

Single-cell RNA sequencing identifies aberrant transcriptional profiles of cellular populations and altered alveolar niche signalling networks in Chronic Obstructive Pulmonary Disease (COPD)

Sauler M^{1*}, McDonough JE^{1*}, Adams TS¹, Kothapalli N¹, Schupp JC¹, Nouws J¹, Chioccioli, M¹, Omote N¹, Cosme Jr C¹, Poli S², Ayaub EA³, Chu SG³, Jensen KH⁴, Gomez-Villalobos J¹, Britto CJ¹, Raredon MSB^{5,6}, Timshel PN⁴, Kaminski N^{1*}, Rosas IO^{7*}

¹ Pulmonary, Critical Care and Sleep Medicine, Yale School of Medicine, New Haven, CT, USA

² Department of Internal Medicine, Mount Sinai Medical Center, Miami, FL, USA.

³ Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

⁴ Intomics A/S, Lyngby, Denmark

⁵ Department of Biomedical Engineering, Yale University, New Haven, CT, USA

⁶ Medical Scientist Training Program, Yale School of Medicine, New Haven, CT, USA

⁷ Pulmonary, Critical Care and Sleep Medicine, Baylor College of Medicine, Houston, TX, USA

* Contributed equally to this work

Corresponding Authors:

Maor Sauler MD

300 Cedar Street,

New Haven, CT 06519 USA

email: maor.sauler@yale.edu

John E. McDonough PhD

300 Cedar Street,

New Haven, CT 06519 USA

email: john.e.mcdonough@yale.edu

ABSTRACT

Chronic Obstructive Pulmonary Disease (COPD) pathogenesis involves a failure to maintain alveolar homeostasis due to repetitive injury and inflammation. In order to improve our understanding of cell-specific mechanisms contributing to COPD pathogenesis, we analysed single-cell RNA sequencing (scRNAseq) profiles of explanted parenchymal lung tissue from 17 subjects with advanced COPD requiring transplant and 15 control donor lungs. We identified a subpopulation of alveolar type II epithelial cells that uniquely express *HHIP* and have aberrant stress tolerance profiles in COPD. Amongst endothelial cells, we identified overlapping and unique shifts in transcriptional profiles of endothelial subtypes that may contribute to vascular inflammation and susceptibility to injury. We also identified population composition changes amongst alveolar macrophages. Finally, application of integrative analyses to our scRNAseq data identified cell-specific contributions to COPD heritability and dysfunctional cell-cell communication pathways that occur within the COPD alveolar niche. These findings provide cell type-specific resolution of transcriptional changes associated with advanced COPD that may underlie disease pathogenesis.

INTRODUCTION

Chronic Obstructive Pulmonary Disease (COPD) is characterized by persistent airflow obstruction and the pathologic findings of small airway inflammation and parenchymal tissue destruction (i.e. emphysema).^{1,2} While COPD is commonly caused by aerosolized pollutants such as cigarette smoke (CS), the clinical and biologic effects of CS are heterogenous. Therefore, complex interactions of genetic factors, host responses, and environmental exposures underlie COPD pathogenesis.

Previous studies of COPD pathogenesis based on murine models of disease, genetic association studies, or analyses of bulk human lung tissue have greatly expanded our understanding of disease pathogenesis.³⁻⁵ The pathogenic cascade of COPD is initiated by repetitive insults to epithelial and endothelial cells within the distal airways and alveolar niche. This engenders activation of diverse cellular processes including immune cell infiltration, extracellular matrix proteolysis, cellular metabolic dysfunction, loss of proteostasis, DNA damage, autophagy, cellular senescence, and activation of regulated cell death pathways.^{4,6,7} Consequently, there is an inability to maintain alveolar homeostasis, chronic inflammation, and alveolar septal destruction, particularly in advanced COPD.⁴ However there remains limited knowledge of cell type-specific mechanisms or complex interactions amongst the multiple lung cell types that contribute to disease progression.

Recent studies have used single-cell RNA sequencing (scRNAseq) to obtain gene expression profiles at single-cell resolution and identify novel cellular phenotypes and disease mechanisms.⁸⁻¹² Herein, we analysed scRNAseq profiles of parenchymal tissue obtained from explanted lungs of patients with advanced COPD requiring lung transplant and control donor lungs. Through this approach, we characterized aberrant transcriptional profiles of epithelial, endothelial, and immune cell populations. We also integrated genome-wide association study (GWAS) data with our scRNAseq analysis to determine cell types contributing to COPD heritability and applied ligand-receptor connectome analyses to identify dysregulated cell-cell interactions. These findings suggest novel cell-type-specific mechanisms that may contribute to disease pathogenesis.

METHODS

Sample collection and single-cell RNA sequencing: Tissue procurement, sample processing and single-cell sequencing was previously performed, and methods are further detailed in supplemental methods.¹³ Briefly, scRNAseq was performed on parenchymal lung tissue, either from subjects with advanced COPD undergoing lung transplantation or from rejected donor lungs without disease that were used as controls. Study protocols were approved by Partners Healthcare Institutional Board Review (IRB Protocol 2011P002419).

Clustering, differential cell expression and regulon analysis: Clustering and differential expression was performed using the Seurat package (v.3.2.0) in R. Overall marker genes for each cell type were identified by calculating the diagnostics odds ratio (DOR) for each gene, and cell types were annotated using canonical marker genes as previously described.¹³ We identified regulons (i.e. imputed transcription factors predicted to regulate gene expression) using pySCENIC (v.0.10.2).¹⁴ Further details are available in supplemental methods.

CELL-type Expression-specific integration for Complex Traits (CELLECT): We used CELLECT to quantify associations between cell-type specificity of expressed genes and findings from genome-wide association studies (GWAS) of lung function and COPD.^{15,16} CELLECT has been previously described¹⁷ and further detailed in supplemental methods. Briefly, CELLECT generates a genetic prioritization scores for each gene based on cell type-expression specificity and GWAS summary statistics for related genetic variants. It then prioritizes cell types enriched for genes with high genetic prioritization scores. As input to CELLECT, we used summary statistics derived from genetic studies of UK Biobank data for presence of COPD¹⁵ and lung function.¹⁶

Lung connectome to identify cell-cell interactions: Methods used to generate the lung connectome have been previously described and are detailed in supplemental methods.¹⁸ Briefly, average expression values for every gene within cell types were calculated and mapped using the FANTOM5 database of known ligand-receptor pairs. A node was defined as a cell type. An edge connecting two nodes was generated if > 5% of the cells within each node expressed the cognate molecule of the ligand-receptor pair, and edge-weights represented the product of the average expression values of the ligand and receptor. For the alveolar niche network, both non-directional and directional (i.e. incoming and outgoing) cumulative edge-weights were calculated using the sum of edge-weights computed from the average scaled expression values of all edges between two nodes. For the pathway centrality analysis, we only used ligand-receptor pairs preassigned to specific signaling modes (i.e. signaling pathways). Cumulative out- and incoming edge-weights were calculated using average unscaled expression values computed for each individual signalling mode. For these analyses, we also calculated Kleinberg hub and authority scores which are metrics of outgoing and incoming centrality that take into account the number of edges and edge-weights connecting a node within a network.

RESULTS

We analysed scRNAseq profiles of explanted lung tissue parenchyma obtained from 17 patients with advanced COPD, and 15 age-matched donor lungs as controls (**Figure 1A**). Demographics and pulmonary function test results are shown in **Supplemental Table 1**. There were eight females in both groups and the median age of all subjects was 62 years old (range 41–80). All advanced COPD subjects had radiographic evidence of emphysema. COPD subjects were former smokers; four of the donors were either current or former smokers. The final dataset consisted of 49,976 cells from control donor lungs and 61,564 cells from COPD lungs, and we identified 37 distinct cell types in both control and COPD lungs based on representative marker genes (**Figures 1B**). Canonical markers of identified cell types are shown. (**Figure 1C–E**). Data has been deposited in the Gene Expression Omnibus (GSE136831) and can be explored using our online portal (www.copdcellatlas.com).

Epithelial cells from advanced COPD lungs exhibit transcriptional evidence of aberrant differentiation and impaired stress responses: We identified all major epithelial cell types including alveolar type I (AT1) and alveolar type II (AT2) cells (**Figure 2A**). Amongst all epithelial cells, there were no significant cell proportion differences between COPD and control lung tissue except for a small population of 33 aberrant basaloid cells, from 9 COPD subjects (range: 1–10 cells per subject), which we recently identified in IPF (**Figure 2B**).¹³ These cells do not express basal markers *KRT5* and *KRT15* but do express *TP63*, *KRT17*, *LAMC2*, and senescence-related genes (*CDKN1A*, *CDKN2A*, *GDF15*).

