nature

Accelerated Article Preview

A molecular single-cell lung atlas of lethal COVID-19

Johannes C. Melms, Jana Biermann, Huachao Huang, Yiping Wang, Ajay Nair, Somnath Tagore, Igor Katsyv, André F. Rendeiro, Amit Dipak Amin, Denis Schapiro, Chris J. Frangieh, Adrienne M. Luoma, Aveline Filliol, Yinshan Fang, Hiranmayi Ravichandran, Mariano G. Clausi, George A. Alba, Meri Rogava, Sean W. Chen, Patricia Ho, Daniel T. Montoro, Adam E. Kornberg, Arnold S. Han, Mathieu F. Bakhoum, Niroshana Anandasabapathy, Mayte Suárez-Fariñas, Samuel F. Bakhoum, Yaron Bram, Alain Borczuk, Xinzheng V. Guo, Jay H. Lefkowitch, Charles Marboe, Stephen M. Lagana, Armando Del Portillo, Emmanuel Zorn, Glen S. Markowitz, Robert F. Schwabe, Robert E. Schwartz, Olivier Elemento, Anjali Saqi, Hanina Hibshoosh, Jianwen Que & Benjamin Izar

This is a PDF file of a peer-reviewed paper that has been accepted for publication. Although unedited, the content has been subjected to preliminary formatting. Nature is providing this early version of the typeset paper as a service to our authors and readers. The text and figures will undergo copyediting and a proof review before the paper is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers apply.

Article

A molecular single-cell lung atlas of lethal COVID-19

Johannes C. Melms[1,2,29], Jana Biermann[1,2,29], Huachao Huang[3,4,29], Yiping Wang[1,2,29], Ajay Nair[4,29], Somnath Tagore[5,29], Igor Katsyv[6,29], André F. Rendeiro[7,8,29], Amit Dipak Amin[1,2,29], Denis Schapiro[9,10], Chris J. Frangieh[10,11], Adrienne M. Luoma[12], Aveline Filliol[4], Yinshan Fang[3,4], Hiranmayi Ravichandran[8,13,14], Mariano G. Clausi[15], George A. Alba[16], Meri Rogava[1,2], Sean W. Chen[1,2], Patricia Ho[1,2], Daniel T. Montoro[17,18], Adam E. Kornberg[2], Arnold S. Han[2], Mathieu F. Bakhoum[19], Niroshana Anandasabapathy[8,20], Mayte Suárez-Fariñas[21], Samuel F. Bakhoum[22], Yaron Bram[23], Alain Borczuk[24,25], Xinzheng V. Guo[15], Jay H. Lefkowitch[6], Charles Marboe[6], Stephen M. Lagana[6], Armando Del Portillo[6], Emmanuel Zorn[2], Glen S. Markowitz[6], Robert F. Schwabe[4,26], Robert E. Schwartz[23,30], Olivier Elemento[7,8,14,30], Anjali Saqi[6,30], Hanina Hibshoosh[6,30], Jianwen Que[3,4,27,30 ✉] & Benjamin Izar[1,2,27,28,30 ✉]

Respiratory failure is the leading cause of death in patients with severe SARS-CoV-2 infection[1,2], yet the host response at the lung tissue-level is poorly understood. Here, we performed single-nucleus RNA-sequencing of ~116,000 nuclei of lungs from 19 COVID-19 decedents who underwent rapid autopsy and 7 control lungs. Integrated analyses revealed significant alterations in cellular composition, transcriptional cell states, and cell-to-cell interactions, providing insights into the biology of lethal COVID-19. COVID-19 lungs were highly inflamed with dense infiltration of aberrantly activated monocyte-derived macrophages and alveolar macrophages, but demonstrated impaired T cell responses. Monocyte/macrophage-derived IL-1β and epithelial cell-derived IL-6 were unique features of SARS-CoV-2 infection compared to other viral and bacterial causes of pneumonia. Alveolar type 2 cells adopted an inflammation-associated transient progenitor cell state and failed to undergo full transition into alveolar type 1 cells resulting in impaired lung regeneration. Furthermore, we identified expansion of recently described *CTHRC1*[+] pathological fibroblasts[3] contributing to rapidly ensuing pulmonary fibrosis in COVID-19. Inference of protein activity and ligand-receptor interactions identified putative drug targets to disrupt deleterious circuits. This atlas enables dissection of lethal COVID-19, may inform our understanding of long-term complications of COVID-19 survivors, and provides an important resource for therapeutic development.

Globally, the novel coronavirus disease 2019 (COVID-19) pandemic, resulting from infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has led to more than 131 million cases (31 million in the US) and 2.85 million deaths (550,000 in the US) to date[1]. Approximately 15% of infected individuals develop severe disease which can manifest as acute respiratory distress syndrome (ARDS) and is associated with substantial morbidity and mortality[2,4].

[1]Department of Medicine, Division of Hematology/Oncology, Columbia University Irving Medical Center, New York, NY, USA. [2]Columbia Center for Translational Immunology, New York, NY, USA. [3]Columbia Center for Human Development and Division of Digestive and Liver Diseases, New York, NY, USA. [4]Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA. [5]Department of Systems Biology, Columbia University Irving Medical Center, New York, NY, USA. [6]Department of Pathology and Cell Biology, Columbia University Irving Medical Center, New York, NY, USA. [7]Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY, USA. [8]Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY, USA. [9]Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA, USA. [10]Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [11]Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, MA, USA. [12]Department of Cancer Immunology and Virology, Dana-Farber Cancer Center, Boston, MA, USA. [13]Department of Physiology and Biophysics, Weill Cornell Medical College, New York, NY, USA. [14]WorldQuant Initiative for Quantitative Prediction, New York, USA. [15]Human Immune Monitoring Core, Columbia University Irving Medical Center, New York, NY, USA. [16]Department of Medicine, Division of Pulmonary and Critical Care, Massachusetts General Hospital, Boston, MA, USA. [17]Cell Circuits, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [18]Systems Biology, Harvard Medical School, Boston, MA, USA. [19]Department of Ophthalmology, University of California San Diego, La Jolla, CA, USA. [20]Department of Dermatology and Meyer Cancer Center, Weill Cornell Medical College, New York, NY, USA. [21]Department of Genetics and Genomic Science and Department of Population Health Science and Policy, Icahn School of medicine at Mount Sinai, New York, NY, USA. [22]Human Oncology and Pathogenesis Program and Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [23]Division of Gastroenterology and Hepatology, Department of Medicine, Weill Cornell Medicine, New York, NY, USA. [24]Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY, USA. [25]Department of Medicine, Weill Cornell Medicine, New York, NY, USA. [26]Institute of Human Nutrition, Columbia University, New York, NY, USA. [27]Herbert Irving Comprehensive Cancer Center, New York, NY, USA. [28]Program for Mathematical Genomics, Columbia University, New York, NY, USA. [29]These authors contributed equally: Johannes C. Melms, Jana Biermann, Huachao Huang, Yiping Wang, Ajay Nair, Somnath Tagore, Igor Katsyv, André F. Rendeiro, Amit Dipak Amin. [30]These authors jointly supervised this work: Robert E. Schwartz, Olivier Elemento, Anjali Saqi, Hanina Hibshoosh, Jianwen Que, Benjamin Izar. ✉e-mail: jq2240@cumc.columbia.edu; bi2175@cumc.columbia.edu

Article

Previously, single-cell RNA-sequencing (scRNA-seq) analyses of healthy individuals revealed the tissue distribution of host receptors required for SARS-CoV-2 entry[5–7], and examination of bronchoalveolar lavage fluid and blood from patients with varying COVID-19 disease severity revealed the impact of SARS-CoV-2 infection on immune responses and cytokine dysregulation[8–12]. However, due to the practical limitations of accessing patient tissues, effects at a lung tissue-level remained elusive. A series of autopsy studies examining formalin-fixed, paraffin-embedded (FFPE) tissue sections of COVID-19 decedents extended our understanding of organotropism, but these studies were limited in their discovery potential by low-plex assays (e.g. immunohistochemistry) and/or prolonged post-mortem intervals (PMI) which adversely affect RNA quality[13–15].

Here, we established a rapid autopsy program, and under Institutional Review Board approved protocols, collected snap-frozen organ specimens from COVID-19 decedents within hours of death. We performed single-nucleus RNA-sequencing (snRNA-seq) on lung tissues from COVID-19 decedents and controls to build an atlas that enables insights into COVID-19 pathophysiology and provides a critical resource for further investigation.

The lung cellular landscape in COVID-19

The COVID-19 cohort consisted of 19 patients, including 12 males and 7 females, who died at a median age of 72 years (58 to >89) (Supplementary Table (ST)1, Extended Data Figure (ED)1a) and underwent rapid autopsy with a median post-mortem interval (PMI) of 4 hours (range, 2–9 hours). All COVID-19 decedents had underlying co-morbidities associated with an increased risk for severe COVID-19[16] (ST1). The control cohort included 7 individuals, 4 males and 3 females, with a median age of 70 years (range, 67 to 79 years), who underwent lung resection/biopsy in the pre-COVID-19 era (ST1).

Using snRNA-seq[17] and an integrated quality control pipeline (**Methods**), we generated a lung atlas profiling 116,314 nuclei, including 79,636 from COVID-19 and 36,678 from control lungs (**Figure** (Fig. 1a)). We used a three-pronged approach for cell type identification, including unbiased identification of cluster markers, discovery of cell types using signatures from reported atlases, and manual curation to sub-stratify cell populations and cell states using expert knowledge (**Methods**). We report cell type assignment with three levels of granularity: major cell types, intermediate, and fine granularity (ST2). We visualized data with dimensionality reduction using uniform manifold approximation and projection (UMAP) (Fig. 1b, c, ED1b-d). We identified 9 major cell types, including epithelial cells (n=30,070 cells), myeloid cells (n=29,632), fibroblasts (n=22,909), endothelial cells (n=5,386), T and Natural Killer (NK) lymphocytes (n=16,751), B lymphocytes and plasma cells (n=7,236), neuronal cells (n=2,017), mast cells (n=1,464), and antigen-presenting cells (APCs, primarily dendritic cells, DCs) (n=849). At the most granular level, we identified 41 different cell types (ST2).

We found significant differences in cell fractions between COVID-19 and control lungs both globally (Fig. 1d) and within the immune and non-immune compartments (ED2a-c). There was a reduction in the epithelial cell compartment, due to loss of both alveolar type II (AT2) and type I (AT1) cells, and an increase in monocytes/macrophages, fibroblasts, and neuronal cells; these observations were independent of donor sex (ED3a,b).

We found no major differences in the expression of *ACE2*, *CD147 (BSG)*, *NPR1*, *TMPRSS2*, *FURIN*, and *CTSL* between COVID-19 and control lungs (ED3c-f). This indicates that changes in cell type proportions were unrelated to receptor expression or putative proteases important for viral entry, although we cannot exclude the possibility that virus-mediated cell death selectively depletes cells with high expression of these genes. We detected SARS-CoV-2 reads in two patients (ST3), including one with human immunodeficiency virus/acquired immunodeficiency syndrome (CD4+ T cell count 29/mm³ on hospital admission; 662 unique

molecular identifiers detected in 28 cells), suggesting that viral reads can, in principle, be captured.

Aberrant activation of myeloid cells

Myeloid cells represented a major cellular constituent in COVID-19 lungs and were more prevalent than in controls (Fig. 1d, ED2a,c, ED4a). We identified monocytes (n=3,176), monocyte-derived macrophages (MDMs) (n=9,534), transitioning MDMs (n=4,203), and resident alveolar macrophages (AM) (n=12,511), which were recovered as distinct trajectories in diffusion component (DC) analysis and were more frequent in COVID-19 (Fig. 2a-c, ED4b-i, ST2,4,5). Myeloid cells from COVID-19 decedents were highly and aberrantly activated. For example, MDMs from COVID-19 decedents differentially expressed genes of activation (e.g., *CTSB*, *CTSD*, *CTSZ*, *PSAP*) and two long non-coding RNAs, *NEAT1* and *MALAT1*, associated with aberrant macrophage activation and impaired T cell immunity[18] (ED5a, ST5). AMs, which arise from foetal monocytes and have the capacity for self-renewal[19], were enriched and highly activated in COVID-19 lungs (Fig. 2c, ED5a). Notably, COVID-19 AMs showed strongly decreased mRNA and protein expression of the tumor-associated macrophage receptor *AXL* (Fig. 2d, ED5b,c), a receptor tyrosine kinase important for coordinated clearance of apoptotic cells (efferocytosis) and subsequent anti-inflammatory regulation during tissue regeneration[20]. These data suggest that myeloid cells are a major source of dysregulated inflammation in COVID-19.

Plasma and T cell responses

To gain insights into humoral immunity against SARS-CoV-2 infection in the lung, we identified plasma cells (ED6a-c) and reconstructed immunoglobulins by determining mRNA co-expression of the variable heavy (*IGHV*) and light (*IGLV*) chains and isotypes on a per cell basis (**Methods**) (ED6d-k, ST6). IGHV1-18/IGLV3-20, which gives rise to a neutralizing antibody (S309)[21] against the receptor binding domain (RBD) of SARS-CoV-2 spike protein, was among the commonly identified IGHV/IGLV combinations, suggesting that a coordinated antibody response occurred (Fig. 2e, f, ED6l,m). In the T/NK cell compartment (Fig. 2g), we distinguished CD8+ T cells (n=3,561), T regulatory cells (Tregs) (n=649), other CD4+ T cells (n=7,586), and NK cells (n=2,141). We found no significant increase in T cell abundances in COVID-19 and only modest upregulation of cytokines and programs associated with activation and tissue residency (Fig. 2g–i, ED7a-i). While immune response patterns were highly variable (ED7j,k), these data suggest that an impaired T cell response may contribute to lethal outcomes in COVID-19 in the context of a principally preserved humoral immune response.

Impaired alveolar epithelial regeneration

Within the epithelial compartment, we identified alveolar epithelial cells (AT1 and AT2 cells) (n=20,949), airway epithelial cells (basal, ciliated, club, goblet, and mucous cells) (n=7,223), a cluster characterised by expression of inflammatory and cell cycle genes, including *IRF8*, *B2M*, *MKI67* and *TOP2A* ("cycling epithelium") (n=609), and a cluster showing high expression of extracellular matrix (ECM) components *COL6A3*, *COL1A2*, *and COL3A1* ("ECM^high epithelium") (n=1,179) (Fig. 3a, b, ED8a-c, ST2,7).

AT2 cells serve as progenitors for AT1 cells during lung regeneration[22]. AT2 and AT1 cells in control lungs formed distinct clusters (Fig. 3a, b) and demonstrated expected changes in differential gene expression (DGE) analysis, including lineage markers *SFTPC* and *SFTPB* in AT2 cells, and *CLIC5* and *AGER* in AT1 cells (Fig. 3c, ST7). In contrast, clustering of AT2 and AT1 cells in COVID-19 decedents was less discrete, with a significant portion of cells not overlapping with their control counterparts (Fig. 3b). We found that both AT2 and AT1 cells from COVID-19

decedents showed decreased overall expression of defining markers (Fig. 3c). COVID-19 AT2 cells displayed decreased expression of *ETV5* (Fig. 3d), a transcription factor required for maintaining AT2 cell identity. Decreased *ETV5* expression is associated with differentiation towards AT1 cells[23], indicating that AT2 cells initiated a regeneration program (Fig. 3d, ED8d). *CAV1*, a marker of late AT1 maturation[24], was expressed at significantly lower levels in AT1 cells from COVID-19 lungs (Fig. 3e). Overall, these data suggest an incomplete transition of AT2 to AT1 cells in COVID-19.

Recent studies showed that inflammation may induce a cell state characterised by failure to fully transition to AT1 cells termed "damage-associated transient progenitors" (DATP), "alveolar differentiation intermediate" (ADI), or "pre-AT1 transitional cell state" (PATS)[25–27] (hereafter summarised as DATP). Based on expression of the DATP marker genes (*KRT8/CLDN4/CDKN1A*)[25], we first developed a DATP signature (**Methods)** (ED8e-h, ST8) and found that alveolar epithelial cells from COVID-19 decedents scored significantly higher than those from controls (Fig. 3f, g, ED8i). DC analysis separated a main trajectory from AT2 to AT1 cells, while DATPs were primarily localized between AT2 and AT1 cells (Fig. 3h, ED8j-n). Gene set enrichment analysis (GSEA) of DATP compared to differentiated AT2/AT1 cells showed enrichment for TNF-α and p53 signalling and hypoxia response via HIF-1α (ED8o), consistent with pathways implicated in DATP in murine models[27]. Consistent with overrepresentation of p53 signalling, the majority of DATPs did not undergo cell division (ED8p), suggesting that they arrest in the DATP cell state.

DATPs were more frequent in COVID-19 than control lungs (Fig. 3i). Immunofluorescence (IF) staining of corresponding tissues validated an increased frequency of KRT8+ and CLDN4+ DATPs in COVID-19 lungs (Fig. 3j, ED8r-s), and we observed progressive loss of AT1 cell abundance over time (ED8t). Overall, these data suggest that in addition to direct destruction of the alveolar epithelium by viral infection, lung-regenerative processes are impaired in COVID-19 decedents.

We next determined the sources of inflammation that contribute to the DATP cell state, and more generally, to the hyperinflammatory environment in COVID-19 lungs. Capture of the inflammatory cytokine IL-1β (and others) at an mRNA level may be limited, as the bioactive form of IL-1β, which plays a major role in triggering DATPs[25], is generated by cleavage from a pro- IL-1β upon inflammasome activation; thus, protein-level assessment provides complementary information. For this purpose, we leveraged a recently released high-plex imaging mass-cytometry data set profiling 237 tissue regions from 23 individuals, including healthy controls; patients with influenza pneumonia, bacterial pneumonia, or ARDS; and 10 patients who died from COVID-19[28]. IL-1β was more strongly expressed in monocytes and macrophages of COVID-19 patients compared to healthy individuals and patients in the other disease groups (Fig. 3k, ED9a-c). IL-6, another key inflammatory cytokine invoked in the pathophysiology of COVID-19, was more abundant in epithelial cells from COVID-19 patients, but was not differentially expressed among macrophages compared to healthy controls and other disease groups (Fig. 3l, ED9d-f). Finally, we found strong expression of type I interferons and interferon response genes among various cell types, including AT2 cells, monocytes, and macrophages in COVID-19 compared to control donors (ED9g,h). Together, these data suggest that myeloid-derived IL-1β may be a distinguishing feature of COVID-19 compared to other viral or bacterial pneumonias and may contribute to the induction and maintenance of the DATP cell state.

