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 Ingleheim 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@vanderbilt.edu

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 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 allowed for 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, with term infants at approximately the same corrected gestational age as the preterm infants for comparison.

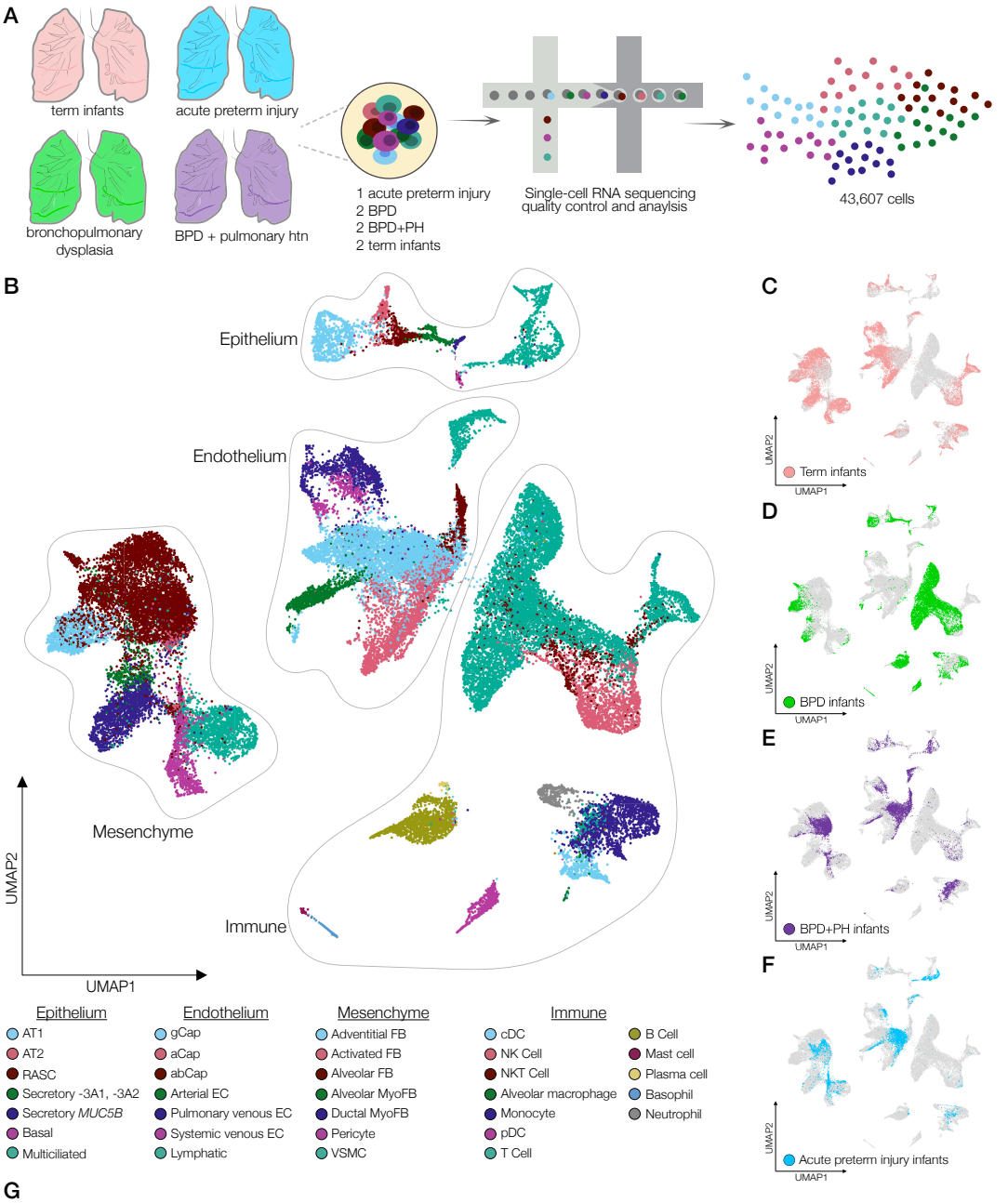
While we anticipate that this transcriptional dataset will be a resource for future work in neonatal lung injury and BPD in general, we were especially curious about differences in the endothelial cell populations in infants with BPD+PH and how they might differ from corrected gestational age-matched controls and patients with BPD at a similar age. In analysis of these sequencing data, we identify a capillary endothelial cell-state highly specific to patients with BPD+PH and spatially localize these capillary cells to the alveolar parenchyma. Further, we employ predictive signaling analysis to generate hypotheses about the role of semaphorin guidance-cue signaling during alveolar formation, a pathway we identify as deficient in both BPD and BPD+PH patients transcriptomically and spatially in alveolar regions of lung tissue samples. The 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 connections in 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.

