nature portfolio

Peer Review File

Robust mucosal SARS-CoV-2-specific T cells effectively combat COVID-19 and establish polyfunctional resident memory in patient lungs

Corresponding Author: Professor Jincun Zhao

Version 0:

Decision Letter:

17th Jul 2024

Dear Professor Zhao,

Your Article, "Robust mucosal SARS-CoV-2-specific T cells effectively combat COVID-19 and establish polyfunctional resident memory in patient lungs" has now been seen by 2 referees. You will see from their comments copied below that while they find your work of considerable potential interest, they have raised quite substantial concerns that must be addressed. In light of these comments, we cannot accept the manuscript for publication, but would be very interested in considering a revised version that addresses these serious concerns.

We hope you will find the referees' comments useful as you decide how to proceed. If you wish to submit a substantially revised manuscript, please bear in mind that we will be reluctant to approach the referees again in the absence of major revisions.

We consider that it is very important that you clarify the exact number and type of samples that are being analysed in every figure, add information about the vaccine received by the subjects and other related information and clarifications as requested by the reviewers, and improve the statistics.

If you choose to revise your manuscript taking into account all reviewer and editor comments, please highlight all changes in the manuscript text file in Microsoft Word format.

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

If revising your manuscript:

* Include a “Response to referees” document detailing, point-by-point, how you addressed each referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the referees along with the revised manuscript.

* If you have not done so already please begin to revise your manuscript so that it conforms to our Article format instructions at http://www.nature.com/ni/authors/index.html. Refer also to any guidelines provided in this letter.

* Include a revised version of any required reporting checklist. It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review. A revised checklist is essential for re-review of the paper.

The Reporting Summary can be found here:
https://www.nature.com/documents/nr-reporting-summary.pdf

When submitting the revised version of your manuscript, please pay close attention to our
href="https://www.nature.com/nature-portfolio/editorial-policies/image-integrity">Digital Image Integrity Guidelines. and to the following points below:

-- that unprocessed scans are clearly labelled and match the gels and western blots presented in figures.
-- that control panels for gels and western blots are appropriately described as loading on sample processing controls

-- all images in the paper are checked for duplication of panels and for splicing of gel lanes.

Extended Data figures and tables are online-only (appearing in the online PDF and full-text HTML version of the paper), peer-reviewed display items that provide essential background to the Article but are not included in the printed version of the paper due to space constraints or being of interest only to a few specialists. A maximum of ten Extended Data display items (figures and tables) is typically permitted. When re-submitting your manuscript, please ensure that any supplementary figures and tables that are more critical to the manuscript's conclusions are converted to Extended data to increase these data's visibility.

Finally, please ensure that you retain unprocessed data and metadata files after publication, ideally archiving data in perpetuity, as these may be requested during the peer review and production process or after publication if any issues arise.

You may use the link below to submit your revised manuscript and related files:

Link Redacted

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

If you wish to submit a suitably revised manuscript we would hope to receive it within 6 months. If you cannot send it within this time, please let us know. We will be happy to consider your revision so long as nothing similar has been accepted for publication at Nature Immunology or published elsewhere.

Nature Immunology is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

Please do not hesitate to contact me if you have any questions or would like to discuss the required revisions further.

Thank you for the opportunity to review your work.

Sincerely,
Paula

Paula Jauregui, PhD
Senior Editor
Nature Immunology

Referee expertise:

Referee #1: T cells and respiratory viral infections.

Referee #2: Response to SARS-CoV-2 and scRNAseq.

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
The manuscript describes an analysis of PBMC and BAL samples from patients with COVID-19. A series of functional (ICS) and profiling assays (high parameter flow cytometry, scRNA-Seq) are performed on the samples to identify specific T cell responses and then characterize these responses by comparison between BAL and PBMC, as correlations with various clinical outcomes, and longitudinally pre and post viral clearance. In general, they observe robust effector-like signatures followed by the development of Trm profiles. Functional CD8 T cells were associated with protective features, and activation was higher in BALF than PBMCs.
The strengths of the manuscript include the valuable sample set and the careful profiling. As a reference dataset or resource this could be quite useful for the broader field. However, I have some concerns about the conclusions drawn and the representation of the data in the current form.
1) It is very difficult to understand how many subjects are represented in each figure, whether there were multiple time points from a subject plotted on the same graph, The description at the beginning of the result is very confusing. It's clear there's 24 patients with BALF plus PBMC (which will be the focus for most of the paper), and another 95 with just BALF. On line 124, it

