Title

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

Running title

Single-cell landscape of 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], Arjun Nair[4], 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
4. Department of Radiology, University College London Hospitals NHS Foundation Trust, 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.

Author contributions

Conceived the study: GT, JCP, RCC, MZN

Collected data: EKD, PM, KBW, MY, AN, JCP, GT

Provided resources or analysis tools: BSMDDE, CT, KF, MM, MZN, MP, JSB, BMC, MN

Data analysis: BSMDDE, KF, AM, MN, BMC, GT

Prepared the manuscript draft: GT, BSMDDE, EKD

Reviewed and contributed to the final manuscript: All authors

Funding

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

Total word count: 3291

Abstract word count: 246

Scientific knowledge on the subject

The immune pathogenesis of persistent pulmonary radiological abnormalities following COVID-19 is poorly understood. Whether post-COVID radiological inflammation and fibrosis represent distinct disease entities with different molecular

mechanisms of tissue injury is not known.

What this study adds to the field

Single-cell bronchoalveolar transcriptomes of inflammatory and fibrotic post-COVID lung disease closely resemble each other across all cell types, but CD4 central memory and CD8 effector memory T cells are more abundant in the inflammatory group. Marked T cell receptor clustering, suggestive of antigen-specific responses is evident in both phenotypes. These two radiological patterns likely represent different manifestations of the same disease process, which may benefit from therapies which target T cells.

ABSTRACT

Rationale: Persistent pulmonary sequelae are evident in many survivors of acute coronavirus disease 2019 (COVID-19) but the molecular mechanisms responsible are incompletely understood. Post-COVID radiological lung abnormalities comprise two broad categories, organising pneumonia and reticulation, interpreted as indicative of subacute inflammation and fibrosis, respectively. Whether these two patterns represent distinct pathologies, likely to require different treatment strategies is not known.

Objectives: 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 of post-COVID lung disease (PCLD).

Measurements and Main Results: Inflammatory and fibrotic PCLD single-cell transcriptomes closely resembled each other across all cell types. However, CD4 central memory T cells (TCM) and CD8 effector memory T cells (TEM) were significantly more abundant in inflammatory PCLD. A greater proportion of CD4 TCM also exhibited clonal expansion in inflammatory PCLD. High levels of clustering of similar TCRs from multiple donors was a striking feature of both PCLD phenotypes, consistent with tissue localised antigen-specific immune responses, but there was no enrichment for known SARS-CoV-2 reactive TCRs.

Conclusions: There is no evidence that radiographic organising pneumonia and reticulation in PCLD are associated with differential immmunopathological pathways. Inflammatory radiology is characterised by greater bronchoalveolar T cell

accumulation. Both groups show evidence of shared antigen-specific T cell responses, but the antigenic target for these T cells remains to be identified.

Word count: 246

KEY WORDS

Inflammation, fibrosis, single-cell RNA sequencing, T cell receptor, SARS-CoV-2

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) (1). Current understanding of the immunopathogenic mechanisms responsible for post-COVID lung disease (PCLD) is very limited (2). Elevated numbers of airway CD4 and CD8 T cells have been reported (3, 4) and the post-COVID airway proteome displays evidence of ongoing epithelial injury that resolves with time (4). 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 are two major radiological patterns in PCLD: organizing pneumonia and ground glass opacities, thought to represent subacute inflammation, and reticulation, widely interpreted as fibrosis (1, 5). 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 (1). We hypothesised that these radiological changes reflect heterogeneity in 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 immune cells from convalescent individuals with predominant CT features of inflammation or fibrosis, infected during the first or second waves of the pandemic.

We showed that in comparison to fibrotic PCLD, the bronchoalveolar environment of

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

METHODS

Ethics statement

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

Study design

Immune cells from the site of disease were obtained from individuals undergoing bronchoscopy for clinical investigation of persistent respiratory symptoms and radiological abnormalities consistent with post-COVID-19 pulmonary inflammation (n=5) or fibrosis (n=5) 3-12 months after the acute illness. Inclusion criteria are provided in the online data supplement. Thoracic CT scans were classified as inflammatory or fibrotic by consensus opinion of the interstitial lung disease (ILD) multi-disciplinary team, which included thoracic radiologists with ILD expertise.

Isolation of bronchoalveolar cells

Flexible fibreoptic bronchoscopy was used to obtain bronchoalveolar lavage (BAL) samples by instillation of up to 240 ml of warmed normal saline into a lung segment affected by the predominant radiological abnormality. A detailed description of the cell isolation procedure and processing for single-cell sequencing are provided in the online data supplement.

Data analysis

Details of the computational analysis of scRNAseq and scTCRseq data are given in the online data supplement. Briefly, the Seurat R package was used to perform quality control, normalization, integration, clustering and two-dimensional visualisation of single-cell transcriptomes (6). Automatic cell annotation was performed with the Semi-supervised Category Identification and Assignment (SCINA) R package and literature-based markers were used for manual annotation (7–10). The scran R package was used to identify differentially expressed genes and differentially abundant cell populations. Pathway enrichment analysis and upstream regulator analysis of differentially expressed genes and TCR clustering were performed as previously described (11–13). TCRs annotated for SARS-CoV-2, cytomegalovirus (CMV) and Epstein-Barr virus (EBV) were obtained from the VDJdb database (14).

Data availability

scRNAseq and scTCRseq data will be available in the Gene Expression Omnibus (GEO) database at the time of peer-reviewed publication of the manuscript.

