Title

Single-cell analysis of bronchoalveolar cells in inflammatory and fibrotic post-COVID lung disease

Authors

Puja Mehta[2]*, Blanca Sanz-Magallón Duque de Estrada[1]*†, Emma K Denneny[2]*, Kane Foster[3], Carolin T Turner[1], Andreas Mayer[1], Martina Milighetti[1], Manuela Platé[2], Kaylee B Worlock[2], Masahiro Yoshida[2], Jeremy S Brown[2], Marko Z Nikolić[2], Benjamin M Chain[1], Mahdad Noursadeghi[1], Rachel C Chambers[2]*, Joanna C Porter[2]*, Gillian S Tomlinson[1]*†

*These authors contributed equally.

†Corresponding authors.

Affiliations

1. Division of Infection and Immunity, University College London, London, UK
2. UCL Respiratory, University College London, London, UK
3. UCL Cancer Institute, University College London, London, UK

Corresponding authors

Dr Gillian S Tomlinson and Dr Blanca Sanz-Magallón Duque de Estrada, Infection and Immunity, University College London, Cruciform Building, Gower Street, London WC1E 6BT, United Kingdom. Email: g.tomlinson@ucl.ac.uk, blanca.estrada.16@ucl.ac.uk

ABSTRACT

Background Persistent radiological lung abnormalities are evident in many survivors of acute coronavirus disease 2019 (COVID-19). Consolidation and ground glass opacities are interpreted to indicate subacute inflammation whereas reticulation is thought to reflect fibrosis. We sought to identify differences at molecular and cellular level, in the local immunopathology of post-COVID inflammation and fibrosis.

Methods We compared single-cell transcriptomic profiles and T cell receptor (TCR) repertoires of bronchoalveolar cells obtained from convalescent individuals with each radiological pattern, targeting lung segments affected by the predominant abnormality.

Results Single-cell transcriptomes of inflammatory and fibrotic post-COVID lung disease closely resembled each other across all cell types. However, CD4 central memory T cells and CD8 effector memory T cells were significantly more abundant in those with inflammatory radiology. Clustering of similar TCRs from multiple donors was a striking feature of both phenotypes, consistent with tissue localised antigen-specific immune responses. There was no enrichment for known SARS-CoV-2-reactive TCRs, raising the possibility of T cell-mediated immunopathology driven by failure in immune self-tolerance.

Conclusions We found no evidence that post-COVID radiographic inflammation and fibrosis are associated with differential immmunopathological pathways. Both show evidence of shared antigen-specific T cell responses, suggesting a role for therapies targeting T cells in limiting post-COVID lung damage.

INTRODUCTION

Persistent functional and radiological lung abnormalities are evident at one year in approximately 20% of people who survive acute coronavirus disease 2019 (COVID-19) (Fabbri et al., 2023). Current understanding of the immunopathogenic mechanisms responsible for post-COVID lung disease (PCLD) is very limited (Mehta et al., 2022). Elevated numbers of airway CD4 and CD8 T cells have been reported (Cheon et al., 2021; Vijayakumar et al., 2022) and the post-COVID airway proteome displays evidence of ongoing epithelial injury that resolves with time (Vijayakumar et al., 2022). It is imperative to address this knowledge deficit to inform therapeutic interventions that could expedite resolution of pathology and minimize irreversible tissue damage, to reduce long term morbidity secondary to PCLD and the attendant burden on healthcare services.

There has been the impression of two major radiological patterns in PCLD: consolidation and ground glass opacities, thought to represent subacute inflammation, and reticulation, widely interpreted as fibrosis (Fabbri et al., 2023; Stewart et al., 2022). Inflammation predominates during acute COVID-19, but fibrosis is evident on 32% of computed tomography (CT) scans during hospitalisation. Follow-up imaging within the first year suggests radiological sequelae reduce with time, with less marked improvement in fibrosis (Fabbri et al., 2023). We hypothesised that these distinct radiological changes reflect distinct pathogenic mechanisms, which may require different treatments. We sought to evaluate the molecular characteristics of cellular function at the site of disease in PCLD, by single-cell RNA sequencing (scRNAseq) of bronchoalveolar cells from convalescent individuals infected during the first or second waves of the pandemic, with predominant CT features of inflammation or fibrosis at

68 the time of sampling.

69 We showed that in comparison to fibrotic PCLD, the bronchoalveolar environment of
70 inflammatory PCLD was enriched for CD4 T central memory cells (TCM) and CD8 T
71 effector memory cells (TEM). Consistent with this finding, a higher proportion of CD4
72 TCM clones were expanded in the inflammatory phenotype. Both inflammatory and
73 fibrotic PCLD bronchoalveolar T cells exhibited high levels of T cell receptor (TCR)
74 clustering, indicative of an antigen-specific immune response, but there was no
75 enrichment for known severe acute respiratory syndrome coronavirus 2 (SARS-CoV-
76 2)-reactive sequences. No major differences were evident in any of the cell type-
77 specific transcriptomic profiles of the two radiological phenotypes, suggesting that they
78 may represent different manifestations of the same disease process.

79 **RESULTS**

80 **Increased abundance of bronchoalveolar T cells in inflammatory PCLD**

81 Individuals undergoing bronchoscopy for clinical investigation of persistent respiratory
82 symptoms and predominant radiological features of either inflammation or fibrosis
83 following acute COVID-19, with no previous evidence of interstitial lung disease (ILD)
84 were recruited (Figure 1a). Bronchoalveolar lavage (BAL) samples from lung
85 segments affected by the predominant abnormality were obtained from five subjects
86 with each radiological pattern (Figure 1a). Clinical and demographic characteristics
87 are provided in Table 1 and Supplementary Table 1. Evaluation of the composition of
88 post-COVID BAL by scRNAseq revealed that macrophages dominated in all subjects,
89 with smaller T cell, NK T cell, dendritic cell, epithelial and B cell populations also
90 identified (Figure 1b-d and Supplementary Figure 1a-d). Each of the cell types
91 expressed high levels of independently established marker genes (Adams et al., 2020;

92 Cheng et al., 2021; David et al., 2021; Davies et al., 2013; Davis and Wypych, 2021;
93 Gardell and Parker, 2017; Glass et al., 2020; Murray and Wynn, 2011; Tsyklauri et al.,
94 2023; van Aalderen et al., 2021; Villani et al., 2017), validating our annotations (Figure
95 1d). The notable difference between the two phenotypes was significantly higher
96 abundance of CD4 T cells and CD8 T cells in inflammatory PCLD (Figure 1e). The
97 relative proportions of the other seven cell types present were not found to differ
98 between the two radiological phenotypes, providing confidence our analysis was not
99 confounded by compositional bias, where alterations in the proportion of one cell type
100 lead to many other cell types being falsely identified as differentially abundant.

101 To explore differences in cell type-specific transcriptional profiles between the two
102 radiological phenotypes, we first aggregated gene expression count data for each cell
103 type for each donor to form “pseudobulks”. We leveraged the ability of pseudobulk
104 statistical approaches to account for variability of biological replicates, allowing
105 detection of genuine differential gene expression whilst minimizing false discoveries
106 (Squair et al., 2021). Very few cell type-specific differentially expressed genes were
107 identified (Supplementary Table 2), which precluded further bioinformatic analysis and
108 suggested the transcriptomes of inflammatory and fibrotic PCLD were similar for all
109 cell types.

110 Detection of transcriptomic differences between the two groups using pseudobulk
111 statistical analysis may have been limited by small sample size. Therefore, we
112 repeated the comparison of inflammatory and fibrotic PCLD at the level of individual
113 cells. Single-cell statistical analysis identified thousands of cell type-specific
114 differentially expressed genes in each radiological phenotype. However, single-cell
115 differential gene expression analysis has a propensity for false positive results (Squair

116 et al., 2021). To mitigate against this, we sought to assess whether differentially
117 expressed gene lists represented differentially enriched biological pathways in the two
118 study groups. This was not evident at the level of biological processes or upstream
119 regulator analysis (Supplementary Figure 2 and 3), which suggested enrichment of
120 overlapping processes and pathways despite apparent differential gene expression.
121 Hence, our single-cell analysis also supports the outcome of the pseudobulk analysis.

122 **CD4 central memory and CD8 effector memory are the predominant T cell**
123 **subsets in PCLD**

124 Given that greater abundance of T cells in inflammatory cases was the only robust
125 difference between the two PCLD radiological phenotypes, we re-clustered these
126 populations alone to undertake a more detailed analysis. This revealed CD4 TCM and
127 CD8 TEM as the two predominant T cell subsets in the PCLD bronchoalveolar
128 environment, and a smaller population of regulatory T cells (Treg). A small mixed T
129 cell cluster, which expressed high levels of interferon-stimulated genes and an NK cell
130 cluster were also present in all samples. A minor population of gamma delta T cells
131 was identified in one individual with inflammatory PCLD (Figure 2a-c). There was no
132 difference in the relative proportions of any T cell subset between the two PCLD
133 phenotypes (Figure 2d).

134 Analysis of T cell subset pseudobulks revealed no genes as differentially expressed
135 in either group (Supplementary Table 2), consistent with our earlier observation of few
136 differences between the transcriptional programmes of the two PCLD phenotypes at
137 the level of broad cell types defined using the full dataset. In single-cell differential
138 expression analysis of T cells, genes expressed at a significantly higher level in CD4
139 TCM and CD8 TEM in inflammatory PCLD exhibited weak enrichment for immune

140 response signalling pathways (Supplementary Figure 4a). Robust enrichment of
141 cellular pathways in anti-viral responses was evident for genes expressed at a
142 significantly higher level in CD4 TCM and CD8 TEM in fibrotic PCLD (Supplementary
143 Figure 4b), but this was not supported by enrichment for Type I interferon signalling in
144 upstream regulator analysis (Supplementary Figure 5b). Interestingly, interleukin (IL)2
145 and IL15 were the most statistically significant upstream regulators of CD4 TCM and
146 CD8 TEM differentially expressed genes in inflammatory PCLD but not fibrotic PCLD,
147 consistent with the notion of cytokine driven proliferation leading to increased
148 abundance of T cells in the former phenotype (Supplementary Figure 5a-c).

