natureresearch

Peer Review Information

Reviewer Comments & Decisions:

Decision Letter, initial version:

Subject: Decision on Nature Immunology submission NI-RS30805
Message: 26th Oct 2020

Dear Dr Raychaudhuri,

Your Resource, "Multimodal profiling of 500,000 memory T cells from a tuberculosis cohort identifies cell state associations with demographics, environment, and disease" has now been seen by 3 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.

While the dataset is comprehensive and generally felt to the a high quality analysis (though there are some seemingly important questions from Ref. 1 in particular about the analysis), none of the Refs. is convinced with a key contention of the data i.e. the relevance of C12. At the moment the Referees feel it's unclear whether this cluster is functionally relevant to the Mtb response. In particular Ref. 2 requests analysing the Ag-specificity of this cluster. Many in the TB field are actively searching for phenotypes of protective T cells to target in TB vaccine development, and the the Refs. felt the data stops just shy of asking if these cells are pathogen-specific or likely just an irrelevant correlating bystander. MTB-peptide megapools might allow for the careful detection of mycobacteria-specific T cells in diverse human populations. Unfortunately without further insights into the understanding the significance of C12 I strongly suspect these Referees will not be supportive.

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

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 Resource 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-research/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.

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: [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,

Zoltan Fehervari, Ph.D.
Senior Editor
Nature Immunology

The Macmillan Building
4 Crinan Street
Tel: 212-726-9207
Fax: 212-696-9752
z.fehervari@nature.com

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
In their manuscript, Nathan et al., describe the results of a study that aimed to define "memory T cell states associated with demographic factors, environment, and baseline TB progression risk".
The primary analyses utilized CITE-seq to identify characteristics of memory T cells (CD45RO+) in healthy (referred to as "steady state") Peruvian participants who had a prior episode of tuberculosis, diagnosed >4 years before the current analysis as part of a longitudinal study that followed contacts of TB patients, and a set of controls from the same longitudinal study, who did not develop disease during follow-up. They identified 31 clusters of memory T cells by analysing about 500,000 cells from 259 individuals and sought to find associations between abundance of these clusters and 38 demographic and environmental covariates. These analyses showed that age, gender, season of blood draw and ancestry were associated with abundance of certain memory T cell clusters. The primary result was identification of a cluster of memory T cells (called C-12) that was phenotypically consistent with polyfunctional Th17 cells based on a combination of surface protein and mRNA markers and which was reduced in abundance and function in participants who had previously had TB.
To follow up this discovery, they devised a strategy to identify this C-12 cell subset in published whole transcriptome datasets such as (1) whole blood RNA microarrays from individuals who were healthy (uninfected), latently infected, or had active disease or (2) RNA-seq data from sorted T cells from South African M.tb-infected adolescents who either progressed to TB or did not. This they did by defining a C-12 gene expression score (and a

C-11 "negative control signature") using their single cell CITE-seq gene expression data from sorted memory T cells and then applying this gene expression score to compute abundance within the whole transcriptome datasets. These data supported the finding that the C-12 cell subset is lower in abundance during active disease than in healthy (uninfected) or latently infected individuals, and that this subset fully "recovers" to the level of BCG+ healthy donors following antibiotic TB treatment, and showed that progressors have lower abundance of the C-12 cell subset than non-progressors. They complete the story with immunological assessments to show that these cells can be defined as CD4+CD26+CD161+CCR6+ memory T cells and that they expressed IL-17 and IL-22 along with IFNg and TNF (and some other cytokines). The paper addresses an important and potentially clinically significant issue, namely immune responses that correlate with protection against TB. It is well written and the data are mostly presented clearly and comprehensively. The supplementary information is very useful to fully understand the extensive analyses and results. The paper has a number of weaknesses that are pointed out below:

It is not clear why the authors frame their paper around the concept of baseline TB progression risk. Critically, several lines of argument are presented to support the inference that the reduction in C-12 causally increases TB risk rather than the disease itself causing lower C-12. The problem is that the study design and the data as presented do not sufficiently support this claim for the following reasons:
By comparing people who have had TB before with those who have not does not mean that the findings can be inferred to have existed before the TB diagnosis (which would be required to support the baseline TB progression risk claim). In fact, it is well known that individuals who have had TB are more likely to develop disease again. While many factors could underlie the higher risk of TB, previous studies have demonstrated immunological sequelae that persist for years after completion of TB treatment (e.g. DOI: 10.1164/rccm.201706-1208OC). The authors seem to rule out this possibility when it is much more likely than persistence of a subtle difference in a minor T cell subset that existed before TB was diagnosed or treated.
The authors set out to show that patients with active TB have significantly lower levels of C-12 cells by inferring this using a newly developed C-12 gene expression score in microarray data (which has its own limitations - see further below). This analysis does indeed show that this was observed in TB patients. However, the same analysis shows that the gene expression score fully recovers to levels observed in healthy controls after successful TB treatment. If these cells fully recover after TB treatment, why would the effect of prior TB disease still be detectable more than 4 years in their discovery cohort? This result is at odds with their primary discovery and weakens the strength of the discovery.
The TB risk claim would be markedly strengthened if their analysis of TB progressors were to show that the C-12 subset was depleted before evidence of TB disease. In fact they state in line 459 that "we demonstrate in previously published cohorts
that C-12's gene expression signature is reduced in active TB cases compared to latently infected controls even prior to progression". However, they do not provide any detail or evidence about when the disease process commences in the progressors. This missing information dramatically weakens their argument because onset of disease is likely to start months (perhaps more than a year?) before TB is diagnosed (in fact, this was a main finding in the cited paper by Scriba at al., Plos Pathogens 2017) and is therefore certainly present and able to influence the =composition of memory T cells. The analysis of these progressor data could be made stronger by only considering progressor samples far from TB diagnosis, when the effect of TB on the immune system is either not

present (as there is no TB) or less (as TB is less advanced and/or been present for less time).

Regardless of this, it should be further pointed out that the difference in C-12 gene expression signature between progressors and controls was very modest, and apparently not present when discrete time windows before Tb diagnosis were analyzed (Suppl Fig 9e).

The C-12 gene expression signature is central to the authors' ability to compute abundance of C-12 in the public whole transcriptome datasets. The signature was defined from gene expression data from sorted, single memory T cells and then applied to whole blood transcriptomes or sorted T cell transcriptomes. But how do the authors know that this signature actually faithfully detects the abundance of C-12 in these datasets with high specificity? The fact that the C-11 gene expression signature behaves differently to the C-12 gene expression signature does not demonstrate that the C-12 signature specifically detects only C12 cells when gene expression in the much more complex whole blood is analyzed. This issue further contributes to the concerns raised in 1 above.

