

Reproducible single cell annotation of programs underlying T cell subsets, activation states, and functions

Corresponding Author: Soumya Raychaudhuri

This file contains all editorial decision letters in order by version, followed by all author rebuttals in order by version.

A version of this paper was originally rejected for publication by Nature Methods, however that decision was reconsidered after appeal by the authors.

Version 0:

Decision Letter:

28th Aug 2024

Dear Soumya,

One again, we apologize for the delay in reaching a decision on your manuscript. As I explained we had to assign a fourth review in order to be able to fairly assess your manuscript.

Your Resource entitled "Reproducible single cell annotation of programs underlying T-cell subsets, activation states, and functions" has now been seen by the fourth ref, whose comments are attached. While they find your work of potential interest, they have raised serious concerns which in our view are sufficiently important that they preclude publication of the work in Nature Methods, at least in its present form.

As you will see, the reviewers raise concerns about validation and the utility of this resource. We also thought, the manuscript would benefit from emphasizing the value this Resource can provide. Should further experimental data allow you to fully address these criticisms we would be willing to look at a revised manuscript (unless, of course, something similar has by then been accepted at Nature Methods or appeared elsewhere). This includes submission or publication of a portion of this work somewhere else. We hope you understand that until we have read the revised paper in its entirety we cannot promise that it will be sent back for peer-review.

If you are interested in revising this manuscript for submission to Nature Methods in the future, please contact me to discuss your appeal before making any revisions. Otherwise, we hope that you find the reviewers' comments helpful when preparing your paper for submission elsewhere.

Sincerely,
Madhura

Madhura Mukhopadhyay, PhD
Senior Editor
Nature Methods

Reviewers' Comments:

Reviewer #1:

Remarks to the Author:

1. gene expression program GEP needs to be defined and explained in the introduction
2. The abbreviation of the tool is confusing -- with "***". Not sure how it will be searchable on the web etc
3. Not sure what wildcard character stands for in the name of the tool

4. In the results section, the author says that the batch effect could not hinder the identification of irreproducible GEP. This needs to be explained in more detail
5. Line 135. Authors claimed that reference datasets shared only 90% of genes. This needs to be explained and explained why this is expected
6. Line 153. It would be helpful to provide the table with a description of all datasets being used with details about those databases. Also explain what criteria was used to select those public datasets
7. Line 151. Which other tissues in addition to blood were used
8. Line 160. p-value was not provided
9. Line 181. Mentions IGKV gene. The relationship to the T cell needs to be explained
10. Line 222. Mentioned ground truth. It needs to be explained why manual gating of surface proteins will provide ground truth. Also, it would be helpful to discuss if such a dataset indeed can be considered as ground truth or gold standard See this paper discussing the type of gold standard datasets <https://www.nature.com/articles/s41467-019-09406-4>

Reviewer #2:

Remarks to the Author:

In this paper by Kotlia et al., the authors effectively apply their previously published cNMF algorithm to create an annotator for T-cell gene expression programs, with successful results. Annotating T cell subsets and unifying prior descriptions of T cell functional states is a very common and important task for many studies involving single-cell sequencing analysis of T cells. Their aggregation and consolidation of 46 GEPs describing possible T cell states into a single easy-to-use package is potentially quite useful. Although this package and method does seem to be generally useful, it is not clear to me that this method is highly differentiated from standard approaches. For the examples provided, there also does not seem to be major new insights gained based on the use of this approach.

Comments:

It's a bit unclear why the authors chose to only select two GEPs to turn into a binary classification variable; for an annotator, it would be expected and helpful to have binary classifications for all programs, otherwise it's unclear how this would be an improvement over simply using a calculated gene score with a set of marker genes for a given phenotype/differentiation state. The multinomial labeling tool does this somewhat, but despite the high number of GEPs here, the level of granularity of T cell phenotypes identified by the multinomial labeling tool seems quite limited. As those phenotypes are all well-characterized by a few marker genes, it's unclear how this tool is a significant improvement over existing methods (e.g. Azimuth).

Lines 129-147; Fig S1:

Removing only 4 GEPs out of 16 for performance testing purposes seems low; what if a dataset had only CD4 or CD8 T cells for example? How would *CAT perform on such a dataset? Figure S1 is hard to interpret quantitatively; extracted accuracy values would be helpful.

Fig 2A:

The UMAP in Fig 2A does not appear to effectively separate or cluster the identified cell types; since these labels are being used as ground truth this may be concerning and should be addressed. A corresponding UMAP showing the multinomial label predictions would be helpful here.

Fig 2C, Fig S3C:

The figure description in S3C does not match the legend in the figure; is multilabel prediction the light blue color or the dark blue color? If the legend is correct, then for 2C and S3C, why does the multilabel prediction perform better than the single cGEP version (the explanation for how this was calculated is also somewhat unclear)?

Minor comments:

- The text in some figures is quite small and almost unreadable (e.g. Fig 1A, S2B)
- Color schemes in figures are completely mixed throughout; it would be clearer if the same color schemes were used for representing the same data types
- Typo in line 984

Reviewer #4:

Remarks to the Author:

A. Summary of the key results

This paper proposes a new way to analyse single T cell expression as being the sum of a variety of independent gene expression profiles, i.e. instead of an individual T cell being described as exhausted or expanding it can be thought of as occupying a state somewhere between both (and indeed up to 46 different GEPs).

To do this the authors applied the recently published cNMF algorithm, but first had to design a new batch normalisation strategy to allow the data from 7 previously published studies to be amalgamated in such a way as to be compatible with cNMF.

They find several GEPs to be represented in all the T cell datasets considered, though these include trivial GEPs such as cell cycle and cytoskeletal associated GEPs that should be in all cells which flatters the comparison. GEPs were annotated by manual inspection of the top weighted genes, and poorly supported GEPs removed.

TCAT GEPs were benchmarked against the previous tool NMFproj and clustering approaches using an independent CITEseq experiment. TCAT was then applied to the COMBAT covid database and shown to be able to deconvolute subsets of proliferating T cells which had previously been grouped together, and showed that the previous dataset contained many cells of poor quality that had not been filtered out. TCAT also discriminates between CD8+ T cells polarisation states.

The authors introduce their AIMseq technique, looking at TCR dependent T cell activation and cGEPs that correlate with antigen reactivity. Using the cGEP enriched in activated T cells they then generate an ASA score by which the antigen reactivity of a cell can be assessed, and use this to analyse T cells derived from various cancer samples (in part by determining a ASA threshold at which to specify a cell as antigen reactive). The authors show that tumor mutation burden and the proportion of antigen-specific T cells correlate as would be expected, and then further analyse the cGEP phenotype of bystander T cells.

B. Originality and significance: if not novel, please include reference

Originality: NMF approaches have recently been applied to circulating CD4+ T cells (Single-cell transcriptome landscape of circulating CD4+ T cell populations in autoimmune diseases - ScienceDirect); this new study seeks to define gene expression profiles more broadly and determines a larger number of GEPs in an effort to make a shared reference representation of cell states, in particular T cell states.

