**Bronchopulmonary Dysplasia with Pulmonary Hypertension Associates with Loss of Semaphorin Signaling and Functional Decrease in FOXF1 Expression**

Shawyon P. Shirazi[1,2], Nicholas M. Negretti[1, 2], Christopher S. Jetter[1], Alexandria L. Sharkey[1], Shriya Garg[1], Meghan E. Kapp[3], Devan Wilkins[1], Gabrielle Fortier[2], Saahithi Mallapragada[4], Nicholas E. Banovich[4,5], Christopher V. E. Wright[2,5,6], David B. Frank[7], Jonathan A. Kropski[2,5,6,8,9]*, Jennifer M. S. Sucre[1,2,5,6]*

[1] Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN

[2] Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN

[3] Department of Pathology, Case Western Reserve University Hospitals, Cleveland, OH

[4] Division of Bioinnovation and Genome Science, Translational Genomics Research Institute, Phoenix, AZ

[5] Biodevelopmental Origins of Lung Disease (BOLD) Center, Vanderbilt University Medical Center

[6] Program in Developmental Biology, Vanderbilt University, Nashville, TN

[7] Department of Cardiology, Children's Hospital of Philadelphia, Philadelphia, PA

[8] Department of Medicine, Vanderbilt University Medical Center, Nashville, TN

[9] Department of Veterans Affairs Medical Center, Nashville, TN

*denotes equal contribution

**Disclosures:** JAK reports grant funding from Boehringer Ingelheim and Bristol-Myers Squibb outside the scope of this work, and serves on the Scientific Advisory Board for APIE and ARDA, outside of the scope of this work.

Address correspondence to

Jennifer M.S. Sucre, MD

jennifer.sucre@vumc.org

Jonathan Kropski, MD

jon.kropski@vumc.org

**Abstract**

Lung injury in preterm infants leads to structural and functional respiratory deficits, with a risk for bronchopulmonary dysplasia (BPD) that in its most severe form is accompanied by pulmonary hypertension (PH). To examine cellular and molecular dynamics driving evolving BPD in humans, we performed single-cell RNA sequencing of preterm infant lungs in early stages of BPD and BPD+PH compared to term infants. Analysis of the endothelium revealed a unique aberrant capillary cell-state primarily in BPD+PH marked by *ANKRD1* expression. Predictive signaling analysis identified deficits in the semaphorin guidance-cue signaling pathway and decreased expression of pro-angiogenic transcription factor *FOXF1* within the alveolar parenchyma in neonatal lung samples with BPD/BPD+PH. Loss of semaphorin signaling was replicated in a murine BPD model and in humans with alveolar capillary dysplasia (ACDMPV), suggesting a mechanistic link between the developmental programs underlying BPD and ACDMPV and a critical role for semaphorin signaling in normal lung development.

## Introduction

In infants born preterm, abnormal lung development leading to chronic lung disease is a leading complication, resulting in bronchopulmonary dysplasia (BPD)[1]. In its most severe form, BPD is associated with life-threatening pulmonary hypertension (BPD+PH), which is thought to be caused by abnormal development of the pulmonary vasculature, with decreased numbers of alveolar-capillary units formed for gas exchange[2,3]. Despite increasing awareness of the environmental and perinatal risk factors that increase the risk of BPD, there are very few disease-modifying therapies, and our understanding of the molecular mechanisms and potential therapeutic targets remains limited.

The advent of single-cell multiomics technology as applied to human lung samples has enabled an increasingly nuanced understanding of the cell populations and cell-states under homeostatic and diseased conditions in adults[4–6]. However, because neonatal biopsies are infrequently performed and infant autopsies are rare in general, lung tissue from preterm infants and healthy neonates is relatively under-represented in most of the currently available large sequencing series. Here we report single-cell transcriptomic findings from former preterm infants with evolving BPD, and infants with both BPD and PH, compared to term infants at approximately the same corrected gestational age as the preterm infants.

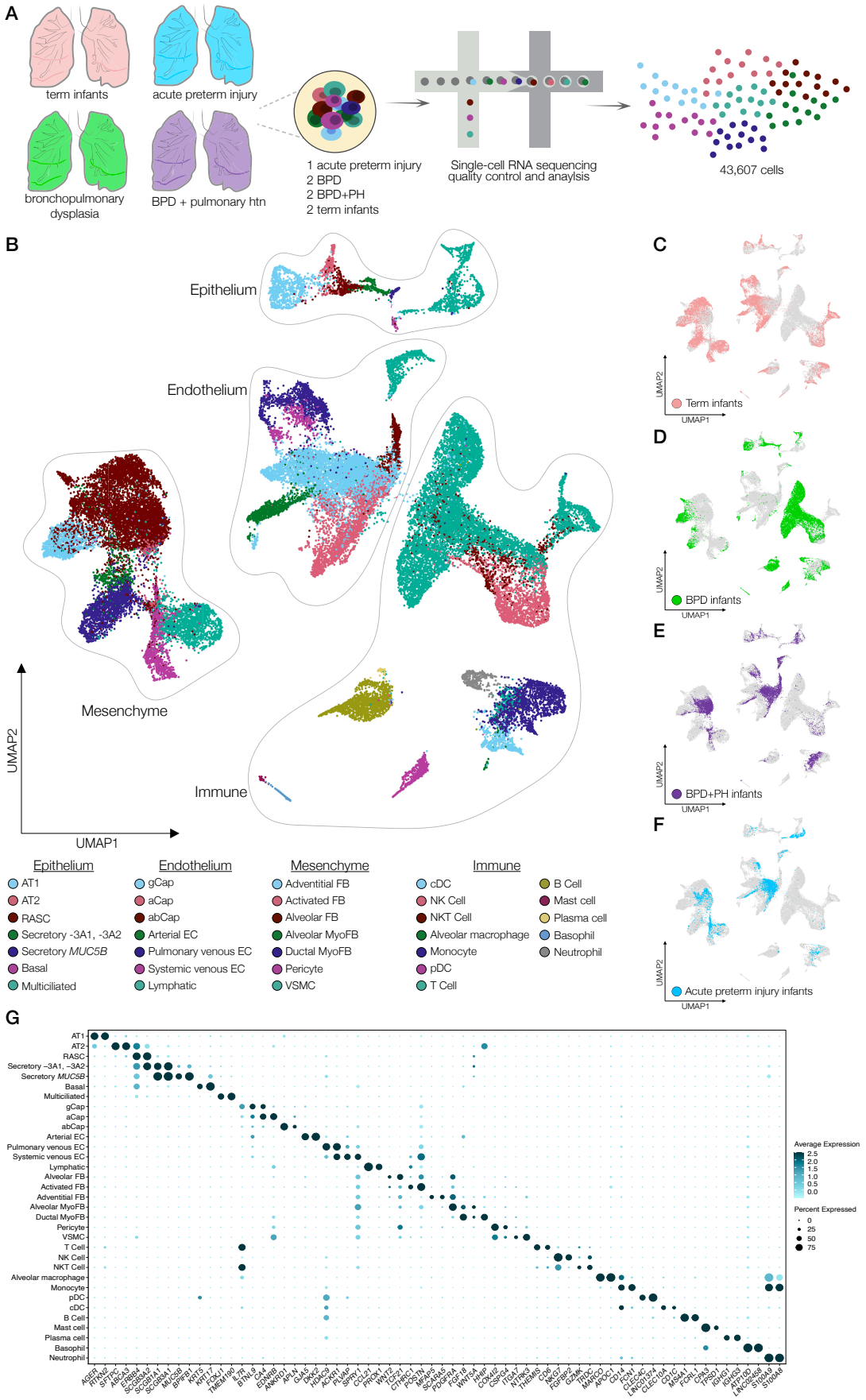
While we anticipate that this transcriptional dataset will be a resource for future work in neonatal lung injury and BPD in general, we were particularly interested in the endothelial cell populations in infants with BPD+PH and how they might differ from corrected gestational age-matched controls and patients with BPD at a similar age. In these sequencing data, we identified a capillary endothelial cell-state highly specific to patients with BPD+PH and spatially localized these capillary cells to the distal alveolar parenchyma, and not in more proximal tissue areas. Further, we employed predictive signaling analysis to generate hypotheses about the role of semaphorin guidance-cue signaling during alveolar formation, a pathway we identified as deficient in both BPD and BPD+PH patients transcriptomically, and then spatially localized to the developing alveolar regions. Our parallel identification of semaphorin signaling deficits in alveolar epithelial and endothelial cells in an established mouse model of BPD suggests that this pathway is central to normal alveologenesis. Further, we find evidence of dysregulated semaphorin signaling in patients with alveolar capillary dysplasia with misaligned pulmonary veins (ACDMPV), implicating semaphorin signaling as a potential mechanism driving abnormal lung development after preterm birth.

47 **Results**

48 To examine the cellular dynamics and candidate mediators of neonatal lung injury and
49 the development of severe BPD, we used the 10x Genomics Chromium Platform to perform
50 single-cell RNA sequencing (scRNAseq) of human neonatal lungs at varying stages of lung
51 injury and chronic lung disease: one acute preterm infant with lung injury, two BPD samples
52 at term corrected gestational age (CGA) , two BPD+PH at term CGA, and two term infants in
53 the first three weeks of life for age-matched comparison (**Fig. 1A**, **Supplementary Table 1**).
54 After filtering and quality control to remove doublets and ambient RNA (see materials and
55 methods), we recovered 43,607 single-cell transcriptomes with a range of 2,199 to 8,339 cells
56 per subject, a median of 4,128 unique molecular identifiers (UMI) (interquartile range 1,188 –
57 4,670) per cell, and a median of 1,795 genes (interquartile range 728 – 1,944) per cell (**Fig. 1A-**
58 **B**). Unbiased clustering with cell-type annotation informed by prior human, fetal, and
59 neonatal studies identified seven epithelial, seven endothelial, seven mesenchymal, and twelve
60 immune-cell clusters (**Fig. 1B-G**)[4–12]. To specifically focus on potential vascular cell differences
61 in BPD patients with and without PH, the endothelial cluster was extracted, re-embedded, and
62 separately analyzed to identify cell-type-specific differences across homeostasis and disease.
63 These sequencing data are available through an interactive data explorer at
64 www.sucrelab.org/lungcells.

