Single-cell Transcriptome Analysis Reveals an Anomalous Epithelial Variation and Ectopic Inflammatory Response in Chronic Obstructive Pulmonary Disease

Naoaki Watanabe[1,2+], Jun Nakayama[1+], Yu Fujita[2]*, Yutaro Mori[3], Tsukasa Kadota[2], Iwao Shimomura[1], Takashi Ohtsuka[4], Koji Okamoto[3], Jun Araya[2], Kazuyoshi Kuwano[2], Yusuke Yamamoto[1]*

[1]Division of Cellular Signaling, National Cancer Center Research Institute, Tokyo, Japan.
[2]Division of Respiratory Diseases, Department of Internal Medicine, The Jikei University School of Medicine, Tokyo, Japan.
[3]Division of Cancer Differentiation, National Cancer Center Research Institute, Tokyo, Japan.
[4]Division of Thoracic Surgery, Department of Surgery, The Jikei University School of Medicine, Tokyo, Japan.

+These authors equally contributed to this work.

*Correspondence to:
Yusuke Yamamoto, National Cancer Center, 5-1-1 Tsukiji, Chuo-ku, Tokyo, 104-0045, Japan, Phone: (+81) 3-3542-2511 (Ext. 3664), Fax: (+81)3-3543-9305, E-mail: yuyamamo@ncc.go.jp

Yu Fujita, The Jikei University School of Medicine, 3-25-8 Nishi-Shimbashi, Minato-ku, Tokyo, 105-8461, Japan, Phone: 81-3-3433-1111 Ext.3271, Fax: 81-3-3433-1020, E-mail: yuugot@jikei.ac.jp

Competing Interests statement: The authors have declared that no conflict of interest exists.

Abstract

Chronic obstructive pulmonary disease (COPD) is a progressive disease that obstructs the airflow from the lungs, and tobacco smoking is the major cause of COPD. Here, we applied single-cell RNA sequencing to analyze COPD pathogenesis in COPD patients, non-COPD smokers and never-smokers and investigated the disease progression at single-cell resolution. By single-cell transcriptome analysis, COPD was characterized by shifts in the stromal, immune system and epithelial cell compositions. While epithelial components in never-smokers were relatively uniform, the smoker groups presented with extensive heterogeneity in epithelial cells, particularly in the alveolar type II (AT2) lineages. We identified a subpopulation of AT2 epithelial cells that emerged in smokers, such as COPD patients, and specifically expressed a series of chemokines and PD-L1. A trajectory analysis revealed that the inflammatory AT2 cell subpopulation followed a unique differentiation path, and a prediction model of cell-to-cell interactions inferred increased intercellular networks of inflammatory AT2 cells with immune and stromal cell populations. Thus, our analysis reveals a unique cellular differentiation pathway and function underlying the biological and clinical characteristics of COPD pathogenesis.

Introduction

Chronic obstructive pulmonary disease (COPD) is a chronic inflammatory disease mainly caused by long-term cigarette smoke (CS) exposure. It has emerged as the tenth leading cause of loss of years of life and the eighth leading cause of years lived with disability globally [1,2]. COPD is characterized by persistent airflow limitation, which is progressive even after smoking cessation. The inflammatory process of COPD is predominantly localized to the small airways and surrounding lung parenchyma with concomitant systemic inflammation [3]. Phenotypic alterations in the lung epithelium caused by CS exposure have been widely implicated in COPD pathogenesis [4], but the precise mechanisms orchestrating this persistent inflammatory process remain unknown.

Alveolar type 2 (AT2) cells are multifunctional and polarized epithelial cells spatially restricted to distal adult human lungs. AT2 cells play an important role as progenitors, such as self-renewal and differentiation into alveolar type 1 (AT1) cells, rendering AT2 crucial to the maintenance of alveolar homeostasis [5]. AT2 dysfunction has been implicated in the pathogenesis of a variety of parenchymal lung diseases, including COPD, acute respiratory distress syndrome (ARDS), and idiopathic pulmonary fibrosis (IPF) [4, 6, 7]. However, despite recent improvements in cell yields and purities, the functions of AT2 cells in both physiological and pathological settings have not been elucidated due to the difficulties in culturing and expanding primary AT2 cells, which maintain their original characteristics in lung tissues. Therefore, little is known about not only AT2 cell type-specific mechanisms but also cell-cell interactions among the multiple lung resident cell types and inflammatory cells that contribute to disease progression during COPD pathogenesis.

The development of high-throughput single-cell RNA sequencing (scRNA-seq) has provided a new research tool to obtain gene expression profiles at single-cell resolution and identify novel cellular phenotypes in lung diseases, especially IPF [8-11]. Although some scRNA-seq studies have revealed the diversity of aberrant lung resident cell landscapes in COPD lungs [11,12], the AT2 cell heterogeneity and cell type-specific mechanisms of COPD progression have not been fully investigated. To improve our understanding of the cell-specific mechanisms contributing to COPD pathogenesis, we used scRNA-seq to reveal the mechanism of COPD pathogenesis and elucidate differences in the proportions and transcriptional phenotypes of epithelial cells between COPD patients, non-COPD smokers, and never-smoker controls. Notably, we identified inflamed epithelial populations, including AT2 and the basal lineage. A novel AT2 cell subset, which harbors inflammatory features, differentiates from a novel AT2 cell lineage. Furthermore, analysis of intercellular communications revealed that this AT2 cell subset had the strongest interactions with immune cells, and inflammatory signaling in COPD patients was remarkably higher than that in never-smoker controls. We generated novel insights into phenotypic changes in epithelial cells and altered communication patterns between the AT2 subset and immune cells in COPD pathogenesis.

Results

Single-cell composition in the lungs of patients with COPD.

We initially performed single-cell transcriptome profiling from human lung tissues of 6 COPD, 3 non-COPD smoker, and 3 never-smoker samples (**Fig. 1A**). COPD patients were diagnosed with the Global Initiative for Chronic Obstructive Lung Disease criteria, and their lung tissues were obtained from pneumonectomy or lobectomy specimens of primary lung cancer. Pathohistological observation clearly showed morphological differences between COPD, non-COPD smokers, and never-smokers (**Fig. 1B**). Patient demographics and clinical data are shown in **Fig. 1C** and **Supplementary Table 1**. The scRNA-seq information of each sample is summarized in **Supplementary Table 2**. COPD, non-COPD smoker, and never-smoker samples yielded an average of 5,226 cells, 4,568 cells, and 4,232 cells per donor, respectively (a total of 57,755 single cells). In the process of quality control with the Seurat R package, the sequencing data of the estimated dead cells and doublet cells were computationally excluded, and then all data from 12 samples (6 COPD, 3 non-COPD smoker, and 3 never-smoker samples) were combined for further analysis.

Unsupervised clustering followed by t-distributed stochastic neighbor embedding (t-SNE) revealed 17 distinct molecular clusters (**Fig. 1D and Supplementary Table 3**). A similar result was obtained by clustering with two-dimensional uniform manifold approximation and projection (UMAP) (**Extended Data Fig. 1A**), although principal component analysis (PCA) mapping failed to appropriately segregate the clusters (**Extended Data Fig. 1B**). According to the gene expression profiles, this clustering analysis identified 7 clusters of immune-type cells, such as T cells (marked by CD3D, CD2, and CD69), dendritic cells (TGFBI and HLA-DMB), and mast cells (KIT and TPSAB2); 5 clusters of epithelial lineages, such as alveolar type 2 cells (SFTPD, SFTA2, and ABCA3), club cells (BIRC5, KRT19, and CXCL17) and ciliated cells (CAPS, RSPH1, and TSPAN11); and stromal-type cells, such as fibroblasts (COL1A2 and COL6A2) and endothelial cells (CALCRL and RAMP2) (**Fig. 1D**). When remapping single cells with the patient states (COPD, non-COPD smoker, and never-smoker), although the variation in cell distribution was evident, all clusters contained cells from three patient states (**Fig. 1E**). The number of cells in each cluster varied among patients and states, and most clusters were distinctly represented in each patient, as summarized in **Fig. 1F**.

Clustering with t-SNE represents the differences in gene expression profiles in the clusters. Thus, we further investigated cluster-specific gene signatures and selected the top 10 specifically expressed genes in each of the clusters (**Fig. 1G**). The heatmaps displayed distinct expression patterns of cluster-specific gene signatures (**Fig. 1G**), and the expression of selected genes plotted in the t-SNE map confirmed the cell types of epithelial, immune, and stromal clusters (**Extended Data Fig. 1C**). For example, FoxJ1, a ciliated cell marker, was specifically expressed in the ciliated cell cluster, and MARCO, a macrophage marker, was specifically expressed in the macrophage cluster. Thus, these

data reveal that our single-cell transcriptome successfully represented the cellular clusters in COPD, non-COPD smokers and never-smokers.

Differential cellular distribution of the immune and stromal cell components in COPD lungs. Given the importance of immune cell populations in COPD pathogenesis, we initially focused on single-cell transcriptome profiling of immune cells in patients with COPD. As the largest cell cluster was T cells (**Fig. 1D**), our dataset comprised a number of immune cell-related clusters, representing 62.0% (35,816 cells) of human lung cells. To scrutinize the immune cell population, single-cell data of immune clusters based on t-SNE mapping (**Fig. 1D**) were selected, and a t-SNE plot with all immune cells was redeveloped (**Extended Data Fig. 2A**). After contaminated, unrelated cell data were removed based on their gene expression profiles, t-SNE visualization of immune cells revealed 14 clusters (**Extended Data Fig. 2A**). Similar to the t-SNE mapping of all lung cells (**Fig. 1E**), the t-SNE plot of immune-related cells revealed the overall proportion of immune cells among the patient states (**Extended Data Fig. 2B-D**).

In 14 immune cell clusters, we found two discrete cell populations for each neutrophils and mast cells and three cell populations for macrophages (**Extended Data Fig. 2A and D**). Neutrophil-1 and -2 cells were both characterized by high expression of FCN1, S100A8, and LILRA5. Neutrophil-1 specifically expressed S100A12, and in contrast, neutrophil-2 expressed CDKN1C (**Extended Data Fig. 2E**). Macrophage receptors with collagenous structures (MARCOs) were highly expressed in all macrophage-1, -2, and -3 clusters (**Extended Data Fig. 2F**). INHBA was highly expressed in macrophage-1 and -3 but not in macrophage-2. TK1 was specifically expressed in macrophage-3. Both mast cell-1 and -2 cells were defined as high expression of TPSB2 and RGS10, but CEBPD expression was restricted in mast cell-2 (**Extended Data Fig. 2G**). Comparative analysis of cell number in the clusters among the patient states revealed that NK cell and myeloid DC numbers were significantly decreased in COPD and non-COPD smoker samples (**Extended Data Fig. 2H**). In the clustering of scRNA-seq datasets, we could not distinguish Treg cells from the $CD4^+$ T cell cluster; however, the t-SNE plot with Foxp3 expression strikingly showed a small population of Tregs in the CD4 T cell cluster (**Extended Data Fig. 2I**).

Next, we sought to investigate stromal cell populations in COPD lungs. The t-SNE plot of all lung cells contained four stromal cell clusters, including fibroblast, endothelium, activated endothelium and smooth muscle, representing 23.1% (13,346 cells) **(Fig. 1D)**. For further analysis of stromal cells at single-cell resolution, stromal cell clusters from the t-SNE plot of all lung cells were selected, and a t-SNE plot with stromal cells was redeveloped in the same manner as the immune cell analysis. Stromal cells were reclassified into 11 discrete clusters: capillary EC 1-3, vein EC, artery EC, lymphatic EC, fibroblasts 1-3, and smooth muscle 1 and 2 (**Extended Data Fig. 3A**). These stromal cell clusters were annotated among the patient states, and distinct cell populations

were investigated (**Extended Data Fig. 3B-D**).

In the t-SNE plot of stromal cells, we found three distinct cell populations in fibroblasts and two cell populations in smooth muscle (**Extended Data Fig. 3A and D**). Fibroblasts 1-3 all expressed general fibroblast markers such as DCN and PDGFRA. A difference in the expression levels of a small number of genes defined the fibroblast clusters. For example, ITGA8 was highly expressed in fibroblast-1, and GPNMB was specifically expressed in fibroblast-3 (**Extended Data Fig. 3E**). In the clusters of smooth muscle-1 and -2, ACTA2 was highly expressed in both; NOTCH3 was detected in only smooth muscle-1, and DES was expressed in only the smooth muscle-2 cluster (**Extended Data Fig. 3F**). Comparative analysis of cell number in the clusters among patient states revealed that fibroblast-1 and capillary EC 1 were significantly decreased; on the other hand, the cell number in the vein EC cluster was increased in COPD and non-COPD smoker samples (**Extended Data Fig. 3G**).

Our scRNA-seq analysis of both immune and stromal cells apparently showed cellular states in COPD, non-COPD smokers and never-smokers and found a difference in cell populations between the COPD and never-smoker samples. However, our current analysis failed to show evident differences in the cell populations between COPD and non-COPD smoker samples.

Precise and extensive classification of epithelial cell lineages in COPD lungs.

A total of 8,047 single epithelial cells were clustered in a t-SNE plot in the same manner as the immune and stromal cell lineages (**Fig. 2A**). The t-SNE plot with epithelial lineages revealed 18 clusters, including three alveolar type 1 (AT1) cell clusters, seven alveolar type 2 (AT2) cell clusters, four ciliated cell clusters, three club cell clusters, and one basal cell cluster. In our epithelial cell datasets, we could not find goblet cells, which were mainly located in the upper airway and were not in the lower airway. Few neuroendocrine cells, which are marked with CHGA, ASCL1, and INSM1, were found in the cluster of club cells (**Extended Data Fig. 4A**). Presumably due to too few cells, they did not form a unique cluster. Additionally, FOXI1, an ionocyte marker, could not mark a unique cell cluster (**Extended Data Fig. 4B**). We did not identify specific clusters of tuft cells, whose markers were distributed in multiple clusters (**Extended Data Fig. 4C**). The t-SNE plot marked with the patient states revealed nonhomogeneous distribution of multiple epithelial cell lineages, e.g., the AT2-1 and AT2-5 clusters (**Fig. 2B**). The density plot of t-SNE mapping showed that a large number of cells were present in the AT2-2 and AT1-1 clusters (**Fig. 2C**). As expected, the cell numbers of AT1 and AT2 clusters in COPD samples tended to decrease compared with those in non-COPD samples (**Extended Data Fig. 4D, E, and Supplementary Table 4**). For each of the clusters, cell populations among the patient states are summarized in **Fig. 2D and E**, which exhibit remarkable differences in epithelial cell composition.

We identified seven epithelial cell clusters of AT2 cells. SFTPC, a typical marker for AT2

cells, was strongly expressed in all seven AT2 clusters; however, subtle variations in gene expression among the AT2 clusters were observed (**Extended Data Fig. 4F**). For example, AT2-4 cluster cells coexpressed SFTPC and AGER, which are also known as AT1 cell markers, presumably indicating that AT2-4 is an intermediate between AT1 and AT2 (**Extended Data Fig. 4G**). Likewise, variations in gene expression were also found in epithelial lineages, such as Club cells, although typical markers were commonly detected (**Extended Data Fig. 4H**). These data suggest epithelial cell diversity in human lung cells. The top 10 differentially expressed genes in each of the clusters are summarized (**Fig. 2F and G**) and exhibit the diversity of epithelial lineages.

As our single-cell transcriptome profiling provided high-resolution expression data, we explored COPD genome-wide association study (GWAS) gene expression from the epithelial scRNA-seq data to show the broad contribution of airway epithelial cell types to COPD susceptibility. As a first step, the expression levels of 82 COPD GWAS genes were examined in all lung cells **(Extended Data Fig. 4I)** [13], and COPD GWAS genes were cell type-selectively expressed. Then, we investigated COPD GWAS gene expression in the 18 epithelial cell clusters (**Fig. 2H**). High expression of COPD GWAS genes was mainly observed in the AT1, AT2 and ciliated cell clusters. The expression of AGER, which mediates responses to cell injury induced by a variety of stimuli, including oxidative stress, aging and hypoxia, was found in AT1 clusters and was presumably associated with the development of COPD. Likewise, SFTPD is involved in surfactant metabolism and plays a role in protecting the lungs against inhaled microorganisms and chemicals. A SFTPD mutation expressed in AT2 cells is related to COPD pathogenesis.

Unique epithelial cell populations and an inflammatory AT cluster in the COPD lung.

We next sought to explore the unique or increased cell populations in the COPD lung epithelia. The cell populations of the AT1 subtypes, AT2 subtypes and basal cell lineages were separately compared among COPD, non-COPD smoker and never-smoker samples (**Fig. 3A**). In the AT1 subtypes, the AT1-1 cluster is the major cell cluster in COPD (47.5% in AT1 subtypes of COPD); however, this cluster represents a small population in never-smokers (18.6% in the AT1 subtypes of never-smokers). In contrast, the AT1-2 clusters, which specifically expressed COL4A2 and COL4A3 (**Extended Data Fig. 5A**), decreased during progression (never-smoker: 56.4%; non-COPD smoker: 41.0%; COPD: 18.3%) (**Fig. 3A, top panel**). In the AT2 subtypes, the cell population of the AT2-1 cluster (marked with CPB2, CD36 and HHIP, **Extended Data Fig. 5A**) obviously decreased in the smoker groups (COPD and non-COPD smoker) (**Fig. 3A, middle panel**). Interestingly, the AT2-5 and AT2-6 clusters appeared in the smoker groups (COPD: AT2-5, 23.6% and AT2-6, 12.3%; non-COPD smoker: AT2-5, 30.9% and AT2-6, 15.2%; never-smoker: AT2-5, 0.33% and AT2-6, 3.21%;). In basal cell lineages, the cell populations of club cell-3 (marked with TPSB2, CD69 and PTPRC) and basal cell (marked with KRT17, KRT5, and TP63) clusters were

higher in COPD samples than in never-smoker samples (**Fig. 3A, bottom panel, and Extended Data Fig. 5A**). Furthermore, we found that the percentage of the subpopulations in AT1-2 and AT2-1 were significantly lower in the COPD groups than in the other groups (**Fig. 3B, upper panels**). Although not statistically significant, we observed that the percentage of the club-3, basal cell and AT2-5 clusters were obviously higher in the COPD group than in the never-smoker group (**Fig. 3B, lower panels, and 3C**). The cell distribution of these subtype clusters was also confirmed in the t-SNE plot of epithelial lineages (**Extended Data Fig. 5B**).

