Cellular and molecular heterogeneities and signatures, and pathological trajectories of fatal COVID-19 lungs defined by spatial single-cell transcriptome analysis

Short title: Spatial single-cell transcriptome analysis of COVID-19 lungs

Arun Das[1,2]†, Wen Meng[1,3]†, Zhentao Liu[1,4], Md Musaddaqul Hasib[1,2], Hugh Galloway[1,4], Suzane Ramos da Silva[1,3], Luping Chen[1,3], Gabriel L Sica[5], Alberto Paniz-Mondolfi[6], Clare Bryce[6], Zachary Grimes[6], Emilia Mia Sordillo[6], Carlos Cordon-Cardo[6], Karla Paniagua Rivera[7], Mario Flores[7], Yu-Chiao Chiu[2,8], Yufei Huang[1,2,4]*, and Shou-Jiang Gao[1,3]*

[1]Cancer Virology Program, UPMC Hillman Cancer Center, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[2]Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[3]Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[4]Department of Electrical and Computer Engineering, Swanson School of Engineering, University of Pittsburgh, Pittsburgh, PA, USA

[5]Department of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[6]Department of Pathology, Molecular and Cell-Based Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA

[7]Department of Electrical and Computer Engineering, KLESSE School of Engineering and Integrated Design, University of Texas at San Antonio, San Antonio, TX, USA

[8]Cancer Therapeutics Program, UPMC Hillman Cancer Center, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

†These authors contributed equally to this work.

*Corresponding authors. Emails: yuh119@pitt.edu; gaos8@upmc.edu

Abstract

Despite intensive studies during the last 3 years, the pathology and underlying molecular mechanism of coronavirus disease 2019 (COVID-19) remain poorly defined. Here, we examined postmortem COVID-19 lung tissues by spatial single-cell transcriptome analysis (SSCTA). We identified 18 major parenchymal and immune cell types, all of which are infected by SARS-CoV-2. Compared to the non-COVID-19 control, COVID-19 lungs have reduced alveolar cells (ACs), and increased innate and adaptive immune cells. Additionally, 19 differentially expressed genes in both infected and uninfected cells across the tissues mirror the altered cellular compositions. Spatial analysis of local infection rates revealed regions with high infection rates that are correlated with high cell densities (HIHD). The HIHD regions express high levels of SARS-CoV-2 entry-related factors including ACE2, FURIN, TMPRSS2, and NRP1, and co-localized with organizing pneumonia (OP) and lymphocytic and immune infiltration that have increased ACs and fibroblasts but decreased vascular endothelial cells and epithelial cells, echoing the tissue damage and wound healing processes. Sparse non-negative matrix factorization (SNMF) analysis of neighborhood cell type composition (NCTC) features identified 7 signatures that capture structure and immune niches in COVID-19 tissues. Trajectory inference based on immune niche signatures defined two pathological routes. Trajectory A progresses with primarily increased NK cells and granulocytes, likely reflecting the complication of microbial infections. Trajectory B is marked by increased HIHD and OP, possibly accounting for the increased immune infiltration. The OP regions are marked by high numbers of fibroblasts expressing extremely high levels of COL1A1 and COL1A2. Examination of single-cell RNA-seq

data (scRNA-seq) from COVID-19 lung tissues and idiopathic pulmonary fibrosis (IPF) identified similar cell populations primarily consisting of myofibroblasts. Immunofluorescence staining revealed the activation of IL6-STAT3 and TGF-β-SMAD2/3 pathways in these cells, which likely mediate the upregulation of COL1A1 and COL1A2, and excessive fibrosis in the lung tissues. Together, this study provides an SSCTA atlas of cellular and molecular signatures of fatal COVID-19 lungs, revealing the complex spatial cellular heterogeneity, organization, and interactions that characterized the COVID-19 lung pathology.

Key words: SSCTA, spatial single-cell transcriptome analysis; COVID-19, coronavirus disease 2019; SARS-CoV-2; SNMF, sparse non-negative matrix factorization analysis; NCTC, neighborhood cell type composition analysis; trajectory inference; organizing pneumonia; fibrosis; IL6-STAT3; TGF-β-SMAD2/3

Introduction

Multiple single-cell RNA-seq (scRNA-seq) analyses of coronavirus disease 2019 (COVID-19) patients with different severities have improved our understanding of cellular diversity associated with infection and provided important molecular insights into the host immune response [1-5]. Despite intensive studies in the last 3 years, the pathology and underlying molecular mechanism of COVID-19 remain unclear. Severe COVID-19 is often accompanied by diffuse alveolar damage (DAD) that presents complex pathological manifestations and is heterogeneous within infected tissues and across patients [6]. The dissociation of tissue localization of cells in scRNA-seq has become a bottleneck to decoding the pathology features at the molecular and cellular levels and failed to reveal the immune signatures of the microenvironment for severe COVID-19.

Spatial single-cell transcriptome analysis (SSCTA) promises to reveal the molecular basis of cellular heterogeneity, organization, and interactions in tissues and organs [7, 8]. However, analysis of these complex datasets including defining the spatial cellular organizations, immune microenvironment patterns, cell-cell interactions, and molecular signatures associated with disease pathophysiology remains a daunting task.

Here, we utilized SSCTA to examine postmortem lung tissues from 5 cases with severe COVID-19 and one case without COVID-19. From 10,414,863 detected transcripts of cellular 221 genes from six tissues, we identified 1,719,459 cells that were mapped to 18 major parenchymal and immune cell types, all of which are infected by SARS-CoV-2. We further identified the spatial cellular and molecular signatures that define the patterns of SARS-CoV-2 infection, structural and pathological presentations,

and associated immune microenvironments, which project the trajectories of disease progression. Together, this study provides an atlas of cellular and molecular signatures of fatal COVID-19 lungs and reveals the complex spatial cellular heterogeneity, organization, and interactions that characterize COVID-19 lung pathology.

Results

Spatial single-cell transcriptome analysis, cell segmentation, cell typing, and spatial mapping of cells.

Postmortem lung tissues from five COVID-19 autopsies and one postmortem case without COVID-19 were subjected to SSCTA (Fig. 1A). All five COVID-19 cases contracted SARS-CoV-2 in the first wave of the pandemic and had underlying conditions [9], including hypertension (cases 1-3), HIV infection and asthma (case 2), and Parkinson's and chronic kidney diseases (cases 4 and 5). Hematoxylin and eosin (H&E) stains revealed various degrees of diffuse alveolar diseases (DAD), pulmonary thromboembolism, and lymphocytic infiltration in all cases (Fig. S1). Tissues 1 and 2 (1-2C and 2-1A) had prominent organizing pneumonia (OP) or organizing diffuse alveolar damage while edema, hyaline membrane, and fibrin clot or microthrombi were prominent in tissues 3, 4, and 5 (3-1A, 4-3B and 5-3B). We designed probes for detecting the SARS-CoV-2 genome and 221 cellular genes covering markers of common lung parenchymal and immune cells, and immune and inflammatory genes induced by viral infections (Table S1). Following hybridization and gene decoding by *in-situ* sequencing, we stained the tissues with 4′,6-diamidino-2-phenylindole (DAPI) (Fig. S2) to facilitate cell segmentation (Fig. S2).

A workflow was developed to analyze the SSCTA data (Fig. 1B). We detected 869,453 to 3,424,675 reads in the six lung tissues for a total of 10,414,863 reads (Table S2 and Fig. S3A). Using the Baysor algorithm [10], we segmented ~89%-95% of reads into cells based on DAPI staining and the spatial distribution of reads (Fig. S4). We identified 186,659 to 470,294 cells for each sample for a total of 1,719,459 cells in these tissues (Table S2 and Fig. S3B). Over 99% of the cells harbored at least 5-15 reads (Fig. S5). We further filtered the cells deemed low quality (see Methods) and retained a total of 779,137 cells for the subsequent analyses.

We summarized the reads for each gene and normalized the gene reads of segmented cells using *scTransform* [11]. We assigned cell types to 70-84% of the segmented cells based on the expressions of cell type-specific markers (Table S1) and identified a total of 18 cell types including 11 types of parenchymal cells and 7 types of immune cells in all six tissues (Fig. 1C and S6). We used Uniform Manifold Approximation and Projection (UMAP) to visualize the relationship of gene expressions of individual cells associated with different cell types in all tissues (Fig. 1D and S7). There were negligible batch differences as cells from different samples mixed well in individual cell clusters (Fig. 1D and S7). We noticed separated clusters for major parenchymal cells including alveolar cells (ACs) and fibroblasts. By contrast, immune cells were grouped together and mixed with vascular endothelial cells (VECs) in several clusters (Fig. 1D). The poor separation of immune cell types was largely due to highly expressed immunoglobulin kappa light chain (IGKC) and cathepsin L (CTSL) in these cell types (Fig. S8).

We then mapped the individual cells to their spatial locations in the tissues (Fig. S9) and performed the neighborhood cell type composition (NCTC) analysis by computing a vector of 18 cell-type percentages in the neighborhood of individual cells. NCTC reveals spatial variations of the local cell type composition and informs local interplays among cell types. In agreement with the observed pathology that showed DAD in all COVID-19 tissues (Fig. S1), NCTC uncovered disorganized distributions of various cell types, especially the major parenchymal cells such as ACs, VECs, and fibroblasts (Fig. 1E, S1, and S10-S14). In contrast, organized structures and orderly distributions of these cell types were observed in the non-COVID-19 tissue (PBC-PR) (Fig. 1E, S1, and S15). Examination of regions with blood vessels revealed the lining of VECs along the vessels together with other vessel-associated cells including ACs, fibroblasts, and smooth muscle cells (SMCs) (Fig. 1E and S16). Furthermore, we observed abundant infiltrating immune cells including macrophage and monocytes (MMs), natural killer (NK) cells, and T- and B-cells along regions of blood vessels (Fig. 1E and S10-S14). We also observed the expected cell compositions of other structures including bronchiole, capillary, and endothelium (Fig. S17). These results validated the marker-based cell typing approach.

SARS-CoV-2 infection alters the cell compositions of major parenchymal cells and induces immune infiltrations.

In agreement with the results of our previous study [9], we found SARS-CoV-2 reads in diverse cell types with infection rates ranging from 0.6% to 5% except that no infected cell was detected in the small number of erythrocytes (55) identified in tissue 1

(Fig. 1D, 1F, S7 and Table S3). Most infected cells were identified as ACs, fibroblasts, VECs, and MMs as they were the major cell types in the COVID-19 tissues (Fig. 1F).

Consistent with the tissue pathology, we observed significant alterations in the compositions of different cell types in COVID-19 tissues compared to the non-COVID-19 tissue (Fig. 1E and 1F). Among the parenchymal cells, there were reduced numbers of ACs from 44.08% to 12.81-28.23% and SMCs from 5.22% to 1.72-3.05%, and increased numbers of epithelial cells (ECs) from 2.08% to 2.40-8.42%, ionocytes from 0.22% to 0.41-1.36% and basal cells from 0.25% to 0.31-1.58%, suggesting their likely involvements in COVID-19 lung pathology (Fig. 1E, 1F, and Table S3). Furthermore, we observed increases in cell numbers of most types of the identified immune cells in COVID-19 tissues (Fig. 1E, 1F, and Table S3). In particular, there were increased inflammatory cells including MMs from 9.4% to 17.51-21.37% and NK cells from 1.00% to 2.32-9.19%, and adaptive immune cells including T cells from 2.40% to 4.34-6.81% and B cells from 1.91% to 2.70-10.51%, which were consistent with the reported inflammatory and cellular immune response following SARS-CoV-2 infection [12]. However, these changes were more subtle for some cell types in tissue 1, which had the least infected cells and mildest pathological manifestations with the best integrity of parenchymal cells among all COVID-19 tissues (Fig. 1E, 1F, S1, and S10).

SARS-CoV-2 infection induces global differential gene expressions that marked pathological damages and inflammation with spatial cellular features.

We examined the differential gene expression between COVID-19 and non-COVID-19 lung tissues, and identified 9 upregulated and 10 downregulated genes,

182 respectively (Fig. 2A, first panel). Comparison of SARS-CoV-2-infected cells in COVID-
183 19 tissues with all cells in the non-COVID-19 tissue confirmed the differential
184 expressions in 16 of these 19 genes (Fig. 2A, second panel). Interestingly, differential
185 gene expressions were also observed in 18 of these 19 genes between uninfected cells
186 in COVID-19 tissues and all cells in the non-COVID-19 tissue (Fig. 2A, third panel). In
187 contrast, there were only 3 differentially expressed genes between SARS-CoV-2-
188 infected and -uninfected cells in the COVID-19 tissues (Fig. 2A, fourth panel). These
189 results indicated that there were strong indirect effects such as those mediated by
190 cytokines and complement activation induced by SARS-CoV-2 infection that likely
191 contributed to the dysregulated gene expressions and pathology in COVID-19 lung
192 tissues. Thus, we focused on comparing COVID-19 tissues with the non-COVID-19
193 tissue in subsequent analyses of dysregulated genes.

194 The expressions of the 19 dysregulated genes in different cell types and their
195 tissue distributions had significant variations suggesting their complex involvements in
196 different aspects of SARS-CoV-2 infection and COVID-19 lung pathology (Fig. 2B, 2C,
197 S8, and S18-S23). We determined the contributions of individual cell types to the
198 expressions of these genes across all the tissues (Fig. 2D). Of the 19 genes, 10 of them
199 were cell markers (Table S1). All 6 parenchymal cell markers including SFTPC,
200 SFTPA1, and advanced glycosylation end product (AGE) receptor (AGER) for ACs,
201 INMT for fibroblasts, smooth muscle actin alpha 2 (ACTA2) for SMCs, and claudin 5
202 (CLDN5) for VECs were downregulated, and mainly expressed in their respective cell
203 types (Fig. 2A, first panel, 2C, 2D, S8 and S18-S23), of which ACs and SMCs had
204 decreased cell numbers in COVID-19 tissues (Fig. 1E, 1F and S7). Both SFTPC and

SFTPA1 are implicated in lung homeostasis and functions, and their mutations and dysregulation are associated with pulmonary fibrosis [13-15]. Additionally, SFTPA1, a C-type lectin, which binds to specific carbohydrate moieties on lipids and the surface of microorganisms, is essential in the defense against respiratory pathogens [16-18] while the pulmonary-associated surfactant protein C encoded by SFTPC is a marker for COVID-19 patients with high viral loads [19]. AGER, predominantly expressed in ACs (Fig. 2C), is a multiligand receptor interacting with AGE and other molecules implicated in lung homeostasis, development and inflammation, and certain diseases such as diabetes and Alzheimer's disease, and regulates diverse pathways including MyD88-dependent, nuclear receptors, TNF-α, ERK1/2 and p38 MAPK, and p53/TP53 pathways [20-23]. Following interaction with S100A12, AGER triggers the activation of mononuclear phagocytes, lymphocytes, and endothelium by generating key pro-inflammatory mediators [24]. Indeed, AGER-related pathways are activated by SARS-CoV-2 infection and are implicated in the COVID-19 lung pathology [25, 26]. ACTA2 encodes a smooth-muscle actin involved in lung functions including cell motility, tissue structure and integrity, and intercellular signaling [27]. Dysregulation of ACTA2 is linked to a variety of vascular diseases as well as multisystemic smooth muscle dysfunction syndrome [28, 29]. CLDN5 is an integral membrane protein and component of tight junction strands regulating the integrity of epithelial or endothelial cell sheets, and immune cell transmigration [30-32]. SARS-CoV-2 infection suppressed CLDN5 expression contributing to the disruption of respiratory vascular barriers [33]. Thus, downregulations of parenchymal cell markers including SFTPC, SFTPA1 and AGER in

ACs, ACTA2 in SMCs, and CLDN5 in VECs might be involved in SARS-CoV-2 infection and COVID-19 lung pathology (Fig. 2D).

Several other downregulated genes including caveolin 1 (CAV1) and alpha-2-macroglobulin (A2M) are implicated in numerous lung functions (Fig. 2A). CAV1, a component of caveolae involved in multiple pathways such as integrins-mediated and Ras-ERK signaling, is essential for lung functions and during the host response to infections [34, 35]. A2M is an inhibitor of broad spectrum of proteases including trypsin, thrombin, and collagenase [36, 37]. It disrupts inflammatory cascades by inhibiting inflammatory cytokines, is implicated in Alzheimer's disease, and regulates extracellular matrix organization and platelet cytosolic Ca2+ [36, 37]. CAV1 and A2M were downregulated and mainly expressed by ACs, fibroblasts, VECs, and MMs (Fig. 2A, first panel, 2D, S8, and S18-S23), of which ACs and fibroblasts had decreased while MMs had increased cell numbers in COVID-19 tissues (Fig. 1F).

Three upregulated immune marker genes including complement component C1q, A chain (C1QA) and B chain (C1QB), and neural cell adhesion molecule 1 (NCAM1) are involved in innate immune response and inflammation (Fig. 2A, first panel) [38, 39]. C1QA and C1QB were predominantly expressed in MMs (Fig. 2D, S8 and S18-S23). SARS-CoV-2 infection induced robust complement activation, contributing to tissue damage and COVID-19 lung pathology [40-42]. Both cathepsin L (CTSL) and granzyme K (GZMK) could be involved in complement activation [43, 44]. CTSL was induced after SARS-CoV-2 infection and enhanced SARS-CoV-2 infection [45] while GZMK, a serine protease from the cytoplasmic granules of cytotoxic lymphocytes (CTL) and NK cells that recognize and lyse specific target cells [46], could limit the spread of SARS-Co-V-2.

Both CTSL and GZMK were upregulated and mainly expressed by ACs, fibroblasts, MMs, and VECs while CTSL exhibited a comparable expression level but with extremely high levels in localized regions in COVID-19 tissues (Fig. 2A, 2B, 2C, 2D, S8 and S18-S23). NCAM1, the marker of NK cells, is a member of the immunoglobulin superfamily involved in cell-to-cell interactions as well as cell-matrix interactions [47]. The upregulations of C1QA, C1QB, CTSL, GZMK, and NCAM1 were consistent with the increased numbers of MMs and NK cells in COVID-19 tissues (Fig. 1F, S8, and S18-S23). Thus, increased infiltrations of MMs and NK cells were likely involved in COVID-19 lung complement activation and inflammation.

Two immune-modulating cytokines were upregulated including colony-stimulating factor 3 (CSF3), a member of the IL-6 superfamily of cytokines that controls the production, differentiation, and function of granulocytes involved in the innate immune response [48, 49], and inducible T cell costimulator ligand (ICOSLG or B7-H2) involved in positive regulation of interleukin-4 production and CD28 signaling in T-helper cells [50, 51]. The major cells that expressed CSF3 and its receptor CSF3R, as well as ICOSLG and its receptor ICOS, were ACs, fibroblasts, VECs, and MMs (Fig. 2D).

Numerous other upregulated genes in COVID-19 tissues are involved in inflammatory and immune responses (Fig. 2A). Leukocyte-specific transcript 1 (LST1), a myeloid leukocyte-specific transmembrane adaptor protein recruiting protein tyrosine phosphatases SHP-1 and SHP-2 to the plasma membrane, inhibits the proliferation of lymphocytes, and its expression is enhanced by lipopolysaccharide (LPS), interferon-gamma (IFN-γ) and bacteria [52, 53]. Immunoglobulin kappa light chain (IGKC) is essential for antibody production but is also expressed in non-lymphoid cells [54-56].

Both LST1 and IGKC were mainly expressed in ACs, fibroblasts, VECs and MMs (Fig. 2B-E, S8 and S18-S23), which might mediate dysregulation of immune cells in COVID-19 lung tissues. Finally, fibulin 1 (FBLN1), a secreted glycoprotein incorporated into a fibrillar extracellular matrix in a calcium-dependent manner and mediates platelet adhesion via binding fibrinogen [57, 58], was mainly expressed in ACs, fibroblasts, VECs, and MMs (Fig. 2D).

We examined differential gene expressions in cell types that had significant changes in cell numbers (Fig. 1F). Compared to non-COVID-19 tissue, ACs from COVID-19 tissues had 20 differentially expressed genes, of which 14 were identified at the whole tissue level (Fig. 2A, first panel and 2E). The most upregulated genes were LST1, CTSL, IGKC, ICOSLG, C1QB, GZMK, and NCAM1 while the most downregulated genes were SFTPC, A2M, SFTPA1, CLDN5, and CAV1 (Fig. 2E), which likely contribute to the decreased cell numbers in COVID-19 tissues (Fig. 1F). Several differentially expressed genes were also identified in fibroblasts, VECs and ECs (Fig. 2E). Among the immune cells, MMs had the most differentially expressed genes including upregulation of C1QB, NCAM1, and IGKC, and downregulation of A2M, CD68, TYROBP and PRG4 (Fig. 2E).

Spatial analysis identifies regions with high SARS-CoV-2 infection rates that match high cell densities, high levels of viral entry-related factors, and localized pathology.

Even though we did not observe notable global differences in gene expression between infected and uninfected cells within each COVID-19 tissue, the spatial nature

of the SSCTA permitted the identification of the local impact of SARS-CoV-2 infection on lung pathology. We performed the local infection rate analysis to examine the spatial distributions of local SARS-CoV-2 infection (Fig. 3A) and identified regions with high and low infection rates in each tissue (Fig 3B; local Moran's I, p-value < 0.05). SARS-CoV-2 infects host cells by binding to its entry receptor, angiotensin-converting enzyme 2 (ACE2), and subsequent engagement of host proteases and other entry-related factors including FURIN, transmembrane protease, serine 2 (TMPRSS2), and neuropilin 1 (NRP1) [59]. We assessed the spatial correlation between high-infection regions and cells expressing transcripts of these viral entry-related proteins. We found a significant correlation between high infection rates and high expressions of ACE2, FURIN, TMPRSS2, and NRP1 (Fig. 3C; bivariate local Moran's I, p-value < 0.05), supporting the essential roles of these cellular proteins in SARS-CoV-2 infection in lung tissues.

Previous studies reported increased densities of both parenchymal and immune cells in COVID-19 lung tissues [6]. Here, we observed a positive spatial correlation between cell densities and SARS-CoV-2 infection rates especially in tissues 1-4 (Fig. 3D, Table S4), suggesting that increased cell densities might be associated with high infection rates. We segmented regions with high infection rates and high cell densities (HIHD) (Fig. 3E; bivariate local Moran's I, p-value < 0.05). Compared with H&E staining (Fig. S1), we found a close association between HIHD and OP, which was most prominent in tissues 1 and 2, as well as lymphocytic and immune infiltration, which was most severe in tissues 3 and 4 (Fig. 3E). Both are the most common lung pathological manifestations in COVID-19 lungs [12]. Hence, high infection rates might cause persistent damages to the lung tissues, resulting in the infiltration of immune cells, and

increased wound repair and fibrosis [60, 61]. We also observed HIHD areas adjacent to blood vessels in tissues 1 and 3 (Fig. 3E and S1), which might be due to the homing of the circulating virus and immune cells in the bloodstream to these sites or that these cells were more susceptible to SARS-CoV-2 infection [9, 62].

Regions with high SARS-CoV-2 infection rates have unique cellular and gene expression features.

We next examined cell type compositions of the high and low SARS-CoV-2 infection regions. Compared to low-infection regions across the tissues, the compositions in high-infection regions exhibited significant increases in fibroblasts, echoing the observed OP in HIHD areas, and an increase in ACs, possibly due to the proliferation of ACs response to injury repair of alveoli and alveoli capillaries (Fig. 3E and 4A). However, there were decreases in ECs and VECs, which might reflect the damages caused by the infection (Fig. 3E and 4A). These results suggested that high-infection regions might suffer more damages. However, except for T cells and granulocytes, the compositions of major immune cell types including MMs between high and low infection regions remained similar (Fig. 4A), suggesting induction of a broad immune infiltration by SARS-CoV-2 infection, as well as the observed broad impact on the tissue-wide gene expression (Fig. 2A).

The patchwork of high infection regions in tissue 1 prompted us to inspect patterns of local cell-type compositions. Interestingly, NCTC analysis showed three clusters in high-infection regions, of which two were distinctly enriched with prominent local compositions of fibroblasts and ACs, respectively (Fig. 4B). Examining their spatial

342 distributions confirmed a negative *in-situ* correlation between local compositions of
343 fibroblasts and ACs (-0.6, Pearson correlation), with regions showing high (low)
344 compositions of fibroblasts accompanied with low (high) compositions of ACs (Fig. 4B
345 and C). Strikingly, the local regions with high fibroblasts but low ACs were mostly
346 annotated OP regions (Fig. S1).

347 We next investigated the gene expression patterns in the high- and low-infection
348 regions. Compared with the normal tissue, the high and low infection regions presented
349 highly consistent differential expression patterns as they shared 18 of the 20
350 differentially expressed genes (Fig. 4D), confirming the strong indirect effect of SARS-
351 CoV-2 infection (Fig. 2A). One of the two uniquely differentially expressed genes in high
352 infection regions was again cytokine CSF3, which was upregulated in all infected
353 tissues. We, therefore, investigated spatial co-expression of CSF3 and its receptor
354 CSF3R (Fig. 3F) and found that their co-expression patterns also presented spatial
355 correlations in high infection regions (Fig. 3G; bivariate local Moran's I, p-value < 0.05),
356 suggesting that the CSF3-CSF3R axis might modulate the immune response in high-
357 infection regions. In contrast to the consistent differentially expressed patterns of both
358 high- and low-infection regions when compared with the normal control, direct
359 comparison between high- and low-infection regions revealed distinct differentially
360 expressed patterns. Particularly, we observed, in high- *vs.* low-infection regions,
361 upregulation of AC markers SFTPA1 or SFTPA2 in tissues 1, 3, and 5 (Fig. 4E and 2C)
362 and fibroblast markers COL1A1 or COL1A2 in all COVID-19 tissues (Fig. 4E and 4F),
363 despite that SFTPA1 and COL1A2 displayed down-regulation in either high- or low-
364 infection regions in COVID-19 tissues when compared to the normal tissue (Fig. 4D).

The increased expressions of these AC and fibroblast markers were also consistent with the observed increases in ACs and fibroblasts in high-infection regions in the corresponding tissues (Fig. 4A). Cell type-specific differential expressions further confirmed the upregulation of SFTPA1 or SFTPA2 in ACs, and COL1A2 in fibroblasts in infected tissues (Fig. 4E and S24). However, different cell types exhibited complex differential expression patterns between high *vs.* low infection regions in individual COVID-19 tissues (Fig. S24), likely due to the distinct COVID-19 pathologies or stages associated with the individual infected tissues.

Sparse non-negative matrix factorization analysis identifies seven cell composition signatures that recapitulate different healthy and disease statuses.

Although different DAD phases can inform COVID-19 severity, they are confounded by many factors including the underlying conditions of the patients as illustrated by our five COVID-19 cases (Fig. S1). Importantly, the complex tissue microenvironments of parenchymal and immune cells associated with different stages of DAD within and across tissues have not been well characterized. Therefore, we determined the extent to which tissue-independent and tissue-specific spatial patterns in spatially resolved cell-type compositions are associated with different DAD stages. For this purpose, we performed the sparse non-negative matrix factorization (SNMF) [63] of the NCTC vectors of all cells in all tissues. SNMF decomposed these NCTC vectors into a sparse linear combination of cell-type composition signatures that define spatial cell-type patterns in these tissues. We obtained a factorization of seven NCTC signatures after assessing the compactness and biological meanings of the factorization results for

388 different maximum numbers of factors (Fig. 5). Close examination of cell type
389 compositions of the signatures (Fig. 5A and Table 5) and their spatial prevalence across
390 all tissues (Fig. 5B) defined by the factor loading revealed their associations with normal
391 structures, and broad and tissue-specific infection. Specifically, Signature 1 likely
392 defined the cell type composition of “normal-like lung alveoli” because it has 56.4%
393 ACs, 18.5% VECs, 11.9% fibroblasts, and 5.3% SMs, which were comparable to the
394 percentages of the non-COVID-19 tissue (Fig. 1F and Table S3). Thus, the high-loading
395 regions of Signature 1 in COVID-19 tissues might mark the less damaged regions.
396 Naturally, the non-COVID-19 tissue had a high prevalence of Signature 1 (Fig. 5B and
397 S25). Tissue 1 had the highest prevalence of Signature 1 among all COVID-19 tissues
398 reflecting its least damaged pathology, followed by tissue 4, 3, 2, and 5 (Fig. 5B, S1,
399 and S25). Signature 2 described a composition of 61% ECs, 14% ciliated cells, and
400 10% basal cells. It had sparse but high-intensity loadings in localized regions identified
401 as bronchial tubes (Fig. 5B and S1). In contrast to Signatures 1 and 2 which
402 characterized cell-type compositions of normal structures, Signature 3 captured the
403 “broad immune infiltration” as it contained 28.6% B cells, 22.3% MMs, 17.5% T cells,
404 9.6% granulocytes, and 8.7% ECs (Fig. 5A and Table S5). As expected, it was the least
405 common in the non-COVID-19 tissue (Fig. 5B and S25). Its loading distributions in
406 COVID-19 tissues were consistent with the lowest immune infiltration in tissue 1,
407 followed by 4, 3, 2, and 5 (Fig. 5B and S25), closely corroborating the order of damage
408 revealed by Signature 1. Trajectory analysis of the cell type compositions with
409 Signatures 1 and 3 using *Slingshot* [64] defined a trajectory of reduced normal alveoli
410 and increased immune infiltration from the non-COVID-19 tissue to tissue 1, 4, 3, 5, and

2 (Fig. 5C), which was in agreement with the order of their increased severity of pathology observed in H&E staining (Fig. S1). Remarkably, Signatures 4 to 7 captured four distinct types of immune microenvironments or niches in these COVID-19 tissues. Signature 4 was characterized by high ACs (64.8%) with MMs (27.3%), likely reflecting the infiltration of alveolar MMs and proliferation of ACs in wound healing, all of which are features of exudative DAD [65]. The relatively higher prevalence of Signature 4 in the less severe tissues 1 and 4 suggested its association with early phases of DAD (Fig. 5B and S25). There were also overlaps in regions with higher loadings of Signature 4 with the annotated regions of immune infiltration in the H&E image in tissue 4 (Fig. 5B and S1). Signatures 5 and 6 depicted inflammations around VECs, where Signature 5 was associated with extensive infiltration of MMs (46.7%) and T cells (8.7%) around VEC (26.3%), whereas Signature 6 was associated with substantial infiltration of NK cells (31.8%) and granulocytes (13.6%) around VECs (27.6%) and ECs (12.2%). These results were consistent with the influx of MMs and NK cells (Fig. 1E and F), which was likely due to the activation of complement and coagulation systems (Fig. 2A and C), resulting in damages and therefore reduced numbers of ACs and VECs (Fig. 1E and F). Signature 5 was more common in the less severe tissues 1, 3 and 4 but Signature 6 appeared predominantly in 3 (Fig. 5B and S25), particularly in low infection region (Fig. 3A). Indeed, lymphocytic infiltrations were also identified in the regions associated with Signatures 5 and 6 (Fig. S1). Signature 7 denotes extremely high fibroblasts (73.4%) with infiltration of MMs (15.8%). Advanced disease in COVID-19 patients often has an excessive accumulation of fibroblasts and may develop pathological fibrosis due to chronic inflammation from the infiltrated immune cells and dysregulation of the

434 extracellular matrix triggered (ECM) remodeling [66]. As expected, there was a high
435 correlation between the COL1A1 expression level (Fig. 4F) with Signature 7. The
436 loadings of this signature were high in tissues 1 and 2, which were annotated as OP
437 regions, and in tissue 5, which were annotated as fibrin and hyaline membrane in
438 addition to OP regions (Fig. 5B, S1 and S25), suggesting a connection with OP, fibrin
439 and hyaline membrane.
440
441 **Unique cell composition signatures define two separate pathological trajectories.**
442 We then applied Signatures 4 to 7 to delineate the trajectories of DAD
443 characterized by the immune responses against viral infection. We added Signature 1
444 “normal-like alveoli” so that the non-COVID-19 tissue could be included as a reference.
445 We used these 5 signatures to redefine the landscape of spatially resolved cell type
446 compositions in these tissues. The UMAP visualization (Fig. 6A) recapitulated the tissue
447 severity inferred by Signatures 1 and 3 (Fig. 5C) but presented more complex patterns,
448 with each immune signature defining a largely distinct group of cell populations (Fig.
449 6B). We applied *Slingshot* to infer pseudo-progressions of cells defined by these
450 signatures and identified two trajectories (Fig. 6A). Clustering of cells based on pseudo-
451 time values further segmented the trajectories into 6 stages (Fig. 6A). At the onset, the
452 two trajectories shared a common path including stages T1 and most T2, which
453 traversed the cells enriched by Signature 1 “normal-like alveoli” and Signature 4 “ACs
454 with MMs infiltration” in T1 and then those enriched by Signature 4 and Signature 5
455 “VECs with MMs infiltration” in T2 (Fig. 6A). These stages were consistent with the less
456 damaged presentations of tissues 1 and 4 (Fig. 6D and S1) and showed a reduction of

457 ACs due to likely lung injury and an increase of MMs possibly promoted by innate
458 immune responses (Fig. 6C). Inspecting associated tissue regions in the H&E images
459 showed characteristics of early DAD manifested by shedding of ACs, capillary leakage,
460 lymphocytic/immune infiltration, and yet preserved alveoli structure (Fig. 6D and E, T1
461 and T2). After the shared section, the trajectories diverged into two paths. The first path
462 (Trajectory A) included T3a and T4a and progressed through cells mostly from tissues 3
463 and 4 enriched in Signature 5 “VEC with MMs infiltration” ending in Signature 6 “VEC
464 with NK cell infiltration”, which was almost exclusively from tissue 3 (Fig. 6A, B and D).
465 We observed a persistent decrease in ACs but an increase in VECs, NK cells and B
466 cells (Fig. 6C). The H&E staining of tissue regions for T3a and T4a showed hyaline
467 membrane, fibrin, lymphocytic infiltration, and collapsed alveoli and capillary structure
468 (Fig. 6E, T3a and T4a, and S1). These features are consistent with a more advanced
469 DAD. The close association of this path with tissues 3 and 4 suggested that they
470 encompassed similar DAD-related tissue patterns and immune niches. Indeed, the
471 postmortem report showed similar findings for these two patients including bilateral
472 pulmonary consolidation [9].

473 The second path (Trajectory B) included two stages, i.e., T3b and T4b, which are
474 enriched by Signature 6 in T3b and Signature 7 “high fibroblasts with MMs infiltration” in
475 T4b highlighted by increased fibroblasts and immune infiltration of especially MMs (Fig.
476 6A, B, and C). Their associated tissue regions were mostly in tissues 1, 2, and 5 (Fig.
477 6D), suggesting potentially similar DAD tissue patterns and underlying immune niches
478 in these tissues. The postmortem findings indeed revealed massive pulmonary emboli
479 that were the main cause of death for cases 1 and 2 [9]. Close examination of these

regions in the H&E staining showed edema, increased fibrin, and OP (Fig. 6D, E, T3b and T4b, and S1), which are pathological signatures of DAD. There was a striking association between regions in tissues 1, 2 and 5 in T4b with the identified OP, fibrin and hyaline membrane regions and a significant positive correlation between the progression of Trajectory B and cell density (0.504, Spearman correlation, *p-value* < 0.05) suggesting a more advanced stage of T4b than T3b with tissue 1 manifested in a more localized while tissues 2 and 5 in a more systematic manner. These results demonstrated a high degree of inter and intra-tissue heterogeneity in DAD-associated patterns. While tissue 1 contained regions affiliated with both trajectories, tissues 2 and 5 were mostly associated with the trajectory featured by high fibroblasts and their reorganization, and tissues 3 and 4 were with the trajectory featured by increased VEC and immune infiltration.

Next, we investigated the genes whose expressions in a cell type were correlated with the trajectories. We found 53 and 67 genes significantly correlated with Trajectory A and B in at least one cell type, respectively (Spearman correlation, *p-value*<0.05, permutation test; Table S6). In both trajectories, we obtained a negative correlation for ACs marker SFTPC in ACs (Fig. S26A, S26B, Table S6, and Table S7), suggesting a decrease in its expression, which echoes the reductions in ACs in both trajectories (Fig. 6C). We also found a positive correlation of NCAM1 in NK cells with Trajectory A and positive correlations of COL1A1 in Fibroblasts and C1QB in MMs with Trajectory B (Fig. S26A, S26B, Table S6, and Table S7), all of which were consistent with the observed changes of cell type compositions and DAD pathologies along the trajectories (Fig. 6C).