Two distinct sub-clusters of AT2 cells (AT2_A and AT2_B) were observed in COPD and control donor lungs (**Figure 2C**). While both clusters expressed AT2 markers such as *SFTPC* and *SFPTA1*, the most striking difference between the two clusters was a markedly greater expression of these genes in the AT2_B population. AT2_B cells also expressed higher levels of tissue protease *PGC* and the chemokine *CXCL2*, and almost uniquely express *HHIP*, a common GWAS association for COPD and emphysema.¹⁹ Regulon analysis of imputed transcription factors suggested increased activity of *ETV1* and *ETV5*, which are involved in maintaining AT2 cell identity.²⁰ In contrast, AT2_A cells demonstrated a relative increase in the expression of genes implicated in AT2 stem cell function including *CDH1*,²¹ WNT pathway gene *TNIK*,²² tyrosine kinase receptor *ERBB4*,²³ and decreased expression of the WNT inhibitory protein *WIF1*. Regulon analysis suggested increased activity of transcription factors that potentiate or mediate HIPPO/YAP and WNT signalling pathways including *FOXP1*, *FOXP2*, *TEAD1/3/4*, *NFATC2*, and *STAT3*. Therefore, AT2_A resemble WNT-responsive AT2 epithelial progenitor cells (AEPs) and AT2_B resemble non-AEP AT2 cells that express higher levels of canonical AT2 cell markers.²⁴⁻²⁶

Amongst all epithelial cells, AT2_B cells had the greatest number of differentially expressed genes (DEGs) between control and COPD lung tissue, followed by ciliated, AT1, and AT2_A cells (**Supplemental table 2**). AT2_B cells from COPD lungs had decreased expression of genes involved in mitigating oxidative and toxin-

mediated stress (*MGST*, *AKR1A1*, *TXND17*, *GSTP1*, *ALDH2*) and genes encoding S100 family proteins (**Figure 2D**). The most significantly decreased expressed gene was *NUPRI* (fold change (fc)=0.35; adj. P=2.75x10⁻¹⁷), a stress-induced protein that orchestrates different aspects of the cellular stress response.²⁷ In AT2_A cells, we identified decreased gene expression and regulon activity related to cytoskeletal rearrangement (*KAZN*, *ANK3*, *UTRN*, *KANK1*) and epithelial differentiation (*NHSL1*, *ELF3*, *GPC6*, *MEF2D*) (**Figure 2E**). Together, these data highlight transcriptional profile shifts that may contribute to impaired AT2_B cellular stress responses and impaired AT2_A epithelial renewal with advanced COPD.

COPD-specific features of endothelial cells: Endothelial injury is implicated in COPD pathogenesis, but the specific changes in the human lung are not thoroughly understood. Clustering identified two distinct populations of capillary cells (denoted as capillary A and B), as well as arterial, venous, lymphatic, and systemic peri-bronchial endothelial cells (**Figure 3A**). No significant differences in the proportions of endothelial cells between disease and control were detected (**Supplemental Figure 1**). We compared DEGs between COPD and control cells for each endothelial cell type independently. While the majority of DEGs were unique to each cell type, we observed that many DEGs overlapped across multiple endothelial cell types. (**Figure 3B**). Notably, 30 genes were differentially expressed across three or more endothelial cell types, including increased expression of metallothionines (*MTIX*, *MTIM*, *MT2A*, *MT1E*) and AP-1 subunits (*FOS*, *FOSB*, *JUND*), and decreased expression of genes related to angiogenesis (*ID1*, *ID3*, *LDB2*) (**Figure 3C**). The largest set of overlapping endothelial DEGs occurred between both capillary cell types and involved 20 DEGs demonstrating increased expression of genes involved in inflammatory signalling (*CX3CLI*, *IL32*), cellular stress responses (*GADD45B*), and vesicular trafficking (*WASHC2C*, *WASHC2A*), and decreased expression of genes that can promote endothelial repair (*SEMA6A*, *WNT2B*).^{28,29} Capillaries also exhibited cell type-specific increases in the expression of inflammatory genes such as *TNFRSF10D*, *ITPKC*, and *IRF1* in capillary A and *TNFAIP3*, *IFI5*, and *IL6* in capillary B, as well as cell type-specific decreases in the expression of genes that promote endothelial angiogenesis and homeostasis such as *HIF3* and *MEF2C* in capillary A and *TFPI* in capillary B.³⁰ These data demonstrate overlapping and distinct DEGs amongst lung endothelial cell types that may contribute to pathologic endothelial activation or impaired endothelial repair in COPD.

Transcriptional heterogeneity of alveolar macrophages: We explored the immunologic landscape of the distal lung and identified 18 distinct immune cells. (**Supplemental Figure 2**) Between control and COPD, we observed significant differences in cell compositions amongst interstitial macrophages (fc=0.48; P=4.5x10⁻⁴), non-classical monocytes (fc=2.18; P=0.026), plasmacytoid dendritic cells (fc=3.46; P=8.1x10⁻⁴), conventional dendritic cells (fc=2.79; P=0.010), mature dendritic cells (fc=1.89; P=0.046), and mast cells (fc=7.70; P=4.5x10⁻⁴) (**Supplemental Figure 1**). The largest population of cells were alveolar macrophages which have been implicated in alveolar inflammation and tissue destruction.^{31,32} To characterize the heterogeneity of

alveolar macrophages, we re-embedded the alveolar macrophage cluster in UMAP space and re-clustered these cells into eight clusters (**Figure 4A**). We then compared the relative abundance of the eight clusters between control and COPD and identified changes in alveolar macrophage population composition between COPD and control; cluster-0 macrophages were enriched amongst controls while cluster-5 macrophages were enriched amongst patients with COPD (adjusted $p < 0.05$) (**Figure 4B**). The corresponding cluster markers for cluster-5 and cluster-0 are shown in **Figure 4C**. Cluster-0 markers included heat shock proteins (*HSPA1A*, *HSPA1B*, *HSPH1*, *HSPD1*, *HSP90AA1*, *HSPE1*), while the top cluster-5 markers were metallothionines (*MT1G*, *MT1X*, *MT1E*, *MT2A*, *MT1M*, *MT1F*, *MT1H*, *MT1A*, *MT1L*). There were also multiple DEGs between COPD and control macrophages that were not represented in any specific cluster. The most significant DEGs included genes associated with macrophages chemotaxis and inflammation (*THBS1*, *PEL1I*, and *CDC42*) (**Figure 4D**). These results highlight changes in macrophage population composition as well as common DEGs across the multiple macrophages subpopulations.

COPD genetic variants enrich for alveolar niche cells: The cell types through which genetic variants mediate COPD susceptibility are not well-defined.³³ We sought to address this by using CELLECT, a computational tool that colocalizes scRNAseq and GWAS data. Using this approach, we identified multiple cell types enriched for GWAS-associated genes including AT2, smooth muscle, and fibroblast populations (**Figure 5A**). Notably, we did not identify immune cell populations with high prioritization scores despite adequate representation of immune cells in our dataset. For reference, the top genes commonly associated with COPD with genetic prioritization scores > 0.5 are shown (**Figure 5B**). Collectively, these data suggest that much of COPD heritability is mediated by non-hematopoietic lung resident cells.

Alveolar niche network and pathway centrality analysis reveal changes in alveolar signalling topology: We performed connectome analyses to characterize alveolar signalling topologies and identify differences in cell-cell communication between control and COPD lungs. We generated network-level maps based on predicted ligand-receptor interactions amongst alveolar cell types (**Figure 6A**). To quantify changes in cell-cell interactions, we measured differences in directed cumulative edge-weights (i.e. cumulative outgoing or cumulative incoming edges) between control and COPD (**Figure 6B**). In COPD, there were increases in directed edge-weights between nodes of capillaries, monocytes, and macrophage, with the greatest increase in edge-weight occurring between non-classical monocytes and capillaries. We then sought to determine the contribution of individual signalling pathways to COPD-related changes in alveolar signalling topology. Ligand-receptor interactions were preassigned to canonical signalling pathways and centrality metrics were calculated, including incoming and outgoing edge-weights and Kleinberg hub and authority scores. These centrality rankings prioritize specific cell types responsible for cell-cell interactions through these individual pathways. We found changes in 16 out of 22 signalling pathways, including inflammatory pathways (e.g. CC-

motif chemokines, CXCL-motif chemokines, interleukins, and TNF) and canonical signalling pathways previously implicated in COPD pathogenesis (VEGF, PDGF, MET, WNT and TGF- β) (**Figure 6C**). We did not identify changes in FGF, Ephrin, and NOTCH signalling (**Supplemental Figure 3**). To quantify changes in cell-cell interactions, we generated individualized pathway centrality networks for each subject, and compared edge-weights for each signalling mode across subjects between control and COPD. The pathway with the most significant differences in cumulative edge-weights was CXCL-motif chemokine signalling. There was an increase in outgoing CXCL-motif edge-weights from adventitial fibroblasts (fc=2.4; P=2.0x10⁻⁴) and capillary B cells (fc=4.2; P=3.1x10⁻⁴), and an increase in incoming CXCL-motif edge-weights towards alveolar macrophages (fc=3.2; P=7.2x10⁻⁴), interstitial macrophages (fc=3.0; P=4.6x10⁻³), AT2_B cells (fc=3.6; P=4.0x10⁻³), and AT1 cells (fc=4.2; P=3.5x10⁻²). (**Figure 6D**).¹⁴ Collectively, these findings suggest that increased interactions between capillaries, monocytes, and macrophages may be major drivers of alveolar inflammation in COPD, and demonstrate potential pathway-specific changes in cell-cell communication, particularly increases in CXCL- chemokine signalling, occurring within the COPD alveolar niche.