Because the cells located in the AT2-5 cluster were hardly detected in the never-smoker group, we further scrutinized the cell population and characterized the gene expression profile of the AT2-5 cluster in the smoker groups. As shown in **Fig. 3C**, the percentage of AT2-5 cluster cells in the total cell number of human lung tissues was very low; even in the COPD group, it was approximately 0.83%. Notably, a number of inflammation-related genes, including CD274 (PD-L1), CXCL1, and CCL2, were specifically expressed in the AT2-5 cluster, and we designated the population inflammatory AT2 (iAT2) (**Extended Data Fig. 6A**). Additionally, we identified novel markers to classify the AT2-5 cluster, such as CLDN1 and TNC (**Extended Data Fig. 6B**). The correlation coefficients revealed a correlation of these inflammation-related gene expression in the AT2-5, although it's not so strong but most of them are positive (**Fig. 3D**). CXCL1 and CXCL8, which are closely related to COPD pathogenesis, presented a moderate correlation; CD274 (PD-L1) was also synchronously expressed with genes such as CCL2 and TNC (**Fig. 3E**). The specific expression of the inflammatory genes and the markers that we identified were confirmed with immunofluorescence (**Fig. 3F**). CD274 was not detected in the AT2 cells of the never-smoker group; however, very few ABCA3-positive AT2 cells in the COPD groups coexpressed CD274. Likewise, we found coexpression of CXCL1 and SFTPC, CCL20 and ABCA3, CCL2 and SFTPC, TNC and ABCA3, and CXCL8 and SFTPC in COPD lung alveolar tissues (**Fig. 3F, and Extended Data Fig. 7A**). In addition, AT1-3 and club cell-3 cells, which were predominantly detected in the COPD lungs, also possessed high inflammatory signals (**Extended Data Fig. 7B, C**).

Distinct differentiation program and increased cellular variation in the COPD epithelia.

To investigate how the AT2-5 cluster appears in the COPD lung, we sought to dissect the inferred differentiation trajectories of epithelial cell lineages among the COPD, non-COPD smoker and never-smoker groups. For this purpose, we performed pseudotime analyses for alveolar (**Fig. 4A-C**) and basal lineages (**Extended Data Fig. 8A-C**), separately. Because AT2 cells generally serve as the progenitors of AT1 cells, the plot for alveolar differentiation revealed a trajectory from AT2 to AT1 clusters (**Fig. 4A, right panel**). In the smoker groups (COPD and non-COPD smokers), the AT2-5 cluster was located at the end of the AT2 clusters (opposite the AT1 cluster), suggesting that the inflammatory AT2-5 cluster did not emerge during the differentiation from AT2 to AT1 but followed

a different differentiation process (**Fig. 4C**). In basal lineages, we observed bidirectional differentiation into either ciliated cell clusters or club cell clusters in all groups (**Extended Data Fig. 8D**), although the cell composition differed among the disease statuses, e.g., increased populations of club cell-3 and basal cells in the COPD group (**Extended Data Fig. 8C, and Fig. 3B, lower panels**).

To understand the biological processes contributing to the pseudotime components, we analyzed gene expression that cooperatively changed with pseudotime. From the marker genes selected by the Seurat program, significantly altered gene groups (729 genes for alveolar clusters and 742 genes for basal clusters, false discovery rate (FDR) < 1%) were identified in the processes of alveolar and basal lineage differentiation (**Fig. 4D, and Extended Data Fig. 8E**). In line with the distribution of the clusters, the AT2 cell markers, such as CD74, were expressed in the predicted early phase, while the CDKN2B, KRT7 and AT1 cell markers were highly detected in the late phase (**Fig. 4D**). Likewise, we observed that basal cell markers, such as TP63, were expressed early and that ciliated cell markers, such as FOXJ1 and the DNAH family members, were expressed late (**Extended Data Fig. 8E**).

We further investigated cluster-specific gene signatures by performing pathway enrichment analysis based on the Reactome database for alveolar and basal lineages (**Fig. 4E, and Extended Data Fig. 8F**). In **Fig. 4E**, we found clear differences in the pathway enrichment between the AT1 and AT2 clusters. While three AT1 clusters exhibited relatively uniform patterns of pathway enrichment, more heterogeneity was observed within the AT2 clusters. In addition, diversity in pathway enrichment was also found among the patient states (COPD, non-COPD smoker, and never-smoker). For example, in the AT2-5 clusters, “L1 CAM interactions” were enriched in COPD and non-COPD smokers but not in never-smokers; “MAP kinase activation” was detected in only AT2-5 (**Fig. 4E**). Next, to clarify the cellular diversity within the AT2 clusters, the variations in cell localization on the t-SNE plot were calculated. As shown in **Fig. 4F**, the cells in all AT2 clusters showed a large variation in the COPD group. Among the seven clusters, AT2-1, AT2-2, AT2-3, AT2-5, and AT2-7 exhibited the largest variation in the COPD group, suggesting that COPD pathogenesis increased epithelial cell diversity in AT2 cells. Although not as pronounced as in alveolar lineages, similar trends were also observed in the basal cell lineages; the pathway enrichment in basal cells showed a number of enriched pathways in only the COPD and non-COPD smoker groups (**Extended Data Fig. 8F**).

Anomalous cell-to-cell networks in the COPD epithelial clusters.

Because our data demonstrated the increased cellular variation in AT2 clusters in COPD patients, we examined the closeness centrality of whole epithelial cells among COPD, non-COPD smokers and never-smokers (**Fig. 5A**). Based on the gene expression profiles of each epithelial cell, closeness

centrality was calculated as the reciprocal of the sum of the length of the shortest paths between each of two cells (**Fig. 5B**). The data revealed a decreased closeness centrality in the COPD epithelial cells than in the non-COPD smoker and never-smoker cells, suggesting increased diversity in the COPD epithelia.

As COPD epithelial cells exhibited increased variations in expression profiles and ectopic inflammatory expression, we explored receptor–ligand pairing between epithelial cells and immune cells in both COPD patients and never-smokers. To this end, we focused on immunological responses, extracted chemokines and inflammatory cytokines from the Cell-Cell Interaction Database (https://baderlab.org/), and established an immune-response network between epithelial and immune cells. The interaction scores were calculated by the expression levels of receptors and ligands and the positive ratio in each cluster (**Methods**). We identified 1,789 predicted interactions (509 pairs between ligands in the epithelial cluster and receptors in the immune cluster and 1280 pairs between ligands in the immune cluster and receptors in the epithelial cluster) in COPD patients (**Fig. 5C**) and 1,021 predicted interactions (357 pairs between ligands in the epithelial cluster and receptors in the immune cluster and 664 pairs between ligands in the immune cluster and receptors in the epithelial cluster) in never-smokers (**Extended Data Fig. 9A**). In the interaction of epithelial ligands with immune receptors, our data suggested that the club cell-3 cluster, whose population was increased in COPD, considerably influenced the immune cell clusters (**Fig. 5D, left panel**). Among the AT2 clusters, the AT2-5 cluster exerted the strongest effect on immune cells (**Fig. 5C**); the effect in COPD patients was remarkably higher than that in never-smokers (**Fig. 5D, left panel**). On the other hand, in the interaction of immune ligands with epithelial receptors, several clusters, such as NK cells, myeloid dendritic cells (mDCs), $CD8^+$ T cells and $CD4^+$ T cells, exhibited stronger interactions with epithelial clusters in COPD patients than never-smokers (**Fig. 5C and D, right panel**). While many interactions were detected in both groups, some were specific to the COPD or never-smoker groups (**Supplementary Table 5**). In addition, when focusing on the clusters of club cell-3, AT2-5 and AT1-3, which showed obvious differences in the interaction between epithelial ligand and immune receptor pairs, the data apparently showed similarities and differences in their interaction atlas in COPD patients and never-smokers (**Fig. 5E**).

Then, we also examined a cell-to-cell interaction with the inflammatory cytokine network between epithelial and stromal cells in COPD and never-smokers (**Extended Data Fig. 9B, C)**. Analysis of the predicted cell-to-cell interactions between epithelial and stromal cells, such as fibroblasts and endothelial cells, revealed distinct cell-to-cell interaction patterns between the COPD and never-smoker groups. Similar to immune cells, the AT2-5 and club cell-3 clusters were predicted to tightly interact with stromal cells (**Extended Data Fig. 9B, C)**. Collectively, epithelial cells, particularly the AT2-5 and club cell-3 clusters in COPD lungs, have increased predicted interactions with immune and stromal cells.

Inflammatory response in COPD basal cells.

The number of cells in the basal cell clusters increased in only the COPD lungs but not in the non-COPD smoker or never-smoker lungs (**Fig. 3B**). Given the importance of basal cells in tissue regeneration, we sought to scrutinize individual cells in the basal cell cluster among COPD patients, non-COPD smokers and never-smokers. Clustering analysis of differentially expressed genes (COPD patients vs. non-COPD smokers and never-smokers, 22 genes 1.5-fold or greater and $p < 0.05$) showed a rough segregation between COPD patients and others (**Fig. 6A**). Although separation among the patient states was not obvious in the t-SNE plot of basal cells (**Fig. 6B**), we found that typical basal markers, KRT5, KRT19 and S100A2, were highly expressed in COPD basal cells, while the expression levels of an AT2 cell marker, SFTPC, and a club cell marker, SCGB1A1 (also known as CC10), were low in COPD lungs (**Fig. 6C**). Because we reasoned that basal cells in COPD are likely to be more immature, we examined the cellular interactions between basal cells and fibroblasts, which are well known to interact with each other [14]. The analysis showed that Notch signaling was more frequently detected in basal cells from non-COPD samples than from COPD samples (**Extended Data Fig. 9C**). As Notch signaling is important for differentiation in the airway, it might be one of the possible explanations that COPD lungs contain a greater number of basal cells.

To further investigate the functional properties of basal cells in COPD, we enzymatically isolated small airway epithelial cells from surgical samples of COPD or never-smoker lungs and expanded the cells *in vitro*. The patient information is shown in **Supplementary Table 6**. Morphologically, small airway epithelial cells from COPD or never-smokers were similar (**Fig. 6D**), and the expression patterns of basal cell markers such as Krt5 and p63 were also similar (**Fig. 6E**). Total RNA was extracted from early passaged cells (less than passage 3), and bulk RNA-seq was performed to characterize these cells in detail. Although PCA mapping did not separate COPD and never-smoker samples (**Extended Data Fig. 10A**), 762 genes were differentially expressed between them (1.5-fold or greater and $p < 0.05$, **Fig. 6F and Supplementary Table 7**) and included immune-related and basal/stem-related genes (**Extended Data Fig. 10B, C**). Gene set enrichment analysis (GSEA) in basal cells between COPD and never-smokers revealed that gene sets such as the IFNγ response, PI3K/AKT/mTOR signaling and positive regulation of epithelial cell proliferation were significantly enriched in COPD basal cells (**Fig. 6G, Extended Data Fig. 10D)**. Ingenuity pathway analysis (IPA) clearly demonstrated activated immune pathways in COPD basal cells. The top 20 canonical pathways included several immune-related pathways, such as the Th2 and TLR signaling pathways (**Extended Data Fig. 10E and Supplementary Table 8).** IL13 and IL33 are well known to be associated with COPD pathogenesis [15], and analysis of upstream signals in IPA revealed that genes regulated by IL13 and IL33 were significantly enriched in COPD basal cells (**Fig. 6H**). Consistent with the cell-cell interaction analysis (**Extended Data Fig. 9C**), Notch signaling

was significantly suppressed in COPD basal cells (**Fig. 6H**). Additionally, several other immune-related signals were positively regulated in COPD basal cells (**Fig. 6H, I)**. These findings suggest that the COPD basal cells retained the inflammatory response, even after a few passages in culture, and it might cause basal cell expansion *in vivo* and give rise to inflammatory lineages such as club cell-3.

Discussion

Chronic inflammation has been widely implicated in airway remodeling and emphysema development during COPD pathogenesis [16]. This comprehensive single-cell transcriptome profiling of human lung tissues elucidates the presence of novel epithelial cell populations that are specifically responsible for inflammatory processes in COPD lungs. We identified 17 clusters of resident cell landscapes, including immune-type cells, epithelial lineages, and stromal-type cells, such as fibroblasts and endothelial cells. Based on further extensive classification, differential cellular distributions of the immune and stromal cell components were detected among the patient groups. Among immune-type cells, significant decreases in NK cell and mDC numbers were demonstrated in COPD and non-COPD smoker lungs. Although stromal-type cells also show significantly different cell numbers in decreased fibroblast-1 expressing ITGA8 and capillary EC 1 but increased vein EC in COPD and non-COPD smoker lungs, both immune and stromal-type cells failed to show clear difference in cellular distribution between COPD and non-COPD smoker lungs, indicating that those alterations can be mainly attributed to CS exposure but may not directly reflect COPD pathogenesis. Extensive classification of epithelial lineages identified 18 clusters, including three AT1 cell clusters, seven AT2 cell clusters, four ciliated cell clusters, three club cell clusters, and one basal cell cluster, and elucidated not only the diversity of epithelial lineages but also marked differences in cell composition between COPD and non-COPD smoker lungs. In AT1 clusters, the AT1-2 cluster specifically expressing COL4A2 and COL4A3 was decreased along with COPD progression, but the AT1-3 cluster with high inflammatory signals was increased in COPD lungs. Among the AT2 clusters, AT2-1, AT2-2, AT2-3, AT2-5, and AT2-7 exhibited the largest variation in the COPD group by means of pathway enrichment analysis, suggesting the participation of AT2 diversity in COPD pathogenesis. Both the AT2-5 and AT2-6 clusters were present in the smoker groups, and the highest percentage of the AT2-5 cluster was demonstrated in the COPD group. The AT2-5 cluster represents inflammatory AT2 (iAT2) cells expressing inflammation-related genes accompanied by bidirectional interactions with immune cells. The physiological participation of iAT2 cells in COPD pathogenesis was supported by the existence of AT2-5 cells in COPD lungs by detecting cell type-specific markers, including CD274 (PD-L1), in AT2 cells. In the basal cell lineage, increased populations of club cell-3 and basal cell clusters were detected specifically in COPD samples. Club cell-3 also showed increased bidirectional interactions with immune cells. COPD basal cells demonstrated activated immune pathways, such as the Th2 pathway and TLR

signaling. Taken together, these inflamed populations of airway and alveolar epithelial cells can be involved in the mechanisms for the development of COPD lesions by exaggerating the inflammatory process.

Although COPD is the representative inflammatory lung disease induced by CS inhalation, the exact mechanism for persistent activation of lung inflammatory pathways even after smoking cessation remains elusive [17]. Heterogeneity in airway inflammatory patterns has also been demonstrated in COPD lungs, which contain neutrophils associated with inflammasome activation, Th1- and Th17-mediated immunity, and Th2-mediated immunity [16]. During airway remodeling in COPD development, the airway epithelium undergoes profound changes in the number, proportion of cell types, and differentiation profiles [18,19]. Furthermore, phenotypic alterations to cellular senescence in both the airway and alveoli have been implicated in COPD pathogenesis through excessive cytokine production of the senescence-associated secretory phenotype (SASP) [20-23]. Accordingly, heterogeneity in epithelial phenotypic alterations with respect to inflammation may have a pivotal role in regulating not only CS-mediated transient responses but also the persistent nature of inflammatory processes in COPD lungs. Our comprehensive scRNA-seq analysis, especially focusing on epithelial phenotypic alterations, uncovered the large variation in gene expression profiles in COPD lungs (**Fig. 4F, 5A**) and the existence of inflamed airway and alveolar epithelial cell clusters, which are capable of mediating inflammatory processes through bidirectional interactions with immune cells, including macrophages, monocytes, $CD4^+$ T cells, $CD8^+$ T cells, neutrophils, and NK cells (**Fig. 5C, D, E**). The cell-cell interactions and chemokine expression may partly explain the role of epithelial cells in CS-mediated activation of inflammatory signaling with a concomitant increased recruitment of neutrophils, $CD4^+$ and $CD8^+$ T cells into the lungs [17]. The GWAS gene indicates a genetic component affecting disease susceptibility and severity[13], and high expression of COPD GWAS genes was demonstrated in the epithelial clusters (**Fig. 2H**), indicating that lung epithelial cells are not only the first line of defense against CS exposure but also have a regulatory role in COPD pathogenesis. We obtained lung samples from lung cancer patients who had all ceased smoking for at least two months before the surgery to ensure a safe operation. Hence, it is likely that those inflamed epithelial clusters do not reflect the simple effect of CS exposure but rather intrinsic changes in phenotype. In addition, bulk RNA-seq analysis showed that both the proliferative properties and activated immune-related pathways were retained in basal cells isolated from COPD lungs even after a few passages in culture, further supporting the notion that inflamed epithelial populations do not transiently change in response to CS exposure but may commit to persistent inflammation after smoking cessation in COPD pathogenesis.

AT2 has been recognized to play an essential role in maintaining alveolar homeostasis by not only secreting surfactant but also behaving as a tissue resident progenitor cell for AT1 repopulation in response to lung injury. The pathogenic implication of alveolar epithelial cells in

COPD development is not entirely clear, and the cells have been recognized to be passive sufferers of repetitive smoking stress with respect to cumulative cell loss mainly via regulated cell death by apoptosis, necroptosis, and ferroptosis[24-26]. Furthermore, exaggerated loss of alveolar epithelial cells cannot be compensated for by an increase in proliferation of AT2 in COPD lungs, which is attributed to loss of stemness through enhanced cellular senescence induced by repetitive smoking stress [21], resulting in lung tissue destruction from emphysema development. Although recent advances in scRNA-seq technology have uncovered potential heterogeneity of the AT2 population in terms of tolerability to cellular stress [27], we first identified iAT2 (AT2-5) cells with increased chemokine expression, represented by CXCL8, CXCL1, CCL2, and CCL20, which are involved in neutrophilic, lymphocytic, and monocytic inflammation. Cell-cell interaction analysis in terms of the immune-response network clarified the strong bidirectional interaction between iAT2 and immune cells, including neutrophils, $CD8^+$ T cells, and monocytes (**Fig. 5C, D, E**). In addition, iAT2 cells showed increased expression of NFKB2, a component of NF-κB, and CSF3, an IL-6 superfamily member, further supporting the inflammatory nature of AT2 (**Extended Data Fig. 6**). In contrast, the fraction of AT2-1 clusters without an inflammatory phenotype was significantly decreased in COPD patients but occupied a major part of the AT2 population in never smokers, indicating that AT2-1 is involved in a physiological role in normal lungs. Importantly, the increased existence of iAT2 cells was confirmed by immunostaining analysis of COPD lungs by cell type-specific markers, including CD274 (PD-L1), CXCL1, and CCL20, in AT2 cells (**Fig. 3F**), indicating the pathogenic participation of iAT2 cells in COPD development. The physiological involvement of iAT2 cell-mediated inflammation in the loss of parenchymal structure may be further strengthened by a recent report, which showed that the existence of β1 integrin-deficient AT2 with upregulation of NF-κB-dependent chemokines was sufficient for developing accelerated inflammation accompanied by emphysematous lung destruction [28]. Pseudotime analyses of alveolar epithelial cells indicated that iAT2 cells were located at the end of the AT2 clusters and opposite to the AT1 cluster, suggesting that phenotypic alteration to iAT2 cells is a specific cellular process that does not induce canonical differentiation from AT2 to AT1. Regarding the AT1 subtypes, the AT1-2 cluster specifically expressing COL4A2 and COL4A3 decreased along with COPD progression, but the AT1-3 cluster with high inflammatory signals was increased in COPD lungs. Thus, it is plausible that alterations in the AT1 population are involved in the progression of lung emphysema development via not only inflammation but also impaired ECM production. However, pathway enrichment analysis based on the Reactome database showed relatively uniform patterns in the AT1 clusters, but heterogeneity and diversity were mainly detected in the AT2 clusters. Moreover, MAP kinase activation, which is responsible for the inflammatory process in COPD pathogenesis, was detected in only iAT2 cells. Accordingly, we speculate that iAT2 cells are mainly responsible for the inflammatory process in alveolar lesions and that alterations in AT1 clusters may mainly contribute to impaired recovery from

basement membrane damage during emphysema development. A recent preprint paper suggested the existence of a novel subpopulation of AT2 with high PD-L1 but low SFTPC expression in IPF lungs [29]. Despite the common characteristic of high PD-L1 expression, we recognize substantial differences between our iAT2 and AT2 subpopulations. Although the AT2 subpopulation with high PD-L1 but low SFTPC expression was responsible for the process of lung regeneration in IPF pathogenesis, we identified that iAT2 cells clearly exhibited an inflammatory phenotype associated with potential loss of the lung structure of the COPD phenotype. However, the physiological role of high PD-L1 expression in iAT2 cells remains uncertain in terms of regulating cell survival, T cell function, and potential association with smoking-related lung cancer development.