Ectopic tuft-like cells in COVID-19

Among captured airway epithelial cells, we recovered four distinct trajectories including *KRT5*+*TP63*+ basal (n=534), club (n=1,232), and goblet cells (n=1,757), and one with fewer cells (n=110), primarily encountered in COVID-19 decedents, which we identify as putative tuft-like cells (ED10a-e). Tuft cells are involved in airway inflammation

and intestinal tissue regeneration[29], but their role in viral pneumonia remains unclear. The numbers of tuft cells (ChAT+ or POU2F3+) were increased 3-fold in the upper airways, and they were ectopically present in the lung parenchyma of COVID-19 but not control lungs (ED10f-k). To begin elucidating a putative role of tuft cells in viral pneumonia, we infected both wildtype and *Pou2f3−/-* mice, which lack tuft cells, with PR8, a laboratory adapted H1N1 strain (**Methods**). Compared to controls, *Pou2f3−/-* lungs demonstrated decreased infiltration of macrophages and decreased expression of chemotaxis genes (including *Ccl3* and *Ccl8*) that are also involved in myeloid cell recruitment to the lungs of COVID-19 decedents (ED9g,h, ED11a-l**)**. While their role needs to be further examined, we detected ectopic tuft-like cells that may contribute to the pathophysiology of COVID-19.

Pathologial fibroblasts and lung fibrosis

We observed a significant increase in the frequency of fibroblasts in COVID-19 lungs compared to controls (Fig. 1d) and validated this finding using immunohistochemistry staining for α-smooth muscle actin (α-SMA) (ED12a-d). The degree of fibrosis (determined by a Sirius red fibrosis score, **Methods**) was correlated with disease duration (Fig. 4a), indicating ensuing lung fibrosis over time in COVID-19. We identified five different fibroblast subtypes: alveolar (n=4,670), adventitial (n=3,773), pathological (n=2,322), intermediate pathological (n=8,779), and other fibroblasts (n=1,099) (Fig. 4b, ED12e). The major driver of differences in the fibroblast cluster was the increased frequency of pathological/intermediate pathological fibroblasts (henceforth summarised as pFB) in COVID-19 decedents compared to controls (Fig. 4c, ED12f). pFB strongly expressed *CTHRC1*, a recently described hallmark gene defining these cells, and genes of pathologic extracellular matrix (ECM), including *COL1A1* and *COL3A1* (ED12e, ST9)[3]. pFB are critical drivers of lung fibrosis in murine models and in patients with idiopathic pulmonary fibrosis (IPF) and scleroderma[3]. Their increased frequency suggests a role of pFB in promoting rapidly evolving fibrosis in COVID-19.

Given the importance of fibroblasts in remodeling of the lung ecosystem, we next explored ligand-receptor interactions across all major cell types, including fibroblasts (**Methods**). Among the enriched inferred ligand-receptor interactions across all cells were TGF-β1:TGF-β receptor 2 and BMP6:ACVR1 (ED12g-i, ST10), which belong to the TGF-β family and superfamily, respectively. TGF-β signaling plays an important role in promoting lung fibrosis and has been implicated in fibroblast-mediated maintenance of the ADI[27], which is closely related to the DATP cell state. To explore potential therapeutic strategies directed against pFB, we inferred protein activity from single-nucleus transcriptomes followed by comparison of pFB with other fibroblasts. This analysis predicted increased activities of JunB and JunD in pFB (ED12j, ST11), which induce lung fibrosis in murine models via enhanced TGF-β and STAT3 signalling and are associated with increased production of IL-1β[30]. Finally, we inferred druggable targets in pFB (**Methods**) and identified MMP14 and STAT3 as potential targets to abrogate detrimental programs in pFB (ED12j, ST11).

Discussion

We generated a single-cell transcriptome lung atlas of COVID-19 using short PMI autopsy specimens and control lungs. Our analysis provides a broad census of the cellular landscape, cell programs, and cell circuits of lethal COVID-19. The additional inference of protein activity and cell-to-cell interactions, and analysis of inflammatory cytokines across various cell types using imaging mass cytometry data, provides a granular perspective of the detrimental consequences of SARS-CoV-2 infection in the lung.

Our analyses suggest interplay among aberrantly activated monocytes/macrophages producing IL-1β, inflammation-induced

Article

impairment of alveolar epithelial regeneration, and expansion of pathological fibroblasts that promote fibrosis and may impair regeneration (ED12f,k, Supplementary Discussion). In addition to these deleterious events, our data suggest that despite a potentially sufficient humoral immune response (Supplementary Discussion), there was an inadequate T cell response. A recent study in cancer patients showed that impaired B cell function was not associated with increased mortality[31], however lack of an adequate CD8+ T cell response (even in the presence of adequate humoral immunity) was associated with worse viral control and increased mortality[31]. Although our COVID-19 cohort did not include cancer patients, these data suggest that while humoral immunity may be dispensable in the context of adequate T cell immunity against SARS-CoV-2, lack of appropriate T cell responses in our patients likely contributed to fatal outcomes.

Although our study yields important insights into host responses to lethal SARS-CoV-2 infection, it is limited by small sample size. However, through coordinated efforts, our work will contribute to a collection of studies, including the companion paper by *Delorey et al.*, with streamlined protocols and harmonised meta-data to enable integration and combined analyses, and will help to account for important co-variables. Furthermore, because our analysis is focussed on lung tissues from patients who died of COVID-19, we examine only a subset of potential disease phenotypes. Nonetheless, several observations, such as the rapid development of pulmonary fibrosis (Supplementary Discussion), are likely relevant for patients who survive severe COVID-19, and may inform our understanding of long-term complications seen in these individuals[32].

In conclusion, we have generated a molecular single-cell lung atlas from short PMI tissue specimens and uncovered pathological circuits of lethal COVID-19. This atlas establishes an important resource to investigate host responses to SARS-CoV-2, understand potential long-term pulmonary sequelae resulting from COVID-19 and provides a basis for therapeutic development for severe disease.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03569-1.

1. Dong, Y. *et al.* Epidemiology of COVID-19 Among Children in China. *Pediatrics*, https://doi.org/10.1542/peds.2020-0702 (2020).
2. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* **395**, 1054-1062, https://doi.org/10.1016/S0140-6736(20)30566-3 (2020).
3. Tsukui, T. *et al.* Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis. *Nat Commun* **11**, 1920, https://doi.org/10.1038/s41467-020-15647-5 (2020).
4. Bellani, G. *et al.* Epidemiology, Patterns of Care, and Mortality for Patients With Acute Respiratory Distress Syndrome in Intensive Care Units in 50 Countries. *JAMA* **315**, 788-800, https://doi.org/10.1001/jama.2016.0291 (2016).
5. Muus, C. *et al.* Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics. *Nat Med* **27**, 546-559, https://doi.org/10.1038/s41591-020-01227-z (2021).
6. Sungnak, W. *et al.* SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat Med* **26**, 681-687, https://doi.org/10.1038/s41591-020-0868-6 (2020).
7. Ziegler, C. G. K. *et al.* SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. *Cell*, https://doi.org/10.1016/j.cell.2020.04.035 (2020).
8. Schulte-Schrepping, J. *et al.* Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment. *Cell* **182**, 1419-1440 e1423, https://doi.org/10.1016/j.cell.2020.08.001 (2020).
9. Wilk, A. J. *et al.* A single-cell atlas of the peripheral immune response in patients with severe COVID-19. *Nat Med* **26**, 1070-1076, https://doi.org/10.1038/s41591-020-0944-y (2020).
10. Xu, G. *et al.* The differential immune responses to COVID-19 in peripheral and lung revealed by single-cell RNA sequencing. *Cell Discov* **6**, 73, https://doi.org/10.1038/s41421-020-00225-2 (2020).
11. Hadjadj, J. *et al.* Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **369**, 718-724, https://doi.org/10.1126/science.abc6027 (2020).
12. Blanco-Melo, D. *et al.* Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*, https://doi.org/10.1016/j.cell.2020.04.026 (2020).
13. Ackermann, M. *et al.* Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. *N Engl J Med* **383**, 120-128, https://doi.org/10.1056/NEJMoa2015432 (2020).
14. Puelles, V. G. *et al.* Multiorgan and Renal Tropism of SARS-CoV-2. *N Engl J Med*, https://doi.org/10.1056/NEJMc2011400 (2020).
15. De Michele, S. *et al.* Forty Postmortem Examinations in COVID-19 Patients. *Am J Clin Pathol* **154**, 748-760, https://doi.org/10.1093/ajcp/aqaa156 (2020).
16. Goyal, P. *et al.* Clinical Characteristics of Covid-19 in New York City. *N Engl J Med* **382**, 2372-2374, https://doi.org/10.1056/NEJMc2010419 (2020).
17. Slyper, M. *et al.* A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat Med* **26**, 792-802, https://doi.org/10.1038/s41591-020-0844-1 (2020).
18. Hewitson, J. P. *et al.* Malat1 Suppresses Immunity to Infection through Promoting Expression of Maf and IL-10 in Th Cells. *J Immunol* **204**, 2949-2960, https://doi.org/10.4049/jimmunol.1900940 (2020).
19. Hoeffel, G. *et al.* C-Myb(+) erythro-myeloid progenitor-derived fetal monocytes give rise to adult tissue-resident macrophages. *Immunity* **42**, 665-678, https://doi.org/10.1016/j.immuni.2015.03.011 (2015).
20. Doran, A. C., Yurdagul, A., Jr. & Tabas, I. Efferocytosis in health and disease. *Nat Rev Immunol* **20**, 254-267, https://doi.org/10.1038/s41577-019-0240-6 (2020).
21. Pinto, D. *et al.* Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature*, https://doi.org/10.1038/s41586-020-2349-y (2020).
22. Barkauskas, C. E. *et al.* Type 2 alveolar cells are stem cells in adult lung. *J Clin Invest* **123**, 3025-3036, https://doi.org/10.1172/JCI68782 (2013).
23. Zhang, Z. *et al.* Transcription factor Etv5 is essential for the maintenance of alveolar type II cells. *Proc Natl Acad Sci U S A* **114**, 3903-3908, https://doi.org/10.1073/pnas.1621177114 (2017).
24. Little, D. R. *et al.* Transcriptional control of lung alveolar type 1 cell development and maintenance by NK homeobox 2-1. *Proc Natl Acad Sci U S A* **116**, 20545-20555, https://doi.org/10.1073/pnas.1906663116 (2019).
25. Choi, J. *et al.* Inflammatory Signals Induce AT2 Cell-Derived Damage-Associated Transient Progenitors that Mediate Alveolar Regeneration. *Cell Stem Cell* **27**, 366-382 e367, https://doi.org/10.1016/j.stem.2020.06.020 (2020).
26. Kobayashi, Y. *et al.* Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis. *Nat Cell Biol* **22**, 934-946, https://doi.org/10.1038/s41556-020-0542-8 (2020).
27. Strunz, M. *et al.* Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. *Nat Commun* **11**, 3559, https://doi.org/10.1038/s41467-020-17358-3 (2020).
28. Rendeiro, A. F. *et al.* The spatial landscape of lung pathology during COVID-19 progression. *Nature*, https://doi.org/10.1038/s41586-021-03475-6 (2021).
29. Westphalen, C. B. *et al.* Long-lived intestinal tuft cells serve as colon cancer-initiating cells. *J Clin Invest* **124**, 1283-1295, https://doi.org/10.1172/JCI73434 (2014).
30. Cui, L. *et al.* Activation of JUN in fibroblasts promotes pro-fibrotic programme and modulates protective immunity. *Nat Commun* **11**, 2795, https://doi.org/10.1038/s41467-020-16466-4 (2020).
31. Bange, E. M. *et al.* CD8 T cells compensate for impaired humoral immunity in COVID-19 patients with hematologic cancer. *Reseacrh Square*, https://doi.org/10.21203/rs.3.rs-162289/v1 (2021).
32. Nalbandian, A. *et al.* Post-acute COVID-19 syndrome. *Nat Med*, https://doi.org/10.1038/s41591-021-01283-z (2021).
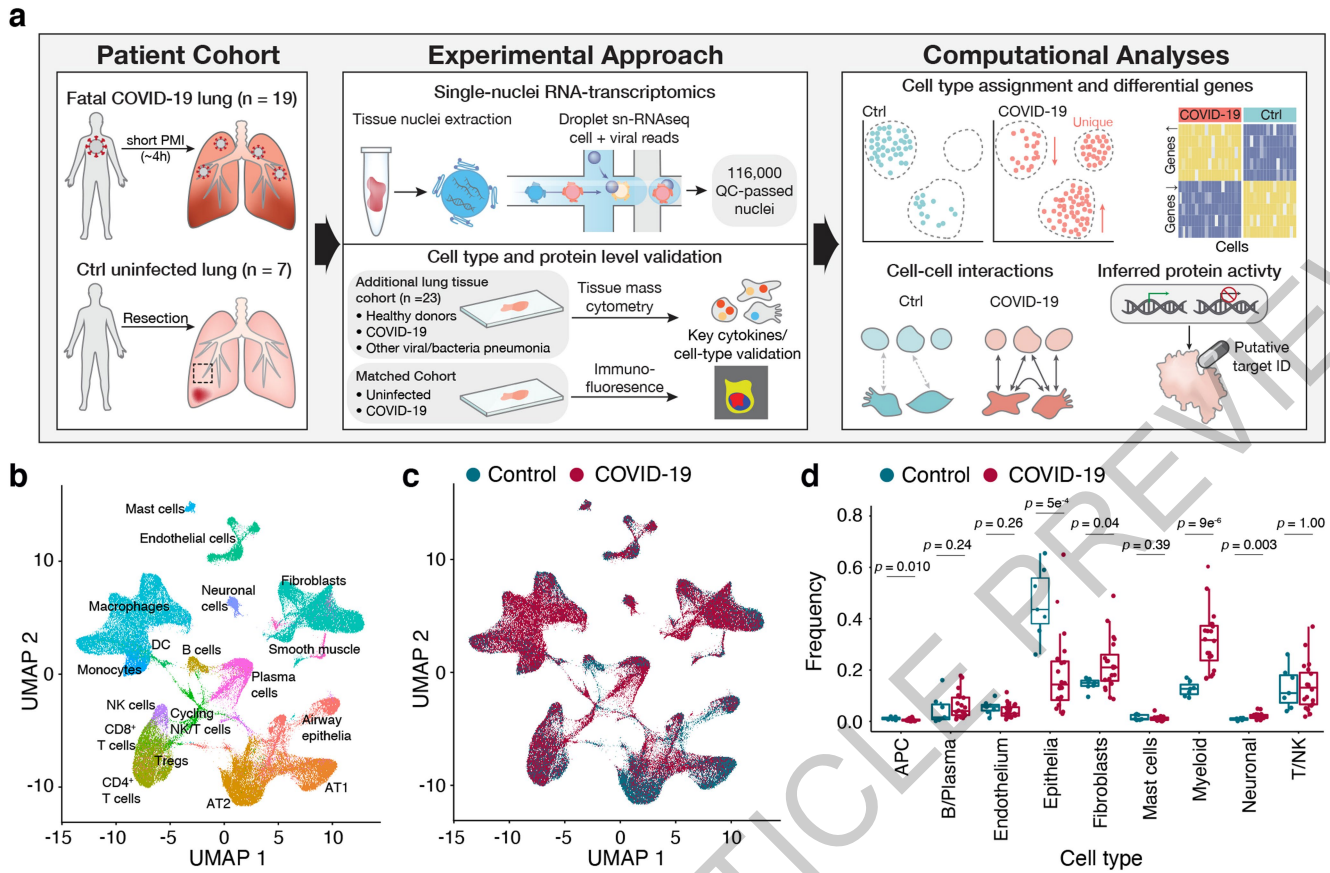
Fig. 1 | Study design and cellular landscape. a. Overview of study design. **b.** Major clusters and respective cell type assignment in UMAP and **c.** Origin of cells with same embedding as in (b). **d.** Fraction of major cell types in controls (n=7) and COVID-19 (n=19). Middle line: median; box edges: 25th and 75th percentiles; whiskers: most extreme points that do not exceed ±1.5 times interquartile range (IQR). Wilcoxon rank-sum test, p values indicated on top for respective comparisons.
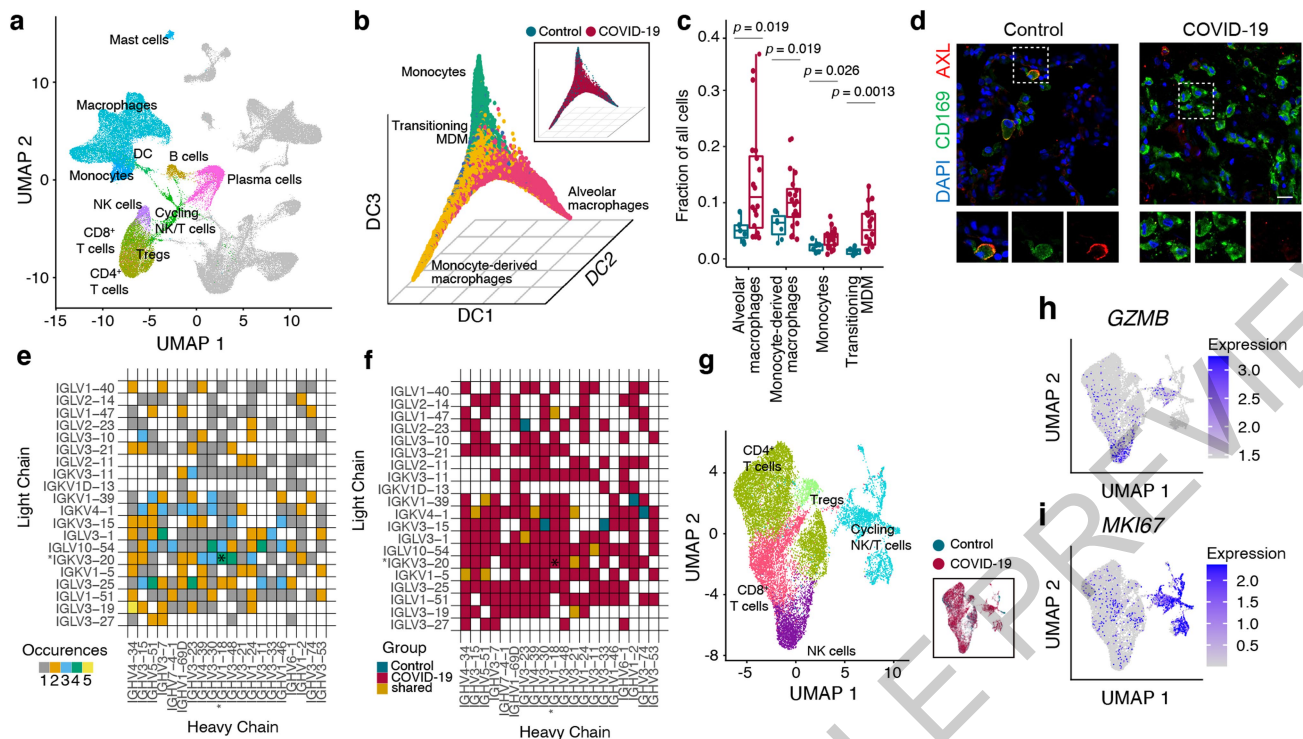
Fig. 2 | Immune responses in COVID-19. a. UMAP projection highlighting immune cell clusters. **b**. Visualization of myeloid cells using the first 3 diffusion components (DCs). Insert indicates group assignment. **c**. Fraction of myeloid cells in controls (n=7) and COVID-19 (n=19). Middle line: median; box edges: 25th and 75th percentiles; whiskers: most extreme points that do not exceed ±1.5xIQR. Wilcoxon rank-sum test, p values indicated on top. **d**. Representative IF staining of CD169, AXL and DAPI (large image) in control and COVID-19 lung tissues; selected area with overlay and individual channels (bottom). Scale bar=20μm. **e**. Top 20 recurrently detected *IGHV/IGLV* combinations in COVID-19 and **f**. corresponding group annotation. *combination for previously described anti-RBD-antibody[21]. **g**. UMAP of T/NK cells, insert indicates group assignment **h**. RNA expression (log-normalised expression values) of *GZMB* and **i**. *MKI67* in same embedding as (g).
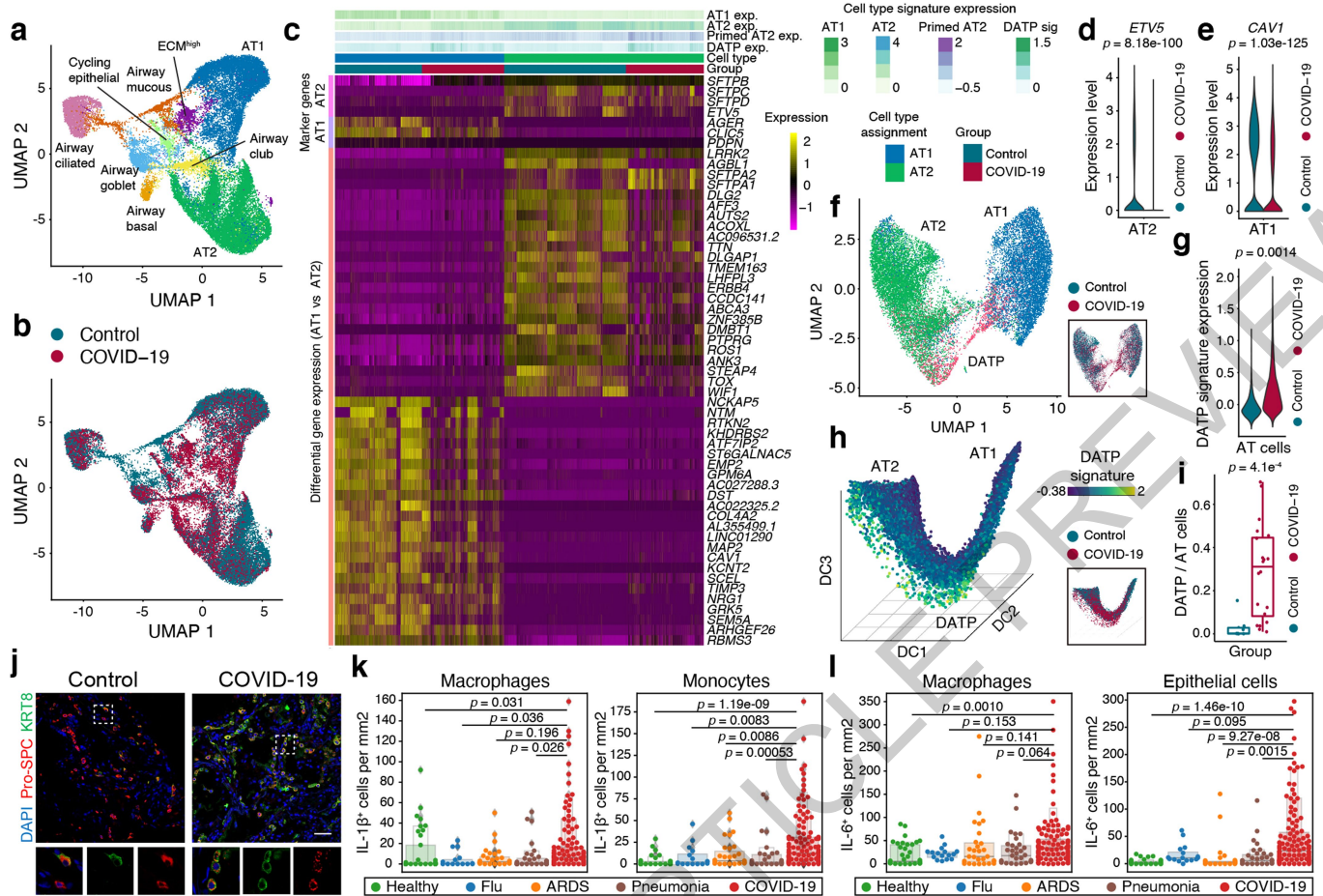
Fig. 3 | Impaired lung regeneration and sources of inflammation. a. UMAP of investigated alveolar and airway epithelial cells and **b**. corresponding group assignment. **c**. Differential gene expression (log-normalised, scaled, **Methods**) of AT1/AT2 cells across COVID-19 and control lungs. Single-cells (columns) and expression of top-regulated genes (rows). Left bar: lineage markers for AT1 (purple) and AT2 (pink) markers. Color-coded top lanes indicate expression strength of signatures (log-normalised, **Methods**) and group assignment as outlined in legends on the right. **d,e**. Violin plots of *ETV5* and *CAV1* mRNA expression (log-normalised) in AT2 and AT1 cells, respectively; Wilcoxon signed-rank test with Bonferroni correction. **f**. UMAP embedding of AT1/AT2 and identified DATPs; insert indicates group assignment. **g**. Violin plots of DATP signature expression (log-normalised) in AT1/AT2 cells. Wilcoxon singed-rank test. **h**. First three DCs showing main trajectories of AT2/AT1/

