nature research

Corresponding author(s): Zheng Zhang

Last updated by author(s): Apr 16, 2020

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 COVOD-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

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

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.*