Unbiased single cell spatial analysis localises inflammatory clusters of immature neutrophils-CD8 T cells to alveolar progenitor cells in fatal COVID-19 lungs


Praveen Weeratunga*, Laura Denney*, Joshua A. Bull*, Emmanouela Repapi*, Martin Sergeant*, Rachel Etherington†, Chaitanya Vuppusetti†, Gareth D.H. Turner†, Colin Clelland§, Amy Cross∥, Fadi Issa†, Carlos Eduardo de Andrade†, Ignacio Melero Bermejo‡, David Sims§, Simon McGowan§, Yasemin-Xiomara Zurke§, David J. Ahern§, Eddie C Gamez§, Justin Whalley§, Duncan Richards∥∥, Paul Klenerman∥∥, Claudia Monaco§, Irina A. Udalova§, Tao Dong†, Graham Ogg†, Julian C. Knight∥∥, Helen M. Byrne∥∥∥, Stephen Taylor§, Ling-Pei Ho∥∥∥.

*Joint first authors

Corresponding authors

Stephentaylorbioinformatics@gmail.com
Ling-pei.ho@imm.ox.ac.uk

1. MRC Human Immunology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, UK
2. Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, UK
3. MRC WIMM Computational Biology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, UK
4. Department of Cellular Pathology and Radcliffe Department of Medicine, Oxford University Hospitals NHS Foundation Trust, UK
5. Anatomic Pathology, Weill Cornell Medical College, Doha, Qatar
6. Nuffield Department of Surgical Sciences, University of Oxford, UK
7. Navarra Institute for Health Research, Pamplona, Spain
8. Kennedy Institute for Rheumatology, University of Oxford, UK
9. Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, UK
10. Nuffield Department of Orthopedics, Rheumatology and Musculoskeletal Diseases, University of Oxford, UK
11. Nuffield Department of Medicine, University of Oxford, UK
12. Ludwig Institute for Cancer Research, University of Oxford, UK

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Single cell spatial interrogation of the immune-structural interactions in COVID-19 lungs is challenging, mainly because of the marked cellular infiltrate and architecturally distorted microstructure. To address this, we developed a suite of mathematical tools to search for statistically significant co-locations amongst immune and structural cells identified using 37-plex imaging mass cytometry. This unbiased method revealed a cellular map interleaved with an inflammatory network of immature neutrophils, cytotoxic CD8 T cells, megakaryocytes and monocytes co-located with regenerating alveolar progenitors and endothelium. Of note, a highly active cluster of immature neutrophils and cytotoxic CD8 T cells, was found spatially linked with alveolar progenitor cells, and temporally with the diffuse alveolar damage stage. These findings provide new insights into how immune cells interact in the lungs of severe COVID-19 disease. We provide our pipeline [Spatial Omics Oxford Pipeline (SpOOx)] and visual-analytical tool, Multi-Dimensional Viewer (MDV) software, as a resource for spatial analysis.
Introduction

Since the first reports of COVID-19 cases in Dec 2019, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused more than 6 million deaths worldwide (Wang et al., 2022), mainly from respiratory failure. Similarities between COVID-19 and other viral infections of the lungs like SARS and influenza have been noted, but there are specific differences which may be indicative of underlying disease mechanisms unique to COVID-19. In particular, patients with COVID-19 have excess incidence of thromboembolic disease, endothelial damage, and greater acute and long-term impact on organs other than lungs (Gorog et al., 2022; Lamers and Haagmans, 2022; Nalbandian et al., 2021; Pustake et al., 2022). High-resolution immune studies in the blood have shed light on the potential mechanisms for severe COVID-19 disease, with evidence supporting myeloid cell overactivation and dysregulation, T cell exhaustion and cytokine hyperactivation (COMBAT, 2022; Kusnadi et al., 2021; Mann et al., 2020; Schulte-Schrepping et al., 2020; Unterman et al., 2022). Our recent comprehensive multi-modal study of circulating immune cells (COMBAT study) and several other major studies have also concluded that a key hallmark of severity was emergency myelopoiesis (Chen et al.; COMBAT, 2022; Mann et al., 2020; Schulte-Schrepping et al., 2020; Wang et al., 2021b), characterized by raised circulating immature neutrophils, cycling monocytes, and raised haematopoietic progenitors. However, it is not known how these findings in blood relate to damaged lung structural cells and other immune cells in the lungs, nor if they formed injurious immune entities.

Interrogation of the immune response in COVID-19 lungs have lagged behind studies in peripheral blood. Our understanding of the immune response in the lungs is derived mostly from several single cell and single nucleus RNA sequencing studies which have provided valuable insights on a transcriptomic level (Bharat et al., 2020; de Rooij et al., 2022; Delorey et al., 2021; Melms et al., 2021; Wang et al., 2021a; Wendisch et al., 2021). However, these are limited by a lack of high resolution (cell level) spatial context. Transcriptomics studies are also restricted by a lower detection rate for neutrophils as these cells possess relatively low RNA content and high levels of RNases and other inhibitory compounds which confound their identification. Studies in intact COVID-19 lung tissue are challenging, owing in part to the complex and distorted lung micro-architecture which makes deciphering cellular connectivity and organisation challenging. An initial evaluation of COVID-19 lung tissue using imaging mass cytometry by Rendeiro et al concluded that there was greater spatial proximity between macrophages, stromal cells and fibroblasts in patients whose lung
samples were obtained later in infection (Rendeiro et al., 2021). This was in keeping with a repair response in the later stages of disease.

In this study, we adopted an approach which focused on unbiased identification of co-locating cells and their spatial organisation in relation to immunopathology in the lungs. Our aim was to deconvolute a highly disordered immune and structural landscape by adapting mathematical tools usually used in network science. To do this, we developed a bespoke mathematical package to identify statistically significant cellular-level co-location and contact-based cellular networks, and enumerate the strength of this co-location, without a priori specification of neighbourhood environment. Applying these methods, we identified a cluster of closely apposed immature neutrophils and CD8 T cells with high immune activity, which were spatiotemporally associated with proliferating alveolar epithelium and lungs with extensive alveolar damage. Additional associations were also evident between monocytes and this nidus of inflammation, and between monocytes, megakaryocytes and endothelium. These findings raised the possibility of an injurious entity generated by the interaction between immature neutrophils and a specific subset of CD8 T cells, and a key role for these cells and classical monocytes in the immunopathology of severe COVID-19 pneumonitis.
RESULTS

An integrated pipeline to uncover and quantify spatial association amongst cells in the lungs

Our first task was to establish a method to quantify statistically significant spatial correlations between highly-resolved immune and structural cell types in our lung tissue sections. To do this, we developed a new analytical pipeline which combined an immunology-centric annotation approach with a 3-step spatial association analysis [quadrat correlation matrices (QCMs), cross-pair correlation functions (cross-PCFs) and adjacency cell network (ACN)] to provide a set of statistically rigorous spatial analytical output (Figure 1A and Figure S1, and described in detail in Methods). In brief, we first used the QCM to identify cell pairs that are statistically significantly correlated in cell counts. Correlated cell pairs, of types A and B say, were then examined for co-location above random spatial association (using cross-PCF). If the cross-PCF, \(g(r)\), is greater than 1, then cells of type B are observed more frequently at distance \(r\) from cells of type A than would be expected under complete spatial randomness (CSR). We considered \(g(r=20)\), the value of the cross-PCF at \(r=20\), as a means of quantifying how many more cells of type B are observed at distance 20\(\mu\)m from an anchor cell of type A than under CSR. We then examined whether the co-locating cell pairs were physically in contact with each other using a spatially embedded ‘adjacency cell network’ (ACN). Using the ACN, we computed the proportion of cells of type A that were in contact with at least one cell of type B, (full description found in Methods).

These work packages were integrated computationally into a workflow of Python and R based command line tools which may be run individually or as an automated pipeline (Spatial Omics Oxford pipeline; SpOOx) (Figure 1A). The pipeline is supported by a visualisation platform [Multi-Dimensional Viewer (MDV)] (Video V1). Both are available as an open access online resource (see Methods for link).

Lung sections from fatal COVID-19 pneumonitis can be divided into temporally progressive histopathology states
We started by examining formalin-fixed paraffin-embedded (FFPE) lung sections from a cohort of patients who died from PCR-positive COVID-19 pneumonitis from one hospital (University of Navarra, Spain) (n=12). Samples were obtained at the point of death and fixed immediately, markedly reducing post-mortem tissue deterioration (Melero et al., 2022). All samples were collected during the first wave of the pandemic in 2020, before vaccination and repeat infection with SARS-CoV-2. Sections from healthy control (HC) lungs (n=2) were obtained from the Oxford Radcliffe Biobank (Oxford University Hospitals NHS Foundation Trust, UK).

Six of 12 patients were mechanically ventilated (range 6-23 days). All but three were receiving corticosteroids at the point of death (Table S1). In all patients, thoracic CT scans closest to the day of death demonstrated typical and extensive COVID-19 pneumonitis comprising ground glass changes and consolidation (Table S1). Five of 12 lung sections showed evidence of both PCR and immunostaining for SARS-CoV2 Nucleocapsid protein; 3 were PCR+ but protein negative (Table S2). Four sequential lung sections (6 μm thick) were used for haematoxylin and eosin (H&E), 37-plex metal-tagged antibody panel staining, and selected immunofluorescence validation sequentially.

Initial histopathology analysis (independently performed by two senior pathologists with expertise in lung and infectious disease, and a senior respiratory clinician) revealed a highly distorted lung architecture with extensive cellular infiltrate in all samples, changes previously observed in post mortem studies of COVID-19 lung sections(Angelas Montero-Fernandez and Pardo-Garcia, 2021; Carsana et al., 2020; Mauad et al., 2021; Menter et al., 2020; Xu et al., 2020; Youd and Moore, 2020). However, all our sections can be categorised into three formal histopathology classifications of alveolitis (ALV), diffuse alveolar damage (DAD) or organising pneumonia (OP)(Borczuk et al., 2020; Vadasz et al., 2021) (Figure 1B and Figure S2). ALV was characterised by thickened alveolar epithelial wall and septae with immune cell infiltrate and congestion of alveolar walls; DAD, by widespread alveolar epithelial lining injury accompanied by hyaline membrane, regenerating/proliferating Type II alveolar epithelium and interstitial oedema, while OP depicts a repair state typified by presence of fibroblasts, proliferation of alveolar epithelium and collagen presence around bronchial epithelium (Vadasz et al., 2021). In keeping with this, patients with dominant OP histopathology showed a trend of being sampled furthest away from their first symptoms (Figure 1C), had longer periods of stay in hospital and were mechanically ventilated for longer (Table S1 and Figure S3) (no statistical difference observed). All 5 patients with evidence of dual SARS-CoV-2 N protein and PCR expression had lung sections that showed DAD. No sections with OP were positive for SARS-CoV-2 protein staining (Table S2; Figure S4). There were no associations between histopathology states and clinical
features (age, drugs used, co-morbidities, or C-reactive protein (CRP) nearest the point of death) (Table S1, Figure 1D).

These results provide a histopathology-based temporally progressive states for further analysis. 2-3 regions of interest (ROIs) per patient (total of 4 mm² area per patient), selected as representative areas for the dominant histopathology state, were drawn for ablation.

Identification of a specific immature neutrophils - CD8 T cell cluster with high immune activity.

After preliminary staining with an initial panel (Table S3, Figure S5) on a ‘sentinel’ cohort, we designed a final panel which incorporated the most abundant structural and immune cell types (Figure S7A). Single cell segmentation performed using Mesmer library from the DeepCell algorithm (Greenwald et al., 2022) resulted in 677,623 single cells from all ROIs, [ALV (n=10 ROIs), DAD (n=8), OP (n=8) and HC (n=4)] (Figure 2A and Figure S7 and S12). Cell clusters were derived using Phenograph and annotation was performed using a combination of expression heat map analyses and expression density plots (Figure S7). Final annotation was refined with Pseudotime analysis of selected groups of cell clusters, examination of distributions of the cell clusters in all samples, and cross-checking with H&E and MCD images against known structural cell location and cell morphology (Figure S7). This produced a final list of 37 identifiable cell clusters (26 immune cell types and 11 structural) (Figure 2B-C). Expanded description of the annotated cells is provided in Table S4. For clarity of terminology, once the cell clusters were annotated, they were termed ‘cell types’ unless there were more than two cell types in the annotation.

