

# A topographic lung cell atlas reveals regional variation in cell-type specific gene programs and identifies healthy and diseased cellular neighborhoods

**Christos Samakovlis**

`christos.samakovlis@su.se`

Stockholm University, Science for Life Laboratory

**Alexandra Firsova**

Stockholm University, Science for Life Laboratory

**Sergio Marco Salas**

Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden <https://orcid.org/0000-0002-4636-0322>

**Louis Kümmerle**

Helmholtz Zentrum München

**Xesus Abalo**

Science for Life Laboratory, Kungliga Tekniska Högskolan

**Ludvig Larsson**

Science for Life Laboratory, Kungliga Tekniska Högskolan

**Krishna Mahubani**

University of Cambridge <https://orcid.org/0000-0002-1327-2334>

**Alexandros Sountoulidis**

Science for Life Laboratory, Stockholm University

**Jonas Theelke**

Science for Life Laboratory, Stockholm University <https://orcid.org/0000-0002-5074-1793>

**Zaneta Andrusivova**

Science for Life Laboratory, Kungliga Tekniska Högskolan

**Leire Alonso Galicia**

Science for Life Laboratory, Kungliga Tekniska Högskolan

**Andreas Lontos**

Science for Life Laboratory, Stockholm University

**Tamas Balassa**

HUN-REN Biological Research Centre

**Ferenc Kovacs**

HUN-REN Biological Research Centre

**Peter Horvath**

Helmholtz Zentrum München

**Yuexin Chen**

German Center for Lung Research, Helmholtz Zentrum München

**Janine Gote-Schniering**

German Center for Lung Research, Helmholtz Zentrum München

**Mircea-Gabriel Stoleriu**

University Hospital of the Ludwig-Maximilians University, Asklepios Medical Center, Helmholtz-Zentrum München

**Jürgen Behr**

University Hospital of Ludwig-Maximilians-University, Department of Internal Medicine V, Munich, Member of the German Center for Lung Research (DZL), Germany

**Kerstin Meyer**

Wellcome Sanger Institute

**Wim Timens**

University of Groningen, University Medical Center Groningen

**Herbert Schiller**

German Center for Lung Research, Helmholtz-Zentrum München, Ludwig-Maximilians University

**Malte Lücken**

Helmholtz Zentrum München, Institute of Lung Health and Immunity, Comprehensive Pneumology Center

**Fabian Theis**

Helmholtz Munich <https://orcid.org/0000-0002-2419-1943>

**Joakim Lundeberg**

KTH-Royal Institute of Technology <https://orcid.org/0000-0003-4313-1601>

**Mats Nilsson**

Science for Life Laboratory <https://orcid.org/0000-0001-9985-0387>

**Martijn Nawijn**

University Medical Center Groningen <https://orcid.org/0000-0003-3372-6521>

---

**Article**

**Keywords:**

**Posted Date:** October 7th, 2024

**DOI:** <https://doi.org/10.21203/rs.3.rs-5046381/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** **Yes** there is potential Competing Interest. MDL contracted for the Chan Zuckerberg Initiative and received speaker fees from Pfizer and Janssen Pharmaceuticals. JL and ZA are the scientific consultants for 10x Genomics Inc which holds intellectual property rights to the spatial transcriptomics technology. FJT is a scientific consultant for Immunai Inc., Singularity Bio B.V., CytoReason Ltd, Cellarity, Curie Bio Operations, LLC, and has an ownership interest in Dermagnostix GmbH and Cellarity. MN is co-founder of CARTANA.

---

# Abstract

Integration of single cell mRNA sequencing data from millions of cells revealed a high diversity of cell types in the healthy and diseased human lung. In a large and complex organ, which is also constantly exposed to external agents, it is crucial to understand the influence of lung tissue topography or external factors on gene expression variability within each cell type. Here, we applied three spatial transcriptomics approaches, to: (i) localize the majority of lung cell types, including rare epithelial cells within the tissue topography, (ii) describe consistent anatomical and regional variability in gene expression within and across cell types, and (iii) reveal distinct cellular neighborhoods for specific anatomical regions and examine gene expression variations in them. We thus provide a spatially resolving tissue reference atlas including cell type composition and gene expression variations in three representative regions of the healthy human lung. We further demonstrate its utility by defining previously unknown imbalances of epithelial cell type compositions in diseased tissue from patients with stage II COPD. Our topographic atlas enables a precise description of characteristic regional cellular responses upon experimental perturbations or during disease progression.

## Introduction

Recent advances in single cell-omics have led to extensive reference datasets of cell types from various human organs, including the lung, harboring the respiratory system with its multitude of cell types and states [1–5]. Single cell mRNA sequencing (scRNA-seq) unveiled previously unknown cell types in healthy human lung, such as ionocytes, hillock-like and tuft cells, neuroendocrine (NE) cell states and arocytes [3, 4, 6–8]. Multiple subtypes of fibroblasts, immune and endothelial cells were also characterized based on their gene expression programs and inferred tissue distribution [4]. A limitation of these datasets deriving from dissociated tissue is that cell types and their annotation often lack the information about cellular location relative to tissue landmarks or along relevant axes in the tissue such as the proximo-distal axis in lung with gradients in for instance oxygen tension or airway diameter that are likely to impact on cell type variation. Moreover local intercellular interactions can only be predicted based on selectivity of ligand/receptor pair expression at the mRNA level in cell-type pairs, but lack information about spatial proximity which is especially relevant at higher resolutions of cell subset annotation. On the other hand, creation of spatially resolved gene expression maps can highlight physical proximity of cells and establish consistent cellular neighborhoods allowing focused analysis of cell-cell interactions. Such topographic atlases of healthy tissues in comparison to corresponding maps of diseased tissues can reveal potentially causative alterations in the cellular landscape in diseased organs.

The large size of the adult human lung precludes in-scale mapping of the full tissue. Instead, available atlases of the adult human lung rely on sampling of distinct anatomic regions from different donors [1–4, 6, 9–14]. For example, extensive proximal airway epithelial sampling allowed identification of nasal cell types, including nasal-specific serous, goblet and club cells [3]. Another thorough characterization of distal airways focused on bronchial secretory cell populations, and defined the AT0 intermediate cell

state, characterized by the co-expression of bronchial secretory and alveolar epithelial Type 2 (AT2) markers [9]. A recent integration of scRNA-seq data has created a comprehensive cellular catalog of the healthy human lung, counting a total of 61 major cell clusters, 58 of those from the trachea and the lung. This defined 35 major cell types with 51 subtypes and seven cell states [2]. Most subtypes were annotated by sampling location and others by differential expression of single or few distinguishing markers. Cell states of different cell types were further classified as resting, proliferating, activated, or intermediate, suggesting the presence of cells in distinct steps of differentiation progression. The potentially distinct cellular environments of such transition states remain unknown.

Here, we generated a representative topographic atlas of the healthy adult lung by combining sections from distinct anatomic locations of the respiratory system from four donors of different age and gender. We used three different, multiplexed spatially-resolved transcriptomic (SRT) approaches with high resolution to obtain complementary results. We confirmed and deeply characterized consistent, location-related gene expression variability within and across cell types. Finally, we used the topographic atlas as a reference to define changes in cell-type and cell-state abundance and their distribution in distal lung samples from three stage-II COPD patients revealing distinct cellular niches at an early stage of disease progression.

## Results

# A HybISS-based cell type map reveals specific cellular neighborhoods

We collected tissue samples from six donors targeting five discrete anatomical regions, congruent to the previously described locations of cells in scRNA-seq datasets [1, 15], and grouped them into three major anatomical regions: trachea (ventral side of the airway with surrounding mesenchyme), proximal lung (generation 2–3 intralobar bronchus with surrounding mesenchyme and occasionally alveoli) and distal lung (distal/terminal bronchioli and alveolar tissue close to the edge of the lobes). After histology-based assessment, two out of six donors were excluded due to multiple signs of pathology, including fibrosis or large immune infiltrations. Samples from the remaining four donors (Suppl. Table 1) were subjected to mRNA quality controls to reject the samples with low or diffuse RNA signal (Methods, Suppl. Figure 1A). We selected high-quality samples from different locations and applied three different complementary SRT technologies (Fig. 1A, Suppl. Table 2). First, we used RNA-rescue Spatial Transcriptomics (RRST) for unbiased mapping of gene expression on the tissue sections. This broad regional mapping was complemented by SCRINSHOT and HybISS [16, 17], aimed to simultaneously detect either 23 cell types with additional intra-population gene expression variability (SCRINSHOT), or all 35 cell types (HybISS) at cellular resolution. We classified cells based on previously published scRNA-seq data (Suppl. Figure 1B) [1, 3] and generated probe panels for targeted methods using Spapros, as previously described, and partially validated using SCRINSHOT [18]. Cross-validation between methods, including the unbiased SRT

method (Visium/RRST) on serial tissue sections, demonstrated consistent cell type marker gene expression patterns, and therefore the accuracy of probe selection (Suppl. Figure 1C).

First, we identified cell types and their spatial distribution by profiling 14 sections from four donors by HybISS using a gene panel consisting of 162 genes (Suppl. Table 3). After decoding, cells were segmented based on DAPI-stained nuclei using an AI-based deep neural network for segmentation in the BIAS software (Methods). Fluorescent signals were assigned to cells (regions of interest, ROIs) using Baysor (Methods) [18, 19]. We excluded cells with low transcript counts and finally processed a total of 260,398 cells for further analysis and clustering based on their expression profiles. This separated the cells into six major classes, assigned according to marker gene expression: airway epithelial, immune, alveolar epithelial, endothelial, stromal, and submucosal gland (SMG) (Suppl. Tables 3, 4). The cells in these classes mapped to their expected histological locations (Suppl. Figure 2A). By further subclustering of each class, we revealed and annotated 28 cell types (Fig. 1B, Suppl. Table 4), corresponding to the majority of the adult lung cell types, described in previous scRNA-seq studies [2, 4]. Based on positivity for corresponding cell type marker genes in the RNA-seq atlases [1–3], we manually annotated seven additional cell types that could not be assigned by the unsupervised sub-clustering of the HybISS data either due to their low abundance or sparse gene expression (T lymphocytes, NK cells, a mixed group of T and NK cells, ionocytes, tuft cells, rare tuft-like cells, squamous-like cells and aerocytes; Suppl. Figure 2B, Suppl. Table 5). Therefore, our analysis resulted in identification of a total of 35 cell types that were mapped onto the tissue topography *in situ* (Fig. 1B, Suppl. Figure 2C-D). All data are deposited in an open access searchable browser that visualizes primary signals, cell type annotation, gene expression levels and histological stainings (see Data Availability in viewers for HybISS Atlas).

Complementing the HybISS datasets, we analyzed sequential sections of the same tissue blocks using the RNA-rescue Spatial Transcriptomics (RRST) modification of the Visium protocol due to the tissue-specific challenges presented in profiling lung sections [20]. This technique allows targeting the mRNA sequences directly instead of their poly-A tails. We used the Stereoscope method to deconvolve the cell type composition of each spot using the finest annotation from scRNA-seq dataset from Madisson et al [1] as a reference. Lastly, we profiled sequential sections by a highly sensitive mRNA detection method (SCRINSHOT) [16], employing a gene panel of 64 marker genes also selected using Spapros to assign cells to clusters according to marker gene positivity (Suppl. Table 6). The location of assigned cell types within the tissue was consistent between all three methods (Suppl. Figure 3), confirming the specificity of each of the three technologies and the robustness of the combinatorial approach, which overcomes limitations of individual spatial mapping protocols, such as resolution or limited gene panel.

