



Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19

Mingfeng Liao^{1,6}, Yang Liu^{1,6}, Jing Yuan^{2,6}, Yanling Wen¹, Gang Xu¹, Juanjuan Zhao¹, Lin Cheng¹, Jinxiu Li², Xin Wang¹, Fuxiang Wang², Lei Liu^{1,3}✉, Ido Amit^{1,4}✉, Shuye Zhang^{1,5}✉ and Zheng Zhang^{1,3}✉

Respiratory immune characteristics associated with Coronavirus Disease 2019 (COVID-19) severity are currently unclear. We characterized bronchoalveolar lavage fluid immune cells from patients with varying severity of COVID-19 and from healthy people by using single-cell RNA sequencing. Proinflammatory monocyte-derived macrophages were abundant in the bronchoalveolar lavage fluid from patients with severe COVID-19. Moderate cases were characterized by the presence of highly clonally expanded CD8⁺ T cells. This atlas of the bronchoalveolar immune microenvironment suggests potential mechanisms underlying pathogenesis and recovery in COVID-19.

The underlying immune responses to severe acute respiratory syndrome (SARS)-CoV2 remain unclear. Abnormal immune responses have been reported in coronavirus animal infection models^{1,2} and lymphopenia is a prominent feature of severe COVID-19 (ref. ³). Bronchoalveolar lavage fluid (BALF) contains microenvironment information on bronchioles and lung alveoli. Here, we performed single-cell RNA sequencing (scRNA-seq) on BALF cells from three patients with moderate COVID-19 (M1–M3), six patients with severe/critical infection (S1–S6), three healthy controls (HC1–HC3) and a publicly available BALF (HC4)⁴ sample (Supplementary Tables 1 and 2). Clustering analysis showed 31 distinct clusters composed of macrophages (CD68), neutrophils (FCGR3B), myeloid dendritic cells (mDCs) (CD1C, CLEC9A), plasmacytoid dendritic cells (pDCs) (LILRA4), natural killer (NK) cells (KLRD1), T cells (CD3D), B cells (MS4A1), plasma cells (IGHG4) and epithelial cells (TPPP3, KRT18) (Fig. 1a), identified by signature genes (Extended Data Fig. 1a,b). Major cell types, including mDCs, pDCs, mast cells, NK cells, T cells and B cells, contained cells from most samples, whereas macrophages showed specific enrichment in different groups (Extended Data Fig. 1c–e). BALFs of patients with severe/critical COVID-19 infection contained higher proportions of macrophages and neutrophils and lower proportions of mDCs, pDCs and T cells than those with moderate infection (Fig. 1b and Extended Data Fig. 1f).

We re-clustered 20 macrophage subclusters to further dissect their heterogeneity (Extended Data Fig. 2a). According to recent classification criteria⁴ and typical markers (Extended Data Fig. 2b), the macrophages were grouped by FCN1, SPP1 and FABP4 expression patterns including the FCN1^{hi} group 1, FCN1^{lo}SPP1⁺ group 2, SPP1⁺ group 3 and FABP4⁺ group 4 macrophages (Extended Data Fig. 2c).

The macrophage compartments differed largely in both composition and expression of FCN1, SPP1 and FABP4 in different cell groups (Fig. 1c and Extended Data Fig. 2d). FABP4 was preferentially expressed by controls and patients with moderate infection, while FCN1 and SPP1 were highly expressed by patients with severe/critical infection (Fig. 1d). We conducted differentially expressed gene (DEG) analysis (Extended Data Fig. 2e), gene ontology (GO) analysis and gene set enrichment analysis (GSEA) (Extended Data Fig. 2f) between cell groups. Group 1 expressed the peripheral monocyte-like markers S100A8, FCN1 and CD14, and group 2 expressed high levels of the chemokines CCL2, CCL3, CXCL10 and other genes. Both groups had gene expression patterns suggestive of classic M1-like macrophages. Group 3 expressed the immunoregulatory genes A2M, GPR183 and CCL13 and the profibrotic genes TREM2, TGFB1 and SPP1, suggestive of alternative M2-like macrophages, a reparative but profibrotic subset⁵. Group 4 expressed the alveolar macrophage (AM) genes FABP4, APOC1, MARCO and genes associated with lipid metabolic functions⁶ and was enriched in BALFs of controls and moderate cases but not in severe cases (Extended Data Fig. 2f,g). Using SCENIC, we found enhanced expression of the transcription factor genes NFκB, STAT1, STAT2 and interferon (IFN) regulatory factors in group 1 and 2, M2-promoting TFEB, NR1H3, PPARA and CREB1 activities in group 3 and AM-specific PPARG and CEBPB in group 4 macrophages (Extended Data Fig. 2h and Methods). These findings suggest that a highly proinflammatory macrophage microenvironment is present in the lungs of patients with severe COVID-19, which is consistent with previous knowledge of macrophage populations during steady-state, inflammation and recovery⁷.

Memory T cell responses have been observed in patients with SARS⁸ and are necessary for resolving SARS-CoV infection in mice^{9,10}. We identified six major clusters of T and NK lymphocytes (Extended Data Fig. 3a) based on expression of canonical genes (Extended Data Fig. 3b,c). Lower CD8⁺ T cell and higher proliferating T cell proportions were observed in patients with severe/critical infection than in patients with moderate infection (Extended Data Fig. 3d,e). These CD8⁺ T cells expressed higher levels of GZMA, GZMK and FASLG and the tissue-residence markers ITGA1, CXCR6 and JAML (Extended Data Fig. 3f), with upregulation of genes involved in activation, migration and cytokine-related pathways in moderate cases and translation initiation, cell homeostasis and nucleoside metabolic pathways in severe cases (Extended Data Fig. 3g).

¹Institute for Hepatology, National Clinical Research Center for Infectious Disease, Shenzhen Third People's Hospital, Shenzhen, China. ²Department for Infectious Diseases, Shenzhen Third People's Hospital, Shenzhen, China. ³The Second Affiliated Hospital, School of Medicine, Southern University of Science and Technology, Shenzhen, China. ⁴Department of Immunology, Weizmann Institute of Science, Rehovot, Israel. ⁵Shanghai Public Health Clinical Center, Fudan University, Shanghai, China. ⁶These authors contributed equally: Mingfeng Liao, Yang Liu, Jing Yuan. ✉e-mail: liulei3322@aliyun.com; ido.amit@weizmann.ac.il; zhangshuye@shphc.org.cn; zhangzheng1975@aliyun.com