The significance is hard to determine: the authors are correct that having a well characterised reference panel of GEPs that (in theory) captures all possible T cell GEPs would be of value in mapping studies (especially smaller studies) onto a common representation of cell states is of value to the field of single cell sequencing. The downside is that the cNMF approach is conceptually harder to understand than simpler approaches using reference maps such as Azimuth and ProjectILs, which may limit adoption. However the authors have made an admirable effort to make their tool available on github, present appropriate reference implementations, and are working on a web based tool that can be deployed on remote compute resources, which should facilitate implementation.

As the costs of sequencing continue to fall, and the size and availability of single cell datasets increases, this tool will likely be of increasing value (in particular to meta analyses)

While the authors present a huge amount of data, this paper would benefit from less data and a more compelling use case – for example the Sakaguchi group paper puts a clear focus on autoimmunity, I'd encourage the authors to put more emphasis on the last third of their paper looking at cancer to emphasise the novelty or additional value (though I appreciate that *CAT has uses beyond TCAT).

C. Data & methodology: validity of approach, quality of data, quality of presentation

As a general point the paper seems undecided as to whether it's a technical paper for a specialist audience with a high background level of understanding, or a newly generalisable method that might be of interest to less specialist readers who have single cell sequencing data. Given the journal is Nature Methods the figure legends are excessively terse which can make the data hard to understand: for example Fig 1A providing the overview of the whole approach is described in a single line of text. The figure itself could be expanded upon (or at least made larger) to help the reader grasp the core concept of this paper, while Fig1B is difficult to read and could be better linked to Fig1A_step2.

Compared to the previous NMF paper this paper has a more detailed methods section and a similar or more detailed supplementary data. The main figures are less crowded than the supplementals and have been prepared with care commensurate for a top tier journal.

D. Appropriate use of statistics and treatment of uncertainties

I do not have the requisite background in statistics to comment on the mathematical suitability of the new normalisation techniques used to combine the 7 datasets. Some figures lack statistics, and/or the statistical tools to compare these types of highly complex single cell datasets are not yet well established.

The authors have made the code underlying every figure available to allow readers to repeat their analyses which is admirable and underscores their high level of technical ability. If one were to criticise this might benefit from a little more curation - for example Analysis/TCAT/Polarization_MarkerHeatmaps.ipynb contains a number of KeyErrors which suggest that this might not be the final version used.

E. Conclusions: robustness, validity, reliability

The authors demonstrate a number of use cases for *CAT using TCAT to analyse diverse T cell repertoires to show that they can more accurately deconvolute T cell subsets.

F. Suggested improvements: experiments, data for possible revision

In general review figures and legends for clarity.

Fig S5B not mentioned in text

Section 1

I would also suggest tying the results of single cell sequencing/TCAT more closely to what the scientific audience know about T cells – for example in Fig1D how are known markers like ENTPD1 and CXCL13 that happen not to be among the top genes

per GEP distributed? Some information here – such as that HLA-DRA and HLA-DRB are top genes in the HLA cGEP – doesn't add much information.

I was surprised to see CXCL13 as a top weighted gene in T follicular helped cells, as this has been reported by many groups (Platten/Green, Wu, Rosenberg) as being a key marker for tumor reactive T cells

Section 3

Fig 2B – Accuracy does not reflect false positive assessments, a metric like G-Mean might better show the improved performance of the TCAT multilabel approach vs the single cGEP especially for CD8 EM. Is this what the authors mean in Fig2C with “balanced accuracy”? Possibly this is stated in the methods, but would be easier to state this in the legend/results.

Section 4

While TCAT does deconvolute the proliferating T cell cluster well, many studies perform regression of cell cycle genes so this is more a limitation of the COMBAT Covid-19 dataset than a feature of TCAT (though it nicely shows how previous data benefits from mapping onto a common reference).

Looking at Fig S4 comparing the two covid databases COMBAT with UK-Covid, while TCAT can deconvolute proliferating T cells into constituent subsets to show the important role of the T peripheral helper cGEP in covid patients vs healthy, there is little explanation of the large differences between these two covid datasets. Could the authors comment on this – I believe this is important given that a core argument is that datasets can (should?) be mapped to a common reference panel of GEPs. Dataset quality seems to be part of the answer given that the authors later comment on the high incidence of a cGEP indicative of poor cell quality in the COMBAT study. Why is the proportion of CD8trm cGEP cells in the proliferative state so different between studies? I would speculate that some cGEP are dependent on genes with low expression, such that poor sample handling (e.g. long delay between sample acquisition and single cell capture) depletes some cGEP more than others? The authors used an overdispersion metric to identify key genes for the cNMF analysis, but does this come at a price of robustness? I would like to see some discussion of these differences

When the authors state “Cells with high usage of the CD4-Naïve cGEP expressed CD4 naïve markers including CD45RA protein and SELL RNA, confirming that clustering misclassified them as memory T-cells (Extended Data 4)” some memory T cells also express CD45RA and SELL (e.g. TEMRA or Tcm)

Section 5 & 6

By definition naïve cells should not have encountered their cognate antigen. Clearly the immune system is constantly generating new TCR clonotypes, so indeed some CEFX reactive T cells could be naïve, however can the authors rule out a false positive result? By sequencing TCR clonotypes (as in section 6) one would expect that clonotypes in the naïve population should be less clonal/more diverse, and that T cells sharing the same clonotypes should not be found in both groups. Some analysis of TCR clonality would support this analysis and give additional confidence in the scoring (some of which is described in Fig S6).

While I would agree that proliferation is a core response of TCR activation, using it to validate cGEP choice seems less useful – I believe Fig 4F is one of the least convincing in the paper, with very large error bars and no stats. This should be better explained.

The ASA data is very interesting, but again large differences between COMBAT and UK-covid data in terms of absolute percentage are not explained. I would like to see more on this part of the paper – for example why are some cGEPs specific to particular cancers, presumably due to cancer microenvironment factors or similar? Fig 5J starts to get at this with a correlation between tumor mutational load and exhaustion, though seeing how ASA correlates with survival of individual patients would be more convincing (though such datasets are only just becoming available).

A large number of recent papers have shown CXCL13 to be key marker for tumor-reactive T cells (i.e. not just T peripheral helper cells), while virus reactive T cells seem to show a slightly difference activation pattern. See previous points on CXCL13, I'd like to see this explored more: could it be that the use of a viral peptide pool in AIMseq gives slightly different cGEP enrichment than true tumor reactivity?

G. References: appropriate credit to previous work?

Yes

H. Clarity and context: lucidity of abstract/summary, appropriateness of abstract, introduction and conclusions

The abstract, introduction and discussion sections are very well written.

Although we cannot publish your paper, it may be appropriate for another journal in the Nature Portfolio. If you wish to explore the journals and transfer your manuscript please use our [manuscript transfer portal](#). You will not have to re-supply manuscript metadata and files, unless you wish to make modifications. For more information, please see our http://www.nature.com/authors/author_resources/transfer_manuscripts.html?WT.mc_id=EMI_NPG_1511_AUTHORTRANSF&WT.ec_id=AUTHOR manuscript transfer FAQ page.