We defined cell type compositions across tissue locations by calculating the relative frequency of each cell type within each profiled region (Fig. 1C). We treated distal regions as a single location due to similarity in cellular composition. Several cell types exhibited a regional preference, for example AT1 and AT2 epithelial cells were mostly present in distal lung, whereas B plasma and SMG cells mainly occupied tracheal regions (Fig. 1C-D). To further dissect the relative spatial distributions of cell types and describe consistent cellular colocalizations across the mature lung, we performed neighborhood proximity

enrichment analysis of all datasets. This revealed multiple consistent cellular colocalizations, which included most cell types, except tuft and lymphatic endothelial cells. These colocalizations were combined into larger neighborhoods, which were characterized by defined histological features (Fig. 1D). In addition to the SMG and airway epithelial neighborhoods, we also revealed a group of cell types in proximity to AT1, AT2 and aerocytes, which included stromal, endothelial, monocytes and NK cells (Fig. 1E). This neighborhood was therefore labeled alveolar parenchyma. A distinct neighborhood composed of adventitial fibroblasts, venous and immune cells, was revealed in both peri-bronchial and peri-SMG locations (Fig. 1D-E). In summary, our combined data provide an overview of 35 cell types and their occurrences in the topographic map of healthy adult lung, and define distinct cellular neighborhoods based on cell type proximity. These neighborhoods as well as their gene expression profiles were presented in more detail below.

## Multiple cell states with distinct topologies in the airway epithelium

We first focused on the topography of cell diversity in the airway epithelium. Among bronchial epithelial cells, a total of 11 cell types were identified, including basal, suprabasal, ciliated, deuterosomal, and neuroendocrine cells, as well as manually-assigned ionocytes, tuft (brush), rare tuft-like and squamous-like cells (Fig. 1B, Suppl. Figure 2B, Suppl. Table 4). The remaining bronchial epithelial cells were split into two groups. First, secretory cells expressing the *AGR2*, *SFTPB*, *WFDC2*, *MUC5AC* markers and comprising 33% of total bronchial epithelial cells and second a smaller group comprising 12% of total bronchial epithelial cells, which were positive for the general epithelial marker genes *SLPI* (Suppl. Table 4). These cells were found spread along the airways, but also occasionally in SMGs and alveoli. However, they were negative for the characteristic epithelial cell type markers, such as mucins or secretoglobins and were designated not annotated 'nan' cells. They could represent less differentiated epithelial cells or unknown cell states (Fig. 1B, Suppl. Figure 2B, Suppl. Table 4).

To investigate cell type composition diversity along the airway proximal-distal axis, we further characterized the composition of the airway epithelium in tracheal, proximal and distal airway sections from individual donor samples by HybISS. In the trachea, the epithelium was dominated by suprabasal cells, whereas in the intralobar airways the epithelium mainly composed of secretory and ciliated cells. Gene expression comparison across the regions confirmed this distribution, with basal (*KRT5*, *KRT15*), squamous (*SPRR3/1B*) and suprabasal (*S100A2*) genes expressed predominantly in the trachea. In addition, this analysis revealed further variability across the regions, including, for example, mesothelin (*MSLN*) expression in trachea, trefoil factor 3 (*TFF3*) and *SLPI* in the proximal lung, and surfactant protein B genes (*SFTPB*) in the distal lung (Fig. 2A, B). These variable genes could mainly be attributed to distinct secretory cell populations or regional variations in the secretory cell transcriptomes. Statistical analysis of all four donors confirmed a significant dominance of *AGR2*-positive populations in the trachea compared to other regions, and the higher abundance of *SFTPB*-positive populations in distal lung, compared to the trachea (Suppl. Figure 4A). This analysis identifies consistent differences in epithelial composition between three anatomical regions along the proximo-distal axis of the airway tree.

To further define the location of the major secretory cell types, we quantified gene expression by SCRINSHOT, targeting characteristic cell type markers for goblet, club and pre-terminal bronchiole epithelial cells (pre-TB or TASC or RAS) [10, 13, 21] in three different locations (Suppl. Tables 6, 7). We found club cells in all three anatomical regions but localized goblet cells in trachea and proximal lung and pre-TB cells only in distal lung [10, 22] (Suppl. Figure 4B). Interestingly, some of the club cells in proximal regions co-expressed low levels of mucins, and some of the distally-located ones expressed *SCGB3A2*. Moreover, in contrast to pre-TB cells, the distally-located club cells expressed genes encoding antimicrobial proteins (*LTF*, *LCN2*, and *BPIFB1*, see online atlas for SCRINSHOT), suggesting a specialized role in epithelial immunity. Both distal club and pre-TB cells were located in small clusters along distal bronchi and respiratory bronchioles (Fig. 2C). We also detected terminal respiratory bronchiolar (TRB) secretory and alveolar type 0 (AT0) cells in peri-bronchial and alveolar regions respectively (Fig. 2C). AT0 cells were defined by co-expression of the alveolar type II cell marker *NAPSA*, and low but evident levels of either *SCGB3A2* [9], or *SCGB3A1* or *LCN2*, suggesting additional heterogeneity in this cell type (Fig. 2C).

The analysis so far revealed large gene expression heterogeneity in the thin distal airway epithelium and a dominant abundance of suprabasal epithelial cells in the thicker tracheal epithelium (Suppl. Figure 4B-C). To investigate the suprabasal cell type further and define its topological relationships within the pseudostratified tracheal epithelium, we investigated its distribution and gene expression in relation to the distance from the basal membrane to the lumen. In the HybISS dataset, basal and suprabasal cells were enriched close to the epithelial basement membrane. In contrast, secretory and ciliated cells were enriched in more apical positions, ciliated cells being closest to the airway lumen (Fig. 2D).

Quantification of the mRNA signals along the distance from the basal membrane to the lumen defined basally-enriched (*KRT15*, *IFITM1* and *IFITM2*), and apically-enriched mRNAs (*AGR2*, *BPIFB1*, *CAPS*), as well as an intermediately located gene expression program (*SERPINB3*, *SPRR1B*, *SPRR3*, and *HSPB1*) [10, 23] (Fig. 2E, Suppl. Figure 4D). To explore the cellular co-expression of these genes, we performed basal and suprabasal cell subclustering of the HybISS dataset, and identified *KRT5* and *KRT15* double positive and *KRT5* single-positive basal cells, *S100A2* and *KRT5* double positive suprabasal cells, and three intermediately located cell clusters, which expressed low levels of the apically-enriched gene *AGR2* together with one of the following: *KRT5* or *SERPINB3* or *S100A8*. Among these three groups of cells, the *KRT5* and *SERPINB3* positive ones were commonly observed in the intermediate layer of tracheal epithelium, whereas the *S100A8* expressing cells were found dispersed in both intermediate and luminal epithelial layers. The frequency of *S100A8* cells was highly variable among donors (Suppl. Figure 4E-F), possibly due to local responses of the airway epithelium involving *S100A8* and *S100A9* transcription [24]. SCRINSHOT analysis of consecutive tracheal sections from two donors supported this suggestion as *S100A9* expression was similarly observed in the intermediate and apical layers. A subset of *S100A9* cells also expressed *KRT13*, which was mainly found in the intermediate layer of the tracheal epithelium as either patches of cells or solitary cells (Fig. 2F). This supports that the *KRT13* positive cells might be Hillock cells, a distinct tracheal cell state, in line with recent observations [25]. SCRINSHOT analysis also confirmed the distinct distributions of *KRT15* and *S100A2* expressed in the basal, *SERPINB3* in the

intermediate, *LCN2*, *SCGB1A1* and *SCGB3A1* in subluminal, and *MUC5AC*, *MUC5B* and *CAPS* in the luminal layer of tracheal epithelium (Fig. 2F, Suppl. Figure 4C and Data viewer for SCRINSHOT Atlas). This suggests that the characteristic gene expression programs in intermediate layers reflect progressive differentiation states of the tracheal epithelial cells with the most differentiated cells (ciliated and secretory cells) facing the lumen. Our spatial analysis reveals multiple cell states located in distinct layers of the pseudostratified tracheal epithelium. Overall, there is a strong correspondence between gene expression and cellular localization along the proximal-distal and apical-basal axis of the airway epithelium (Suppl. Figure 4G).

## Rare cell type mapping reveals region-specific neuroendocrine cells

The variability of gene expression patterns in the apical epithelial cells along the proximo-distal axis could partly be explained by differential exposure to external factors. Certain environmental factors are sensed by specialized rare cells located in the airways. Ionocytes and tuft cells have been identified in the nasal epithelium and distal airways, whereas neuroendocrine cells were predominantly located in trachea and intermediate airways [3]. Our HybISS-based analysis allowed mapping most of these cell types. Yet, the expression levels of their markers were low, making it difficult to extract safe conclusions regarding their differential distribution. To explore the potential variation and location of rare cells, we developed a marker panel targeting pulmonary ionocytes, tuft cells and neuroendocrine (NE) cells, based on the previously integrated human lung cell atlas from scRNA-seq [2], as well as specific airway epithelial cell types [3] and embryonic single cell atlas studies [26, 27]. Since neuroendocrine cells of adult lung are diverse, a precise selection of markers was performed to uncover potential heterogeneity in neuroendocrine cell phenotypes, targeting the four most abundant adult NE genes, as well as two genes marking a NE population discovered predominantly in the developing embryonic lung [8, 26]. We used these markers in SCRINSHOT and located rare epithelial cell types manually by positivity for the expected markers. We visually analyzed samples from three regions of four donors and selected samples with large parts of the airway (covering a continuous airway length of at least 2 mm per section) for further quantification. We assessed gene expression in 158 rare epithelial cells, and clustered these cells, identifying at least four groups of neuroendocrine cells: (1) NE-GRP, expressing *GRP* and low levels of *ASCL1*, (2) NE-ASCL1 expressing *ASCL1* and low levels of *GRP*, (3) NE-GHRL positive for *GHRL* and *CFC1*, and (4) NE-PCSK1N expressing variable levels of *PCSK1N*, *GRP* and *ASCL1* (Fig. 3A-C). All NE groups sparsely expressed variable levels of *CHGB* and were represented in each donor. *ASCL3* expression defined ionocytes, and cells expressing variable levels of *POU2F3*, *RGS13* and *CRYM* were annotated as tuft cells, the latter including a rare tuft-like cell population expressing previously published markers *NREP* and *HES6* [3] (Fig. 3A-C). In order to create a uniform regional annotation of their positions disregarding the variable epithelial thickness, we quantified rare cell types per length of basal membrane from the selected samples from at least two donors per region (Fig. 3D). Ionocytes were observed in all anatomical locations, preferentially in trachea and proximal bronchi (Fig. 3A, D). Tuft and rare tuft-like cells were mostly located in proximal bronchi, but were observed in other locations along the airway, occasionally in close proximity to other rare cells, but also solitary

(Fig. 3A, C, D). The three neuroendocrine cell identities were observed across locations, but interestingly, GHRL-positive NE cells only appeared in distal bronchioles of three donors, and were not observed in trachea or proximal lung (Fig. 3A, D). GHRL-positive NE cells have previously been detected in embryonic and pediatric datasets and these cells were hypothesized to gradually disappear in adulthood [27, 28]. Our results indicate that targeted spatially-resolved methods allow the detection of low abundant or very rare cell populations with high efficiency, enabling the re-evaluation of the roles of these cells in the lung.

