
























Integrated histopathology, spatial and single cell transcriptomics resolve cellular drivers of early and late alveolar damage in COVID-19

Received: 13 February 2024

Accepted: 21 January 2025

Published online: 10 March 2025

 Check for updates

Jimmy Tsz Hang Lee ^{1,16}, Sam N. Barnett ^{2,3,16}, Kenny Roberts ¹, Helen Ashwin ⁴, Luke Milross⁵, Jae-Won Cho⁶, Alik Huseynov², Benjamin Woodhams ^{1,7}, Alexander Aivazidis¹, Tong Li ¹, Joaquim Majo⁸, Patricia Chaves ², Michael Lee ², Antonio M. A. Miranda ², Zuzanna Jablonska ², Vincenzo Arena⁹, Brian Hanley¹⁰, Michael Osborn¹⁰, Virginie Uhlmann ⁷, Xiao-Ning Xu ¹¹, Gary R. McLean^{2,12}, Sarah A. Teichmann ^{1,13}, Anna M. Randi ^{2,3}, Andrew Filby¹⁴, Paul M. Kaye ⁴, Andrew J. Fisher ^{5,15,17} , Martin Hemberg ^{6,17} , Michela Nosedà ^{2,3,17}  & Omer Ali Bayraktar ^{1,17} 

The most common cause of death due to COVID-19 remains respiratory failure. Yet, our understanding of the precise cellular and molecular changes underlying lung alveolar damage is limited. Here, we integrate single cell transcriptomic data of COVID-19 and donor lung tissue with spatial transcriptomic data stratifying histopathological stages of diffuse alveolar damage. We identify changes in cellular composition across progressive damage, including waves of molecularly distinct macrophages and depletion of epithelial and endothelial populations. Predicted markers of pathological states identify immunoregulatory signatures, including IFN-alpha and metallothionein signatures in early damage, and fibrosis-related collagens in late damage. Furthermore, we predict a fibrinolytic shutdown via endothelial upregulation of *SERPINE1*/PAI-1. Cell-cell interaction analysis revealed macrophage-derived *SPPI*/osteopontin signalling as a key regulator during early steps of alveolar damage. These results provide a comprehensive, spatially resolved atlas of alveolar damage progression in COVID-19, highlighting the cellular mechanisms underlying pro-inflammatory and pro-fibrotic pathways in severe disease.

Since the outbreak of the COVID-19 pandemic in late 2019, SARS-CoV-2 infection remains prevalent, with nearly 700 million cases recorded worldwide and almost 7 million deaths¹. Whilst being a respiratory illness, the severity across infected patients is variable, with critical cases manifesting as a systemic disease with hyperinflammation and cytokine storm,

leading to multiple organ damage and dysfunction. Crucially, endothelial damage and associated coagulopathy contribute to severe forms of the disease². A better understanding of the cellular and molecular mechanisms underlying the devastating lung alveolar damage in COVID-19 could inform novel therapies to the benefit of patients with severe symptoms.

A full list of affiliations appears at the end of the paper.  e-mail: a.j.fisher@newcastle.ac.uk; mhemberg@bwh.harvard.edu; m.nosedà@imperial.ac.uk; ob5@sanger.ac.uk

The predominant histological lung injury pattern in COVID-19 is termed diffuse alveolar damage (DAD). DAD presents with heterogeneous histopathological features and stages, where early or exudative DAD (EDAD) is characterised by hyaline membrane deposition and inflammation, while late or organising DAD (ODAD) is marked by extensive fibrosis, with intermediate states showing mixed pathological features (MDAD)³. Hence, EDAD and ODAD display increasingly severe patterns of lung injury and are thought to represent the temporal progression of disease pathology^{4–7}. However, beyond initial reports documenting expanded immune cells and fibroblast populations in ODAD⁷, the cellular and molecular differences across alveolar damage progression are not known. Furthermore, distinct alveolar damage stages can be spatially intermixed in a given donor's lung samples^{4,5}, obfuscating their molecular signatures in tissue dissociation based single-cell and bulk RNA-sequencing datasets.

Previously, a wide range of assays, including single cell/single nucleus RNA sequencing (sc/snRNA-seq), spatial transcriptomics (ST), and imaging mass cytometry, were successfully applied to study bronchoalveolar fluid and post-mortem lung tissue samples from COVID-19 patients^{8–10}. Moreover, targeted profiling using subsets of genes and proteins demonstrated early immune cell recruitment and inflammatory pathway activation, followed by fibrosis, but did not provide a comprehensive overview of the cellular and molecular changes^{10,11}. Consequently, the exact drivers of tissue remodelling within DAD stages remain incompletely understood.

Here, we combine sc/snRNA-seq and ST to provide a comprehensive cellular and molecular characterisation of alveolar damage progression stages in COVID-19. We integrated 11 datasets to create a large multi-study sc/snRNA-seq atlas, with a newly generated ST dataset profiling histologically defined DAD stages across autopsy lung tissue samples from 33 patients. For each DAD stage, we identified distinct molecular biomarkers, pathological cell states, tissue micro-environments, as well as cell-cell interactions (Fig. 1A). In particular, we identify waves of macrophage subtypes accumulating through progressive alveolar damage and the enrichment of COVID-19-specific *SPPI* (encoding osteopontin - OPN) signalling from macrophage subpopulations in early damage to specific lung tissue niches. Furthermore, we link *SPPI*/OPN signalling to fibrinolytic shutdown via endothelial upregulation of PAI-1 (encoding *SERPINE1*) upon OPN treatment. By combining multimodal transcriptomics data and integrating histopathological definitions of tissue damage, our study provides a framework that can be applied to other organs in health and disease.

Results

Single cell transcriptomic atlas of the healthy and COVID-19 lung

To maximise the number of COVID-19 lung cells and individuals studied, we generated an integrated sc/snRNA-seq atlas comprising newly generated snRNA-seq data and 10 publicly available sc/snRNA-seq datasets (Fig. 1B, C and Supplementary Data 1, 2). Lung tissue from organ donors ($n=77$) were used as a control in comparison with COVID-19 patients ($n=51$). After sample processing and quality control, the resulting integrated object comprised 514,756 cells and nuclei (Fig. 1D, E). Integration of transcriptomic data was performed accounting for variations from donor, cell/nuclei, and 10x Genomics chemistry (Fig. 1D and Supplementary Fig. 1A–C). Leiden clustering was performed, and we identified 12 coarse-grained cell states within four major cellular compartments based on curated lineage-specific gene markers and unbiased differential gene expression (DGE) analysis (Fig. 1F). This included epithelial (EP), endothelial (EC), stromal and immune cells, including myeloid (MYE) and lymphoid lineages (Fig. 1E).

Further subclustering defined 33 distinct cell states (Fig. 1G, Supplementary Fig. 1C, D and Supplementary Data 3, 4), including four EP states and six EC states. The latter included the recently defined

EC.Aerocyte (*HPGD*, *EDNRB*, *ILIRLI*) and two *ACKR1+* venous EC populations (systemic and pulmonary - EC.Venous.Sys, EC.Venous.Pul) distinguished by expression or lack of *COL15A1*, respectively^{12,13} (Fig. 1F and Supplementary Fig. 1D). Vascular smooth muscle cells (Vascular.SMC), pericytes and mesothelial cells were also profiled, along with four fibroblasts (FB) populations within the stromal compartment (Fig. 1F and Supplementary Fig. 1D). Within immune cells, we identified B plasma cells expressing either *IGHA1* or *IGHG1* (B.Plasma.IgA, B.Plasma.IgG), and five macrophage populations, including an alveolar macrophage subset co-expressing markers from 'macro-alv-MT' (including metallothionein (MT) related *MTIF* and *MTIH*) and 'macro-alv-CCL' (including chemokines *CCL4* and *CCL20*) identified in previous studies¹⁴, here termed Macro.Alv.Meta.CCL. (Fig. 1F and Supplementary Fig. 1D). Furthermore, we profiled four lymphoid populations, in addition to two proliferating immune populations (Proliferating.T.NK, Proliferating.MYE). Consistent with previous reports¹⁰, we observed widespread gene expression differences in these cell states between healthy and COVID-19 lungs, including EP.AT1 and EP.AT2 cell states that showed dysregulation of genes associated with interferon response, and as described below, EC states that showed dysregulation of the coagulation cascade (Supplementary Fig. 1E–G and Supplementary Data 5–7).

Transcriptome-wide spatial atlas of alveolar damage progression

Given that DAD stages can be spatially intermixed in COVID-19 lung tissue^{4,5}, we used ST to identify transcriptional differences across histopathologically defined alveolar damage patterns. For this purpose, we examined post-mortem lung tissue samples from a multi-centre patient cohort that we recently characterised using targeted antibody panels^{4,5}. Our cohort included 33 individuals, mostly males, across different ages (22–98 years; median 27 years), who died from severe COVID-19 during the first (ancestral strain) and second wave (B.1.1.7 or alpha variant) of the pandemic, and spanned diverse ethnic backgrounds and clinical histories, including both hospital (70% of patients) and community (30%) deaths (Supplementary Fig. 2 and Supplementary Data 8).

Here, we performed unbiased ST characterisation of alveolar damage stages across our patient cohort using the NanoString Whole Transcriptome assay (WTA), profiling a total number of 326 tissue regions of interest (ROIs) sized 400 μm^2 each (Fig. 2A). As before, we defined DAD stages strictly based on published histological criteria¹⁵ and guided by expert pulmonary pathologists^{4,5} (Fig. 2B, "Methods"). We profiled tissue regions with EDAD, MDAD and ODAD, as well as tissue areas with preserved (PRES) lung morphology contrasting DAD. To identify the molecular and cellular signatures specific to DADs, we also examined COVID-19 patients with pulmonary oedema consistent with acute cardiac failure, as well as superimposed bacterial bronchopneumonia. Although all patients were confirmed SARS-CoV-2-positive via RT-PCR of nasopharyngeal and/or lung tissue swabs, viral abundance was not quantified in the NanoString WTA Profiling. Imaging mass cytometry from a related study found no SARS-CoV-2 spike or nucleocapsid proteins in lung tissue⁴. Finally, as healthy controls, we also sampled lung sections of two patients who died from non-COVID-19 disease and included a publicly available lung WTA dataset of three non-COVID-19 patients¹⁰ in our analysis. In addition, we conducted Visium profiling on six FFPE lung tissue sections obtained from three patients, to validate the findings of our WTA analysis.

We applied the standard WTA data processing workflow, using stringent gene filters (> 6000 genes per ROI) for quality control and background correction using the CountCorrect algorithm that leverages negative WTA control probes¹⁶. This resulted in 260 high-quality spatial transcriptome profiles with an average of 11,423 genes and 689 nuclei detected per ROI (Supplementary Fig. 3). Our final

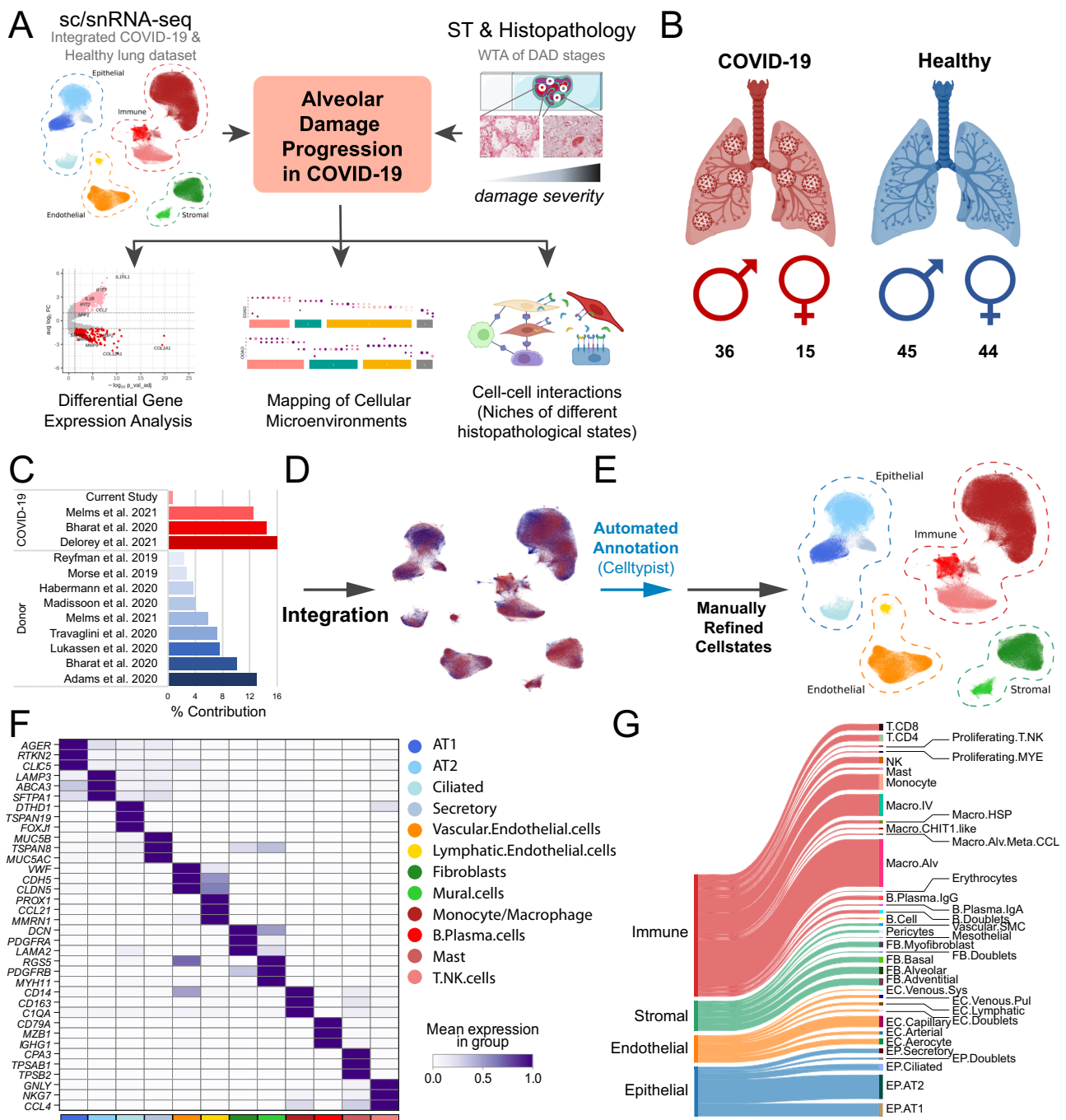


Fig. 1 | Single cell transcriptomic atlas of the healthy and COVID-19 lung.