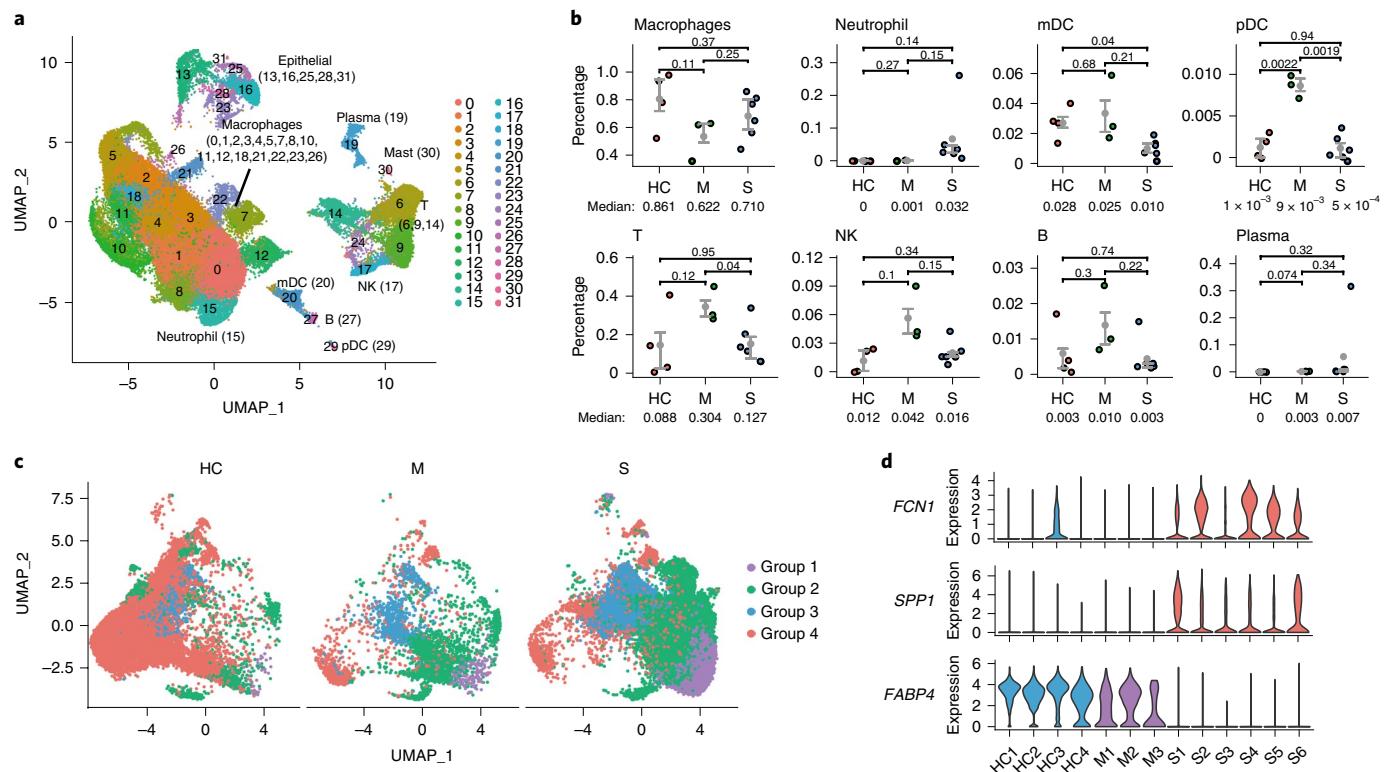


Fig. 1 | Dysregulated bronchoalveolar immune landscapes in patients with severe COVID-19 infection. **a**, UMAP presentation of major cell types and associated clusters in BALFs ($n=13$). **b**, The proportions of the major BALF immune cell types among healthy controls (HC) and patients with moderate (M) and severe (S) COVID-19 infection (using two-sided Student's *t*-test for pairwise comparisons). **c**, UMAP projection of four macrophage groups among controls ($n=4$) and patients (moderate, $n=3$; severe, $n=6$). **d**, Expression of *FCN1*, *SPP1* and *FABP4* by BALF macrophages from each sample.

We further assessed T cell clonal expansion using single-cell T cell receptor sequencing (scTCR-seq) of seven T cell clusters from BALF from patients with COVID-19 (Fig. 2a and Extended Data Fig. 4a). TCR clonotype was obtainable from 66.9% (26.6–87.3%) of T cells in each patient and from 60.9% (45.2–84.2%) of cells in each T cell cluster (Supplementary Table 2 and Extended Data Fig. 4b). On the uniform manifold approximation and projection (UMAP), projections of clonally expanded T cells showed an aggregative distribution, indicating transcriptional homogeneity in patients with moderate infection versus a diffusive distribution indicating transcriptional heterogeneity in patients with severe/critical infection (Fig. 2b). Cell numbers and clonal expansion differed between CD8⁺ T cell clusters in patients with moderate versus severe/critical infection (Fig. 2c). More than 50% of ZNF683⁺ CD8⁺ T cells from patients with moderate infection belonged to expanded clones, showing a higher amplification index (> 5 cells) than those from patients with severe/critical infection (Fig. 2c). Among the top 20 individual clones (Supplementary Table 3), the majority of cells from patients with moderate infection belonged to the ZNF683⁺ cluster, but exhibited various expression patterns in patients with severe/critical infection (Extended Data Fig. 4c). ZNF683⁺-expanded clones that were enriched in patients with moderate infection could represent SARS-CoV-2-specific CD8⁺ T cells. These expanded clones preferentially expressed tissue-residence genes, including *XCL1*, *CXCR6* and *ITGAE* (Extended Data Fig. 4d). A higher score of core tissue-residence genes¹¹ was seen in expanded CD8⁺ T cells versus nonexpanded cells in patients with moderate infection and in total CD8⁺ T cells in moderately versus severely/critically infected patients (Fig. 2d). Collectively, our data suggest that CD8⁺ T cells in BALFs from patients with severe/critical infection were less expanded, more proliferative and more phenotypically

heterogeneous, whereas a larger proportion of CD8⁺ T cell effectors with tissue-resident and highly expanded features were present in BALFs from patients with moderate infection.

Finally, we measured several cytokines and chemokines in BALF of patients with COVID-19 (Supplementary Tables 1 and 4). Compared to patients with moderate COVID-19 infection, patients with severe/critical infection had much higher levels of inflammatory cytokines, particularly interleukin (IL)-8, IL-6 and IL-1 β , in their BALFs (Extended Data Fig. 5a). Three patients (S1, S8 and S9) with critical infection displayed persistently high levels of cytokines. In scRNA-seq data, *IL1B*, *IL6*, *TNF* and various chemokines (*CCL2*, *CCL3*, *CCL4* and *CCL7*) were expressed at higher levels in lung macrophages from patients with severe COVID-19 infection. *CXCL9*, *CXCL10* and *CXCL11* levels were much higher in both COVID-19 groups than in healthy people, but *CXCL16*, whose product binds CXCR6, was more highly expressed in patients with moderate than severe infection (Extended Data Fig. 5b,c). We also examined the corresponding receptors to other upregulated chemokines in patients with COVID-19 (Extended Data Fig. 5d). These data suggest that lung macrophages in patients with severe COVID-19 infection may contribute to local inflammation by recruiting inflammatory monocytic cells and neutrophils through CCR1 and CXCR2, whereas macrophages in patients with moderate COVID-19 infection produce more T cell-attracting chemokines through engaging CXCR3 and CXCR6.

Together with previous studies^{3,12-14}, our data suggest that cytokine storm is associated with disease severity in COVID-19. Our study has several shortcomings, including limited and heterogeneous patients and a lack of longitudinal samples collected before and after infection. In addition, the B cell response could not be analyzed due to low cell numbers. The effects of age, pre-existing