To gain further insight into potential ligand-receptor interactions that underlie the immune patterns associated with the trajectories, we correlated the ligand-receptor co-expressions with the trajectories. We found that CD40LG-CD40, CD80-CD28 and CXCL10-CXCR3 had the highest positive correlations with Trajectory A in all key cell types including especially VECs, ACs and Fibroblasts (Fig. S26C) while CD40LG-CD40, CSF3-CSF3R and CXCL10-CXCR3 were the top positive correlated pairs with Trajectory B in ACs, Fibroblasts, VECs and MMs (Fig. S26D). Thus, these ligand-receptor pairs might contribute to the chemoattraction of the immune cells as a result of SARS-CoV-2 infection.

IL6-STAT3 and TGF-β-SMAD2/3 pathways mediate COVID-19 lung fibrosis and organizing pneumonia.

Because Trajectory B progressed into tissue regions that manifested extensive OP, we investigated the spatial patterns and molecular signatures associated with OP, which was mostly observed in tissues 1 and 2 and to a less extent in tissue 5, which also contained fibrin and hyaline membrane (Fig. 7A and S1). Analysis of cell compositions in regions with OP revealed an overall increase in cell density, and obvious decreases in cell numbers of ACs and VECs but an increase in cell numbers of fibroblasts compared to other regions in the same tissues and the non-COVID-19 tissue (Fig. 7B, C and D). As in the whole tissues, we observed downregulations of AGER, CAV1, SFTPA1, SFTPC, and A2M, and upregulation of IGKC in OP regions compared to the normal tissue (Fig. 7E). The changes of AGER, SFTPA1, SFTPC, and A2M were primarily observed in fibroblasts, MMs and VECs with downregulation of A2M also being

observed in ACs (Fig. 7F, G and H). The most striking observation is the upregulation of genes encoding for collagen type I alpha 1 and 2 chains (COL1A1 and COL1A2) in OP in both tissues compared to the normal tissue or to regions without OP in the same COVID-19 lung tissues (Figure 7E and G). Furthermore, both COL1A1 and COL1A2 were upregulated in fibroblasts in OP regions compared to fibroblasts from either normal tissue or non-OP regions from the same tissue (Fig. 7G, H and I). COL1A1 and COL1A2 form the triple helix of a fibril-forming collagen, which is essential for lung homeostasis. Dysregulation of COL1A1 and COL1A2 leads to lung inflammation and fibrosis [67, 68]. Both COL1A1 and COL1A2 were mainly expressed in fibroblasts (Fig. 7G and 7H), whose cell numbers were doubled in OP regions in both tissues compared to non-OP regions (Fig. 7D). Interestingly, the expression levels of COL1A1 and COL1A2 were highly heterogeneous in the individual cells in OP regions with extremely higher levels in 10-20% of the fibroblasts, which were absent in fibroblasts in non-OP regions or any other types of cells in the same COVID-19 tissues (Fig. 7J). COL1A1 and COL1A2 were not expressed in high levels in any types of cells in normal lung tissue (Fig. 7J). These results indicated that COVID-19 OP regions had severe dysregulation of fibroblasts with upregulation of COL1A1 and COL1A2, and abnormal ACs and VECs as a result of downregulation of AGER, SFTPA1, and SFTPC.

To further confirm that increased expressions of COL1A1 and COL1A2 were hallmarks of COVID-19 OP, we examined their expressions in different cell types in single-cell RNA-seq (scRNA-seq) data from COVID-19 lung tissues [4]. Consistent with our observation, we found that COL1A1 and COL1A2 expressed mainly in fibroblasts and also showed varying expression levels (Fig. S27A). Specifically, they showed a

548 higher expression level in a subset of fibroblasts enriched with myofibroblasts ($p <$
549 0.001, Gene Set Enrichment Analysis (GSEA): Fig. S27A and B). Myofibroblasts are
550 differentiated fibroblasts, whose dysregulation may lead to idiopathic pulmonary fibrosis
551 (IPF) including COVID-19 lung fibrosis, a lung disease that exhibits OP [69-71]. To
552 provide a linkage between the OP regions and IPF, we examined the scRNA-seq data
553 from lung tissue patients with IPF and nonfibrotic controls [72]. Like in the COVID-19
554 data, we observed a subpopulation of fibroblasts with high expressions of COL1A1 and
555 COL1A2 (Fig. S27C) and they were also enriched in myofibroblasts (p-values<0.001,
556 GSEA; Fig. S27D). We further examined the differentially expressed genes in IPF
557 fibroblast cells with high COL1A1 and COL1A2 expressions *vs.* fibroblast cells with low
558 COL1A1 and COL1A2 expressions or nonfibrotic cells, respectively (Fig. S27E and
559 S27F). The two lists of significant differentially expressed genes (DEGs) were similar (p-
560 values<0.001, GSEA; Fig. S27G), suggesting that the subpopulation with low COL1A1
561 and COL1A2 expressions could serve as normal control. As a result, we compared the
562 fibroblasts with a high *vs.* low COL1A1 and COL1A2 expressions in the COVID-19
563 scRNA-seq data and obtained the DEGs (Fig. S27H). These DEGs are enriched by two
564 DEG lists in IPF fibroblasts with high COL1A1 and COL1A2 expressions ($p< 0.001$,
565 GSEA; Fig. S27I and S27J), suggesting that these cells likely had similar molecular
566 characteristics as those in IPF. Indeed, functional enrichment analysis using the
567 Ingenuity Pathway Analysis reported “Pulmonary Fibrosis Idiopathic Signaling
568 Pathways” as the 2nd top-ranked pathway (Fig. S27K). Several other top enriched
569 pathways related to extracellular matrix remodeling were also identified. Examination of
570 the networks of these pathways identified IL6-STAT3 and TGF-β-SMAD2/3 pathways

that could directly regulate the expressions of COL1A1 and COL1A2 (Fig. 7K). Multi-color indirect immunofluorescence antibody assay (IFA) staining indeed revealed that cells with a high COL1A1 expression level also had high levels of IL6 receptor-α (IL6Rα), phospho-STAT3 (p-STAT3), TGF-βR2 and phospho-SMAD2/3 (p-SMAD2/3), confirming the activation of these pathways and their potential roles in the upregulation of COL1A1 (Fig. 7L). Both IL6 and TGF-β are highly induced, and implicated in fibrosis and lung injury [73, 74] as well as upregulation of COL1A1 and other extracellular matrix proteins in COVID-19 lung tissues [75, 76]. These results demonstrated the important roles of IL6 and TGF-β in the induction of severe COVID-19 lung OP and fibrosis.

Discussion

Using *in-situ* sequencing and our spatial single-cell analysis pipeline, we have painted an atlas of cellular and molecular signatures of lung tissues from severe COVID-19 cases. The atlas describes the cellular heterogeneity, organization, and interactions associated with inflammation, damage, and immune responses in COVID-19 tissues. It provides molecular and cellular insights into the mechanisms underlying SARS-CoV-2 infection and the pathological manifestations in COVID-19 lungs.

We detected a total of 10,414,863 transcripts in five COVID-19 and one non-COVID-19 lung tissues. Spatial single-cell transcriptomics analysis at this scale for COVID-19 and other pathological conditions is still very limited. We have overcome considerable challenges and developed a pipeline of robust, scalable, interpretable computational tools and visualization methods for targeted and exploratory analyses emphasizing on spatial discovery. The first critical component of this pipeline is a

transcript-based cell segmentation strategy that integrates cell nuclei in the matching DAPI staining identified by *CellPose* with spatially localized reads using *Baysor* and iteratively segments reads into cells of potentially different sizes [10, 77]. Using this strategy, we segmented 93% of total transcripts into 1,719,459 cells including 186,659 to 470,294 cells for different tissues. A customized filter was further applied to retain only a subset of high-quality cells.

Cell typing of these filtered cells based on marker genes showed that SARS-CoV-2 infected all the 18 identified cell types, a result consistent with those of our previous study [9] and others [4]. However, only a small fraction of these cells (<5%) are infected. These results agree with those of genome-wide studies that detected small numbers of viral reads in blood, lung, and nasopharyngeal samples from severe COVID-19 patients [4, 78-81]. Despite the detected low infection rates, SARS-CoV-2 inflicts similar effects on the infected and uninfected cells across the tissues as shown by differential expression analysis of genes in infected or uninfected cells in COVID-19 tissues against cells in the non-COVID-19 lung tissue (Fig. 2A). Thus, indirect effects such as those induced by the immune response, inflammation, cytokines, and complement activation might play a significant role on lung pathology in COVID-19 patients. In fact, we have observed vast pulmonary microthrombi, thrombosis, and immune infiltrations in these COVID-19 lung tissues (Fig. S1). The major parenchymal cells including ACs, fibroblasts, and VECs have the highest numbers of infected cells. The COVID-19 lung tissues have reduced numbers of ACs and SMCs, but an increased number of ECs accompanied by vast immune infiltrations of innate immune cells

including MMs, granulocytes, and NK cells, and adaptive immune cells including T and B cells.

Despite these global and consistent impacts of SARS-CoV-2 infection on the lungs, there are extensive spatial heterogeneities in addition to varying cell type percentages across the tissues. To reveal spatial dynamics of cellular features presented in considerable spatial discontinuity due to sparse measurements, we adopted a strategy to compute local feature distributions in the neighboring region of the cell with a fixed radius or cell number. This approach assesses the spatial gene expression maps, local infection rates, cell densities, and co-expression maps of ligand-receptor pairs. To further characterize spatial patterns of an individual feature or joint pattern relationship between two features, we extensively applied the Moran's I statistics, a method widely adopted in geoscience but scarcely in spatial transcriptomics analysis. Using these novel strategies, we identified regions with high infection rates, which are found in regions with high cell densities expressing higher levels of SARS-CoV-2 entry-related factors including ACE2, FURIN, TMPRSS2, and NRP1. Thus, SARS-CoV-2 may preferentially infect cells expressing these factors and the infected cells could also aggregate to these high-density regions following infection. Importantly, these regions are mapped to OP (tissues 1 and 2) or regions with fibrin, hyaline membrane, and edema (tissues 3, 4, and 5), and manifest cellular and molecular patterns resembling tissue damage and wound healing including infiltration of immune cells and pattern of fibrosis. We have found that OP harbors high densities of fibroblasts expressing high levels of COL1A1 and COL1A2, a phenomenon common in fibrosis and wound healing, which are unique to OP regions unseen in other regions of COVID-19

639 lung tissues or in the non-COVID-19 tissue. We have identified similar cell populations
640 enriched by myofibroblasts in 10X scRNA-seq data from COVID-19 lung samples and
641 IPF samples, suggesting that SARS-CoV-2 might preferentially infect myofibroblasts or
642 reprogram other cell type(s) into myofibroblasts. Pathway enrichment analysis revealed
643 the enrichment of multiple pathways related to IPF signaling and extracellular matrix
644 remodeling including IL6-STAT3 and TGF-b-SMAD2/3 pathways. Indeed, we have
645 detected high levels of IL6Ra, p-STAT3, TGF-bRII, and p-SMAD2/3 in cells expressing
646 myofibroblast markers in OP regions. These results suggest that SARS-CoV-2 infection
647 might induce aggregation of fibroblasts, and IL6 and TGF-b to promote wound healing
648 in OP regions.

649 To define spatial patterns associated with COVID-19 lung pathology, we
650 performed the NCTC analysis, or the niche analysis, commonly used for probing
651 interactions of cells such as in immune microenvironments. However, the COVID-19
652 tissues contain complex pathology manifestations resulting in highly heterogenous
653 cellular organizations and thus disparate NCTC patterns. To tackle this challenge, we
654 applied SNMF, owing to its scalability and interpretability, to NCTCs of all cells in six
655 tissues to reveal potentially latent features underlying these seemingly less organized
656 NCTCs. We indeed identified 7 latent signatures of spatial cell type compositions that
657 define normal lung structures and SARS-CoV-2 infection-induced immune niches. While
658 many existing studies have revealed the global immune landscape of SARS-CoV-2
659 infection [1, 2, 5], few report spatial signatures of immune niches. We showed that these
660 niches are mostly tissue-independent, but COVID-19 tissues presented heterogeneous
661 distributions of these niches: while tissues 1, 2, 4, and 5 contain multiple types of

niches, tissue 3 has predominately a single type. We further drew the connections between these niches and different stages of DAD. Such a linkage prompted us to apply the trajectory analysis to NCTCs of these signatures. While the trajectory analysis is most commonly applied to gene expressions to predict a pseudo-progression of the cells, here, this innovative use of trajectory analysis defined the relative severities of damages among COVID-19 tissues and constructed two different pathological routes of progression in COVID-19 patients. Both routes start with enriched “ACs with MM infiltration” and “VECs with MMs infiltration” niches but Route A is marked by increased numbers of NK cells and granulocytes, likely reflecting complications of microbial infections, while Route B is characterized by increased HIHD and OP, marked by increased fibrosis. Both routes are also correlated with multiple cytokine-cytokine receptor pairs such as CD40LG-CD40 and CXCL10-CXCR3 that are likely to mediate the chemoattraction of the immune cells (Fig. S26C and S26D). By mapping these routes to individual cells, we further revealed considerable inter and intra-tissue heterogeneity in the inferred progression of COVID-19 pathology with tissue 3 and 4 associated with Route A, tissue 2 and 5 with Route B, and tissue 1 with both routes.

In summary, we have presented an atlas of spatial patterns of different cellular features that characterize SARS-CoV-2 infection, its induced immune infiltrations, inflammation, and damages in severe COVID-19 lungs, in addition to changes in gene expression in multiple cell types, including immune cells and lung parenchymal cells. These results provide insights into the spatial mechanisms at both molecular and cellular levels, which characterize the development of ARDS in COVID-19 patients. Overall, this study demonstrates the power of spatial single-cell transcriptomics and

685 enables spatial computational analyses in the study of COVID-19 lung pathology. The
686 developed innovative methods can benefit spatial single-cell analyses for other healthy
687 and diseased conditions.
688
689 **Methods**
690 **Lung tissue samples**
691 COVID-19 lung tissues were collected from 5 adults with fatal SARS-CoV-2
692 infection by the Autopsy Service of the Department of Pathology, Molecular and Cell-
693 Based Medicine at the Icahn School of Medicine at Mount Sinai, and non-COVID-19
694 lung tissue was obtained from Pitt Biospecimen Core [9].
695
696 **Study approval**
697 Specimens obtained at autopsy do not meet the definition of a living individual
698 per Federal Regulations 45 CFR 46.102, and as such, research using specimens
699 obtained at autopsy does not meet the requirements for Institutional Review Board
700 (IRB) review or oversight under the Icahn School of Medicine Program for the Protection
701 of Human Subjects. The University of Pittsburgh IRB determined that the study is not
702 research involving human subjects as defined by DHHS and FDA regulations and
703 waived of ethical oversight (STUDY20050085).
704
705 **In-situ sequencing (ISS)**
706 HS Library Preparation kit (P/N 1110-02, CARTANA AB, part of 10x Genomics)
707 was used to prepare the library according to the manufacturer's instruction with minor

modification [82]. Four µm FFPE tissue sections were baked for 1 hour at 60°C and deparaffinized by incubating in xylene twice for 7 minutes (min) each. The sections were then rehydrated by incubating in 100% ethanol (EtOH) for 5 min, followed by 70% EtOH for 5 min, and nuclease-free deionized distilled water (SH30538.02, HyClone) for 2 min. Sections were then permeabilized by incubating in citrate buffer pH 6.0 (C9999, Sigma Aldrich) for 3 hours at 95°C. Chimeric padlock probes (from 1 custom panel (Table S1) and 4 predesigned panels: hImmune I1A, hImmune I3D, hLung L1C, hLung L2E, CARTANA AB) directly targeting RNA and containing an anchor sequence as well as a gene-specific barcode were hybridized overnight at 37°C, then ligated overnight at 30°C. Quality control of the library preparation was performed by applying anchor probes which labeled by Cy5 to simultaneously detect all rolling circle amplification products from all genes in the panel. All incubations were performed in SecureSeal™ hybridization chambers (621502, Grace Biolabs). Slow Fade Antifade Mountant (S36936, ThermoFisher) was used for mounting. All samples passed the quality control and were sent to CARTANA AB (part of 10x Genomics) for *in-situ* barcode sequencing, imaging, and data processing. Briefly, the fluorescent signals for quality control were stripped. Adapter probe pool 1 and a sequencing pool containing 4 different fluorescent labels were hybridized to the *in-situ* libraries to detect gene-specific barcodes. The scanning was fulfilled by using epifluorescence microscopy, and raw data consisting of 20x magnification images from 5 fluorescent channels (DAPI, Alexa Fluor® 488, Cy3, Cy5 and Alexa Fluor® 750) and individual z-stacks, were flattened to 2D using maximum intensity projection with a Nikon Ti2 Microscope (software NIS elements) utilizing a Zyla 4.2 camera. In total, 6 sequencing cycles were achieved for full decoding

of all designed genes. After image processing, which includes image stitching, background filtering, and a sub-pixel object registration algorithm, true signals were scored based on signal intensities from individual multicolor images. The results were summarized in a CSV file and gene plots were generated using MATLAB. Each ISS spot provided a unique fluorescent barcode identifying the targeted RNA of a gene marker. The experiment generated a map of gene expressions of selected genes recorded on the natural morphology of the tissue at a single-cell level. The metadata generated from the images were further analyzed to obtain data containing the reads and their 2D (X, Y) positions of each gene, and the DAPI images. TissuuMaps [83, 84] was used to visualize spatial read locations of multiple genes against DAPI or histology images.

Hematoxylin-eosin staining (H&E) and whole slide scanning

H&E was carried out with the same slides subjected to *in-situ* sequencing using Hematoxylin & Eosin Stain Kit (H-3502, Vector Laboratories) according to the manufacturer's instructions. The slides were then scanned with VS200 Slide Scanner (Olympus).

Indirect immunofluorescence antibody assay (IFA)

IFA was carried out as previously described [9]. Primary antibodies included CoraLite® Plus 647-conjugated smooth muscle actin (1:200, Proteintech, CL647-67735), CoraLite® Plus 488-conjugated Collagen Type I (1:100, Proteintech, CL488-67288), IL6Rα (1:500, Proteintech, 23457-1-AP), p-STAT3 (Tyr705) (1 to 20, Invitrogen,

754 710093), p-SMAD2/3 (p-SMAD2-S465/467 and p-SMAD3-S423/425, 1:500, ABclonal,
755 AP0548), and TGF-βRII (1:10, R&D Systems, AF-241-NA). Secondary antibodies
756 included Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa
757 Fluor™ Plus 555 (1:400, Invitrogen, A32732), Donkey anti-Mouse IgG (H+L) Highly
758 Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 568 (1:400, Invitrogen, A10037).
759

760 **Cell segmentation**

761 Reads in a tissue sample were segmented into corresponding cells using the
762 Baysor cell segmentation algorithm [10]. Baysor applies a Markov Random Field model
763 to identify spatial clustering of reads from the same cell. It can perform cell
764 segmentation based on read coordinates alone but can also incorporate nuclear
765 information from DAPI staining. We first utilized the DAPI-stained images of the tissues
766 and performed the nuclei segmentation using the *CellPose* anatomical segmentation
767 algorithm with the default setting of parameters [77]. Segmented nuclei masks with
768 locations were provided to Baysor to define its scale and standard deviation
769 parameters. Baysor was able to segment 71%- 91% of reads after the first run for all
770 samples/tissues. However, we found that among the unsegmented, the so-called noisy
771 background reads determined by Baysor, there is a considerable portion that could still
772 be assigned to cells visually. To address this issue, we applied Baysor again to these
773 noisy reads from the first Baysor run. We optimized the scale and standard deviation
774 parameter for the second Baysor run to maximize reads assignment to cells. Including
775 the second run significantly improved the percentages of the segmented reads to 88%-

776 95% for all samples/tissues (Fig. S3, S4 and Table S2). We implemented only two runs
777 because the remaining unsegmented reads were mostly noisy background reads.
778
779 **Normalization and filtering**
780 Segmented gene reads assigned to cells were then filtered using multiple criteria.
781 First, we removed cells that did not contain reads from any marker genes. Second, we
782 removed cells with total read counts <5 or less than 4 genes with reads. Of the
783 remaining cells, read counts for each gene within a cell were tallied and gene
784 expressions were normalized using the *scTransform* algorithm [11] to further remove
785 biases due to technical variability. The normalized expressions of 220 genes were used
786 for subsequent analysis.
787
788 **Cell typing**
789 We determined the cell type of each cell using an in-house curated list of marker
790 genes for 18 cell types (Table S1). For each cell, we calculated the average expression
791 of marker genes for each of the 18 cell types and then assigned the cell type with the
792 highest average expression to this cell. We further annotated infected cells as those
793 with at least one SARS-CoV-2 read. We used UMAP to visualize the expression pattern
794 of cells associated with the identified cell types in a lower dimension by using the
795 visualization pipeline in the Seurat package [85].
796
797 **Image registration**

798 The DAPI images of tissues that were used as a prior for cell segmentation had
799 translational and rotational differences with their corresponding H&E stained images.
800 For single-cell resolution datasets, small spatial geometric changes in the available raw
801 morphology images could have dramatic differences in downstream analysis for
802 studying the pathology signatures using annotated regions in the H&E image. To
803 mitigate these spatial geometric differences, we registered the DAPI and H&E images
804 by maximizing the phase correlation between both images. Here, we selected the DAPI
805 image as a fixed reference because we used DAPI images as references for reads
806 mapping and cell segmentation. To register both images, the H&E stained image is
807 transformed by changing the scale, rotation, and translation parameters and moved
808 across the DAPI image to improve the phase correlation of both 2D images. Once a
809 peak correlation value was found, the optimization algorithm returned the 2D geometric
810 transformations required to warp the H&E-stained image and registered it to the
811 corresponding DAPI image. We applied the same image registration algorithm to all six
812 samples used in this study. Pathology annotations were carried out on the registered
813 H&E images.
814
815 **Cell density analysis**
816 The abundance of cells relative to the tissue context could reveal biology related
817 to cell proliferation, cell damage, as well as other pathologies related to infection. In our
818 spatial analysis pipeline (Fig. 1B), we defined cell density based on the number of cells
819 within a neighborhood of a fixed radius of 200 units (200 x 0.32 μm/unit = 64 um)
820 centering a cell, covering an area of 0.013 mm^2. By calculating the number of cells

821 within the fixed neighborhoods of each cell, we summarized the global cell density
822 patterns across the tissue. We further presented the cell density as a contour plot by
823 fitting a 2D kernel density estimator (KDE) on the cell density feature (Fig. 3).
824
825 **Neighboring cell type composition analysis**
826 Local cell type distributions can reveal tissue structures and different COVID-19
827 pathology. We assessed the local cell type distribution by computing the neighboring
828 cell type composition (NCTC) of each cell. NCTC defined a cellular neighborhood of 200
829 cells and counts the number of different cell types in the neighborhood to create a
830 neighborhood count vector for each cell. This neighborhood count vector was
831 normalized to obtain the percentages of cell types. NCTC summarized the distributions
832 of each cell type in the vicinity of a cell and was further used to analyze the spatial
833 correlation of different cell types in a region-of-interest (ROI), study spatial gene
834 expression patterns of local neighborhoods, identify local hotspots of individual cell
835 types, and more. Furthermore, the neighborhood composition of individual cell types is
836 spatially mapped by painting the neighborhood count/percentage of the selected cell
837 type with a continuous “rainbow” color scheme where deeper violet to deeper reds
838 indicates a range of low-to-high neighborhood cell composition (Fig. S10-S15).
839
840 **Spatial gene expression map**
841 SSCT captures gene expression of single cells in the context of intact tissue
842 structures. In our study, we investigated the spatial organization of cells in tissue and
843 their gene expression by studying their spatial gene expression maps (SGM). For each

gene, its SGM is defined as the sum of the expression of that gene in its 200 nearest neighboring cells. Thus, SGM represents a spatial gradient by summarizing the regions with high or low gene expression patterns. We then painted the SGM using the continuous rainbow colormap similar to the NCTC analysis, generating the spatial gene expression maps illustrated in Fig. 2B, 2C, and 4F. By comparing the SGM with available pathology annotations, pathology signatures with relevant expression patterns can be localized to specific spatial regions in the tissue. Additionally, novel pathology signatures or ROIs can be discovered using the proposed SGM analysis.

Local infection rate analysis

We used the local infection rate analysis to study the spatial distribution of SARS-Cov-2 infection. It was similar to the NCTC analysis and computed the percentage of infected cells in a fixed physical area of 1.286 mm^2 or a radius of 2,000 units (640 μm) around each cell across the tissue. These spatial rates can be further visualized using the same rainbow color scheme as NCTC (Fig. 3A).

Ligand-receptor coexpression map

Ligand-receptor interactions among cells in a defined location can define spatial patterns of immune microenvironments due to SARS-CoV-2 infection. To infer the interactions, we examine the spatial co-expression between a ligand-receptor pair. Inspired by Moran's I spatial cross-correlation, we computed the spatial coexpression between ligand x and receptor y for cell i as

$$E_i = x_i \sum_{j \in \mathcal{N}_i} y_j$$

where $x_i = 1$ ($y_j = 1$) if ligand x (receptor y) expressed in cell i (j) and 0, otherwise and $\mathcal{N}_i$ defines a set of neighboring cells of cell i, which are cells within a radius of 1,000 units (320 μm) or a fixed physical area of 0.322 mm^2 around cell i. Then, we visualized the coexpression in a tissue using a fixed monochrome color scheme (Fig 3F).

Moran's I analysis of spatial patterns

Spatial global autocorrelation of one variable or more variables can be summarized using Moran's I score. However, calculating the global spatial autocorrelation using Moran's I across the whole tissue assumes homogeneity of the studied variable and yields only one statistic that summarizes the complete spatial pattern across the tissue disregarding their difference over space. Since our SARS-CoV-2 infected tissue samples show spatial heterogeneity, in this study, we used Local Indicators of Spatial Association (LISA) [86] to evaluate the spatial autocorrelation and the statistical significance of a study variable in each location using Local Moran's I.

Consider x_i as, e.g., the neighboring infection rate of a cell at location i, the univariate spatial autocorrelation can be found as the degree of linear association between x_i and a weighted average of the neighboring cells x_j, based on a spatial weight w_{ij} between cells at location i and j. Thus, formally, Moran's I for each location i is given by

$$I_i = \frac{x_i - \bar{x}}{m_2} \sum_{j=1}^{N} w_{ij}\left(x_j - \bar{x}\right)$$

where N defines the number of cells, $\bar{x}$ defines the mean of all the cells, and m_2 is given by

889 $$m_2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}$$

890 I_i ranges from -1 to 1 with -1 indicating highly negative and +1 indicating highly positive
891 spatial autocorrelation. Such formulation can be used to identify spatial regions with
892 high or low autocorrelation. For example, in Fig. 3B, we identified regions with high and
893 low infection rates in each tissue.

894 Under the same formulation, bivariate spatial autocorrelation can be used to
895 assess spatial cross-correlation of a feature x_i, e.g., the neighboring infection rate of a
896 cell at location i and another feature y_j, e.g., the cell density of a neighboring cell j as

897 $$I_i = \frac{x_i - \bar{x}}{m_2} \sum_{j=1}^{N} w_{ij}(y_j - \bar{y})$$

898 where I_i defines the degree of linear association between the neighboring infection rate
899 of a cell at location i and a weighted average of the cell density of the neighboring cells.
900 Bivariate spatial associations can uncover cross-correlations between any such variable
901 x at location i and another variable y at neighboring locations by ignoring correlations
902 between x and y for cases where $i = j$. In our study, the bivariate LISA method is used
903 to calculate the Local Moran's I to identify HIHD regions described in Fig. 3E.
904

905 **Sparse Non-negative matrix factorization (SNMF) of neighboring cell type**
906 **compositions**

907 SNMF was applied to extract interpretable, tissue-specific, and tissue-
908 independent signatures from NCTC vectors from all tissue samples [63]. Let $\boldsymbol{A} \in$
909 $(\mathbf{0}, \mathbf{1})^{N \times M}$ represent the NCTC matrix of N cell types and M cells and it is factored as
910

$$\boldsymbol{A} \approx \boldsymbol{SDH}$$

where $\boldsymbol{S} \in (\boldsymbol{0}, \boldsymbol{1})^{N \times K}$ is the matrix of K NCTC signatures with each column denoting a signature, $\boldsymbol{H} \in (\boldsymbol{0}, \boldsymbol{1})^{K \times M}$ is the signature loading matrix with mth column denoting the contributions of each signature to the NCTC vector of cell m, and $\boldsymbol{D} \in \mathbb{R}^{K \times K}$ is a scaling matrix.

R package *RcppML* was used for the SNMF analysis [63], which implemented an alternating least squares (ALS) algorithm to minimize the mean squared error (MSE) between $\boldsymbol{A}$ and $\boldsymbol{SDH}$. L_1 regularization was introduced to promote a compact, sparse signature loading.

Determining the best number of signatures K is important for uncovering meaningful signatures. While an underestimated number could miss critical signatures, an overestimated number would include many noisy signatures. To this end, we evaluated different maximum numbers of factors from 4 to 12 and examined the meaning regarding its cell type composition and spatial distributions of loadings of each signature and chose $K = 7$ for the SNMF analysis.

Trajectory analysis

The goal of trajectory analysis is to infer a progression or 'pseudotime' of infection severity and associated tissue damages based on NCTC patterns in all tissues. We applied the trajectory analysis to uncover the pseudotime of 1) infection severity among tissues and 2) tissue damage defined by immune microenvironments. For 1), we used loadings of Signature 1 "Normal-like Alveoli" and Signature 3 "Broad

Immune Infiltration" obtained from SNMF and defined tissue samples as clusters. For 2), we used the loadings of Signature 1 and four immune microenvironment-related signatures, *i.e.*, Signature 4-7, and performed Louvain clustering to define the clusters. The R package *Slingshot* was used to infer the trajectory [64]. The loadings, clusters, and the signatures or the 2-D UMAP matrix were fed into *Slingshot* to obtain so-called 'pseudotime' scores for each cell. For both cases, the cluster enriched by cells from the non-infected tissue (PBC-PR) cells was set as the origin.

The pseudotime scores of the cells in each trajectory can correlate with gene expression to uncover genes that correlate with tissue damage. Additionally, ligand-receptor coexpressions that correlate with each trajectory could inform potential ligand-receptor interactions that govern immune responses. To this end, we computed the Spearman correlation of the pseudotime score with the expression of each gene per cell type in each trajectory. We also computed the correlation of ligand-receptor coexpressions with the pseudotime score of each trajectory. In both cases, to properly assess *p-values*, we performed a permutation test to obtain the empirical null distribution of the correlation coefficients and chose $p\text{-value}<0.5$ as the significant level. We then plotted heatmaps to illustrate the cell-type-wise correlations of the genes in each trajectory (Fig. S26A-D).

Publicly available scRNA-seq datasets

Two scRNA-seq datasets were obtained and processed as follows.

1. COVID-19 scRNA-seq dataset [4]. The dataset (SCP1052, lung.h5ad.gz) was downloaded from the Single Cell Portal

(https://singlecell.broadinstitute.org/single_cell). The processed data file lung.h5ad was accessed using the Python package Scanpy (Scanpy 1.9.1). UMAP coordinates for the processed data were accessed from the file upload.scp.X_umap.coords.txt. No follow-up processing was performed on the processed *anndata* object after it was loaded into Scanpy.

2. IPF scRNA-seq dataset [72]. The dataset (GSE135893, GSE135893_ILD_annotated_fullsize.rds.gz) was downloaded from Gene Expression Omnibus (GEO). The Seurat object file was converted into the h5ad file format using the SeuratData package (Seurat 4.3.0, SeuratData 0.2.2). After conversion, the data was loaded in Scanpy (Scanpy 1.9.1) and all cells not originating from an IPF sample or control sample were removed.

Statistics and reproducibility

Reads-based cell segmentation was performed using *Baysor* (version 0.5.2). Nuclei segmentation from DAPI images was performed using *CellPose* (version 2.1.1). All other image analyses including imaging registration were performed using MATLAB (version 2022a). Unless otherwise specified, p-value < 0.05 was considered significant. The differential expression analysis was performed in R (version 3.6.3) using the Wilcoxon rank-sum test with Bonferroni's correction of multiple testing as necessary. The Moran's I analyses for spatial patterns were performed using GeoDa (version 1.20). Pearson correlation coefficients were calculated using function *pearsonr* in Python Scipy (version 1.9.3). UMAP dimension reduction and visualization were performed

using function *DimPlot* in Seurat (version 4.2.0). Trajectory analyses were performed using R package *Slingshot* (version 2.6.0).

Reference

1. Ren, X., et al., *COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas.* Cell, 2021. **184**(23): p. 5838.
2. Melms, J.C., et al., *A molecular single-cell lung atlas of lethal COVID-19.* Nature, 2021. **595**(7865): p. 114-119.
3. Wilk, A.J., et al., *A single-cell atlas of the peripheral immune response in patients with severe COVID-19.* Nat Med, 2020. **26**(7): p. 1070-1076.
4. Delorey, T.M., et al., *COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets.* Nature, 2021. **595**(7865): p. 107-113.
5. Stephenson, E., et al., *Single-cell multi-omics analysis of the immune response in COVID-19.* Nat Med, 2021. **27**(5): p. 904-916.
6. Erjefalt, J.S., et al., *Diffuse alveolar damage patterns reflect the immunological and molecular heterogeneity in fatal COVID-19.* EBioMedicine, 2022. **83**: p. 104229.
7. Asp, M., J. Bergenstrahle, and J. Lundeberg, *Spatially Resolved Transcriptomes-Next Generation Tools for Tissue Exploration.* Bioessays, 2020. **42**(10): p. e1900221.
8. Larsson, L., J. Frisen, and J. Lundeberg, *Spatially resolved transcriptomics adds a new dimension to genomics.* Nat Methods, 2021. **18**(1): p. 15-18.
9. Ramos da Silva, S., et al., *Broad Severe Acute Respiratory Syndrome Coronavirus 2 Cell Tropism and Immunopathology in Lung Tissues From Fatal Coronavirus Disease 2019.* J Infect Dis, 2021. **223**(11): p. 1842-1854.
10. Petukhov, V., et al., *Cell segmentation in imaging-based spatial transcriptomics.* Nat Biotechnol, 2022. **40**(3): p. 345-354.
11. Hafemeister, C. and R. Satija, *Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression.* Genome Biol, 2019. **20**(1): p. 296.
12. Arish, M., et al., *COVID-19 immunopathology: From acute diseases to chronic sequelae.* J Med Virol, 2022. **95**(1): p. e28122.
13. Whitsett, J.A., S.E. Wert, and T.E. Weaver, *Alveolar surfactant homeostasis and the pathogenesis of pulmonary disease.* Annu Rev Med, 2010. **61**: p. 105-19.
14. Lin, Z., et al., *Genetic Association of Pulmonary Surfactant Protein Genes, SFTPA1, SFTPA2, SFTPB, SFTPC, and SFTPD With Cystic Fibrosis.* Front Immunol, 2018. **9**: p. 2256.
15. Floros, J., et al., *Human Surfactant Protein SP-A1 and SP-A2 Variants Differentially Affect the Alveolar Microenvironment, Surfactant Structure, Regulation and Function of the Alveolar Macrophage, and Animal and Human Survival Under Various Conditions.* Front Immunol, 2021. **12**: p. 681639.
16. Mikerov, A.N., et al., *Inhibition of hemagglutination activity of influenza A viruses by SP-A1 and SP-A2 variants expressed in CHO cells.* Med Microbiol Immunol, 2008. **197**(1): p. 9-12.
17. Garcia-Laorden, M.I., et al., *The role of mannose-binding lectin in pneumococcal infection.* Eur Respir J, 2013. **41**(1): p. 131-9.