A Schematic overview. A multi-study sc/snRNA-seq dataset was integrated with histopathology driven spatial whole transcriptome analysis. The cell2location tool was used to map cell types/states to spatial transcriptomic data, with DGE, cell-colocalization, abundance analysis, and cell-cell interaction interrogation performed downstream. Created in BioRender. Barnett, S. (2025) <https://BioRender.com/n08s131>. **B** Number of COVID-19 patients / donors contributing to the integrated sc/snRNA-seq dataset. Created in BioRender. Barnett, S. (2025) <https://>

<https://BioRender.com/h64x272>. **C** Percentage contribution of sc/snRNA-seq datasets from organ donors and COVID-19 patients. **D** UMAP representation of integrated COVID-19 (red) and healthy control (blue) datasets contributing to the final sc/snRNA-seq object. **E** UMAP representation of the global object with broad cell type (dotted) and mid-level annotation. **F** Heatmap representation of markers used for mid-level annotation. **G** Sankey plot visualisation of cell state level annotations derived from subclustering of the broad cell type compartments (Supplementary Fig. 1C, D and Supplementary Data 3, 4).

dataset captures transcriptomic profiles of each DAD stage across multiple patients (Fig. 2C), where multiple DAD stages can often be observed in a given patient (Fig. 2B, C). Hence, our dataset enables a robust comparison of DAD stages accounting for both patient variability and technical batch effects, providing the most comprehensive transcriptomic profiles of alveolar damage progression to date. Our ST

data, along with our sc/snRNA-seq atlas, can be accessed at <https://covid19-multiomicatlas.cellgeni.sanger.ac.uk/>.

Distinct transcriptional signatures of alveolar damage stages

As a first step to determine the molecular signatures associated with different alveolar damage states, we used a random forest classifier for

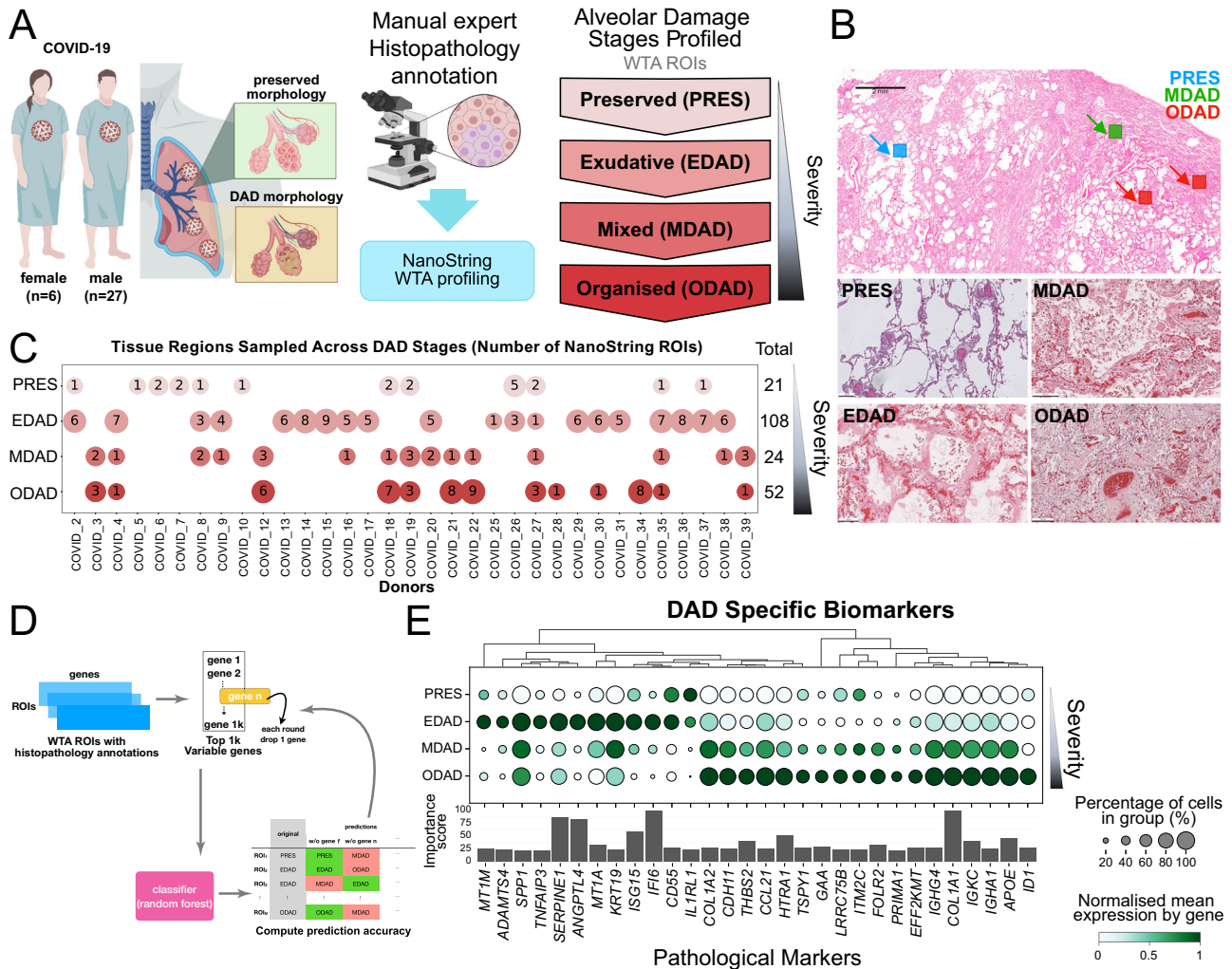


Fig. 2 | Transcriptome-wide spatial profiling of alveolar damage progression in COVID-19. **A** Number of COVID-19 patients contributing to the spatial WTA dataset (left) and histopathological regions of interest (ROI). Created in BioRender. Barnett, S. (2025) <https://BioRender.com/m70f077>. **B** Representative images from H&E stained tissue sections illustrating preserved (PRES) tissue structure and ROIs including different DAD histopathological features. Upper: Indication of multiple ROIs taken from the same tissue section. Lower: Tissue morphology of PRES ($n=21$), EDAD ($n=108$), MDAD ($n=24$) and ODAD ($n=52$) states. **C** Number of

histopathological ROIs for each COVID-19 patient used in the study. **D** Schematic representing random forest classifier approach for predicting pathology-associated gene signatures. **E** Dotplot representation of histopathological state-associated gene signatures obtained from random forest classifier analysis. Dot colour = scaled mean gene expression within a column. Dot size = percentage expressed in group. Importance score = ratio of correct pathology classification in all ROIs.

the identification of predictive biomarkers of PRES, EDAD, MDAD, and ODAD (Fig. 2D and Supplementary Fig. 4A, “Methods”). Within EDAD ROIs, the classifier identified genes associated with inflammatory response (*SPPI*, *IFI6*, *ISG15*, and *TNFAIP3*), regulation of fibrinolysis (*SERPINE1*), and metallothionein-related genes (*MTIA* and *MTIM*) (Fig. 2E). Whilst metallothionein (MT) genes are typically attributed to metal ion homeostasis and oxidative stress alleviation, they are also involved in early-stage inflammatory responses¹⁷, concordant with the pathophysiological phenotype of EDAD in acute lung injury. Further, EDAD is enriched for *ADAMTS4* (aggrecanase-1), a protease upregulated in severe influenza infection, which disrupts lung tissue integrity to enable early immune infiltration by degrading extracellular matrix proteins, including versican¹⁸. Due to a lack of viral particles in our lung samples, these patterns of gene expression in EDAD more likely reflect the inflammatory phenotype of DAD rather than direct viral presence, supporting the “hit and run” hypothesis⁴. Conversely, in ODAD, we observed markers associated with fibrosis and TGF-beta pathway (*COL1A1*, *COL1A2*, *ID1*, *HTRA1*), as well as anti-inflammatory response (*FOLR2*)¹⁹ (Fig. 2E). These predicted biomarkers identified from WTA

data displayed consistent expression patterns across Visium spots annotated as PRES, EDAD, and ODAD histopathological states (Supplementary Fig. 4B). In addition, the top 100 biomarkers exhibited a high correlation between WTA and Visium datasets, with Spearman’s correlation coefficients of 0.56 (p -value = 1.4×10^{-9}) and 0.52 (p -value = 2.6×10^{-8}) in EDAD and ODAD, respectively (Supplementary Fig. 4C).

DGE analysis of EDAD, MDAD and ODAD compared to PRES histopathological states revealed a trend of largely downregulated markers, including the shared downregulation of 371 genes, as well as 34 shared upregulated genes (Supplementary Fig. 5 and Supplementary Data 9–12). EdgeR pseudobulk analysis of EDAD vs ODAD ROIs further highlighted inflammatory and fibrotic signatures within COVID-19 tissue (Fig. 3A). Within EDAD, we observed upregulation of interleukin-related genes *IL1A*, *IL1B* and *IL6*, interferon alpha and gamma-related *IFNG*, *IFIT1*, and *MX1*, and proliferation-related G2/M checkpoint markers *CDK1*, *CDC6*, and *CDC45* (Fig. 3B, C). Furthermore, EDAD enriched for *SPPI*, encoding the matricellular protein OPN, involved in leukocyte recruitment and immune cell activation²⁰, which was previously shown

