1 **Single cell spatial analysis reveals inflammatory foci of immature neutrophil**
2 **and CD8 T cells in COVID-19 lungs**
3

4 Praveen Weeratunga[1], Laura Denney[1], Joshua A. Bull[2], Emmanouela Repapi[3], Martin Sergeant[3], Rachel

5 Etherington[1], Chaitanya Vuppussetty[1], Gareth D.H. Turner[4], Colin Clelland[5], Jeongmin Woo[1], Amy

6 Cross[6], Fadi Issa[6], Carlos Eduardo de Andrea[7], Ignacio Melero Bermejo[7], David Sims[3], Simon

7 McGowan[3], Yasemin-Xiomara Zurke[8], David J. Ahern[8], Eddie C Gamez[9], Justin Whalley[9], Duncan

8 Richards[10], Paul Klenerman[11], Claudia Monaco[8], Irina A. Udalova[8], Tao Dong[1,12], Agne Antanaviciute[1],

9 Graham Ogg[1,12], Julian C. Knight[9,12], Helen M. Byrne[2, 13], *Stephen Taylor[3],& *Ling-Pei Ho[1,12,14].

10

11 These authors contributed equally: Praveen Weeratunga, Laura Denney, Joshua A. Bull and

12 Emmanouela Repapi

13

14 Corresponding authors

15 Stephen.Taylor@well.ox.ac.uk

16 Ling-pei.ho@imm.ox.ac.uk

17

18 **Affiliations**

19 1. MRC Human Immunology Unit, MRC Weatherall Institute of Molecular Medicine, University
20   of Oxford, UK
21 2. Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, UK
22 3. MRC WIMM Computational Biology Unit, MRC Weatherall Institute of Molecular Medicine,
23   University of Oxford, UK
24 4. Department of Cellular Pathology and Radcliffe Department of Medicine, Oxford University
25   Hospitals NHS Foundation Trust, UK
26 5. Anatomic Pathology, Weill Cornell Medical College, Doha, Qatar
27 6. Nuffield Department of Surgical Sciences, University of Oxford, UK
28 7. Navarra Institute for Health Research, Pamplona, Spain
29 8. Kennedy Institute for Rheumatology, University of Oxford, UK
30 9. Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of
31   Oxford, UK
32 10. Nuffield Department of Orthopedics, Rheumatology and Musculoskeletal Diseases,
33   University of Oxford, UK
34 11. Nuffield Department of Medicine, University of Oxford, UK
35 12. Chinese Academy of Medical Science (CAMS) Oxford Institute (COI), University of Oxford,
36   Oxford, UK
37 13. Ludwig Institute for Cancer Research, University of Oxford, UK
38 14. Respiratory Medicine, Nuffield Department of Medicine, University of Oxford, UK.
39
40
41
42
43

44   **Abstract**

45

46
47   Single cell spatial interrogation of the immune-structural interactions in COVID -19 lungs is challenging,

48   mainly because of the marked cellular infiltrate and architecturally distorted microstructure. To

49   address this, we develop a suite of mathematical tools to search for statistically significant co-locations

50   amongst immune and structural cells identified using 37-plex imaging mass cytometry. This unbiased

51   method reveals a cellular map interleaved with an inflammatory network of immature neutrophils,

52   cytotoxic CD8 T cells, megakaryocytes and monocytes co-located with regenerating alveolar

53   progenitors and endothelium. Of note, a highly active cluster of immature neutrophils and CD8 T cells,

54   is found spatially linked with alveolar progenitor cells, and temporally with the diffuse alveolar damage

55   stage. These findings offer further insights into how immune cells interact in the lungs of severe

56   COVID-19 disease. We provide our pipeline [Spatial Omics Oxford Pipeline (SpOOx)] and visual-

57   analytical tool, Multi-Dimensional Viewer (MDV) software, as a resource for spatial analysis.

58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85

86

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[1], 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 [2-5]. 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 [6-10]. Our recent comprehensive multi-modal study of circulating immune cells (COMBAT study)[6] and several other major studies have also concluded that a key hallmark of severity was emergency myelopoiesis[6,8,9,11,12], 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[13-18]. 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. Notwithstanding these limitations, studies in intact COVID-19 lung tissue are also challenging, due to the distorted lung micro-architecture and massive cellular infiltrate, making it difficult to unravel cellular connectivity and organisation. 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 lung samples obtained later in infection but did not identify reveal further insight into the cause of severe alveolar damage [19].

In this study, we develop a bespoke mathematical package to identify statistically significant co-location between different cells, including structural cells, at the level of single cell resolution. We identify a cluster of closely apposed immature neutrophils and CD8 T cells with high immune activity,

121   which are spatio-temporally associated with proliferating alveolar epithelium in tissue sections
122   demonstrating diffuse alveolar damage.   These findings raise the possibility of an injurious entity
123   generated by the interaction between immature neutrophils and a specific subset of CD8 T cells in
124   severe COVID-19 pneumonitis.

125

126

127   **RESULTS**

128

129   **An integrated pipeline to uncover and quantify spatial association amongst cells**

130

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

147

148   These work packages were integrated computationally into a workflow of Python and R based
149   command line tools which may be run individually or as an automated pipeline (Spatial Omics Oxford
150   pipeline; SpOOx) (Fig. 1A). The pipeline is supported by a visualisation platform [Multi-Dimensional
151   Viewer (MDV)] (Video V1). Both are available as an open access online resource (see Methods for link).
152   To differentiate this lung-based from our blood-based study (COMBAT)[6], we have called this the
153   COSMIC (COVID-19 Lung Single Cell Mass Cytometry Imaging Consortium) study.

154

155

156

157 **Histopathology states of inflammation, damage and repair are found in lung sections at point of**

158 **death**

159

160 We started by examining formalin-fixed paraffin-embedded (FFPE) lung sections from a cohort of

161 patients who died from PCR-positive COVID-19 pneumonitis from one hospital (University of Navarra,

162 Spain) (n=12). Samples were obtained at the point of death and fixed immediately, markedly reducing

163 post-mortem tissue deterioration [20]. All samples were collected during the first wave of the pandemic

164 in 2020, before vaccination and repeat infection with SARS-CoV-2. Healthy lung sections from patients

165 undergoing lobectomy for early, isolated lung cancer (HC) lungs (n=2) were used as comparators

166 (Demographics in Supplementary Table 1); obtained from the Oxford Radcliffe Biobank (Oxford

167 University Hospitals NHS Foundation Trust, UK).

168

169 Six of 12 patients were mechanically ventilated (range 6-23 days). All but three were receiving

170 corticosteroids at the point of death (Supplementary Table 1). In all patients, thoracic CT scans closest

171 to the day of death demonstrated typical and extensive COVID-19 pneumonitis comprising ground

172 glass changes and consolidation (Supplementary Table 1). Five of 12 lung sections showed evidence

173 of both PCR and immunostaining for SARS-CoV2 Nucleocapsid protein; 3 were PCR+ but protein

174 negative (Supplementary Table 2). Four sequential lung sections (6 μm thick) were used for

175 haematoxylin and eosin (H&E), 37-plex panel staining (35 metal-tagged antibodies and 2 DNA

176 chelators), and selected immunofluorescence validation sequentially.

177

178 Initial histopathology analysis (independently performed by two senior pathologists with expertise in

179 lung and infectious disease, and a senior respiratory clinician) revealed a highly distorted lung

180 architecture with extensive cellular infiltrate in all samples, changes previously observed in post

181 mortem studies of COVID-19 lung sections[21-26]. However, all our sections can be categorised into three

182 formal histopathology classifications of predominantly alveolitis (ALV), diffuse alveolar damage (DAD)

183 or organising pneumonia (OP) (n=4 patients in each category)[27,28] (Fig. 1B and Supplementary Fig. 2).

184 ALV was characterised by thickened alveolar epithelial wall and septae with immune cell infiltrate and

185 congestion of alveolar walls; DAD, by widespread alveolar epithelial lining injury accompanied by

186 hyaline membrane, regenerating/proliferating Type II alveolar epithelium and interstitial oedema,

187 while OP depicts a repair state typified by presence of fibroblasts, proliferation of alveolar epithelium

188 and collagen presence around bronchial epithelium [28]. In keeping with this, patients with dominant

189 OP histopathology showed a trend of being sampled furthest away from their first symptoms (Fig. 1C),

190 had longer periods of stay in hospital and were mechanically ventilated for longer (Supplementary

191 Table 1 and Supplementary Fig. 3) (no statistical difference observed). All 5 patients with evidence of

192 dual SARS-CoV-2 N protein and PCR expression had lung sections that showed DAD. No sections with

193 OP were positive for SARS-CoV-2 protein staining (Supplementary Table 2; Supplementary Fig. 4).

194 There were no associations between histopathology states and clinical features (age, drugs used, co-

195 morbidities, or C-reactive protein (CRP) nearest the point of death) (Supplementary Table 1, Fig. 1D).

196

197 These results provide a histopathology-based temporally progressive states for further analysis. 2-3

198 regions of interest (ROIs) per patient (total of 4 mm^2 area per patient), selected as representative

199 areas for the dominant histopathology state, were drawn for ablation.

200

201 **Identification of immature neutrophils - CD8 T cell clusters with high immune activity.**

202

203 After preliminary staining with an initial panel (Supplementary Table 3, Supplementary Fig. 5) on a

204 'sentinel' cohort, we designed a final panel which incorporated the most abundant structural and

205 immune cell types (Supplementary Fig. 7A). Single cell segmentation performed using Mesmer library

206 from the DeepCell algorithm [29] resulted in 677,623 single cells from all ROIs, [ALV (n=10 ROIs), DAD

207 (n=8), OP (n=8) and HC (n=4)] (Fig. 2A, Supplementary Fig. 7 and 12). Cell clusters were derived using

208 Phenograph and annotation was performed using a combination of expression heat map analyses and

209 expression density plots (Supplementary Fig. 7). Final annotation was refined with Pseudotime

210 analysis of selected groups of cell clusters, examination of distributions of the cell clusters in all

211 samples, and cross-checking with H&E and MiniCAD Design (MCD) images (generated by the Hyperion

212 imaging system) against known structural cell location and cell morphology (Supplementary Fig. 7).

213 This produced a final list of 37 identifiable cell clusters (26 immune cell types and 11 structural) (Fig.

214 2B-C). Expanded description of the annotated cells is provided in Supplementary Table 4. For clarity

215 of terminology, once the cell clusters were annotated, they were termed 'cell types' or 'cells' unless

216 there were more than two cell types in the annotation.

217

218 Compared with healthy lungs (without dividing into different histopathology states), monocytes were

219 the most abundant immune cells (Fig. 2B). Amongst the annotated cell types, five were found to co-

220 express defining markers of different immune or structural cells, reflecting closely located or apposed

221 cell groups (Neutrophil and CD8 T cell, Monocyte and CD31$^+$ cells, Monocyte- PAI-1$^+$ cells, CD8 T cells

222 – PAI-1$^+$ cells and IFN-γ^{hi} cells and RAGE$^+$ cells). We labelled these 'adjacent' (ADJ) cell types.

223   Immunofluorescence staining confirmed presence of two different adjacent cells for the
224   Neut_CD8_ADJ (immature neutrophil and CD8 T cells) (Fig. 2D, and Supplementary Fig. 7L).
225   Mono_CD31_ADJ comprised both monocytes that were found adjacent to endothelial cells and CD31-
226   expressing monocytes (Fig. 2E).

227

228   Of note, the Neut_CD8_ADJ cells contained the most immature neutrophil cell type (CD71^{hi}
229   neutrophils) coupled with CD45RO$^+$ CD107a$^-$CD8 T cells (Fig. 2F). The cluster also had the highest
230   expression of Granzyme B (GZB), CD172a (SIRPA), IFN-β and IFN-γ (Fig. 2F, G-I). Within the cluster, the
231   GZB expression was found on the CD8 T cells, indicating these as cytotoxic CD8 T cells (Fig. 2I). The
232   monocyte subset in the Mono_CD31_ADJ cluster was the least differentiated (to macrophage)
233   monocyte subset; similar to the Mono_1 cell type (Fig. 2D, Supplementary Fig. 7L).

234

235   **High innate immune cell numbers found in all histopathology states**

236

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

246

247   Across the three histopathology states, we found high numbers of classical monocytes, immature
248   neutrophils, and some subsets of MAIT, CD4 and CD8 T cells. The most significant progressive increase
249   in numbers across the histopathological states (compared to healthy lungs) was observed for CD8 T
250   cell subsets and CD8 containing ADJ cell clusters, and CD107a$^+$ CD4 T cells (Fig. 3B-D) (see
251   Supplementary Table 4 for expanded phenotype description of immune cell types). Neut_CD8_ADJ
252   cluster was increased from the earliest histopathology state and remained high in all states. Apart from
253   IFN-γ^{lo} MAIT cells, there were only small numbers of MAIT and NK cell subsets (Fig. 3C). Cycling (Ki67$^+$)
254   monocytes were not found in the lungs.

255

256 Overall, innate cell numbers in the infiltrate did not decline despite disease progression and was

257 accompanied by increasing numbers of CD8 T cells [even though viral protein was absent in OP (repair)

258 samples]. Some immune correlates of severity in the blood observed in other studies (cycling

259 monocytes, NK cells, and activated MAIT cells) were not found in significant numbers in the lungs.