RESULTS

Increased abundance of bronchoalveolar T cells in inflammatory PCLD

Individuals undergoing bronchoscopy for clinical investigation of persistent respiratory symptoms and predominant radiological features of either inflammation or fibrosis following acute COVID-19 were recruited (Figure 1A, B). BAL samples were obtained from five subjects with each radiological pattern (Figure 1B). Clinical and demographic characteristics are provided in Table 1 and E1. Evaluation of the composition of PCLD BAL by scRNAseq revealed that macrophages dominated in all subjects, with smaller T cell, NK T cell, dendritic cell, epithelial and B cell populations also identified (Figure 1C-E and E1A-E). Detected cell types expressed high levels of established marker genes (8–10, 15–22), validating our annotation strategy (Figure 1E). The notable difference between the two phenotypes was significantly higher abundance of CD4 T cells and CD8 T cells in inflammatory PCLD (Figure 1F).

To explore differences in cell type-specific transcriptional profiles between the two radiological phenotypes we first aggregated gene expression count data for each cell type for each donor to form “pseudobulks”. We leveraged the ability of pseudobulk statistical approaches to account for variability of biological replicates, allowing detection of genuine differential gene expression whilst minimizing false discoveries (23). Very few cell type-specific differentially expressed genes were identified (Table E2), which precluded further bioinformatic analysis and suggested the transcriptomes of inflammatory and fibrotic PCLD were similar for all cell types.

Detection of transcriptomic differences between the two groups using pseudobulk statistical analysis may have been limited by small sample size. Therefore, we repeated the comparison of inflammatory and fibrotic PCLD at the level of individual

cells. Single-cell statistical analysis identified thousands of cell type-specific differentially expressed genes in each radiological phenotype. However, single-cell differential gene expression analysis has a propensity for false positive results (23). To mitigate against this, we sought to assess whether differentially expressed gene lists represented differentially enriched biological pathways in the two study groups. This was not evident at the level of biological processes or upstream regulator analysis (Figure E2-E3), which suggested enrichment of overlapping processes and pathways despite apparent differential gene expression. Hence, our single-cell analysis also supports the outcome of the pseudobulk analysis.

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

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

Analysis of T cell subset pseudobulks revealed no genes as differentially expressed in either group (Table E2), consolidating our earlier observation of few differences between the transcriptional programmes of the two PCLD phenotypes at the level of broad cell types defined using the full dataset. In single-cell differential expression

analysis of T cells, genes expressed at a significantly higher level in CD4 TCM and CD8 TEM in inflammatory PCLD exhibited weak enrichment for immune response signalling pathways (Figure E4A). Robust enrichment of cellular pathways in anti-viral responses was evident for genes expressed at a significantly higher level in CD4 TCM and CD8 TEM in fibrotic PCLD (Figure E4B), but this was not supported by enrichment for Type I interferon signaling in upstream regulator analysis (Figure E5B). Interestingly, interleukin (IL)2 and IL15 were the most statistically significant upstream regulators of CD4 TCM and CD8 TEM differentially expressed genes in inflammatory PCLD but not fibrotic PCLD, consistent with cytokine driven proliferation leading to increased abundance of T cells in the former phenotype (Figure E5A-C).

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

Macrophages constituted the predominant cell type in PCLD BAL samples and have been implicated in the pathogenesis of fibrosis associated with severe COVID-19 infection and other ILDs (24, 25). We therefore reclustered macrophages to further dissect their potential role in immune pathogenesis of PCLD. Two populations with similar expression of macrophage markers and comparable transcriptomic profiles, representantive of the transcriptional spectrum of resident healthy alveolar macrophages AM (26) were combined for subsequent analysis (Figure 3A,B and E6A,B). We identified three small specialized AM subsets detected in healthy individuals (26), characterized by high levels of proinflammatory 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 (27) (Figure 3A-E).

Monocyte-like cells are present in healthy airspaces, and suggested by trajectory inference to differentiate into AM, implying constant trafficking of monocytes into the lung (26). Consistent with this, we identified CD14, Ficolin 1 (FCN1) expressing classical monocytes (FCN1-Mono) (26, 28) characterized by high levels of CCR2, the receptor for monocyte chemoattractant protein, suggestive of recent recruitment from the peripheral blood (29) (Figure 3D,E). A second matrix-associated legumain (LGMN), osteopontin (SPP1) expressing subset (LGMN-Mono) which exists as a rare population in healthy airspaces (26), was also present (Figure 3D,E). In contrast to acute severe COVID-19 where proinflammatory monocytes and profibrotic SPP1, LGMN expressing macrophages are abundant (24, 30), monocytes represented a minor constituent of PCLD BAL. Neither subset expressed high levels of inflammatory mediators (Figure 3D,E) and very few cells expressed a gene signature characteristic of profibrotic macrophages in idiopathic pulmonary fibrosis (21), suggesting little evidence of exaggerated fibrotic activity (Figure 3F and Figure E7A).

No differences in the relative proportions of any of the myeloid populations were evident between the two PCLD phenotypes (Figure 3G). Very few gene expression differences were detected between the two groups by pseudobulk analysis of macrophage or monocyte subsets (Table E2). Similar to the analysis of the full dataset stratified by broad cell type, reclustering macrophages to identify more discrete cell subtypes did not reveal transcriptomic differences between inflammatory and fibrotic PCLD.

Single-cell differential gene expression analysis revealed weak enrichment for immune response and cellular metabolism pathways in both groups (Figure E7B,C). Minor enrichment for TGFβ-mediated signaling was evident for genes expressed at

higher levels in AM, Inflam-AM and FCN1-Mono in inflammatory PCLD (Figure E7B). Proinflammatory cytokines, T cell activation factors, osteopontin and TGFβ were identified as more statistically enriched upstream regulators of differentially expressed genes in both monocyte populations in inflammatory PCLD (Figure E8A,B,D,E). However, as for analysis at the level of broad cell types and refined T cell subsets, there was considerable overlap between molecules predicted to drive gene expression differences in each PCLD phenotype (Figure E8C,F), suggesting that between group differential gene expression in this analysis did not represent differential biology between radiological patterns of disease.