DATPs, expression of DATP signature and group assignment (insert). **i**. Fraction of DATP/AT cells in controls (n=7) and COVID-19 (n=19). Middle line: median; box edges: 25th and 75th percentiles; whiskers: most extreme points that do not exceed ±1.5xIQR. Wilcoxon rank-sum test. **j**. Representative IF staining of Pro-SPC, KRT8 and DAPI in control and COVID-19 lung tissues; representative area with overlay and small images with individual channels of selected area. Scale bar=50µm. **k,l**. Tissue mass cytometric quantification of (k) IL-1β and (l) IL-6 across healthy lung controls and donors with different infectious etiologies (groups per legend); examined cell types indicated on top. Each dot represents quantification of IL-1β and IL-6 in a region of interest (ROI); two-sided Mann-Whitney test, adjusted p values with the Benjamini-Hochberg FDR adjustment.
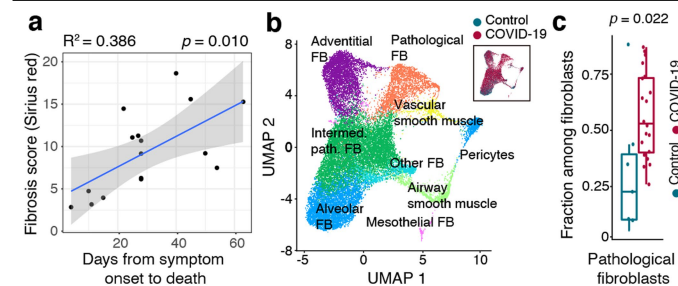
Fig. 4 | Pathological fibroblasts and ensuing fibrosis in COVID-19.
a. Coefficient of determination (R^2) of days of symptom onset to death and fibrosis score in COVID-19 samples (**Methods**). Error bands: 95% standard error interval on the Pearson correlation, p value indicated on top. **b**. UMAP of fibroblast sub-populations, insert indicataes group assignment. **c**. Fraction of pathological fibroblasts among all fibroblasts in controls (n=7) and COVID-19 (n=19) lungs. Middle line: median; box edges: 25th and 75th percentiles; whiskers: most extreme points that do not exceed ±1.5xIQR. Wilcoxon rank-sum test, p value indicated on top.

Methods

Tissue collection

All tissue specimens of lethal COVID-19 (with confirmed SARS-CoV-2 infection by reverse transcription polymerase chain reaction (RT-PCR)) and controls were collected at New York Presbyterian Hospital/ Columbia University Medical Center under IRB approved protocols (AAAB2667, AAAT0785, AAAS7370). Appropriate consent was obtained from patients or their next of kin. All procedures performed on patient samples were in accordance with the ethical standards of the IRB and the Helsinki Declaration and its later amendments. Samples were selected based on pathological review of corresponding haematoxylin and eosin (H&E) stained formalin-fixed paraffin-embedded (FFPE) tissue slides showing pathological involvement of the selected biopsy region from donors with a post-mortem incision time <10 hours. The donor age was 59 to >89 years. Tissues of ~1 cm^3 were snap-frozen embedded in Tissue-Tek optimal cutting temperature (OCT) compound (Sakura Finetek USA Inc, Torrance, CA) and stored at -80 °C until processing. For all decedents included in this study, affected lung tissues were harvested, and additionally, for a subset of individuals, matching tissues from kidney and heart were collected (*see Delorey et al.*, companion paper). Seven lung control tissues were collected from non-COVID-19 patients. The data set analysed and presented in this manuscript focuses on lung specimens from 19 COVID-19 decedents (profiled in 20 experiments) and 7 control (non-COVID-19) lungs.

Sample processing and preparation of single nuclei suspensions

All samples were processed in a biosafety cabinet equipped to comply with Columbia University safety measures established for working with COVID-19 specimens. Samples were processed as described previously[17] with the following specifications and modifications. For tissue dissociation we used tween with salts and tris (TST) buffer. For all wash steps we used salt and tris (ST) buffer, and all buffers were supplemented with 40 U/mL RNAse inhibitor (Thermo Fisher Scientific, Waltham, MA). All buffers were pre-chilled on ice and samples were kept on ice throughout the process to further prevent RNA degradation. Briefly, a fraction of the OCT embedded snap-frozen tissue was broken off and put in a pre-cooled 50 mL tube (Corning, NY) in a large volume of ice-cold phosphate buffered saline (PBS) and inverted until OCT was fully dissolve. Tissue was then collected by centrifuging at 300 x g for 2 minutes at 4 °C. PBS was decanted, and the tissue was resuspended in 2 mL cold TST buffer, mechanically dissociated using fine scissors and pipettes with decreasing orifice size, and incubated on ice for 5-10 minutes. The TST was quenched with 8 mL ST buffer, and the suspension was filtered through a 70 μM cell strainer. The tissue/nuclei suspension was pelleted by centrifuging at 500 x g for 5 minutes at 4 °C. The supernatant was decanted, and the nuclei were resuspended in 200-1000 μL ST buffer, filtered through a 40 μM cell strainer attached to a FACS tube (Corning, NY), counted, and immediately processed for single-nuclei RNA sequencing.

Single-nucleus RNA library preparation and sequencing

Single-nuclei suspensions were counted with disposable counting chambers (Bulldog Bio, Portsmouth, NH) on a Leica DMi 1 microscope by a second investigator not involved in tissue processing. A total of 15,000-20,000 nuclei were loaded per channel on a Chromium controller using Chromium Next GEM Single Cell 3′v3.1 reagents (10x Genomics, Pleasanton, CA) placed inside the bio-safety cabinet and single-nucleus RNA-sequencing libraries were prepared as per manufacturer instructions increasing the recommended initial cDNA amplification cycles by one to account for lower amounts of RNA from nuclei compared to whole cells. Single-nucleus RNA libraries were analyzed and quantified using the TapeStation D1000 screening tapes (Agilent, Santa Clara, CA) and Qubit HS DNA quantification kit (Thermo Fisher Scientific). Libraries were pooled equimolarly and quantified using quantitative PCR. Libraries were sequenced on a NovaSeq 6000 with S4 flow cell (Illumina, San Diego, CA) using paired end, single index sequencing with 28 cycles for read 1, 8 cycles for i7 index, and 91 cycles for read 2.

Generating single nucleus gene expression matrices

Raw 3′ snRNA-seq data were demultiplexed using Cell Ranger (v5.0) 'mkfastq' followed by 'count' to align the sequencing reads and generate a counts matrix. Transcripts were aligned to the human GRCh38 reference genome which was appended with the entire SARS-CoV-2 genome (severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome, GenBank MN908947.3) as an additional chromosome to the human reference genome. Subsequently, the customized 'GRCh38_SARSCoV2' reference genome was indexed using 'cellranger_mkref'.

Removal of background noise in gene expression matrices

We used the 'remove-background' function of CellBender (v.0.2.0) to remove technical ambient RNA counts and empty droplets from the gene expression matrices[33]. Cell Ranger-generated 'raw_feature_ bc_matrix.h5' files served as input for CellBender. The parameter 'expected-cells' was obtained from the Cell Ranger metric 'Estimated Number of Cells', while the parameter 'total-droplets-included' was set to a value between 18,000-24,000 representing a point within the plateau of the barcode rank plot in all samples.

Quality control and filtering

The resulting expression matrices were processed individually in R (v.4.0.2) by using Seurat (v.3.2.3)[34]. Filters were applied to keep nuclei with 200-7,500 genes, 400-40,000 unique molecular identifiers (UMIs), and less than 10% of mitochondrial reads. Additionally, Scrublet was applied to identify and remove doublets with an expected doublet rate ranging from 4-9.6% based on the loading rate[35]. Samples containing less than 1,000 nuclei after filtering were excluded from further analyses. Filtered gene-barcode matrices were normalized with the 'Normalize-Data' function using 'LogNormalize' and the top 2,000 variable genes were identified using the 'vst' method in 'FindVariableFeatures'. Gene expression matrices were scaled and centred using the 'ScaleData' function. Next, we performed principal component analysis (PCA) as well as UMAP using the first 30 principal components. UMAPs of individual samples were inspected prior to integration.

Integration of individual samples

Individual samples were integrated in Seurat using the reciprocal PCA (RPCA) pipeline to remove batch effects in large data sets. The 'SelectIntegrationFeatures' function was applied to choose the features ranked by the number of data sets they were detected in. Next, the 'FindIntegrationAnchors' function selected a set of anchors between different samples using the top 50 dimensions from the RPCA to specify the neighbour search space. Six samples were specified as a reference, including three controls (C51ctr, C52ctr, C53ctr) and three COVID-19 (L01cov, L12cov, L16cov) samples. 'IntegrateData' was then applied to integrate the data sets using the pre-computed anchors and the integrated data set was scaled using 'ScaleData'. PCA and UMAP dimension reduction based on the top 30 principal components were performed. Nearest-neighbour graphs using the top 30 dimensions of the PCA reduction were calculated and clustering was applied with a resolution of 0.8. Harmony[36] was run on the PCA matrix above using default parameters with patient ID as the batch key and 10 iterations.

