling pathway
K:VEGF signaling pathway
K:Hedgehog signaling pathway
R:MAP kinase activation
R:Regulation of RUNX2 expression
K:Notch signaling pathway
R:Signaling by WNT
Gene Ratio
p.adjust
obs/exp

Figure 4

Figure 4. Mosaic mesodermal *pik3ca*^{PROS} expression induces pan-lineage dysregulation

A) Schematic of experimental workflow for scRNA-seq. *Tg(fli1:eGFP)* embryos were injected with *tbxta:mScarlet* (control) or *tbxta:pik3ca*^{E545K}*-mScarlet* (PROS) and grown until 19 hpf, dissociated, and all live cells sorted by FACS before processing for 10x sequencing.
B) UMAP showing 40 cell clusters found the aggregated dataset, split by experimental condition.
C) UMAP of aggregated dataset coloured by mesodermal or other germ layer origin
D) FeaturePlots split by experimental condition, showing expression levels of *mScarlet* (red scale), *eGFP* (yellow scale) and *pik3ca* (blue scale)
E) The numbers of cells per cluster was expressed as a percentage of total cells for each experimental condition, and then the difference between percentages in PROS and control clusters visualised in a UMAP heat plot. Blue indicates a higher percentage of control cells relative to PROS, and red indicates a higher proportion of PROS relative to control cells.
F) Deseq2 DE analysis and gene enrichment analyses using GO (G:), KEGG (K:) and Reactome (R:) databases were performed to compare PROS mosaic vs control cells in various cluster groups, or in *mScarlet*+ cells across the dataset.