Compared with healthy lungs (without dividing into different histopathology states), monocytes were the most abundant immune cells (Figure 2B). Amongst the annotated cell types, five were found to co-express defining markers of different immune or structural cells, reflecting closely located or apposed cell groups (Neutrophil and CD8 T cell, Monocyte and CD31+ cells, Monocyte-PAI-1+ cells, CD8 T cells – PAI-1+ cells and IFN-γ+ cells and RAGE+ cells). We labelled these 'adjacent' (ADJ) cell types. Immunofluorescence staining confirmed presence of two different adjacent cells for the Neut_CD8_ADJ (immature neutrophil and CD8 T cells) (Figure 2D, and Figure S7L). Mono_CD31_ADJ comprised both monocytes that were found adjacent to endothelial cells and CD31-expressing monocytes (Figure 2E).
Of note, the Neut_CD8_ADJ cells contained the most immature neutrophil cell type (CD71$^\text{hi}$ neutrophils) coupled with CD45RO$^+$ CD107$^-$ CD8 T cells (Figure 2F). The cluster also had the highest expression of Granzyme B (GZB), CD172a (Sirpa), IFN-β and IFN-γ (Figure 2F, G-I). Within the cluster, the GZB expression was found on the CD8 T cells, indicating these as cytotoxic CD8 T cells (Figure 2I). The monocyte subset in the Mono_CD31_ADJ cluster was the least differentiated (to macrophage) monocyte subset; similar to the Mono_1 cell type (Figure 2D, Figure S7L).

**Increased innate immune cell numbers found in all histopathology states, from injury to repair**

We next sought to understand how immune cell abundance changed as the overall histology progresses from injury to repair. Firstly, we observed that changes in numbers of structural and relevant immune cells supported the temporal progression of histopathology states from inflammation to damage and subsequent repair (Figure 3A). There was a progressive increase in numbers of all subsets of macrophages, fibroblasts, proliferating fibroblasts and myofibroblasts from ALV to DAD to OP, consistent with transition from tissue injury to repair. Endothelial and proliferating endothelial cells, proliferating bronchial epithelium and bronchial epithelium also increased progressively. Changes in abundance of macrophages over the three histopathology states reflected accumulation of macrophages as monocytes differentiate into macrophages with progression of disease.

Across the three histopathology states, we found high numbers of classical monocytes, immature neutrophils, and some subsets of MAIT, CD4 and CD8 T cells. The most significant progressive increase in numbers across the histopathological states (compared to healthy lungs) was observed for CD8 T cell subsets and CD8 containing ADJ cell clusters, and CD107$^+$ CD4 T cells (Figure 3B-D) (see Table S4 for expanded phenotype description of immune cell types). Neut_CD8_ADJ cluster was increased from the earliest histopathology state and remained high in all states. Apart from IFN-γ$^b$ MAIT cells, there were only small numbers of MAIT and NK cell subsets (Figure 3C). Cycling (Ki67$^+$) monocytes were not found in the lungs.

Overall, innate cell numbers in the infiltrate did not decline despite disease progression and was accompanied by increasing numbers of CD8 T cells [even though viral protein was absent in OP (repair) samples]. Some immune correlates of severity in the blood observed in other studies (cycling monocytes, NK cells, and activated MAIT cells) were not found in significant numbers in the lungs.
Distinct spatial organisation observed amongst immune cells despite extensive infiltration of cells and disordered tissue architecture

To determine if the cells showed spatial association and organisation amongst themselves, we employed spatial statistical algorithms (Figure 1A and S1) to (a) understand which immune cells were found co-located with injured structural cells, (b) explore how immune cells organise themselves amongst themselves, and (c) for the immune cells implicated in severe disease from COMBAT (monocyte, megakaryocyte, MAIT, CD4, CD8 and neutrophil subsets), if these were co-located or physically interacting with other immune or structural cells.

In total, we found 3888 non-replicate pairs of cell types (mono1:mono1 and CD15^hi iNeut:CD15^hi iNeut were examples of pairs of identical cell types, and filtered out) in the three histopathological states (ALV, DAD and OP). Using our three-step spatial analysis, 357 pairs of cell types were identified as statistically correlated in the QCM analysis (FDR < 0.05). These cell pairs were submitted for cross-PCF analysis, with one cell type in the pair defined as the ‘anchor cell’ - the cell against which statistically significant connections were quantified. By pre-analysis consensus, pairs of cell types with borderline statistical significance i.e. FDR values between 0.05 and 0.10 were also submitted to prevent loss of biologically relevant data from hard mathematical cut-off. The resulting co-located cell pairs were divided into ‘structure:immune’ pairs (structural cells were designated the ‘anchor’ cell type) (n=33) and ‘immune:immune’ pairs (one cell type in one of the duplicate pairs was designated the anchor cell type; e.g. for the CD107^+CD8:mono1 pair and mono1:CD107^+CD8 pair; CD107^+CD8 in the former pair was made the anchor cell, and the latter pair was excluded) (n=117) (Figure 4A). ‘Structure’ cells of interest were the key structural cells that were known to be inflamed or damaged in COVID-19 pneumonitis – endothelium (‘Endothelium’ and ‘proliferating endothelium’) and larger blood vessels (‘blood vessels’), alveolar epithelial cells (‘proliferating alveolar epithelium’) and bronchial epithelial cells (‘HLA DR^hi bronchial epithelium’ ‘HLA DR^b bronchial epithelium’ and ‘bronchial epithelium’). We were particularly interested in ‘proliferating alveolar epithelium’ as their markers and location in the lung sections suggest they were likely the type II alveolar epithelial cells, the purported progenitors (or stem cells) of alveolar epithelium (Figure S7J and K). The ability of these cells to differentiate to type 1 alveolar epithelium is critical to normal repair and alveolar regeneration after viral induced damage (Barkauskas et al., 2013; Choi et al., 2020; Zacharias et al., 2018).
Amongst the immune cells, the strongest co-location, depicted by $g(r=20)>2$ (i.e., >2 times more cells of type B observed at 20 μm from cells of type A (anchor cell) than expected under complete spatial randomness), was observed for pairs of immune cell types that belonged to the same immune phenotype, e.g., Mac1 and Mac2 (macrophages), and CD4 and CD8 T cells (CD3 T cells) (Figure 4B). This was expected biologically and provided a degree of validation for the mathematical analysis. We note that despite the lack of viral protein staining, CD107a+ CD8 T cells and IFN-γ+ CD4 T cells showed co-location [$g(r=20)>1$] in OP, with a trend of increase in these values in OP compared to DAD (Figure S9).

These results signify presence of specific spatial organisation for several immune and structural cells despite appearance of disorder in tissue. The strongest co-location between all cells was found between CD4 and CD8 T cell subsets, particularly active effector memory CD4 T cells (IFN-γ+ CD4 T cells) and cytotoxic CD8 T cells (CD107a+ CD8 T cells), which did not lessen with progression to repair, and despite absence of viral proteins.

**Neutrophil-CD8 T clusters are spatially co-located with proliferating alveolar epithelium in areas with diffuse alveolar damage**

For the significantly co-located pairs of cells, we next questioned which immune cells were found co-located with injured structural cells. To provide a composite view of the multiple output from our spatial analysis, we generated a ‘spatial connectivity plot’ to show all cell types that were statistically co-located with a designated ‘anchor cell type’. Each spatial connectivity plot displayed the strength of co-location [$g(r=20)$] and the average count for the immune cell types in the histopathology state (Figure 5A, B). The proportions of co-locating cell types which were in direct contact with the anchor cell type were calculated with the ACN analysis (see Methods) and shown in the accompanying histograms (Figure 5C and D).

Our main structural cell types of interest were the Ki67+ proliferating alveolar epithelial cell and endothelial cells. Designating proliferating alveolar epithelium as the anchor cell, we found CD15hi iNeut, Mono_CD31_ADJ and Neut_CD8_ADJ to be significantly co-located with proliferating alveolar epithelium in DAD (Figure 5A, B) [$g(r=20)>1$]. Of these cells, proliferating alveolar epithelium was most in contact with Mono_CD31_ADJ (average of 17.6% of proliferating alveolar epithelial cells in DAD) and Neut_CD8_ADJ (8.9% of proliferating alveolar epithelial cells in DAD) (Figure 5C). There was also a small number of IFN-γhi_RAGE_ADJ cells found co-located with proliferating alveolar
epithelium in all histopathology states, which could be resident alveolar macrophages found along alveolar epithelium.

For endothelial cells (which encompassed the smaller capillaries and the larger blood vessels in the lungs), the co-locating cell types with highest $g(r=20)$ in DAD were Mono_CD31_ADJ (2.1) and Mono_PA1-1_ADJ (1.6) clusters (Figure 5B,F). ACN analysis showed more of the endothelial cells were physically in contact with the Mono_PA1-1_ADJ cluster (21.2%) than Mono_CD31_ADJ (16.5%) in DAD (Figure 5D). Mono_CD31_ADJ cells showed significant spatial association with endothelial cells across all histopathology states.

Next, we designed a ‘radial connectivity map’ to provide an overview of all immune cells that were significantly co-located with all structural cells and their corresponding histopathology states (Figure 5G). Using this map, and focusing on proliferating alveolar epithelium and endothelial cells, we observed that while the monocytes (and their subsets and ADJ clusters) were mainly found co-located with both alveolar epithelium and endothelial cells, immature neutrophils were found predominantly with proliferating alveolar epithelium. We also observed that besides proliferating alveolar epithelium, the Neut_CD8_ADJ cluster was not found with any other structural cell types.

Finally, we developed a topographical correlation map (TCM) (Methods, Figure S14) to visualise how the spatial correlation between Neut_CD8_ADJ and proliferating alveolar epithelium changed across an ROI (Figure 5H). We observed marked heterogeneity in the strength of correlation for this pair of cell types across the tissue. As the lung sections were not the same sections that were used for viral staining, we were unable to accurately match viral staining with TCM sections. However, there is a suggestion of similarity between the heterogeneity found in the TCM and with the viral N protein staining (Figure 5H).

One other cell type of interest was the megakaryocyte. These CD34+ platelet precursors, a product of emergency myelopoiesis, were the most abundant immune correlate in the blood in the COMBAT study (COMBAT, 2022). Examining their spatial connections with our two structural cells of interest, we observed that megakaryocytes were associated with endothelium in DAD (Figure 5G).

Drawing these data together, our spatial analysis identified Neut_CD8_ADJ and Mono_CD31_ADJ clusters as key spatial correlates with proliferating alveolar epithelium in DAD. A visual exemplar of
this co-location of Neut\_CD8\_ADJ and alveolar epithelium is shown in Figure 5I. Mono\_CD31\_ADJ and Mono\_PAI-1\_ADJ were the strongest spatial correlates with endothelial cells, the former was the case across all states. No immature neutrophils (alone or in an ADJ cluster with CD8 T cells) were found with endothelial cells in any histopathology states. It is noteworthy that there was no significant co-location between any immune cells and the larger blood vessels; nor between CD107\(^+\) CD8 T cells and IFN-\(\gamma\) CD4 T cells with proliferating alveolar epithelium or endothelial cells despite relatively high abundance in the tissue. In addition, despite a correlation with disease severity in the blood, NK and MAIT cells did not co-locate with any structural cells. Further, even though macrophage subsets were the most abundant cells in lungs, there was also no statistically significant co-location between these cells and damaged structural cells.

All data, the spatial connectivity plot, radial connectivity map, and topographical correlation map functions are available as open resources on MDV (https://mdv.medbiol.ox.ac.uk/, Figure S10, Methods and Video1).

**Immature neutrophils have a predilection for spatial co-location with CD8 T cells**

We next examined how immune cells connected to other immune cells by interrogating the 91 pairs of immune cells with g(r=20)>1 across the three histopathology states (Figure 6A-C).

We observed that as single entities (as opposed to those found within ADJ clusters), immature neutrophils only co-located with CD8 T cells or CD8-ADJ clusters (Figure 6A), regardless of histopathology state. However, immature neutrophils within the Neut\_CD8\_ADJ cluster, co-localised with Mono\_CD31\_ADJ clusters in DAD and other monocyte subsets in OP (Figure 6A and D).

Therefore, in DAD, proliferating alveolar epithelium not only co-located with Neut\_CD8\_ADJ, but also with a further network of co-locating immune cell types linked to the Neut-CD8\_ADJ cluster, forming a super network of Neut\_CD8\_ADJ and Mono\_CD31\_ADJ clusters around the proliferating alveolar epithelial cells. This can be seen in the ACN analysis (Figure 6F) and an MCD image view of the cells in the tissue (Figure 6G).