149 **Alveolar macrophage and monocyte subsets in PCLD are consistent with**
150 **healthy airspace myeloid populations**

151 Macrophages are the most abundant cell types in healthy airspaces (Mould et al.,
152 2021) and were also found to be the most abundant cell type in PCLD BAL samples
153 in the present study. Macrophage subpopulations have been implicated in the
154 pathogenesis of fibrosis associated with severe COVID-19 and idiopathic pulmonary
155 fibrosis (Nouno et al., 2019; Wendisch et al., 2021). We therefore re-clustered
156 macrophages alone to further our understanding of macrophage subpopulations in
157 PCLD. Two populations with similar expression of macrophage markers and
158 comparable transcriptomic profiles, likely representative of the transcriptional
159 spectrum of resident healthy alveolar macrophages (AM) (Mould et al., 2021) were
160 combined for subsequent analysis (Figure 3a,b and Supplementary Figure 6a,b). We
161 identified three further small specialized AM subsets previously detected in healthy
162 individuals (Mould et al., 2021), characterized by high levels of proinflammatory
163 molecule expression, “Inflam AM”, metal-binding metallothioneins, “MT-AM”, or

interferon-stimulated genes “IFN stim AM”. Proliferating macrophages were delineated by high expression of a gene module representing the cellular proliferation response (Chandran et al., 2022) (Figure 3a-d).

Monocyte-like cells are present in healthy airspaces, and suggested by trajectory inference and increasing expression of macrophage marker genes over pseudotime to differentiate into AM, implying constant trafficking of monocytes into the lung (Mould et al., 2021). Consistent with this, we identified *CD14, FCN1* expressing classical monocytes (*FCN1*-Mono) (Mould et al., 2021; Wauters et al., 2021) characterized by high levels of *CCR2*, the receptor for monocyte chemoattractant protein, suggestive of recent recruitment from the peripheral blood (Auffray et al., 2009) (Figure 3c,d). A second *LGMN* and *SPP1* expressing subset (*LGMN*-Mono), recently identified as a rare population defined by expression of cell-matrix interaction genes in healthy airspaces (Mould et al., 2021), was also present (Figure 3c,d). In contrast to acute severe COVID-19 where proinflammatory monocytes and profibrotic *SPP1, LGMN* expressing macrophages have been reported to be abundant (Liao et al., 2020; Wendisch et al., 2021), monocytes represented a minor constituent of PCLD BAL. Neither subset expressed high levels of inflammatory mediators (Figure 3c,d) and very few cells expressed a gene signature characteristic of profibrotic macrophages recently identified in idiopathic pulmonary fibrosis (Adams et al., 2020), (Figure 3e and Supplementary Figure 7a).

No differences in the relative proportions of any of the myeloid populations were evident between the two PCLD phenotypes (Figure 3f). Very few gene expression differences were detected between the two phenotypes by pseudobulk analysis of macrophage or monocyte subsets (Supplementary Table 2). Similar to the analysis of

188 the full dataset stratified by broad cell type, re-clustering macrophages to identify more
189 discrete sub-populations did not reveal transcriptomic differences between
190 inflammatory and fibrotic PCLD.

191 Single-cell differential gene expression analysis revealed weak enrichment for
192 immune response and cellular metabolism pathways in both groups (Supplementary
193 Figure 7b,c). Minor enrichment for TGFβ-mediated signalling was evident for genes
194 expressed at higher levels in AM, Inflam-AM and *FCN1*-Mono in inflammatory PCLD
195 (Supplementary Figure 7b). Proinflammatory cytokines, T cell activation factors, SPP1
196 and TGFβ were identified as statistically enriched upstream regulators of differentially
197 expressed genes in both monocyte populations in inflammatory PCLD
198 (Supplementary Figure 8a,b,d,e). However, as for analysis at the level of broad cell
199 types and refined T cell subsets, there was considerable overlap between molecules
200 predicted to drive gene expression differences in each PCLD phenotype
201 (Supplementary Figure 8c,f), suggesting that between group differential gene
202 expression in this analysis did not represent differential biology between radiological
203 patterns of disease.

204 **Highly related TCRs indicate antigen-specific immune responses in PCLD**

205 To further evaluate T cell responses in PCLD we undertook single-cell TCR
206 sequencing (scTCRseq) and compared the TCR repertoire in each radiological
207 phenotype. We successfully obtained scTCRseq data for five fibrotic and three
208 inflammatory cases. Increased T cell abundance in inflammatory PCLD may reflect
209 increased numbers of unique T cell clones or increased expansion of individual clones.
210 Expanded clonotypes identified by being present with a frequency of greater than one,
211 were evident within the three major T cell subsets in both inflammatory and fibrotic

212 PCLD (Figure 4a). A greater proportion of CD4 TCM clones were expanded in the
213 inflammatory phenotype, consistent with our observation of the higher abundance of
214 this subset in this group. However, the proportion of expanded CD8 TEM and Treg
215 clones was similar in both phenotypes (Figure 4b).

216 As T cell clonal expansion was evident in both PCLD phenotypes, we next sought to
217 identify related TCRs for each group based on the similarity of their antigen specificity-
218 determining complementarity determining region (CDR)3 amino acid sequences, on
219 the premise that clusters of related TCRs recognize similar epitopes (details of the
220 analysis are provided in the methods). We hypothesized that the inflammatory cases
221 would exhibit more clustering than the fibrotic cases, given the trend towards a greater
222 proportion of expanded CD4 TCM clones and increased abundance of both CD4 TCM
223 and CD8 TEM in the inflammatory group. However, high levels of clustering were
224 evident in both inflammatory and fibrotic PCLD. In benchmarking, this level of
225 clustering was similar to that observed for expanded peripheral blood TCR clones
226 detected following non-severe SARS-CoV-2 infection, and exceeded that in non-
227 expanded TCRs from non-infected individuals from the same cohort (Chandran et al.,
228 2022; Milighetti et al., 2023) (Figure 4c,d). In both radiological groups of PCLD the
229 vast majority of clusters contained TCRs from multiple donors (Figure 4e) but there
230 was minimal sharing of identical CDR3 sequences between different individuals
231 (Supplementary Table 3). As a further comparison of inflammatory and fibrotic PCLD
232 T cell repertoires we clustered CDR3 amino acid sequences from both groups
233 together. Strikingly, most clusters contained TCRs from both inflammatory and fibrotic
234 samples and multiple donors, suggesting the presence of T cell clones that recognize
235 similar antigens across both phenotypes. A few small clusters composed uniquely of

236 either inflammatory or fibrotic PCLD TCRs were evident, indicative of subtle
237 differences between the two repertoires (Figure 4f).

238 The high level of relatedness between TCRs in PCLD is suggestive of antigen-specific
239 immune responses. We therefore sought to determine whether these repertoires were
240 enriched for T cells specific for SARS-CoV-2 by comparison to other common viruses.
241 Of the SARS-CoV-2-reactive TCRs identified in the VDJdb database, six were present
242 in TCR data from inflammatory PCLD and eight in fibrotic PCLD. None were identified
243 in equivalent sized healthy peripheral blood repertoires. However, fewer SARS-CoV-
244 2-specific-TCRs were detected than Epstein-Barr virus (EBV)-specific or
245 cytomegalovirus (CMV)-specific sequences, indicating no enrichment for SARS-CoV-
246 2-specific T cells at the site of disease. There was no enrichment for SARS-CoV-2-
247 reactive or EBV-reactive TCRs in inflammatory compared to fibrotic PCLD. However,
248 CMV-specific TCRs were significantly enriched in the fibrotic group (Figure 4g). TCRs
249 found in clusters composed uniquely of either PCLD phenotype were not enriched for
250 known virus reactive-sequences (Supplementary Table 4).

251 **DISCUSSION**

252 We report the first comparative molecular analysis of cells sampled by bronchoalveolar
253 lavage from patients displaying either predominantly inflammatory or fibrotic
254 pulmonary radiological sequelae following COVID-19. The bronchoalveolar
255 environment of inflammatory PCLD was characterised by significantly increased
256 abundance of CD4 central memory and CD8 effector memory T cells compared to
257 fibrotic PCLD. Clustering of similar TCRs from different donors, far exceeding that
258 observed in healthy peripheral blood, was evident in both radiological phenotypes,
259 suggestive of an antigen-specific immune response localised to the lung. The

260 transcriptomes of post-COVID radiological inflammation and fibrosis were highly
261 similar for all bronchoalveolar cell types, dominated by cellular processes involved in
262 inflammatory and immune responses. We found no robust evidence of enhanced
263 activity of tissue damage or wound repair pathways in those with fibrotic radiological
264 changes. Although many cell type-specific differentially expressed genes were
265 identified between the two radiological phenotypes by single-cell methods, there were
266 no systematic differences in enriched biological pathways or their upstream regulators
267 among these differentially expressed genes. Furthermore, almost no differences in
268 gene expression between the fibrotic and the inflammatory groups were detected if
269 gene expression within cell types was aggregated by donor before differential analysis
270 (Squair et al., 2021). Our data lead us to propose that the two radiological phenotypes
271 represent distinct manifestations of a similar pathological process.

272 T cell infiltration has been a consistent finding in recent studies which have examined
273 the post-COVID-19 airspaces; almost exclusively within 3-6 months of the acute insult,
274 when radiological inflammation is more common than fibrosis (Cheon et al., 2021;
275 Gagiannis et al., 2023; Ravaglia et al., 2022; Vijayakumar et al., 2022). The cellular
276 proportions of our fibrotic PCLD samples, harvested at 9-12 months after acute illness,
277 were akin to those reported in BAL samples from healthy individuals, with
278 macrophages comprising greater than 80% and lymphocytes less than 15% of cells
279 (Mould et al., 2021; The BAL Cooperative Group Steering Committee, 1990). This is
280 in keeping with the reported repopulation of the airspaces by AM in the later stages of
281 COVID-19 acute respiratory distress syndrome (ARDS) (Wendisch et al., 2021).
282 Minimal data currently available for longitudinal samples obtained at approximately
283 one year post-infection also suggest a trajectory of gradual normalization of molecular

284 abnormalities and airspace cellular composition (Vijayakumar et al., 2022).
285 Nonetheless there is considerable interest in therapeutic intervention to expedite
286 resolution of subacute inflammation, to prevent progression to fibrosis with
287 concomitant irreversible tissue damage in susceptible individuals. Our findings
288 suggest therapies which target T cells, rather than anti-fibrotic agents could be
289 beneficial in PCLD, particularly in those with radiological inflammation. In support of
290 this premise, one small uncontrolled clinical study demonstrated improvement in
291 clinical symptoms, physiological parameters and radiological abnormalities, following
292 three weeks of corticosteroid treatment instituted approximately three months post-
293 COVID-19 (Myall et al., 2021).