The authors mention that studies which focus on "antigen-specific cells miss the broader immune context that predisposes people to disease", framing this as a weakness. However, it is not clear from the discussion what the mechanism of control conferred by C-12 cells is, if it is not a Mtb-specific cell subset? All functional analyses are based on polyclonal stimulation and do not reveal what the specificity of this subset is. The finding that the signature-inferred C-12 subset is lower in abundance during TB disease and during progression is consistent with the well-described general lymphocytopenia observed during inflammation and there is little evidence to rule this out. As an additional QC exercise it would be useful to see a comparison of the distributions of post-QC cell yields stratified into cases and controls, to confirm that they are equivalent.

The authors computed a gamma statistic summarizing each potential confounder's influence on memory T cell composition. The authors do not state what the actual statistic is, but we assume that it is $-\sum_{k=1}^{K} \ln(P_k)$, since this has the claimed $Ga(n,1)$ distribution when all $P_k \sim U(0,1)$ and are independent, as they would be when each P_k is a p-value sampled under the null hypothesis. However, as stated such a gamma statistic assumes that p-values are independent of one another. For example, this independence criterion means that the p-value for the association of age with cluster C1 is independent of the p-value for the association of Age with cluster C2. This is very unlikely given that abundances of T cell subsets are often highly correlated. If that is true then the null distribution $Ga(n,1)$ is incorrect and the following could be done:

convince the reader that an empirically-derived null for the gamma statistic yields the same results (these are not available), OR

Use an FDR procedure and conclude that age (for example) is associated with T cell composition if at least one of the clusters are associated with age after controlling the FDR (as the BH procedure works in the presence of dependent p-values), OR

Use a multivariate model that allows for correlation in the cluster responses between different clusters.

Minor comments:
Page 9, 214. Clearly state from which model the p-values are taken (univariate, intermediate or full).

Reviewer #2:
Remarks to the Author:
In thus study, Nathan et al perform a large CITE-SEQ analysis of CD45RA-negative T cells from the PBMC of healthy individuals with latent Mtb infection or a distant history of active TB disease. The cohort is from a large household contact study performed in Peru that has been very well characterized. By choosing a time point where all individuals have all returned to health, the authors are able to look for differences in resistant and susceptible populations at immune homeostasis, and they are not confounded by active disease. Over 30 cell states are identified, and correlated with a number of covariates including age, sex, season, European ancestry, etc. One particular state (annotated here as C12), shows a significant reduction in individuals with a history of TB, as well a slight reduction with age, reduction in males, and increase in winter. The C12 state contains CD4 T cells with a very interesting mixed Th17-like phenotype. It is also shown that a smaller fraction of the cells with this phenotype isolated from cases produce cytokine compared to controls. The authors conclude that the decrease in this cell type or its function likely predisposes to the development of TB. This study both provides a valuable basic resource and attacks a major question in tuberculosis immunology. This is clearly an important data set is for the community interested in human T cells. Moreover, the identification of T cell correlates of protection is a major goal for the TB field, and the subset of T cells identified here will further stimulate the field's developing interest in the host-protective role of T cells with Th17-like characteristics. The lack of information regarding the antigen specificity of the C12 subset, however, seems a fundamental oversight that should be addressed.

1.) Are the C12 cells specific for Mtb? Bystander cells have rarely, if ever, been shown to play a major role in host resistance. The authors went so far as to FACS purify these cells from individuals with known Mtb exposure and stimulate them but didn't test if these cells respond to Mtb Ags. This should be done. Previous studies have shown that the majority of Mtb-specific T cells in healthy individuals express markers inconsistent with the C12 subset (e.g. CXCR3+CCR6+), and the authors point this out, indicating these cells may not be Mtb-specific. If these cells are bystanders that recognize unrelated antigens, it seems most likely, in my opinion, that this subset is just an irrelevant population whose frequency in the blood happens to correlate with TB history, along with age, sex and wintertime. The importance for tuberculosis research will be limited if the net result of this impressive analysis of this exciting cohort is an association with a cell that cannot directly contribute to control of Mtb infection. If these cells are not specific for Mtb, then the authors should provide a very convincing argument as to why they propose these cells actually contribute to outcome of Mtb infection.

2.) Are the C12 cells specific for peptide or non-peptide antigens? These cells resemble the CD45RA- IL-17+ IL-22+ CD4 T cells described by Scriba and Hanekom more than a decade ago. Interestingly, the cells reported in that manuscript produced IL-17 and 22 only to whole bacterial stimulation (not to peptides) in whole blood assays (not in PBMC), indicating they may be unconventional T cells.

3.) gdT cells were identified, but were the authors able to identify populations of unconventional abT cells? They do not show up in clusters and are not discussed.

Reviewer #3:

Remarks to the Author:
The manuscript by Nathan et al. describes the intricate profiling of memory T cells from carefully characterized individuals with Mtb infection. It is a well-written manuscript with many important controls included in the analysis (technical replicates for flow, the correlation between CITE-seq and flow cytometry, batch correction, etc.). It was a joy to read! I have a few comments for clarification of the last few figures and corresponding results. I agree that C-12 is an interesting population, but it seems like CD26, CD161, and CCR6 surface stain capture not only C-12.

The CD26+CD161+CCR6+ surface markers capture C-12, but these markers also likely capture some of the other clusters that express these markers, like C-19, C-20, C-16, etc.? How does that influence the results in figure 5e and after? Also, considering that C-12 had reduced transcripts for IFNG but they produce the cytokine just fine (Fig 6a), along with TNF.
If all of the clusters that express these markers are included in, for example, figure 7a, is the correlation better?

The Mtb infection status is unknown in the Boston donor samples. While they are most likely negative, it should be easy enough to run a T-spot.TB or measure reactivity against ESAT-6 and CFP10 to check?

One outstanding question is whether the identified cell population C-12 (+ other cells expressing CD26, CD161, and CCR6) produces the same cytokines in response to specific antigens? They have the capability, clearly, of producing IL-17, IL-22, etc. when stimulated with polyclonal stimuli, but whether they actually do in response to infection remains to be determined. Can the authors comment on that aspect?

Author Rebuttal to Initial comments

We were pleased to see that we had addressed the points raised by Reviewers #2 and #3. We wanted to take the opportunity to respond to Reviewer #1, who raised points about how we described the cohort, and our claims around the results of analyzing public data sets. We believe that these points are fully addressable, and to some extent derive from a misunderstanding about our analysis of public data.

The authors have done a lot of good work to comprehensively address many of the concerns raised by the reviewers. They should be commended for the manner in which they addressed the questions around whether the C-12 cell population includes Mtb-specific cells, which improves the paper substantially. However, there remain three issues that were not appropriately addressed. Addressing these in a satisfactory manner is essential, in our view, for the integrity and interpretation of the results.