**Fig. 1: Analysis of neonatal human lung single-cell transcriptomics identifies distinct populations of epithelial, endothelial, mesenchymal, and immune cells**

A) Experimental design overview for single-cell sequencing of two term control infants, one acute neonatal lung injury, two infants with severe bronchopulmonary dysplasia, and two patients with severe bronchopulmonary dysplasia and pulmonary hypertension. B) Uniform Manifold Approximation and Projection (UMAP) embedding of the neonatal human lung dataset (n = 43,607 cells). Cells labeled by lineage and by cell type. C-F) Breakdown of the dataset's UMAP embedding by disease-state condition annotation. G) Dot plot of the hallmark-identifying genes for each cell cluster. Dot size corresponds to the percentage of cells within a cluster expressing a certain gene. Color intensity corresponds to the relative average expression level.



*Bronchopulmonary dysplasia with pulmonary hypertension is associated with a specific population of capillary endothelial cells*

Analysis and re-embedding of the endothelial cells identified a population primarily obtained from BPD+PH patients with transcriptional hallmarks common to both alveolar capillaries (aCaps, CAP2) and general capillaries (gCaps, CAP1) (**Fig. 2A-B**, **S1A**). This group of 647 cells is marked by high expression of *ANKRD1*, a gene which is nearly absent among all other endothelial cell clusters, with limited detection in alveolar type 1 (AT1) cells (**Fig. 2C**). This aberrant capillary cell-state (abCap) expresses some aCap-associated genes, e.g. *APLN*, *EMCN*, *S100A4*, but has markedly reduced expression of aCap hallmark genes *EDNRB* and *SOSTDC1* when compared to other aCaps. Similarly, abCaps express some genes associated with gCaps, e.g., *PTPRB* and *EDN1*, but lack expression of canonical gCap genes *FCN3*, *IL7R*, and *BTNL9* (**Fig. 2D**). The abCaps also express certain pan-capillary markers *PDGFB*, *RGCC*, and *ADGRL2*, but not *CA4*, *CD36*, and *AFF3*. In addition to the YAP/TAZ signaling target *ANKRD1*, this population expresses relatively high levels of other YAP targets *CCN1* (*CYR61*), *CCN2* (*CTGF*), *SERPINE1*, and *CRIM1*, compared to other endothelial cells (**Fig. 2D**). Notably, the abCaps present a gene-expression profile similar to previously described endothelial tip cells[13,14] , a population present in early lung development that directionally guides the growth and migration of capillary endothelium. Relative to other endothelial cells, abCaps have increased expression of tip cell hallmarks *ESM1*, *ANGPT2*, *APLN*, *MCAM*, *COTL1*, but lower expression of tip cell–associated receptors such as *KDR*, *FLT4*, *UNC5B*, and *NRP1,* relative to other endothelial cells (**Fig. 2E**).

To spatially validate and localize the abCap cell-state, we performed RNA *in situ* hybridization (RNA ISH) using formalin-fixed paraffin-embedded (FFPE) lung tissue blocks from preterm infants with BPD with and without PH, as well as from term infant control lungs (**Fig. 2F-H**). These FFPE tissues are from an expanded set of samples, independent of those interrogated by scRNAseq (**Supplementary Table 2**). In lung tissue from patients with BPD+PH, *ANKRD1+/PECAM1+* endothelial cells were localized to the alveolar parenchyma, near *SFTPC+* alveolar type 2 (AT2) cells, consistent with their expression of capillary hallmark genes. Quantification demonstrated a marked enrichment of this population in patients with BPD+PH when compared with age-matched BPD patients and normal term infant controls of approximately the same corrected gestational age **(Fig. 2I**).

96 *The abCap endothelial cell state does not transcriptionally resemble endothelial cells previously*
97 *described in different human lung diseases*

98 To understand abCaps in the context of previously reported atlases of human
99 pulmonary endothelium under normal and disease conditions, we performed analysis with
100 SCimilarity[15], a computational tool that allows comparison of a given cell population with cell
101 populations from nearly 400 previously published human datasets, including 41 from the
102 lung. Comparing abCaps across the 1.12-million endothelial cells in the 22.7-million cell corpus
103 of SCimilarity, we selected the top 100 human samples (contained in 34 source studies) with
104 the highest SCimilarity score to our BPD+PH capillary cells. Of these 34 studies, we excluded
105 seven studies of iPSCs, 14 that did not contain lung tissue, two with samples from patients in
106 the $7^{th}$ – $8^{th}$ decade of life, two with a sample size < 2, and two with inaccessible scRNAseq data
107 files. This subtraction left six studies with 178 samples for comparison (**Fig. S1B**). We
108 compared the abCap BPD+PH endothelial cell state to the endothelial cells of healthy controls
109 and five disease states – COVID, LAM, IPF, SSc-ILD, and PAH. Using the genes most specific
110 for abCaps, we created a gene-module score to identify cells with expression patterns
111 characteristic of abCaps (see methods for details). Among these six studies, fewer than 12% of
112 cells in each endothelial sub-population were showed abCap similarity (**Fig. 2J**). As a positive
113 control that the selected module score is inherently selective of the abCap population,
114 applying this metric to our entire dataset confirmed that the cells containing the genes with the
115 highest gene-module score were within the BPD+PH associated abCap population.

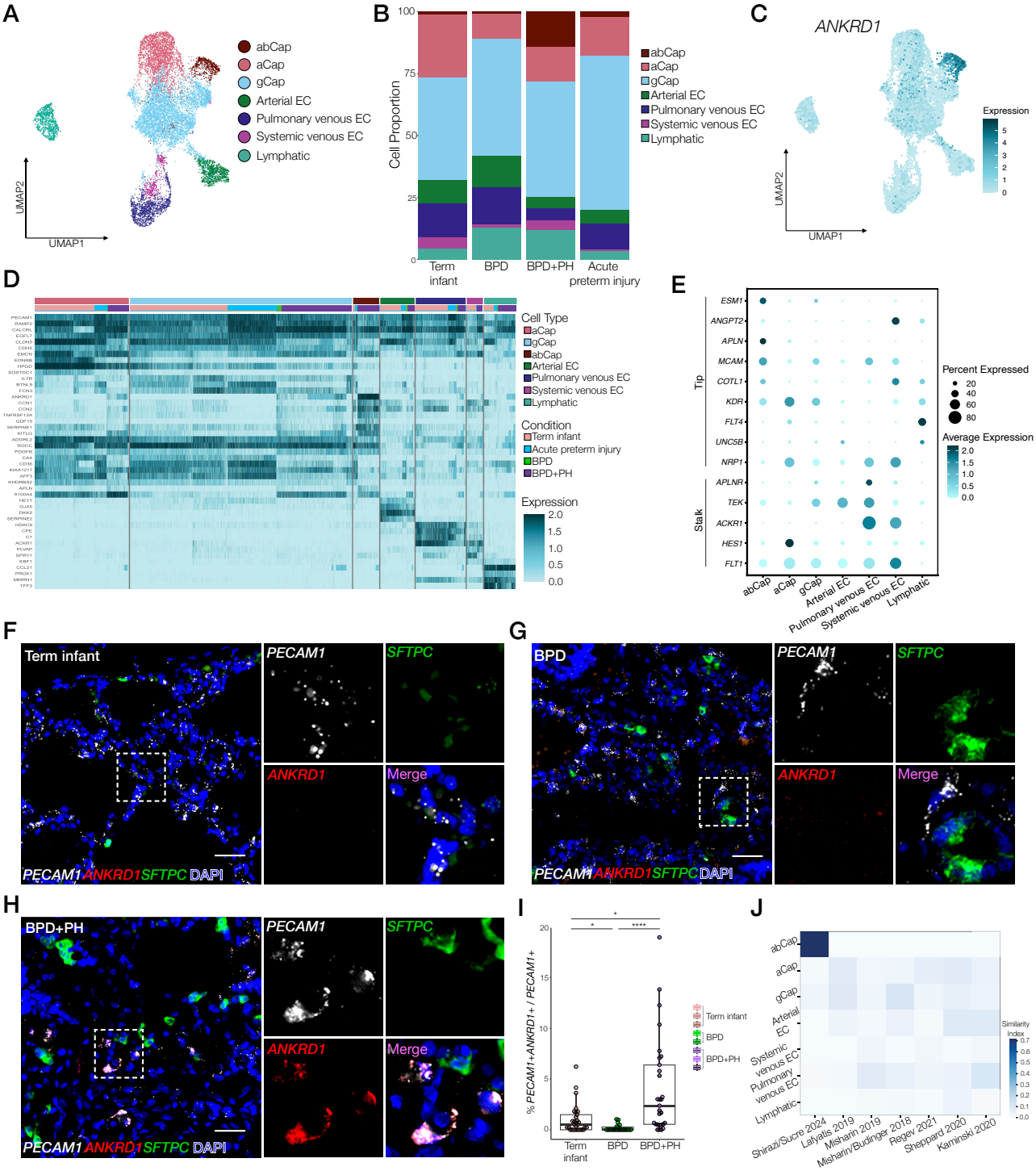



**Fig. 2: Analysis of the endothelium reveals a cluster of aberrant capillary cells (abCaps) enriched in patients with BPD+PH**

A) UMAP embedding of the neonatal human lung endothelial cells (n = 11,816) labeled by cell type. B) Stacked area plot of cellular proportions across the four disease-state conditions. C) UMAP of the expression level of *ANKRD1*, the highest expressing gene in the abCaps. D) Heatmap of hallmark genes of the seven identified endothelial cell type clusters, demonstrating a distinct expression profile for abCaps when compared with aCaps and gCaps. The cell types are subdivided by disease condition. Maximum expression values are clipped at 2. E) Dot plot of the hallmark genes for endothelial tip and stalk cells. Dot size corresponds to the percentage of cells within a cell type cluster expressing a certain gene. Color intensity increases with relative average expression level. F-H) RNA ISH for *ANKRD1* (red), *PECAM1* (white), and *SFTPC* (green) demonstrating co-localization of *ANKRD1* within *PECAM1*+ cells in BPD+PH patients in the alveolar parenchyma. Scale bar = 30 μm. I) RNA ISH quantification using HALO of the percentage of *PECAM1*+ cells that are also *ANKRD1*+ over the total amount of *PECAM1*+ cells (*PECAM1*+*ANKRD1*+/ all *PECAM1*+) across disease condition, * *p*-value<0.05, **** *p*-value<0.0001, Kruskal-Wallace one-way ANOVA. J) Heatmap showing relative homology from top 6 datasets identified by SCimiliarity, with intensity of color indicating greater similarity, abCap cell included in upper right at positive control.