## Specific cell states in distinct tissue compartments

The neighborhood analysis predicts cell niches based on cell proximities of all cell types in the entire tissue. A common classification of tissue compartments uses histologic landmarks and cellular morphology. To complement predicted neighborhoods, we related in the same sections, cellular morphologies in hematoxylin-eosin (H&E) staining with cell-type annotations and gene expression. We defined (i) the SMG and (ii) peri-SMG mesenchyme by selecting the tubular structures located between the airway epithelium and the cartilage, and their surrounding connective tissue (usually 50–100  $\mu\text{m}$  from the basal membrane of the tubular structures, Fig. 4A). The peri-bronchial compartment (iii), which was thick in the trachea (up to 400  $\mu\text{m}$ ) and thinner in distal airways (100–200  $\mu\text{m}$ ), was defined by subepithelial mesenchymal cells, smooth muscle fibers and connective tissue. The alveolar compartments (iv) were defined by alveolar structures, which were not in direct contact with large vessels or airways (Fig. 4A). The remaining histologic regions lacked epithelial structures and were distinguished either by the presence of large vessels or cartilage structures. Vessel compartments were divided into (v) peri-venous and (vi) peri-arterial, according to the histology of the surrounding mesenchyme (including smooth muscle layer or tunica adventitia), which is usually thicker (up to 300  $\mu\text{m}$ ) in the arteries than in veins (up to 100  $\mu\text{m}$ ). Finally, (vii) the peri-chondrial compartment included cartilage and its surrounding peri-chondrial connective tissue (extending up to 100  $\mu\text{m}$ ). These histological subdivisions were largely in agreement with the calculated neighborhoods (Fig. 1E) and covered most of the tissue area. We mapped the 35 cell types and their subtypes in relation to histologically defined tissue compartments, assessing compartment-specific gene expression by three SRT methods.

First, we focused on the epithelial cell types in the submucosal gland structure, which includes a duct protruding from the airway lumen branching into the tubules and acini composed of mucous and serous cells [29]. The acini are sheathed by myoepithelial cells enabling mucus ejection into airway lumen. In the acini and small tubules, we detected SMG mucous and serous cells expressing their corresponding markers (Suppl. Tables 5 and 6). These cells either intermingled with each other or were found in continuous patches of either mucous or serous cells (Suppl. Figure 5B). Additionally, *BPIFB1* was expressed in a subset of mucous and serous cells adjacent to each other (Suppl. Figure 5B, Data viewer for HybISS and SCRINSHOT Atlas), and *SCGB3A2* was expressed in a subpopulation of serous cells, usually located in small tubules and not in the duct [10, 30] (Suppl. Figure 5B). Myoepithelial cells were sparse and located around the SMG acini and tubules (Data viewer for HybISS and SCRINSHOT Atlas). In

the ducts, we detected both mucous and serous cells, surrounded by the layer of basal cells. These duct cells also expressed characteristic airway secretory cell markers (*LCN2*, *ALDH1A3*, *SCGB3A1*) together with either serous or mucous markers (Fig. 1B, Suppl. Figure 5B), and were therefore called SMG intermediate (Suppl. Figure 5A; Data viewer for HybISS Atlas). Overall, we located the major SMG cell types and uncovered additional heterogeneity in the expression of secretory cell markers.

The previously reported description of an SMG immune niche [1], as well as our neighborhood analysis (Fig. 1E) suggests the location of specific cell types around the gland. To extend the description of the SMG niche, we defined all non-epithelial cells of the peri-SMG compartment in the trachea and proximal lung. In the Visium dataset, these cells were represented with large accumulations of *JCHAIN* expressing B plasma cells, intermingled with rare B lymphocyte and T/NK cells and macrophages, as well as *PLA2G2A* positive fibroblasts (annotated adventitial), but also other fibroblasts and venous cells (Suppl. Figure 5C). Fibroblasts, endothelial cells and macrophages could be further split into subclusters by gene expression in the SCRINSHOT dataset, and varied in the different anatomical regions. SMGs in the trachea and proximal bronchi were surrounded by fibroblasts expressing *FBLN1*, as well as smooth muscle and immune cells, which could not be more precisely annotated (nan), due to the absence of B plasma cell markers in the panel (Fig. 4A, B). Interestingly, endothelial cells around the tracheal SMG expressed both *SPARCL1* and *CLDN5*, whereas in the lobes we found either *SPARCL1* or *CLDN5* positive cells potentially corresponding to venous or capillary cells, respectively (Fig. 4A, B). As expected from the neighborhood analysis, the peri-airway compartment contained very similar cell type combinations as the peri-SMG one (Fig. 4A, B, Suppl. Figure 5C-E). However, the peri-bronchial compartments varied in different anatomic locations. For example, we only found ganglia with *VIM* positive cells (annotated Schwann cells according to their morphology) in proximal peri-bronchial and peri-SMG regions. (Fig. 4B, Data viewer for SCRINSHOT Atlas). Additionally, smooth muscle cells were most abundant in proximal bronchi, whereas the distinct populations of endothelial cells expressing either *SPARCL1* (aerocyte or arterial) or *CLDN5* (capillary or arterial) were only found around bronchioles together with *RGCC* expressing fibroblasts and *APOE* expressing macrophages (Fig. 4A, B).

The alveolar parenchyma was defined by the presence of AT1 and AT2 epithelial cells and was dominated by capillaries (including aerocytes), endothelial cells expressing *RAMP2* and non-adventitial (general) fibroblasts, (Suppl. Figure 5E). In comparison to other compartments, the alveolar parenchyma had the highest proportion of *CLDN5* positive endothelial cells (most likely corresponding to alveolar capillaries), and *APOE* macrophages (alveolar macrophages (Fig. 4A, B) [1, 2]. Fibroblasts positive for *RGCC* (alveolar fibroblasts) were dominating in the distal lung. In the HybISS dataset, fibroblasts in the distal lung also expressed higher *FN1* and *RGCC*, and lower *PLA2G2A* and *C3*, compared to the fibroblasts in the other regions (Suppl. Figure 5F). This suggests that gene expression patterns reveal the existence of multiple fibroblast subtypes located in different peri-epithelial tissue compartments.

Large vessels and cartilage, were surrounded by endothelial cells expressing both *CLDN5* and *SPARCL1*, and *FBLN1* positive fibroblasts. The peri-arterial compartment was distinguished by the increased proportion of smooth muscle cells (Fig. 4A, C). The peri-venous compartment was contained small

proportions of all mesenchymal cell types. Chondrocytes only were detected in the Visium dataset (Suppl. Figure 3A). The peri-chondrial regions were composed of *FBLN1* and *PLA2G2A* positive fibroblasts, and occasionally capillaries, pericytes, and mast cells (Data viewer for HybISS Atlas).

Location-specific distributions of cell types and cell states with distinct gene expression patterns in different compartments define cell type niches and inform on potential cell-to-cell signaling domains. Our data reveal an enrichment of *APOE* macrophages and endothelial cells highly expressing *CLDN5* in alveoli. Peri-bronchial and peri-SMG regions on the other hand, were composed of *FBLN1* fibroblasts, and *JCHAIN* plasma cells, with *PLA2G2A* fibroblasts enriched around the gland and cartilage (Suppl. Figure 5G), which were also confirmed by our Visium dataset (Suppl. Figure 6). The definition of regional gene expression variation in non-epithelial cell types, such as fibroblasts, immune and endothelial cells in the healthy lung provides a basis for the precise comparison of the same regions in the diseased states. This may distinguish the regional gene expression variations from the disease-associated ones.

## Spatial analysis of early-stage COPD patients demonstrates AT0 cell state alterations

We further explored the utility of our topographic atlas as a reference to detect deviations in cellular proportions, gene expression and local cell interactions in diseased lung tissue. We focused on a common lung disease, chronic obstructive pulmonary disease (COPD), using samples from 3 patients with COPD GOLD stage II obtained from the most distal lung locations (corresponding to region 3c in Fig. 1A). These samples were derived from the tumor-free regions from lung cancer surgeries. Two healthy atlas samples together with one histologically normal tumor-free lung sample of a COPD-free cancer patient were processed side-by-side for comparison. Samples contained variable airway sizes (large, medium, small bronchioles and respiratory bronchioles). We applied a SCRINSHOT panel to test the expression of the 41 most selective genes in order to define major cell types.

In this topographic lung tissue atlas of COPD we analyzed 84,631 high-quality cells and defined 20 major cell types according to their markers (Suppl. Figure 7A). The major cell classes were equivalently represented in all analyzed samples (Suppl. Figure 7B), however the proportion of AT1 cells was decreased and the proportion of T lymphocytes increased in COPD samples (Fig. 5A). Previous extensive scRNA-seq studies on COPD patients reported a shift in the expression of epithelial secretory cell gene programs, where proximal airway gene expression levels gradually increased in distal epithelial cells of COPD airways [31] leading to a decrease in the proportion of pre-TB (TASC) secretory cells [10]. Another recent publication reported an increase in bronchial secretory cell type marker expression in the AT2 cells from COPD patients [21]. We therefore subclustered both bronchial secretory and AT2 cells and defined a population of cells co-expressing AT2 (*NAPSA*, *SFTPC*) and airway (*SCGB3A1*, *SCGB3A2*, *LTF*) markers, which we annotated as AT0 cells (Suppl. Figure 7A, C). We found that proportion of these AT0 cells was significantly increased in all COPD samples (Fig. 5A, B). These AT0 cells were mostly in the alveolar regions in proximity to the large and small airways, and near accumulations of lymphatic

immune cells. This *in situ* increase in AT0 state is in line with the scRNA-seq analysis arguing for a general upregulation of the proximal secretory cell type program in epithelial cells, not only in the airways, but also in the alveoli [21, 31].

We extended our analysis to find the COPD-specific cellular niches. First, we compared healthy and COPD peri-bronchial and alveolar non-epithelial cells, and found no alterations in their proportions in COPD samples, apart from the increase in T lymphocytes in both COPD compartments (Suppl. Figure 7D). Following this, we performed neighborhood analysis (Methods) [32], and clustered the COPD-cellular neighborhoods together with the ones from the healthy atlas, which contained the coordinates of 218,496 cells grouped into 30 cell types, from three anatomic locations (3–4 donor samples per location). The integrated data separated into twelve neighborhood clusters with two of them corresponding to the SMG, and ten of them matching the distal lung regions (Fig. 5C). Three of these neighborhoods were composed predominantly of cells deriving from COPD samples of all three patients (Fig. 5D, arrows). The first COPD-cell neighborhood (termed, T-E) located in terminal bronchioles and alveoli, contained cells from all disease-samples expressing TRB and AT0 markers, an unannotated secretory epithelial cell type, AT1 cells and endothelial cells. This neighborhood was particularly increased in proportion in one of the patients. The second COPD-cell neighborhood (Imm-P) was composed of T lymphocytes and other immune cells, fibroblasts, endothelial and AT2 epithelial cells, and was consistently increased in all three patients analyzed. (Fig. 5D-E, Suppl. Figure 7D, E). This is in accordance with the known inflammatory nature of the COPD. Finally, the third COPD-specific cellular neighborhood (AT0-Alv) was composed of cells expressing AT0, AT1 and AT2 cell markers, fibroblasts and endothelial cells. This neighborhood mainly contained cells from one COPD patient. Moreover, we identified two neighborhoods (AT2-Alv) and (Cap-Alv), composed of alveolar epithelial cells, endothelial cells, macrophages and fibroblasts, which were decreased in all three COPD patients compared to the healthy lung tissue samples (Fig. 5D-E, Suppl. Figure 7D, E). This is consistent with the onset of alveolar simplification, which is an important component of the COPD pathology. This spatial analysis based on the topography of the healthy lung describes deviations in the cellular locations and neighborhood composition in the diseased lungs. We detected a reduction in the alveolar epithelial neighborhoods (AT2-Alv and Cap-Alv). Instead, AT0 cells in COPD patients were increased and contributed to different COPD-specific neighborhoods, AT0-Alv, T-E and Imm-P (Fig. 5E). The neighborhood analysis reveals a consistent shift in the balance of the distal airway and alveolar cell phenotypes. We conclude that the usage of the spatially resolved healthy lung cell atlas as a reference aids the detection of cellular composition and cellular environments of tissue samples derived from diseased lungs.