294 The bronchoalveolar TCR repertoire from both PCLD phenotypes exhibited high levels
295 of relatedness, suggestive of antigen-specific immune responses. In addition, the high
296 proportion of clusters containing TCRs from multiple donors suggests immune
297 responses against similar antigens, despite almost no sharing of identical TCR
298 sequences between different individuals. Aberrant immune responses to persistent
299 reservoirs of virus have been posited as drivers of post-acute sequelae of SARS-CoV-
300 2 infection (Cheung et al., 2022; Merad et al., 2022; Stein et al., 2022). However, even
301 including all detected TCRs, to offset the inherent sparsity of single-cell data, there
302 was no enrichment for known SARS-CoV-2-specific T cell clones compared to TCRs
303 specific for other common viruses. Hence, based on the current compendium of
304 SARS-CoV-2-reactive TCRs (Bagaev et al., 2020), which may not be comprehensive,
305 there was no evidence for viral persistence at the site of disease in our cohort. A more
306 plausible hypothesis, given reports of cross-reactivity between SARS-CoV-2 and
307 human antigens (Nunez-Castilla et al., 2022; Vojdani and Kharrazian, 2020), is that

308 the PCLD TCR repertoire is directed against as yet unknown respiratory autoantigens.
309 Identifying the antigenic targets of the PCLD T cell repertoire may present
310 opportunities for more specific therapeutic interventions.

311 Pro-fibrotic macrophages have been implicated in the pathogenesis of severe COVID-
312 19 ARDS, which has been associated with rapid onset pulmonary fibrosis, which
313 improves over time (Wendisch et al., 2021). Minimal interstitial fibrosis and the
314 presence of pro-fibrotic macrophages have also been reported in transbronchial lung
315 biopsies from a subset of individuals sampled at least 12 weeks after mild COVID-19
316 (Gagiannis et al., 2023). However, there was no difference between the frequency at
317 which mild fibrotic features were detected in the PCLD group and pre-pandemic
318 autopsy samples from individuals who had died of non-respiratory causes (Gagiannis
319 et al., 2023). In our study, which encompassed a broad range of severity of acute
320 disease, including several individuals with COVID-19 ARDS, the phenotype of myeloid
321 cells in both radiological groups was consistent with the spectrum of macrophages
322 and monocytes found in healthy airspaces (Mould et al., 2021); with no convincing
323 evidence for exaggerated activity of pro-fibrogenic pathways. The heterogeneity of
324 disease phenotype and longer time interval after acute illness at which our cohort were
325 sampled may account for this discrepancy. Of note, CMV-specific T cell clones were
326 enriched in individuals with fibrotic PCLD, consistent with the repeated association of
327 CMV with pulmonary fibrosis (Moore and Moore, 2015). However, this observation is
328 of uncertain significance, given the lack of existing evidence for a direct role for CMV
329 in causing human pulmonary fibrosis (Moore and Moore, 2015).

330 Our study has some limitations. We acknowledge a small sample size, heterogeneous
331 patient cohort, and lack of specimens for histological correlation and orthogonal

validation of our single-cell data. Our findings provide early mechanistic insights, are hypothesis-generating and will require further validation in larger cohorts. Those with fibrotic radiological appearances were sampled later after acute disease than those with radiological inflammation. Ideally, we would have more closely matched the interval after acute COVID-19 at which both groups were sampled. However, the introduction of dexamethasone treatment early in the second wave of the pandemic was coincident with reduced numbers of individuals with persistent respiratory symptoms and radiological abnormalities. Consequently, we extended the interval after acute illness within which individuals were eligible for sampling. Since our aim was to evaluate whether single-cell profiling of the bronchoalveolar environment of individuals with inflammatory and fibrotic radiological patterns of PCLD at the time of sampling would reveal different immunopathogenic mechanisms for these phenotypes, we believe our approach remains valid. It also provided invaluable opportunities to assess respiratory tract samples from the later stages of PCLD which have received minimal attention to date. This was a cross-sectional evaluation of bronchoalveolar immune cells, which did not allow assessment of the temporal evolution of the immune response in PCLD within individuals. Future studies encompassing longitudinal monitoring of the bronchoalveolar environment might provide important insights into the molecular mechanisms driving resolution and whether this trajectory is a universal phenomenon for both inflammation and fibrosis following COVID-19. Augmented neutrophil-associated immune signatures have been described in plasma and nasal samples in individuals with post-COVID pulmonary sequelae, however, neutrophils were not detected in our samples, possibly due to the Chromium 10x Genomics sample processing conditions used at the time of our

356 analysis (George et al., 2022). Finally, due to technical limitations because of the low
357 numbers of T cells present in BAL samples, we were unable to obtain TCR data for
358 two inflammatory cases. Nonetheless, our data provide strong support for the notion
359 that the TCR repertoire of PCLD reflects antigen-directed immune responses.

360 Our observations that inflammatory PCLD is characterised by airway T cell infiltration
361 and that antigen-specific T cell responses are evident in both radiological phenotypes,
362 highlight opportunities for early intervention with therapies targeting T cells.
363 Understanding the timing and duration of intervention, stratification of those at high
364 risk of irreversible tissue damage and the potential use of more targeted T cell
365 immunomodulators all merit further investigation.

366 METHODS

367 Ethics statement

368 The study was approved by the North London Research Ethics Committee
369 (13/LO/0900). Written informed consent was obtained from all participants. Subject
370 identifiers were not known to anyone outside the research group.

371 Study design and eligibility

372 Immune cells from the site of disease were obtained from adults (≥18 years)
373 undergoing bronchoscopy for clinical investigation of persistent respiratory symptoms
374 and CT abnormalities consistent with pulmonary inflammation (n=5) or fibrosis (n=5)
375 at least 12 weeks after acute COVID-19, confirmed by a positive SARS-CoV-2
376 antibody or polymerase chain reaction (PCR) test. PCLD was defined in this cohort
377 based on the following criteria: 1) new, persistent respiratory symptoms following
378 SARS-CoV-2 infection at least 12 weeks previously, 2) post-COVID-19 residual lung
379 abnormalities with more than 10% lung involvement on CT and 3) breathlessness in

380 keeping with the CT changes and not explained by other causes. Thoracic CT scans
381 were classified as predominantly inflammatory or fibrotic by consensus opinion of the
382 ILD multi-disciplinary team, which included thoracic radiologists with ILD expertise.
383 Radiological inflammation was defined as consolidation or ground glass opacities
384 without reticulation or parenchymal distortion and fibrosis defined as reticulation or
385 traction bronchiectasis. Individuals with evidence of ILD prior to COVID-19, those with
386 coincident malignancy, human immunodeficiency virus infection, bacterial, viral or
387 fungal respiratory tract infection, taking immunomodulatory therapy, or unable to give
388 informed consent were excluded.

389 **Isolation of bronchoalveolar cells**

390 Flexible fibreoptic bronchoscopy was used to obtain BAL samples by instillation of
391 180-240 ml of warmed normal saline into a lung segment affected by the predominant
392 radiological abnormality. Aspirated BAL fluid was cooled to 4°C and filtered through a
393 cell strainer to remove particulate debris before centrifugation. After removal of the
394 supernatant, cells were resuspended in PBS. Cell count and viability were determined
395 by Trypan blue staining and erythrocytes removed where indicated, using ammonium-
396 chloride-potassium red cell lysis buffer. Cells were resuspended at 2 x10^6 per ml for
397 immediate downstream processing.

398 **scRNAseq and scTCRseq library preparation and sequencing**

399 20,000 cells per sample were loaded on to the Chromium controller (10x Genomics)
400 to generate single-cell gel beads in emulsion (GEMs). Single-cell partitioning, reverse
401 transcription, cDNA amplification and library construction were performed using the
402 Chromium Single-cell 5' Reagent kits v1.1 and v2 (10x Genomics) according to the
403 manufacturer's instructions. T cell receptor (TCR) V(D)J segments were enriched from

404 amplified cDNA using Chromium Single-Cell V(D)J Enrichment kits v1.1 and v2 (10x
405 Genomics) per the manufacturer's protocol. Libraries were quality checked and
406 quantified using the High Sensitivity DNA kit and 4200 TapeStation (Agilent).
407 Sequencing was performed in paired end mode with SP100, P2 and P3 flow cells (100
408 cycles) using NovaSeq 6000 and NextSeq 2000 systems (Illumina).

409 **scRNAseq data analysis**

410 ***Cell Ranger***

411 Raw sequencing files were demultiplexed using BCL Convert v3.7.5 (Illumina) or Cell
412 Ranger version 6.1.1 using the "mkfastq" script. Transcript alignment and quantitation
413 against the GRCh38 human genome assembly was performed using Cell Ranger
414 "multi" for samples with gene expression and T cell VDJ data or Cell Ranger "count"
415 for samples with gene expression data only.

416 ***Quality control***

417 Initial processing of Cell Ranger output data was performed using Seurat v4.1.0 in R
418 4.1.1 (Stuart et al., 2019). Cell Ranger output files were loaded using the "Read10x"
419 function. Low quality cells with less than 200 or more than 6000 unique features or
420 more than 10% mitochondrial genes were removed; 57,712 cells were retained for
421 downstream analysis.

422 ***Normalization, feature selection and integration***

423 For each sample data were normalized and highly variable genes selected before
424 integration to remove donor-specific batch effects, using the following Seurat functions
425 implemented with default parameters; NormalizeData (normalization.method =
426 "LogNormalize), FindVariableFeatures (selection.method = "vst", nfeatures = 2000),
427 IntegrateData (anchorset = immune.anchors).

428 ***Dimensional reduction, clustering and annotation***

429 Data were then scaled using the Seurat ScaleData function and dimensional reduction
430 achieved by principal component analysis (PCA) of the most variable genes using
431 RunPCA. The first 25 principal components (PCs) were used to generate the Uniform
432 Manifold Approximation and Projection (UMAP) for two-dimensional visualization of
433 the cells (RunUMAP) and for nearest neighbor graph construction (FindNeighbors)
434 and Louvain clustering (FindClusters) using a resolution of 0.8. This clustering
435 resolution was selected on the basis that it successfully partitioned single-cell
436 transcriptomes into the broad cell types expected to be present in BAL. Automatic cell
437 type annotation of clusters was performed with the Semi-supervised Category
438 Identification and Assignment (SCINA) R package using gene signatures from the
439 Azimuth human lung v1 reference
440 (https://azimuth.hubmapconsortium.org/references/#Human-Lungv1) (Travaglini et
441 al., 2020; Zhang et al., 2019). Additional manual annotation of dendritic cells and B
442 cells was performed using literature-based markers (Cheng et al., 2021; Glass et al.,
443 2020; Villani et al., 2017). A population with low numbers of genes, which could not be
444 annotated by automatic or manual methods and likely represents empty droplets
445 containing ambient RNA, was removed after clustering. To validate our clustering and
446 annotation strategy, we visualized the expression of canonical or published marker
447 genes for each cell type using a dot plot (Adams et al., 2020; Cheng et al., 2021; David
448 et al., 2021; Davies et al., 2013; Davis and Wypych, 2021; Gardell and Parker, 2017;
449 Glass et al., 2020; Murray and Wynn, 2011; Tsyklauri et al., 2023; van Aalderen et al.,
450 2021; Villani et al., 2017).