Cell type identification

The main cell types were identified by manual annotation of differential gene expression (DGE) between clusters. The 'FindAllMarkers' function identified positive markers for each cluster with a minimal fraction of 25% and a log-fold change threshold of 0.25. This initial labelling resulted in the identification of epithelial, endothelial, fibroblast,

neuronal, myeloid, antigen presenting cells (APC), mast, T/NK and B/plasma cell populations as well as one low quality cluster which we removed. Next, we split the Seurat object into subsets of the main labels and reran scaling, PCA, UMAP dimension reduction, clustering and DGE analysis on each subset. The resulting clusters were annotated manually (see fibroblast assignment) or by using cell-type specific single-cell signatures from respective cell atlases, and labels were added to the main object. Additionally, cell cycle phases were scored in the subsets using the 'CellCycleScoring' function, adjusted for individual cut-offs and added to the main object. Within the myeloid subpopulation, two low-quality clusters (characterised by higher expression of mitochondrial reads) were observed and removed leaving a total of 116,314 cells for downstream analyses (of 119,535 initial cells after QC). Signatures and canonical markers (ST4) to identify airway basal, club, ciliated, goblet, mucous, AT1, and AT2 cells were obtained from Travaglini *et al.*[37]. Alveolar macrophages were scored using a signature based on DGE obtained from Travaglini *et al.*[37] and identified as AMs[38] with a module score > 0.15. A tuft-cell signature was obtained from Deprez *et al.*[39]. To further characterise the fibroblast population, fibroblast cells were selected using Seurat's 'subset' function and reanalysed to identify the different fibroblast subtypes. The reanalysis included the standard Seurat workflow with 'RunPCA,' 'FindNeighbours,' 'FindClusters,' and 'RunUMAP' performed on the 'integrated' assay. The number of PCA dimensions used was 15 with a resolution parameter of 0.5. After obtaining the fibroblast cell clusters, the DGE in each cluster were computed with 'FindAllMarkers' on 'RNA' assay (ST9). The fibroblast subtypes were identified by manually curating the cluster DGE with the reported literature such as the single-cell lung-atlas[37], lung-fibroblast atlas[3], single-cell database PanglaoDB[40], and Human Protein Atlas[41–43]. However, these resources were based on scRNA-seq or bulk studies. Therefore, the few reported fibroblast subtype markers were usually not specific or had low expression in snRNA-seq data. Therefore, we compared our subcluster DGE with the literature reported subtype DGE with shared high expression in sn/scRNA-seq data. These manually curated lists of fibroblast subtype specific marker genes were used to identify fibroblast subtypes in our data set (ST4). This procedure was used to identify alveolar fibroblasts, adventitial fibroblasts, pericytes, airway smooth muscle, vascular smooth muscle, and mesothelial fibroblasts. Cell clusters with high expression of *COL1A1* and *CTHRC1* were annotated as 'Pathological fibroblasts' because they are reported to contribute to the leading edge of fibrosis[3]. Clusters with lower expression of *COL1A1* and *CTHRC1* compared to 'Pathological fibroblasts' but without any markers for other fibroblast subtypes in their DGE were annotated as 'Intermediate pathological fibroblasts'. One cell cluster without distinct DGE was annotated as 'Other fibroblasts'. For visualisation purposes, expression scores were plotted in UMAP embeddings or violin plots as log-normalised values (natural logarithm ln (1+x)), and in dot plots as log-normalised values (natural logarithm ln (1+x)) that are furthermore centered on 0 with a variance of 1 (scaled).

Cell Type Frequency Comparison

Unless otherwise noted, we calculated frequencies of cell types in each sample from COVID-19 and control groups, and compared medians of the two groups to determine differences in frequency. Significance was assessed using a Wilcoxon rank-sum test.

Module scores for feature expression

The 'AddModuleScore' function was applied to calculate average expression levels of gene signatures on a single-cell level. Mouse-based signatures to identify DATPs and primed and cycling AT2 cells were obtained from Choi *et al.*[25] and converted to human homolog genes. Three genes (*CLDN4*, *KRT8*, *CDKN1A*) comprised the initial, Choi *et al.*-derived "DATP" signature. AT1 and AT2 cells were subset from the main Seurat object and reintegrated using the Seurat standard integration with 30 dimensions and a k-neighbours filter of 60 in the 'FindIntegrationAnchors' function.

First, all AT1 and AT2 cells were scored for the three-gene signature and cells with a module score > 0.7 were preliminarily labelled as DATPs. Next, we used DGE to identify additional markers that define the DATP program. We then scored our resulting DATP signature including 163 genes to the AT1/AT2 cells and labelled all cells with a module score of > 0.4 as DATPs. T cell scores were obtained by using the Seurat implementation of gene set scoring with 50 bins and a control size equal to the number of genes in the set. Up-regulation and down-regulation programs (TRM, Tact, Tmem Texh), defined by Devi et al. (under revision), were used to infer T cell phenotypes. The up-regulation and down-regulation signatures were scored separately, and the down-regulation score was subtracted cell-wise from the up-regulation score to obtain the composite score. Effect size was calculated using Cohen's D (i.e. the difference of means divided by the pooled standard deviation).

Diffusion component (DC) analysis

We applied diffusion maps as a nonlinear dimensionality reduction technique to examine the major components of variation across subsets of cells. We computed DCs using the 'DiffusionMap' function of the *Destiny* R-package (v3.3.0) with the top 30 principal components used in the k-nearest neighbours algorithm (k-NN)[44]. The epithelial subset consisting of airway basal, club, and goblet cells was reintegrated for the DC analysis using the Seurat standard integration with 30 dimensions and a k-neighbours filter of 50 in the 'FindIntegrationAnchors' function. Samples with < 50 cells were excluded from reintegration which removed a total of 10 samples (one control sample and nine COVID-19 samples). Tuft-like cells were identified as cells with DC1 values > 0.015 based on an overlap with the tuft-cell signature in the diffusion trajectory dominating the first DC.

Differential gene expression (DGE)

DGE was performed by using the Seurat function 'FindMarkers' on normalized count data to identify positive (overexpressed) markers in each population. The Wilcoxon rank-sum test (two-sided) was used to identify differentially expressed genes between two groups of cells and the log-fold change was set to 0.25. The parameter 'min.pct' was set to 0.25 to assure that genes are detected at a minimum fraction of 25% of cells in either of the populations. P values were adjusted using Bonferroni correction unless otherwise stated. Differentially expressed genes were plotted in violin plots using log-normalised expression values (natural logarithm ln(1+x)). For Heatmaps and dot plots expression values were log-normalised (natural logarithm ln(1+x)) that are furthermore centered on 0 with a variance of 1 (scaled).

Differential expression of signature scores

To test differential expression of three immune pathway signatures (Type I interferon abbreviated, inflammasome receptors, and chemotaxis, ST4), we obtained log-normalised expression values (ln(1+x)) for each gene in the signatures, and summed them for each signature. We then used a two-sided Wilcoxon test to test for differential expression of signatures in each cell type, and calculated log2FC.

Geneset enrichment

Geneset enrichment analyses were performed using the hypeR R-package[45]. The background population of genes was set to all detected genes. Geneset over-representation was determined by hypergeometric test.

B cell chain analysis

To analyse the distribution of heavy and light chains in B cells, the data set was subset to include only B cells. For the identification of variable chain regions, we selected the highest expressed heavy and light chain gene of each cell that expressed both, heavy (starting with '*IGHV*') and light (starting with '*IGLV*' or '*IGKV*') chain encoding genes. Next, we identified the highest expressed constant chain region among expressed

genes following the pattern '*IGH*[*G*, *M*, *A*, or *E*][number]'. The resulting pairs of heavy and light chains were visualized as a heatmap using average linkage for hierarchical clustering analysis and cross-referenced with previously described recurrently observed combinations[46].

Master Regulator Analysis and drug target identification

The fibroblast regulatory network in this study was reverse-engineered from snRNA-seq data by the ARACNe-AP[47,48] algorithm. We generated networks for each sub-cluster and integrated the networks by taking a union of the predictions of all networks. p values of Master regulator (MR)-target interactions predicted by the networks were integrated by using Fisher's method. The final fibroblast network contained predictions for 1,341 transcription factors regulating 9,770 target genes through 295,546 interactions. The relative activity of each transcription factor represented in the fibroblast network was inferred using the VIPER[49,50] algorithm, available as a package through Bioconductor. Conceptually, the VIPER algorithm is similar to the Master Regulator Inference Algorithm (MARINA)[48,51], which uses the MR targets inferred by the ARACNe[47,48] algorithm to predict drivers of changes in cellular phenotypes. In addition to calculating the enrichment of ARACNe-predicted targets in the signature of interest, VIPER also considers the regulator mode of action, regulator-target gene interaction confidence, and pleiotropic nature of each target gene regulation. Statistical significance, including p value and normalized enrichment score (NES), was estimated by comparison to a null model generated by permuting the samples uniformly at random 1,000 times. Druggable proteins with VIPER-predicted[49,50,52] aberrant increase in activity were ranked by their -log10 Bonferroni adjusted p value.

Ligand-receptor interaction inference in individual samples

CellPhoneDB[53] is a curated repository of ligand-receptor interactions along with their subunit architectures and integrated in a statistical framework to infer cell-type-enriched ligand-receptor interactions between cell types in single-cell/nucleus transcriptomics data. We used CellPhoneDB to identify ligand-receptor interactions between cell types in each individual Control (n=7) and COVID-19 (n=19) snRNA-seq data set. The ligand-receptor interactions were inferred in each patient separately, as by definition cell-to-cell interactions are only biologically meaningful within an individual. Moreover, separate inference also prevents inferring spurious interactions between patients having heterogeneous disease or health status. After identifying and annotating different cell types in our snRNA-seq datasets, we followed the recommended procedures for the preparation of input files for local implementation of CellPhoneDB v.2.0.0[53]. Briefly, for each individual sample, QC filtered raw counts matrix were normalized to counts per 10,000 and metadata files were obtained from the respective cell type annotations. CellPhoneDB analysis was performed with 'cellphonedb method statistical_analysis' command with default parameters.

Cell-cell interaction differences between COVID-19 and control samples

CellPhoneDB analysis of each sample identified the number of ligand-receptor interactions occurring between all 9 major cell-types in that sample. We analysed these cell-cell interaction counts between control (n=7) and COVID-19 donors (19 individuals, 20 samples) to identify the cellular cross-talk differences between COVID-19 and control samples. The median cell-cell interaction values from all the control samples formed the overall control lung cell-cell interaction counts. Similarly, the overall COVID-19 lung cell-cell interaction counts were the median from all the COVID-19 samples. The overall control and COVID-19 lung interaction counts were visualized as an interactome (Figure 4H and 4I, respectively) using 'igraph' R package with circle layout, where the edge width between two cell types was proportional to the number of interactions between them and the size of a cell-type circle was proportional to its frequency in the snRNA-seq.

Differential enrichment of ligand-receptor interactions between COVID-19 and control samples

CellPhoneDB analysis of each sample identified the significantly enriched ligand-receptor interactions in that sample by computing a mean of the ligand and receptor gene expression for each ligand-receptor interaction together with a corresponding p value. To find differentially regulated ligand-receptor interactions between COVID and control conditions, we first identified the common interactions across all samples. Briefly, we consolidated ligand-receptor expression for controls and COVID-19 separately by taking the median of ligand-receptor mean-expressions from seven control samples or 20 COVID-19 samples (from 19 donors). The minimum value of consolidated ligand-receptor expression in COVID-19 and control samples were set to 0.001 to prevent noise in low expression values from affecting the log fold-change calculations. Log2 of the control median expression was subtracted from the log2 COVID-19 median expression to obtain the log2FC of ligand-receptor expression in COVID-19. To compute p value of the log2FC for each interaction, unpaired two-sided Wilcoxon rank-sum test was performed for each interaction between COVID-19 and control samples. Adjusted p values were obtained by Benjamini-Hochberg procedure. Interactions with log2FC >=|2| and FDR p value<0.1 were reported as the top differentially enriched interactions in COVID-19.

Tissue preparation and processing for imaging

Lung tissues (human and mouse) were fixed with 4% paraformaldehyde (PFA) at 4 °C overnight with rotation. For paraffin sections, tissues were dehydrated through a 70-100% ethanol gradient and then embedded in paraffin. For cryosections, tissues were sequentially incubated with 20% and 30% sucrose and subsequently embedded in O.C.T. compound. 8-10 μm thick cryosections were obtained using a cryostat.

Microscopic imaging and quantification

Paraffin sections were dewaxed and rehydrated. Antigen retrieval was performed by high-pressure heating with a commercial antigen unmasking retrieval solution followed by blocking with 5% normal donkey serum. For IF staining, the sections were then incubated with primary antibodies listed in ST12 at 4 °C overnight. Cryosections were washed twice with PBS, and blocked with 5% normal donkey serum, followed by incubation with primary antibodies as shown in ST12 at 4 °C overnight. Conjugated secondary antibodies (1:500) were added to the sections and incubated for 2 hours at room temperature. The nucleus was stained with DAPI, and images were captured with a Zeiss LSM T-PMT confocal laser-scanning microscope (Carl Zeiss) and Zen 2012 SP1 (black edition) software (Zeiss) . Immunohistochemistry of C4d was performed on a Leica Bond 3 automated staining platform. Briefly, paraffin sections including both healthy control lung and COVID-19 lung tissues were treated with BOND Epitope Retrieval Solution 2 (Leica) for 20 mins and they were incubated with a C4d antibody for 30 mins. IHC signals were developed with Bone Polymer Refine Detection kit (Leica) with treatment of post primary polymer for 20 mins and DAB chromogen for 10 mins. For quantification, cells were counted using tiled stitched 20x images from over five sections per mouse and included at least three individual lobes or from representative areas of at least three human control lungs and COVID-19 lungs. Images were processed and analysed using ZEN blue 2.3 (Zeiss) and Adobe Photoshop Creative Suite 6 (Adobe) software in a blinded fashion. DATPs detected with co-immunostaining of Pro-SPC and KRT8 or HTII-280 and CLDN4. DATP percentages was determined by counting KTR8[hi] Pro-SPC[+] cells over Pro-SPC[+] cells or CLDN4[+] cells over HTII-280[+] cells. Macrophages were quantified by counting the total number of CD45[+]CD64[+] cells over CD45[+] cells. ChAT[+] tuft cells were quantified by counting the total number of ChAT[+] cells over DAPI[+] airway nuclei (for airway tuft cells) or over mm^2 of lung parenchyma.

Article

Multiplexed Immunofluorescence

Multiplexed immunofluorescence staining of lung tissues from COVID-19 decedents and control patients was performed using CD4, CD8, CD19, CD103, CD163 and Granzyme B (GzmB) antibodies (ST12) with the Opal 7-colour IHC kit (Akoya Bioscience) on the Leica Bond RX automated stainer (Leica Biosystems). FFPE tissue sections (5 μm) were baked for 2 hours at 60 °C, followed by automatic deparaffinization, rehydration, and antigen retrieval in BOND Epitope Retrieval Solution 2, pH 9 (Leica Biosystems) for 30 minutes at 95 °C. Immunofluorescence staining with Opal and Tyramide Signal Amplification (TSA) was performed in 6 cycles. In each cycle, the tissue was incubated sequentially with a primary antibody for 30 min at room temperature, the secondary antibody conjugated to Polymeric Horseradish Peroxidase (HRP), an Opal fluorophore in TSA buffer, and BOND Epitope Retrieval Solution 1, pH 6 (Leica Biosystems) for 20 minutes at 95 °C to strip the tissue-bound primary/secondary antibody complexes prior to the next staining cycle. After nuclear counterstaining with DAPI, slides were coverslipped with Vectrashield HardSet Antifade mounting medium (Vector Laboratories) and 12-15 areas per slide were imaged using the Vectra 3 automated multispectral microscope (Akoya/PerkinElmer) with Vectra 3.0.5 software. Regions of interest were chosen by the pathologist for multispectral imaging (MSI) at 20x magnification and spectral unmixing using the InForm v2.4.6 software (Akoya). Demultiplexed images were exported as 32-bit TIFF files for further analysis.

Multiplexed Image analysis

All images were analysed and visualized using QuPath[54]. We used the highest resolution for all described steps. The QuPath project files as well as the additional scripts are available at github.com/IzarLab/CUIMC-NYP_COVID_autopsy_lung/tree/main/code/Vectra_image_analysis. First, images were loaded, renamed and segmented using 'WatershedCellDetection' based on DAPI intensity with a cell expansion of 4μm. Further parameter settings for these steps can be found in the 'Load_and_segmentation.groovy' script. Next, we created classes and the corresponding classifiers for each of the 6 markers of interest: CD4, CD19, GZMB, CD103, CD8 and CD163. The thresholds for the individual classifiers ('ClassifyByMeasurementFunction') were automatically calculated and adjusted for each patient based on visual inspection of the mean marker expression. If no patient specific classifier was created, the classifier with the ending "_04_A6.json" was used. All classifiers can be found in the object classifiers folder as json files. Once performed for all images, the individual assignments for each single cell were exported to a CSV file for downstream analysis and boxplot visualization.

Imaging Mass-cytometry

Imaging mass cytometry data from post-mortem lung tissue of lung infection patients and otherwise healthy donors was used[55]. The dataset is comprised of 237 images from 23 donors, containing 664,006 single cells for which cell type identities were derived from the intensity of 36 markers. All analysis were conducted in Python v3.8.2 with the following programs: numpy v1.18.5, scipy v1.4.1, Tifffile 2020.6.3, Networkx v2.5, Scikit-image v0.17.2, Pingouin v0.3.7, and Scanpy v1.6.0. Single cells were labelled as positive for IL-6 or IL-1β based on their Z-score of intensity using Gaussian Mixture models (scikit-learn[56], version 0.23.0) using model selection based on the Davies–Bouldin index[57]. The number of cells positive for a marker in each ablated region of interest (ROI) was normalized by its area, and mean values per disease group and cell type across all ROIs were visualized as bar charts. To assess the significance of changes across both disease groups and cell types a two-sided Mann-Whitney test and adjusted p values with the Benjamini-Hochberg FDR adjustment using the pingouin package (version 0.3.9)[58] was used. Representative regions within the ROIs were displayed as false-colour images by normalizing the signal intensity to the unit scale after clipping the signal below and above the 3rd and

98th percentiles, respectively. Finally, a Gaussian filter with sigma of one pixel (one micron) was applied to the images.

Sirius Red staining and fibrosis scoring

Paraffin-embedded lung sections were dewaxed, rehydrated and stained during 1.5h with a picrosirius red solution (1.3% Picric acid, 1% Fast red and 1% Fast green). Four to five field at 4x magnification were taken using a polarized light filter on an Olympus IX71S1F-3 microscope with QCapture Suite Plus (v3.1.3.10) software. Images were quantified (% of Sirius red area/total area) using Adobe Photoshop (v 11.0). Pearson correlations between fibrosis score and days from symptom onset to death was calculated for 16 of 19 COVID-19 patients with samples available and time from symptom onset to death reported.

αSMA immunohistochemistry

Antigen retrieval of dewaxed and rehydrated paraffin embedded lung sections was performed with citrate pH6, blocked with 3% BSA and incubated with anti-αSMA-FITC (Sigma, F3777) overnight at 4 °C. After incubation with a biotin-anti-FITC antibody (Abcam, ab6655), detection was performed using the Vectastatin Elite ABC-HRP kit (Vector Laboratories, SP-2001) with DAB Peroxidase Substrate kit (Vector Laboratories, SK-4100) followed by counterstaining with haematoxylin. All reagents and dilutions are listed in ST12. All 7 control slides and 17 available slides from COVID-19 lungs were included in the analysis. Slides were scanned using a Leica SCN400 slide scanner with Leica Scanner Console software (v102.0.7.5) and quantified using the Leica Aperio ImageScope software (v12.4.3.5008) on at least 5 fields at 10x magnification.