260

261 **Distinct spatial organisation found between immune and structural cells**

262

263 To determine if the cells showed spatial association and organisation amongst themselves, we

264 employed spatial statistical algorithms (Fig. 1A and Supplementary Fig. 1) to (a) understand which

265 immune cells were found co-located with injured structural cells, (b) explore how immune cells

266 organise themselves amongst themselves, and (c) for the immune cells implicated in severe disease

267 from COMBAT (monocyte, megakaryocyte, MAIT, CD4, CD8 and neutrophil subsets), if these were co-

268 located or physically interacting with other immune or structural cells

269

270 In total, we found 3888 non-replicate pairs of cell types (mono1:mono1 and CD15^{hi} iNeut:CD15^{hi} iNeut

271 were examples of pairs of identical cell types, and filtered out) in the three histopathological states

272 (ALV, DAD and OP). Using our three-step spatial analysis, 357 pairs of cell types were identified as

273 statistically correlated in the QCM analysis (FDR <0.05). These cell pairs were submitted for cross- PCF

274 analysis, with one cell type in the pair defined as the 'anchor cell' -the cell against which statistically

275 significant connections were quantified. By pre-analysis consensus, pairs of cell types with borderline

276 statistical significance i.e. FDR values between 0.05 and 0.10 were also submitted to prevent loss of

277 biologically relevant data from hard mathematical cut-off. The resulting co-located cell pairs were

278 divided into 'structure:immune' pairs (structural cells were designated the 'anchor' cell type) (n=33)

279 and 'immune: immune' pairs (one cell type in one of the duplicate pairs was designated the anchor

280 cell type; e.g. for the CD107a^{+}CD8:mono1 pair and mono1: CD107a^{+}CD8 pair; CD107^{+}CD8 in the former

281 pair was made the anchor cell, and the latter pair was excluded) (n=117) (Fig. 4A). 'Structure' cells of

282 interest were the key structural cells that were known to be inflamed or damaged in COVID-19

283 pneumonitis – endothelium ('Endothelium' and 'proliferating endothelium') and larger blood vessels

284 ('blood vessels'), alveolar epithelial cells ('proliferating alveolar epithelium') and bronchial epithelial

285 cells ('HLA DR^{hi} bronchial epithelium' 'HLA DR^{lo} bronchial epithelium' and 'bronchial epithelium'). We

286 were particularly interested in 'proliferating alveolar epithelium' as their markers and location in the

287 lung sections suggest they were likely the type II alveolar epithelial cells, the purported progenitors

288 (or stem cells) of alveolar epithelium (Supplementary Fig. 7J and K). The ability of these cells to

289   differentiate to type 1 alveolar epithelium is critical to normal repair and alveolar regeneration after

290   viral induced damage [30-32].

291

292   Amongst the immune cells, the strongest co-location, depicted by g(r=20)>2 [i.e. >2 times more cells

293   of type B observed at 20 μm from cells of type A (anchor cell) than expected under complete spatial

294   randomness], was observed for pairs of immune cell types that belonged to the same immune

295   phenotype, e.g. Mac1 and Mac2 (macrophages), and the CD4 and CD8 T cell types (Fig. 4B). This was

296   expected biologically and provided a degree of validation for the mathematical analysis. For example,

297   close association between helper CD4 T cells (IFN-γ+ CD4 T cells) and cytotoxic CD8 T cells (CD107a+

298   CD8 T cells) is expected as the former plays critical roles in aiding the latter's anti-viral activities.

299   However it is notable that this close physical relationship persists in OP, despite lack of viral protein at

300   this stage of the disease (Supplementary Fig. 9).

301

302   These results signify presence of specific spatial organisation for several immune and structural cells

303   despite appearance of disorder in tissue. The strongest co-location between all cells was found

304   between CD4 and CD8 T cell subsets, particularly active effector memory CD4 T cells (IFN-γ$^+$ CD4 T

305   cells) and cytotoxic CD8 T cells (CD107a$^+$ CD8 T cells), which did not lessen with progression to repair,

306   and despite absence of viral proteins.

307

308   **Immature neutrophil-CD8 T clusters are co-located with proliferating alveolar epithelium in regions**

309   **with maximal alveolar damage**

310

311   For the significantly co-located pairs of cells, we next questioned which immune cells were found co-

312   located with injured structural cells. To provide a composite view of the multiple output from our

313   spatial analysis, we generated a 'spatial connectivity plot' to show all cell types that were statistically

314   co-located with a designated 'anchor cell type'. Each spatial connectivity plot displayed the strength

315   of co-location [g(r=20)] and the average count for the immune cell types in the histopathology state

316   (Fig. 5A, B). The proportions of co-locating cell types which were in direct contact with the anchor cell

317   type were calculated with the ACN analysis (see Methods) and shown in the accompanying histograms

318   (Fig. 5C and D).

319

320   Our main structural cell types of interest were the Ki67$^+$ proliferating alveolar epithelial cell and

321   endothelial cells. Designating proliferating alveolar epithelium as the anchor cell, we found CD15^{hi}

322   iNeut, Mono_CD31_ADJ and Neut_CD8_ADJ to be significantly co-located with proliferating alveolar

323 epithelium in DAD (Fig. 5A, B) [g(r=20)>1]. Of these cells, proliferating alveolar epithelium was most in

324 contact with Mono_CD31_ADJ (average of 17.6% of proliferating alveolar epithelial cells in DAD) and

325 Neut_CD8_ADJ (8.9% of proliferating alveolar epithelial cells in DAD) (Fig. 5C). There was also a small

326 number of IFN-γ^{hi}_RAGE_ADJ cells found co-located with proliferating alveolar epithelium in all

327 histopathology states, which could be resident alveolar macrophages found along alveolar epithelium.

328

329 For endothelial cells (which encompassed the smaller capillaries and the larger blood vessels in the

330 lungs), the co-locating cell types with highest g(r=20) in DAD were Mono_CD31_ADJ (2.1) and

331 Mono_PAI-1_ADJ (1.6) clusters (Fig. 5B,F). ACN analysis showed more of the endothelial cells were

332 physically in contact with the Mono_PAI-1_ADJ cluster (21.2%) than Mono_CD31_ADJ (16.5%) in DAD

333 (Fig. 5D). Mono_CD31_ADJ cells showed significant spatial association with endothelial cells across all

334 histopathology states.

335

336 Next, we designed a 'radial connectivity map' to provide an overview of all immune cells that were

337 significantly co-located with all structural cells and their corresponding histopathology states (Fig. 5G).

338 Using this map, and focusing on proliferating alveolar epithelium and endothelial cells, we observed

339 that while the monocytes (and their subsets and ADJ clusters) were mainly found co-located with both

340 alveolar epithelium and endothelial cells, immature neutrophils were found predominantly with

341 proliferating alveolar epithelium. We also observed that besides proliferating alveolar epithelium, the

342 Neut_CD8_ADJ cluster was not found with any other structural cell types.

343

344 Finally, we developed a topographical correlation map (TCM) (Methods, Supplementary Fig. 15) to

345 visualise how the spatial correlation between Neut_CD8_ADJ and proliferating alveolar epithelium

346 changed across an ROI (Fig. 5H). We observed marked heterogeneity in the strength of correlation for

347 this pair of cell types across the tissue.

348

349 One other cell type of interest was the megakaryocyte. These CD34$^-$ platelet precursors, a product of

350 emergency myelopoiesis, were the most abundant immune correlate in the blood in the COMBAT

351 study [6]. Examining their spatial connections with our two structural cells of interest, we observed

352 that megakaryocytes were associated with endothelium in DAD (Fig. 5G).

353

354 Drawing these data together, our spatial analysis identified Neut_CD8_ADJ and Mono_CD31_ADJ

355 clusters as key spatial correlates with proliferating alveolar epithelium in DAD. A visual exemplar of

356 this co-location of Neut_CD8_ADJ and alveolar epithelium is shown in Fig. 5I. Mono_CD31_ADJ and

357   Mono_PAI-1_ADJ were the strongest spatial correlates with endothelial cells, the former was the case

358   across all states. No immature neutrophils (alone or in an ADJ cluster with CD8 T cells) were found

359   with endothelial cells in any histopathology states. It is noteworthy that there was no significant co-

360   location between any immune cells and the larger blood vessels; nor between CD107a$^+$ CD8 T cells

361   and IFN-γ^+ CD4 T cells with proliferating alveolar epithelium or endothelial cells despite relatively high

362   abundance in the tissue. In addition, despite a correlation with disease severity in the blood, NK and

363   MAIT cells did not co-locate with any structural cells. Further, even though macrophage subsets were

364   the most abundant cells in lungs, there was also no statistically significant co-location between these

365   cells and damaged structural cells.

366

367   All data, the spatial connectivity plot, radial connectivity map, and topographical correlation map

368   functions are available as open resources on MDV (https://mdv.molbiol.ox.ac.uk/, Supplementary Fig.

369   10, Methods).

370

371   **Immature neutrophils have a spatial predilection for CD8 T cells**

372

373   We next examined how immune cells connected to other immune cells by interrogating the 91 pairs

374   of immune cells with g(r=20)>1 across the three histopathology states (Fig. 6A-C).

375

376   We observed that as single entities (as opposed to those found within ADJ clusters), immature

377   neutrophils only co-located with CD8 T cells or CD8-ADJ clusters (Fig. 6A), regardless of histopathology

378   state. However, immature neutrophils within the Neut_CD8_ADJ cluster, co-localised with

379   Mono_CD31_ADJ clusters in DAD and other monocyte subsets in OP (Fig. 6A and D).

380

381   Therefore, in DAD, proliferating alveolar epithelium not only co-located with Neut_CD8_ADJ, but also

382   with a further network of co-locating immune cell types linked to the Neut-CD8_ADJ cluster, forming

383   a super network of Neut_CD8_ADJ and Mono_CD31_ADJ clusters around the proliferating alveolar

384   epithelial cells. This can be seen in the ACN analysis (Fig. 6F) and an MCD image view of the cells in the

385   tissue (Fig. 6G).

386

387   In contrast to neutrophils, there was a less restricted repertoire of co-locating cell partners for

388   monocytes. Monocyte subsets and ADJ clusters were found co-located with NK, MAIT, CD4 and CD8 T

389   cell subsets (Fig. 6B-C). Notably, megakaryocytes were found uniquely associated with

390   Mono_CD31_ADJ in DAD (Fig. 6E).

391

392   Our analyses showed that there were distinct organisations amongst immune cells in COVID-19 lungs,

393   with specific predilection of immature neutrophil for CD8 T cells, and upon connection (as the

394   neutrophil_CD8_ADJ cluster), a further connection with Mono_CD31_ADJ cluster was formed,

395   resulting in a network of Neut_CD8_ADJ and Mono_CD31_ADJ, linked to proliferating alveolar

396   epithelium in diffuse alveolar damage. These were then linked to megakaryocytes via the latter cell

397   type's connection with Mono_CD31_ADJ cluster in DAD. Thus, a spatial network of immature

398   neutrophils, CD8 T cells, classical monocytes and megakaryocyte form a connected web of cells

399   juxtaposed against proliferating alveolar epithelial cells and alveolar capillaries in DAD.

400

401

402   **Projection of the circulating source of lung CD8 T cells, monocytes and immature neutrophils**

403

404   Finally, we returned to our COMBAT data[6] to explore if we can identify the circulating source of the

405   monocytes, CD8 T cells and neutrophils found in the lungs. Using *scmap*, a method which enables label

406   projection by calculating the similarity between cells profiled by two separate studies [33], we examined

407   the phenotypic similarity between monocytes and CD8 T cells in the lungs [this study (COSMIC)] and

408   blood (COMBAT study). For COMBAT, we used the CYTOF dataset from neutrophil-depleted whole

409   blood (Supplementary Fig. 3 in COMBAT)[6].

410

411   Both lung CD107a$^-$ CD8 and CD107a$^+$ CD8 matched to blood 'GZB^{neg} CD8 T cells' in COMBAT (Fig. 7A).

412   Lung IFN-γ+ CD4 T cells matched to COMBAT's 'activated CD4 T cells' subset (which contained CD27$^-$

413   and CD27$^+$ CD4 T cells). All monocyte subsets in the lung [including Mono_CD31_ADJ, Mono_PAI-

414   1_ADJ (but not Mono3)], and all macrophage subsets showed high Jaccard similarity index with HLA

415   DR^{hi} classical monocytes in the blood (Fig. 7B).

416

417   We next interrogated the markers for these two COMBAT cell types (GZB^{neg} CD8 T cells and HLA

418   DR^{hi} classical monocytes) (data found in Supplementary Data 3 in COMBAT). We observed that

419   compared to healthy and disease controls, GZB^{neg} CD8 T cells expressed markers of exhaustion and

420   were KLRG1$^+$ compared to other CD8 T cells. HLA DR^{hi} classical monocytes showed high expression of

421   CLA. Both GZB^{neg} CD8 T cells and HLA DR^{hi} classical monocytes were unique amongst CD8 T cell and

422   monocyte subsets in showing lower abundance in COVID-19 patients compared to healthy

423   volunteers[6], raising the possibility that these were the subsets that have trafficked to the lungs. This

424   is not unprecedented given previous findings in lungs which showed sparse antigen-specific T cells in

425   blood of severe influenza patients but 8 times higher in matched blood-lung samples [34] .

426

427   For neutrophil comparisons between lungs (COSMIC) and blood (COMBAT), we obtained stored whole

428   blood samples and stained these with a 42-marker CYTOF panel (Supplementary Table 5). 8

429   subclusters of neutrophils were evident from dimensionality reduction (UMAP) and unsupervised

430   clustering, and annotated according to maturity – from pro-neutrophil to mature neutrophils (Fig. 7C-

431   D). Compared to the lung neutrophils, 'immature neutrophil 2' in the blood (which expressed the

432   highest level of CD172a amongst the immature CD10$^-$ neutrophil subsets), most closely matched the

433   neutrophil subset in Neut_CD8_ADJ (Fig. 2F). Notably, the abundance of 'immature neutrophil 2'

434   correlated positively with severity of disease (Fig. 7E).

435

436   These findings showed that the lung CD8 T cell subsets matched most closely to a GZB^{neg} KLRG1$^+$ CD8

437   T cell subset in the blood, which also expressed a T cell exhaustion signature. This suggests that this

438   blood CD8 T cell subset is a likely source for the GZB$^+$ CD8 T cells found in the Neut_CD8_ADJ cluster;

439   and that within this cluster, CD8 T cells expressed GZB, possibly with exposure to IFN-β[35]. On the other

440   hand, blood CD172a^{hi} immature neutrophil subset is the likely source for the immature neutrophils in

441   the lungs, including that found in the Neut_CD8_ADJ cluster.

442

443

444   **Discussion**

445

446   In this paper, we deconvoluted a highly disordered immune and structural landscape to provide

447   accurate annotations and abundance metrics for the cellular landscape and then leveraged

448   mathematical techniques to describe co – location and cell contact-based network construction. Our

449   mathematical tools encompassed a range of spatial statistics and methods from network science;

450   some transposed from ecology [36-38]. The pipeline uncovered a hitherto undescribed physical

451   partnership between immature neutrophils and CD8 T cells in COVID-19 lungs linked to proliferating

452   alveolar epithelium in areas with diffuse alveolar damage. This further connected with classical

453   monocytes and megakaryocyte around endothelial cells, forming a super pro-inflammatory network

454   across the alveolar bed in DAD. The observations on neutrophils are especially significant since

455   relatively little is understood of the role of neutrophils in the lungs of patients with COVID-19 due to

456   poor detection with transcriptomic methods [17,39].