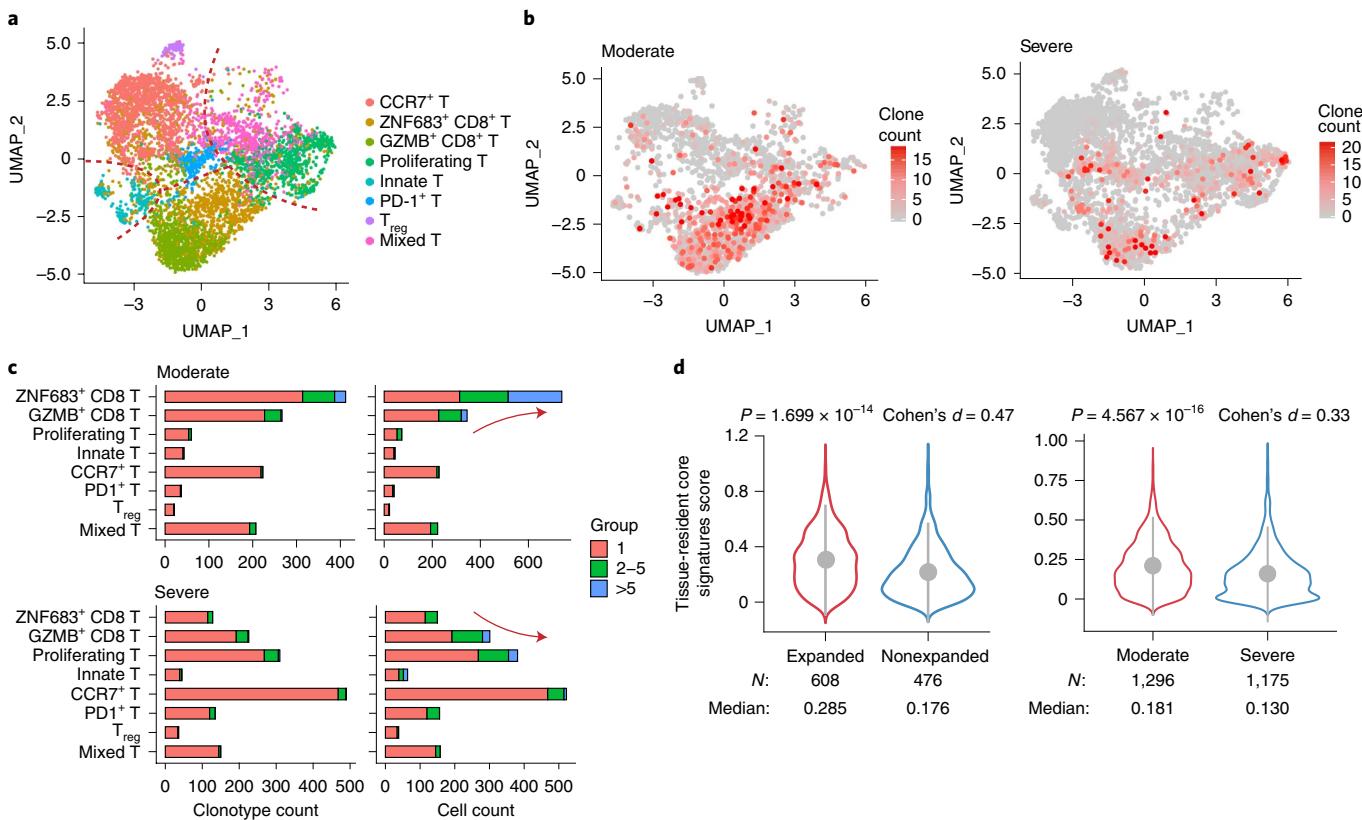


Fig. 2 | T cell clonal expansion in BALF from patients with COVID-19. **a**, The clustering of BALF-derived T cells in patients with COVID-19 ($n=8$, S3 excluded). The dotted lines partition the UMAP into CD4⁺ T cells (upper), CD8⁺ T cells (lower), innate T cells (left) and proliferating T cells (right). **b**, UMAP projection of clonally expanded T cells from patients with COVID-19. **c**, Clonal expansion status of T cell subsets in patients with moderate and severe/critical infection. Clonotype counts (left) and cell counts (right) are shown. **d**, Tissue-resident signature score by the expanded versus nonexpanded BALF CD8⁺ T cells in patients with moderate infection and by CD8⁺ T cells in patients with moderate versus severe/critical infection. In **b-d**, moderate was $n=3$ and severe was $n=5$.

conditions and immunoregulatory therapies could not be fully assessed. Nevertheless, this landscape of bronchoalveolar immune cells reveals aberrant macrophage and T cell responses underlying immunopathogenesis in COVID-19.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-020-0901-9>.

Received: 24 February 2020; Accepted: 23 April 2020;

Published online: 12 May 2020

References

- Channappanavar, R. et al. Dysregulated type I interferon and inflammatory monocyte-macrophage responses cause lethal pneumonia in SARS-CoV-infected mice. *Cell Host Microbe* **19**, 181–193 (2016).
- Liu, L. et al. Anti-spike IgG causes severe acute lung injury by skewing macrophage responses during acute SARS-CoV infection. *JCI Insight* **4**, e123158 (2019).
- Cao, X. COVID-19: immunopathology and its implications for therapy. *Nat. Rev. Immunol.* **20**, 269–270 (2020).
- Morse, C. et al. Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur. Respir. J.* **54**, 1802441 (2019).
- Reyfman, P. A. et al. Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **199**, 1517–1536 (2019).
- Evren, E., Ringqvist, E. & Willinger, T. Origin and ontogeny of lung macrophages: from mice to humans. *Immunology* <https://doi.org/10.1111/imm.13154> (2019).
- Bonnardel, J. & Guilliams, M. Developmental control of macrophage function. *Curr. Opin. Immunol.* **50**, 64–74 (2018).
- Liu, W. J. et al. T cell immunity of SARS-CoV: implications for vaccine development against MERS-CoV. *Antiviral Res.* **137**, 82–92 (2017).
- Channappanavar, R., Fett, C., Zhao, J., Meyerholz, D. K. & Perlman, S. Virus-specific memory CD8⁺ T cells provide substantial protection from lethal severe acute respiratory syndrome coronavirus infection. *J. Virol.* **88**, 11034–11044 (2014).
- Zhao, J., Zhao, J. & Perlman, S. T cell responses are required for protection from clinical disease and for virus clearance in severe acute respiratory syndrome coronavirus-infected mice. *J. Virol.* **84**, 9318–9325 (2010).
- Kumar, B. V. et al. Human tissue-resident memory T cells are defined by core transcriptional and functional signatures in lymphoid and mucosal sites. *Cell Rep.* **20**, 2921–2934 (2017).
- Huang, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).
- Zhou, F. et al. COVID-19 with spontaneous pneumomediastinum. *Lancet* **20**, 384–510 (2020).
- Chen, G. et al. Hepatitis C virus-specific CD4⁺ T cell phenotype and function in different infection outcomes. *J. Clin. Invest.* **130**, 768–773 (2020).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

Patients. *Ethics statement.* This study was conducted according to the principles expressed in the Declaration of Helsinki. Ethical approval was obtained from the Research Ethics Committee of Shenzhen Third People's Hospital (2020-112). All participants provided written informed consent for sample collection and subsequent analyses.

Patients and clinical sample collection. All 13 patients with COVID-19 in this study were enrolled from the Shenzhen Third People's Hospital from January to February, 2020. Disease severity was defined as moderate, severe and critical, according to the 'Diagnosis and Treatment Protocol of COVID-19 (the 7th Tentative Version)' by the National Health Commission of China issued on 3 March 2020 (<http://www.nhc.gov.cn/yzygj/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989.shtml>).

Briefly, moderate cases have fever, respiratory symptoms and pneumonia evidenced by computed tomography (CT) imaging. This study enrolled three patients with moderate infection with bilateral pneumonia, labeled M1 to M3. Patients with severe infection were diagnosed on the basis of one of the following criteria: (1) respiratory distress with respiratory rate ≥ 30 times min $^{-1}$; (2) fingertip oxygen saturation $\leq 93\%$ at resting state; (3) ratio of partial pressure of arterial oxygen to fraction of inspired oxygen ($\text{PaO}_2/\text{FiO}_2 \leq 300$ mm Hg (1 mm Hg = 0.133 kPa)); and (4) obvious progression of lesions in 24–48 h shown by pulmonary imaging $> 50\%$. Two patients with severe infection were enrolled, labeled S1 and S7. These two patients showed $\text{PaO}_2/\text{FiO}_2 \leq 300$ mm Hg. The patients with critical infection (S2–S6 and S8–S10) were diagnosed on the basis of one of the following criteria: (1) respiratory failure and an artificial airway required for invasive mechanical ventilation; (2) shock; and (3) combined failure of other organs that required intensive care unit monitoring. All of the eight patients with critical infection received invasive mechanical ventilation.

The demographic characteristics of the nine patients studied by scRNA-seq are listed in Supplementary Table 1, which includes three moderate cases (M1–M3), one severe case (S1) and five critical cases (S2–S6). The median age was 57 years, and the participants included six male and three female patients. All nine patients had Wuhan exposure history and had cough and/or fever as the first symptom. Diagnosis of SARS-CoV-2 was based on clinical symptoms, exposure history, chest radiography and were SARS-CoV-2 RNA-positive using commercial quantitative PCR with reverse transcription (qRT-PCR) assays in the sputum, nasal swab and/or BALF. Influenza A and B virus infection were excluded at enrollment. The patient M2 had clinical symptoms and CT evidence showing bilateral pneumonia, although the SARS-CoV-2 RNA test was negative. Patient M2 had close contact with patient S2 and was confirmed to have anti-SARS-CoV-2 antibodies using a SARS-CoV-2 total-IgG antibody detection kit (chemiluminescent immunoassay, nCoV-04100C, innoDx). The cutoff value was a cutoff index of 1 for the kit. Nineteen days after the onset of illness, her plasma IgG cutoff index value was increased to 25.39, showing a strong positive antibody signal and confirming the diagnosis of COVID-19. For assessment of cytokine and chemokine levels, BALF samples from additional patients with COVID-19 were examined. Clinical information of the additional patients is shown in Supplementary Table 4. In addition, three healthy controls, previously characterized and labeled as HC1 to HC3, were included in the study. These donors were confirmed to be free of tuberculosis, tumor and other lung diseases through CT imaging and other laboratory tests. The demographic characteristics of the three healthy controls are shown in Supplementary Table 4.

qRT-PCR assay for SARS-CoV-2 RNA. A throat swab, sputum, nasal swab, anal swab and BALF were collected from the patients at various time points after hospitalization and were sent to the diagnosis laboratory. Total nucleic acid was extracted from the samples using the QIAamp RNA Viral kit (Qiagen) and qRT-PCR was performed using a China Food and Drug Administration-approved commercial kit specific for SARS-CoV-2 detection (GeneoDX). Each qRT-PCR assay provided a threshold cycle (Ct) value, indicating the number of cycles surpassing the threshold for a positive test. The specimens were considered positive if the Ct value was ≤ 37 , and otherwise it was negative. Specimens with a Ct value > 37 were repeated. The specimen was considered positive if the repeat results were the same as the initial result or between 37 and 40. If the repeat Ct was undetectable, the specimen was considered to be negative.