18. Sawada, K., et al., *Pulmonary collectins protect macrophages against pore-forming activity of Legionella pneumophila and suppress its intracellular growth.* J Biol Chem, 2010. **285**(11): p. 8434-43.
19. Fassan, M., et al., *Multi-Design Differential Expression Profiling of COVID-19 Lung Autopsy Specimens Reveals Significantly Deregulated Inflammatory Pathways and SFTPC Impaired Transcription.* Cells, 2022. **11**(6): p. 1011.
20. Perkins, T.N., M.L. Donnell, and T.D. Oury, *The axis of the receptor for advanced glycation endproducts in asthma and allergic airway disease.* Allergy, 2021. **76**(5): p. 1350-1366.
21. Yue, Q., et al., *Receptor for Advanced Glycation End Products (RAGE): A Pivotal Hub in Immune Diseases.* Molecules, 2022. **27**(15): p. 4922.
22. Koerich, S., et al., *Receptors for Advanced Glycation End Products (RAGE): Promising Targets Aiming at the Treatment of Neurodegenerative Conditions.* Curr Neuropharmacol, 2022. Sep 22. doi: 10.2174/1570159X20666220922153903.Online ahead of print
23. Yamaguchi, K., et al., *Association of the RAGE/RAGE-ligand axis with interstitial lung disease and its acute exacerbation.* Respir Investig, 2022. **60**(4): p. 531-542.
24. Hofmann, M.A., et al., *RAGE mediates a novel proinflammatory axis: a central cell surface receptor for S100/calgranulin polypeptides.* Cell, 1999. **97**(7): p. 889-901.
25. Kapandji, N., et al., *Importance of Lung Epithelial Injury in COVID-19-associated Acute Respiratory Distress Syndrome: Value of Plasma Soluble Receptor for Advanced Glycation End-Products.* Am J Respir Crit Care Med, 2021. **204**(3): p. 359-362.
26. Chiappalupi, S., et al., *Hyperactivated RAGE in Comorbidities as a Risk Factor for Severe COVID-19-The Role of RAGE-RAS Crosstalk.* Biomolecules, 2021. **11**(6): p. 876.
27. Zhang, K. and S.H. Phan, *Cytokines and pulmonary fibrosis.* Biol Signals, 1996. **5**(4): p. 232-9.
28. Zhu, Y., et al., *Calcium in Vascular Smooth Muscle Cell Elasticity and Adhesion: Novel Insights Into the Mechanism of Action.* Front Physiol, 2019. **10**: p. 852.
29. Ringvold, H.C. and R.A. Khalil, *Protein Kinase C as Regulator of Vascular Smooth Muscle Function and Potential Target in Vascular Disorders.* Adv Pharmacol, 2017. **78**: p. 203-301.
30. Morita, K., et al., *Endothelial claudin: claudin-5/TMVCF constitutes tight junction strands in endothelial cells.* J Cell Biol, 1999. **147**(1): p. 185-94.
31. Schlingmann, B., S.A. Molina, and M. Koval, *Claudins: Gatekeepers of lung epithelial function.* Semin Cell Dev Biol, 2015. **42**: p. 47-57.
32. Williams, M.R., et al., *Gene expression of endothelial cells due to interleukin-1 beta stimulation and neutrophil transmigration.* Endothelium, 2008. **15**(1): p. 73-84.
33. Hashimoto, R., et al., *SARS-CoV-2 disrupts respiratory vascular barriers by suppressing Claudin-5 expression.* Sci Adv, 2022. **8**(38): p. eabo6783.
34. Pang, L., et al., *Role of caveolin-1 in human organ function and disease: friend or foe?* Carcinogenesis, 2022. **43**(1): p. 2-11.

35. Gokani, S. and L.K. Bhatt, *Caveolin-1: A Promising Therapeutic Target for Diverse Diseases.* Curr Mol Pharmacol, 2022. **15**(5): p. 701-715.
36. Lagrange, J., et al., *Alpha-2-macroglobulin in hemostasis and thrombosis: An underestimated old double-edged sword.* J Thromb Haemost, 2022. **20**(4): p. 806-815.
37. Vandooren, J. and Y. Itoh, *Alpha-2-Macroglobulin in Inflammation, Immunity and Infections.* Front Immunol, 2021. **12**: p. 803244.
38. Mathison, R., J.S. Davison, and A.D. Befus, *Neural regulation of neutrophil involvement in pulmonary inflammation.* Comp Biochem Physiol C Comp Pharmacol Toxicol, 1993. **106**(1): p. 39-48.
39. Sahu, S.K., et al., *Emerging roles of the complement system in host-pathogen interactions.* Trends Microbiol, 2022. **30**(4): p. 390-402.
40. Magro, C., et al., *Complement associated microvascular injury and thrombosis in the pathogenesis of severe COVID-19 infection: A report of five cases.* Transl Res, 2020. **220**: p. 1-13.
41. Mastaglio, S., et al., *The first case of COVID-19 treated with the complement C3 inhibitor AMY-101.* Clin Immunol, 2020. **215**: p. 108450.
42. Afzali, B., et al., *The state of complement in COVID-19.* Nat Rev Immunol, 2022. **22**(2): p. 77-84.
43. Freeley, S., et al., *Asparaginyl Endopeptidase (Legumain) Supports Human Th1 Induction via Cathepsin L-Mediated Intracellular C3 Activation.* Front Immunol, 2018. **9**: p. 2449.
44. Zizzo, G., et al., *Immunotherapy of COVID-19: Inside and Beyond IL-6 Signalling.* Front Immunol, 2022. **13**: p. 795315.
45. Zhao, M.M., et al., *Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development.* Signal Transduct Target Ther, 2021. **6**(1): p. 134.
46. Trapani, J.A. and M.J. Smyth, *Killing by cytotoxic T cells and natural killer cells: multiple granule serine proteases as initiators of DNA fragmentation.* Immunol Cell Biol, 1993. **71 (Pt 3)**: p. 201-8.
47. Berezin, V., et al., *Targeting of ECM molecules and their metabolizing enzymes and receptors for the treatment of CNS diseases.* Prog Brain Res, 2014. **214**: p. 353-88.
48. Pimentel, E., *Colony-stimulating factors.* Ann Clin Lab Sci, 1990. **20**(1): p. 36-55.
49. Venet, F., et al., *Myeloid cells in sepsis-acquired immunodeficiency.* Ann N Y Acad Sci, 2021. **1499**(1): p. 3-17.
50. Roussel, L. and D.C. Vinh, *ICOSL in host defense at epithelial barriers: lessons from ICOSLG deficiency.* Curr Opin Immunol, 2021. **72**: p. 21-26.
51. Panneton, V., et al., *Inducible T-cell co-stimulator: Signaling mechanisms in T follicular helper cells and beyond.* Immunol Rev, 2019. **291**(1): p. 91-103.
52. Mulcahy, H., et al., *LST1 and NCR3 expression in autoimmune inflammation and in response to IFN-gamma, LPS and microbial infection.* Immunogenetics, 2006. **57**(12): p. 893-903.
53. Heidemann, J., et al., *Regulated expression of leukocyte-specific transcript (LST) 1 in human intestinal inflammation.* Inflamm Res, 2014. **63**(7): p. 513-7.

54. Chen, Z., X. Qiu, and J. Gu, *Immunoglobulin expression in non-lymphoid lineage and neoplastic cells.* Am J Pathol, 2009. **174**(4): p. 1139-48.
55. Vettermann, C. and M.S. Schlissel, *Allelic exclusion of immunoglobulin genes: models and mechanisms.* Immunol Rev, 2010. **237**(1): p. 22-42.
56. Napodano, C., et al., *Mono/polyclonal free light chains as challenging biomarkers for immunological abnormalities.* Adv Clin Chem, 2022. **108**: p. 155-209.
57. Godyna, S., M. Diaz-Ricart, and W.S. Argraves, *Fibulin-1 mediates platelet adhesion via a bridge of fibrinogen.* Blood, 1996. **88**(7): p. 2569-77.
58. Harikrishnan, K., et al., *Cell Derived Matrix Fibulin-1 Associates With Epidermal Growth Factor Receptor to Inhibit Its Activation, Localization and Function in Lung Cancer Calu-1 Cells.* Front Cell Dev Biol, 2020. **8**: p. 522.
59. Jackson, C.B., et al., *Mechanisms of SARS-CoV-2 entry into cells.* Nat Rev Mol Cell Biol, 2022. **23**(1): p. 3-20.
60. Felsenstein, S., et al., *COVID-19: Immunology and treatment options.* Clin Immunol, 2020. **215**: p. 108448.
61. Guizani, I., et al., *SARS-CoV-2 and pathological matrix remodeling mediators.* Inflamm Res, 2021. **70**(8): p. 847-858.
62. Zhang, F., et al., *SARS-CoV-2 pseudovirus infectivity and expression of viral entry-related factors ACE2, TMPRSS2, Kim-1, and NRP-1 in human cells from the respiratory, urinary, digestive, reproductive, and immune systems.* J Med Virol, 2021. **93**(12): p. 6671-6685.
63. DeBruine, Z.J., K. Melcher, and T.J. Triche Jr, *Fast and robust non-negative matrix factorization for single-cell experiments.* bioRxiv, 2021: p. 2021.09. 01.458620.
64. Street, K., et al., *Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics.* BMC Genomics, 2018. **19**(1): p. 477.
65. Sweeney, R.M. and D.F. McAuley, *Acute respiratory distress syndrome.* Lancet, 2016. **388**(10058): p. 2416-2430.
66. Wynn, T.A., *Cellular and molecular mechanisms of fibrosis.* J Pathol, 2008. **214**(2): p. 199-210.
67. Turpin, E.A., et al., *Respiratory syncytial virus infection reduces lung inflammation and fibrosis in mice exposed to vanadium pentoxide.* Respir Res, 2010. **11**: p. 20.
68. Elewa, Y.H.A., et al., *Histopathological Correlations between Mediastinal Fat-Associated Lymphoid Clusters and the Development of Lung Inflammation and Fibrosis following Bleomycin Administration in Mice.* Front Immunol, 2018. **9**: p. 271.
69. Nagata, N., et al., *Features of idiopathic pulmonary fibrosis with organizing pneumonia.* Respiration, 1997. **64**(5): p. 331-5.
70. Phan, S.H., *The myofibroblast in pulmonary fibrosis.* Chest, 2002. **122**(6 Suppl): p. 286S-289S.
71. Ortiz-Zapater, E., et al., *Lung Fibrosis and Fibrosis in the Lungs: Is It All about Myofibroblasts?* Biomedicines, 2022. **10**(6).
72. Habermann, A.C., et al., *Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis.* Sci Adv, 2020. **6**(28): eaba1972.

73. Laloglu, E. and H. Alay, *Role of transforming growth factor-beta 1 and connective tissue growth factor levels in coronavirus disease-2019-related lung Injury: a prospective, observational, cohort study.* Rev Soc Bras Med Trop, 2022. **55**: p. e06152021.
74. Vaz de Paula, C.B., et al., *COVID-19: Immunohistochemical Analysis of TGF-beta Signaling Pathways in Pulmonary Fibrosis.* Int J Mol Sci, 2021. **23**(1): p. 168.
75. Fan, T., et al., *Discovery of 9O-Substituted Palmatine Derivatives as a New Class of anti-COL1A1 Agents via Repressing TGF-beta1/Smads and JAK1/STAT3 Pathways.* Molecules, 2020. **25**(4): p.773.
76. Malik, M., et al., *Cross-talk between Janus kinase-signal transducer and activator of transcription pathway and transforming growth factor beta pathways and increased collagen1A1 production in uterine leiomyoma cells.* F S Sci, 2020. **1**(2): p. 206-220.
77. Stringer, C., et al., *Cellpose: a generalist algorithm for cellular segmentation.* Nat Methods, 2021. **18**(1): p. 100-106.
78. Qi, F., et al., *Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses.* Biochem Biophys Res Commun, 2020. **526**(1): p. 135-140.
79. Sungnak, W., et al., *SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes.* Nat Med, 2020. **26**(5): p. 681-687.
80. Xu, G., et al., *The differential immune responses to COVID-19 in peripheral and lung revealed by single-cell RNA sequencing.* Cell Discov, 2020. **6**: p. 73.
81. Zou, X., et al., *Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection.* Front Med, 2020. **14**(2): p. 185-192.
82. Sounart, H., et al., *Dual spatially resolved transcriptomics for SARS-CoV-2 host-pathogen colocalization studies in humans.* bioRxiv, 2022: p. 2022.03.14.484288.
83. Pielawski, N., et al., *TissUUmaps 3: Interactive visualization and quality assessment of large-scale spatial omics data.* BioRxiv, 2022. doi: https://doi.org/10.1101/2022.01.28.478131.
84. Solorzano, L., G. Partel, and C. Wahlby, *TissUUmaps: interactive visualization of large-scale spatial gene expression and tissue morphology data.* Bioinformatics, 2020. **36**(15): p. 4363-4365.
85. Satija, R., et al., *Spatial reconstruction of single-cell gene expression data.* Nat Biotechnol, 2015. **33**(5): p. 495-502.
86. Anselin, L., *Local indicators of spatial association—LISA.* Geographical analysis, 1995. **27**(2): p. 93-115.

Acknowledgments

We thank members of Drs. Shou-Jiang Gao and Yufei Huang laboratories for technical assistance and discussions. This study was supported by grants from the National Institutes of Health (CA096512 and CA124332 to S.-J. Gao), UPMC Hillman Cancer Center Startup Funds to S.-J. Gao and Y. Huang, and in part by award P30CA047904. This research was also supported in part by the University of Pittsburgh Center for Research Computing, RRID:SCR_022735, through the resources provided. Specifically, this work used the HTC cluster, which is supported by the National Institutes of Health (S10OD028483).

Author contributions

S.-J. Gao and Y. Huang conceived, designed, supervised and managed the project. W. Meng, S.R. da Silva and L.P. Chen performed the experiments. G.L. Sica and W. Meng performed the pathological examination. A. Das, Z.T. Liu, D.M. Hasib, H. Galloway and Y. Huang developed the computational pipeline and performed the analyses. A. Das, W. Meng, Z.T. Liu, D.M. Hasib, H. Galloway, S.R. da Silva, L.P. Chen, Y.F., K.P. Rivera, M. Flores, Y.-C. Chiu, Y. Huang and S.-J. Gao interpreted the data and participated in the discussions throughout the analysis. A. Paniz-Mondolfi, C. Bryce, Z. Grimes, E.M. Sodillo and C. Cordon-Cardo obtained the COVID-19 samples. Y. Huang and S.-J. Gao wrote the manuscript with input from all the authors. All the authors read, reviewed and approved the manuscript.

Competing interests

The authors declare no competing interests.

Data availability

The data supporting findings in this study including all the original spatial single-cell transcriptome data, segmented and filtered reads of all cells, and DAPI and H&E staining images will be deposited to Zenodo. The spatial single-cell transcriptome data will also be deposited in GEO.

Code availability

All codes will be available at Zenodo.

Supplementary information

The online version contains supplementary material available at:

Correspondence and requests for materials should be addressed to Shou-Jiang Gao or Yufei Huang.

Figure Legend

Figure 1: Spatial organization of parenchymal and immune cells in COVID-19 lung tissues revealed by spatial single-cell transcriptome analysis (SSCTA). A. Schematic illustration of the spatial single-cell sequencing pipeline used to generate the target dataset comprised of healthy control (non-COVID-19 tissue, n=1) and COVID-19 tissues (n=5). Samples were hybridized and decoded by sequencing. DAPI and H&E staining were carried out. All samples were annotated by an expert pathologist (Fig. S1). **B.** SSCTA workflow illustrating cell segmentation, cell typing, neighborhood cell-type composition (NCTC) analysis, and tissue pathology analysis. **C.** Average expression of cell type markers of all segmented cells across identified cell types in the SSCTA dataset. **D.** The UMAP projection of the gene expressions of segmented cells colored based on cell type, infection, and sample identifier. The UMAP shows negligible batch differences between the samples. **E.** Spatial visualization of identified cell types in two COVID-19 tissues and the non-COVID-19 tissue along with individual spatial plots illustrating the neighborhood cell type composition of specific cell types. We identified DAD in all COVID-19 tissues and orderly distribution of major parenchymal cells in non-COVID-19 tissue (PBC-PR). Further plots are provided in Fig. S1. **F.** Bar plots showing the percentages of identified cell types and their percentages of SARS-CoV-2 infected cells.

Figure 2: SARS-CoV-2 infection induces global differential gene expression that mediates pathological damages and inflammation with cell type and spatial features. A. Differential gene expressions between the identified cells in the COVID-19

non-COVID-19 tissue with various conditions are illustrated. The title indicates the conducted differential expression analysis. Within each row, the bubble size indicates the -log10 of the corrected p-value and the color indicates the log2 fold change of the corresponding gene. Each column indicates the different COVID-19 tissues. The first three panels indicate differential expression analysis between COVID-19 tissues and non-COVID-19 tissue while the fourth panel indicates the differential expression analysis between the infected and uninfected cells in the respective COVID-19 tissues. **B.** Spatial gene expression map of highly expressed CTSL and IGKC. **C.** Spatial gene expression maps of the rest of the differentially expressed genes. **D.** Pie chart illustrating the percentages of cell types in which the genes are expressed in each tissue. **E.** Differential expression analysis of COVID-19 tissues compared to non-COVID-19 tissue in cell types that had significant changes in cell numbers based on Fig. 1F.

Figure 3: Spatial analysis of local SARS-CoV-2 infection rates identifies regions with high infection rates that match high cell densities, high levels of viral entry-related factors, and localized pathology. A. Spatial distribution of the SARS-CoV-2 infected cells. The spatial visualization of the infection rates revealed infection hotspots in the COVID-19 tissues. **B.** Moran's I analysis of spatial patterns revealed spatially distinct high and low infection regions in each COVID-19 tissue. **C.** Spatial visualization of the significant local regions identified by Bivariate Moran's I analysis, in which infection rates and the expressions of TMPRSS2, NRP1, FURIN and ACE2 are spatially correlated. **D.** Spatial visualization of the kernel density plot for the cell densities

revealed highly dense cellular hotspots in the COVID-19 samples. **E.** Spatial visualization of the significant local regions identified by Bivariate Moran's I analysis, in which cell densities and the infection regions are spatially correlated, identified as high-density high-infection (HIHD) regions. **F.** Spatial visualization of the ligand-receptor coexpression maps for the CSF3-CSF3R pair, which are highly expressed in the high infection regions. **G.** Spatial visualization of the significant local regions identified by Bivariate Moran's I analysis, in which both high infection and high CSF3-CSF3R coexpression regions are spatially correlated.

Figure 4: Spatial analysis reveals cellular and gene expression features in regions with high SARS-CoV-2 infection rates. A. A bar plot showing the cell type composition of all identified cell types in the high and low infection regions. The increase in alveolar cells (ACs) might suggest the proliferation of ACs due to injury repair of alveoli and alveoli capillaries and a decrease in epithelial cells (ECs) and vascular endothelial cells (VECs) might suggest the damages caused by the infection. **B.** UMAP projection of the NCTC analysis results of high infection regions (refer to figure 3B) revealed three distinct clusters out of which two were distinctly enriched in local compositions of ACs and fibroblasts, which are illustrated on the right with their spatial visualizations. **C.** Spatial visualization of the compositions in high infection regions revealed negative correlation of fibroblasts and ACs (-0.6, Pearson correlation) which overlapped with annotated OP regions (high fibroblasts, low ACs). **D.** Differential expressions of high and low infection regions of the COVID-19 tissues compared to the non-COVID-19 tissue, which share 18 of the 20 differentially expressed genes,

confirming strong indirect effect of SARS-CoV-2 infection. **E.** Differential expressions between the high and low infection regions of the COVID-19 tissues revealed distinct differentially expressed genes showing upregulation of AC markers SFTPA1 and SFTPA2 in tissues 1, 3, and 5, and fibroblast markers COL1A1 or COL1A2 in all COVID-19 tissues. **F.** Spatial visualization of the spatial gene expression map of COL1A1 and COL1A2 markers.

Figure 5: Sparse non-negative matrix factorization (SNMF) identifies seven cell composition signatures that recapitulate different healthy and disease statuses. **A.** SNMF revealed seven NCTC signatures which are illustrated in a logo plot. **B.** Spatial NCTC signatures visualized after factor loading for each tissue revealed their associations with normal structures and broad and tissue-specific infection. **C.** Trajectory analysis of the cell type compositions with Signatures 1 and 3 generated a trajectory of reduced normal alveoli and increased immune infiltration from the non-COVID-19 tissue to COVID-19 tissues 1, 4, 3, 5, and 2. The trajectory agrees with the order of the severity of pathology in the tissues identified in the H&E stained images (Fig. S1).

Figure 6: Unique cell composition signatures define two separate pathological trajectories. A. The UMAP projection of the Signatures 1 (for non-COVID-19 tissue reference) and Signatures 4-7 recapitulated the severity of pathological progression identified in figure 5C. Two trajectories were derived from the pseudo-progressions of cells defined by these signatures. The trajectories showed a common shared onset T1

and T2 which traverses through Signature 1 “normal-like-alveoli” and Signature 4 “ACs with MMs infiltration” in T1 and then those enriched by Signature 4 and Signature 5 “VECs with MMs infiltration” in T2. The trajectory then diverged into two paths T3a and T4a, and T3b and T4b. Path/Trajectory A defined by T3a and T4a progressed through cells mostly from tissues 3 and 4 enriched in Signature 5 “VEC with MMs infiltration” ending in Signature 6 “VEC with NK cell infiltration”, which was almost exclusively from tissue 3. Path/Trajectory B defined by T3b and T4b are enriched by Signature 6 in T3b and Signature 7 “high fibroblasts with MMs infiltration” in T4b. **B.** UMAP projections of the five signatures colored based on their individual signature loadings. **C.** Line plots illustrating the cell type percentages in band T1, T2, T3(a, b), and T4(a, b). **D.** Spatial visualization of the cells by cell type in each band is shown along with an identified representative ROI (red box with dotted lines) for each COVID-19 tissue. Column 7 illustrates the H&E morphology image of the COVID-19 tissues with black dotted lines highlighting the high infection regions. **E.** Zoomed-in images of the ROIs selected are shown. A close examination revealed characteristics of early DAD in T1 and T2, hyaline membrane, fibrin, lymphocytic infiltration, collapsed alveoli and capillary structure in T3a and T4a, and edema, increased fibrin, and organizing pneumonia (OP) in T3b and T4b suggesting a more advanced stage of DAD.

Figure 7: IL6-STAT3 and TGF-β-SMAD2/3 pathways mediate COVID-19 lung fibrosis and organizing pneumonia. A. Zoomed in OP region (bottom, red color border) and other non-OP regions (top, black color border) from the H&E image of COVID-19 tissue 1-2C are illustrated. **B.** Identified cells colored based on the cell types

revealed differences in cell composition signatures between OP and other regions. **C.** Bar plot showing an increase in cell density in OP regions compared to other regions. **D.** Bar plot showing the percentage of cells per cell type in COVID-19 and non-COVID-19 tissues. There is a decrease in ACs and VECs, and an increase in fibroblasts in the OP regions of the COVID-19 tissue. An obvious increase in ACs is seen in the non-COVID-19 sample. **E.** Differential gene expressions between the identified cells in the OP regions of the COVID-19 tissue sample with cells from the non-COVID-19 tissue and cells from the other non-OP regions of the same COVID-19 tissue is illustrated. We observed an upregulation of genes encoding for collagen type I alpha 1 and 2 chains (COL1A1 and COL1A2) in OP in both tissues compared to the normal tissue or to regions without OP in the same COVID-19 lung tissues. **F.** Pie chart illustrating the percentages of cell types in which the selected genes are expressed in tissues 1 and 2. **G.** Spatial visualization of differentially expressed genes from figure 7E. The size of the dots indicates the expression level, and the color indicates the cell type of the cell where the gene is expressed. **H.** Differential gene expressions between the identified cells in the OP regions of the COVID-19 tissue with cells from the non-COVID-19 tissue as well as non-OP regions from the same COVID-19 tissue based on their cell types. Both COL1A1 and COL1A2 were upregulated in fibroblasts in OP regions compared to fibroblasts from either normal tissue or non-OP regions from the same tissue. **I.** The density plot of the COL1A1 and COL1A2 expressions in OP and other non-OP regions in COVID-19 tissue samples 1-2C, 2-1A, and non-COVID-19 tissue is illustrated. **J.** Examination of the networks associated with the Pulmonary Fibrosis Idiopathic Signaling Pathways reported by the Ingenuity Pathway Analysis is illustrated. Both IL6-

STAT3 and TGF-β-SMAD2/3 pathways mediate the expression of COL1A1 and COL1A2. **K.** Multi-color IFA staining revealed that cells with a high COL1A1 expression level also had high levels of IL6Rα, p-STAT3, TGF-βR2 and SMAD2/3, confirming the activation of these pathways and their potential roles in the upregulation of COL1A1.

Supplemental Information

Supplemental Tables

Table S1. Gene annotation, cell type markers, and total cells expressing individual genes and total reads of the individual genes in each lung tissue after segmentation.

Table S2. Summary of cell segmentation results.

Table S3. Summary of cell typing results and SARS-CoV-2 infection status.

Table S4. Global spatial correlations between local SARS-CoV-2 infection rates and cell densities by global Moran's I.

Table S5. Cell composition signatures identified by sparse non-negative matrix factorization (SNMF).

Table S6. Spearman correlations of pseudotime Trajectory A versus gene expressions.

Table S7. Spearman correlations of pseudotime Trajectory B versus gene expressions.

Supplementary Figures

Supplementary Figure 1: Hematoxylin and eosin (H&E) stained images of the five COVID-19 tissues (1-2C, 2-1A, 3-1A, 4-3B, 5-3B) and the non-COVID-19 tissue (PBC-PR) along with pathology annotations are illustrated. We observe various degrees of diffuse alveolar damage (DAD) in the COVID-19 tissues. Tissues 1 and 2 (1-2C and 2-1A) had prominent organizing pneumonia (OP) while edema, hyaline membrane, and fibrin clot were prominent in tissues 3, 4, and 5 (3-1A, 4-3B, and 5-3B).

Supplementary Figure 2: 4′,6-diamidino-2-phenylindole (DAPI) staining of all tissues under study revealed the nuclei location. We used the DAPI-stained images to facilitate cell nuclei segmentation using the CellPose algorithm.

Supplementary Figure 3: A. Bar plot showing the total number of reads and the reads that are segmented to cells using the Baysor algorithm for all tissues under study. We segmented ~89%-95% of the reads into cells. **B.** Bar plot showing the total number of segmented cells and the identified cells with a cell-marker read. A total of 1,719,459 cells were identified across the tissues.

Supplementary Figure 4: A representative image showing cell boundary polygons of the cells segmented using the Baysor algorithm. Baysor used a binary mask of the DAPI image as a prior to guide the cell segmentation.

Supplementary Figure 5: Bar plots showing the distribution of the number of transcripts (reads) that are segmented to each cell. Over 99% of the cells harbored at least 5-15 reads.

Supplementary Figure 6: Heatmaps illustrating the average expressions of marker genes in each cell type within each tissue. In all tissues, a total of 18 cell types were identified including 11 types of parenchymal cells and 7 types of immune cells.

Supplementary Figure 7: Uniform Manifold Approximation and Projection (UMAP) of the gene expressions of individual cells revealed separated clusters for major parenchymal cells including alveolar cells (ACs) and fibroblasts. By contrast, immune cells were grouped together and mixed with vascular endothelial cells (VECs) in several clusters. The poor separation of immune cell types was largely due to highly expressed Immunoglobulin kappa light chain (IGKC) and Cathepsin L (CTSL) in these cell types, which is further illustrated in Fig. S8.

Supplementary Figure 8: The UMAP of the gene expressions of all segmented and cell-typed cells across all tissues under study is illustrated in the bottom-right panel. Here, the color of the dots indicates the tissue samples. The rest of the UMAPs illustrate the expressions of different genes colored based on their expression in each cell. We observe that IGKC and CTSL are spread across many clusters, resulting in the poor separation of immune cell types which are mixed with VECs.

Supplementary Figure 9: Spatial visualization of the segmented cells painted based on their cell types along with the location of the SARS-CoV-2 reads marked with an 'X'.

Supplementary Figure 10: The neighborhood cell type composition (NCTC) analysis for COVID-19 tissue 1-2C is illustrated based on the cell type.

Supplementary Figure 11: The neighborhood cell type composition (NCTC) analysis for COVID-19 tissue 2-1A is illustrated based on the cell type.

Supplementary Figure 12: The neighborhood cell type composition (NCTC) analysis for COVID-19 tissue 3-1A is illustrated based on the cell type.

Supplementary Figure 13: The neighborhood cell type composition (NCTC) analysis for COVID-19 tissue 4-3B is illustrated based on the cell type.

Supplementary Figure 14: The neighborhood cell type composition (NCTC) analysis for COVID-19 tissue 5-3B is illustrated based on the cell type.

Supplementary Figure 15: The neighborhood cell type composition (NCTC) analysis for non-COVID-19 tissue PBC-PR is illustrated based on the cell type.

Supplementary Figure 16: Spatial visualization of regions with blood vessels revealed the lining of VECs along the vessels together with other vessel-associated cells

including ACs, fibroblasts and smooth muscle cells (SMCs). Each dot indicates a cell, and the color of each dot indicates the cell type.

Supplementary Figure 17: Spatial visualization of a representative region with bronchioles revealed the lining of epithelial cells (ECs), smooth muscle cells (SMCs) and VECs along the bronchiole together with other bronchiole-associated cells including ACs and fibroblasts. Each dot indicates a cell, and the color of each dot indicates the cell type.

Supplementary Figure 18: Spatial visualization of the expressions of the 19 dysregulated genes in different cell types and their tissue distributions in COVID-19 tissue 1-2C are illustrated. We observe significant variations in the expression patterns suggesting their complex involvements in different aspects of SARS-CoV-2 infection and COVID-19 lung pathology. The color of each dot indicates the cell type, and the size indicates their expression level.

Supplementary Figure 19: Spatial visualization of the expressions of the 19 dysregulated genes in different cell types and their tissue distributions in COVID-19 tissue 2-1A are illustrated. We observe significant variations in the expression patterns suggesting their complex involvements in different aspects of SARS-CoV-2 infection and COVID-19 lung pathology. The color of each dot indicates the cell type, and the size indicates their expression level.

Supplementary Figure 20: Spatial visualization of the expressions of the 19 dysregulated genes in different cell types and their tissue distributions in COVID-19 tissue 3-1A are illustrated. We observe significant variations in the expression patterns suggesting their complex involvements in different aspects of SARS-CoV-2 infection and COVID-19 lung pathology. The color of each dot indicates the cell type, and the size indicates their expression level.

Supplementary Figure 21: Spatial visualization of the expressions of the 19 dysregulated genes in different cell types and their tissue distributions in COVID-19 tissue 4-3B are illustrated. We observe significant variations in the expression patterns suggesting their complex involvements in different aspects of SARS-CoV-2 infection and COVID-19 lung pathology. The color of each dot indicates the cell type, and the size indicates their expression level.

Supplementary Figure 22: Spatial visualization of the expressions of the 19 dysregulated genes in different cell types and their tissue distributions in COVID-19 tissue 5-3B are illustrated. We observe significant variations in the expression patterns suggesting their complex involvements in different aspects of SARS-CoV-2 infection and COVID-19 lung pathology. The color of each dot indicates the cell type, and the size indicates their expression level.

Supplementary Figure 23: Spatial visualization of the expressions of the 19 dysregulated genes in different cell types and their tissue distributions in non-COVID-19

tissue PBC-PR are illustrated. We observe significant variations in the expression patterns suggesting their complex involvements in different aspects of SARS-CoV-2 infection and COVID-19 lung pathology. The color of each dot indicates the cell type, and the size indicates their expression level.

Supplementary Figure 24: Bubble plots of cell-type specific differential expression analysis of the high infection versus the low infection regions of the same COVID-19 tissues is illustrated. The analysis further confirmed the upregulation of SFTPA1 or SFTPA2 in ACs and COL1A2 in fibroblasts in infected tissues. Within each row, the bubble size indicates the -log10 of the corrected p-value and the color indicates the log2 fold change of the corresponding gene.

Supplementary Figure 25: Box plots of signature loadings for each identified signature generated by the SNMF decomposition of NCTC vectors are illustrated.

Supplementary Figure 26: A. Heatmap illustrating the Spearman correlation of the pseudotime scores of Trajectory A and the gene expression. **B.** Heatmap illustrating the Spearman correlation of the pseudotime scores of Trajectory B and the gene expression. **C.** Heatmap illustrating the Spearman correlation of the pseudotime scores of Trajectory A and the ligand-receptor co-expression values of each cell. **D.** Heatmap illustrating the Spearman correlation of the pseudotime scores of Trajectory B and the ligand-receptor co-expression values of each cell. In all plots, deeper red indicates a high positive correlation and deeper blue indicates a high negative correlation. We also

apply a p-value threshold (p-values<0.5) to remove the genes with lesser correlations and the removed genes are painted white to indicate no correlation.

Supplementary Figure 27: Analysis of COVID-19 and IPF single-cell RNA-seq datasets. A. UMAP projections showing expression of COL1A1 and COL1A2, paired with fibroblast and myofibroblast cell subtypes from single-cell RNA-seq COVID-19 tissue atlas. Violin plots showing expression of COL1A1 and COL1A2 in cells from COVID-19 patients. **B.** GSEA enrichment of myofibroblasts in cells ranked by expressions of COL1A1 and COL1A2 from the single-cell COVID-19 tissue atlas with a resulting p-value < 0.001. The significant enrichment and the large number of highly ranked hits indicate that most myofibroblasts from the single-cell COVID-19 tissue atlas have high expressions of COL1A1 and COL1A2. **C.** UMAP projections showing expressions of COL1A1 and COL1A2 next to UMAP of fibroblast cell subtypes from single-cell RNA-seq pulmonary fibrosis dataset. Violin plots showing expressions of COL1A1 and COL1A2 in IPF and control patients. The UMAPs and violin plots only represent cells from IPF patients or control. **D.** GSEA enrichment of myofibroblast in cells from IPF and control donors ranked by expressions of COL1A1 and COL1A2 with a resulting p-value < 0.001. The significant enrichment and high correlation between the myofibroblast cell type and high expressions of COL1A1 and COL1A2 indicate that most myofibroblasts from the single-cell pulmonary fibrosis dataset have high expressions of COL1A1 and COL1A2. **E.** A heatmap showing expressions of significantly differentially expressed genes from the ‘IPF High’ run in cells with high and low expressions of COL1A1 and COL1A2. **F.** A heatmap showing expressions of

significantly differentially expressed genes from the 'IPF Diagnosis' cells with high and low expressions of COL1A1 and COL1A2. **G.** GSEA enrichment of significantly differentially expressed genes from the 'IPF Diagnosis' run against significantly differentially expressed genes ordered by fold change from the 'IPF High' run with a resulting p-value < 0.001. The high enrichment of upregulated genes and downregulated genes from the 'IPF Diagnosis' run in the 'IPF High' run indicates that the two differential expression analyses are capturing similar expression patterns. **H.** A heatmap showing expression of differentially expressed genes in cells with high and low COL1A1 and COL1A2 expressions from the single-cell COVID-19 dataset. **I.** GSEA enrichment of significantly differentially expressed genes from the 'IPF High' run in significantly differentially expressed genes from the COVID-19 dataset with a resulting p-value < 0.001. The high ranking of the 'IPF High' genes indicates that similar genes co-express with COL1A1 and COL1A2 in both COIVID-19 and IPF when considering IPF and control samples. **J.** GSEA enrichment of differentially expressed genes from 'IPF Diagnosis' against differentially expressed genes from the 'COVID High' run ordered by fold change with a resulting p-value < 0.001. The high ranking of the 'IPF Diagnosis' genes indicates that the genes that co-express with COL1A1 and COL1A2 in COVID-19 have similar co-expression patterns in IPF. **K.** Top 10 enriched pathways from IPA pathway analysis for the differentially expressed genes from the COVID-19 single-cell RNA-seq dataset.