457

458 Our study did not elucidate how neutrophil-CD8 clustering might contribute to disease pathogenesis.

459 However, evidence from other diseases provide some insight. Neutrophils and CD8 T cells aggregation

460 in colorectal cancer and graft vs host disease have been shown to enhance T-cell receptor–triggered

461 activation of CD8$^+$ T cells [40] causing neutrophil-mediated tissue damage by the release of reactive

462 oxygen species [41]. Neutrophils can also act as antigen presenting cell which cross present antigen to

463 CD8 T cells, further enhancing activation [42,43]. CD8 T cells with a similar effector memory and GZB$^+$

464 profile as that found in the Neut_CD8_ADJ cluster have also been implicated in immunopathology of

465 COVID-19 in other organs. Imaging mass cytometry studies in COVID-19 brain tissue showed intriguing

466 spatial associations with microglia, which also sustained immune activation and neuroinflammation

467 [44].

468

469 The presence of viral antigen could be the trigger for these foci of immature neutrophils and CD8 T

470 cells, possibly initiated by recognition of viral antigen by CD8 T cells. However, we note abundant

471 Neut_CD8_ADJ cluster in the OP state (Fig. 3B) where there were no viral proteins or RNA. One

472 explanation is that these CD8 T cells were self-proliferating, as suggested by Liao's study using single

473 cell RNA sequencing of lung-lavaged cells in COVID-19 patients [45]. Supporting this, Neut_CD8_ADJ

474 cluster showed the highest Ki67 expression (Fig. 2F), with MCD imaging isolating this expression to

475 CD8 T cells (Fig. 2I). Organising pneumonia is not a natural sequela of all viral infection or alveolar

476 inflammation. Indeed, many patients who do well do not progress to consolidation on computed

477 tomographic (CT) scans. Thus, a potential deleterious effect of these foci of inflammation could be the

478 obliteration of regenerative potential in type II alveolar epithelial cells, the purported stem cells for

479 the alveolar unit [46], and development of organising pneumonia (OP).

480

481 Another cluster highlighted by our analyses was the Mono_CD31_ADJ cluster, which was spatially

482 associated with Neut_CD8_ADJ cluster, and with proliferating alveolar epithelial cells. Proliferating

483 alveolar epithelial cells are the nominal stem cells for the alveoli and key to replenishment of type 1

484 alveolar epithelial cells. Its health, and ability to function optimally, is a key requirement for repair of

485 infected and damaged alveoli.   A consequence could be that the production of type I IFN, [and other

486 monocyte-specific cytokines like IL-6 and TNF- α (as reviewed by, [47,48]], combined to impact on

487 regeneration of alveolar epithelium. It is also possible that type I IFN production from these monocytes

488 causes upregulation of ACE2, thereby sustaining viral entry and alveolar epithelial damage[49]. This

489 agrees with observation from transcriptomic studies of the lungs where type II alveolar epithelium

490 were found in an inflammation-associated intermediate state rather than progressing via normal

491 regeneration to type I alveolar epithelium [13,15,17].

492

493

494   The tight association between a large number of monocytes and endothelial cells in all histopathology

495   states could result in excess inflammation and also predispose to small vessel thrombosis, particularly

496   with further presence of megakaryocytes at the point of maximal injury (DAD) (Fig. 5B). Single cell

497   transcriptomic analyses in COVID lungs have demonstrated upregulation of endothelial-damage

498   markers, including VWF, ICAM1 and VCAM1, and transcriptional programs suggesting altered vessel

499   wall integrity and widespread activation of coagulation pathway associated genes in endothelial

500   cells.[13,16,50]. In addition, autopsy studies have shown high numbers of megakaryocytes and platelet

501   rich thrombi in the lungs with COVID-19 pneumonitis [51].

502

503   Beyond these key messages, other findings clarified the importance of immune cell numbers and

504   phenotype in blood of patients with severe COVID-19. There was no significant spatial co-location

505   between activated NK cells and MAIT cells with any structural cells although the numbers for MAIT

506   cells were increased, in keeping with blood levels. With the ability to identify single cells of CD4 and

507   CD8 T cells, and quantify their abundance per mm^2 of lungs, we also showed definitively that levels

508   of CD4 and CD8 T cells were high in lung samples in contrast to studies which inferred their depletion

509   from gene expression profiles [15]. Immature cycling monocytes, one of the most striking observations

510   in the blood of patients with severe compared to mild COVID-19 disease[6,8], were not found in lung

511   tissue. This suggests that immature monocytes are unlikely to be involved in tissue damage, and

512   unlike immature neutrophils, probably differentiated rapidly to mature monocytes and

513   macrophages.

514

515   Our findings refined our earlier work on a smaller subset of COVID-19 lungs (n=3) using targeted

516   transcriptomic analysis (GeoMx) in specified sections in the lungs linked to alveolar damage [52]. In that

517   work, we deconvoluted cells detected by gene expression profile using limited protein markers and

518   showed that CD8 T cells and macrophages with IFN-γ signature correlated with areas of lungs with

519   alveolar damage. Interestingly, areas of severe damage exhibited consistent expression of IFNG-

520   regulated chemokines such as *CXCL9/10/11* that may promote CXCR3-mediated chemotaxis or

521   retention of CD8 T effector lymphocytes. Further to the findings from this paper, we performed

522   additional analyses to determine if we can provide a transcriptomic view of the immature neutrophils

523   and CD8 T cell cluster. This strengthened but did not reveal further findings (described in

524   Supplementary Fig. 11).

525

526   Another earlier work in the same lung samples showed significant presence of neutrophil extracellular

527   traps (NETS) in the lung samples which correlated with areas of low CD8 T cell levels. Re-examining

528   the number of NETS per lung section, we observed widespread presence with no significant difference

529   between the three histopathology states (Supplementary Fig. 3C). As NETS production is a feature of

530   mature rather than immature neutrophils [20], one explanation is that there is a CD8-directed immature

531   neutrophil localisation to proliferating alveolar epithelium, which is separate from the relatively less

532   discriminate NETS expression by mature neutrophils.

533

534   The key limitation of our study is that it is an observation of association, albeit that there was clear

535   comparison between histopathology characterisations of alveolitis, damage and repair. Thus, it is not

536   possible to elucidate cause or effect. Further functional studies will strengthen the findings. Our

537   cohort was also small though this was counterbalanced by uniquely fresh samples from lungs, with

538   minimal effect of degradation due to the sampling methods at the point of death. Finally, our study

539   was led by specific questions. To that end, the antibody panels, and analyses were targeted to those

540   questions and cellular identities were constrained to that linked to the antibody panel.

541

542   We conclude that statistically rigorous analyses of spatial associations of immune and structural cells

543   in lungs of those with fatal COVID-19 identified an inflammatory nidus of immature neutrophils and

544   CD8 T cells with high immune activity and proliferating capabilities that were linked to alveolar

545   progenitor cells in areas with greatest alveolar damage. It establishes the importance of emergency

546   myelopoiesis in lung immune pathology, with potential roles for immature neutrophils and

547   megakaryocytes in alveolar damage, aberrant alveolar regeneration, and excess thrombogenesis. The

548   findings support the evaluation of therapeutics that target monocytes and immature neutrophils,

549   potentially earlier in disease to limit its impact on progression to widespread alveolar damage and

550   organising pneumonia. It also means that drugs that increase the longevity or survival of CD8 T cells

551   require further assessment given the potential contribution of CD8 T cells to lung damage.

552

553

554

555

556

557   **Methods**

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

559 All antibodies, their catalogue numbers, final dilutions, and source are documented in Supplementary

560 Data 1.

561

562 **Patients, samples, and ethical approvals**

563

564 Lung samples were obtained from collaborators from the University of Navarra, Spain and comprised

565 those patients who died in hospital after admission with COVID-19. The only inclusion criteria were

566 (i) hospitalisation, (ii) evidence for COVID-19 pneumonitis, defined as presence of ground glass

567 changes +/- consolidation and peri bronchial shadowing in mid to peripheral distribution on thoracic

568 CT scan begore death, (iii) PCR+ results for nucleocapsid (N) and/or envelope protein I in lung or liver

569 tissue sample and (iv) negative bacterial culture from blood and lung within 3 days of death. The study

570 was approved by the Ethics Committee of the University of Navarra, Spain (Approval 2020.192). Tissue

571 collections were obtained with consent from a first-degree relative, following a protocol approved by

572 the ethics committee of the University of Navarra (Protocol 2020.192p); and stored under Spain's

573 Human Tissue Authority regulations. Samples were collected during the first wave of pandemic (2020)

574 via an intercostal space incision, using core biopsy methods (BioPince Full Core Biopsy Instrument kit)

575 immediately after death [20,53]. Tissues were immediately fixed in neutral buffered formalin for over 24

576 hours, and then paraffin-embedded. These samples were also shared with other collaborators and

577 studies carried out independently [20,52].

578

579 Healthy lung controls were obtained from the Oxford Centre for Histopathology Research and the

580 Oxford Radcliffe Biobank based at the Oxford University NHS Hospitals Foundation Trust. Ethics

581 approval was received from Oxford A South-Central NHS REC (ref 19/SC/0173). The inclusion criteria

582 were that lung sections had to be obtained away from localised lung cancer site on lung imaging; they

583 had to have normal lung histopathology as agreed by two independent histopathologists, aged

584 between 50-90y and had no concomitant lung diseases. Altogether 8 such patients were identified,

585 their lung sections stained with H & E and two representative patients selected to proceed to IMC

586 staining. H & E stained sections are shown in Supplementary Fig. 4.

587

588 We have considered sex balance in selection of samples. There are 5 females and 7 males in our

589 cohort. Patients and relatives were not financially compensated.

590

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

592

593   RNA extraction from biopsies was performed using the QIAamp Viral RNA Mini Kit (Qiagen) and the

594   identification of SARS-CoV-2 transcripts encoding nucleocapsid (N) and an envelope protein I was

595   performed using a commercial kit (SARS-CoV-2 Real Time PCR Kit, Vircell), both according to

596   manufacturer recommendations, at the Microbiology Laboratory of the Clinica Universidad de

597   Navarra (ref). Samples with amplification of both targets with Ct values below 35 were considered

598   positive for SARS-COV-2. Ct threshold was selected based on comparison between Ct values and

599   presence of viral DNA on nasopharyngeal-swab standards.

600

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

602

603   Slides were deparaffinised and heat-induced epitope retrieval were performed on the Leica BOND-

604   RXm using BOND Epitope Retrieval Solution 2 (ER2, pH 9.0) for 30 minutes at 95°C. Staining was

605   conducted with the Bond Polymer Refine Detection kit, a rabbit anti-SARS-CoV-2 nucleocapsid

606   antibody (Sinobiological; clone: #001; dilution: 1:5000) and counterstained with haematoxylin.

607

608   **Region of interest (ROI) selection**

609

610   H&E stained sections were examined by two senior pathologists independently and a pulmonologist

611   and data compiled with consensus at the third iteration. ROIs were selected based on size (2x 2mm

612   squares or equivalent surface areas) to represent the dominant histopathology findings for the

613   section. Slides were imaged on AxioScan Z1 slide scanner [Zeiss] and viewed using QuPath [54]

614

615   **Imaging mass cytometry (IMC) staining**

616

617   Sequential 6µm thick FFPE lung tissue section slides were incubated for 2 hours at 60°C on a slide

618   warmer, dewaxed twice in Histo-clear II (National Diagnostics) for 10 minutes before rehydration

619   through serial alcohols; 100%, 100%, 95%, 70% ethanol and MilliQ water. Slides were then incubated

620   for 30 minutes at 96°C in EDTA Target Retrieval Solution, pH 9 (Agilent) and cooled to 70°C before

621   washing twice in MilliQ water. Slides were blocked in 3% BSA solution in Maxpar PBS (Standard

622   BioTools; previously Fluidigm) for 45min. Sections were then stained with metal-conjugated

623   antibodies in Maxpar PBS containing 0.5% BSA overnight. Antibodies conjugated in house were

624   conjugated with MaxPar X8 antibody labelling kits (Standard BioTools) or Lightning-Link kits (Abcam)

625   according to manufacturer's instructions. Slides were washed in 0.2% Triton X-100 then twice in

626   Maxpar PBS. Intercalator-Ir (Standard BioTools) diluted in Maxpar PBS was used to stain DNA (30min),

627   slides were washed in MilliQ water then air dried.

628

629   Ablation of the relevant regions of interest (ROIs) was carried out on Standard BioTools Hyperion

630   Imaging System using CyTOF7 Software v7.0 (Standard BioTools) and visualized using MCD Viewer

631   (Standard BioTools). Images were processed for publication using FIJI [55] to de-speckle and sharpen

632   the images.

633

634   **Antibody validation and optimization**

635

636   Antibody clones were selected which had previously been published and validated in IMC studies as

637   well as antibodies frequently utilized for immunofluorescence or immunohistochemistry studies with

638   FFPE tissues. Staining validation for IMC markers was performed in healthy control lung and tonsil as

639   well as in some COVID-19 infected lung (Supplementary Fig. 5,6 and 12). During optimisation, we

640   checked that (i) mutually exclusive expression pattern were found in key immune and structural

641   lineage markers i.e. CD68, Epcam, CD3 and CD19 (ii) markers showed appropriate subcellular location

642   expression i.e. transcription factors Foxp3 and Ki67 were nuclear, whereas CD68 expression was

643   cytoplasmic and cell membrane. (iii) structural cell identities defined by IMC lineage marker expression

644   are compatible with cell morphology and location in H&E. Adjacent H&E-stained slides and structural

645   markers expression was examined e.g. α-SMA expression around vessels and bronchi, EpCAM

646   expression on bronchial and alveolar epithelial cells. (iv) Non-biological sense expression e.g. CD4 and

647   CD8 co-expression and biologically expected and coherent co-expression patterns eg. cells expressing

648   CD45, CD3, CD8 and CD45RO were examined (v.) Expression for the following key markers was

649   validated by immunofluorescence staining in adjacent slides – CD4, CD8, CD14, CD15, CD31, CD172a,

650   CD206, ProSPC, PAI-1, Epcam and Ki67.

651

652   Antibody clones that did not perform well i.e. those with weak signal, high background, or nonspecific

653   staining were discarded. Antibody titration was performed to maximise signal to noise ratio in both

654   lung and tonsil tissues and panels were designed to minimise the already low levels of signal spill over

655   see in IMC [less than 1-5%] [56].

656

657

658

659

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 Fischer 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 [55].