** For Nature Portfolio general information and news for authors, see <http://npg.nature.com/authors>.

Version 1:

Decision Letter:

18th Dec 2024

Dear Soumya,

Thank you for your letter asking us to reconsider our decision on your Resource, "Reproducible single cell annotation of programs underlying T-cell subsets, activation states, and functions". After careful consideration we have decided that we are willing to consider a revised version of your manuscript as you have outlined in your plan that highlights the technical advance and usability of your method..

When revising your paper:

- * include a point-by-point response to our referees and to any editorial suggestions
- * please underline/highlight any additions to the text or areas with other significant changes to facilitate review of the revised manuscript
- * address the points listed described below to conform to our open science requirements
- * ensure it complies with our general format requirements as set out in our guide to authors at www.nature.com/naturemethods
- * resubmit all the necessary files electronically by using the link below to access your home page

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.

We hope to receive your revised paper within ten weeks. If you cannot send it within this time, please let us know. In this event, we will still be happy to reconsider your paper at a later date so long as nothing similar has been accepted for publication at Nature Methods or published elsewhere.

OPEN SCIENCE REQUIREMENTS

REPORTING SUMMARY AND EDITORIAL POLICY CHECKLISTS

When revising your manuscript, please submit reporting summary and editorial policy checklists.

Reporting summary: <https://www.nature.com/documents/nr-reporting-summary.zip>

Editorial policy checklist: <https://www.nature.com/documents/nr-editorial-policy-checklist.zip>

If your paper includes custom software, we also ask you to complete a supplemental reporting summary.

Software supplement: <https://www.nature.com/documents/nr-software-policy.pdf>

Please submit these with your revised manuscript. They will be available to reviewers to aid in their evaluation if the paper is re-reviewed. If you have any questions about the checklist, please see <http://www.nature.com/authors/policies/availability.html> or contact me.

Please note that these forms are dynamic 'smart pdfs' and must therefore be downloaded and completed in Adobe Reader. We will then flatten them for ease of use by the reviewers. If you would like to reference the guidance text as you complete the template, please access these flattened versions at <http://www.nature.com/authors/policies/availability.html>.

DATA AVAILABILITY

Please include a "Data availability" subsection in the Online Methods. This section should inform readers about the availability of the data used to support the conclusions of your study, including accession codes to public repositories, references to source data that may be published alongside the paper, unique identifiers such as URLs to data repository entries, or data set DOIs, and any other statement about data availability. At a minimum, you should include the following statement: "The data that support the findings of this study are available from the corresponding author upon request", describing which data is available upon request and mentioning any restrictions on availability. If DOIs are provided, please include these in the Reference list (authors, title, publisher (repository name), identifier, year). For more guidance on how to write this section please see: <http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf>

CODE AVAILABILITY

Please include a "Code Availability" subsection in the Online Methods which details how your custom code is made available. Only in rare cases (where code is not central to the main conclusions of the paper) is the statement "available upon request"

allowed (and reasons should be specified).

We request that you deposit code in a DOI-minting repository such as Zenodo, Gigantum or Code Ocean and cite the DOI in the Reference list. We also request that you use code versioning and provide a license.

For more information on our code sharing policy and requirements, please see:

<https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-computer-code>

MATERIALS AVAILABILITY

As a condition of publication in Nature Methods, authors are required to make unique materials promptly available to others without undue qualifications.

Authors reporting new chemical compounds must provide chemical structure, synthesis and characterization details. Authors reporting mutant strains and cell lines are strongly encouraged to use established public repositories.

More details about our materials availability policy can be found at <https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards#availability-of-materials>

SUPPLEMENTARY PROTOCOL

To help facilitate reproducibility and uptake of your method, we ask you to prepare a step-by-step Supplementary Protocol for the method described in this paper. We [encourage authors to share their step-by-step experimental protocols](https://www.nature.com/nature-research/editorial-policies/reporting-standards#protocols) on a protocol sharing platform of their choice and report the protocol DOI in the reference list. Nature Portfolio's protocols.io is a free-to-use and open resource for protocols; protocols deposited onto protocols.io are citable and can be linked from the published article. More details can be found at [protocols.io](https://www.protocols.io/help/publish-articles).

ORCID

Nature Methods 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. This applies to primary research papers only. 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 www.springernature.com/orcid.

We look forward to hearing from you soon.

Happy Holidays !

Sincerely,
Madhura

Madhura Mukhopadhyay, PhD
Senior Editor
Nature Methods

Version 2:

Decision Letter:

Our ref: NMETH-RS56554B

18th Feb 2025

Dear Soumya,

Thank you for submitting your revised manuscript "Reproducible single cell annotation of programs underlying T-cell subsets, activation states, and functions" (NMETH-RS56554B). 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 Methods, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements within two weeks or so. Please do not upload the final materials and make any revisions until you receive this additional information from us.

TRANSPARENT PEER REVIEW

Nature Methods offers a transparent peer review option for new original research manuscripts submitted from 17th February 2021. 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](https://www.nature.com/documents/nr-transparent-peer-review.pdf).

ORCID

IMPORTANT: Non-corresponding authors do not have to link their ORCIDs but are encouraged to do so. Please note that it will not be possible to add/modify ORCIDs at proof. Thus, please let your co-authors know that if they wish to have their ORCID added to the paper they must follow the procedure described in the following link prior to acceptance: <https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research>

Thank you again for your interest in Nature Methods. Please do not hesitate to contact me if you have any questions. We will be in touch again soon.

Sincerely,
Madhura

Madhura Mukhopadhyay, PhD
Senior Editor
Nature Methods

Reviewer #2 (Remarks to the Author):

Reviewer #4 (Remarks to the Author):

In this revised manuscript the authors have added additional material to previous manuscript describing a *CAT, a new cell annotator pipeline that replaces discrete cell states (a hangover from the days of FACS analysis where cells were gated into bins) with flexible GEP (gene expression programs) that reflect how cells can have an intermediate or uncommitted state.

The authors are applying recent techniques (cNMF) into a new package that is readily accessible online. Compared to the previous version of this manuscript, the authors have addressed reviewer points in considerable detail: amongst others the authors have made the use of *CAT by the wider community even simpler by providing a website implementation (<https://immunogenomics.io/starcat/>), have performed benchmarking against existing tools showing superior performance, and have illustrated the utility of *CAT on real-world datasets concerning immune checkpoint inhibition. This resolves questions as to whether the authors' pre-trained GEPs will add value to the T cell community. While the *CAT results for the different clinical ICI datasets do not always generate the same results, there is considerable similarity which gives confidence in the robustness of the GEPs, and the authors reasonably point out that T cell states are affected by the tumor microenvironment, and may persist after ICI treatment, leading to different results in different tumor entities.

The paper is well written throughout, and now contains appropriate statistics for the few points where they were previously missing. The authors have also clarified figures and expanded methods where appropriate. As before the authors' code is fully accessible to the community, and extraneous or confusing code has been tidied up for this revised manuscript.