In contrast to neutrophils, there was a less restricted repertoire of co-locating cell partners for monocytes. Monocyte subsets and ADJ clusters were found co-located with NK, MAIT, CD4 and CD8
T cell subsets (Figure 6B-C). Notably, megakaryocytes were found uniquely associated with Mono_CD31_ADJ in DAD (Figure 6E).

Our analyses showed that there were distinct organisations amongst immune cells in COVID-19 lungs, with specific predilection of immature neutrophil for CD8 T cells, and upon connection (as the neutrophil_CD8_ADJ cluster), a further connection with Mono_CD31_ADJ cluster was formed, resulting in a network of Neut_CD8_ADJ and Mono_CD31_ADJ, linked to proliferating alveolar epithelium in diffuse alveolar damage. These were then linked to megakaryocytes via the latter cell type’s connection with Mono_CD31_ADJ cluster in DAD. Thus, a spatial network of immature neutrophils, CD8 T cells, classical monocytes and megakaryocyte form a connected web of cells juxtaposed against proliferating alveolar epithelial cells and alveolar capillaries in DAD.

**Phenotype mapping of lung immune cell to blood immune cells links CD8, monocytes and immature neutrophils to disease severity-associated circulating cell subsets**

Finally, we returned to our COMBAT data to explore if we can identify the circulating source of the monocytes, CD8 and neutrophils found in lung tissue. Using scmap, a method which enables label projection by calculating the similarity between cells profiled by two separate studies (Kiselev et al., 2018), we examined the phenotypic similarity between lung monocytes and CD8 T cells in our study with blood-equivalent cells from the COMBAT study. From COMBAT, we selected the suspension single cell mass cytometry (CYTOF) of CD66 (neutrophil)-depleted whole blood for comparison to our tissue single cell mass cytometry (Hyperion) data set.

All lung monocyte subsets including Mono_CD31_ADJ, Mono_PAI-1_ADJ (but not Mono3), and all macrophage subsets showed high Jaccard similarity index with HLA DR\textsuperscript{hi} classical monocytes in the blood (Figure 3B). Both lung CD107\textsuperscript{-}CD8 and CD107\textsuperscript{+}CD8 matched to blood ‘GZB\textsuperscript{neg} CD8 T cells’ in COMBAT (Figure 7A). Lung IFN-γ\textsuperscript{+} CD4 T cells matched to COMBAT’s ‘activated CD4 T cells’ subset (which contained CD27\textsuperscript{−} and CD27\textsuperscript{+} CD4 T cells).

For the two COMBAT cell types of relevance to our findings in tissue, (GZB\textsuperscript{neg} CD8 T cells and HLA DR\textsuperscript{hi} classical monocytes) we further interrogated their marker expression in comparison to healthy and disease controls (data found in Supplementary Data S3 in COMBAT). We observed that the GZB\textsuperscript{neg} CD8 T cells’ expressed markers of exhaustion and were KLRG1\textsuperscript{+} compared to other CD8 T cells.
in the CYTOF analysis from COMBAT. For HLA DRhi classical monocytes, some notable features were high expression of HLADR and CLA. More importantly, both GZBneg CD8 T cells and HLA DRhi classical monocytes were unique amongst CD8 T cell subsets and other monocyte subsets in showing relatively lower abundance compared to healthy volunteers in COMBAT(COMBAT, 2022), raising the possibility that these were the subsets that have trafficked to the lungs. This is not unprecedented given previous findings in lungs which showed sparse antigen-specific T cells in blood of severe influenza patients but 8 times higher in matched blood-lung samples (Zhao et al., 2012).

For neutrophil comparisons between our findings in the lungs and those in blood from COMBAT, we obtained stored whole blood samples to perform more detailed examination of sub-clusters of immature neutrophils to match markers used in the lungs, using a 42-marker CYTOF panel (Table S5). 8 subsets of neutrophil were obtained with clustering and UMAP, which was classified according to maturity of these cells, from pro-neutrophil to mature neutrophils (Figure 7C-D). When compared to the lung neutrophils, we found that ‘immature neutrophil 2’ in the blood (Figure 7C-D), which expressed the highest level of CD172a amongst the immature CD107- neutrophil subsets was most closely matched to the neutrophil subset in Neut_CD8_ADJ (Figure 2F). When were re-examined the abundance of this circulating ‘Immature neutrophil 2’ in blood of COVID-19 patients, we observed that their levels correlated positively with severity of disease (Figure 7E).

These findings show that the CD107- CD8 T cells, implicated in the inflammatory network in the lungs with immature neutrophils, matched most closely to a GZBneg KLRG1+ CD8 T cell subset in the blood which also expressed a T cell exhaustion signature, and a likely source for the CD8 T cells found in the Neut_CD8_ADJ cluster. A highly increased circulating CD172ahi immature neutrophil subset is the likely source for the immature neutrophils found in Neut_CD8_ADJ cluster.
Discussion

In this paper, we deconvoluted a highly disordered immune and structural landscape to provide accurate annotations and abundance metrics for the cellular landscape and then leveraged mathematical techniques to describe co-localization and cell contact-based network construction. Our mathematical tools encompassed a range of spatial statistics and methods from network science; some transposed from ecology (Baddeley et al., 2014; Guseva et al., 2022; Morueta-Holme et al., 2016). The pipeline uncovered a hitherto undescribed physical partnership between immature neutrophils and CD8 T cells in COVID-19 lungs linked to proliferating alveolar epithelium in areas with diffuse alveolar damage. This further connected with classical monocytes and megakaryocyte around endothelial cells, forming a super pro-inflammatory network across the alveolar bed in DAD. The observations on neutrophils are especially significant since relatively little is understood of the role of neutrophils in the lungs of patients with COVID-19 due to poor detection with transcriptomic methods (Delorey et al., 2021; Slyper et al., 2020).

Our study did not elucidate how neutrophil-CD8 clustering might contribute to disease pathogenesis. However, evidence from other diseases provide some insight. Neutrophils and CD8 T cells aggregation in colorectal cancer and graft vs host disease have been shown to enhance T-cell receptor-triggered activation of CD8+ T cells (Governa et al., 2017) causing neutrophil-mediated tissue damage by the release of reactive oxygen species (Schwab et al., 2014). Neutrophils can also act as antigen presenting cell which cross present antigen to CD8 T cells, further enhancing activation (Samadi et al., 2019; Takashima and Yao, 2015). CD8 T cells with a similar effector memory and GZB profile as that found in the Neut_CD8_ADJ cluster have also been implicated in immunopathology of COVID-19 in other organs. Imaging mass cytometry studies in COVID-19 brain tissue showed intriguing spatial associations with microglia, which also sustained immune activation and neuroinflammation (Schwabenland et al., 2021).
The presence of viral antigen could be the trigger for these foci of immature neutrophils and CD8 T cells, possibly initiated by recognition of viral antigen by CD8 T cells. However, we note abundant Neut_CD8_ADJ cluster in the OP state (Figure 3B) where there were no viral proteins or RNA. One explanation is that these CD8 T cells were self-proliferating, as suggested by Liao’s study using single cell RNA sequencing of lung-lavaged cells in COVID-19 patients (Liao et al., 2020). Supporting this, Neut_CD8_ADJ cluster showed the highest Ki67 expression (Figure 2F), with MCD imaging isolating this expression to CD8 T cells (Figure 2I). Organising pneumonia is not a natural sequela of all viral infection or alveolar inflammation. Indeed, many patients who do well do not progress to consolidation on computed tomographic (CT) scans. Thus, a potential deleterious effect of these foci of inflammation could be the obliteration of regenerative potential in type II alveolar epithelial cells, the purported stem cells for the alveolar unit (Olajuyin et al., 2019), and development of organising pneumonia (OP).

Another cluster highlighted by our analyses was the Mono_CD31_ADJ cluster, which was spatially associated with Neut_CD8_ADJ cluster, and with proliferating alveolar epithelial cells. Proliferating alveolar epithelial cells are the nominal stem cells for the alveoli and key to replenishment of type 1 alveolar epithelial cells. Its health, and ability to function optimally, is a key requirement for repair of infected and damaged alveoli. A consequence could be that the production of type I IFN, [and other monocyte-specific cytokines like IL-6 and TNF-α (as reviewed by, (Bert et al., 2022; Costa et al., 2019)], combined to impact on regeneration of alveolar epithelium. It is also possible that type I IFN production from these monocytes causes upregulation of ACE2, thereby sustaining viral entry and alveolar epithelial damage(Ziegler et al., 2020). This agrees with observation from transcriptomic studies of the lungs where type II alveolar epithelium were found in an inflammation-associated intermediate state rather than progressing via normal regeneration to type I alveolar epithelium (Delorey et al., 2021; Melms et al., 2021; Wang et al., 2021a).

The tight association between a large number of monocytes and endothelial cells in all histopathology states could result in excess inflammation and also predispose to small vessel thrombosis, particularly with further presence of megakaryocytes at the point of maximal injury (DAD) (Figure 5B). Single cell transcriptomic analyses in COVID lungs have demonstrated upregulation of endothelial-damage markers, including VWF, ICAM1 and VCAM1, and transcriptional programs suggesting altered vessel wall integrity and widespread activation of coagulation pathway
associated genes in endothelial cells (Chua et al., 2020; de Rooij et al., 2022; Wang et al., 2021a). In addition, autopsy studies have shown high numbers of megakaryocytes and platelet rich thrombi in the lungs with COVID-19 pneumonitis (Rapkiewicz et al., 2020).

Beyond these key messages, other findings clarified the importance of immune cell numbers and phenotype in blood of patients with severe COVID-19. There was no significant spatial co-location between activated NK cells and MAIT cells with any structural cells although the numbers for MAIT cells were increased, in keeping with blood levels. With the ability to identify single cells of CD4 and CD8 T cells, and quantify their abundance per mm² of lungs, we also showed definitively that levels of CD4 and CD8 T cells were high in lung samples in contrast to studies which inferred their depletion from gene expression profiles (Melms et al., 2021). Immature cycling monocytes, one of the most striking observations in the blood of patients with severe compared to mild COVID-19 disease (COMBAT, 2022; Mann et al., 2020), were not found in lung tissue. This suggests that immature monocytes are unlikely to be involved in tissue damage, and unlike immature neutrophils, probably differentiated rapidly to mature monocytes and macrophages.

Our findings refined our earlier work on a smaller subset of COVID-19 lungs (n=3) using targeted transcriptomic analysis (GeoMx™) in specified sections in the lungs linked to alveolar damage (Cross et al., 2022). In that work, we deconvoluted cells detected by gene expression profile using limited protein markers and showed that CD8 T cells and macrophages with IFN-γ signature correlated with areas of lungs with alveolar damage. Interestingly, areas of severe damage exhibited consistent expression of IFNG-regulated chemokines such as CXCL9/10/11 that may promote CXCR3-mediated chemotaxis or retention of CD8+ T effector lymphocytes. Another earlier work in the same lung samples showed significant presence of neutrophil extracellular traps (NETS) in the lung samples which correlated with areas of low CD8 T cell levels. Re-examining the number of NETS per lung section, we observed widespread presence with no significant difference between the three histopathology states (Figure S3C). As NETS production is a feature of mature rather than immature neutrophils (Melero et al., 2022), one explanation is that there is a CD8-directed immature neutrophil localisation to proliferating alveolar epithelium, which is separate from the relatively less discriminate NETS expression by mature neutrophils.

The key limitation of our study is that it is an observation of association, albeit that there was clear comparison between histopathology characterisations of alveolitis, damage and repair. Thus, it is not possible to elucidate cause or effect. Our cohort was also small though this was counterbalanced by
uniquely fresh samples from lungs, with minimal effect of degradation due to the sampling methods at the point of death. Finally, our study was led by specific questions. To that end, the antibody panels, and analyses were targeted to those questions and cellular identities were constrained to that linked to the antibody panel.