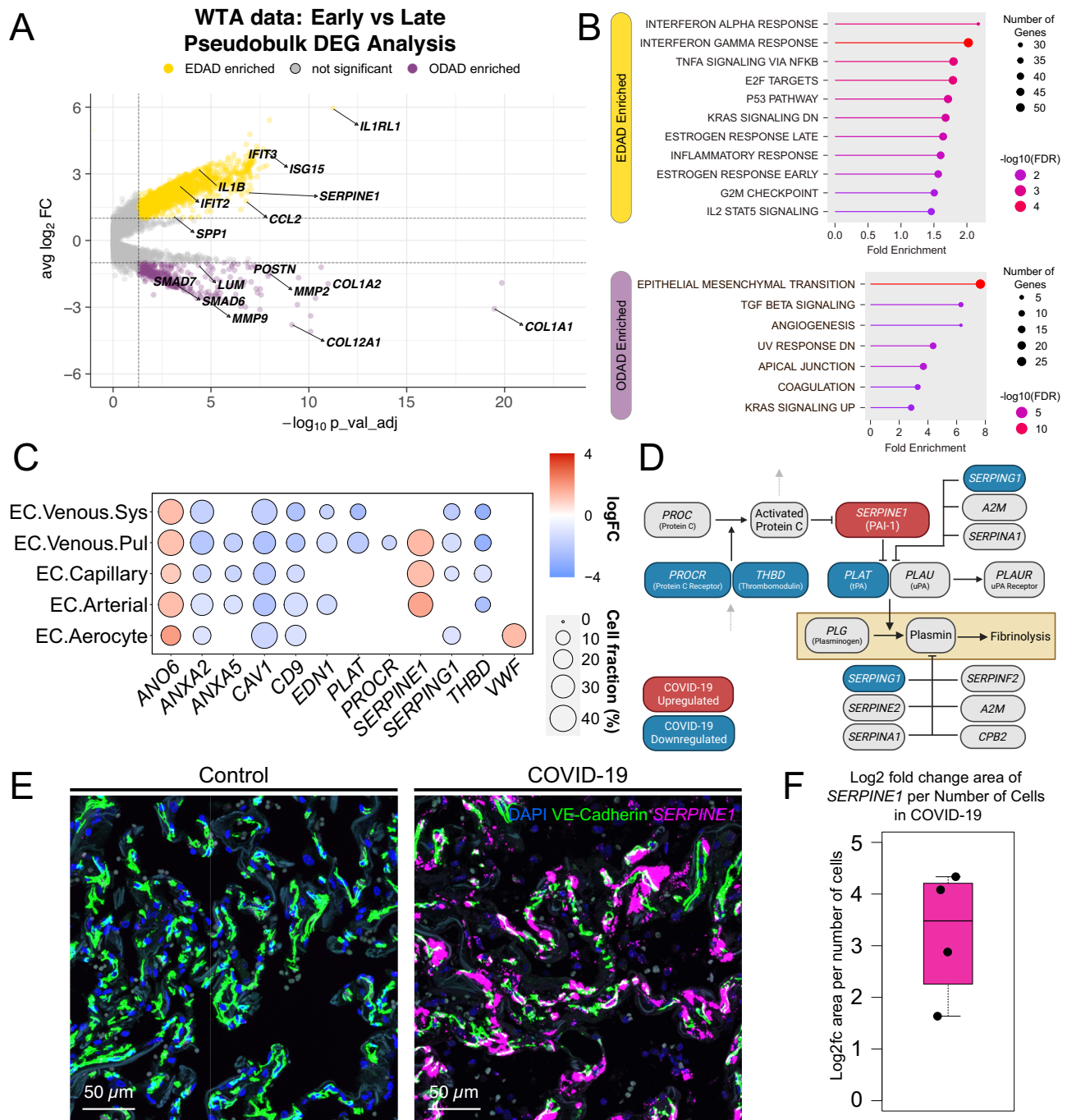


Fig. 3 | Transcriptional progression of alveolar damage progression. **A** Volcano plot representing genes upregulated in EDAD (yellow) and ODAD (purple). Differentially expressed genes were calculated using Pseudobulk EdgeR. **B** Hallmark MSigDB gene set enrichment of genes upregulated in EDAD (top) vs. ODAD (bottom). **C** Dot plot illustrates DGE between COVID-19 and controls related to fibrinolysis and coagulation in the sc/snRNA-seq dataset calculated using Pseudobulk EdgeR with QLF test. Red = upregulated ($\log_{2}FC \geq 1$) in COVID-19, blue = downregulated ($\log_{2}FC \leq -1$) in COVID-19. All data points shown are significant ($FDR < 0.05$). For a complete list of DGE, see Supplementary Data 5. **D** Schematic

representation of the fibrinolysis pathway. Genes upregulated (red) and downregulated (blue) in COVID-19 EC compared to healthy controls are shown. Grey arrows indicate upstream and downstream factors that are not visualised. **E** Healthy control and COVID-19 lung parenchyma samples stained by smFISH for *SERPINE1* (magenta), and CDH5 (VE-Cadherin) protein (green) counterstained with DAPI (blue). **F** Boxplot shows log₂fc area of *SERPINE1* staining per number of cells in COVID-19 lung parenchyma ($n = 4$ patients \times 1 region) compared to healthy controls ($n = 2$ donors \times 1 region). Boxplot elements: centre line, median; box limits, upper and lower quartiles; whiskers, 1.5 \times interquartile range.

to be increased in macrophages of patients with idiopathic pulmonary fibrosis^{21,22}. In contrast, ODAD ROIs enriched for genes associated with extracellular matrix (ECM) turnover, including matrix metalloproteinases (*MMP2*, *MMP9*, and *MMP14*), as well as collagen deposition (*COL1A1*, *COL1A2*, *COL5A1*, *COL6A1*, and *COL6A2*), the pathological fibroblast marker *CTHRC1*²³, and TGF-beta related genes (*SMAD6*,

SMAD7, and *ID1*) (Fig. 3B, C). This signature infers activation of profibrotic and ECM remodelling pathways in ODAD. Taken together, these data present biomarkers to stratify alveolar damage stages and highlight molecular pathways underlying the progression of an inflammatory phenotype in early damage to a profibrotic phenotype observed histologically in late damage.

Dysregulation of the coagulation cascade in COVID-19

With coagulopathy being a major complication in severe COVID-19^{24,25}, and the frequent presence of microthrombi in the lung microvasculature²⁶, we interrogated the expression of genes encoding factors contributing to clot formation and resolution. DGE analysis of the sc/snRNA-seq dataset revealed differential regulation of several genes directly and indirectly involved in the fibrinolysis pathway, including upregulation of *SERPINE1*, encoding the fibrinolysis inhibitor PAI-1, in arterial, capillary, and pulmonary venous EC (Fig. 3C and Supplementary Data 5). Concomitant downregulation of *PLAT*, encoding the fibrinolytic factor tPA, which is inhibited by *SERPINE1*/PAI-1, likely contributes to enhancing anti-fibrinolytic effects (Fig. 3D and Supplementary Data 5). We also observed downregulation of *ANXA2* (Annexin A2), which typically acts as a receptor for *PLAT*/tPA to promote fibrinolysis²⁷, as well as downregulation of *PROCR* (protein C receptor) and *THBD* (thrombomodulin), which together are required for activation of protein C, and downstream inhibition of *SERPINE1*/PAI-1²⁸ (Fig. 3C, D). Other differentially regulated coagulation signatures included downregulation of anticoagulant *ANXA5* (Annexin A5), and upregulation of procoagulant *ANO6* (TMEM16F) (Fig. 3C), which inhibit and promote exposure of phosphatidylserine residues in the phospholipid bilayer, respectively, and in turn, enable prothrombin complex formation^{29,30}. We also observe downregulation of *CAVI* (Caveolin-1), which regulates the anticoagulant activity of TFPI (Tissue Factor Pathway Inhibitor), *EDNI* (Endothelin-1) shown to promote Tissue Factor production³¹, and *SERPING1* (Complement 1 inhibitor) which has both pro- and anticoagulant properties³². Furthermore, while aerocytes are defined transcriptionally by an absence of *VWF*^{32,13}, we observe their upregulation of *VWF* in COVID-19, suggesting a phenotypic shift in the disease.

We confirmed the upregulation of *SERPINE1* in COVID-19 by single-molecule fluorescence in situ hybridisation (smFISH) on COVID-19 post-mortem lung parenchyma tissue (Fig. 3E), observing a several-fold increase in *SERPINE1* mRNA area staining compared to healthy control tissue (Fig. 3F). Collectively, this data suggests that *SERPINE1*, in conjunction with other coagulation related factors, may act as a key player in the fibrinolytic shutdown and persistency of microthrombi observed in COVID-19 lung endothelium from early stages of alveolar damage, particularly since being classified as a marker of EDAD, and upregulated in EDAD compared to ODAD (Figs. 2E, 3A)^{25,33}. In addition, while *SERPINE1* is not differentially expressed in ODAD compared to PRES (Supplementary Data 11), previous studies have demonstrated its overexpression leads to ECM accumulation, as well as fibroblast and AT2 cell senescence in a model of idiopathic pulmonary fibrosis^{34,35}, suggesting long term implications of *SERPINE1*/PAI-1 in fibrosis and advanced DAD states.

Distinct cellular composition of alveolar damage stages

To reveal the cellular composition changes in COVID-19 and across alveolar damage progression, we computationally deconvolved cell states in our ST data via integration with sc/snRNA-seq using the cell2location-WTA model^{16,36} (Fig. 4A). Given the pervasive transcriptional changes in COVID-19¹⁰ (Supplementary Fig. 1E), we derived gene expression signatures of cell states from both healthy and COVID-19 donors from our integrated sc/snRNA-seq atlas and mapped them separately in our WTA dataset (“Methods”). Initially, we compared the abundance of coarse-grained cell states between healthy and COVID-19 ROIs. We observed an increase of immune subtypes (monocyte/macrophages, T & NK, and mast cells) as well as a depletion of epithelial subtypes (EP.AT1, EP.AT2, EP.Ciliated) in COVID-19 tissue (Fig. 4B and Supplementary Fig. 6A). Differential abundance analysis of our sc/snRNA-seq dataset using MiloR³⁷ similarly revealed an enrichment of immune and depletion of epithelial related subpopulations, further confirming destruction of the alveolar bed and increased inflammation observed clinically in COVID-19 patients⁷ (Fig. 4C and Supplementary Fig. 6B, C). In addition, smFISH was used to confirm the loss of AT1 and

AT2 cells in COVID-19 lung parenchyma as indicated by reduced expression of *AGER* and *SFTPC*, respectively (Supplementary Fig. 6D).

Next, we leveraged our integrated sc/snRNA-seq and ST datasets to identify cellular abundance changes across alveolar damage stages (Fig. 4D, E and Supplementary Fig. 7A). First, we examined whether our analysis recovered previously known DAD-associated cellular patterns^{4,5}. Initially, we observed that the healthy gene expression signatures of epithelial subtypes were depleted in both PRES and DAD tissue locations compared to healthy controls (Fig. 4D), consistent with our previous report on disease phenotypes preceding morphological changes associated with DAD⁴. Comparing the distribution of COVID-19 cell state signatures across DAD stages, we found that fibroblast subtypes (FB.Basal and FB.Alveolar) involved in normal alveolar function²³, as well as monocytes associated with early stages of inflammation³⁸ were enriched in EDAD compared to PRES and ODAD (Fig. 4E and Supplementary Fig. 7A). Conversely, fibroblast populations (FB.Myofibroblast and FB.Adventitial) implicated in fibrosis³⁹, were more abundant in ODAD, consistent with ODAD representing a pro-fibrotic phenotype⁶. In addition, lymphoid populations, including T cells (T.CD4, T.CD8) and B cells (B.Plasma.IgA, B.Plasma.IgG) (Fig. 4E and Supplementary Fig. 7A), were enriched in ODAD, recapitulating previous reports⁷, including our previous study on this patient cohort based on proteomic profiling⁴. Of note, while AT1 and AT2 cells were overall depleted in COVID-19 compared to healthy lung tissue (Fig. 4C, D and Supplementary Fig. 6), we observed that their disease cell state gene expression signatures became more abundant across progressive DAD (Supplementary Fig. 7). In concordance with AT1/2 cell disease phenotypes becoming more prominent across DAD, we found that many genes upregulated in AT1/2 cells in COVID-19 in our sc/snRNA-seq dataset (Supplementary Fig. 1E–G) showed increased expression across DAD progression in our WTA data (Supplementary Fig. 8). These results validate our spatial transcriptomic mapping approach and ability to stratify fine-grained cell types across alveolar damage progression.

Beyond previously characterised cell state changes, we observed that distinct macrophage subtypes accumulated through progressive alveolar damage. Macro.HSP, distinguished by heat shock protein-related markers, and Macro.Alv.Meta.CCL, were enriched in EDAD compared to PRES tissue areas (Fig. 4E). The presence of these *MT* + macrophages in early DAD is consistent with our biomarker and DGE analyses (Figs. 2E, 3A). These macrophage populations were slightly reduced in MDAD and ODAD, whereas alveolar macrophages (Macro.Alv) and *CHIT1* + macrophages (Macro.CHIT1.like) increased in abundance (Fig. 4E). Interestingly, these four macrophage populations were either absent or had significantly lower abundances in COVID-19 patients that died from acute cardiac failure, but were present in patients with bronchopneumonia with the exception of Macro.Alv (Supplementary Fig. 7C). This suggests that the superimposed bacterial infection seen in the lungs of some patients with severe COVID-19 develops on the immunological background of virus-driven EDAD. Taken together, our results define specific cellular signatures of DAD stages and identify changes in macrophage subpopulations across alveolar damage progression.

Cellular niches and cell-cell signalling across alveolar damage progression

To explore the changes in cell-cell signalling occurring across alveolar damage stages, we combined sc/snRNA-seq analysis with identification of spatially co-localised cell types and differentially expressed ligand-receptor pairs across DAD stages in the WTA data (Fig. 5A). We first used CellChat⁴⁰ to infer cell-cell communication across all cell states within our sc/snRNA-seq dataset and identify differentially regulated pathways between COVID-19 and healthy control lung samples (Fig. 5B). Broadly, we found enrichment of pro-inflammatory (SPP1) and pro-fibrotic (COLLAGEN, TGFβ) pathways in COVID-19 (Fig. 5B).