Isolation of BALF cells. Approximately 20 ml of BALF was obtained and placed on ice. BALF was processed within 2 h and all operations were performed in a BSL-3 laboratory. After passage of BALF through a 100- μm nylon cell strainer to remove clumps and debris, the supernatant was centrifuged and the cells were re-suspended in cooled RPMI 1640 complete medium. BALFs of patients with severe/critical COVID-19 infection contained a larger number of cells (1.25×10^6 to 2.25×10^6 cells ml $^{-1}$) than those from moderate cases (6.63×10^3 to 5.6×10^4 cells ml $^{-1}$) and healthy controls (1.17×10^4 to 2.1×10^4 cells ml $^{-1}$). Then cells were counted in 0.4% Trypan blue, centrifuged and re-suspended at a concentration of 2×10^6 ml $^{-1}$ for further use.

Cytokine measurement by cytometric bead array. The cytokines including IL-1 β , IL-2, IL-4, IL-5, IL-6, IL-8, IL-10, IL-12p70, IL-17, IFN- α , IFN- γ , tumor necrosis factor- α were detected according to the instruction (Uni-medica, cat. no. 503022).

In brief, BALF was centrifuged at 1,000g for 10 min and supernatant was taken. Then 25 μl of sonicated beads, 25 μl of BALF supernatant and 25 μl of detection antibodies were mixed and placed on a shaker at 500 r.p.m. for 2 h at room temperature. Then 25 μl of SA-PE was added to each tube directly. The tubes were placed on a shaker at 500 r.p.m. for 30 min. The data were obtained by flow cytometry (Canto II, BD) and were analyzed with LEGENDplex v.8.0 (VigeneTech).

ScRNA-seq and TCR-seq capturing, library construction and sequencing.

Single-cell capturing and downstream library constructions were performed using Chromium Single Cell 3g Reagent kits v.2 (10x Genomics; PN-120237, PN-120236 and PN-120262) for HC1 and HC2, kit v.3 (10x Genomics; PN-1000075, PN-1000073 and PN-120262) for HC3 and Chromium Single Cell V(D)J Reagent kits (10x Genomics; PN-1000006, PN-1000014, PN-1000020, PN-1000005) for all patients with COVID-19, according to the manufacturer's protocol. Briefly, a total 11 μl of single cell suspension (containing 20,000 cells), 40 μl of barcoded gel beads and partitioning oil were loaded to Chromium Chip A/B to generate single-cell gel bead-in-emulsion. The polyadenylated transcripts were reverse-transcribed inside each gel bead-in-emulsion afterwards. Full-length cDNA along with cell barcode identifiers were PCR-amplified and sequencing libraries were prepared and normalized. The constructed library was sequenced on a BGI MGISEQ-2000 or Illumina platform.

Publicly available healthy control data. In addition to data for the three healthy control BALFs, scRNA-seq data from BALF from one additional healthy control were acquired from the Gene Expression Omnibus (GEO) database under accession code GSE128033 (ref. ⁴), which contains data of one fresh BALF (GSM3660650) from a lung transplant donor generated using a 3' V2 chemistry kit on Chromium Single Cell Controller (10x Genomics). For the purpose of cross-reference, control data (labeled HC4) were analyzed together with our own dataset.

ScRNA-seq data alignment and sample aggregating.

The Cell Ranger Software Suite (v.3.1.0) was used to perform sample de-multiplexing, barcode processing and single-cell 5' unique molecular identifier (UMI) counting. To detect SARS-CoV-2 reads, a customized reference was built by integrating human GRCh38 and SARS-CoV-2 genome (severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome, GenBank MN908947.3). Specifically, splicing-aware aligner STAR¹⁵ was used in FASTQs alignment. Cell barcodes were then determined based on the distribution of UMI counts automatically. The following criteria were then applied to each cell of all nine patients and four healthy controls: gene number between 200 and 6,000, UMI count $> 1,000$ and mitochondrial gene percentage < 0.1 . After filtering, a total of 66,452 cells were left for the following analysis. Finally, a filtered gene-barcode matrix of all samples was integrated with Seurat v.3 (ref. ¹⁶) to remove batch effects across different donors. In parameter settings, the first 50 dimensions of canonical correlation analysis (CCA) and principal-component analysis (PCA) were used.

Dimensionality reduction and clustering. The filtered gene-barcode matrix was first normalized using 'LogNormalize' methods in Seurat v.3 with default parameters. The top 2,000 variable genes were then identified using the 'vst' method in Seurat FindVariableFeatures function. Variables 'nCount_RNA' and 'percent.mito' were regressed out in the scaling step and PCA was performed using the top 2,000 variable genes. Then tSNE was performed on the top 50 principal components for visualizing the cells. Meanwhile, graph-based clustering was performed on the PCA-reduced data for clustering analysis with Seurat v.3. The resolution was set to 1.2 to obtain a finer result.

Macrophage, NK and T cell re-integration. Macrophages and T cells were re-integrated and re-clustered, respectively, with Seurat v.3. Specifically, macrophages of all samples were integrated using the first 50 dimensions of canonical correlation analysis and PCA. The parameter k.filter was set to 115 to account for the small cell number in sample M3. In the clustering step, parameter resolution was set to 0.8. For all NK and T cells, samples HC2 and S3 were excluded in the integration (Extended Data Fig. 3) due to small T cell numbers in both samples. The parameter k.filter was set to 175 in the integration step. Moreover, the T cells in patients with COVID-19 were re-integrated by excluding sample S3, leaving eight samples in the following TCR analysis (Fig. 2 and Extended Data Fig. 4). The parameter k.filter was set to 140 in this integration step.

Differential gene expression analysis.

MAST¹⁷ in Seurat v.3 (FindAllMarkers function) was used to perform differential gene expression analysis. For each cluster of macrophage and T cells, DEGs were generated relative to all of the other cells. DEGs of expanded versus nonexpanded CD8 $^{+}$ T cells in moderate cases ($n=3$) were calculated using MAST with default parameters. A gene was considered significant with adjusted $P < 0.05$ (P values were adjusted by false discovery rate in MAST). To compare the differential expression in CD8 $^{+}$ T cells between moderate and severe cases, we first selected upregulated genes in T cells compared to macrophages derived in patients with severe and moderate COVID-19 infection, respectively, using the MAST method with default parameters, to remove higher background noises of macrophage-specific transcripts in T cells. Only upregulated genes in T cells were considered for further DEG

analysis between moderate and severe group comparisons. The DEG results are summarized in Supplementary Table 5.

Regulatory network inference. A single-cell regulatory network for four macrophage groups was constructed with SCENIC¹⁸. Specifically, GRNBoost2 (<https://github.com/tmoerman/arboreso>) in pySCENIC v.0.9.19 was applied to infer gene regulatory networks from raw count data. Then potential direct-binding targets (regulons) were selected based on DNA-motif analysis. Finally, gene regulatory network activity for individual cells was identified. To find the regulators for macrophage groups, the regulon activity was averaged. A regulon-group heat map was generated with the R pheatmap package.