says “a total of 122 BALF patients were collected from these patients” (where does that number come from? Repeat sampling?) including 27 samples with paired PBMC (again it’s 24 patients, so….is it repeat sampling?) and 5 longitudinal BALF samples from two individuals (is that the three extra?). Then 44 patients with PBMC samples are mentioned, but there are only 43 by the numbers. On line 141, they include an additional 114 BALF and 272 PBMC samples….where are these from? Is the 114 a subset of the 122 above?
2) I think the paper is primarily focused on 24 patients then, two of whom had BAL at more than one time point. The description on line 50 in the abstract that they have examined blood and BAL from 159 patients is thus highly misleading.
3) Throughout the rest of the paper, there is no reference to how many subjects are being discussed or analyzed—in the figure legends or in the results. This information should be clearly stated for each analysis, ideally in both places.
4) A major set of conclusions in the paper is the correlation analysis in figure 2. The figure legends and the methods do not discuss if multiple comparisons correction across all analyses were performed—I think it’s unlikely because Graphpad doesn’t do this very well, in which case the conclusions may not be robust to this correction.
5) In several figures, I think it’s possible that multiple time points from one individual are plotted in a single category—if this is the case, a mixed effects linear model should be used, or something similar to control for the repeat sampling. It’s possible I’m misunderstanding what’s plotted though in terms of sample numbers.
6) Several analyses explore the response in vaccinated vs. unvaccinated individuals, finding little differences. This contradicts several papers, including those that had access to respiratory samples. I cannot find any reference to what vaccine the subjects received—this should be very clear, as mRNA vaccines and the adeno-vectored vaccines were dominant in other studies.
7) The T cell receptor analysis has some interesting features, including the comparison of PBMC and BAL cells with the same clonotype. I think it should be clarified though—it seems that a cell was called specific if it matched a TCR from the stimulation and sequencing experiment from any subject. Is that right? That is, individuals were not just matched to their own TCRs, they were matched to anyone’s TCRs and this was largely drawn from the PBMC only group?
8) The analysis from lines 275-287 does not add very much to the manuscript and should be removed. The figure is inscrutable and without any kind of analysis for HLA this information is largely irrelevant. It could just represent HLA biases or overrepresentation by a few individuals.
9) Given the vast amount of public TCR data, it would seem like a substantially more useful analysis could be performed matching to these published data—much of it is deposited in vdjdb or IEDB.
10) Another numbers confusion: on lines 254 and 255 we learn there are 217 overlapping CD4 clones and 130 overlapping CD8 clones in BALF and PBMC, corresponding to 1186 cell and 690 cells with “high quality transcriptome information”. On line 343 it’s stated of “total of 49 overlapped T cell clones were identified”….what happened to all the other clones?
11) The RNA+ and RNA- analysis seems like it might be confounded heavily by sampling time—these are not from the same people I think since there was very limited longitudinal sampling. I think sampling time post-symptom onset should be explicitly modeled in the analysis (using a linear model or similar approach).

Reviewer #2:
Remarks to the Author:
There is a great deal of work in this manuscript which is impressive to see.

There are several different cohorts but the synergism of these could be brought out more clearly. The size of the cohort needs to be accurately defined - i.e. in reality only 24 have BALF and PBMC.

Details of the donors could be expanded. Is vaccination status known only for 46% (Fig 1). It appears the other donors are assumed not vaccinated.

All but 1 fatal cases are vaccinated and relatively young. Were they co-morbidities?
Which vaccine was used?
Is there prior demographic information or data on prior infection?

From the ‘current’ infection there are no details regarding what variant this was.

For their functional assays the authors are comparing a large pool of peptides, which is across multiple antigens, and state that NI and BI are comparable, but this does not take into account that pre-existing Spike-specific T cells (from vaccine) may form a different proportion of the whole responses following vaccine.
Also as this is a relatively late time point i.e. +10 days from symptom onset, it is not a surprise that the NI would catch up with an pre-existing resident or circulating S-specific T cells. Likewise given the timepoint it is less surprising that T cells are at the site of infection.

Version 1:

Decision Letter:

Our ref: NI-A38064A

19th Nov 2024

Dear Dr. Zhao,

Thank you for submitting your revised manuscript "Robust mucosal SARS-CoV-2-specific T cells effectively combat COVID-19 and establish polyfunctional resident memory in patient lungs" (NI-A38064A). It has now been seen by the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Immunology, pending minor revisions to comply with our editorial and formatting guidelines.

We will now perform detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements in about a week. Please do not upload the final materials and make any revisions until you receive this additional information from us.

If you had not uploaded a Word file for the current version of the manuscript, we will need one before beginning the editing process; please email that to immunology@us.nature.com at your earliest convenience.

Thank you again for your interest in Nature Immunology Please do not hesitate to contact me if you have any questions.