Highly related TCRs indicate antigen-specific immune responses in PCLD

To further evaluate T cell responses in PCLD we undertook single-cell TCR sequencing (scTCRseq) and compared the TCR repertoire in each radiological phenotype. We successfully obtained scTCRseq data for five fibrotic and three inflammatory cases. Increased T cell abundance in inflammatory PCLD may reflect increased numbers of unique T cell clones or increased expansion of individual clones. Expanded clonotypes present with a frequency of greater than one were evident within the three major T cell subsets in both inflammatory and fibrotic PCLD (Figure 4A). A greater proportion of CD4 TCM clones were expanded in the inflammatory group, consistent with our observation of its higher abundance of this T cell subset. However, the proportion of expanded CD8 TEM and Treg clones was similar in both groups (Figure 4B).

As T cell clonal expansion was evident in both PCLD phenotypes, we next sought to identify related TCRs for each group based on the similarity of their antigen specificity-determining CDR3 amino acid sequences, on the premise that clusters of related

TCRs recognize similar epitopes (details of the analysis are provided in the online data supplement). We hypothesized that inflammatory cases would exhibit more clustering than fibrotic cases, given the trend towards a greater proportion of expanded CD4 TCM clones and increased abundance of both CD4 TCM and CD8 TEM in the former group. However, high levels of clustering were evident in both inflammatory and fibrotic PCLD. In benchmarking, this level of clustering was similar to that observed for expanded peripheral blood TCR clones detected following non-severe SARS-CoV-2 infection, and exceeded that in non-expanded TCRs from non-infected individuals from the same cohort (27, 31) (Figure 4C,D). In both radiological groups of PCLD the vast majority of clusters contained TCRs from multiple donors (Figure 4E) but there was minimal sharing of identical CDR3 sequences between different individuals (Table E3). As a further comparison of inflammatory and fibrotic PCLD T cell repertoires we clustered CDR3 amino acid sequences from both groups together. Strikingly, most clusters contained TCRs from both inflammatory and fibrotic samples and multiple donors, suggesting the presence of T cell clones that recognize similar antigens across both phenotypes. A few small clusters composed uniquely of either inflammatory or fibrotic PCLD TCRs were evident, indicative of subtle differences between the two repertoires (Figure 4F).

The high level of relatedness between TCRs in PCLD is suggestive of antigen-specific immune responses. We therefore sought to determine whether these repertoires were enriched for T cells specific for SARS-CoV-2 by comparison to other common viruses. Of SARS-CoV-2 reactive TCRs identified in the VDJdb database, six were present in TCR data from inflammatory PCLD and eight in fibrotic PCLD. None were identified in equivalent sized healthy peripheral blood repertoires. However, fewer SARS-CoV-

specific-TCRs were detected than EBV-specific or CMV-specific sequences, indicating no enrichment for SARS-CoV-2-specific T cells at the site of disease. There was no enrichment for SARS-CoV-2-reactive or EBV-reactive TCRs in inflammatory compared to fibrotic PCLD, however, CMV-specific TCRs were significantly enriched in the fibrotic group (Figure 4G). TCRs found in clusters composed uniquely of either PCLD phenotype were not enriched for known virus reactive-sequences (Table E4).

DISCUSSION

We report the first comparative molecular analysis of cells from the site of disease of the two predominant pulmonary radiological sequelae observed following COVID-19 (1). Contrary to our hypothesis that these two phenotypes exhibit different molecular characteristics, the transcriptomes of post-COVID radiological inflammation and fibrosis were highly similar for all bronchoalveolar cell types, dominated by cellular processes involved in inflammatory and immune responses. We found no robust evidence of enhanced activity of epithelial damage or wound repair pathways in those with fibrotic radiological changes. Although many cell type-specific differentially expressed genes were identified between the two groups by single-cell methods, there were no systematic differences in enriched biological pathways or their upstream regulators among these differentially expressed genes. Furthermore, almost no differences in gene expression between the fibrotic and the inflammatory groups were detected if gene expression within cell types was aggregated by donor before differential analysis (23). Thus, these two entities likely represent distinct manifestations of the same disease process.

In comparison to fibrotic PCLD, the bronchoalveolar environment of inflammatory PCLD was characterised by significantly increased abundance of CD4 central memory and CD8 effector memory T cells. T cell infiltration has been a consistent finding in the few recent studies which have examined the post-COVID-19 airspaces; almost exclusively within 3-6 months of the acute insult, when radiological inflammation is more common than fibrosis (3, 4, 32). The cellular proportions of our fibrotic PCLD samples, harvested at 9-12 months after acute illness, were akin to those reported in BAL samples from healthy individuals, macrophages comprising greater than 80% and lymphocytes less than 10% of cells (26, 33). This is in keeping with the reported repopulation of the airways by AM in the later stages of COVID-19 acute respiratory distress syndrome (ARDS) (24). Minimal data currently available for longitudinal samples obtained at approximately one year post-infection also suggest a trajectory of gradual normalization of molecular abnormalities and airspace cellular composition (4). Nonetheless there is considerable interest in therapeutic intervention to expedite resolution of subacute inflammation, to prevent progression to fibrosis with concomitant irreversible tissue damage in susceptible individuals. Our findings suggest therapies which target T cells, rather than anti-fibrotic agents could be beneficial in PCLD. In support of this premise, one small clinical study demonstrated improvement in clinical symptoms, physiological parameters and radiological abnormalities, following three weeks of corticosteroid treatment instituted approximately three months post-COVID-19 (34).