Mice

Mouse studies were approved by Columbia University Medical Center (CUMC) Institutional Animal Care and Use Committees (IACUC). *Pou2f3*[-/-] mouse strain was described previously[59]. All mice were maintained on a C57BL/6 and 129SvEv mixed background and housed in the mouse facility at Columbia University according to institutional guidelines. The facility provides a 12h light/dark cycle, 18-23 °C room temperature and 40-60% humidity. All animal studies used a minimum of three mice per group and sample size was based on pilot experiments and previous experience. Mice were randomized to experiments and 8-12-week-old animals of both sexes were used in equal proportions.

Influenza infection mouse model

A total of 260 plaque forming units (pfu) of influenza A/Puerto Rico/8/1934 H1N1 (PR8) virus (a gift from Dr. Jie Sun at Mayo Clinics, Cleveland) dissolved in 40 μL of RPMI medium was pipetted onto the nostrils of anesthetized mice, whereupon mice aspirated the fluid directly into their lungs. For all procedures, administration of the same volumes of vehicle (RPMI medium) was used as control.

Flow cytometry analysis

At 14 days post infection, mice were euthanized and transcardially perfused with 10 ml of cold PBS. The lungs were then perfused with 1ml of PBS with 2mg/mL Dispase I and 0.5mg/mL DNase I and incubated in 5 mL of above buffer for digestion with gentle shaking for 60 minutes at room temperature. Lung lobes were removed and physically disassociated, followed by filtering through a 40 μm cell strainer. Cells were pelleted and resuspended in 1 mL lyse RBC buffer followed by incubation on ice for 5 min to remove red blood cells. After washing with FACS buffer (5% FBS, 0.2 mM EDTA in PBS), single cells were collected and immunostained with Fc blocking antibody (5 μg/mL) and live/dead cell stain kit at room temperature for 10 min. Cells were then washed and incubated with the following antibodies for one hour: PE/Cyanine7 anti-mouse CD45 (1:100), FITC anti-mouse CD64 (1:100), and APC anti-mouse F4/80 (1:100). Samples were analysed on LSR II (BD, Biosciences) with four lasers (405 nm, 488 nm, 561 nm, and 635 nm). Data were analysed with FlowJo software (Treestar).

Quantitative RT-PCR (qRT-PCR)

To quantitively measure the indicated cytokines, human lung tissues (three donors for both healthy and COVID-19 samples) or mouse lungs (a minimum of 3 mice per genotype) were individually homogenized in Trizol and total RNA was extracted using a RNeasy Plus Mini Kit (Qiagen) following the manufacture's instruction. cDNA was synthesized with Superscript-IV First-Strand Synthesis System (Invitrogen) and the gene-specific primers were mixed with cDNA templates and iTaq™ Universal SYBRR Green supermix (Bio-Rad). qPCR was carried out on a CFX Connect™ real-time PCR detection system (Bio-Rad) in a total volume of 20 μL. Three technical and biological replicates were performed. Relative fold change was determined by normalizing to *ß-actin* mRNA for mouse or to *GAPDH* mRNA for human. The primers for qPCR are listed in ST13.

Statistical analysis of imaging and qRT-PCR data

Imaging and qPCR data are presented as means with standard deviations of measurements unless stated otherwise. Individual values are plotted and represent independent biological samples unless stated otherwise. Statistical differences between samples are assessed with unpaired Student's t test using GraphPad Prism 9.0 (GraphPad Software Inc., San Diego, CA). p values below 0.05 are considered significant.

For multiplexed immunofluorescent images, cell fractions (% total or % parental population) were computed for each field of view individually using Excel 16.45 (Microsoft). After calculating the mean on a per sample basis, values were plotted using GraphPad Prism 9.0 (GraphPad Inc. San Diego, CA) and presented as means with standard deviations of measurements. Statistical differences between samples are assessed with unpaired Student's t test using GraphPad Prism 9.0 (GraphPad Software Inc., San Diego, CA). p values below 0.05 are considered significant.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Processed data are made available via the single-cell portal: https://singlecell.broadinstitute.org/single_cell/study/SCP1219. Processed data is also deposited in GEO (GSE171524)). Raw data are made available on the Broad Data Use and Oversight System: https://duos.broadinstitute.org (study ID DUOS-000130). Source data are provided with this paper.

Code availability

Code is made publicly available via: https://github.com/IzarLab/CUIMC-NYP_COVID_autopsy_lung.

33. Fleming, S. J., Marioni, J. C. & Babadi, M. CellBender remove-background: a deep generative model for unsupervised removal of background noise from scRNA-seq datasets. *bioRxiv*, 791699, https://doi.org/10.1101/791699 (2019).
34. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821, https://doi.org/10.1016/j.cell.2019.05.031 (2019).
35. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281-291 e289, https://doi.org/10.1016/j.cels.2018.11.005 (2019).
36. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289-1296, https://doi.org/10.1038/s41592-019-0619-0 (2019).
37. Travaglini, K. J. *et al.* A molecular atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619-625, https://doi.org/10.1038/s41586-020-2922-4 (2020).
38. Mould, K. J., Jackson, N. D., Henson, P. M., Seibold, M. & Janssen, W. J. Single cell RNA sequencing identifies unique inflammatory airspace macrophage subsets. *JCI Insight* **4**, https://doi.org/10.1172/jci.insight.126556 (2019).
39. Deprez, M. *et al.* A Single-Cell Atlas of the Human Healthy Airways. *Am J Respir Crit Care Med* **202**, 1636-1645, https://doi.org/10.1164/rccm.201911-2199OC (2020).
40. Franzen, O., Gan, L. M. & Bjorkegren, J. L. M. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)* **2019**, https://doi.org/10.1093/database/baz046 (2019).
41. Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419, https://doi.org/10.1126/science.1260419 (2015).
42. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* **356**, https://doi.org/10.1126/science.aal3321 (2017).
43. Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science* **357**, https://doi.org/10.1126/science.aan2507 (2017).
44. Angerer, P. *et al.* destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32**, 1241-1243, https://doi.org/10.1093/bioinformatics/btv715 (2016).
45. Federico, A. & Monti, S. hypeR: an R package for geneset enrichment workflows. *Bioinformatics* **36**, 1307-1308, https://doi.org/10.1093/bioinformatics/btz700 (2020).
46. Yuan, M. *et al.* Structural basis of a shared antibody response to SARS-CoV-2. *Science* **369**, 1119-1123, https://doi.org/10.1126/science.abd2321 (2020).
47. Basso, K. *et al.* Reverse engineering of regulatory networks in human B cells. *Nat Genet* **37**, 382-390, https://doi.org/10.1038/ng1532 (2005).
48. Lachmann, A., Giorgi, F. M., Lopez, G. & Califano, A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics* **32**, 2233-2235, https://doi.org/10.1093/bioinformatics/btw216 (2016).
49. Alvarez, M. J. *et al.* Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet* **48**, 838-847, https://doi.org/10.1038/ng.3593 (2016).
50. Ding, H. *et al.* Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. *Nat Commun* **9**, 1471, https://doi.org/10.1038/s41467-018-03843-3 (2018).
51. Lefebvre, C. *et al.* A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol Syst Biol* **6**, 377, https://doi.org/10.1038/msb.2010.31 (2010).
52. Alvarez, M. J. A precision oncology approach to the pharmacological targeting of mechanistic dependencies in neuroendocrine tumors. *Nat Genet* **50**, 979-989, https://doi.org/10.1038/s41588-018-0138-4 (2018).
53. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* **15**, 1484-1506, https://doi.org/10.1038/s41596-020-0292-x (2020).
54. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878, https://doi.org/10.1038/s41598-017-17204-5 (2017).
55. Rendeiro, A. F. *et al.* The spatio-temporal landscape of lung pathology in SARS-CoV-2 infection. *medRxiv*, https://doi.org/10.1101/2020.10.26.20219584 (2020).
56. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825-2830 (2011).
57. Davies, D. L. & Bouldin, D. W. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1**, 224-227, https://doi.org/10.1109/TPAMI.1979.4766909 (1979).
58. Vallat, R. Pingouin: statistics in Python. *Journal of Open Source Software* **3**, 1026, https://doi.org/10.21105/joss.01026 (2018).
59. Gerbe, F. *et al.* Intestinal epithelial tuft cells initiate type 2 mucosal immunity to helminth parasites. *Nature* **529**, 226-230, https://doi.org/10.1038/nature16527 (2016).

Article

a

Group	ID	Age (years)	Sex	Symptom onset to death (days)	Lung pathology	PMI (h)
COVID-19	L01cov	70	M	< 4	Congestion	4.5
COVID-19	L03cov*	73	M	16	DAD	4
COVID-19	L04cov *†	NA	F	11	DAD	2.5
COVID-19	L05cov*	78	F	28	DAD	5
COVID-19	L06cov	80	F	22	DAD	9
COVID-19	L07cov*	80	M	> 28	DAD	4
COVID-19	L08cov	83	M	10	DAD	3.5
COVID-19	L09cov*	79	M	15	DAD	5.5
COVID-19	L10cov*	84	M	NA	DAD	3
COVID-19	L11cov*	71	M	27	DAD	4
COVID-19	L12cov*	80	F	27	DAD	4
COVID-19	L13cov	63	M	28	DAD	2.5
COVID-19	L15cov	68	M	28	DAD	2
COVID-19	L16cov	70	F	50	DAD	2.5
COVID-19	L17cov	65	M	25	DAD	6
COVID-19	L18cov	72	M	54	DAD	4
COVID-19	L19cov	69	F	> 40	DAD	3
COVID-19	L21cov	68	F	45	Congestion	5
COVID-19	L22cov	58	M	63	DAD	6.75
Control	C51ctr	70	F	-	-	-
Control	C52ctr	69	F	-	-	-
Control	C53ctr	79	M	-	-	-
Control	C54ctr	72	F	-	-	-
Control	C55ctr	69	M	-	-	-
Control	C56ctr	75	M	-	-	-
Control	C57ctr	68	M	-	-	-

* concurrently profiled heart, and/or kidney tissue in companion study *Delorey et.al*
† Profiled in two experiments; DAD = diffuse alveolar damage

b

c

d

Extended Data Fig. 1 | Patient information and alternative batch-correction. Basic demographics of COVID-19 and control donors. *indicates decedents with concurrently profiled heart and/or kidney tissue in companion study *Delorey et al.;* † indicates a decedent with two independent lung specimens profiled. **b**. Effect of post-mortem interval (PMI, in hours) on clustering. **c**. Cell type labels overlaid on UMAP embedding resulting from the batch-corrected PCA matrix using Harmony (**Methods**). **d**. Same embedding as (c) with annotation of COVID-19 and control groups.
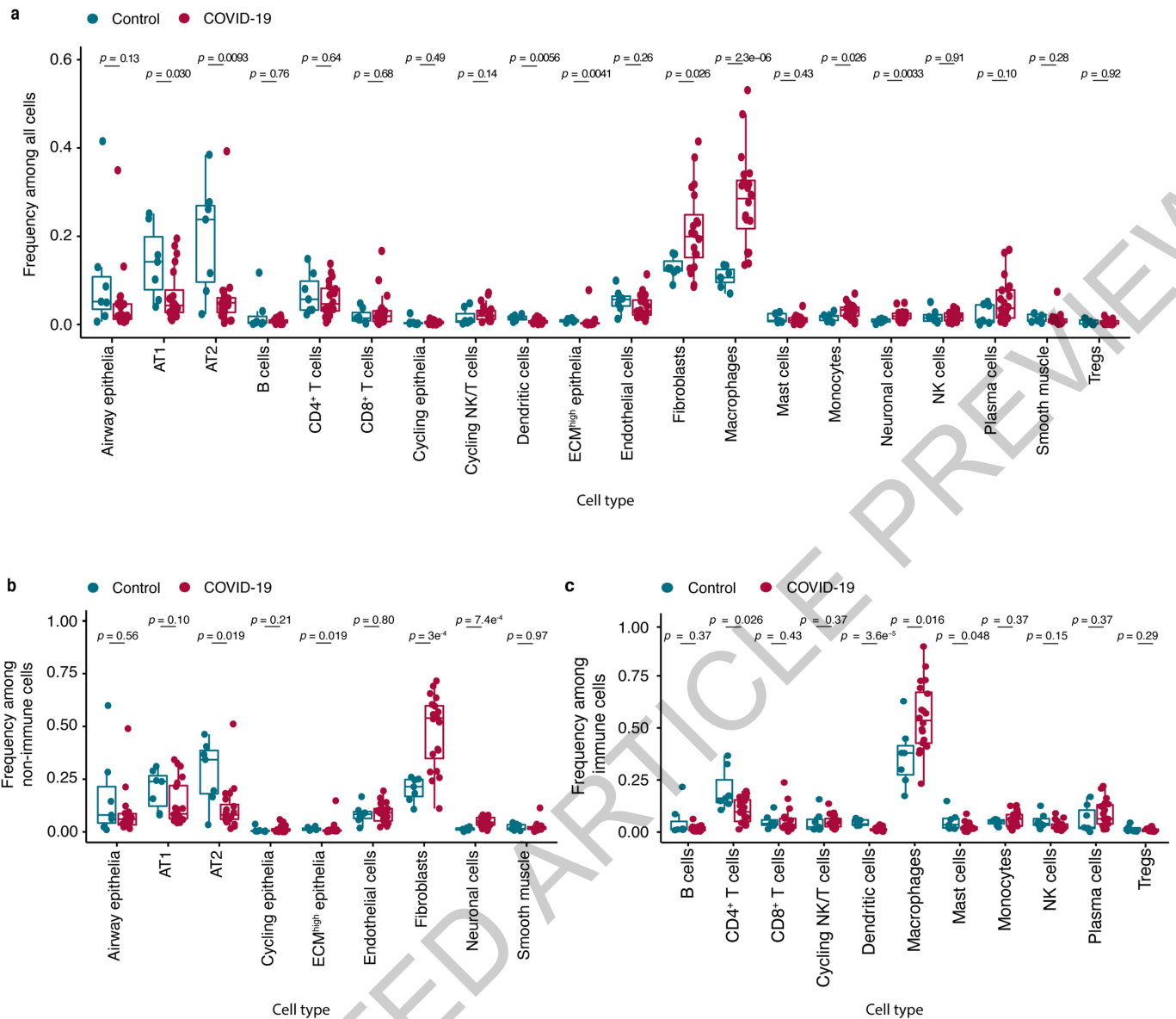
Extended Data Fig. 2 | Changes in celluar composition. a. Fraction of cell types in COVID-19 and controls across all cells (intermediate granularity). **b**. Fraction of cell types in COVID-19 and controls among non-immune cells only. **c**. Fraction of cell types in COVID-19 and controls among immune cells only. For a-c, controls (n=7 donors) and COVID-19 (n=19 donors examined over 20 experiments). Middle line: median; box edges: 25th and 75th percentiles; whiskers: most extreme points that do not exceed ±1.5xIQR. Wilcoxon rank-sum test, p values indicated on top.
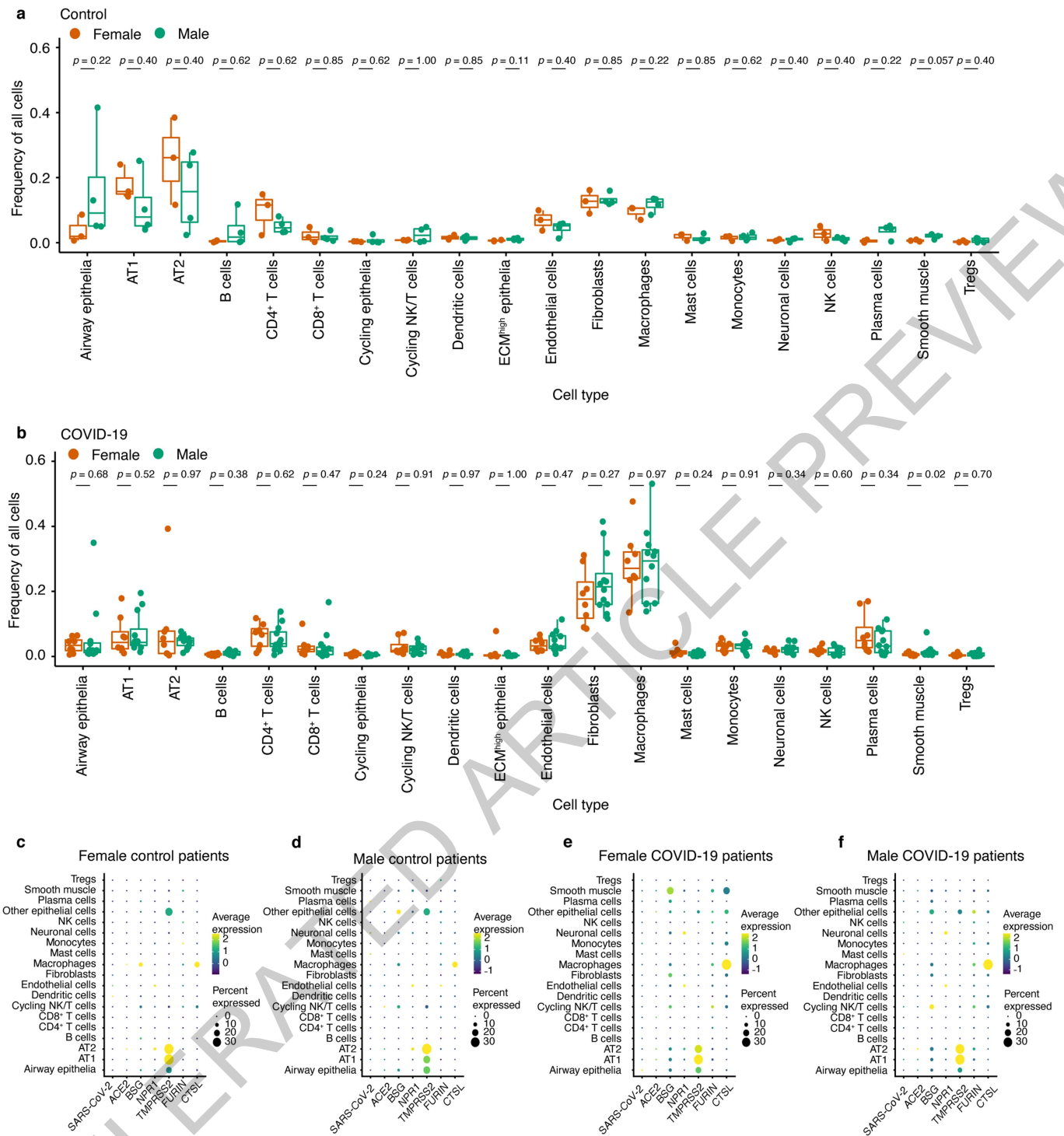
Extended Data Fig. 3 | Impact of sex on cellular composition and host receptor expression. a. Cell fractions in female and male individuals in control (n=7 donors) and **b**. COVID-19 decedents (n=19 donors examined over 20 experiments). Middle line: median; box edges: 25th and 75th percentiles; whiskers: most extreme points that do not exceed ±1.5xIQR. Wilcoxon rank-sum test, p values indicated on top. **c,d**. Log-normalised and scaled