Results

To examine the cellular dynamics and candidate mediators of neonatal lung injury and the development of severe BPD, we used the 10x Genomics Chromium Platform to perform single-cell RNA sequencing of human neonatal lungs at varying stages of lung injury and chronic lung disease: 1 acute preterm infant with lung injury, 2 BPD samples at term corrected gestational age (CGA) , 2 BPD+PH at term CGA, and 2 term infants in the first 3 weeks of life for age-matched comparison (**Figure 1A**, **Supplementary Table 1**). After filtering and quality control to remove doublets and ambient RNA (see materials and methods), we recovered 43,607 single-cell transcriptomes with a range of 2,199 to 8,339 cells per subject, a median of 4,128 unique molecular identifiers (UMI) (interquartile range 1,188 – 4,670) per cell, and a median of 1,795 genes (interquartile range 728 – 1,944) per cell (**Figure 1A-B**). Unbiased clustering with cell-type annotation informed by prior human, fetal, and neonatal studies identified 7 epithelial, 7 endothelial, 7 mesenchymal, and 12 immune clusters (**Figure 1B-G**)[4–12]. To specifically focus on potential vascular cell differences in BPD patients with and without PH, the endothelial cluster was extracted, re-embedded, and separately analyzed to identify cell-type specific differences across homeostasis and disease. All the sequencing data are available as an interactive data explorer at www.sucrelab.org/lungcells.

Figure 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 2 term control infants, 1 acute neonatal lung injury, 2 infants with severe bronchopulmonary dysplasia, and 2 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) (**Figure 2A-B**, **S1A**). This group of 647 cells is marked by high expression of *ANKRD1*, which has almost no expression in any endothelial cell cluster, with rare expression in some alveolar type 1 cells (**Figure 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 lacks expression of canonical gCap genes *FCN3*, *IL7R*, and *BTNL9* (**Figure 2D**). The abCaps also express certain pan-capillary markers *PDGFB*, *RGCC*, and *ADGRL2*, but not *CA4*, *CD36*, and *AFF3*. In addition to YAP/TAZ signaling target *ANKRD1*, this population expresses relatively high levels of YAP targets *CCN1* (*CYR61*), *CCN2* (*CTGF*), *SERPINE1*, and *CRIM1*, when compared to other endothelial cells (**Figure 2D**). Notably, the abCaps also present a gene expression profile similar to previously described markers of endothelial tip cells[13,14] with increased expression of *ESM1*, *ANGPT2*, *APLN*, *MCAM*, *COTL1*, relative to all other endothelial cells, but lower gene expression for tip cell-associated receptors such as *KDR* (*VEGFR2*), *FLT4* (*VEGFR3*), *UNC5B*, and *NRP1* (**Figure 2E**).

To spatially validate and localize the abCap cell-state, we performed RNA *in situ* hybridization using formalin-fixed paraffin-embedded (FFPE) lung tissue blocks preterm infants with BPD, with and without PH as well as from term infant control lungs (**Figure 2F-H**). These FFPE tissues are from an expanded set of samples, independent of the samples interrogated by scRNAseq (**Supplementary Table 2**). In lung tissue from patients with BPD+PH, *ANKRD1+/PECAM1+* endothelial cells 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 **(Figure 2I**).

This endothelial cell state identified in patients with BPD and PH does not transcriptionally resemble endothelial cells previously described in different human lung diseases

To understand abCaps in the context of previously reported atlases of human pulmonary endothelium under normal and disease conditions, we performed analysis with SCimilarity[15], a computational tool that allows for comparison of a given cell population with cell populations from nearly 400 previously published human datasets, including 41 from the lung. Comparing abCaps across the 1.12M endothelial cells in the 22.7M cell corpus of SCimilarity, we selected the top 100 human samples (contained in 34 source studies) with the highest SCimilarity score to our BPD+PH capillary cells. Of these 34 studies, we excluded 7 studies of iPSCs, 14 that did not contain lung tissue, 2 with samples in the 7th – 8th decade, 2 with a sample size of < 2, and 2 that had inaccessible scRNAseq data files, leaving 6 studies with 178 samples for comparison (**Figure S1B**).

With these 6 studies, we compared the BPD+PH capillary endothelial cell state (abCaps) with the endothelial cells defined in healthy controls and in the 5 lung disease states described by these studies (COVID, LAM, IPF, SSc-ILD, and PAH). Using the genes most-specific for abCaps, we created a cell module score to identify cells in these datasets with expression patterns that were characteristic of abCaps (see methods for details). Examining the endothelial cells in these 6 studies, fewer than 12% of cells in each endothelial sub-population were similar to the characteristic gene expression of the BPD+PH associated capillary cells **(Figure 2J).** As a positive control, applying this same metric to our own data identified the genes with the highest cell module score were contained within the BPD+PH associated abCap population.