The TCR repertoire from both PCLD phenotypes exhibited high levels of relatedness, suggestive of antigen-specific immune responses. In addition, the high proportion of clusters containing TCRs from multiple donors suggests immune responses against

similar antigens, despite almost no sharing of identical TCR sequences between different individuals. Aberrant immune responses to persistent reservoirs of virus have been posited as drivers of post-acute sequelae of SARS-CoV-2 infection (35–37). However, even including all detected TCRs, to offset the inherent sparsity of single-cell data, there was no enrichment for known SARS-CoV-2-specific T cell clones. Hence, based on the current compendium of SARS-CoV-2-reactive TCRs (14), which may not be comprehensive, there was no evidence for viral persistence at the site of disease in our cohort. A more plausible hypothesis, given reports of cross-reactivity between SARS-CoV-2 and human antigens (38, 39), is that the PCLD TCR repertoire is directed against as yet unknown respiratory autoantigens. Identifying the antigenic targets of the PCLD T cell repertoire may open opportunities for more specific therapeutic interventions.

Pro-fibrotic macrophages have been implicated in the pathogenesis of severe COVID-19 ARDS, which has been associated with rapid onset pulmonary fibrosis, which improves over time (24). In our study, which encompassed a broad range of disease severity including several individuals with COVID-19 ARDS, the phenotype of myeloid cells in both radiological groups was consistent with the spectrum of macrophages and monocytes found in healthy airspaces (26); with no convincing evidence for exaggerated activity of pro-fibrogenic pathways. The heterogeneity of disease phenotype and longer time interval after acute illness at which our cohort were lavaged may account for this discrepancy. Of note, CMV-specific T cell clones were enriched in individuals with fibrotic PCLD, consistent with the repeated association of CMV with pulmonary fibrosis (40).

Limitations of the study

We recognize that small sample size will necessitate validation of our observations in larger cohorts. Those with fibrotic radiological appearances were sampled significantly later after acute disease than those with radiological inflammation. The introduction of dexamethasone treatment early in the second wave of the pandemic reduced the number of individuals with persistent respiratory symptoms and radiological abnormalities. Consequently, we extended the interval after acute illness within which individuals were eligible for sampling. This provided invaluable opportunities to assess 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 airway environment would provide important insights into the molecular mechanisms driving resolution and whether this trajectory is a universal phenomenon for both inflammation and fibrosis. Finally, due to technical limitations because of the low numbers of T cells present in BAL samples, we were unable to obtain TCR data for two inflammatory cases. Nonetheless, our data clearly demonstrate that the TCR repertoire of PCLD reflects antigen-directed immune responses.

Conclusion

Our observations that inflammatory PCLD is characterised by airway T cell infiltration and that antigen-specific T cell responses persist longer term, highlight opportunities for early intervention with therapies targeting T cells. Understanding the timing and duration of intervention, stratification of those at high risk of irreversible tissue damage and the potential use of more targeted T cell immunomodulators all merit further

investigation.

REFERENCES

1. Fabbri L, Moss S, Khan FA, Chi W, Xia J, Robinson K, *et al.* Parenchymal lung abnormalities following hospitalisation for COVID-19 and viral pneumonitis: a systematic review and meta-analysis. *Thorax* 2023;78:191–201.
2. Mehta P, Rosas IO, Singer M. Understanding post-COVID-19 interstitial lung disease (ILD): a new fibroinflammatory disease entity. *Intensive Care Med* 2022;48:1803–1806.
3. Cheon IS, Li C, Son YM, Goplen NP, Wu Y, Cassmann T, *et al.* Immune signatures underlying post-acute COVID-19 lung sequelae. *Sci Immunol* 2021;6:eabk1741.
4. Vijayakumar B, Boustani K, Ogger PP, Papadaki A, Tonkin J, Orton CM, *et al.* Immuno-proteomic profiling reveals aberrant immune cell regulation in the airways of individuals with ongoing post-COVID-19 respiratory disease. *Immunity* 2022;55:542-556.e5.
5. Stewart I, Jacob J, George PM, Molyneaux PL, Porter JC, Allen RJ, *et al.* Residual Lung Abnormalities Following COVID-19 Hospitalization: Interim Analysis of the UKILD Post-COVID Study. *Am J Respir Crit Care Med* 2022;doi:10.1164/rccm.202203-0564OC.
6. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, *et al.* Comprehensive Integration of Single-Cell Data. *Cell* 2019;177:1888-1902.e21.
7. Zhang Z, Luo D, Zhong X, Choi JH, Ma Y, Wang S, *et al.* SCINA: A Semi-Supervised Subtyping Algorithm of Single Cells and Bulk Samples. *Genes* 2019;10:531.
8. Villani A-C, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and

progenitors. *Science* 2017;356:eaah4573.

9. Cheng S, Li Z, Gao R, Xing B, Gao Y, Yang Y, *et al.* A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* 2021;184:792-809.e23.

10. Glass DR, Tsai AG, Oliveria JP, Hartmann FJ, Kimmey SC, Calderon AA, *et al.* An Integrated Multi-omic Single-Cell Atlas of Human B Cell Identity. *Immunity* 2020;53:217-232.e5.

11. Fang H, Knezevic B, Burnham KL, Knight JC. XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits. *Genome Med* 2016;8:129.

12. Turner CT, Brown J, Shaw E, Uddin I, Tsaliki E, Roe JK, *et al.* Persistent T Cell Repertoire Perturbation and T Cell Activation in HIV After Long Term Treatment. *Front Immunol* 2021;12:634489.

13. Joshi K, de Massy MR, Ismail M, Reading JL, Uddin I, Woolston A, *et al.* Spatial heterogeneity of the T cell receptor repertoire reflects the mutational landscape in lung cancer. *Nat Med* 2019;25:1549–1559.

14. Bagaev DV, Vroomans RMA, Samir J, Stervbo U, Rius C, Dolton G, *et al.* VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Res* 2020;48:D1057–D1062.