The authors have to my mind appropriately addressed all reviewer comments, and the extensive new analyses and comparisons underscore the value of the method. I am particularly happy to see a simple web implementation of *CAT making it accessible to the scientific community, in contrast to so many computational methods that gain little traction due to complex installation or use requirements. I am therefore happy to recommend accepting the manuscript in its current form

Reviewer #4 (Remarks on code availability):

I previously reviewed the code and see that the authors have made suggested edits, though I have not this time run the code - the new website the authors present makes this unnecessary.

Version 3:

Decision Letter:

11th Jul 2025

Dear Soumya,

I am pleased to inform you that your Resource, "Reproducible single cell annotation of programs underlying T cell subsets, activation states, and functions", has now been accepted for publication in Nature Methods. The received and accepted dates will be 24 May 2024 and 11 Jul 2025. This note is intended to let you know what to expect from us over the next month or so, and to let you know where to address any further questions.

Over the next few weeks, your paper will be copyedited to ensure that it conforms to Nature Methods style. Once your paper is typeset, 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. It is extremely important that you let us know now whether you will be difficult to contact over the next month. If this is the case, we ask that you send us the contact information (email, phone and fax) of someone who will be able to check the proofs and deal with any last-minute problems.

After the grant of rights is completed, 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.

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates.

If your research is supported by a funder that requires immediate open access (e.g. according to [Plan S principles](https://www.springernature.com/gp/open-science/plan-s-compliance) or the [NIH public access policy](https://www.springernature.com/gp/open-science/us-federal-agency-compliance)) then you should select the gold OA route, and we will direct you to the compliant route where possible. Because authors warrant under our subscription licensing terms that they haven't committed to licensing any version of their article under a licence inconsistent with the terms of our agreement – including the applicable embargo period – publication under the subscription model isn't suitable for authors whose funders require no embargo.

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

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

You may wish to make your media relations office aware of your accepted publication, in case they consider it appropriate to organize some internal or external publicity. Once your paper has been scheduled you will receive an email confirming the publication details. This is normally 3-4 working days in advance of publication. If you need additional notice of the date and time of publication, please let the production team know when you receive the proof of your article to ensure there is sufficient time to coordinate. Further information on our embargo policies can be found here: <https://www.nature.com/authors/policies/embargo.html>

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.

If you are active on Twitter/X or Bluesky, please e-mail me your and your coauthors' handles so that we may tag you when the paper is published.

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.

Please note that you and any of your coauthors will be able to order reprints and single copies of the issue containing your article through Nature Portfolio's reprint website, which is located at <http://www.nature.com/reprints/author-reprints.html>. If there are any questions about reprints please send an email to author-reprints@nature.com and someone will assist you.

Please feel free to contact me if you have questions about any of these points.

Best regards,
Madhura

Madhura Mukhopadhyay, PhD
Senior Editor
Nature Methods

** Visit the Springer Nature Editorial and Publishing website at www.springernature.com/editorial-and-publishing-jobs for more information about our career opportunities. If you have any questions please click here.**

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source. The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

1 We thank the reviewers and the editor for their thoughtful feedback on our manuscript. Overall,
2 we were encouraged by the comments. For example, reviewer #2 noted “*Annotating T cell*
3 *subsets and unifying prior descriptions of T cell functional states is a very common and*
4 *important task for many studies involving single-cell sequencing analysis of T cells.*” Reviewer
5 #4 noted “*having a well characterised reference panel of GEPs that (in theory) captures all*
6 *possible T cell GEPs would be of value in mapping studies (especially smaller studies)... [and] is*
7 *of value to the field of single cell sequencing... the authors have made an admirable effort to*
8 *make their tool available... which should facilitate implementation.*”
9

10 Based on the feedback from the reviewers, we have made four major revisions to emphasize
11 the unique value of the method and to address concerns about validation:
12

13 **Key Revision 1.** We analyzed 3 new cancer datasets showing that TCAT reveals gene
14 expression programs (GEPs) associated with immune checkpoint inhibitor responsive
15 and resistant tumors (**new Section 8, Figure 6, R2.1, R4.2, R4.15**).
16

17 **Key Revision 2.** We performed additional benchmarking against Azimuth, ProjectTILs,
18 and Symphony. TCAT out-performed all three for T-cell lineage prediction (**new Figure**
19 **2B-C, S3B, R2.1, R2.2, R4.2**).
20

21 **Key Revision 3.** We benchmarked TCAT’s activation prediction against 9 literature-
22 derived gene sets. TCAT out-performed 9/9 for prediction *in vivo* and 7/9 *ex vivo* (**new**
23 **Figure S6F-G, R2.1, R2.2**).
24

25 **Key Revision 4.** We created a web tool to expand starCAT beyond T-cells to any cell
26 type or tissue. We currently include published myeloid and bone marrow hematopoiesis
27 references (**R4.2**).
28

29 We describe the four major revisions in detail below and respond point by point to the reviewer
30 comments after that. We greatly appreciate the opportunity to submit a revised manuscript and
31 look forward to your response.
32
33
34
35
36
37
38
39
40
41
42
43
44

45 Key revision 1 - Biological novelty

46
47 We added a more clinically impactful analysis to address reviewer 2's concern that there "does
48 not seem to be major new insights gained based on the use of this approach" and reviewer 4's
49 comment that "seeing how ASA correlates with survival of individual patients" would be more
50 convincing. Using T-CellAnnoTator, we demonstrated important characteristics of CD4 T-cells
51 that are consistently associated with immune-checkpoint inhibitor (ICI) therapy response across
52 three cancer types (Section 8 and new figure 6 in the revised text). We re-analyzed
53 published scRNA-Seq from melanoma, colorectal cancer, and non-melanoma skin cancer using
54 TCAT. We found consistent up-regulation of antigen-specific activation and cell cycle cGEPs in
55 ICI non-responders and consistent up-regulation of naive T-cell cGEPs in ICI responders. These
56 findings are biologically and clinically impactful since they demonstrate new properties of tumor
57 immunology that are strongly and consistently associated with response to ICI. TCAT was
58 essential for this analysis since it enables fast and rigorous cross-comparisons of T cell gene
59 expression programs across datasets. For convenience, we reproduce the new section of the
60 manuscript and the corresponding figures below:

61 8. Characterizing immune checkpoint inhibitor response using TCAT

62 *We demonstrate an application of TCAT by identifying cGEPs that predict tumor*
63 *response to immune checkpoint inhibitors (ICIs). While ICIs are state of the art for*
64 *treating many cancer types, five-year survival following treatment remains poor for over*
65 *half of treated patients¹. There remains much to learn about the properties of tumor-*
66 *infiltrating T-cells that determine ICI response. We therefore applied TCAT to*
67 *melanoma², non-melanoma skin cancer (NMSC)³, and colorectal cancer (CRC)^{3,4}*
68 *datasets. Each of these datasets contained pre- and post-treatment tumors with*
69 *information about therapeutic response. This enabled us to define T-cell cGEPs that*
70 *potentially predict ICI clinical response.*

71
72 *To define the prominent T-cell states within tumors, we first examined the melanoma*
73 *dataset², which had the most pre-treatment and total samples (N=19 and 48*
74 *respectively). Application of TCAT revealed T-cell populations expressing antigen-*
75 *specific activation (ASA), exhaustion, cell cycle, and CD4-Naive signatures (Figure 6A).*
76 *We also noted a prominent subset of cells expressing TCF7 which was previously*
77 *associated with good ICI response².*