We conclude that statistically rigorous analyses of spatial associations of immune and structural cells in lungs of those with fatal COVID-19 identified an inflammatory nidus of immature neutrophils and CD8 T cells with high immune activity and proliferating capabilities that were linked to alveolar progenitor cells in areas with greatest alveolar damage. It establishes the importance of emergency myelopoiesis in lung immune pathology, with potential roles for immature neutrophils and megakaryocytes in alveolar damage, aberrant alveolar regeneration, and excess thrombogenesis. The findings support the evaluation of therapeutics that target monocytes and immature neutrophils, potentially earlier in disease to limit its impact on progression to widespread alveolar damage and organising pneumonia. It also means that drugs that increase the longevity or survival of CD8 T cells require further assessment given the potential contribution of CD8 T cells to lung damage.
Figure legends

Figure 1. Spatial analysis pipeline and histopathology categorisation of samples

A. Overview of the workflow and SpO0x pipeline for high resolution phenotyping and enumeration of cells in tissue and statistical quantification of spatial co-location of immune and structural cells. The steps of the analysis are presented in Figure S1 in more detail.

B. Representative H&E section from COVID-19 tissue section showing formal histopathology features of alveolitis (ALV), diffuse alveolar damage (DAD) and organizing pneumonia with their corresponding MCD file image showing staining for 5 antibodies (α-SMA, EpCAM, PanCK, Col 1a and CD31). Within figures ‘a’-‘c’ – characteristic features of ALV, DAD and OP. ‘a’ - thickened alveolar epithelial wall and septae with immune cell infiltrate and congestion of alveolar walls ‘b’ widespread presence of hyaline membrane, and regenerating/proliferating Type II alveolar epithelium and ‘c’ - fibroblasts and collagen presence around bronchial epithelium. See also Figure S2.

C. Point when samples were obtained from the first day of symptoms and corresponding histopathology states in lung sections

D. C-reactive protein (CRP) levels closest to the point of sampling and corresponding histopathology state in lung sections

Figure 2. High definition immunophenotyping of lung cells and identification of tissue structure

A. UMAP representation of myeloid, lymphocyte and structural cell ‘mega clusters’ from all regions of interest (ROI) (k = 30) (COVID-19 and HC). See also Figure S7 for extended pathway of analysis.
B-C. Number of cells per mm² of lung tissue sections in all COVID-19 samples (n=12 patients, 2-3 ROIs' each) compared to healthy control (HC) samples. Median (IQR) shown; statistical analysis performed after division into separate histopathology states; see Figure 3D.

D. Immunofluorescence staining validation for the adjacent cell cluster, Neut_CD8_ADJ. with high magnification confocal image (right top inset). Broken circles show CD8 (white, CD8 T cell) and CD15(green, neutrophil) couplets throughout lungs. See also Figure S6 for negative controls, and as immaturity of neutrophil is in part contingent on non-expression of CD10, positive controls are provided in Figure S6.

E. Immunofluorescence staining validation for the adjacent cell cluster, Mono_CD31_ADJ showing CD14 staining (white, monocytes) with expression of CD31 (red) or adjacent to CD31-expressing endothelial cells. Broken circles show intimate proximity between CD14-expressing monocytes (white) and CD31-expressing endothelium (red).

F. Heatmaps of median scaled intensity for each marker for all cell clusters of the ‘Myeloid’ mega cluster. n_cells - average number of cells in all COVID-19 ROIs. UD — undefined cluster

G. MCD image from a DAD sample showing co-expression of CD8 (green), CD15(red) and CD8_CD15_ADJ (green and red, making yellow staining), with some CD8 without adjacent neutrophils (green).

H. Same MCD images as (G) but with IFN-β channel ‘open’ (white) showing IFN-β expression on Neut_CD8_ADJ (yellow)

I. Higher magnification of a set of 3 MCD panels - ‘none’ - Neut_CD8_ADJ (yellow) only (arrows); ‘IFN-γ’ - with ‘IFN-γ’ (white) channel opened on MCD viewer showing expression on Neut_CD8_ADJ (yellow) (arrows) and some CD8 (green); ‘GZB’ - with ‘GZB’ (cyan) channel opened, and showing expression on Neut_CD8_ADJ (yellow) (arrows). CD172a panel shows confocal immunofluorescence staining (white) on CD15 and CD8 adjacent to each other.

**Figure 3. Enumeration of immune and structural cells in COVID-19 lungs**

A-C. Cell abundance plots for immune cells (myeloid and lymphoid cells) and structural cells in lung tissue, adjusted for surface area in COVID lungs categorised into those with histopathology states of alveolitis, DAD and OP compared to healthy control. See Table S4 for extended phenotypic description for all cell types and clusters.

D. Differential abundance analysis of COVID-19 samples (alveolitis, DAD and OP) vs healthy controls (HC). Asterisks are adjusted p-values, *p<0.05 **p<0.01 ***p<0.001. Arrow refers
to immune cells that showed progressive increase in abundance with progression histopathology states from ALV to OP.

**Figure 4. Spatial analysis of immune and structural cells in COVID-19 lungs**

A. Schematic representation of the sequential spatial analysis of cellular co-location, starting with quadrat correlation matrix (QCM), then cross pair correlation function (cross-PCF) analysis, interrogation of cross-PCF output and organization according to main questions. QCM output is provided in Figure S8.

B. $g(r=20)$ heatmaps showing statistically significant correlated pairs of cells derived from QCM and cross-PCF analysis. Red boxes indicate groups of cell subsets from the same immune phenotype – neutrophils (Group 1), monocytes and macrophages (Group 2), CD3 T cells (Group 3) and MAIT cells (Group 4).

**Figure 5. Spatial organization of immune cells around structural cells in COVID-19 lungs**

A-B. Spatial connectivity plots for proliferating alveolar epithelium, showing immune cells that are significantly co-located to proliferating alveolar epithelium (designated ‘anchor cell’) in the three histopathology states. The size of the nodes (filled-in circle) represents mean cell counts (abundance) for the specified cell cluster for all the ROIs in the histopathology state (scale shown in grey), and colour of nodes relate to histopathology state. Connecting lines indicate a statistically significant co-location between the two cell types derived from QCM and cross-PCF analyses. The thickness of the lines relates to the $g(r=20)$ value relative to each pair in the plot – the thicker the line, the higher the $g(r=20)$ and therefore greater strength of co-location between the immune cell type and anchor cell.

C-D. Histogram shows % of anchor cells – proliferating alveolar epithelial (PAE) cells (C) and endothelial cells (D) in contact with specified immune cell type.

E-F. cross-PCF profiles for the two most abundant co-located structure:immune cell pairs in DAD

G. Radial connectivity map depicting all statistically significant pairs of structure:immune cells in all histopathology states; anchor cells (structural cells) are in smaller, inner circle. ‘DRhi BE’ – HLA DR$^h$ bronchial epithelium; ‘DRlo BE’ – HLA DR$^h$ bronchial epithelium; “Endo’- endothelial cells; ‘PAE’ – ‘proliferating alveolar epithelium’, ‘PBE’ – ‘proliferating bronchial epithelium’; ‘PE’ – ‘proliferating endothelium’ ‘BV’ – blood vessels’. Numerical values indicate $g(r=20)$ for that pair in that state (coloured bar), and % indicates proportion of anchor cells that are co-located with the specified immune cells.
H. Topographical correlation map showing distribution of the co-located Neut_CD8_ADJ cluster and proliferating alveolar epithelial cells (left panel) accompanied by same area of lung section showing immunohistochemistry staining for SARS-CoV-2 N protein (right panel).

I. MCD images showing Neut_CD8_ADJ clusters amidst single CD8⁺ T cells, CD15⁺ immature neutrophils and epithelial markers (EpCAM and PanOK).Couplets of CD8⁺ and CD15⁺ cells - Neut_CD8_ADJ clusters (red and green merging to form yellow cells) (arrows) are most clearly visible in DAD amongst the epithelial markers.

**Figure 6. Spatial organization amongst immune cells in COVID-19 lungs**

A-C. Radial connectivity map depicting all statistically significant pairs of immune to immature neutrophil subsets (including ADJ subsets) (A) and to monocyte subsets (B and C, separated for clarity). Cells in all histopathology states. Anchor cells (immature neutrophil and monocyte subsets) are in smaller, inner circle. Numerical values indicate g(r=20) for that pair in that state (coloured bar), and % indicates proportion of anchor cells that are co-located with the specified immune cells.

D-E. Spatial connectivity plots for Neut_CD8_ADJ (D) and Mono_CD31_ADJ (E), showing immune cells that are statistically significant co-located to proliferating alveolar epithelium (designated 'anchor cell') in the three histopathology states. Size of nodes (filled-in circle) represent mean cell counts for the specified cell cluster for all the ROIs in the histopathology state, and colour of nodes relate to histopathology state. Connecting lines indicate a statistically significant co-location between the two cell types derived from QCM and cross-PCF analyses. Thickness of line relate to value of g(r=20) relative to each pair in the plot – the thicker the line, the higher the g(r=20) and strength of co-location between the immune cell type and anchor cell.

F. Adjacency cell network (ACN) map showing contact between the Mono_CD31_ADJ cluster, Neut_CD8_ADJ cluster and proliferating alveolar epithelial. Cell segmentation masks generated by DeepCell were used to produce this spatially-embedded network in which nodes represent cell centres and are labelled according to their cell type. Nodes are connected by a line if the corresponding cells in the segmentation mask share a border.

G. MCD image showing CD8 (green), CD15(red) and Neutrophil-CD8 T cell couplets (yellow) amidst endothelial cells (green) and monocytes (CD14, yellow) in a lung section with DAD on histopathology analysis.

**Figure 7**
A. SCMAP matching heatmaps representing the Jaccard indices of similarity between COMBAT (blood) and COSMIC (lung) lymphocyte clusters. CD107$^+$ CD8 T cell and CD107$^-$ CD8 T cell in COSMIC matched to blood GZB$^-$ CD8 T cells in COMBAT. IFN-$\gamma$ CD4 T cells matched to COMBAT's 'activated CD4 T cells'

B. SCMAP matching heatmaps representing the Jaccard indices of similarity between COMBAT (blood) and COSMIC (lung) myeloid clusters. Mono_CD31_ADJ and Mono_PA1-1_ADJ and all macrophage subsets matched with HLA DR$^+$ classical monocytes in the blood from COMBAT data

C. UMAP representation of neutrophils from controls and COVID-19 infected patients (down sampled to 100 000 cells per condition) obtained from COMBAT consortium, showing 8 subsets of neutrophils

D. Heatmap showing median marker expression for 8 neutrophil subsets demonstrating high similarity of marker expression in immature neutrophil 2 (iNeut2); comparable to the neutrophil subset in Neut_CD8_ADJ observed in the lungs (Figure 2F).

E. Percentage abundance of the 8 neutrophil subsets from healthy volunteers (HV), mild, severe and critical COVID-19 patients from the COMBAT consortium show a progressive increase in immature 2 neutrophils with increasing COVID-19 disease severity

**Supplementary Figure legends**

**Figure S1. Sequential steps in overall segmentation and spatial analysis pipeline with description of each step**

**Figure S2. Representative H&E staining showing three histopathology states – Alveolitis, DAD and OP**

A-C. 2x2mm regions of interest (ROI) sections from same lung samples used for IMC, representative sections from three patients each from those showing alveolitis, diffuse alveolar damage (DAD) and organising pneumonia (OP) on their lung sections

**Figure S3. Correlation between clinical data and histopathology states**

A. Graph showing total days in ICU up to the point of sampling and corresponding histopathology state for lung sections (n=12 patients)
B. Graph showing number of days of hospitalization up to the point of sampling and corresponding histopathology state for lung sections

C. Graph showing the presence of NETs in slides (as a percentage of surface area; data from (Melero et al., 2022) with corresponding histopathology state

Figure S4. SARS-CoV-2 Nucleocapsid protein staining in lungs

Representative SARS-CoV-2 Nucleocapsid protein staining by immunohistochemistry in regions of interest from lung sections with the histopathology states of alveolitis, DAD and OP, and healthy control (HC).

Figure S5. MCD images from the first ‘sentinel’ cohort of n=3 patients (one each with lung sections showing ALV, DAD and OP), and one healthy control

A. MCD images showing expression of structural cell markers (Col1a, EpCAM, PanCK, α-SMA) and the viral integration receptors ACE2 and TMPRSS2. There was widespread ACE2 and TMPRSS2 expression in all histopathology states.

B. MCD images showing expression of key immune cell populations in same sections as (A); B cell (CD20); hematopoietic cells (CD45), CD8 T cells (CD8), regulatory T cells/Tregs (FoxP3), macrophages (CD68). There were very low numbers of B cells and Tregs in lung sections from all three patients. Scale bar 200µm.