1. As written, the paper still gives the impression that discovery of memory T cell states was

done in a primary progression cohort and not several years after TB diagnosis and cure, which is misleading and unnecessary. In fact, when reading the abstract there is no clear description that the analyses were done in people who had previously progressed to TB disease and were successfully treated before sampling. In the first paragraph of the results section, where the cohort is described, the description also does not make this clear and the first time this is acknowledged in the main text of the paper is in line 301. Perhaps the authors are so familiar with the cohort that this issue is not easily recognised, but to a reader who is unfamiliar with the cohort it just does not reflect a clear and accurate picture of the primary study design. The abstract, introduction and results section should clearly state that the progressors were profiled years after TB disease diagnosis and treatment, so that the reader is aware of this important feature of the study design. The language around this issue should be edited throughout the manuscript to make it less misleading.

We have tried to be clear about this issue in the text, but looking at the manuscript, it does appear in a cursory read, it might be possible to miss the fact that we are querying individuals in the discovery data set years after TB. We are happy to revise the manuscript to make this more explicit in the abstract, introduction, and results. We do think some of this language was already present in the manuscript. For example:

Abstract:

We computationally integrated high-dimensional single-cell RNA and surface protein marker data to define an atlas of 31 memory T cell states spanning 500,089 memory T cells from 259 individuals in a Peruvian tuberculosis (TB) progression cohort profiled at immune steady state > 4 years after infection.

Results:

Participants diagnosed with microbiologically confirmed TB were classified as cases; household contacts who were tuberculin skin test (TST)-positive and had not developed TB disease by time of re-recruitment (4.72–6.60 years [median: 5.7] after initial recruitment) were controls. During this time, cases were treated for active disease, which has an estimated cure rate of at least 95%, so they were expected to have returned to immune steady state without disease-driven perturbation.

We acknowledge that these sentences could be even more explicit about the absence of TB disease at time of assay. We will modify this language to be clear that case individuals are

being profiled after treatment (abstract and introduction), and that both cases and controls were profiled years after disease (results).

2. This paper states that it identified "a polyfunctional Th17-like effector state reduced in abundance and function in individuals who had progressed from Mycobacterium tuberculosis (M.tb) infection to active TB disease" and "we also demonstrated that its depletion may precede and persist beyond active disease". The evidence provided by the authors to discount the possibility that the associations they identify with their high-dimensional single-cell RNA and surface protein marker data analyses as being the product of persisting effects of either TB disease or TB treatment is, as it stands, not sufficiently strong. We appreciate the challenges of accessing a prospective study of infected or exposed individuals to identify enough progressors, but this does not justify their discounting the likely effects of previous disease or treatment. Given the limitations of the C-12 gene expression score in predicting abundance of the C-12 population in whole blood(see point 3 below) and the sample size issue they face in their analyses of donors followed through TB treatment (in Suppl Fig 10, which limits their ability to rule out a real difference due to TB treatment and/or disease), it is surprising that they claim so definitively that "C-12 reduction observed in progressors after 4-7 years and anti-mycobacterial treatment is not because of treatment and, furthermore, is likely distinct from depletion observed during disease" in the Results section, or "Reduction during ongoing active disease is of a significantly larger magnitude than that which is observed after 12 months of treatment, during which we infer that C-12 frequencies increase toward their original levels as patients recover." Absence of evidence due to sample size limitations is not evidence of absence of an effect. This is especially relevant since the posited pre-disease difference between progressors and non-progressors is very small.

In the primary analysis we examine a cohort 4-7 years after TB disease and treatment. To address the question of whether the differences that we observe are a result of disease or treatment, we analyzed data from Scriba, et al. (2017) which precede TB by at least one year. Using a statistical model, we predict C-12 frequencies based on their RNA-seq *of sorted T cells*, and see significant differences preceding TB, consistent with our primary analysis. On that basis we are able to claim that the differences we see after disease and treatment likely preceded TB disease and treatment. We believe this evidence is stronger than the reviewer claims it is, especially given that the cell types are well matched (see Point #3 response).

9

We had already acknowledged in the Discussion that we cannot exclude the possibility of disease-related sequelae:

> Although our study design does not allow us to conclusively disentangle such sequelae from baseline differences in T cell states.

> Further prospective studies profiling the immune system prior to *M.tb* infection are required in order to conclusively establish a causal link between the C-12 state at baseline and response to *M.tb* infection.

We will expand on this to discuss the other possibilities in greater depth in the text (discussion).

Regarding treatment, we would like to clarify that sample size limitations discussed in the revised manuscript (Supp Fig 10e) specifically affect our ability to detect a difference between C-12 in BCG+ Healthy donors and former progressors after 12 months of treatment in the Berry et al. (2010) dataset. However, this is not directly pertinent to the question of whether differences in C-12 abundance are due to disease or treatment. In fact, in spite of having a limited sample size, we were still able to measure significant increase in estimated C-12 proportion over the 12 month treatment course that the donors in Berry et al. (2010) underwent (Supp. Fig. 10d, two-sided t test p = 7.27 x 10^{-3}.) On the other hand, the difference in C-12 that we describe in our primary analysis was a decrease — the opposite direction of what is observed during treatment. While we did profile our cohort 4-7 years after disease, which is outside the scope of Berry et al., these data still show that during treatment C-12 levels are not reduced, so treatment is likely not the cause of the reduction we observed. (There is a lack of published data on the long-term effects of anti-mycobacterial treatment on immune cell states years after treatment is completed, so we will not claim this conclusively.)

Following the reviewer's suggestion, we will revise our statement about the potential roles of disease and treatment on C-12 abundance to be less definitively stated. We also recognize that the results from the Berry et al. treatment data are worded confusingly, so we will also revise that to be more clear about the finding.

3. The inferences about abundance of the C-12 population in the pre-disease state (abundance in

individuals a significant time before they are diagnosed with TB or treated) hinges on a gene expression score that predicts frequencies of cells with the C-12 phenotype. The authors use this to estimate C-12 abundance in the whole blood datasets from Berry et al., 2010 and Scriba et al. 2017. The corresponding inferences in these datasets depend entirely on whether the gene expression score predicts C-12 frequencies well in whole blood. In their revised manuscript they now make two attempts to train a model that demonstrably generates such a gene expression score. The first attempt is by testing their original model, trained only on T cell data, to PBMC data. The model predicted poorly. The second attempt is to retrain the model on both PBMC and T cell data and predict on PBMCs again, which improves performance. The claim then is that because they can fit a model that predicts well on PBMCs, it therefore must predict well on whole blood. However, the finding that a model trained using PBMC data predicts reasonably well on PBMC data does not demonstrate that it would necessarily predict well on whole blood. The key question is rather whether a model trained on data from sorted (or isolated in the case of PBMC) cells has sufficiently accurate prediction on data from unsorted (whole blood, predominated by granulocytes, which are of course absent from sorted T cells and PBMC) cells. Clearly, successfully predicting using data from cells with the same degree of sorting/isolation as the cells providing the data used to train the model does not demonstrate this. Moreover, and more importantly, when the model was trained only on sorted T cells, it predicted poorly on PBMCs, which undercuts the claim of generalisability to whole blood and strengthens the concern that a gene expression score trained on sorted/isolated cells is unlikely to predict C-12 well on whole blood, since accuracy deteriorated severely when generalising from T cells to PBMCs. Therefore, whilst their model may identify a gene signature that is different between progressors long before TB disease and controls, the fact that the only available statistical evidence undercuts its generalisability to whole blood means that it is difficult to interpret this transcriptomic difference as reflecting changes in the memory T cell subset of interest.