Sincerely,
Paula

Paula Jauregui, PhD
Senior Editor
Nature Immunology

Reviewer #1 (Remarks to the Author):

The reviewers have clarified most of the points of confusion from the original manuscript. I appreciate the additional analyses and the more rigorous statistical tests applied throughout. I think this will be a useful dataset for the field.

Reviewer #2 (Remarks to the Author):

This is an interesting paper with technology that is at the forefront of current capability.

The strengths include novel insights into the properties of virus-specific tissue resident cells. As detailed in the manuscript these include effector phenotype, potent function, no correlation with peripheral populations and functional capacity despite checkpoint expression. There is a correlation of the CD8+ subset with viral control.
The cohort is very extensive, albeit somewhat complex.
The new figure 1 is helpful.

Reviewer 1

The manuscript describes an analysis of PBMC and BAL samples from patients with COVID-19. A series of functional (ICS) and profiling assays (high parameter flow cytometry, scRNA-Seq) are performed on the samples to identify specific T cell responses and then characterize these responses by comparison between BAL and PBMC, as correlations with various clinical outcomes, and longitudinally pre and post viral clearance. In general, they observe robust effector-like signatures followed by the development of Trm profiles. Functional CD8 T cells were associated with protective features, and activation was higher in BALF than PBMCs.

The strengths of the manuscript include the valuable sample set and the careful profiling. As a reference dataset or resource this could be quite useful for the broader field. However, I have some concerns about the conclusions drawn and the representation of the data in the current form.

1. It is very difficult to understand how many subjects are represented in each figure, whether there were multiple time points from a subject plotted on the same graph, The description at the beginning of the result is very confusing. It's clear there's 24 patients with BALF plus PBMC (which will be the focus for most of the paper), and another 95 with just BALF. On line 124, it says "a total of 122 BALF patients were collected from these patients" (where does that number come from? Repeat sampling?) including 27 samples with paired PBMC (again it's 24 patients, so….is it repeat sampling?) and 5 longitudinal BALF samples from two individuals (is that the three extra?).

Then 44 patients with PBMC samples are mentioned, but there are only 43 by the numbers. On line 141, they include an additional 114 BALF and 272 PBMC samples….where are these from? Is the 114 a subset of the 122 above?

Response: We recognize that our initial explanation may not have been sufficiently clear. To enhance the readability of patient and sample information, we have provided a summary as shown in Fig.1 (now illustrated as Fig. 1b in the revised manuscript). We have also updated this information in the revised text as follows: Twenty-four patients (PT1–PT24) provided 27 paired BALF and PBMC samples, with PT1 contributing samples at three timepoints and PT7 at two timepoints; 95 patients (PT25–PT119) provided only BALF samples, each with a single timepoint; and 44 patients (PT1, PT4, PT5, PT8, and PT120–PT159) contributed a total of 253 longitudinal blood samples. (lines 156-160)

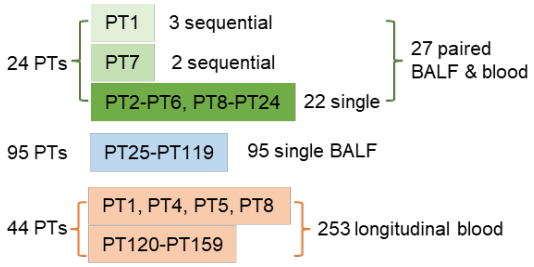

Fig. 1 Patient and sample information. Summary of patients who provided BALF and PBMC samples included in this study. For additional details, see Extended Data Table 1. PTs represents patients.

Additionally, we have added a summary to indicate the number of samples processed for each assay as shown in Fig.2 (now illustrated as Extended Data Fig. 1a in the revised manuscript).

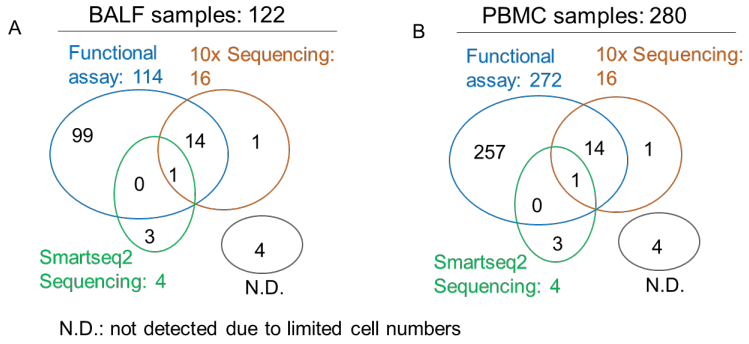

Fig. 2 Schematic overview of samples processed for assays in the study. Schematic overview illustrating the comprehensive analysis workflow of BALF (**A**) and PBMC (**B**) samples. The analyses include flow cytometry, 10x single-cell RNA and TCR sequencing, and Smart-seq2 single-cell sequencing.