Figure S6. Selected controls for IMCs and immunofluorescence

A. MCD images show expression of CD10 and CD71 on CD15-expressing neutrophils in lung and bone marrow from a healthy control and a lung sample from a patient with DAD histopathology state. Scale bar 200µm

B-C Isotype control for immunofluorescence staining for CD8/CD15 staining (Figure 2B) and CD14/CD31/CD15 staining (Figure 2C) performed on sections from same sections as Figure 2B and 2C.

Figure S7. Cluster annotation workflow

A-E Overall view of work flow for annotation of clusters obtained from single cell segmentation. 37-plex antibody panel and their respective metal tags (A) is shown with overall UMAP of clusters from 677,623 single cells (B) from all patients (n=12) and healthy controls (n=4). Cells that did not show any antibody staining were first filtered from further analysis. The remaining
cells were grouped into three mega-clusters termed Structural, Myeloid and Lymphoid based on presence and/or absence of CD45, EPCAM, PanCK, CD31, α-SMA, CD56, Vα7.2, CD3, CD14, CD68, PF4 (for megakaryocyte) and CD15 expression. The three mega-clusters were then re-clustered based on immunological and structural markers into Structural (C), myeloid (D) and lymphoid (E) mega clusters. The clusters generated were annotated according to heatmaps showing (i) median marker expression, (ii) expression density histograms which allowed better delineation of the range of marker expression, specifically differentiating low and negative expression levels and (iii) cluster distribution plots which showed the frequency of each cluster in different samples.

F-J. To further define cluster identities, the spatial location of clusters was visualised using cell centroid plots and mapped onto an adjacent H&E slide (F-J). A series of selected markers from MCD images of a samples with DAD (F), its corresponding H&E section (G) and the cell centroid map showing the corresponding final identity of the cells (H). (I-J) show examples of final identity of annotated cells – endothelial cells, blood vessels and proliferating endothelial cells in healthy lungs and DAD (I) and bronchial epithelium and proliferating alveolar epithelium (J).

K-M. High resolution and larger figures from (C-E).

N-P. Final annotations were also aided by Pseudotime analysis of selected immune cells where differentiation trajectories are expected e.g. from monocytes to macrophages thus early monocytes (less differentiated monocytes to more differentiated monocytes or ‘monocyte-macrophage’ (N); and from less mature to more mature macrophages (O) and from very immature neutrophils to less immature neutrophils (P). No directionality is inferred in these Pseudotime analyses, rather connectivity from one to another.

**Figure S8 Output from quadrat correlation matrix analysis**

A-C. Heat map showing correlation coefficient derived using partial correlation (PC) methods and standardised effect size of correlation for pairs in sections of lungs showing ALV, DAD and OP that are statistically significantly correlated (FDR q<0.05) after adjusting for multiple comparisons. For all histopathology states, pairs with borderline FDR q values (i.e. 0.05 to 0.10) were also examined as agreed by the consortium pre-analysis.

**Figure S9. Cross- PCF output [g(r=20)] for selected pairs of cells**
Graphs show \( g(r=20) \) values for ROIs from all patients placed into histopathology states (ALV, DAD and OP) \((n=12\) patients\) for cell types in CD3 T cell subsets.

**Figure S10. MCD software**

Screenshots demonstrating the functionality of Multi-Dimensional Viewer (MDV), which is introduced as a public resource and platform in visualization and analysis of imaging mass cytometry data. Following data upload to MDV it provides a drop-down menu of several interactive views encompassing quality control metrics, clustering, visualization of cell centroid maps and relevant spatial statistics. A video walkthrough of the functionality exemplar of MDV is provided (https://drive.google.com/file/d/1E0n5iwuexNGTFCFvw8UgURS_ri63AZEYa/view).

A. Sample information view with principal component analyses of included samples, cell numbers per sample and disease state, information on regions of interest included in the analysis including surface are ablated, false colour images and representative H and E images

B. Quality control view showing median marker staining intensity per sample. Data is visualized on histograms, heatmaps showing the median marker expression per sample as well as UMAP displays coloured by sample

C. Global clustering view showing high level cluster information visualized with UMAP, heatmaps showing median expression density, and annotation into structural, myeloid, lymphocyte and megakaryocyte sub clusters.

D, E, F. Representative example of visualizations for sub-clustering of structural cells. D. annotated cluster distribution plots showing abundance of each cluster per sample, E. UMAP of structural cell clusters, with accompanying annotations and heatmap showing median marker expression. F. Scaled marker expression density plots which show the range of marker expression per defined cluster. All these views are interactive, enabling the user to closely examine the features of a cell cluster/clusters of interest.

G. Cell centroid representations of annotated cell clusters with regions of interest in the histopathological state diffuse alveolar damage (DAD) as an exemplar. This view allows the user to examine the spatial distribution of selected clusters for further robustness and accuracy of annotation as well as obtain a qualitative view of cell type co-location

H. Spatial analysis views showing the \( g(r=20) \) heatmaps for 2 histopathological states (DAD and OP) for identification of co-locating cell pairs. Interactive representation of this co-locating can be obtained by mapping of the cell pair spatial location on to representative H&E images on the left panel.
I. Exemplar spatial analysis views showing spatial connectivity radial map for structural cells and spatial connectivity diagrams for proliferating alveolar epithelial cells and endothelial cells in the different histopathological states.

**Figure S11. Optimisation staining for 37-plex panel in tonsil and lung.**
Representative raw IMC staining shown for immune and structural markers in healthy control tonsil and lung. ROI size 2mmx2mm. MK – megakaryocyte, HC – healthy control.

**Figure S12. Single cell segmentation workflow with DeepCell**
Images that are representative of nuclear and cytoplasmic sections were selected and Z projected to create nuclear and cytoplasmic single images. These were contrast adjusted and processed using DeepCell to generate a label matrix file of segmentation masks for each ROI (cropped version shown for illustrative purposes).

**Figure S13. Pair correlation function**
Derivation of the pair-correlation function showing demonstrative cross-PCF profiles for co-location and anti-co-location.

**Figure S14. Process of creating a topographical correlation map (TCM).**
A. Locations of points with labels A (blue) and B (orange) are identified within the region of interest.
B. The number of cells of type B within a 100μm radius of each cell of type A is calculated. This number is compared with the number of cells of type B that would be found within this radius under complete spatial randomness (CSR) to calculate the mark $m_{ab}$.
C. Cells of type A coloured according to the mark $m_{ab}$.
D. Sketch showing how the normalised mark $M_{ab}$ depends on the mark $m_{ab}$. The magnitude of $M_{ab}$ describes the strength of interaction while the sign indicates whether the interaction is correlation ($M_{ab} > 0$) or anti-correlation ($M_{ab} < 0$).
E. Cells of type A coloured according to the normalised mark $M_{ab}$.
F. A kernel (Gaussian) of height $M_{ab}$ is centred on each cell of type A. We sum the kernels to construct the topographical correlation map $\Gamma_{ab}$. High values of $\Gamma_{ab}$ indicate close spatial proximity of cells of type B to cells of type A; large negative values indicate that there are few cells of type B in proximity to cells of type A.
S15. Composition of final merged clusters from CYTOF data in COMBAT for scmap matching to lung clusters of interest

A. COMBAT cell clusters or ‘centroid’ (centroid is defined in scmap as a vector of the median values of expression for each gene) from CYTOF analysis of blood in all COVID-19 patients (n=64) showing myeloid clusters - before merging on left and after merging on right.

B. COMBAT cell clusters (‘centroid’) from CYTOF analysis of blood in all COVID-19 patients (n=64) showing lymphoid clusters - before merging on left and after merging on right.

In both A and B, coloured boxes on left correspond to the ones on right panel. Scored out clusters on left panel were removed as described in Methods. Those without boxes were not merged. Our eventual cell cluster of interest, GZB- activated CD8 T cells is marked with an asterisk, and it was not merged.

Methods

Key resources

Table of antibodies and reagents used in imaging mass cytometry and immunofluorescence

Table S6

Patients, samples, and ethical approvals

Lung samples were obtained from collaborators from the University of Navarra, Spain and comprised those patients who died in hospital after admission with COVID-19. The only inclusion criteria were (i) hospitalisation, (ii) evidence for COVID-19 pneumonitis, defined as presence of ground glass changes +/- consolidation and peri bronchial shadowing in mid to peripheral distribution on thoracic CT scan before death, (iii) PCR+ results for nucleocapsid (N) and/or envelope protein (E) in lung or liver tissue sample and (iv) negative bacterial culture from blood and lung within 3 days of death.

The study was approved by the Ethics Committee of the University of Navarra, Spain (Approval 2020.192). Tissue collections were obtained with consent from a first-degree relative, following a protocol approved by the ethics committee of the University of Navarra (Protocol 2020.192p); and
stored under Spain's Human Tissue Authority regulations. Samples were collected during the first wave of pandemic (2020) via an intercostal space incision, using core biopsy methods (BioPince™ Full Core Biopsy Instrument kit) immediately after death (Melero et al., 2022; Recalde-Zamacona et al., 2020). Tissues were immediately fixed in neutral buffered formalin for over 24 h, and then paraffin-embedded. These samples were also shared with other collaborators and studies carried out independently (Cross et al., 2022; Melero et al., 2022).

Healthy lung controls were obtained from the Oxford Centre for Histopathology Research and the Oxford Radcliffe Biobank based at the Oxford University NHS Hospitals Foundation Trust. Ethics approval was received from Oxford A South-Central NHS REC (ref 19/SC/0173).

**RNA extraction and quantitative RT-PCR for viral genes**

RNA extraction from biopsies was performed using the QIAamp Viral RNA Mini Kit (Qiagen) and the identification of SARS-CoV-2 transcripts encoding nucleocapsid (N) and an envelope protein (E) was performed using a commercial kit (SARS-CoV-2 Real Time PCR Kit, Vircell), both according to manufacturer recommendations, at the Microbiology Laboratory of the Clinica Universidad de Navarra (ref). Samples with amplification of both targets with Ct values below 35 were considered positive for SARS-COV-2. Ct threshold was selected based on comparison between Ct values and presence of viral DNA on nasopharyngeal-swab standards.

**SARS-CoV-2 Nucleocapsid protein staining**

Slides were deparaffinised and heat-induced epitope retrieval were performed on the Leica BOND-RXm using BOND Epitope Retrieval Solution 2 (ER2, pH 9.0) for 30 minutes at 95°C. Staining was conducted with the Bond Polymer Refine Detection kit, a rabbit anti-SARS-CoV-2 nucleocapsid antibody (Sinobiological; clone: #001; dilution: 1:5000) and counterstained with haematoxylin.

**ROI selection**

H&E stained sections were examined by two senior pathologists independently and a pulmonologist and data compiled with consensus at the third iteration. ROIs were selected based on size (2x 2mm squares or equivalent surface areas) to represent the dominant histopathology findings for the section. Slides were imaged on AxioScan Z1 slide scanner (Zeiss) and viewed using QuPath (Bankhead et al., 2017).
Imaging mass cytometry (IMC) staining

Sequential 6µm thick FFPE lung tissue section slides were incubated for 2hrs at 60°C on a slide warmer, dewaxed twice in Histo-clear II (National Diagnostics) for 10 minutes before rehydration through serial alcohols; 100%, 100%, 95%, 70% ethanol and MilliQ water. Slides were then incubated for 30 minutes at 96°C in EDTA Target Retrieval Solution, pH 9 (Agilent) and cooled to 70°C before washing twice in MilliQ water. Slides were blocked in 3% BSA solution in Maxpar PBS (Standard BioTools; previously Fluidigm) for 45min. Sections were then stained with metal-conjugated antibodies in Maxpar PBS containing 0.5% BSA overnight. Antibodies conjugated in house were conjugated with MaxPar X8 antibody labelling kits (Standard BioTools) or Lightning-Link kits (Abcam) according to manufacturer’s instructions. Slides were washed in 0.2% Triton X-100 then twice in Maxpar PBS. Intercalator-Ir (Standard BioTools) diluted in Maxpar PBS was used to stain DNA (30min), slides were washed in MilliQ water then air dried.

Ablation of the relevant regions of interest (ROIs) was carried out on Standard BioTools Hyperion Imaging System using CyTOF7 Software v7.0 (Standard BioTools) and visualized using MCD Viewer (Standard BioTools). Images were processed for publication using FIJI (Schindelin et al., 2012) to de-speckle and sharpen the images.