In our study we identified a reduction in a C-12 Th17-like effector cell state in individuals that had previously progressed to TB (assayed 4-7 years after they had TB). We analyzed public data collected prior to disease progression (Scriba et al.) to make the case that these differences preceded TB disease. In order to estimate C-12 in the public data, we trained and applied a model that can use whole-transcriptome gene expression to infer C-12 proportion. We demonstrated using cross-validation analysis of samples from our own cohort that this model is effective on sorted T cells and PBMCs. The reviewer questioned these important analyses, and suggested that our classifier would not work on whole blood transcriptomic data. **Very importantly, the Scriba et al. analysis (the basis of the pre-disease claim) was based on sorted T cell transcriptional data, and not whole-blood transcriptional data. We think the reviewer misunderstood this distinction.** We are happy to revise this manuscript to clarify this. From the reviewer comments, it appears that they acknowledge that our model will work on sorted T cells, and thus our results based on Scriba et al. are in fact valid.

The reviewer is correct that we do, however, apply the model to whole blood data from Berry, et al. (2010) to determine whether treatment affects C-12 proportions. To reiterate, this analysis was **not** used to make the claim that the C-12 reduction preceded disease, only to assess C-12 *during* disease and treatment. It is possible that our model may be less accurate in this context. It is also true, that we don't know how well our model would fare on whole blood. We believe that the model would do reasonably well because the most highly weighted genes are T-cell-specific and not expressed in neutrophils.

We note however, that we use the Berry et al. data in a limited way in our analysis. We use the Barry data to argue that C-12 frequencies are depleted during active disease (suggesting the disease relevance of the state) and recover with treatment. We acknowledge that we need to be more circumspect in this claim. We will modify the language to say that the C-12 score has not been validated in whole blood, so there is chance of higher error in the whole blood estimates due to the presence of granulocytes, erythrocytes, and platelets. As discussed in our response to Point 2, we will make the claims based on the Berry et al. data less definitive. We will also include the highest and lowest weighted genes in the C-12 score a supplementary table to assuage concerns that the score may be capturing non-T-cell signatures.

Decision Letter, first revision:

Subject: Decision on Nature Immunology submission NI-RS30805A

Message: Dear Dr Raychaudhuri,

Thank you for your response to the reviewers' comments on your manuscript "Multimodal profiling of 500,000 memory T cells from a tuberculosis cohort identifies cell state associations with demographics, environment, and disease". We are happy to inform you that if you revise your manuscript appropriately in response to the referees' comments and our editorial requirements your manuscript should be publishable in Nature Immunology.

Please revise your manuscript according with the reviewers' comments and as outlined in your letter. At resubmission, please include a point-by-point response to the referees' comments, noting the pages and lines where the changes can be found in the revision. Please highlight the changes in the revised manuscript as well. Once we receive that I'll start the edits at this end.

We are trying to improve the quality and transparency of methods and statistics reporting in our papers (please see our editorial in the May 2013 issue). Please update the Life Sciences Reporting Summary, and supplements if applicable, with any information relevant to any new experiments and upload it (as a Related Manuscript File) along with the files for your revision. If nothing in the checklist has changed, please upload the current version again.

TRANSPARENT PEER REVIEW

Nature Immunology offers a transparent peer review option for new original research manuscripts submitted from 1st December 2019. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters if the authors agree. Such peer review material is made available as a supplementary peer review file. Please state in the cover letter 'I wish to participate in transparent peer review' if you want to opt in, or 'I do not wish to participate in transparent peer review' if you don't. Failure to state your preference will result in delays in accepting your manuscript for publication.

Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our FAQ page.

ORCID

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. For more information please visit www.springernature.com/orcid.

Before resubmitting the final version of the manuscript, if you are listed as a corresponding author on the manuscript, please follow the steps below to link your account on our MTS with your ORCID. If you don't have an ORCID yet, you will be able to create one in minutes. If you are not listed as a corresponding author, please ensure that the corresponding author(s) comply.

1. From the home page of the MTS click on 'Modify my Springer Nature account' under 'General tasks'.
2. In the 'Personal profile' tab, click on 'ORCID Create/link an Open Researcher Contributor ID(ORCID)'. This will re-direct you to the ORCID website.
3a. If you already have an ORCID account, enter your ORCID email and password and click on 'Authorize' to link your ORCID with your account on the MTS.
3b. If you don't yet have an ORCID, you can easily create one by providing the required information and then click on 'Authorize'. This will link your newly created ORCID with your account on the MTS.

IMPORTANT: All authors identified as 'corresponding authors' on the manuscript must follow these instructions. Non-corresponding authors do not have to link their ORCIDs, but please note that it will not be possible to add/modify ORCIDs at proof. Thus, if they wish to have their ORCID added to the paper, they must also follow the above procedure prior to acceptance.

To support ORCID's aims, we only allow a single ORCID identifier to be attached to one account. If you have any issues attaching an ORCID identifier to your Manuscript Tracking System account, please contact the Platform Support Helpdesk.

We hope that you will support this initiative and supply the required information. Should you have any query or comments, please do not hesitate to contact me.

Nature Immunology has now transitioned to a unified Rights Collection system which will allow our Author Services team to quickly and easily collect the rights and permissions required to publish your work. Once your paper is accepted, you will receive an email in approximately 10 business days providing you with a link to complete the grant of rights. If you choose to publish Open Access, our Author Services team will also be in touch at that time regarding any additional information that may be required to arrange payment for your article.

Please note that you will not receive your proofs until the publishing agreement has been received through our system.

For information regarding our different publishing models please see our Transformative Journals page. If you have any questions about costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com..

In recognition of the time and expertise our reviewers provide to Nature Immunology's editorial process, we would like to formally acknowledge their contribution to the external peer review of your manuscript entitled "Multimodal profiling of 500,000 memory T cells from a tuberculosis cohort identifies cell state associations with demographics, environment, and disease". For those reviewers who give their assent, we will be publishing their names alongside the published article.