## Discussion

We generated a reference topographic atlas of three representative regions in the adult human lung, combining Visium unbiased transcriptomics with expression maps at cellular resolution for 162 genes with HybISS and 96 genes with SCRINSHOT. The targeted methods mapped 478,894 cells defining the location of 35 major cell types and discovered additional region-related cell variability. This variability includes new suprabasal intermediate cell states, such as the S100A8/9 cell state in the trachea, which

included a *KRT13*-positive cell subset in the intermediate layer of the epithelium. In the SMG, we discovered an *SCGB3A2*-expressing serous cell state and revealed a *GHRL*-expressing neuroendocrine cell group selectively in the distal airways. The utilization of complementary methods with different sensitivity, multiplexity and cellular resolution was also suitable to localize previously known but underrepresented cell types, including neuronal cells and lipofibroblasts (fibroblasts that co-express *APOD*). However, a few of the expected cell types and subtypes could not be identified either due to the sparse signal from some probes, or due to the absence of particular cell types in the collected sections. Our analysis was also limited by the absence of certain cell types in the scRNAseq datasets (such as Schwann and adipocytes). We detected these cells in the histological sections, but cannot confidently annotate them by gene expression. Our further analysis of cell neighborhoods relates gene expression levels with cell proximities and tissue histology and thereby facilitates future work to reveal homeostatic cell communication patterns by mining the extensive scRNA-seq datasets from healthy lung tissues [2]. A major aim of the Human Cell Atlas projects is to generate reference maps and databases that also enable the study of diseased cell states [32, 33]. We generated an additional SCRINSHOT dataset including two distal lung regions from the atlas, three from COPD stage II patients and a corresponding healthy region from another patient to explore the utility of our spatial atlas in identifying alterations in cell composition in disease. In this integration we used a general cell type panel, without aiming to target COPD-specific gene expression changes but to define a diseased phenotype purely by cellular neighborhoods and regional cell composition. Extensive previous work with samples from large cohorts indicated that COPD pathology is linked with multiple perturbed cell types [10, 21]. Our relatively small experiment identified gene expression alterations including the increased proportion of AT0 and a relative reduction in AT1 cells in the diseased lungs, in line with previous knowledge of the COPD histopathology [34]. We further constructed the proximity-based COPD cellular neighborhoods and compared them with the atlas revealing aberrant, COPD-characteristic cell niches involving fibroblasts, macrophages, epithelial, endothelial and immune cells. Our identification of the cell types involved in these aberrant disease-specific niches provides a basis for hypothesis building and further exploration of available RNA-seq data from the corresponding cell types, as well as Visium data from COPD patients [10, 21, 35, 36]. The extent and open availability of our atlas for exploration and data mining together with the sensitivity of our targeted methods suggest their further application in early disease and pathology detection in heterogeneous lung tissues.

## Materials And Methods

### 1. Sample collection and screening

#### 1.1. Donor information

For the healthy adult lung atlas, four deceased organ donors of various age, gender and smoking status were used. Informed consent from the families and approval from NRES Committee of East of England, Cambridge South, was obtained (15/EE/0152). Donors were anonymized and numbered randomly. Donor information can be found in Suppl. Table 1. For the diseased sample analysis, tissues were collected

with written informed consent from all patients and with ethics approval from the ethics committee of the Ludwig Maximilian University of Munich (#330-10, #19-629 and #19-630). Patients admitted to the hospital due to lung cancer were either diagnosed with COPD or were COPD-free. Lobectomy of the lobe with the tumor was performed, and tissues were collected from peri-tumor (tumor-free) regions. Patients were anonymized and numbered randomly. Patient information can be found in Suppl. Table 1.

### 1.2. Sample collection, freezing, and histopathological assessment of donor samples

Tissues for the healthy lung atlas were collected from the lungs of deceased organ donors. Lung tissue samples (approximately 0.5-1 cm<sup>3</sup>) were collected from the following locations: (1) ventral side of trachea in proximity to carina, (2) left bronchus 2-3 generation, (3a) bottom part of left upper lobe, 1-2 cm from pleura, (3b) top part of left upper lobe, 1-2 cm from pleura, (3c) bottom part of left lower lobe, 1-2 cm from pleura. The samples were rinsed in PBS and stored until freezing in OCT within 2 hours from collection. The samples from COPD and non-COPD patients were collected from the following locations: (3c) left lower lobe, and (3d) right lower lobe. Then samples were briefly rinsed in PBS and dried, then refrigerated until freezing in OCT within 18 hours from collection. Frozen blocks of tissue were sectioned using a cryostat at 10 µm thickness, fixed with 4% PFA and stained using hematoxylin and eosin (H&E). Imaged sections were evaluated by a histopathologist, samples with signs of severe inflammation were excluded. A total of 55 tissue samples from six donors were sectioned and analyzed histologically. We excluded two donors as unhealthy due to observed pathological landmarks in alveolar regions. The remaining samples (from four donors described in Suppl. Table 1) were further screened using H&E staining for the presence of the airways, submucosal glands, alveoli and blood vessels, as well as the absence of freezing artifacts and other more subtle pathological conditions, such as inflammation or fibrosis. Within the patient cohort, six tissue samples from six patients (three COPD and three non-COPD) were scanned. Two non-COPD samples were excluded due to poor quality and lack of airways. One remaining non-COPD sample together with two healthy donor samples were used for the experiment side-by-side with COPD samples. We processed morphologically suitable samples using SCRINSHOT and screened for RNA integrity with the presence of mRNA signals of well-characterized cell type marker genes.

### 1.3. mRNA quality assessment

For Visium/RRST eight 10 µm sections of each tissue were collected for the total RNA integrity (RIN) values. Samples with RIN values above 5 were processed further. For targeted spatial analysis one section per morphologically assessed tissue sample was processed with SCRINSHOT as described previously [16] for main cell type marker genes. The cell type marker panel was used for the detection of major cell types, as described previously [18]. Samples or regions with low or sparse signal were considered to be of unsuitable quality. Samples with specific patterns of gene expression and strong (>10 dots per cell) SCRINSHOT signal were considered suitable for further analysis. Selected samples are summarized in Supplementary Table 2.

## 2. Gene panel selection

The cell type probe panels were designed using Spapros, as described previously [18]. For the SCRINSHOT panel, we utilized a precursor version of the method, which exhibited distinct characteristics compared to the current versions v0.1.0-v0.1.4. Specifically, during the training for binary cell type classification, individual decision trees were computed instead of generating multiple trees and subsequently choosing the most optimal one. Moreover, secondary trees aimed at enhancing classification performance for finely annotated cell states, which are challenging to discern, were not included. The implemented version combined gene set selections for the lung regions: proximal (airway) and distal (alveoli, parenchyma) lung. This process involved choosing common genes for shared cell types and unique genes for region-specific cell types. Fifty genes for each region were chosen from the scRNA-seq lung data from previous publication [1]. Selections were based on log-normalised data post scran normalisation [37]. We omitted Donor A47 due to missing location annotations. An internal marker list was provided as input for Spapros, ensuring that genes from marker list groups that were not well captured with the initial selection were automatically included. For the SCRINSHOT panel gene selection, 26 clusters (23 cell types and 3 subtypes) were targeted. To filter out genes that might be below the detection threshold, an expression penalty was applied. This penalty employed a smoothed rectangular function to penalise genes with 0.99 expression quantiles below 2 and above 6 (set parameters on cpm log-normalised data), which translates to 0.75 and 4.3 respectively when adjusted to scran normalisation. Conversely, for HybISS selections, Spapros v0.1.0 was used on the same dataset but with more detailed cell type annotations covering 52 cell types (36) and subtypes (16). No region-specific selection or expression restrictions were applied. As before, the internal lung marker list was supplied for selection.

### **3. *In Situ* Sequencing (HybISS)**

#### 3.1 HybISS mRNA detection

Cell type markers were used as previously reported [15], replacing some of the markers by alternative markers when the design of specific padlock probes was not possible. The final panel of genes profiled can be found at Supplementary Table 3. The CARTANA High-Sensitivity library preparation kit was employed, following the manufacturer's instructions, with customized backbones as described in Supplementary Table 3. In the experimental process, tissue sections were fixed and then subjected to an overnight incubation with the probe mix in a hybridization buffer. Subsequently, stringent washing was performed, followed by incubation with the ligation mix. After further washes, RCA (Rolling Circle Amplification) was conducted overnight. For detection, labeling was performed according to the procedure described in the protocols.io website (<https://doi.org/10.17504/protocols.io.xy4fpyw>).

#### 3.2 HybISS imaging

RCPs were detected using 4 different fluorophores (Cy3, Cy5, AF750 and AF488) across five imaging rounds. DAPI staining was imaged on each cycle to identify cell nuclei. All images were acquired using a Leica DMI8 epifluorescence microscope, which was equipped with various accessories. The microscope setup included an external LED light source called Lumencor® SPECTRA X light engine, an automatic

multi-slide stage (LMT200-HS), a high-quality sCMOS camera named Leica DFC9000 GTC, and different objectives such as HC PL APO 10X/0.45, HC PL APO 20X/0.80, and HCX PL APO 40X/1.10 W CORR. For capturing multispectral images, the microscope was equipped with specialized filter cubes capable of separating 6 different dyes. Additionally, an external filter wheel (DFT51011) was used to enhance the imaging capabilities further. The image scanning process involved outlining Regions of Interest (ROIs) that could be saved for multi-cycle imaging, employing tiled imaging with a 10% overlap. To capture the depth of the tissue, Z-stack imaging was performed, covering 10  $\mu\text{m}$  at 0.5  $\mu\text{m}$  intervals.

### 3.3 HybISS preprocessing and decoding

The initial preprocessing of microscope images involved several steps. Z-stacks were subjected to maximum intensity projection, tiles were aligned between imaging cycles, and image stitching was performed. The code for this preprocessing can be accessed at [https://github.com/Moldia/ISS\\_preprocessing](https://github.com/Moldia/ISS_preprocessing). During imaging, the images and their accompanying metadata were exported. The images were then formatted into OME tiff files and stitched and aligned using ASHLAR [38]. To reduce computational requirements, the stitched images were sliced into 6000 by 6000-pixel sections for decoding. Due to the heightened sensitivity of the High Sensitivity Cartana kit, we encountered challenges related to overlapping RCPs (optical crowding) in the 2D projected data, which was essential for decoding. To address this issue, we employed a content-aware image restoration (CARE) approach, which had previously been trained on pairs of raw-deconvolved RCP images of multiple tissues ([https://github.com/Moldia/ISS\\_CARE](https://github.com/Moldia/ISS_CARE)) [39]. As a result, the RCPs in the CARE-processed images exhibited significantly enhanced sharpness, leading to a reduction in overlapping RCPs and improved decoding results. The transcript decoding process relied on the Python package called starfish (<https://spacetxstarfish.readthedocs.io/en/latest/>). The decoding code can be found at [https://github.com/Moldia/ISS\\_decoding](https://github.com/Moldia/ISS_decoding). In brief, the images were registered and underwent white top hat filtering. The channel intensities were then normalized across channels. Spots, representing transcripts, were located in a composite maximum intensity projected image of signal images from the same field-of-view and sequencing round. The decoding of spots was achieved using the PerRoundMaxChannel method. For each spot and each base, the highest intensity channel was determined and matched to a corresponding barcode in the codebook. Additionally, a quality metric was assigned to each spot in every cycle. This metric was defined as the called channel intensity divided by the sum of all other channels, with values ranging between 0.25 and 1.