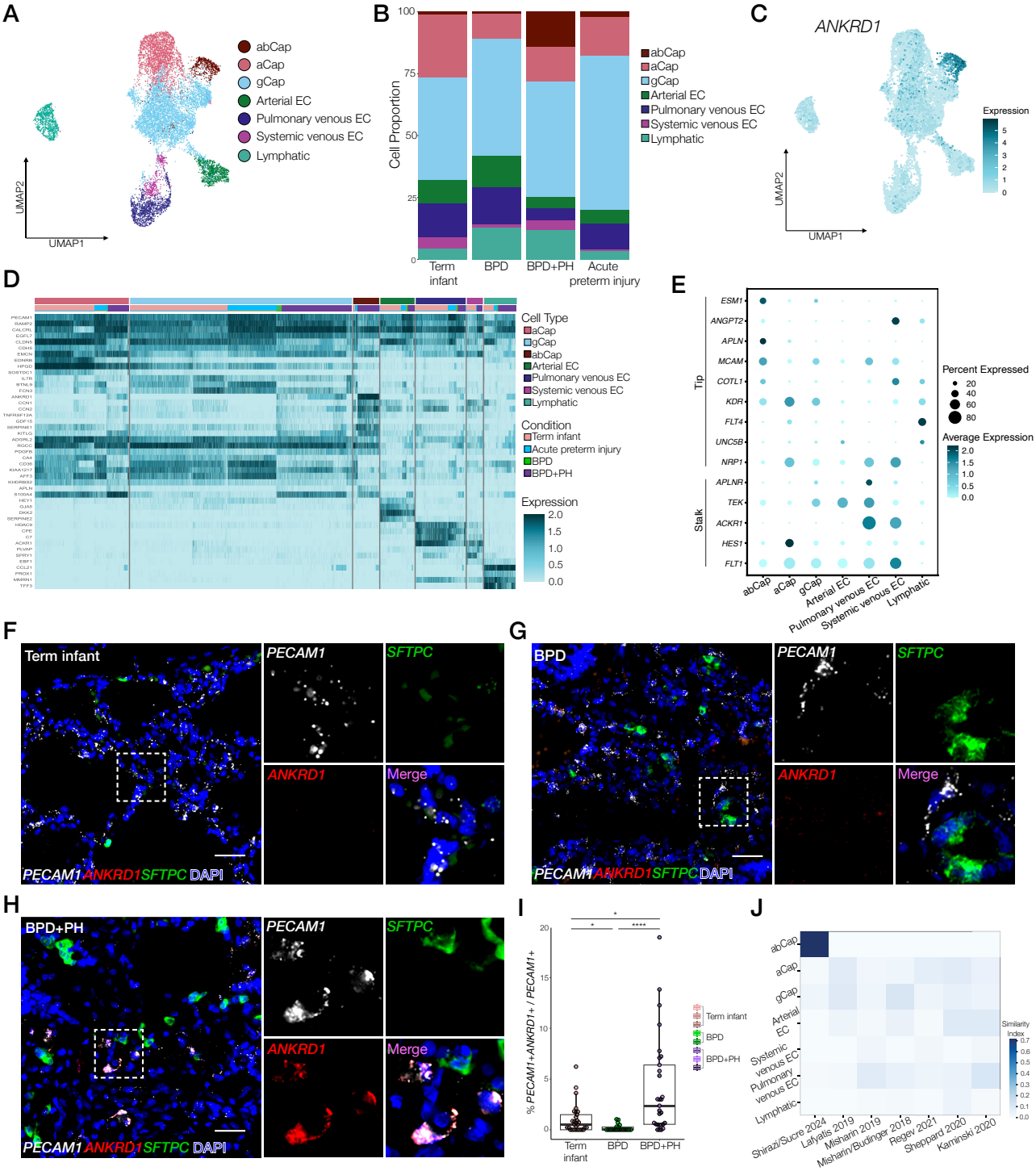


Figure 2: Analysis of the endothelium reveals a cluster of aberrant capillary cells (abCaps) enriched in patients with BPD+PH

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 SCimiliary, with intensity of color indicating greater similarity, abCap cell included in upper right at positive control.

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

To characterize how this specialized population of endothelial cells was interacting with other cells in the alveolar niche, we interrogated our sequencing data with CellChat, which predicts signaling interactions by expression of ligand-receptor pairs. In intracellular communication, cells can be classified as having more outgoing or incoming signaling information[16]. Broadly, CellChat identified that the endothelial capillaries act predominantly as "receivers" of cell-signaling inputs, with AT1 cells and mesenchymal cells having a higher "outgoing" score (**Figure 3A**). Examining predicted signaling pathways, CellChat identified semaphorin signaling as a leading differentially expressed candidate pathway differentiating term controls and infants with BDP and BPD+PH, with significantly decreased interactions in both epithelial-capillary and mesenchymal-capillary interactions (**Figure 3B**). Interrogation of the transcriptomic dataset for the specific ligands and receptors of this pathway revealed that patients with BPD+PH, had markedly decreased expression of *SEMA3B* by AT1 cells, *SEMA3C* by alveolar and ductal myofibroblasts, and *SEMA6A* by endothelial capillary cells, with some decreased expression of these ligands in patients with severe BPD without PH, compared to controls (**Figure 3C**). Expression of semaphorin receptors *NRP1* and *PLXNA2* was also decreased (**Figure 3C**) in BPD and BPD+PH patients. Some of these predicted expression changes were spatially validated by RNA ISH, with decreased expression of *SEMA3B* in *AGER*-expressing AT1 cells and decreased *SEMA6A* in *PECAM1+* alveolar endothelial cells (**Figure 3D-E**).