Expression (methods) of selected receptors/putative receptors and proteases/putative proteases involved in SARS-CoV-2 entry in different cell types of control samples from female and male donors. Dot size indicates fraction of cells and colour indicates expression level. **e,f**. Same as (c,d) in lungs from COVID-19 decedents.
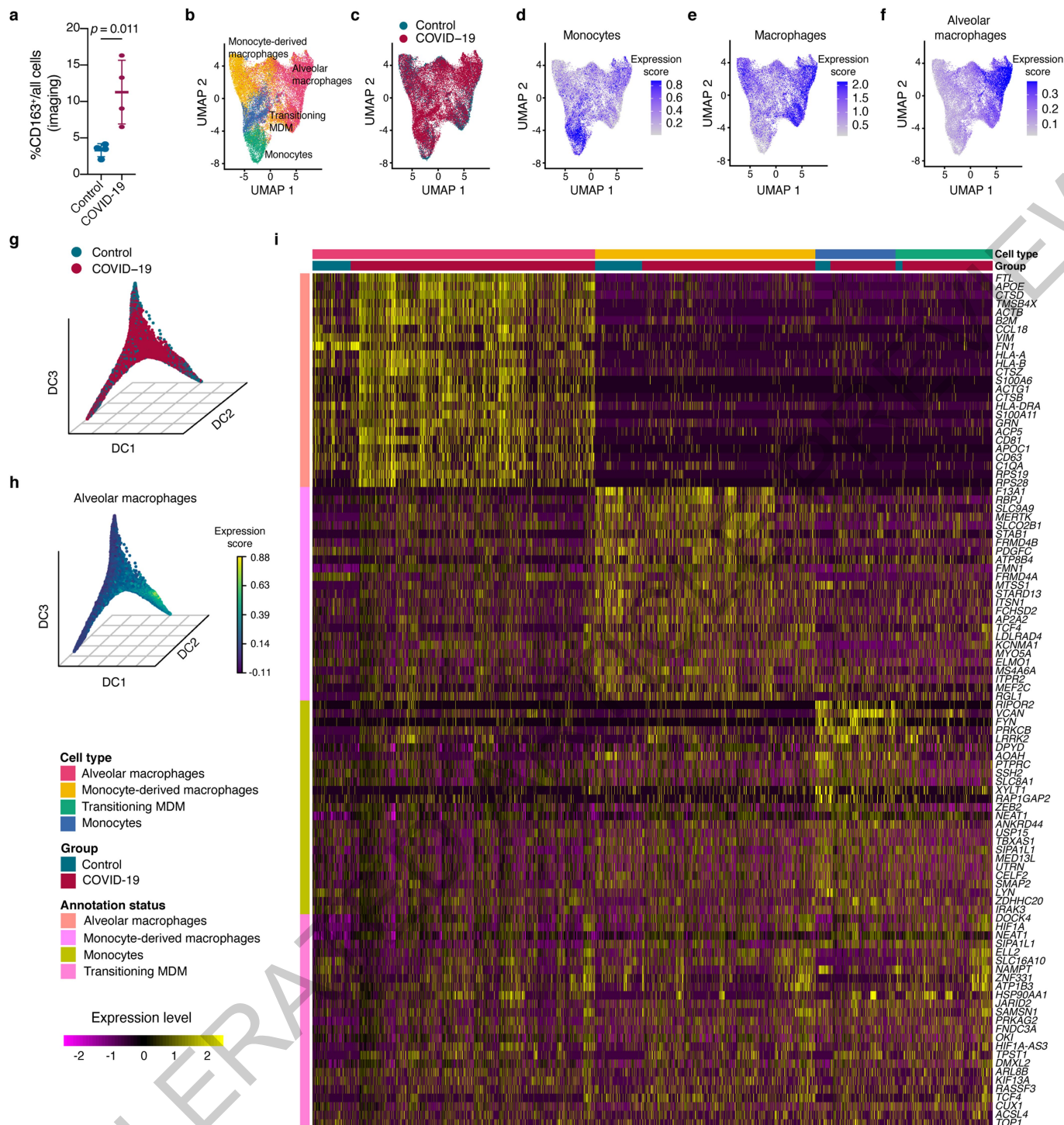
Extended Data Fig. 4 | Global changes in myeloid cells. a. Quantification of cells with CD163+ staining as percentage of all cells in a subset of control and COVID-19 tissues (n=4 donors per group). Mean±SD, t test, p values indicated on top. **b**. UMAP embedding with myeloid cell type assignment and **c**. group assignment. **d-f**. Expression scores (log-normalised) for monocyte, macrophage and alveolar macrophage signatures in same UMAP ambedding as

(b,c). **g**. First three DCs with annotation of control and COVID-19. **h**. First three DCs with expression of the alveolar macrophage signature **i**. Heatmap of top differentially regulated genes among indicated myeloid sub-populations. Left most bar indicates genes differentially regulated in respective cell types. Top lanes indicate cell type and group. Rows indicate log-normalised and scaled expression of genes (methods).
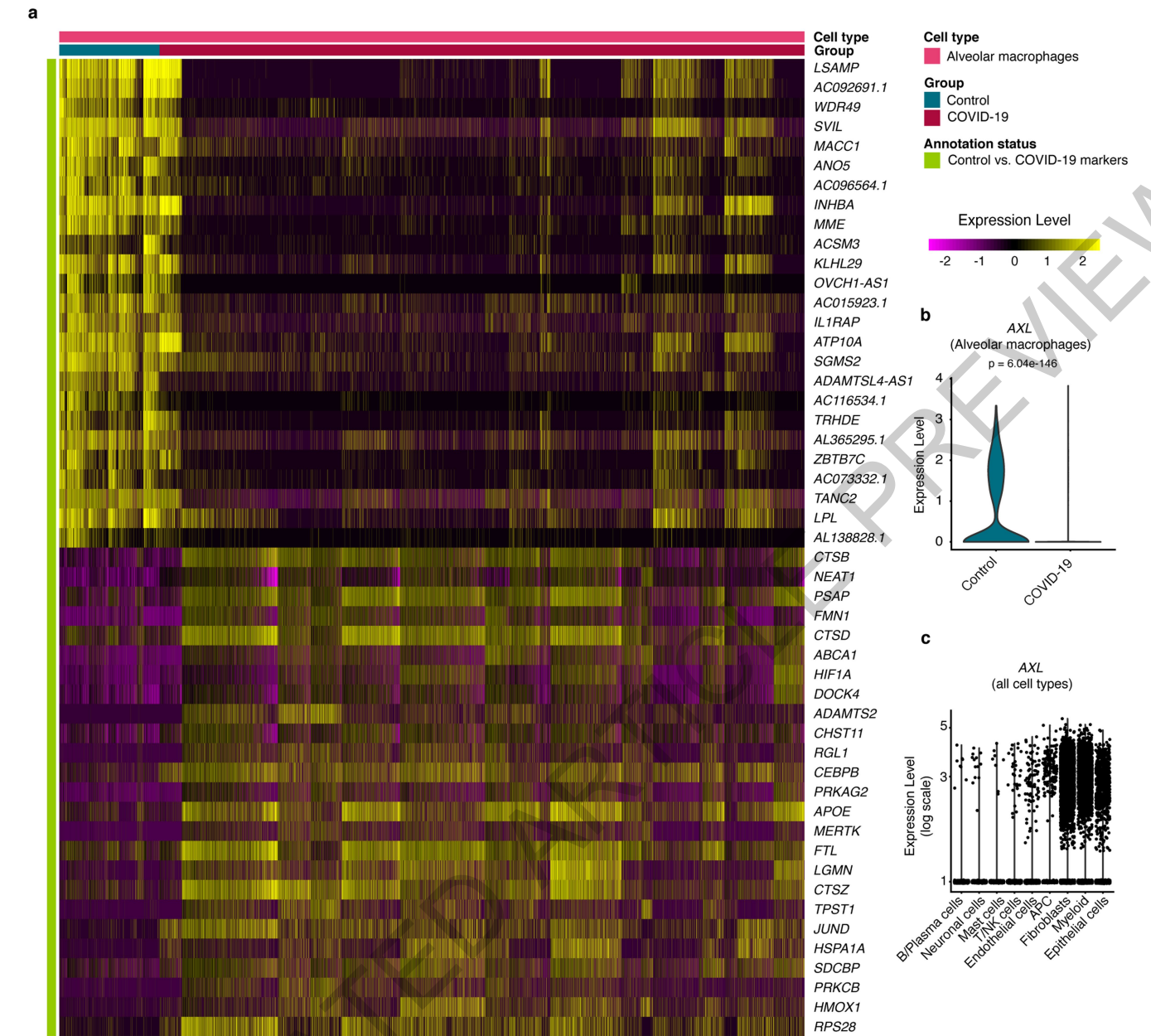
Extended Data Fig. 5 | Differential gene expression in alveolar macrophages. a. Heatmap of top differentially regulated genes (log-normalised and centered, methods) among indicated alveolar macrophages in COVID-19 and control samples. Top lane indicates cell type and group. Rows indicate expression of genes. **b**. Violin plot of *AXL* expression (log-normalised) in alveolar macrophages from controls and COVID-19 tissues. Wilcoxon rank-sum test with Bonferroni adjusted p value indicated on top. **c**. Expression of *AXL* (log-normalised) among major cell types. Expression of this gene was nearly exclusive to fibroblasts, myeloid, and epithelial cells.

Extended Data Fig. 6 | See next page for caption.

Article

Extended Data Fig. 6 | Inferred immunoglobulins in plasma cells a. UMAP embedding of cells within the B/plasma cell cluster and **b.** corresponding group assignment. **c.** Selected genes that define cells within the B/plasma cell cluster. Dot size indicates fraction of cells and colour indicates log normalized and scaled expression level (methods). **d.** Heatmap illustrating the number of cells with combinations of variable heavy (x axis) and light (y axis) chains recovered in plasma cells across all patients. Average linkage was used for hierarchical clustering analysis. The colour of each square indicates the number of cells detected for each specific pair (colour key). **e.** Same as (d), but indicating number of control samples with each combination detected (ST6). **f.** Same as in (e) but indicating isotype usage in control donors alone (ST6). **g.** Same as (e), demonstrating isotype usage in COVID-19 patients (corresponding to Fig. 3e, f; shown are the top 20 commbinations; **complete list in** ST6). **h.** Frequency (y axis) of variable heavy chains (x axis) in COVID-19 and control donors. **i.** Same as (h) for variable light chain usage. **j.** Frequency (y axis) of variable heavy chains (x axis) on a per donor basis. **k.** Same as (j) for variable light chain usage. **l.** Exemplary H&E image (n=19 donors evaluated) outlining different immune cell types. Scale bar=100μm. **m.** C4d immunohistochemistry in representative control (left) and COVID-19 (right) samples (n=6 donors per group). Scale bar=100μm.
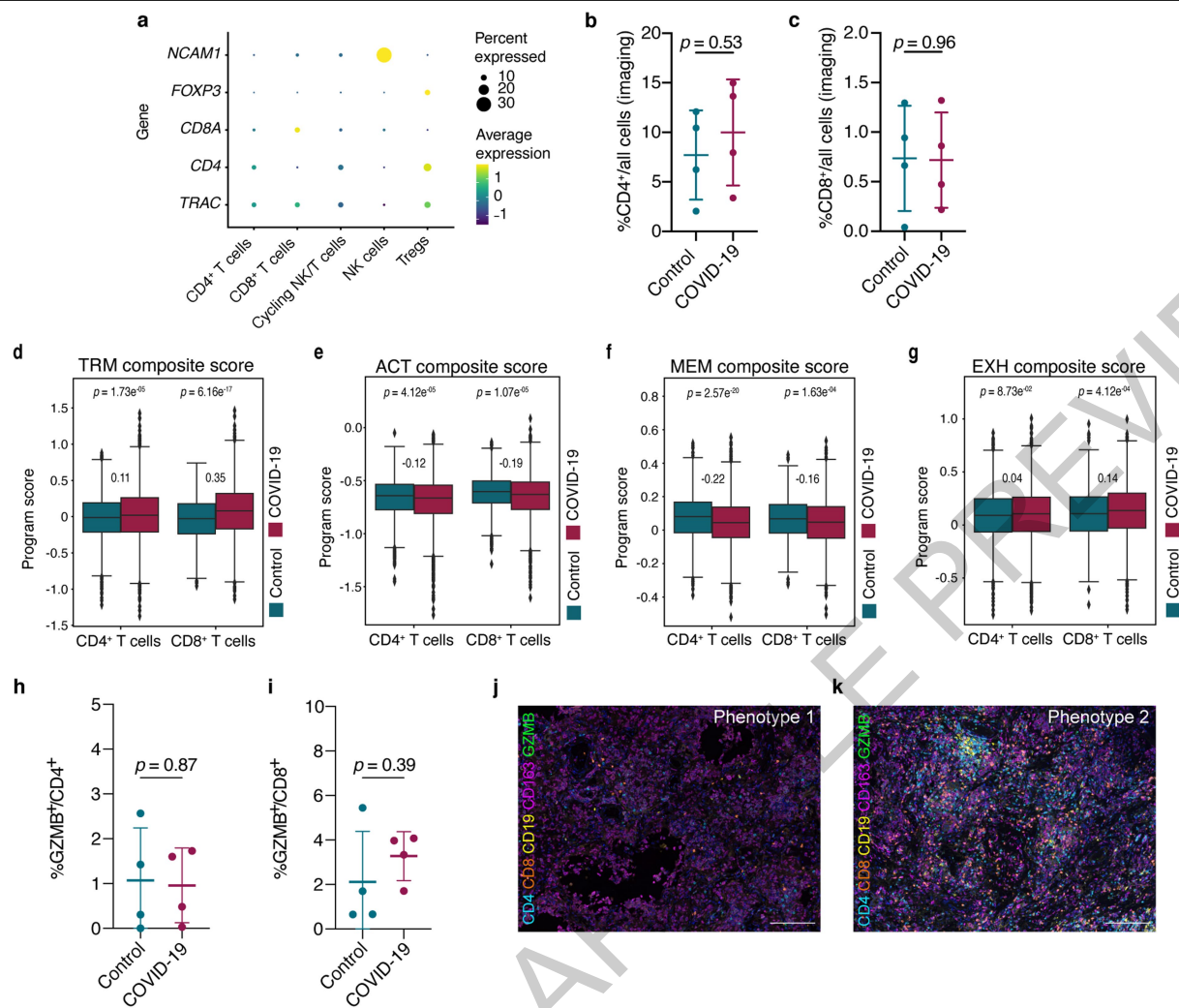
Extended Data Fig. 7 | Activation, residency and dysfunction cell states in T cells. a. Expression of selected genes in cells of the T/NK cell compartment. Dot size indicates fraction of cells and colour indicates expression level. **b.** Quantification of cells with CD4+ staining as percentage of all cells (y axis) in control and COVID-19 donors (n=4 donors/group). **c.** Same as (b) for CD8+ T cells. Mean±SD, t test, p values indicated on top. **d-g.** Expression of different Program Scores (tissue residency memory program, activation score, memory score and exhaustion score, all from *Devi et al.*, **Methods**) in CD4+ T cells (left) and CD8+ T cells (right) among control and COVID-19 donors. Middle line: median; box edges: 25th and 75th percentiles; whiskers: most extreme points that do not exceed ±1.5xIQR. Wilcoxon rank-sum test, p value indicated on top. Cohen's D is indicated on the bottom for each comparison (COVID-19 vs. control). **h.** Quantification of CD4+GzmB+ T cells as percentage of CD4+ T cells (y axis) in control and COVID-19 donors (n=4 donors/group). **i.** Same as (h) for CD8+ T cells. Mean±SD, t test, p values indicated on top. **j.** Representative multiplexed IF of a COVID-19 patients with a pure myeloid infiltrate and **k.** with a mixed myeloid and lymphoid infiltrate, n=4 donors/group. Scale bar=200µm.
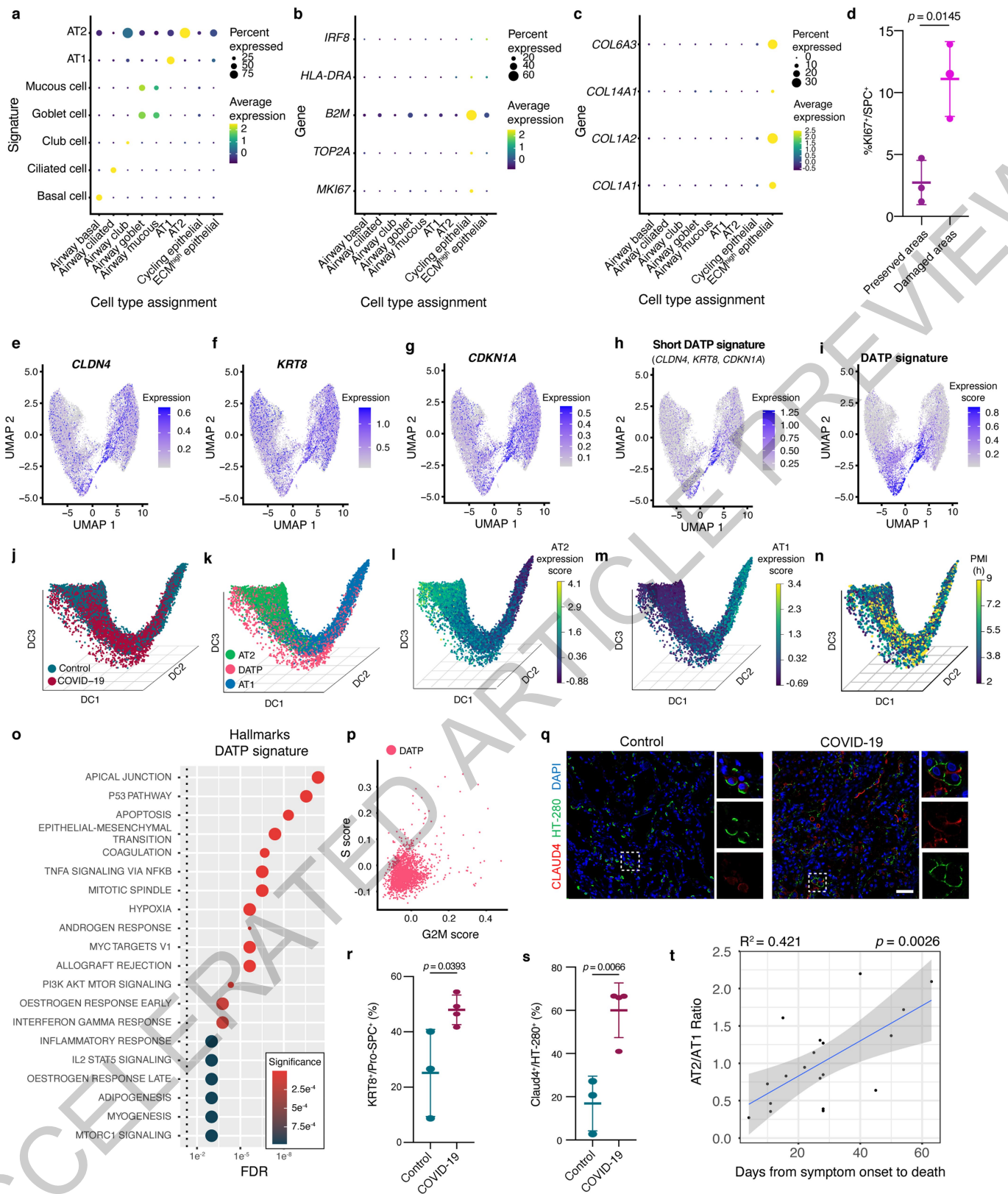
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | DATPs and lung regeneration. a. Expression of selected, previously established cell-type specific signatures (y axis) in cell types defined in this study (x axis). Dot size indicates fraction of cells and colour indicates expression level. **b,c**. Expression of selected genes (y axis) in different cell types (x axis) highlighting high expression of B2M in cycling epithelial cells (b) and collagen genes in ECM^{high} epithelial cells (c). **d**. Fraction of KI67$^+$ among Pro-SPC$^+$ cells in structurally preserved vs damaged areas (n=3 distinct areas each) of a COVID-19 sample. Mean±SD, t test, p values indicated on top. **e-g**. UMAP embedding of alveolar epithelium and expression of selected genes that define the DATP signature. **h**. Composite expression of the three-gene DATP signature. **i**. Expression of the refined DATP signature (**Methods**). **j-n**. First three DCs showing j. group assignment, k. cell/cell state assignment, l. expression of AT2 signature, m. AT1 signature (log-normalised, methods), and n. impact of PMI. **o**. Gene set enrichment analysis in DATPs (compared to AT1/AT2 cells). Rows indicate pathways in descending order of enrichment/significance, x axis indicates FDR. Colour indicating significance per legend. **p**. Inference of G2/M and S phase of individual DATPs (dots) (**Methods**). **q**. Representative IF staining (DATP marker CLDN4 and AT2 cell marker HTII-280) in control and COVID-19 lung tissue sections. Dashed boxes indicate area that is highlighted to the right of each image. Scale bar=50μm. **r,s**. Quantification of KRT8$^+$ q. and CLDN4$^+$ r. cells in a subset of tissue section in control and COVID-19 donors. Mean±SD, t test, p values indicated on top. q-s. control, n=3 donors; COVID-19, n=4 donors. **t**. Coefficient of determination (R^2) of days of symptom onset to death and AT2/AT1 ratio; Error bands: 95% standard error interval on the Pearson correlation (n=18 donors), p value indicated on top.