Gene functional annotation. For DEGs, GO and GSEA¹⁹ were performed with clusterProfiler²⁰, which supports statistical analysis and visualization of functional profiles for genes and gene clusters. In GSEA, 50 hallmark gene sets in MSigDB²¹ were used for annotation.

TCR V(D)J sequencing and analysis. Full-length TCR V(D)J segments were enriched from amplified cDNA from 5' libraries via PCR amplification using a Chromium Single-Cell V(D)J Enrichment kit according to the manufacturer's protocol (10x Genomics). TCR sequences for each single T cell were assembled by Cell Ranger vdj pipeline (v.3.1.0), leading to the identification of CDR3 sequence and the rearranged TCR gene. Cells with both TCR α- and β-chains were kept and cells with only one TCR chain were discarded. For cells with more than one α- or β-chain, the chain with higher UMIs or reads was kept. Then the clonality of the different T cells defined by scRNA-seq data was explored.

Calculations of the composite signature scores by single cells. Tissue-resident core signature scores was calculated for CD8⁺ T cells by a modified version of Single-Cell Signature Scorer²², which considers a core list of 31 upregulated and downregulated genes by tissue-resident T cells¹¹.

Statistics. The Student's *t*-test (t.test in R 3.6.2, two-sided, unadjusted for multiple comparisons) was used to compare tissue-resident core signature scores of the CD8⁺ T cells belonging to expanded versus nonexpanded clonotypes and patients with moderate versus severe COVID-19 infection. Effect size (Cohen's *d*) of the comparisons was calculated with R lsr v.0.5. Differences of median percentage between healthy controls, moderate and severe groups of all cell types were compared using a Student's *t*-test (two-sided, unadjusted for multiple comparisons) with R ggpubr v.0.2.5.

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

Data availability

All data used in this study, including scRNA-seq and scTCR-seq raw data, filtered expression matrix and scTCR-seq contig annotation that support the findings of this study can be accessed in GEO under the accession number [GSE145926](https://www.ncbi.nlm.nih.gov/geo/study/?study_id=GSE145926).

Code availability

The source code and software pipeline to reproduce our analyses can be accessed at https://github.com/zhangzlab/covid_balf.

References

15. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
16. Stuart, T. et al. Comprehensive Integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
17. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
18. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
19. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
20. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
21. Liberzon, A. et al. The molecular signatures database hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
22. Pont, F., Tosolini, M. & Fournie, J. J. Single-cell signature explorer for comprehensive visualization of single cell signatures across scRNA-seq datasets. *Nucleic Acids Res.* **47**, e133–e133 (2019).

Acknowledgements

We acknowledge the work and contribution of all the health providers from Shenzhen Third People's Hospital. This study was supported by the Science and Technology Innovation Committee of Shenzhen Municipality (202002073000002) and the National Natural Science Foundation of China (91442127 to Z.Z.; 81700540 to Y.L.).

Author contributions

Z.Z., S.Z. and I.A. contributed to the study design and the manuscript preparation. J.Y., J.L., F.W. and L.L. contributed to the collection of clinical specimens. M.L., G.X. and X.W. performed scRNA-seq and TCR-seq capturing and library construction. J.Z. performed qRT-PCR. L.C. contributed to cytometric bead array detection. Y.L., Y.W. and I.A. contributed to data analysis. All authors approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

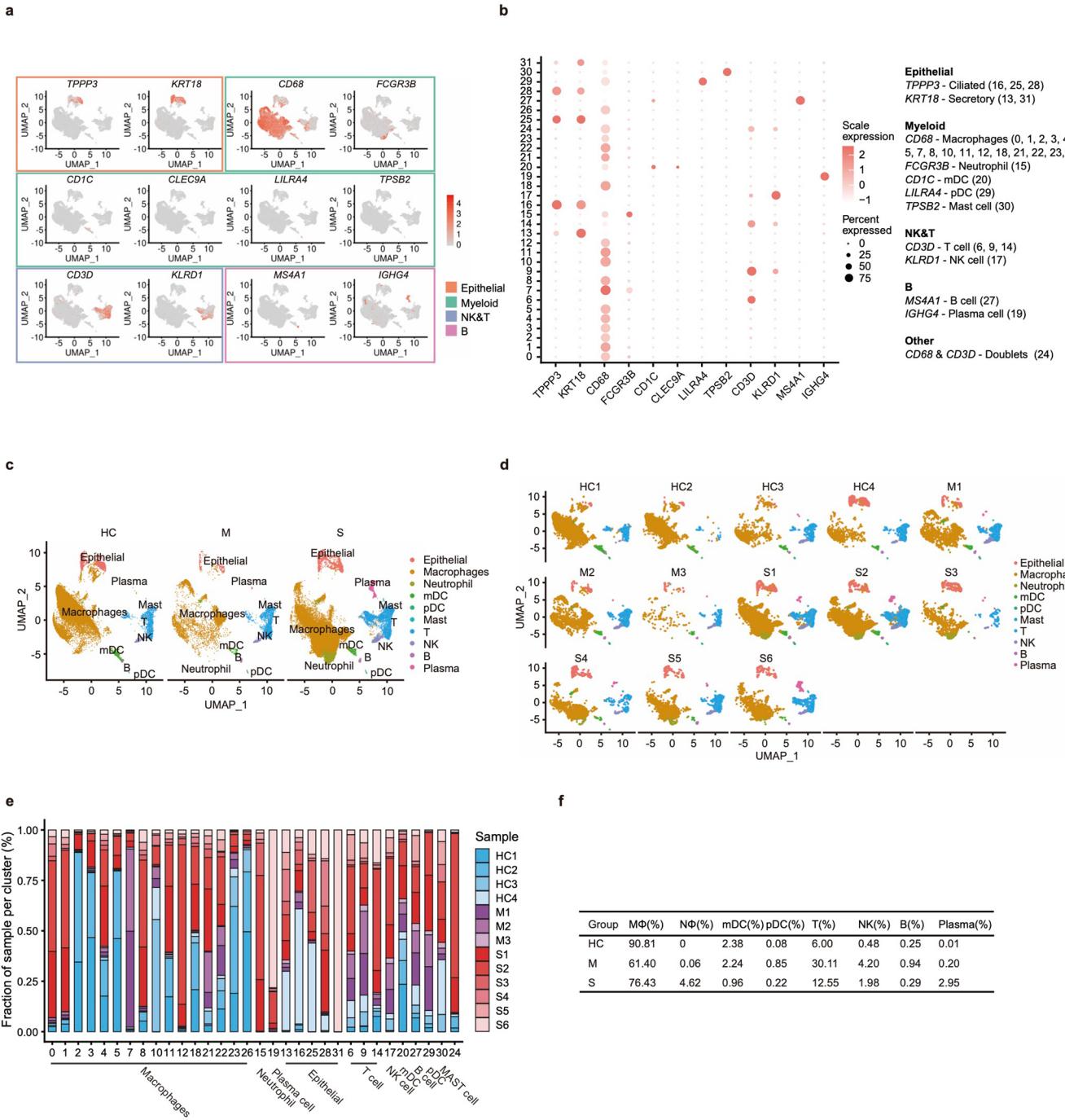
Extended data is available for this paper at <https://doi.org/10.1038/s41591-020-0901-9>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41591-020-0901-9>.