In the revised manuscript, we have included detailed information about samples and corresponding subject data in the relevant figure and result descriptions. Furthermore, we have revised the demographic section (lines 144-160) and updated the patient enrollment and sample collection descriptions in the *Materials and Methods* (lines 1526-1542) for better readability and comprehension.

Therefore, the statement "A total of 122 BALF samples were collected from these patients, including 27 samples with paired PBMC and 5 longitudinal BALF samples from 2 individuals" on line 124, refers to the 27 samples from PT1-PT24 and 95 samples from PT25-PT119.

Regarding the 44 patients with longitudinal PBMC samples mentioned on line 127, these include PT1, PT4, PT5, PT8, and PT120-PT159. We inadvertently omitted PT8 in the previous version, which has now been corrected (lines 156-160).

On line 141, the "114 BALF and 272 PBMC samples" are from the 159 COVID-19 patients (PT1-PT159). The 114 BALF samples are a subset of the 122 BALF samples mentioned earlier. We have updated this description in the revised manuscript (lines 197-200).

2. I think the paper is primarily focused on 24 patients then, two of whom had

BAL at more than one time point. The description on line 50 in the abstract that they have examined blood and BAL from 159 patients is thus highly misleading.

Response: We recognize that the description in the abstract was misleading. The subject and sample details for our cohort have been summarized in the revised Fig 1b and 1c, as noted in response #1. To avoid confusion, we have revised the text as follows: "Here, we conducted integrated single-cell profiling of SARS-CoV-2-specific T cells in 122 bronchoalveolar lavage fluid (BALF) and 280 blood samples from 159 patients, including 27 paired BALF and blood samples from 24 patients." (lines 50-52)

3. Throughout the rest of the paper, there is no reference to how many subjects are being discussed or analyzed—in the figure legends or in the results. This information should be clearly stated for each analysis, ideally in both places.

Response: We thank the reviewer for his/her suggestion. In response, we have included detailed subject information within each result section and figure legend in the revised manuscript, ensuring clarity and transparency throughout our analysis.

4. A major set of conclusions in the paper is the correlation analysis in figure 2. The figure legends and the methods do not discuss if multiple comparisons correction across all analyses were performed—I think it's unlikely because Graphpad doesn't do this very well, in which case the conclusions may not be robust to this correction.

Response: We acknowledge the importance of correlation analysis between the specific T cell responses and the viral loads as well as clinical parameters in understanding the role of airway T cell function in infection and disease development. We agree with reviewer's comments that the analysis and conclusions should be approached with caution.

To address the statistical concerns raised, we consulted with a professional statistician to review and analyze all the statistical data and methods used in the manuscript. In the revised manuscript, we have employed Spearman correlation analysis and applied the Benjamini-Hochberg (BH) correction method using R (version4.4.0) software to appropriately address multiple comparisons, enhancing the robustness of our conclusions. (Fig. 2k-l and legends, lines 1870-1879, 2147-2150)

5. In several figures, I think it's possible that multiple time points from one individual are plotted in a single category—if this is the case, a mixed effects linear model should be used, or something similar to control for the repeat

sampling. It's possible I'm misunderstanding what's plotted though in terms of sample numbers.

Response: We appreciate the reviewer's insightful feedback. In response, we have revised all figures in the manuscript and modified the statistical analyses as suggested by the reviewer. To model the dynamics of antigen-specific T cells over time, we fitted a linear mixed-effects model using restricted maximum likelihood (REML), applying t-tests based on Satterthwaite's approximation for degrees of freedom. Furthermore, we have included subject numbers alongside sample counts and provided a more detailed description in each figure and legend. (Fig 2d-e, Fig 6a and legends; lines 1870-1879)

6. Several analyses explore the response in vaccinated vs. unvaccinated individuals, finding little differences. This contradicts several papers, including those that had access to respiratory samples. I cannot find any reference to what vaccine the subjects received—this should be very clear, as mRNA vaccines and the adeno-vectored vaccines were dominant in other studies.

Response: We acknowledge the oversight regarding vaccine information in the manuscript. All vaccinated subjects in our study received inactivated COVID-19 vaccines (*CoronaVac* or *BBIBP-CorV*), and we have added this information in the revised version (lines 150-153).