116 *Analyzing putative ligand-receptor interactions identifies deficient semaphorin signaling between*
117 *abCaps and the alveolar niche in BPD and BPD+PH*

118 To characterize how abCaps might interact with other cells in the alveolar niche, we
119 interrogated our scRNAseq data with CellChat, which predicts signaling interactions by
120 expression of ligand-receptor pairs. In intracellular communication, cells can be classified as
121 having more outgoing or incoming signaling information[16]. Broadly, CellChat identified
122 endothelial capillaries as acting predominantly as "receivers" of cell-signaling inputs, with
123 AT1 cells and mesenchymal cells having a higher "sender" score (**Fig. 3A**). CellChat identified
124 semaphorin signaling as a leading differentially expressed candidate pathway differentiating
125 term controls and infants with BPD and BPD+PH, with significantly decreased interactions in
126 both epithelial-capillary and mesenchymal-capillary interactions (**Fig. 3B**). Interrogation of the
127 transcriptomic dataset for the specific ligands and receptors of this pathway revealed that
128 BPD+PH patients had markedly decreased expression of *SEMA3B* by AT1 cells, *SEMA3C* by
129 alveolar and ductal myofibroblasts, and *SEMA6A* by endothelial capillary cells, with some
130 decreased expression of these ligands in patients with severe BPD without PH, compared to
131 controls (**Fig. 3C**). Expression of semaphorin receptors *NRP1* and *PLXNA2* was also decreased
132 (**Fig. 3C**) in BPD and BPD+PH patients. Some of these predicted expression changes were
133 spatially validated by RNA ISH, showing decreased expression of *SEMA3B* in *AGER*-
134 expressing AT1 cells and decreased *SEMA6A* in *PECAM1+* alveolar endothelial cells (**Fig. 3D-**
135 **E**).

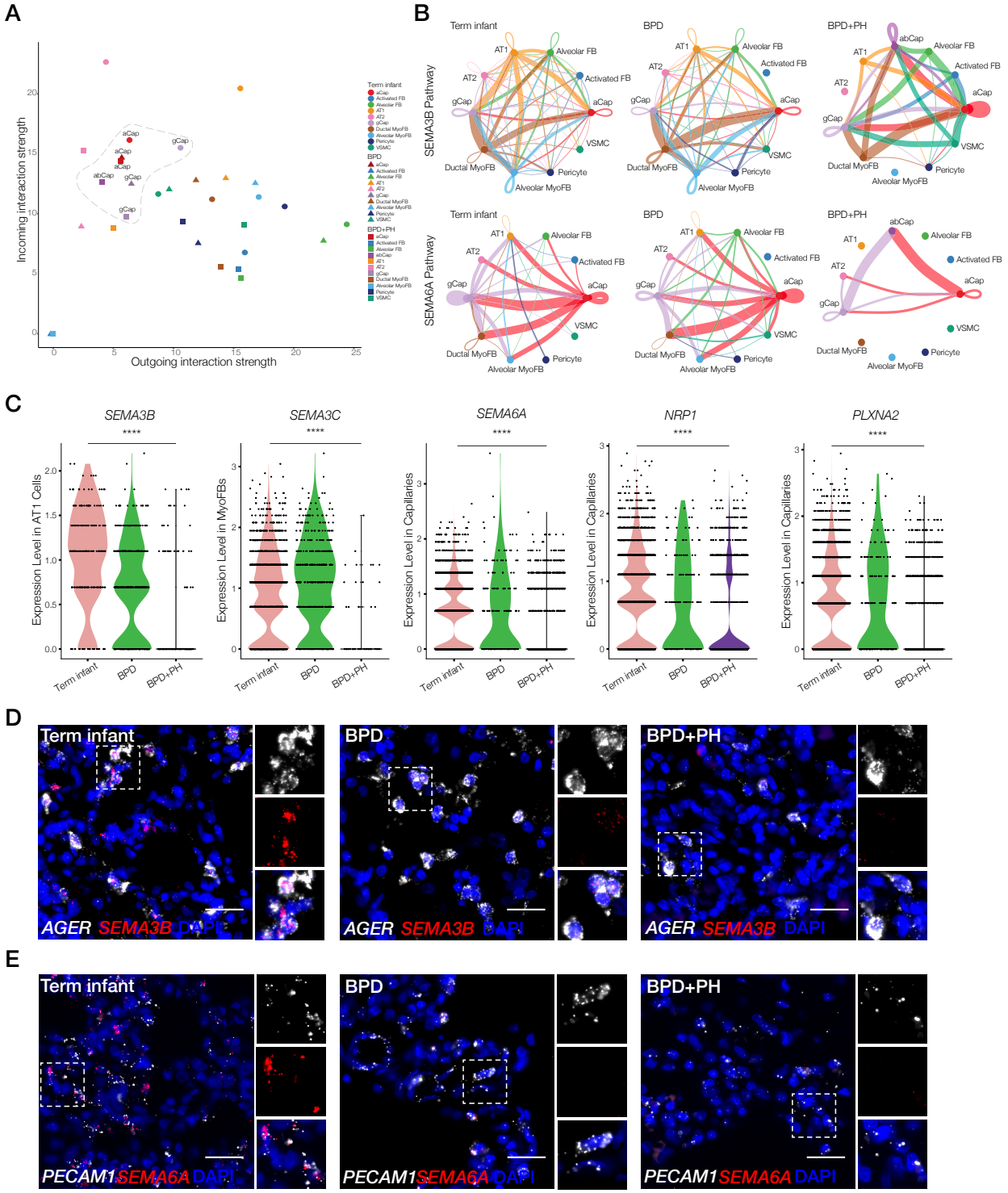



**Fig. 3: Analysis of predicted ligand-receptor interactions by single-cell sequencing identifies decreased semaphorin signaling between endothelial cells and alveolar niche in the setting of BPD and BPD + PH**

A) Cell types plotted by predicted outgoing vs incoming information averaged for all detected pathway interactions from the CellChat ligand and receptor database. B) Circle plots for the SEMA3 and SEMA6 pathways displayed by disease state. The line weights indicate signal interaction strength from gene expression of ligands and receptors for each respective pathway. C) Violin plots for expression levels of semaphorin pathway genes *SEMA3B*, *SEMA3C*, *SEMA6A*, *NRP1*, and *PLXNA2* by disease state, with cell-type specificity indicated on y-axis. **** *p*-value<0.0001 by one-way ANOVA. D) RNA ISH for *SEMA3B* (red) and *AGER* (white), demonstrating co-localization of *SEMA3B* within *AGER*+ AT1 cells in term controls, with decreased expression in patients with BPD and BPD+PH. Scale bar = 25 μm. E) RNA ISH for *SEMA6A* (red) and *PECAM1* (white), demonstrating co-localization of *SEMA6A* within *PECAM1*+ endothelial cells in term controls, with decreased expression in patients with BPD and BPD+PH. Scale bar = 25 μm.

136 *Decreased semaphorin ligand expression replicated in a murine BPD model*

137 To determine whether the decreased semaphorin signaling seen in human tissues could
138 be modeled *in vivo*, we examined a murine model of saccular-stage injury for semaphorin
139 ligand expression. Injury by exposure to hyperoxia (70% $O_2$) and inflammation (by intranasal
140 LPS) in the saccular stage (P1-P5) mimics many of the features of BPD, including impaired
141 alveologenesis (as measured by increased mean linear intercept, a measurement of terminal
142 airspace subdivision) (**Fig. 4A-C**)[17]. RNA ISH on samples from this model demonstrated
143 decreased expression of semaphorin ligands *Sema3b* and *Sema6a* in a cell-type specific pattern
144 (**Fig. 4D-E**) similar to that observed in human patients (**Fig. 3D-E**). Notably, by RNA ISH,
145 there was abundant expression of *Sema3b* and *Sema6a* in uninjured AT1 cells and endothelial
146 cells, respectively, suggesting a conserved role of semaphorin signaling in early alveologenesis
147 under normal developmental conditions.