Small airways are the major site of airway obstruction in COPD, and airway remodeling comprises detrimental changes in structural tissues and an increased fractional volume of bronchiolar epithelial cells accompanied by epithelial cell alterations, including squamous metaplasia and goblet cell metaplasia [30,31]. Metaplastic alterations have been implicated in small airway remodeling through inflammatory cytokines and excessive mucus secretion[31-33]. In our analysis, distal airway epithelial cells were separated into four ciliated cell clusters, three club cell clusters, and one basal cell cluster based on the t-SNE plot with epithelial lineages (**Fig. 2A to D**). Among ciliated cell clusters, no apparent COPD dominant cluster was detected (**Fig. 2D**). In contrast, significantly increased club-3 cell and basal cell clusters were demonstrated (**Fig. 3A, B**). Club cells have a potential protective role in airway pathologies in COPD through anti-inflammation and immunomodulation by secreting CCSP [30,34]. A reduced number of club cells accompanied by CCSP reduction and CCSP polymorphism have been implicated in peribronchiolar fibrosis during COPD pathogenesis [34]. Club cells are known to have high plasticity and heterogeneity and can dedifferentiate into basal cells with multipotent stem cells to repair airway epithelial injury and may also promote alveolar regeneration, suggesting that phenotypic alterations in club cells may have a critical role in both airway and alveolar pathology[35,36]. Cell-cell interaction analysis clarified the strongest epithelial ligand-to-immune receptor interaction in club-3 cell clusters, including neutrophils, $CD4^+$ T cells, $CD8^+$ T cells, NK cells, and monocytes (**Fig. 5C, D, E**). Accordingly, it is plausible that some club cell populations may not be protective but aggressively contribute to enhancing airway inflammation during COPD pathogenesis. A recent paper demonstrated the heterogeneity of distal airway epithelial progenitor cells (p63 and KRT5) in COPD lungs and the existence of variants stably committed to those metaplastic fates, indicating that heterogeneity of distal airway epithelial cells, including basal cells, participates in small airway pathogenesis during COPD development [12]. In basal cell lineages, stemness was confirmed by bidirectional differentiation into either ciliated cell or club cell clusters (**Extended Data Fig. 8C, 8D**). Intriguingly, the number of cells in basal cell clusters clearly increased in only COPD lungs but not in non-COPD smoker or never-smoker lungs (**Fig. 3B**), indicating the more specific implication of

basal cells in airway pathology in COPD. With respect to the mechanisms of basal cell proliferation, cellular interaction analysis between basal cells and fibroblasts showed a paucity of Notch signaling in COPD samples **(Extended Data Fig. 9C)**. Notch signaling levels can be a critical determination between differentiation and self-renewal in basal cells [18], and COPD basal cells were relatively immature with respect to the expression levels of differentiation markers (**Fig. 6C and Extended Data Fig. 10C**). Accordingly, the exact mechanism remains unclear, but immatureness conferred by lower Notch signaling levels might be one of the possible explanations for the expansion of basal cells in COPD lungs (**Extended Data Fig. 9C**). Functionally, bulk RNA-seq analysis using basal cells after a few passages in culture isolated from COPD lungs showed activation of immune-related pathways such as the Th2 and TLR signaling pathways. Both IL13 and IL33 may be involved in the activation of the Th2 pathway (**Fig. 6H, 6I, Extended Data Fig. 10E**). Although the majority of COPD patients have predominantly neutrophilic inflammation and Th1-mediated immunity, some have eosinophilic inflammation, which can be governed by Th2 cells and type 2 innate lymphoid cells (ILC2s) though the release of IL-33 from epithelial cells [15]. It is likely that increased basal cells participate in not only the TLR-mediated innate immune response but also eosinophilic inflammation in COPD pathogenesis.

There are several limitations to our present study. Although our single-cell transcriptome profiling clarified the existence of novel epithelial cell clusters with inflammatory phenotypes, we obtained data sets from 12 samples from a single center, which may not be sufficient for comprehensively evaluating COPD pathogenesis with high clinical heterogeneity. As we observed the potential involvement of Notch signaling in basal cell expansion, epithelial-mesenchymal interactions are a critical determinant in regulating the proliferation and differentiation of epithelial cells, including basal cells, club cells, and alveolar epithelial cells[18,37,38]. Although the specific involvement of the stromal-type cell population was not revealed in the present form of evaluation, phenotypic alterations in epithelial cells denote the potential existence of pathogenic stromal-type cell clusters participating in COPD pathogenesis through aberrant cell-cell interactions. Chronic CS exposure is a major cause of COPD, but approximately 20% of smokers develop clinically diagnosable COPD. The number of iAT2 cells was increased in both COPD (0.83%) and non-COPD (0.50%) lungs; hence, it is not deniable that iAT2 cells can exhibit a simple phenotypic alteration to AT2 cluster cells in response to chronic CS exposure. The pathogenic role of the iAT2 cell-mediated inflammatory process as a distinctive AT2 subpopulation in COPD pathogenesis remains uncertain. The clinical implications of alterations in the basal cell lineage also remain elusive, especially in terms of airway remodeling and eosinophilic inflammation. The clinical implications of inflammatory epithelial cells should be examined by using appropriate mouse models in future studies.

In conclusion, our comprehensive single-cell transcriptome profiling of COPD lung tissues identified novel inflammatory epithelial cell clusters, including alveolar epithelial cell, club cell, and basal cell clusters. These inflamed epithelial cell populations may explain the mechanisms for persistent activation of inflammatory pathways after smoking cessation and have a pivotal role in orchestrating inflammation through bidirectional interactions with a variety of inflammatory cells, resulting in airway remodeling and emphysema development in COPD pathogenesis.

Methods

Patient recruitment and ethical approval.

Human lung tissue from never-smokers, non-COPD smokers, and patients with COPD was obtained from patients undergoing pneumonectomy or lobectomy for primary lung cancer. The diagnosis of COPD was established using the Global Initiative for Chronic Obstructive Lung Disease criteria. Informed consent was obtained from all surgical participants as part of an approved ongoing research protocol by the ethical committee of The Jikei University School of Medicine (#20-153 (5443)).

Lung tissue resection.

Lung tissue fragments were mechanically minced and dissociated with enzymes according to a Lung Dissociation Kit protocol (Miltenyi Biotec, Bergisch Gladbach, Germany). A Miltenyi Gentle MACS dissociator was used for mincing, followed by incubation with enzymes for 30 minutes at 37 °C. Cells were then washed in phosphate-buffered saline (PBS) supplemented with 1% bovine serum albumin (BSA, Sigma-Aldrich, MO, USA), passed through 70-μM and 40-μM filters, centrifuged at 300 g for 10 minutes, and resuspended in PBS supplemented with 1% BSA.

Chromium 10x Genomics library preparation and sequencing.

Cells were counted and resuspended in PBS supplemented with 1% BSA for loading for single-cell library construction on the 10x Genomics platform (10x Genomics, CA, USA). Experiments up to 4th samples were performed with the Chromium Single Cell 3′ GEM, Library & Gel Bead Kit v3 and Chromium Single Cell B Chip Kit according to the manufacturer's instructions in the Chromium Single Cell 3′ Reagents Kits V3 User Guide. Experiments after the 5th sample were performed with the Chromium Next GEM Single Cell 3′ GEM, Library & Gel Bead Kit v3.1 and Chromium Next GEM Chip G Single Cell Kit according to the manufacturer's instructions in the Chromium Next GEM Single Cell 3′ Reagents Kits V3.1 User Guide. Briefly, 6,000 cells were loaded into each channel and then partitioned into gel beads in emulsion in the Chromium Controller, where cell lysis and barcoded reverse transcription of RNA occurred, followed by amplification, fragmentation and 5′ adaptor and sample index attachment. Sequencing was performed on a HiSeq X (Illumina, CA, USA) at a median sequencing depth of 112,482 reads/cell.

Single-cell RNA sequencing data alignment.

Raw sequencing data were processed with the CellRanger pipeline (version 3.0.2, 10x Genomics) and mapped to the human genome (GRCh38) to generate matrices of gene counts by cell barcodes.

Data quality control and normalization.

scRNA-seq analyses, including normalization, scaling, clustering of cells and identifying cluster marker genes, were performed using the R software package Seurat version 3.2. [39]. We extracted single cells with nFeature_RNA > 100 and percent.mt < 75 for doublet and dead cell removal. Log normalization by function 'NormalizeData' from Seurat. Samples were log normalized and scaled for the number of genes and percentage of mitochondrial reads.

Data clustering and dimensionality reduction.

We performed principal component analysis (PCA) for dimensionality reduction in R with Seurat. Clustering of a single cell was performed by the functions 'FindNeighbors' and 'FindClusters' from Seurat. The dimensionality-reduced cell clustering is shown as a tSNE plot by the function 'runTSNE'.

Trajectory analysis.

We used the 'monocle' package for trajectory analysis of alveolar cells and bronchiolar cells (Trapnell et al., Nat Biotechnol., 2014). Dimensionality reduction in the trajectory plot was performed by the 'DDRTree' method with the top 10% genes of the identified cluster markers.

GO and pathway enrichment analyses.

Pathway datasets were downloaded from the Reactome database. We performed enrichment analysis against the signature gene list from each cell cluster by the 'ClusterProfiler' and 'ReactomePA' packages in R (Yu et al., OMICS, 2012; Yu et al., Mol Biosyst., 2016). For enrichment analysis, gene symbols were converted to ENTREZ IDs using the 'org.Hs.eg.db' package (Carison M, R package version 3.10.0., 2019). GO enrichment analysis using the 'enrichGO' function and pathway enrichment analysis using the 'enrichPathway' function were performed by the BH method.

Network analysis.

Correlational network analysis was performed by the 'igraph' package as previously described [40]. We calculated correlational coefficients between epithelial single cells in R, where r is the absolute value of Pearson's correlational coefficient and n is the cell number of the cluster.

$$Normalized\ Closeness\ Centrarility = \frac{n-1}{\sum min(1-r)}$$

Network visualization was performed by 'Y Organic Edge Router Layout' in yfile programs using Cytoscape v3.7.1. [41] [42]. For Cytoscape visualization, we extracted CSN > 0.9 edges in epithelial correlation of sif format.

Ligand-receptor cell-cell interaction (CCI) analysis.

Ligand-receptor interactions between epithelia and immune/stromal cell subpopulations were performed using the interaction database of Bader's laboratory from Toronto University (https://baderlab.org/CellCellInteractions#Download_Data) in R software. We selected the genes that were categorized as 'interferons', 'interleukins' and 'TNFSF superfamily' in the HUGO Gene Nomenclature Committee (https://www.genenames.org/). We calculated the cell number of subpopulations that expressed a value greater than 2. Only subpopulations whose expression cell ratio exceeded 10% were extracted for CCI network analysis, and the CCI score was calculated between each subpopulation. Next, a circos plot was drawn using the 'circlize' package in R [43]. The circle plot used the sum of the CCI scores from subpopulation A to subpopulation B for all combinations.

L: Ligand subpopulation (ligand gene expression > 2), R: receptor subpopulation (receptor gene expression > 2), n: cell number.

$$CCI\ score = \frac{n(L\ exp)}{n\ (total\ L)} \times \frac{n\ (R\ exp)}{n\ (total\ R)}$$

Primary cell culture of human small airway epithelial cells (HSAECs).

HSAECs were isolated from distal portions of lung tissue and dissociated in the same way as above. The dissociated cell suspension was cultured on rat tail collagen type I-coated dishes at 37 °C in 5% CO2 using Small Airway Epithelial Cell Growth Medium (SAGM; basal medium plus growth supplements, Lonza, CC-3118) containing 1% AA.

Bulk RNA-seq preparation and sequencing.

Total RNA was extracted from cultured cells using QIAzol and the miRNeasy Mini Kit (Qiagen, CA, USA) according to the manufacturer's protocol. After amplification of the total RNA, RNA quantity and quality were evaluated with a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, MA, USA) and an Agilent Bioanalyzer (Agilent Technologies, CA, USA). cDNA libraries for RNA sequencing were established from total RNA using NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, MA, USA) for selecting poly-A mRNA followed by strand-specific library preparation using MGIEasy RNA Directional Library Prep Set V2.0 (MGI tech, Shenzhen, China). Paired-end sequencing with a lead length of 150 bases was performed on a DNBSEQ-G400 (MGI tech) platform following the manufacturer's instructions.

Bulk transcriptome computational analysis.

Raw sequence qualities were checked by FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The human reference was 'Homo_sapiens.GRCh37.75.cdna.all.fa' from the Ensembl ftp site (http://asia.ensembl.org/info/data/ftp/index.html). Alignment and transcripts per million (TPM) calculations were performed using Kallisto [44] with a human cDNA reference. Output files of Kallisto were converted to an expression matrix by the 'tximport' and 'biomaRt' packages in R [45] [46] [47]. The transcriptome data were analyzed with gene set enrichment analysis (GSEA, https://www.gsea-msigdb.org/gsea/index.jsp), ingenuity pathway analysis (IPA, https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-ipa/), and Morpheus (https://software.broadinstitute.org/morpheus/).

Data availability

scRNA-seq data are now prepared to upload into the public database. Bulk RNA-seq data were submitted to the NCBI GEO database accession number GSE162154.

Code availability

The source code of scRNA-seq analysis is available on GitHub (https://github.com/JunNakayama/scRNAseq-COPD).

Immunohistochemical staining.

Each lung was fixed in 4% paraformaldehyde overnight and embedded in paraffin. Lung sections were dewaxed with xylene and rehydrated with ethanol (100–70%). Antigen retrieval was performed by boiling the specimens in Immunosaver (Nissin EM, Tokyo, Japan) diluted 1:200 for 45 minutes at 98 °C. The sections were permeabilized with 0.1% Triton X-100 (Sigma-Aldrich) for 15 minutes. After blocking with Dako blocking reagent for 30 minutes, sections were incubated with primary antibodies overnight at 4°C in a humidified box. Sections were then incubated with secondary antibodies labeled with Alexa Fluor 488 and 594 (Invitrogen). Slides were mounted with VECTASHIELD mounting medium with DAPI (Vector Labs, CA, USA). Stained sections were imaged using a BZ-X710 microscope (Keyence, Osaka, Japan).

Immunofluorescence staining

The cells were fixed in 4% paraformaldehyde (Fujifilm, Tokyo, Japan) for 15 minutes and blocked in 0.1% BSA containing PBS for 30 minutes. The cells were incubated with primary antibodies diluted in 0.1% BSA containing PBS overnight at 4 ℃. The cells were then incubated with Alexa Fluor fluorescent secondary antibodies. DAPI (Vector Labs) staining was performed immediately before imaging. All staining was observed using a BZ-X710 microscope (Keyence, Osaka, Japan)

Antibodies and reagents

Antibodies and reagents were as follows: anti-PD-L1 (rabbit, 1:1000, #ab205921, Abcam, Cambridge, UK); anti-TNC (rabbit, 1:100, #HPA004823, Sigma-Aldrich); anti-CCL2 (rabbit, 1:100, #HPA019163, Sigma-Aldrich); anti-RAGE (rabbit, 1:1000, #ab216329, Abcam); anti-SFTPC (rabbit, 1:1000, #HPA010928, Sigma-Aldrich); anti-ABCA3 (mouse, 1:1000, #WMAB-ABCA3-17, Seven Hills Bioreagents, OH, USA); anti-macrophage inflammatory protein 3 alpha (rabbit, 1:1000, #ab224188, Abcam); anti-CXCL1 (mouse, 1:100, #MAB275, R&D Systems, MN, USA); anti-CXCL8 (mouse, 1:100, #MAB208, R&D Systems); anti-p63 (mouse, 1:250, #ab735, Abcam); anti-Cytokeratin5 (rabbit, 1:1000, #ab53121, Abcam); and anti-Hoechst 33242 (1:1000, #H342, Sigma-Aldrich).

Statistical Analysis

Student’s t-test was used for comparison of two data sets. Correlation coefficients were calculated by Spearman correlation. Analysis of variance (ANOVA) was used for multiple comparisons, followed by Tukey's or Dunnett's multiple comparison to find where the differences lie. Significance was defined as $P < 0.05$. The statistical software used was Prism version 8 (GraphPad Software, Inc., CA, USA).

Acknowledgments

This publication is part of the Single Cell Medical Network of Japan. This work was supported by Project for MEXT KAKENHI (Grant-in-Aid for Young Scientists (A); grant number: 17H04991, Grant-in-Aid for JSPS Fellows: 20J01794), Research grant from The Naito Foundation, GSK Japan Research Grant, Mochida Memorial Foundation for Medical and Pharmaceutical Research, and Takeda Science Foundation. Schematics in figures were made using an academic license of Biorender software.

Author contributions

N.W., J.N., Y.F. and Y.Y. conceived and designed the study. N.W. and Y.M. performed the experiments. N.W., J.N. and Y.Y. performed the data analysis. N.W., T.K., I.S., K.O., J.A., and K.K. interpreted the data. T.O. provided the patient samples. N.W., J.N., Y.F., J.A., and Y.Y. wrote the manuscript. Y.Y. supervised this project. All authors reviewed and edited the manuscript.

Figure legends

Figure 1 | Charting the human lung cell landscape of COPD patients, non-COPD smokers and never-smokers by scRNA-seq. A. Experimental flowchart for scRNA-seq. Patient lung tissue samples were enzymatically digested into single cells. cDNA synthesis was performed using a 10x genomics Chromium controller system, and libraries were sequenced on Illumina platforms. **B.** Representative H&E images of COPD, non-COPD smoker and never-smoker lung sections. **C.** Table with the details of patient ID, classification, age, sex and cell numbers present in this figure. **D.** t-SNE plot displaying the major clusters of human lung cells. **E.** t-SNE plot classifying patient states such as COPD, non-COPD smokers and never-smokers. **F.** Fraction of cells (x-axis) from each sample (left) and patient state (middle) in each cluster from the t-SNE plot (y-axis). The cell numbers of each cluster are shown (right). **G.** Heatmaps of differentially expressed genes. The average expression (log2[TPM + 1]) of the top 10 genes (row) in each cluster (columns) is shown. Immune-related clusters, epithelial-related clusters and stromal-related clusters are separately shown.