451 ***T cell and macrophage re-clustering***

452 For separate analyses of T cells only and macrophages only, cells in the relevant
453 clusters were subsetted and normalization, variable feature selection, integration, data
454 scaling, PCA, UMAP generation and Louvain clustering repeated. The first 25 PCs
455 were used for data integration and clustering for both macrophages and T cells. The
456 k.weight parameter was reduced to 70 for the T cell integration step in order to take
457 account of the low number of cells in sample 12. Resolution 0.4 was used for re-
458 clustering T cells and resolution 0.7 for re-clustering macrophages. T cell subsets were
459 annotated using SCINA with Azimuth human PBMC reference marker genes
460 https://azimuth.hubmapconsortium.org/references/#Human-PBMC and additional
461 published CD4 T cell signatures (Cano-Gamez et al., 2020; Zhang et al., 2019).
462 Macrophages were annotated using SCINA with Azimuth human lung v1 reference
463 marker genes and additional signatures derived from published literature (Mould et al.,
464 2021; Travaglini et al., 2020; Wendisch et al., 2021; Zhang et al., 2019).

465 ***Calculation of gene module scores***

466 The Seurat function AddModuleScore was used to calculate enrichment of a cyclin D1
467 (CCND1) regulated module representing the cellular proliferation response (Chandran
468 et al., 2022) and a profibrotic macrophage gene module derived from idiopathic
469 pulmonary fibrosis (Adams et al., 2020). Module scores represent the average
470 expression of gene signatures of interest with subtraction of the average expression
471 of control gene sets randomly selected from each expression bin which contains a
472 module gene.

473 ***Differential gene expression and differential cell type abundance analysis***

474 For comparison of gene expression in individual cells (single-cell method) the scran R

475 package findMarkers function was implemented using the following settings: test.type
476 = "wilcox", direction = ‘‘up’’, pval.type = "all”, lfc = 0. For identification of cell type
477 (cluster) marker genes, sample was included as a blocking factor. Genes with an
478 adjusted p-value <0.05 were considered significantly differentially expressed. For
479 comparison of gene expression at cell type level (pseudobulk method) the scuttle R
480 package aggregateAcrossCells function was used to aggregate counts for each cell
481 type within each sample to create “pseudobulks” before statistical analysis was
482 performed. Differentially expressed genes were identified using the scran
483 pseudoBulkDGE function, implementing the edgeR negative binomial generalized
484 linear model with quasi-likelihood F test (GLM-QLF) (Robinson et al., 2010). Genes
485 with an adjusted p-value <0.05 were considered significant. Differences in cell type
486 abundance between inflammatory and fibrotic PCLD were identified using the edgeR
487 GLM-QLF test implemented by scran. Cell types with an adjusted p-value <0.05 were
488 considered differentially abundant.

489 ***Pathway enrichment analysis***

490 The biological pathways represented by differentially expressed genes were identified
491 by Reactome pathway enrichment analysis using XGR as previously described (Fang
492 et al., 2016; Turner et al., 2021). For visualization, 15 pathway groups were identified
493 by hierarchical clustering of Jaccard indices to quantify similarity between the gene
494 compositions of each pathway. For each group the pathway term with the largest
495 number of annotated genes was then selected as representative of the enriched
496 biology.

497 ***Upstream regulator analysis***

498 Ingenuity Pathway Analysis (Qiagen) was used to identify upstream transcriptional

499 regulation of differentially expressed genes. This analysis was restricted to molecules
500 annotated with the following functions: cytokine, growth factor, transmembrane
501 receptor, kinase and transcriptional regulator, representing the canonical components
502 of pathways which execute transcriptional reprogramming in immune and tissue repair
503 responses. Enriched molecules with an adjusted p-value <0.05 were considered
504 statistically significant. Area-proportional Venn diagrams visualizing the overlap
505 between molecules predicted to regulate cell type-specific differentially expressed
506 genes in inflammatory PCLD and fibrotic PCLD were generated using BioVenn
507 (Hulsen et al., 2008).

508 ***TCR quantitation and CDR3 clustering***

509 TCR sequences were assembled by the Cell Ranger multi pipeline (v6.1.1). For single-
510 cell TCR analysis, TCR clonotype abundance information was imported directly from
511 Cell Ranger "filtered_contig_annotations" output files, where clonotype identity was
512 determined as cells with identical V(D)J and CDR3 sequences. Clonotypes were
513 assigned to single cells using index barcodes. TCRs found more than once were
514 defined as expanded. Density plots were calculated using the UMAP coordinates of
515 every expanded cell with a detectable TCR using geom_density_2d from the ggplot2
516 package.

517 For clustering analysis of each PCLD phenotype individually, all detected alpha and
518 beta chain sequences in the inflammatory samples were included, to take account of
519 the sparsity of single-cell data. Including larger numbers of TCR sequences leads to
520 more clustering (Madi et al., 2017); fibrotic PCLD and control group TCR sequences
521 were therefore subsampled to match the number of sequences in the inflammatory
522 PCLD group, which contained the smallest repertoire. We included expanded

523 peripheral blood TCRs from individuals with non-severe SARS-CoV-2 infection as a
524 positive control dataset known to exhibit high levels of clustering and non-expanded
525 TCRs randomly selected from uninfected individuals from this same cohort, as a
526 negative control dataset, not expected to cluster highly (Chandran et al., 2022;
527 Milighetti et al., 2023). For the combined analysis of both PCLD phenotypes all
528 detected CDR3 sequences were included. TCR clustering was performed as
529 previously described (Joshi et al., 2019). Briefly, CDR3 amino acid sequences were
530 deconstructed into overlapping series of contiguous triplets. Pairwise similarity
531 between two CDR3s was calculated as the normalized string (triplet) kernel using the
532 Kernlab R package (Karatzoglou et al., 2004). The resulting TCR similarity matrix was
533 converted into a network diagram in which CDR3s with a pairwise similarity above a
534 designated threshold were connected by an edge using the iGraph R package (Csardi
535 and Nepusz, 2006). We visualized thresholds at which PCLD TCRs exhibited the
536 largest increase in the percentage of TCRs clustering compared to negative control
537 TCR data for clustering of PCLD groups individually. For combined clustering of
538 inflammatory and fibrotic PCLD we visualized thresholds at which the largest clusters
539 composed uniquely of one phenotype were present.

540 ***Virus specific TCR enrichment analysis***

541 TCRs annotated for SARS-CoV-2, CMV and EBV were obtained from the VDJdb
542 database (Bagaev et al., 2020) (https://vdjdb.cdr3.net/), accessed on 1st November
543 2021. The number of annotated sequences for each virus in VDJdb either matching
544 or not matching TCRs detected in inflammatory and fibrotic PCLD or the negative
545 controls was used to calculate the odds ratio (Fisher's exact test) for enrichment of
546 virus-specific TCRs in each group.

547 **Data availability**

548 scRNAseq and scTCRseq data will be available in the Gene Expression Omnibus
549 (GEO) database at the time of peer-reviewed publication of the manuscript under
550 accession number GSE228236.

551 **ACKNOWLEDGEMENTS**

552 We thank Dr Arjun Nair for oversight of the ILD MDT discussions to classify radiological
553 changes as predominantly inflammatory or fibrotic and for identification of lung
554 segments to be targeted for sampling by BAL.

555 **REFERENCES**

556 Adams TS, Schupp JC, Poli S, Ayaub EA, Neumark N, Ahangari F, Chu SG, Raby
557 BA, Deluliis G, Januszyk M, Duan Q, Arnett HA, Siddiqui A, Washko GR,
558 Homer R, Yan X, Rosas IO, Kaminski N. 2020. Single-cell RNA-seq reveals
559 ectopic and aberrant lung-resident cell populations in idiopathic pulmonary
560 fibrosis. *Sci Adv* **6**:eaba1983. doi:10.1126/sciadv.aba1983

561 Auffray C, Sieweke MH, Geissmann F. 2009. Blood monocytes: development,
562 heterogeneity, and relationship with dendritic cells. *Annu Rev Immunol*
563 **27**:669–692. doi:10.1146/annurev.immunol.021908.132557

564 Bagaev DV, Vroomans RMA, Samir J, Stervbo U, Rius C, Dolton G, Greenshields-
565 Watson A, Attaf M, Egorov ES, Zvyagin IV, Babel N, Cole DK, Godkin AJ,
566 Sewell AK, Kesmir C, Chudakov DM, Luciani F, Shugay M. 2020. VDJdb in
567 2019: database extension, new analysis infrastructure and a T-cell receptor
568 motif compendium. *Nucleic Acids Res* **48**:D1057–D1062.
569 doi:10.1093/nar/gkz874

570 Cano-Gamez E, Soskic B, Roumeliotis TI, So E, Smyth DJ, Baldrighi M, Willé D,
571 Nakic N, Esparza-Gordillo J, Larminie CGC, Bronson PG, Tough DF, Rowan
572 WC, Choudhary JS, Trynka G. 2020. Single-cell transcriptomics identifies an
573 effectorness gradient shaping the response of CD4+ T cells to cytokines. *Nat*
574 *Commun* **11**:1801. doi:10.1038/s41467-020-15543-y

575 Chandran A, Rosenheim J, Nageswaran G, Swadling L, Pollara G, Gupta RK, Burton
576 AR, Guerra-Assunção JA, Woolston A, Ronel T, Pade C, Gibbons JM, Sanz-
577 Magallon Duque De Estrada B, Robert de Massy M, Whelan M, Semper A,
578 Brooks T, Altmann DM, Boyton RJ, McKnight Á, Captur G, Manisty C, Treibel
579 TA, Moon JC, Tomlinson GS, Maini MK, Chain BM, Noursadeghi M,
580 COVIDsortium Investigators. 2022. Rapid synchronous type 1 IFN and virus-
581 specific T cell responses characterize first wave non-severe SARS-CoV-2
582 infections. *Cell Rep Med* **3**:100557. doi:10.1016/j.xcrm.2022.100557

583 Cheng S, Li Z, Gao R, Xing B, Gao Y, Yang Yu, Qin S, Zhang L, Ouyang H, Du P,

584 Jiang L, Zhang B, Yang Yue, Wang X, Ren X, Bei J-X, Hu X, Bu Z, Ji J,
585 Zhang Z. 2021. A pan-cancer single-cell transcriptional atlas of tumor
586 infiltrating myeloid cells. *Cell* **184**:792-809.e23. doi:10.1016/j.cell.2021.01.010