DISCUSSION

In this study, we leveraged the cell-specific resolution of scRNAseq to improve our understanding of aberrant transcriptional changes in the lung parenchyma that occur with advanced COPD. We identified two subpopulations of alveolar type II epithelial cells, an AT2_B population that expresses *HHIP* with transcriptional changes suggestive of impaired stress tolerance profiles and an AT2_A subpopulation with transcriptional changes suggestive of impaired epithelial differentiation. We identified augmented stress and inflammatory expression profiles in lung endothelial cells in emphysema and distinct macrophage subpopulations enriched in control and emphysematous lung tissue. Integration of GWAS with our scRNAseq dataset identified lung resident cells as the primary drivers of COPD heritability, while our connectome analysis revealed striking differences in cell-cell interactions, particularly CXCL-chemokine signalling, within the alveolar niche between control and COPD. Together, these findings suggest cell-specific molecular changes that may underlie COPD pathogenesis.

COPD pathogenesis is associated with a failure of AT2 cells to maintain alveolar homeostasis. Previous studies have demonstrated two unique populations of AT2 cells: AEPs are AT2 cells that can function as facultative progenitor cells to regenerate alveolar epithelium and non-AEPs express higher levels of canonical AT2 cell markers.²⁴⁻²⁶ Similarly, we identified two AT2 populations in both control and COPD lungs that correspond to these populations; AT2_A cells have a small relative increase in the expression WNT-responsive genes and regulons while AT2_B cells have higher expression of surfactant proteins, *HHIP*, and other AT2 cellular markers. Interestingly, the largest difference between AT2_A and AT2_B cells was the expression of genes related to surfactant production and other homeostatic AT2 functions, suggesting that the contrasting properties of these two cell populations may extend beyond the eponymous functions of AEPs and non-AEPs.

Both AT2_A and AT2_B populations have divergent transcriptional changes in advanced COPD. The largest number of DEGs was surprisingly identified amongst AT2_B cells, that have decreased expression of proteins involved in mitigating oxidative and toxin-mediated stress responses. Additionally, we found that AT2_B cells in both COPD and control lungs almost uniquely express *HHIP*, a gene repeatedly associated with COPD and emphysema in genetic association studies.¹⁹ Previous studies showed haploinsufficient *Hhip* mice are prone to emphysema in part through their susceptibility to oxidative stress,³⁴ with the caveat that the cellular distribution of *HHIP* is different between mice and humans. Collectively, these findings suggest unique roles for AT2_A and AT2_B cell populations in COPD pathogenesis, in which AT2_B cells have a diminished capacity to mitigate cellular stress while AT2_A cells have an impaired capacity to mediate repair.

Our analysis provides a comprehensive overview of transcriptional changes amongst endothelial cells in advanced COPD.³⁵⁻⁴⁰ We found evidence for endothelial activation (i.e. pro-inflammatory, pro-coagulant phenotype) as demonstrated by increased endothelial expression of AP1 transcription factor, inflammatory cytokines, cellular stress response genes, and pro-coagulant signalling genes. We also found evidence for a diminished capacity for endothelial repair as suggested by downregulation of genes associated with angiogenesis. The causes of these transcriptional changes may be a direct consequence of oxidative stress from environmental exposures. However, our connectome analyses suggest the possibility that increased crosstalk between capillaries and myeloid-derived cells may also be important contributors to endothelial transcriptional changes in advanced COPD. Using our connectome analyses, we identified dysregulated alveolar signalling topologies and prioritized cell types responsible for canonical signalling pathways previously implicated in COPD pathogenesis. One of the most striking findings was increased endothelial CXCL-motif chemokine activity and increased signalling amongst monocytes, macrophages, and capillaries, particularly from non-classical monocytes to capillary A cells. Therefore, increased signalling between capillaries and myeloid-derived cell populations may be important contributors to endothelial activation in advanced COPD.

In contrast to solely identifying phenotypic shifts amongst endothelial cells with advanced COPD, we observed both phenotypic shifts and population composition changes amongst alveolar macrophages. A subset of macrophages in control lungs was enriched for heat shock proteins while lungs from subjects with advanced COPD were enriched with a cluster expressing metallothionines. Metallothionines are induced by oxidative stress and inflammation and can mediate protective effects by sequestering intracellular metals such as zinc and copper.⁴¹ Despite their homeostatic function, little is known about the consequences of chronic metallothionein upregulation in the setting of advanced COPD where redox homeostasis and heavy metal metabolism are disrupted.⁴² For instance, Menkes disease, a congenital disorder of copper deficiency causes emphysema,⁴³ additionally, dysregulated zinc homeostasis can cause impaired phagocytosis and an abnormal inflammatory response in macrophages.⁴⁴ Therefore, understanding metallothionein regulation of zinc and copper in COPD may improve our understanding of disease pathogenesis.

We found enrichment for genetic variants associated with lung function in COPD amongst epithelial and mesenchymal cells but not immune cells. This novel finding suggests that lung resident cells mediate much of the heritability of COPD and suggest the intriguing hypothesis that aberrant immunologic phenotypes in COPD are largely acquired rather than genetically programmed. An effective therapeutic strategy may be targeting epithelial, endothelial, and fibroblast cells in at-risk individuals in early COPD; however, future studies will be necessary to elucidate cell-type specific pathways that underlie COPD heritability.

There are a few limitations that need to be considered. First, COPD is a heterogeneous disorder, yet the subjects in this study represented a distinct subtype of COPD patients with advanced disease, requiring transplant,

radiographic evidence of emphysema, and who were not actively smoking due to their upcoming lung transplant, and therefore our findings cannot be generalized to all COPD sub-phenotypes or earlier stages of disease. However, our analysis likely reflects the pathologic changes that persist despite smoking cessation. Additionally, while all subjects had radiographic emphysema, we do not know the degree of emphysema within the tissue sample obtained from a given subject. However, given the large number of cells and subjects included in this study, the spectrum of pathologic changes in the lung parenchyma with advanced emphysema were likely captured in our scRNAseq analysis. Other limitations are common to scRNAseq experiments. While transcriptomic changes are informative, they do not always reflect protein concentration, nor do they necessarily impact cellular function. There may also be dissociation bias due variable cellular sensitivities to digestive enzymes and differences in cellular embedding in the extracellular matrix. However, all samples were processed using a previously validated protocol. Finally, findings from the study, particularly the *in silico* CELLECT and connectome analyses, will require further validation.

Overall, scRNAseq identified novel transcriptional profiles of heterogenous populations of AT2, macrophage, and endothelial cell types within the emphysematous lung. Additionally, we demonstrated a key role for lung resident cells in COPD heritability and dysregulation of canonical signalling pathways within the alveolar niche of the emphysematous lung. Together, these findings highlight the complexity of cellular injury, inflammation, and impaired tissue repair that underlies COPD pathogenesis. Future studies evaluating outcomes of specific transcriptional dysregulations identified herein will provide novel insights into mechanisms that contribute to disease.

Figure 1 Profiling of cell types in COPD using scRNAseq data. (A) Overview of project design. Tissue from 17 lungs with advanced COPD and 15 control donor lungs were dissociated into single cell suspensions. Individual cells were barcoded and sequenced for analysis. (B) Uniform Manifold Approximation and Projection (UMAP) representation of 111,540 single cells grouped into 37 distinct cell types (left) with identification of COPD and control cells (right). Violin plots of normalized expression values for canonical cell-specific marker genes for (C) epithelial, (D) endothelial, and (E) stromal cells.