Targeted transcriptomic analysis of specific areas of interest with matched IMC staining and analyses

We extracted the RNA sequence data from AOIs (n=46) in three COVID lung sections as described in our previous paper (Cross, A.R. et al 2022) and organised these into enhanced histopathology

695 classification as described in this paper – ALV, DAD and OP. We then compiled the differential

696 expressed gene list between the three states (using DESeq2) and performed a pathway analysis using

697 Reactome [57] (Supplementary Fig. 11). Here, we found upregulation of genes associated with

698 neutrophil activation when comparing DAD to OP and ALV as observed in Cross A.R. et al. In particular,

699 S100A8 (highly expressed in neutrophils and a feature of degranulation) and CXCL10 (chemokine

700 related to neutrophils trafficking) were highly upregulated, supporting trafficking of neutrophil to the

701 tissue at the DAD phase[58]. High expression of CXCL9 a key chemokine in T cell extravasation into tissue

702 supports finding of T cells (e.g. CD8 T cells) in these AOIs.

703
704
705 **Data analysis**

706 **Software and algorithms**

707 All software and algorithms used are documented in Table 1.

708

709 **The Spatial Omics Oxford (SpOOx) Analysis Pipeline**

710
711 The SpOOx pipeline is a computational framework that brings together the methods we have used to

712 derive final spatial interpretation for the COVID-19 lung sections. It incorporates a suite of Python and

713 R based command line tools which may be run individually or as a semi-automated pipeline. We have

714 implemented SpOOx using the Ruffus framework [59]. Ruffus allows encapsulation of the workflow and

715 parameters to enable reproducibility, transparency and code reuse. All steps discussed in the Methods

716 are encapsulated in the SpOOx pipeline and example commands to achieve the step are shown below.

717 An overview of the pipeline can be found in Fig. 1a and Supplementary Fig. 1. Full detailed

718 documentation and a tutorial are included on the SpOOx GitHub page (https://github.com/Taylor-

719 CCB-Group/SpOOx). SpOOx produces a series of output directories and files that may be uploaded to

720 the Multi-Dimensional Viewer (MDV) software (see below). MDV has been developed based on the

721 Multi Locus View [60] framework and has been heavily modified and extended to allow visualisation and

722 analysis of large multidimensional data sets, images and the resulting spatial statistics. The code to

723 upload data to MDV is available on GitHub at https://github.com/Taylor-CCB-Group/MDV . Both the

724 SpOOx and MDV are open source under the GPL 3.0 license with these links – SpOOx is available for

725 install at https://github.com/Taylor-CCB-Group/SpOOx and MDV at https://github.com/Taylor-CCB-

726 Group/MDV.   The   project   data   analysis   is   available   online   within   MDV   at

727 https://mdv.molbiol.ox.ac.uk/projects/hyperion/6567 .

728

729   ***Conversion of MCD files to TIFF.*** MCD files were checked for problems with ablation or staining

730   using the MCD viewer (provided by Standard BioTools). Once these initial checks were

731   completed, the images were converted to OME-TIFF format for segmentation.

732

733   *Commands: python hyperion_pipeline.py make mcd_to_tiff and python hyperion_pipeline.py make*

734   *tiff_to_histocat*

735

736   ***Segmentation and cell mask generation*** Cell segmentation was performed with the Mesmer library

737   in DeepCell [61], Nuclear markers (DNA1 and DNA3) and cytoplasmic markers (a-SMA, CCR2, CCR6,

738   CD107a, CD10, CD114, CD115, CD14, CD15, CD16, CD172a, CD31, CD3, CD45, CD45RO, CD4, CD71,

739   CD8a, Collagen1, DAP12, EpCAM, GZB, HLA DR, IFN-β, IFN-γ, PAI1, PanCK, PF4 and RAGE) were

740   extracted to TIFF files and Z projected to single channel nuclear and cytoplasmic single TIFF images (

741   Supplementary Fig. 13). These images were contrast adjusted (--contrast 5) and passed to the Mesmer

742   library (pixel size adjusted to 1 micron) as nuclear and cytoplasm channels. From these, cell

743   segmentation masks were generated for each ROI.

744

745   *Command: python hyperion_pipeline.py make deepcell*

746

747

748   **Extraction of signal intensities for each cell.** The intensity of each marker within each labelled cell was

749   extracted from the data using the segmentation masks using the mean arcsinh-transformed (with –

750   cofactor 5) pixel intensity for each. The data were recorded as a table, each row representing a cell

751   with a unique id for the ROI). Shape features such as area, perimeter, eccentricity, and centroid were

752   also extracted from the masks. All cells were then filtered using a cell area greater than 50 μm and

753   less than 300 μm to exclude poorly segmented cells and cell debris. Further QC was performed within

754   MDV by plotting the distribution of marker intensity across each ROI.

755

756   *Command: python hyperion_pipeline.py make signal_extraction*

757

758   **Dimensionality reduction and cluster analysis.** For all downstream analysis the intensity values were

759   arcsinh transformed with a cofactor 5. Clustering was performed using the Phenograph algorithm [62]

760   through the implementation of the Rphenograph R package (version 0.99.1) with parameter k=30.

761   Using MDV, the clusters were first visualised using interactive UMAP scatter plots and heatmaps

762   (showing the median marker intensities per cluster) then manually annotated to define the cell

763   phenotypes at the cell level. The clustering was performed at two levels: a sample level (on the

764   trimmed [q=0.001] and scaled values) and per condition after having integrated the data with

765   Harmony (version 1.0) [63], using the default parameters with the option do_pca = TRUE. The integration

766   of the data was performed per condition to remove variation from different patients and to better

767   define common populations of cells. The annotations before and after integration were compared to

768   ensure that no biologically meaningful populations were missed when integrating the data. The

769   heatmaps, PCA and UMAP plots were done using the functions from the CATALYST R package (version

770   1.16.0).

771

772   *Command: python hyperion_pipeline.py make phenoharmonycluster.*

773

774   **Annotation workflow**. Cells were first examined for antibody staining and those cells that did not

775   show any antibody staining were filtered from further analysis. The remaining cells were grouped into

776   three mega-clusters termed Structural, Myeloid or Lymphocyte based on presence and/or absence of

777   CD45, EPCAM, PanCK, CD31, α–SMA, CD56, Vα7.2, CD3, CD14, CD68, PF4 and CD15 expression. The

778   three mega-clusters were then re-clustered using protein markers selected on immunological basis

779   (Supplementary Fig. 7). The resultant final clusters were annotated using an integrated approach. In

780   the first step, we defined clusters using (i) heatmaps showing median marker expression (ii) expression

781   density histograms which allow better delineation of the range of marker expression, specifically

782   differentiating low and negative expression levels and (iii) cluster distribution plots which showed the

783   frequency of each cluster in different samples. Phenotypic similarity of clusters was interrogated via

784   UMAP and cluster dendrograms. To further define cluster identities, the spatial location of clusters

785   was visualised using cell centroid plots and mapped onto an adjacent H&E slide with the same ROI.

786   Based on these analyses, some clusters were excluded under the following criteria: a) clusters with

787   uniformly low/negative expression of markers, b) clusters only found in one sample, and c) Undefined

788   clusters (where the combination of markers did not amount to a subset which could be defined). These

789   clusters were not submitted for spatial analysis. Sub-clusters with very similar expression profiles were

790   merged and those which contained 2 or more clusters were annotated as such. A small number of

791   clusters demonstrated expression of markers normally associated with disparate cell populations (e.g.

792   Neutrophil_CD8 adjacent), which can be attributed to closely apposed cell types. These adjacent cell

793   populations were validated via high resolution immunofluorescence microscopy. To aid final

794   annotation, we also performed Pseudotime inference for selected populations.

795

796   Final annotated clusters were then sense-checked against the MCD images by an independent

797   investigator not involved in annotating the clusters, and some key clusters of interest were further

798   examined by immunofluorescence staining with confocal microscopy.

799

800   **Pseudotime analysis.** The Pseudotime analysis was performed on the macrophage, monocytes and

801   neutrophils populations (Supplementary Fig. 6). Their arcsinh transformed values were integrated

802   using Harmony with the same parameters as in the main analysis, followed by dimensionality

803   reduction using UMAP. Then the Pseudotime inference was performed by applying the Slingshot

804   algorithm [64] to the UMAP dimensions using the default parameters and the above annotations as

805   clusterLabels.

806

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

808   *Command: R slingshot.R <parameters>*

809

810   **Differential cell abundance analysis**. Differential abundance analysis between conditions was

811   performed using code from the diffcyt R package (version 1.8.8) with the option testDA_edgeR. To

812   account for the differences in area between the ablated samples, the area was used as a normalising

813   factor. The dispersion was estimated using the option trend.method="none" and the negative

814   binomial generalized log-linear model was used for the analysis (with the glmFit and glmLRT

815   functions). The BH (Benjamini-Hochberg) method was used to adjust p-values for multiple testing.

816

817   **Cell centroid maps.** For each ROI, the cell centroids were plotted and coloured according to cell type

818   to produce a cell centroid map which forms the basis of subsequent analyses. These were overlaid

819   with ROI images in MDV so cell types may be located by colour.

820

821   **Spatial analyses**

822

823   A suite of mathematical tools for spatial analyses is incorporated in SpOOx (see below under QCM,

824   cross-PCF and ACN). The following command runs all the spatial analysis methods in SpOOx:

825

826   *Command: python hyperion_pipeline.py make spatialstats*

827

828   It is also possible to run each spatial function separately and to adjust parameters (see

829   https://github.com/Taylor-CCB-Group/SpOOx/tree/main/src/spatialstats for details). The command

830   line option that can be appended to the basic command above is stated after each method is

831   described.

832

833   **Quadrat Correlation Matrix (QCM).** The "Quadrat Correlation Matrix" (QCM) describes correlations

834   between counts of different cell types within square quadrats with edge length 100μm (resulting in

835   between 100 and 400 quadrats per ROI), following an approach used by [38] to identify statistically

836   significant co-occurrences ($p < 0.05$) and applied to multiplex images of cancer by [65].

837   We construct the QCM by first generating a matrix O whose entries O_{ij} record the number of cells of

838   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

839   m is the number of quadrats. We use O to generate 1000 matrices $N^1, \dots, N^{1000}$ which form a

840   distribution of "observations" in which the number of cells of each type and the number of cells in

841   each quadrat are the same as in O, but spatial correlations between cell types are removed by shuffling

842   cell labels. Each matrix N^k is such that, for each j:

$$\sum_i N_{ij}^k = \sum_i O_{ij}, \tag{1}$$

843   and for each i:

$$\sum_j N_{ij}^k = \sum_j O_{ij}, \tag{2}$$

844   We construct each matrix N^k as follows. We fix $N^{k,0} = O$, and define rules which permute the entries

845   of $N^{k,s}$ to obtain a new matrix $N^{k,s+1}$. This is accomplished by selecting two rows *(a,b)* and two

846   columns *(c,d)* of $N^{k,s}$ at random. For some integer p sampled uniformly at random from the interval

847     $[0, \min(N_{bc}^{k,s}, N_{ad}^{k,s})]$, we then fix:

$$N_{ac}^{k,s+1} = N_{ac}^{k,s} + p, \tag{3}$$

$$N_{bc}^{k,s+1} = N_{bc}^{k,s} - p \tag{4}$$

$$N_{bd}^{k,s+1} = N_{bd}^{k,s} + p \tag{5}$$

848   and

$$N_{ad}^{k,s+1} = N_{ad}^{k,s} - p. \tag{6}$$

849   This process is repeated for $s = 0, 1, \dots 10{,}000$ to ensure that the final matrix $N^k = N^{k,10000}$ is well

850   shuffled.

851

852   Partial correlation matrices $\boldsymbol{C_O}$ and $\boldsymbol{C_{N^1}} \ldots \boldsymbol{C_{N^{1000}}}$ are then calculated for $\boldsymbol{O}$ and $\boldsymbol{N^1}, \ldots, \boldsymbol{N^{1000}}$

853   respectively. Standard effect sizes (SES) are determined by rescaling the partial correlations in $\boldsymbol{C_O}$ by

854   the element-wise mean μ and standard deviation σ of the $\boldsymbol{C_{N^k}}$, such that

$$SES_{ij} = (\boldsymbol{C_{O_{ij}}} - \mu\left[\boldsymbol{C_{N^k}}\right]_{ij}) \, / \, (\sigma\left[\boldsymbol{C_{N^k}}\right]_{ij}). \tag{7}$$

855

856   Non-significant associations are identified by calculating a 2-tailed p-value for each pair of cell types

857   and applying a Benjamini-Hochberg correction, with false discovery rate FDR = 0.05. Non-significant

858   entries of SES are set to 0 in order to generate the QCM, a cell association matrix whose non-zero

859   entries identify standardised effect sizes of pairs of cell types that are statistically significantly

860   correlated within the ROI.

861

862   The average QCM across Q ROIs is obtained by concatenating the relevant observation matrices.

863   Denoting by $\boldsymbol{O}_q$ the observation matrix from *ROI q*, we concatenate $\boldsymbol{O}_1, \ldots, \boldsymbol{O}_Q$ to form a combined

864   observation matrix $\boldsymbol{O} = (\boldsymbol{O}_1 \quad \boldsymbol{O}_2 \quad \ldots \quad \boldsymbol{O}_Q)$, an ($n$ x ($m_1 + m_2 + \ldots + m_Q$)) matrix, where m_q denotes the

865   number of quadrats in *ROI q*. Similarly, we concatenate $\boldsymbol{N}_1^k, \ldots, \boldsymbol{N}_Q^k$ to form $\boldsymbol{N}^k = (\boldsymbol{N}_1^k \quad \boldsymbol{N}_2^k \quad \ldots \boldsymbol{N}_Q^k)$.

866   Standard partial correlation matrices are then calculated and then the process described above for a

867   single ROI is used to compute the average QCM for multiple ROIs.

868

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

870

871   **Cross pair correlation functions (cross-PCF).** Significant correlations identified at length scales in the

872   range 0-100µm via the QCM are further assessed by using cross pair correlation functions (cross-PCFs

873   – see, e.g., Bull 2020). Cross-PCFs quantify clustering and dispersal of pairs of cell populations across

874   a range of length scales (here 0-300µm). The cross-PCF considers pairs of cells which are separated by

875   distances $r \in [r_k, r_{k+1})$, where $r_0 = 0$ and $r_k = r_{k-1} + 10$ for $k = 1, \ldots, 30$.