### 3.4 HybISS data analysis

Cell segmentation was performed on each sample using the BIAS lite software (Single Cell Technologies Ltd), using pre-trained deep neural network model (DiscovAIR Segmentation, v.1.4), which provided precise delineation of individual cells in the imaging data. Approximately  $10^4$ - $10^5$  nuclei were segmented per sample. To ensure comprehensive coverage of each cell area, we expanded the cell masks by two micrometers. The critical task of assigning reads to cells was accomplished using Baysor [19]. Taking the segmentation mask generated with BIAS and the location of every decoded read as an input, Baysor

enabled us to efficiently associate the sequencing reads with their respective cells, facilitating downstream analyses. By implementing these steps, we were able to establish a robust and reliable cell-to-read assignment pipeline for our research, providing a solid foundation for further investigations and insights into the biological processes under study.

### 3.5. Clustering and subclustering method.

With the aim of identifying cell populations present in the tissue, cell-by-gene matrices from different regions and donors were pooled together. Cells with less than 3 genes and less than 5 transcripts per gene were excluded from analysis. The counts of the remaining cells were then normalized and log-transformed prior to clustering. Main clusters identified were further subclustered, occasionally with excluding 1-2 non-marker genes which interfered with the clustering in order to guarantee the division of cells based on their cell identity. All details can be found in Suppl. Table 4.

### 3.6. Cell type annotation.

Manual doublet exclusion was performed for each cluster based on scRNA-seq data and previously published annotations [1]. Gene detection levels ranged from 1 to 20 dots per cell (each dot representing mRNA molecule), consistent with the differences in the expected gene expression levels [1]. Approximately 1-2% of misannotation was observed, allowing few AT1, AT2 and aerocyte-annotated cells to be mapped to tracheal regions, some SMG cells to the parenchymal regions, and stromal cells to epithelium. For overall analysis in the current study this misannotation was insignificant, but it should be considered for refined cell type mapping using our HybISS-based atlas. Annotated cell types can be found in the web data viewer (<https://adult-lung-iss.serve.scilifelab.se/>), which was created using TissUmaps [40].

### 3.7. Measurement of baso-luminal cell type distribution

With the aim of assessing the baso-luminal gradients present in the human lung airways, we first manually segmented individual airways based on DAPI and complementary H&E stainings, keeping only cells detected within the epithelium for further analysis. Next, we manually defined the basal layer using TissUmaps and we computed the minimum distance from each transcript and cell detected with ISS to the basal layer. Finally, since airways present a different width along the proximodistal axis of the lung, we normalized the distance to the basal layer of each cell/transcript per airway, which resulted in each transcript/cell presenting a relative proximity to the basal layer (0-1). In this context, the closest transcripts/cells to the basal layer will have a distance of 0, whereas the most distant ones will present a relative distance of 1.

## 4. SCRINSHOT.

### 4.1. Probe design.

All padlock probes for SCRINSHOT were designed with a unique barcode, as described previously [26]. In order to detect five genes per hybridization cycle using DAPI for nuclei, detection probe oligos were conjugated to one of the five different fluorophores: FITC, Cy3, TexasRed, Cy5 and Cy7. List of SCRINSHOT probes is available in Supplementary Table 8.

#### 4.2. Probe application and detection

Experimental procedure in SCRINSHOT was performed as previously described [16, 18, 26]. Briefly, tissue sections were fixed for 10 minutes in 4% paraformaldehyde, treated with 1M HCl, blocked and incubated with padlock probe mix. Probes of highly expressed genes were reduced to 1-2 padlocks per gene, the rest was applied at 3-4 padlocks per gene. Ligation with SplintR, and probe amplification steps were followed by fixation of cDNA product and detection cycles. Detection probes were applied at 30 °C in 30% formamide solution, followed by washes at 30 °C in 20% formamide solution. All samples were processed with cell type panel (Suppl. Table 9) in three experiments with variation in probe list for *ZG16B*, *GRP*, *ZFP36L2*, *NKX2.1*, *CD69* and *BPIFB1*, and slight changes in concentration of padlocks for highly expressed genes. Rare cell type panel was composed of the epithelial cell type markers and rare cell type markers (Suppl. Table 9). It was applied to samples from four donors, but only two samples per location were selected for further analysis (Suppl. Table 2). Additionally, *SERPINB3* was detected in samples from donors 1 and 4 together with other epithelial markers. Images were taken at 20x magnification using as a Z-stack with 10-11 steps of 0.8 µm (to cover the whole 10 µm thickness) at a widefield microscope (Zeiss Axio Observer Z.2, Carl Zeiss Microscopy GmbH, with a Colibri led light source, equipped with a Zeiss AxioCam 506 Mono digital camera and an automated stage).

#### 4.3 Data analysis.

Analysis of SCRINSHOT data was performed as described previously [16, 18, 26]. Projection and stitching, followed by image export was performed in Zen (2.3 lite). Tiling of SCRINSHOT images was performed in Fiji (ImageJ 1.53c), SCRINSHOT signal (dot) detection was done in CellProfiler (3.1.9). Automated nuclei segmentation was performed in BIAS on tiled images (via Image Filters function) using deep neural network model (DiscovAIR Segmentation, v.1.0) in Segmentation function with scaling 1.50-2.00, detection confidence 1%, contour confidence 50%. Manual correction of segmented nuclei shapes in areas with compact tissue, such as bronchial epithelium, was used via Manual Segmentation function. Approximately  $10^4$ - $10^5$  nuclei were segmented per sample. Nuclei shapes were then expanded by 0.5 µm using Mask Operators function, overlapping shapes were removed at tile borders using Remove Duplicate function at 10% overlap threshold, then nuclei shapes were eroded back by 0.5 µm in order to create a gap between regions of interest using Mask Operators function. Images of nuclei masks were exported using Scan function, and used to define regions of interest (ROIs) in Fiji. Nuclei regions were expanded by 2 µm (without overlaps) to recapitulate cell ROIs using CellProfiler. Assignment of dots to cells was performed in Fiji, as described previously [16, 26]. Cell type panel data was used for clustering for cell type confirmation, epithelial cell quantification, epithelial cell type mapping, SMG mapping, muscle and fibroblast gene expression level comparison, and immune cell

mapping. Rare cell panel was used for manual annotation and quantification of rare cells. COPD panel was used for selected cell type annotation, as in Supplementary Table 7, quantification of cell types and neighborhood enrichment analysis.

#### 4.4. Clustering data and mapping cell types

Gene detection levels ranged from 1 to above 50 dots per cell (each dot representing mRNA molecule), consistent with the differences in the expected gene expression levels [1, 18]. For cell type clustering with 64 genes, negative and low positive cells were excluded by filtering out cells with less than 25 total dot count, leaving around 40% of cells for further analysis. For a shortened cell type panel of 41 genes used for COPD analysis the filtering of low positive cells was performed by excluding cells with less than eight counts per cell. Normalization by total count was the only data conversion applied. For cell type maps data was clustered separately for each sample. For comparison of healthy and COPD samples, datasets were combined. Different resolutions and principle components (pc) were tested on a representative dataset, so that most expected cell types appeared as clusters. Leiden clustering with 20 nearest neighbors, resolution 1.5 and pc number 7 was used for all remaining datasets. Subclustering was performed for selected clusters to retrieve smaller groups of cells with 10 nearest neighbors, resolution 0.4 - 1.5 and pc number 0. Clusters were manually annotated according to expected marker genes. Manual cell type assignment for rare cell types was performed by positivity (<20% of maximum counts per cell for each gene) for any of the rare cell type marker genes. The rare cells were then clustered as described above. Annotated cell types were mapped by ROI coordinates using TissuUmaps [40], and can be found in the web data viewers.

#### 4.5. Airway length measurement

Airway length or circumference was measured by drawing a line along the basal membrane (based on the border of KRT5 and KRT15 signal) in Fiji and measuring its length in micrometers.

### 5. Visium and RRST

Visium and RRST were performed as described previously [20]. Proximal lung sample from donor 4 was used for standard Visium in 8 sections 100  $\mu$ m apart from each other. RRST was used in proximal and distal lung sections from donor 4 and distal lung section from donor 1. Cell type annotation was performed using stereoscope [41]. Graph panels were created using RNA rescue app and the annotated dataset available online: [https://github.com/ludvigla/DiscovAir\\_data\\_explorer/](https://github.com/ludvigla/DiscovAir_data_explorer/)

### 6. Neighborhood enrichment analysis

To explore the cellular environment of each cell and define cellular niches, each cell was re-defined based on the local neighborhood of each profiled cell. For each cell, its 20 closest cells were considered its microenvironment and used to create a cell-by-neighboring cell types matrix. For every cell, we quantified the amount of cells of each cell type present in its defined neighborhood, as done previously [33]. Cell-by-neighborhood matrices were then preprocessed following standard single cell preprocessing

steps including library size-based normalization. To define cellular neighborhoods in COPD vs healthy dataset, graph-based clustering was performed using Leiden clustering. Neighborhoods were further represented via UMAP low dimensional representation. Clusters resulting from this process represent tissue neighborhoods, defined as groups of cells that present the same local microenvironment. Since some cellular neighborhoods were sample, or even cell specific, local clusters with less than 50 cells were excluded from the analysis, as they did not represent general neighborhoods, but rather unique rare microenvironments.

## **7. Statistical analysis**

Data normality for percentage values was reached using logit-transformation, variance was tested in using F test. Statistical tests for each corresponding dataset are indicated in figure legends. When normality could not be reached (for example, due to true 0 values), non-parametric tests were used for data comparison. For too low/undetectable values the 0 were replaced half of the minimum detection values. Statistic comparisons were performed and graphs were created in GraphPad Prism 8.3.0 (538), or in Microsoft Excel (14.0.7268.5000).

## **Declarations**

## **Acknowledgements**

We acknowledge discovAIR grant agreement 874656, and all members of discovAIR consortium, especially Karen van Eunen, Pascal Barbry, Laure-Emmanuelle Zaragosi, Amanda Oliver and Elo Madisson. We gratefully acknowledge the provision of human biomaterial and clinical data from the CPC-M bioArchive and its partners at the Asklepios Biobank Gauting, the Klinikum der Universität München and the Ludwig-Maximilians-Universität München. We also kindly acknowledge the ISS facility in Science for Life Laboratories, Stockholm, Sweden, including Chika Yokota and Amitha Raman, for performing the HybISS experiments. Finally, we acknowledge Jan-Olov Persson at The Statistical research group (SFG), Department of Mathematics, Stockholm University, for consulting in statistical data analysis. FK and PH acknowledge support from the TKP2021-EGA09, Horizon-BIALYMPH, Horizon-SYMMETRY, Horizon-SWEEPICS, Horizon-Fair-CHARM, HAS-NAP3, OTKA-SNN no. 139455/ARRS, and Finnish Cancer Society. This work was supported by the EU Horizon Program (DiscovAIR). MN acknowledges Cancerfonden grant CAN 2021/1726. MCN was supported by the Chan Zuckerberg Initiative, LLC Seed Network grant CZF2019-002438 “Lung Cell Atlas 1.0”. CS laboratory is supported by VR 2019-04893 from Cancerfonden 21 1794 P=1H and Erling-Persson Foundation 2023-0035.