Article
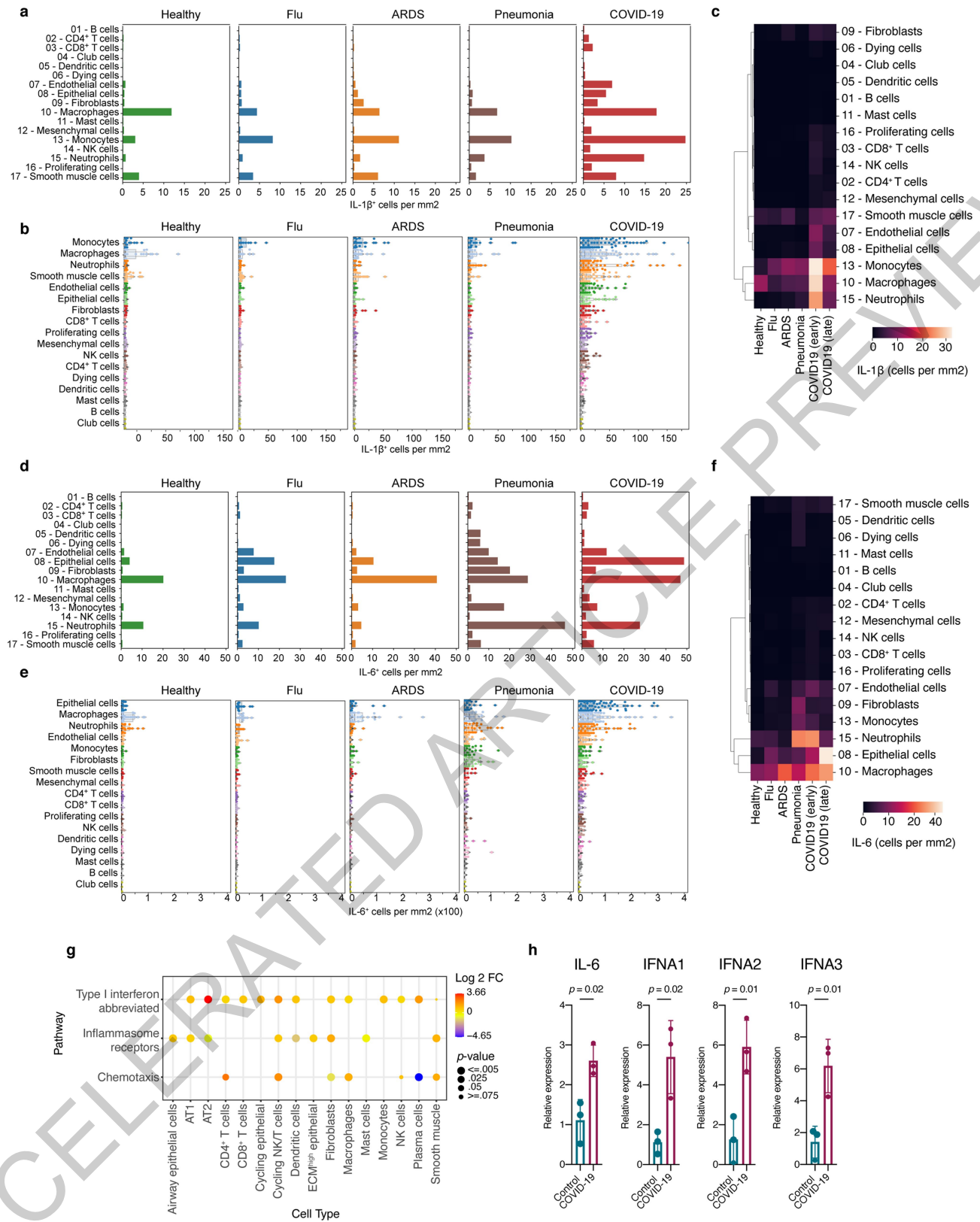
Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Cellular sources of inflammatory cytokine.
a. Average frequency of cell types expressing IL-1β across healthy and disease conditions. **b**. Quantification of IL-1β across cell types in controls and different infectious etiologies (indicated on top of each box). Each dot represents a single region of interest (ROI). **c**. Quantification of IL-1β across healthy controls and different infectious etiologies and cell types, including separation of patients with early (within 14 days) and late (within 30 days) of symptom onset to death from COVID-19. **d**. Average frequency of cell types expressing IL-6 across healthy and disease conditions. **e**. Quantification of IL-6 across cell types in controls and different infectious etiologies (indicated on top of each

box). Each dot represents a single region of interest (ROI). **f**. Quantification of IL-6 across healthy controls and different infectious etiologies and cell types, including separation of patients with early (within 14 days) and late (within 30 days) of symptom onset to death from COVID-19. **g**. Expression of selected manually curated gene sets of chemotaxis, inflammasome receptors and type I interferon (response) genes across different cell types (y axis). Dot size indicates significance level and colour indicates expression level (log2FC). **h**. qRT-PCR comparing the *IFNA1*, *IFNA2*, *IFNB1*, and *IL-6* mRNA levels in COVID-19 and control lungs (n=3 donors for each group). Mean±SD, t test, p values indicated on top.
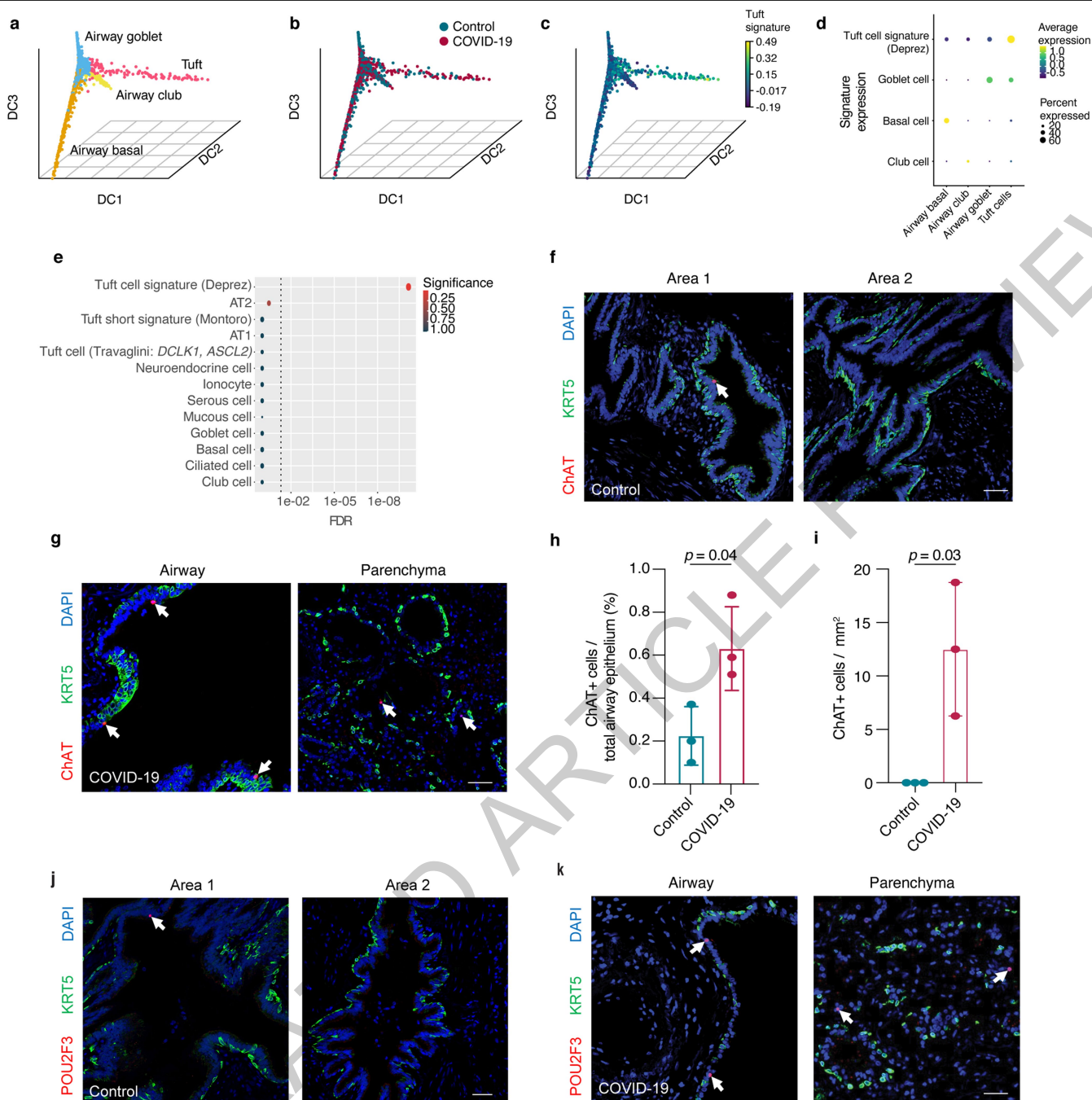
Extended Data Fig. 10 | Identification of ectopic tuft-like cells. a. First three DCs of airway epithelial cells with group annotation with cell-type assignment, **b.** group assignment and **c.** indicating expression of tuft cell signature in same projections. **d.** Expression of previously established signatures identifying cell types in cell types assigned in this study. Dot size indicates fraction of cells and colour indicates expression level (log-normalised and scaled, methods). **e.** Expression of selected cell type specific signatures of airway and alveolar epithelium from prior studies in cells identified as tuft-like cells in this study. Signatures in descending order of enrichment/significance, x axis indicates false discovery rate (FDR). Colour indicating significance as indicated by the legend. **f,g.** Representative immunofluorescence staining of f. control lungs (two areas) and g. COVID-19 (airway and parenchyma) staining for KRT5 and ChAT. Arrows highlight ChAT+ cells. Scale bar=50μm. **h.** Quantification of ChAT+ cells in the upper airway epithelium of control and COVID-19 lung tissues. Mean±SD, t test, p values indicated on top. **i.** Quantification of ChAT+ cells in the alveolar epithelium of control and COVID-19 lung tissues (n=3 donors/ group). Mean±SD, t test, p values indicated on top. **j,k.** IF staining for KRT5 and POU2F3 of j. control lungs and k. COVID-19 samples, including upper airway (left) and parenchyma (right). White arrows indicate POU2F3+ cells. Scale bar=50 μm. f-k. n=3 donors/group.
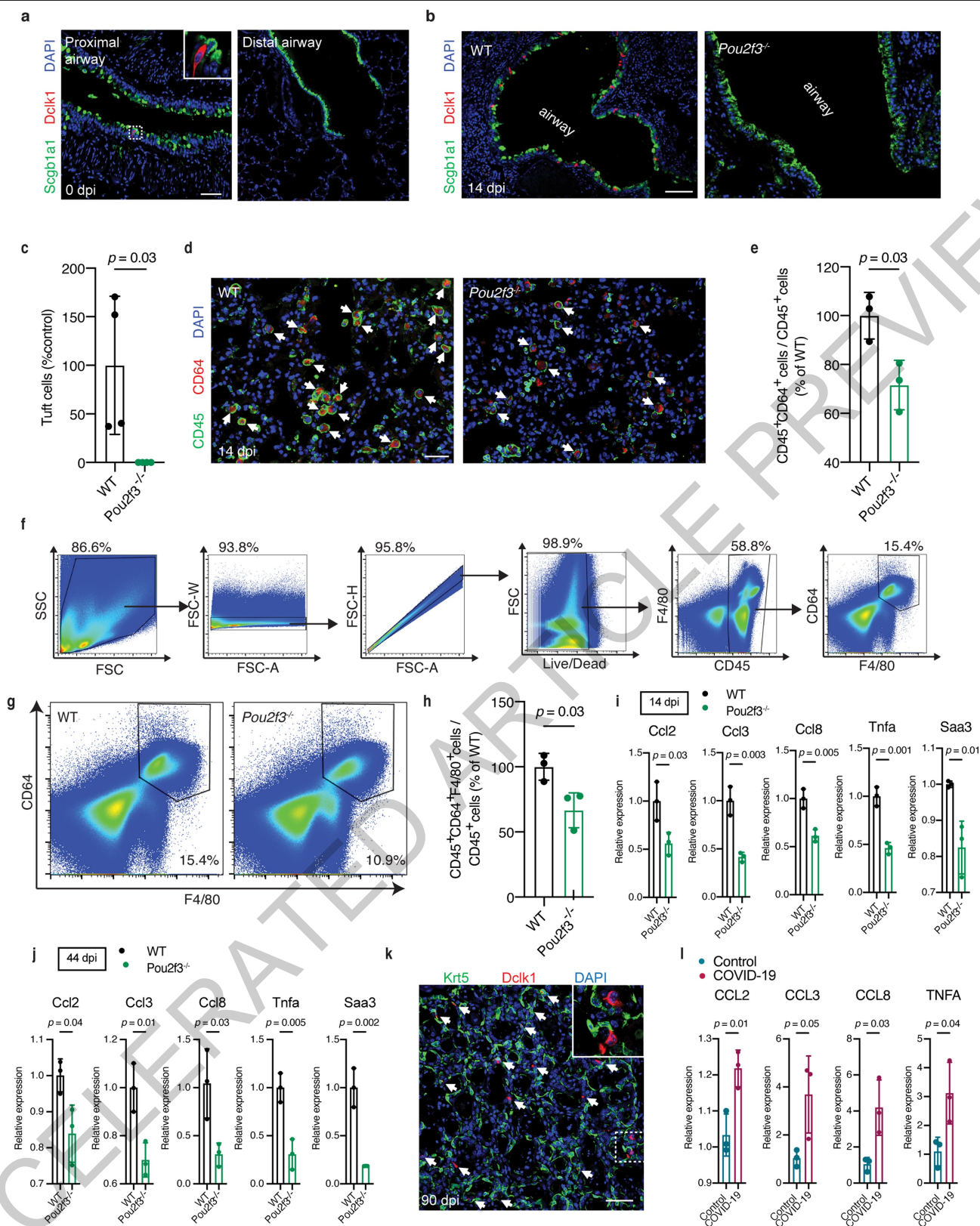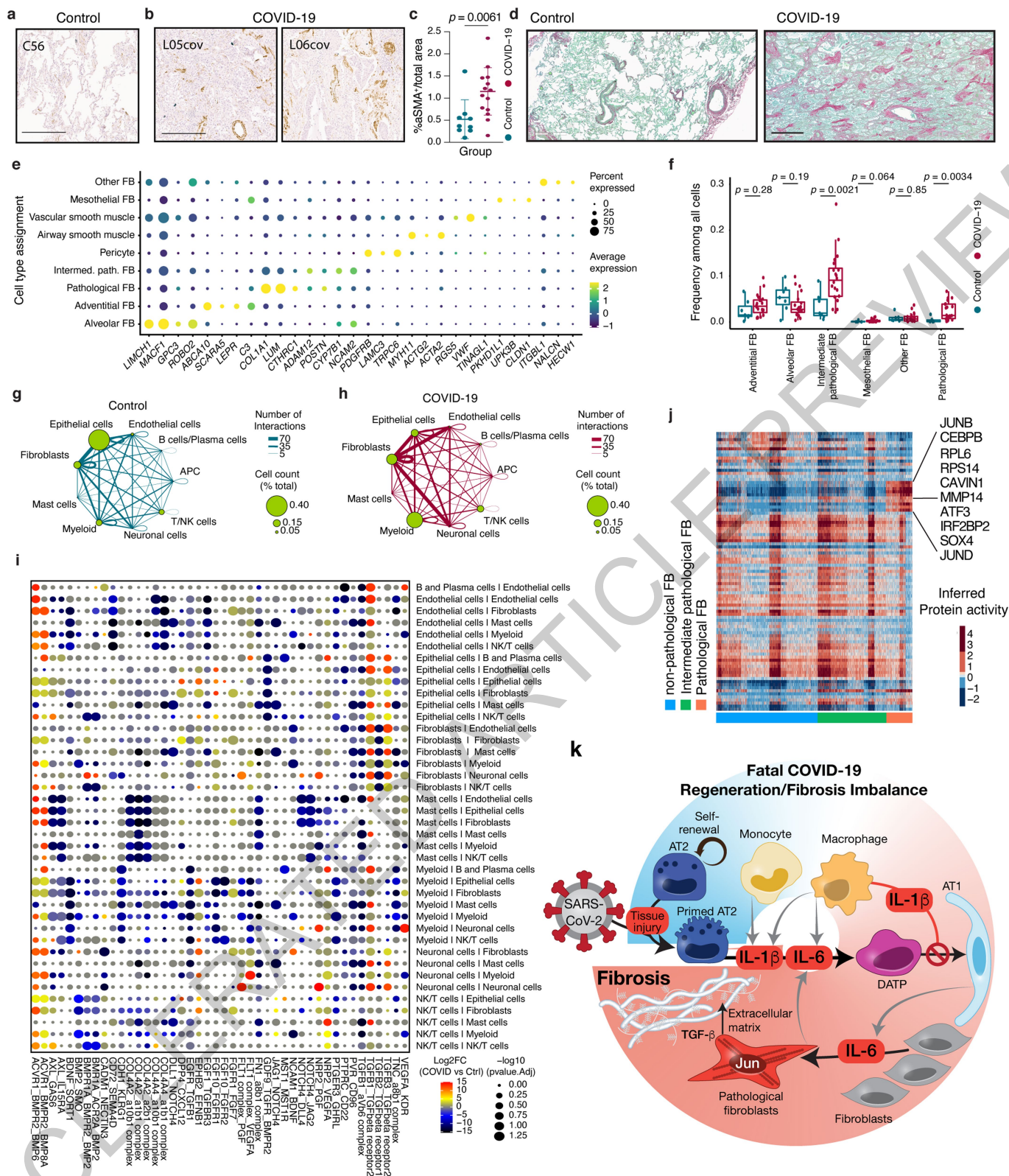
Extended Data Fig. 11 | See next page for caption.