Figure 1

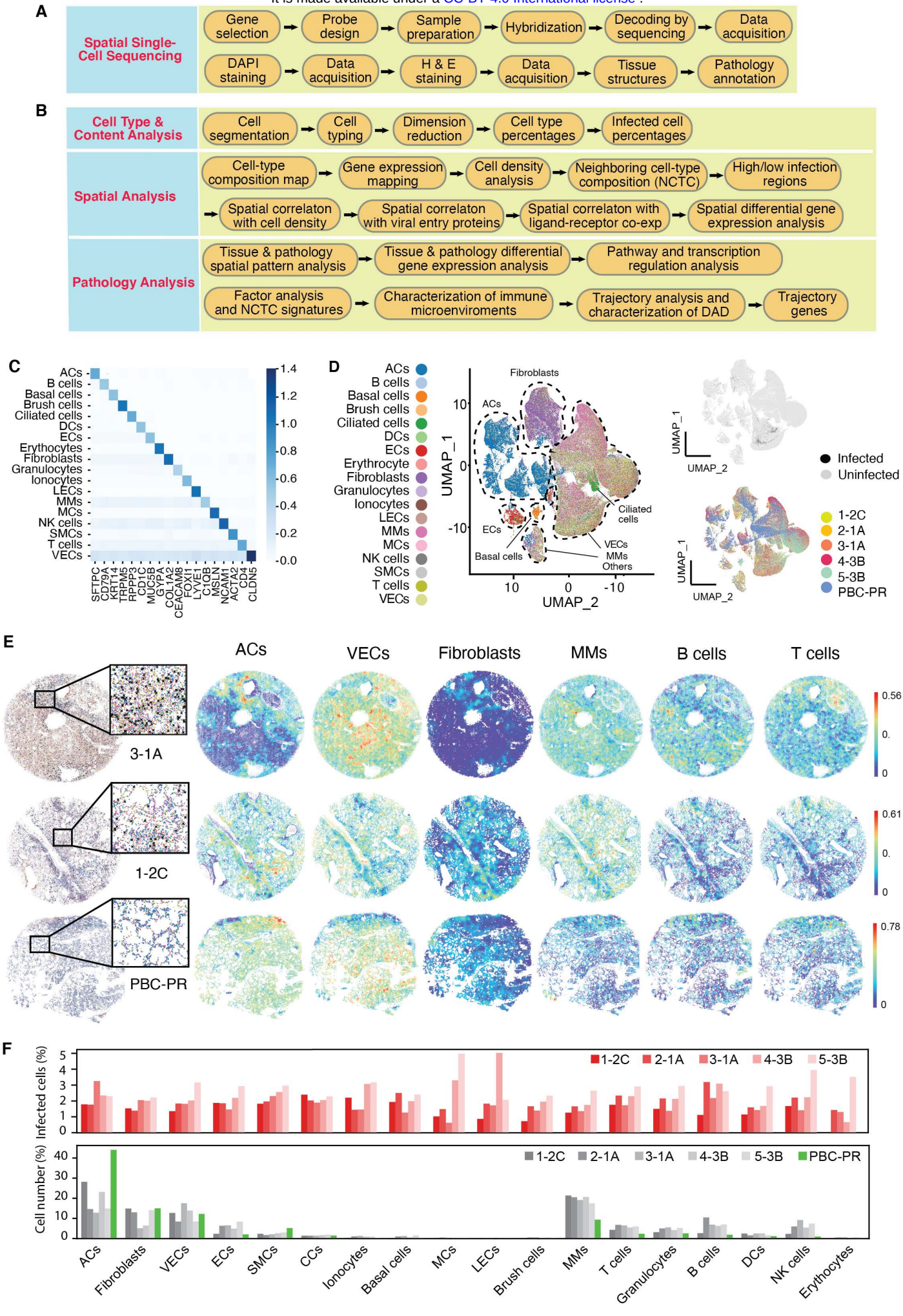


Figure 2

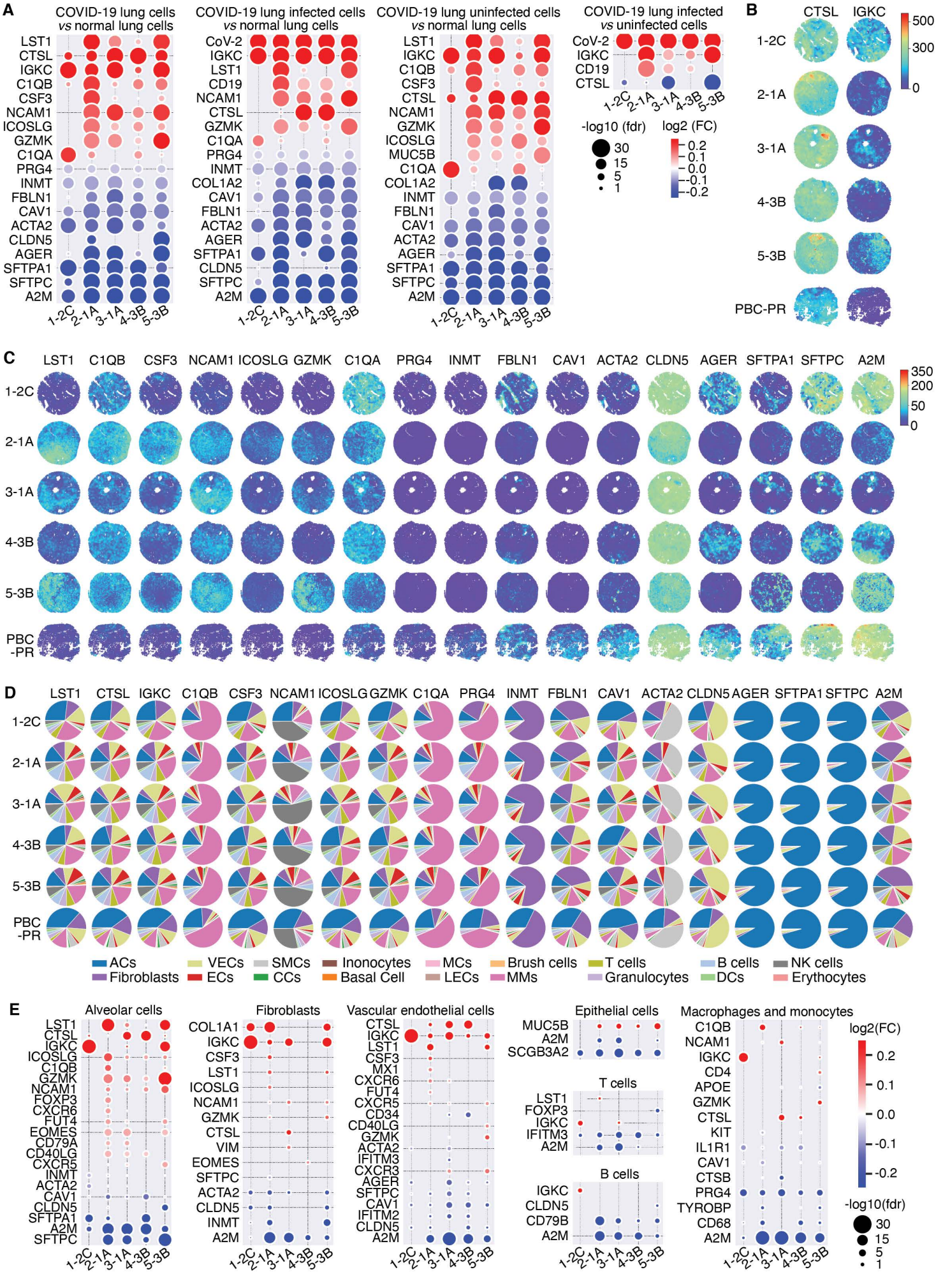


Figure 3

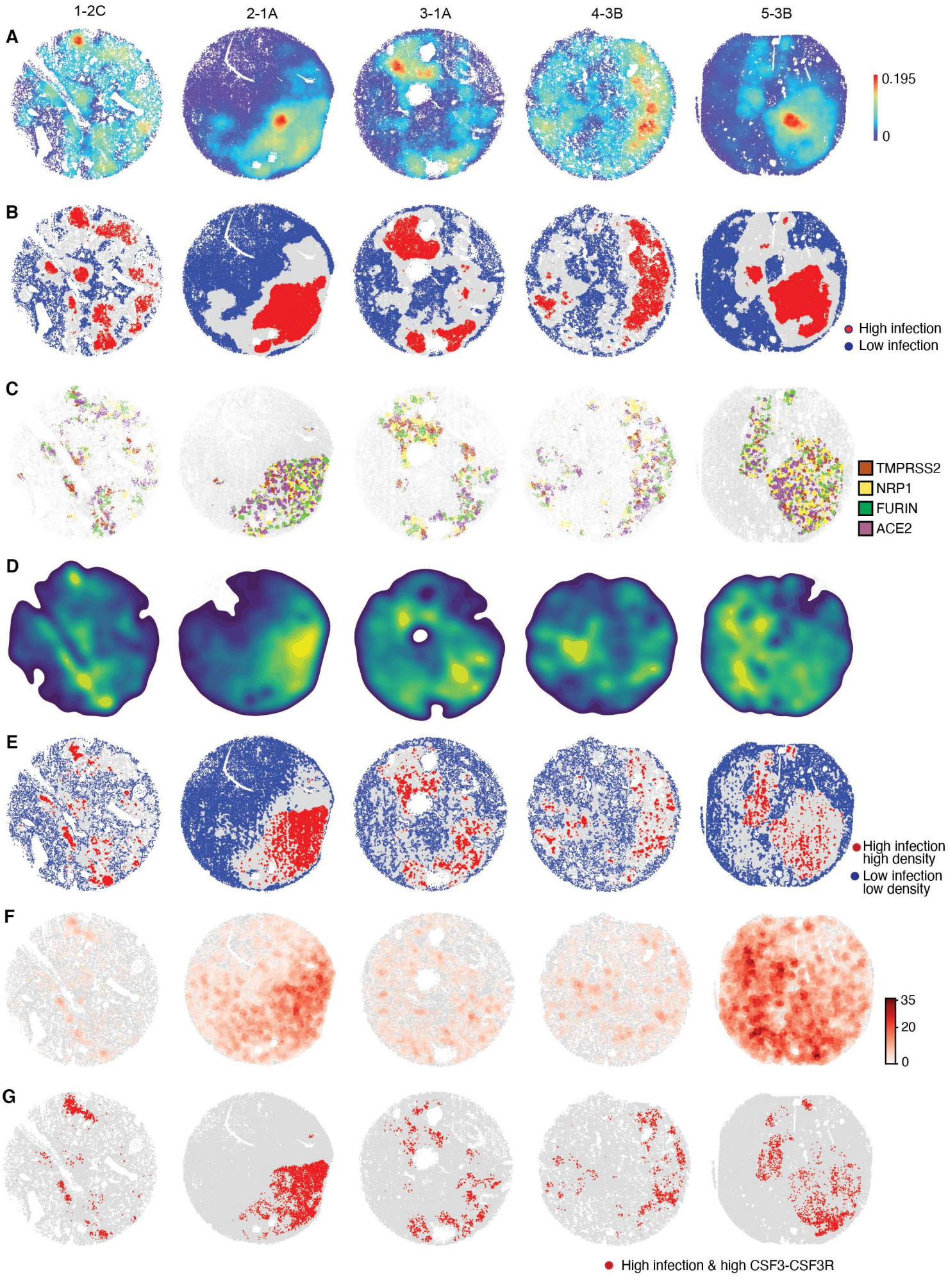


Figure 4

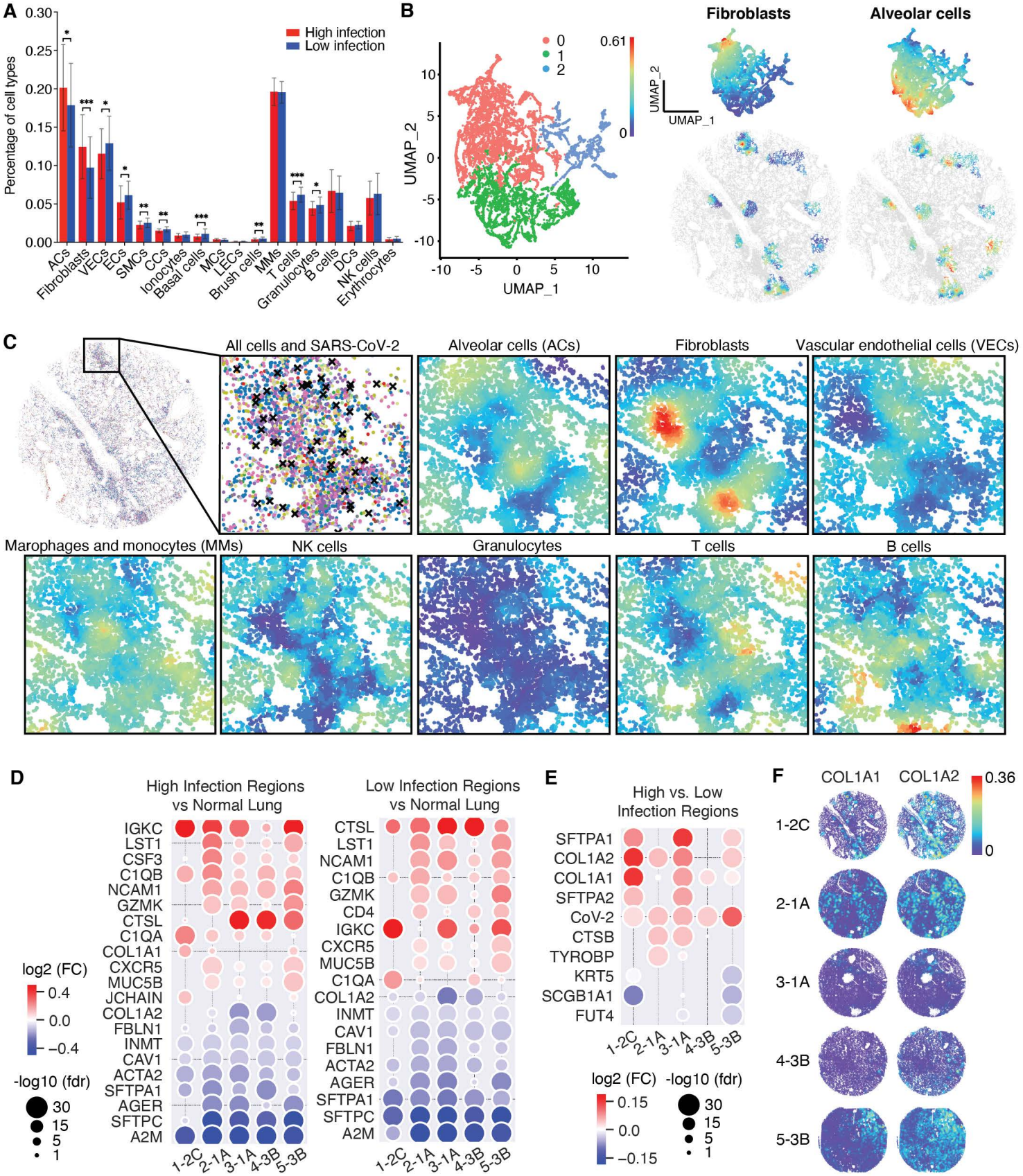


Figure 5

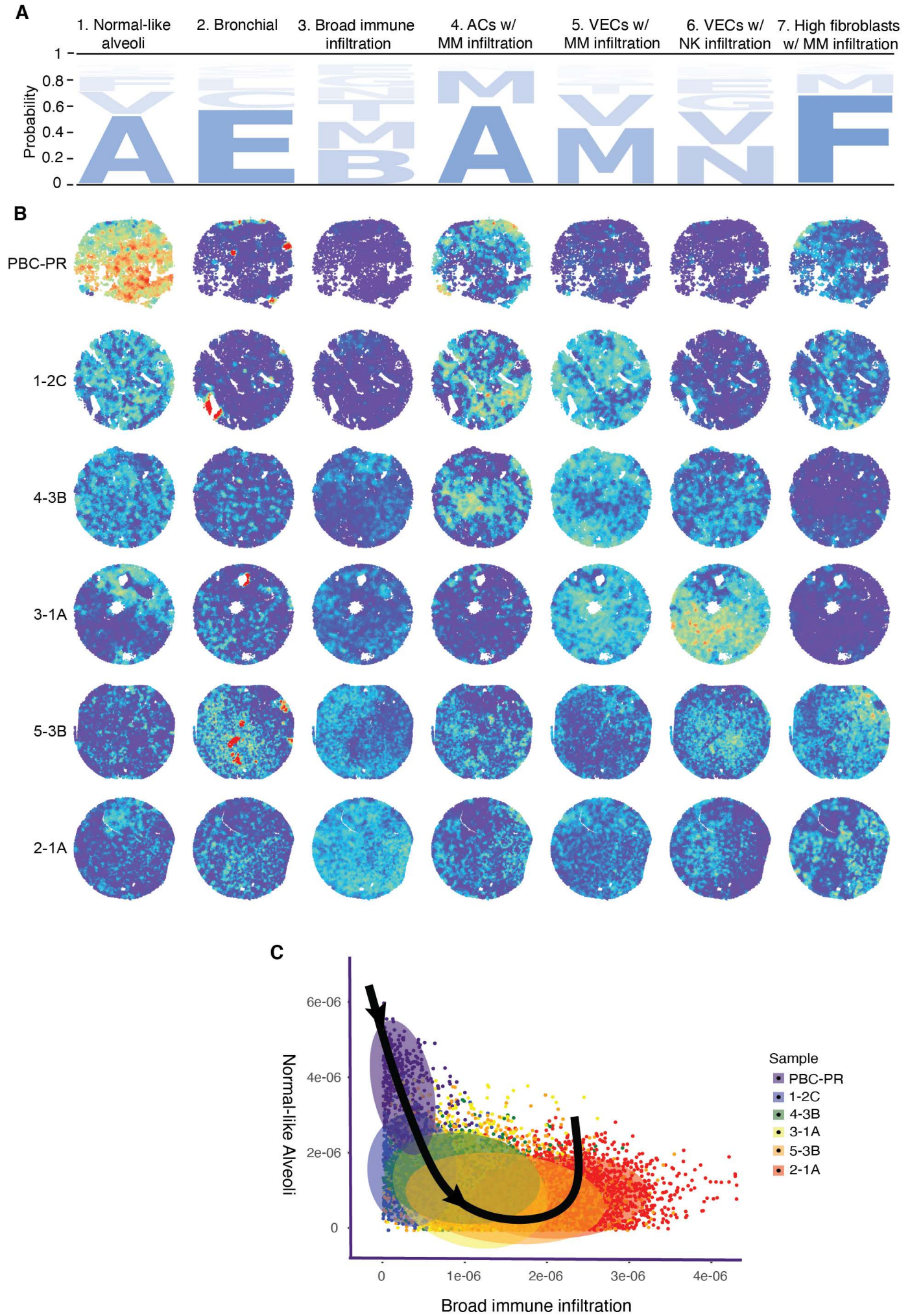


Figure 6

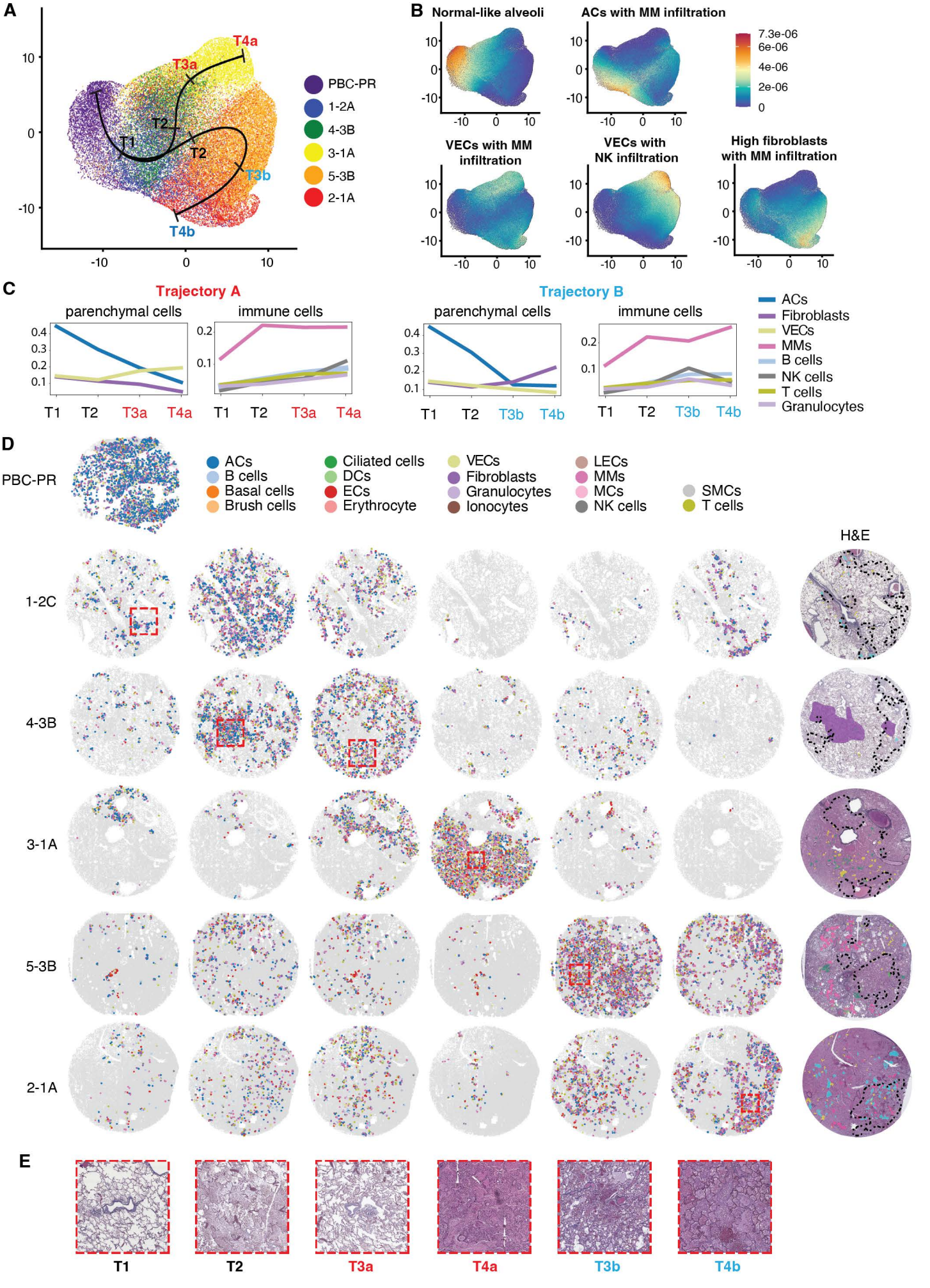


Figure 7

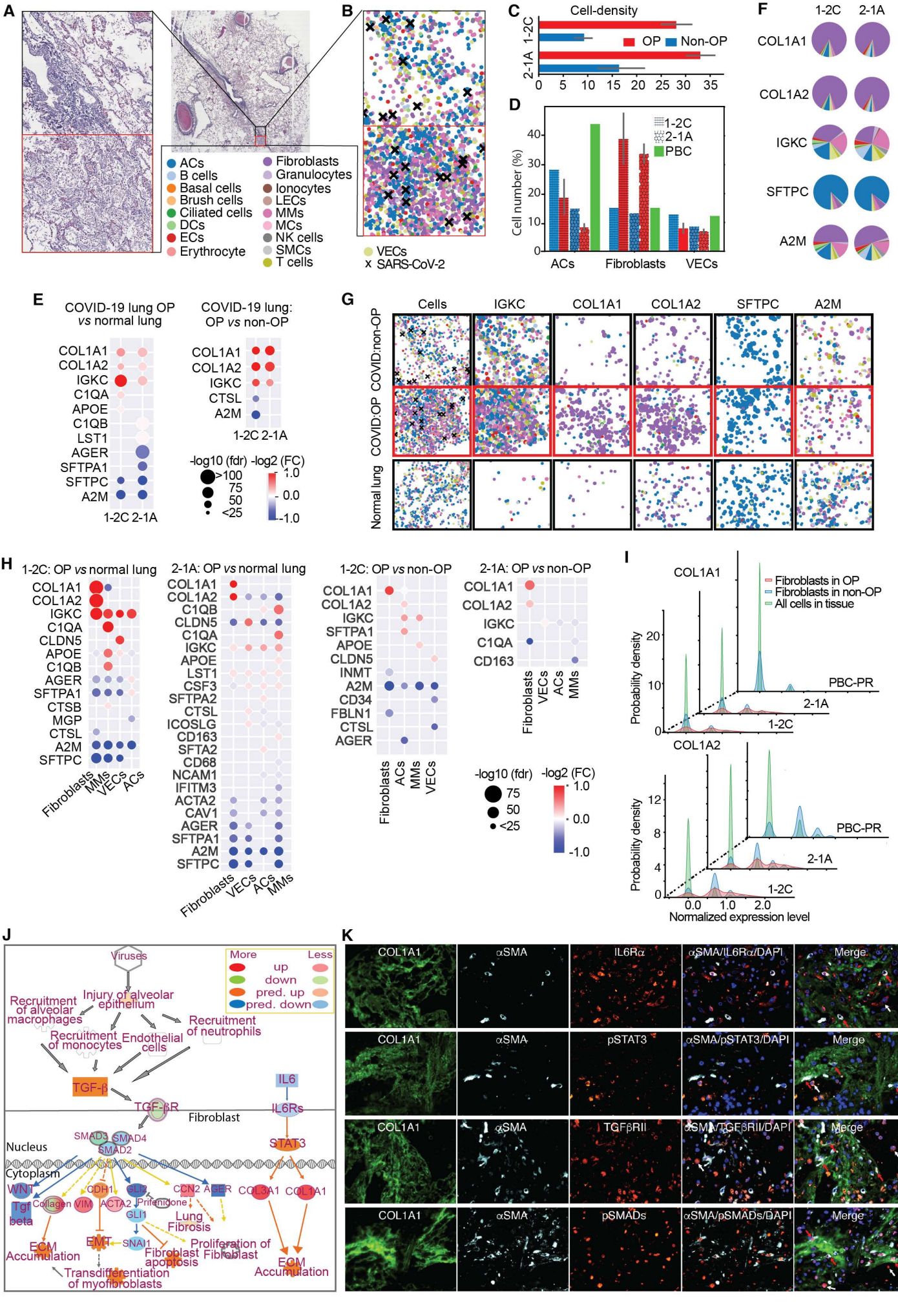


Figure S1

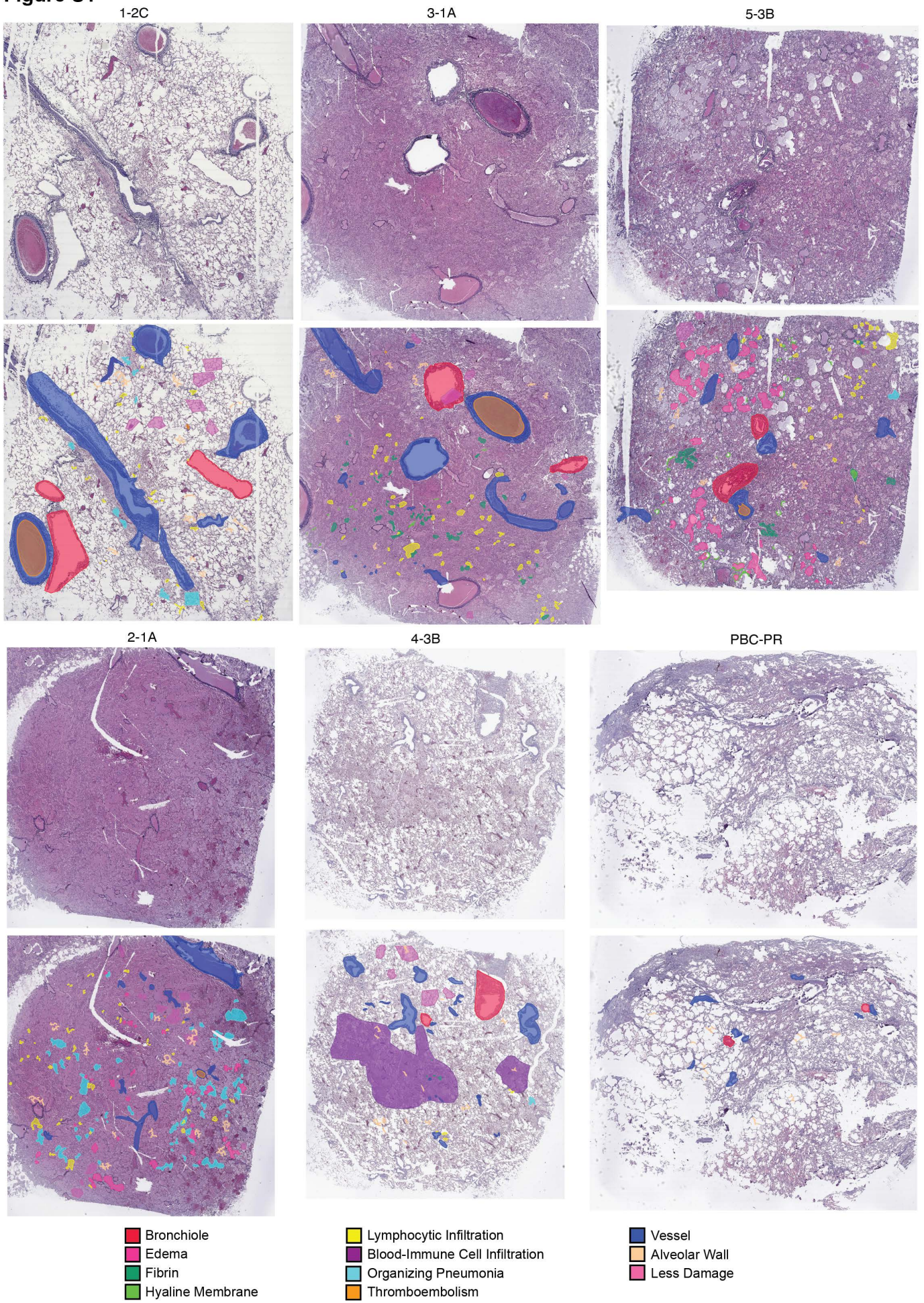
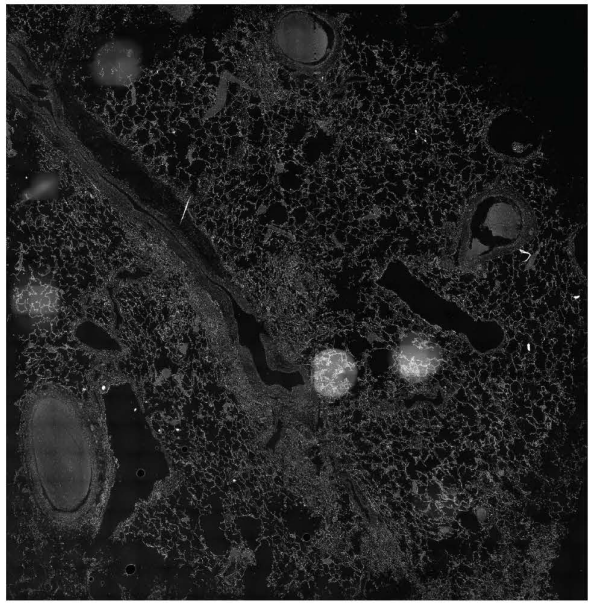
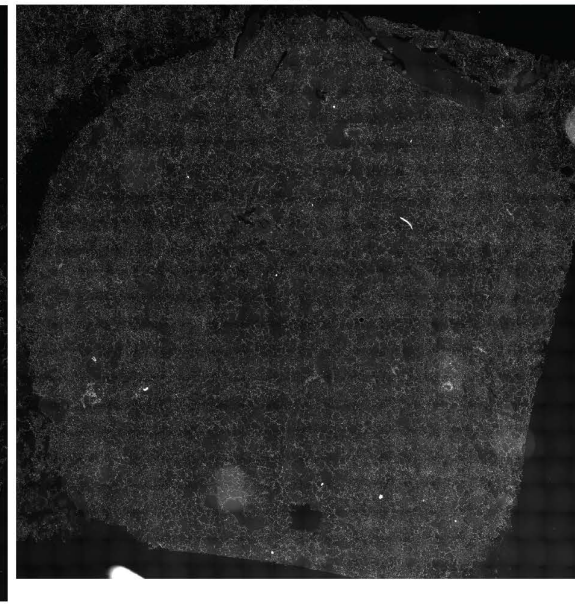