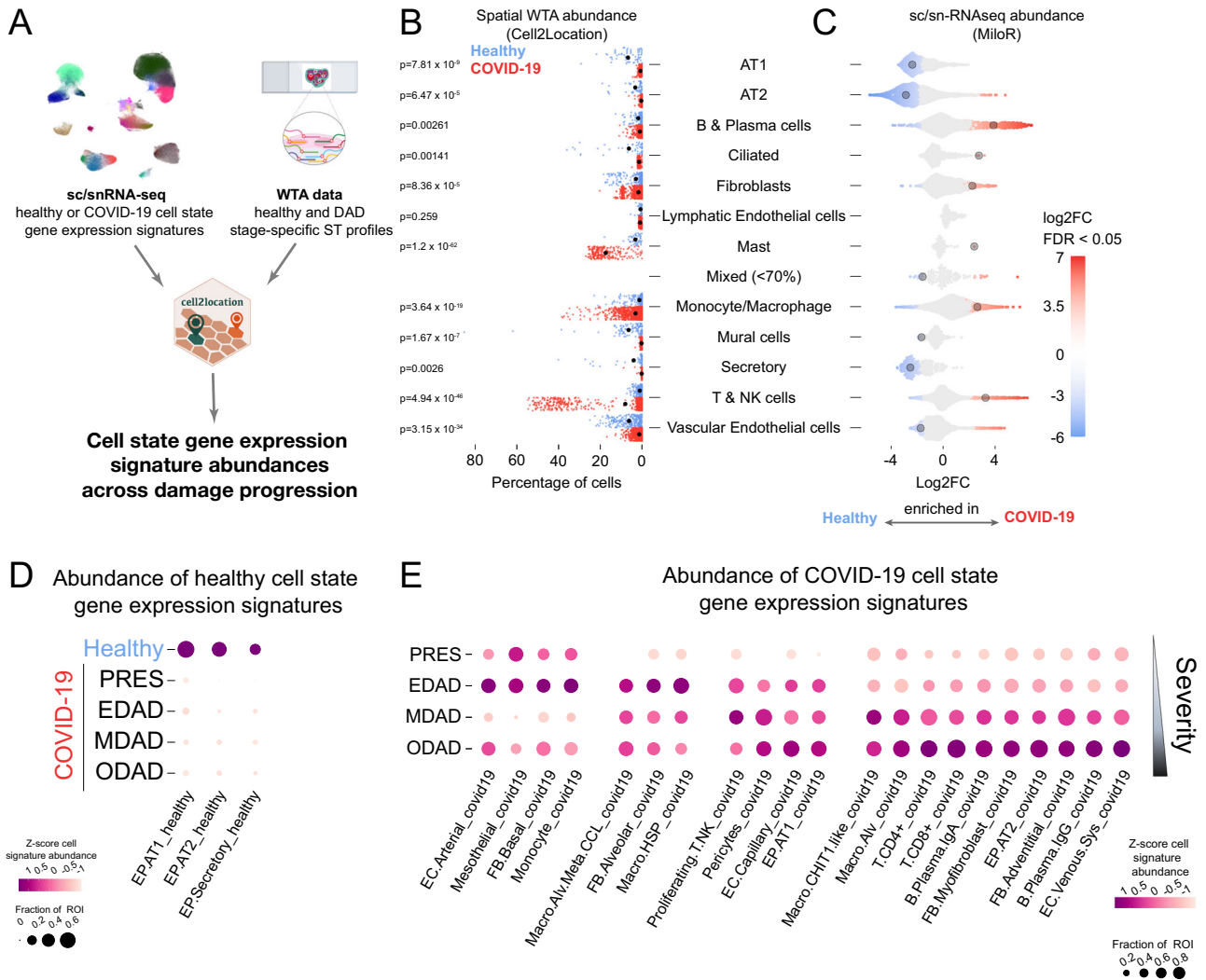


Fig. 4 | Evolving cellular composition across alveolar damage progression.

A Schematic representing the integration of sc/snRNA-seq data with ST data for cell state deconvolution. **B, C** Cell state abundances in WTA and sc/snRNA-seq data. **B** Healthy and COVID-19 cell state gene expression signatures were mapped to healthy and COVID-19 ROIs using cell2location, respectively. The scatter plot shows the percentage of estimated cell abundance enriched in ROIs of COVID-19 (red) or healthy (blue) samples, black spot indicating the median of abundance. A two-sided Welch’s *t* test with the Benjamini-Hochberg adjustment for multiple comparisons

was used. **C** Beeswarm plots illustrate the enrichment (red) or decrease (blue) of neighbourhoods in COVID-19 for each indicated cell type calculated using MiloR (FDR < 0.05). **D, E** The distribution of selected healthy (**D**) and COVID-19 (**E**) cell state signatures across DAD pathologies. Dot plots show cell abundance values that were z-score normalised per cell type across rows/pathologies (colour) and the fraction of ROIs (percent) with cell abundance above the average value of all ROIs in a given pathology (size).

We also observed down-regulation of EC cell-cell adhesion signalling via ESAM, potentially reflecting loss of barrier function observed within the vasculature of COVID-19 patients, as well as downregulation of IL1, IL6 and IFN-II (interferon gamma) signalling, which may reflect an end-stage phenotype in post-mortem tissues processed for sc/snRNA-seq.

Next, we mapped pathological cellular niches across DAD progression, by assigning spatially deconvolved DAD cell states in WTA data into distinct tissue microenvironments. We used non-negative matrix factorisation (NMF) on cell2location estimated abundances to identify spatially co-localising cell states across each DAD stage (i.e. cell states that recurrently co-occur in tissue ROIs across a given pathology). We identified three major pathological niches across the stages of DAD (Fig. 5C). In niche #1, we observed epithelial, mesothelial and immune cell states. In niches #2 and #3, we observed distinct patterns of myeloid-vascular cell colocalization. In niche #2, Macro-Alv.Meta.CCL was found to co-localise with both EC.Aerocyte and EC.Capillary cells, as well as with FB.Alveolar and EP.AT1 cells. While in

niche #3, Macro.Alv co-localised with EC.Venous.Pul and EC.Arterial cells, alongside FB.Myofibroblast and EP.AT2 cells (Fig. 5C). Comparing the abundances of niches across DAD progression, we observed that niches #1 and #2 were established in EDAD and persisted into ODAD. In contrast, niche #3 along with a fourth distinct cell compartment, enriching for FB.Myofibroblast, were more prominent in ODAD (Fig. 5C). These findings present distinct tissue microenvironments across the early and late stages of DAD.

Finally, we examined whether any COVID-19 cellular pathways were distinctly associated with early or late alveolar damage tissue microenvironments. Initially, we selected cellular pathways enriched in COVID-19 in our sc/snRNA-seq dataset (Fig. 5B) and examined the expression pattern of their receptor and ligands in EDAD vs ODAD pathologies in ST data by pseudobulk DGE analysis (Fig. 3A and Supplementary Data 9). This revealed elevated expression of the *SPP1* ligand (encoding OPN) in early DAD (Fig. 3A and Supplementary Fig. 9), which we had previously identified as a biomarker of EDAD (Fig. 2E). *SPP1* receptors and downstream interferon genes⁴¹ were also

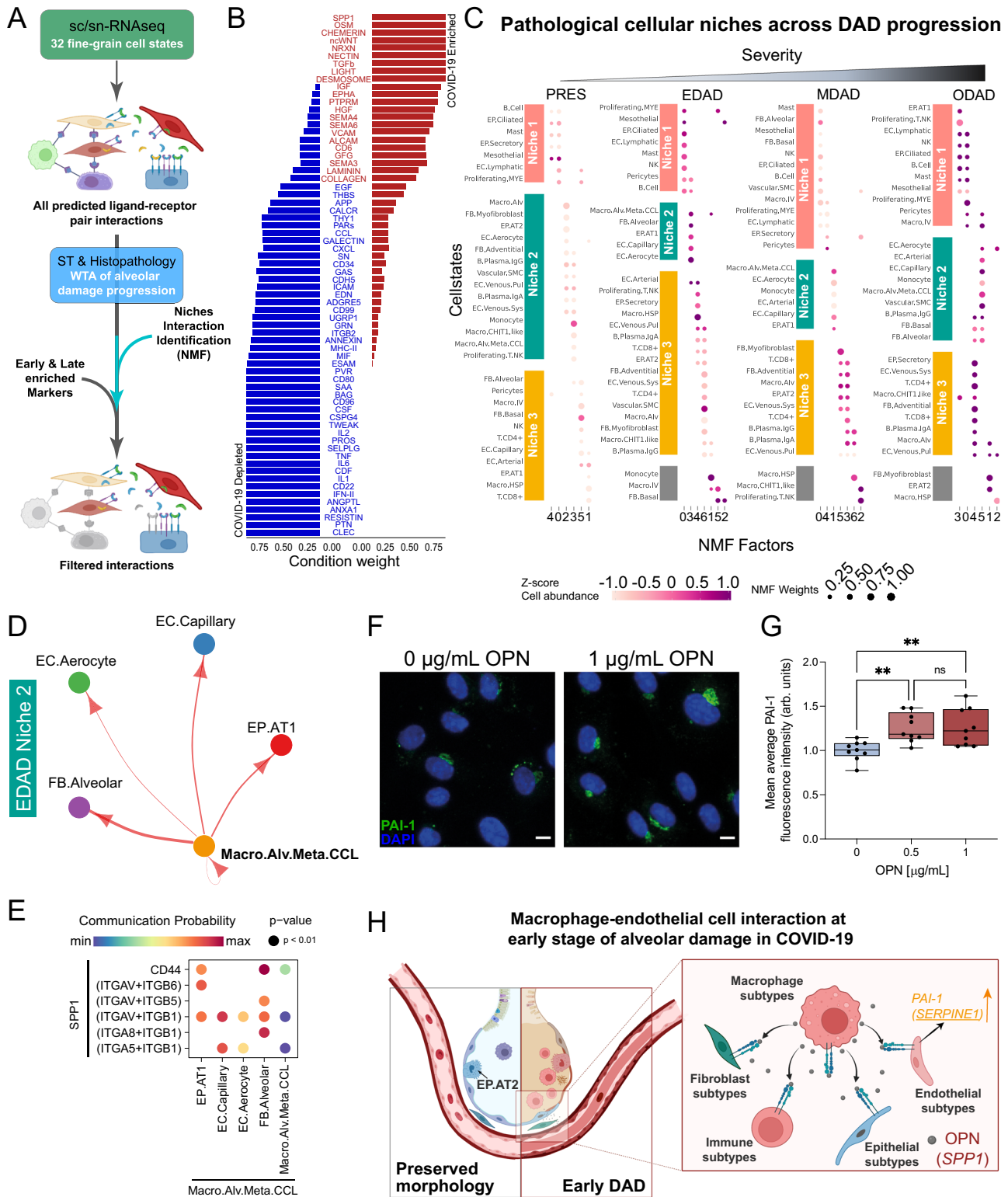


Fig. 5 | Cellular niches and cell-cell signalling across alveolar damage stages. **A** Diagram of analysis approach for mapping cell-cell communication across early versus late DAD. Created in BioRender. Barnett, S. (2025) <https://BioRender.com/n08s131>. **B** Waterfall plot visualisation of global pathway analysis between healthy (blue) and COVID-19 (red) sc/snRNA-seq, with significant pathways highlighted in blue and red respectively. **C** NMF analysis across histopathological states. **D** SPP1 signalling within the COVID-19 sc/snRNA-seq compartment, mapped to EDAD niche 2. Arrows indicate the directionality of ligand (SPP1) signalling to receptors between cell states. **E** Dotplot visualisation of SPP1 signalling to specific receptors across EDAD niche 2 cell states. *P*-values are computed from a one-sided

permutation test. **F** Representative confocal images of endothelial cells treated with 1 μ g/mL rhOPN or vehicle (0 μ g/mL) for 24 h. Scale bar = 10 μ m. **G** Box-whiskers plot of PAI-1 average fluorescence intensity. Values normalised to vehicle control (0 μ g/mL rhOPN). *n* = 3 independent experiments, performed in triplicate. One-way ANOVA; ** adjusted *p* < 0.01. 0 vs. 0.5 adjusted *p*-value = 0.0093. 0 vs. 1 adjusted *p*-value = 0.0086. Boxplot elements: centre line, median; box limits, upper and lower quartiles; whiskers, 1.5 x interquartile range. Arb. units = arbitrary units. **H** Summary of Macrophage subtypes contributing to pro-thrombotic and anti-fibrinolytic states in early DAD through SPP1 signalling. Created in BioRender. Barnett, S. (2025) <https://BioRender.com/fl3w044>.

expressed in EDAD (Supplementary Fig. 9A). We then examined *SPPI* signalling across EDAD cell states and cellular niches. *SPPI* expression was enriched in macrophage subtypes in our COVID-19 sc/snRNA-seq data, including Macro.Alv.Meta.CCL and Macro.HSP that accumulates in EDAD (Supplementary Fig. 9B). In EDAD tissue micro-environments, we identified different types of macrophages as a potential source of *SPPI*: Macro.Alv.Meta.CCL in niche #2 and Macro.Alv and Macro.HSP in niche #3 (Fig. 5D, E and Supplementary Fig. 9C, D). Predicted cells receiving the signals included a range of epithelial, endothelial, mesenchymal and immune populations. Furthermore, we validated enrichment of *SPPI* upregulation in EDAD (Supplementary Fig. 4B), and more specifically, in niches characterised by co-localisation of macrophages and endothelial populations (Supplementary Fig. 10) using Visium ST datasets. These observations present macrophage *SPPI* signalling as an early event during alveolar damage progression.