When you are ready to submit your revised manuscript, please use the URL below to submit the revised version: [REDACTED]

We hope to receive your revised manuscript in 10 days, by 22nd Mar 2021. Please let us know if circumstances will delay submission beyond this time. If you have any questions please do not hesitate to contact me.

Sincerely,

Zoltan Fehervari, Ph.D.
Senior Editor
Nature Immunology

The Macmillan Building
4 Crinan Street
Tel: 212-726-9207
Fax: 212-696-9752
z.fehervari@nature.com

Reviewer #1 (Remarks to the Author):

The authors have done a lot of good work to comprehensively address many of the concerns raised by the reviewers. They should be commended for the manner in which they addressed the questions around whether the C-12 cell population includes Mtb-specific cells, which improves the paper substantially. However, there remain three issues that were not appropriately addressed. Addressing these in a satisfactory manner is essential, in our view, for the integrity and interpretation of the results.

1. As written, the paper still gives the impression that discovery of memory T cell states was done in a primary progression cohort and not several years after TB diagnosis and cure, which is misleading and unnecessary. In fact, when reading the abstract there is no clear description that the analyses were done in people who had previously progressed to TB disease and were successfully treated before sampling. In the first paragraph of the results section, where the cohort is described, the description also does not make this clear and the first time this is acknowledged in the main text of the paper is in line 301. Perhaps the authors are so familiar with the cohort that this issue is not easily recognised, but to a reader who is unfamiliar with the cohort it just does not reflect a clear and accurate picture of the primary study design. The abstract, introduction and results section should clearly state that the progressors were profiled years after TB disease diagnosis and treatment, so that the reader is aware of this important feature of the study design. The language around this issue should be edited throughout the manuscript to make it less misleading.

2. This paper states that it identified "a polyfunctional Th17-like effector state reduced in abundance and function in individuals who had progressed from Mycobacterium tuberculosis (M.tb) infection to active TB disease" and "we also demonstrated that its depletion may precede and persist beyond active disease". The evidence provided by the authors to discount the possibility that the associations they identify with their high-dimensional single-cell RNA and surface protein marker data analyses as being the product of persisting effects of either TB disease or TB treatment is, as it stands, not sufficiently strong. We appreciate the challenges of accessing a prospective study of infected or exposed individuals to identify enough progressors, but this does not justify their discounting the likely effects of previous disease or treatment. Given the limitations of the C-12 gene expression score in predicting abundance of the C-12 population in whole blood (see point 3 below) and the sample size issue they face in their analyses of donors followed through TB treatment (in Suppl Fig 10, which limits their ability to rule out a real difference due to TB treatment and/or disease), it is surprising that they claim so definitively that "C-12 reduction observed in progressors after 4-7 years and anti-mycobacterial treatment is not because of treatment and, furthermore, is likely distinct from depletion observed during disease" in the Results section, or "Reduction during ongoing active disease is of a significantly larger magnitude than that which is observed after 12 months of treatment, during which we infer that C-12 frequencies increase toward their original levels as patients recover." Absence of evidence due to sample size limitations is not evidence of absence of an effect. This is especially relevant since the posited pre-disease difference between progressors and non-progressors is very small.

3. The inferences about abundance of the C-12 population in the pre-disease state (abundance in individuals a significant time before they are diagnosed with TB or treated) hinges on a gene expression score that predicts frequencies of cells with the C-12 phenotype. The authors use this to estimate C-12 abundance in the whole blood datasets from Berry et al., 2010 and Scriba et al. 2017. The corresponding inferences in these

datasets depend entirely on whether the gene expression score predicts C-12 frequencies well in whole blood. In their revised manuscript they now make two attempts to train a model that demonstrably generates such a gene expression score. The first attempt is by testing their original model, trained only on T cell data, to PBMC data. The model predicted poorly. The second attempt is to retrain the model on both PBMC and T cell data and predict on PBMCs again, which improves performance. The claim then is that because they can fit a model that predicts well on PBMCs, it therefore must predict well on whole blood. However, the finding that a model trained using PBMC data predicts reasonably well on PBMC data does not demonstrate that it would necessarily predict well on whole blood. The key question is rather whether a model trained on data from sorted (or isolated in the case of PBMC) cells has sufficiently accurate prediction on data from unsorted (whole blood, predominated by granulocytes, which are of course absent from sorted T cells and PBMC) cells. Clearly, successfully predicting using data from cells with the same degree of sorting/isolation as the cells providing the data used to train the model does not demonstrate this. Moreover, and more importantly, when the model was trained only on sorted T cells, it predicted poorly on PBMCs, which undercuts the claim of generalisability to whole blood and strengthens the concern that a gene expression score trained on sorted/isolated cells is unlikely to predict C-12 well on whole blood, since accuracy deteriorated severely when generalising from T cells to PBMCs. Therefore, whilst their model may identify a gene signature that is different between progressors long before TB disease and controls, the fact that the only available statistical evidence undercuts its generalisability to whole blood means that it is difficult to interpret this transcriptomic difference as reflecting changes in the memory T cell subset of interest.

Reviewer #2 (Remarks to the Author):

The authors have adequately addressed all of my concerns and questions. Most importantly, they have provided data indicating that the C12 subset does in fact contain Mtb specific T cells. This is an important contribution.

Author Rebuttal, first revision:

Response to Reviewers

We thank the reviewers and the editor for carefully reviewing our revised manuscript. We are pleased to see that we had addressed the points raised by Reviewers #2 and #3, especially around *M.tb* antigen specificity (as noted specifically by Reviewer #2). We want to respond to Reviewer #1, who raised points about how we described the cohort, and our claims around the results of analyzing public data sets. Here, we offer a point-by-point response to Reviewer #1's further comments:

Specific Comments:

1. As written, the paper still gives the impression that
discovery of memory T cell states was done in a primary
progression cohort and not several years after TB diagnosis and
cure, which is misleading and unnecessary. In fact, when reading
the abstract there is no clear description that the analyses
were done in people who had previously progressed to TB disease
and were successfully treated before sampling. In the first
paragraph of the results section, where the cohort is described,
the description also does not make this clear and the first time
this is acknowledged in the main text of the paper is in line
301. Perhaps the authors are so familiar with the cohort that
this issue is not easily recognised, but to a reader who is
unfamiliar with the cohort it just does not reflect a clear and
accurate picture of the primary study design. The abstract,
introduction and results section should clearly state that the
progressors were profiled years after TB disease diagnosis and
treatment, so that the reader is aware of this important feature
of the study design. The language around this issue should be
edited throughout the manuscript to make it less misleading.