Figure S2

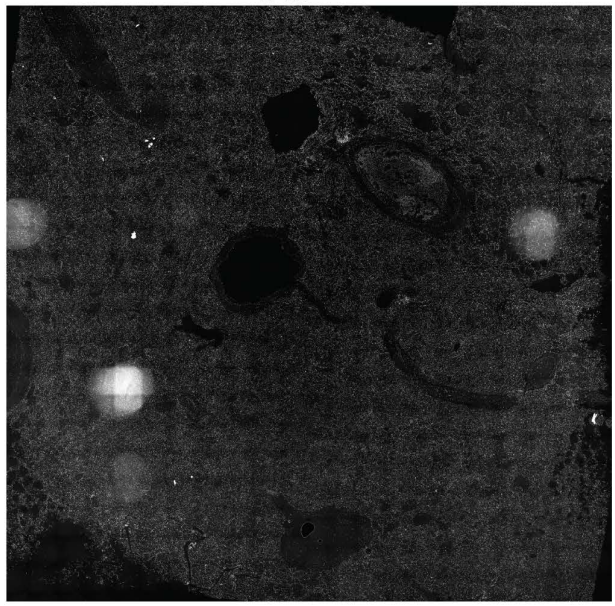
1-2C



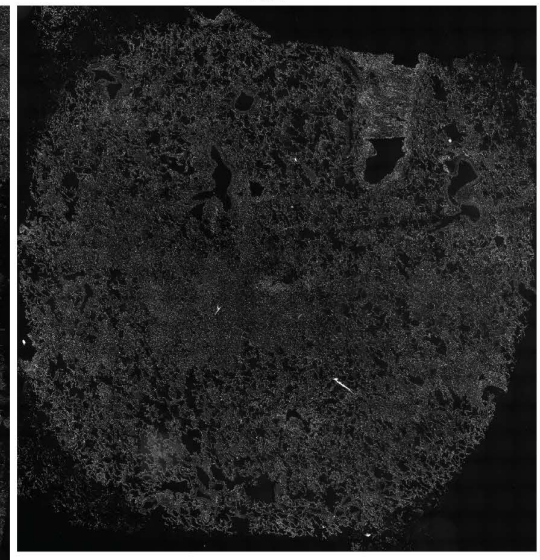
2-1A



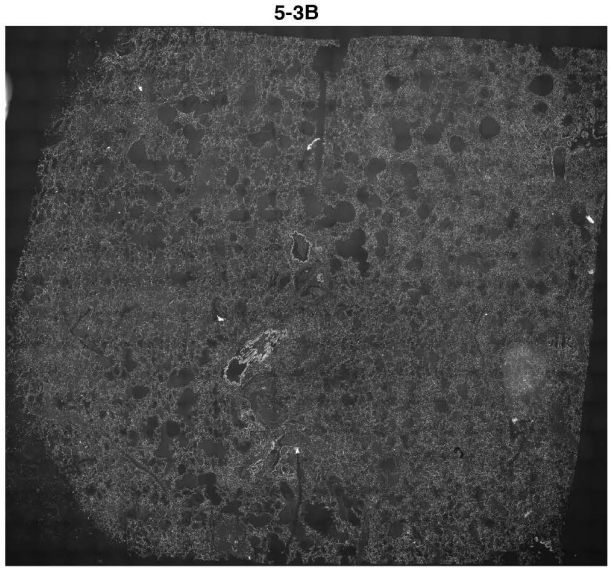
3-1A



4-3B



5-3B



PBC-PR

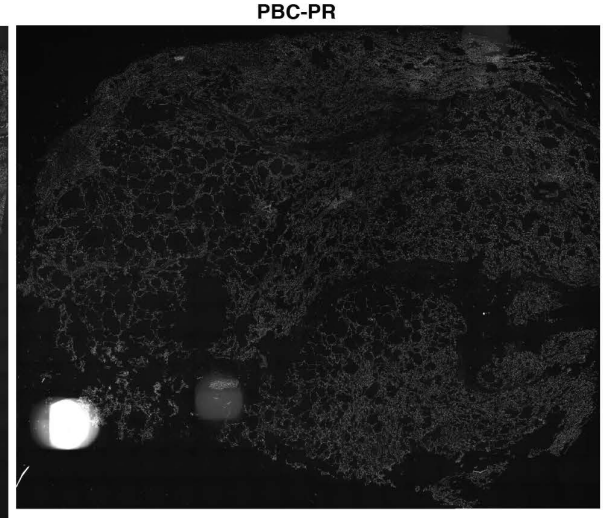


Figure S3

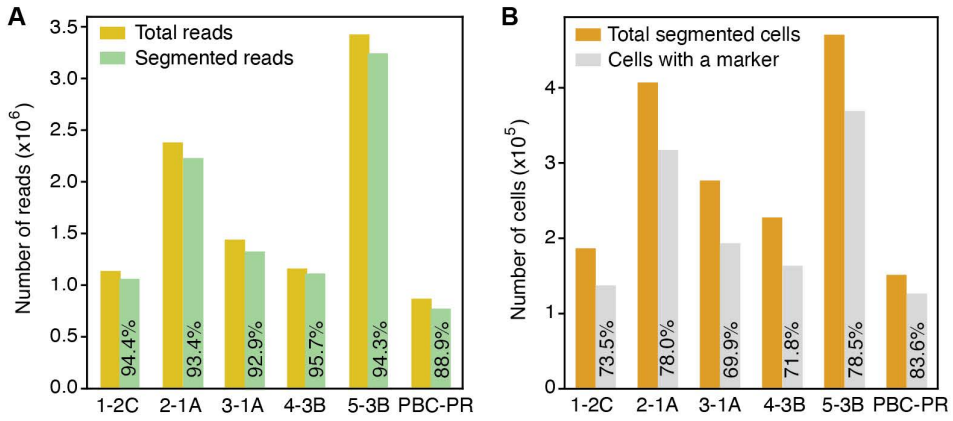


Figure S4

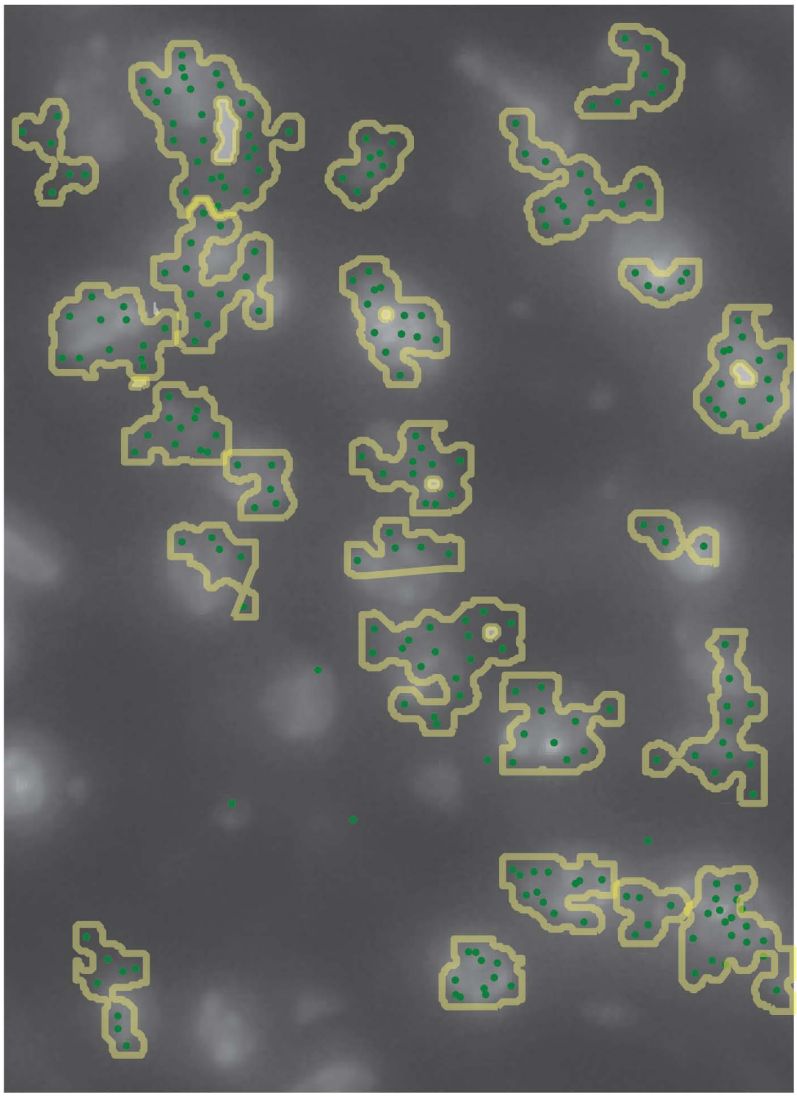


Figure S5

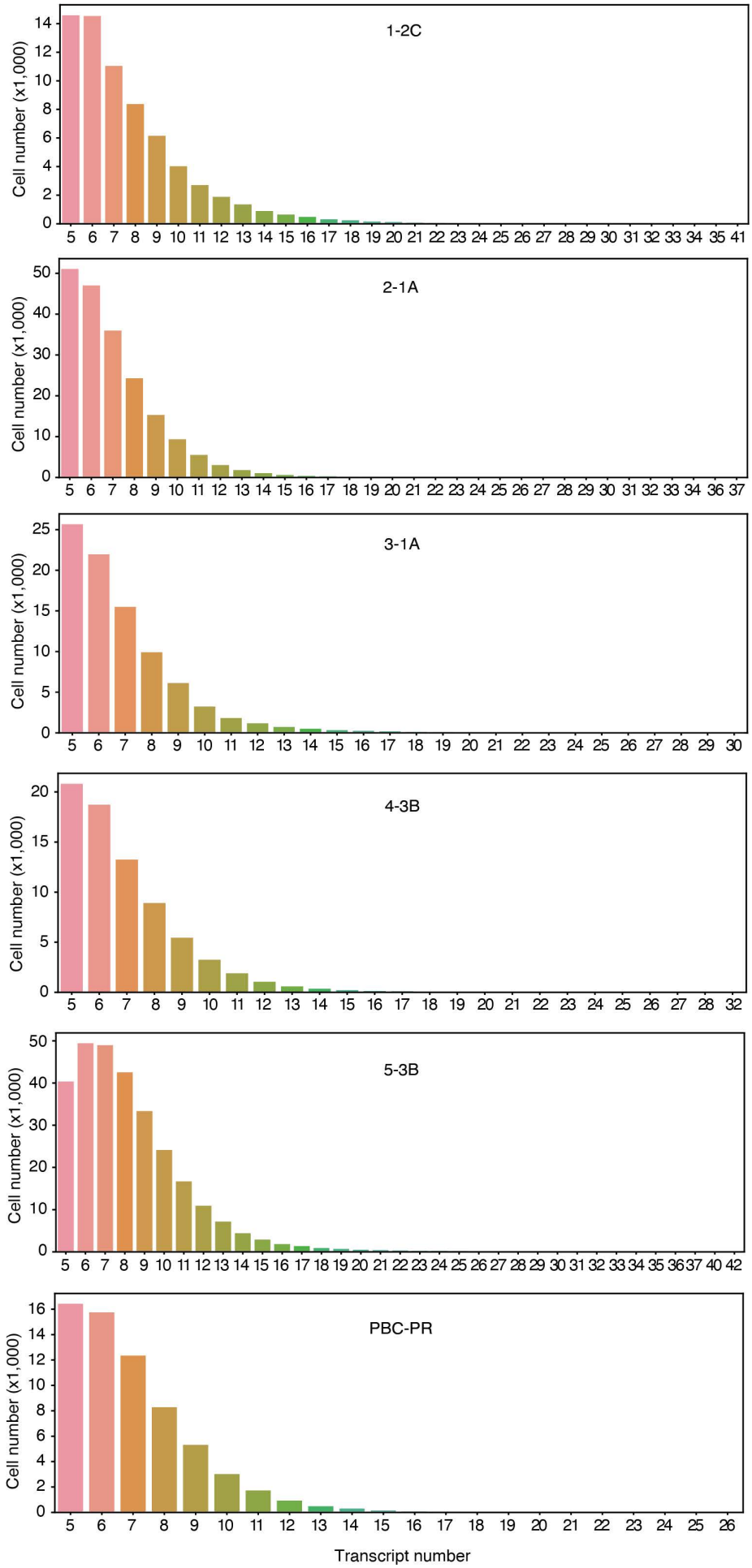


Figure S6

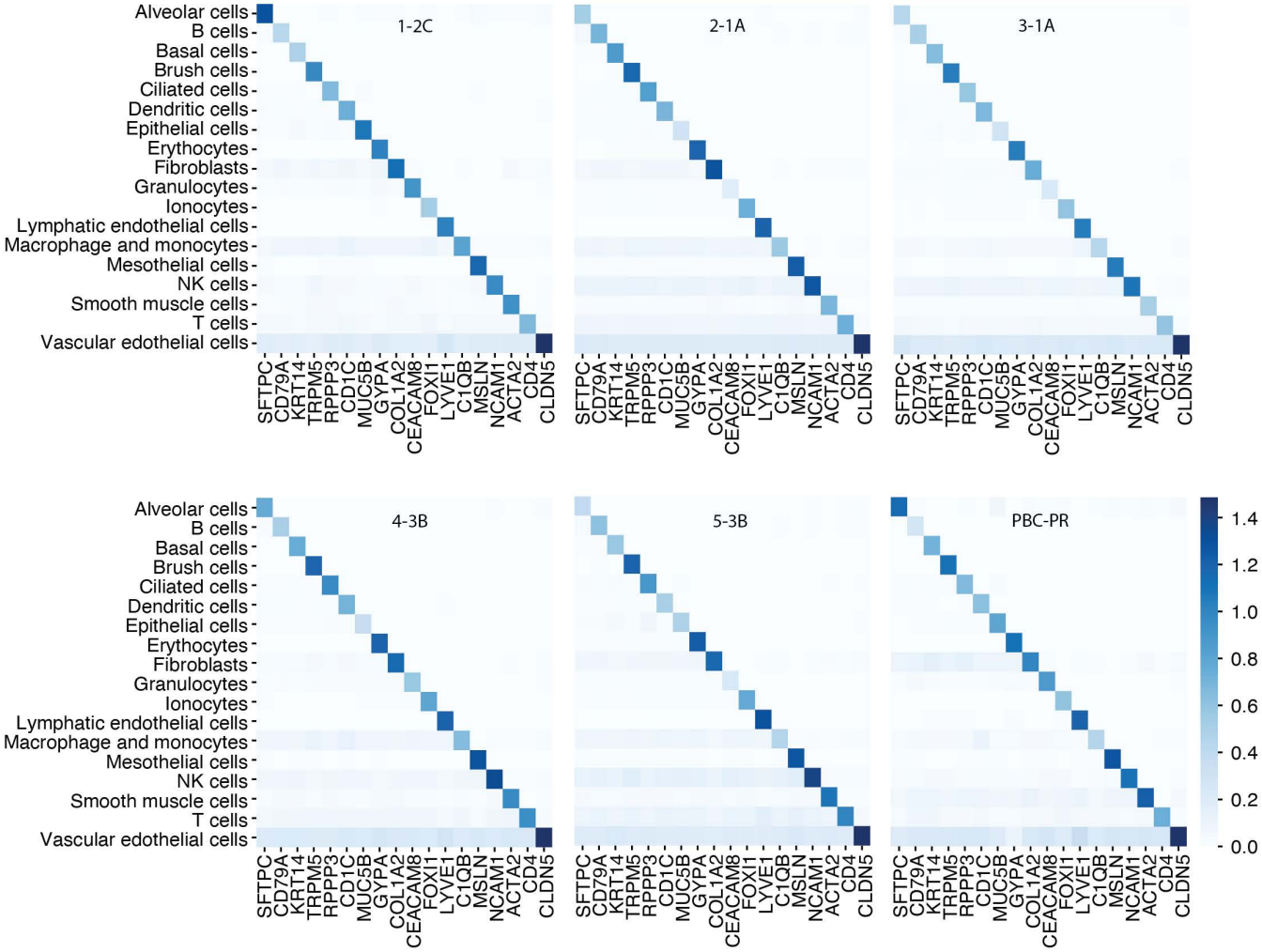


Figure S7

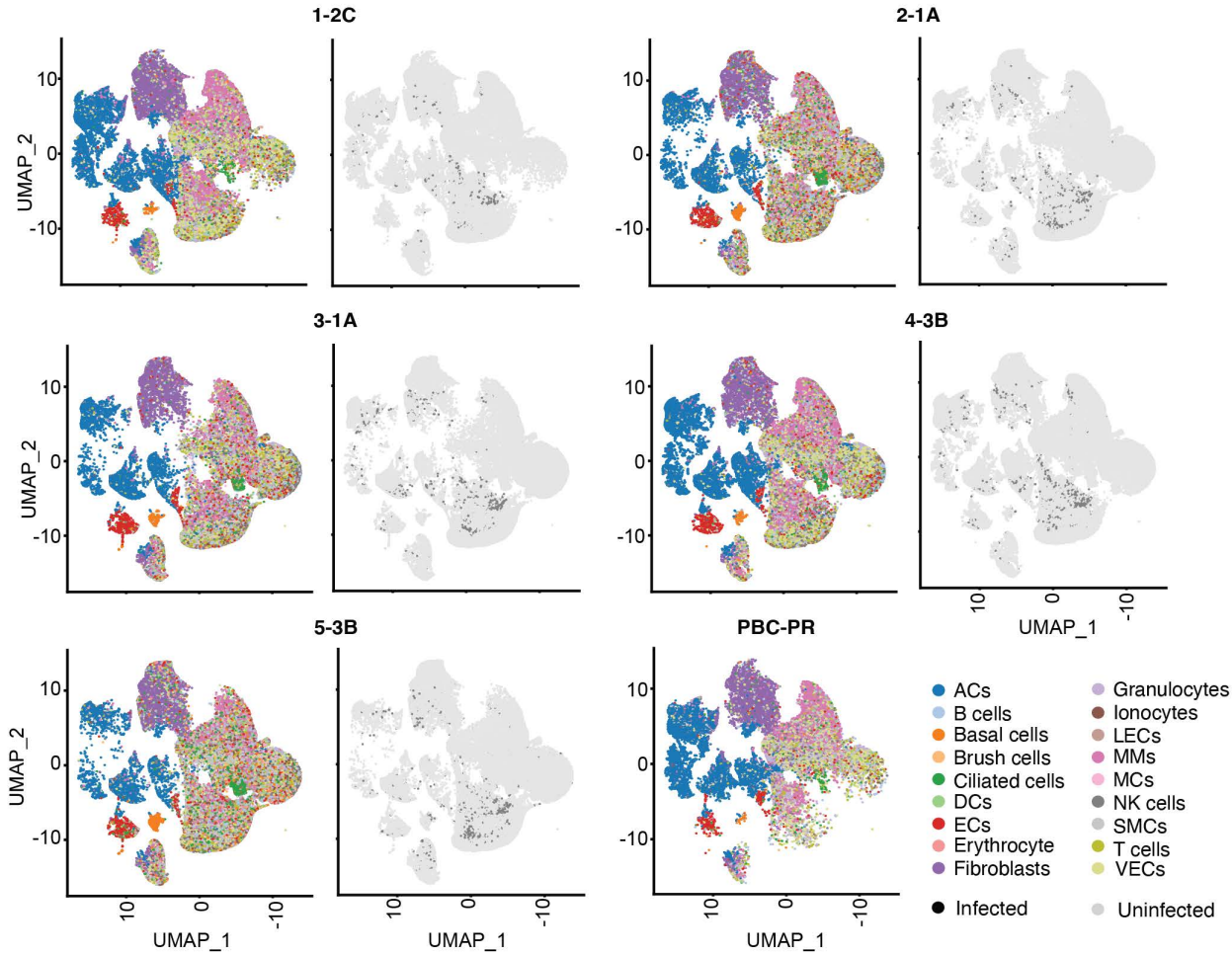


Figure S8

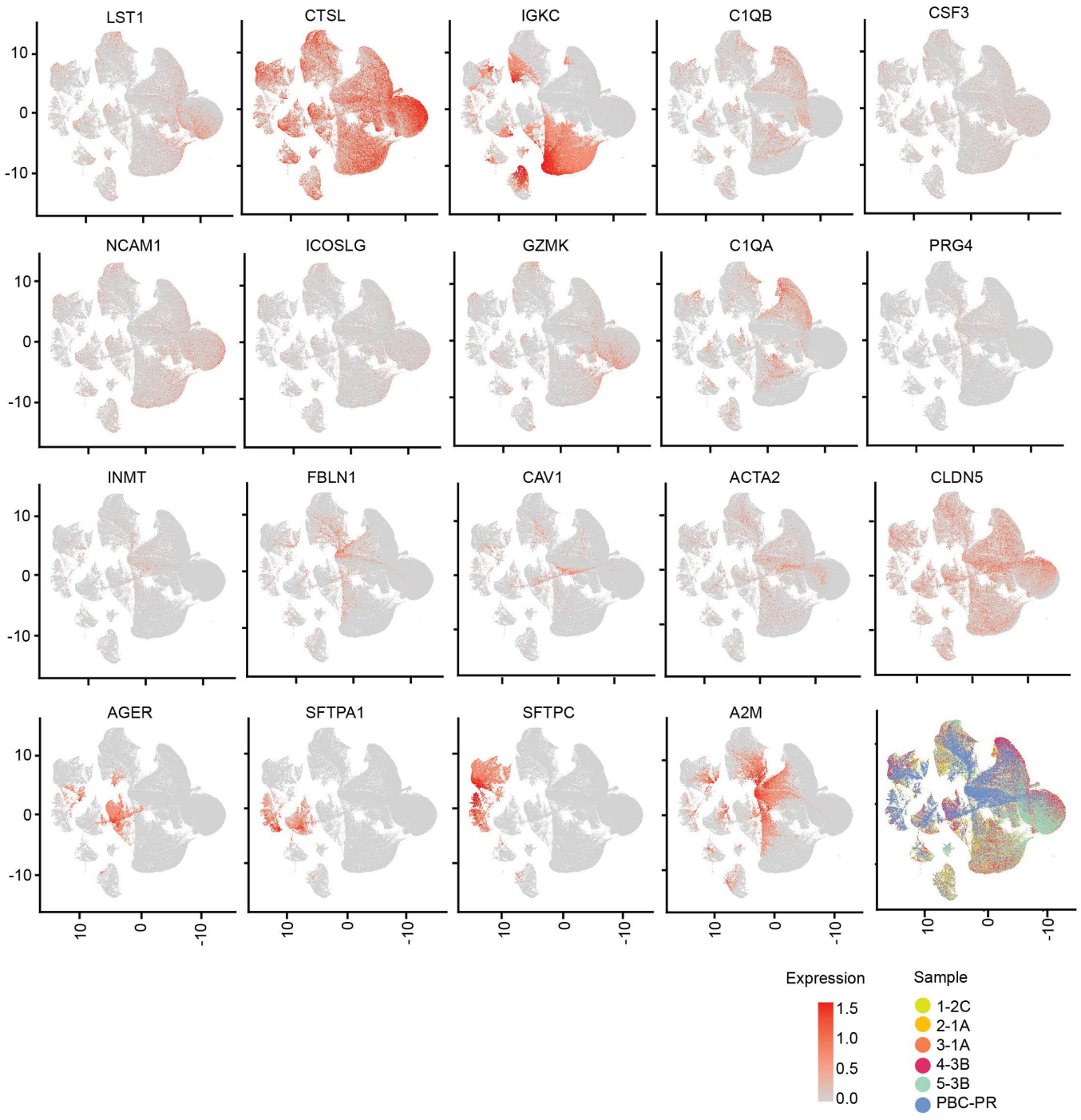


Figure S9

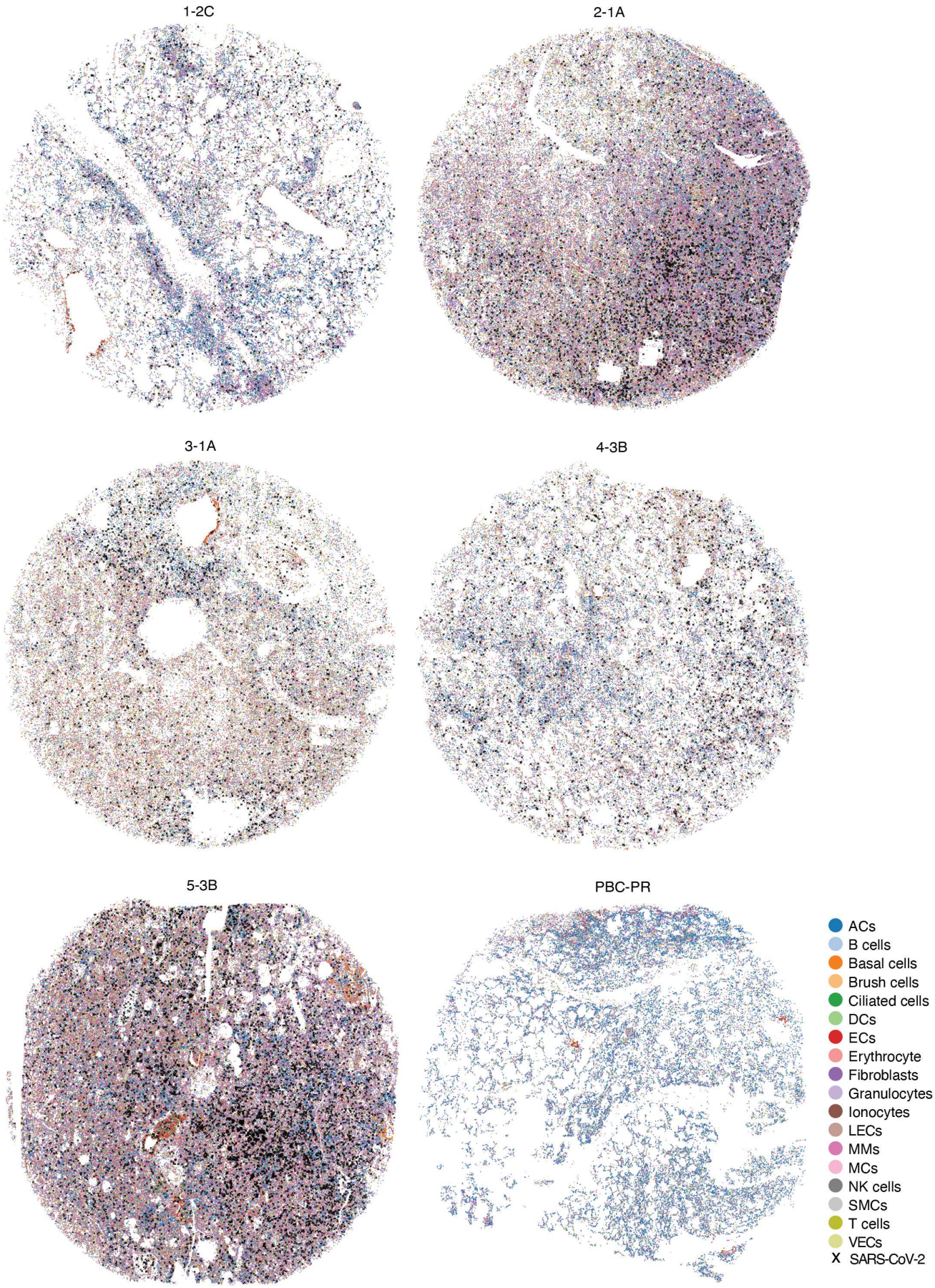


Figure S10 - Sample 1-2C Spatial Map of Cell Types

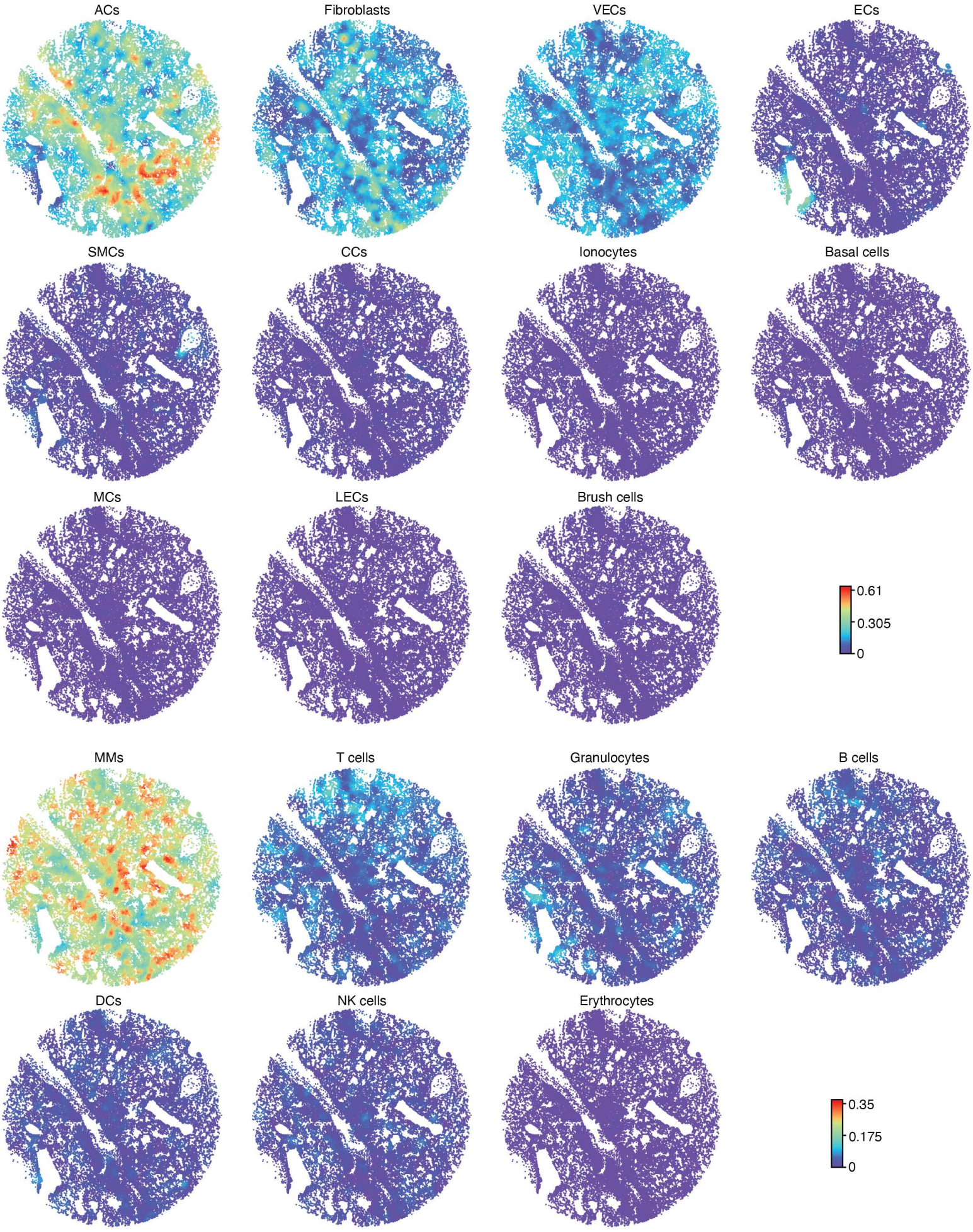


Figure S11 - Sample 2-1A Spatial Map of Cell Types

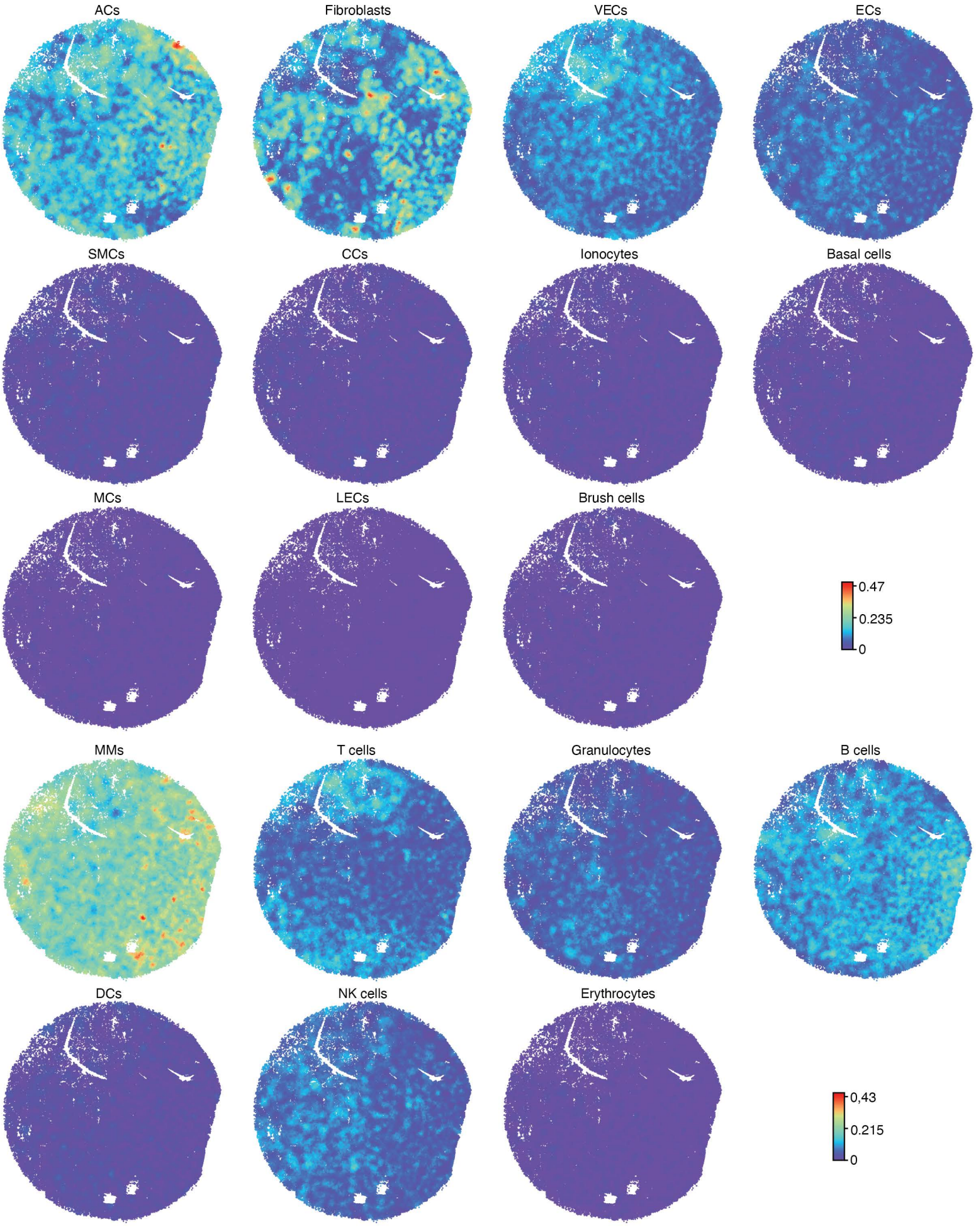


Figure S12 - Sample 3-1A Spatial Map of Cell Types

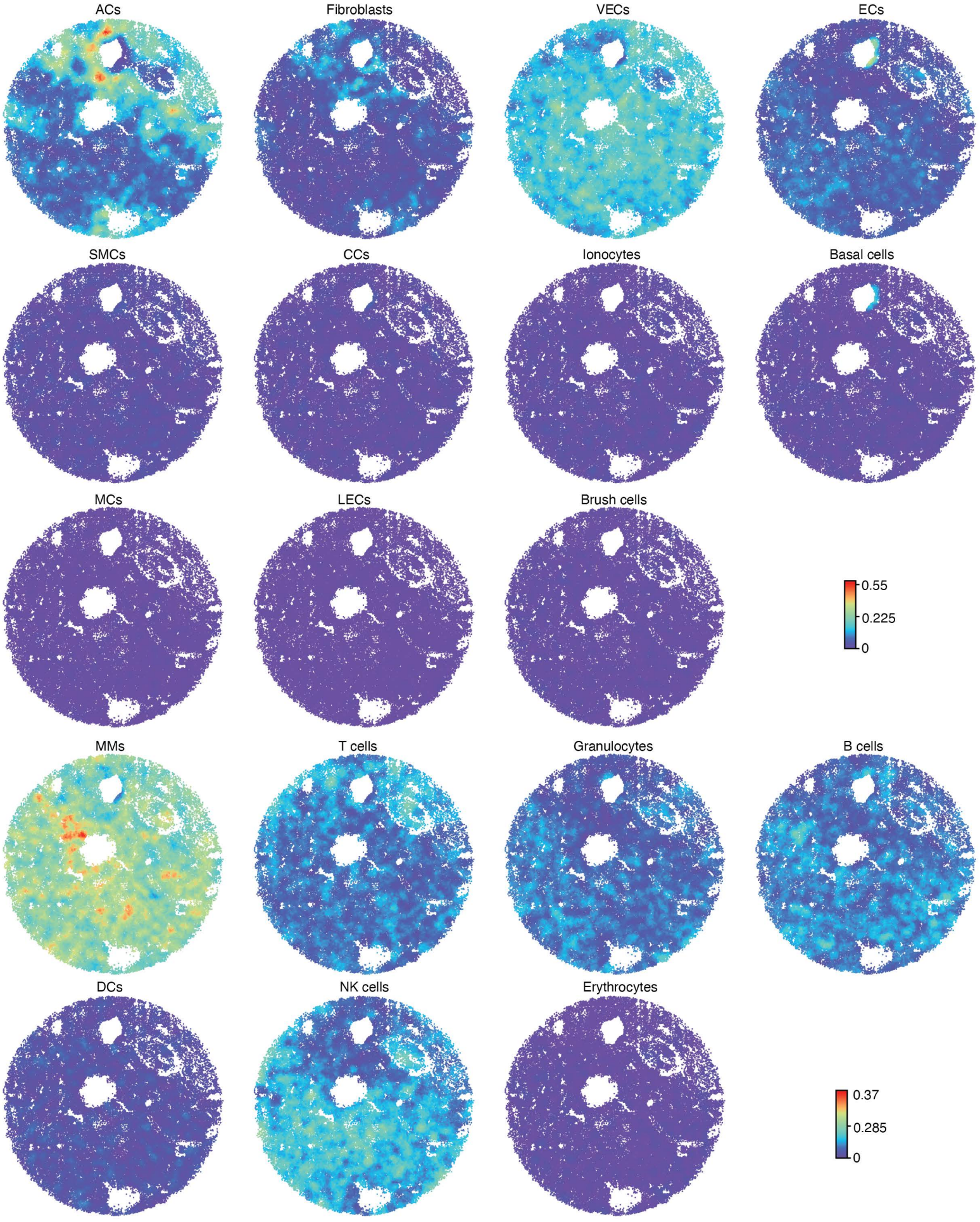


Figure S13 - Sample 4-3B Spatial Map of Cell Types

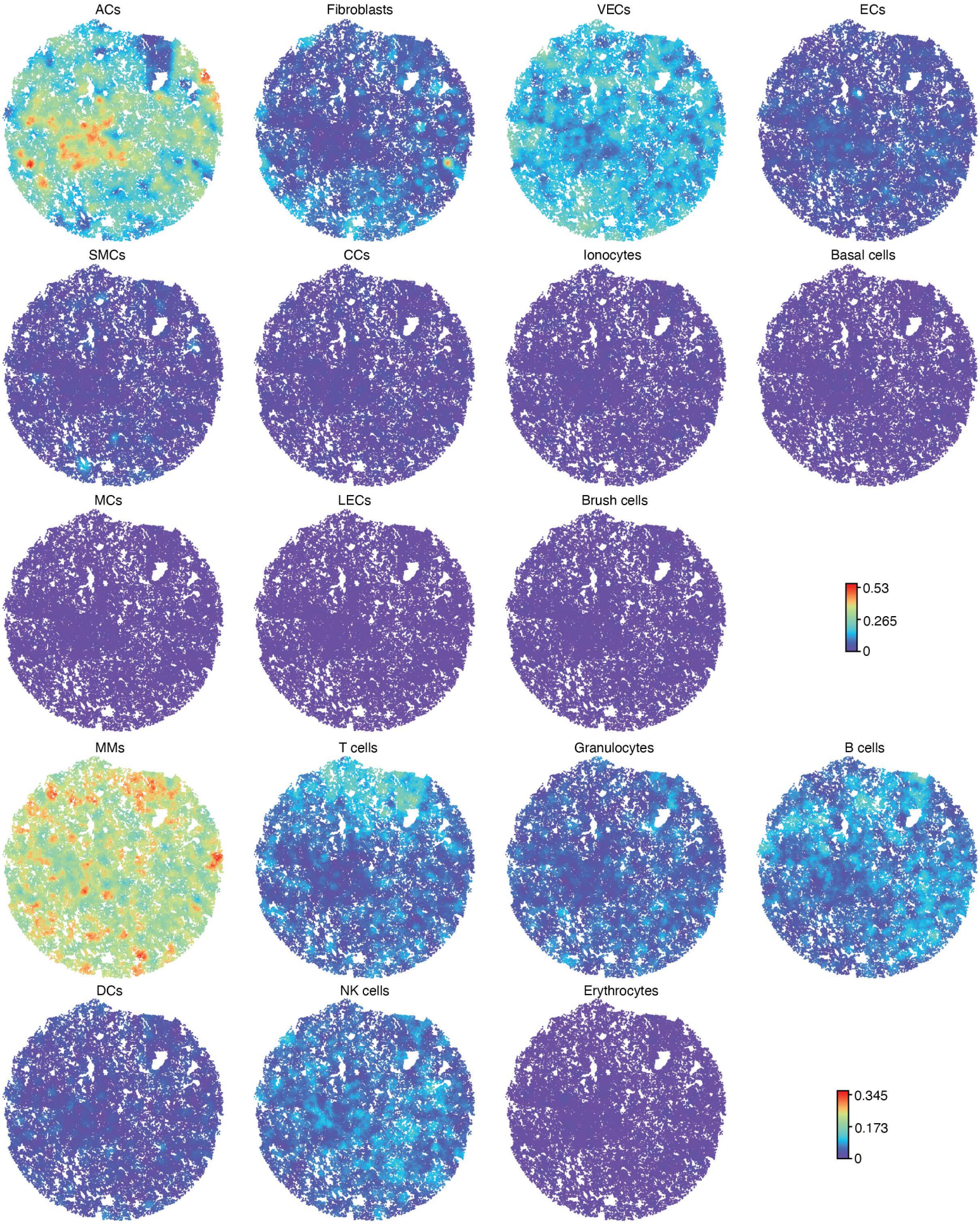


Figure S14 - Sample 5-3B Spatial Map of Cell Types

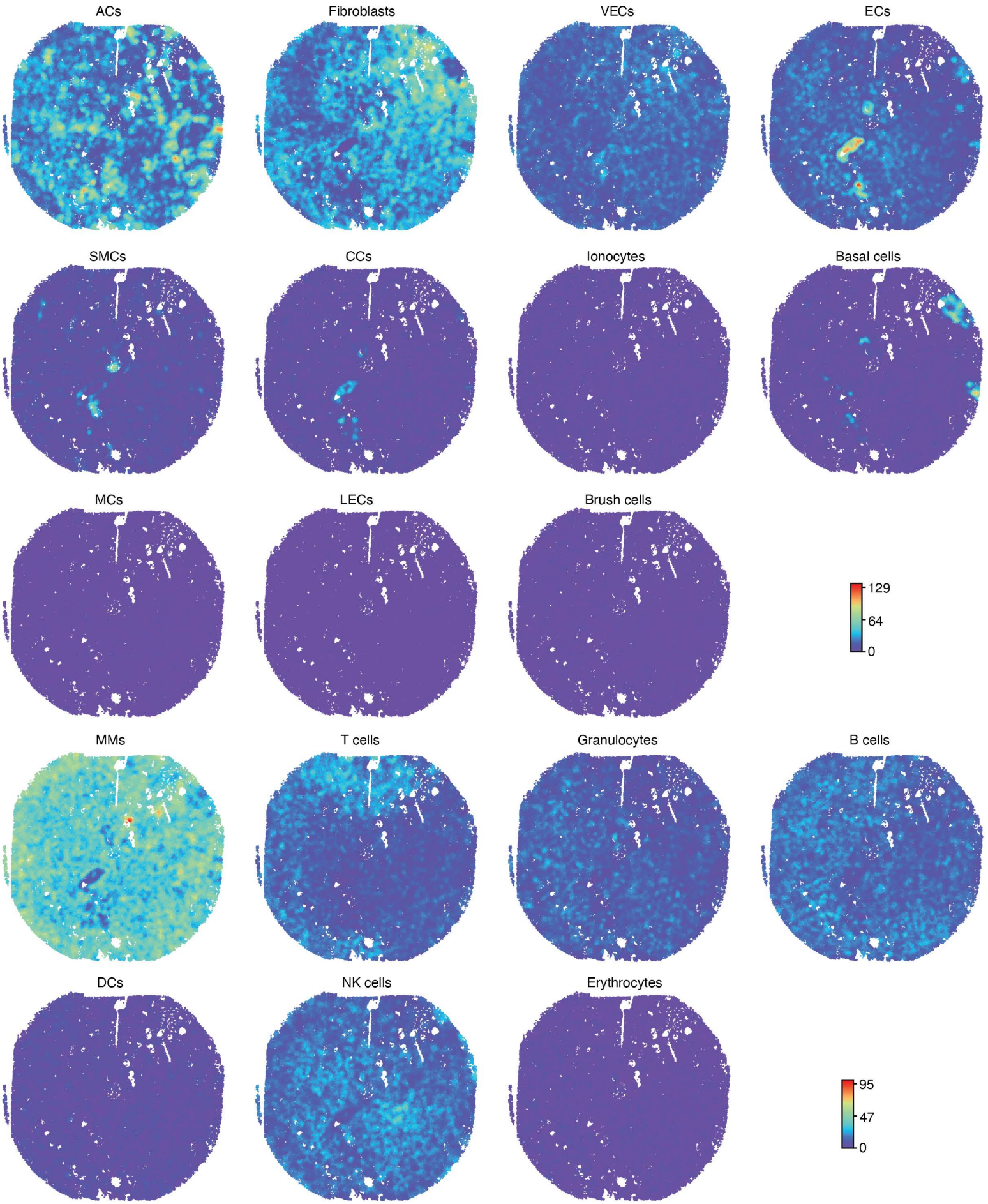


Figure S15 - Sample PBC-PR Spatial Map of Cell Types

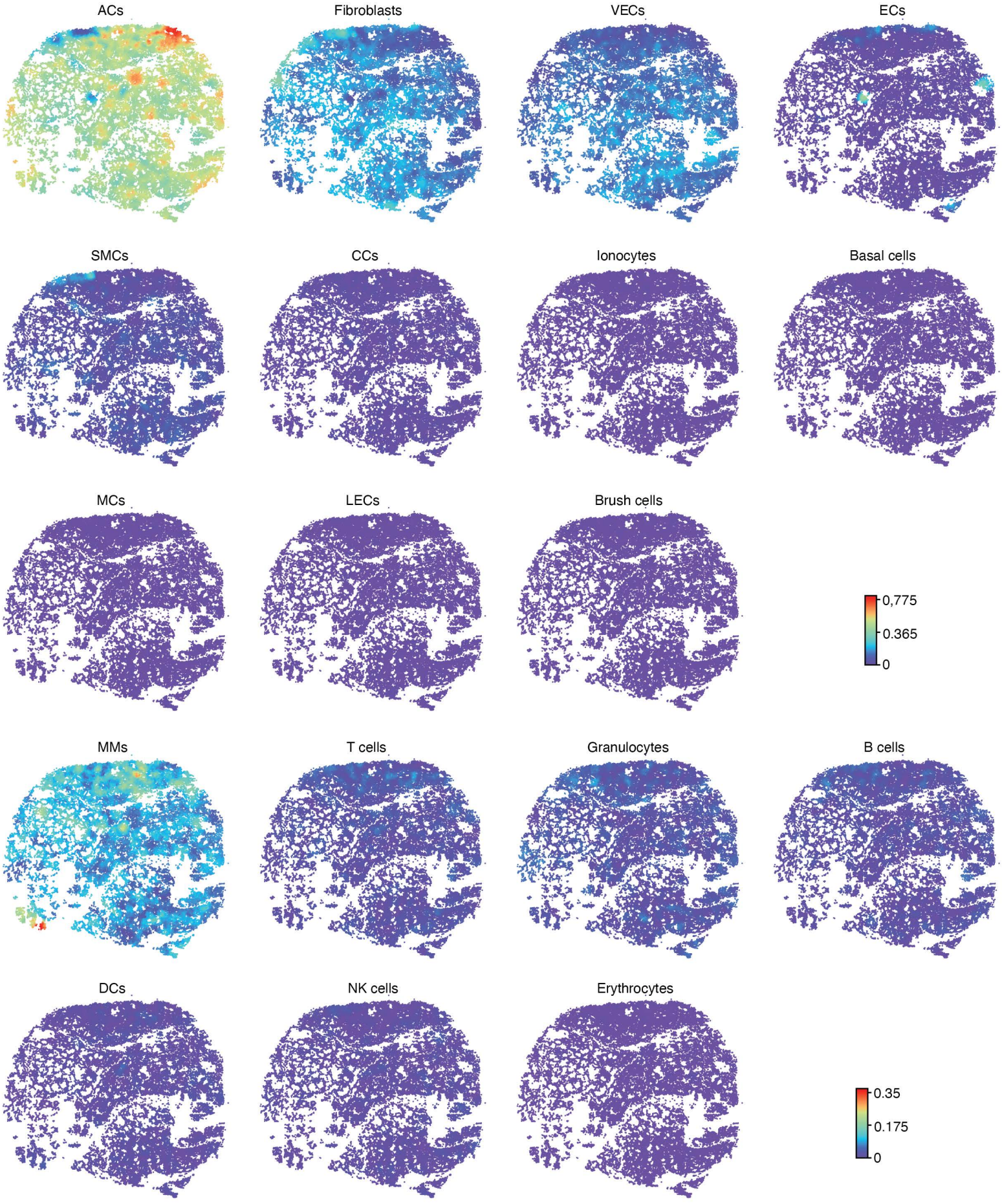


Figure S16

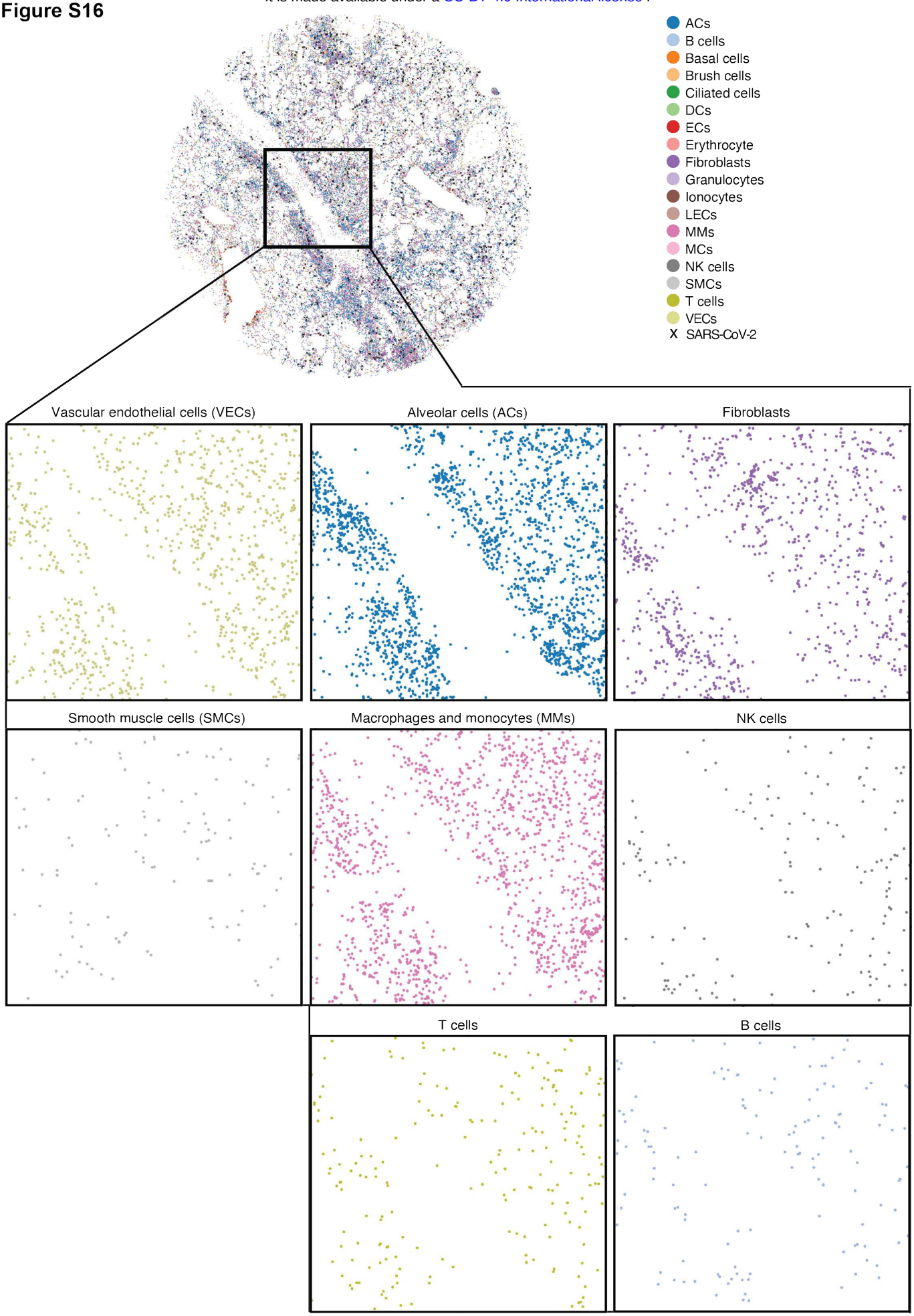


Figure S17

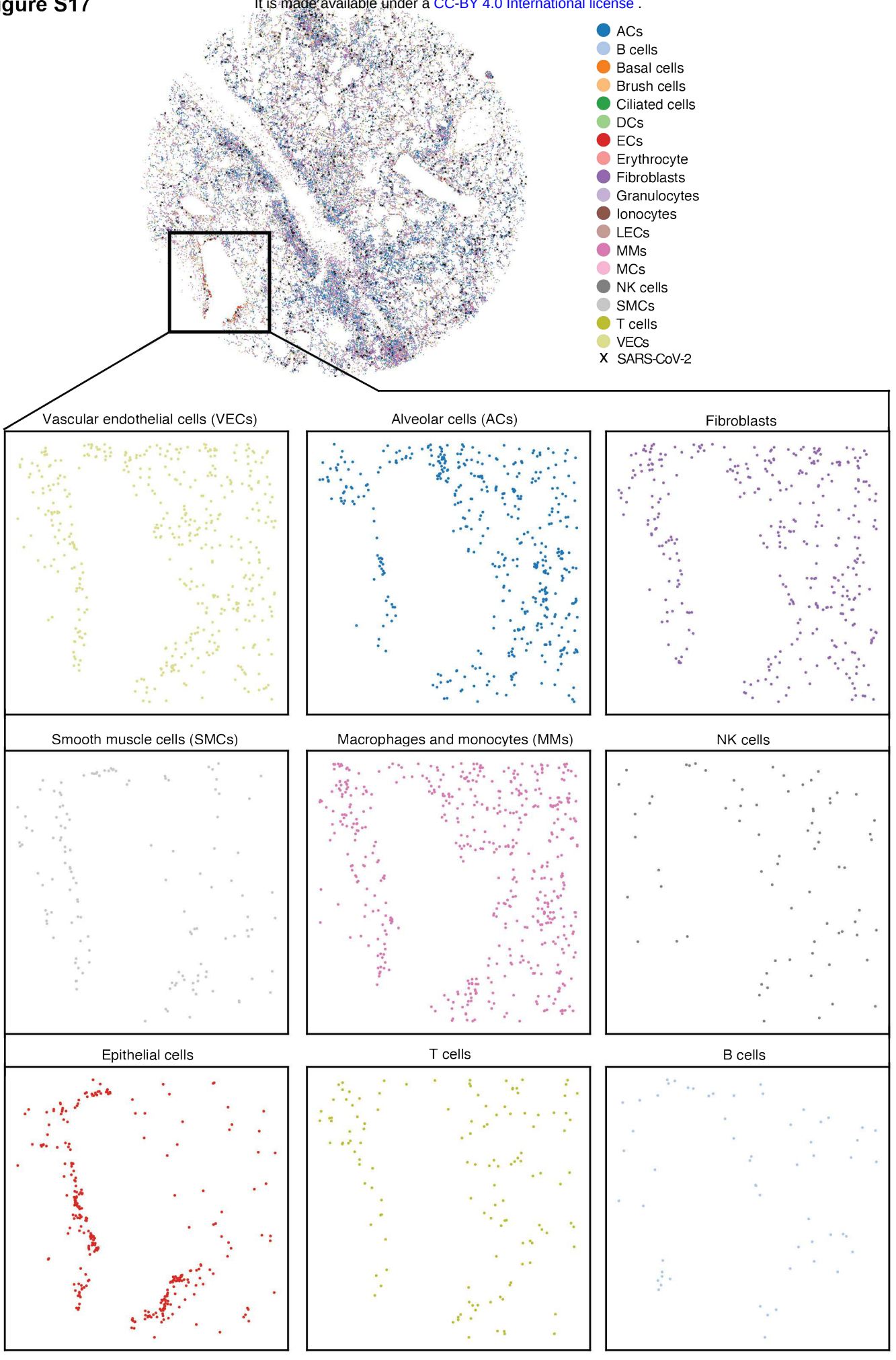


Figure S18: 1-2C

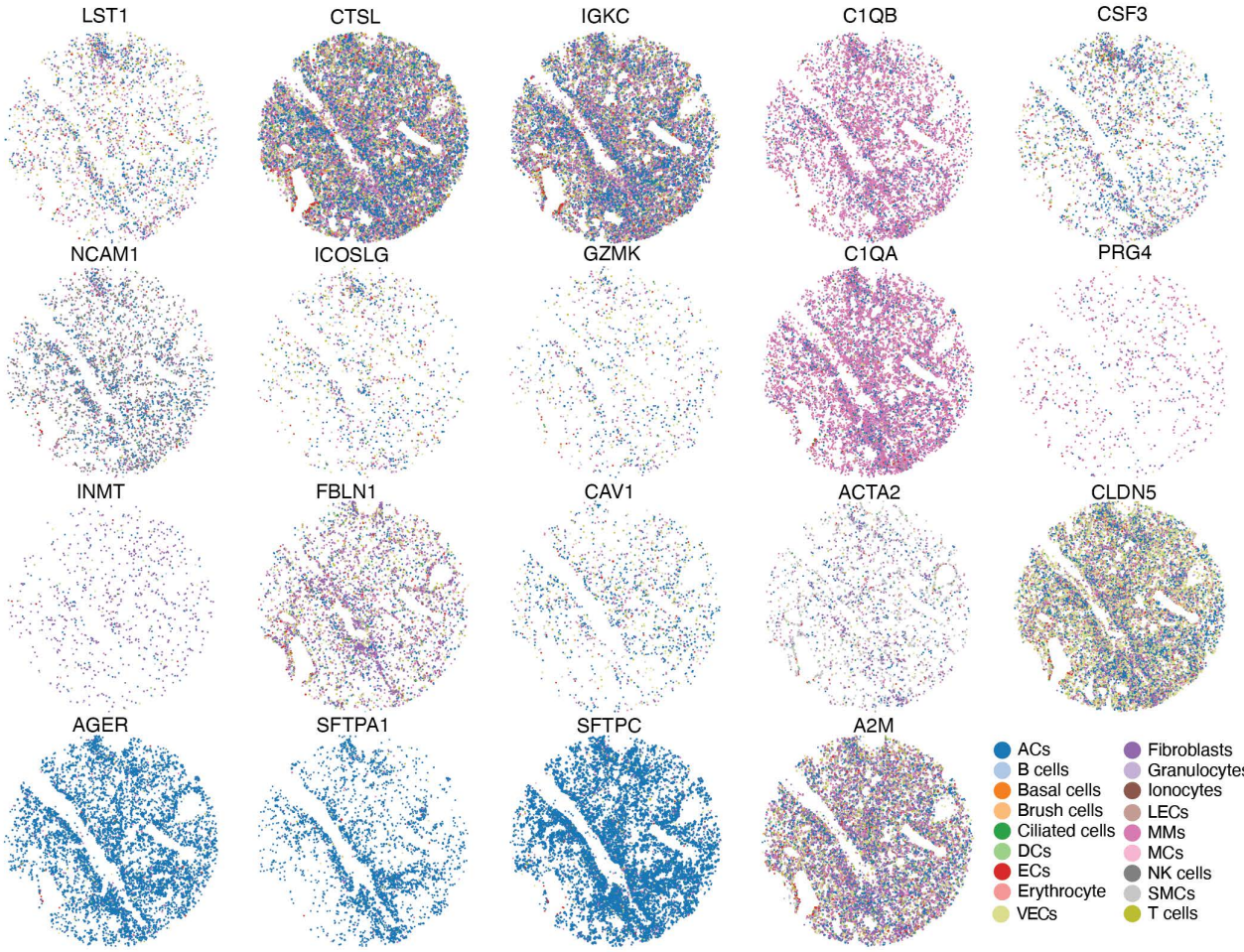


Figure S19: 2-1A

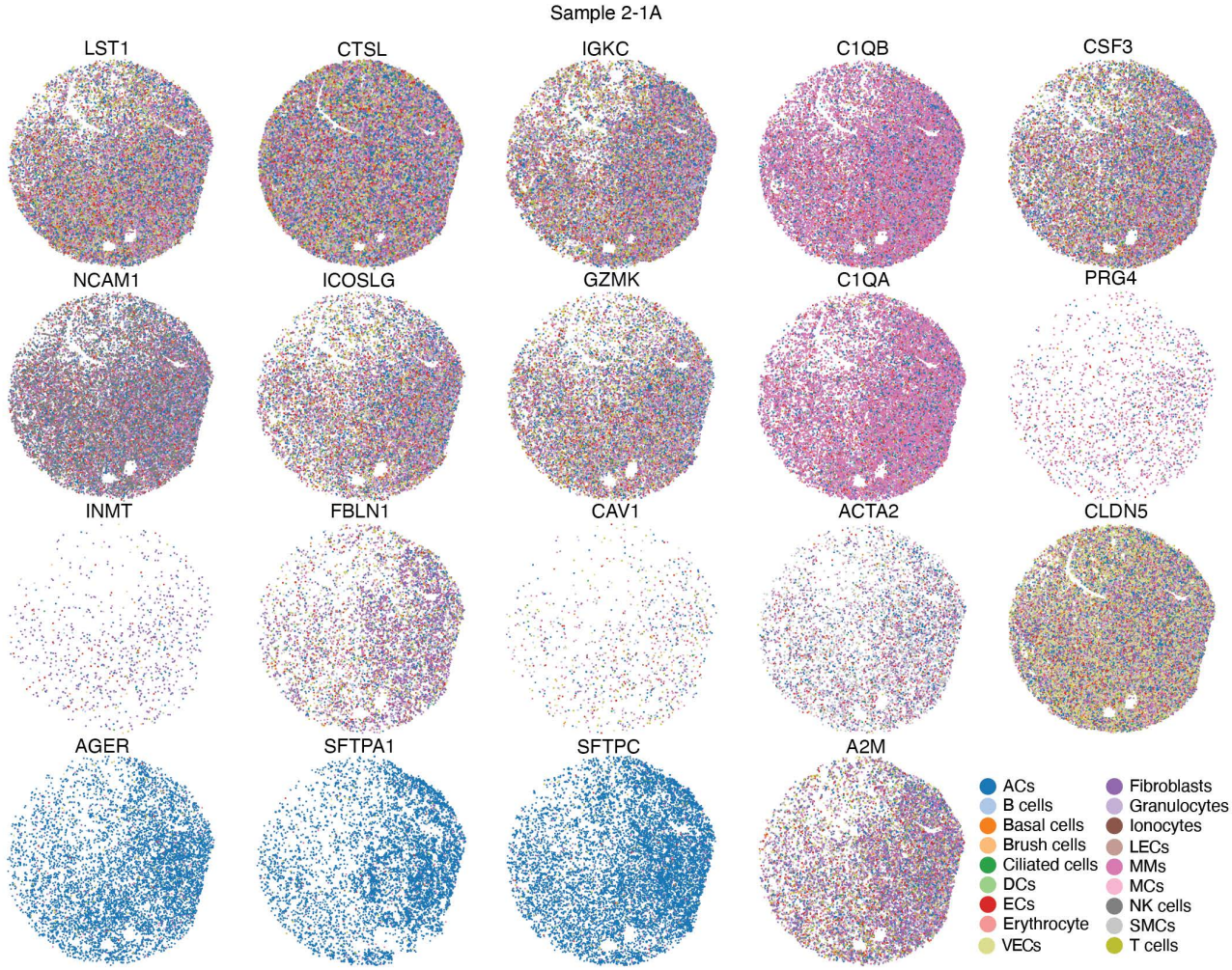


Figure S20: 3-1A

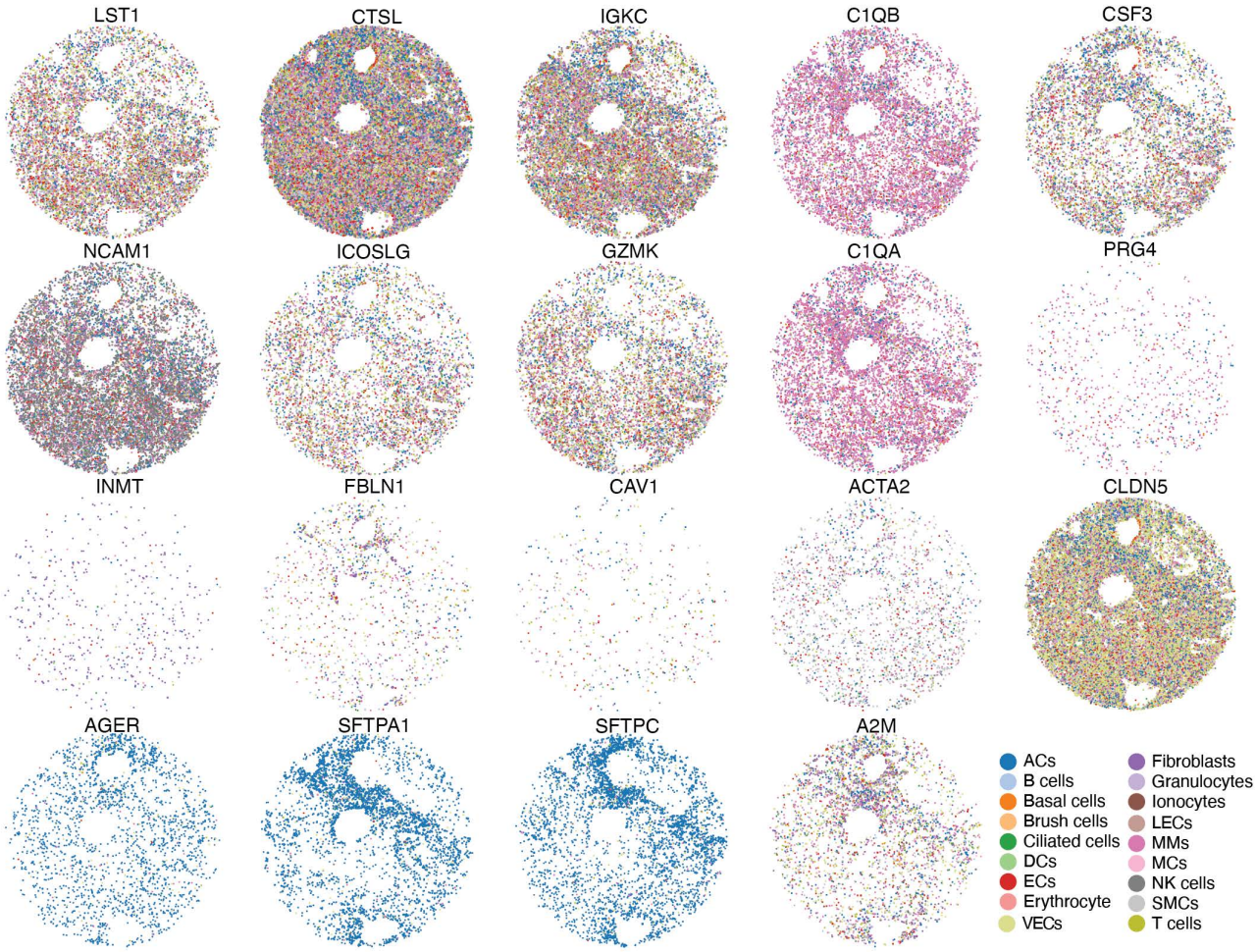


Figure S21: 4-3B

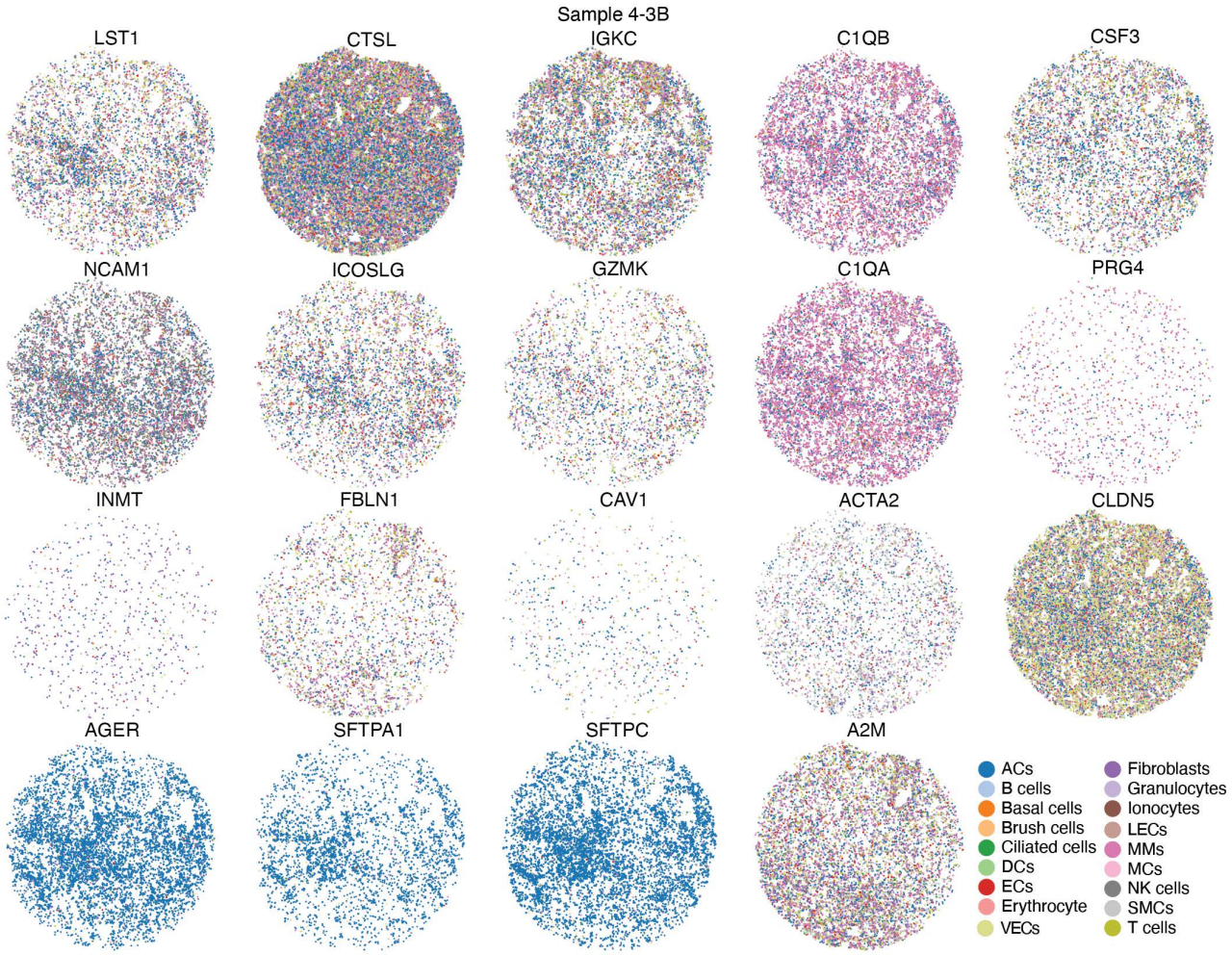


Figure S22: 5-3B

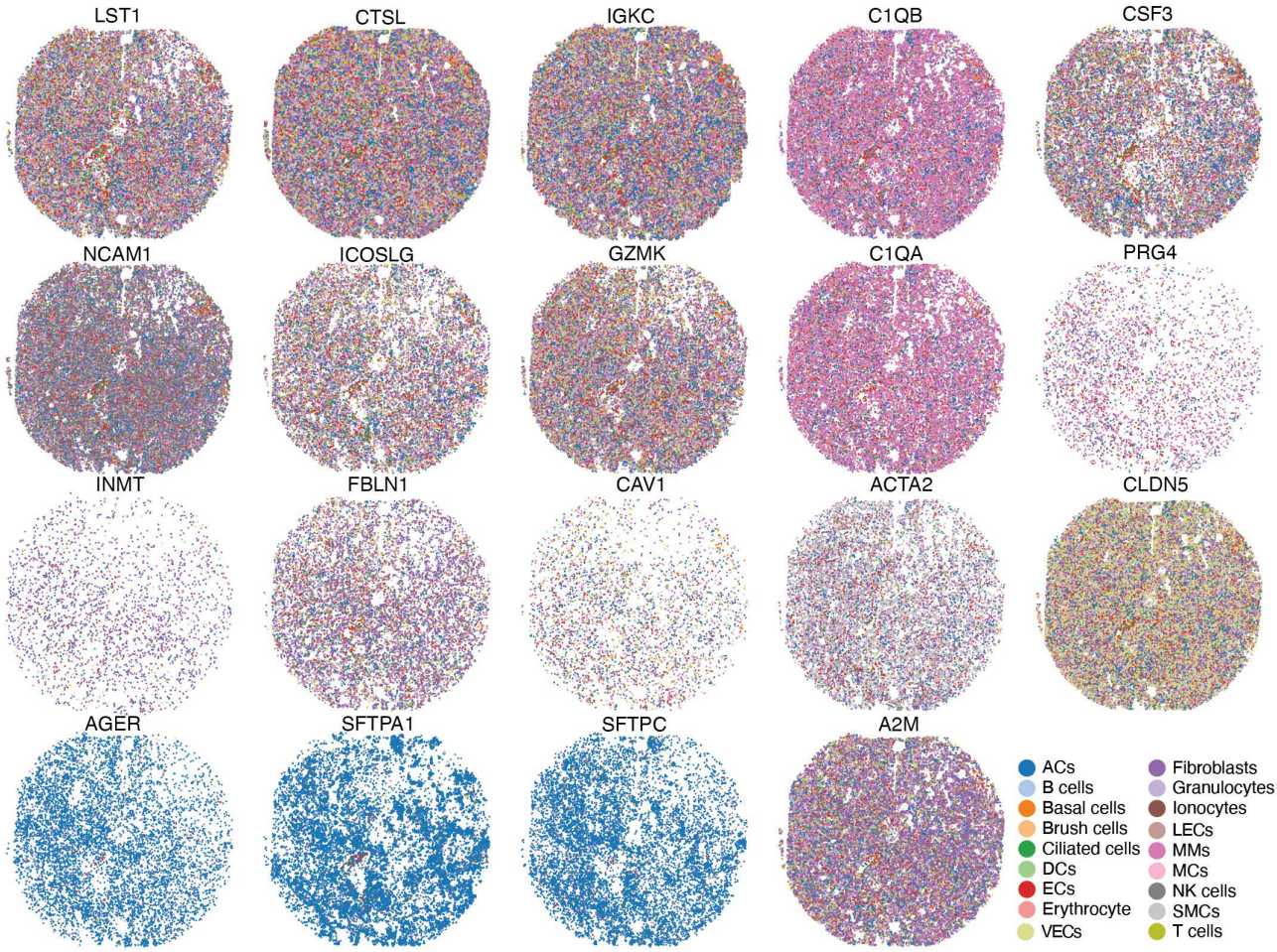


Figure S23: PBC-PR

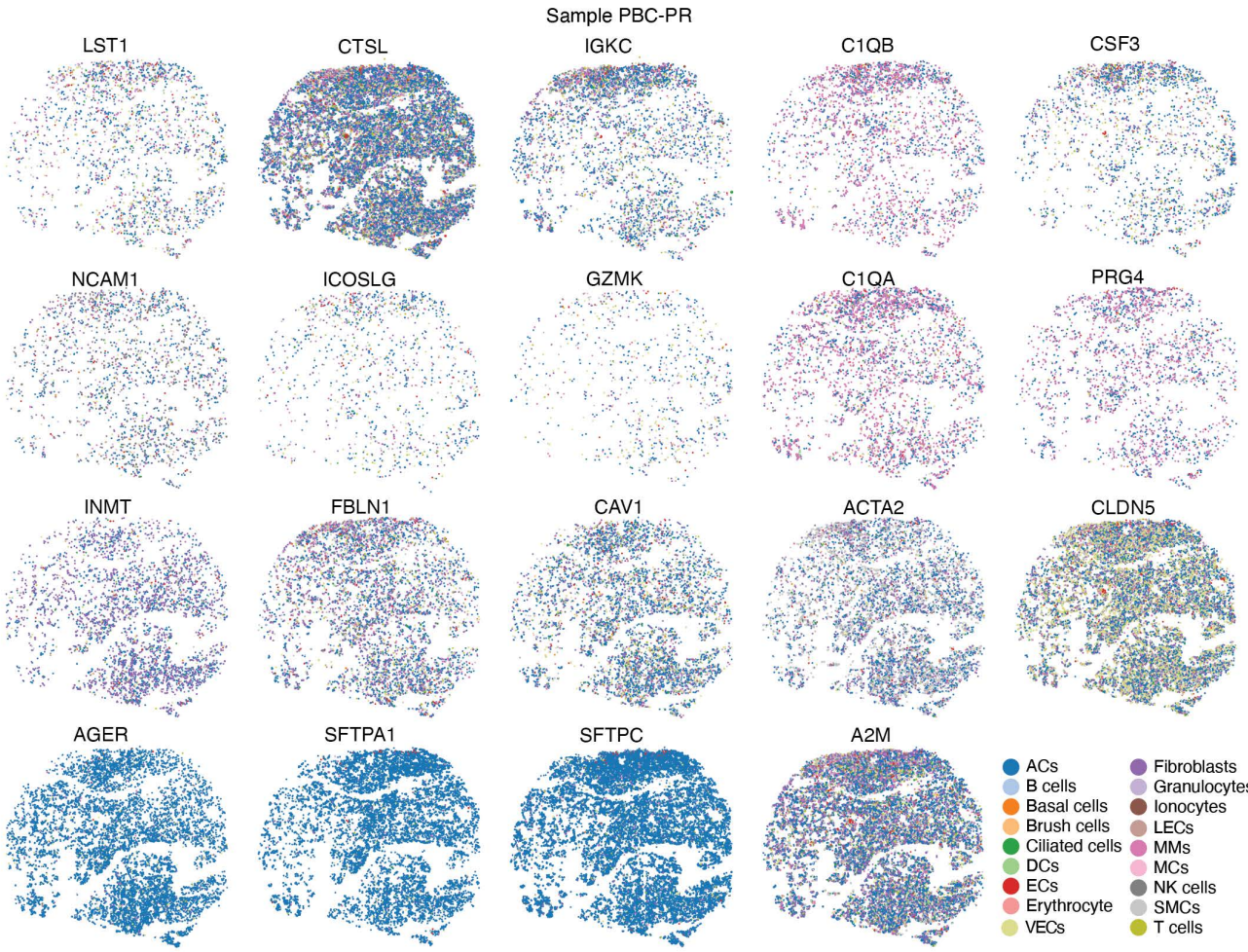


Figure S24

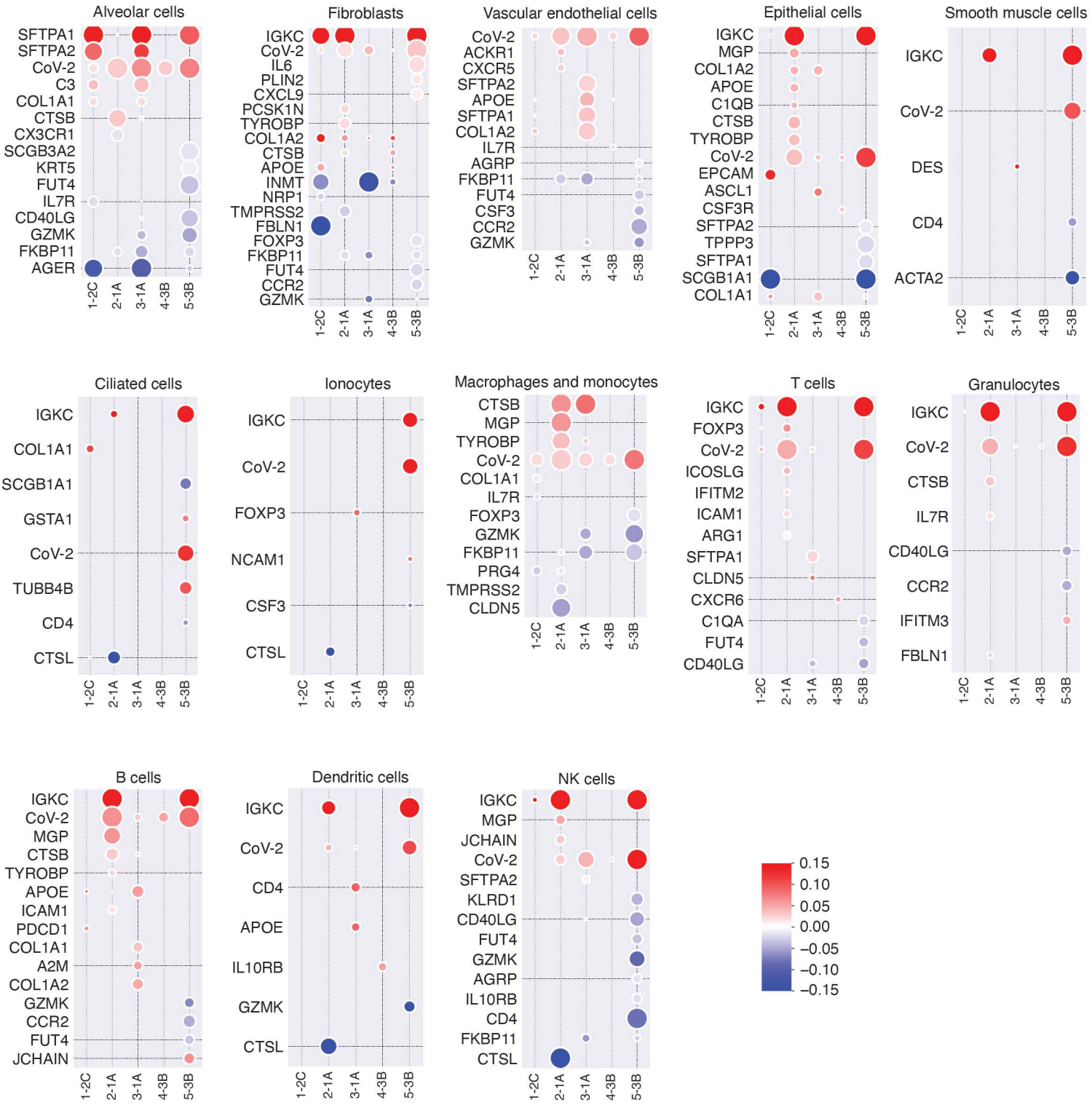


Figure S25

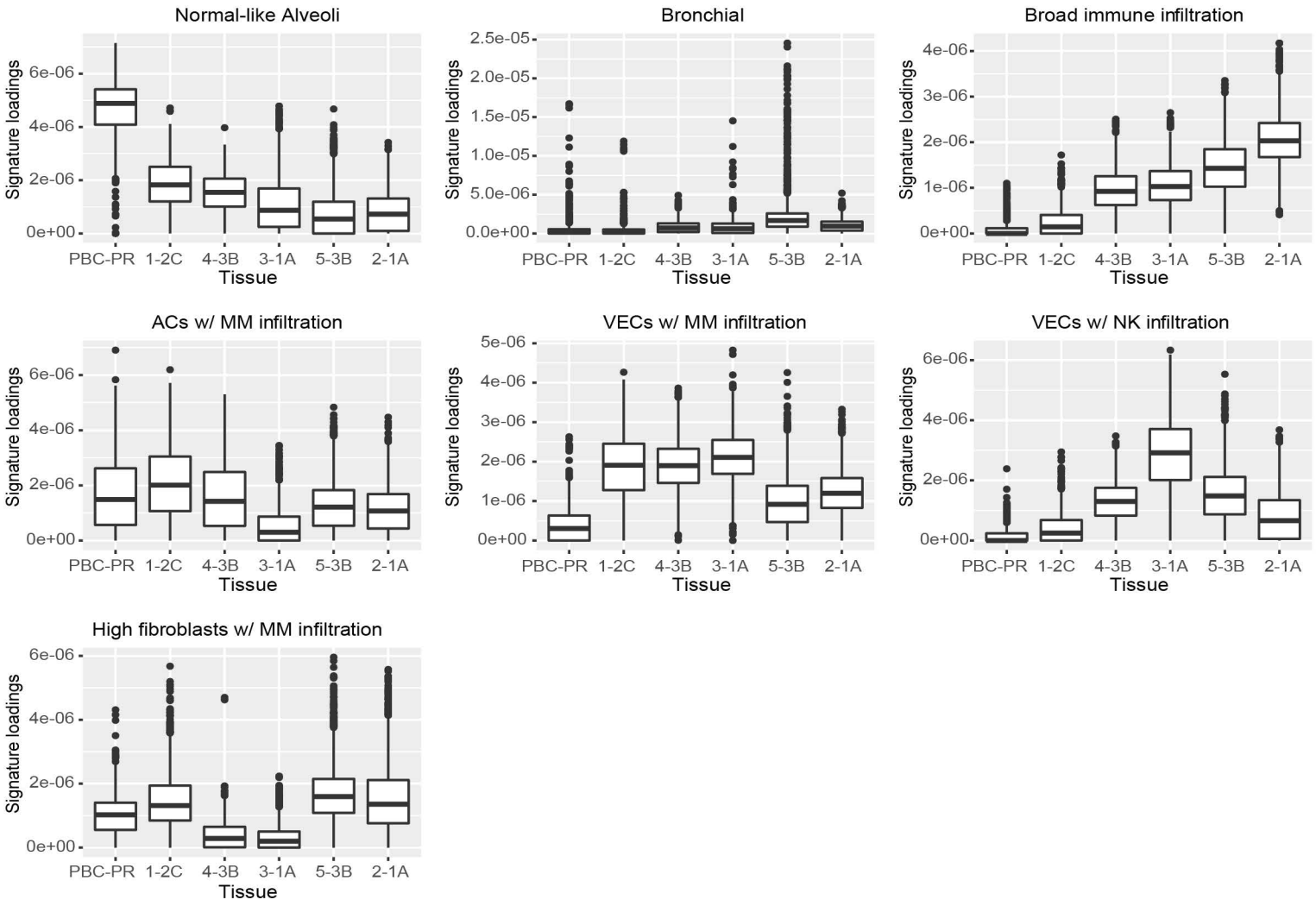


Figure S26

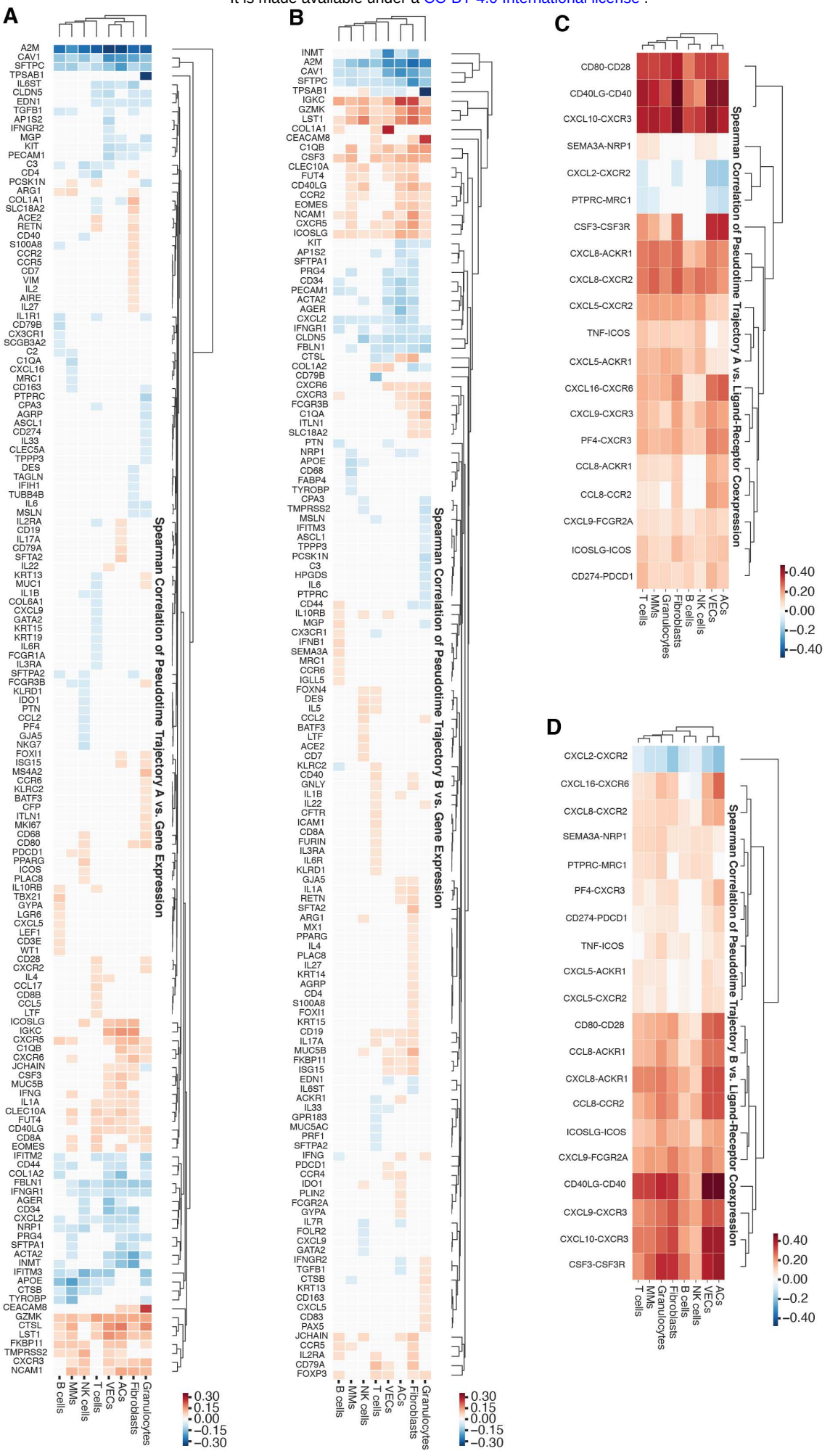


Figure S27

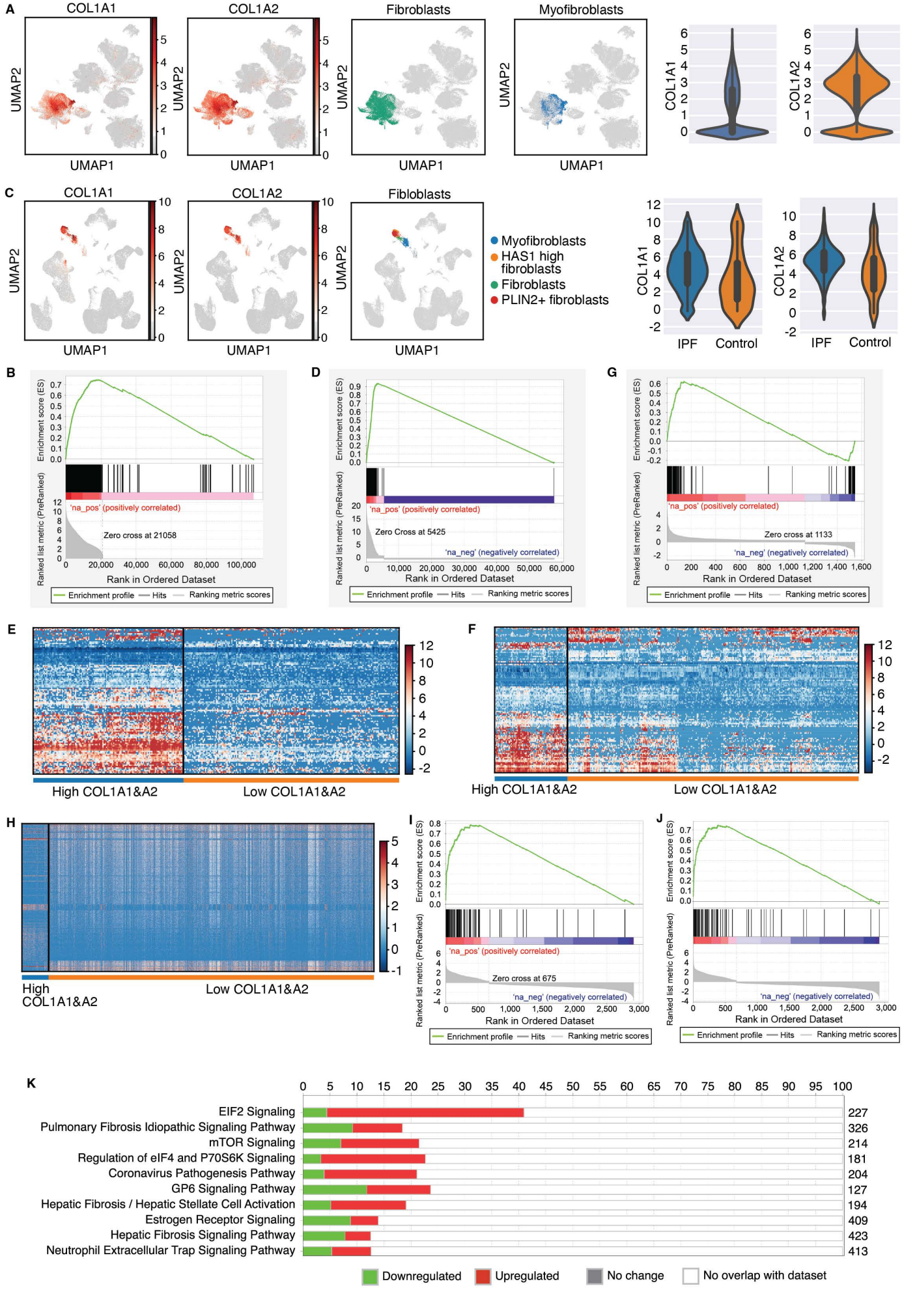


Table S1. Gene annotation, cell type markers, and total cells expressing individual genes and total reads of the individual genes in each lung tissue after segmenta

			Sample name	1-2C		2-1A		3-1A		4-3B		5-3B		PBC-PR	
Gene	HGNC ID	Name	Annotation of Cell Typ	Total C	Total R	Total C	Total R	Total C	Total R	Total C	Total R	Total C	Total R	Total C	Total Re
A2M	HGNC:7	alpha-2-macroglobulin		42643	80341	28959	37365	9830	11626	24699	33682	74687	1E+05	50795	83748
ACE2	HGNC:1355	angiotensin I converting enzyme 2		213	221	2281	2344	1274	1310	667	694	3353	3471	176	178
ACKR	HGNC:4035	atypical chemokine receptor 1 (Duffy blood g	Vascular endothelial ce	1283	1366	5750	5926	3307	3428	2622	2729	9958	10555	2301	2614
ACTA:	HGNC:130	actin alpha 2, smooth muscle	Smooth muscle cells	5164	6156	9006	9442	4110	4296	6756	7571	20041	23428	14003	16904
AGER	HGNC:320	advanced glycosylation end-product specific	Alveolar cells	16263	27549	10958	11939	4034	4371	16636	25020	10146	11282	17701	22204
AGRP	HGNC:330	agouti related neuropeptide		406	427	2963	3067	1601	1655	1154	1208	5717	5966	499	523
AIRE	HGNC:360	autoimmune regulator		196	200	1116	1130	660	671	549	566	1506	1539	321	333
AP1S:	HGNC:560	adaptor related protein complex 1 subunit sigma 2		926	962	1135	1166	806	832	591	609	2509	2570	1954	2042
APOE	HGNC:613	apolipoprotein E		12592	17576	21172	27036	6671	7410	22642	32501	7374	7863	5771	6545
ARG1	HGNC:663	arginase 1	Macrophages and mon	385	399	5884	6076	1952	2018	1534	1594	5139	5357	322	332
ASCL1	HGNC:738	achaete-scute family bHLH transcription factor	Epithelial cells	404	418	3640	3738	1531	1583	1713	1793	3938	4080	301	308
ASCL:	HGNC:740	achaete-scute family bHLH transcription factor	Ionocytes	320	333	1475	1544	1194	1254	588	604	1717	1768	203	217
BATF:	HGNC:2891	basic leucine zipper ATF-like transcription factor 3		541	553	2619	2671	1098	1118	983	1013	2611	2673	437	446
BPIFB	HGNC:1610	BPI fold containing family B member 1	Epithelial cells	94	98	333	352	337	348	196	216	1188	1296	81	83
C1QA	HGNC:1241	complement C1q A chain	Macrophages and mon	20720	27989	37792	45071	17616	20129	20676	24133	36007	40708	6548	7757
C1QB	HGNC:1242	complement C1q B chain	Macrophages and mon	12297	14766	51493	58737	17542	19111	18151	20123	47053	53019	5310	5947
C2	HGNC:1248	complement C2		738	783	896	927	497	515	548	576	1156	1211	291	305
C3	HGNC:1318	complement C3		5183	6047	3892	4086	1834	2012	2738	2956	4882	5191	2884	3135
C4BP	HGNC:1325	complement component 4 binding protein alpha		382	402	1406	1452	650	668	475	487	1589	1642	704	735
CAV1	HGNC:1527	caveolin 1		3742	3999	1940	2007	1098	1129	1671	1737	5318	5526	11368	12788
CCL1	HGNC:1061	C-C motif chemokine ligand 17		249	261	1159	1246	1024	1083	424	447	1566	1683	300	311
CCL2	HGNC:1061	C-C motif chemokine ligand 2		1131	1237	4239	4377	1970	2072	589	615	11301	12757	1651	1759
CCL5	HGNC:1063	C-C motif chemokine ligand 5		758	795	938	961	995	1024	537	554	1338	1396	493	510
CCL8	HGNC:1063	C-C motif chemokine ligand 8		254	270	1143	1161	578	595	403	417	2857	3001	146	149
CCR2	HGNC:1603	C-C motif chemokine receptor 2		2594	2753	19388	20595	8426	8872	7505	7945	29142	31621	3473	3643
CCR4	HGNC:1605	C-C motif chemokine receptor 4		408	425	4728	4880	2670	2808	1730	1809	6289	6509	491	507
CCR5	HGNC:1606	C-C motif chemokine receptor 5		1710	1785	12777	13338	4988	5223	4221	4416	15676	16525	1685	1767
CCR6	HGNC:1607	C-C motif chemokine receptor 6		136	142	1197	1219	661	682	428	442	1195	1223	206	212
CCR7	HGNC:1608	C-C motif chemokine receptor 7		577	597	2161	2210	1791	1853	630	654	3593	3686	565	579
CD14	HGNC:1628	CD14 molecule	Macrophages and mon	1773	1852	1668	1700	1539	1575	759	782	2163	2197	498	514
CD16	HGNC:1631	CD163 molecule	Macrophages and mon	2316	2483	6400	6949	3026	3275	3698	4010	9813	10817	1383	1488
CD19	HGNC:1633	CD19 molecule	B cells	761	785	10762	11125	3179	3273	3477	3608	8204	8483	576	590
CD1C	HGNC:1636	CD1c molecule	Dendritic cells	3239	3373	4063	4182	3639	3760	3177	3294	3550	3629	1215	1249
CD20	HGNC:1641	CD209 molecule		384	401	1780	1826	1131	1172	717	736	2537	2620	399	410
GYPA	HGNC:4702	glycophorin A (MNS blood group)	Erythrocyte	161	169	2335	2409	1410	1447	462	477	3642	3775	101	102