15. Murray PJ, Wynn TA. Protective and pathogenic functions of macrophage subsets. *Nat Rev Immunol* 2011;11:723–737.

16. Davies LC, Jenkins SJ, Allen JE, Taylor PR. Tissue-resident macrophages. *Nat Immunol* 2013;14:986–995.

17. Gardell JL, Parker DC. CD40L is transferred to antigen-presenting B cells

during delivery of T-cell help. *Eur J Immunol* 2017;47:41–50.

18. van Aalderen MC, van Lier RAW, Hombrink P. How to Reliably Define Human CD8+ T-Cell Subsets: Markers Playing Tricks. *Cold Spring Harb Perspect Biol* 2021;13:a037747.

19. David G, Willem C, Legrand N, Djaoud Z, Mérieau P, Walencik A, *et al.* Deciphering the biology of KIR2DL3+ T lymphocytes that are associated to relapse in haploidentical HSCT. *Sci Rep* 2021;11:15782.

20. Tsyklauri O, Chadimova T, Niederlova V, Kovarova J, Michalik J, Malatova I, *et al.* Regulatory T cells suppress the formation of potent KLRK1 and IL-7R expressing effector CD8 T cells by limiting IL-2. *eLife* 2023;12:e79342.

21. Adams TS, Schupp JC, Poli S, Ayaub EA, Neumark N, Ahangari F, *et al.* Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv* 2020;6:eaba1983.

22. Davis JD, Wypych TP. Cellular and functional heterogeneity of the airway epithelium. *Mucosal Immunol* 2021;14:978–990.

23. Squair JW, Gautier M, Kathe C, Anderson MA, James ND, Hutson TH, *et al.* Confronting false discoveries in single-cell differential expression. *Nat Commun* 2021;12:5692.

24. Wendisch D, Dietrich O, Mari T, von Stillfried S, Ibarra IL, Mittermaier M, *et al.* SARS-CoV-2 infection triggers profibrotic macrophage responses and lung fibrosis. *Cell* 2021;184:6243-6261.e27.

25. Nouno T, Okamoto M, Ohnishi K, Kaieda S, Tominaga M, Zaizen Y, *et al.* Elevation of pulmonary CD163 + and CD204 + macrophages is associated with the

clinical course of idiopathic pulmonary fibrosis patients. *J Thorac Dis* 2019;11:.

26. Mould KJ, Moore CM, McManus SA, McCubbrey AL, McClendon JD, Griesmer CL, *et al.* Airspace Macrophages and Monocytes Exist in Transcriptionally Distinct Subsets in Healthy Adults. *Am J Respir Crit Care Med* 2021;203:946–956.

27. Chandran A, Rosenheim J, Nageswaran G, Swadling L, Pollara G, Gupta RK, *et al.* Rapid synchronous type 1 IFN and virus-specific T cell responses characterize first wave non-severe SARS-CoV-2 infections. *Cell Rep Med* 2022;3:100557.

28. Wauters E, Van Mol P, Garg AD, Jansen S, Van Herck Y, Vanderbeke L, *et al.* Discriminating mild from critical COVID-19 by innate and adaptive immune single-cell profiling of bronchoalveolar lavages. *Cell Res* 2021;31:272–290.

29. Auffray C, Sieweke MH, Geissmann F. Blood monocytes: development, heterogeneity, and relationship with dendritic cells. *Annu Rev Immunol* 2009;27:669–692.

30. Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, *et al.* Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med* 2020;26:842–844.

31. Milighetti M, Peng Y, Tan C, Mark M, Nageswaran G, Byrne S, *et al.* Large clones of pre-existing T cells drive early immunity against SARS-COV-2 and LCMV infection. *bioRxiv* 2022;2022.11.08.515436 [preprint]. 2022 Nov 8. Available from https://doi.org/10.1101/2022.11.08.515436.

32. Ravaglia C, Doglioni C, Chilosi M, Piciucchi S, Dubini A, Rossi G, *et al.* Clinical, radiological and pathological findings in patients with persistent lung disease following SARS-CoV-2 infection. *Eur Respir J* 2022;60:2102411.

33. The BAL Cooperative Group Steering Committee. Bronchoalveolar lavage constituents in healthy individuals, idiopathic pulmonary fibrosis, and selected comparison groups. *Am Rev Respir Dis* 1990;141:S169-202.

34. Myall KJ, Mukherjee B, Castanheira AM, Lam JL, Benedetti G, Mak SM, *et al.* Persistent Post-COVID-19 Interstitial Lung Disease. An Observational Study of Corticosteroid Treatment. *Ann Am Thorac Soc* 2021;18:799–806.

35. Merad M, Blish CA, Sallusto F, Iwasaki A. The immunology and immunopathology of COVID-19. *Science* 2022;375:1122–1127.

36. Stein SR, Ramelli SC, Grazioli A, Chung J-Y, Singh M, Yinda CK, *et al.* SARS-CoV-2 infection and persistence in the human body and brain at autopsy. *Nature* 2022;612:758–763.

37. Cheung CCL, Goh D, Lim X, Tien TZ, Lim JCT, Lee JN, *et al.* Residual SARS-CoV-2 viral antigens detected in GI and hepatic tissues from five recovered patients with COVID-19. *Gut* 2022;71:226–229.

38. Nunez-Castilla J, Stebliankin V, Baral P, Balbin CA, Sobhan M, Cickovski T, *et al.* Potential Autoimmunity Resulting from Molecular Mimicry between SARS-CoV-2 Spike and Human Proteins. *Viruses* 2022;14:1415.

39. Vojdani A, Kharrazian D. Potential antigenic cross-reactivity between SARS-CoV-2 and human tissue with a possible link to an increase in autoimmune diseases. *Clin Immunol* 2020;217:108480.

40. Moore BB, Moore TA. Viruses in Idiopathic Pulmonary Fibrosis. Etiology and Exacerbation. *Ann Am Thorac Soc* 2015;12 Suppl 2:S186-192.