Figure 2 | Single-cell transcriptome analysis of the epithelial components of human distal lung tissues among COPD, non-COPD smokers and never-smokers. A. t-SNE plot displaying the major epithelial clusters of human lung cells. **B.** t-SNE plot of epithelial components classifying patient states such as COPD, non-COPD smoker and never-smoker. **C.** Density plot of epithelial components in human lung cells. **D.** Fraction of cells (x-axis) of epithelial components from the patient state (left) in each epithelial cluster (y-axis). The cell numbers of each cluster of epithelial components are shown (right). **E.** Pie charts displaying epithelial cell compositions among COPD, non-COPD smokers and never-smokers. **F.** Heatmap of differentially expressed genes across epithelial clusters. The average expression (log2[TPM + 1]) of the top 10 genes (row) in each epithelial cluster (columns) is shown. **G.** Dot plot heatmap showing the average expression levels and percentage of expression of selected marker genes in the epithelial cluster. **H.** Heatmap showing COPD GWAS genes in the epithelial clusters.

Figure 3 | Variations in epithelial cell components among COPD, non-COPD smokers and never-smokers. A. 100% stacked bar charts of percentage of cell populations for AT1 subtype clusters, AT2 subtype clusters and basal lineage clusters. **B.** Bar charts displaying the percentages of subpopulations such as AT1-2, AT2-1, club-3 and basal clusters across patient states. *****$p < 0.05$ by Tukey's test. **C.** t-SNE plot of epithelial cells focusing on the AT2-5 cluster (left). Cell population of the AT2-5 cluster (right). The y-axis represents the ratio of cells in the AT2-5 cluster across patient states. Brackets () represent the percentage of cells of the AT2-5 cluster based on all lung cells in each of the patient states. **D.** Correlation plots of specifically expressed genes in the AT2-5 cluster.

Among the 30 genes, 15 genes were related to the immune response. **E.** Scatter plots displaying the correlations between CXCL1 and CXCL8, CXCL1 and CSF3, CD274 and CCL2, and CD274 and TNC. The p value and correlation coefficient were calculated by Spearman correlation. **F.** Representative images immunostained with CD274 (PD-L1), CXCL1 and CXCL20 with AT2 cell markers (SFTPC or ABCA3) in COPD and never-smoker lung sections. Scale bars, 25 μm (left), 12.5 μm (right).

Figure 4 | Pseudotime prediction of alveolar differentiation and epithelial variations in COPD. A. Pseudotime developmental trajectory analysis of alveolar lineages. Left panel: state, middle panel: group (COPD, non-COPD smoker and never-smoker), right panel: alveolar clusters. **B.** Pseudotime time plots for COPD, non-COPD smokers and never-smokers. **C.** Cluster distribution density along pseudotime for COPD, non-COPD smoker and never-smoker. **D.** Expression heatmap showing differentially expressed genes between two transition states (FDR < 0.01). Selected genes are highlighted. **E.** Pathway enrichment analysis of alveolar clusters among COPD, non-COPD smokers and never-smokers. Color scale represents FDR q value. Dot size represents likelihood ratio. **F.** Box plots showing the variation of clusters in COPD, non-COPD smokers and never-smokers. Statistics are shown in **Supplementary Table 9**.

Figure 5 | Inferred cell-to-cell interaction networks in the COPD epithelia. A. Representative images of topological positions of cell networks for closeness centrality across COPD, non-COPD smokers and never-smokers. The degrees of closeness centrality were normalized to the cell number of patient states. **B.** Box plot of normalized closeness centrality across COPD, non-COPD smokers and never-smokers. **C.** Circos plot of cell-to-cell interactions featuring inflammatory cytokines and chemokines in COPD. Red lines represent predicted immune cell-ligand to epithelial cell-receptor interactions. Blue lines represent predicted epithelial cell-ligand to immune cell-receptor interactions. Ma: macrophage, MC: mast cell, Neu: neutrophil, mDC: myeloid dendritic cell and pDC: plasmacytoid dendritic cell. **D.** Stacked bar charts showing cell-to-cell interaction counts. Left: epithelial cell-ligand to epithelial cell-receptor interactions. C: COPD, and N: never-smoker. **E.** Bubble chart displaying cell-to-cell interactions of club-3, AT2-5 or AT1-3 clusters with immune cells. Red: COPD and blue: never-smoker. Circle size represents the sum of the CCI score.

Figure 6 | Increased population of basal cells and their specific function in COPD epithelia. A. Heatmap showing differentially expressed genes in individual basal cells from COPD, non-COPD smokers and never-smokers. Twenty-two genes were selected as differentially expressed genes between COPD patients and others (non-COPD smokers and never-smokers). Fold change > 1.5-fold and p value < 0.05 by Student's t-test. **B.** t-SNE plot of basal cells in COPD, non-COPD smokers

and never-smokers. **C.** Box plots of selected differentially expressed genes (KRT5, KRT19, S100A2, SFTPC and SCGB1A1). **D.** Representative images of cultured human small airway epithelial cells (HSAECs) from COPD or never-smoker lung tissues. Scale bars, 100 μm. **E**. Representative images stained with p63 and Krt5 in cultured HSAECs. Scale bars, 50 μm. **F.** Heatmap of differentially expressed genes between COPD and never-smokers (762 genes, fold change > 1.5 and $p < 0.05$). **G**. GSEA with HSAEC between COPD and never smoker. NES: normalized enrichment score. **H.** IPA data of upstream signals. Selected signals ($p < 0.01$) are shown in the graph. **I.** Signal interaction network significantly enriched in the COPD HSAECs. Orange: activated; blue: inhibited.

Extended Data Figure legends

Extended Data Figure 1 | Unbiased clustering of human lung tissue cells from COPD patients, non-COPD smokers, and never-smokers. A. UMAP with human lung tissue cells for unbiased clustering. **B.** PCA mapping with human lung tissue cells. **C.** t-SNE showing the expression of individual genes used for cell-type assignment of different cell subsets, such as epithelial cells, immune cells and stromal cells.

Extended Data Figure 2 | Single-cell transcriptome analysis of the immune components of human distal lung tissues from COPD patients, non-COPD smokers and never-smokers. A. t-SNE plot displaying the major immune cell clusters of human lung cells. **B.** t-SNE plot of immune components classifying the patient states, such as COPD, non-COPD smoker and never-smoker. **C.** Pie charts displaying immune cell compositions among COPD, non-COPD smokers and never-smokers. **D.** Fraction of cells (x-axis) of immune components from the patient state (left) in each immune cluster (y-axis). The cell numbers of each cluster of immune components are shown (right). **E-G.** Violin plots for typical markers and differentially expressed genes among clusters. **E.** Neutrophil clusters (neutrophil-1 and -2); **F.** macrophage clusters (macrophage-1, -2 and -3); and **G.** mast cell clusters (mast cell-1 and -2). **H.** Bar charts displaying percentages of subpopulations for NK cell and myeloid DC clusters across the patient states. $p < 0.05$ by Tukey's test (COPD vs. never smoker). **I.** t-SNE plot of Foxp3 for depicting Tregs in the CD4 T cluster.

Extended Data Figure 3 | Single-cell transcriptome analysis of the stromal components of human distal lung tissues among COPD, non-COPD smokers and never-smokers. A. t-SNE plot displaying the major stromal clusters of human lung cells. **B.** t-SNE plot of stromal components classifying patient states such as COPD, non-COPD smokers and never-smokers. **C.** Pie charts displaying stromal cell compositions among COPD, non-COPD smokers and never-smokers. **D.** Fraction of cells (x-axis) of stromal components from the patient state (left) in each stromal cluster

(y-axis). The cell numbers of each cluster of stromal components are shown (right). **E, F.** Violin plots for typical markers and differentially expressed genes among clusters. **E.** Fibroblast clusters (fibroblasts-1, -2 and -3); and **F.** smooth muscle clusters (smooth muscle-1 and -2). **G.** Bar charts displaying percentages of subpopulations for fibroblast-1, vein EC and capillary EC 1 clusters across the patient states. $p < 0.05$ by Tukey's test (COPD vs. never smoker).

Extended Data Figure 4 | Expression profiles of specific epithelial marker genes and COPD GWAS genes in the epithelial clusters. A. t-SNE plot of CHGA, ASCL1 and INSM1 for depicting neuroendocrine cells. **B.** t-SNE plot of FOXI1 for depicting ionocytes. **C.** t-SNE plot of DCLK1 and ASCL2 for depicting tuft cells. **D.** Number of AT1 and AT2 cells in COPD and non-COPD (non-COPD smoker and never-smoker) patients. **E**. Ratio of AT1/AT2 in COPD and non-COPD patients. **F**. Violin plots of typical and specific markers for characterizing seven AT2 clusters. **G**. Representative images of AT2-4 cells stained with AGER and SFTPC. Scale bars, 25 μm (left), 12.5 μm (right). **H.** Violin plots of typical and specific markers for characterizing three club cell clusters. **I.** Heatmap showing COPD GWAS genes [13] across the clusters (all lung tissue cells) in Fig. 1D.

Extended Data Figure 5 | t-SNE plots representing the epithelial clusters. A. t-SNE plots of typical markers depicting AT1-2, AT2-1, club-3 and basal cell clusters. **B.** t-SNE plots showing basal, club-3, AT1-2 and AT2-1 clusters, marked with the patient states. Red: COPD, orange: non-COPD smoker, and blue: never-smoker.

Extended Data Figure 6 | Expression profiles of identified inflammatory AT2-5 cluster markers. A. t-SNE plots of inflammatory-related genes displaying the inflammatory AT2-5 cluster. **B.** t-SNE plots of selectively expressed genes in the inflammatory AT2-5 cluster.

Extended Data Figure 7 | Confirmation of inflammatory responses in epithelial cell clusters. A. Representative images immunostained with CCL2, TNC, and CXCL8 with AT2 cell markers (SFTPC or ABCA3) in COPD and never-smoker lung sections. B, C. Expression levels of immune-related genes in AT1 clusters and club cell clusters.

Extended Data Figure 8 | Pseudotime prediction of basal lineage differentiation. A. Pseudotime developmental trajectory analysis of basal cell lineages. Left panel: state, middle panel: group (COPD, non-COPD smoker and never-smoker), right panel: basal clusters. **B.** Pseudotime time plots for COPD, non-COPD smokers and never-smokers. **C.** Cluster distribution density along pseudotime for COPD, non-COPD smoker and never-smoker. **D**. 3D image of pseudotime developmental trajectory analysis of basal cell lineages showing bidirectional differentiation into either ciliated cell

clusters or club cell clusters. **E.** Expression heatmap showing differentially expressed genes between two transition states (FDR < 0.01). Selected genes are highlighted. **F.** Pathway enrichment analysis of basal cell clusters among COPD, non-COPD smokers and never-smokers. Color scale represents FDR q value. Circle size represents the likelihood ratio.

Extended Data Figure 9 | Inferred cell-to-cell interaction networks in never-smokers and for stromal cell interactions with epithelial cell clusters. A. Circos plot of cell-to-cell interactions featuring inflammatory cytokines and chemokines in never-smokers. Red lines represent predicted immune cell-ligand to epithelial cell-receptor interactions. Blue lines represent predicted epithelial cell-ligand to immune cell-receptor interactions. Ma: macrophage, MC: mast cell, Neu: neutrophil, mDC: myeloid dendritic cell, and pDC: plasmacytoid dendritic cell. **B.** Circos plots of cell-to-cell interactions between epithelial and stromal clusters. **C.** Cell-to-cell interaction between basal cells and fibroblast clusters focusing on Notch signals in the COPD and non-COPD (non-COPD smoker and never-smoker) groups. Circle size represents the likelihood ratio.

Extended Data Figure 10 | Bulk RNA-seq analysis of HSAECs in COPD and never smoker lungs. A. PCA mapping with 3 COPD and 3 never smoker samples. Red: COPD and blue: never smoker HSAEC. **B, C**. Representative heatmaps for immune-related genes (**B**) and basal/stem cell markers (**C**). **D**. GSEA with HSAEC between COPD and never smoker. NES: normalized enrichment score. **E.** Top 20 canonical pathways enriched in 762 genes that were highly expressed in the COPD HSAECs based on IPA.

Supplementary Table

Supplementary Table 1：**Patient demographics and clinical data**
Supplementary Table 2：**scRNA-seq information of each sample**
Supplementary Table 3：**Number of cells per cluster per donor**
Supplementary Table 4：**Ratio of AT1 and AT2 cells to total cell count**
Supplementary Table 5A：**CCIScore: Epithelial cells to Immune cells**
Supplementary Table 5B：**CCIScore: Immune cells to Epithelial cells**
Supplementary Table 6：**Patient demographics and clinical data for HSAEC culture**
Supplementary Table 7：**Differentially expressed genes in Basal Cells betweem COPD and Never smoker**
Supplementary Table 8：**Canonical pathways enriched in 762 genes which were highly expressed in COPD**
Supplementary Table 9：**Statistical analysis of AT2 cluster variation**

Reference

1. Mortality, G.B.D. & Causes of Death, C. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1459-1544 (2016).
2. Global Burden of Disease Study, C. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **386**, 743-800 (2015).
3. Rabe, K.F. & Watz, H. Chronic obstructive pulmonary disease. *Lancet* **389**, 1931-1940 (2017).
4. Tuder, R.M. & Petrache, I. Pathogenesis of chronic obstructive pulmonary disease. *J Clin Invest* **122**, 2749-2755 (2012).
5. Beers, M.F. & Moodley, Y. When Is an Alveolar Type 2 Cell an Alveolar Type 2 Cell? A Conundrum for Lung Stem Cell Biology and Regenerative Medicine. *Am J Respir Cell Mol Biol* **57**, 18-27 (2017).
6. Selman, M. & Pardo, A. Revealing the pathogenic and aging-related mechanisms of the enigmatic idiopathic pulmonary fibrosis. an integral model. *Am J Respir Crit Care Med* **189**, 1161-1172 (2014).
7. Garcia, O., *et al.* Targeted Type 2 Alveolar Cell Depletion. A Dynamic Functional Model for Lung Injury Repair. *Am J Respir Cell Mol Biol* **54**, 319-330 (2016).
8. Plasschaert, L.W., *et al.* A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377-381 (2018).
9. Reyfman, P.A., *et al.* Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. *Am J Respir Crit Care Med* **199**, 1517-1536 (2019).
10. 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).
11. 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).
12. Rao, W., *et al.* Regenerative Metaplastic Clones in COPD Lung Drive Inflammation and Fibrosis. *Cell* **181**, 848-864 e818 (2020).
13. Sakornsakolpat, P., *et al.* Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat Genet*

51, 494-505 (2019).

14. Shin, K., *et al.* Hedgehog/Wnt feedback supports regenerative proliferation of epithelial stem cells in bladder. *Nature* 472, 110-114 (2011).

15. Barnes, P.J. Inflammatory mechanisms in patients with chronic obstructive pulmonary disease. *J Allergy Clin Immunol* 138, 16-27 (2016).

16. Brightling, C. & Greening, N. Airway inflammation in COPD: progress to precision medicine. *Eur Respir J* 54(2019).

17. Crotty Alexander, L.E., Shin, S. & Hwang, J.H. Inflammatory Diseases of the Lung Induced by Conventional Cigarette Smoke: A Review. *Chest* 148, 1307-1322 (2015).

18. Rock, J.R., *et al.* Notch-dependent differentiation of adult airway basal stem cells. *Cell Stem Cell* 8, 639-648 (2011).

19. Shi, W., Chen, F. & Cardoso, W.V. Mechanisms of lung development: contribution to adult lung disease and relevance to chronic obstructive pulmonary disease. *Proc Am Thorac Soc* 6, 558-563 (2009).

20. Saito, N., *et al.* Involvement of Lamin B1 Reduction in Accelerated Cellular Senescence during Chronic Obstructive Pulmonary Disease Pathogenesis. *J Immunol* 202, 1428-1440 (2019).

21. Tsuji, T., Aoshiba, K. & Nagai, A. Alveolar cell senescence in patients with pulmonary emphysema. *Am J Respir Crit Care Med* 174, 886-893 (2006).

22. Takasaka, N., *et al.* Autophagy induction by SIRT6 through attenuation of insulin-like growth factor signaling is involved in the regulation of human bronchial epithelial cell senescence. *J Immunol* 192, 958-968 (2014).

23. Ito, S., *et al.* PARK2-mediated mitophagy is involved in regulation of HBEC senescence in COPD pathogenesis. *Autophagy* 11, 547-559 (2015).

24. Demedts, I.K., Demoor, T., Bracke, K.R., Joos, G.F. & Brusselle, G.G. Role of apoptosis in the pathogenesis of COPD and pulmonary emphysema. *Respir Res* 7, 53 (2006).

25. Mizumura, K., *et al.* Mitophagy-dependent necroptosis contributes to the pathogenesis of COPD. *J Clin Invest* 124, 3987-4003 (2014).

26. Yoshida, M., *et al.* Involvement of cigarette smoke-induced epithelial cell ferroptosis in COPD pathogenesis. *Nat Commun* 10, 3145 (2019).

27. medRxiv, Sauler, M., *et al.* Single-cell RNA sequencing identifies aberrant transcriptional profiles of cellular populations and altered alveolar niche signalling networks in Chronic Obstructive Pulmonary Disease (COPD). Preprint at https://doi.org/10.1101/2020.09.13.20193417(2020).

28. Plosa, E.J., *et al.* beta1 Integrin regulates adult lung alveolar epithelial cell

inflammation. *JCI Insight* **5**(2020).

29. bioRxiv, Ahmadvand, N., *et al.* Identification a novel subset of alveolar type 2 cells expanding following pneumonectomy and enriched in PD-L1. Preprint at https://doi.org/10.1101/2020.05.20.106005 (2020).
30. Randell, S.H. Airway epithelial stem cells and the pathophysiology of chronic obstructive pulmonary disease. *Proc Am Thorac Soc* **3**, 718-725 (2006).
31. Hogg, J.C. & Timens, W. The pathology of chronic obstructive pulmonary disease. *Annu Rev Pathol* **4**, 435-459 (2009).
32. Araya, J., *et al.* Squamous metaplasia amplifies pathologic epithelial-mesenchymal interactions in COPD patients. *J Clin Invest* **117**, 3551-3562 (2007).
33. Araya, J. & Nishimura, S.L. Fibrogenic reactions in lung disease. *Annu Rev Pathol* **5**, 77-98 (2010).
34. Gamez, A.S., *et al.* Supplementing defect in club cell secretory protein attenuates airway inflammation in COPD. *Chest* **147**, 1467-1476 (2015).
35. Kathiriya, J.J., Brumwell, A.N., Jackson, J.R., Tang, X. & Chapman, H.A. Distinct Airway Epithelial Stem Cells Hide among Club Cells but Mobilize to Promote Alveolar Regeneration. *Cell Stem Cell* **26**, 346-358 e344 (2020).
36. Tata, P.R., *et al.* Dedifferentiation of committed epithelial cells into stem cells in vivo. *Nature* **503**, 218-223 (2013).
37. Lee, J.H., *et al.* Anatomically and Functionally Distinct Lung Mesenchymal Populations Marked by Lgr5 and Lgr6. *Cell* **170**, 1149-1163 e1112 (2017).
38. Nabhan, A.N., Brownfield, D.G., Harbury, P.B., Krasnow, M.A. & Desai, T.J. Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science* **359**, 1118-1123 (2018).
39. Stuart, T., *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).
40. Nakayama, J., Ito, E., Fujimoto, J., Watanabe, S. & Semba, K. Comparative analysis of gene regulatory networks of highly metastatic breast cancer cells established by orthotopic transplantation and intra-circulation injection. *Int J Oncol* **50**, 497-504 (2017).
41. Saito, R., *et al.* A travel guide to Cytoscape plugins. *Nat Methods* **9**, 1069-1076 (2012).
42. Shannon., P., *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 2498-2504 (2003).
43. Gu, P. & Chen, H. Modern bioinformatics meets traditional Chinese medicine. *Brief Bioinform* **15**, 984-1003 (2014).