CD27	HGNC:1763	CD274 molecule		237	241	1010	1023	498	505	413	429	1149	1175	192	198
CD28	HGNC:1653	CD28 molecule		257	266	2219	2310	1178	1238	792	821	2634	2731	287	297
CD30	HGNC:2887	CD300e molecule		266	269	942	969	639	654	713	733	1574	1612	191	200
CD34	HGNC:1662	CD34 molecule	Vascular endothelial ce	3195	3371	4713	4856	2423	2511	2047	2132	5684	5884	5157	5513
CD3D	HGNC:1673	CD3d molecule	T cells	1154	1216	5081	5261	3234	3360	2383	2496	8192	8618	1223	1282
CD3E	HGNC:1674	CD3e molecule	T cells	451	464	1159	1177	860	879	383	388	1567	1599	603	618
CD3G	HGNC:1675	CD3g molecule	T cells	44	44	109	109	16	16	17	17	54	54	61	61
CD4	HGNC:1678	CD4 molecule	T cells	9244	9854	32619	34880	15017	15908	15346	16465	45532	50512	6899	7329
CD40	HGNC:1191	CD40 molecule		1284	1323	6802	7021	2935	3042	2489	2584	9651	10041	1100	1128
CD40	HGNC:1193	CD40 ligand		785	827	12160	12659	6193	6460	3159	3307	25458	27608	633	663
CD44	HGNC:1681	CD44 molecule (Indian blood group)		3032	3218	4303	4473	2278	2369	2170	2271	12624	13517	5280	5692
CD68	HGNC:1693	CD68 molecule	Macrophages and mon	6117	6731	15311	16792	12141	13295	8779	9502	27401	30488	6359	7170
CD7	HGNC:1695	CD7 molecule		645	676	4375	4464	1954	2005	1387	1432	3516	3618	455	470
CD79	HGNC:1698	CD79a molecule	B cells	1664	1762	21479	22599	6222	6448	4727	4911	15235	15962	800	842
CD79	HGNC:1699	CD79b molecule	B cells	1428	1506	4402	4647	3115	3286	2752	2877	6391	6771	1968	2061
CD80	HGNC:1700	CD80 molecule		653	676	3873	3981	3291	3383	1595	1652	5186	5335	514	534
CD83	HGNC:1703	CD83 molecule		562	588	3093	3184	1163	1202	1034	1077	3142	3256	666	692
CD86	HGNC:1705	CD86 molecule		1	1	6	6	5	5	6	6	7	7	4	4
CD8A	HGNC:1706	CD8a molecule	T cells	690	717	3340	3433	2467	2550	943	971	3632	3738	515	534
CD8B	HGNC:1707	CD8b molecule	T cells	177	182	638	653	684	710	236	247	762	788	173	182
CEAC/	HGNC:1820	CEA cell adhesion molecule 8	Granulocytes	592	623	12368	12895	5524	5723	2825	2937	15840	16754	477	493
CFP	HGNC:8864	complement factor properdin		249	259	866	886	522	537	253	259	836	857	168	172
CFTR	HGNC:1884	CF transmembrane conductance regulator	Ionocytes	76	78	664	685	362	372	315	318	935	948	88	92
CLDN	HGNC:2047	claudin 5	Vascular endothelial ce	46145	58039	81977	95546	76968	95416	49549	59676	86131	1E+05	38423	49829
CLEC1	HGNC:1691	C-type lectin domain containing 10A		1432	1507	13952	14589	8036	8443	4628	4847	18709	19859	1189	1257
CLEC5	HGNC:2054	C-type lectin domain containing 5A		155	164	606	618	561	567	330	338	1186	1225	249	255
CLEC9	HGNC:2670	C-type lectin domain containing 9A	Dendritic cells	87	89	1244	1337	761	796	326	337	1788	1901	132	139
COL1	HGNC:2197	collagen type I alpha 1 chain	Fibroblasts	7620	10687	21508	31708	4084	4627	2387	2792	25152	32237	4447	4884
COL1	HGNC:2198	collagen type I alpha 2 chain	Fibroblasts	17445	26898	37213	55927	7678	8625	9522	11572	48486	64941	20266	24613
COL6	HGNC:2211	collagen type VI alpha 1 chain		4358	4725	6794	7095	4101	4285	4035	4278	15730	16919	4344	4608
CPA3	HGNC:2298	carboxypeptidase A3	Granulocytes	172	184	784	830	528	540	334	352	1152	1190	536	568
CR1	HGNC:2334	complement C3b/C4b receptor 1 (Knops blood group)		149	156	535	551	552	567	266	273	924	960	177	183
CSF3F	HGNC:2439	colony stimulating factor 3 receptor		1207	1255	4135	4515	2462	2652	2288	2417	5857	6299	978	1027
CTSB	HGNC:2527	cathepsin B		7288	8303	13176	14716	6499	7033	9760	11249	19731	22099	6933	7722
CTSL	HGNC:2537	Cathepsin L		88357	2E+05	2E+05	4E+05	2E+05	4E+05	1E+05	3E+05	3E+05	6E+05	59979	95863
CX3C	HGNC:2558	C-X3-C motif chemokine receptor 1		2180	2289	4816	4977	5229	5456	3811	3979	6462	6702	2088	2175
CXCL:	HGNC:1063	C-X-C motif chemokine ligand 10		503	515	2813	2887	2280	2352	1265	1305	5770	6047	581	615
CXCL:	HGNC:1664	C-X-C motif chemokine ligand 16		925	960	2117	2184	1503	1560	1365	1405	3750	3860	1436	1507
CXCL:	HGNC:1923	C-X-C motif chemokine ligand 17		244	253	1440	1465	781	803	443	453	2546	2644	416	429
CXCL:	HGNC:4603	C-X-C motif chemokine ligand 2		433	459	406	415	498	514	474	484	1440	1501	2463	2681

CXCL!	HGNC:1064	C-X-C motif chemokine ligand 5		115	124	802	821	690	707	380	390	1272	1307	300	318
CXCL!	HGNC:6025	C-X-C motif chemokine ligand 8		584	613	6944	7172	2635	2737	2163	2224	7876	8142	593	612
CXCL!	HGNC:7098	C-X-C motif chemokine ligand 9		910	989	1887	1962	962	994	366	379	2635	2831	1031	1327
CXCR	HGNC:6027	C-X-C motif chemokine receptor 2		384	400	1190	1233	1065	1097	660	680	1723	1785	351	361
CXCR	HGNC:4540	C-X-C motif chemokine receptor 3		1580	1652	13483	14118	7375	7700	3969	4144	15472	16266	1170	1219
CXCR	HGNC:2561	C-X-C motif chemokine receptor 4		409	426	447	453	430	435	363	368	907	927	677	705
CXCR	HGNC:1060	C-X-C motif chemokine receptor 5		1262	1299	22082	23246	6761	6969	4409	4551	26834	28867	608	619
CXCR	HGNC:1664	C-X-C motif chemokine receptor 6		1706	1803	18547	19774	7365	7789	7532	8136	14835	15713	886	911
CYP2I	HGNC:2632	cytochrome P450 family 2 subfamily F member 1		126	131	580	585	359	365	233	238	742	755	148	148
DDX5	HGNC:1910	DExD/H-box helicase 58		348	351	970	975	704	723	468	478	2246	2293	398	408
DES	HGNC:2770	desmin	Smooth muscle cells	2105	2251	6468	6681	4351	4512	2619	2745	11073	12074	3562	4077
EDN1	HGNC:3176	endothelin 1		968	1019	1213	1232	810	824	654	682	3617	3760	2183	2341
EOME	HGNC:3372	eomesodermin		872	913	12299	12803	6617	6955	3231	3372	21051	22484	619	644
EPCAI	HGNC:1152	epithelial cell adhesion molecule	Epithelial cells	473	498	1318	1362	756	781	686	718	1665	1730	566	584
EREG	HGNC:3443	epiregulin		151	152	900	952	696	726	412	435	1556	1636	222	234
F13A:	HGNC:3531	coagulation factor XIII A chain		647	686	2258	2430	1186	1244	773	816	2851	3037	467	484
FABP4	HGNC:3559	fatty acid binding protein 4		44	46	565	576	347	356	226	232	724	742	241	254
FBLN:	HGNC:3600	fibulin 1		13160	18118	12658	14019	3117	3539	7320	8448	19702	22383	11966	14313
FCGR:	HGNC:3613	Fc fragment of IgG receptor Ia		136	136	452	455	186	188	82	82	317	320	60	62