We thank the review for sharing this suggestion. Some of this language was already present in the manuscript, but upon further review, we recognize that the description of the cohort can be clearer. We have modified the text in the sections recommended by the reviewer to more explicitly state that the cohort was profiled after cases had disease and were treated:

Abstract (lines 48-50):

> … from 259 individuals in a Peruvian tuberculosis (TB) progression cohort profiled at immune steady state >4 years after infection and disease resolution.

(lines 53-55):
> … reduced in abundance and function in individuals who had previously progressed from Mycobacterium tuberculosis (*M.tb*) infection to active TB disease.

Introduction (lines 99-100):

… profile memory T cells at single-cell resolution from a TB progression cohort at post-disease immune steady state (i.e., after treatment and TB disease resolution) …

Results (lines 115-122):

We re-recruited donors 4.72–6.60 years (median: 5.7) after initial recruitment. Participants who had originally been diagnosed with microbiologically confirmed TB and subsequently treated were classified as cases; household contacts who had originally been tuberculin skin test (TST)-positive and had not developed TB disease by time of re-recruitment were controls. By the time of re-recruitment and sample collection, cases had been treated for active disease, which has an estimated cure rate of at least 95%, so they were expected to have returned to immune steady state without disease-driven perturbation[23].

2. This paper states that it identified "a polyfunctional Th17-like effector state reduced in abundance and function in individuals who had progressed from Mycobacterium tuberculosis (M.tb) infection to active TB disease" and "we also demonstrated that its depletion may precede and persist beyond active disease". The evidence provided by the authors to discount the possibility that the associations they identify with their high-dimensional single-cell RNA and surface protein marker data analyses as being the product of persisting effects of either TB disease or TB treatment is, as it stands, not sufficiently strong. We appreciate the challenges of accessing a prospective study of infected or exposed individuals to identify enough progressors, but this does not justify their discounting the likely effects of previous disease or treatment. Given the limitations of the C-12 gene expression score in predicting abundance of the C-12 population in whole blood (see point 3 below) and the sample size issue they face in their analyses of donors followed through TB treatment (in Suppl Fig 10, which limits their ability to rule out a real difference due to TB treatment and/or disease), it is surprising that they claim so definitively that "C-12 reduction observed in progressors after 4-7 years and anti-mycobacterial treatment is not because of treatment and, furthermore, is likely distinct from depletion observed during disease" in the Results section, or "Reduction during ongoing active disease is of a significantly larger magnitude than that which is observed after 12 months of treatment, during which we infer that C-12 frequencies increase toward their original levels as patients recover." Absence of evidence due to sample size limitations is not evidence of absence of an effect. This is especially relevant since the

```
posited pre-disease difference between progressors and non-
progressors is very small.
```

Some of the uncertainty expressed by the reviewer regarding the claim that C-12 reduction "may precede and persist beyond active disease" may be mitigated by our response below to Comment #3, where we note that the pre-disease analysis used public gene expression from sorted T cells — the same cell type used in our primary analysis. We trained a linear predictor to predict C-12 frequencies based on gene expression from T cells and PBMCs, and observed consistent reduction in progressors in both pre- and post-disease datasets. On that basis, we claim that the differences we see after TB disease and treatment likely preceded disease. We still cannot state this conclusively, as acknowledged in the Discussion, and we have further edited the text to reflect the limitations of this analysis:

Results (lines 356-359):

> This demonstrates that C-12 reduction observed in progressors after 4-7 years and anti-mycobacterial treatment is likely not due to treatment. Furthermore, while we observe depletion during disease, it also does not conclusively explain the post-disease reduction, necessitating further analysis prior to disease.

Discussion (lines 540-543):

> First, we demonstrate in previously published cohorts that despite potential modulation during ongoing TB disease or treatment, C-12's gene expression signature is elevated in latently infected controls compared to active TB cases even prior to progression.

(See response to Comment #3 below for further discussion of the C-12 expression score.)

Furthermore, we would like to clarify that sample size limitations in the Berry et al. (2010) dataset raised by the reviewer (Supplementary Fig. 10e) do not affect our ability to detect a difference during disease (Supplementary Fig. 10b) or treatment (Supplementary Fig. 10d).
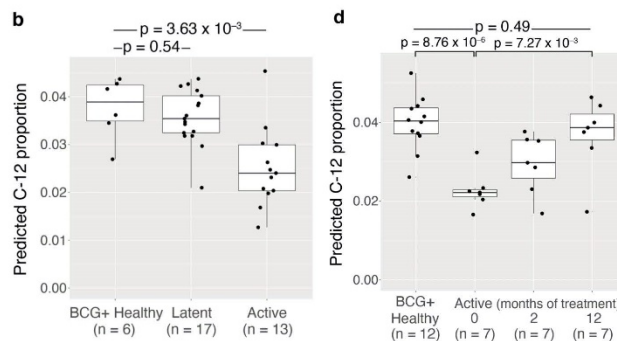
In fact, in spite of the sample size, we were still able to measure significant increase in estimated C-12 proportion over the 12-month treatment that the donors in Berry et al. underwent (two-sided t test p = 7.27×10^{-3}.) On the other hand, the difference in C-12 that we describe in our primary analysis was a decrease — the opposite direction of what is observed during treatment. While we did profile our cohort 4-7 years after disease, which is outside the scope of Berry et al., these data still show that during treatment C-12 levels are not reduced, so treatment is likely not the cause of the reduction we observed. (There is a lack of published data on the long-term effects of anti-mycobacterial treatment on immune cell states years after treatment is completed, so we will not claim this conclusively.)

The power analysis that we had added to this figure in our first revision was in response to original reviewer comments about the lack of significant difference between estimated C-12 in BCG+ Healthy donors vs. former progressors after 12 months of treatment (given that we observe a difference between formerly active and latent donors in our primary analysis of the post-disease CITE-seq cohort). However, the two comparison groups are not equivalent and we are underpowered. Furthermore, this is not directly pertinent to the question of whether differences in C-12 abundance are due to treatment, and the pre-disease analysis (using data from Scriba, et al. (2017)) more directly addresses whether the differences precede disease. We recognize that this distinction was not clear in our manuscript, so we have revised the following text:

Results (lines 342-359):

> In a small subset of donors (n = 7) followed through treatment, C-12 frequencies are initially significantly lower than in BCG-vaccinated healthy donors (before treatment: mean = 0.023, SE = 6.9×10^{-4}; BCG+ healthy: mean = 0.040, SE = 5.7×10^{-4}; two-sided t test [BCG+ healthy vs. 0 months] p = 8.76×10^{-6}, **Supplementary Fig. 10d**). However, during 12 months of anti-mycobacterial treatment estimated C-12 frequencies increase (12 months: mean = 0.037, SE = 1.4×10^{-3}; two-sided t test [0 months vs. 12 months] p = 7.27×10^{-3}), which is the opposite effect than what would be expected if the reduced C-12 frequency in progressors at post-disease steady state were due to treatment. In this smaller dataset, at the end of treatment we no longer not observe a difference of similar magnitude or significance to that observed prior to treatment between cases and healthy controls (two-sided t test [BCG+ healthy vs. 12 months] p = 0.49). Although these two groups are not exact proxies for the cases and controls in the post-disease cohort, at this sample size (n = 7 cases vs. 12 controls) we are also underpowered to replicate the same 20% reduction in C-12 (power = 0.15, **Supplementary Fig. 10e**). This demonstrates that C-12 reduction observed in progressors after 4-7 years and anti-mycobacterial treatment is likely not due to treatment. Furthermore, while we observe depletion during disease, it also does not conclusively explain the post-disease reduction, necessitating further analysis prior to disease.