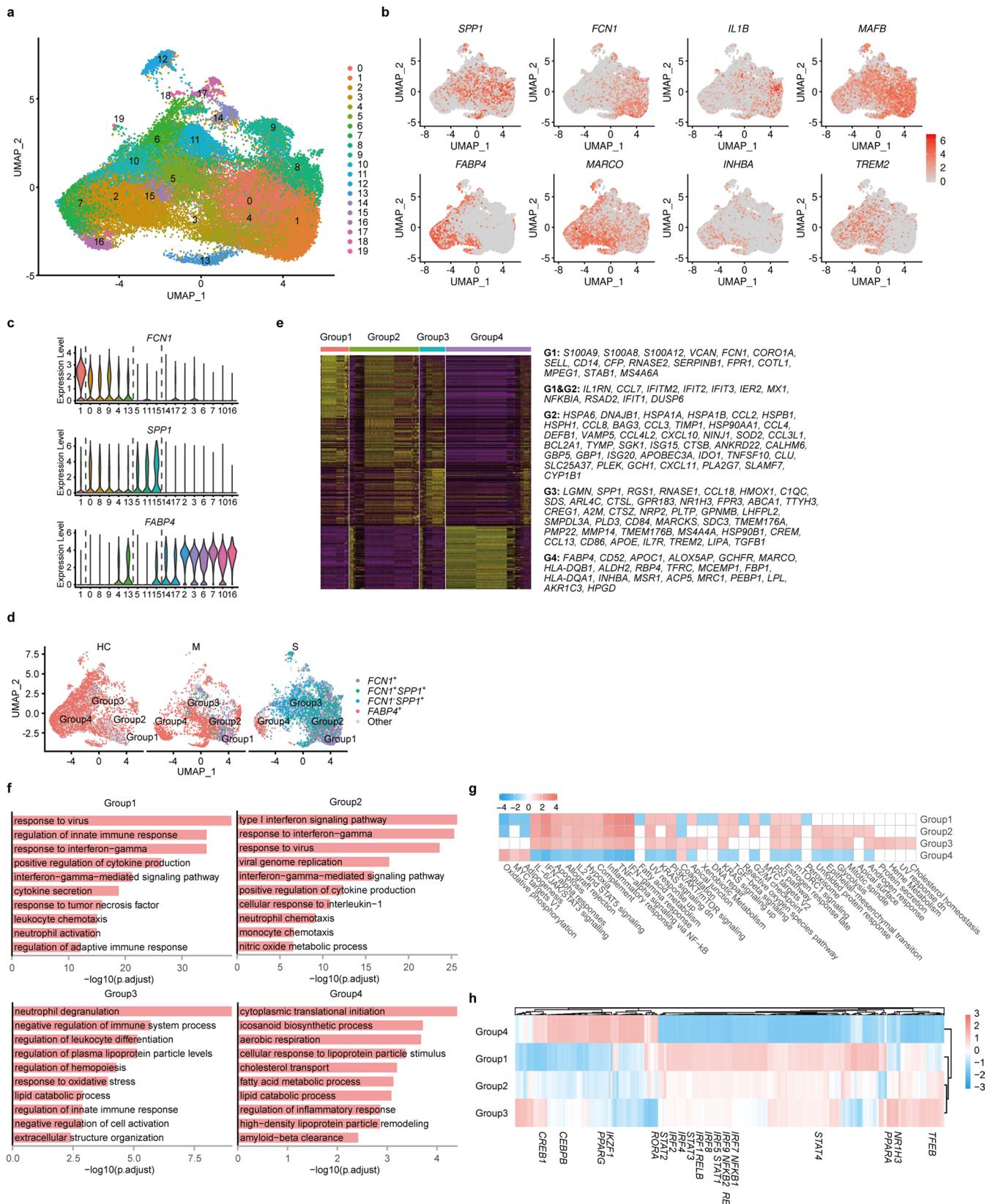
Correspondence and requests for materials should be addressed to L.L., I.A., S.Z. or Z.Z.

Peer review information Saheli Sadanand was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.

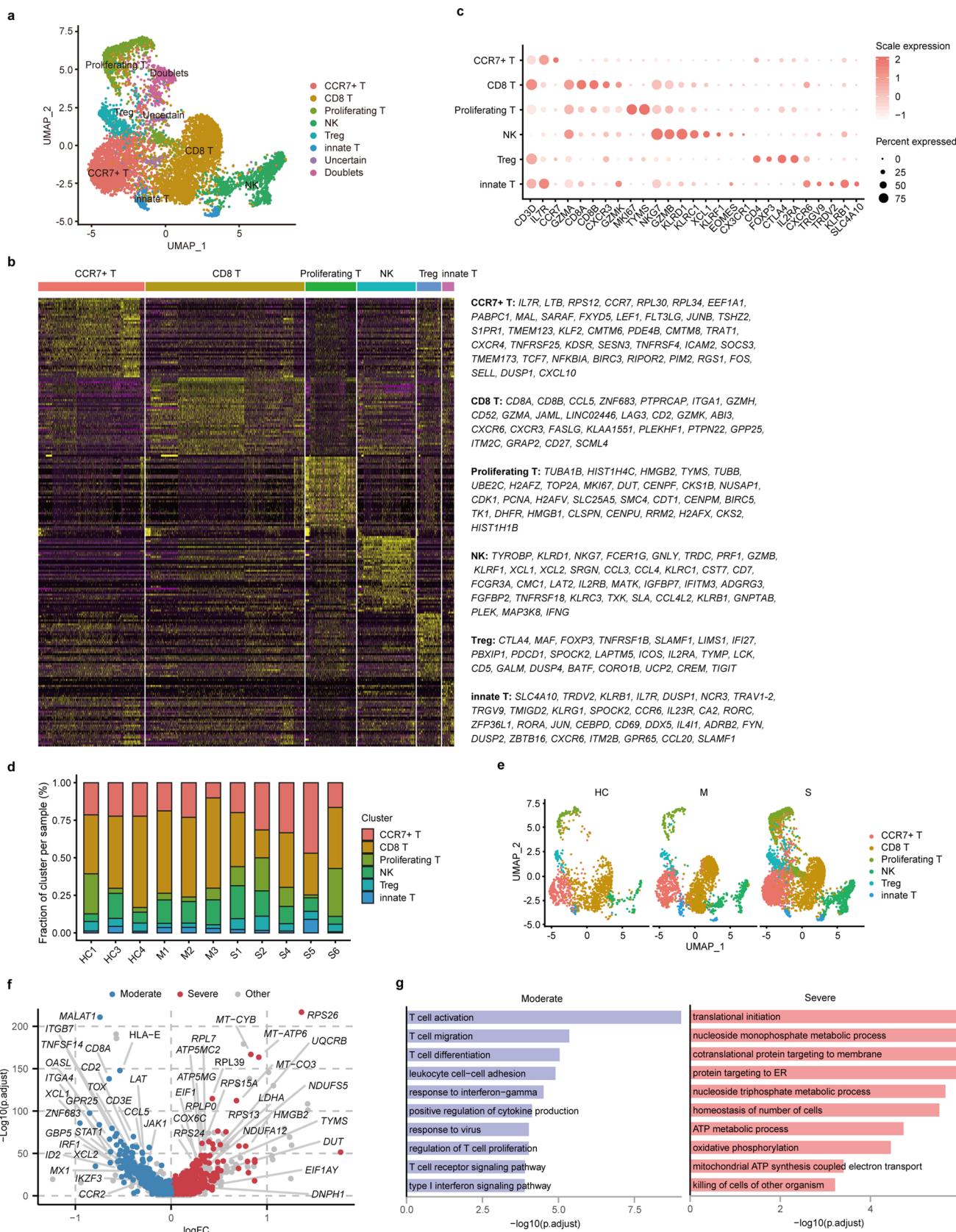


Extended Data Fig. 1 | The clustering information of immune cell landscape in the BALF. (a) UMAP projection of canonical markers for different cell types as indicated ($n = 13$). (b) A heatmap showing expression of hallmark genes by different cell clusters. The makers and their corresponding cell clusters are listed on the right side. (c) The alignment of BALF cell clusters by UMAP across control ($n = 4$), moderate ($n = 3$) and severe/critical ($n = 6$) COVID-19 cases. (d) The UMAP plots of each studied subject, including 4 controls (HC1 to HC4), 3 moderate cases (M1 to M3), and 6 severe / critical cases (S1 to S6). (e) The bar plot shows the relative contributions of each cluster by each studied subject. (f) The table shows the percentages of various immune cell populations in the BALFs among different groups of studied subjects. HC1-HC4: healthy controls, M1-M3: moderate cases, S1-S6: severe/critical cases.