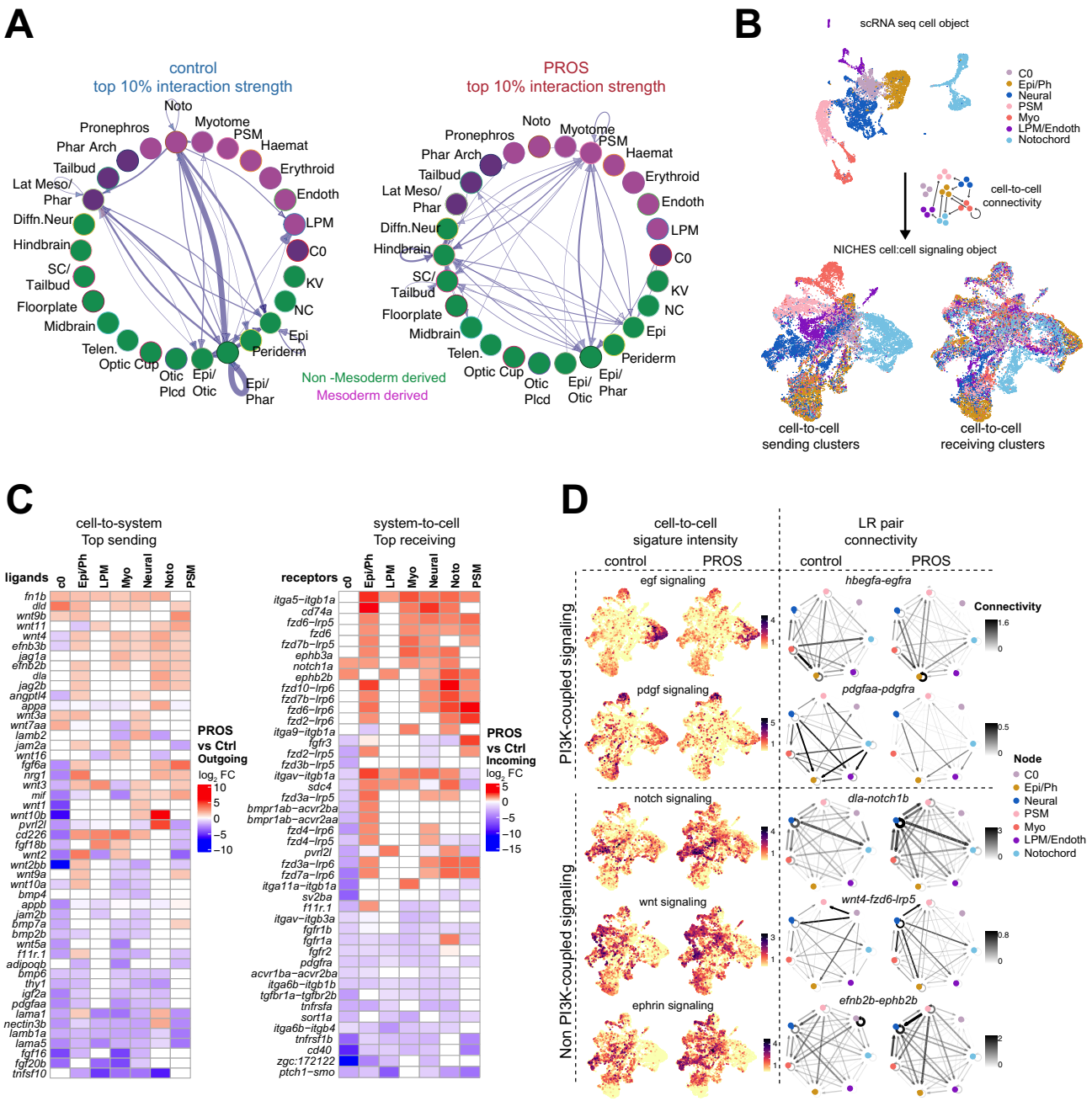

A
control
top 10% interaction strength
PROS
top 10% interaction strength
Noto
Pronephros
Myotome
PSM
Haemat
Phar Arch
Tailbud
Erythroid
Lat Meso/
Phar
Endoth
Diffn.Neur
LPM
Hindbrain
C0
SC/
Tailbud
KV
Floorplate
NC
Midbrain
Epi
Telen.
Periderm
Optic Cup
Otic
Plcd
Epi/
Otic
Epi/
Phar
Non -Mesoderm derived
Mesoderm derived
B
scRNA seq cell object
C0
Epi/Ph
Neural
PSM
Myo
LPM/Endoth
Notochord
cell-to-cell
connectivity
NICHES cell:cell signaling object
cell-to-cell
sending clusters
cell-to-cell
receiving clusters
C
cell-to-system
Top sending
system-to-cell
Top receiving
ligands
receptors
c0
Epi/Ph
LPM
Myo
Neural
Noto
PSM
PROS
vs Ctrl
Outgoing
log2 FC
PROS
vs Ctrl
Incoming
log2 FC
D
cell-to-cell
sigature intensity
LR pair
connectivity
control
PROS
control
PROS
PI3K-coupled signaling
Non PI3K-coupled signaling
egf signaling
pdgf signaling
notch signaling
wnt signaling
ephrin signaling
hbegfa-egfra
pdgfaa-pdgfra
dla-notch1b
wnt4-fzd6-lrp5
efnb2b-ephb2b
Connectivity
Node
C0
Epi/Ph
Neural
PSM
Myo
LPM/Endoth
Notochord

Figure 5

Figure 5. *pik3ca*^{PROS} promotes widespread transcriptional changes in ligand-receptor expression patterns

A) CellChat was used to identify potential interactions between ligand and receptor pairs in clusters. The top 10% strongest pathway interactions of 'control' cells relative to 'PROS' and vice versa shown, normalised to cell number. Blue arrows indicating whether a cluster is a sender or a receiver of the signal, with arrow thickness proportional to connection strength. Abbreviations: LPM=lateral plate mesoderm, PSM, presomitic mesoderm, Noto = notochord, Endoth=endothelial, KV = kuppfer's vesicle, NC=neural crest, Epi=epiderm, Epi/Phar epidermal/pharyngeal, Otic Plcd=otic placode, Telen=telencephalon, SC=spinal cord, Diffn. Neuron = differentiating neurons,

B) UMAP showing merged data to illustrate the methodology of NICHES, whereby a subset of clusters from Figure 4 is reclustered (Top UMAP). Ligand-receptor interactions between each group were quantified using NICHES, converting it into a signaling object of cell:cell interactions (Bottom UMAPs, grouped by either sending cluster type (left) or receiving cluster type (right). Colours in interaction maps were retained from the original clusters in the top UMAP. Epi/Ph = epiderm/pharyngeal mix, PSM = presomitic mesoderm, Myo=myotome, LPM/Endothelial = grouped lateral plate mesoderm and endothelial clusters.

C) Heatmap depicting $\log_2$ fold change in expression of the top differentially expressed "sending" ligand (left) and "receiving" receptor (right) genes in PROS relative to control embryos, clustered by row. Ligand-receptor mechanisms were aggregated by common ligand or common receptor (unaggregated data are shown in **Supplemental Figure 4C**). Only differences for which p.adj<0.05 by Wilcoxon test between PROS and WT are shown.

D) Examples of pathways showing dysregulated expression of ligand-receptor (LR) pairs. The expression values of LR pairs for each pathway were averaged as a module score. FeaturePlots (left) show intensity of this module for each cell:cell interaction, split by experimental condition. Examples of circus plots for each pathway are shown on the right. Signal direction between clusters (nodes) is shown by arrows, with darker and thicker arrows corresponding to greater connectivity strength. Connectivity scale intensity is matched between control and PROS plots for each LR pair.