3. The inferences about abundance of the C-12 population in the pre-disease state (abundance in individuals a significant time before they are diagnosed with TB or treated) hinges on a gene expression score that predicts frequencies of cells with the C-12 phenotype. The authors use this to estimate C-12 abundance in the whole blood datasets from Berry et al., 2010 and Scriba et al. 2017. The corresponding inferences in these datasets depend entirely on whether the gene expression score predicts C-12 frequencies well in whole blood. In their revised manuscript they now make two attempts to train a model that demonstrably generates such a gene expression score. The first attempt is by testing their original model, trained only on T cell data, to PBMC data. The model predicted poorly. The second attempt is to retrain the model on both PBMC and T cell data and predict on PBMCs again, which improves performance. The claim then is that because they can fit a model that predicts well on PBMCs, it therefore must predict well on whole blood. However, the finding that a model trained using PBMC data predicts reasonably well on PBMC data does not demonstrate that it would necessarily predict well on whole blood. The key question is rather whether a model trained on data from sorted (or isolated in the case of PBMC) cells has sufficiently accurate prediction on data from unsorted (whole blood, predominated by granulocytes, which are of course absent from sorted T cells and PBMC) cells. Clearly, successfully predicting using data from cells with the same degree of sorting/isolation as the cells providing the data used to train the model does not demonstrate this. Moreover, and more importantly, when the model was trained only on sorted T cells, it predicted poorly on PBMCs, which undercuts the claim of generalisability to whole blood and strengthens the concern that a gene expression score trained on sorted/isolated cells is unlikely to predict C-12 well on whole blood, since accuracy deteriorated severely when generalising from T cells to PBMCs. Therefore, whilst their model may identify a gene signature that is different between progressors long before TB disease and controls, the fact that the only available statistical evidence undercuts its generalisability to whole blood means that it is difficult to interpret this transcriptomic difference as reflecting changes in the memory T cell subset of interest.

Here, the reviewer raises concerns with our pre-disease analysis, where we estimated C-12 frequency in public data collected prior to disease progression (Scriba et al. 2017) to make the case that C-12 reduction in progressors preceded TB disease. Using our own sorted T cell and PBMC data, we trained and cross-validated a model that uses whole-transcriptome gene expression to infer C-12 proportion and then applied this model to the public data. The reviewer's concern is about the applicability of this model to whole blood RNA-seq data, but very importantly, the Scriba et al. data (the basis of the pre-disease claim) was sorted T cell RNA-seq data, not whole blood. We think the reviewer misunderstood this distinction. In the reviewer comments, they acknowledge that our model works on sorted T cells, and thus they would agree that our results based on Scriba et al. are valid. We have revised the manuscript to clarify the tissue source for the training data and public data:

Results (lines 363-365):

> Using the C-12 score trained on T cells and PBMCs, we estimated the abundance of C-12 in longitudinal bulk RNA-seq of sorted T cells collected from a South African cohort…

The reviewer is correct that we do, however, apply the model to whole blood data from Berry, et al. (2010) to determine whether latent infection or treatment affect C-12 proportions. To reiterate, this analysis was **not** used to make the claim that the C-12 reduction preceded disease. We want to emphasize that we use the Berry et al. data in a *limited* way in our manuscript. We use it to provide potential evidence against alternative hypotheses explaining the C-12 reduction that we observe at post-disease steady state, i.e., to argue that C-12 frequencies are not increased due to latent infection or decreased due to treatment. But our claims about pre-disease reduction of C-12 in progressors are independent of these analyses, and any uncertainty in the Berry et al. analysis does not invalidate our observations of C-12 reduction 2+ years before and 4+ years after disease.

We did not validate the C-12 score on whole blood, but we are encouraged that the model would be reasonably accurate, given published studies demonstrating high correlation between gene expression in whole blood and PBMCs (He, et al. 2019, *Allergy Asthma Clin Immunol*) and the transferability of published RNA-based TB risk signatures between whole blood and PBMCs (Zak, et al. 2016, *Lancet*). Nevertheless, there are expression differences due to the presence of granulocytes, erythrocytes, and platelets in whole blood, so there may be some loss of accuracy in our C-12 score when applied to whole blood. We have edited the text to reflect this:

Results (lines 330-333):

> This suggests that C-12 abundance is not increased by latency, although we recognize that there may be some error in C-12 estimation due to differences in tissue type, processing, and RNA assay between the training data and public data.

Discussion (lines 531-539)

To address this, we first trained a linear model on T cell and PBMC data and estimated C-12 frequencies in public data collected during disease, although the interpretation of these results is limited by sample size and gene expression differences between whole blood and PBMCs. Our subsequent analysis of public data collected pre-disease has both a large sample size and a nearly identical cell type (sorted T cells) as our training data, motivating confidence in pre-disease inferences. Although our study design still does not allow us to conclusively disentangle disease sequelae from baseline differences in T cell states, there is evidence suggesting that reduction in C-12 may precede disease.

Decision Letter, second revision:

Subject: Your manuscript, NI-RS30805B
Message: Our ref: NI-RS30805B

1st Apr 2021

Dear Dr. Raychaudhuri,

Thank you for your patience as we've prepared the guidelines for final submission of your Nature Immunology manuscript, "Multimodal profiling of 500,000 memory T cells from a tuberculosis cohort identifies cell state associations with demographics, environment, and disease" (NI-RS30805B). Please carefully follow the step-by-step instructions provided in the personalised checklist attached, to ensure that your revised manuscript can be swiftly handed over to our production team.

We would like to start working on your revised paper, with all of the requested files and forms, as soon as possible (preferably within two weeks). Please get in contact with us if you anticipate delays.

When you upload your final materials, please include a point-by-point response to any remaining reviewer comments and please make sure to upload your checklist.