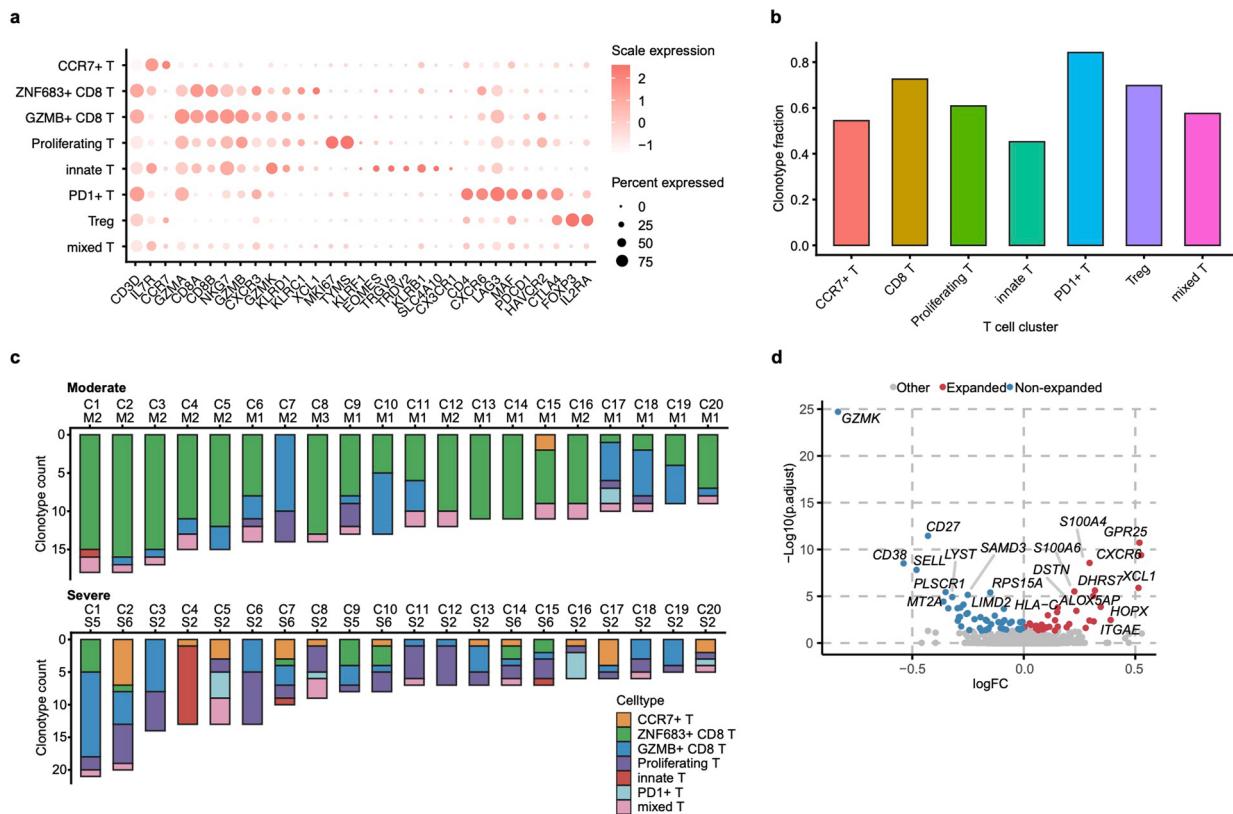
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | The macrophage markers and pathway analysis. (a) The UMAP presentation of the heterogenous clusters of BALF macrophages. Cluster 12 ($CD68^+CD3D^+$), 18 ($CD68^+IGHG4^+$) and 19 ($CD68^+IGHG4^+$) were excluded in the following analysis ($n = 13$). (b) UMAP plots showing the expression of several markers on BALF macrophages. The markers on top and bottom panels indicate the existence of heterogeneous macrophages ($n = 13$). (c) The BALF macrophage clusters were classified to 4 groups according to *FCN1*, *SPP1* and *FABP4* expression levels, indicated by grey dashed lines ($n = 13$). (d) The UMAP projection of *FCN1*, *SPP1*, and *FABP4* expressing cells among controls ($n = 4$) and patients ($n = 3$ for moderate patients, $n = 6$ for severe/critical patients). (e) The heatmaps of hierarchically clustered top 50 differentially-expressed genes (DEGs) across 4 groups of macrophages. The gene names were listed to the right. (f) The GO analysis of up-regulated DEGs showing some highlighted pathways in the four groups of macrophages ($n = 13$, hypergeometric test, adjusted p-values obtained by Benjamini-Hochberg procedure). (g) The GSEA NES heatmap of the 50 MSigDB hallmark gene sets across 4 groups of macrophages. The gene list for each group was ordered based on signal-to-noise ratio of this group compared with all other cells. Default parameters in GSEA function of clusterProfiler were used (except nPerm = 100, 000, Kolmogorov-Smirnov test). (h) Average regulon activities of 4 groups of macrophages calculated with SECENIC. PySCENIC was used to infer co-expression modules, prune modules for targets with cis regulatory footprints and calculate cellular regulon enrichment matrix with default parameters.