876

877   For cell populations A and B, the cross-PCF, $g(r_k)$, is defined as follows:

$$g(r_k) = \frac{1}{N_A} \sum_{a=1}^{N_A} \sum_{b=1}^{N_B} \frac{I_{[r_k, r_{k+1})}(|x_a - x_b|)}{\rho_B A_{r_k}(x_a)}, \tag{8}$$

878   where N_A and N_B are the numbers of cells of types A and B, $A_{r_k}(\boldsymbol{x})$ is the area of that portion of an

879   annulus centred at $\boldsymbol{x} = (x, y)$ with inner radius r_k and outer radius r_{k+1} which falls within the ROI, $\boldsymbol{x}_a$

880   and $\boldsymbol{x}_b$ are the spatial coordinates of cells *a* and *b* (of types A and B respectively), $I_{[r_k, r_{k+1})}(r)$ is an

881 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 ρ_B is the

882 density of cells of type B in the ROI.

883

884 A cross-PCF with $g(r) > 1$ means that cells of type A are observed more frequently at distance r from

885 cells of type B than would be expected under complete spatial randomness (CSR), and is indicative of

886 clustering at distance r. Conversely, a cross-PCF with $g(r) < 1$ means that cells of type A are observed

887 less frequently at distance r from cells of type B than would be expected under CSR, and is indicative

888 of exclusion.

889

890 For individual ROIs, 95% confidence intervals are obtained via bootstrapping. The spatial dependence

891 of resampled points is accounted for by resampling grid sites within a 20μm square lattice, following

892 [66].

893

894 To aid comparison between the clustering and dispersal of different pairs of cell populations, we

895 frequently report cross-PCF values at $r_k = 20$, corresponding to length scales in the range $r \in [20,30)$

896 μm. We focus on $r_k = 20$ since it approximates the distance between the centroids of cells which are

897 in physical contact. For notational simplicity, we denote this value as $g(r = 20)$.

898

899 *Command option: --function paircorrelationfunction*

900

901 **Topographical Correlation Map.** The cross-PCF quantifies clustering and dispersal of pairs of cell

902 populations at different length scales within an ROI. We also introduce the Topographical Correlation

903 Map (TCM), to visualise how the spatial correlation between cells of types A and B, say, changes across

904 an ROI.

905

906 In order to define Γ_{ab}, the TCM for cells of types A and B, we first associate a mark m_{ab} with each cell

907 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

908 of cell a, to the expected number of cells of type B if they were distributed according to CSR:

$$m_{ab} = \sum_{j=1}^{N_B} \frac{I_{[0,100)}(|x_a - x_j|)}{\rho_B A_{100}(x_a)}, \tag{9}$$

909 where ρ_B is the density of cells of type B in the ROI, $A_{100}(x_a)$ is the area of that portion of a circle

910 with radius 100μm centred at $x_a = (x_a, y_a)$ which falls within the ROI, $I_{[0,100)}(r)$ is an indicator

911 function ($I_{[0,100)}(r) = 1$ when $0 \le r < 100$ and $I_{[0,100)}(r) = 0$ otherwise), and N_B is the total

912 number of cells of type B within the ROI. We interpret values of m_{ab} in a manner similar to that used

913   for cross-PCFs: $m_{ab} < 1$ indicates anti-correlation between cells of types A and B within a distance of

914   100μm, and $m_{ab} > 1$ indicates correlation.

915

916   To facilitate visualization and interpretation, we normalize the mark m_{ab} by introducing the

917   transformed mark, M_{ab}, where:

$$M_{ab}(m_{ab}) = 1 \qquad \text{if } m_{ab} \geq \alpha, \tag{10}$$

$$M_{ab}(m_{ab}) = \frac{m_{ab} - 1}{\alpha - 1} \qquad \text{if } 1 < m_{ab} \leq \alpha, \tag{11}$$

$$M_{ab}(m_{ab}) = \frac{1 - \frac{1}{m_{ab}}}{\alpha - 1} \qquad \text{if } \frac{1}{\alpha} < m_{ab} < 1, \tag{12}$$

$$M_{ab}(m_{ab}) = -1 \qquad \text{if } m_{ab} \leq 1/\alpha. \tag{13}$$

918   The constant α defines a threshold for extreme clustering. If $m_{ab} > \alpha$ then we have strong clustering

919   and we fix $M_{ab} = 1$; if $m_{ab} \leq 1/\alpha$ then we have strong exclusion and we fix $M_{ab} = -1$.

920   A sketch of M_{ab} is presented in Supplementary Fig. 14.

921

922   We note the following properties of the transformed mark, M_{ab}. First, $M_{ab}(m_{ab}) = -M_{ab}(\frac{1}{m_{ab}})$, so

923   that dispersal and clustering are measured on the same scales. For example, $m_{ab} = 2$ indicates the

924   presence of twice as many cells of type B as expected under CSR, while $m_{ab} = 1/2$ indicates the

925   presence of half as many cells of type B as expected under CSR. Secondly, the magnitude of M_{ab}

926   describes the strength of the spatial interaction. Finally, the sign of M_{ab} identifies whether there is

927   clustering ($M_{ab} > 0$) or exclusion ($M_{ab} < 0$) between cell a (of type A) and cells of type B.

928

929   The parameter α characterises the most extreme clustering or exclusion which can be resolved in each

930   kernel, with extremal values being mapped to 1 and -1 respectively. We use $\alpha = 5$, so clustering or

931   exclusion stronger than 5x is interpreted as the strongest clustering/exclusion that we can distinguish.

932

933   After calculating M_{ab} for each cell of type a across the ROI, we centre a Gaussian kernel, with standard

934   deviation $\sigma = 50\mu m$, and maximum height M_{ab}, at $\boldsymbol{x_a}$. We sum the kernels associated with all cells

935   of type A to generate the TCM, $\Gamma_{ab}(\boldsymbol{x})$:

$$\Gamma_{ab}(\boldsymbol{x}) = \sum_{a=1}^{N_A} \frac{M_{ab}}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}|x - x_a|^2}. \tag{14}$$

936

937   The TCM permits identification of spatial locations in which cells of type A are positively ($\Gamma_{ab} \gg 0$) or

938   negatively ($\Gamma_{ab} \ll 0$) associated with cells of type B. For computational efficiency, when calculating

939   Γ_{ab}, we assume that each kernel has compact support, being centred in a square region of edge length

940   300μm.

941

942   Finally, we note that $\Gamma_{ab} \neq \Gamma_{ba}$, since the kernels used to construct Γ_{ab} are centered on cells of type

943   A (and vice versa). While areas in which cells of type A and type B are co-located should be identified

944   by both Γ_{ab} and Γ_{ba}, their values will differ in regions rich in one cell type and poor in another. We

945   therefore stress that Γ_{ab} describes locations in which cells of type A are correlated or anti-correlated

946   with cells of type B, and that the presence or absence of cells of type B cannot be inferred from regions

947   in which Γ_{ab} is close to 0.

948

949   *Command option: --function localclusteringheatmaps*

950

951   **Adjacency Cell Networks**. We use the cell segmentation masks generated by DeepCell to produce a

952   spatially-embedded adjacency cell network (ACN), whose nodes represent cell centres and are

953   labelled according to their cell type. Nodes are connected by an edge if the corresponding cells in the

954   segmentation mask share a border. To ensure that small perturbations in cell boundaries do not lead

955   to errors in cell connections, we expand the border of each segmented cell by 5 pixels before

956   generating the network.

957

958   We use the ACN to define two statistics for each pairwise combination of cell types A and B. First, we

959   compute ϕ_{AB}, the proportion of cells of type A which are in contact with at least one cell of type B:

$$\phi_{AB} \ = \ \frac{1}{N_A} \sum_{a=1}^{N_A} I_B(a), \quad\quad\quad (15)$$

960   where N_A is the number of cells of type A and $I_B(a)$ is an indicator function ($I_B(a) = 1$ if cell a is

961   connected with a cell of type B and $I_B(a) = 0$ otherwise). Secondly, we calculate Φ_{AB}, the average

962   number of cells of type B that are in contact with a cell of type A:

$$\Phi_{AB} = \frac{1}{N_A} \sum_{a=1}^{N_A} \eta_B(a), \quad\quad\quad (16)$$

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

964

965   In this paper, we used the ACN to calculate ϕ_{AB}, the proportion of cells of type A that have at least

966   one cell of type B in contact with them, and Φ_{AB}, the average number of cells of type B that are in

967   contact with a cell of type A in the ROI.

968

969   *Command option: --function networkstatistics*

970

971   **Multi-Dimensional Viewer (MDV)**

972

973   MDV is a comprehensive spatial analytics platform that facilitates the interrogation of large complex

974   data sets and includes various interactive dashboards to facilitate quality control, interactive

975   clustering, phenotyping and spatial analysis. It is an open source web application which can be

976   downloaded and installed locally or used on the publicly available web site

977   http://mdv.molbiol.ox.ac.uk. Users register to use the site and projects can private, shared with other

978   users or made public. Full documentation and tutorial videos are provided on the MDV website but

979   we provide an overview here.

980

981   MDV allows output generated by the SpOOx pipeline to be loaded at different states. Data locations

982   are specified in a yaml format file which can be edited by the user (command: python mdvupload.py

983   myconfig.yaml). Examples of data tables that may be uploaded are:

984

985   ● Image data (PNGs/OME-TIFF stacks): ROI image stacks, H and E images binary cell

986   masks.

987   ● Cell data (tab separated file): one cell per row, including size, size, shape, phenograph

988   clusters identification, UMAP coordinates, marker signal intensities.

989   ● Spatial Statistics data (tab separated file): one row containing cell to cell interaction

990   data and associated statistics.

991   ● Data related to the disease states (JSON file): allowing grouping of samples for high

992   level analysis.

993

994   Once uploaded the data are presented in MDV as a series of views that contain multiple interactive

995   charts corresponding to different analytical methods from clustering, annotation, cell centroid

996   visualisation and spatial analytical methods. Each view focuses on a particular aspect of the pipeline.

997   View contents can be adjusted and added to by adding other chart types and saved as a new view.

998   Chart types can be D3 components (https://d3js.org/) but we have also written custom chart types

999   for performance reasons. For example, MDV scatterplot chart can visualise and interrogate at least 10

1000   million data points. We also integrate Viv viewer [67] to visualise composite image stacks.

1001

1002   The complete analysis and data set were published by publicly sharing the data at

1003   https://mdv.molbiol.ox.ac.uk/projects/hyperion/6356.

1004

1005   **COMBAT data mapping**

1006

1007   **COMBAT CyTOF data generation and processing**. Cell suspension mass cytometry (CyTOF) data were

1008   generated by the COMBAT consortium as previously described[6]. In brief, whole blood from COVID-19

1009   patients was stabilised using a Cytodelics fixative solution, red blood cells were lysed, cellular material

1010   was fixed, and samples were run in a Helios CyTOF machine. Importantly, samples were enriched for

1011   mononuclear cells before profiling by performing magnetic depletion of $CD66^+$ granulocytes.

1012

1013   After acquisition, data were formatted into a single-cell protein abundance table and annotated into

1014   cell types based on marker expression [6]. For the analyses in the present study, this expression matrix

1015   was split into two subsets: one containing T and NK cell types, and a second one containing myeloid

1016   cell types (i.e. monocyte subsets).

1017

1018   **Mapping cells from lung tissue to blood cells from the COMBAT study[6]**. Cells in the lung dataset were

1019   matched to the most closely related cell types in blood using scmap, a method which enables label

1020   projection by calculating the similarity between cells profiled by two separate studies [33]. In brief,

1021   CyTOF and CITE-seq expression matrices from the COMBAT study were used to build index references

1022   for label projection. First, proteins which were detected in both studies were identified. This resulted

1023   in a panel of 13 and 18 proteins shared between our study and the COMBAT CyTOF and CITE-seq

1024   panels, respectively. Next, these proteins were used as a basis for cell type classification with the

1025   scmapCluster function. Classification accuracy was tested by splitting the COMBAT data into training

1026   and test sets containing 80% and 20% of cells, respectively. The training set was used to generate the

1027   scmap reference index, while the test set was used to assess cell type prediction accuracy [33]. Given

1028   the reduced set of markers shared between studies, not all COMBAT cell populations could be

1029   accurately predicted. Thus, in order to maximise predictive accuracy similar subpopulations of the

1030   same cell type were merged into a single group and any cell types known to be absent from our lung

1031   data, such as B cells and plasmablasts, were removed. This approach achieved over 70% accuracy for

1032   CyTOF data (71% and 78% predictive accuracy for myeloid and lymphoid cell types, respectively) and

1033 85% accuracy for CITE-seq data and components of final merged clusters are shown in Supplementary

1034 Fig. 16.

1035

1036 Indexed references were next used to match cells in the lung to the most similar clusters in blood

1037 using the scmapCluster() function. To do so, cell type labels were predicted for each cell in the lung

1038 based on the CyTOF and CITE-seq reference sets. Any unassigned cells were discarded. Cluster overlap

1039 between studies was visualised using Sankey diagrams [68] and quantified using Jaccard indexes [69].

1040

1041 **Neutrophil subset analysis from stored COMBAT samples**. Whole blood samples frozen in whole

1042 blood cell stabilizer (Cytodelics) were obtained from COMBAT consortium storage for healthy

1043 volunteers (n=11), health care workers (n=12), COVID-19 (n=93) and Sepsis (n=48). Pre-processed

1044 $CD45^+$ gated FCS files of granulocyte containing whole blood samples were analysed with R (v4.0.0).

1045 50000 cells per sample were integrated using Harmony (v1.0)[2] and the CATALYST package

1046 (v1.14.0)[3] was used for downstream analysis. $CD45^+$ cells were clustered based on the FlowSOM and

1047 ConsensusClusterPLus algorithms using the cluster () function. 50 metaclusters (xdim=100, ydim=100,

1048 k=50) were then assigned to major cell types (T cells, B cells, plasmablasts, mononuclear phagocytes

1049 and neutrophils). Neutrophils were selected and reclustered based on CD45, CD15, CD38, CD64, CD16,

1050 CD43, CD66b CD10, CD33, KI67, CD172a/b, CD141, CD71, CD114, CD371 and CD274 expression. 30

1051 neutrophil metaclusters (xdim=100, ydim=100, k=30) were manually merged to 8 neutrophil clusters

1052 (proNeut, preNeut, iNeut1-3, mNeut1-3) based on median marker expression.