Figure 2 Epithelial cells in COPD demonstrate dysfunction in AT2 subtypes. (A) UMAP of epithelial cells from COPD and control donor lungs. Samples are colour labelled by cell type (*left*) and disease category (*right*). Alveolar type II (AT2) cells can be distinguished as two clusters denoted as AT2_A and AT2_B. (B) Distribution of subject-specific epithelial cell types as a fraction of the total number of epithelial cells per disease category. Boxes represent interquartile ranges (IQRs); whiskers are 1.5 x IQR; dots represent subjects (C) Dot plot of z-scores for marker gene expression values (*top*) and predicted transcription factor (i.e. regulon) activity (*bottom*) of AT2_A and AT2_B cells. Dot size reflects percentage of cells with gene expression or regulon activity; colour corresponds to degree of expression (gene) or activity (regulon). (D, E) Heatmaps of z-scores for differentially expressed genes (yellow/purple) and predicted transcription factor (i.e. regulon) activity (red/blue) between control and COPD for AT2_B and AT2_A cells (Wilcoxon rank sum test, FDR <0.05). Each column represents expression values for an individual cell. Columns are hierarchically ordered by disease phenotype and subject, in which disease category and individual subject are represented by unique colours. z-scores were calculated across all epithelial cells.

Figure 3 COPD endothelial cell types demonstrate universal and cell type-specific transcriptional aberrations. (A) UMAPs of all vascular endothelial (VE) and lymphatic endothelial cells from control and COPD subjects. UMAPs are colour labelled by cell type (*top*) and disease status (*bottom*). UpSet plot visualizing the properties of intersecting and unique sets of differentially expressed (DE) genes between COPD and control amongst endothelial (Wilcoxon rank sum test, $p < 0.001$, $f_c > 0.5$). (C) Heatmap of corresponding differentially expressed genes between COPD and control amongst six subtypes of endothelial cells. Each column represents expression values for an individual cell. Columns are hierarchically ordered by endothelial subtype, disease phenotype, and then subject. Grey row (*top*): expression values for top marker genes are unity normalized between 0 and 1 across all endothelial subtypes. Orange row (*middle*): z-scores of differentially expressed genes in three or more endothelial cell types between control and COPD. Yellow row (*middle*): z-scores of differentially expressed genes in both capillary A and capillary B cells between control and COPD.

Blue and green row (*bottom*): z-scores of differentially expressed genes unique to capillary A (blue) or capillary B (green). Unity normalization and z-score calculations were performed using all endothelial subtypes.

Figure 4 Changes in alveolar macrophage population composition in COPD. (A) UMAPs of control and COPD alveolar macrophage cells, colour labelled by Louvain cluster (*left*) and disease status (*right*). (B) Percent makeup of alveolar macrophage across all nine Louvain clusters per subject, grouped by disease state. Boxes represent interquartile ranges (IQRs); whiskers are 1.5 x IQR; dots represent subjects. The percent of cluster 0 and cluster 5 alveolar macrophages are different between control and COPD (Wilcoxon rank sum test, FDR <0.05). (C) Heatmap of the distribution of z-scores of marker genes for cluster 0 and cluster 5 alveolar macrophages. Columns represent expression values from individual cells and are hierarchically ordered by macrophages cluster, disease status, and subject. (D) Density plots of *THBS1*, *PEL1* and *CDC42* gene expression across all alveolar macrophage cells, grouped by control (blue) and COPD (red) subjects. Within each disease group, subjects are ordered top to bottom from highest within-disease group expression average to lowest.

Figure 5 CELLECT analysis identifies cellular basis for COPD heritability. (A) Plot of negative log adjusted p-values for cell type-specific enrichment for GWAS-identified genes associated with lung function. Red line denotes threshold of significance (FDR < 0.05). Data showing high enrichment scores for non-hematopoietic cells are associated with COPD heritability traits. (B) Dot plot of z-scores for commonly associated GWAS genes that were highly ranked in CELLECT analysis. Dot size reflects percentage of cells with gene expression; colour corresponds to degree of expression.

Figure 6 Alveolar niche networks and pathway centrality analyses. (A) Network plots of the alveolar niche in control (*left*) and COPD (*right*). Nodes represent cell types. Node size is proportional to the Kleinberg hub score, and the thickness of each edge is proportional to the sum of all edge-weights (non-directed) between two nodes. Individual cell types are labelled by colour and number. (B) Heatmap shows relative changes between control and COPD in cumulative directed edge-weights between outgoing (source) and incoming (target) nodes. (C) Centrality analysis comparing the alveolar connectome for each mode of signalling between control and COPD. Panel shows outgoing edge-weights and Kleinberg hub scores (*left*) and incoming edge-weights and Kleinberg authority scores (*right*). Cumulative edge-weights (*x-axis*) have been scaled for each mode of signalling. Dot size is proportional to the Kleinberg scores for each cell type within a given mode of signalling. Individual cell types are colour labelled as in Figure 6A, and numbers shown identify cell types with the largest cumulative edge-weights. P-values computed using the Durbin test to compare control and COPD across cell types for each signalling mode. (D) Box and whisker plots of cumulative edge-weights were computed using

patient-level connectomes for CXCL-motif signalling. P-values computed using Wilcoxon rank sum tests to compare cumulative edge-weights between control and COPD subjects. AT1=alveolar type I epithelial cells, AT2=alveolar type II epithelial cells, cMonocyte=classical monocytes, ncMonocyte=non-classical monocytes, VE=vascular endothelial, Fibro=fibroblast, Semaph=semaphorins. MØ=interstitial macrophages, MØ alveolar=alveolar macrophage.

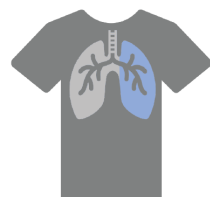
REFERENCES

1. Hogg JC, Chu F, Utokaparch S, et al. The nature of small-airway obstruction in chronic obstructive pulmonary disease. *N Engl J Med* 2004; **350**(26): 2645-53.
2. Rabe KF, Watz H. Chronic obstructive pulmonary disease. *Lancet* 2017; **389**(10082): 1931-40.
3. Silverman EK. Genetics of COPD. *Annu Rev Physiol* 2020; **82**: 413-31.
4. Tudor RM, Petrache I. Pathogenesis of chronic obstructive pulmonary disease. *J Clin Invest* 2012; **122**(8): 2749-55.
5. Barnes PJ, Burney PG, Silverman EK, et al. Chronic obstructive pulmonary disease. *Nat Rev Dis Primers* 2015; **1**: 15076.
6. Sauler M, Bazan IS, Lee PJ. Cell Death in the Lung: The Apoptosis-Necroptosis Axis. *Annu Rev Physiol* 2019; **81**: 375-402.
7. Sauler M, Lamontagne M, Finnemore E, et al. The DNA repair transcriptome in severe COPD. *Eur Respir J* 2018; **52**(4).
8. Marcus JS, Anderson WF, Quake SR. Microfluidic single-cell mRNA isolation and analysis. *Anal Chem* 2006; **78**(9): 3084-9.
9. Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* 2019; **20**(2): 163-72.
10. Habermann AC, Gutierrez AJ, Bui LT, et al. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Science Advances* 2020; **6**(28): eaba1972.
11. Reyfman PA, Walter JM, Joshi N, et al. Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2018.
12. Plasschaert LW, Zilionis R, Choo-Wing R, et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* 2018; **560**(7718): 377-81.
13. Adams TS, Schupp JC, Poli S, et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Science Advances* 2020; **6**(28): eaba1983.
14. Aibar S, Gonzalez-Blas CB, Moerman T, et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* 2017; **14**(11): 1083-6.
15. Sakornsakolpat P, Prokopenko D, Lamontagne M, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat Genet* 2019; **51**(3): 494-505.
16. Shrine N, Guyatt AL, Erzurumluoglu AM, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet* 2019; **51**(3): 481-93.
17. Timshel PN, Thompson JJ, Pers TH. Mapping heritability of obesity by brain cell types. *bioRxiv* 2020: 2020.01.27.920033.
18. Raredon MSB, Adams TS, Suhail Y, et al. Single-cell connectomic analysis of adult mammalian lungs. *Sci Adv* 2019; **5**(12): eaaw3851.
19. Zhou X, Baron RM, Hardin M, et al. Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. *Hum Mol Genet* 2012; **21**(6): 1325-35.
20. Zhang Z, Newton K, Kummerfeld SK, et al. Transcription factor Etv5 is essential for the maintenance of alveolar type II cells. *Proc Natl Acad Sci U S A* 2017; **114**(15): 3903-8.
21. Reddy R, Buckley S, Doerken M, et al. Isolation of a putative progenitor subpopulation of alveolar epithelial type 2 cells. *Am J Physiol Lung Cell Mol Physiol* 2004; **286**(4): L658-67.
22. Mahmoudi T, Li VS, Ng SS, et al. The kinase TNK1 is an essential activator of Wnt target genes. *EMBO J* 2009; **28**(21): 3329-40.
23. Desai TJ, Brownfield DG, Krasnow MA. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature* 2014; **507**(7491): 190-4.
24. Zacharias WJ, Frank DB, Zepp JA, et al. Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. *Nature* 2018; **555**(7695): 251-5.