FCGR:	HGNC:3616	Fc fragment of IgG receptor IIa		920	953	4115	4245	2500	2572	1171	1201	6552	6751	986	1026
FCGR:	HGNC:3618	Fc fragment of IgG receptor IIb		499	524	1164	1200	842	858	487	507	1804	1852	442	456
FCGR:	HGNC:3620	Fc fragment of IgG receptor IIIb		2732	2890	5795	5991	2908	3009	1843	1929	9812	10392	1547	1656
FKBP:	HGNC:1862	FKBP prolyl isomerase 11		1764	1864	10293	10777	10756	11504	4846	5096	20117	21717	1634	1706
FOLR:	HGNC:3793	folate receptor beta		1084	1129	1803	1837	1177	1206	792	814	2267	2311	661	679
FOXI1	HGNC:3815	forkhead box I1	Ionocytes	409	428	3688	3804	1886	1963	1337	1384	4358	4513	343	354
FOXJ1	HGNC:3816	forkhead box J1	Ciliated cells	3	3	3	3	3	3	2	2	18	18		
FOXN	HGNC:2139	forkhead box N4		123	128	489	503	432	437	228	238	624	645	138	141
FOXP	HGNC:6106	forkhead box P3	T cells	2743	2853	24051	25729	6922	7186	4868	5058	13528	14139	2291	2367
FURIN	HGNC:8568	furin, paired basic amino acid cleaving enzyme		2131	2233	10658	11076	4796	4976	5064	5313	14298	14921	2722	2845
FUT4	HGNC:4015	fucosyltransferase 4		969	1012	15533	16926	7541	8103	2887	3034	19700	21747	886	917
GATA:	HGNC:4171	GATA binding protein 2	Granulocytes	749	779	3271	3331	1402	1431	832	848	3459	3506	790	812
GATA:	HGNC:4172	GATA binding protein 3		339	350	2641	2707	1199	1231	701	713	3269	3374	329	338
CSF3	HGNC:2438	colony stimulating factor 3		6513	7005	56710	64391	16965	18195	14800	15962	39064	43311	4914	5203
GJA5	HGNC:4279	gap junction protein alpha 5		4072	4323	14580	15272	6626	6995	5248	5544	26618	28678	3351	3669
GNLY	HGNC:4414	granulysin		1299	1436	5564	5748	2419	2498	1598	1676	7443	7725	622	671
GPR1	HGNC:3128	G protein-coupled receptor 183		157	159	444	455	547	560	266	276	663	680	238	248
GSTA1	HGNC:4626	glutathione S-transferase alpha 1	Ciliated cells	305	315	1094	1120	949	972	679	702	2015	2069	418	431
GZMK	HGNC:4711	granzyme K		2384	2493	33257	35891	18686	20077	7559	7977	66279	78630	1577	1632
HLA-D	HGNC:4944	major histocompatibility complex, class II, DQ beta 1		1320	1390	1267	1335	790	823	970	1015	1431	1507	1331	1423
HLA-D	HGNC:4954	major histocompatibility complex, class II, DR beta 6 (pseudogene)		769	799	3451	3546	2105	2175	1457	1492	4225	4382	597	621
HPGD	HGNC:1789	hematopoietic prostaglandin D synthase		91	92	470	486	495	514	208	214	673	697	169	172
ICAM	HGNC:5344	intercellular adhesion molecule 1		2515	2710	7157	7385	2243	2313	2848	2976	6938	7172	2268	2381
ICOS	HGNC:5351	inducible T cell costimulator		201	202	1277	1310	586	601	384	401	1182	1213	210	215
ICOSL	HGNC:1708	inducible T cell costimulator ligand		2901	3035	32600	35066	12841	13580	12265	13179	27279	29152	2238	2328
IDO1	HGNC:6059	indoleamine 2,3-dioxygenase 1		640	683	965	986	777	802	637	668	5481	5824	337	361
IFIH1	HGNC:1887	interferon induced with helicase C domain 1		421	434	728	743	433	446	373	380	1621	1670	424	435
IFITM	HGNC:5413	interferon induced transmembrane protein 2		8012	8583	16017	16883	7018	7363	9795	10406	28790	31093	8062	8677
IFITM	HGNC:5414	interferon induced transmembrane protein 3		16452	18575	17428	18617	8998	9577	19021	21300	64561	77842	12481	13753
IFNA2	HGNC:5423	interferon alpha 2		33	33	314	326	281	293	133	135	450	458	58	60
IFNB1	HGNC:5434	interferon beta 1		139	143	1292	1319	641	656	431	453	1382	1421	134	136
IFNE	HGNC:1816	interferon epsilon		54	57	258	264	185	188	98	102	210	219	57	57
IFNG	HGNC:5438	interferon gamma		525	555	4161	4328	3203	3374	1638	1713	5996	6342	357	372
IFNGF	HGNC:5439	interferon gamma receptor 1		1795	1876	1465	1496	724	744	1654	1711	3837	3974	3034	3194
IFNGF	HGNC:5440	interferon gamma receptor 2		713	748	1920	1983	932	959	879	916	3236	3362	1227	1273
IGKC	HGNC:5716	immunoglobulin kappa constant		74828	2E+05	74799	1E+05	55923	1E+05	28740	50571	1E+05	3E+05	10118	15596
IGLL5	HGNC:3847	immunoglobulin lambda like polypeptide 5		2009	2643	7002	7721	4199	4514	3505	3873	11634	13561	1240	1439
IL10	HGNC:5962	interleukin 10		79	81	385	388	209	211	193	199	395	401	72	75
IL10R	HGNC:5965	interleukin 10 receptor subunit beta		833	871	5833	6008	3274	3368	2142	2241	9732	10091	979	1033
IL17A	HGNC:5981	interleukin 17A		378	391	4253	4397	1900	1966	956	979	4913	5080	290	297
IL1A	HGNC:5991	interleukin 1 alpha		488	501	6524	6789	2731	2864	1406	1468	7135	7411	392	403
IL1B	HGNC:5992	interleukin 1 beta		711	735	3369	3432	1390	1419	1180	1226	3375	3448	448	458
IL2	HGNC:6001	interleukin 2		127	131	2079	2133	1121	1190	735	767	2614	2702	159	163
IL22	HGNC:1490	interleukin 22		275	278	3039	3120	1750	1813	1074	1115	3001	3090	232	239
IL27	HGNC:1915	interleukin 27		239	249	2350	2411	924	950	798	819	2139	2190	185	191
IL2RA	HGNC:6008	interleukin 2 receptor subunit alpha		2346	2431	10989	11338	3998	4103	4454	4625	12931	13390	1482	1532
IL33	HGNC:1602	interleukin 33		197	199	394	399	218	221	176	180	620	635	369	384
IL3RA	HGNC:3563	interleukin 3 receptor subunit alpha		946	982	5462	5666	2165	2228	1478	1530	6942	7183	1201	1246
IL4	HGNC:6014	interleukin 4		473	491	3969	4146	1787	1862	1503	1577	3354	3519	319	325
IL5	HGNC:6016	interleukin 5		343	355	1668	1720	964	987	578	592	1304	1355	289	299
IL6	HGNC:6018	interleukin 6		319	333	1430	1471	1062	1094	527	546	4160	4399	1266	1381
IL6R	HGNC:6019	interleukin 6 receptor		554	573	3306	3366	1139	1155	1114	1146	2536	2600	685	717
IL6ST	HGNC:6021	interleukin 6 signal transducer		688	715	1378	1415	632	649	685	709	3566	3700	2331	2450
IL7	HGNC:6023	interleukin 7		287	297	1906	1950	1214	1250	802	827	2714	2797	338	350
IL7R	HGNC:6024	interleukin 7 receptor		3600	3920	4640	4800	1612	1653	2247	2367	3929	4051	2181	2327
INMT	HGNC:6069	indolethylamine N-methyltransferase	Fibroblasts	1312	1396	1323	1387	924	959	1260	1323	3237	3408	5878	6438
IRF4	HGNC:6119	interferon regulatory factor 4		902	921	4295	4375	1593	1623	1110	1134	3447	3509	454	464
ISG15	HGNC:4053	ISG15 ubiquitin like modifier		1784	1866	7142	7480	4105	4337	4103	4309	15226	16219	1369	1443
ITGAX	HGNC:6152	integrin subunit alpha X	Dendritic cells	915	949	1519	1546	1110	1130	1644	1706	2850	2930	730	765
ITLN1	HGNC:1825	intelectin 1		250	259	2774	2848	1509	1544	741	766	3299	3378	248	251