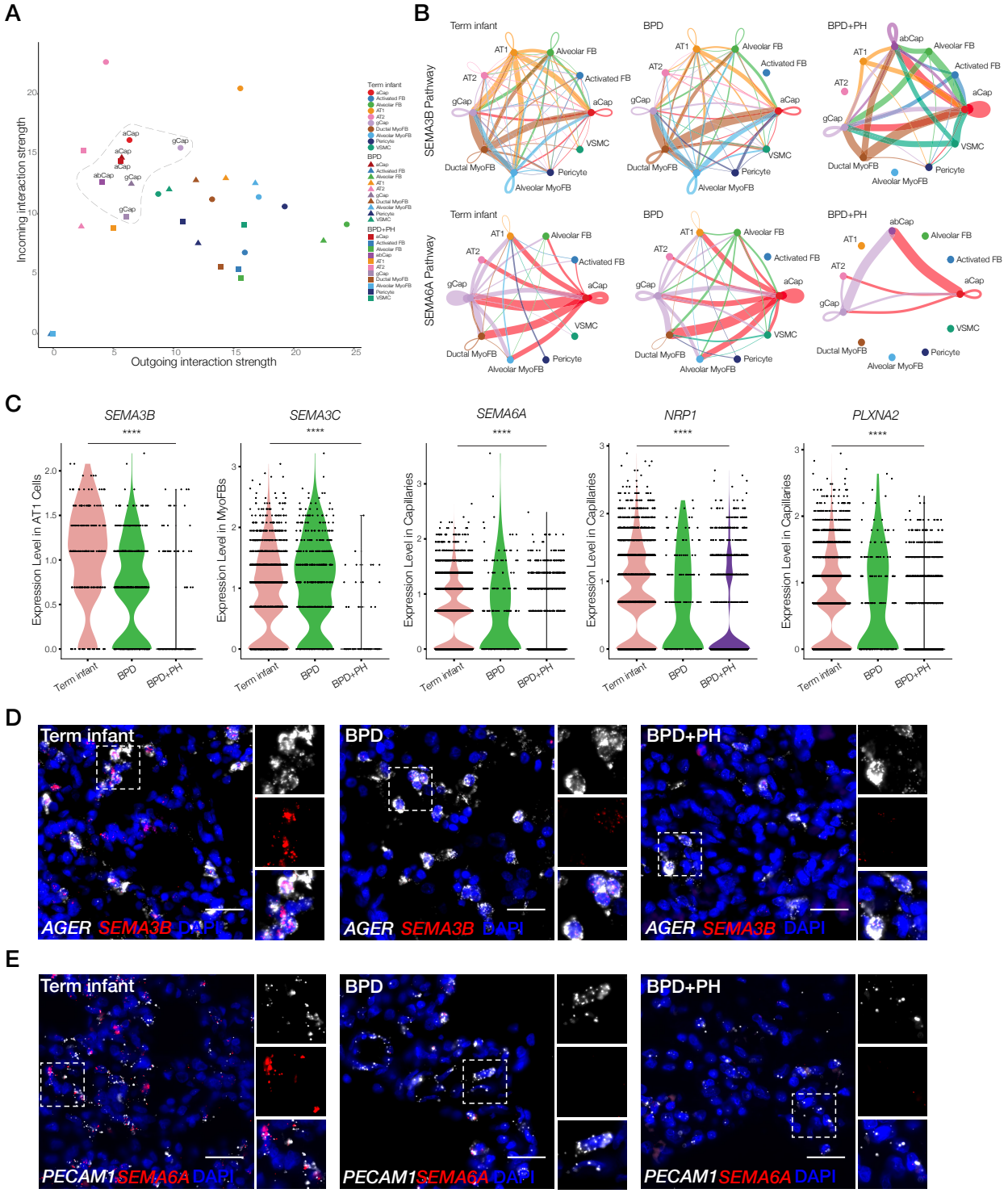


Figure 3: Analysis of predicted ligand-receptor interactions by single-cell sequencing identifies decreased semaphorin signaling between endothelial cells and alveolar niche in the setting of BPD and BPD + PH

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. ****adjusted 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 *PECAM*+ endothelial cells in term controls, with decreased expression in patients with BPD and BPD+PH. Scale bar = 25 μm.