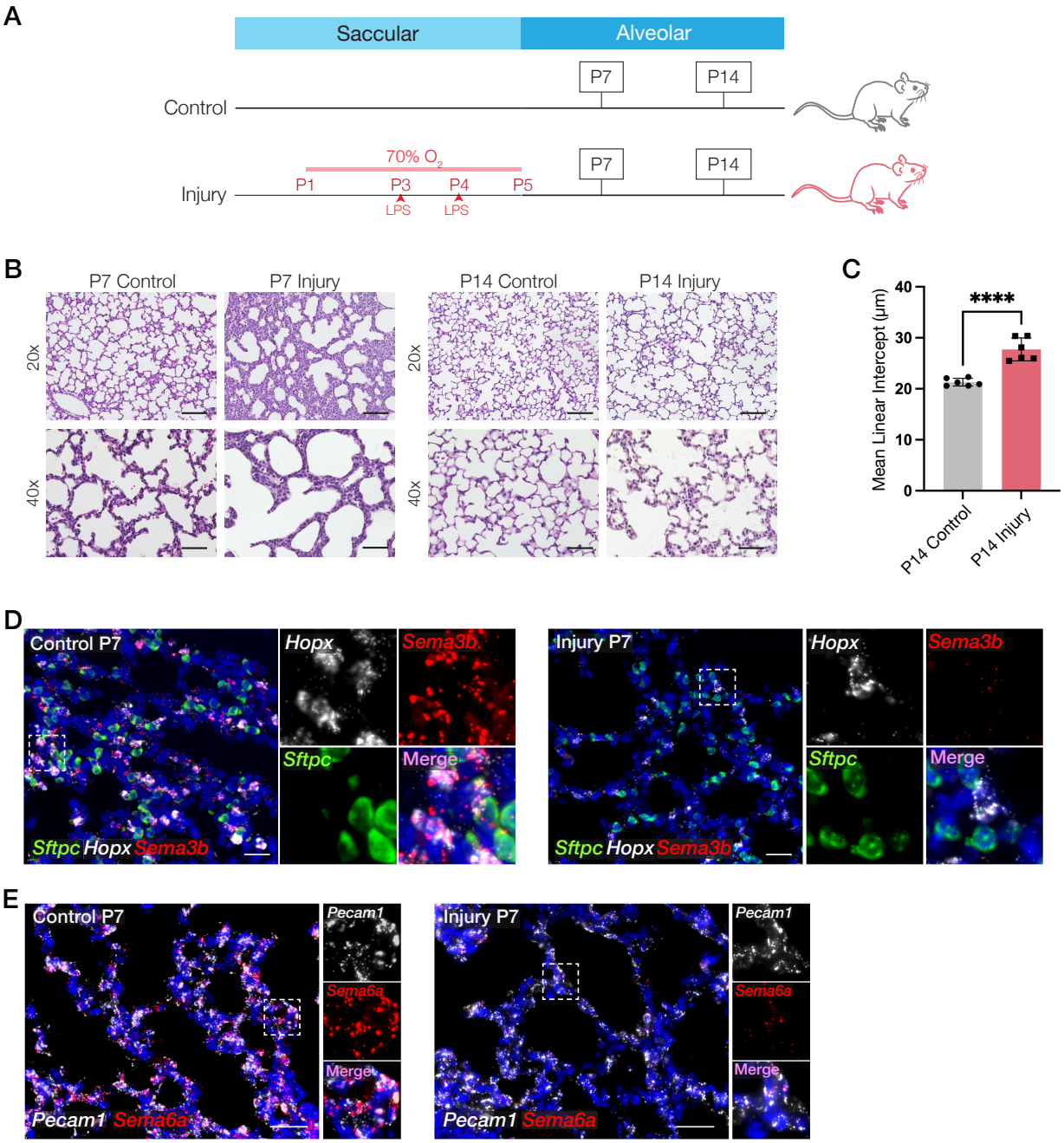



**Fig. 4: Murine mouse injury models of BPD and BPD+PH display decreased semaphorin signaling**

A) Schematic of murine BPD model with hyperoxia exposure from P1-5 and intranasal LPS on P3 and 4. B) H&E of the BPD murine mouse injury model and control mice at P7 and P14 at 20x and 40x magnification. Scale bar = 110 µm (20x), 55 µm (40x). C) Mean linear intercept on the P14 control and injured mouse H&E images at 40x magnification, **** *p*-value<0.0001 by Welch's *t*-test. D) RNA ISH of *Sftpc* (green), *Hopx* (white), and *Sema3b* (red) on the P7 control and injured mouse lungs exhibiting co-localization of *Sema3b*+ within *Hopx*+ epithelial cells in the alveolar parenchyma. Scale bar = 20 µm. E) RNA ISH of *Pecam1* (white) and *Sema6a* (red) on the P7 control and injured mouse lungs exhibiting co-localization of *Sema6a*+ within *Pecam1*+ endothelial cells in the alveolar parenchyma. Scale bar = 35 µm.

*Human infants with ACDMPV demonstrate decreased semaphorin ligand and receptor expression*

Previous work has linked semaphorin signaling to regulation by endothelial-cell transcription factor FOXF1[18], a known upstream mediator of angiogenesis, and mutations in *FOXF1* result in alveolar capillary dysplasia with misalignment of pulmonary veins (ACDMPV), a rare and lethal form of developmental lung disease with severe pulmonary hypertension. To further explore the relationship between FOXF1 and semaphorin signaling, we interrogated a recently published single-nucleus RNA sequencing dataset from infants and children with ACDMPV and age-matched controls[11] (**Fig. 5A, S2**) for expression of semaphorin ligands and receptors. While the expression deficits in semaphorin signaling are not identical in ACDMPV and BPD+PH (**Fig. 5B**), we note that patients with ACDMPV have decreased *SEMA3B* expression in AT1 cells and decreased *SEMA6A* signaling in alveolar capillaries (aCaps + gCaps), reminiscent of the deficits in BPD+PH.

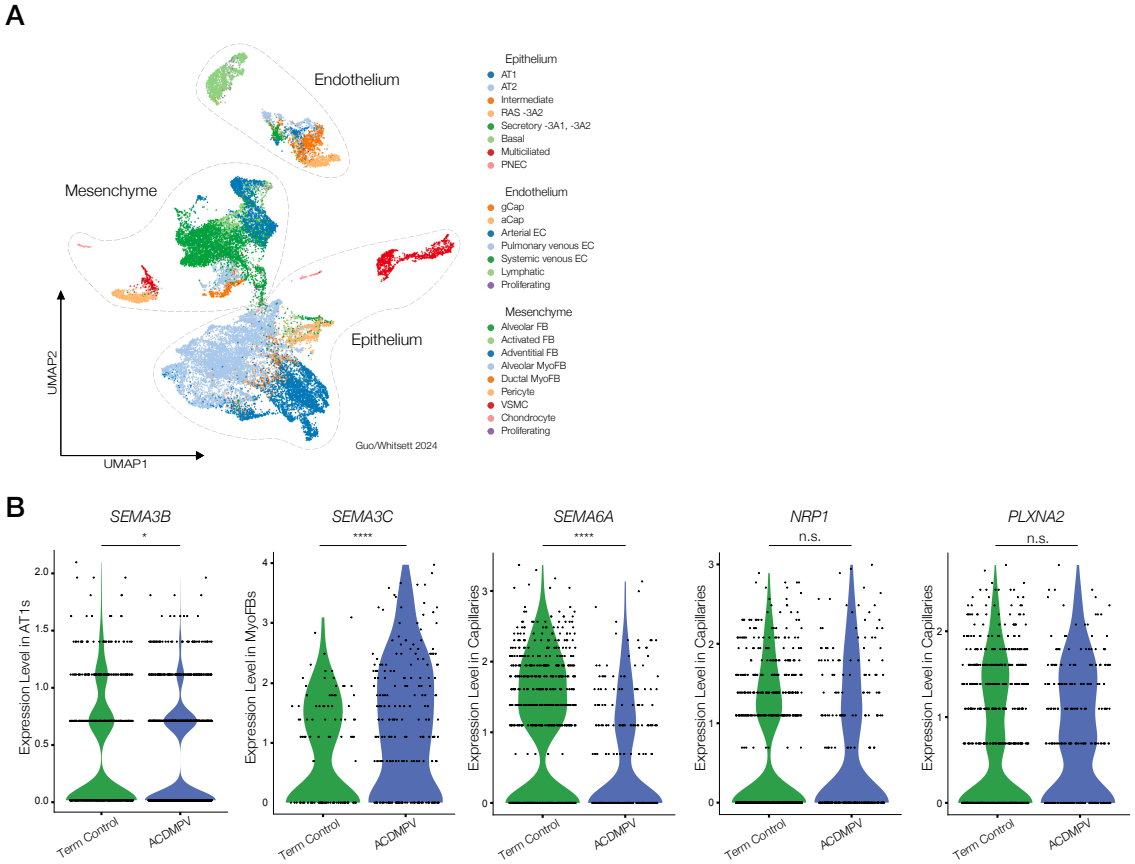



**Fig. 5: Analysis of human infant ACDMPV dataset identifies decreased semaphorin signaling**
A) UMAP embedding of single-nucleus RNA sequencing of five ACDMPV neonatal lung samples[11], three preterm, and three term control lung samples with an age range of 29 weeks gestational age to 3.5 years old annotated by lineage and cell type. B) Violin plots for expression levels of semaphorin pathway genes *SEMA3B*, *SEMA3C*, *SEMA6A*, *NRP1*, and *PLXNA2* by disease state, with cell-type specificity indicated on y-axis. * *p*-value<0.05, **** *p*-value<0.0001 by Welch's *t*-test.

*BPD and BPD+PH lung tissues demonstrate decreased FOXF1 expression*

Interrogation of our neonatal transcriptomic dataset demonstrated significantly decreased *FOXF1* in the capillary endothelium in BPD and BPD+PH conditions (**Fig. 6A**). Immunofluorescence for FOXF1 and the capillary hallmark protein ERG on BPD, BPD+PH, and term infant control lung tissue substantiates the decreased nuclear localization of FOXF1 within endothelial cells of both disease states (**Fig. 6B-D**), suggesting a functional deficit of this critical transcription factor in BPD and BPD+PH.