Previous studies indicate that prior vaccination can induce memory T cells in periphery, which are rapidly recalled upon breakthrough infections (PMID: 35172129, 37735592, 37037791, 36582222). While mRNA and adeno-vectored vaccines are known to enhance preexisting mucosal T cell responses from prior infections, several studies have shown a failure to induce pathogen-specific T cell memory in mucosal areas following intramuscular vaccination in uninfected individuals (PMID: 35972472, 35857583, 37884506), aligning with our findings. This likely results from the intramuscular route primarily promoting systemic immune responses, which we discussed in detail in the original manuscript (lines 1462-1470).

Additionally, as suggested by reviewer #2 (question #8), differences in airway specific T cell responses between vaccinated and unvaccinated individuals may manifest early in the infection but become overshadowed by the infection itself, leading to minimal observable differences. To explore this hypothesis, we selected 12 BALF samples from 12 patients sampled within 10 days post onset (3-10 dpo), comprising 6 unvaccinated (NI) and 6 vaccinated (BI) samples. Our results indicated no significant differences between the NI and BI groups, consistent with our prior findings, as shown in Fig.3.

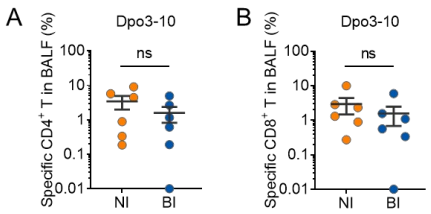

Fig. 3 SARS-CoV-2-specific T cells in airway at early infection stage. Comparison of airway SARS-CoV-2-specific CD4$^+$ (**A**) and CD8$^+$ (**B**) T cells between NI and BI groups.

To further substantiate this, we analyzed NI and BI samples stratified by mild and severe symptoms across early (0-2 weeks post-onset) and late (>2 weeks post-onset) infection stages and updated the result as shown in Fig.4 (now illustrated as Fig 2h-i in the revised manuscript). A Kruskal–Wallis test revealed no statistical differences among these groups. (lines 275-277)

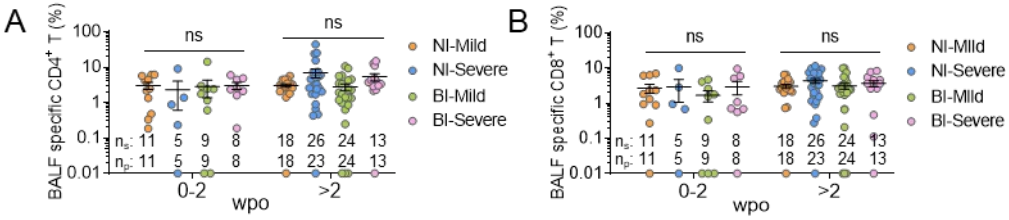

Fig. 4 Comparison of SARS-CoV-2-specific T cells in airway among various groups. Comparison of specific CD4$^+$ (**A**) and CD8$^+$ (**B**) T cell responses in BALF between mild and severe patients with natural infection (NI) and breakthrough infection (BI). Data are shown as mean ± SEM. The number of samples (n_s) and patients (n_p) is indicated in each panel. Statistical analysis was conducted using Kruskal–Wallis test.

7. The T cell receptor analysis has some interesting features, including the comparison of PBMC and BAL cells with the same clonotype. I think it should be clarified though—it seems that a cell was called specific if it matched a TCR from the stimulation and sequencing experiment from any subject. Is that right? That is, individuals were not just matched to their own TCRs, they were matched to anyone's TCRs and this was largely drawn from the PBMC only group?

Response: We would like to clarify that a T cell is considered "specific" if it corresponds to a T cell receptor (TCR) identified in the stimulation and sequencing experiments of the same subject but not in those of other subjects. We have provided a more detailed description of this methodology in the revised version of the manuscript. (Fig. 4a and legends, lines 519-552, 1818-1821).

8. The analysis from lines 275-287 does not add very much to the manuscript and should be removed. The figure is inscrutable and without any kind of

analysis for HLA this information is largely irrelevant. It could just represent HLA biases or overrepresentation by a few individuals.

Response: We thank the reviewer's advice. We have removed these results in the revised version.

9. Given the vast amount of public TCR data, it would seem like a substantially more useful analysis could be performed matching to these published data—much of it is deposited in vdjdb or IEDB.