Decreased semaphorin ligand expression replicated in murine model of BPD

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

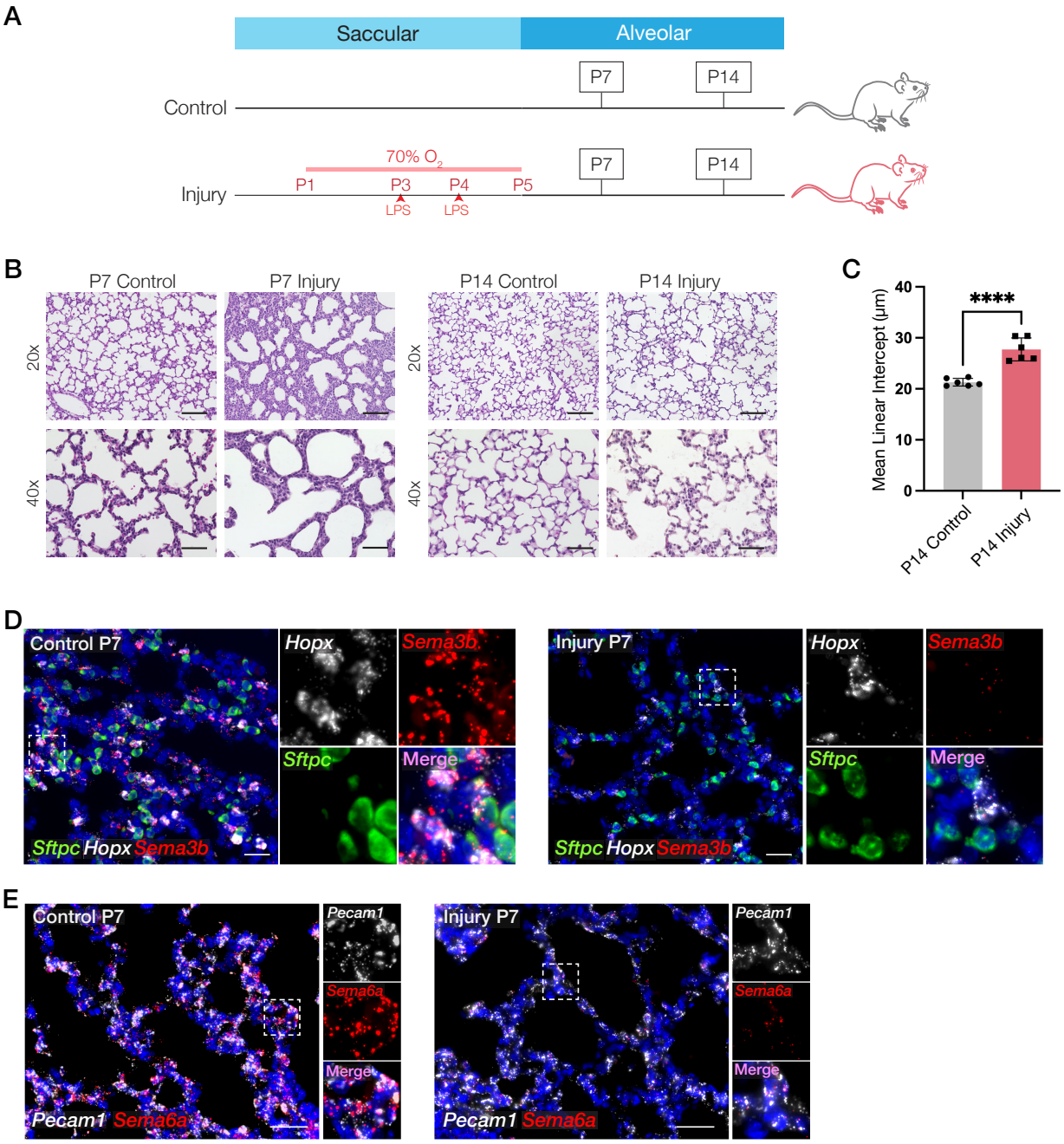


Figure 4: Murine mouse injury models of BPD and BPD+PH display decreased semaphorin signaling

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, ****adjusted *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 transcription factor FOXF1[18], 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 function and semaphorin signaling, we interrogated a recently published single-nucleus RNA sequencing dataset from infants and children with ACDMPV and age-matched controls[11] (**Figure 5A, S2**) for expression of ligands and receptors in the semaphorin signaling pathway. While the expression deficits in semaphorin signaling are not identical in ACDMPV and BPD+PH, (**Figure 5B**), we note that patients with ACDMPV have decreased *SEMA3B* expression in AT1 cells and decreased *SEMA6A* signaling in alveolar capillaries (aCaps + gCaps), similar to deficits in BPD+PH.