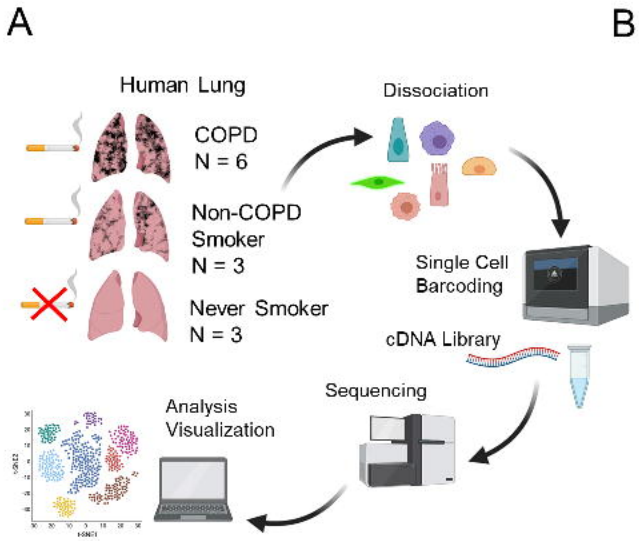
44. Bray, N.L., Pimentel, H., Melsted, P. & Pachter, L. Erratum: Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 888 (2016).

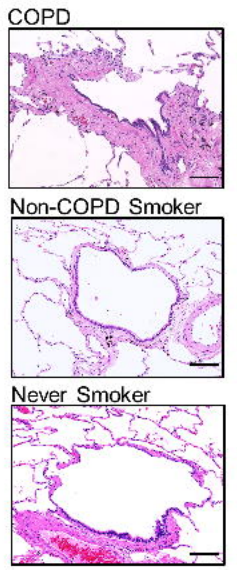
45. Soneson, C., Love, M.I. & Robinson, M.D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* **4** (2015).

46. Durinck, S., Spellman, P.T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**, 1184-1191 (2009).

47. Durinck, S., *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439-3440 (2005).

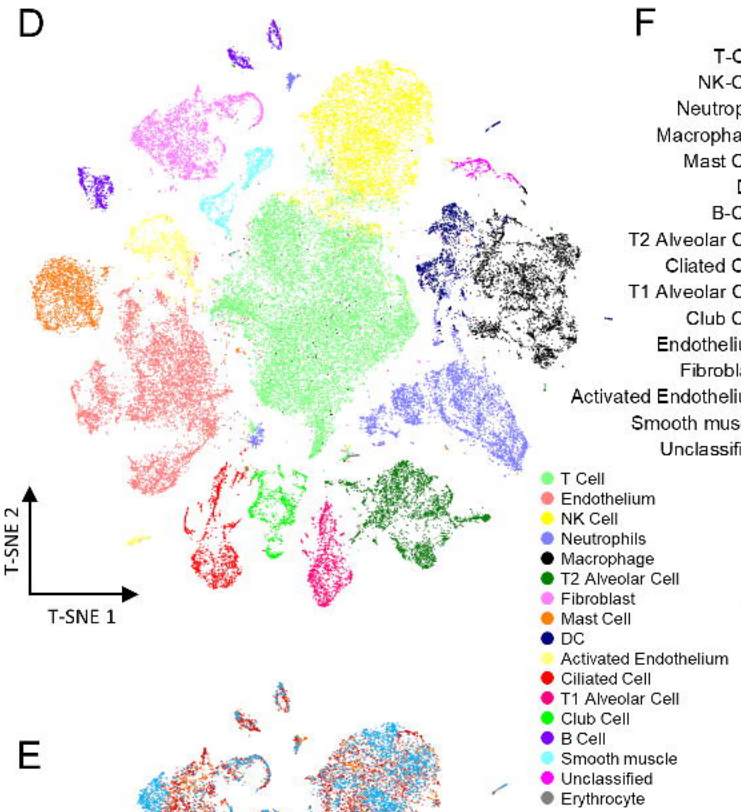
Figure 1

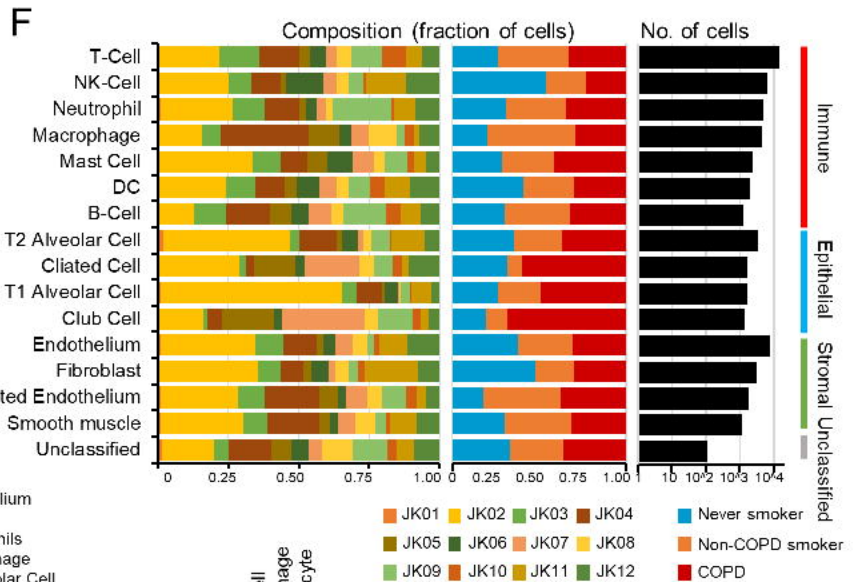


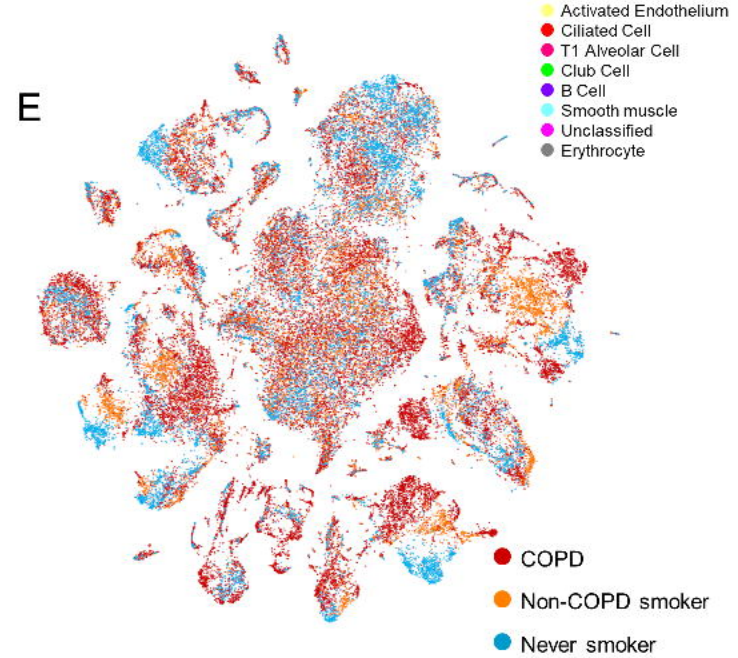


C

Donor	Classification	Age	Gender	Cell No.
JK02	COPD	60's	M	15777
JK05	COPD	70's	F	2694
JK07	COPD	80's	M	3229
JK08	COPD	80's	M	2852
JK09	COPD	70's	M	4636
JK10	COPD	80's	M	2165
JK01	Non-COPD	50's	M	438
JK03	Non-COPD	60's	M	5557
JK04	Non-COPD	60's	M	7710
JK06	Never smoker	60's	M	3440
JK11	Never smoker	60's	F	4586
JK12	Never smoker	30's	M	4671







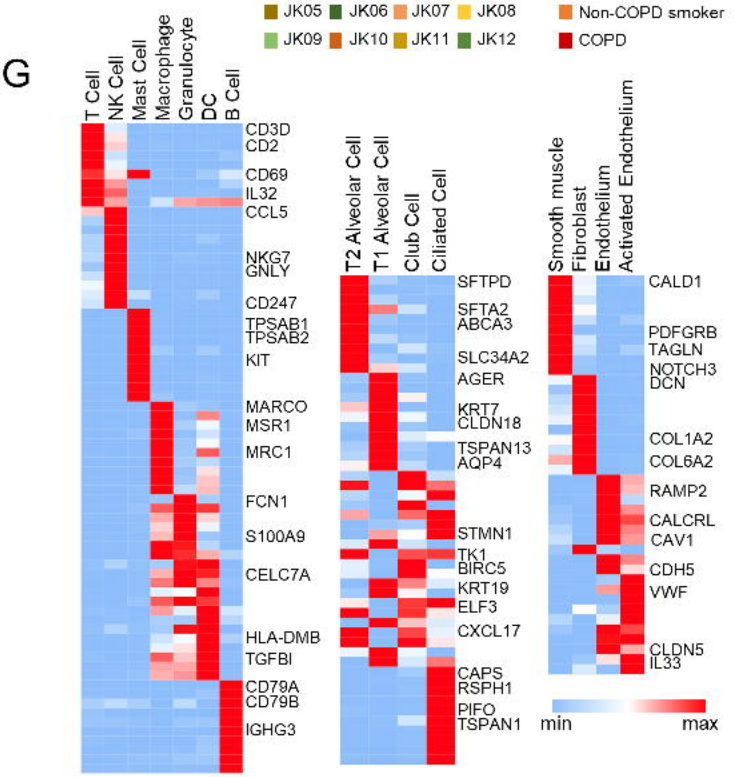


Figure 2

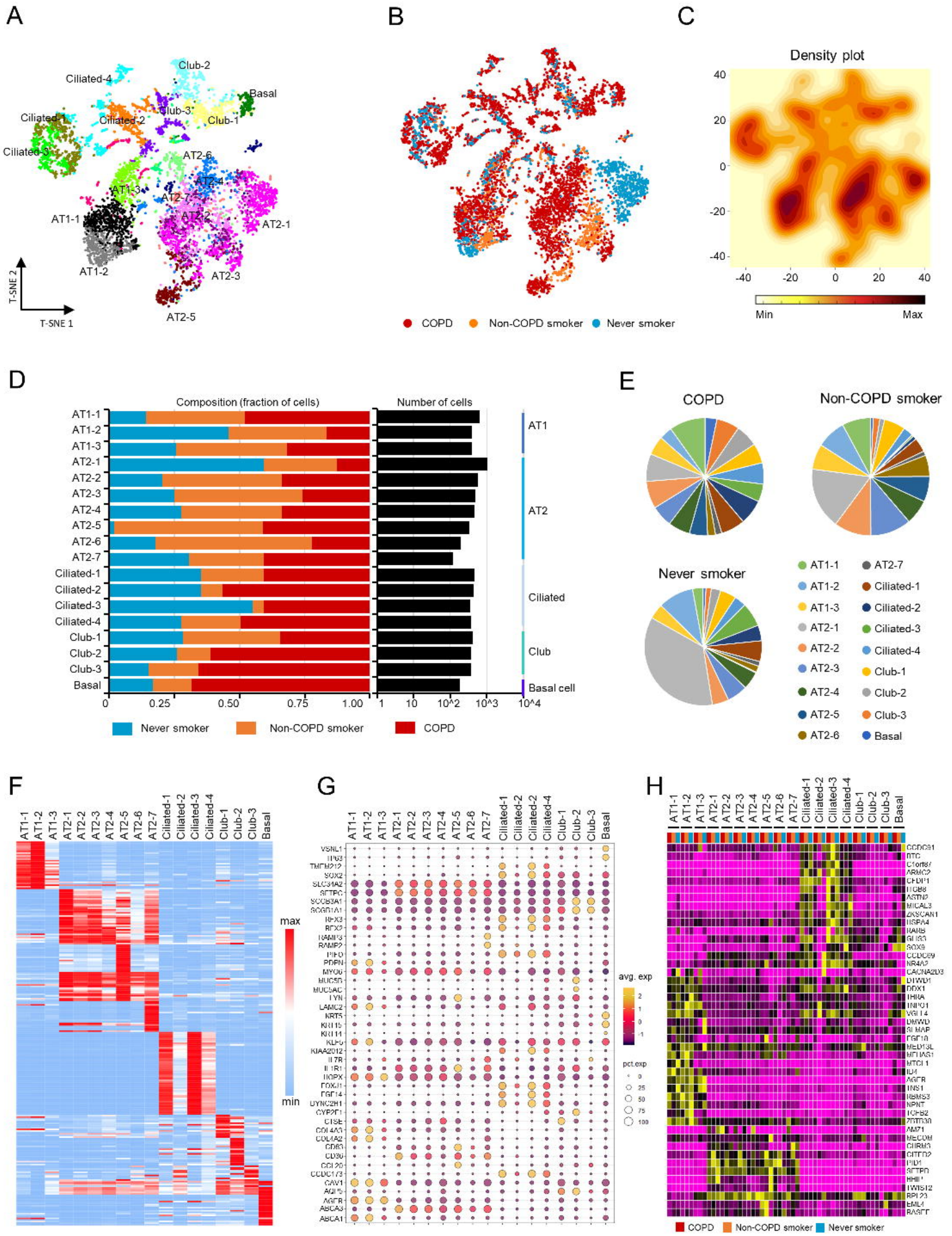

A
Ciliated-4
Club-2
Basal
Ciliated-1
Ciliated-2
Club-3
Club-1
Ciliated-3
AT2-6
AT2-4
AT1-3
AT2-7
AT1-1
AT2-2
AT2-1
AT1-2
AT2-3
AT2-5
T-SNE 2
T-SNE 1
B
COPD
Non-COPD smoker
Never smoker
C
Density plot
Min
Max
D
Composition (fraction of cells)
Number of cells
AT1
AT2
Ciliated
Club
Basal cell
Never smoker
Non-COPD smoker
COPD
E
COPD
Non-COPD smoker
Never smoker
F
G
H
max
min
avg. exp
pct.exp
COPD
Non-COPD smoker
Never smoker

Figure 3

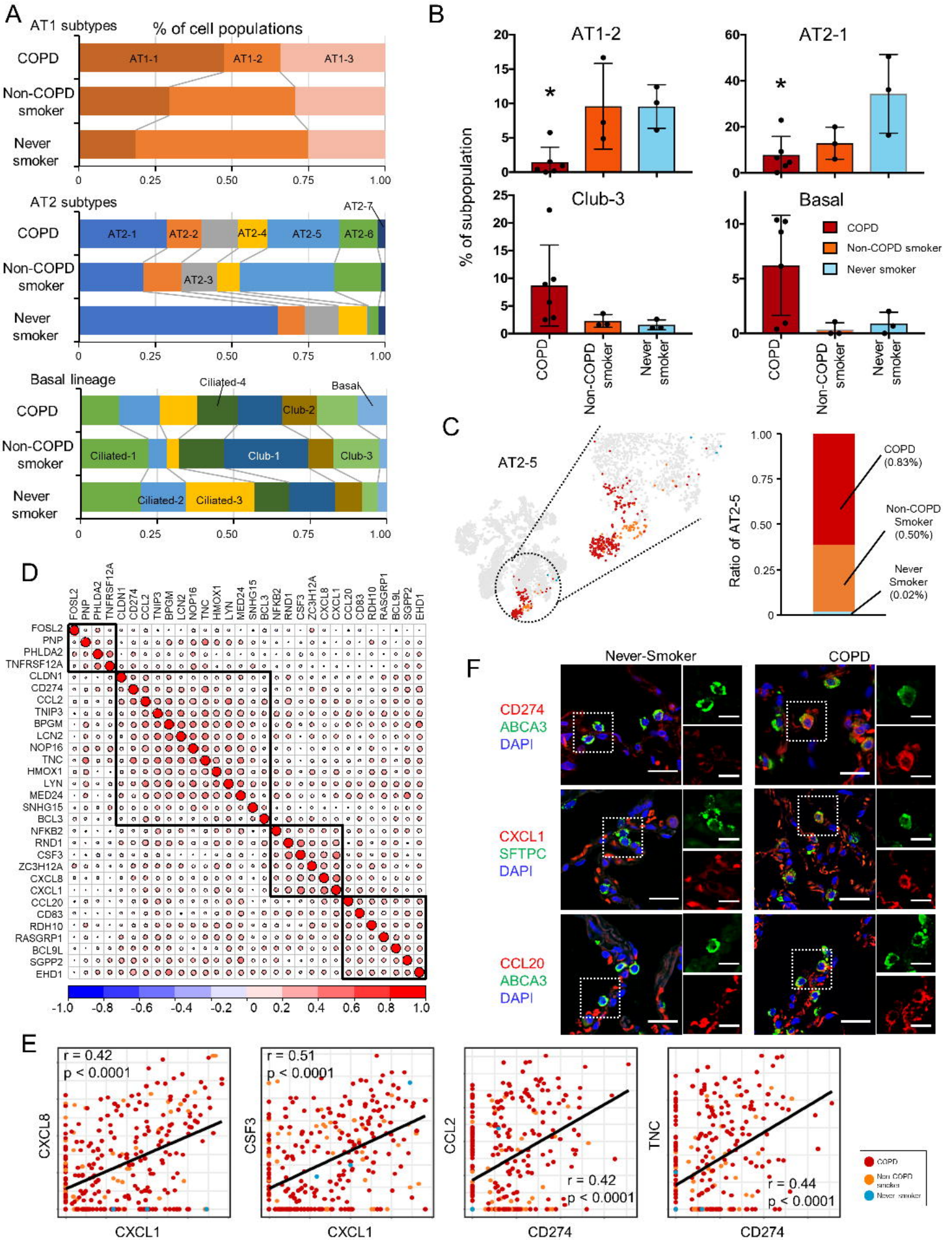

A
AT1 subtypes
% of cell populations
COPD
Non-COPD smoker
Never smoker
AT1-1
AT1-2
AT1-3
0
0.25
0.50
0.75
1.00
AT2 subtypes
AT2-1
AT2-2
AT2-3
AT2-4
AT2-5
AT2-6
AT2-7
Basal lineage
Ciliated-1
Ciliated-2
Ciliated-3
Ciliated-4
Club-1
Club-2
Club-3
Basal
B
AT1-2
AT2-1
Club-3
Basal
% of subpopulation
COPD
Non-COPD smoker
Never smoker
C
AT2-5
Ratio of AT2-5
COPD (0.83%)
Non-COPD Smoker (0.50%)
Never Smoker (0.02%)
D
FOSL2
PNP
PHLDA2
TNFRSF12A
CLDN1
CD274
CCL2
TNIP3
BPGM
LCN2
NOP16
TNC
HMOX1
LYN
MED24
SNHG15
BCL3
NFKB2
RND1
CSF3
ZC3H12A
CXCL8
CXCL1
CCL20
CD83
RDH10
RASGRP1
BCL9L
SGPP2
EHD1
-1.0
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1.0
F
Never-Smoker
COPD
CD274
ABCA3
DAPI
CXCL1
SFTPC
DAPI
CCL20
ABCA3
DAPI
E
r = 0.42
p < 0.0001
CXCL8
CXCL1
r = 0.51
p < 0.0001
CSF3
CXCL1
CCL2
CD274
r = 0.42
p < 0.0001
TNC
CD274
r = 0.44
p < 0.0001
COPD
Non-COPD smoker
Never smoker