Response: We appreciate the reviewer's advice. In our study, we matched TCR sequences to publicly available SARS-CoV-2-specific TCR repertoires from the IEDB database. As most published TCR sequences originate from single-chain data, we utilized TRBV sequences for alignment, applying a total of 9,978 TRBV sequences for $CD4^+$ T cells and 133,118 for $CD8^+$ T cells from public sources. The matching analysis revealed that 25 $CD4^+$ and 31 $CD8^+$ T cell clones in BALF samples, as well as 19 $CD4^+$ and 33 $CD8^+$ clones in PBMC samples, shared TRBV sequences with the public repertoire. Among the virus-specific clones identified via autologous short-term T cell line assays, 15 $CD4^+$ and 3 $CD8^+$ clones in BALF, as well as 7 $CD4^+$ and 5 $CD8^+$ clones in PBMC, shared TRBV sequences with the public database, as shown in the figure below. We would like to emphasize that our definition of specific T cell clones is based on the matching of both TRAV and TRBV chains, thereby enhancing the specificity of our findings. This result has been shown in Fig. 5 and also included in the revised Fig 4i. (lines 569-651)

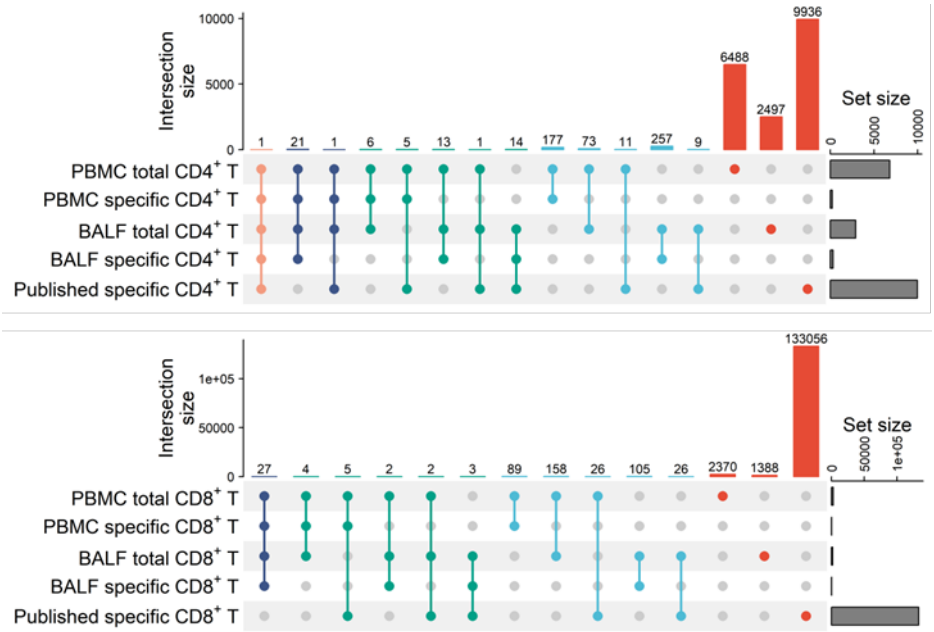

Fig. 5 Upset plots of shared TRBV clonotypes. Upset plots illustrating the TRBV clonotypes shared among virus-specific and total T cells identified from single-cell sequencing data in BALF and PBMC samples, along with published SARS-CoV-2-specific TCR sequences. Colored dots connected by lines represent

shared clonotypes. The horizontal bar graph shows the total number of TCR clonotypes for each cluster intersection, while the vertical bar graph indicates the number of clonotypes shared across overlapping clusters.

10. Another numbers confusion: on lines 254 and 255 we learn there are 217 overlapping CD4 clones and 130 overlapping CD8 clones in BALF and PBMC, corresponding to 1186 cell and 690 cells with "high quality transcriptome information". On line 343 it's stated of "total of 49 overlapped T cell clones were identified"….what happened to all the other clones?

Response: We apologize for the confusion. Our current definition of specific T cell clones employs a stringent criterion requiring simultaneous alignment of both the α and β TCR sequences. Following this, we implemented rigorous quality control measures (detailed in the methods, lines 1790-1803) on the transcriptome sequencing data to ensure the accuracy and quality of gene expression analysis for specific T cells. Clones that did not meet these matching criteria or failed the transcriptome quality filters were systematically excluded from our subsequent analyses. Upon careful review, we have updated the overlapping clone numbers between BALF and PBMC for both total T cells and virus-specific T cells, as shown in fig. 6 below. These corrections are also reflected in the revised manuscript (Fig 4b-c, lines 514-518, 552-555).

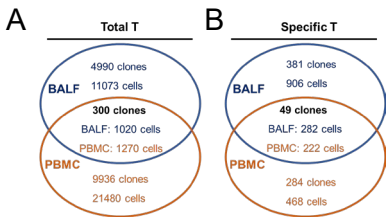

Fig. 6 Overlapping clones and corresponding cell numbers between BALF and PBMC. Venn diagrams illustrating the number of overlapping and unique TCR clones, along with the corresponding total T cell counts (**A**) and SARS-CoV-2-specific T cell counts (**B**), between BALF and PBMC.