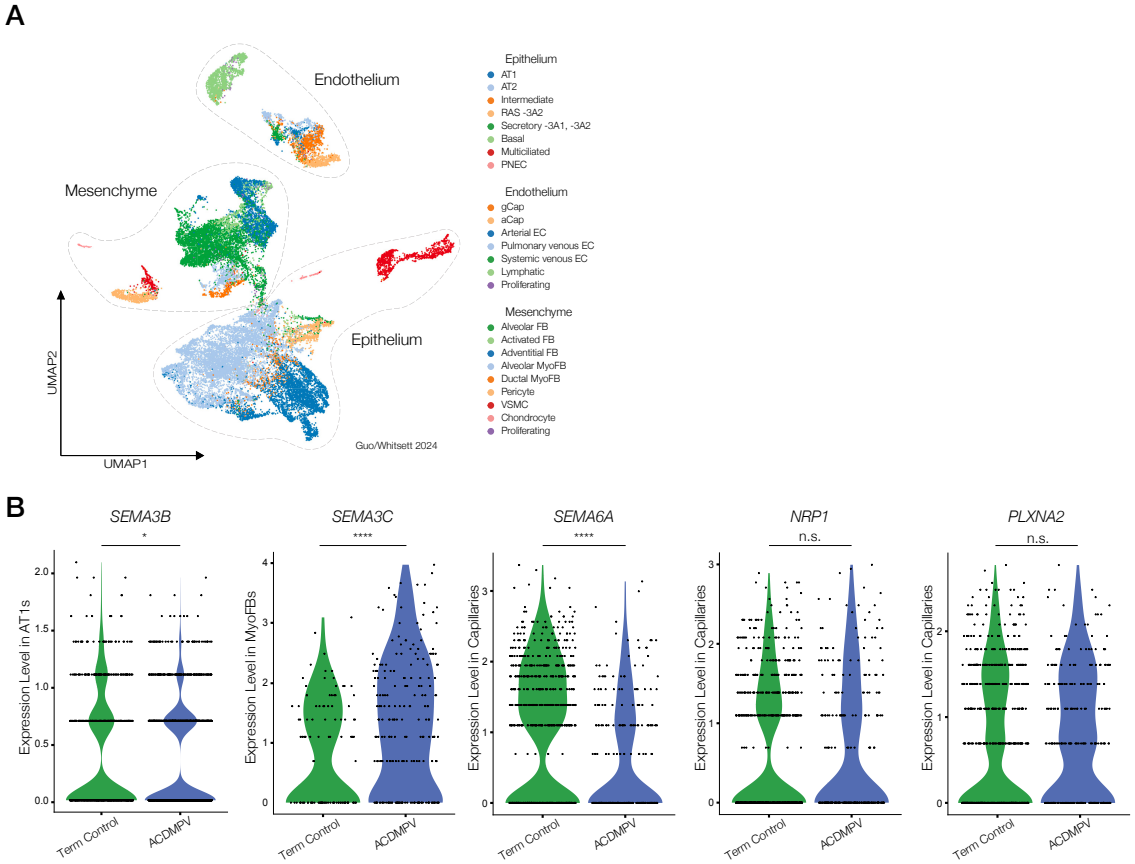


Figure 5: Analysis of human infant ACDMPV dataset identifies decreased semaphorin signaling
A) UMAP embedding of single-nucleus RNA sequencing of five ACDMPV neonatal lung samples[11], three preterm, and three term control lung samples with an age range of 29 weeks gestational age to 3.5 years old annotated 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.

Lungs from infants with BPD and BPD+PH demonstrate decreased expression in FOXF1, a known mediator of angiogenesis and semaphorin signaling in the developing lung

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

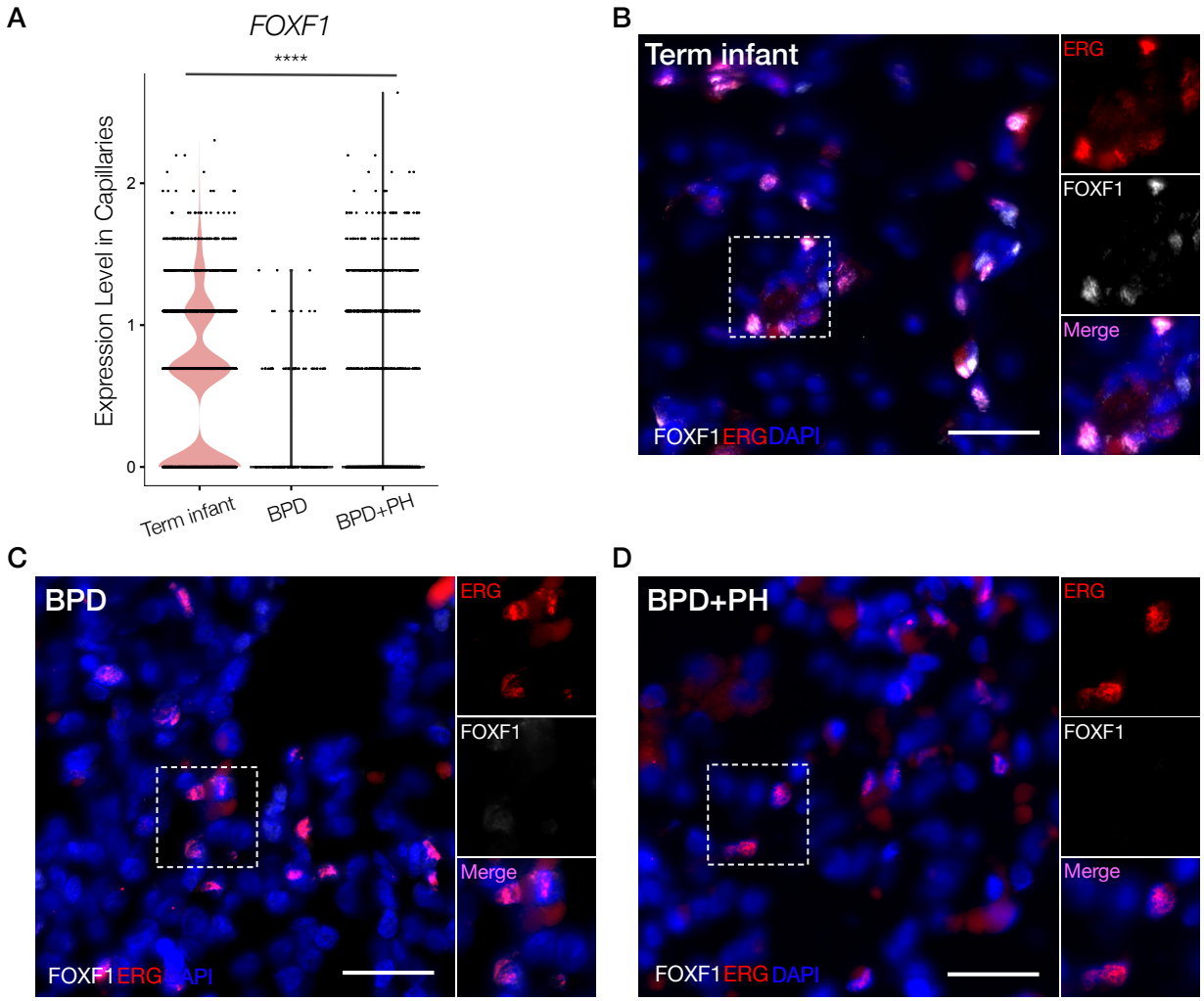


Figure 6: FOXF1 has decreased expression in human neonatal lung injury

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.