Figure 4

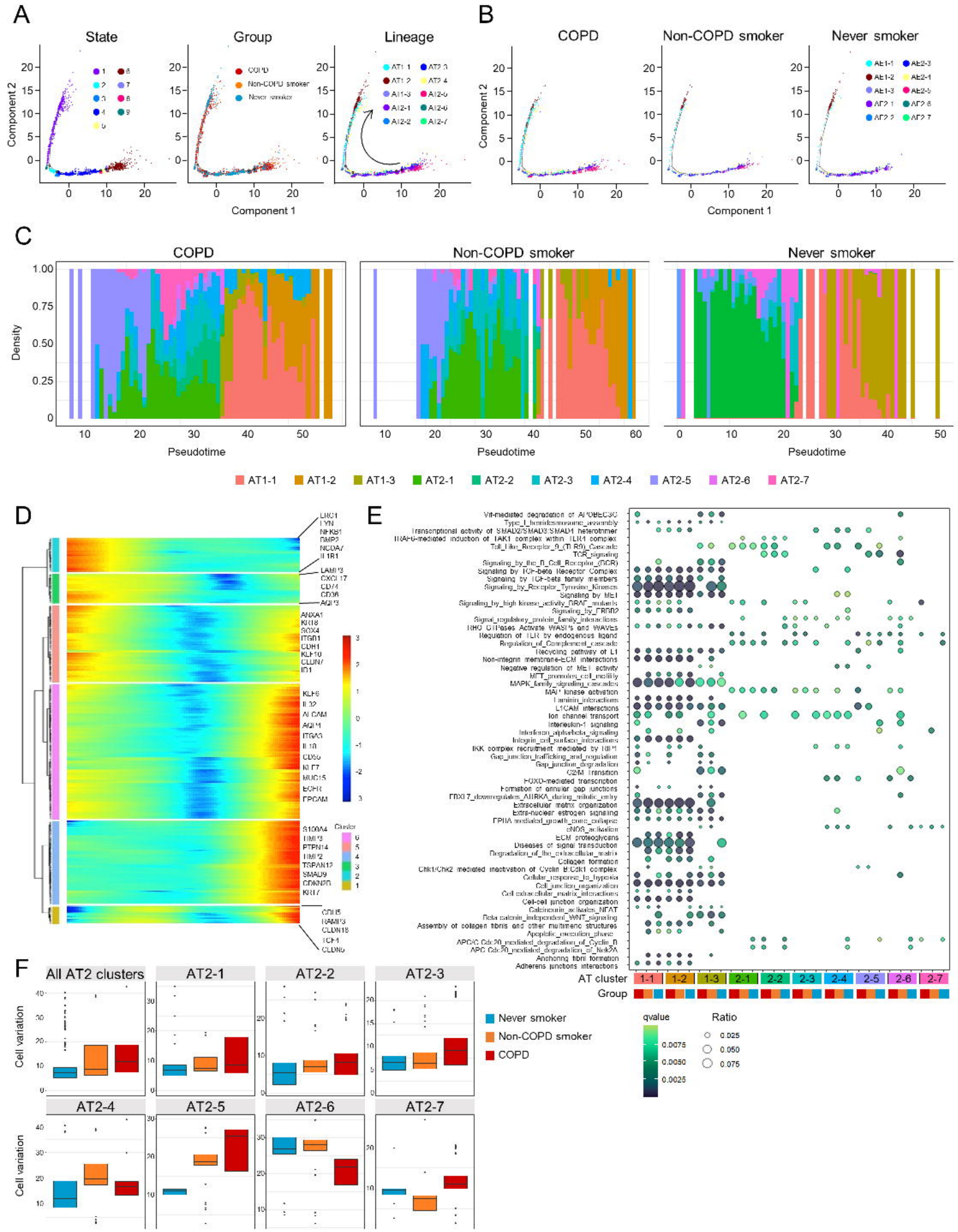

A
State
Group
Lineage
COPD
Non-COPD smoker
Never smoker
Component 1
Component 2
B
COPD
Non-COPD smoker
Never smoker
Component 1
Component 2
C
COPD
Non-COPD smoker
Never smoker
Density
Pseudotime
AT1-1
AT1-2
AT1-3
AT2-1
AT2-2
AT2-3
AT2-4
AT2-5
AT2-6
AT2-7
D
Cluster
E
AT cluster
Group
F
All AT2 clusters
AT2-1
AT2-2
AT2-3
AT2-4
AT2-5
AT2-6
AT2-7
Cell variation
Never smoker
Non-COPD smoker
COPD
qvalue
Ratio

Figure 5

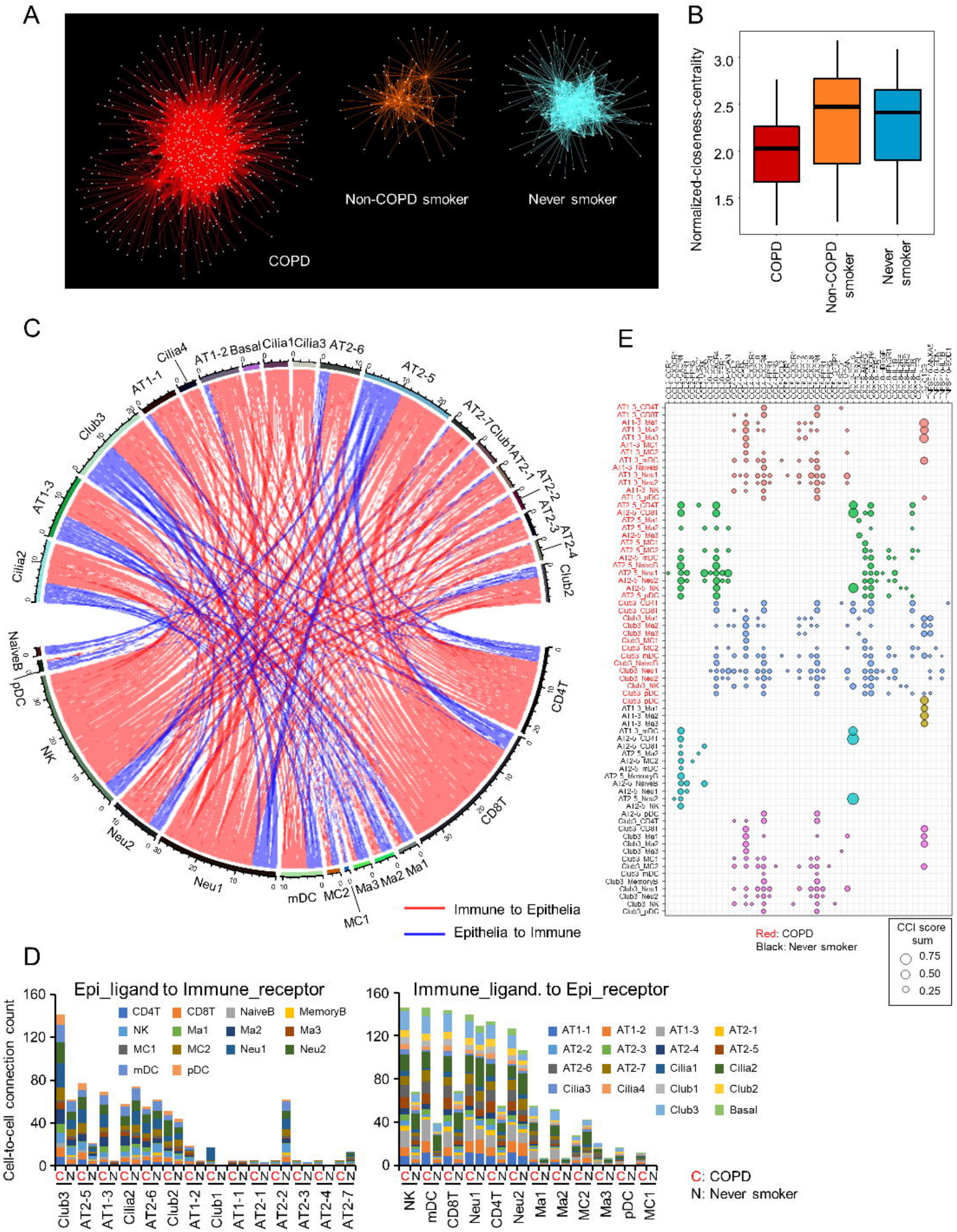
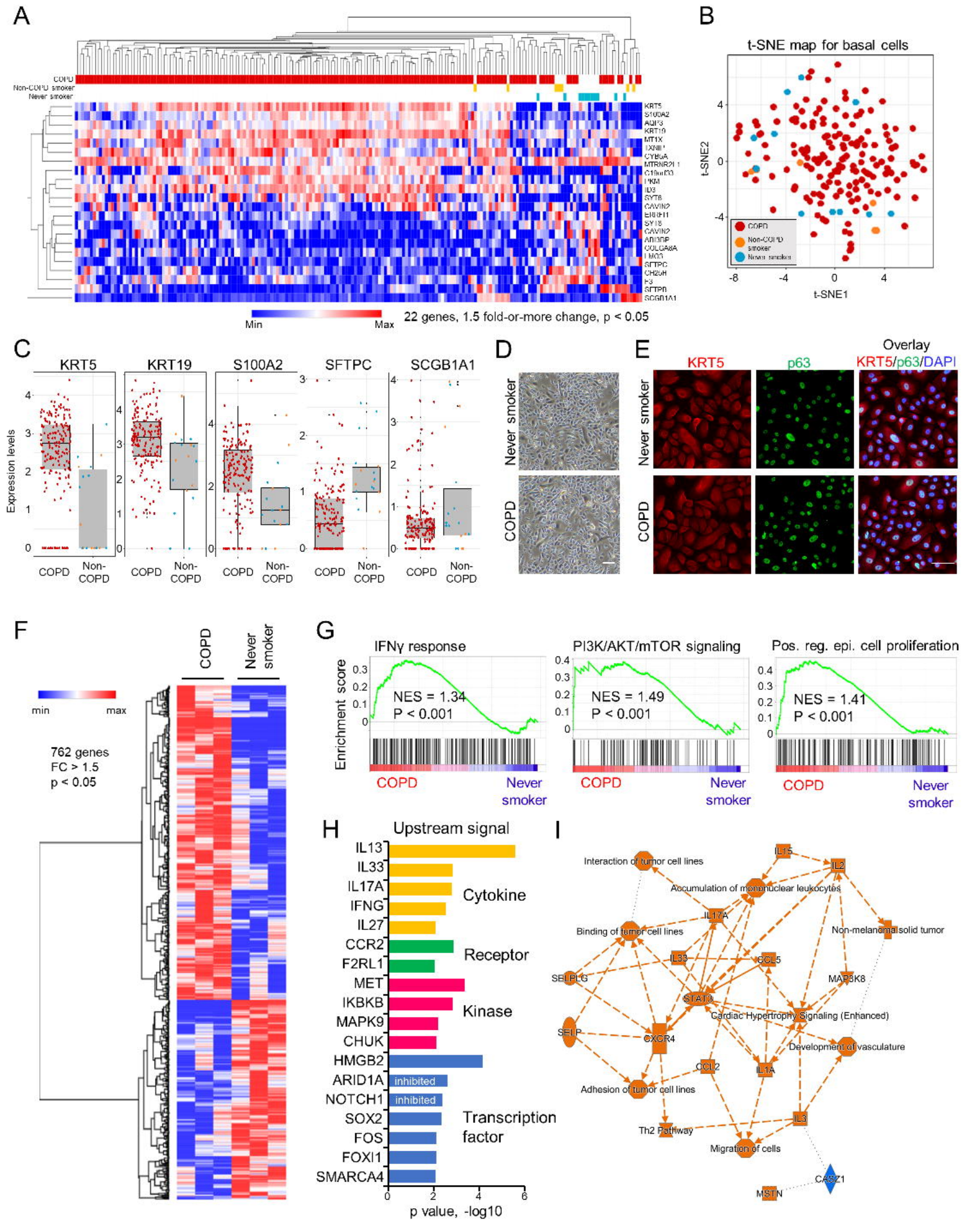
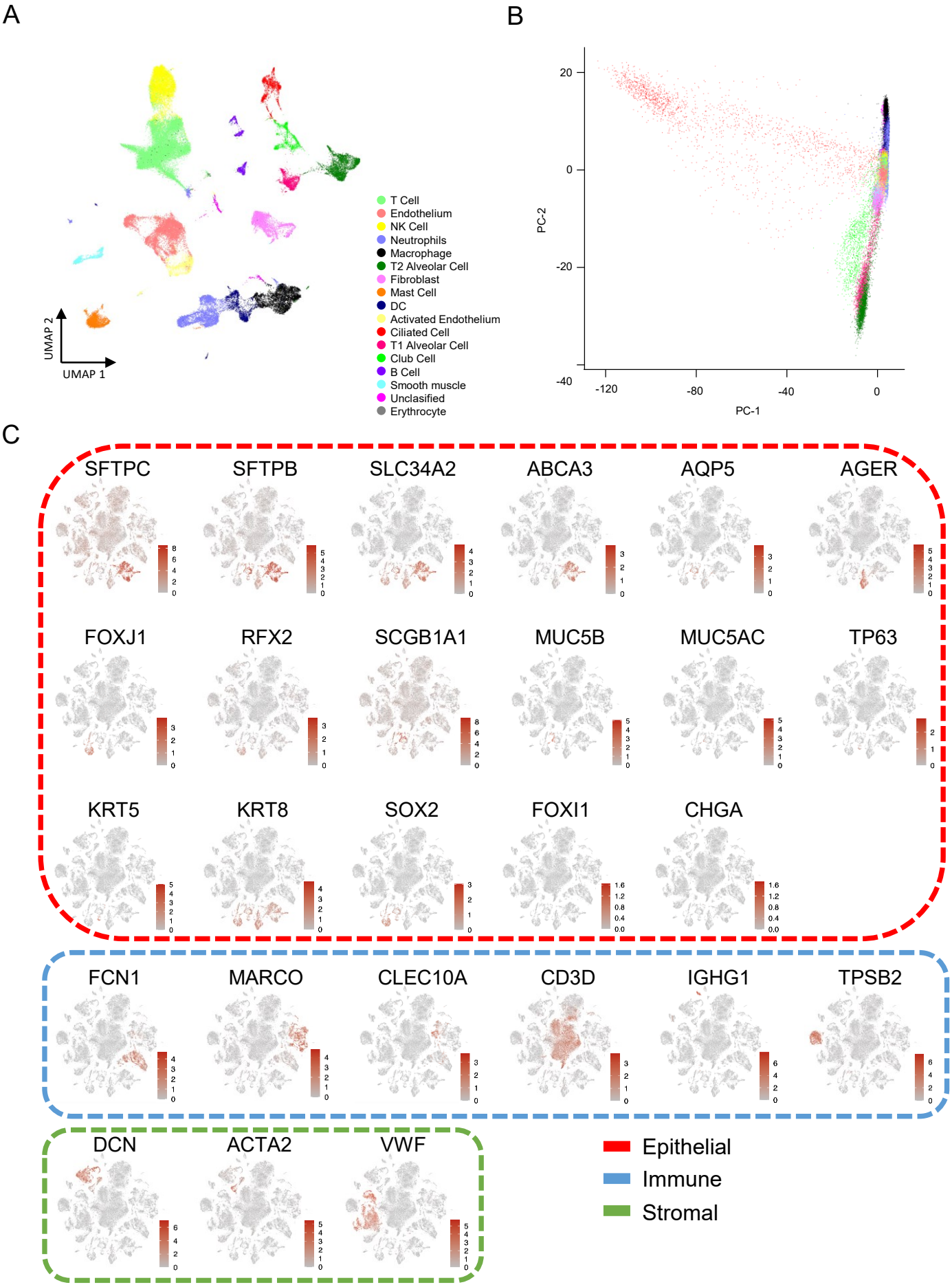

A
COPD
Non-COPD smoker
Never smoker
B
Normalized-closeness-centrality
3.0
2.5
2.0
1.5
COPD
Non-COPD smoker
Never smoker
C
Cilia4
AT1-2
Basal
Cilia1
Cilia3
AT2-6
AT2-5
AT2-7
Club1
AT2-1
AT2-2
AT2-3
AT2-4
Club2
CD4T
CD8T
Ma1
Ma2
Ma3
MC2
MC1
mDC
Neu1
Neu2
NK
pDC
NaiveB
Cilia2
AT1-3
Club3
AT1-1
Immune to Epithelia
Epithelia to Immune
E
Red: COPD
Black: Never smoker
CCI score sum
0.75
0.50
0.25
D
Epi_ligand to Immune_receptor
Cell-to-cell connection count
CD4T
CD8T
NaiveB
MemoryB
NK
Ma1
Ma2
Ma3
MC1
MC2
Neu1
Neu2
mDC
pDC
Club3
AT2-5
AT1-3
Cilia2
AT2-6
Club2
AT1-2
Club1
AT1-1
AT2-1
AT2-2
AT2-3
AT2-4
AT2-7
Immune_ligand. to Epi_receptor
AT1-1
AT1-2
AT1-3
AT2-1
AT2-2
AT2-3
AT2-4
AT2-5
AT2-6
AT2-7
Cilia1
Cilia2
Cilia3
Cilia4
Club1
Club2
Club3
Basal
NK
mDC
CD8T
Neu1
CD4T
Neu2
Ma1
Ma2
MC2
Ma3
pDC
MC1
C: COPD
N: Never smoker