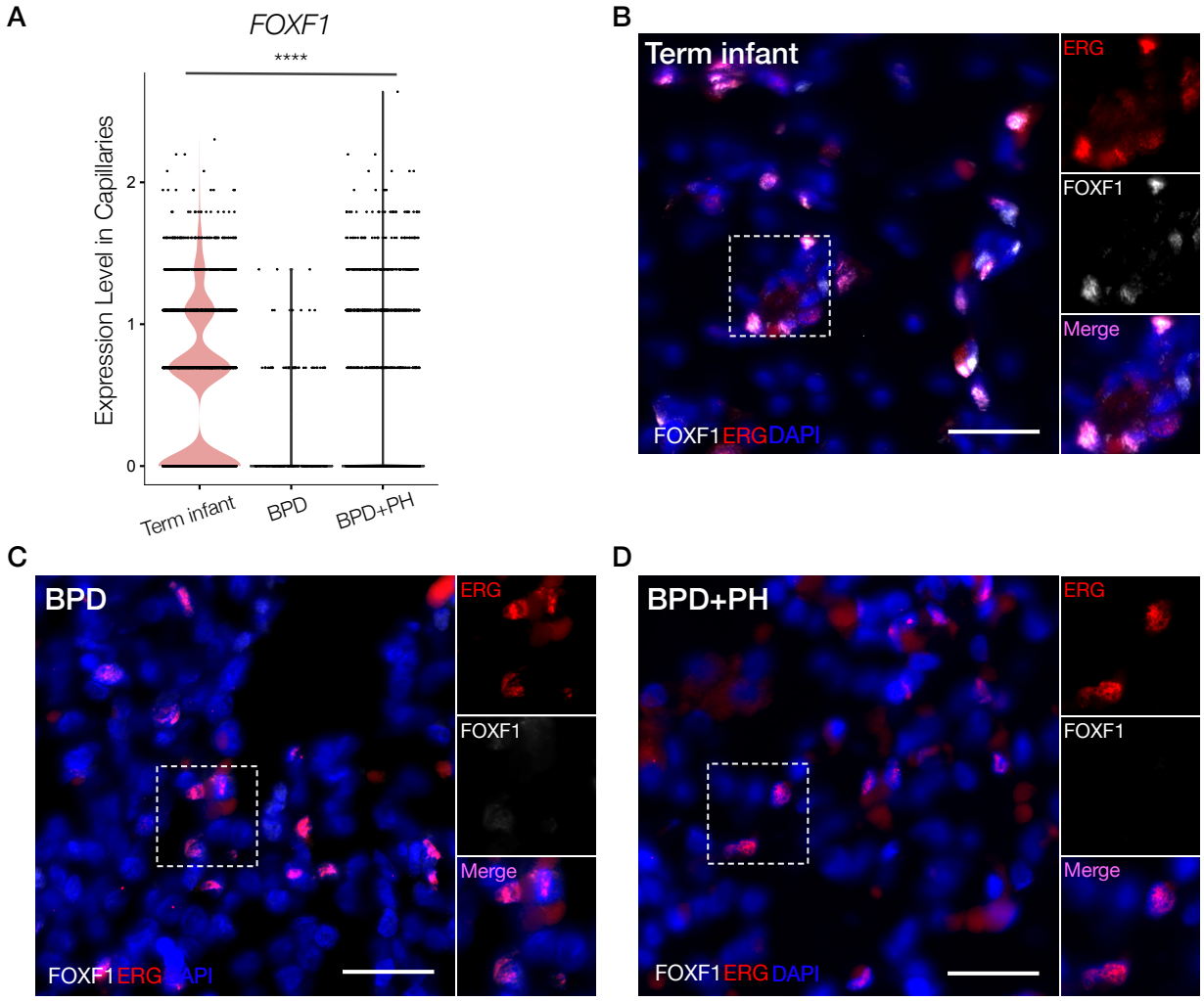



Fig. 6: FOXF1 has decreased expression in human neonatal lung injury

A) Violin plots for expression levels of endothelial transcription factor *FOXF1* in the human neonatal BPD and BPD+PH single-cell RNA sequencing dataset comparing disease state in capillaries, **** *p*-value<0.0001 by one-way ANOVA. B) RNA ISH for ERG (red) and FOXF1 (white) demonstrating co-localization of FOXF1 within ERG+ endothelial cells in the distal lung parenchyma in term controls, with decreased expression of FOXF1 in patients with (C) BPD and (D) BPD+PH. Scale bar = 25 μm.

167 **Discussion**

168 In this single-cell transcriptomic analysis of human neonatal lungs across stages of
169 injury, we identified a previously undefined endothelial cell state, abCap, that is enriched in
170 lungs from infants with bronchopulmonary dysplasia and pulmonary hypertension
171 (BPD+PH). Using predictive analysis tools and tissue validation, we discovered deficits in
172 semaphorin signaling and an apparent functional decrease in FOXF1 in infant tissue
173 representing BPD and BPD+PH. The latter suggests a possible link between the rare disease
174 ACDMPV and the severe forms of abnormal lung development that occur with preterm birth.
175 The concordance of these findings in a murine model of BPD provide foundational data that
176 support a new role for semaphorin signaling as a critical mediator of normal lung
177 development whose disruption may lead to the abnormal alveologenesis and vascular
178 development observed after preterm birth[19].

179 There are potential limitations in this study. While transcriptomic data from these
180 neonates at varying stages of development and evolving injury are informative in generating
181 new hypotheses about molecular mechanisms that drive abnormal lung development after
182 preterm birth, they come from a limited number of subjects, and autopsy samples in general
183 have an inherent bias toward the sickest individuals who have died from BPD and BPD+PH.
184 Nonetheless, these data fill a critical gap regarding the deep transcriptomic analysis of normal
185 and injured lungs across the lifespan, because there are so very few published reports on
186 neonatal premature injured and uninjured lung samples.

187 A central strength of our study is the transcriptomic integration of these rare samples
188 with samples from ACDMPV patients. The observed deficits in semaphorin signaling common
189 to BPD+PH and ACDMPV suggest that the functional deficits in *FOXF1* associated with the
190 developmental phenotypes of BPD and BPD+PH may be mechanistically related to the
191 observed deficits in ACDMPV. Emerging technologies in spatial transcriptomic analysis offer
192 increasing potential for interrogation of archival FFPE tissue blocks, which we anticipate will
193 facilitate rapid growth in studies across the neonatal and pediatric populations. Making these
194 emergent data accessible and searchable could promote discovery of plausible mechanistic
195 targets for promoting lung growth and regeneration and allow for a larger representation of
196 samples across institutions and platforms.

197 While further work is required to define when abCaps emerge in disease and the role
198 they may play in the pathogenesis of BPD, the discovery and validation of the abCap cell-state
199 broadens our understanding of the endothelium as a receiver and coordinator of cell signaling

200 and cell movements during this specific developmental stage. Analysis across hundreds of
201 published datasets reveals that abCaps are distinct from the diversity of endothelial cell types
202 and states observed in other diseases (e.g. adult PH, COVID, ARDS), and suggests that abCaps
203 may be a cell state specific to BPD+PH. Deeper characterization of these cells demonstrates
204 increased expression of YAP signaling targets[20–22] (e.g., *CCN1, CCN2, SERPINE*1, and *CRIM1*,
205 along with abCap hallmark and YAP target gene *ANKRD1*). While the VEGF-induced
206 YAP/TAZ signaling responsiveness in endothelial cells is required for vasculogenesis[20-23],
207 elevated YAP/TAZ target expression in abCaps in BPD+PH patients could indicate an
208 attempted cell-autonomous compensatory response within abCaps themselves or arising as a
209 response to abnormal cell-signaling by injury-affected adjacent cells. Indeed, in murine models
210 of hyperoxia-induced injury in neonates and hypoxia-induced PH in adults, an increase in
211 endothelial CCN gene expression and *Ankrd1* expression was reported[24–26]. Understanding
212 how failed injury repair after preterm birth leads to severe BPD+PH, and how abCaps affect
213 and respond to the complex signaling milieu of the post-injury alveolar niche is critical to
214 untangling the mechanisms of abnormal repair in BPD and BPD+PH[27].

215 CellChat analysis predicted alveolar capillaries as receivers of signaling information
216 from adjacent alveolar epithelial and mesenchymal cells. In BPD and BPD+PH, and murine
217 hyperoxia BPD models, we noted dramatic reductions in expression of factors involved in
218 semaphorin signaling. The semaphorin pathway is well defined in neuronal axon movement
219 and synapse coordination, especially in the cellular targeting of growth cones toward target
220 cells[28]. While the role of this pathway in lung development has generally received little
221 attention, a global Sema3a inactivation resulted in high perinatal lethality with marked
222 impairment in alveologenesis in rare surviving KO pups[29], global knockdown of *Sema3C*
223 impaired alveologenesis[30], and ectopic *Sema3a* overexpression attenuates apoptosis in neonatal
224 rats exposed to hyperoxia[31]. Further, genetic inactivation of semaphorin receptor Nrp1
225 (neuropilin-1) in the adult mouse impairs airspace remodeling after smoke exposure[32] ,
226 suggesting a role for semaphorins in repair after injury. This study builds from these
227 foundational papers to dissect the cell-type specific expression of distinct semaphorins in the
228 setting of alveologenesis and injury-repair. Our findings on semaphorins add to previous
229 work identifying axonal-guidance signaling among several pathways differentially methylated
230 in BPD[33] as well as those made on another axon-guidance pathway (Netrin/DCC/Unc) in fine-
231 tuning the size and shape of emerging epithelial buds during early lung development[34],
232 although the latter study is of a far earlier developmental stages than the alveolar stage
233 reported here.

234 A central role of semaphorins is in regulating movement of cells toward or away from
235 each other. Recent 4D-live imaging experiments suggest that alveologenesis results from
236 precisely coordinated migration and shape change of epithelial and endothelial cells as they
237 move away from the alveolar duct to form a common functional unit of gas exchange[35] with
238 AT1 and aCap cells tightly aligned on a shared basement membrane, and our data strongly
239 implicate the semaphorin pathway as a regulator of these complex cellular movements and
240 shape changes required for alveolar formation. In BPD+PH, whether abCaps still positively
241 contribute to alveolar function or are the major contributor to deficit is unclear. Although
242 abCaps present with characteristics notably distinct from prior published cell populations,
243 they do share some transcriptional homology with endothelial tip cells, a specialized epithelial
244 cell that provides guidance for the growth and migration of capillaries during angiogenesis.
245 Expressing some but not all features of endothelial tip cells in evolving BPD+PH may be
246 indicative of compensatory response during repair or representative of a misaligned
247 developmental program in abCaps as a result of decreased cues from semaphorins and other
248 intercellular signals. Together, the decreased expression of *SEMA3B* by AT1 cells, *SEMA6A* by
249 aCaps and gCaps, *SEMA3C* in the mesenchyme and decreased *NRP1* expression in endothelial
250 cells under arrested lung development brings semaphorin signaling to the foreground in the
251 injured and normal contexts. A deeper characterization of the timing and cell specificity of
252 semaphorin signaling during lung development may reveal critical windows in which this
253 pathway could be targeted to promote regrowth and structural repair after injury.