Antibody validation and optimization

Antibody clones were selected which had previously been published and validated in IMC studies as well as antibodies frequently utilized for immunofluorescence or immunohistochemistry studies with FFPE tissues. Staining validation for IMC markers was performed in healthy control lung and tonsil as well as in some COVID-19 infected lung (Figure S11). During optimisation, we checked that (i) mutually exclusive expression pattern were found in key immune and structural lineage markers i.e. CD68, Epcam, CD3 and CD19 (ii) markers showed appropriate subcellular location expression i.e. transcription factors Foxp3 and Ki67 were nuclear, whereas CD68 expression was cytoplasmic and cell membrane. (iii) structural cell identities defined by IMC lineage marker expression are compatible with cell morphology and location in H&E. Adjacent H&E-stained slides and structural markers expression was examined e.g. α-SMA expression around vessels and bronchi, EpCAM expression on bronchial and alveolar epithelial cells. (iv) Non-biological sense expression e.g. CD4 and CD8 co-expression and biologically expected and coherent co-expression patterns eg cells
expressing CD45, CD3, CD8 and CD45RO were examined (v.) Expression for the following key markers was validated by immunofluorescence staining in adjacent slides - CD4, CD8, CD14, CD15, CD31, CD172a, ProSPC, PAI-1, Epcam and Ki67.

Antibody clones that did not perform well i.e. those with weak signal, high background, or nonspecific staining were discarded. Antibody titration was performed to maximise signal to noise ratio in both lung and tonsil tissues and panels were designed to minimise the already low levels of signal spill over seen in IMC [less than 1-5%] (Chevrier et al., 2018).

**Immunofluorescence**

Paraffin-embedded human lung tissue sections were deparaffinized and each section was pre-treated using heat-mediated antigen epitope retrieval with sodium citrate buffer (pH 6) for 20 minutes. Then sections were blocked in 10% normal goat serum (Thermo Fisher Scientific, 50062Z) for 20 minutes and then incubated with CD14 antibody 1:100 dilution (Abcam, AB183322), CD15 antibody 1:200 dilution (Cell signalling Technology, 4744S), CD31 Antibody 1:100 dilution (LS Bio, LS-B15507-LSP), CD8 Antibody 1:100 dilution (Cell signalling Technology, 90257SF), CD172a, Anti-SIRP-Alpha Antibody 1:100 dilution (Abcam AB19149), Pro-Surfactant Protein C Antibody 1:100 dilution (Abcam AB90716), overnight at 4°C. Each section is washed three times in TBS-T (0.1% Tween) and stained with Alexa Fluor 568 or 647 conjugated Goat anti Rabbit IgG or Alexa Fluor 488 or 568 conjugated goat anti-mouse IgM secondary antibody or Alexa Fluor 488, 568 or 647 conjugated goat anti-mouse IgG1 for 30 minutes and washed three times in TBS-T (0.1% Tween) and mounted with Prolong platinum antifade Mountant with DAPI (Fischer Scientific) and the section slides were imaged using a Nikon Ti2 microscope (Nikon Instruments, Japan) attached to an Andor Dragonfly 200 spinning disk confocal microscope (Oxford Instruments, Belfast).

**Imaging of Fluorescent labeled Tissue Sections**

Slides were imaged using a Nikon Ti2-E microscope (Nikon Instruments, Japan) attached to an Andor Dragonfly 200 spinning disk confocal unit (Oxford Instruments, Belfast). Using Andor Fusion software, the microscope was configured for DAPI (Excitation 405 nm: Emission 450/50 nm), GFP (Excitation 488 nm: Emission 525/50 nm), Red (Excitation 561 nm: Emission 600/50 nm) and Far Red (Excitation 647 nm: Emission 700/75 nm). A 10x 0.45 NA objective was initially selected to provide an overview of the entire area of the tissue section. Relevant areas (or the whole section) were then
selected using the software for higher resolution scanning, utilizing either a Nikon Plan Fluor 40x 1.3 NA oil objective with 1 um z-slice sectioning or a Nikon Plan Apo Lambda 100x 1.45 NA oil objective with 0.13 um z-slice sectioning, this ensured that the whole thickness of the tissue would be imaged. Images were saved on a computer for further processing using custom Fiji/Image J macros (Schindelin et al., 2012).

Data analysis
Software and Algorithms

<table>
<thead>
<tr>
<th>Name of software</th>
<th>Source</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>imctools</td>
<td><a href="https://github.com/BodenmillerGroup/imctools">https://github.com/BodenmillerGroup/imctools</a></td>
<td>RRID:SCR_017132</td>
</tr>
<tr>
<td>Deepcell</td>
<td><a href="https://vanvalen.github.io/about/">https://vanvalen.github.io/about/</a></td>
<td>RRID:SCR_022197</td>
</tr>
<tr>
<td>Phenograph</td>
<td><a href="https://github.com/JinmiaoChenLab/Phenograph">https://github.com/JinmiaoChenLab/Phenograph</a></td>
<td>RRID:SCR_016919</td>
</tr>
<tr>
<td>Harmony</td>
<td><a href="https://github.com/slowkow/harmony">https://github.com/slowkow/harmony</a></td>
<td>RRID:SCR_022206</td>
</tr>
<tr>
<td>Slingshot</td>
<td><a href="https://github.com/lstreet13/slingshot">https://github.com/lstreet13/slingshot</a></td>
<td>RRID:SCR_017012</td>
</tr>
<tr>
<td>Ruffus</td>
<td><a href="http://www.ruffus.org.uk/">http://www.ruffus.org.uk/</a></td>
<td>RRID:SCR_022196</td>
</tr>
<tr>
<td>QuPath</td>
<td><a href="https://qupath.github.io/">https://qupath.github.io/</a></td>
<td><a href="https://doi.org/10.1038/s41598-017-17204-5">https://doi.org/10.1038/s41598-017-17204-5</a></td>
</tr>
<tr>
<td>MCD</td>
<td><a href="https://www.standardbio.com/pro">https://www.standardbio.com/pro</a></td>
<td>RRID:SCR_023007</td>
</tr>
<tr>
<td>Name of software</td>
<td>Source</td>
<td>Identifier</td>
</tr>
<tr>
<td>------------------</td>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>imctools</td>
<td><a href="https://github.com/BodenmillerGroup/imctools">https://github.com/BodenmillerGroup/imctools</a></td>
<td>RRID:SCR_017132</td>
</tr>
<tr>
<td>DeepSea</td>
<td><a href="https://vanvalen.github.io/about/">https://vanvalen.github.io/about/</a></td>
<td>RRID:SCR_022197</td>
</tr>
<tr>
<td>Phenograph</td>
<td><a href="https://github.com/JinmiaoChenLab/Rphenograph">https://github.com/JinmiaoChenLab/Rphenograph</a></td>
<td>RRID:SCR_016919</td>
</tr>
<tr>
<td>Harmony</td>
<td><a href="https://github.com/slowkow/harmonypy">https://github.com/slowkow/harmonypy</a></td>
<td>RRID:SCR_022206</td>
</tr>
<tr>
<td>Slingshot</td>
<td><a href="https://github.com/kstreet13/slingshot">https://github.com/kstreet13/slingshot</a></td>
<td>RRID:SCR_017012</td>
</tr>
<tr>
<td>Harmony</td>
<td><a href="https://github.com/immunogenomics/harmony">https://github.com/immunogenomics/harmony</a></td>
<td>RRID:SCR_022206</td>
</tr>
</tbody>
</table>

### The Spatial Omics Oxford (SpOOx) Analysis Pipeline

The SpOOx pipeline is a computational framework that brings together the methods we have used to derive final spatial interpretation for the COVID-19 lung sections. It incorporates a suite of Python and R based command line tools which may be run individually or as a semi-automated pipeline. We have implemented SpOOx using the Ruffus framework (Goodstadt, 2010). Ruffus allows encapsulation of the workflow and parameters to enable reproducibility, transparency and code reuse. All steps discussed in the Methods are encapsulated in the SpOOx pipeline and example commands to achieve the step are shown below. An overview of the pipeline can be found in Figure 1a and S1. Full detailed documentation and a tutorial are included on the SpOOx GitHub page ([https://github.com/Taylor-CCB-Group/SpOOx](https://github.com/Taylor-CCB-Group/SpOOx)). SpOOx produces a series of output directories and files that may be uploaded to the Multi-Dimensional Viewer (MDV) software (see below). MDV has been developed based on the Multi Locus View (Sergeant et al., 2021) framework and has been heavily modified and extended to allow visualisation and analysis of large multidimensional data sets, images and the resulting spatial statistics. The code to upload data to MDV is available on GitHub at [https://github.com/Taylor-CCB-Group/MDV](https://github.com/Taylor-CCB-Group/MDV). Both the SpOOx and MDV are open source.
under the GPL 3.0 license with these links - SpOOx is available for install at https://github.com/Taylor-CCB-Group/SpOOx and MDV at https://github.com/Taylor-CCB-Group/MDV. The project data analysis is available online within MDV at https://mdv.molbiol.ox.ac.uk/projects/hyperion/6567.

**Conversion of MCD files to TIFF.** MCD files were checked for problems with ablation or staining using the MCD viewer (provided by Standard BioTools). Once these initial checks were completed, the images were converted to OME-TIFF format for segmentation.

*Commands:* `python hyperion_pipeline.py make mcd_to_tiff` and `python hyperion_pipeline.py make tiff_to_histocat`

**Segmentation and cell mask generation** Cell segmentation was performed with the Mesmer library in DeepCell (Van Valen et al., 2016), Nuclear markers (DNA1 and DNA3) and cytoplasmic markers (a-SMA, CCR2, CCR6, CD107a, CD10, CD114, CD115, CD14, CD15, CD16, CD172a, CD31, CD3, CD4, CD45RO, CD4, CD71, CD8a, Collagen1, DAP12, EpCAM, GranzymeB, HLADR, IFN-γR, IFN-γR, PAI1, PanCK, PF4 and RAGE) were extracted to TIFF files and Z projected to single channel nuclear and cytoplasmic single TIFF images (Figure S12). These images were contrast adjusted (—contrast 5) and passed to the Mesmer library (pixel size adjusted to 1 micron) as nuclear and cytoplasm channels. From these, cell segmentation masks were generated for each ROI.

*Command:* `python hyperion_pipeline.py make deepcell`

**Extraction of signal intensities for each cell.** The intensity of each marker within each labelled cell was extracted from the data using the segmentation masks using the mean arcsinh-transformed (with —cofactor 5) pixel intensity for each. The data were recorded as a table, each row representing a cell with a unique id for the ROI. Shape features such as area, perimeter, eccentricity, and centroid were also extracted from the masks. All cells were then filtered using a cell area greater than 50 \( \mu \text{m} \) and less than 300 \( \mu \text{m} \) to exclude poorly segmented cells and cell debris. Further QC was performed within MDV by plotting the distribution of marker intensity across each ROI.

*Command:* `python hyperion_pipeline.py make signal_extraction`
**Dimensionality reduction and cluster analysis.** For all downstream analysis the intensity values were arcsinh transformed with a cofactor 5. Clustering was performed using the Phenograph algorithm (Levine et al., 2015) through the implementation of the Rphenograph R package (version 0.99.1) with parameter k=30. Using MDV, the clusters were first visualised using interactive UMAP scatter plots and heatmaps (showing the median marker intensities per cluster) then manually annotated to define the cell phenotypes at the cell level. The clustering was performed at two levels: a sample level (on the trimmed [q=0.001] and scaled values) and per condition after having integrated the data with Harmony (version 1.0) (Korsunsky et al., 2019), using the default parameters with the option do_pca = TRUE. The integration of the data was performed per condition to remove variation from different patients and to better define common populations of cells. The annotations before and after integration were compared to ensure that no biologically meaningful populations were missed when integrating the data. The heatmaps, PCA and UMAP plots were done using the functions from the CATALYST R package (version 1.16.0).

Command: python hyperion_pipeline.py make phenoharmonycluster.