1053

1054 **Data availability**

1055 The spatial mass cytometry dataset (MCD) files and results of analysis by the Spatial Omics Oxford

1056 pipeline are available at https://doi.org/10.5281/zenodo.6513508. The analysis results are also

1057 presented as a dynamic online resource in Multi-Dimensional Viewer (MDV)

1058 (https://mdv.molbiol.ox.ac.uk/projects/hyperion/6567). All source data are found in

1059 https://doi.org/10.5281/zenodo.6513508; and also within the hyperion 6567 project in the MDV link.

1060 Specific source data for graphs are also provided in Source Data File. Source data are provided with

1061 this paper.

1062

1063 **Code availability**

1064 The complete code for the Spatial Omics Oxford pipeline is available as a GitHub repository under

1065 the GPL license: https://github.com/Taylor-CCB-Group/SpOOx. In addition, the SpOOx pipeline has

1066 been deposited at Zenodo (https://zenodo.org/record/8320986). The Multi-Dimensional Viewer

1067   code is available under the GPL license: https://github.com/Taylor-CCB-Group/MDV. This package

1068   has also been deposited at Zenodo (https://zenodo.org/record/8324918).

1069

References

1070

1071

1072   1   Wang et al. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis
1073   of COVID-19-related mortality, 2020-21. *Lancet* **399**, 1513-1536 (2022).
1074   https://doi.org:10.1016/S0140-6736(21)02796-3
1075   2   Gorog, D. A. *et al.* Current and novel biomarkers of thrombotic risk in COVID-19: a Consensus
1076   Statement from the International COVID-19 Thrombosis Biomarkers Colloquium. *Nat Rev*
1077   *Cardiol* **19**, 475-495 (2022). https://doi.org:10.1038/s41569-021-00665-7
1078   3   Lamers, M. M. & Haagmans, B. L. SARS-CoV-2 pathogenesis. *Nat Rev Microbiol* **20**, 270-284
1079   (2022). https://doi.org:10.1038/s41579-022-00713-0
1080   4   Nalbandian, A. *et al.* Post-acute COVID-19 syndrome. *Nat Med* **27**, 601-615 (2021).
1081   https://doi.org:10.1038/s41591-021-01283-z
1082   5   Pustake, M., Tambolkar, I., Giri, P. & Gandhi, C. SARS, MERS and CoVID-19: An overview and
1083   comparison of clinical, laboratory and radiological features. *J Family Med Prim Care* **11**, 10-17
1084   (2022). https://doi.org:10.4103/jfmpc.jfmpc_839_21
1085   6   COMBAT. A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. *Cell*
1086   **185**, 916-938 e958 (2022). https://doi.org:10.1016/j.cell.2022.01.012
1087   7   Kusnadi, A. *et al.* Severely ill COVID-19 patients display impaired exhaustion features in SARS-
1088   CoV-2-reactive CD8(+) T cells. *Sci Immunol* **6** (2021).
1089   https://doi.org:10.1126/sciimmunol.abe4782
1090   8   Mann, E. R. *et al.* Longitudinal immune profiling reveals key myeloid signatures associated
1091   with COVID-19. *Sci Immunol* **5** (2020). https://doi.org:10.1126/sciimmunol.abd6197
1092   9   Schulte-Schrepping, J. *et al.* Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell
1093   Compartment. *Cell* **182**, 1419-1440 e1423 (2020). https://doi.org:10.1016/j.cell.2020.08.001
1094   10   Unterman, A. *et al.* Single-cell multi-omics reveals dyssynchrony of the innate and adaptive
1095   immune system in progressive COVID-19. *Nature Communications* **13**, 440 (2022).
1096   https://doi.org:10.1038/s41467-021-27716-4
1097   11   Wang, X. *et al.* Dysregulated hematopoiesis in bone marrow marks severe COVID-19. *Cell*
1098   *Discovery* **7**, 60 (2021). https://doi.org:10.1038/s41421-021-00296-9
1099   12   Chen, S. T. *et al.* A shift in lung macrophage composition is associated with COVID-19 severity
1100   and recovery. *Science Translational Medicine* **14**, eabn5168
1101   https://doi.org:10.1126/scitranslmed.abn5168
1102   13   Wang, S. *et al.* A single-cell transcriptomic landscape of the lungs of patients with COVID-19.
1103   *Nature Cell Biology* **23**, 1314-1328 (2021). https://doi.org:10.1038/s41556-021-00796-6
1104   14   Wendisch, D. *et al.* SARS-CoV-2 infection triggers profibrotic macrophage responses and lung
1105   fibrosis. *Cell* **184**, 6243-6261.e6227 (2021). https://doi.org:10.1016/j.cell.2021.11.033
1106   15   Melms, J. C. *et al.* A molecular single-cell lung atlas of lethal COVID-19. *Nature* **595**, 114-119
1107   (2021). https://doi.org:10.1038/s41586-021-03569-1
1108   16   de Rooij, L. *et al.* The pulmonary vasculature in lethal COVID-19 and idiopathic pulmonary
1109   fibrosis at single cell resolution. *Cardiovascular research* (2022).
1110   https://doi.org:10.1093/cvr/cvac139
1111   17   Delorey, T. M. *et al.* COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets.
1112   *Nature* **595**, 107-113 (2021). https://doi.org:10.1038/s41586-021-03570-8
1113   18   Bharat, A. *et al.* Lung transplantation for patients with severe COVID-19. *Sci Transl Med* **12**
1114   (2020). https://doi.org:10.1126/scitranslmed.abe4282
1115   19   Rendeiro, A. F. *et al.* The spatial landscape of lung pathology during COVID-19 progression.
1116   *Nature* **593**, 564-569 (2021). https://doi.org:10.1038/s41586-021-03475-6
1117   20   Melero, I. *et al.* Neutrophil Extracellular Traps, Local IL-8 Expression, and Cytotoxic T-
1118   Lymphocyte Response in the Lungs of Patients With Fatal COVID-19. *Chest* **162**, 1006-1016
1119   (2022). https://doi.org:10.1016/j.chest.2022.06.007

1120   21   Carsana, L. *et al.* Pulmonary post-mortem findings in a series of COVID-19 cases from northern
1121   Italy: a two-centre descriptive study. *The Lancet. Infectious diseases* **20**, 1135-1140 (2020).
1122   https://doi.org:10.1016/s1473-3099(20)30434-5

1123   22   Youd, E. & Moore, L. COVID-19 autopsy in people who died in community settings: the first
1124   series. *Journal of clinical pathology* **73**, 840-844 (2020). https://doi.org:10.1136/jclinpath-
1125   2020-206710

1126   23   Menter, T. *et al.* Postmortem examination of COVID-19 patients reveals diffuse alveolar
1127   damage with severe capillary congestion and variegated findings in lungs and other organs
1128   suggesting vascular dysfunction. *Histopathology* **77**, 198-209 (2020).
1129   https://doi.org:10.1111/his.14134

1130   24   Xu, Z. *et al.* Pathological findings of COVID-19 associated with acute respiratory distress
1131   syndrome. *The Lancet. Respiratory medicine* **8**, 420-422 (2020).
1132   https://doi.org:10.1016/s2213-2600(20)30076-x

1133   25   Mauad, T. *et al.* Tracking the time course of pathological patterns of lung injury in severe
1134   COVID-19. *Respiratory research* **22**, 32 (2021). https://doi.org:10.1186/s12931-021-01628-9

1135   26   Angeles Montero-Fernandez, M. & Pardo-Garcia, R. Histopathology features of the lung in
1136   COVID-19 patients. *Diagnostic histopathology (Oxford, England)* **27**, 123-127 (2021).
1137   https://doi.org:10.1016/j.mpdhp.2020.11.009

1138   27   Borczuk, A. C. *et al.* COVID-19 pulmonary pathology: a multi-institutional autopsy cohort from
1139   Italy and New York City. *Mod Pathol* **33**, 2156-2168 (2020). https://doi.org:10.1038/s41379-
1140   020-00661-1

1141   28   Vadasz, I. *et al.* Severe organising pneumonia following COVID-19. *Thorax* **76**, 201-204 (2021).
1142   https://doi.org:10.1136/thoraxjnl-2020-216088

1143   29   Greenwald, N. F. *et al.* Whole-cell segmentation of tissue images with human-level
1144   performance using large-scale data annotation and deep learning. *Nat Biotechnol* **40**, 555-565
1145   (2022). https://doi.org:10.1038/s41587-021-01094-0

1146   30   Barkauskas, C. E. *et al.* Type 2 alveolar cells are stem cells in adult lung. *The Journal of clinical
1147   investigation* **123**, 3025-3036 (2013). https://doi.org:10.1172/jci68782

1148   31   Zacharias, W. J. *et al.* Regeneration of the lung alveolus by an evolutionarily conserved
1149   epithelial progenitor. *Nature* **555**, 251-255 (2018). https://doi.org:10.1038/nature25786

1150   32   Choi, J. *et al.* Inflammatory Signals Induce AT2 Cell-Derived Damage-Associated Transient
1151   Progenitors that Mediate Alveolar Regeneration. *Cell stem cell* **27**, 366-382.e367 (2020).
1152   https://doi.org:10.1016/j.stem.2020.06.020

1153   33   Kiselev, V. Y., Yiu, A. & Hemberg, M. scmap: projection of single-cell RNA-seq data across data
1154   sets. *Nat Methods* **15**, 359-362 (2018). https://doi.org:10.1038/nmeth.4644

1155   34   Zhao, Y. *et al.* High levels of virus-specific CD4+ T cells predict severe pandemic influenza A
1156   virus infection. *Am J Respir Crit Care Med* **186**, 1292-1297 (2012).
1157   https://doi.org:10.1164/rccm.201207-1245OC

1158   35   Lanzavecchia, A. & Sallusto, F. Understanding the generation and function of memory T cell
1159   subsets. *Current opinion in immunology* **17**, 326-332 (2005).
1160   https://doi.org:10.1016/j.coi.2005.04.010

1161   36   Baddeley, A. *et al.* On tests of spatial pattern based on simulation envelopes. **84**, 477-489
1162   (2014). https://doi.org:https://doi.org/10.1890/13-2042.1

1163   37   Guseva, K. *et al.* From diversity to complexity: Microbial networks in soils. *Soil Biology and
1164   Biochemistry* **169**, 108604 (2022).
1165   https://doi.org:https://doi.org/10.1016/j.soilbio.2022.108604

1166   38   Morueta-Holme, N. *et al.* A network approach for inferring species associations from co-
1167   occurrence data. *Ecography* **39**, 1139-1150 (2016). https://doi.org:10.1111/ecog.01892

1168   39   Slyper, M. *et al.* Author Correction: A single-cell and single-nucleus RNA-Seq toolbox for fresh
1169   and frozen human tumors. *Nat Med* **26**, 1307 (2020). https://doi.org:10.1038/s41591-020-
1170   0976-3

40   Governa, V. *et al.* The Interplay Between Neutrophils and CD8(+) T Cells Improves Survival in Human Colorectal Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **23**, 3847-3858 (2017). https://doi.org:10.1158/1078-0432.Ccr-16-2047

41   Schwab, L. *et al.* Neutrophil granulocytes recruited upon translocation of intestinal bacteria enhance graft-versus-host disease via tissue damage. *Nat Med* **20**, 648-654 (2014). https://doi.org:10.1038/nm.3517

42   Takashima, A. & Yao, Y. Neutrophil plasticity: acquisition of phenotype and functionality of antigen-presenting cell. *Journal of leukocyte biology* **98**, 489-496 (2015). https://doi.org:10.1189/jlb.1MR1014-502R

43   Samadi, N. *et al.* T-cell-derived cytokines enhance the antigen-presenting capacity of human neutrophils. *European Journal of Immunology* **49**, 1441-1443 (2019). https://doi.org:https://doi.org/10.1002/eji.201848057

44   Schwabenland, M. *et al.* Deep spatial profiling of human COVID-19 brains reveals neuroinflammation with distinct microanatomical microglia-T-cell interactions. *Immunity* **54**, 1594-1610.e1511 (2021). https://doi.org:10.1016/j.immuni.2021.06.002

45   Liao, M. *et al.* Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med* **26**, 842-844 (2020). https://doi.org:10.1038/s41591-020-0901-9

46   Olajuyin, A. M., Zhang, X. & Ji, H. L. Alveolar type 2 progenitor cells for lung injury repair. *Cell Death Discov* **5**, 63 (2019). https://doi.org:10.1038/s41420-019-0147-9

47   Bert, S., Nadkarni, S. & Perretti, M. Neutrophil-T cell crosstalk and the control of the host inflammatory response. *Immunol Rev* (2022). https://doi.org:10.1111/imr.13162

48   Costa, S., Bevilacqua, D., Cassatella, M. A. & Scapini, P. Recent advances on the crosstalk between neutrophils and B or T lymphocytes. *Immunology* **156**, 23-32 (2019). https://doi.org:10.1111/imm.13005

49   Ziegler, C. G. K. *et al.* SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. *Cell* **181**, 1016-1035.e1019 (2020). https://doi.org:10.1016/j.cell.2020.04.035

50   Chua, R. L. *et al.* COVID-19 severity correlates with airway epithelium–immune cell interactions identified by single-cell analysis. *Nature Biotechnology* **38**, 970-979 (2020). https://doi.org:10.1038/s41587-020-0602-4

51   Rapkiewicz, A. V. *et al.* Megakaryocytes and platelet-fibrin thrombi characterize multi-organ thrombosis at autopsy in COVID-19: A case series. *EClinicalMedicine* **24**, 100434 (2020). https://doi.org:10.1016/j.eclinm.2020.100434

52   Cross, A. R. *et al.* Spatial transcriptomic characterization of COVID-19 pneumonitis identifies immune circuits related to tissue injury. *JCI Insight* (2022). https://doi.org:10.1172/jci.insight.157837

53   Recalde-Zamacona, B. *et al.* Histopathological findings in fatal COVID-19 severe acute respiratory syndrome: preliminary experience from a series of 10 Spanish patients. *Thorax* **75**, 1116-1118 (2020). https://doi.org:10.1136/thoraxjnl-2020-215577

54   Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878 (2017). https://doi.org:10.1038/s41598-017-17204-5

55   Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682 (2012). https://doi.org:10.1038/nmeth.2019

56   Chevrier, S. *et al.* Compensation of Signal Spillover in Suspension and Imaging Mass Cytometry. *Cell Syst* **6**, 612-620 e615 (2018). https://doi.org:10.1016/j.cels.2018.02.010