Previous studies have shown that *SPPI/OPN* can induce TGF-beta expression⁴², which in turn can upregulate *SERPINE1/PAI-1*⁴³. Given the enrichment of both *SPPI* and *SERPINE1* expression in EDAD (Fig. 2E) and the dysregulation of the coagulation cascade in endothelial cells in COVID-19 (Fig. 3), we explored whether *SPPI/OPN* can upregulate *SERPINE1/PAI-1* expression in vasculature in vitro. We first validated our sc/snRNA-seq data by confirming the expression of *SPPI/OPN* receptors integrin alphaV-beta1 and alpha5-beta1 in cultured endothelial cells using immunofluorescence staining (Supplementary Fig. 9E). We then treated endothelial cells with 0.5 and 1 µg/mL rhOPN for 24 h and observed an upregulation of PAI-1 (Fig. 5F, G). Taken together, we demonstrate that macrophage-derived *SPPI/OPN* may be acting at the intersection of pro-inflammatory and anti-fibrinolytic pathways via *SERPINE1/PAI-1* upregulation in early stages of alveolar damage (Fig. 5G, H).

Discussion

One of the critical questions in understanding COVID-19 pathology is the advancing transcriptional regulation of early to late-stage alveolar damage. In this study, we have generated the most comprehensive single-cell and spatial transcriptomic study of lung pathology in severe COVID-19 patients to date, providing a characterisation of the cell states, tissue microenvironments, and cellular interactions that underlie different stages of alveolar damage. By utilising a multiomics approach guided by histological analysis, these data begin to highlight the intricate gene expression changes in COVID-19 lung across various histopathological microenvironments which underlie coagulation, inflammation, and fibrosis, highlighting biomarkers for predicting disease severity and for therapeutic targeting. Our study provides a unique, open-access resource comprising sc/snRNA-seq from COVID-19 and donor cohorts, as well as whole transcriptome spatial RNA-seq data across histopathological states of alveolar damage.

To enable a multi-study analysis of the COVID-19 lung, we present an integrated sc/snRNA-seq atlas comprising newly generated snRNA-seq data and 10 publicly available datasets spanning 128 individuals including COVID-19 patients and donors. Our integration strategy differs from the Human Lung Cell Atlas by incorporating Seurat v4's reciprocal PCA to correct batch effects. Subsequently, we employ Harmony to adjust the Seurat-integrated counts, addressing minor batch effects that remain due to donor and 10X version variations. This method allows us to precisely identify and refine the integration for each specific purpose. From our cell type annotation, we successfully transferred and harmonised the cell state labels from a major lung cell atlas¹⁴ and recovered 33 distinct cell states. We also identified rare macrophage populations, including metallothionein (MT) and chemokine gene-enriched alveolar macrophages (Macro.Alv.Meta.CCL). While we attempted to recapitulate previously described cell state annotations, this was complicated by technical and biological differences between datasets. The difficulties that we faced highlight the consequences of choices relating to how the data was processed and

integrated, as inconsistencies can affect biological interpretations. Addressing these challenges requires community-driven efforts to establish consensus on cell annotations, alongside comprehensive benchmarking to optimise integration methodologies and ensure reproducibility.

By exploring regulators of the coagulation cascade, we demonstrate upregulated *SERPINE1* as a potential key regulator of the hypercoagulable state / fibrinolytic shutdown exhibited by COVID-19 patients in our sc/snRNA-seq DGE analysis. DGE analysis of the WTA data also demonstrated *SERPINE1* upregulation across EDAD regions, suggesting a response in the acute phase of the disease. A similar result was obtained using a random forest classifier trained on WTA ST data to predict markers of disease severity. Previous bulk analyses of lung tissue revealed an increase in *SERPINE1/PAI-1* expression in COVID-19 lung autopsy samples⁴⁴, and increased PAI-1 protein levels in COVID-19 patient blood plasma⁴⁵. Interestingly, *SERPINE1* gene / PAI-1 protein upregulation has been linked to ACE2 inhibition via angiotensin-2 production⁴⁶, potentially suggesting a direct SARS-CoV-2 virus-mediated mechanism for upregulated *SERPINE1/PAI-1*. As such, *SERPINE1/PAI-1* may represent a crucial biomarker for COVID-19 infection, and for targeting macro- and microthrombi therapeutically in specific vessel beds along the vascular axis.

Next, we looked at the changes in the distribution of cell states across alveolar damage progression. Our multi-omic strategy, which integrated ST data with finely annotated cell types from sc/snRNA-seq, uniquely allowed the identification of fine-grained cellular changes across DAD progression and revealed distinct macrophage subtypes in early versus late damage. One limitation of our integrated mapping approach is that our sc/snRNA-seq reference likely represents a mixture of DAD pathological states, which would likely dilute cell state signatures associated with each phenotype.

Lastly, we focused on the molecular signalling pathways within cellular niches of alveolar damage stages and identified a role for *SPPI/OPN* signalling in specific macrophage populations. *SPPI+* macrophages have been associated with severe COVID-19 in recent human lung cell atlas integration efforts^{47,48}, as well as idiopathic pulmonary fibrosis²². Here, we leveraged our ST data to extend these findings and demonstrate that macrophage *SPPI* signalling is most highly enriched in EDAD, though continues through to ODAD and targets various endothelial, epithelial, mesenchymal and immune cells. Interestingly, *SPPI/OPN* targets a number of integrin subunits upstream of TGF-beta activation, potentially suggesting a role for this in the activation of pro-fibrotic pathways. Our analysis also suggests a functional role for *SPPI/OPN* in thrombosis observed in COVID-19 via PAI-1 upregulation. Macrophage-derived OPN has also been implicated in various cancers^{49,50}, suggesting that greater exploration is required to understand the mechanism of action in disease contexts. Furthermore, *SPPI/OPN* inhibition in a mouse model of idiopathic pulmonary fibrosis demonstrated decreased fibrosis, suggesting a candidate for therapeutic targeting of progression to ODAD fibrotic phenotypes in COVID-19 patients²¹. This targeting in the early stages of the disease may also extend to suppression of fibrinolytic shutdown mediated by endothelial *SERPINE1/PAI-1*⁵¹.

While we examined DAD progression in severe COVID-19, our study profiled a limited number of non-COVID-19 patients and was supplemented by published WTA data¹⁰ to enable comparison of early-stage DAD to healthy controls. Due to sample availability, our spatial analysis is limited to post-mortem tissue sampled from predominantly older individuals, with a bias towards male patients. The sample size for the spatial data was insufficient for robust conclusions to determine the contribution of co-morbidities and medical history. For orthogonal validation of our tissue niche assignments from WTA data, we provide a small Visium dataset. Future studies utilising more extensive Visium profiling or high-resolution spatial transcriptomics technologies⁵² can provide a deeper view of tissue niche organisation and cellular interactions underlying DAD progression.

Taken together, our study provides a unique resource to investigate the cellular and molecular landscape of alveolar damage progression within COVID-19 lung tissue at single cell and spatial resolution. This discovery study also provides mechanistic inferences and a baseline to explore novel putative therapeutic targets for early and late stages of alveolar damage. Our data is available for interactive browsing and download at our webportal under <https://covid19-multiomicatlas.cellgeni.sanger.ac.uk/>.

Methods

Human lung tissue procurement and ethics

Human lung samples from patients who died with severe COVID-19 were obtained from 3 UK-based biobanks. The Newcastle Hospitals CEPA Biobank (NHS ethics 17/NE/0070). The Imperial College Healthcare Tissue Bank (ICHTB), supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London (NHS ethics 22/WA/0214). The ICECAP tissue bank at the University of Edinburgh (NHS ethics 16/ED/0084). Work on these samples at the University of York was approved by the Hull York Medical School Ethics Committee (20/52). Additional samples from control donors were obtained from Cambridge Biorepository for Translational Medicine (CBTM), Addenbrooke's Hospital, Cambridge, under approval from the East of England - Cambridge South National Research Ethics Service Committee (REC 15/EE/0152). The collection of clinical metadata is described in detail in our previous studies^{4,5} and the metadata is presented in Supplementary Data 8.

Tissue processing

Fresh human post-mortem tissue samples were either snap-frozen in liquid nitrogen and stored at -80°C for single nuclei transcriptomics, or fixed in 10% neutral-buffered formalin for 24–72 h before transfer to 70% ethanol and processed to paraffin (FFPE) for WTA profiling.

Single nuclei extraction and library preparation from COVID-19 lung post-mortem tissue

Snap-frozen COVID-19 patient post-mortem lung was mechanically dissociated using a pre-cooled pestle and mortar. Dissociated tissues were transferred to a pre-cooled Dounce homogeniser containing Homogenisation Buffer (HB) (250 mM Sucrose (Sigma Aldrich, Cat#S0389); 25 mM KCl (ThermoFisher, Cat#AM9640G); 5 mM MgCl₂ (ThermoFisher, Cat#AM9530G); 10 mM Tris Buffer pH 8 (ThermoFisher, Cat#AM9855G); 1 μM DTT (ThermoFisher, Cat#P2325); 1xEDTA-Free Protease Inhibitor (Sigma Aldrich, Cat#I1873580001); 0.4U/ μL RNaseIn Plus (Promega, Cat#N2611); 0.2U/ μL SUPERasein (ThermoFisher, Cat#AM2696); 0.1% Triton X-100 (Sigma Aldrich, Cat#T8787) and Nuclease Free Water (Sigma Aldrich, Cat#W4502). Samples were further dissociated with 20 strokes of a loose and tight pestle. Cell suspensions were filtered through a 40 μm strainer (Corning, Cat#352340). Nuclei suspensions were centrifuged at 500 $\times g$ for 5 mins at 4 $^{\circ}\text{C}$. The supernatant was aspirated, and the pellet resuspended Storage Buffer (DPBS- (Gibco, Cat#14190094); 4% Bovine Serum Albumin (Sigma Aldrich, Cat#A3059); 0.2U/ μL Protector RNaseIn (Sigma Aldrich, Cat#03335402001). Nuclei were stained using NucBlue™ Hoechst 33342 (ThermoFisher, Cat#R37605) and incubated for 10 mins on ice. Nuclei were sorted by FACS (FACSAria™III, Beckton Dickinson), gated on size and DNA-containing events (NucBlue™ positive in the 405-450/50 channel) (Supplementary Fig. 11) and collected in fresh Storage Buffer. Nuclei were counted, centrifuged at 500 $\times g$ for 5 mins at 4 $^{\circ}\text{C}$ and resuspended in fresh Storage Buffer at a concentration of 1000 nuclei/ μL . Single-cell 3' gene expression libraries were obtained from single nuclei using 10x Chromium Next GEM Single Cell v3.1 kit. Quality control of cDNA and libraries was performed using Agilent 2100 Bioanalyser High Sensitivity DNA analysis

(Agilent). Libraries were sequenced using the NovaSeq sequencing platform, targeting 50,000 reads per nucleus.

Lung tissue histology and pathology annotation

Formalin-fixed paraffin-embedded lung blocks were obtained from multiple lung regions from each patient and were serially cut and mounted onto slides and stained with haematoxylin and eosin (H&E). The primary slide from each block was imaged using brightfield microscopy and the images were uploaded onto an OMERO webserver to serve as a reference slide for interactive annotation. Regions of interest (ROIs) were selected by a consultant histopathologist with cardiothoracic expertise with sizes ranging from 0.25 mm² (500 μm \times 500 μm) to 1 mm² (1000 μm \times 1000 μm), each being the selection target for a collaborative co-application of multiple advanced pathology technologies conducted across multiple academic centres. An ROI classification framework was developed based on experience gained through an early pandemic pilot population³³ and nomenclature reflected existing published literature^{7,53}. ROI classifications included the temporal phases of DAD - exudative DAD (EDAD), organising DAD (ODAD) and mixed (or 'intermediate') DAD ('MDAD') - as well as bronchopneumonia ('BRON') and pulmonary oedema consistent with acute cardiac failure ('ACFF'). Detailed histological selection criteria and methodology is outlined further in Milross et al.⁴.