## **Author contributions**

The DiscovAir Consortium (work package 2), led by CS and MCN, conceived the project. ABF managed and coordinated the project. LBK and MDL designed cell type marker panels using Spapros and analyzed scRNA-seq datasets. LBK assisted with additional marker gene selection and data interpretation. KTM

collected and froze the donor tissue samples for the Atlas dataset. YC and JGS froze the patient samples for healthy peri-tumor and COPD cohort. ABF, AS, AL and JT designed SCRINSHOT probes. ABF, XMA, ZA and LAG performed tissue cryosectioning and histological staining. WT performed histological and pathological assessment of sections. XMA, ZA and LAG performed RIN analysis, Visium and RRST. LL analyzed Visium and RRST datasets. ABF performed SCRINSHOT experiments and analyzed SCRINSHOT dataset. AS assisted with SCRINSHOT image alignment and data visualization preparation. SMS analyzed HybISS dataset. TB and FK optimized the automated nuclei segmentation design, algorithm training dataset of manually-drawn nuclei was provided by ABF and AS. ABF and SMS performed data clustering and cell type annotations. SMS performed measurements of proximo-distal and baso-luminal cell type and gene expression distribution and neighborhood enrichment analysis in healthy (HybISS) and COPD (SCRINSHOT) datasets. ABF performed measurements of rare cell types, tissue compartments, and COPD-related cell type changes. CS, MCN, KBM, MN, JL, FT, MDL, HBS, WT and PH supervised the project. ABF, SMS and CS wrote the manuscript, and all authors reviewed it.

## Competing interests statement

MDL contracted for the Chan Zuckerberg Initiative and received speaker fees from Pfizer and Janssen Pharmaceuticals. JL and ZA are the scientific consultants for 10x Genomics Inc which holds intellectual property rights to the spatial transcriptomics technology. FJT is a scientific consultant for Immunai Inc., Singularity Bio B.V., CytoReason Ltd, Cellarity, Curie Bio Operations, LLC, and has an ownership interest in Dermagnostix GmbH and Cellarity. MN is co-founder of CARTANA.

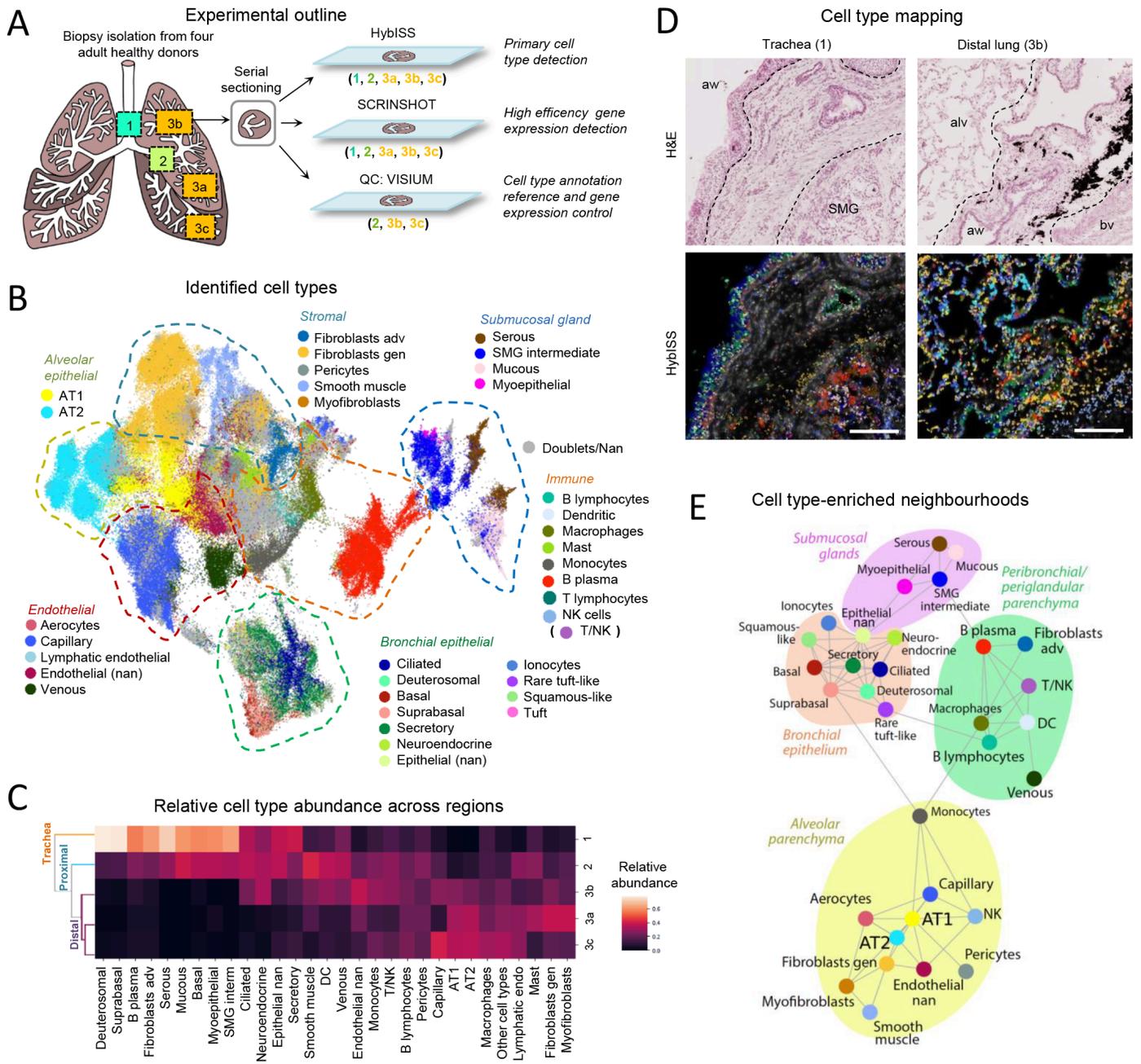
## References

1. Madisson E et al (2023) A spatially resolved atlas of the human lung characterizes a gland-associated immune niche. *Nat Genet* 55(1):66–77
2. Sikkema L et al (2023) An integrated cell atlas of the lung in health and disease. *Nat Med* 29(6):1563–1577
3. Deprez M et al (2020) A Single-Cell Atlas of the Human Healthy Airways. *Am J Respir Crit Care Med* 202(12):1636–1645
4. Travaglini KJ et al (2020) A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* 587(7835):619–625
5. Tabula Sapiens C et al (2022) The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* 376(6594):eabl4896
6. Plasschaert LW et al (2018) A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* 560(7718):377–381
7. Montoro DT et al (2018) A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* 560(7718):319–324
8. Kuo CS et al (2022) Neuroendocrinology of the lung revealed by single-cell RNA sequencing. *Elife*, 11

9. Murthy KL (2022) Human distal lung maps and lineage hierarchies reveal a bipotent progenitor. *Nature* 604(7904):111–119
10. Rustam S et al (2023) A Unique Cellular Organization of Human Distal Airways and Its Disarray in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 207(9):1171–1182
11. Vieira Braga FA et al (2019) A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat Med* 25(7):1153–1163
12. Reyfman PA et al (2019) Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. *Am J Respir Crit Care Med* 199(12):1517–1536
13. Habermann AC et al (2020) *Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis*. *Sci Adv*, 6(28): p. eaba1972
14. Morse C et al (2019) Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur Respir J*, 54(2)
15. Luecken MD et al (2022) The discovAIR project: a roadmap towards the Human Lung Cell Atlas. *Eur Respir J*, 60(2)
16. Sountoulidis A et al (2020) SCRINSHOT enables spatial mapping of cell states in tissue sections with single-cell resolution. *PLoS Biol* 18(11):e3000675
17. Gyllborg D et al (2020) Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue. *Nucleic Acids Res* 48(19):e112
18. Kuemmerle LB et al (2022) *Probe set selection for targeted spatial transcriptomics*. *bioRxiv*, : p. 2022.08.16.504115
19. Petukhov V et al (2022) Cell segmentation in imaging-based spatial transcriptomics. *Nat Biotechnol* 40(3):345–354
20. Mirzazadeh R et al (2023) Spatially resolved transcriptomic profiling of degraded and challenging fresh frozen samples. *Nat Commun* 14(1):509
21. Basil MC et al (2022) Human distal airways contain a multipotent secretory cell that can regenerate alveoli. *Nature* 604(7904):120–126
22. Zuo W-L et al (2018) Ontogeny and Biology of Human Small Airway Epithelial Club Cells. *Am J Respir Crit Care Med* 198(11):1375–1388
23. Lunardi F et al (2011) Overexpression of SERPIN B3 promotes epithelial proliferation and lung fibrosis in mice. *Lab Invest* 91(6):945–954
24. Skronska-Wasek W et al (2021) The antimicrobial peptide S100A8/A9 produced by airway epithelium functions as a potent and direct regulator of macrophage phenotype and function. *Eur Respir J*, : p. 2002732
25. Lin B et al (2024) Airway hillocks are injury-resistant reservoirs of unique plastic stem cells. *Nature* 629(8013):869–877
26. Sountoulidis A et al (2023) A topographic atlas defines developmental origins of cell heterogeneity in the human embryonic lung. *Nat Cell Biol*

27. He P et al (2022) A human fetal lung cell atlas uncovers proximal-distal gradients of differentiation and key regulators of epithelial fates. *Cell* 185(25):4841–4860e25
28. Volante M et al (2002) Ghrelin expression in fetal, infant, and adult human lung. *J Histochem Cytochem* 50(8):1013–1021
29. Meyrick B, Sturgess JM, Reid L (1969) A reconstruction of the duct system and secretory tubules of the human bronchial submucosal gland. *Thorax* 24(6):729–736
30. Tachihara-Yoshikawa M et al (2008) Expression of secretoglobulin3A2 (SCGB3A2) in primary pulmonary carcinomas. *Fukushima J Med Sci* 54(2):61–72
31. Yang J et al (2017) Smoking-Dependent Distal-to-Proximal Repatterning of the Adult Human Small Airway Epithelium. *Am J Respir Crit Care Med* 196(3):340–352
32. Dann E et al (2023) Precise identification of cell states altered in disease using healthy single-cell references. *Nat Genet*
33. Kukanja P et al (2024) Cellular architecture of evolving neuroinflammatory lesions and multiple sclerosis pathology. *Cell* 187(8):1990–2009e19
34. Hogg JC, Timens W (2009) The pathology of chronic obstructive pulmonary disease. *Annu Rev Pathol* 4:435–459
35. Olvera N et al *Lung Tissue Multi-Layer Network Analysis Uncovers the Molecular Heterogeneity of COPD*. *American Journal of Respiratory and Critical Care Medicine*. 0(ja): p. null
36. Rojas-Quintero J et al (2024) Spatial Transcriptomics Resolve an Emphysema-Specific Lymphoid Follicle B Cell Signature in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 209(1):48–58
37. Lun L, Bach ATK, Marioni JC (2016) Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol* 17(1):75
38. Muhlich JL et al (2022) Stitching and registering highly multiplexed whole-slide images of tissues and tumors using ASHLAR. *Bioinformatics* 38(19):4613–4621
39. Langseth CM et al (2023) *Single cell-resolution <em>in situ</em> sequencing elucidates spatial dynamics of multiple sclerosis lesion and disease evolution*. *bioRxiv*, : p. 2023.06.29.547074
40. Pielawski N et al (2023) TissUUmeps 3: Improvements in interactive visualization, exploration, and quality assessment of large-scale spatial omics data. *Heliyon* 9(5):e15306
41. Andersson A et al (2020) Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Commun Biol* 3(1):565

## Figures



**Figure 1**

HybISS-based cell type map reveals cell type distribution and neighborhoods. A. Experimental outline including the location of sample collection from donated healthy human lungs, and the methods used for the mRNA-based cell type mapping. 1 – trachea, 2 – proximal bronchi, 3a – bottom part of upper left lobe, 3b – top part of upper left lobe, 3c – bottom part of lower left lobe. B. UMAP of cells after leiden clustering profiled with HybISS, colored by the assigned cell type (35 cell types presented). Gen – general, adv – adventitial, nan – not annotated. C. Heatmap of relative abundance of clustered cell types between locations demonstrating their frequency across the profiled regions. Similarity in cell type composition between three distal lung regions can be assessed by hierarchical clustering dendrogram

on the left side. D. Representative histological images of an analysed trachea (donor 4) and a distal lung (donor 1) biopsies with hematoxylin and eosin (H&E) staining (up) coupled to the maps of cell types identified by HybISS on top of nuclei (DAPI, white) in the same sections (down). Spots represent detected transcripts, colored according to the corresponding cell type of the cell they were assigned to. Colors as in Figure 1B. Dashed lines indicate the approximate borders of histologic compartments. SMG – submucosal gland, aw – airway, alv – alveolar region, bv – blood vessel. Scale bar 200  $\mu$ m. E. Cell type neighborhood enrichment graph representing cell types as nodes, and edges indicating a positive neighborhood enrichment (>2) between cell types across the profiled sections. Suggested neighborhoods are shown as bubbles. Node colors as in Figure 1B.