11. The RNA+ and RNA- analysis seems like it might be confounded heavily by sampling time—these are not from the same people I think since there was very limited longitudinal sampling. I think sampling time post-symptom onset should be explicitly modeled in the analysis (using a linear model or similar approach).

Response: We appreciate the reviewer's suggestion. In response, we have plotted the specific T cell responses at RNA+ and RNA- stages over time post-symptom onset, as shown in Fig.7. A linear mixed-effects model fitted by REML, with t-tests based on Satterthwaite's method, was employed for the analysis. Results show that specific T cell responses are higher during the RNA- stage compared to the RNA+ stage, even after accounting for the sampling time

factor. This difference is particularly notable during the 14-28 dpo interval, where sample numbers in the two groups are comparable. Updated analysis is included in the revised manuscript (Fig 6a-b, lines 1005-1009).

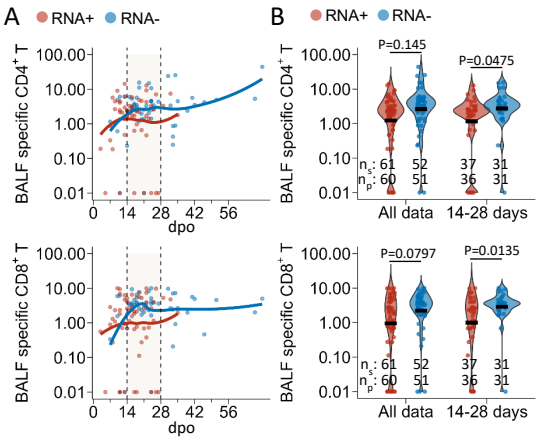

Fig. 7 Virus-specific T cell responses at RNA+ and RNA- stages over time post-symptom onset. (**A**)Kinetics of specific T cells in BALF over time post-symptom onset (dpo) during viral shedding (RNA+, red, ns/np: 61/60) and after viral clearance (RNA-, blue, ns/np: 52/51). Data were analyzed using a linear mixed-effects model fit by REML, with t-tests based on Satterthwaite's method. (**B**) Comparison of specific T cells in BALF between RNA+ and RNA− stages, encompassing all data from panel a, as well as samples collected during 14–28 dpo. Data are presented as mean ± SEM, with the number of samples (ns) and patients (np) indicated. Statistical comparisons were performed using a linear mixed-effects model fit by REML, with t-tests based on Satterthwaite's method.

Reviewer 2

(Remarks to the Author)
There is a great deal of work in this manuscript which is impressive to see.

1. There are several different cohorts but the synergism of these could be brought out more clearly. The size of the cohort needs to be accurately defined - i.e. in reality only 24 have BALF and PBMC.

Response: We recognize that our initial explanation was not sufficiently clear, as also noted by Reviewer 1. In response, we have revised the description and added a summary figure to present comprehensive subject and sample information in our cohort. Please refer to the updated description in response to Reviewer 1, Question #1 for further details.

2. Details of the donors could be expanded. Is vaccination status known only for 46% (Fig 1). It appears the other donors are assumed not vaccinated.

Response: In our study, we thoroughly documented the vaccination history of all patients, as summarized in Extended data Table 1. Overall, 34.6% (55/159) of the patients received inactivated COVID-19 vaccines (*CoronaVac* or *BBIBP-CorV*), while 65.4% (104/159) were unvaccinated at the time of natural infection. Our cohort was established between January 2020 and January 2023. Patients enrolled before July 2021 had not received any COVID-19 vaccine, whereas those enrolled between 2022 and 2023 comprised both vaccinated and unvaccinated individuals. We have updated and expanded donors' information in the revised manuscript (lines 150-153, 1526-1533).

3. All but 1 fatal cases are vaccinated and relatively young. Were they co-morbidities?

Response: Upon further review of the results, we have compiled the comorbidity information and vaccination history for the nine fatal cases, presented in the table below. Among these cases, eight had underlying medical conditions or comorbidities, which are detailed in the table. The remaining patient, PT14, experienced superinfection with bacteria and fungi, leading to microbial sepsis. This comprehensive comorbidity information has been updated in the revised Fig 1a and Extended Data Table 1.