57   Griss, J. *et al.* ReactomeGSA - Efficient Multi-Omics Comparative Pathway Analysis. *Molecular & cellular proteomics : MCP* **19**, 2115-2125 (2020). https://doi.org:10.1074/mcp.TIR120.002155

1220   58   Ichikawa, A. *et al.* CXCL10-CXCR3 enhances the development of neutrophil-mediated
1221   fulminant lung injury of viral and nonviral origin. *Am J Respir Crit Care Med* **187**, 65-77 (2013).
1222   https://doi.org:10.1164/rccm.201203-0508OC

1223   59   Goodstadt, L. Ruffus: a lightweight Python library for computational pipelines. *Bioinformatics*
1224   **26**, 2778-2779 (2010). https://doi.org:10.1093/bioinformatics/btq524

1225   60   Sergeant, M. J. *et al.* Multi Locus View: an extensible web-based tool for the analysis of
1226   genomic data. *Commun Biol* **4**, 623 (2021). https://doi.org:10.1038/s42003-021-02097-y

1227   61   Van Valen, D. A. *et al.* Deep Learning Automates the Quantitative Analysis of Individual Cells
1228   in Live-Cell Imaging Experiments. *PLoS Comput Biol* **12**, e1005177 (2016).
1229   https://doi.org:10.1371/journal.pcbi.1005177

1230   62   Levine, J. H. *et al.* Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that
1231   Correlate with Prognosis. *Cell* **162**, 184-197 (2015). https://doi.org:10.1016/j.cell.2015.05.047

1232   63   Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony.
1233   *Nat Methods* **16**, 1289-1296 (2019). https://doi.org:10.1038/s41592-019-0619-0

1234   64   Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics.
1235   *BMC Genomics* **19**, 477 (2018). https://doi.org:10.1186/s12864-018-4772-0

1236   65   Prabhakaran S., G. C., Robertson-Tessi M., Beg A. A., Gray J., Antonia S., Gatenby R. A.,
1237   Anderson A. R. A. Distinct tumor-immune ecologies in NSCLC patients predict progression and
1238   define a clinical biomarker of therapy response. *bioRxiv preprint* (2022).
1239   https://doi.org:doi:10.1101/2022.10.22.513219

1240   66   Loh, J. M. A valid and fast spatial bootstrap for correlation functions. *Astrophys J* **681**, 726-734
1241   (2008). https://doi.org:Doi 10.1086/588631

1242   67   Manz, T. *et al.* Viv: multiscale visualization of high-resolution multiplexed bioimaging data on
1243   the web. *Nat Methods* **19**, 515-516 (2022). https://doi.org:10.1038/s41592-022-01482-7

1244   68   Schmidt, M. The sankey diagram in energy and material flow management. *J Ind Ecol* **12**, 82-
1245   94 (2008). https://doi.org:10.1111/j.1530-9290.2008.00004.x

1246   69   Jaccard, P. The distribution of the flora in the alpine zone. *New Phytol* **11 (2)** (1912).

1247   70.   Büttner, M., Ostner, J., Müller, C. L., Theis, F. J. & Schubert, B. scCODA is a Bayesian model
1248   for compositional single-cell data analysis. *Nat. Commun. 2021 121* **12**, 1–10 (2021).
1249
1250

1251   **Acknowledgements**

1258   **Author Contributions Statement**
1259

1260   PW analysed the data, contributed to development of MDV, spatial analyses, and interpretation of

1261   mathematical output and all spatial data, and writing of the paper. LD contributed to analysis and

1262   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 DA 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, FI interpreted data, discussed annotations and immunological analysis, and contributed to writing of the paper. DS and SMG performed early code testing for the pipeline. 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. JMW, AA, PW and LPH performed and analysed all the transcriptomic data from previous studies and publicly deposited data, and revisions of paper. 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.

Our authorship is diverse, inclusive and equal. There is representation from both sexes, high and low income countries of origin, age, and minorities.

Competing interests Statement

PK has acted as a consultant for Biomunex, Infinitopes, Astra Zeneca and UCB. The remaining authors declare no competing interests.

1297

1298 **Figure legends**

1299 **Fig. 1. Spatial analysis pipeline and histopathology categorisation of samples**

1300   A. Overview of the workflow and SpOOx pipeline. The steps of the analysis are presented in
1301   Supplementary Fig. 1 in more detail. n=677,623 single cells refer to segmented cells, without
1302   filtering for cells with no antibody staining, and 'undefined' clusters. IMC – imaging mass
1303   cytometry.

1304   B. H&E section from COVID-19 tissue section showing formal histopathology features of alveolitis
1305   (ALV), diffuse alveolar damage (DAD) and organizing pneumonia with their corresponding MCD
1306   file image showing staining for 5 of 35 antibodies (α-SMA, EpCAM, PanCK, Col 1a and CD31).
1307   'a'-'c' in figure refer to characteristic features of ALV, DAD and OP. 'a' - thickened alveolar
1308   epithelial wall and septae with immune cell infiltrate and congestion of alveolar walls 'b'
1309   widespread presence of hyaline membrane, and regenerating/proliferating Type II alveolar
1310   epithelium and 'c' -fibroblasts and collagen presence around bronchial epithelium. See also
1311   Supplementary Fig. 2. Representative H&E and MCD images is for ROI from n=10 ROIs for ALV,
1312   n=8 ROIs (DAD), n=8 ROIs (OP); n=12 patients. H&E staining performed once per tissue section.
1313   37-plex staining was performed once for each lung sample.

1314   C. Point when samples were obtained from the first day of symptoms and corresponding
1315   histopathology states in lung sections. Mean and S.D. shown, p value calculated using one-way
1316   ANOVA test with Tukey's multiple comparison test; normality tested with d'Agostino & Pearson
1317   test. n=4 patients in each histopathology group (ALV, DAD and OP).

1318   D. C-reactive protein (CRP) levels closest to the point of sampling and corresponding
1319   histopathology state in lung sections. Median and IQR shown, p value calculated using Kruskal-
1320   Wallis test with Dunn's multiple comparison test. n=4 patients in each histopathology group
1321   (ALV, DAD and OP). Source data are provided in the Source Data File.

1322

1323 **Fig. 2. High definition immunophenotyping of lung cells and identification of tissue structure**

1324

1325   A. UMAP representation of myeloid, lymphocyte and structural cell 'mega clusters' from all
1326   regions of interest (ROI) (k = 30) (COVID-19 and HC). See also Supplementary Fig. 7 for extended
1327   analysis steps. HC – healthy control.

1328   B-C. Number of cells per mm^2 of lung tissue sections in all COVID-19 samples (n=12 patients, 30
1329   ROIs' in total) compared to healthy control (HC) samples (n=2 individuals, 4 ROIs in total).
1330   Median shown, error bars are IQR. n=524,552 cells in total for COVID-19 samples, n= 30,053

1331   cells for HC. Statistical analysis performed after samples grouped into histopathology states

1332   (see Fig. 3D). Source data are provided in the Source Data File.

1333

1334   D-E.   Immunofluorescence (IF) staining validation for Neut_CD8_ADJ cell cluster and

1335   Mono_CD31_ADJ cell clusters. Small panels are high magnification confocal images showing

1336   CD8 and CD15 staining (top small panel), and CD14 and CD31 staining (bottom small panel) on

1337   adjacent cells. Broken yellow circles show CD8 T cell (white- CD8) – neutrophil (green-CD15)

1338   couplets throughout lungs (D); and CD14-staining cells next to CD31-expressing cells

1339   (endothelium) in lung tissue (E). See also Supplementary Fig. 6 for negative controls. IF images

1340   shown are representative of lung sections from n=3 patients; staining experiment performed

1341   once per lung sections.

1342   F.   Heatmap of median scaled intensity for each marker for all cell clusters in the 'Myeloid' mega

1343   cluster. 'n_cells' - average number of cells in all COVID-19 ROIs. Total cells – 171, 777. UD –

1344   undefined cluster

1345   G-H. Exemplar MCD image from 37-plex imaging mass cytometry (IMC) staining of a DAD ROI

1346   showing expression of CD8 T cells (CD8 -green), neutrophils (CD15-red) and CD8_CD15_ADJ

1347   cell clusters (green and red co- expression, making yellow). Image is one of n=26 ROIs, some

1348   of which do not have the CD8_CD15_ADJ cell clusters – see Fig. 3B for number of ROIs showing

1349   presence of this cell cluster in all ROIs (n=26 COVID-19; n=4 HC). H - Same MCD images as (G)

1350   but with IFN-β channel 'open' (white) showing IFN-β expression on Neut_CD8_ADJ (yellow).

1351   I.   Higher magnification of a set of 3 MCD panels - 'none' - Neut_CD8_ADJ (yellow) only (arrows);

1352   'IFN-γ' – with 'IFN-γ' (white) channel opened on MCD viewer showing expression on

1353   Neut_CD8_ADJ (yellow) (arrows) and some CD8 (green); 'GZB' - with 'GZB' (cyan) channel

1354   opened and showing expression on Neut_CD8_ADJ (yellow) (arrows). CD172a panel shows

1355   confocal immunofluorescence staining (white) on CD15 and CD8 adjacent to each other. IF

1356   images shown are representative of lung sections from n=3 patients; staining experiment

1357   performed once per lung sections. ALV – alveolitis, DAD – diffuse alveolar damage, OP -

1358   organising pneumonia. MCD images from all 26 ROIs (n=10 ALV, n=8 DAD and n=8 OP) were

1359   analysed and median expression intensity for all ROIs shown in (F) and Supplementary Fig. S7.

1360   All scale bars in μm.

1361

1362   **Fig. 3. Quantification of immune and structural cells in COVID-19 lungs**

1363   A-C.   Cell abundance plots for immune cells (myeloid and lymphoid cells) and structural cells in lung

1364   tissue, adjusted for surface area in COVID lungs categorised into those with histopathology

1365   states of alveolitis(ALV) (n=4 patients, 10 ROIs), diffuse alveolar damage (DAD) (n=4 patients,
1366   8 ROIs) and organising pneumonia (OP) (n=4 patients, 8 ROIs), compared to healthy control
1367   (n=2 individuals, 4 ROIs). Line in figure represents median. See Supplementary Table 4 for
1368   extended phenotypic description for all cell types and clusters. Source data are provided in the
1369   Source Data File.

1370

1371   D.   Heatmap of fold change (FC) difference in abundance of cell types for COVID-19 samples (ALV,
1372   DAD and OP) vs healthy controls (HC) depicted in (A). Asterisks show those with significant
1373   differences - adjusted p-values are *p<0.05 **p<0.01 ***p<0.001, calculated using code from
1374   the diffcyt R package (version 1.8.8) with the option testDA_edgeR; two-sided analysis
1375   employed, and multiple comparisons adjusted using Benjamini-Hochberg method . Arrow
1376   refers to immune cells that showed progressive increase in abundance with progression
1377   histopathology states from ALV to OP.

1378

1379   **Fig. 4. Spatial analysis of immune and structural cells in COVID-19 lungs**

1380   A.   Schematic representation of the sequential spatial analysis of cellular co-location, starting
1381   with quadrat correlation matrix (QCM), then cross pair correlation function (cross-PCF)
1382   analysis, interrogation of cross-PCF output and organization according to main questions.
1383   QCM output is provided in Supplementary Fig. 8.
1384   B.   g(r=20) heatmaps showing statistically significant correlated pairs of cells derived from QCM
1385   and cross-PCF analysis (see Methods for full description). n=479,349 single cells from n=12
1386   COVID patients' lung sections (n=26 ROIs); in total, n=144,937 cells in ALV, n=146333 in DAD
1387   and n=163,506 in OP. Red boxes indicate groups of cell subsets from the same immune
1388   phenotype – neutrophils (Group 1), monocytes and macrophages (Group 2), CD3 T cells
1389   (Group 3) and MAIT cells (Group 4).

1390

1391   **Fig. 5. Spatial organization of immune cells around structural cells in COVID-19 lungs**

1392   A-B. Spatial connectivity plots for proliferating alveolar epithelium, showing immune cells that are
1393   significantly co-located to proliferating alveolar epithelium (designated 'anchor cell') in the three
1394   histopathology states. The size of the nodes (filled-in circle) represents mean cell counts
1395   (abundance) for the specified cell cluster for all the ROIs in the histopathology state (scale shown
1396   in grey), and colour of nodes relate to histopathology state. Connecting lines indicate a
1397   statistically significant co-location between the two cell types derived from QCM and cross-PCF
1398   analyses. The thickness of the lines relates to the g(r=20) value relative to each pair in the plot –

1399   the thicker the line, the higher the g(r=20) and therefore greater strength of co-location between

1400   the immune cell type and anchor cell. n=479,349 single cells from n=12 COVID patients' lung

1401   sections (n=10 ROIs for ALV; n= 8 DAD; n= 8 OP); n=144,937 cells in ALV, n=146333 in DAD and

1402   n=163,506 in OP.

1403   C-D. Histogram shows % of two anchor cells – proliferating alveolar epithelial (PAE) cells (C) and

1404   endothelial cells (D) that are in contact with specified immune cell type. Source data are provided

1405   in the Source Data File.

1406   E-F. Cross-PCF profiles for the two most abundant co-located structure:immune cell pairs in DAD.

1407   Curves show the change in g(r) along the radius(r) from anchor cells [proliferating alveolar

1408   epithelium (prolif alv epit) and endothelial cells (endo)] for Neut-CD8_ADJ cell clusters and

1409   Mono_CD31_ADJ cell clusters respectively. Blue coloured area around curve is the 95%

1410   confidence interval for n=8 ROIs with DAD.

1411   G. Radial connectivity map depicting all statistically significant pairs of structure:immune cells in all

1412   histopathology states; anchor cells (structural cells) are in smaller, inner circle. n=479,349 single

1413   cells from n=12 COVID patients' lung sections (n=10 ROIs for ALV; n= 8 DAD; n= 8 OP). 'DRhi BE'

1414   – HLADR^{hi} bronchial epithelium; 'DRlo BE' – HLA DR^{lo} bronchial epithelium; "Endo"- endothelial

1415   cells; 'PAE'- 'proliferating alveolar epithelium', 'PBE' – 'proliferating bronchial epithelium'; 'PE' –

1416   'proliferating endothelium' 'BV" –'blood vessels'. Numerical values indicate g(r=20) for that pair

1417   in that state (coloured bar), and % indicates proportion of anchor cells that are co-located with

1418   the specified immune cells.