25. Nabhan AN, Brownfield DG, Harbury PB, Krasnow MA, Desai TJ. Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science* 2018; **359**(6380): 1118-23.
26. Travaglini KJ, Nabhan AN, Penland L, et al. A molecular cell atlas of the human lung from single cell RNA sequencing. *bioRxiv* 2020: 742320.
27. Cano CE, Hamidi T, Sandi MJ, Iovanna JL. Nupr1: the Swiss-knife of cancer. *J Cell Physiol* 2011; **226**(6): 1439-43.
28. Segarra M, Ohnuki H, Maric D, et al. Semaphorin 6A regulates angiogenesis by modulating VEGF signaling. *Blood* 2012; **120**(19): 4104-15.
29. Goss AM, Tian Y, Tsukiyama T, et al. Wnt2/2b and beta-catenin signaling are necessary and sufficient to specify lung progenitors in the foregut. *Dev Cell* 2009; **17**(2): 290-8.
30. Lin Q, Lu J, Yanagisawa H, et al. Requirement of the MADS-box transcription factor MEF2C for vascular development. *Development* 1998; **125**(22): 4565-74.
31. Shapiro SD. The macrophage in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1999; **160**(5 Pt 2): S29-32.
32. Hautamaki RD, Kobayashi DK, Senior RM, Shapiro SD. Requirement for macrophage elastase for cigarette smoke-induced emphysema in mice. *Science* 1997; **277**(5334): 2002-4.
33. Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, Laird NM. Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. *Am J Respir Crit Care Med* 2013; **188**(8): 941-7.
34. Lao T, Jiang Z, Yun J, et al. Hhip haploinsufficiency sensitizes mice to age-related emphysema. *Proc Natl Acad Sci U S A* 2016; **113**(32): E4681-7.
35. Casanova C, de Torres JP, Navarro J, et al. Microalbuminuria and hypoxemia in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2010; **182**(8): 1004-10.
36. Kasahara Y, Tudor RM, Taraseviciene-Stewart L, et al. Inhibition of VEGF receptors causes lung cell apoptosis and emphysema. *J Clin Invest* 2000; **106**(11): 1311-9.
37. Thomashow MA, Shimbo D, Parikh MA, et al. Endothelial microparticles in mild chronic obstructive pulmonary disease and emphysema. The Multi-Ethnic Study of Atherosclerosis Chronic Obstructive Pulmonary Disease study. *Am J Respir Crit Care Med* 2013; **188**(1): 60-8.
38. Clauss M, Voswinckel R, Rajashekhar G, et al. Lung endothelial monocyte-activating protein 2 is a mediator of cigarette smoke-induced emphysema in mice. *J Clin Invest* 2011; **121**(6): 2470-9.
39. Petrache I, Natarajan V, Zhen L, et al. Ceramide upregulation causes pulmonary cell apoptosis and emphysema-like disease in mice. *Nat Med* 2005; **11**(5): 491-8.
40. Gonzalez S, Hards J, van Eeden S, Hogg JC. The expression of adhesion molecules in cigarette smoke-induced airways obstruction. *Eur Respir J* 1996; **9**(10): 1995-2001.
41. Coyle P, Philcox JC, Carey LC, Rofe AM. Metallothionein: the multipurpose protein. *Cell Mol Life Sci* 2002; **59**(4): 627-47.
42. Rahman I, MacNee W. Oxidant/antioxidant imbalance in smokers and chronic obstructive pulmonary disease. *Thorax* 1996; **51**(4): 348-50.
43. Mizuno S, Yasuo M, Bogaard HJ, et al. Copper deficiency induced emphysema is associated with focal adhesion kinase inactivation. *PLoS One* 2012; **7**(1): e30678.
44. Gao H, Zhao L, Wang H, et al. Metal transporter Slc39a10 regulates susceptibility to inflammatory stimuli by controlling macrophage survival. *Proc Natl Acad Sci U S A* 2017; **114**(49): 12940-5.

FIGURE 1

A Human Lung Tissue Dissociation and Barcoding Sequencing and Analysis



Control n = 15 (8 F; 7 M)
COPD n = 17 (8 F; 9 M)

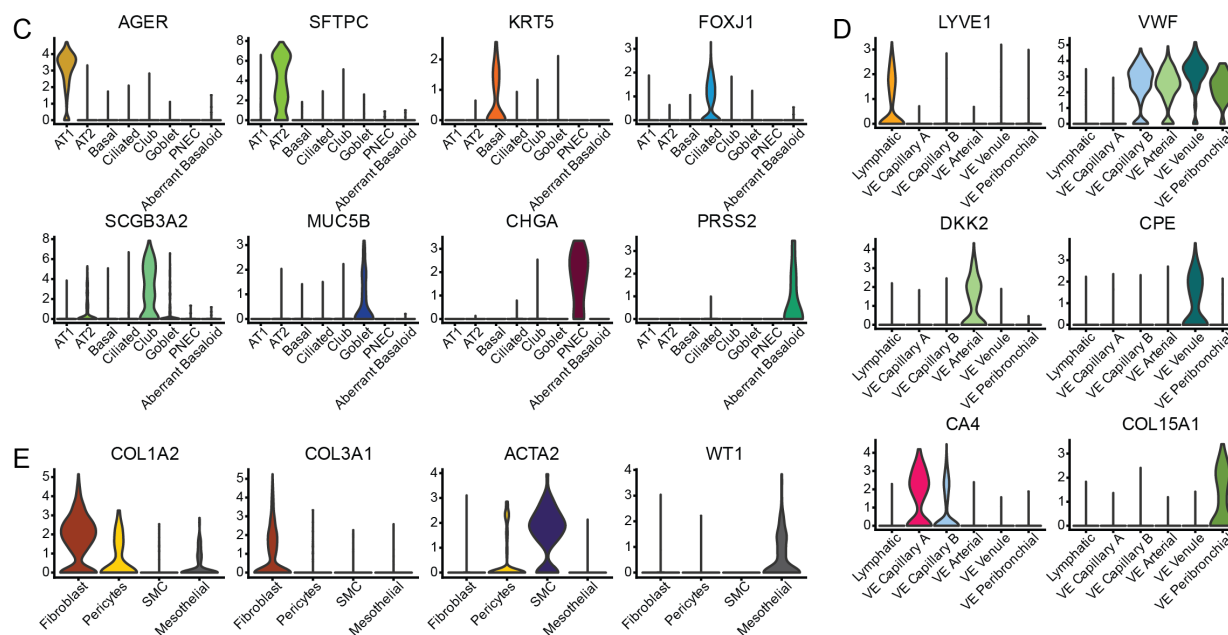
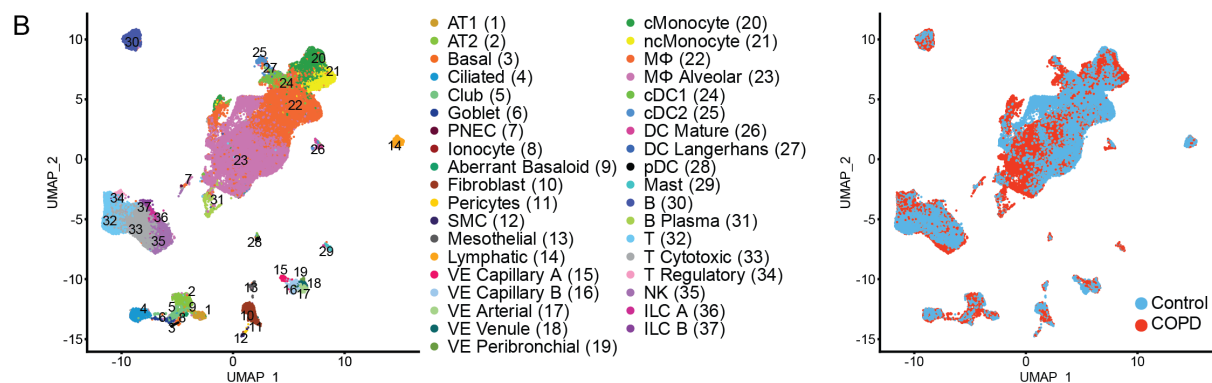
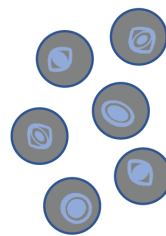


Figure 1 Profiling of cell types in COPD using scRNAseq data. (A) Overview of project design. Tissue from 17 lungs with advanced COPD and 15 control donor lungs were dissociated into single cell suspensions. Individual cells were barcoded and sequenced for analysis. **(B)** Uniform Manifold Approximation and Projection (UMAP) representation of 111,540 single cells grouped into 37 distinct cell types (left) with identification of COPD and control cells (right). Violin plots of normalized expression values for canonical cell-specific marker genes for **(C)** epithelial, **(D)** endothelial, and **(E)** stromal cells.