78
79 *Next, to find predictors of ICI response in melanoma, we systematically associated*
80 *cGEPs with ICI response in pre-treatment tumors (nine responders and ten non-*
81 *responders) for CD4 and CD8 T-cells, separately. In CD4 T-cells, non-responders had*
82 *significantly higher usage of cell cycle and activation cGEPs such as TIMD4/TIM3,*
83 *OX40/EBI3, CellCycle-Late-S, and CellCycle-G2M (P<0.05 two-tailed T-test, Table S8,*
84 *Figure 6B). To assess if similar associations were seen in an independent dataset, we*
85 *examined a non-melanoma skin cancer (NMSC) dataset³ containing seven non-*

86 responders and six responders, pre-treatment. There was significant concordance
87 between the melanoma and NMSC datasets in the direction of the associations for the
88 52 TCAT signatures, suggesting generally consistent signals (sign test $P=0.0016$,
89 **Figure S8A**). Focusing specifically on the melanoma treatment response cGEPs, we
90 observed a consistent association of cell cycle and activation-associated cGEPs such as
91 CellCycle-Late-S and TIMD4/TIM3 (one-tailed T-test $P=0.037$ and $P=0.046$, respectively,
92 **Figure 6C**). Meta-analysis of the associations between these two datasets was
93 statistically significant for many activation and cell cycle cGEPs including TIMD4/TIM3,
94 CellCycle-Late-S, Exhaustion, and ICOS/CD38 ($P<0.05$ for all, **Table S8**). Associations
95 were similarly significant for the combined ASA and cell cycle scores (Meta-analysis
96 $P=0.0072$ and $P=0.0036$ for ASA and cell cycle, respectively, **Figure 6E-F**). These
97 findings suggest that having a greater amount of TCR activated and/or proliferating T-
98 cells in pre-treatment skin cancers predicts a worse response to ICI.

99
100 We also observed higher usage of the CD4-Naive cGEP in responders for pre-treatment
101 melanoma and NMSC tumors (meta-analysis $P=0.0063$, **Figure 6G, Table S8**). This is
102 consistent with a previous observation that TCF7 expression is associated with a
103 positive response to ICI, as TCF7 is the 5th most strongly associated gene with the
104 CD4-Naive cGEP (**Table S2**). In melanoma, cells annotated as CD4 Naive represented
105 a distinct subset of the TCF7 expressing cells; this subset was especially enriched for
106 ICI responsiveness (**Figure 6A**). Consistent with this, we found increased expression of
107 the naive T-cells markers CCR7 and SELL in pseudobulked CD4 T-cells from pre-
108 treatment responders (meta-analysis $P=0.024$, 0.0016 , and 0.00047 for TCF7, SELL,
109 and CCR7, respectively, **Figure S8B-D**). Furthermore, the percentage of CD4 T-cells
110 classified as naive by TCAT's lineage prediction was also significantly associated with
111 ICI responsiveness (meta-analysis $P=0.016$, **Figure S8E**). These findings suggest that
112 increased tumor infiltrating naive CD4 T-cells may predict a good response to ICI.

113
114 We sought to replicate these analyses in a similar CRC dataset, but unfortunately, it only
115 had one pre-treatment non-responder sample. However, we note that there is strong
116 overall agreement in association results from samples that are pre-treatment or post
117 treatment initiation (**Table S8**). For example, there was a significant correlation in the
118 association effect sizes between pre-treatment and post-treatment samples for all three
119 datasets ($R=0.64$, 0.51 , 0.56 for melanoma, NMSC, and CRC respectively, $P<1\times 10^{-4}$,
120 **Figure S8F-H**). In addition, six out of ten cGEPs that were significantly increased in pre-
121 treatment non-responder melanomas were also significant in post-treatment samples,
122 including many cell cycle and activation cGEPs (**Figure 6B** and **Extended Data 10A**,
123 Fisher exact test $P=0.0039$). This suggests that some key immune states are driven by
124 the tumor environment and persist after ICI treatment.

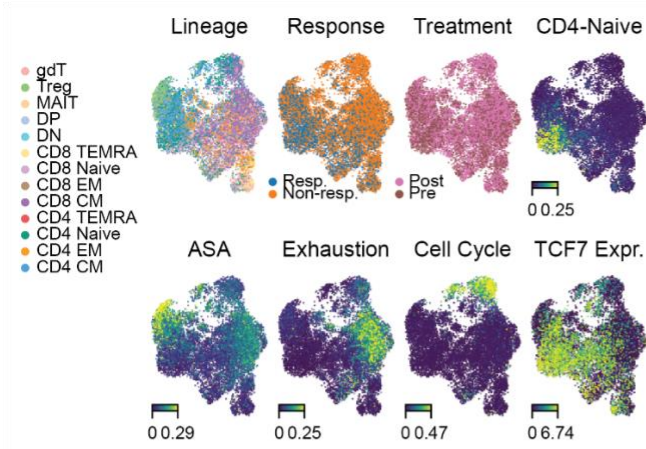
125
126 Making the assumption that pre and post-treatment samples largely share many immune
127 states, we repeated the associations using linear regression in combined pre- and post-
128 treatment samples, modeling treatment status as a fixed effect and patient of origin with
129 random effects (**Methods**). With this approach, we observed results that were consistent

130 with the prior analyses. For CRC, this model showed increased activation and cell cycle
131 cGEPs in ICI non-responders (e.g. $P < 1 \times 10^{-6}$ for OX40/EBI3 and CellCycle-Late-S,
132 **Figure 6D**). CRC pre-treatment tumors followed the expected trend of lower ASA and
133 cell cycle scores in responders than non-responders (**Extended Data Figure 10B**).
134 Application of this model to the melanoma and NMSC datasets was consistent with the
135 previous analyses in pre-treatment tumors (**Extended Data Figure 10C-D**). These
136 distinct analyses support that our findings are robust and reproducible in three cancers.

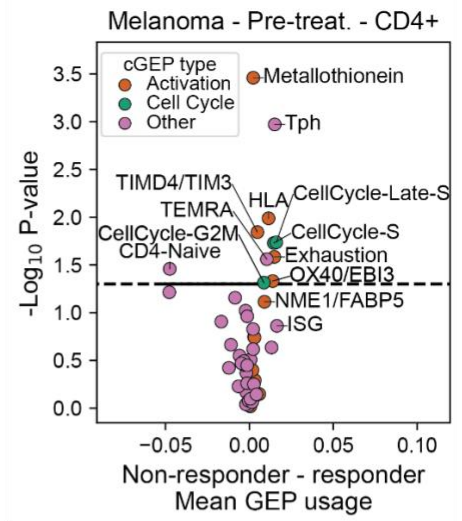
137
138 Lastly, we repeated these analyses in CD8 T-cells (**Figure S8I-K**). As was the case for
139 CD4s, there was increased activation and cell cycle cGEPs usage in pre-treatment non-
140 responders (**Table S8**, meta-analysis $P < 0.05$ for CellCycle-G2M, CellCycle-Late-S,
141 TIMD4/TIM3, Exhaustion, ICOS/CD38, OX40/EBI3, and HLA). These cGEPs were also
142 significant in the mixed linear models analysis of CD8 T-cells from CRC, modeling
143 treatment status as a fixed effect and patient of origin as random effects ($P < 0.05$ for all).
144 Similarly, the CD8-Naive cGEP was significantly associated with a positive response to
145 ICI in pre-treatment samples (meta-analysis $P = 0.032$). Thus, the above findings from
146 CD4s are consistent in CD8s as well. These analyses illustrate how TCAT can help
147 reveal consistent and clinically significant patterns of GEPs across multiple datasets.