587 Cheon IS, Li C, Son YM, Goplen NP, Wu Y, Cassmann T, Wang Z, Wei X, Tang J, Li
588 Y, Marlow H, Hughes S, Hammel L, Cox TM, Goddery E, Ayasoufi K,
589 Weiskopf D, Boonyaratanakornkit J, Dong H, Li H, Chakraborty R, Johnson
590 AJ, Edell E, Taylor JJ, Kaplan MH, Sette A, Bartholmai BJ, Kern R, Vassallo
591 R, Sun J. 2021. Immune signatures underlying post-acute COVID-19 lung
592 sequelae. *Sci Immunol* **6**:eabk1741. doi:10.1126/sciimmunol.abk1741

593 Cheung CCL, Goh D, Lim X, Tien TZ, Lim JCT, Lee JN, Tan B, Tay ZEA, Wan WY,
594 Chen EX, Nerurkar SN, Loong S, Cheow PC, Chan CY, Koh YX, Tan TT,
595 Kalimuddin S, Tai WMD, Ng JL, Low JG-H, Yeong J, Lim KH. 2022. Residual
596 SARS-CoV-2 viral antigens detected in GI and hepatic tissues from five
597 recovered patients with COVID-19. *Gut* **71**:226–229. doi:10.1136/gutjnl-2021-
598 324280

599 Csardi G, Nepusz T. 2006. The igraph software package for complex network
600 research. *Complex Syst* **1695**:1–9.

601 David G, Willem C, Legrand N, Djaoud Z, Mérieau P, Walencik A, Guillaume T,
602 Gagne K, Chevallier P, Retière C. 2021. Deciphering the biology of KIR2DL3+
603 T lymphocytes that are associated to relapse in haploidentical HSCT. *Sci Rep*
604 **11**:15782. doi:10.1038/s41598-021-95245-7

605 Davies LC, Jenkins SJ, Allen JE, Taylor PR. 2013. Tissue-resident macrophages.
606 *Nat Immunol* **14**:986–995. doi:10.1038/ni.2705

607 Davis JD, Wypych TP. 2021. Cellular and functional heterogeneity of the airway
608 epithelium. *Mucosal Immunol* **14**:978–990. doi:10.1038/s41385-020-00370-7

609 Fabbri L, Moss S, Khan FA, Chi W, Xia J, Robinson K, Smyth AR, Jenkins G,
610 Stewart I. 2023. Parenchymal lung abnormalities following hospitalisation for
611 COVID-19 and viral pneumonitis: a systematic review and meta-analysis.
612 *Thorax* **78**:191–201. doi:10.1136/thoraxjnl-2021-218275

613 Fang H, Knezevic B, Burnham KL, Knight JC. 2016. XGR software for enhanced
614 interpretation of genomic summary data, illustrated by application to
615 immunological traits. *Genome Med* **8**:129. doi:10.1186/s13073-016-0384-y

616 Gagiannis D, Hackenbroch C, Bloch W, Zech F, Kirchhoff F, Djudjaj S, von Stillfried
617 S, Bülow R, Boor P, Steinestel K. 2023. Clinical, Imaging, and
618 Histopathological Features of Pulmonary Sequelae Following Mild COVID-19.
619 *Am J Respir Crit Care Med*. doi:10.1164/rccm.202302-0285LE

620 Gardell JL, Parker DC. 2017. CD40L is transferred to antigen-presenting B cells
621 during delivery of T-cell help. *Eur J Immunol* **47**:41–50.
622 doi:10.1002/eji.201646504

623 George PM, Reed A, Desai SR, Devaraj A, Faiez TS, Laverty S, Kanwal A, Esneau
624 C, Liu MKC, Kamal F, Man WD-C, Kaul S, Singh S, Lamb G, Faizi FK,
625 Schuliga M, Read J, Burgoyne T, Pinto AL, Micallef J, Bauwens E, Candiracci
626 J, Bougoussa M, Herzog M, Raman L, Ahmetaj-Shala B, Turville S, Aggarwal
627 A, Farne HA, Dalla Pria A, Aswani AD, Patella F, Borek WE, Mitchell JA,
628 Bartlett NW, Dokal A, Xu X-N, Kelleher P, Shah A, Singanayagam A. 2022. A
629 persistent neutrophil-associated immune signature characterizes post-COVID-

630 19 pulmonary sequelae. *Sci Transl Med* **14**:eabo5795.
631 doi:10.1126/scitranslmed.abo5795

632 Glass DR, Tsai AG, Oliveria JP, Hartmann FJ, Kimmey SC, Calderon AA, Borges L,
633 Glass MC, Wagar LE, Davis MM, Bendall SC. 2020. An Integrated Multi-omic
634 Single-Cell Atlas of Human B Cell Identity. *Immunity* **53**:217-232.e5.
635 doi:10.1016/j.immuni.2020.06.013

636 Hulsen T, de Vlieg J, Alkema W. 2008. BioVenn - a web application for the
637 comparison and visualization of biological lists using area-proportional Venn
638 diagrams. *BMC Genomics* **9**:488. doi:10.1186/1471-2164-9-488

639 Joshi K, de Massy MR, Ismail M, Reading JL, Uddin I, Woolston A, Hatipoglu E,
640 Oakes T, Rosenthal R, Peacock T, Ronel T, Noursadeghi M, Turati V,
641 Furness AJS, Georgiou A, Wong YNS, Ben Aissa A, Sunderland MW, Jamal-
642 Hanjani M, Veeriah S, Birkbak NJ, Wilson GA, Hiley CT, Ghorani E, Guerra-
643 Assunção JA, Herrero J, Enver T, Hadrup SR, Hackshaw A, Peggs KS,
644 McGranahan N, Swanton C, TRACERx consortium, Quezada SA, Chain B.
645 2019. Spatial heterogeneity of the T cell receptor repertoire reflects the
646 mutational landscape in lung cancer. *Nat Med* **25**:1549–1559.
647 doi:10.1038/s41591-019-0592-2

648 Karatzoglou A, Smola A, Hornik K, Zeileis A. 2004. kernlab - An S4 Package for
649 Kernel Methods in R. *J Stat Softw* **11**:1–20. doi:10.18637/jss.v011.i09

650 Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, Cheng L, Li J, Wang X, Wang F, Liu L,
651 Amit I, Zhang S, Zhang Z. 2020. Single-cell landscape of bronchoalveolar
652 immune cells in patients with COVID-19. *Nat Med* **26**:842–844.
653 doi:10.1038/s41591-020-0901-9

654 Madi A, Poran A, Shifrut E, Reich-Zeliger S, Greenstein E, Zaretsky I, Arnon T,
655 Laethem FV, Singer A, Lu J, Sun PD, Cohen IR, Friedman N. 2017. T cell
656 receptor repertoires of mice and humans are clustered in similarity networks
657 around conserved public CDR3 sequences. *eLife* **6**:e22057.
658 doi:10.7554/eLife.22057

659 Mehta P, Rosas IO, Singer M. 2022. Understanding post-COVID-19 interstitial lung
660 disease (ILD): a new fibroinflammatory disease entity. *Intensive Care Med*
661 **48**:1803–1806. doi:10.1007/s00134-022-06877-w

662 Merad M, Blish CA, Sallusto F, Iwasaki A. 2022. The immunology and
663 immunopathology of COVID-19. *Science* **375**:1122–1127.
664 doi:10.1126/science.abm8108

665 Milighetti M, Peng Y, Tan C, Mark M, Nageswaran G, Byrne S, Ronel T, Peacock T,
666 Mayer A, Chandran A, Rosenheim J, Whelan M, Yao X, Liu G, Felce SL,
667 Dong T, Mentzer AJ, Knight JC, Balloux F, Greenstein E, Reich-Zeliger S,
668 Pade C, Gibbons JM, Semper A, Brooks T, Otter A, Altmann DM, Boyton RJ,
669 Maini MK, McKnight A, Manisty C, Treibel TA, Moon JC, Noursadeghi M,
670 Chain B. 2023. Large clones of pre-existing T cells drive early immunity
671 against SARS-COV-2 and LCMV infection. *iScience* **26**.
672 doi:10.1016/j.isci.2023.106937

673 Moore BB, Moore TA. 2015. Viruses in Idiopathic Pulmonary Fibrosis. Etiology and
674 Exacerbation. *Ann Am Thorac Soc* **12 Suppl 2**:S186-192.
675 doi:10.1513/AnnalsATS.201502-088AW

676 Mould KJ, Moore CM, McManus SA, McCubbrey AL, McClendon JD, Griesmer CL,
677 Henson PM, Janssen WJ. 2021. Airspace Macrophages and Monocytes Exist
678 in Transcriptionally Distinct Subsets in Healthy Adults. *Am J Respir Crit Care*
679 *Med* **203**:946–956. doi:10.1164/rccm.202005-1989OC

680 Murray PJ, Wynn TA. 2011. Protective and pathogenic functions of macrophage
681 subsets. *Nat Rev Immunol* **11**:723–737. doi:10.1038/nri3073

682 Myall KJ, Mukherjee B, Castanheira AM, Lam JL, Benedetti G, Mak SM, Preston R,
683 Thillai M, Dewar A, Molyneaux PL, West AG. 2021. Persistent Post-COVID-19
684 Interstitial Lung Disease. An Observational Study of Corticosteroid Treatment.
685 *Ann Am Thorac Soc* **18**:799–806. doi:10.1513/AnnalsATS.202008-1002OC

686 Nouno T, Okamoto M, Ohnishi K, Kaieda S, Tominaga M, Zaizen Y, Ichiki M,
687 Momosaki S, Nakamura M, Fujimoto K, Fukuoka J, Shimizu S, Komohara Y,
688 Hoshino T. 2019. Elevation of pulmonary CD163 + and CD204 +
689 macrophages is associated with the clinical course of idiopathic pulmonary
690 fibrosis patients. *J Thorac Dis* **11**. doi:10.21037/jtd.2019.09.03

691 Nunez-Castilla J, Stebliankin V, Baral P, Balbin CA, Sobhan M, Cickovski T, Mondal
692 AM, Narasimhan G, Chapagain P, Mathee K, Siltberg-Liberles J. 2022.
693 Potential Autoimmunity Resulting from Molecular Mimicry between SARS-
694 CoV-2 Spike and Human Proteins. *Viruses* **14**:1415. doi:10.3390/v14071415