FIGURE LEGENDS

Figure 1. Higher abundance of bronchoalveolar T cells in inflammatory post-COVID-19 lung disease (PCLD). (A) Schematic representation of the timing of recruitment of the study cohort in relation to the COVID-19 pandemic waves. Colors represent the predominant radiological manifestation of PCLD. Numbers below each figure represent subject identity and those above, interval between onset of acute illness and bronchoalveolar lavage (BAL). Asterisks indicate COVID-19 treatment received. (B) Representative computed tomography (CT) images for each individual. (C) Uniform manifold approximation and projection (UMAP) embedding of 55,776 bronchoalveolar single-cell transcriptomes obtained from five individuals with radiological features of pulmonary inflammation and five individuals with radiological features of pulmonary fibrosis following COVID-19 infection, split by PCLD phenotype, color coded by cell type. Cell type annotation was achieved by assignment of Azimuth human lung reference gene signatures using the SCINA R package and in the case of dendritic cells and B cells, expression of literature-based markers. Prolif = proliferating cells, identified by their high module score for a gene signature representing the cellular proliferation response. (D) Cellular composition of each BAL sample defined by single-cell RNA sequencing (scRNAseq). Color indicates cell type and bar height represents proportion. (E) Dot plot visualization of the expression of known marker genes for each cell type; “Mac” = macrophage. Dot size represents the percentage of cells expressing the gene in each cell type, color shows the z-scores of average log-normalized expression for each cell type compared to the entire data set. (F) Comparison of the proportions of each cell type in inflammatory and fibrotic PCLD. Horizontal lines indicate median, box limits the interquartile range and whiskers the 5^{th} to 95^{th}

percentiles, * = FDR<0.05.

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 1C, 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) UMAP from (A) split by donor, colored by T cell subset. (C) Relative proportions of each T cell subset in each subject, colors represent cell type. (D) 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. (E) 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 1C, 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) UMAP from (A) split by subject, colored by the identified macrophage and monocyte populations. (C) Relative proportions of each myeloid subset within each individual, colors represent cell type. (D) Expression of marker genes for PCLD macrophage and monocyte populations, colored by scaled, log-normalized counts, projected on to the myeloid cell UMAP. (E)

Dot plot visualization of the expression of selected marker genes for each macrophage and monocyte subset. Dot size represents the percentage of cells expressing the gene in each myeloid subset, color shows the z-scores of average log-normalized expression for each subset compared to the entire data set. (F) Expression of a profibrotic macrophage gene signature derived from idiopathic pulmonary fibrosis, calculated on a single-cell level, colored by module score and projected on to the macrophage UMAP. (G) Comparison of the bronchoalveolar myeloid populations in inflammatory and fibrotic PCLD. Horizontal lines indicate median, box limits the interquartile range and whiskers the 5th to 95th percentiles.

Figure 4. Highly similar T cell receptors (TCRs) characterize both inflammatory and fibrotic post-COVID lung disease (PCLD). (A) T cell clonal expansion is visualized on the T cell UMAP split by radiological PCLD phenotype. TCR sequences detected at a frequency of greater than one are colored blue and contour lines provide a 2D representation of TCR density overlaid in red for inflammatory PCLD and blue for fibrotic PCLD. (B) Comparison of the proportion of expanded TCR sequences, defined as those detected at a frequency greater than one in the three largest T cell subsets identified in Figure 2A in the two PCLD phenotypes. (C) Percentage of complementarity determining region (CDR)3 alpha and beta chain amino acid sequences clustering and number of clusters generated over a range of thresholds above which two TCRs are considered similar, for inflammatory and fibrotic PCLD bronchoalveolar (BAL) lavage samples, negative control PBMC samples not expected to cluster highly and positive control PBMC samples known to cluster highly, analyzed separately. (D) Representative network diagrams of TCR β chain clusters present in each group described in (C). Nodes represent TCRs, related TCRs are connected by

an edge and colors represent the groups. (E) Network diagrams of related TCR β chains in each PCLD phenotype in which nodes are colored by donor. (F) Network diagrams visualizing TCR β chain clusters identified by combined analysis of the two PCLD phenotypes. Nodes are colored by donor (right) or radiological phenotype and membership of clusters composed of one or both PCLD groups (left), circular nodes represent inflammatory PCLD and square nodes fibrotic PCLD. (G) Number of TCR sequences (α and β genes combined) annotated for SARS-CoV-2, cytomegalovirus (CMV), and Epstein-Barr virus (EBV) in VDJdb either detected or not detected in TCR sequences from individuals with each PCLD phenotype and negative controls, giving the odds ratio ±95% confidence interval (Fisher's exact test) for enrichment of antigen-specific TCR sequences in each instance.

TABLES

Table 1. Clinical and demographic information summary.

	Inflammatory (n=5)	Fibrotic (n=5)
Age (Median + IQR)	62 (21.5)	59 (8.5)
Sex (Male/Female)	2/3	4/1
Ethnicity (White/Asian)	3/2	3/2
BMI (Median + IQR)	31.8 (15.6)	27.3 (6.1)
Smoking (Never/Former)	4/1	2/3
Respiratory Support		
I&V*†	2	4
CPAP‡	1	0
HFNO§	1	1
Nil	1	0
COVID-19 treatment		
Steroid	3	5
Tocilizumab	1	1
Days to BAL‖ (Median + IQR)	116 (50)	316 (50.5)

Summary of clinical and demographic information for the study cohort. *One patient also received extracorporeal membrane oxygenation (ECMO). †I&V = Intubation and ventilation, ‡CPAP = continuous positive airways pressure; §HFNO = high flow nasal oxygen ‖ BAL = bronchoalveolar lavage.