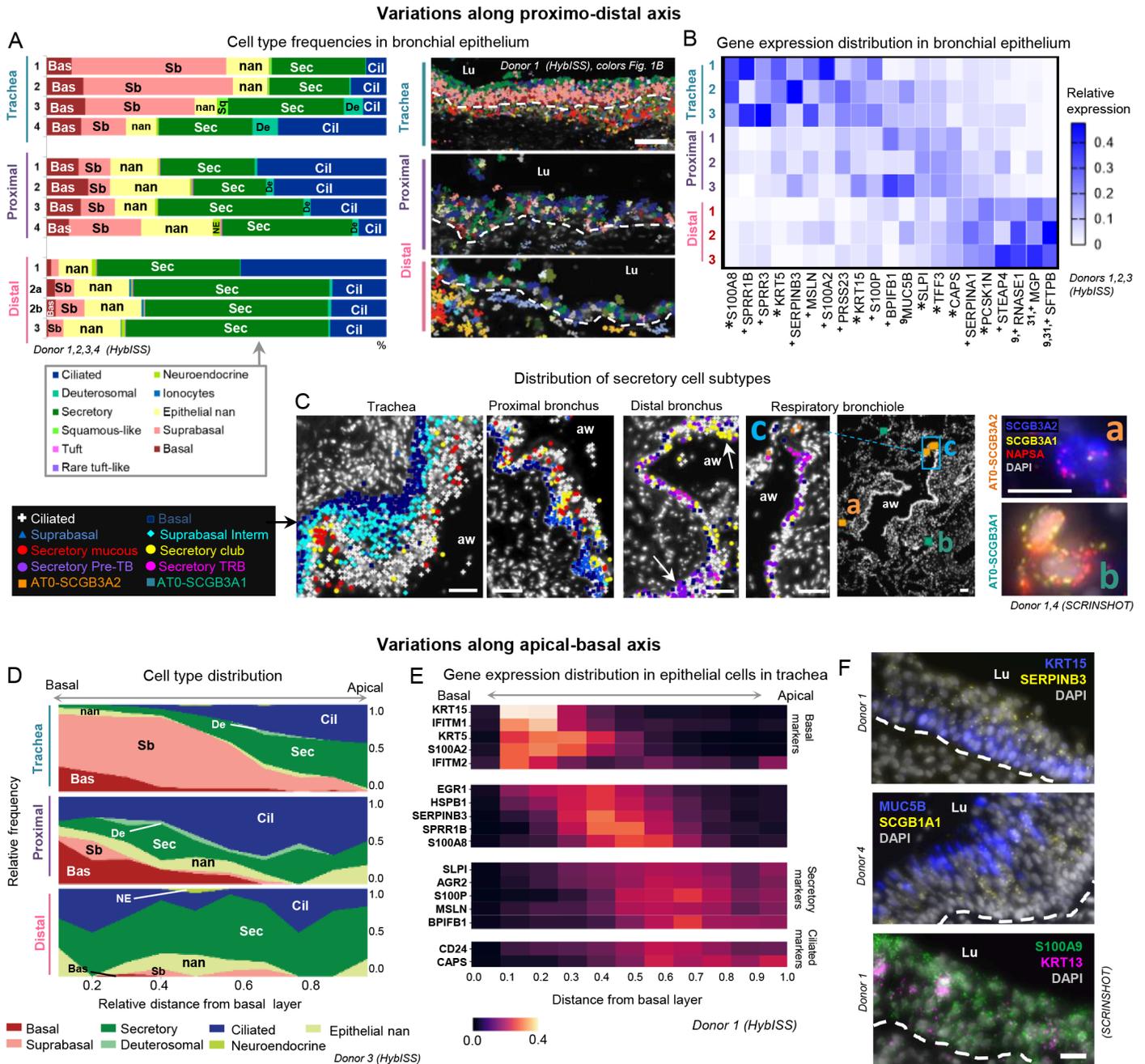
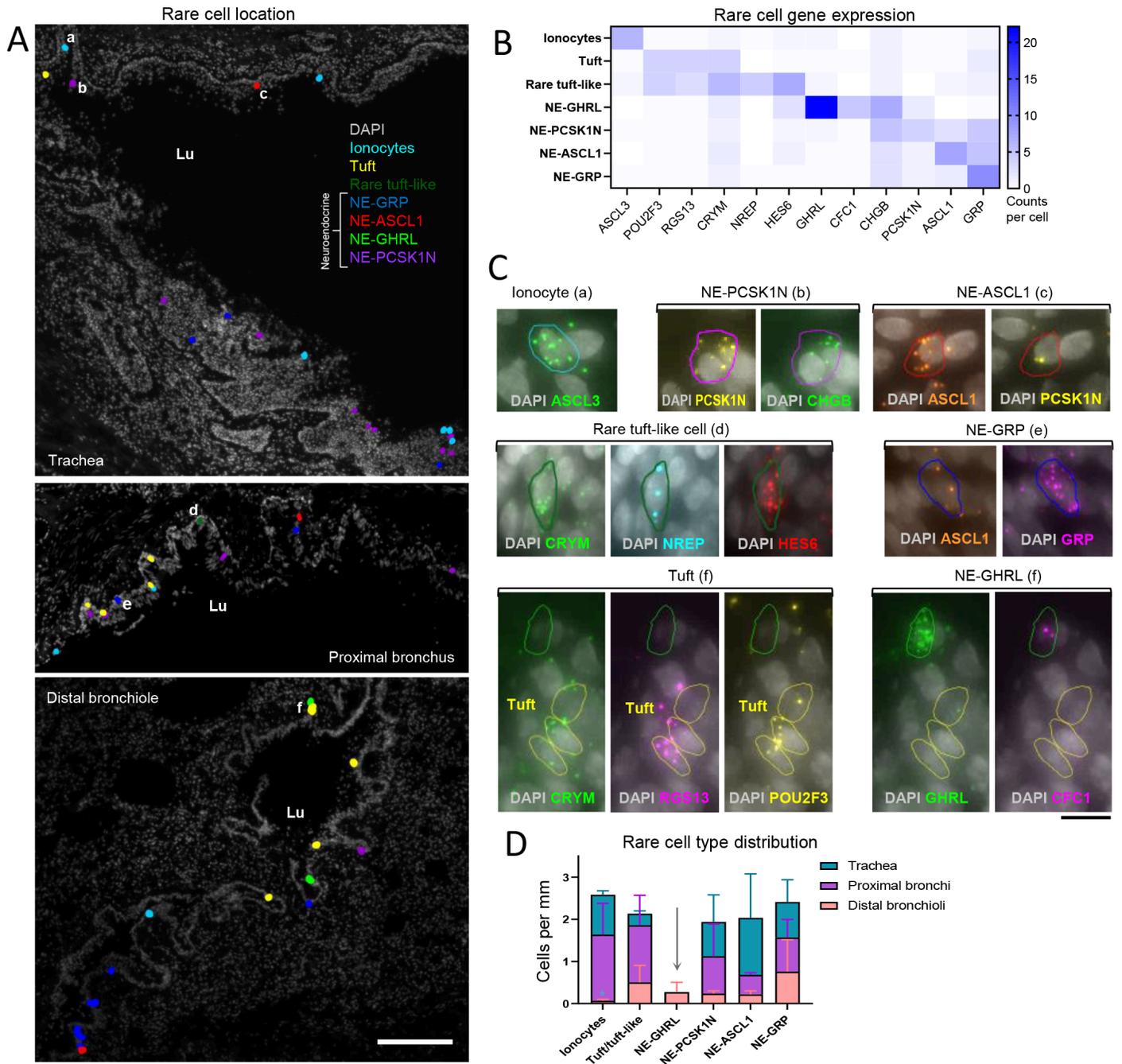