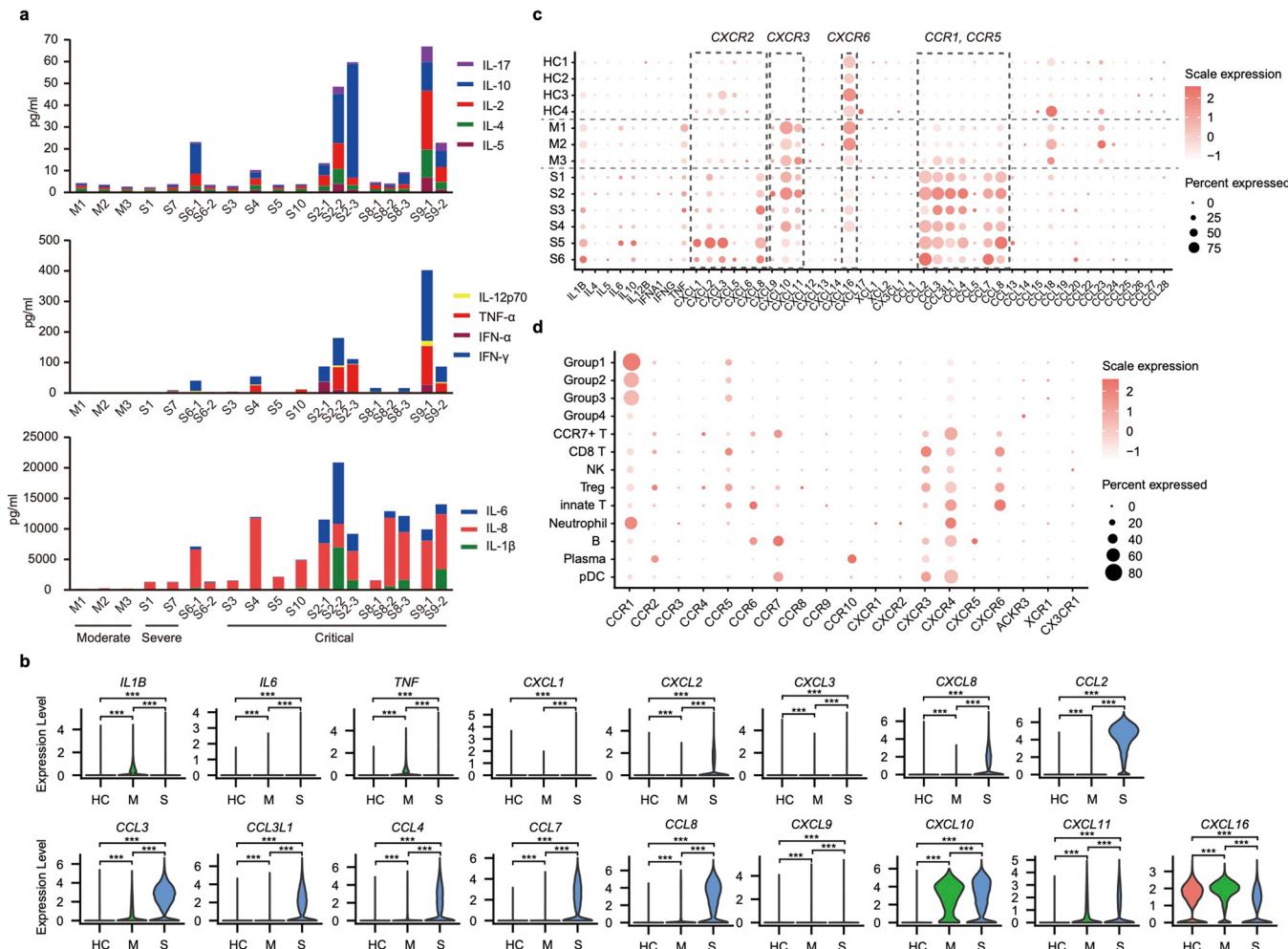


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Analysis of the T and NK lymphocytes in BALFs of COVID-19 patients. (a) The UMAP plot shows the clustering of T and NK cells in BALFs ($n = 11$, HC2 and S3 excluded). (b) Heatmap showing the top 50 differentially expressed genes by various T and NK cell clusters in the BALFs. The gene names and their corresponding cell types were listed to the right. (c) The heatmap shows the expression levels of immune typing markers on different T and NK clusters. Doublets clusters were identified based on marker expression ($CD68^+CD3D^+$ and $FCGR3B^+CD3D^+$). (d) The bar plot shows the percentages of T and NK clusters in each studied subject. (e) The UMAP plots align the T and NK cell clusters across different groups of patients. ($n = 3$ for controls, $n = 3$ for moderate patients, $n = 6$ for severe/critical patients) (f) The volcano plot shows the selected DEGs of CD8⁺ T cells from moderate vs. severe/critical COVID-19 patients. Only genes specifically upregulated in T cells compared to macrophage cells are kept. And a gene is considered significant with $\text{adjust } P < 0.05$ (P -value adjusted by false discovery rate in MAST). Other indicates non-significant genes. (g) The GO-BP enrichment analysis of DEGs of CD8⁺ T cells between moderate vs. severe/critical COVID-19 patients shows enriched pathways (selected among pathways upregulated in moderate and severe/critical cases, $\text{adjust } P < 0.05$ (hypergeometric test, adjusted P -values obtained by Benjamini-Hochberg procedure). In (f-g), moderate: $n = 3$, severe: $n = 5$.



Extended Data Fig. 4 | TCR clonotype analysis in BALFs of COVID-19 patients. **(a)** A heatmap showing the hallmark genes expressed by various T cell subsets in the BALFs. **(b)** The bar plot shows the fraction of cells in each T cell subset for which both transcriptomes and TCR clonotypes data can be characterized. **(c)** The bar plots show the individual top 20 expanded T cell clones (each bar represents a single T cell clone) in the BALFs of moderate (upper panel) and severe/critical (lower panel) COVID-19 patients. The 20 clones were aligned from left to right according to their cell numbers in each clone. The colored segments of the bar indicating which subsets the cells of that clone belong to. **(d)** The volcano plot shows the selected DEGs of expanded vs. non-expanded CD8⁺ T cells in BALFs of moderate COVID-19 cases ($n = 3$). A gene is considered significant with $\text{adjust } P < 0.05$ (P -value adjusted by false discovery rate in MAST). Other indicates non-significant genes.



Extended Data Fig. 5 | Characterization of levels of cytokine/chemokine and chemokine receptors in BALFs of COVID-19 patients. (a) The levels of selected cytokines and chemokines in the BALF were measured by cytometric bead array. Patient S6 was previously in critical condition (S6-1) but improved later (S6-2). S2, S8 and S9 are critical COVID-19 cases cared in ICU with longitudinal BALF samples, as labeled sequentially as S2-(1, 2, 3), S8-(1, 2, 3) and S9-(1, 2). The detailed information of studied subjects was described in Supplementary Tables 1 and 4. (b) The gene expression levels of selected cytokines and chemokines in macrophages from health controls ($n = 4$), moderate cases ($n = 3$) and severe/critical cases ($n = 6$). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, P -value adjusted by false discovery rate in MAST. (c) The gene expression levels of indicated cytokines and chemokines in macrophages of each studied COVID-19 patient based on the scRNA-seq data. The horizontal dashed lines separate controls, moderate and severe/critical patients. Some ligand-receptor pairs are indicated by dashed boxes. (d) The gene expression levels of chemokine receptors by various immune cell subsets, including myeloid, B cell, plasma cell, NK cell and T cell subsets based on the scRNA-seq data ($n = 13$).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Cell Ranger v3.1.0 (10X Genomics) was used to align and generate expression matrices for downstream analysis.

Data analysis

Seurat v3.1.3 was used to perform sample aggregating, dimension reduction and clustering. MAST v1.12.0 in Seurat was used to perform differential expression analysis. PySCENIC v0.9.19 was applied to infer gene regulatory networks. Gene ontology (GO), KEGG pathway analyses and Gene Set Enrichment Analysis (GSEA) were performed with R package clusterProfiler v3.14.3.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data used in this study, including scRNA-seq raw data, expression matrix and scTCR-seq contig annotation that support the findings of this study can be accessed in GSE145926. The source code and software pipeline to reproduce our analyses can be assessed at https://github.com/zhangzlab/covid_balf.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This is a exploratory study using novel scRNA-seq and scTCR-seq technology during a urgent situation. For this study, sample size of patients were limited by the ethic issues, biosafety issues, reagent availability, and the difficulty of obtaining and processing fresh BALF samples during the urgent COVID-19 outbreak. However, despite that only 3 samples for moderate group and 6 samples for severe / critical groups were analyzed, we found the changes consistently observed across individual patients, strengthening our conclusions.
Data exclusions	When performing DEG analysis between T cell subsets of moderate disease vs. severe disease, we found numerous upregulated macrophage specific genes in T cell subsets from severe COVID-19 patients. We assumed that these macrophage associated genes in T cells were from ambient contaminants and were background noise. Thus, as we stated in Methods, we selected the T cell specific and exclude the macrophage specific genes from our DEG analysis of T cell subsets between moderate vs. severe cases.
Replication	Since one healthy control dataset was borrowed from a previous published study (Morse et al., 2019), it provides us the opportunity to test the consistency between the current and previous analysis. Indeed, we found the all the previously reported clusters were identified by similar frequencies in our study. For the script identifying TCR clonotypes through CellRanger software, we confirmed its reliability are follows: First, across T cell subsets, 45%-84% of T cell subsets with overlapping scRNA and scTCR data were obtained. Second, among different cell subsets, the CCR7+ T population, showed little clonal expansion levels, consistent with the naïve phenotype. Thus, these data corroborated that our TCR analysis is correct and reproducible.
Randomization	No randomization of subjects was performed in this study. The disease severity were defined to be moderate, severe or critical, according to the "Novel coronavirus guidelines for diagnosis and treatment, Seventh Edition" by National Health Commission of China issued on March 3rd, 2020. Patients in severe and critical conditions are coded S1 to S10, while patients with moderate diseases are coded M1 to M3 in the current study.
Blinding	Patient samples were obtained during the first wave of COVID-19 patients in Shenzhen. Sample preparation was performed by laboratory personnel who were blinded to the patients status and computational analysis. The integration of patients samples and published healthy control data were performed without bias about expected results.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms		
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants		
<input type="checkbox"/>	<input checked="" type="checkbox"/> Clinical data		

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	The 9 patients examined by scRNA-seq and scTCR-seq are as follows: M1, age 36, male; M2, age 37, female; M3, age 35, male; S1, age 62, male; S2, age 66, male; S3, age 63, male; S4, age 65, Female; S5, age 57, Female; S6, age 46, Male. The 13 patients examined by CBA are as follows: in addition to M1-M3, S1-S6, other 4 patients are S7, age 52, male; S8, age 73, male; S9, age 67, male; S10, age 65, male. The 3 health control examined by scRNA-seq and scTCR-seq are as follows: HC1, age 38, female; HC2, age 24, male; HC3, age 22, male. Other clinical characteristics of the enrolled patients are described in the Methods, Supplementary Table 1 and Supplementary Table 4.
----------------------------	---

Recruitment

The study enrolled a total of 13 confirmed COVID-19 patients. The BALF samples of these patients were obtained from Shenzhen Third People's Hospital, the city's designated hospital for treating emerging infectious diseases.

Ethics oversight

The study was approved by the Ethical Committee of the Shenzhen Third People's Hospital (2020-112), China and conducted in strict accordance with the Chinese government rules and regulations for the protection of human subjects. The written informed consents were obtained from each participated subjects.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.

Study protocol

Note where the full trial protocol can be accessed OR if not available, explain why.

Data collection

Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.

Outcomes

Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.