148
149
150

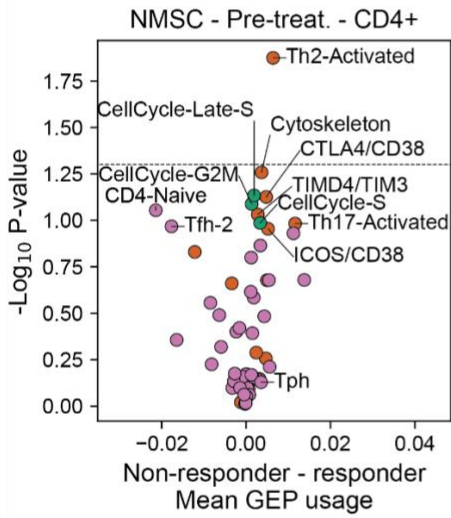
A



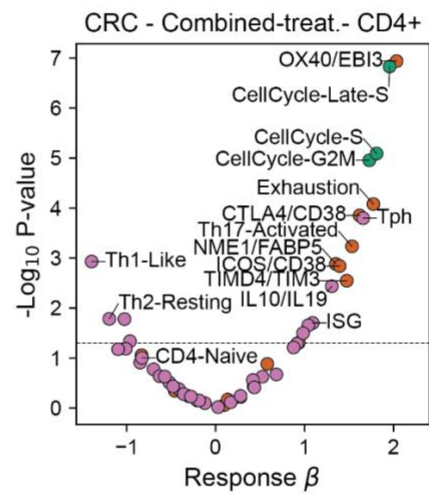
B



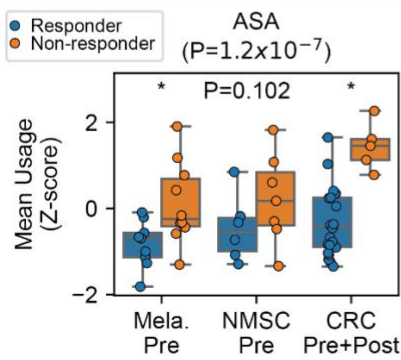
C



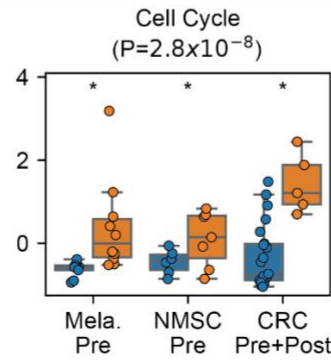
D



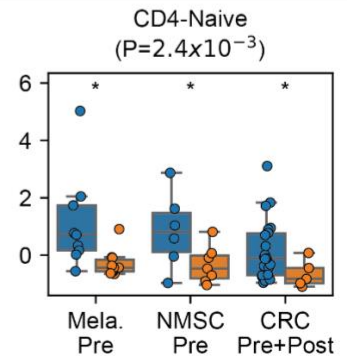
E



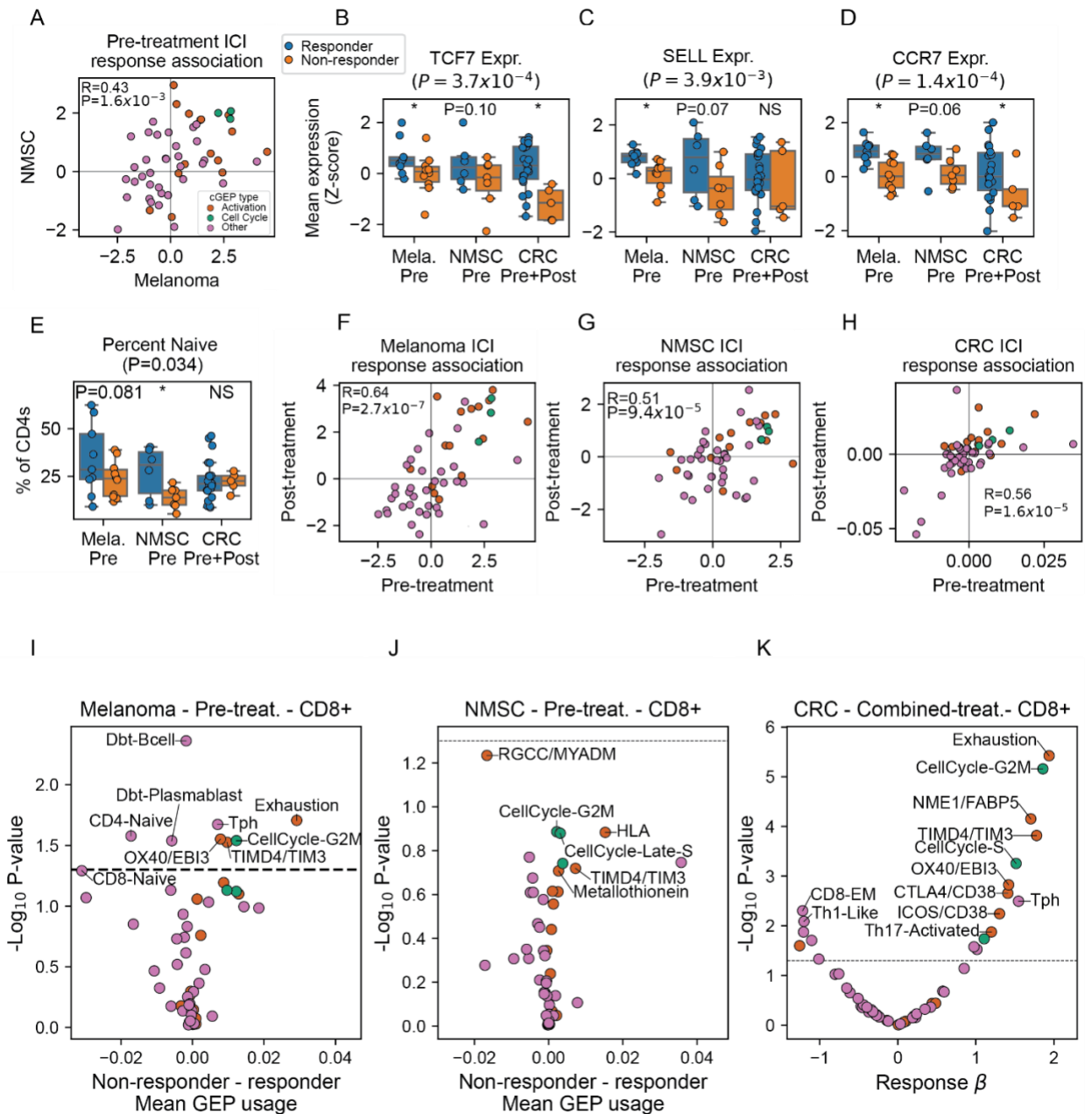
F



G



152 **New Figure 6. cGEPs associated with immune checkpoint inhibitor response.** (A)
153 UMAPs of the melanoma dataset showing TCAT predicted lineage; the CD4-Naive and
154 Exhaustion cGEPs; the ASA and cell cycle scores; TCF7 expression; and treatment
155 status and response. (B-D) Associations of cGEP usage with ICI response in CD4 T-
156 cells of (B) pre-treatment melanoma, (C) pre-treatment non-melanoma skin cancer
157 (NMSC) and (D) combined pre- and post- treatment colorectal cancer (CRC) controlling
158 for treatment timepoint and donor of origin. Dots are colored by cGEP type. X-axis
159 shows the average difference in usage between non-responders and responders or the
160 regression coefficient. Y-axis shows the $-\text{Log}_{10}$ two-tailed P-value. (E-G) Average ASA
161 score, Cell Cycle score, and CD4-Naive cGEP usage in pre-treatment melanomas and
162 NMSCs and combined pre- and post-treatment CRCs. The average scores are mean
163 and variance normalized. P-values are one-tailed T-tests for melanoma and NMSC and
164 mixed linear regression P-values for CRC. P-value in the title is a meta-analysis of the P-
165 values across the three cancer types. * indicates $P < 0.05$. Boxes show interquartile
166 range.
167



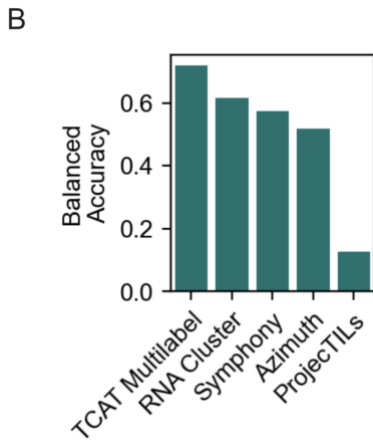
168
 169
 170
 171
 172
 173
 174
 175
 176
 177

New Figure S8. cGEPs associated with immune checkpoint inhibitor response. (A) T-statistics from pre-treatment melanoma and NMSC associations. R denotes Pearson correlation. Dots are colored by cGEP type. (B-D) Average expression of naive T-cell marker genes. Melanoma and NMSC P -values are from one-tailed T-tests and CRC P -values are from a mixed linear regression likelihood ratio tests. P -value in the title is from a