Discussion

In this single-cell transcriptomic analysis of human neonatal lungs in varying stages of injury, we have identified a previously-undefined endothelial cell state, abCap, that associates with bronchopulmonary dysplasia and pulmonary hypertension. Using predictive analysis tools and tissue validation, we discovered deficits in semaphorin signaling and an apparent functional decrease in FOXF1 in infants with BPD and BPD+PH. The latter suggests that a possible link between the rare disease ACDMPV and the severe forms of abnormal lung development that occur with preterm birth. These findings are largely concordant with findings in a murine model of BPD, providing foundational data that support a new role for semaphorin signaling as a critical mediator of normal lung development whose disruption may lead to the abnormal alveologenesis and vascular development observed after preterm birth[19].

While further work is required to define when abCaps emerge in disease and the role they may play in the pathogenesis of BPD, the discovery and validation of the abCap cell-state broadens our understanding of the endothelium as a receiver and coordinator of cell signaling and cell movements during this specific developmental stage. Analysis across hundreds of published datasets reveals that abCaps are distinct from the diversity of endothelial cell types and states observed in other diseases (e.g. adult PH, COVID, ARDS), and suggests that abCaps may be a cell state specific to BPD+PH. Deeper characterization of these cells demonstrates increased expression of YAP signaling targets[20–22] (e.g., *CCN1*, *CCN2*, *SERPINE*1, and *CRIM1*, along with abCap hallmark and YAP target gene *ANKRD1*). While the VEGF-induced YAP/TAZ signaling responsiveness in endothelial cells is required for vasculogenesis[20-23], elevated YAP/TAZ target expression in abCaps in BPD+PH patients could indicate an attempted compensatory response either within abCaps themselves or from signaling by adjacent cells. Indeed, both hyperoxia mouse models of neonatal injury and hypoxia PH models in mice have been reported to have increased endothelial CCN gene expression and *Ankrd1* expression[24–26]. Although abCaps did not align with features from prior published cell populations, abCaps do share some transcriptional homology with endothelial tip cells in early development. Understanding how failed injury repair after preterm birth leads to severe BPD+PH and the role of abCaps in the complex signaling milieu of the post-injury alveolar niche is critical to untangling the mechanisms that drive the abnormal repair in BPD and BPD+PH[27].

CellChat analysis predicted that the alveolar capillaries tend to be receivers of signaling information from adjacent alveolar epithelial and mesenchymal cells. In both BPD and BPD+PH and hyperoxia mouse models of BPD, we note a dramatic drop in semaphorin signaling. The semaphorin pathway is well-defined in growing neurons guiding axon movement and synapse coordination and in regulating axonal growth cones toward target cells during development[28]. While the role of this pathway in lung development is heretofore underappreciated, a global Sema3a knock out resulted in high perinatal lethality with marked impairment in alveologenesis in rare surviving KO pups[29] and ectopic *Sema3a* overexpression attenuates apoptosis in neonatal rats exposed to hyperoxia[30]. Further, genetic knockout of semaphorin receptor Nrp1 (neuropilin-1) in the adult mouse impairs airspace remodeling after smoke exposure[31]. Central to the role of semaphorins is their regulation of cellular movements toward and away from each other. Recent 4D-live imaging experiments suggest that alveologenesis results from cellular migration and shape change of epithelial and endothelial cells to form a functional unit of gas exchange[32], with the AT1 cell and aCap forming a shared basement membrane. Together, the decreased expression of *SEMA3B* by AT1 cells, *SEMA6A* by aCaps and gCaps, *SEMA3C* in the mesenchyme and decreased *NRP1* expression in endothelial cells under arrested lung development relative to controls, supports the live-imaging observation that alveolar formation requires the epithelial, mesenchymal, and endothelial cells of the neoalveolus to move away from the alveolar duct and toward each other with precisely coordinated movements. A deeper characterization of the timing and cell-specificity of semaphorin signaling during lung development may reveal critical windows in which this pathway could be targeted to promote lung growth after injury.

No study is without limitations. While transcriptomic data from a relatively underrepresented patient population are informative in generating new hypotheses about molecular mechanisms driving abnormal lung development after preterm birth, these data represent a relatively small number of patients, with autopsy samples having an inherent bias toward the sickest patients who have died from BPD and BPD+PH. That said, these data fill a critical gap in the currently available transcriptomic analysis of the normal and diseased lung across the lifespan. There exists an abundance of previously published adult normal and diseased lung data as well as multiple recent atlases of the human fetal lung, but very few published reports of neonatal premature injured and uninjured lung and term infant controls.

One strength of this study is its integration of these rare samples with recently published transcriptomic data from ACDMPV patients. The transcriptional deficits in semaphorin signaling common to both BPD+PH and ACDMPV suggest that the functional

deficits in *FOXF1* associated with the developmental phenotype in BPD and BPD+PH may have a mechanistic relationship with the observed deficits in ACDMPV, providing a foundation for further exploration of semaphorin signaling as a driver of normal alveolar development. With emerging technologies in spatial transcriptomic analysis from previously collected FFPE tissue blocks, we anticipate a rapid growth in samples across the neonatal and pediatric populations, with a goal of making these and future data accessible and searchable to promote discovery of long-awaiting targetable mechanisms that will promote lung growth and regeneration.