If you have not done so already, please alert us to any related manuscripts from your group that are under consideration or in press at other journals, or are being written up for submission to other journals (see: https://www.nature.com/nature-research/editorial-policies/plagiarism#policy-on-duplicate-publication for details).

In recognition of the time and expertise our reviewers provide to Nature Immunology's editorial process, we would like to formally acknowledge their contribution to the external peer review of your manuscript entitled "Multimodal profiling of 500,000 memory T cells from a tuberculosis cohort identifies cell state associations with demographics, environment, and disease". For those reviewers who give their assent, we will be publishing their names alongside the published article.

Nature Immunology offers a Transparent Peer Review option for new original research manuscripts submitted after December 1st, 2019. As part of this initiative, we encourage our authors to support increased transparency into the peer review process by agreeing to have the reviewer comments, author rebuttal letters, and editorial decision letters published as a Supplementary item. When you submit your final files please clearly state in your cover letter whether or not you would like to participate in this initiative. Please note that failure to state your preference will result in delays in accepting your manuscript for publication.

Cover suggestions

As you prepare your final files we encourage you to consider whether you have any images or illustrations that may be appropriate for use on the cover of Nature Immunology.

Covers should be both aesthetically appealing and scientifically relevant, and should be supplied at the best quality available. Due to the prominence of these images, we do not

generally select images featuring faces, children, text, graphs, schematic drawings, or collages on our covers.

We accept TIFF, JPEG, PNG or PSD file formats (a layered PSD file would be ideal), and the image should be at least 300ppi resolution (preferably 600-1200 ppi), in CMYK colour mode.

If your image is selected, we may also use it on the journal website as a banner image, and may need to make artistic alterations to fit our journal style.

Please submit your suggestions, clearly labeled, along with your final files. We'll be in touch if more information is needed.

Nature Immunology has now transitioned to a unified Rights Collection system which will allow our Author Services team to quickly and easily collect the rights and permissions required to publish your work. Approximately 10 days after your paper is formally accepted, you will receive an email in providing you with a link to complete the grant of rights. If your paper is eligible for Open Access, our Author Services team will also be in touch regarding any additional information that may be required to arrange payment for your article.

You will not receive your proofs until the publishing agreement has been received through our system.

Please note that <i>Nature Immunology</i> is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. Find out more about Transformative Journals.

If you have any questions about costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com.

<B>Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. For submissions from January 2021, if your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route our standard licensing terms will need to be accepted, including our self-archiving policies. Those standard licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

Please use the following link for uploading these materials: [REDACTED]

If you have any further questions, please feel free to contact me.

Best regards,

Elle Morris
Editorial Assistant
Nature Immunology
Phone: 212 726 9207
Fax: 212 696 9752
E-mail: immunology@us.nature.com

On behalf of

Zoltan Fehervari, Ph.D.
Senior Editor
Nature Immunology

Final Decision Letter:

Subject: Decision on Nature Immunology submission NI-RS30805C

Message: In reply please quote: NI-RS30805C

Dear Dr. Raychaudhuri,

I am delighted to accept your manuscript entitled "Multimodally profiling memory T cells from a tuberculosis cohort identifies cell state associations with demographics, environment, and disease" for publication in an upcoming issue of Nature Immunology.

The manuscript will now be copy-edited and prepared for the printer. Please check your calendar: if you will be unavailable to check the galley for some portion of the next month, we need the contact information of whom will be making corrections in your stead. When you receive your galleys, please examine them carefully to ensure that we have not inadvertently altered the sense of your text.

Acceptance is conditional on the data in the manuscript not being published elsewhere, or announced in the print or electronic media, until the embargo/publication date. These restrictions are not intended to deter you from presenting your data at academic meetings and conferences, but any enquiries from the media about papers not yet scheduled for publication should be referred to us.

Please note that <i>Nature Immunology</i> is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. Find out more about Transformative Journals.

<B>Authors may need to take specific actions to achieve <a

href="https://www.springernature.com/gp/open-research/funding/policy-compliance-faqs">compliance with funder and institutional open access mandates. For submissions from January 2021, if your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route our standard licensing terms will need to be accepted, including our self-archiving policies. Those standard licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

In approximately 10 business days you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

You will not receive your proofs until the publishing agreement has been received through our system.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

Once your manuscript is typeset and you have completed the appropriate grant of rights, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately. Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

Your paper will be published online soon after we receive your corrections and will appear in print in the next available issue. Content is published online weekly on Mondays and Thursdays, and the embargo is set at 16:00 London time (GMT)/11:00 am US Eastern time (EST) on the day of publication. Now is the time to inform your Public Relations or Press Office about your paper, as they might be interested in promoting its publication. This will allow them time to prepare an accurate and satisfactory press release. Include your manuscript tracking number (NI-RS30805C) and the name of the journal, which they will need when they contact our office.

About one week before your paper is published online, we shall be distributing a press release to news organizations worldwide, which may very well include details of your work. We are happy for your institution or funding agency to prepare its own press release, but it must mention the embargo date and Nature Immunology. Our Press Office will contact you closer to the time of publication, but if you or your Press Office have any enquiries in the meantime, please contact press@nature.com.

Also, if you have any spectacular or outstanding figures or graphics associated with your manuscript - though not necessarily included with your submission - we'd be delighted to consider them as candidates for our cover. Simply send an electronic version (accompanied by a hard copy) to us with a possible cover caption enclosed.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

If you have not already done so, we strongly recommend that you upload the step-by-step protocols used in this manuscript to the Protocol Exchange. Protocol Exchange is an open online resource that allows researchers to share their detailed experimental know-how. All uploaded protocols are made freely available, assigned DOIs for ease of citation and fully searchable through nature.com. Protocols can be linked to any publications in which they are used and will be linked to from your article. You can also establish a dedicated page to collect all your lab Protocols. By uploading your Protocols to Protocol Exchange, you are enabling researchers to more readily reproduce or adapt the methodology you use, as well as increasing the visibility of your protocols and papers. Upload your Protocols at www.nature.com/protocolexchange/. Further information can be found at www.nature.com/protocolexchange/about .

Please note that we encourage the authors to self-archive their manuscript (the accepted version before copy editing) in their institutional repository, and in their funders' archives, six months after publication. Nature Research recognizes the efforts of funding bodies to increase access of the research they fund, and strongly encourages authors to participate in such efforts. For information about our editorial policy, including license agreement and author copyright, please visit www.nature.com/ni/about/ed_policies/index.html

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. Please let your coauthors and your institutions' public affairs office know that they are also welcome to order reprints by this method.

Sincerely,

Zoltan Fehervari, Ph.D.
Senior Editor
Nature Immunology

The Macmillan Building
4 Crinan Street
Tel: 212-726-9207
Fax: 212-696-9752
z.fehervari@nature.com