**Annotation workflow.** Cells were first examined for antibody staining and those cells that did not show any antibody staining were filtered from further analysis. The remaining cells were grouped into three mega-clusters termed Structural, Myeloid or Lymphocyte based on presence and/or absence of CD45, EPCAM, PanCK, CD31, α−SMA, CD56, Vcα7.2, CD3, CD14, CD68, PF4 and CD15 expression. The three mega-clusters were then re-clustered using protein markers selected on immunological basis (Figure S7). The resultant final clusters were annotated using an integrated approach. In the first step, we defined clusters using (i) heatmaps showing median marker expression (ii) expression density histograms which allow better delineation of the range of marker expression, specifically differentiating low and negative expression levels and (iii) cluster distribution plots which showed the frequency of each cluster in different samples. Phenotypic similarity of clusters was interrogated via UMAP and cluster dendrograms. To further define cluster identities, the spatial location of clusters was visualised using cell centroid plots and mapped onto an adjacent H&E slide with the same ROI. Based on these analyses, some clusters were excluded under the following criteria: a) clusters with uniformly low/negative expression of markers, b) clusters only found in one sample, and c) Undefined clusters (where the combination of markers did not amount to a subset which could be defined). These clusters were not submitted for spatial analysis. Sub-clusters with very similar expression profiles were merged and those which contained 2 or more
clusters were annotated as such. A small number of clusters demonstrated expression of markers normally associated with disparate cell populations (e.g. Neutrophil_CD8 adjacent), which can be attributed to closely apposed cell types. These adjacent cell populations were validated via high resolution immunofluorescence microscopy. To aid final annotation, we also performed Pseudotime inference for selected populations.

Final annotated clusters were then sense-checked against the MCD images by an independent investigator not involved in annotating the clusters, and some key clusters of interest were further examined by immunofluorescence staining with confocal microscopy.

**Pseudotime analysis.** The Pseudotime analysis was performed on the macrophage, monocytes and neutrophils populations (Figure S6). Their arcsinh transformed values were integrated using Harmony with the same parameters as in the main analysis, followed by dimensionality reduction using UMAP. Then the Pseudotime inference was performed by applying the Slingshot algorithm (Street et al., 2018) to the UMAP dimensions using the default parameters and the above annotations as clusterLabels.

This analysis is not part of the SpOOx pipeline but code is available in GitHub.

*Command: R slingshot R <parameters>*

**Differential cell abundance analysis.** Differential abundance analysis between conditions was performed using code from the diffcyt R package (version 1.8.8) with the option testDA_edgeR. To account for the differences in area between the ablated samples, the area was used as a normalising factor. The dispersion was estimated using the option trend.method="none" and the negative binomial generalized log-linear model was used for the analysis (with the glmFit and glmLRT functions). The BH method was used to adjust p-values for multiple testing.

**Cell centroid maps.** For each ROI, the cell centroids were plotted and coloured according to cell type to produce a cell centroid map which forms the basis of subsequent analyses. These were overlaid with ROI images in MDV so cell types may be located by colour.

**Spatial analyses**

A suite of mathematical tools for spatial analyses is incorporated in SpOOx (see below under QCM, cross-PCF and ACN). The following command runs all the spatial analysis methods in SpOOx:
**Command:** python hyperion_pipeline.py make spatial stats

It is also possible to run each spatial function separately and adjust parameters (see https://github.com/Taylor-CCB-Group/SpOOx/tree/main/src/spatialstats for details). The command line option is shown after each method below that can be appended to the basic command above.

**Quadrat Correlation Matrix (QCM).** The “Quadrat Correlation Matrix” (QCM) describes correlations between counts of different cell types within square quadrats with edge length 100μm (resulting in between 100 and 400 quadrats per ROI), following an approach used by (Morueta-Holme et al., 2016) to identify statistically significant co-occurrences (p < 0.05) and applied to multiplex images of cancer by (Prabhakaran S., 2022).

We construct the QCM by first generating a matrix $O$ whose entries $O_{ij}$ record the number of cells of type $i$ in quadrat $j$, for $1 \leq i \leq n$ and $1 \leq j \leq m$, where $n$ is the number of cell types in the ROI and $m$ is the number of quadrats. We use $O$ to generate 1000 matrices $N^1, ..., N^{1000}$ which form a distribution of “observations” in which the number of cells of each type and the number of cells in each quadrat are the same as in $O$, but spatial correlations between cell types are removed by shuffling cell labels. Each matrix $N^k$ is such that, for each $j$:

$$\sum_i N_{ij}^k = \sum_i O_{ij},$$

and for each $i$:

$$\sum_j N_{ij}^k = \sum_j O_{ij}.$$

We construct each matrix $N^k$ as follows. We fix $N^{k,0} = O$, and define rules which permute the entries of $N^{k,s}$ to obtain a new matrix $N^{k,s+1}$. This is accomplished by selecting two rows ($a,b$) and two columns ($c,d$) of $N^{k,s}$ at random. For some integer $p$ sampled uniformly at random from the interval $[0, \min(N_{bc}^{k,s}, N_{ad}^{k,s})]$, we then fix:

$$N_{ac}^{k,s+1} = N_{ac}^{k,s} + p,$$

$$N_{bc}^{k,s+1} = N_{bc}^{k,s} - p,$$

$$N_{bd}^{k,s+1} = N_{bd}^{k,s} + p$$
and

\[ N_{ad}^{k,s+1} = N_{ad}^{k,s} - p. \]

This process is repeated for \( s = 0, 1, \ldots, 10,000 \) to ensure that the final matrix \( N^{k} = N_{1000}^{k} \) is well shuffled.

Partial correlation matrices \( C_0 \) and \( C_{N^1} \ldots C_{N^{1000}} \) are then calculated for \( O \) and \( N^1, \ldots, N^{1000} \) respectively. Standard effect sizes (SES) are determined by rescaling the partial correlations in \( C_0 \) by the mean \( \mu \) and standard deviation \( \sigma \) of the \( C_{N^k} \), such that

\[
SES_{ij} = (C_{0ij} - \mu[C_{N^k}]) / (\sigma[C_{N^k}]).
\]

Non-significant associations are identified by calculating a 2-tailed p-value for each pair of cell types and applying a Benjamini-Hochberg correction, with false discovery rate FDR = 0.05. Non-significant entries of SES are set to 0 in order to generate the QCM, a cell association matrix whose non-zero entries identify standardised effect sizes of statistically significantly correlated cell types within the ROI.

The average QCM across \( Q \) ROIs is obtained by concatenating the relevant observation matrices. Denoting by \( O_q \) the observation matrix from ROI \( q \), we concatenate \( O_1, \ldots, O_q \) to form a combined observation matrix \( O = (O_1 \ O_2 \ldots O_q) \), an \( (n \times (m_1 + m_2 + \ldots + m_Q)) \) matrix, where \( m_q \) denotes the number of quadrats in ROI \( q \). Similarly, we concatenate \( N^1, \ldots, N^{1000}_q \) to form \( N^k = (N^k_1 \ N^k_2 \ldots N^k_Q) \).

Standard partial correlation matrices are then calculated and the process described above for a single ROI used to compute the average QCM for multiple ROIs.

**Command option:** –function morueta-holme

**Cross pair correlation functions (cross-PCF).** Significant correlations identified at length scales in the range 0-100\( \mu \text{m} \) via the QCM are further assessed by using cross pair correlation functions (cross-PCFs - see, e.g., Bull 2020). Cross-PCFs quantify clustering and dispersal of pairs of cell populations across a range of length scales (here 0-300\( \mu \text{m} \)). The cross-PCF is calculated at discrete values \( r_k \in [0,10,20, \ldots, 300] \), and considers pairs of cells separated by distances \( r \in [r_k, r_{k+1}) \).

For cell populations A and B, the cross-PCF, \( g(r_k) \), is defined as follows:
where $N_A$ and $N_B$ are the numbers of cells of types A and B, $A_{rk}(x)$ is the area of an annulus (intersecting with the ROI) centred at location $x = (x, y)$ with inner radius $r_k$ and outer radius $r_{k+1}$, $x_a$ and $x_b$ are the spatial coordinates of cells $a$ and $b$ (of types A and B respectively), $I_{[r_k, r_{k+1}]}(r)$ is an indicator function ($I_{[r_k, r_{k+1}]}(r) = 1$ if $r \in [r_k, r_{k+1})$ and $I_{[r_k, r_{k+1}]}(r) = 0$ otherwise), and $\rho_B$ is the density of cells of type B in the ROI.

A cross-PCF with $g(r)>1$ means that cells of type A are observed more frequently at distance $r$ from cells of type B than would be expected under CSR, and is indicative of clustering at distance $r$. Conversely, a cross-PCF with $g(r) < 1$ means that cells of type A are observed less frequently at distance $r$ from cells of type B than would be expected under CSR, and is indicative of exclusion.

For individual ROIs, 95% confidence intervals are obtained via bootstrapping. The spatial dependence of resampled points is accounted for by resampling grid sites within a 20μm square lattice, following (Loh, 2008).

To aid comparison between the clustering and dispersal of different pairs of cell populations, we frequently report cross-PCF values at, $r_k = 20$, corresponding to length scales in the range $r \in [20, 30]$ μm. We focus on $r_k = 20$ since it approximates the distance between the centroids of cells which are in physical contact. For notational simplicity, we denote this value as $g(r=20)$.

**Command option:** --function paircorrelationfunction

**Topographical Correlation Map.** The cross-PCF quantifies clustering and dispersal of pairs of cell populations at different length scales within an ROI. We also introduce the Topographical Correlation Map (TCM), to visualise how the spatial correlation between cells of types A and B, say, changes across an ROI.

In order to define $\Gamma_{ab}$, the TCM for cells of types A and B, we first associate a mark $m_{ab}$ with each cell $a$ of type A. The mark $m_{ab}$ is defined to be the ratio of $b$, the number of cells of type B within 100μm of cell $a$, to the expected number of cells of type B if they were distributed according to CSR:
\[ m_{ab} = \frac{N_B}{\rho_B A_{100}(x_a)} \int_{[0,100]} \left[ I_{[0,100]}(r) \right] \, dr \]

where \( \rho_B \) is the density of cells of type B in the ROI, \( A_{100}(x_a) \) is the area of a circle with radius 100\( \mu \)m centred at \( x_a = (x_a, y_a) \) which falls within the ROI, \( I_{[0,100]}(r) \) is an indicator function \( (I_{[0,100]}(r) = 1 \) when \( 0 \leq r < 100 \) and \( I_{[0,100]}(r) = 0 \) otherwise), and \( N_B \) is the total number of cells of type B within the ROI. We interpret values of \( m_{ab} \) in a manner similar to that used for cross-PCFs: \( m_{ab} < 1 \) indicates anti-correlation between cells of types A and B within a distance of 100\( \mu \)m, and \( m_{ab} > 1 \) indicates correlation.

To facilitate visualization and interpretation, we normalize the mark \( m_{ab} \) by introducing the transformed mark, \( M_{ab} \), where:

\[
M_{ab}(m_{ab}) = \begin{cases} 
1 & \text{if } m_{ab} \geq \alpha \\
\frac{m_{ab} - 1}{\alpha - 1} & \text{if } 1 < m_{ab} \leq \alpha \\
\frac{1 - m_{ab}}{\alpha - 1} & \text{if } \frac{1}{\alpha} < m_{ab} < 1 \\
-1 & \text{if } m_{ab} \leq 1/\alpha 
\end{cases}
\]

The constant \( \alpha \) defines a threshold for extreme clustering. If \( m_{ab} > \alpha \) then we have strong clustering and we fix \( M_{ab} = 1 \); if \( m_{ab} \leq 1/\alpha \) then we have strong exclusion and we fix \( M_{ab} = -1 \). The shape of \( M_{ab} \) is presented in Figure S13.

After calculating \( M_{ab} \) for each cell of type A, we centre a Gaussian kernel, with standard deviation \( \sigma = 50 \mu \)m, and maximum height \( M_{ab} \), at \( x_a \). We sum the kernels associated with all the cells of type A to generate the TCM, \( \Gamma_{ab} \):

\[
\Gamma_{ab} = \sum_{\alpha=1}^{N_A} \frac{M_{ab}}{\sqrt{2\pi}} e^{-\frac{1}{2}(x - x_a)^2}
\]

The TCM permits identification of spatial locations in which cells of type A are positively (\( \Gamma_{ab} \gg 1 \)) or negatively (\( \Gamma_{ab} \ll 1 \)) associated with cells of type B. For computational efficiency, when calculating \( \Gamma_{ab} \) we assume that each kernel has compact support, centred in a square region of edge length 300\( \mu \)m.
We note the following properties of the transformed mark, $M_{ab}$. First, $M_{ab}(m_{ab}) = -M_{ab}\left(\frac{1}{m_{ab}}\right)$, so that dispersal and clustering are measured on the same scales. For example, $m_{ab} = 2$ indicates the presence of twice as many cells of type B as expected under CSR, while $m_{ab} = 1/2$ indicates the presence of half as many cells of type B as expected under CSR. Secondly, the magnitude of $M_{ab}$ describes the strength of the spatial interaction. Finally, the sign of $M_{ab}$ identifies whether there is clustering ($M_{ab} > 0$) or exclusion ($M_{ab} < 0$) between cell $a$ (of type A) and cells of type B.