Methods

Human Infant Lung Repository

At the time of autopsy/biopsy, fresh lung tissue was isolated and processed for single-cell sequencing (scRNAseq) as described previously[5] with additional details below. In addition, we studied formalin-fixed paraffin embedded tissue blocks from human infant lung samples from autopsies and biopsies in varying stages of lung injury and without evidence of lung injury (as determined by a clinical pathologist) have been collected at Vanderbilt University Medical Center. All samples are de-identified with the exception of sex, gestational age at birth, age at time of death, cause of death, and whether or not the patient had evidence of BPD (defined as a need for supplemental oxygen at 36 weeks correct gestational age) or pulmonary hypertension (as determined by clinical echocardiography). A table with the clinical characteristics of the infant lungs used for FFPE tissue and scRNAseq is found in the supplement (**Supplementary Tables 1-2**). This infant lung repository has been reviewed and approved by the Vanderbilt University Institutional Review Board.

Single-cell sequencing

Human lung tissue from infants who died with BPD, BPD+PH, acute preterm lung injury, and term infant controls were processed into single-cell suspension as described previously[5]. scRNAseq libraries were generated using the 10x Chromium platform 3' or 5′ library preparation kits (10x Genomics) following the manufacturer's recommendations and targeting 10,000-20,000 cells per sample. Next-generation sequencing was performed using an Illumina Novaseq 6000. Reads with read quality less than 30 were filtered out and CellRanger Count v7.1.0 (10x Genomics) was used to align reads onto the GRCh38 reference genome. Full code used in the analysis of these data is available at

https://github.com/SucreLab/HumanNeonatal. A table with the clinical characteristics of infants whose lungs were sequenced is in the supplement (**Supplementary Table 1**).

Single-cell RNA sequencing analysis

The ambient RNA was removed from the CellRanger count matrix h5 file outputs using CellBender v0.2.1[33]. The scRNAseq data was analyzed using Seurat (v4.4.0)[34] in R (v4.3.2). The quality control of the ambient RNA corrected cells filtered out cells with >10 % mitochondrial mRNA mapped UMIs, <500 identified genes, and >7000 identified genes. The data was normalized using SCTransform (v0.4.1)[35] with the glmGamPoi (v1.14.0)[36] method by stabilizing the variance of the total molecular counts. The default parameters were used with SCTranform except for: batchvar set to correct batch effects from the tissue samples processed, vars.to.regress set to regress out the effects on differential gene expression analysis due to the presence of increased mitochondrial mRNAs and ribosomal rRNAs, and ncells set to NULL. The reduction of dimensionality was calculated via PCA. The PCA embeddings were further integrated with Harmony (v1.2.0)[37] to account for any remaining batch effects from the samples. The Harmony reduction was then applied in obtaining the UMAP embeddings and the shared nearest neighbor graph which subsequently allowed for the Louvain algorithm to find the cell clusters.

The clusters were split into the four main cell subtypes of epithelial, endothelial, mesenchymal, and immune based on respective hallmark genes (**Supplementary Table 3**). Each subtype went through an iterative process of SCTransform, PCA, Harmony, UMAP, shared nearest neighbors, and clustering to isolate and remove any remaining doublet or nonbiological cells. For each iteration, the cluster resolution would start in the range of $1 < n <= 2$, and with each iteration, n would be brought down slightly until the last iteration which would have $n = 1$. Cell type labels were assigned after quality control and cleaning of the data from their respective hallmark genes (**Supplementary Tables 3-4**). After annotation, the four subtype-group Seurat objects were merged to produce the final Seurat object.

CellChat[16] was used to predict possible ligand-receptor pair interactions between cell types of the same condition to generate hypotheses about possible signaling mechanisms associated with BPD and BPD+PH. We employed SCimilarity[15] to query cell populations or states that could have similar gene expression as the BPD+PH endothelial cell state (abCaps). To do this, we used a gene signature list curated from the intersection of the top Wilcoxon rank sum method ranked genes, top log2 fold change expression genes, and genes with the highest proportion expressed within abCaps versus all other cell types: *ANKRD1, CCN2,*

KITLG, SERPINE1, TNFRSF12A, CRIM1, GDF15, CCND1, EDN1, RBFOX2, CD151, PLS3, CD59, GBE1, PODXL. We then applied this gene module score to all cell populations or clusters in each study independently, with a cutoff of a cell score greater than two standard deviations from the mean cell score in the dataset. In summary, we looked for cells with features of the expression pattern identified in the abCap gene module. We then plotted the percentage of cells within each cluster that passed this cutoff.

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

Murine model of neonatal lung injury

C57BL/6 mice were exposed to 70% O_2 (hyperoxia) or 21% O_2 normoxia conditions from P1 until 5 days of age (P5), with hyperoxia animals receiving intranasal LPS at 1 μg/g on P3 and P4. On P7 and P14, lungs were inflation-fixed by gravity filling with 10% buffered formalin and paraffin embedded as described previously[38]. This protocol was approved by the Institutional Animal Care and Use Committee of Vanderbilt University and was in compliance with the Public Health Services policy on humane care and use of laboratory animals.

RNA in situ hybridization

RNAScope technology (ACDBio) was used to perform all RNA ISH experiments according to the manufacturer's instructions as described previously . RNAScope probes to the following human genes were used (*SEMA3B, AGER, SFTPC, SEMA6A, PECAM1).* RNAScope 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 [39], 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 figure 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), R56HL167937 (DBF), U01HL175444 (JAK, NEB, JMSS). We are grateful to Anne Hilgendorff and Vivian Siegel for thoughtful discussions and insights.

.

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. 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).

31. 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).

32. 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.

33. 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).

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

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

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

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

38. 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).

39. 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.

40. 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).