Figure 1

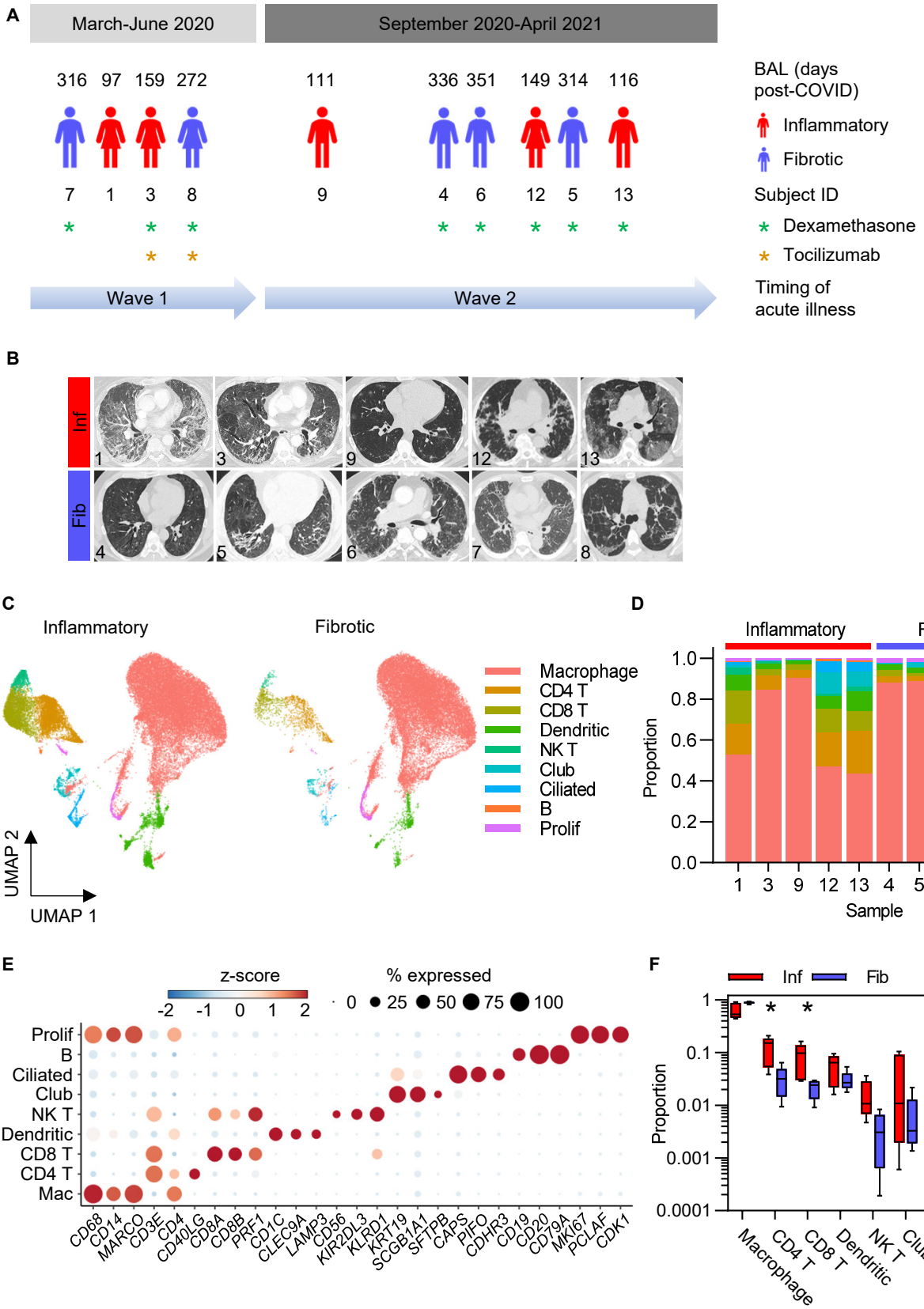

Figure 2