1419   H. Topographical correlation map showing distribution of the co-located Neut_CD8_ADJ cluster and

1420   proliferating alveolar epithelial cell pair (left panel) in an exemplar tissue (an ROI with DAD). Cells

1421   of type A (e.g. Neut_CD8_ADJ) are positively ($\Gamma_{ab} \gg 0$) or negatively ($\Gamma_{ab} \ll 0$) associated with

1422   cells of type B (e.g. Proliferating alveolar epithelium) (see Methods)

1423   I. MCD images showing Neut_CD8_ADJ clusters amidst single CD8$^+$ T cells, CD15$^+$ immature

1424   neutrophils and epithelial markers (EpCAM and PanCK). Couplets of CD8$^+$ and CD15$^+$ cells -

1425   Neut_CD8_ADJ clusters (red and green merging to form yellow cells) (arrows) are most clearly

1426   visible in DAD. Exemplar section is shown from analyses of n=10 ALV ROIs, n=8 DAD ROIs and

1427   n=8 OP ROIs (n=12 patients). Sections were stained once with 37 plex panel.

1428

1429   **Fig. 6. Spatial organization amongst immune cells in COVID-19 lungs**

1430   A-C. Radial connectivity map depicting all statistically significantly co-located pairs of immune-

1431   immature neutrophil subsets (including ADJ subsets) (A) immune-monocyte subsets (B and C,

1432   separated for clarity) cells in all histopathology states (n=10 ALV, n=8 DAD and n=8 OP). Anchor

1433   cells (immature neutrophil and monocyte subsets) are in smaller, inner circle. Numerical values

1434   indicate g(r=20) for that pair in that state (coloured bar), and % indicates proportion of anchor

1435   cells that are co-located with the specified immune cells. These significantly co-located pairs of

1436   cells are derived from n=479,349 single cells in all ROIs from n=12 COVID patients' lung sections

1437   (n=10 ROIs for ALV; n= 8 DAD; n= 8 OP); n=144,937 cells in ALV, n=146333 in DAD and n=163,506

1438   in OP (see Methods for 3-step mathematical algorithm for determining statistical significance of

1439   co-location).

1440

1441 D-E. Spatial connectivity plots for Neut_CD8_ADJ (D) and Mono_CD31_ADJ (E), showing immune

1442   cells that are statistically significant co-located to proliferating alveolar epithelium (designated

1443   'anchor cell') in the three histopathology states (see Methods for 3-step mathematical algorithm

1444   for determining statistical significance of co-location). Size of nodes (filled-in circle) represent

1445   mean cell counts for the specified cell cluster for all the ROIs in the histopathology state, and

1446   colour of nodes relate to histopathology state. Connecting lines indicate a statistically significant

1447   co-location between the two cell types derived from QCM and cross-PCF analyses. Thickness of

1448   line relate to value of g(r=20) relative to each pair in the plot – the thicker the line, the higher the

1449   g(r=20) and strength of co-location between the immune cell type and anchor cell.

1450 F. Adjacency cell network (ACN) map showing contact between the Mono_CD31_ADJ cluster,

1451   Neut_CD8_ADJ cluster and proliferating alveolar epithelial. Cell segmentation masks generated

1452   by DeepCell were used to produce this spatially-embedded network in which nodes represent

1453   centres of cell types (e.g. green – Neut_CD8_ADJ cell cluster). Nodes are connected by a line if

1454   the corresponding cells in the segmentation mask share a border.

1455 G. MCD image showing CD8 (green), CD15(red) and CD15 and CD8 co-staining (yellow) (representing

1456   Neut_CD8_ADJ cell clusters) amidst endothelial cells (CD31 staining in turquoise) and monocytes

1457   (CD14 staining in purple) in a lung section with DAD on histopathology analysis. Exemplar ROI is

1458   shown for (F) and (G), out of 26 ROIs stained, from 12 patients (n=10 ALV ROI, n=8 DAD and n=8
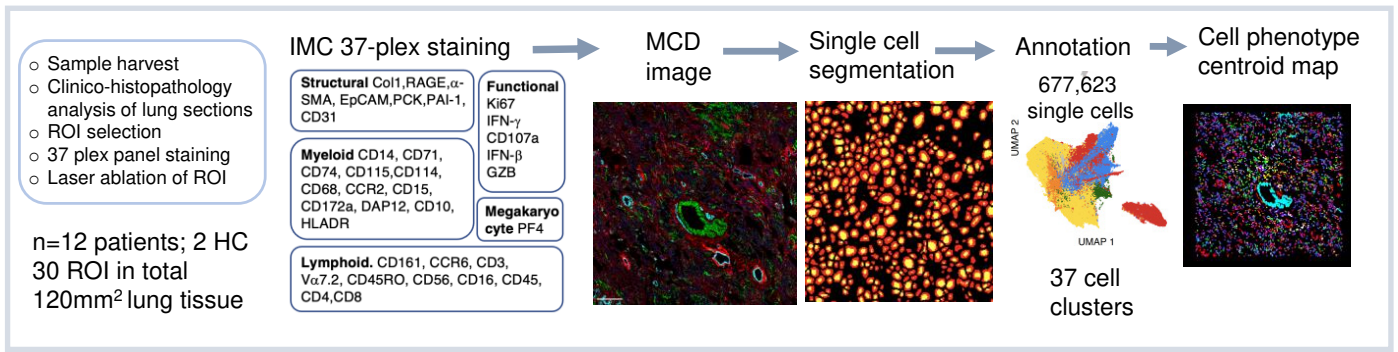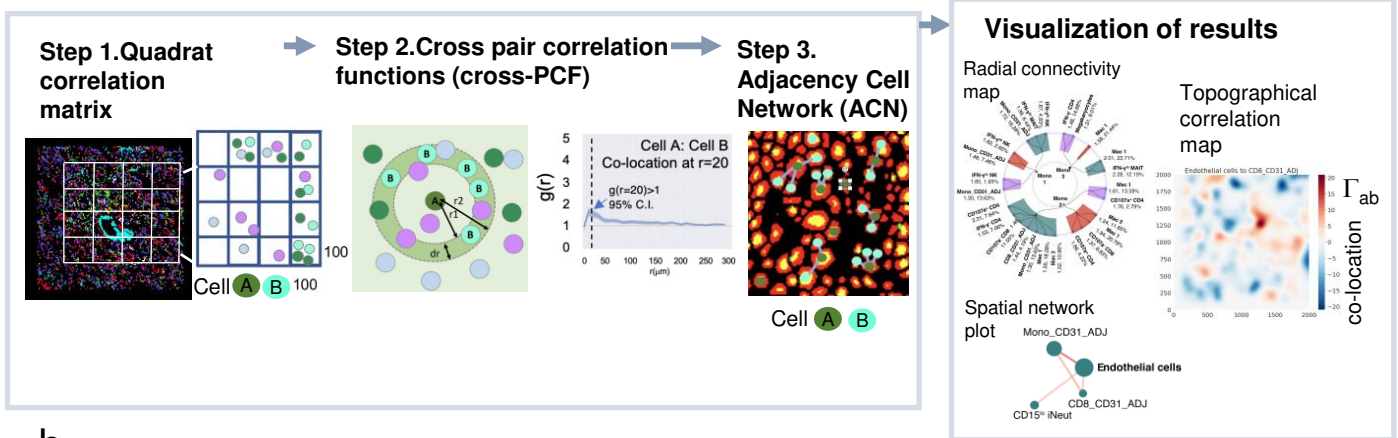
1459   OP).

1460

1461 **Fig. 7**

1462 A. SCMAP matching heatmaps representing the Jaccard indices of similarity between COMBAT

1463   (blood)[6] and COSMIC (lung) lymphocyte clusters. CD107a$^-$ CD8 T cell and CD107a$^+$ CD8 T cell in

1464   COSMIC matched to blood GZB$^-$ CD8 T cells in COMBAT. IFN-γ^+ CD4 T cells matched to COMBAT's

1465   'activated CD4 T cells'

1466   B. SCMAP matching heatmaps representing the Jaccard indices of similarity between COMBAT

1467   (blood) and COSMIC (lung) myeloid clusters. Mono_CD31_ADJ and Mono_PAI-1_ADJ and all

1468   macrophage subsets matched with HLA DR^{hi} classical monocytes in the blood from COMBAT data

1469   C. UMAP representation of neutrophils from controls and COVID-19 infected patients (n=2,776,928

1470   single cells from n=77 COVID-19 patients and 11 healthy volunteers (HV), down sampled to 100

1471   000 cells per condition) obtained from COMBAT consortium, showing 8 subsets of neutrophils.

1472   D. Heatmap showing median marker expression for genes (selected to match COSMIC's key protein

1473   expression on neutrophils) on the 8 neutrophil subsets, demonstrating high similarity of marker

1474   expression in immature neutrophil 2 (iNeut2) in COMBAT (blood) with Neut_CD8_ADJ in COSMIC

1475   (lung) (See also Fig. 2F).

1476   E. Abundance of the 8 neutrophil subsets in blood as % of total neutrophils, from healthy volunteers

1477   (HV), mild, severe and critical COVID-19 patients from the COMBAT consortium showing a

1478   progressive increase in immature 2 neutrophils with increasing COVID-19 disease severity. HV

1479   (n=11), mild (n=18), severe (n=41), critical (n=18) patients, n=1 experiment. The boxplot is

1480   median, with IQR; whiskers are the range or 1.5*IQR (whichever is smaller). Composition

1481   analysis was performed using scCODA with inbuilt adjustment for multiple comparison[70].

1482   Credible compositional changes were identified comparing all groups to HV and FDR<0.1

1483   is marked with #. Source data are provided in Source Data File.
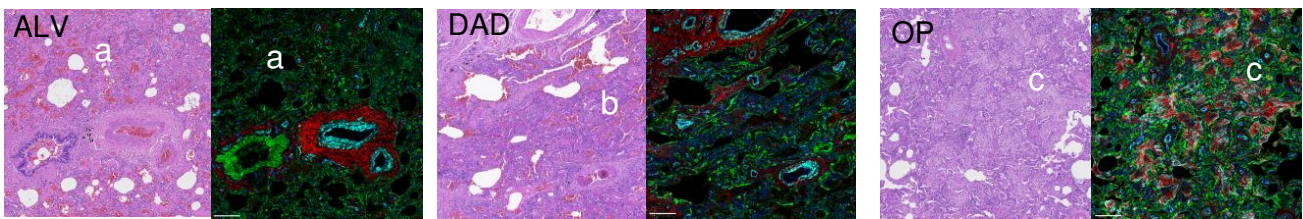
1484

1485

1486

1487

a.

IMC 37-plex staining

- Sample harvest
- Clinico-histopathology analysis of lung sections
- ROI selection
- 37 plex panel staining
- Laser ablation of ROI

n=12 patients; 2 HC
30 ROI in total
120mm² lung tissue

Structural Col1,RAGE,α-SMA, EpCAM,PCK,PAI-1, CD31

Functional
Ki67
IFN-γ
CD107a
IFN-β
GZB

Myeloid CD14, CD71, CD74, CD115,CD114, CD68, CCR2, CD15, CD172a, DAP12, CD10, HLADR

Megakaryo cyte PF4

Lymphoid. CD161, CCR6, CD3, Vα7.2, CD45RO, CD56, CD16, CD45, CD4,CD8

MCD image

Single cell segmentation

Annotation
677,623 single cells
37 cell clusters

Cell phenotype centroid map

Statistical quantification of spatial co-location of cells

Step 1.Quadrat correlation matrix

Cell A B 100

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

Step 3. Adjacency Cell Network (ACN)

Cell A: Cell B
Co-location at r=20
g(r=20)>1
95% C.I.

Cell A B

Visualization of results

Radial connectivity map

Topographical correlation map

Spatial network plot

Mono_CD31_ADJ
Endothelial cells
CD8_CD31_ADJ
CD15ʰⁱ iNeut

b.

ALV a / a
DAD b
OP c / c

α-SMA Epcam/PanCK Collagen CD31 Scale Bar 200μm

c.

Days from first symptom — Alv, DAD, OP — p=0.14

d.

CRP — Alv, DAD, OP — p=0.77

Figure 1

a. Structural   Myeloid   Lymphoid

b. Immune cells
c. Structural cells
f.

d. Neut_CD8_ADJ
e. Mono_CD31_ADJ

g.
h.
i.

Figure 2

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.

Group 1.
CD15^{hi} iNeut
CD15^{mid} iNeut
CD15^{lo} iNeut

Group 2.
Mono1   Mac1
Mono2   Mac2
Mono3   Mac3

Group 3.
CD107a$^+$ CD8   IFN-γ$^+$ CD4
CD107a- CD8   IFNg-γ- CD4
                        CD107a+ CD4

Group 4.
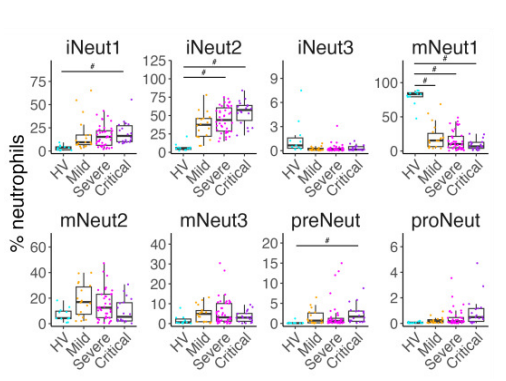IFN-γ^{hi} MAIT
IFN-γ^{lo} MAIT
Vα7.2^{lo} cells

Figure 4

Figure 5

Figure 6

Figure 7

Name of software	Source	Identifier
imctools	https://github.com/BodenmillerGroup/imctools	RRID:SCR_017132
Deepcell	https://vanvalen.github.io/about/	RRID:SCR_022197
Phenograph	https://github.com/JinmiaoChenLab/Rphenograph	RRID:SCR_016919
Harmony	https://github.com/slowkow/harmonypy	RRID:SCR_022206
Slingshot	https://github.com/kstreet13/slingshot	RRID:SCR_017012
Ruffus	http://www.ruffus.org.uk/	RRID:SCR_022196
QuPath	https://qupath.github.io/	https://doi.org/10.1038/s41598-017-17204-5
MCD	https://www.standardbio.com/products-services/software	RRID:SCR_023007
Catalyst R	http://bioconductor.org/packages/CATALYST/	RRID:SCR_017127
Harmony	https://github.com/immunogenomics/harmony	RRID:SCR_022206
diffcyt R package (version 1.8.8)	https://www.bioconductor.org/packages/release/bioc/html/diffcyt.html	RRID:SCR_023006

Table 1 Software and algorithms used in data analysis.