FIGURE 2

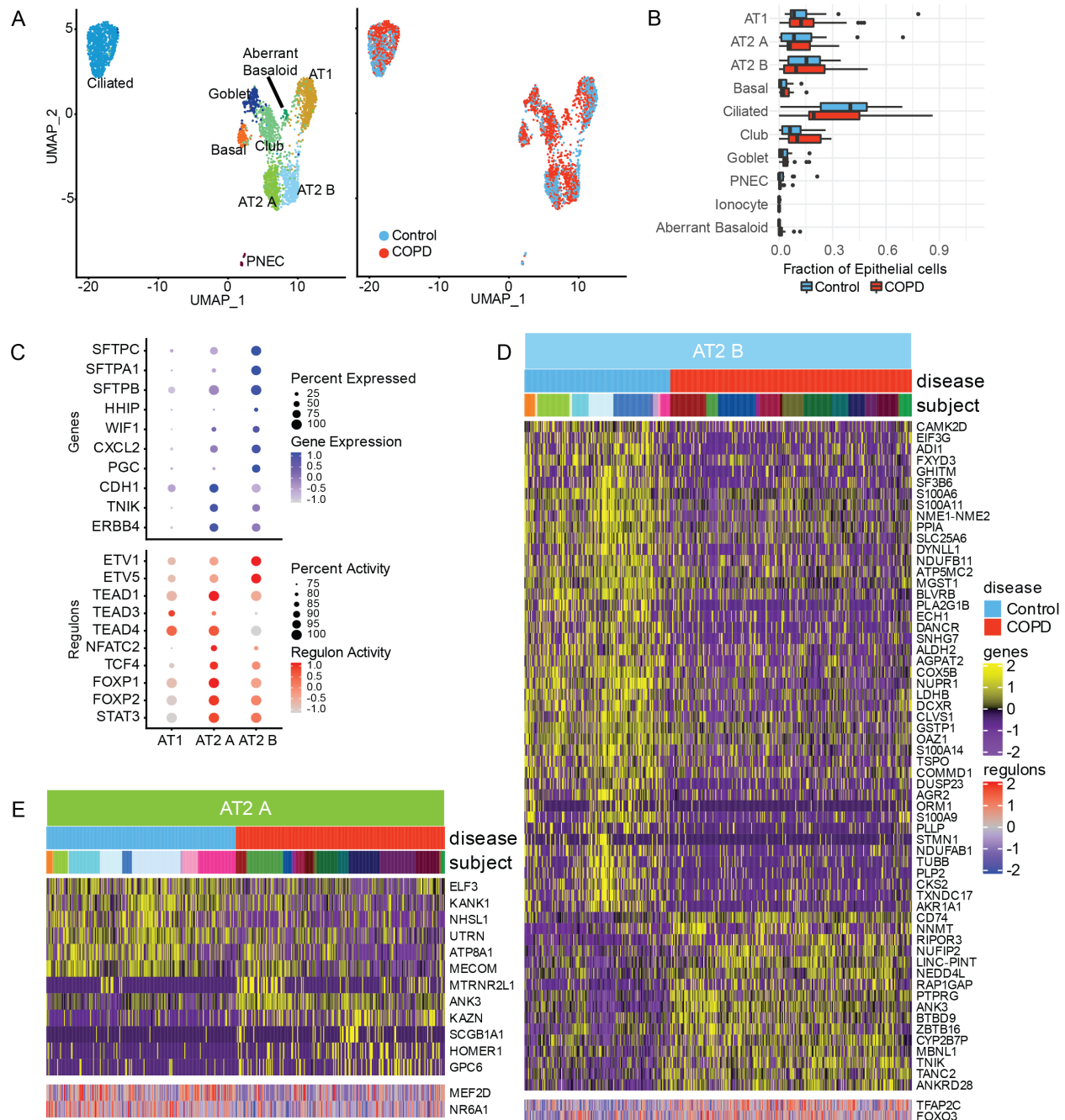


Figure 2 Epithelial cells in COPD demonstrate dysfunction in AT2 subtypes. (A) UMAP of epithelial cells from COPD and control donor lungs. Samples are colour labelled by cell type (*left*) and disease category (*right*). Alveolar type II (AT2) cells can be distinguished as two clusters denoted as AT2_A and AT2_B. (B) Distribution of subject-specific epithelial cell types as a fraction of the total number of epithelial cells per disease category. Boxes represent interquartile ranges (IQRs); whiskers are 1.5 x IQR; dots represent subjects (C) Dot plot of z-scores for marker gene expression values (*top*) and predicted transcription factor (i.e. regulon) activity (*bottom*) of AT2_A and AT2_B cells. Dot size reflects percentage of cells with gene expression or regulon activity; colour corresponds to degree of expression (gene) or activity (regulon). (D, E) Heatmaps of z-scores for differentially expressed genes (yellow/purple) and predicted transcription factor (i.e. regulon) activity (red/blue) between control and COPD for AT2_B and AT2_A cells (Wilcoxon rank sum test, FDR < 0.05). Each column represents expression values for an individual cell. Columns are hierarchically ordered by disease phenotype and subject, in which disease category and individual subject are represented by unique colours. z-scores were calculated across epithelial cells.