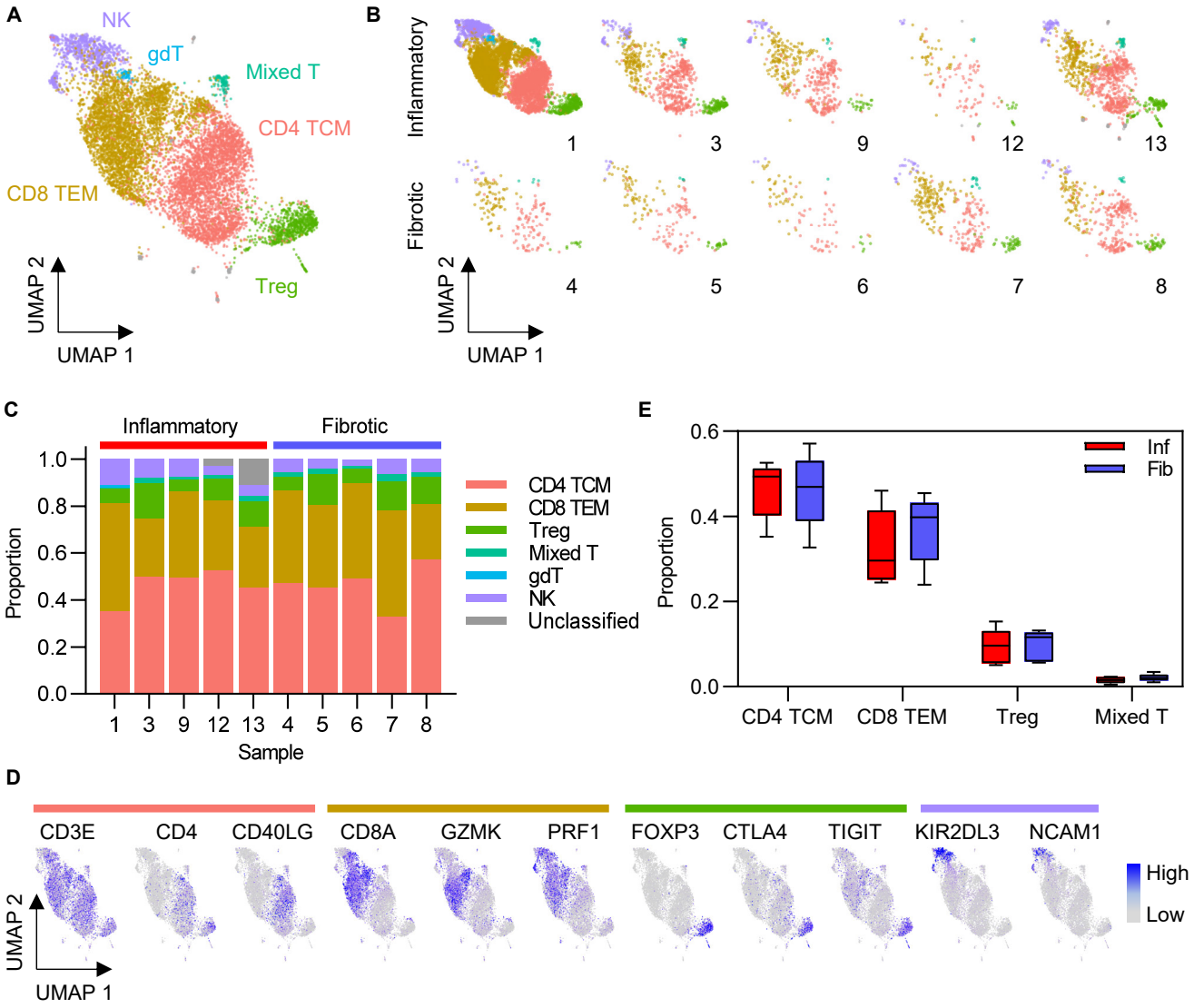

Figure 3

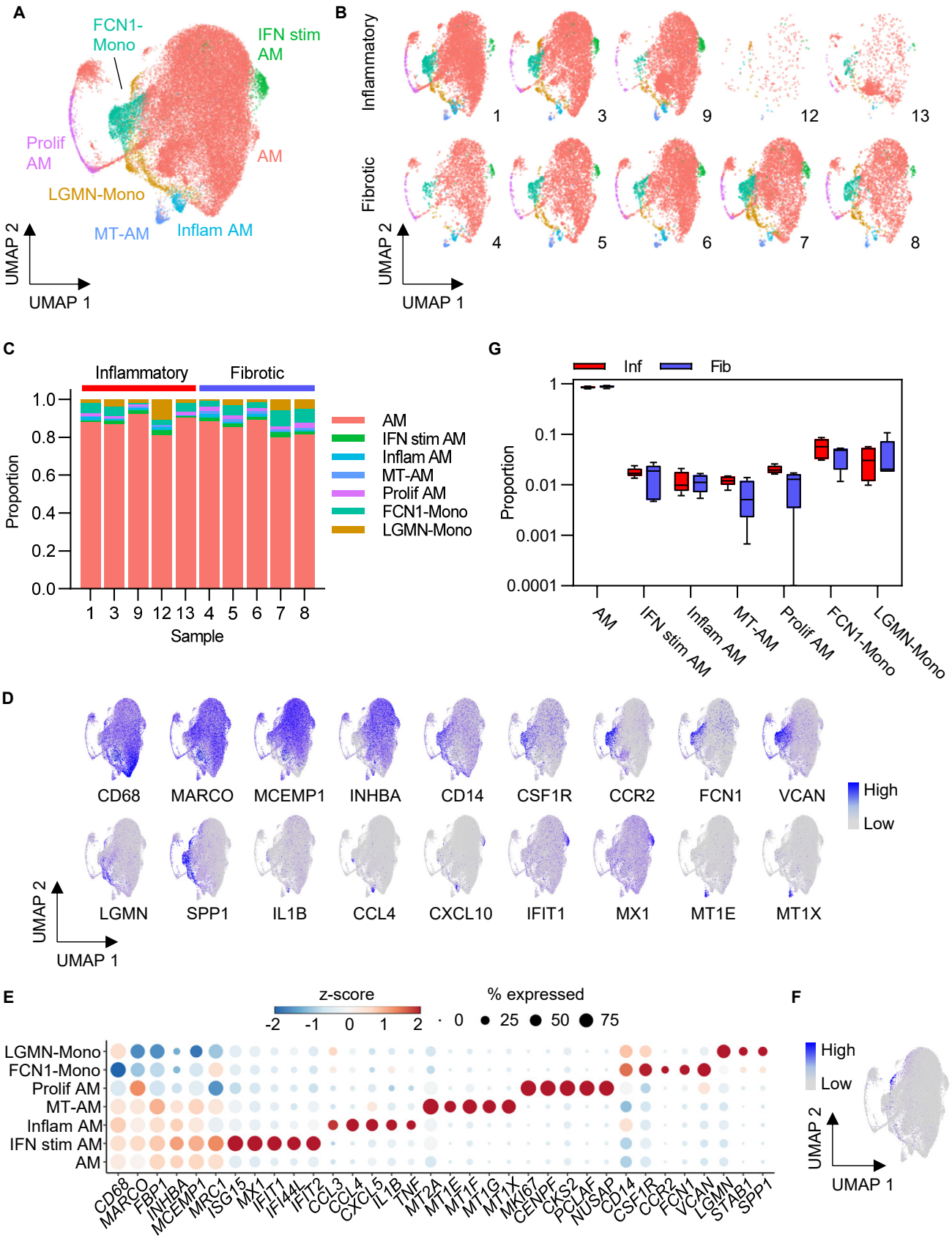

Figure 4