Article

Extended Data Fig. 11 | Role of tuft cells in macrophage infiltration in murine viral pneumonia model. a. IF staining of Scgb1a1 and Dclk1 of proximal (left) and distal (right) airway in wild-type (WT) mice at baseline. n=3 animals/group. Arrow highlighting a Dclk1+ cell. Scale bar=50 μm. **b**. Same as (a) in WT (left) and *Pou2f3*-/- animals 14 dpi with H1N1 (PR8). **c**. Quantification as percentage of Dclk1+ cells in *Pou2f3*-/- compared to WT animals. Mean±SD, t test, p values indicated on top. b,c. n=4 animals/group. **d**. IF staining of CD45 and CD64 of lung parenchyma of WT (left) and *Pou2f3*-/- (right) mice 14 days following infection with H1N1 (PR8). Arrows indicate CD45+CD64+ macrophages. Scale bar=50μm. **e**. Quantification (percentage of CD45+CD64+/CD45+ cells) in WT and *Pou2f3*-/- animals 14 days following infection with H1N1. Mean±SD, t test, p values indicated on top. d,e. n=3 animals/group. **f**. Gating strategy to identify CD45+CD64+F4/80+ cells. **g**. Identification of CD64+F4/80+ cells (based on gating strategy in f.) in WT (left) and *Pou2f3*-/- animals (right) 14 days following infection with H1N1. **h**. Quantification of flow-cytometric determination of CD45+CD64+F4/80+ cells as percentage of CD45+ in *Pou2f3*-/- relative to WT animals (n=3 for each group). Mean±SD, t test, p values indicated on top. **i**. qRT-PCR comparing relative mRNA expression of indicated chemokines and cytokines in *Pou2f3*-/- and WT animals 14 days following infection with H1N1 (n=3 for each group). Mean±SD, t test, p values indicated on top. **j**. Same as (i) but 44 days following infection with H1N1 (n=3 for each group). **k**. Exemplary IF staining (n=3 animals/group) of Krt5 and Dclk1 in WT mouse 90 days following infection. Arrows indicate Dclk1+ cells. Scale bar=50μm. **l**. Same as (i,j) comparing expression of indicated chemokines/cytokines in control compared to COVID-19 patients (n=3 donors/group). Mean±SD, t test, p values indicated on top.

Extended Data Fig. 12 | See next page for caption.

Article

Extended Data Fig. 12 | Role of fibroblasts, potential drug targets and model of lethal COVID-19. a,b. Exemplary αSMA IHC staining in a. control (C56. n=7 donors) and b. COVID-19 samples (n=17 donors). Scale bar = 500 μm. **c**. Percentage of alpha-SMA positive cells per total area, n as in a,b. Mean±SD, t test, p values indicated on top. **d**. Exemplary Sirius red staining of a control (left, n as in a.) and COVID-19 (right, n as in b.) sample. Scale bar=600μm. **e**. Detailed annotation of fibroblasts in this study and selected marker genes. Dot size indicates fraction of cells and colour indicates expression level (log-normalized and scaled). **f**. Fraction of cell types in COVID-19 (n=19 donors examined over 20 experiments) and controls (n=7 donors) among all cells. Middle line: median; box edges: 25th and 75th percentiles; whiskers: most extreme points that do not exceed ±1.5xIQR. Wilcoxon rank-sum test, p values indicated on top. **g,h**. Inferred cell-to-cell interactions among major cell types (indicated as circles connected by lines) in (g) control and (h) COVID-19 lung samples. Size of the circle corresponds to the frequency of the respective cell type, thickness of the lines connecting circles indicate the absolute number of interactions. **i**. Differential enrichment (COVID-19 vs. control samples) of specific ligand-receptor interactions (rows) between two different cell types (columns). Dot colour indicates the log2FC of inferred ligand-receptor expression in COVID-19 compared to control (unpaired two-sided Wilcoxon rank-sum test), dot size is inversely correlated with the Benjamini-Hochberg adjusted p value (**Methods**). **j**. Inferred protein activity (rows) among cells corresponding to pathological fibroblasts, intermediate pathological fibroblasts, and non-pathologic fibroblasts (columns). Proteins of high activity in pathologic fibroblasts are highlighted. **k**. Model summarising potential mechanisms that contribute to morbidity and mortality in patients with COVID-19 focusing on impaired cellular regeneration and rapidly ensuing fibrosis.

nature research

Corresponding author(s): Benjamin Izar

Last updated by author(s): 4/5/2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Image acquisition: Zeiss Zen 2012 SP1 (black edition) Vectra 3.0.5 QCapture Suite Plus 3.1.3.10 Leica Scanner Console v102.0.7.5 RT-PCR: Bio-Rad CFX Manager v3.1
Data analysis	Single-nuclei RNA-seq analysis: Cell Ranger v5.0 R v4.0.2 Seurat v3.2.3 scrublet v0.2.1 CellBender v0.2.0 Future v1.20.1 Future.apply v1.6.0 Destiny v3.3.0 hypeR v.1.4.0 Python 3.7.9 Pytorch implementation of Harmony v0.1.5 Scanpy v1.4.4.

SciPy v1.4.1
NumPy v1.18.1
CellPhoneDB v2.0.0
SingleCellExperiment v1.10.1
Viper v1.24.0
ARACNe-AP

Image analysis:

Zeiss Zen blue v2.3
Aperio ImageScope software (v12.4.3.5008)
InForm 2.4.6 (spectral unmixing)
Qpath 0.2.3
Adobe Photoshop v11.0

CY-TOFF data analysis:

Python v3.8.2
numpy v1.18.5
scipy v1.4.1
Scikit-image v0.17.2
Tifffile 2020.6.3
Networks v2.5
Scikit-learn v0.23.2
Pingouin v0.3.7
Scanpy v1.6.0

Other:

Bio-Rad CFX Manager v3.1
Microsoft Excel v16.45
GraphPad Prism v9.0

Custom Code:

All custom Code is publicly available via: https://github.com/IzarLab/CUIMC-NYP_COVID_autopsy_lung.:

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Processed data are made available via the single-cell portal: https://singlecell.broadinstitute.org/single_cell/study/SCP1219.
Processed data has been deposited in GEO under GSE171524: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE171524
Raw data will be available on the Broad Data Use and Oversight System, study ID: DUOS-000130: https://duos.broadinstitute.org.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size | The single nuclei RNA sequencing (snRNA-seq) cohort included n=27 samples from n=26 individuals with 116,314 snRNA-seq transcriptomes (nuclei) included in this study. The COVID-19 cohort included n=20 samples from n=19 individuals (79,636 nuclei) and the control cohort included n=7 samples from n=7 individuals (36,636). There was no sample size calculation performed, because this was not an interventional study; in this presented study, we included all patients who underwent autopsy within < 10 hours of death from COVID-19. There were no additional inclusion or exclusion criteria defined. We attempted to include the largest possible number of donors in this study.

Data exclusions	Data was processed using CellRanger V5.0 followed by CellBender v0.2.0 to remove technical ambient RNA counts and empty droplets. The resulting matrix was processed on a per sample basis in R (v4.0.2) using Seurat (v3.2.3) Filters were applied to include cells with 200–7,500 genes, 400–40,000 UMIs, and less than 10% of mitochondrial reads. Additionally, Scrublet was applied to identify and remove doublets with an expected doublet rate ranging from 4–9.6% based on the loading rate. Samples containing less than 1,000 cells after filtering were excluded from further analyses. After data integration and clustering one low quality cluster (defined by low feature count and high mitochondrial reads) was removed and the data was subclustered based on major cell types. Among the myeloid cells we excluded two low quality clusters (defined by low feature count and high mitochondrial reads). A total of 116,314 cells (from 119,535 passing the initial QC) were included in the final analysis.
Replication	No dedicated replicates were performed for single nuclei RNA-seq. For imaging and quantitative PCR analysis >3 biological replicates were randomly selected from the snRNA-seq cohort for each group. For experiments involving mice each group contained 3 or more animals. All attempts at replication were successful.
Randomization	For snRNA-seq no randomization was performed. This study was performed on archival tissue from COVID-19 decedents or control donors, and no intervention was performed, hence, no randomization was applied. COVID-19 specimens were included based on criteria outlined below. Controls were selected form the CUIMC bio-bank to match median age and gender of COVID-19 cases. Samples for imaging and PCR were picked at random from all samples profiled by snRNA-seq. Mice were selected by their respective genotype and randomly allocated into experimental groups with both sexes.
Blinding	No blinding was performed for snRNA-seq processing and analysis, and RT-PCR as the data and analysis is quantitative and not qualitative in nature. Blinding was performed for quantification of imaging experiments.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Primary: Anti-Dclk1, Abcam-ab37994 Anti-Scgb1a1, Santa Cruz-sc-365992 Anti-Krt5 , BioLegend -905901 Anti-ProSPC, Seven Hills Bioreagents -WRAB-9337 Anti-Ki67, Invitrogen-14-5698-82 Anti-KRT8, DSHB-TROMA-I Anti-CLAUDIN4, Invitrogen -36-4800 Anti-HTII-280, Terrace Biotech-TB-27AHT2-280 Anti-ChAT, Millipore-AB144P Anti-Pou2f3, Sigma -HPA019652 Anti-CD45, BD-553078 Anti-mouse CD45 - PE/Cy7, BioLegend-103114 Anti-mouse CD64 - FITC, BioLegend-139316 Anti-mouse F4/80 - APC, BioLegend-123116 Fc blocking antibody, BD-553142 Anti-hAXL, R&D Systems-AF154 Anti-CD169, Novus-NB600-534 Anti-C4D, Cell Marque-PA0792 Anti-aSMA-FITC, Sigma -F3777 Anti-CD19, BT51E, Leica-NCL-L-CD19-163 Anti-CD8, 4B11, Leica-PA0183 Anti-CD163, 10D6, Leica-PA0090 Anti-CD4, EPR6855, Abcam-ab133616 Anti-GzmB, 11F1, Leica-PA0291 Anti-CD103, EPR4166(2), Abcam-ab129202 Secondary:

Alexa Fluor 488 Donkey Anti-Chicken , Jackson Immuno Research -703-546-155
Alexa Fluor 488 Donkey Anti-Mouse , Jackson Immuno Research -715-545-151
Cy5 Donkey Anti-Goat , Jackson Immuno Research-705-175-147
Cy3 Donkey Anti-Rat , Jackson Immuno Research-712-165-150
Cy3 Goat Anti-Rabbit , Jackson Immuno Research-111-165-045
Biotin-anti-FITC, Abcam-ab6655
Opal 7-color IHC staining kit , Akoya Bioscience-NEL821001KT

All antibodies with their respective dilutions are listed in Supplementary table 12.

| Validation | All antibodies were validated for their specific application by the manufacturer and validation data is available on the manufacturer's website(s). Antibodies were in-house validated for their species-specific on typical staining pattern (e.g. subcellular localization or co-staining with other markers). Antibodies for multiplexed imaging were validated individually before using in the multiplexed assay. |

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	8-12 week old Pou2f3-/- KO and WT mice on C57BL/6 and 129SvEv mixed background (both male and female) were used in the experiments. Mice were housed with a 12h light/dark cycle at 18-23C and 40-60% humidity.
Wild animals	No wild animals were used in this study.
Field-collected samples	No field-collected samples were used in this study.
Ethics oversight	Mouse studies were approved by Columbia University Medical Center (CUMC) Institutional Animal Care and Use Committees (IACUC-AC-AAAW1474).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Patient characteristics reflect the age, gender, and race representation of patients admitted to New York Presbyterian Hospital/Columbia University Medical Center with COVID-19. Control samples were selected to reflect median age distribution of COVID-19 cases included in the study and match the gender distribution.
Recruitment	Patients were cared for at New York Presbyterian Hospital/Columbia University Medical Center. COVID-19 cases were collected between March and June, 2020. Inclusion criteria included real-time reverse transcription polymerase chain reaction (RT-PCR) confirmed infection, consent to perform rapid autopsy and post mortem intervals <10 hours. Frozen control tissues were assessed by a pulmonary pathologist and represent "uninvolved" regions of biobanked tumor resections.
Ethics oversight	Tissues were collected under New York Presbyterian Hospital/Columbia University Medical Center IRB approved protocols IRB-AAAT0785, IRB-AAAB2667, and IRB-AAAS7370.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	14 days post infection with PR8, mice were euthanized and trancardially perfused with 10ml cold PBS. The lungs were perfused with 1 ml PBS with 2mg/ml Dispase I and 0.5 mg/ml DNAse I and incubated for 60 minutes with agitation. Lobes were physically disscociated and filtered with a 40 μM strainer. Cells were pelleted, RBCs were lysed and cells were stained for 10 min at RT using FC blocking antibody and LIVE/DEAD™ Fixable Violet Dead Cell Stain Kit (Thermofisher, ref:L34964). After one wash cells were stained with surface antibodies and samples were acquired after two additional washes.
Instrument	LSR-II (BD) equipped with four lasers (405nm, 488nm, 561nm, 635nm)
Software	BD FACSDiva v9.0, FlowJo V10.6.2

Cell population abundance	F4/80 cells were ~15% of all CD45+ cells
Gating strategy	Cells were gated by SSC/FSC. Single cells were identified by FSC-W/FSC-A and FSC-H/FSC-A. Viable cells were identified using the violet cell stain. Gating on CD45 followed by gating on F4/80 and CD64 double positive cells identified the population for quantification.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.