695 Ravaglia C, Doglioni C, Chilosi M, Piciucchi S, Dubini A, Rossi G, Pedica F, Puglisi
696 S, Donati L, Tomassetti S, Poletti V. 2022. Clinical, radiological and
697 pathological findings in patients with persistent lung disease following SARS-
698 CoV-2 infection. *Eur Respir J* **60**:2102411. doi:10.1183/13993003.02411-2021

699 Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for
700 differential expression analysis of digital gene expression data. *Bioinforma*
701 *Oxf Engl* **26**:139–140. doi:10.1093/bioinformatics/btp616

702 Squair JW, Gautier M, Kathe C, Anderson MA, James ND, Hutson TH, Hudelle R,
703 Qaiser T, Matson KJE, Barraud Q, Levine AJ, La Manno G, Skinnider MA,
704 Courtine G. 2021. Confronting false discoveries in single-cell differential
705 expression. *Nat Commun* **12**:5692. doi:10.1038/s41467-021-25960-2

706 Stein SR, Ramelli SC, Grazioli A, Chung J-Y, Singh M, Yinda CK, Winkler CW, Sun
707 J, Dickey JM, Ylaya K, Ko SH, Platt AP, Burbelo PD, Quezado M, Pittaluga S,
708 Purcell M, Munster VJ, Belinky F, Ramos-Benitez MJ, Boritz EA, Lach IA,
709 Herr DL, Rabin J, Saharia KK, Madathil RJ, Tabatabai A, Soherwardi S,
710 McCurdy MT, NIH COVID-19 Autopsy Consortium, Peterson KE, Cohen JI, de
711 Wit E, Vannella KM, Hewitt SM, Kleiner DE, Chertow DS. 2022. SARS-CoV-2
712 infection and persistence in the human body and brain at autopsy. *Nature*
713 **612**:758–763. doi:10.1038/s41586-022-05542-y

714 Stewart I, Jacob J, George PM, Molyneaux PL, Porter JC, Allen RJ, Aslani S, Baillie
715 JK, Barratt SL, Beirne P, Bianchi SM, Blaikley JF, Chalmers JD, Chambers
716 RC, Chaudhuri N, Coleman C, Collier G, Denneny EK, Docherty A, Elneima
717 O, Evans RA, Fabbri L, Gibbons MA, Gleeson FV, Gooptu B, Greening NJ,
718 Guillen Guio B, Hall IP, Hanley NA, Harris V, Harrison EM, Heightman M,
719 Hillman TE, Horsley A, Houchen-Wolloff L, Jarrold I, Johnson SR, Jones MG,
720 Khan F, Lawson R, Leavy O, Lone N, Marks M, McAuley H, Mehta P, Parekh
721 D, Piper Hanley K, Platé M, Pearl J, Poinasamy K, Quint JK, Raman B,

722 Richardson M, Rivera-Ortega P, Saunders L, Saunders R, Semple MG,
723 Sereno M, Shikotra A, Simpson AJ, Singapuri A, Smith DJ, Spears M,
724 Spencer LG, Stanel S, Thickett D, Thompson AAR, Thorpe M, Walsh SL,
725 Walker S, Weatherley ND, Weeks M, Wild JM, Wootton DG, Brightling CE, Ho
726 L-P, Wain LV, Jenkins RG. 2022. Residual Lung Abnormalities Following
727 COVID-19 Hospitalization: Interim Analysis of the UKILD Post-COVID Study.
728 *Am J Respir Crit Care Med*. doi:10.1164/rccm.202203-0564OC

729 Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y,
730 Stoeckius M, Smibert P, Satija R. 2019. Comprehensive Integration of Single-
731 Cell Data. *Cell* **177**:1888-1902.e21. doi:10.1016/j.cell.2019.05.031

732 The BAL Cooperative Group Steering Committee. 1990. Bronchoalveolar lavage
733 constituents in healthy individuals, idiopathic pulmonary fibrosis, and selected
734 comparison groups. *Am Rev Respir Dis* **141**:S169-202.
735 doi:10.1164/ajrccm/141.5_Pt_2.S169

736 Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit RV, Chang S, Conley
737 SD, Mori Y, Seita J, Berry GJ, Shrager JB, Metzger RJ, Kuo CS, Neff N,
738 Weissman IL, Quake SR, Krasnow MA. 2020. A molecular cell atlas of the
739 human lung from single-cell RNA sequencing. *Nature* **587**:619–625.
740 doi:10.1038/s41586-020-2922-4

741 Tsyklauri O, Chadimova T, Niederlova V, Kovarova J, Michalik J, Malatova I,
742 Janusova S, Ivashchenko O, Rossez H, Drobek A, Vecerova H, Galati V,
743 Kovar M, Stepanek O. 2023. Regulatory T cells suppress the formation of
744 potent KLRK1 and IL-7R expressing effector CD8 T cells by limiting IL-2. *eLife*
745 **12**:e79342. doi:10.7554/eLife.79342

746 Turner CT, Brown J, Shaw E, Uddin I, Tsaliki E, Roe JK, Pollara G, Sun Y, Heather
747 JM, Lipman M, Chain B, Noursadeghi M. 2021. Persistent T Cell Repertoire
748 Perturbation and T Cell Activation in HIV After Long Term Treatment. *Front*
749 *Immunol* **12**:634489. doi:10.3389/fimmu.2021.634489

750 van Aalderen MC, van Lier RAW, Hombrink P. 2021. How to Reliably Define Human
751 CD8+ T-Cell Subsets: Markers Playing Tricks. *Cold Spring Harb Perspect Biol*
752 **13**:a037747. doi:10.1101/cshperspect.a037747

753 Vijayakumar B, Boustani K, Ogger PP, Papadaki A, Tonkin J, Orton CM, Ghai P,
754 Suveizdyte K, Hewitt RJ, Desai SR, Devaraj A, Snelgrove RJ, Molyneaux PL,
755 Garner JL, Peters JE, Shah PL, Lloyd CM, Harker JA. 2022. Immuno-
756 proteomic profiling reveals aberrant immune cell regulation in the airways of
757 individuals with ongoing post-COVID-19 respiratory disease. *Immunity*
758 **55**:542-556.e5. doi:10.1016/j.immuni.2022.01.017

759 Villani A-C, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, Griesbeck M,
760 Butler A, Zheng S, Lazo S, Jardine L, Dixon D, Stephenson E, Nilsson E,
761 Grundberg I, McDonald D, Filby A, Li W, De Jager PL, Rozenblatt-Rosen O,
762 Lane AA, Haniffa M, Regev A, Hacohen N. 2017. Single-cell RNA-seq reveals
763 new types of human blood dendritic cells, monocytes, and progenitors.
764 *Science* **356**:eaah4573. doi:10.1126/science.aah4573

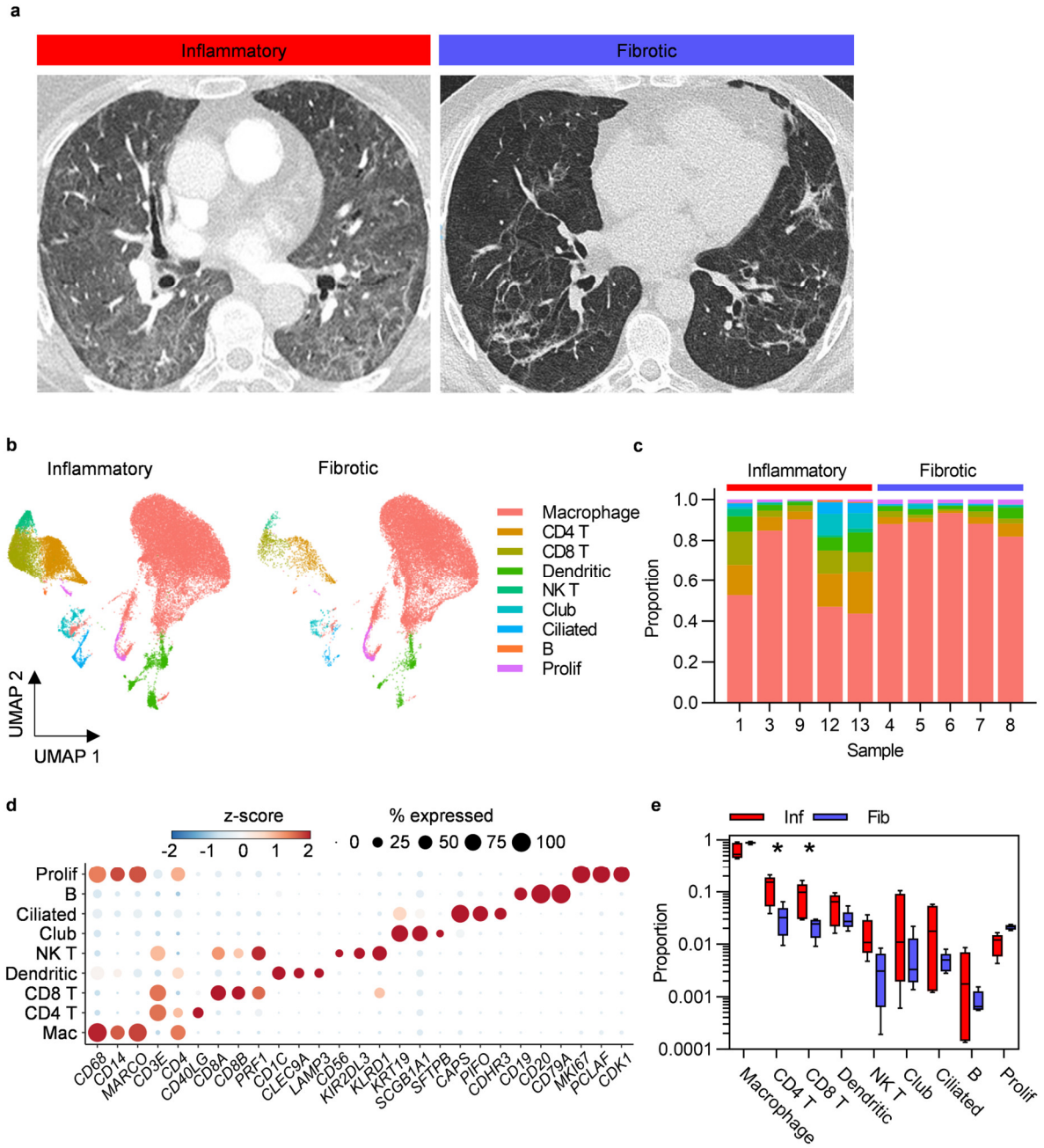
765 Vojdani A, Kharrazian D. 2020. Potential antigenic cross-reactivity between SARS-
766 CoV-2 and human tissue with a possible link to an increase in autoimmune
767 diseases. *Clin Immunol* **217**:108480. doi:10.1016/j.clim.2020.108480