254

255 **Methods**

256 *Human Infant Lung Repository*

257 At the time of autopsy/biopsy, fresh lung tissue was isolated and processed for single-
258 cell sequencing (scRNAseq) as described previously[5] with additional details below. In
259 addition, we studied formalin-fixed paraffin embedded tissue blocks from human infant lung
260 samples from autopsies and biopsies in varying stages of lung injury and without evidence of
261 lung injury (as determined by a clinical pathologist) have been collected at Vanderbilt
262 University Medical Center. All samples are de-identified with the exception of sex, gestational
263 age at birth, age at time of death, cause of death, and whether or not the patient had evidence
264 of BPD (defined as a need for supplemental oxygen at 36 weeks correct gestational age) or
265 pulmonary hypertension (as determined by clinical echocardiography). A table with the
266 clinical characteristics of the infant lungs used for FFPE tissue and scRNAseq is found in the

267 supplement (**Supplementary Tables 1-2**). This infant lung repository has been reviewed and
268 approved by the Vanderbilt University Institutional Review Board.

269 *Single-cell sequencing*

270 Human lung tissue from infants who died with BPD, BPD+PH, acute preterm lung
271 injury, and term infant controls were processed into single-cell suspension as described
272 previously[5]. scRNAseq libraries were generated using the 10x Chromium platform 3' or 5′
273 library preparation kits (10x Genomics) following the manufacturer's recommendations and
274 targeting 10,000-20,000 cells per sample. Next-generation sequencing was performed using an
275 Illumina Novaseq 6000. Reads with read quality less than 30 were filtered out and CellRanger
276 Count v7.1.0 (10x Genomics) was used to align reads onto the GRCh38 reference genome. Full
277 code used in the analysis of these data is available at
278 https://github.com/SucreLab/HumanNeonatal. A table with the clinical characteristics of
279 infants whose lungs were sequenced is in the supplement (**Supplementary Table 1**).

280 *Single-cell RNA sequencing analysis*

281 The ambient RNA was removed from the CellRanger count matrix h5 file outputs using
282 CellBender v0.2.1[36]. The scRNAseq data was analyzed using Seurat (v4.4.0)[37] in R (v4.3.2). The
283 quality control of the ambient RNA corrected cells filtered out cells with >10 % mitochondrial
284 mRNA mapped UMIs, <500 identified genes, and >7000 identified genes. The data was
285 normalized using SCTransform (v0.4.1)[38] with the glmGamPoi (v1.14.0)[39] method by stabilizing
286 the variance of the total molecular counts. The default parameters were used with SCTranform
287 except for: batchvar set to correct batch effects from the tissue samples processed,
288 vars.to.regress set to regress out the effects on differential gene expression analysis due to the
289 presence of increased mitochondrial mRNAs and ribosomal rRNAs, and ncells set to NULL.
290 The reduction of dimensionality was calculated via PCA. The PCA embeddings were further
291 integrated with Harmony (v1.2.0)[40] to account for any remaining batch effects from the
292 samples. The Harmony reduction was then applied in obtaining the UMAP embeddings and
293 the shared nearest neighbor graph which subsequently allowed for the Louvain algorithm to
294 find the cell clusters.

295 The clusters were split into the four main cell subtypes of epithelial, endothelial,
296 mesenchymal, and immune based on respective hallmark genes (**Supplementary Table 3**).
297 Each subtype went through an iterative process of SCTransform, PCA, Harmony, UMAP,
298 shared nearest neighbors, and clustering to isolate and remove any remaining doublet or
299 nonbiological cells. For each iteration, the cluster resolution would start in the range of 1 < n

300 <= 2, and with each iteration, n would be brought down slightly until the last iteration which
301 would have n = 1. Cell type labels were assigned after quality control and cleaning of the data
302 from their respective hallmark genes (**Supplementary Tables 3-4**). After annotation, the four
303 subtype-group Seurat objects were merged to produce the final Seurat object.

304 CellChat[16] was used to predict possible ligand-receptor pair interactions between cell
305 types of the same condition to generate hypotheses about possible signaling mechanisms
306 associated with BPD and BPD+PH. We employed SCimilarity[15] to query cell populations or
307 states that could have similar gene expression as the BPD+PH endothelial cell state (abCaps).
308 To do this, we used a gene signature list curated from the intersection of the top Wilcoxon
309 rank sum method ranked genes, top log2 fold change expression genes, and genes with the
310 highest proportion expressed within abCaps versus all other cell types: *ANKRD1, CCN2,*
311 *KITLG, SERPINE1, TNFRSF12A, CRIM1, GDF15, CCND1, EDN1, RBFOX2, CD151, PLS3, CD59,*
312 *GBE1, PODXL*. We then applied this gene module score to all cell populations or clusters in
313 each study independently, with a cutoff of a cell score greater than two standard deviations
314 from the mean cell score in the dataset. In summary, we looked for cells with features of the
315 expression pattern identified in the abCap gene module. We then plotted the percentage of
316 cells within each cluster that passed this cutoff.

317 Reanalysis of the recently published single-nucleus RNA sequencing dataset of infants
318 who died of ACDMPV and term control infants was performed using the same strategy used
319 above for scRNAseq.

320 *Murine model of neonatal lung injury*

321 C57BL/6 mice were exposed to 70% $O_2$ (hyperoxia) or 21% $O_2$ normoxia conditions
322 from P1 until 5 days of age (P5), with hyperoxia animals receiving intranasal LPS at 1 $\mu$g/g on
323 P3 and P4. On P7 and P14, lungs were inflation-fixed by gravity filling with 10% buffered
324 formalin and paraffin embedded as described previously[41]. This protocol was approved by the
325 Institutional Animal Care and Use Committee of Vanderbilt University and was in compliance
326 with the Public Health Services policy on humane care and use of laboratory animals.

327 *RNA in situ hybridization*

328 RNAScope technology (ACDBio) was used to perform all RNA ISH experiments
329 according to the manufacturer's instructions as described previously . RNAScope probes to the
330 following human genes were used (*SEMA3B, AGER, SFTPC, SEMA6A, PECAM1).* RNAScope
331 probes to the following mouse genes were used*: Sftpc, Hopx, Pecam1, Sema3b, Sema6a.* Positive

control probe (PPIB) and negative control probe (DapB) were purchased from ACDBio and performed with each experimental replicate. When RNAScope and IF are performed together, this was done in accordance with the manufacturer's instructions.

*Immunofluorescence*

Immunofluorescence was performed as described previously , with the following primary antibodies used: anti-PDPN (DSHB, Iowa), anti-FOXF1 (1:100, R&D systems AF4798), anti-ERG1 (1:100, Abcam ab214341) and the following secondary antibodies used: 1:200 Donkey anti-Goat (Invitrogen A21447), 1:200 Donkey anti-Mouse (Invitrogen A31570).

*Image acquisition and analysis*

We acquired fluorescent and brightfield images using a Keyence BZ-X810 with BZ-X Viewer software and 40× objective. Images were collected with the following filter sets: 405, 488, 561 and 647 nm. Automated image analysis was performed with HALO software (Indica Labs) as described previously. After HALO analysis, the rate of false co-localization of non-physiological expression combinations (*Sftpc+Pecam1+*) was low, with a mean of <1%.

*Lung Morphometry*

Mean linear intercept was performed on a minimum of 10 40x images per mouse (n=6 mice per group) as described previously [42], with the use of ImageJ and AlveolEye for automated calculation of MLI measurements (https://github.com/SucreLab/AlveolEye).

*Statistics*

Statistical analysis was performed in R and with Prism Graphpad version 10.0.03, with specific details on statistical tests provided in Fig. legends.

**Data Availability**

The single-cell sequencing data generated for this manuscript are available as a freely searchable tool at www.sucrelab.org/lungcells. As noted above, all code used in analysis and generation of sequencing data display can be found at https://github.com/SucreLab/HumanNeonatal. Raw sequencing data will be deposited at GEO and available for download. We will eagerly provide processed data and objects upon request.

**Acknowledgements**

This work was supported by the Francis Family Foundation (JAK, DBF, JMSS, NMN), R01HL168556 (JMSS), K08HL143051 (JMSS) R01HL145372(JAK/NEB), R56HL167937 (DBF), U01HL175444 (JAK, NEB, JMSS). We are grateful to Anne Hilgendorff and Vivian Siegel for thoughtful discussions and insights.

**Author Contribution**

Conceptualization: JAK, JMSS, DBF, CVEW, Methodology: SPS, JAK, NMN, CSJ, ALS, Validation: SG, GF, DW, Resources: MEK, DBF, Investigation and Interpretation: CSJ, JAK, JMSS, SPS, NMN, BDF, CVEW, SM, NEB Manuscript preparation (original): SPS, JMSS, Manscript preparation (revision): SPS, NMN, CSJ, ALS, SG, MEK, GF, SM, NEB< CVEW, DBF, JAK, JMSS.

## REFERENCES


1. Baraldi, E. & Filippone, M. Chronic lung disease after premature birth. *N Engl J Med* **357**, 1946–55 (2007).

2. Berkelhamer, S. K., Mestan, K. K. & Steinhorn, R. H. Pulmonary hypertension in bronchopulmonary dysplasia. *Semin Perinatol* **37**, 124–31 (2013).

3. Frank, D. B. *et al.* Primary pulmonary vein stenosis during infancy: state of the art review. *J. Perinatol. Off. J. Calif. Perinat. Assoc.* **41**, 1528–1539 (2021).

4. Adams, T. S. *et al.* Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv* **6**, eaba1983 (2020).

5. Habermann, A. C. *et al.* Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci. Adv.* **6**, eaba1972 (2020).

6. Schupp, J. C. *et al.* Integrated Single-Cell Atlas of Endothelial Cells of the Human Lung. *Circulation* **144**, 286–302 (2021).