FIGURE 3

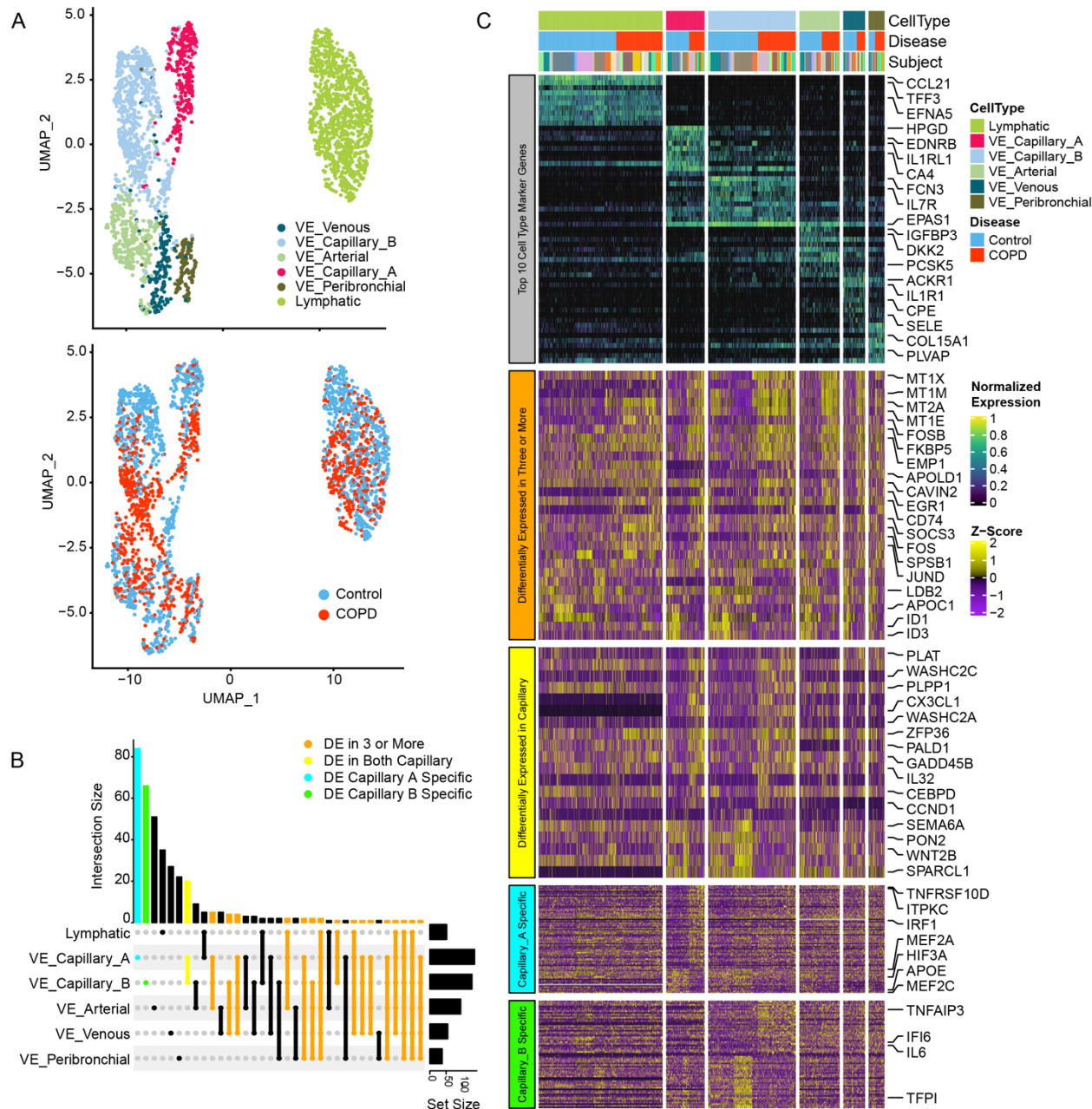


Figure 3 COPD endothelial cell types demonstrate universal and cell type-specific transcriptional aberrations. (A) UMAPs of all vascular endothelial (VE) and lymphatic endothelial cells from control and COPD subjects. UMAPs are colour labelled by cell type (*top*) and disease status (*bottom*). UpSet plot visualizing the properties of intersecting and unique sets of differentially expressed (DE) genes between COPD and control amongst endothelial (Wilcoxon rank sum test, $p < 0.001$, $f_c > 0.5$). **(C)** Heatmap of corresponding differentially expressed genes between COPD and control amongst six subtypes of endothelial cells. Each column represents expression values for an individual cell. Columns are hierarchically ordered by endothelial subtype, disease phenotype, and then subject. Grey row (*top*): expression values for top marker genes are unity normalized between 0 and 1 across all endothelial subtypes. Orange row (*middle*): z-scores of differentially expressed genes in three or more endothelial cell types between control and COPD. Yellow row (*middle*): z-scores of differentially expressed genes in both capillary A and capillary B cells between control and COPD. Blue and green row (*bottom*): z-scores of differentially expressed genes unique to capillary A (blue) or capillary B (green). Unity normalization and z-score calculations were performed using all endothelial subtypes.

FIGURE 4

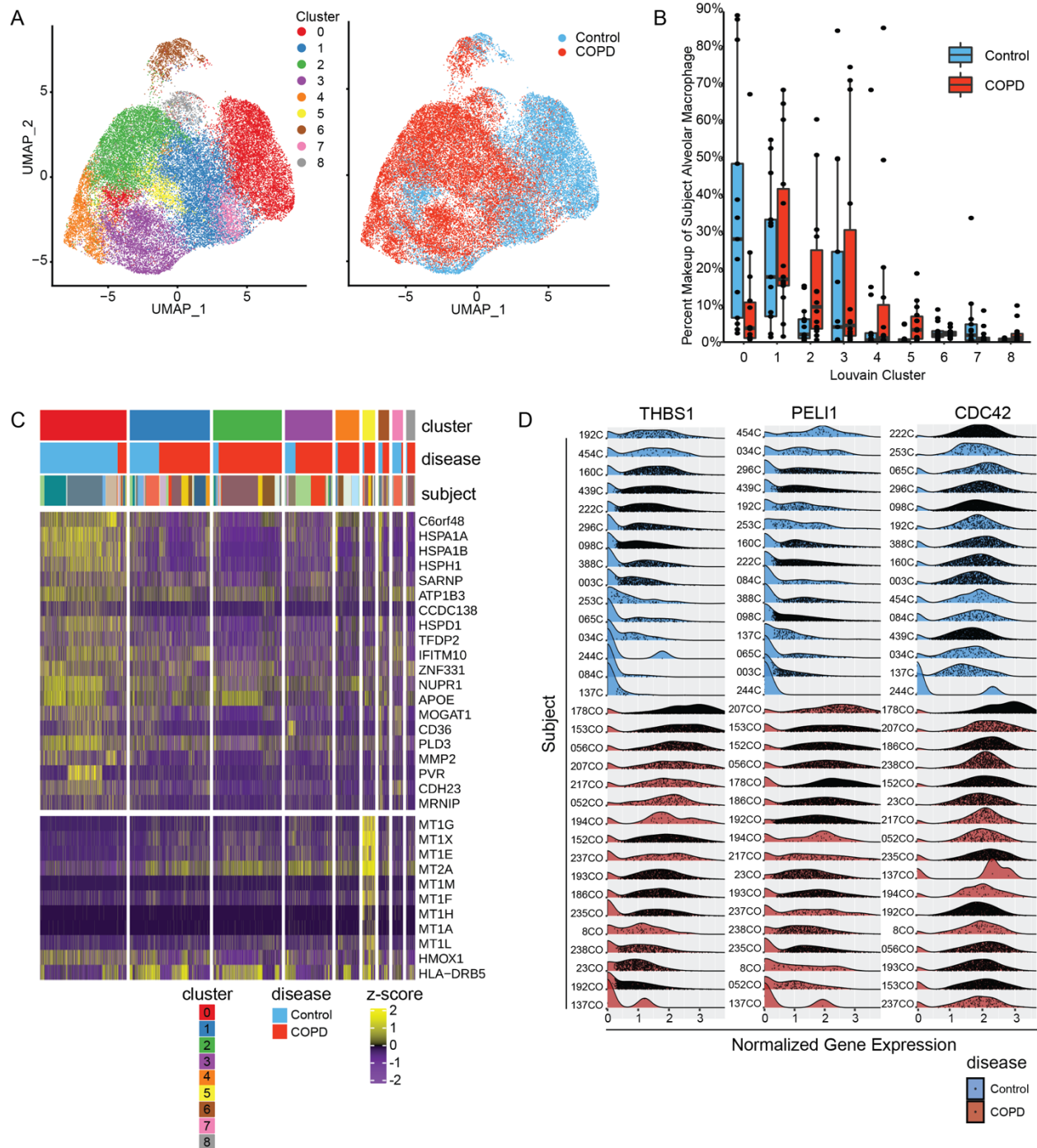


Figure 4 Changes in alveolar macrophage population composition in COPD. (A) UMAPs of control and COPD alveolar macrophage cells, colour labelled by Louvain cluster (*left*) and disease status (*right*). (B) Percent makeup of alveolar macrophage across all nine Louvain clusters per subject, grouped by disease state. Boxes represent interquartile ranges (IQRs); whiskers are $1.5 \times$ IQR; dots represent subjects. The percent of cluster 0 and cluster 5 alveolar macrophages are different between control and COPD (Wilcoxon rank sum test, FDR <0.05). (C) Heatmap of the distribution of z-scores of marker genes for cluster 0 and cluster 5 alveolar macrophages. Columns represent expression values from individual cells and are hierarchically ordered by macrophages cluster, disease status, and subject. (D) Density plots of *THBS1*, *PELI1* and *CDC42* gene expression across all alveolar macrophage cells, grouped by control (blue) and COPD (red) subjects. Within each disease group, subjects are ordered top to bottom from highest within-disease group expression average to lowest.