The parameter $\alpha$ characterises the most extreme clustering or exclusion which can be resolved in each kernel, with extremal values being mapped to 1 and -1 respectively. We use $\alpha = 5$, so clustering or exclusion stronger than 5x is interpreted as the strongest clustering/exclusion that we can distinguish.

Finally, we note that $\Gamma_{ab} \neq \Gamma_{ba}$, since the kernels used to construct $\Gamma_{ab}$ are centered on cells of type A (and vice versa). While areas in which cells of type A and type B are co-located should be identified by both $\Gamma_{ab}$ and $\Gamma_{ba}$, their values will differ in regions rich in one cell type and poor in another. We therefore stress that $\Gamma_{ab}$ describes locations in which cells of type A are correlated or anti-correlated with cells of type B, and that the presence or absence of cells of type B cannot be inferred from regions in which $\Gamma_{ab}$ is close to 0.

Command option: --function locallusteringheatmaps

**Adjacency Cell Networks.** We use the cell segmentation masks generated by DeepCell to produce a spatially-embedded adjacency cell network (ACN), in which nodes represent cell centres and are labelled according to their cell type. Nodes are connected by an edge if the corresponding cells in the segmentation mask share a border. To ensure that small perturbations in cell boundaries do not lead to errors in cell connections, we expand the border of each segmented cell by 5 pixels before generating the network.

We use the ACN to define two statistics for each pairwise combination of cell types A and B. First, we compute $\phi_{AB}$, the proportion of cells of type A which are in contact with at least one cell of type B:

$$\phi_{AB} = \frac{1}{N_a} \sum_{a=1}^{N_A} I_B(a)$$
where $N_A$ is the number of cells of type A and $I_B(a) = 1$ if cell $a$ is connected with a cell of type B and $I_B(a) = 0$ otherwise. Secondly, we calculate $\Phi_{AB}$, the average number of cells of type B that are in contact with a cell of type A:

$$
\Phi_{AB} = \frac{1}{N_A} \sum_{a=1}^{N_A} \eta_B(a)
$$

where $\eta_B(a)$ is the number of cells of type B in contact with cell $a$.

In this paper, we used ACN to calculate the proportion of cells of type A that have at least one cell of type B in contact with them and the average number of cells of type B that is in contact with a cell of type A in the ROI.

*Command option: --function networkstatistics*

**Multi-Dimensional Viewer (MDV)**

MDV is a comprehensive spatial analytics platform that facilitates the interrogation of large complex data sets and includes various interactive dashboards to facilitate quality control, interactive clustering, phenotyping and spatial analysis. It is an open source web application which can be downloaded and installed locally or used on the publicly available web site [http://mdv.molbiol.ox.ac.uk](http://mdv.molbiol.ox.ac.uk). Users register to use the site and projects can private, shared with other users or made public. Full documentation and tutorial videos are provided on the MDV website but we provide an overview here. Supplementary videos are also included to illustrate functionality (See Video).

MDV allows output generated by the SpO0x pipeline to be loaded at different states. Data locations are specified in a yaml format file which can be edited by the user (command: python mdvupload.py myconfig.yaml). Examples of data tables that may be uploaded are:

- Image data (PNGs/OME-TIFF stacks): ROI image stacks, H and E images binary cell masks.
- Cell data (tab separated file): one cell per row, including size, size, shape, phenograph clusters identification, UMAP coordinates, marker signal intensities.
- Spatial Statistics data (tab separated file): one row containing cell to cell interaction data and associated statistics.
Data related to the disease states (JSON file): allowing grouping of samples for high level analysis.

Once uploaded the data are presented in MDV as a series of views that contain multiple interactive charts corresponding to different analytical methods from clustering, annotation, cell centroid visualisation and spatial analytical methods. Each view focuses on a particular aspect of the pipeline. View contents can be adjusted and added to by adding other chart types and saved as a new view. Chart types can be D3 components (https://d3js.org/) but we have also written custom chart types for performance reasons. For example, MDV scatterplot chart can visualise and interrogate at least 10 million data points. We also integrate Viv viewer (Manz et al., 2022) to visualise composite image stacks.

The complete analysis and data set were published by publicly sharing the data at https://mdv.molbiol.ox.ac.uk/projects/hyperion/6356.

**COMBAT data mapping**

**COMBAT CyTOF data generation and processing.** Cell suspension mass cytometry (CyTOF) data were generated by the COMBAT consortium as previously described (COMBAT, 2022). In brief, whole blood from COVID-19 patients was stabilised using a Cytodelics fixative solution, red blood cells were lysed, cellular material was fixed, and samples were run in a Helios CyTOF machine. Importantly, samples were enriched for mononuclear cells before profiling by performing magnetic depletion of CD66+ granulocytes.

After acquisition, data were formatted into a single-cell protein abundance table and annotated into cell types based on marker expression (COMBAT, 2022). For the analyses in the present study, this expression matrix was split into two subsets: one containing T and NK cell types, and a second one containing myeloid cell types (i.e. monocyte subsets).

**Mapping cells from lung tissue to blood cells from the COMBAT study.** Cells in the lung dataset were matched to the most closely related cell types in blood using scmap, a method which enables label projection by calculating the similarity between cells profiled by two separate studies (Kiselev et al., 2018). In brief, CyTOF and CITE-seq expression matrices from the COMBAT study were used to build index references for label projection. First, proteins which were detected in both studies were
identified. This resulted in a panel of 13 and 18 proteins shared between our study and the COMBAT CyTOF and CITE-seq panels, respectively. Next, these proteins were used as a basis for cell type classification with the scmapCluster function. Classification accuracy was tested by splitting the COMBAT data into training and test sets containing 80% and 20% of cells, respectively. The training set was used to generate the scmap reference index, while the test set was used to assess cell type prediction accuracy (Kiselev et al., 2018). Given the reduced set of markers shared between studies, not all COMBAT cell populations could be accurately predicted. Thus, in order to maximise predictive accuracy similar subpopulations of the same cell type were merged into a single group and any cell types known to be absent from our lung data, such as B cells and plasmablasts, were removed. This approach achieved over 70% accuracy for CyTOF data (71% and 78% predictive accuracy for myeloid and lymphoid cell types, respectively) and 85% accuracy for CITE-seq data and components of final merged clusters are shown in Figure S15.

Indexed references were next used to match cells in the lung to the most similar clusters in blood using the scmapCluster() function. To do so, cell type labels were predicted for each cell in the lung based on the CyTOF and CITE-seq reference sets. Any unassigned cells were discarded. Cluster overlap between studies was visualised using Sankey diagrams (Schmidt, 2008) and quantified using Jaccard indexes (Jaccard, 1912).

**Neutrophil subset analysis from stored COMBAT samples**

Whole blood samples frozen in whole blood cell stabilizer (Cytodelics) were obtained from COMBAT consortium storage for healthy volunteers (n=11), health care workers (n=12), COVID-19 (n=93) and Sepsis (n=48). Pre-processed CD45^+ gated FCS files of granulocyte containing whole blood samples were analysed with R (v4.0.0). 50000 cells per sample were integrated using Harmony (v1.0)^2 and the CATALYST package (v1.14.0)^3 was used for downstream analysis. CD45^+ cells were clustered based on the FlowSOM and ConsensusClusterPLus algorithms using the cluster () function. 50 metaclusters (xdim=100, ydim=100, k=50) were then assigned to major cell types (T cells, B cells, plasmablasts, mononuclear phagocytes and neutrophils). Neutrophils were selected and reclustered based on CD45, CD15, CD38, CD64, CD16, CD43, CD66b CD10, CD33, Kl67, CD172a/b, CD141, CD71, CD114, CD371 and CD274 expression. 30 neutrophil metaclusters (xdim=100, ydim=100, k=30) were manually merged to 8 neutrophil clusters (proNeut, preNeut, iNeut1-3, mNeut1-3) based on median marker expression.
Datasets


SpOOx pipeline code and details of how to run it are available at https://github.com/Taylor-CCB-Group/SpOOx

Acknowledgments

We thank patients’ relatives for donation of the patients’ lung tissue. This work was funded by the Oxford University Medical Sciences Division COVID Funds, NIHR Oxford Biomedical Research Centre and the Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Science (GIFMS), China (grant number: 2018-12M- 2-002). LD and CV are supported by the NIHR Oxford Biomedical Research Centre. LPH is supported by MRC Human Immunology Unit (grant MC_UU_00008/1) and the NIHR Oxford Biomedical Research Centre.

Author contributions

PW analysed the data, contributed to development of MDV, spatial analyses, and interpretation of mathematical output and all spatial data, and writing of the paper. LD contributed to analysis and interpretation of data, optimised and performed the staining of the lung sections in conjunction with RE, analysed MCD images and contributed to the writing of the paper. JB performed all the mathematical development and analysis in conjunction with HB and contributed to interpretation of data and writing of the paper. ER performed all bioinformatic analysis and contributed to interpretation of data and writing of the paper. CV performed all the immunofluorescence and imaging of sections. GDHT and CC performed all the histopathology analyses in conjunction with LPH. AC performed the protein immunostaining of the sections and contributed to interpretation of data. CEDA and IMB organised acquisition of patient samples, clinical data and ethical permissions. YXZ and DH optimised, performed CYTOF experiments on neutrophils and analysed data. ECG and JW performed all COSMIC v COMBAT data analysis in conjunction with PW, LPH and JK. DR and PK interpreted data, acquired some of the funding for the study and contributed to writing of the paper. TD, IAU, GO, CM, JK, Fl interpreted data, discussed annotations and immunological analysis, and contributed to writing of the paper. ST oversaw all computational work and code writing for the study, contributed to analysis of data and writing of paper. MS performed all the dataset organisation and spatial analysis set up in MDV, wrote the codes and organised MDV in conjunction with PW, ST, and LPH. HB oversaw all mathematical development and contributed to writing of the
paper. LPH conceptualized, led the study, acquired funding, analysed the data and wrote the manuscript. All authors read and approved the final manuscript.

References


Jaccard, P. (1912). The distribution of the flora in the alpine zone. New Phytol 11 (2)


Histopathology analysis

Tru-cut biopsy at point of death, under ultrasound guidance

Early
Late

Inflammation-Damage-Repair

35-plex metal-tagged antibody staining
IMC

Segment single cell in tissue

Determine cell identity

Mathematical analysis of spatial organization

Map cell with identity back to tissue

SpO2x pipeline on multi-dimensional viewer (MDV) as resource for IMC

Key finding

Nidus of inflammation and injury
A. Statistical quantification of spatial co-location of cells

Step 1. Quadrat correlation matrix

Step 2. Cross pair correlation functions (cross-PCF)

Step 3. Adjacency Cell Network (ACN)

Visualization of results

Radial connectivity map

Topographical correlation map

Spatial network plot

B. ALV, DAD, OP

C. Days from first symptom

D. CRP

Figure 1
Figure 2
A. Structural cells

B. Myeloid cells

C. Lymphoid cells

Figure 3
A.

All possible pairs of cell clusters from all COVID and HC samples. Submit for QCM analyses.

Filter out from QCM output:
- self: self pairs (automatically excluded by QCM analysis), structure: structure pairs, correlation with FDR q > 0.05, all cell clusters annotated as 'undefined' (UD).

Result: 367 COVID-19 pairs of cells.

Submit for cross-PCF analysis.

From cross-PCF output, select co-located pairs defined as:
- pairs with positive correlation coefficient from QCM analysis
- g(r=20)>1 and lower 95% C.I. does not cross g(r=20)=1

Result: n=324 COVID-19 pairs of cells.

Organize co-located pairs into two lists for further interrogation of cross-PCF and ACN output.

- 'structure: immune' cell cluster pairs
  - n=35 COVID-19 pairs of cells
  - only one of reciprocal pairs included in list. Pairs containing any fibroblast subsets are analyzed separately

- 'immune: immune' cell cluster pairs
  - n=100 COVID-19 pairs of cells

B.

Figure 4
Figure 5
Figure 6