7. He, P. *et al.* A human fetal lung cell atlas uncovers proximal-distal gradients of differentiation and key regulators of epithelial fates. *Cell* **185**, 4841-4860.e25 (2022).

8. Guo, M. *et al.* Single cell RNA analysis identifies cellular heterogeneity and adaptive responses of the lung at birth. *Nat Commun* **10**, 37 (2019).

9. Sikkema, L. *et al.* An integrated cell atlas of the lung in health and disease. *Nat. Med.* **29**, 1563–1577 (2023).

10. Duong, T. E. *et al.* A single-cell regulatory map of postnatal lung alveologenesis in humans and mice. *Cell Genomics* **2**, 100108 (2022).

11. Guo, M. *et al.* Single Cell Multiomics Identifies Cells and Genetic Networks Underlying Alveolar Capillary Dysplasia. *Am. J. Respir. Crit. Care Med.* **208**, 709–725 (2023).

12. Tsukui, T. *et al.* Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis. *Nat. Commun.* **11**, 1920 (2020).

13. Chen, W. *et al.* The endothelial tip-stalk cell selection and shuffling during angiogenesis. *J. Cell Commun. Signal.* **13**, 291–301 (2019).

14. del Toro, R. *et al.* Identification and functional analysis of endothelial tip cell–enriched genes. *Blood* **116**, 4025–4033 (2010).

15. Heimberg, G. *et al.* Scalable querying of human cell atlases via a foundational model reveals commonalities across fibrosis-associated macrophages. 2023.07.18.549537 Preprint at https://doi.org/10.1101/2023.07.18.549537 (2023).

16. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).

17. Hilgendorff, A., Reiss, I., Ehrhardt, H., Eickelberg, O. & Alvira, C. M. Chronic lung disease in the preterm infant. Lessons learned from animal models. *Am J Respir Cell Mol Biol* **50**, 233–45 (2014).

18. Karolak, J. A. *et al.* Perturbation of semaphorin and VEGF signaling in ACDMPV lungs due to FOXF1 deficiency. *Respir. Res.* **22**, 212 (2021).

19. Abman, S. H. & Lakshminrusimha, S. Pulmonary Hypertension in Established Bronchopulmonary Dysplasia: Physiologic Approaches to Clinical Care. *Clin. Perinatol.* **51**, 195–216 (2024).

20. Boopathy, G. T. K. & Hong, W. Role of Hippo Pathway-YAP/TAZ Signaling in Angiogenesis. *Front. Cell Dev. Biol.* **7**, (2019).

21. Isago, H. *et al.* The Epithelial Expressions of YAP and TAZ are Sequentially Required in Lung Development. *Am J Respir Cell Mol Biol* (2019) doi:10.1165/rcmb.2019-0218OC.

22. Hooglugt, A., van der Stoel, M. M., Boon, R. A. & Huveneers, S. Endothelial YAP/TAZ Signaling in Angiogenesis and Tumor Vasculature. *Front. Oncol.* **10**, 612802 (2021).

23. Wang, X. *et al.* YAP/TAZ Orchestrate VEGF Signaling during Developmental Angiogenesis. *Dev. Cell* **42**, 462-478.e7 (2017).

24. Lingappan, K. *et al.* Analysis of the Transcriptome in Hyperoxic Lung Injury and Sex-Specific Alterations in Gene Expression. *PLOS ONE* **9**, e101581 (2014).

25. Perkowski, S. *et al.* Gene expression profiling of the early pulmonary response to hyperoxia in mice. *Am. J. Respir. Cell Mol. Biol.* **28**, 682–696 (2003).

26. Lee, S. *et al.* CCN1 Suppresses Pulmonary Vascular Smooth Muscle Contraction in Response to Hypoxia. *Pulm. Circ.* **5**, 716–722 (2015).

27. Vila Ellis, L. & Chen, J. A cell-centric view of lung alveologenesis. *Dev. Dyn.* **250**, 482–496 (2021).

28. Jackson, R. E. & Eickholt, B. J. Semaphorin signalling. *Curr. Biol.* **19**, R504–R507 (2009).

29. Becker, P. M. *et al.* Semaphorin 3A contributes to distal pulmonary epithelial cell differentiation and lung morphogenesis. *PloS One* **6**, e27449 (2011).

30. Vadivel, A. *et al.* The Axonal Guidance Cue Semaphorin 3C Contributes to Alveolar Growth and Repair. *PLoS ONE* **8**, e67225 (2013).

31. Liang, Z., Zhang, X., Liu, Y., Wu, Q. & You, C. SEMA3A protects against hyperoxia-induced lung injury in a bronchopulmonary dysplasia model of newborn rat by inhibiting ERK pathway. *Allergol. Immunopathol. (Madr.)* **49**, 8–15 (2021).

32. Le, A. *et al.* Pulmonary Epithelial Neuropilin-1 Deletion Enhances Development of Cigarette Smoke–induced Emphysema. *Am. J. Respir. Crit. Care Med.* **180**, 396–406 (2009).

33. Cuna, A. *et al.* Alterations in gene expression and DNA methylation during murine and human lung alveolar septation. *Am J Respir Cell Mol Biol* **53**, 60–73 (2015).

34. Liu, Y. *et al.* Novel role for Netrins in regulating epithelial behavior during lung branching morphogenesis. *Curr. Biol. CB* **14**, 897–905 (2004).

35. Negretti, N. M. *et al. Epithelial Outgrowth Through Mesenchymal Rings Drives Alveologenesis*. http://biorxiv.org/lookup/doi/10.1101/2022.10.06.511160 (2022) doi:10.1101/2022.10.06.511160.

36. Fleming, S. J. *et al.* Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nat. Methods* **20**, 1323–1335 (2023).

37. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e29 (2021).

38. Choudhary, S. & Satija, R. Comparison and evaluation of statistical error models for scRNA-seq. *Genome Biol.* **23**, 27 (2022).

39. Ahlmann-Eltze, C. & Huber, W. glmGamPoi: fitting Gamma-Poisson generalized linear models on single cell count data. *Bioinformatics* **36**, 5701–5702 (2021).

40. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

41. Sucre, J. M. S. *et al.* Hyperoxia Injury in the Developing Lung Is Mediated by Mesenchymal Expression of Wnt5A. *Am J Respir Crit Care Med* **201**, 1249–1262 (2020).

42. Schuler, B. A. *et al.* Age-determined expression of priming protease TMPRSS2 and localization of SARS-CoV-2 in lung epithelium. *J Clin Invest* (2020) doi:10.1172/JCI140766.

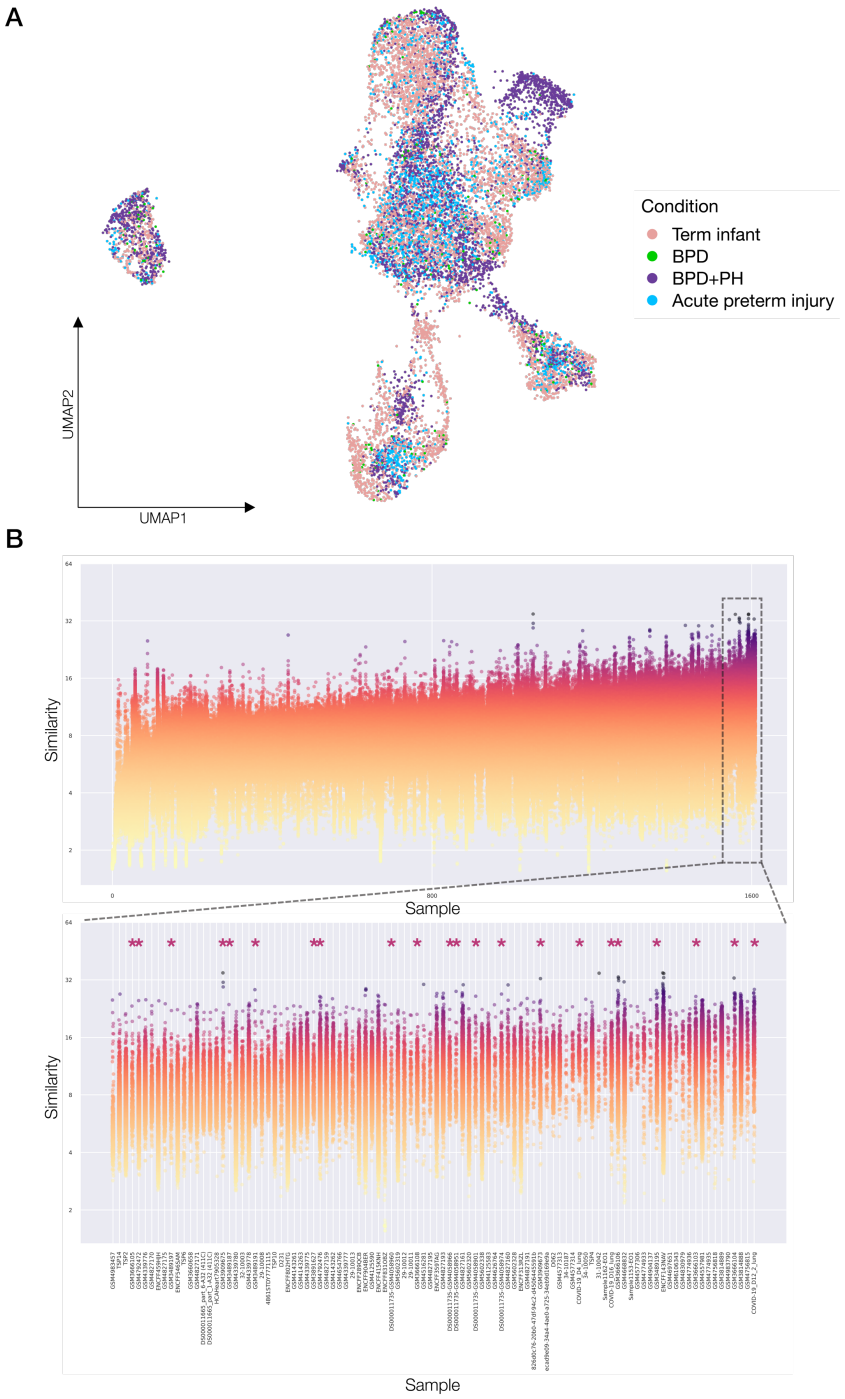
43. Sucre, J. M. S. *et al.* A Shared Pattern of beta-Catenin Activation in Bronchopulmonary Dysplasia and Idiopathic Pulmonary Fibrosis. *Am J Pathol* **188**, 853–862 (2018).