Figure 2

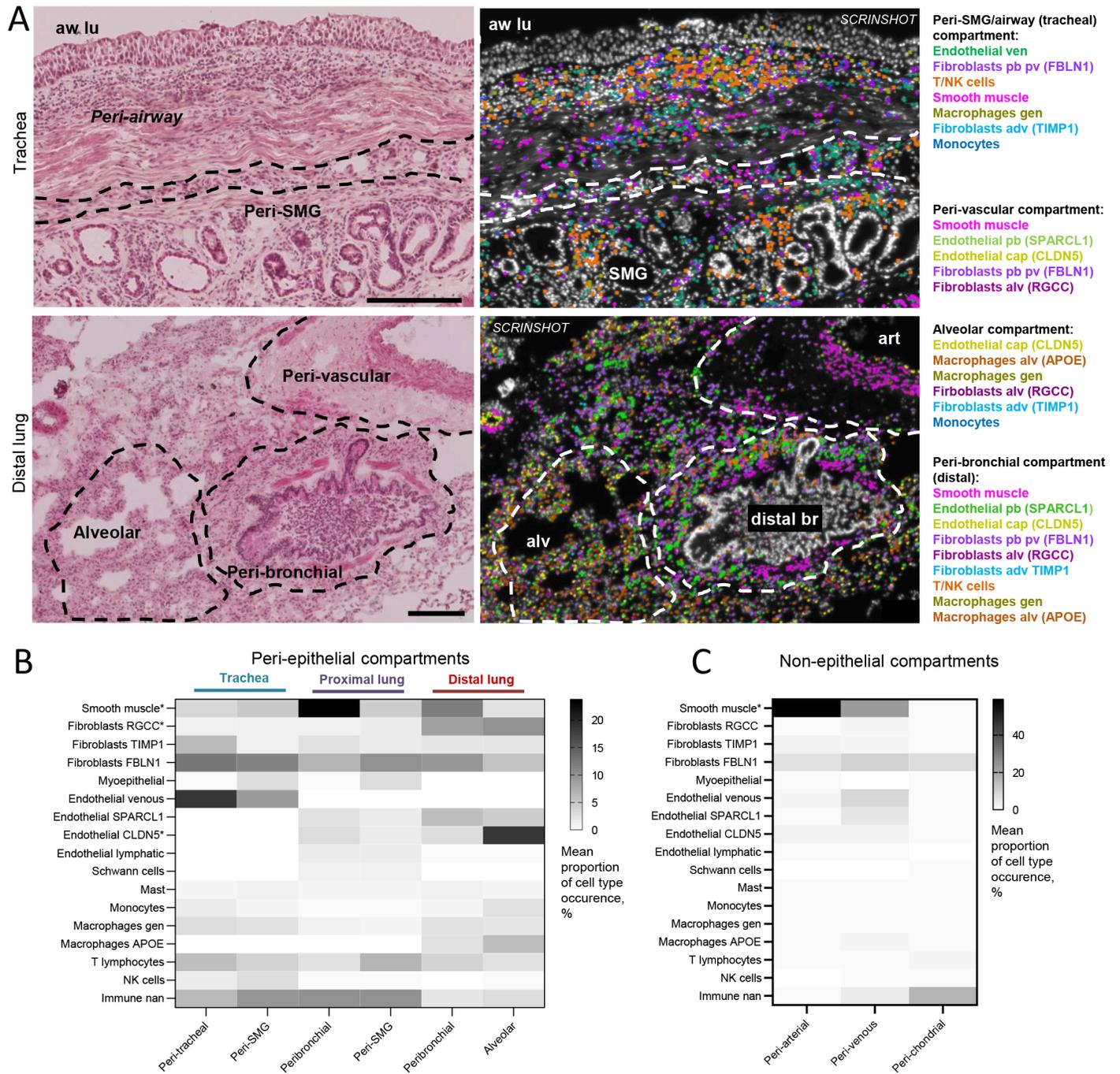
Distribution of cells and gene expression in bronchial epithelium along proximo-distal and apical-basal axes. A. Cell type frequencies in the airway epithelium according to HybISS. Left: stacked bar plot representing relative frequencies of the airway epithelial cell types across regions from different donors, using samples with representative numbers of airway cells. Right: cell type maps in the indicated regions from donor 1. Spots represent detected transcripts, colored according to the corresponding cell type of the cell they were assigned to. Colors as in Figure 1B. Nuclei: gray. Lu – lumen, dashed line: approximate location of basal membrane. Scale bar: 50  $\mu$ m. B. Heatmap of the relative mean gene expression in airway epithelium of variable epithelial markers across regions (colors) in three analysed donors (numbers). The expression is normalized by gene, dividing by sum of values on each row. Superscript numbers: references of previous studies, reporting variable expression of the corresponding marker along the proximal-distal axis. asterisk\*: statistically significant expression differences of the corresponding marker between regions ( $p$ -value<0.05, repeated measures ANOVA with Geisser-Greenhouse correction, followed by Tukey's multiple comparisons test, all 161 detected genes tested). plus+: having highest mean change (labeled with plus+). C. Maps of epithelial cell types detected by SCRINSHOT in the indicated regions of the airways. Arrows: cell clusters. Inserts in respiratory bronchiole map: (a-b) SCRINSHOT images of representative AT0 cells with either SCGB3A2 (a, orange squares) or SCGB3A1 (b, jade squares) dominating expression. (c) Zoomed area. Nuclei: gray. Scale bar in maps: 50  $\mu$ m. Scale bar in SCRINSHOT images 10  $\mu$ m. aw – airway. D. Area plots representing the relative apical-basal cell type distribution across regions according to HybISS (data from one representative donor). X axis: relative distance of cells from the basal membrane. Y axis: relative frequency of cell types. E. Heatmap of the relative mean gene expression of selected markers along the apical-basal axis of the tracheal epithelium (data from one representative donor). The expression is normalized by gene, dividing by sum of values on each row. F. Representative images of detected transcripts of the indicated genes with SCRINSHOT in the tracheal epithelium of different donors. (top) KRT15: basal layer and SERPINB3: suprabasal layer. (middle) SCGB1A1: intermediate layer and MUC5B: luminal layer. (bottom) KRT13 intermediate layer and S100A9: all layers. Lu – lumen, dashed line: approximate location of basal membrane. Scale bar 20  $\mu$ m.



**Figure 3**

Rare cell types and their distribution in the airways. A. Maps of rare cell types detected with SCRINSHOT from three anatomical regions (donor 4). Scale bar 200  $\mu$ m. Lu – lumen. B. Heatmap of gene expression demonstrating unique and overlapping marker genes within the detected rare cell types from at least two donor samples (donors 2 and 4) per anatomical region. (Number of cells quantified: ionocytes – 37, tuft – 23, rare tuft-like – 3, NE-GHRL – 6, NE-PCSK1N – 28, NE-ASCL1 – 29, NE-GRP – 32). C. SCRINSHOT signal of rare cell type marker genes plotted on top of nuclei (DAPI, grey) in the indicated regions in (A). Colored outlines: segmented nuclei expanded to approximate cell borders. Scale bar: 10  $\mu$ m. D. Bar plot

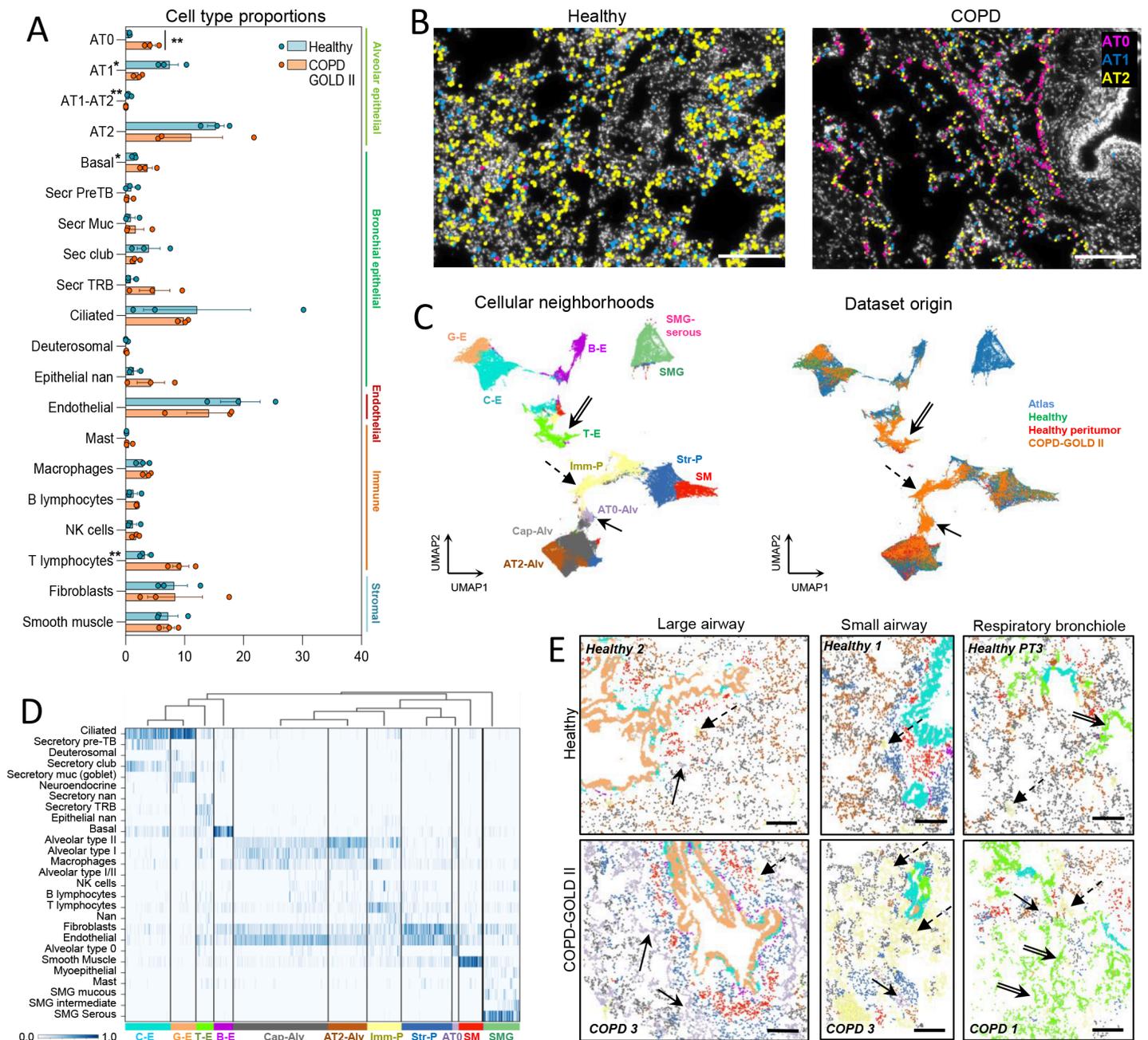
of the average number of the detected cells from (B) per mm length of basal membrane from two donors in each anatomical region, error bars: standard error. Arrow: NE-GHRLpos cells appearing only in distal lung.



**Figure 4**

Distribution of non-epithelial cell subtypes in histological tissue compartments. A. Histological images of sections stained with hematoxylin and eosin (left) after SCRINSHOT analysis and their corresponding maps of cell types/subtypes (right) in trachea (top) and distal lung (bottom). Dotted outlines: peribronchial, SMG and alveolar neighborhoods. Cell type color code at the right side, pb – peribronchial,

adv – adventitial, alv – alveolar, pv – perivascular, cap – capillary, gen – general. Scale bar 200  $\mu$ m. Labels on figure: distal br – distal bronchiole, SMG – submucosal gland, alv – alveoli, art – artery. B. Heatmap of mean cell type/subtype proportions in each of the peri-epithelial compartments. Mean proportion per compartment is indicated in percent (%), n=3 in tracheal regions, n=4 in proximal and distal lung regions. Values were compared in lung regions using Friedman test followed by Dunn’s multiple comparisons test. Significantly changing groups ( $P < 0.05$ ) are indicated by asterisk (\*), data demonstrates the increase in smooth muscle cells in peribronchial region in comparison to alveolar, and in CLDN5 endothelial and RGCC fibroblasts in alveolar region in comparison to peri-SMG and peri-bronchial compartments of proximal lung, respectively. C. Heatmap of average (mean) cell type/subtype proportions in perivascular and peri-chondrial mesenchyme. Peri-arterial (n=9), perivenous (n=7) and peri-chondrial (n=5) compartments from four donors were compared using Kruskal-Wallis test and revealed significant increase in smooth muscle cells in peri-arterial compared to perichondrial compartment.



**Figure 5**

Cell type and neighborhood changes in COPD. A. Box plot of the cell type numbers in healthy and COPD (Median  $\pm$  standard deviation, with individual values). Significant differences are highlighted with asterisk (\*\*,  $p < 0.01$ ), according to multiple t-test (20) of logittransformed data with Holm-Sidak correction,  $n = 3$ . The direct cell type comparison without the correction also revealed significant differences between healthy and COPD samples in the following cell types: AT0, AT1, AT1-AT2, basal cells, T lymphocytes (labeled by asterisks at the cell type name). B. Representative maps of alveolar epithelial cell types detected in healthy and COPD samples, nuclei: gray. Scale bar 200  $\mu\text{m}$ . C. UMAP plots of analyzed cells labeled according to their cellular neighborhoods (left) and their corresponding condition (right).

Neighborhood annotations: Cap-Alv – capillary-enriched alveoli, AT2-Alv – AT2-enriched alveoli, AT0-Alv – AT0-enriched alveoli, C-E – club-enriched epithelium, G-E – goblet-enriched epithelium, T-E – secretory TRB-enriched epithelium, B-E – basal cell-enriched epithelium, Imm-P – immune cell-enriched parenchyma, Str-P – stromal cell-enriched parenchyma, SM – smooth muscle, SMG – submucosal gland, SMG-serous – serous cell-enriched submucosal gland, TRB – terminal respiratory bronchiole. Arrows indicate the clusters that are predominantly composed of COPD-derived cells. D. Heatmap exploring the cell type composition of each neighborhood, with cells (vertical lines) grouped by their assigned neighborhood cluster (y-axis) represented in Fig. 5C (left). Bar color represents the ratio of neighborhood enrichment with each cell type for each cell. E. Maps of cell type neighborhoods in healthy and COPD samples in alveolar region with large and small airway and respiratory bronchioles. Color code as in C (left). Scale bar 200  $\mu\text{m}$ . Simple arrows: AT0-enriched alveolar neighborhood, dashed arrows – immune-enriched parenchyma, double-line arrows – secretory TRB-enriched epithelium.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTables.zip](#)
- [Supplementaryfigures.pdf](#)