Wauters E, Van Mol P, Garg AD, Jansen S, Van Herck Y, Vanderbeke L, Bassez A, Boeckx B, Malengier-Devlies B, Timmerman A, Van Brussel T, Van Buyten T, Schepers R, Heylen E, Dauwe D, Dooms C, Gunst J, Hermans G, Meersseman P, Testelmans D, Yserbyt J, Tejpar S, De Wever W, Matthys P, Neyts J, Wauters J, Qian J, Lambrechts D. 2021. Discriminating mild from critical COVID-19 by innate and adaptive immune single-cell profiling of bronchoalveolar lavages. *Cell Res* **31**:272–290. doi:10.1038/s41422-020-00455-9

Wendisch D, Dietrich O, Mari T, von Stillfried S, Ibarra IL, Mittermaier M, Mache C, Chua RL, Knoll R, Timm S, Brumhard S, Krammer T, Zauber H, Hiller AL, Pascual-Reguant A, Mothes R, Bülow RD, Schulze J, Leipold AM, Djudjaj S, Erhard F, Geffers R, Pott F, Kazmierski J, Radke J, Pergantis P, Baßler K, Conrad Claudia, Aschenbrenner AC, Sawitzki B, Landthaler M, Wyler E, Horst D, Deutsche COVID-19 OMICS Initiative (DeCOI), Hippenstiel S, Hocke A, Heppner FL, Uhrig A, Garcia C, Machleidt F, Herold S, Elezkurtaj S, Thibeault C, Witzenrath M, Cochain C, Suttorp N, Drosten C, Goffinet C, Kurth F, Schultze JL, Radbruch H, Ochs M, Eils R, Müller-Redetzky H, Hauser AE, Luecken MD, Theis FJ, Conrad Christian, Wolff T, Boor P, Selbach M, Saliba A-E, Sander LE. 2021. SARS-CoV-2 infection triggers profibrotic macrophage responses and lung fibrosis. *Cell* **184**:6243-6261.e27. doi:10.1016/j.cell.2021.11.033

Zhang Z, Luo D, Zhong X, Choi JH, Ma Y, Wang S, Mahrt E, Guo W, Stawiski EW, Modrusan Z, Seshagiri S, Kapur P, Hon GC, Brugarolas J, Wang T. 2019. SCINA: A Semi-Supervised Subtyping Algorithm of Single Cells and Bulk Samples. *Genes* **10**:531. doi:10.3390/genes10070531

794 **FIGURES**



795

796 **Figure 1.** Higher abundance of bronchoalveolar T cells in inflammatory post-COVID-
797 19 lung disease (PCLD). **a** Representative computed tomography (CT) images for
798 each radiological phenotype. **b** Uniform manifold approximation and projection
799 (UMAP) embedding of 55,776 bronchoalveolar single-cell transcriptomes obtained
800 from five individuals with radiological features of pulmonary inflammation (33,553

801 cells) and five individuals with radiological features of pulmonary fibrosis (22,223 cells)
802 following COVID-19, split by PCLD phenotype, color coded by cell type. Cell type
803 annotation was achieved by assignment of Azimuth human lung reference gene
804 signatures using the SCINA R package and in the case of dendritic cells and B cells,
805 expression of literature-based markers. Prolif = proliferating cells, identified by their
806 high module score for a gene signature representing the cellular proliferation
807 response. **c** Cellular composition of each BAL sample defined by single-cell RNA
808 sequencing (scRNAseq). Color indicates cell type and bar height represents
809 proportion. **d** Dot plot visualization of the expression of independently established
810 marker genes for each cell type; “Mac” = macrophage. Dot size represents the
811 percentage of cells expressing the gene in each cell type, color shows the z-scores of
812 average log-normalized expression for each cell type compared to the entire data set.
813 **e** Comparison of the proportions of each cell type in inflammatory and fibrotic PCLD.
814 Horizontal lines indicate median, box limits the interquartile range and whiskers the 5th
815 to 95th percentiles, * = FDR<0.05.
816

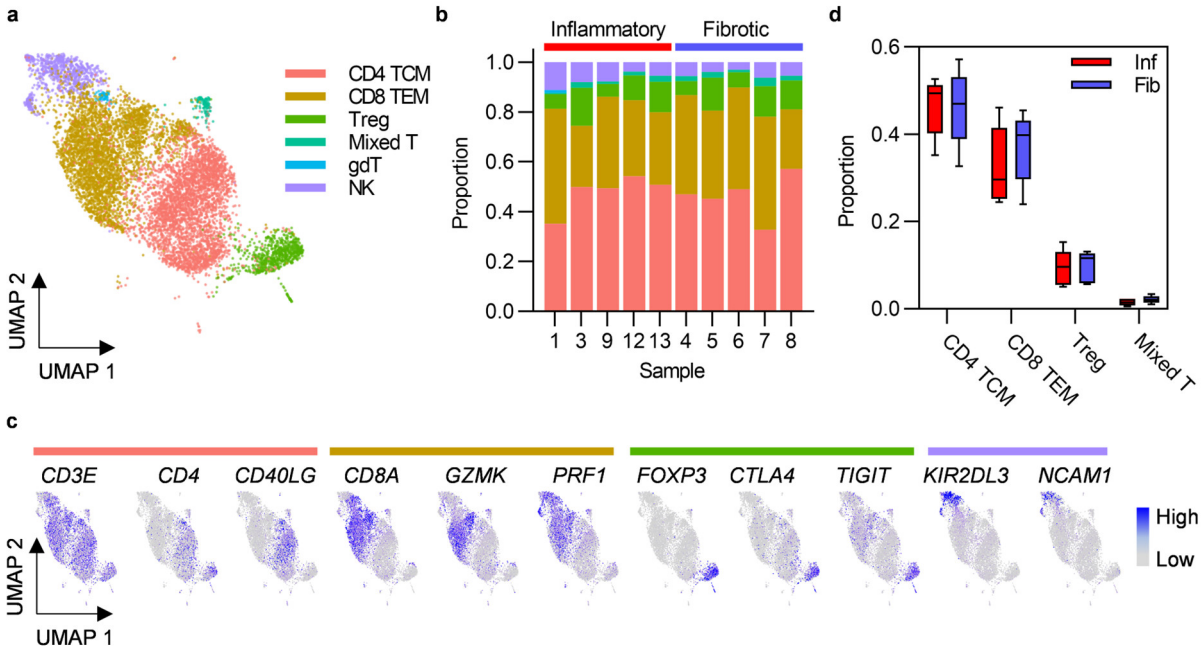


Figure 2. Post-COVID-19 bronchoalveolar T cells are dominated by CD4 central memory and CD8 effector memory subsets. **a** UMAP embedding of 9196 transcriptomes of T cells identified in Figure 1**b**, colored by cell type. T cell subset annotation was based on assignment of Azimuth human PBMC reference marker gene signatures and additional published CD4 T cell signatures by the SCINA R package. **b** Relative proportions of each T cell subset in each subject, colors represent cell type. **c** Feature plots demonstrating the expression of marker genes for the three principal T cell subsets and NK cells, colored by scaled, log-normalized counts, projected on to the T cell UMAP. **d** Comparison of the proportions of T cell subsets found in all individuals in inflammatory and fibrotic post-COVID-19 lung disease (PCLD). Horizontal lines indicate median, box limits the interquartile range and whiskers the 5th to 95th percentiles.

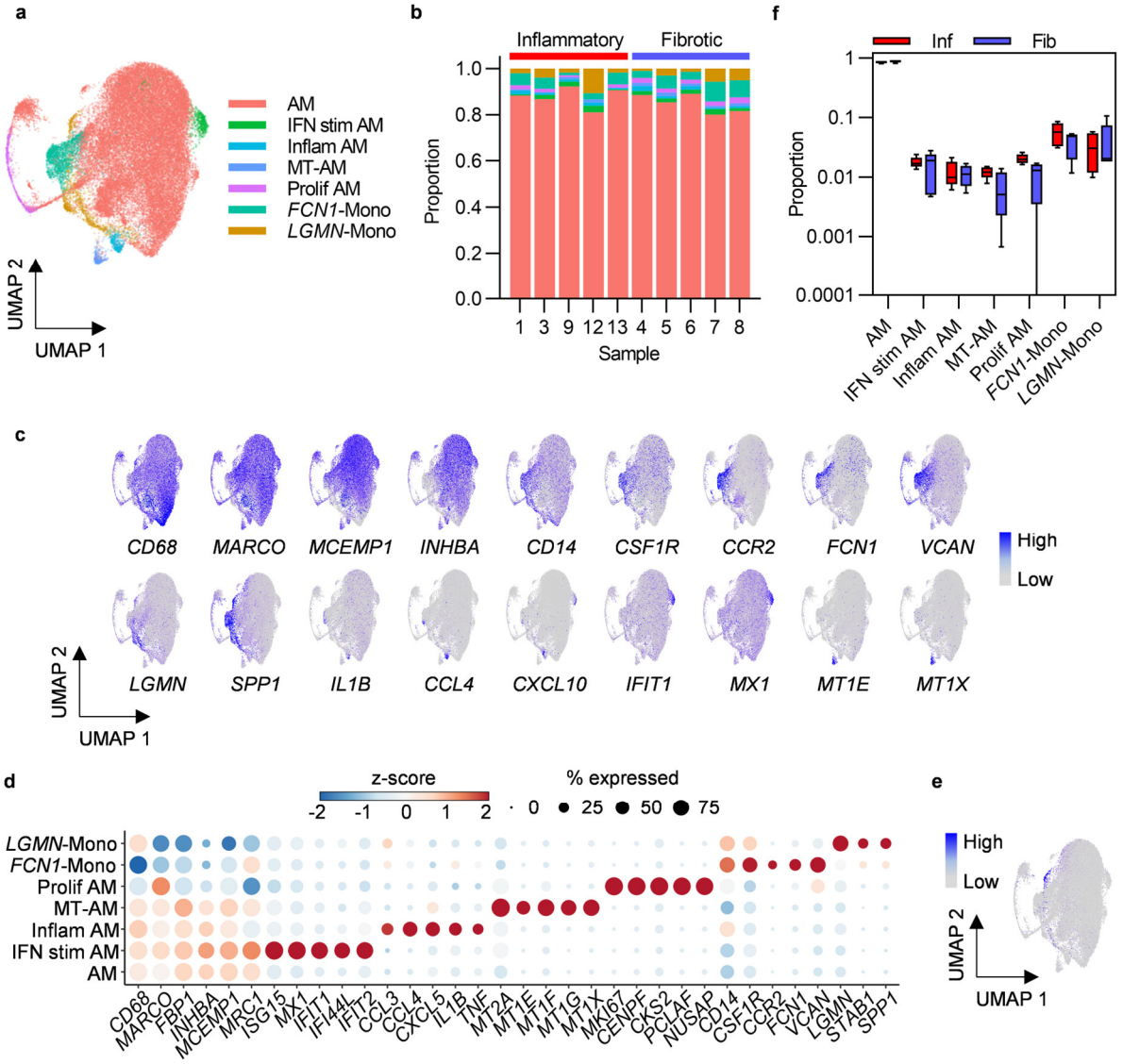


Figure 3. Bronchoalveolar macrophage and monocyte subsets in post-COVID-19 lung disease (PCLD). **a** UMAP embedding of 38,010 macrophage transcriptomes identified in Figure 1**b**, colored by cell type. Myeloid populations were annotated by assignment of Azimuth human lung reference marker genes and additional signatures derived from published literature using the SCINA R package and by visualising expression of canonical marker genes for some subsets. **b** Relative proportions of each myeloid subset within each individual, colors represent cell type. **c** Expression of marker genes for PCLD macrophage and monocyte populations, colored by scaled, log-normalized