| Gestational Age | Age at time of sample collection | Autopsy/Biopsy | Clinical Metadata |
|---|---|---|---|
| 40 wks | 7 days | Biopsy | Normal tissue from infant s/p CPAM resection |
| 37 wks | 20 days | Biopsy | Normal tissue from infant s/p CPAM resection |
| 27 wks | 7 months (52 wks CGA) | Autopsy | BPD + PH |
| 26 wks | 7 months (51 wks CGA) | Autopsy | BPD + PH |
| 27 wks | 7 months (52 wks CGA) | Autopsy | BPD |
| 25 wks | 7 months (50 wks CGA) | Autopsy | BPD |
| 26 wks | 2 wks (28 wks CGA) | Autopsy | Intubated 14 days; Died with severe IVH |

**Supplementary Table 1:** Clinical characteristics of infants whose lung tissue samples were analyzed by single-cell RNA sequencing, CGA = corrected gestational age

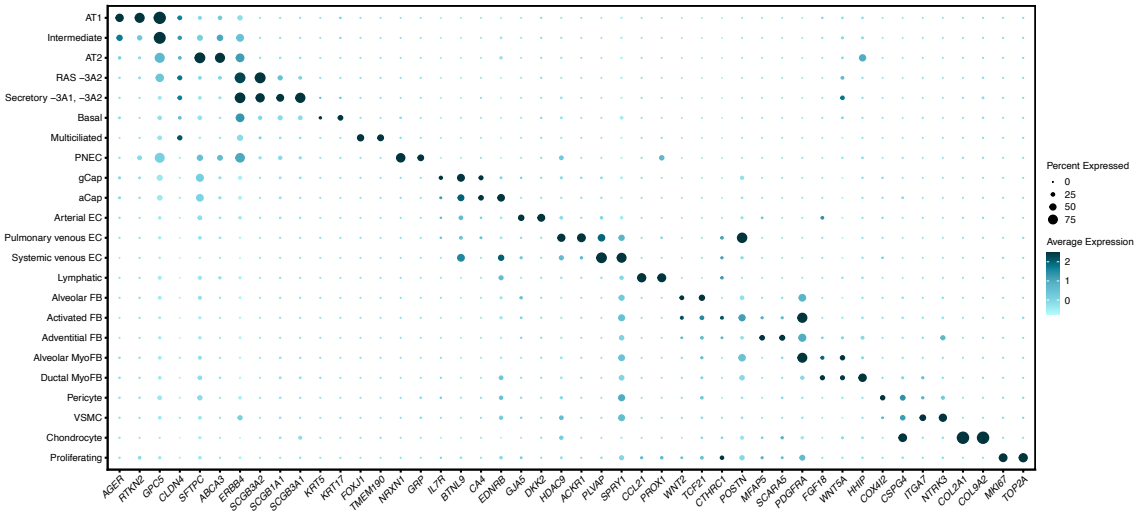
| Gestational Age | Age at time of sample collection | Autopsy/Biopsy | Clinical Metadata |
|---|---|---|---|
| 35 wks | 2 days | Autopsy | HIE |
| 39 wks | 3 days | Autopsy | HIE |
| 39 wks | 1 day | Autopsy | Stillbirth |
| 24 wks | 4 months | Autopsy | BPD |
| 26 wks | 6 months | Autopsy | BPD |
| 27 wks | 7 months | Autopsy | BPD + PH |
| 26 wks | 7 months | Autopsy | BPD + PH |
| 27 wks | 4 months | Autopsy | BPD + PH |

Supplementary Table 2: Clinical characteristics of infants with lung tissue samples in blocks that were studied by RNA *in situ* hybridization and immunofluorescence




Supplemental Figure 1:

A) UMAP embedding of the neonatal human lung endothelial cells labeled by disease-state condition. B) 1,610 identified samples containing endothelial cells filtered by biological factors and by top 100 SCimilarity query score compared to the expression profile of abCaps, identified 22 lung sequencing samples contained within 6 studies of the most comparable cell populations.




## Supplemental Figure 2:

Dot plot of the hallmark-identifying genes for each cell type cluster of the human infant ACDMPV snRNAseq reanalysis dataset. Dot size corresponds to the percentage of cells within a cluster expressing a certain gene. Color intensity corresponds to the relative average expression level.

| Lineage | Cell State | Hallmark Genes |
|---|---|---|
| **Epithelial** | AT1 | *AGER, RTKN2, LAMA3, COL4A3, PDPN* |
| | AT2 | *SFTPC, ABCA3, NAPSA, SFTPD, LAMP3* |
| | RASC | *SCGB3A2,* (neg) *SCGB3A1,* (neg) *SCGB1A1, SGMS2, ERBB4, CDK14* |
| | Secretory -3A1, -3A2 | *SCGB3A2, SCGB3A1, SCGB1A1, (neg) MUC5B, KLK11* |
| | Secretory MUC5B | (neg) *SCGB3A2, SCGB3A1, SCGB1A1, MUC5B, BPIFB1, MSMB, CYP2F1* |
| | Basal | *KRT17, KRT15, KRT5, MIR205HG, TP63* |
| | Multiciliated | *FOXJ1, CAPS, DNAH12, RSPH1, CFAP299, LRRIQ1, SNTN, TMEM190, ERICH3, HYDIN* |
| **Endothelial** | aCap | *EDNRB, SOSTDC1, HPGD, EMCN, TSPAN12, S100A3, APLN* |
| | gCap | *GPIHBP1, FCN3, BTNL9, NTRK2, BMP6, IL7R* |
| | abCap | *ANKRD1, CCN1, CCN2, SERPINE1, KITLG, EDN1, TNFRSF12A, GDF15, APLN* |
| | Arterial EC | *DKK2, GJA5, CXCL12, THSD7A, SERPINE2, BMX, TSPAN2, EGFL8* |
| | Pulmonary venous EC | *ACKR1, HDAC9, CPE, PTGIS, C7, SELP, FAM155A* |
| | Systemic venous EC | *ACKR1, MCTP1, CCL14, PLVAP, ADGRL4, SPRY1, COL15A1* |

| | | |
|---|---|---|
| | Lymphatic | *CCL21, MMRN1, TFF3, PKHD1L1, PROX1, RELN, TBX1* |
| **Mesenchymal** | Alveolar FB | *TCF21, NPNT, INMT, WNT2, BMP5, FAT3, MEOX2, NTNG1, RSPO2, FGFR4* |
| | Activated FB | *CTHRC1, TIMP1, POSTN* |
| | Adventitial FB | *SERPINF1, PI16, MFAP5, SCARA5, CCDC80, FBLN1, C1R, C3, DCLK1, PRRX1, IGFBP6, GSN, PDGFRL* |
| | Alveolar MyoFB | *WNT5A, FGF18, PDGFRA, CXCL14, ASPN, DACH2, LTBP2, CCDC68, COL24A1, LRP4, ANO4* |
| | Ductal MyoFB | *WNT5A, FGF18, HHIP, ZNF385D, HPSE2, MYH11, ACTA2, DOCK3, GREM2, DES, PRUNE2, SCUBE1, LGR6* |
| | Pericyte | *COX4I2, CSPG4, LAMC3, TRPC6, HIGD1B, NDUFAL2, FAM162B, KCNK3, AFF2, ADARB2, LRRTM4* |
| | VSMC | *ACTA2, TAGLN, TPM2, NTRK3, ADIRF, MUSTN1, RCAN2, RGS6, CSMD1, PLN, ECRG4* |
| **Immune** | T Cell | *CD8A, THEMIS, CD6, CAMK4, BCL11B, IL7R, CD2, LEF1, ICOS, CD3D* |
| | NK Cell | *NKG7, CCL4, PRF1, GNLY, KLRD1, KLRB1, KLRF1, SPON2, FGFBP2, MYOM2, LINGO2* |
| | NKT Cell | *KLRB1, CD7, CTSW, TNFRSF18, CCL5, IL7R, GZMK* |

| | |
|---|---|
| Alveolar Macrophage | *C1QA, C1QB, C1QC, APOC1, ACP5, CAPG, MARCO, LYZ, APOE* |
| Monocyte | *S100A12, FCN1, VCAN, KYNU, CD14* |
| pDC | *RHEX, IRF4, PPP1R14B, PTPRS, CCDC50, RUBCN, IRF4, IRF7, PLD4, LINC01478* |
| cDC | *CSF2RA, CPVL, CD86, FCER1A, CD1C, FCGR2B, CLEC10A, HLA-DRA, FPR3* |
| B Cell | *BANK1, MS4A1, IGHM, CD79A, CD79B, EBF1, VPREB3, FCRL1, PAX5, IGHD, COL19A1, TNFRSF13C, CD22* |
| Mast cell | *CPA3, KIT, TPSD1, MS4A2, TPSB2, TPSAB1, HPGDS, CTSG, CADPS* |
| Plasma cell | *MZB1, JCHAIN, SSR4, TXNDC5, SEC11C, FKBP11, PRDX4, IGHG1, IGHG3, POU2AF1* |
| Basophil | *ATP10D, LINC02458, HDC, HSH2D, NLK, INPP5A, ENPP3, TAFA2, TRPM6, IL4* |
| Neutrophil | *S100A8, S100A9, CSF3R, NAMPT, AQP9, LIMK2, FBXL5, LUCAT1, FCGR3B, KCNJ15* |

* These genes were selected from prior published datasets and studies, as well as discovered within this study

Supplementary Table 3: Select hallmark genes of the 33 identified cell type populations