JCHAI	HGNC:5713	joining chain of multimeric IgA and IgM		9916	21499	10626	14866	6728	7953	2628	3794	19176	35488	1454	3375
KIT	HGNC:6342	KIT proto-oncogene, receptor tyrosine kinase		482	499	978	996	730	757	464	480	2147	2212	2120	2278
KLRB:	HGNC:6373	killer cell lectin like receptor B1		311	329	1521	1556	888	906	608	625	1945	1995	304	317
KLRC:	HGNC:6375	killer cell lectin like receptor C2	NK cells	131	137	484	499	494	502	323	339	675	695	129	133
KLRD:	HGNC:6378	killer cell lectin like receptor D1	NK cells	536	562	4258	4482	1868	1931	986	1020	5417	5665	443	456
KRT1:	HGNC:6415	keratin 13		184	202	880	902	534	549	396	411	1116	1158	112	113
KRT1-	HGNC:6416	keratin 14	Basal cells	265	270	2960	3028	1472	1513	746	768	3546	3641	278	288
KRT1!	HGNC:6421	keratin 15		840	901	5259	5485	2733	2867	1606	1684	6785	7147	537	553
KRT1!	HGNC:6436	keratin 19		4295	5157	7262	7681	4066	4336	2867	3044	17270	20219	3220	3432
KRT5	HGNC:6442	keratin 5	Basal cells	186	213	589	602	588	705	275	288	3127	4635	112	113
LEF1	HGNC:6551	lymphoid enhancer binding factor 1		214	220	1782	1824	1187	1238	522	544	3048	3185	603	630
LGR6	HGNC:1971	leucine rich repeat containing G protein-coupled receptor 6		499	516	3062	3191	1629	1696	1002	1054	3142	3296	431	448
LILRA	HGNC:1550	leukocyte immunoglobulin like receptor A4		206	210	1596	1636	827	850	568	577	1548	1576	173	176
LST1	HGNC:1418	leukocyte specific transcript 1		5422	5762	68058	78104	25499	27823	13921	14974	72496	84944	4255	4509
LTF	HGNC:6720	lactotransferrin		147	161	1482	1530	767	795	322	332	2419	2485	136	137
LYVE1	HGNC:1468	lymphatic vessel endothelial hyaluronan rece	Lymphatic endothelial	336	352	608	625	308	317	214	221	376	385	740	786
CSF1	HGNC:2432	colony stimulating factor 1		514	528	1226	1251	474	479	314	320	1096	1113	344	348
MGP	HGNC:7060	matrix Gla protein		9428	10545	34860	38829	10408	10997	6878	7373	22754	24270	10835	11935
MKI6	HGNC:7107	marker of proliferation Ki-67		61	62	180	183	191	194	70	71	348	359	59	61
MRC1	HGNC:7228	mannose receptor C-type 1		1143	1197	5250	5446	2179	2227	1635	1693	6191	6375	1428	1522
MS4A	HGNC:7315	membrane spanning 4-domains A1	B cells	132	133	1053	1067	709	725	400	406	1693	1727	110	114
MS4A	HGNC:7316	membrane spanning 4-domains A2	Granulocytes	299	305	3172	3266	1674	1714	833	857	4015	4129	401	414
MSLN	HGNC:7371	mesothelin	Mesothelial cells	1191	1294	3928	4078	1435	1474	913	964	3332	3449	1717	1799
MUC1	HGNC:7508	mucin 1, cell surface associated		955	1021	662	686	638	663	579	609	2614	2738	1132	1189
MUC5	HGNC:7515	mucin 5AC, oligomeric mucus/gel-forming	Epithelial cells	145	158	799	827	453	462	502	528	987	1014	154	157
MUC5	HGNC:7516	mucin 5B, oligomeric mucus/gel-forming	Epithelial cells	1310	1456	14210	14781	7453	7735	4319	4526	24767	26669	853	878
MX1	HGNC:7532	MX dynamin like GTPase 1		1579	1638	7232	7514	2792	2877	3514	3655	11341	11882	1414	1471
NCAM	HGNC:7656	neural cell adhesion molecule 1	NK cells	6110	6557	40906	44692	27697	30467	17120	18732	58792	66620	3296	3515
NKG7	HGNC:7830	natural killer cell granule protein 7		1352	1492	1162	1192	1060	1110	837	875	2043	2111	758	818
NRP1	HGNC:8004	neuropilin 1		4267	4523	4071	4200	2177	2264	2922	3045	11280	11857	4047	4282
PAX5	HGNC:8619	paired box 5		270	277	2099	2158	1711	1760	566	581	3210	3333	253	264
PCSK:	HGNC:1730	proprotein convertase subtilisin/kexin type 1	inhibitor	4234	4500	18102	19090	9788	10407	9194	9843	17936	18870	3926	4233
PDCD	HGNC:8760	programmed cell death 1		5017	5317	14253	14860	11119	11612	6214	6484	10882	11277	2413	2483
PECAI	HGNC:8823	platelet and endothelial cell adhesion molecu	Vascular endothelial ce	538	551	452	458	308	313	527	533	1013	1028	1484	1525
PF4	HGNC:8861	platelet factor 4		271	309	733	751	584	605	393	424	706	729	94	95
PLAC8	HGNC:1925	placenta associated 8		1599	1703	9248	9601	4682	4869	2901	3038	12029	12703	1279	1329
PLIN2	HGNC:248	perilipin 2		741	766	5024	5171	2095	2158	1803	1869	7434	7721	939	976
PPAR	HGNC:9236	peroxisome proliferator activated receptor gamma		582	602	4635	4783	2662	2744	1738	1814	7013	7320	1229	1305
PRF1	HGNC:9360	perforin 1		672	690	2647	2784	1522	1595	1416	1495	3888	4107	538	551
PRG4	HGNC:9364	proteoglycan 4	Macrophages and mon	1346	1483	2925	3053	1535	1595	1947	2081	4697	4919	3845	4252
PTN	HGNC:9630	pleiotrophin		2909	3202	7876	8310	4531	4858	4619	5019	8072	8546	2408	2577
PTPR(	HGNC:9666	protein tyrosine phosphatase receptor type C		1938	2015	4054	4167	3043	3133	1542	1591	6289	6473	1931	2026
RETN	HGNC:2038	resistin		779	811	7065	7348	3445	3607	2932	3125	9538	10016	715	746
S100/	HGNC:1049	S100 calcium binding protein A8		3361	4237	4766	5142	3103	3395	1839	2036	8888	10537	500	553
SCGB:	HGNC:1252	secretoglobin family 1A member 1	Epithelial cells	1865	4537	5388	6536	2592	3999	2288	2948	9280	16352	1106	3020
SCGB:	HGNC:1839	secretoglobin family 3A member 2	Epithelial cells	534	645	1572	1622	1012	1054	904	988	2634	2825	1687	2580
SEMA	HGNC:1072	semaphorin 3A		72	73	733	749	364	372	245	253	886	914	139	142
SFTA2	HGNC:1838	surfactant associated 2	Alveolar cells	1690	1805	3864	3996	1901	1990	2040	2160	5024	5260	1957	2057
SFTPA	HGNC:1079	surfactant protein A1	Alveolar cells	6106	8052	10922	14182	7860	9860	6527	8187	24407	44290	18841	29304
SFTPA	HGNC:1079	surfactant protein A2	Alveolar cells	4370	5125	10027	11632	4643	5192	4322	4783	14103	17992	7456	8821
SFTPC	HGNC:1080	surfactant protein C	Alveolar cells	25948	69259	16766	21268	8284	10209	16532	26081	16118	19624	35495	74526
SLC1E	HGNC:1093	solute carrier family 18 member A2	Granulocytes	217	223	2954	3056	1801	1871	795	834	4400	4559	291	299
TAGLI	HGNC:1155	transgelin	Smooth muscle cells	345	356	1388	1411	915	931	764	792	1926	1979	1234	1285
TBX2:	HGNC:1159	T-box transcription factor 21		1202	1252	5703	5876	3600	3711	2502	2590	7273	7523	971	1004
TGFB1	HGNC:1176	transforming growth factor beta 1		1562	1650	2225	2276	1735	1770	1231	1280	3875	3965	2271	2374
TMPR	HGNC:1187	transmembrane serine protease 2		9404	10363	22628	24172	23341	25544	10743	11645	19259	20485	6563	7089
TNF	HGNC:1189	tumor necrosis factor		151	153	622	634	529	544	186	188	972	987	197	203
TP63	HGNC:1597	tumor protein p63	Basal cells	82	88	413	449	376	401	140	150	549	583	60	61
TPPP:	HGNC:2416	tubulin polymerization promoting protein fa	Ciliated cells	1484	1763	4075	4233	1630	1767	2161	2474	7056	7902	1962	2064
TPSAE	HGNC:1201	tryptase alpha/beta 1	Granulocytes	4307	5425	3784	3908	2953	3162	4152	4676	6304	6736	3633	4261
TREM	HGNC:1776	triggering receptor expressed on myeloid cells 2		1367	1428	2553	2640	1472	1517	1370	1424	2609	2673	481	500
TRPM	HGNC:1432	transient receptor potential cation channel su	Brush cells	298	307	2062	2103	1322	1362	644	665	2399	2467	394	409
TUBB4	HGNC:2077	tubulin beta 4B class IVb	Ciliated cells	912	956	599	617	512	524	391	397	1620	1668	982	1009
TYROI	HGNC:1244	transmembrane immune signaling adaptor TYROBP		6913	7616	14347	15209	5989	6243	6699	7191	15054	15909	7053	8452
VIM	HGNC:1269	vimentin		930	975	3069	3156	3152	3282	1261	1305	2791	2868	847	873
WT1	HGNC:1279	WT1 transcription factor		62	65	245	246	307	316	138	141	455	460	70	73
IL1R1	HGNC:IL1R1			1780	1867	2331	2400	862	885	1504	1571	11104	11957	2721	2856
CoV v	HGNC:			2054	2222	6879	7865	3393	3577	3356	3720	13596	20379		

Table S2. Summary of cell segmentation results

Sample name	1-2C	2-1A	3-1A	4-3B	5-3B	PBC-PR	Total
Total Reads	1,137,704	2,380,636	1,441,118	1,161,277	3,424,675	869,453	10,414,863
Number of Genes	222	222	222	222	222	221	
Total Segmented Reads	1,074,491	2,224,394	1,339,792	1,112,148	3,231,976	773,102	9,755,903
Percentages of Segmented Reads (%)	94.44381	93.43696	92.96893	95.7694	94.37322	88.9182	93
Total Cells	186,659	406,963	276,676	227,585	470,294	151,282	1,719,459
Total Cells with Cell Type Marker Gene	137,260	317,259	193,293	163,431	369,042	126,531	1,306,816
Percentages of Cells with Cell Type Marker Gene (%)	73.53516	77.9577	69.86258	71.81097	78.47049	83.63916	
Total Cells after Filtering	67,701	196,221	87,701	74,843	287,832	62,674	776,972
Average Transcripts per Cell	5.756469	5.465852	4.842476	4.886758	6.87226	5.110371	

Table S3. Summary of cell typing results and SARS-CoV-2 infection status

Sample Name		Alveolar	Macroph	Fibrobla	Vascula	T cells	Granuloc	B cells	Dendritic	Epitheli	Smooth	NK cells	Ciliated	Ionocyte	Basal ce	Mesothe	Brush ce	Lympha	Erythroc	Total
1-2C	Number of Cells	19111	14470	10097	8621	2936	2165	1831	1695	1622	1612	1574	969	275	209	199	142	118	55	67701
	Cell Type Percentage(%)	28.229	21.373	14.914	12.734	4.3367	3.1979	2.7045	2.5037	2.3958	2.3811	2.3249	1.4313	0.4062	0.3087	0.2939	0.2097	0.1743	0.0812	100
	Number of infected Cells	336	179	152	115	51	32	20	19	30	29	26	23	6	4	2	1	1	0	690
	Infected Cell Percentage (%	1.7581	1.237	1.5054	1.334	1.7371	1.4781	1.0923	1.1209	1.8496	1.799	1.6518	2.3736	2.1818	1.9139	1.005	0.7042	0.8475	0	0.0102
2-1A	Number of Cells	28,779	40357	25625	16635	13360	9856	20629	3120	12470	3392	11793	2792	2113	2018	957	1038	222	1065	167442
	Cell Type Percentage (%)	14.667	20.567	13.059	8.4777	6.8086	5.0229	10.513	1.59	6.3551	1.7287	6.0101	1.4229	1.0768	1.0284	0.4877	0.529	0.1131	0.5428	100
	Number of infected Cells	501	660	350	301	309	210	653	49	228	66	258	56	30	50	14	17	4	15	3771
	Infected Cell Percentage (%	1.7409	1.6354	1.3659	1.8094	2.3129	2.1307	3.1654	1.5705	1.8284	1.9458	2.1877	2.0057	1.4198	2.4777	1.4629	1.6378	1.8018	1.4085	0.0225
3-1A	Number of Cells	11,240	16820	4498	15384	5699	4882	6073	2212	5766	1882	8064	1285	1189	1126	332	584	118	547	87,701
	Cell Type Percentage (%)	12.816	19.179	5.1288	17.541	6.4982	5.5666	6.9247	2.5222	6.5746	2.1459	9.1949	1.4652	1.3557	1.2839	0.3786	0.6659	0.1345	0.6237	100
	Number of infected Cells	363	224	91	276	97	66	131	31	83	43	112	24	17	14	2	8	2	7	1591
	Infected Cell Percentage (%	3.2295	1.3317	2.0231	1.7941	1.7021	1.3519	2.1571	1.4014	1.4395	2.2848	1.3889	1.8677	1.4298	1.2433	0.6024	1.3699	1.6949	1.2797	0.0181
4-3B	Number of Cells	17,337	15485	4825	10434	4132	3179	4698	1931	3683	1974	4080	1278	689	459	183	259	60	157	74,843
	Cell Type Percentage (%)	23.164	20.69	6.4468	13.941	5.5209	4.2476	6.2771	2.5801	4.921	2.6375	5.4514	1.7076	0.9206	0.6133	0.2445	0.3461	0.0802	0.2098	100
	Number of infected Cells	402	267	96	209	94	67	144	32	80	50	90	26	21	9	6	5	3	1	1602
	Infected Cell Percentage (%	2.3187	1.7242	1.9896	2.0031	2.2749	2.1076	3.0651	1.6572	2.1721	2.5329	2.2059	2.0344	3.0479	1.9608	3.2787	1.9305	5	0.6369	0.0214
5-3B	Number of Cells	42,922	50411	40691	24428	17326	15181	20224	4651	24242	8788	21432	5747	2828	4558	929	1435	147	1892	######
	Cell Type Percentage (%)	14.912	17.514	14.137	8.4869	6.0195	5.2743	7.0263	1.6159	8.4223	3.0532	7.446	1.9967	0.9825	1.5836	0.3228	0.4986	0.0511	0.6573	100
	Number of infected Cells	978	1319	893	768	501	443	524	135	705	258	840	130	89	108	46	33	3	66	7839
	Infected Cell Percentage (%	2.2786	2.6165	2.1946	3.1439	2.8916	2.9181	2.591	2.9026	2.9082	2.9358	3.9194	2.262	3.1471	2.3695	4.9516	2.2997	2.0408	3.4884	0.0272
PBC-	Number of Cells	27378	5719	9465	7717	1491	1638	1173	736	1262	3290	629	922	139	159	185	145	199	29	62276
	Cell Type Percentage (%)	43.962	9.1833	15.198	12.392	2.3942	2.6302	1.8836	1.1818	2.0265	5.2829	1.01	1.4805	0.2232	0.2553	0.2971	0.2328	0.3195	0.0466	100

Table S4. Global spatial correlations between local SARS-CoV-2 infection rates and cell densities by global Moran's I.

Sample name	1-2C	2-1A	3-1A	4-3B	5-3B
Coefficient (r)	0.325	0.636	0.287	0.248	0.078

Table S5. Cell composition signatures identified by sparse non-negative matrix factorization (SNMF)

	Normal-like Alveoli	Bronchile	Broad immune infiltration	ACs w/ MM infiltration	VECs w/ MM infiltration	VECs w/ NK infiltration	High fibroblasts w/ MM
n-Alveolar	0.5643179	0.0121138	0	0.6477053	0.0404455	0	0.0336691
n-B	0.0177815	0	0.2864881	0.0001489	0.0357681	0.0298888	0.0017179
n-Basal	0	0.1029554	0.0025063	0	0	0.0135613	0
n-Brush	0	0	0.0019199	0	0	0.0032618	0
n-Ciliated	0.0105512	0.1403089	0.002555	0.002187	0.0131246	0	0.0055952
n-Dendritic	0.0076009	0.006494	0.0079063	0.0128054	0.0373424	0.0272293	0
n-Epithelial	0	0.6141461	0.0867002	0.0092846	0	0.1219462	0
n-Erythrocyte	0	0.0014696	0.0023376	0	0	0.0068894	0
n-Fibroblasts	0.1193691	0	0	0	0.031177	0	0.7342991
n-Granulocytes	0.0125711	0.0494878	0.0963014	0.0253557	0.0136434	0.1360229	0.0080645
n-Ionocytes	0	0	0.0145924	0	0.0018038	0.0202377	0
n-Lymphatic	0	0	0	0	0	0	0
n-Macrophages	0	0.0162048	0.2233146	0.2728252	0.4667314	0	0.1577208
n-Mesothelial	0	0	0	0	0	0	0
n-NK	0	0.0066883	0.0999384	0.0296879	0	0.3184191	0.0179067
n-Smooth	0.0529362	0.0501313	0	0	0.0099377	0.0428492	0.0410268
n-T	0.0294544	0	0.17544	0	0.0869477	0.0040867	0
n-Vascular	0.1854178	0	0	0	0.2630784	0.2756075	0

Table S6. Spearman Correlations of Pseudotime Trajectory A versus Gene Expressions

Gene	ACs	VECs	Fibroblasts	MMs	NK cells	T cells	B cells	Granulocytes
A2M	-0.2765578	-0.285168	-0.238348	-0.182516	-0.229521	-0.249141	-0.223907	-0.2508999
ACE2			0.043395			0.061204		
ACTA2	-0.1070443	-0.084294	-0.149374	-0.054684				-0.07261664
AGER	-0.0464048	-0.124302			-0.069421			
AGRP								-0.06517783
AIRE			0.044831					
AP1S2		-0.085735						
APOE		-0.057446		-0.164077	-0.089098	-0.054176	-0.12528	-0.08344893
ARG1			0.046766	0.063714			0.044833	
ASCL1								-0.0473939
BATF3								0.044194834
C1QA				-0.078339				
C1QB	0.0867708		0.059652					0.061753909
C2				-0.045239			-0.042925	
C3		-0.045129			-0.076592	-0.048363	-0.043823	
CAV1	-0.1734009	-0.188403	-0.157273	-0.090972	-0.120304	-0.05613	-0.11769	-0.12119816
CCL17						0.0635		
CCL2					-0.052658			
CCL5						0.04699		
CCR2			0.051705					
CCR5			0.054213					
CCR6								0.045881367
CD163				-0.046732				-0.04655853
CD19	0.0438856							
CD274								-0.04879447
CD28						0.059283		0.046970801
CD34	-0.0782965	-0.120828	-0.099378		-0.071386			
CD3E							0.0501	
CD4			0.046742		-0.046078	-0.086732		
CD40			0.056504		-0.052999			
CD40LG	0.0655237	0.058233	0.046805			0.069183		0.06671726
CD44	-0.0692463	-0.062723		-0.0675			-0.048534	-0.07242388
CD68					0.053609			0.061876787
CD7			0.053738					
CD79A	0.061052							
CD79B							-0.075542	
CD80			0.058972		0.05314			0.074283622
CD8A			0.049234	0.054959		0.070571		0.063172204
CD8B						0.052538		
CEACAM8	0.0634125		0.063658					0.211105625
CFP								0.043873213
CLDN5		-0.080936	-0.055542			-0.065148		-0.0632258
CLEC10A	0.080319	0.050673	0.050981	0.06959		0.070331		

CLEC5A								-0.04331821
COL1A1	-0.0434716		0.095684			-0.044134		
COL1A2	-0.1288032	-0.077353					-0.071241	-0.04643754
COL6A1						-0.051441		
CPA3						-0.04277		-0.05276376
CSF3	0.0825488	0.055434	0.057477					
CTSB				-0.123315	-0.066372	-0.06243	-0.049898	
CTSL	0.1555903	0.134861	0.082006	0.109746		0.064454	0.060173	0.142779746
CX3CR1							-0.04356	
CXCL16				-0.050227				
CXCL2	-0.0819782	-0.070601	-0.074434		-0.052996	-0.047334	-0.056277	
CXCL5							0.043694	
CXCL9						-0.047127		
CXCR2						0.045496		0.057255351
CXCR3	0.0534724	0.073938	0.07601	0.04483	0.082672			0.091774059
CXCR5	0.0936446	0.055021	0.10011	0.055826			0.073588	0.056048048
CXCR6	0.0586824		0.093284	0.043907				0.051339988
DES			-0.052901					
EDN1	-0.0567181	-0.047026	-0.05669			-0.053867		-0.0593376
EOMES	0.0643587			0.055331		0.069938		0.046845356
FBLN1	-0.0936202	-0.097959	-0.109604	-0.048301	-0.086736	-0.063898		-0.09633971
FCGR1A						-0.04277		
FCGR3B				-0.04382	-0.047063			0.050848563
FKBP11	0.0833244	0.078857		0.078018	0.069366		0.075718	0.049317294
FOXI1	0.0456297							0.049750788
FUT4	0.0546642	0.07476	0.050612	0.04718		0.054029		
GATA2						-0.044676		
GJA5					-0.058265			
GYPA							0.055007	
GZMK	0.1124151	0.11247	0.113746	0.090556	0.067483	0.12196	0.077234	0.116377053
ICOS					0.043273			
ICOSLG	0.0890651	0.080021	0.100679		0.066888			
IDO1					-0.04533			
IFIH1			-0.042804					
IFITM2	-0.0499703	-0.066448				-0.066731	-0.057621	-0.09603561
IFITM3		-0.111749		-0.052976	-0.138866	-0.102936	-0.07538	-0.12173662
IFNG	0.0564185	0.048542	0.066231	0.058263				
IFNGR1	-0.068489	-0.075191	-0.072411	-0.046442	-0.063603	-0.043599		-0.07228366
IFNGR2		-0.053552						
IGKC	0.1358517	0.110124	0.130462					
IL10RB						0.043795	0.050399	
IL17A	0.0427097							
IL1A	0.0579209	0.045525	0.048089			0.052517		
IL1B					-0.042758	-0.046179		
IL1R1						-0.045698	-0.066127	-0.04663377

IL2			0.043067					
IL22		0.046282						
IL27			0.045304					
IL2RA	0.047885					-0.048032		
IL33								-0.04506483
IL3RA						-0.042991		
IL4		0.046986				0.052446		
IL6			-0.0546					-0.05910199
IL6R						-0.042505		
IL6ST	-0.0570582	-0.070271	-0.098127			-0.066205		
INMT	-0.1212277	-0.066216	-0.147888				-0.053286	
ISG15	0.0567547							0.053995743
ITLN1								0.056264167
JCHAIN	0.0478496	0.045442	0.054362					-0.04797501
KIT	-0.0677552	-0.08017	-0.05075					-0.04424782
KLRC2								0.045877299
KLRD1					-0.04341			
KRT13						-0.048733		0.045118996
KRT15						-0.04434		
KRT19						-0.043261		
LEF1							0.042944	
LGR6							0.044754	
LST1	0.11908	0.137493	0.077651	0.07928		0.073122	0.047908	0.08865345
LTF						0.047044		
MGP	-0.0431061	-0.054076						-0.07369367
MKI67								0.052051356
MRC1				-0.048254				
MS4A2								0.092534008
MSLN			-0.04623					-0.04534354
MUC1						-0.057608		0.045367682
MUC5B	0.0802862	0.056127						
NCAM1	0.1151786	0.04776	0.103537	0.097938	0.077103			0.090689078
NKG7					-0.068881			
NRP1	-0.0589314	-0.070173	-0.050651	-0.061653	-0.08547		-0.050626	
PCSK1N				0.043314		0.04325		-0.06915208
PDCD1				0.043172	0.045251			
PECAM1	-0.0425713	-0.084631	-0.04948					
PF4					-0.049737			
PLAC8					0.045747			
PPARG					0.073269			
PRG4	-0.0611091		-0.080717	-0.065106				-0.06883645
PTN					-0.046331			
PTPRC								-0.09296068
RETN			0.067828			0.056703		
S100A8			0.059619				-0.046239	

SCGB3A2							-0.043726	
SFTA2	0.0711606							
SFTPA1	-0.0912305		-0.063382	-0.05158				
SFTPA2			-0.050218		-0.044033		-0.047493	
SFTPC	-0.1716209	-0.117223	-0.088311	-0.104805	-0.057423	-0.082973	-0.097653	-0.11385289
SLC18A2			0.074543			-0.04352		
TAGLN			-0.04517					
TBX21							0.07757	
TGFB1		-0.048865	-0.050631	-0.043419			-0.056812	
TMPRSS2	0.0450597	0.063989		0.058446	0.105499		0.055832	
TPPP3								-0.04383628
TPSAB1								-0.28480969
TUBB4B			-0.04293					
TYROBP				-0.132481			-0.074625	-0.06259818
VIM			0.047509					
WT1							0.047543	

Table S7. Spearman Correlations of Pseudotime Trajectory B versus Gene Expressions

Gene	ACs	VECs	Fibroblasts	MMs	NK cells	T cells	B cells	Granulocytes
A2M	-0.23417	-0.18242	-0.18695	-0.11699	-0.07008	-0.13915	-0.11668	-0.199064005
ACE2						0.047165		
ACKR1		0.048854					-0.04833	
ACTA2	-0.09122	-0.10498	-0.09113			-0.04457		
AGER	-0.0654	-0.12614	-0.04869					
AGRP	0.049945							
AP1S2	-0.04246	-0.07303					-0.04535	
APOE				-0.07856		-0.03913		
ARG1	0.047952					0.046863		
ASCL1								-0.037819398
BATF3						0.037658		
C1QA	0.058691							0.09027356
C1QB	0.117966	0.054073	0.089102	0.096327			0.058879	0.125662283
C3								-0.063321455
CAV1	-0.15927	-0.18698	-0.14654	-0.07419	-0.07115	-0.05335	-0.04429	-0.103846221
CCL2						0.056462		0.041291612
CCR2	0.06541	0.063104		0.046706		0.038923		0.045579992
CCR4		0.05497	0.03879					
CCR5	0.06756			0.037975	0.051752			
CCR6					0.045262			
CD163								0.037295985
CD19	0.060925	0.040203	0.041826				0.044888	
CD34	-0.0689	-0.09669	-0.06878		-0.0509			
CD4	0.049011							
CD40	0.039212						0.058188	
CD40LG	0.107726	0.077691		0.045761		0.08453		0.068746563
CD44	-0.05332				0.052098			-0.050139472
CD68				-0.07668				
CD7						0.052587		
CD79A	0.08843		0.042754				0.092262	
CD79B							-0.12707	
CD83								0.047211302
CD8A							0.038345	
CEACAM8	0.067966						0.055024	0.214210161
CFTR							0.054296	
CLDN5	-0.0661	-0.08001	-0.08032			-0.1043	-0.05523	-0.079253269
CLEC10A	0.068948	0.065709	0.055969	0.057083		0.074607		
COL1A1			0.238038		0.050449		0.063025	
COL1A2	-0.08054		0.074501				0.060741	-0.043885551
CPA3						-0.0388		-0.058816528
CSF3	0.113034	0.10195	0.072741	0.102288	0.071468		0.088778	0.098683964
CTSB				-0.04035				0.055171045
CTSL	0.104649	0.07678	-0.04173				-0.06617	

CX3CR1					0.043399		-0.04299	
CXCL2	-0.07253	-0.0844	-0.05969	-0.05841	-0.06559	-0.05153	-0.04755	
CXCL5								0.03806835
CXCL9						-0.03978		
CXCR3	0.069483	0.048237			0.037495			0.085780591
CXCR5	0.125108	0.091302	0.039747	0.064412			0.053469	0.056036
CXCR6	0.056839	0.055155	0.063929					0.057363906
DES						0.055714	0.041941	
EDN1	-0.04828		-0.04554					
EOMES	0.084288	0.076171		0.045833				0.050475746
FABP4				-0.03939				
FBLN1	-0.07795	-0.06019	-0.0729	-0.04823			-0.06675	-0.099788911
FCGR2A		0.044455						
FCGR3B	0.040975	0.037913						0.072317848
FKBP11	0.063241	0.040355	0.056659					
FOLR2						-0.04681		
FOXI1	0.045565							
FOXN4						0.043416	0.043758	
FOXP3	0.039537		0.050424		0.048612		0.060758	0.056544832
FURIN							0.045908	
FUT4	0.08674	0.072236	0.040281	0.06246		0.044159		
GATA2						-0.04118		
GJA5	0.042105	0.043181						
GNLY	0.049135						0.039547	
GPR183							-0.05082	
GYPA		0.044047						
GZMK	0.162548	0.117448	0.066982	0.07485		0.102471	0.053519	0.121910499
HPGDS								-0.058750952
ICAM1							0.051938	
ICOSLG	0.094865	0.108561	0.059662	0.058779	0.06085	0.041286	0.046443	0.048873673
IDO1		0.039485				0.037285		
IFITM3								-0.0401819
IFNB1					0.064311			
IFNG		0.047643			-0.03724			0.04148565
IFNGR1	-0.06606	-0.06479	-0.04459		-0.06533	-0.0605		-0.043583766
IFNGR2		-0.04789						0.041613098
IGKC	0.200164	0.207234	0.10137	0.08933	0.115094	0.106026	0.061718	0.065923338
IGLL5					0.045037			
IL10RB			0.04243		0.042023	0.043005		
IL17A	0.051116	0.052451					0.049536	
IL1A	0.061588	0.050096						
IL1B		0.038184					0.038756	
IL22							0.041126	0.037526937
IL27	0.040339							
IL2RA	0.073001				0.053015		0.054686	

IL33			-0.03945				-0.03722	
IL3RA							0.043497	
IL4	0.03914							
IL5						0.065467	0.045197	
IL6								-0.053733035
IL6R							0.044341	
IL6ST	-0.05528							
IL7R		-0.04831				-0.05257		
INMT	-0.12332	-0.06816	-0.19792				-0.06262	
ISG15	0.066814	0.038402	0.040259					
ITLN1	0.044288							0.049348458
JCHAIN	0.082298	0.052541			0.063171	0.058752		
KIT	-0.05272	-0.08017						-0.053323915
KLRC2					-0.04922		0.050395	
KLRD1							0.044409	
KRT13								0.041034312
KRT14	0.040772							
KRT15	0.046743							
LST1	0.176173	0.140293	0.075228	0.097155	0.048409	0.150042	0.066195	0.117086543
LTF						0.058064		
MGP					0.060665			-0.047453011
MRC1					0.039171			
MSLN							-0.04002	-0.053807806
MUC5AC							-0.04623	
MUC5B	0.100919	0.037387	0.038258			0.037932		
MX1	0.037898							
NCAM1	0.129607	0.038924	0.070591	0.060254	0.040041			0.054425241
NRP1	-0.0381	-0.04668		-0.05298				
PAX5								0.053371989
PCSK1N								-0.076337283
PDCD1			0.037544					
PECAM1	-0.04893	-0.09513	-0.05137	-0.04777	-0.06127			
PLAC8	0.039513							
PLIN2		0.049445						
PPARG	0.037847							
PRF1							-0.04373	
PRG4	-0.06815	-0.04201	-0.07122	-0.06506				
PTN			-0.03789		-0.06056			
PTPRC								-0.05500837
RETN	0.066851	0.040041						
S100A8	0.048937							
SEMA3A					0.059976			
SFTA2	0.090868							
SFTPA1	-0.07088	-0.05154						
SFTPA2							-0.04342	

SFTPC	-0.18456	-0.09881	-0.07617	-0.07591	-0.05003	-0.06539	-0.06338	-0.11238912
SLC18A2	0.059352							0.053280437
TGFB1		-0.0503						0.042625481
TMPRSS2						-0.04518		-0.07145939
TPPP3								-0.038857276
TPSAB1		-0.04459	-0.04399			0.038693	-0.04669	-0.291267486
TYROBP				-0.04596				