NanoString GeoMX whole transcriptome atlas slide preparation

In addition to the use of RNase-free reagents, surfaces, equipment, and staining containers were cleaned using RNase AWAY Surface Decontaminant (Thermo) throughout slide processing. Following RNAscope staining, slides were processed according to the NanoString GeoMX RNA assay protocol (MAN-10087-02). Sections were briefly rinsed in nuclease-free water and then post-fixed in 10% neutral-buffered formalin for 5 min. Fixation was quenched by incubation in 0.1 M glycine, 0.1 M Tris, twice for 5 minutes each, followed by 5 min washing in PBS, whereafter probes were applied immediately, again according to the assay protocol. The Whole Transcriptome Atlas (WTA) probe reagent was diluted in pre-equilibrated buffer R to a final probe concentration of 4 nM and added to each slide, which was covered with a Hybrislip cover (Grace Bio-Labs) and incubated for 18 h at 37 $^{\circ}\text{C}$ in a HybEZ II System, humidified with 2 \times SSC (saline-sodium citrate) buffer. The following day, slides were de-coverslipped by brief rinsing in 2 \times SSC with 0.05% Tween-20, and then washed twice for 25 minutes each in 2 \times SSC, 50% formamide at 37 $^{\circ}\text{C}$, and twice for 5 minutes each in 2 \times SSC at room temperature. Following washing, slides were counterstained with the DNA dye SYTO13. A SYTO13 stock (5 μM) was clarified by centrifugation at 13,000 $\times g$ for 2 min, and then diluted to 500 nM in buffer W prior to staining in the dark for 30 min. Finally, slides were washed twice for 3 minutes each in 2 \times SSC buffer.

NanoString GeoMX ROI collection

Slides were covered with buffer S and loaded into the GeoMX DSP instrument for imaging and collection. Channels: SYTO13 (ex. 466–494 nm, em. 505–527 nm), CD68 (ex. 579–597 nm, em. 608–638 nm). Reference H&E images annotated by pathologists were overlaid and aligned to images. 500 μm \times 500 μm ROIs were selected and transposed using the polygon tool. ROIs were illuminated by UV light to cleave the barcodes and the aspirate was collected. Following collection of sequencing tag-containing aspirates, wells were dried either at room temperature overnight or at 65 $^{\circ}\text{C}$ for 45 mins in a heating block, and then re-suspended in 10 μL of nuclease-free water (Ambion), in order to minimise any differences due to ambient evaporation.

NanoString GeoMX WTA library preparation and sequencing

ROI-derived oligos were each uniquely dual-indexed using the i5 \times i7 system (Illumina). A 4 μL aliquot of each re-suspended ROI aspirate

containing the photocleaved oligos was amplified in a PCR reaction containing 1 μ M i5 and i7 primers and 1 \times NSTG PCR Master Mix. UDG digestion was carried out at 37 °C for 30 min, and then deactivated at 50 °C prior to denaturation at 95 °C for 3 minutes and 18 cycles of amplification: 95 °C for 15 s, 65 °C for 1 minute, 68 °C for 30 seconds. The final extension was conducted at 68 °C for 5 min. Prior to purification, PCR products were combined into sub-pools, groups of ROIs based upon area, in order to permit biasing of sequencing to ensure sufficient coverage of smaller ROIs. Each subpool of PCR products was purified with two rounds of AMPure XP beads (Beckman Coulter) at 1.2 \times sample volume of beads. Pooled libraries were quantified using an Agilent 2100 Bioanalyzer and High Sensitivity DNA Kit. The libraries were pooled in a biased manner into two sequencing reactions, each of which was sequenced with 30PE reads across all four lanes of an Illumina NovaSeq 6000 S4 flow cell at a concentration of 400 pM, with 5% PhiX spike-in, yielding 26 billion reads.

RNAscope in situ hybridisation and immunohistochemistry

FFPE sections for both RNAscope and NanoString WTA were cut at a thickness of 5 μ m using a microtome, placed onto SuperFrost Plus slides (VWR), and baked overnight at 55 °C to dry and ensure adhesion. Tissue sections were then processed using a Leica BOND RX to automate staining with the RNAscope Multiplex Fluorescent Reagent Kit v2 Assay (Advanced Cell Diagnostics, Bio-Techne), according to the manufacturers' instructions. Automated processing included baking at 60 °C for 30 min and dewaxing, as well as heat-induced epitope retrieval at 95 °C for 30 mins in buffer ER2 and digestion with Protease III for 15 mins. For the visualisation of markers prior to NanoString GeoMX profiling, a 3-plex RNAscope was developed using tyramide signal amplification with Opal 570, Opal 620, and Opal 650 dyes (Akoya Biosciences). No nuclear stain was applied at this stage. For validation staining, 3-plex or 4-plex RNAscope stains were developed using Opal 520, Opal 570, and Opal 650 dyes (Akoya Biosciences), as well as TSA-biotin and streptavidin-conjugated Atto 425 (Sigma). Nuclei were counterstained with DAPI at 167 ng/ml.

Confocal Imaging

Imaging of validation RNAscope-stained slides was performed using a Perkin Elmer Opera Phenix High-Content Screening System, in confocal mode with 1 μ m z-step size, using 20 \times (NA 0.16, 0.299 μ m/pixel) or 40 \times (NA 1.1, 0.149 μ m/pixel) water-immersion objectives. Channels: DAPI (excitation 375 nm, emission 435–480 nm), Atto 425 (ex. 425 nm, em. 463–501 nm), Opal 520 (ex. 488 nm, em. 500–550 nm), Opal 570 (ex. 561 nm, em. 70–630 nm), Opal 650 (ex. 640 nm, em. 650–760 nm).

RNAscope quantification

Quantification of *SERPINE1* in COVID-19 compared to donor tissues was performed using ImageJ. Channels were normalised by subtracting the raw image with a Gaussian blur transformation with sigma=5. The value for area/n_cells was calculated by dividing the number of cells (DAPI positive nuclei) and the *SERPINE1* stain area in each sample. Donor control area/n_cell values were averaged, and used as a benchmark for calculating the log₂fc of individual COVID-19 samples.

sc/snRNA-seq integration and QC

Library samples prepared in this study were aligned to reference genome GRCh38-3.0.0 using Cell Ranger version 3.1.0. Publicly available data were downloaded from relevant sources (Supplementary Data 2). Seurat v4 reciprocal PCA (RPCA) was applied to study batches as previously documented, with SCTransform used to correct for single cell or single nuclei as source per study batch, regressing out percent mitochondrial and ribosomal genes⁵⁴. A second batch correction was performed on Seurat integrated counts using Harmony⁵⁵ (v0.0.6), correcting for single cell or single nuclei as source, donor and 10x version. Doublet removal was performed using Scrublet⁵⁶. Further QC was

performed to filter genes (200 < nFeature_RNA < 7,500), counts (400 < nCount_RNA < 40,000), mitochondrial genes (for single cell: percent_mito < 20%, for single nuclei: percent_mito < 5%) and ribosomal genes (for single cell: percent_ribo < 2.0, for single nuclei: percent_ribo < 5%). Downstream analysis was performed in Python (v3) using SCANPY⁵⁷ (v1.8.2), AnnData (v0.8.0), NumPy (v1.26.4), SciPy (v1.12.0), Pandas (v1.5.3), Seaborn (v0.12.2), and Matplotlib (v3.5.2). Clustering was performed on the global integrated object using the Leiden algorithm⁵⁸. Differentially expressed genes for each cluster were calculated using the Wilcoxon Rank Sum Test with Benjamini-Hochberg correction and used to annotate major cell type populations based on known markers. Clusters with similar DGE profiles were merged.

CellTypist automated annotation

sc/snRNA-seq data was annotated using three reference lung models trained on a logistic regression framework provided with the CellTypist tool (v0.1.9), including 'Cells_Lung_Airway', 'Nuclei_Lung_Airway'¹⁴, and 'Human_Lung_Atlas'^{47,58}. Cells and nuclei were subset and underwent automated annotation separately based on the 'Cells_Lung_Airway' and 'Nuclei_Lung_Airway' respectively. 'Human_Lung_Atlas' was applied to the combined cell/nuclei object. Briefly, an initial prediction step was performed using the 'best_match' approach, assigning one cell type label to each individual cell or nucleus, with highly variable genes not restricted. Next, over-clustering was performed using default parameters of the 'majority_voting' step to refine annotations based on unbiased Leiden clustering. Predicted annotations were used to guide manual subclustering annotation.

Manual annotation of major cellular compartments

Major cellular compartments, including epithelial, endothelial, stromal, myeloid, T/NK cells and B cells were subset separately from the global object. For all compartments, the top 5000 highly variable genes were calculated using the highly_variable_genes function in SCANPY⁵⁷. PCA was performed, and batch correction was performed using HarmonyPy⁵⁵. Neighbours were calculated prior to UMAP and Leiden overclustering. DGE was calculated with the 'rank_gene_groups' function, using Wilcoxon Rank Sum Test with Benjamini-Hochberg correction. Manual annotation was performed guided by previous CellTypist predicted annotations in conjunction with known cell type / state markers, with merging of Leiden clusters with similar transcriptional profiles. Sankey plots were generated using PySankey.

Pseudobulk differential gene expression analysis for sc/snRNA-seq data

DGE analysis by EdgeR pseudobulk was performed for each cell state between COVID-19 and donor samples⁵⁹. Differentially expressed genes were calculated using QLF (Quasi-Likelihood F-test). Significant genes were filtered for $|\log_2fc| > 1$ and FDR < 0.05.

Differential abundance analysis of sc/snRNA-seq data

Differential abundance analysis was performed using MiloR as previously documented at cell type level of sc/snRNA-seq data³⁷. Briefly, a Milo object was generated from the global lung object (RPCA + SCT + Harmony integrated). The KNN graph was constructed with $k = 175$ and $d = 30$, and with neighbourhoods defined and counted, differential abundance testing was performed.

Generation of expression matrices from NanoString GeoMx WTA data

DSP sequencing data were processed with the GeoMx NGS Pipeline as described previously¹⁶. In brief, after sequencing, reads were trimmed, merged, and aligned to a list of indexing oligos to identify the source probe. The unique molecular identifier region of each read was used to remove PCR duplicates and duplicate reads, thus converting reads into digital counts. The limit of detection (LoD) in an ROI was defined based

on the mean and standard deviation (s.d.) of log₂-normalised negative probe counts. On the log scale, the calculation is: $LoD = \text{mean} + (2 \times \text{s.d.})$. All ROI with the number of detected genes lower than 6×10^3 was filtered out.

Count correct of GeoMx WTA data

In the WTA assay, technical noise will be introduced due to non-specific probe binding and incomplete transcript capture on tissue samples. To enhance true biological signals and reduce technical noise, the counts in the target matrix were adjusted with negative probe counts using the Python package 'CountCorrect' (v0.01).

Pathological ROI classifier

To create a classifier for discriminating the 4 stages of DAD pathologies (PRES, EDAD, MDAD, and ODAD) of ROIs by gene expression profile from the GeoMx WTA data, the top 20% highly variable genes of the 'CountCorrect' matrix selected by the 'FindVariableFeatures()' function in the R package Seurat were considered for classification. In each round of classification, one gene was excluded and the importance score, i.e. the ratio of correct pathology classified, was computed. The performance of the Support Vector Machine (SVM), Decision Tree, Naive Bayes Classifier, and Random Forest was compared. For validation, a 5-fold validation strategy was performed. Of all classifiers, Random Forest showed the best performance across all ROIs. The following R packages and functions are used:

- SVM: `library(e1071); svm()`
- Decision Tree: `library(rpart); rpart()`
- Naive Bayes Classifier: `library(e1071); naiveBayes()`
- Random Forest: `library(caret); train(method = 'rf')`

Pseudobulk DGE analysis of GeoMx WTA data

DGE analysis of the GeoMx WTA data was performed by EdgeR pseudobulk. The count matrix generated by 'CountCorrect' was used as input. A pairwise comparison between DAD stages of COVID-19 was performed. Significant genes were filtered for $|\log_2fc| > 1$ and $FDR < 0.05$.

Gene ontology analysis

Gene ontology analysis was performed using the ShinyGO tool⁶⁰, with FDR threshold < 0.05 . Differentially expressed genes from pseudobulk analysis of GeoMx WTA data of EDAD vs ODAD stages were used for the gene input, filtered for $FDR < 0.05$ and either $|\log_2fc| > 1$ for EDAD enriched markers (Fig. 3b, upper panel), or $|\log_2fc| < 1$ for ODAD enriched markers (Fig. 3b, lower panel). Hallmark.MSigDB was selected as the pathway database used to analyse genes in ShinyGO.

Cell state deconvolution and abundance estimation of GeoMx WTA data

To perform cell state deconvolution, the python package `cell2location-WTA`^{16,36} was used. The integrated sc/snRNA-seq dataset was first subsetted into healthy and COVID-19 cells to estimate reference cell state gene expression signatures of the respective conditions, which were then used separately to train `cell2location` models. Reference cell state signatures were estimated by taking the mean of sc/snRNA-seq gene expression profiles per cell state. In the deconvolution step, the cell state signatures were used to decompose mRNA counts in WTA ROIs. The healthy control and COVID-19 WTA ROIs were decomposed using healthy and COVID-19 cell state signatures, respectively, with the exceptions of Fig. 4D and Supplementary Fig. 7B where the healthy signatures were mapped onto COVID-19 data. For the deconvolution of COVID-19 WTA data, processing all ROIs in a single batch produced results where most cell states were enriched in ACFF and BRON samples, whereas few were enriched in PRES samples. To mitigate this, the ROIs were processed in two batches of "normal-like" and "altered" cellular morphologies determined by independent

image analysis. Image texture feature descriptors were extracted from DAPI image channels of each ROI using Local Binary Pattern analysis⁶¹ and embedded with UMAP over the first 50 principal components. The inspection of the UMAP showed two distinct clusters with (i) "normal-like" alveolar morphology, including PRES and ACFF samples, versus (ii) "altered" morphology including ODAD, MDAD and BRON samples. EDAD samples were distributed across both clusters and were assigned between them using a random forest classifier trained on PRES and ODAD to represent normal versus altered morphologies.