PT#	Vaccination history (1: vaccinated; 0: unvaccinated)	Age	Comorbidity/ underlying medical condition
PT9	1	80	Hypertension, Coronary Artery Atherosclerotic Heart Disease post-PCI
PT13	1	85	Type 2 diabetes
PT14	1	45	None
PT15	1	40	Type 1 diabetes
PT16	1	46	Chronic Nephrotic Syndrome, Allograft Kidney Transplantation Status, Hyperuricemia, Hyperlipidemia
PT24	1	67	Type 2 diabetes, Hypertension, End-Stage Renal Disease (ESRD) due to chronic kidney disease (CKD) with Uremia
PT29	0	86	Chronic Kidney Disease (CKD), Hepatic Cyst
PT53	1	84	Renal Angiomyolipoma, Anemia
PT151	0	68	Hypertension, Coronary disease, COPD

4. Which vaccine was used?

Response: All breakthrough infection patients in our cohort received inactivated vaccines (*BBIBP-CorV* or *CoronaVac*). This information has been included in the revised version of the manuscript (lines 150-153).

5. Is there prior demographic information or data on prior infection?

Response: All patients experienced their primary SARS-CoV-2 infection at the time of enrollment. (lines 150-153)

6. From the 'current' infection there are no details regarding what variant this was.

Response: Patients were enrolled between January 2020 and January 2023, encompassing those infected with the WT, Delta, and BA.5 strains of SARS-CoV-2. This detail is now included in the revised manuscript as follow: "Of the 159 patients (PT1–PT159), 44 (PT1, PT4, PT5, PT8, PT120–PT159) were recruited during the Wuhan strain outbreak, 11 (PT2, PT3, PT6, PT7, PT17–PT23) during the Delta variant wave, and 104 (PT9–PT16, PT24–PT119) during the Omicron BA.5 subvariant surge in China.". (lines 145-148, 1526-1533)

7. For their functional assays the authors are comparing a large pool of peptides, which is across multiple antigens, and state that NI and BI are comparable, but this does not take into account that pre-existing Spike-specific T cells (from vaccine) may form a different proportion of the whole responses following vaccine.

Response: In our cohort, breakthrough infection patients were vaccinated with inactivated COVID-19 vaccines. The vaccine-induced specific T cells can recognize not only Spike proteins but also non-Spike proteins such as Nucleocapsid, Membrane, and Envelope proteins. Therefore, employing a multi-antigen peptide library in this study takes into account the pre-existing specific T cell responses from vaccination.

8. Also as this is a relatively late time point i.e. +10 days from symptom onset, it is not a surprise that the NI would catch up with an pre-existing resident or circulating S-specific T cells. Likewise given the timepoint it is less surprising that T cells are at the site of infection.

Response: This issue parallels question #6 raised by Reviewer 1. One potential explanation for the minimal observable differences in airway-specific T cell responses between vaccinated and unvaccinated individuals is that these differences may initially emerge during infection but later be overshadowed by the infection itself. To test this hypothesis, we analyzed 12 BALF samples from

12 patients collected within 10 days post-symptom onset (3-10 days post-onset), comprising 6 unvaccinated (NI) and 6 vaccinated (BI) samples. The results indicated no significant differences between the NI and BI groups, consistent with our previous findings (see Fig. 3).

To further substantiate this, we analyzed NI and BI samples stratified by mild and severe symptoms across early (0-2 weeks post-onset) and late (>2 weeks post-onset) infection stages and updated the result as shown in Fig.4 (now illustrated as Fig 2h-i in the revised manuscript). A Kruskal–Wallis test revealed no statistical differences among these groups. (lines 275-277)

Another plausible factor contributing to the minimal difference between the NI and BI groups could be the administration route of the inactivated COVID-19 vaccines. Intramuscular vaccination primarily induces systemic immune responses rather than local pulmonary T cell memory. Consequently, the benefits of prior vaccination in enhancing the early recall of specific memory T cells may be less evident in the airway. We have also addressed this issue in the revised manuscript (lines 1462-1470). Moreover, since we did not collect samples from the BI group prior to breakthrough infection, we could not determine whether the vaccines had effectively induced peripheral T cell memory in these individuals.

Responses to reviewers' comments

Reviewer #1:
Remarks to the Author:
The reviewers have clarified most of the points of confusion from the original manuscript. I appreciate the additional analyses and the more rigorous statistical tests applied throughout. I think this will be a useful dataset for the field.

Thank you.

Reviewer #2:
Remarks to the Author:
This is an interesting paper with technology that is at the forefront of current capability.

The strengths include novel insights into the properties of virus-specific tissue resident cells. As detailed in the manuscript these include effector phenotype, potent function, no correlation with peripheral populations and functional capacity despite checkpoint expression. There is a correlation of the CD8+ subset with viral control.
The cohort is very extensive, albeit somewhat complex.
The new figure 1 is helpful.

Thank you.