meta-analysis across the three cancer types. (E) Same as B but showing percentage of CD4 Naive T-cells out of total CD4 T-cells based on TCAT lineage classification. (F-G) T-statistics computed in pre- and post- treatment CD4 T-cells for melanoma and non-melanoma skin cancer (NMSC). (H) Same as F but showing mean difference rather than

178 *T*-statistic due to the presence of only one pre-treatment non-responder. (I-K) Same as
179 **Figure 6B-D** but for CD8 T-cells.
180

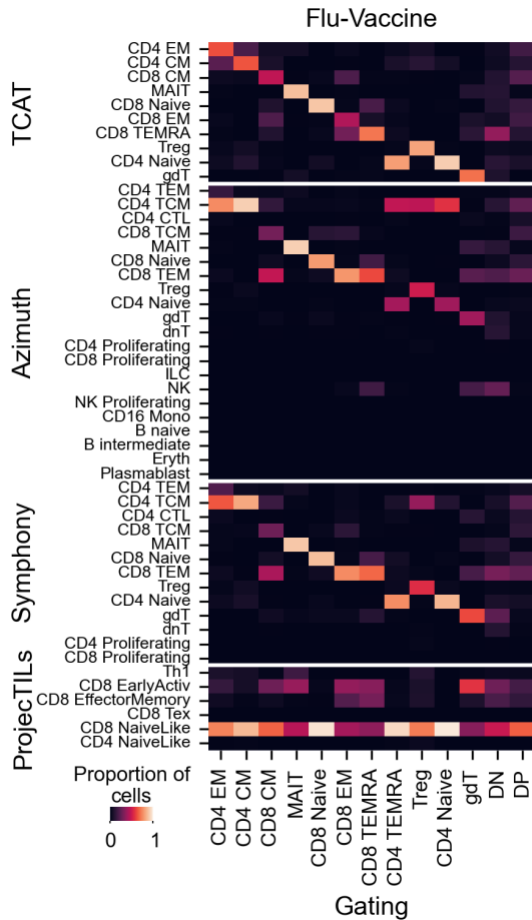
181 **Key revision 2 - Additional benchmarking**

182
183 We wanted to confirm that TCAT quantitatively and qualitatively outperformed commonly used
184 alternative reference mapping strategies. Hence, we benchmarked TCAT's lineage prediction
185 feature against Azimuth⁵ and ProjectTils⁶, the two reference mapping methods specifically
186 mentioned by the reviewers, and Symphony⁷, a commonly used categorical-annotation method
187 developed by our group. We found substantially higher prediction accuracy for classifying T-cell
188 subsets than these alternative methods in our validation Flu-Vaccine dataset (**New Figure 2B,**
189 **S3B**). For example, TCAT greatly outperformed other methods for classification of CD4 effector
190 memory cells in the Flu-Vaccine dataset (accuracy 0.80 compared to 0.57, 0.60, and 0.52 for
191 Azimuth, Symphony, and ProjectTILs respectively). TCAT also predicted TEMRA lineage
192 (accuracy 0.83), which was not predicted by any of these alternative methods. While discrete
193 lineage prediction is only a single facet of TCAT, it was reassuring that TCAT remains state of
194 the art for this task.
195



196
197 **New Figure 2B.** Cross-method comparison of balanced accuracy for prediction of
198 manually gated subsets.
199

200

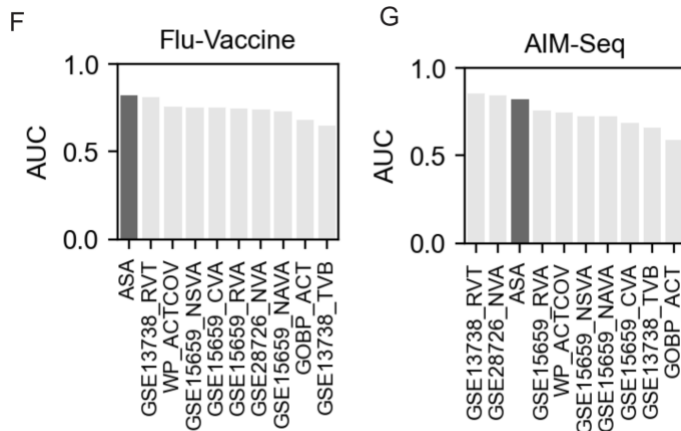


201
 202 **New Figure S3B.** Proportion of cells within each manually gated lineage (columns)
 203 assigned to each lineage annotation (rows). Proportion of cells sums to 1 within each
 204 column, for each annotation method.
 205