Both the cell state gene expression profiles and the WTA data were subsetted to 11,101 common genes. The following parameters were used for the `cell2location` model:

- Training iterations: 20,000
- Learning rate: 0.001
- Prior on cells per location: Mean for each ROI was specified as the nuclei counts estimated by the NanoString software for each ROI, based on DAPI stains on the image. Standard deviation was set to 10% of the mean (CV, representing prior strength, of 0.1).
- Prior on cell types per location: Mean of 6. Default CV of 1.
- Cell type combinations per location: Mean of 5. Default CV of 1.
- Prior to the difference between technologies: Mean of 0.5. SD of 0.125. CV of 0.25 for both.

For visualisation, the abundances of cell state gene expression signatures were normalised by the surface area of ROIs. The alternative approach to normalise cell state abundances by the total number of nuclei in each ROI from nuclei segmentation yielded similar results.

Identifying cell state colocalization and tissue microenvironments

Absolute cell state abundance estimates obtained from `cell2location` were divided by pathology and input for NMF to identify spatially interlaced tissue compartments. For each pathology, NMF implemented in the python package 'scikit-learn' was trained for a range of $R = \{6, \dots, 12\}$, and the decomposition into factors was chosen as a balance between capturing pathological stages, splitting known compartments and the cell state signature enriched in specific pathology in Fig. 4D. To identify the microenvironment niche, an NMF loading threshold of 0.25 was applied to keep cell states that have higher co-localisation likelihood in a given pathology. Here, NMF provides an additional layer of deconvolution to identify spatially co-located cell states (i.e. tissue niches) per DAD condition. The default loss function option, i.e. the Frobenius norm and a random initialisation method were used. The normalisation of the weight matrix of NMF is performed by dividing each element in a cell type by the sum of all the elements across factors per cell type as described previously (Kleshchevnikov et al. 2022) so that we could compare the loading of cell type between factors. Cell types were manually assigned to niches based on their NMF weights per factor (e.g., Macro.Alv.Meta.CCL, FB, Alveolar and others clearly enriched in the same factor which was interpreted as Niche 2 in EDAD, Fig. 5C) as well as prior biological knowledge (e.g., endothelial subtypes, EC, capillary and EC.aerocytes, are expected to co-localise and are assigned to Niche 2 in EDAD, Fig. 5C). Microenvironment niches were identified according to the common cell state pattern shared between all 4 stages of DADs. The shared cell states across niches across DAD stages are interpreted to represent the dynamic nature of tissue state transitions across DAD progression.

Inferring cell-cell communication

Cell-cell interactions in COVID-19 and healthy control samples were determined using CellChat (v1.1.3) as previously documented⁴⁰. Log-transformed, normalised gene counts were used without accounting for population size. Using the CellChat pipeline, all potential interactions were calculated, based on COVID-19 or donor lung cell state expression of ligands and receptors. The 'RankNet' function was used to generate heatmaps to compare pathway enrichment between

COVID-19 and healthy control samples across all cell states. To resolve pathological state specificity, the expression of ligands and receptors contributing to significantly enriched pathways were determined between EDAD and ODAD disease states. To infer likely interactions within pathological states, pathways were mapped to relevant disease state niches (calculated using NMF analysis) and visualised using dot plots (netVisual_bubble function), circle plots (aggregateNet function) and chord plots (netVisual_aggregate function). *P*-values displayed in dotplots were computed from a one-sided permutation test, as per CellChat documentation.

OPN treatment and immunofluorescent staining of endothelial cells

Human umbilical vein endothelial cells (HUVECs; Lonza, C2519A) were treated with recombinant human osteopontin at varying concentrations (0 µg/mL, 0.5 µg/mL, 1 µg/mL) (rhOPN; R&D Systems, 1433-OP-050) for 24 hours in serum and supplement-free growth medium (EGM, Lonza, CC-3162). Following the incubation, cells were fixed in 10% formalin in PBS, then either blocked and permeabilized in 4% (w/v) bovine serum albumin (BSA, Sigma Aldrich, A3059) and 0.2% (v/v) TritonX-100 (Thermo Fisher, 85111) in PBS (Gibco, 20012027) (PAI-1 staining) or blocked in 1% (w/v) BSA and 5% (v/v) normal goat serum (EDM Millipore, S26-100ML) in PBS (ITGAV, ITGB1, ITGA5 staining) for 30 mins at RT. Incubation with primary antibodies diluted in BSA/Triton/PBS (Rabbit anti-PAI-1) or BSA/goat serum/PBS (Rabbit anti-ITGAV, Mouse anti-ITGB1, Rabbit anti-ITGA5) was performed overnight at 4 °C. Isotype controls and secondary antibody only stainings were performed as negative controls. Cells were then stained with secondary antibodies (anti-Rabbit 488, anti-Mouse 647) diluted in BSA/Triton/PBS or BSA/goat serum/PBS was performed for 1 hr at RT. Finally, cell nuclei were stained with DAPI (Invitrogen, D1306) for 15 min at RT. High-throughput image acquisition was carried out using Cellomics ArrayScan VTI platform (ThermoFisher), using the HCS Studio with Cellomics Scan Version 6.4.4 software (ThermoFisher). The automated Zeiss Observer Z1 epifluorescence microscope was used to acquire 12 fields per well at 10x magnification. Fluorescence intensity was recorded in channels 1-3, using the filter sets XF93 Hoechst (DAPI), XF93 FITC (Alexa488), and XF93 Cy5 (Alexa647). Confocal imaging acquisition was performed with a Zeiss LSM-780 inverted microscope, using the EC Plan Neofluar 40x/1.3 oil objective at the Imperial College London Hammersmith FILM facility using 405 nm, 488 nm, and 633 nm lasers for excitation. Image processing was performed in Fiji (v.2.1.0).

Antibodies

Antibodies used in the OPN validation experiments include Cell Signalling (49536S) Anti-PAI-1 (E315H) XP(R) Rabbit (1:100 dilution); Abcam (ab150361) Anti-Integrin alpha 5 [EPR7854] Rabbit (1:200 dilution); Abcam (ab179475) Anti-Integrin alpha V antibody [EPR16800] Rabbit (1:200 dilution); Abcam (ab30394) Anti-Integrin beta 1 antibody [12G10] Mouse (1:200 dilution); Cell Signalling (3900S) Normal rabbit IgG isotype (1:300 dilution); Santa Cruz (sc-3877) Normal mouse IgG1 isotype (1:200 dilution); Cell Signalling (4412S) Goat anti-rabbit Alexa Fluor 488 (1:200 dilution); and Cell Signalling (4410S) Goat anti-mouse Alexa Fluor 647 (1:200 dilution).

Statistics and reproducibility

SERPINE1 smFISH experiments were performed on four COVID-19 patients and two donor samples. In vitro HUVEC experiments were performed in triplicate across three independent experiments.

Reporting summary

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

Data availability

The data generated for this manuscript were submitted to the European Genome-phenome Archive under accession numbers EGAS00001004442 for the snRNA-seq, EGAS00001004441 for the Control WTA data, and EGAS00001005817 for the COVID-19 WTA data. Requests for data access will be referred directly to the Data Access Committee Portal at <https://ega-archive.org/dacs/EGAC00001000205> and the documentation can be found at <https://ega-archive.org/access/data-access-committee/dac-portal/>. The integrated sc/snRNA-seq atlas can be accessed on our webportal at <https://covid19-multiomicatlas.cellgeni.sanger.ac.uk/>. The spatial WTA data will be made available under the same portal. The publicly available sc/snRNA-seq datasets used in this study are documented in Supplementary Table 2 and include the following: Melms et al.⁹ Broad Data Use and Oversight System (DUOS), study ID DUOS-000130 [<https://duos.broadinstitute.org/datalibrary/>]; Bharat et al.⁸ Gene Expression Omnibus (GEO) accession number GSE158127; Delorey et al.¹⁰ Broad DUOS, study IDs DUOS-000126, DUOS-000127, DUOS-000128 and DUOS-000129 [<https://duos.broadinstitute.org/datalibrary/>]; Reyfman et al.⁶² Database of Genotypes and Phenotypes (dbGaP) accession number phs001750.v1.p1 [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001750.v1.p1]; Morse et al.²² GEO accession number GSE128033; Habermann et al.⁶³ GEO accession number GSE135893; Madissoon et al.⁶⁴ European Nucleotide Archive (ENA) project ID PRJEB31843; Travaglini et al.⁶⁵ European Genome-Phenome Archive (EGA) accession number EGAS00001004344; Lukassen et al.⁶⁶ EGA accession number EGAS00001004419; Adams et al.⁶⁷ GEO accession number GSE136831. Source data are provided in this paper.

Code availability

The custom code used for analysis of sc/snRNA-seq and ST data, and for creating the figures are available on GitHub at <https://github.com/thjimmylee/UKCIC-COVID19-paper-figures> [<https://doi.org/10.5281/zenodo.14372128>]. Other codes are available upon request.

References

- Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
- Varga, Z. et al. Endothelial cell infection and endotheliitis in COVID-19. *Lancet* **395**, 1417–1418 (2020).
- Milross, L. et al. Post-mortem lung tissue: the fossil record of the pathophysiology and immunopathology of severe COVID-19. *Lancet Respir. Med.* **10**, 95–106 (2022).
- Milross, L. et al. Distinct lung cell signatures define the temporal evolution of diffuse alveolar damage in fatal COVID-19. *eBioMedicine* **99**, 104945 (2024).
- Ashwin, H. et al. Identification of a protein expression signature distinguishing early from organising diffuse alveolar damage in COVID-19 patients. *J. Clin. Pathol.* **76**, 561–565 (2023).
- Cardinal-Fernández, P., Lorente, J. A., Ballén-Barragán, A. & Matute-Bello, G. Acute respiratory distress syndrome and diffuse alveolar damage. New insights on a complex relationship. *Ann. Am. Thorac. Soc.* **14**, 844–850 (2017).
- Erjefält, J. S. et al. Diffuse alveolar damage patterns reflect the immunological and molecular heterogeneity in fatal COVID-19. *EBioMedicine* **83**, 104229 (2022).
- Bharat, A. et al. Lung transplantation for patients with severe COVID-19. *Sci. Transl. Med.* **12**, <https://doi.org/10.1126/scitranslmed.abe4282> (2020).
- Melms, J. C. et al. A molecular single-cell lung atlas of lethal COVID-19. *Nature* **595**, 114–119 (2021).
- Delorey, T. M. et al. COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. *Nature* **595**, 107–113 (2021).