840 counts, projected on to the myeloid cell UMAP. **d** Dot plot visualization of the
841 expression of selected marker genes for each macrophage and monocyte subset. Dot
842 size represents the percentage of cells expressing the gene in each myeloid subset,
843 color shows the z-scores of average log-normalized expression for each subset
844 compared to the entire data set. **e** Expression of a profibrotic macrophage gene
845 signature derived from idiopathic pulmonary fibrosis, calculated on a single-cell level,
846 colored by module score and projected on to the macrophage UMAP. **f** Comparison
847 of the proportions of bronchoalveolar myeloid populations in inflammatory and fibrotic
848 PCLD. Horizontal lines indicate median, box limits the interquartile range and whiskers
849 the 5th to 95th percentiles.
850

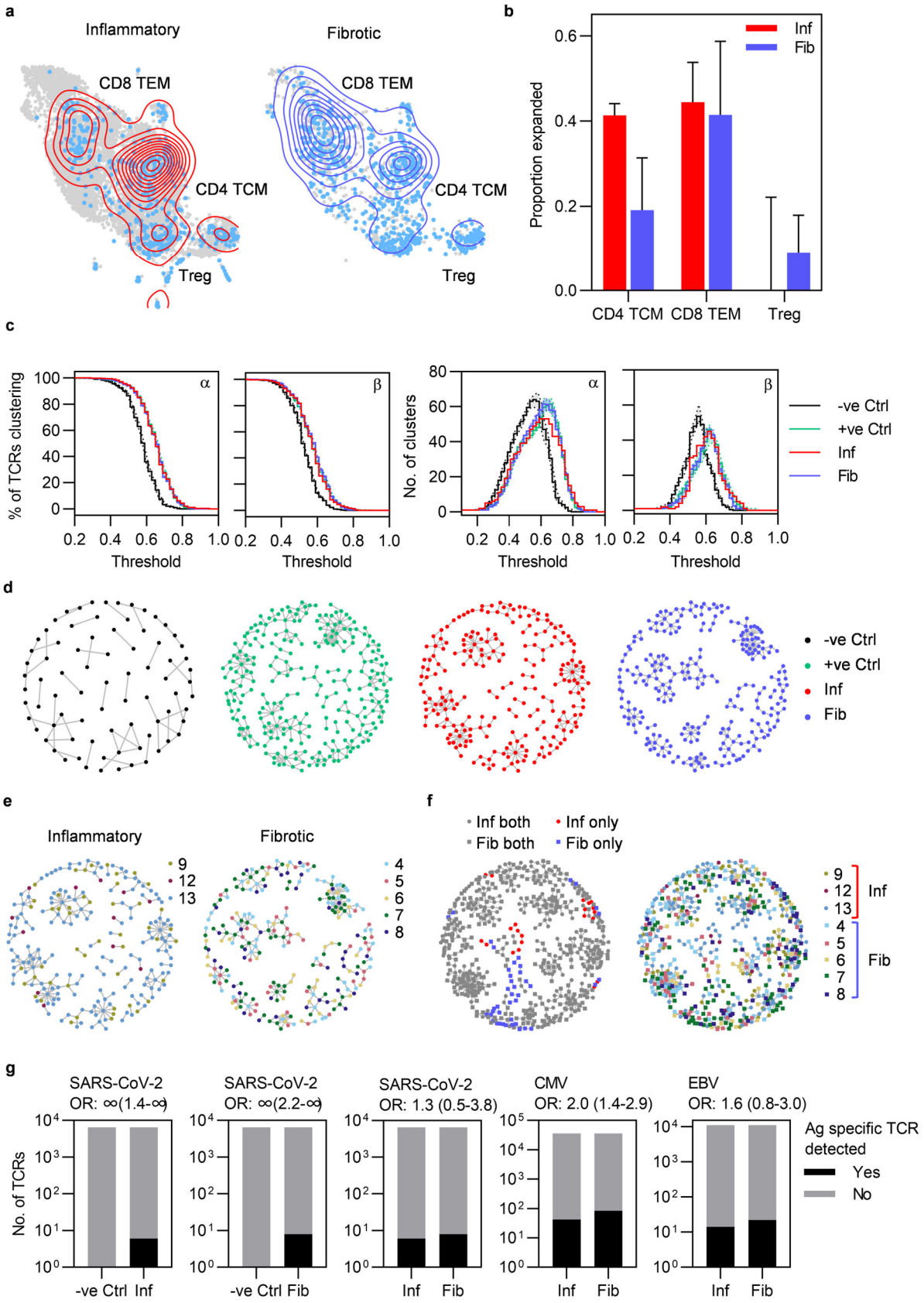


Figure 4. Highly similar T cell receptors (TCRs) characterize both inflammatory and

853 fibrotic post-COVID lung disease (PCLD). **a** T cell clonal expansion is visualized on
854 the T cell UMAP split by radiological PCLD phenotype. TCR sequences detected at a
855 frequency of greater than one are colored light blue and contour lines provide a 2D
856 representation of TCR density overlaid in red for inflammatory PCLD and blue for
857 fibrotic PCLD. **b** Comparison of the proportion of expanded TCR sequences, defined
858 as those detected at a frequency greater than one in the three largest T cell subsets
859 identified in Figure 2**a** in the two PCLD phenotypes. **c** Percentage of complementarity
860 determining region (CDR)3 alpha and beta chain amino acid sequences clustering and
861 number of clusters generated over a range of thresholds above which two TCRs are
862 considered similar, for inflammatory and fibrotic PCLD bronchoalveolar lavage (BAL)
863 samples, negative control PBMC samples not expected to cluster highly and positive
864 control PBMC samples known to cluster highly, analyzed separately. **d** Representative
865 network diagrams of TCR β chain clusters present in each group described in (**c**).
866 Nodes represent TCRs, related TCRs are connected by an edge and colors represent
867 the groups. **e** Network diagrams of related TCR β chains in each PCLD phenotype in
868 which nodes are colored by donor. **f** Network diagrams visualizing TCR β chain
869 clusters identified by combined analysis of the two PCLD phenotypes. Nodes are
870 colored by donor (right) or radiological phenotype and membership of clusters
871 composed of one or both PCLD groups (left), circular nodes represent inflammatory
872 PCLD and square nodes fibrotic PCLD. **g** Number of TCR sequences (α and β genes
873 combined) annotated for SARS-CoV-2, cytomegalovirus (CMV), and Epstein-Barr
874 virus (EBV) in VDJdb either detected or not detected in TCR sequences from
875 individuals with each PCLD phenotype and negative controls, giving the odds ratio
876 ±95% confidence interval (Fisher's exact test) for enrichment of antigen-specific TCR

877 sequences in each instance.

878

879 **TABLES**

880 **Table 1.** Clinical and demographic information summary.

	Inflammatory N=5	**Fibrotic N=5**
Age		
Median (IQR)	62 (21.5)	59 (8.5)
Sex		
Male	2 (40%)	4 (80%)
Female	3 (60%)	1 (20%)
Ethnicity		
White	3 (60%)	3 (60%)
Asian	2 (40%)	2 (40%)
Body mass index (kg/m^2)		
Median (IQR)	31.8 (15.6)	27.3 (6.1)
Smoking Status		
Never	4 (80%)	2 (40%)
Former	1 (20%)	3 (60%)
COVID-19 wave		
1	2 (40%)	2 (40%)
2	3 (60%)	3 (60%)
Respiratory Support		
I&V†	2* (40%)	4 (80%)
CPAP‡	1 (20%)	0
HFNO‡	1 (20%)	1 (20%)
nil	1 (20%)	0
Treatment for COVID-19		
Steroid	3 (60%)	5 (100%)
Tocilizumab	1 (20%)	1 (20%)
Anti-viral	0	0
BAL		
Days post-acute COVID (median (IQR))	116 (50)	316 (50.5)

881 *One patient also received ECMO (extracorporeal membrane oxygenation); †I&V
882 Intubation and ventilation; ‡CPAP Continuous Positive Airways Pressure; ‡HFNO
883 High Flow Nasal Oxygen.

884 **Support statement**

885 PM is supported by a Medical Research Council (MRC)/GlaxoSmithKline
886 Experimental Medicine Initiative to Explore New Therapies network (EMINENT)
887 Clinical Research Training Fellowship. EKD is supported by Breathing Matters. MZN
888 is supported by a UK MRC Clinician Scientist Fellowship (MR/W00111X/1) and a
889 Rutherford Fund Fellowship allocated by the MRC UK Regenerative Medicine Platform
890 2 (MR/5005579/1). MN is supported by the Wellcome Trust (207511/Z/17/Z). GST is
891 supported by a UK MRC Clinician Scientist Fellowship (MR/N007727/1). PM, EKD,
892 MP, JSB, MZN, MN, RCC, JCP and GST are also supported by NIHR Biomedical
893 Research Funding to UCL and UCLH.

894 **Author contributions**

895 GST, JCP, RCC and MZN conceived the study.

896 EKD, PM, KBW, MY, JCP and GST collected the data.

897 PM, EKD, KW, MY 10x Genomics platform.

898 BSMDDE, CTT, KF, MM, MZN, MP, JSB, BMC and MN provided resources or analysis
899 tools.

900 BSMDDE, KF and GST performed the computational data analysis with contributions
901 from AM, MM, BMC and MN.

902 PM, BSMDDE, EKD and GST prepared the manuscript draft.

903 All authors reviewed and contributed to the final manuscript.

904 **COMPETING INTERESTS**

905 All authors declare no competing interests.

906