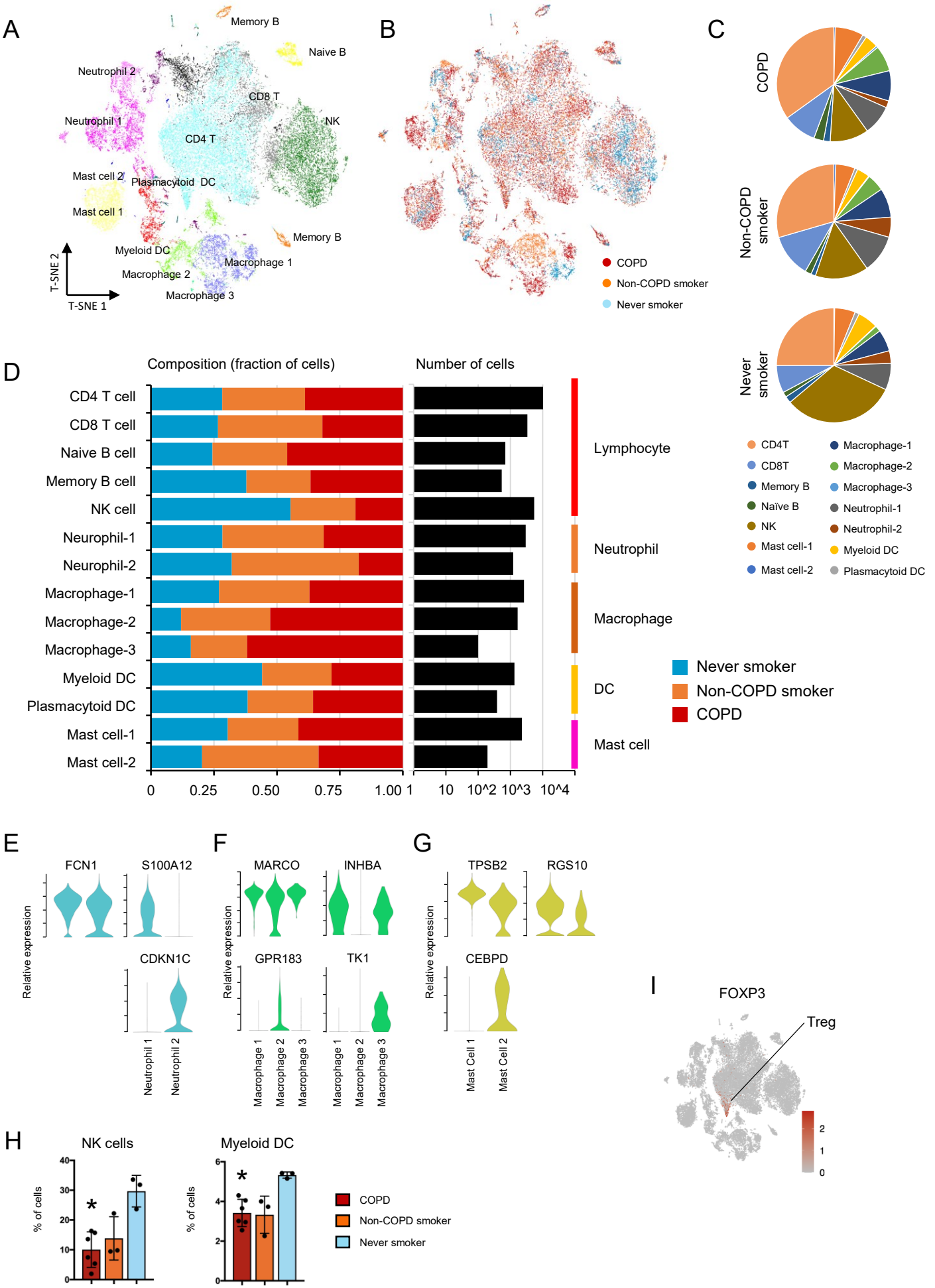
Figure 6


A
COPD
Non-COPD smoker
Never smoker
KRT5
S100A2
AQP3
KRT19
MT1X
TXNIP
CYB5A
MTRNR2L1
C19orf33
IKM
ID3
SYT8
CAVIN2
EIF3H1
SYT8
CAVIN2
ARI3DP
COLGA8A
LMO3
SFTPC
CH25H
F3
SFTPB
SCGB1A1
Min
Max
22 genes, 1.5 fold-or-more change, p < 0.05
B
t-SNE map for basal cells
t-SNE2
t-SNE1
COPD
Non-COPD smoker
Never smoker
C
KRT5
KRT19
S100A2
SFTPC
SCGB1A1
Expression levels
COPD
Non-COPD
D
Never smoker
COPD
E
KRT5
p63
Overlay
KRT5/p63/DAPI
Never smoker
COPD
F
COPD
Never smoker
min
max
762 genes
FC > 1.5
p < 0.05
G
IFNγ response
NES = 1.34
P < 0.001
PI3K/AKT/mTOR signaling
NES = 1.49
P < 0.001
Pos. reg. epi. cell proliferation
NES = 1.41
P < 0.001
Enrichment score
COPD
Never smoker
H
Upstream signal
IL13
IL33
IL17A
IFNG
IL27
CCR2
F2RL1
MET
IKBKB
MAPK9
CHUK
HMGB2
ARID1A
NOTCH1
SOX2
FOS
FOXI1
SMARCA4
inhibited
inhibited
Cytokine
Receptor
Kinase
Transcription factor
p value, -log10
I
Interaction of tumor cell lines
IL15
IL2
Accumulation of mononuclear leukocytes
IL17A
Binding of tumor cell lines
Non-melanoma solid tumor
IL33
CCL5
SELPLG
MAP3K8
STAT3
Cardiac Hypertrophy Signaling (Enhanced)
SELP
CXCR4
Development of vasculature
CCL2
IL1A
Adhesion of tumor cell lines
IL3
Th2 Pathway
Migration of cells
CASZ1
MSTN