11. Rendeiro, A. F. et al. The spatial landscape of lung pathology during COVID-19 progression. *Nature* **593**, 564–569 (2021).
12. Gillich, A. et al. Capillary cell-type specialization in the alveolus. *Nature* **586**, 785–789 (2020).
13. Schupp, J. C. et al. Integrated single-cell atlas of endothelial cells of the human lung. *Circulation* **144**, 286–302 (2021).
14. Madisson, E. et al. A spatially resolved atlas of the human lung characterizes a gland-associated immune niche. *Nat. Genet.* **55**, 66–77 (2022).
15. Katzenstein, A. L., Bloor, C. M. & Leibow, A. A. Diffuse alveolar damage—the role of oxygen, shock, and related factors. A review. *Am. J. Pathol.* **85**, 209–228 (1976).
16. Roberts, K. et al. Transcriptome-wide spatial RNA profiling maps the cellular architecture of the developing human neocortex. Preprint at <https://doi.org/10.1101/2021.03.20.436265> (2021).
17. Dai, H., Wang, L., Li, L., Huang, Z. & Ye, L. Metallothionein 1: A new spotlight on inflammatory diseases. *Front. Immunol.* **12**, 739918 (2021).
18. Boyd, D. F. et al. Exuberant fibroblast activity compromises lung function via ADAMTS4. *Nature* **587**, 466–471 (2020).
19. Puig-Kröger, A. et al. Folate receptor β is expressed by tumor-associated macrophages and constitutes a marker for M2 anti-inflammatory/regulatory macrophages. *Cancer Res.* **69**, 9395–9403 (2009).
20. Kahles, F., Findeisen, H. M. & Bruemmer, D. Osteopontin: A novel regulator at the cross roads of inflammation, obesity and diabetes. *Mol. Metab.* **3**, 384–393 (2014).
21. Hatipoglu, O. F. et al. Osteopontin silencing attenuates bleomycin-induced murine pulmonary fibrosis by regulating epithelial–mesenchymal transition. *Biomed. Pharmacother.* **139**, 111633 (2021).
22. Morse, C. et al. Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur. Respir. J.* **54**, <https://doi.org/10.1183/13993003.02441-2018> (2019).
23. Tsukui, T. et al. Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis. *Nat. Commun.* **11**, 1920 (2020).
24. Conway, E. M. et al. Understanding COVID-19-associated coagulopathy. *Nat. Rev. Immunol.* **22**, 639–649 (2022).
25. Spyropoulos, A. C. & Bonaca, M. P. Studying the coagulopathy of COVID-19. *Lancet* **399**, 118–119 (2022).
26. Hanley, B. et al. Histopathological findings and viral tropism in UK patients with severe fatal COVID-19: a post-mortem study. *Lancet. Microbe* **1**, e245 (2020).
27. Valls, M. D. et al. Annexin A2-Mediated Plasminogen Activation in Endothelial Cells Contributes to the Proangiogenic Effect of Adenosine A2A Receptors. *Front. Pharmacol.* **12** (2021).
28. Barranco-Medina, S. et al. Rational design of protein C activators. *Sci. Rep.* **7**, 1–10 (2017).
29. Reddy, E. C. & Rand, M. L. Procoagulant phosphatidylserine-exposing platelets in vitro and in vivo. *Front. Cardiovasc. Med.* **7**, 505068 (2020).
30. Arndt, M. et al. Structural basis for the activation of the lipid scramblase TMEM16F. *Nat. Commun.* **13**, 1–17 (2022).
31. Kambas, K. et al. Endothelin-1 signaling promotes fibrosis in vitro in a bronchopulmonary dysplasia model by activating the extrinsic coagulation cascade. *J. Immunol.* **186**, 6568–6575 (2011).
32. Prydzial, E. L. G., Leatherdale, A. & Conway, E. M. Coagulation and complement: Key innate defense participants in a seamless web. *Front. Immunol.* **13**, 918775 (2022).
33. Milross, L. et al. The trajectory of COVID-19 cardiopulmonary disease: insights from an autopsy study of community-based, pre-hospital deaths. *ERJ Open Res.* **8**, <https://doi.org/10.1183/23120541.00303-2022> (2022).
34. Adnot, S., Breau, M. & Houssaini, A. PAI-1: A new target for controlling lung-cell senescence and fibrosis? *Am. J. Respir. Cell Mol. Biol.* **62**, 271–272 (2020).
35. Rana, T. et al. PAI-1 Regulation of TGF- β 1-induced alveolar type II cell senescence, SASP secretion, and SASP-mediated activation of alveolar macrophages. *Am. J. Respir. Cell Mol. Biol.* **62**, 319–330 (2020).
36. Kleshchevnikov, V. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
37. Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D. & Marioni, J. C. Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat. Biotechnol.* **40**, 245–253 (2022).
38. Italiani, P. et al. Profiling the course of resolving vs. persistent inflammation in human monocytes: The role of IL-1 family molecules. *Front. Immunol.* **11**, 490312 (2020).
39. Kendall, R. T. & Feghali-Bostwick, C. A. Fibroblasts in fibrosis: novel roles and mediators. *Front. Pharmacol.* **5**, 91491 (2014).
40. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
41. Plataniias, L. C. Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat. Rev. Immunol.* **5**, 375–386 (2005).
42. Kohan, M., Breuer, R. & Berkman, N. Osteopontin Induces Airway Remodeling and Lung Fibroblast Activation in a Murine Model of Asthma. *Am. J. Respir. Cell Mol. Biol.* **41**, 290–296 (2012).
43. Walton, K. L., Johnson, K. E. & Harrison, C. A. Targeting TGF- β mediated SMAD signaling for the prevention of fibrosis. *Front. Pharmacol.* **8**, 277037 (2017).
44. D’Agnillo, F. et al. Lung epithelial and endothelial damage, loss of tissue repair, inhibition of fibrinolysis, and cellular senescence in fatal COVID-19. *Sci. Transl. Med.* **13**, <https://doi.org/10.1126/scitranslmed.abj7790> (2021).
45. Zuo, Y. et al. Plasma tissue plasminogen activator and plasminogen activator inhibitor-1 in hospitalized COVID-19 patients. *Sci. Rep.* **11**, 1–9 (2021).
46. Kellici, T. F., Pilka, E. S. & Bodkin, M. J. Therapeutic potential of targeting plasminogen activator inhibitor-1 in COVID-19. *Trends Pharmacol. Sci.* **42**, 431–433 (2021).
47. Sikkema, L. et al. An integrated cell atlas of the lung in health and disease. *Nat. Med.* **29**, 1563–1577 (2023).
48. Liao, M. et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* **26**, 842–844 (2020).
49. Gao, W. et al. SPP1 is a prognostic related biomarker and correlated with tumor-infiltrating immune cells in ovarian cancer. *BMC Cancer* **22**, 1–12 (2022).
50. Qi, J. et al. Single-cell and spatial analysis reveal interaction of FAP+ fibroblasts and SPP1+ macrophages in colorectal cancer. *Nat. Commun.* **13**, 1–20 (2022).
51. Ghosh, A. K. & Vaughan, D. E. PAI-1 in Tissue fibrosis. *J. Cell. Physiol.* **227**, 493 (2012).
52. Li, T. et al. WebAtlas pipeline for integrated single-cell and spatial transcriptomic data. *Nat. Methods* **22**, 3–5 (2024).
53. Mauad, T. et al. Tracking the time course of pathological patterns of lung injury in severe COVID-19. *Respir. Res.* **22**, 32 (2021).
54. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
55. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
56. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational identification of cell doublets in single-Cell transcriptomic data. *Cell Syst.* **8**, 281–291 (2019).
57. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 1–5 (2018).

58. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 1–12 (2019).
59. Reichart, D. et al. Pathogenic variants damage cell composition and single cell transcription in cardiomyopathies. *Science* **377**, <https://doi.org/10.1126/science.aba1984> (2022).
60. Ge, S. X., Jung, D. & Yao, R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**, 2628–2629 (2020).
61. Ojala, T., Pietikäinen, M. & Harwood, D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognit.* **29**, 51–59 (1996).
62. Reyfman, P. A. et al. Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **199**, 1517–1536 (2019).
63. Habermann, A. C. et al. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci. Adv.* **6**, eaba1972 (2020).
64. Madisson, E. et al. scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. *Genome Biol.* **21**, 1 (2019).
65. Travaglini, K. J. et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020).
66. Lukassen, S. et al. SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. *EMBO J.* **39**, e105114 (2020).
67. Adams, T. S. et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci. Adv.* **6**, <https://doi.org/10.1126/sciadv.aba1983> (2020).

Acknowledgements

We thank the donors and their families for donating tissue samples and enabling this research. We thank Parisa Amjadi and Dominic Smith from the Xu lab (Imperial College London) for helping to establish COVID-19 sample processing under CL3 conditions, and for assistance in FACS sorting single nuclei samples. We thank Liz Tuck, Grant Calder and Tarry Porter for supporting the histology and generation of WTA data and Kristina Sorg, Erica Pawlak and Stijn van Dorgen for supporting WTA data processing. We thank Alessio Alfieri for supporting our smFISH work. We thank L. Lawrence from the Research Histology Facility at the National Heart and Lung Institute of Imperial College London. We thank Kerstin Meyer, Ana-Maria Cujba, Amanda Oliver (Wellcome Sanger Institute), as well as Gisli Jenkins, Alison Johns and Claire Peghaire (Imperial College London) for valuable discussions. We thank Jessica Cox and Martin Prete for establishing the sc/snRNA-seq and ST data portal and website. We thank Olivia Swann and Wendy Barclay for supporting COVID-19 tissue procurement. This research was funded in whole, or in part, by the Wellcome Trust Grants WT206194 and 220540/Z/20/A funding to O.A.B.; the UKRI (MRC) and DHSC (NIHR) for the UK Coronavirus Immunology (UK CIC) consortium award referenced MR/V027638/1 and MR/V028448 to A.F., P.M.K., A.J.F., and O.A.B.; a National Heart and Lung Institute PhD studentship, Imperial College London, to S.N.B.; the Chan Zuckerberg Foundation (2019-002431 and 2019-202666, 2021-237882) to M.N., and S.A.T.; the BHF Centre for Research Excellence Imperial College London (RE/18/4/34215 and RE/24/130023) to S.N.B., A.M.R., and M.N.; the British Heart Foundation and Deutsches Zentrum für Herz-Kreislauf-Forschung (BHF/DZHK: SP/19/1/34461) to M.N. and S.A.T.; the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC) funding to A.M.R., and M.N.; the Imperial College London COVID-19 Response Fund to A.M.R. and M.N.; a British Heart Foundation PhD studentship to Z.J.; a General Sir John Monash Scholarship awarded by the General Sir John Monash Foundation and a Vice-Chancellor's Global Scholarship from Newcastle University to L.M.; startup funds from the Evergrande Centre to J.W.C. and M.H.; the Helmsley Foundation for J.W.C. The Imperial College Healthcare NHS Trust Tissue Bank is funded by the National Institute for Health Research

(NIHR) Biomedical Research Centre. This publication is part of the Human Cell Atlas (www.humancellatlas.org/publications/).

Author contributions

J.T.H.L., S.N.B., A.J.F., P.M.K., M.H., M.N., and O.A.B. conceived the study. S.N.B., P.C., B.H., M.O., A.F., P.M.K., A.J.F., and M.N. procured human lung tissue samples. S.N.B. and P.C. generated snRNA-seq data. L.M., J.M., A.F., P.M.K., and A.J.F. developed the DAD profiling strategy and identified histopathological DAD stages on lung tissue samples. V.A. and J.M. contributed to the pathological assessment of lung tissue samples. K.R. and H.A. generated NanoString WTA data. K.R. generated the smFISH data. S.N.B. and Z.J. performed the SPP1 in vitro experiments. J.T.H.L. and S.N.B. led sc/snRNA-seq and ST data analysis. J.W.C., A.A., A.H., A.M.A.M., and M.L. contributed to sc/snRNA-seq and ST data analysis. S.N.B., T.L., and A.M.A.M. analysed smFISH data. B.W. and V.U. performed an image analysis of WTA data. X.-N.X., G.R.M., S.T., A.M.R., A.F., P.M.K., A.J.F., M.H., M.N., and O.A.B. provided project supervision and data interpretation. J.T.H.L., S.N.B., A.J.F., M.H., M.N., and O.A.B. wrote the manuscript with feedback from all authors.

Competing interests

In the past 3 years, S.A.T. has consulted for or been a member of scientific advisory boards (SABs) at Qiagen, OMass Therapeutics, Xaira Therapeutics and ForeSite Labs, and a non-executive director of 10x Genomics. She is a co-founder and equity holder of TransitionBio and Ensocell, and a part-time employee at GSK. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-56473-x>.

Correspondence and requests for materials should be addressed to Andrew J. Fisher, Martin Hemberg, Michela Nosedà or Omer Ali Bayraktar.

Peer review information *Nature Communications* thanks Martijn Nawijn and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

¹Wellcome Sanger Institute, Hinxton, UK. ²National Heart and Lung Institute, Imperial College London, London, UK. ³British Heart Foundation Centre of Research Excellence, Imperial College London, London, UK. ⁴York Biomedical Research Institute, Hull York Medical School, University of York, York, UK. ⁵Newcastle University Translational and Clinical Research Institute, Newcastle upon Tyne, UK. ⁶The Gene Lay Institute of Immunology and Inflammation, Brigham and Women's Hospital, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. ⁷European Bioinformatics Institute, European Molecular Biology Laboratory (EMBL), Cambridge, UK. ⁸Department of Cellular Pathology, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK. ⁹Department of Woman and Child Health and Public Health, Fondazione Policlinico Universitario A. Gemelli IRCCS, Istituto di Anatomia Patologica, Università Cattolica Del Sacro Cuore, Rome, Italy. ¹⁰Department of Cellular Pathology, Northwest London Pathology, Imperial College London NHS Trust, London, UK. ¹¹Department of Infectious Disease, Imperial College London, London, UK. ¹²London Metropolitan University, London, UK. ¹³Cambridge Stem Cell Institute & Department of Medicine, University of Cambridge, Cambridge, UK. ¹⁴Biosciences Institute and Innovation, Methodology and Application Research Theme, Newcastle University, Newcastle upon Tyne, UK. ¹⁵Institute of Transplantation, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK. ¹⁶These authors contributed equally: Jimmy Tsz Hang Lee, Sam N. Barnett. ¹⁷These authors jointly supervised this work: Andrew J. Fisher, Martin Hemberg, Michela Nosedà, Omer Ali Bayraktar. ✉ e-mail: a.j.fisher@newcastle.ac.uk; mhemberg@bwh.harvard.edu; m.nosedà@imperial.ac.uk; ob5@sanger.ac.uk