FIGURE 5

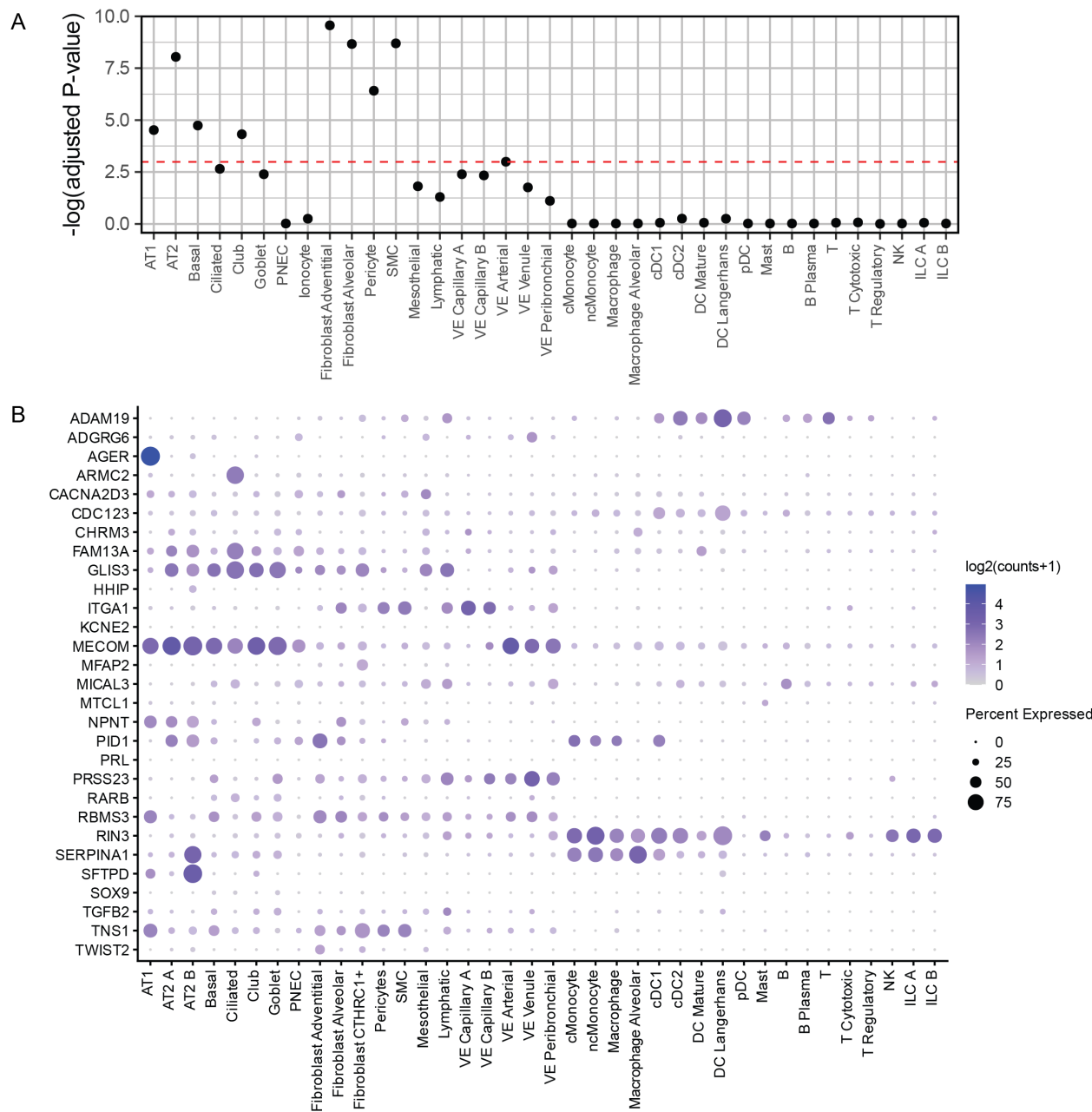


Figure 5 CELLECT analysis identifies cellular basis for COPD heritability. **(A)** Plot of negative log adjusted p-values for cell type-specific enrichment for GWAS-identified genes associated with lung function. Red line denotes threshold of significance (FDR < 0.05). Data showing high enrichment scores for non-hematopoietic cells are associated with COPD heritability traits. **(B)** Dot plot of z-scores for commonly associated GWAS genes that were highly ranked in CELLECT analysis. Dot size reflects percentage of cells with gene expression; colour corresponds to degree of expression.

FIGURE 6

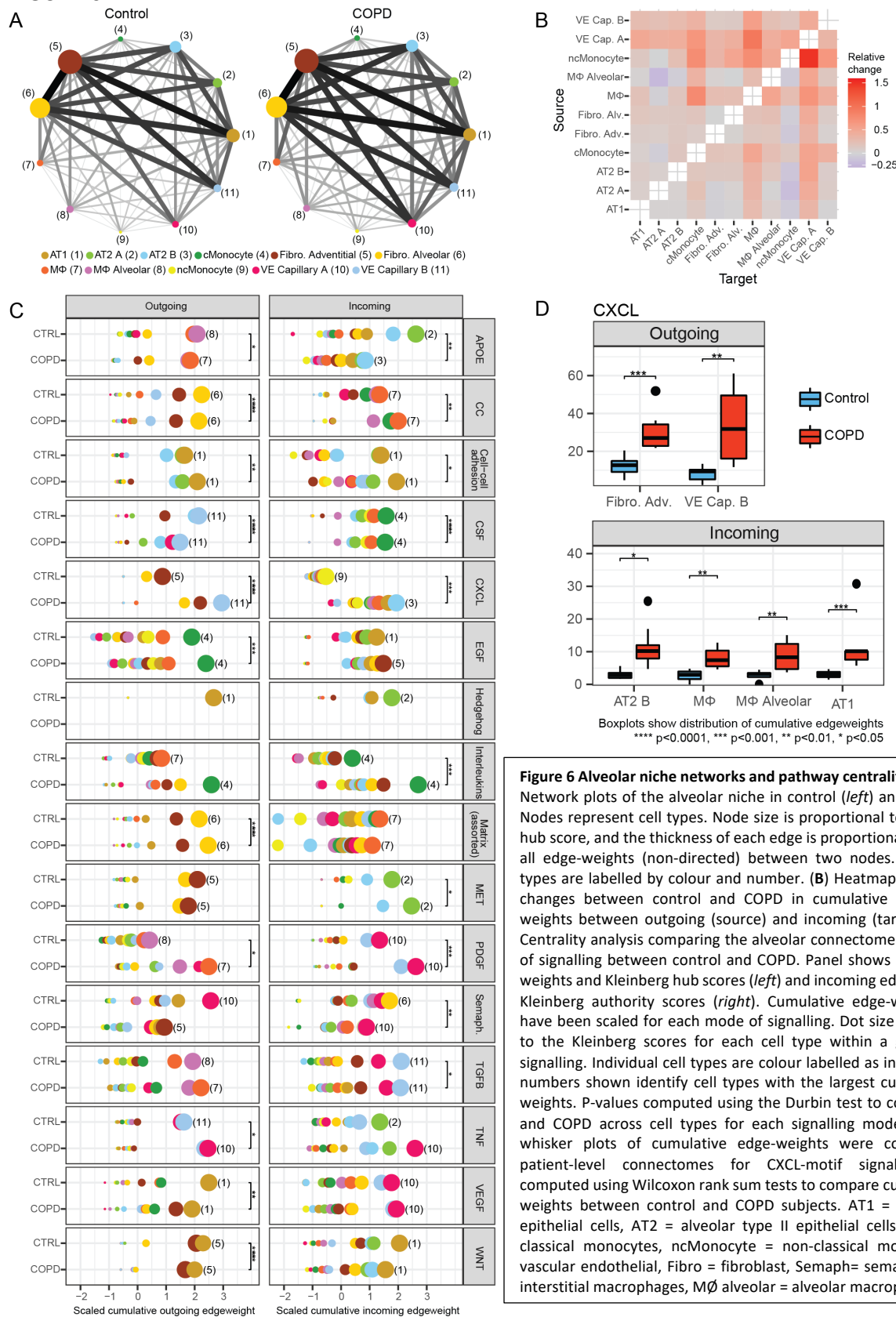


Figure 6 Alveolar niche networks and pathway centrality analyses. (A) Network plots of the alveolar niche in control (left) and COPD (right). Nodes represent cell types. Node size is proportional to the Kleinberg hub score, and the thickness of each edge is proportional to the sum of all edge-weights (non-directed) between two nodes. Individual cell types are labelled by colour and number. **(B)** Heatmap shows relative changes between control and COPD in cumulative directed edge-weights between outgoing (source) and incoming (target) nodes. **(C)** Centrality analysis comparing the alveolar connectome for each mode of signalling between control and COPD. Panel shows outgoing edge-weights and Kleinberg hub scores (left) and incoming edge-weights and Kleinberg authority scores (right). Cumulative edge-weights (x-axis) have been scaled for each mode of signalling. Dot size is proportional to the Kleinberg scores for each cell type within a given mode of signalling. Individual cell types are colour labelled as in Figure 6A, and numbers shown identify cell types with the largest cumulative edge-weights. P-values computed using the Durbin test to compare control and COPD across cell types for each signalling mode. **(D)** Box and whisker plots of cumulative edge-weights were computed using patient-level connectomes for CXCL-motif signalling. P-values computed using Wilcoxon rank sum tests to compare cumulative edge-weights between control and COPD subjects. AT1 = alveolar type I epithelial cells, AT2 = alveolar type II epithelial cells, cMonocyte = classical monocytes, ncMonocyte = non-classical monocytes, VE = vascular endothelial, Fibro = fibroblast, Semaph = semaphorins. MΦ = interstitial macrophages, MΦ alveolar = alveolar macrophage