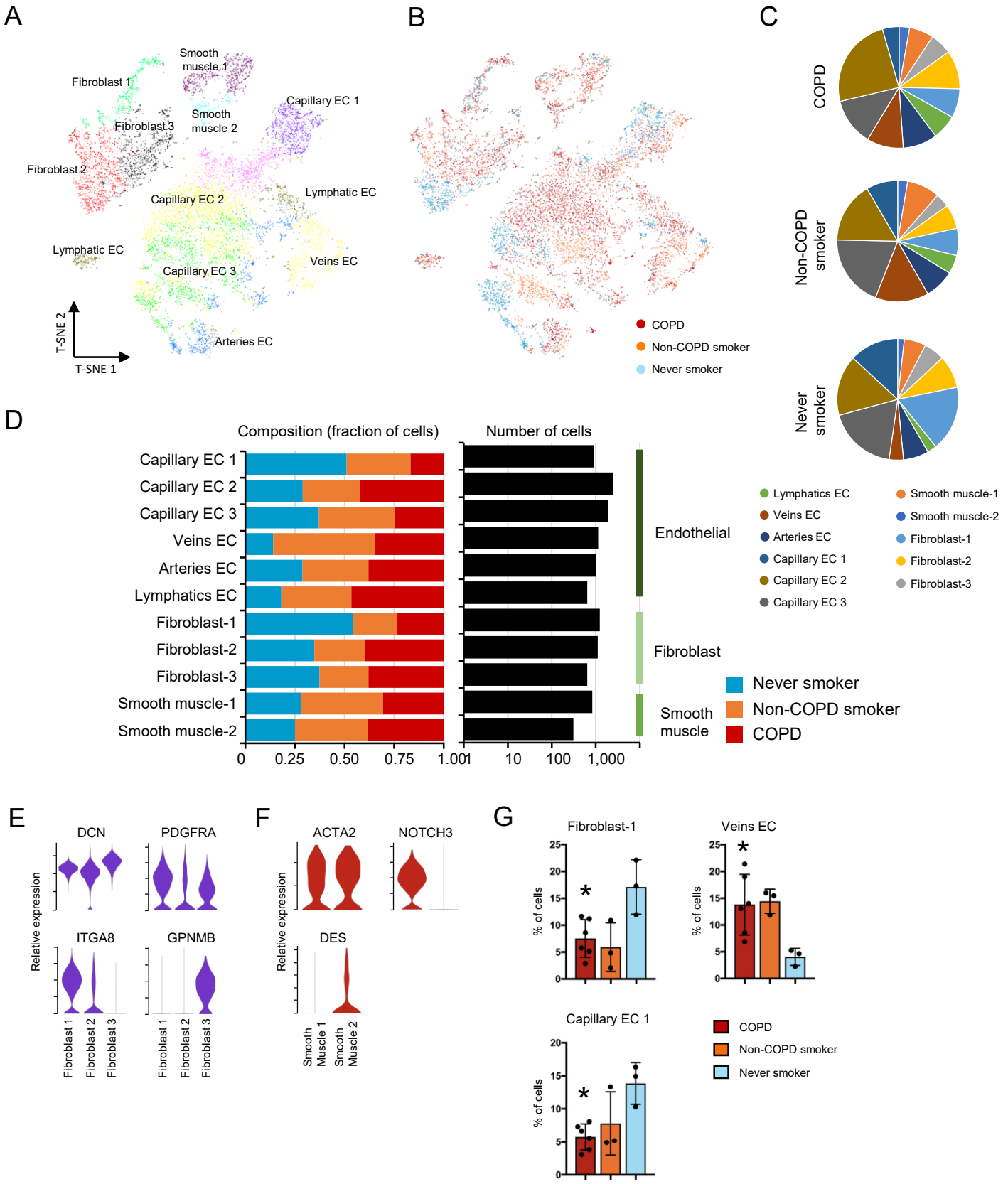
Extended Data Figure 1



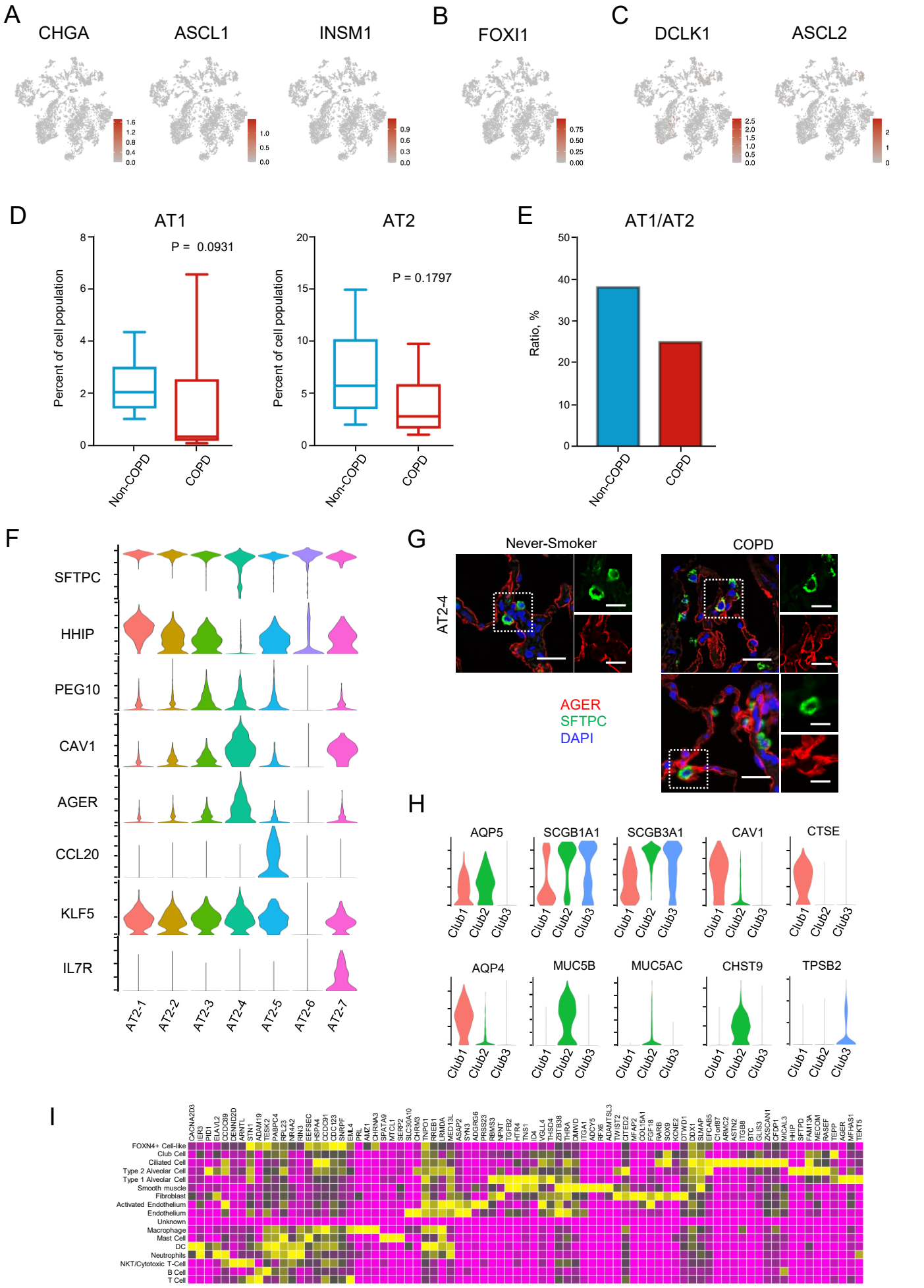
Extended Data Figure 2



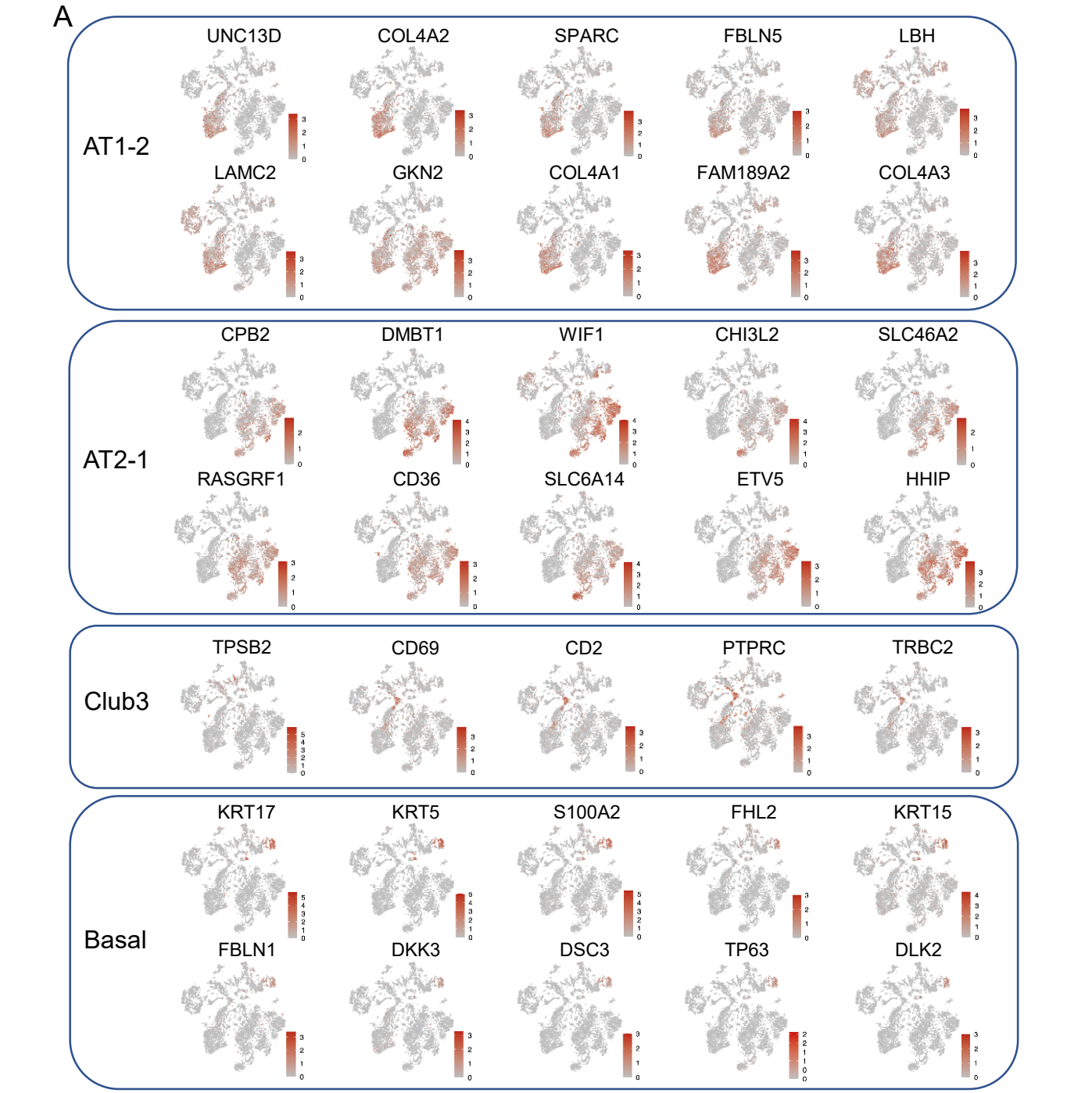
Extended Data Figure 3



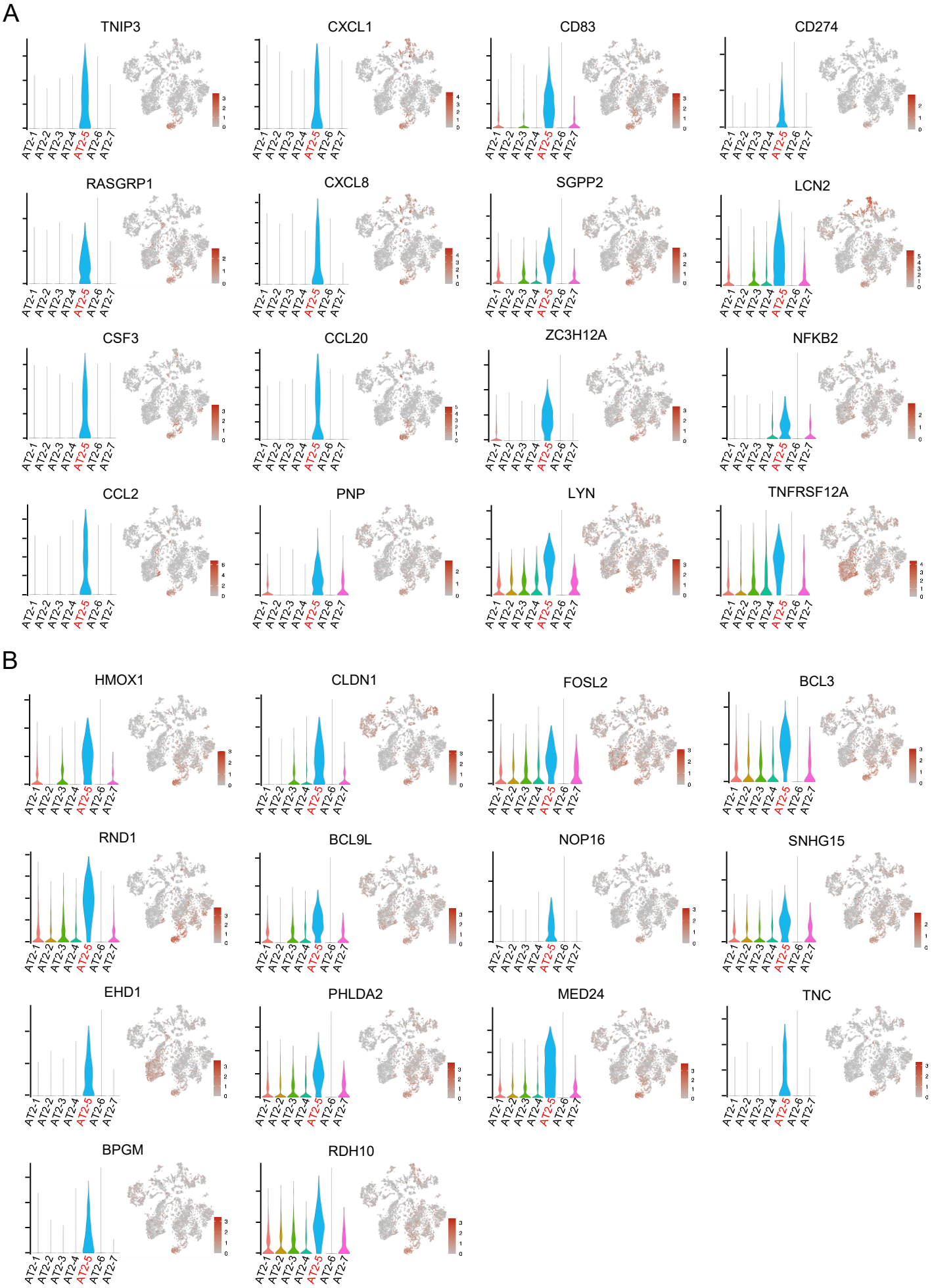
Extended Data Figure 4



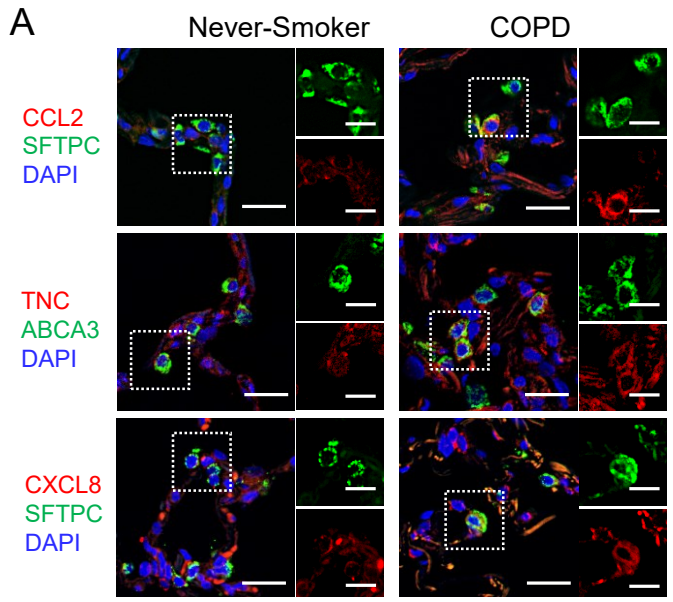
Extended Data Figure 5

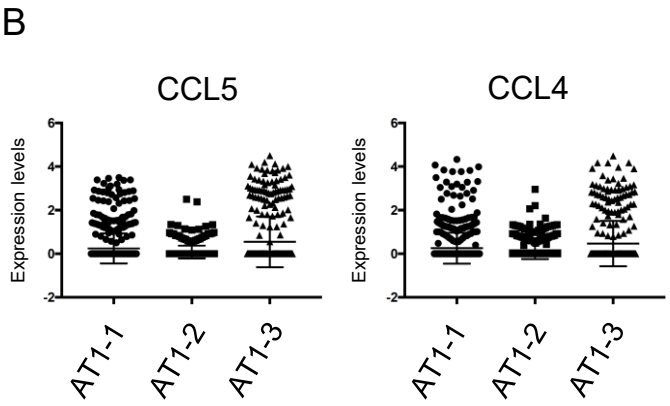


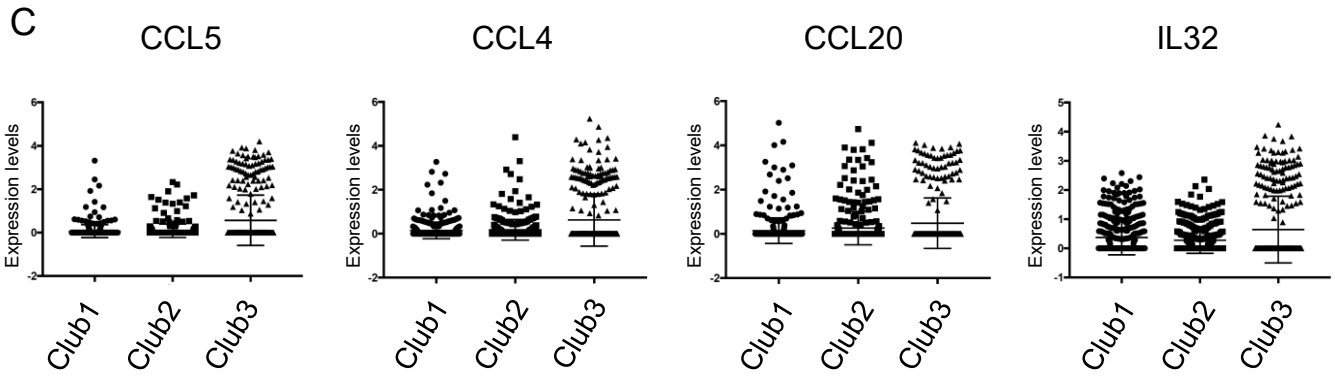
Extended Data Figure 6



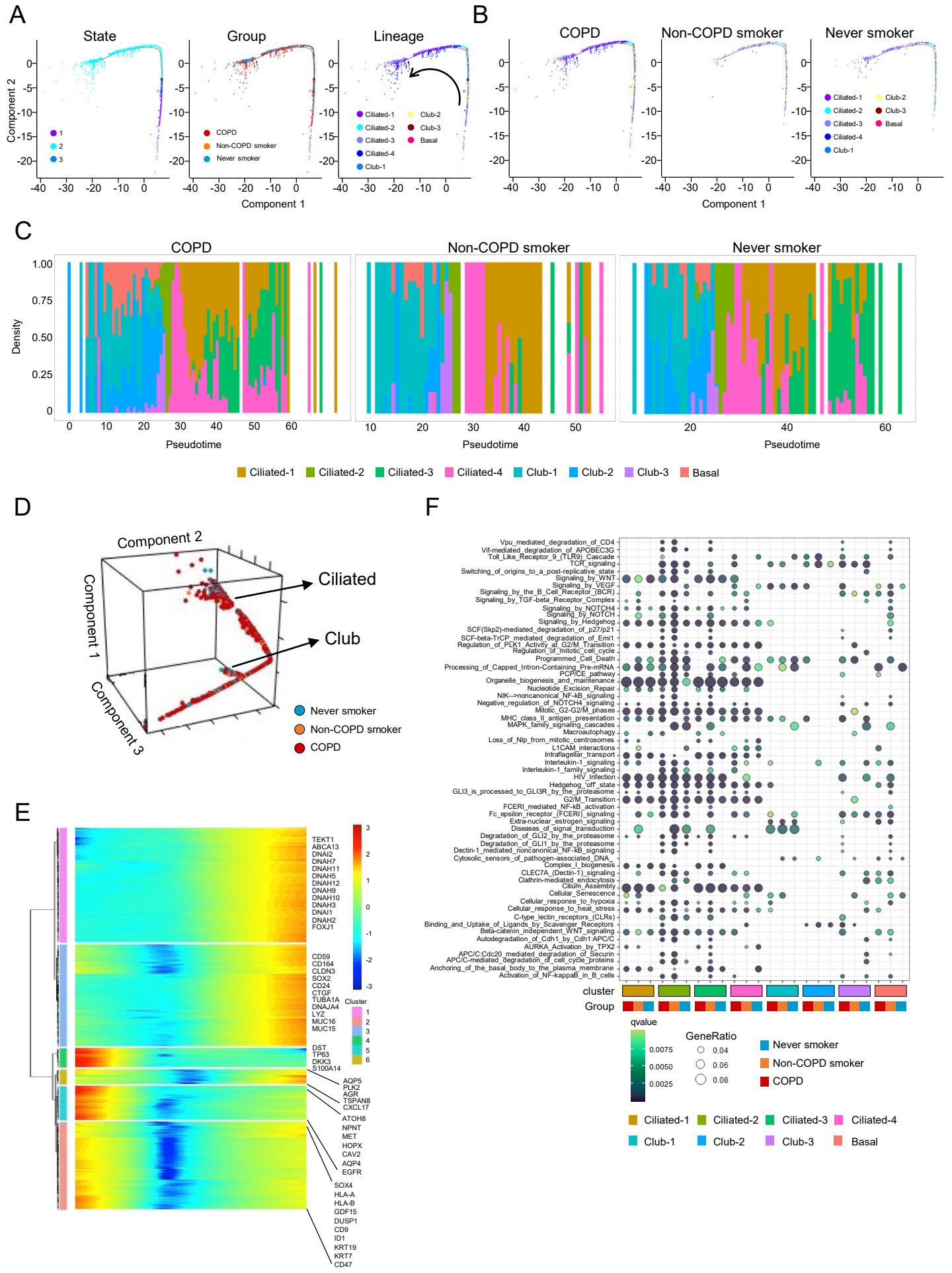
Extended Data Figure 7



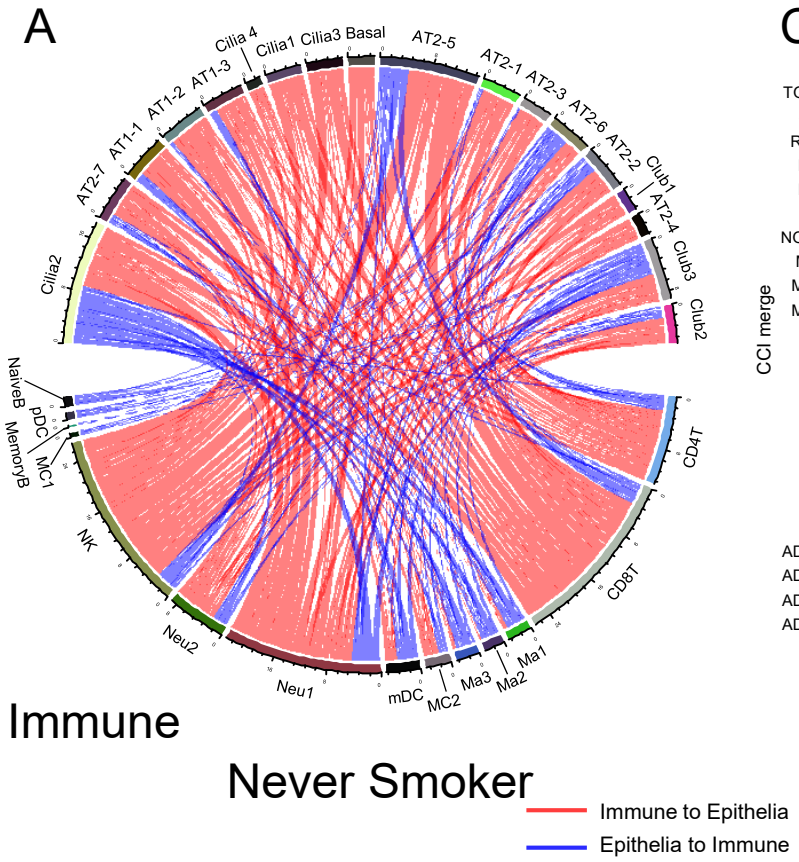


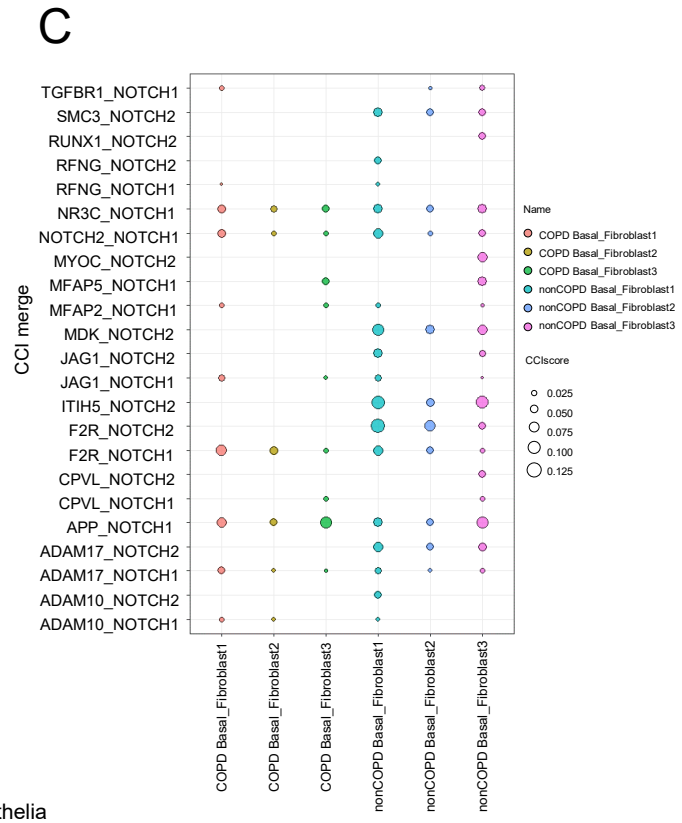


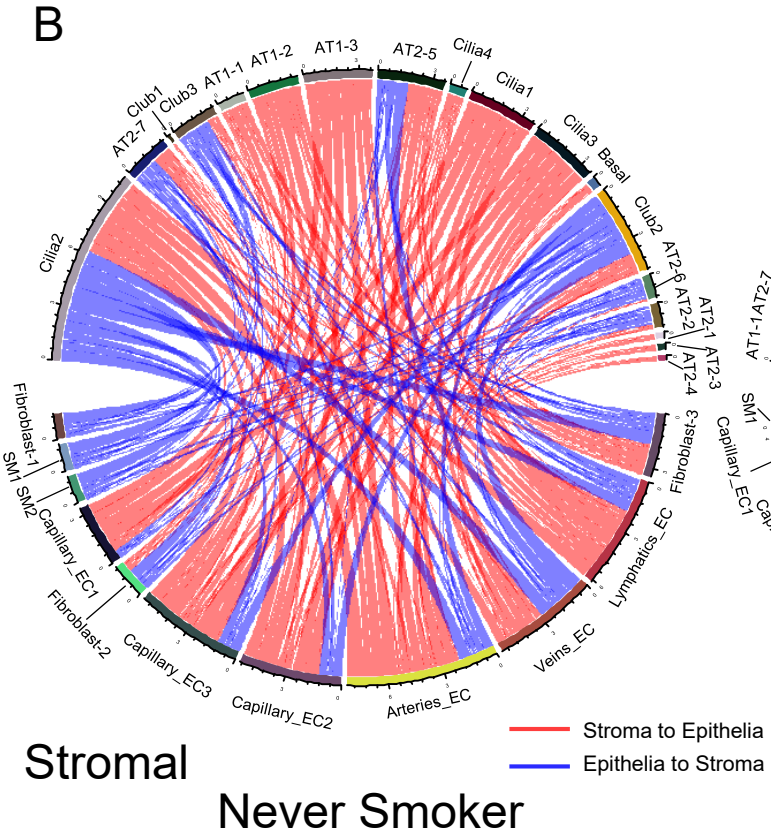
Extended Data Figure 8

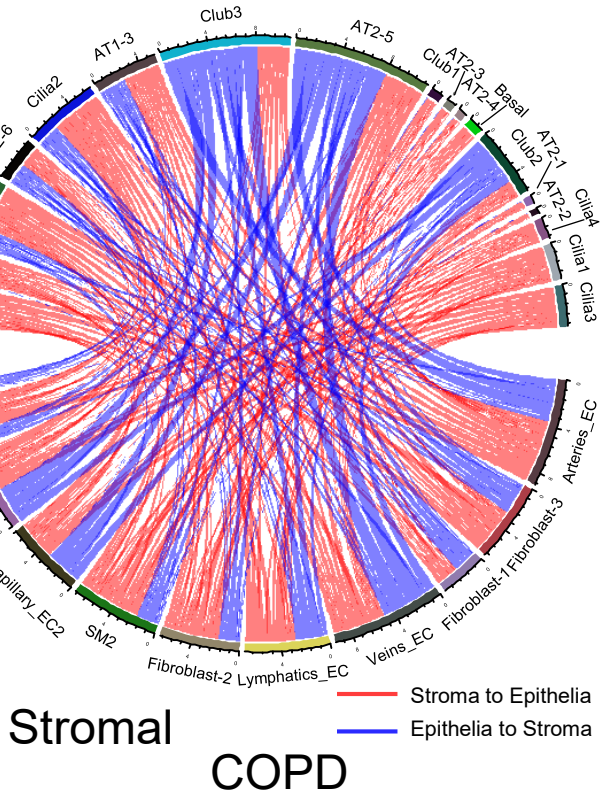
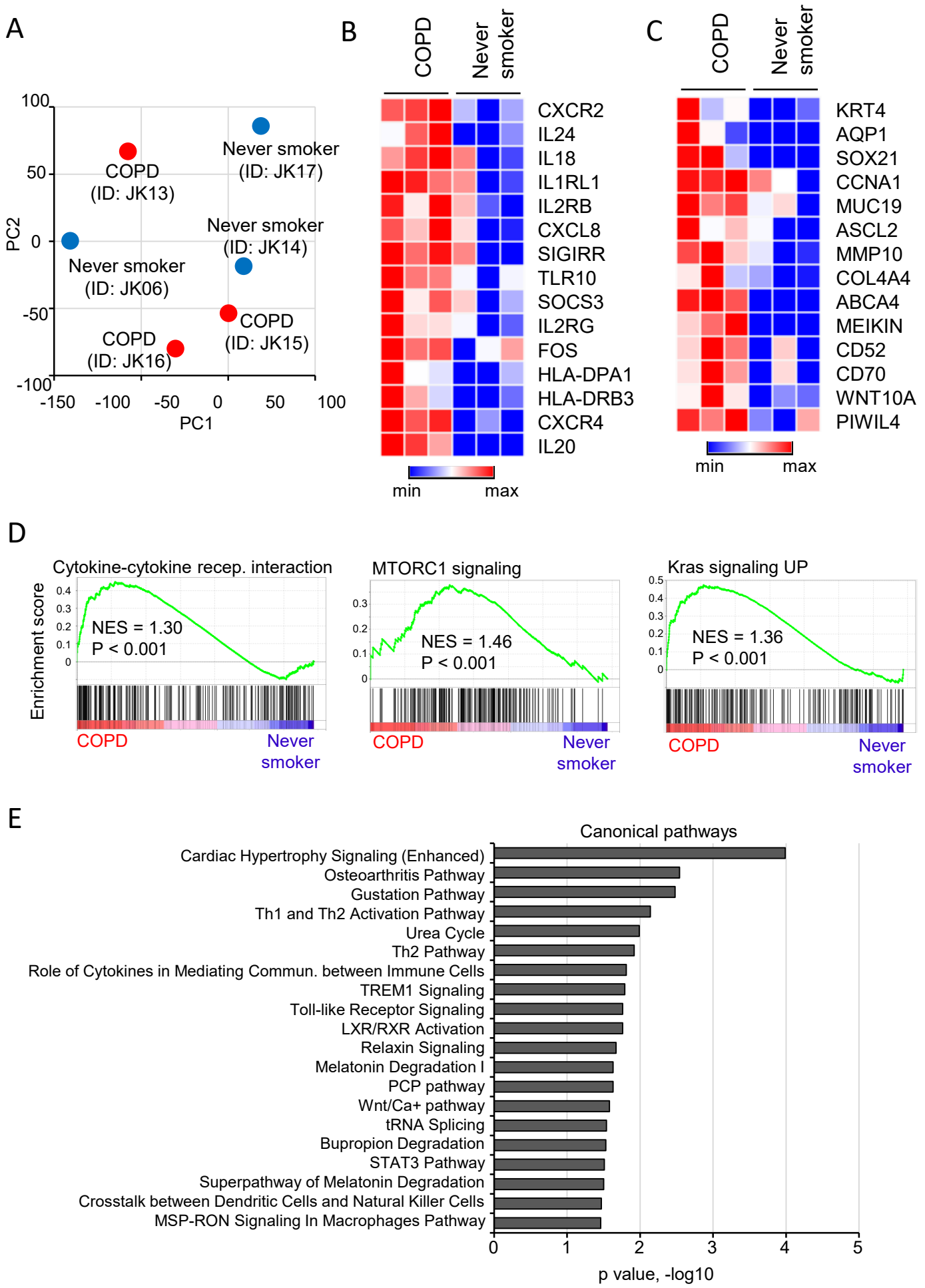


Extended Data Figure 9









Extended Data Figure 10