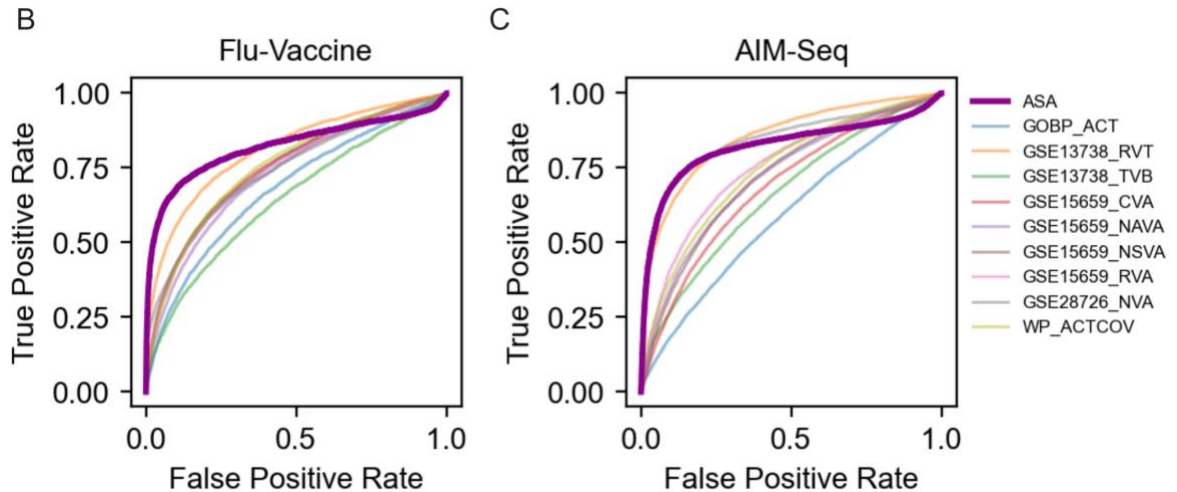
206 **Key revision 3 - Clarifying biological value of the method**

207
 208 Reviewer comments helped identify a lack of clarity about the method which we now specifically
 209 address in the text. Alternative methods such as Azimuth, ProjectTils, and Symphony label cells
 210 with a single mutually exclusive label. In contrast, starCAT simultaneously annotates cells with
 211 multiple distinct features. This avoids confounding of multiple unrelated signals, such as when
 212 cell activation or cell quality obscures T-cell subset, or vice versa. To illustrate the additional
 213 insights enabled by TCAT over these other methods, we added a new benchmark of annotating
 214 antigen-specific activation (ASA), a crucial feature of T-cells that the other reference transfer
 215 methods do not explicitly tackle. To address this task, researchers typically score cells based on
 216 literature-derived activation gene sets. However, *a priori*, it is difficult to know which gene set to
 217 use, and different gene sets can yield widely varying results. Our manuscript adds biological
 218 value by developing two novel benchmarks to address this: (1) prediction of AIM-Seq, an
 219 experimental assay of T-cell activation, and (2) prediction of a CITE-Seq derived activation
 220 label. We compared TCAT's ASA score against 9 literature-derived gene sets for classifying
 221 activation based on surface activation markers in the Flu-Vaccine and AIM-Seq datasets. ASA

222 outperformed all gene sets for activation classification in the Flu-Vaccine dataset and all but 2
 223 gene sets in the AIM-Seq data (**New Figure S6F-G, Extended Data 8B-C**). Thus, in addition to
 224 outperforming reference label transfer methods for lineage classification, it also outperformed
 225 most literature-defined gene sets for classifying T-cell activation. We also added emphasis in
 226 the manuscript that starCAT concurrently annotates a broad range of T-cell features, such as
 227 polarization and exhaustion, which existing reference transfer pipelines currently do not predict.
 228



229
 230 **New Figure S6. (F)** In the Flu-Vaccine dataset, AUCs for prediction of CD71+CD95+ co-
 231 expression by ASA as compared to cell scoring with alternative T-cell gene sets. Full
 232 gene set names are GOBP_ACTIVATED_T_CELL_PROLIFERATION (GOBP_ACT),
 233 GSE13738_RESTING_VS_TCR_ACTIVATED_CD4_TCELL_DN (GSE13738_RVT),
 234 GSE13738_TCR_VS_BYSTANDER_ACTIVATED_CD4_TCELL_UP (GSE13738_TVB),
 235 GSE15659_CD45RA_NEG_CD4_TCELL_VS_ACTIVATED_TREG_DN
 236 (GSE15659_CVA), GSE15659_NAIVE_CD4_TCELL_VS_ACTIVATED_TREG_DN
 237 (GSE15659_NAVA),
 238 GSE15659_NONSUPPRESSIVE_TCELL_VS_ACTIVATED_TREG_DN
 239 (GSE15659_NSVA), GSE15659_RESTING_VS_ACTIVATED_TREG_DN
 240 (GSE15659_RVA), GSE28726_NAIVE_VS_ACTIVATED_CD4_TCELL_DN
 241 (GSE28726_NVA), WP_TCELL_ACTIVATION_SARSCOV2 (WP_ACTCOV). (G) Same
 242 as (F) but prediction of AIM-positivity in the AIM-Seq dataset.
 243



244
245

246 **New Extended Data 8. (B)** ROC curves for ASA prediction of CD71+CD95+ in the Flu-
247 Vaccine dataset, as compared to prediction using per-cell gene scoring with T-cell
248 activation gene sets. **(C)** Same as **(B)** for AIM-positivity prediction in the AIM-Seq

249 dataset. Full gene set names (in order) are

250 GOBP_ACTIVATED_T_CELL_PROLIFERATION,

251 GSE13738_RESTING_VS_TCR_ACTIVATED_CD4_TCELL_DN,

252 GSE13738_TCR_VS_BYSTANDER_ACTIVATED_CD4_TCELL_UP,

253 GSE15659_CD45RA_NEG_CD4_TCELL_VS_ACTIVATED_TREG_DN,

254 GSE15659_NAIVE_CD4_TCELL_VS_ACTIVATED_TREG_DN,

255 GSE15659_NONSUPPRESSIVE_TCELL_VS_ACTIVATED_TREG_DN,

256 GSE15659_RESTING_VS_ACTIVATED_TREG_DN,

257 GSE28726_NAIVE_VS_ACTIVATED_CD4_TCELL_DN,

258 WP_TCELL_ACTIVATION_SARSCOV2.

259

260 Key revision 4 - Application to other cell types

261

262 Additionally, we created a web tool to expand the application of starCAT beyond T-cells to any

263 cell type or tissue. Since its publication in 2019, consensus non-negative matrix factorization

264 (cNMF) has become increasingly popular for characterizing cell states. Each time it is used

265 creates the opportunity to develop a reference that can be applied to new query datasets,

266 enabling researchers to annotate their datasets using a shared set of biological GEPs. To

267 broaden the use of starCAT, we created a website <https://immunogenomics.io/starcat> to host

268 such references and enable users to annotate their data from the browser. We use

269 WebAssembly, a recently developed, powerful software infrastructure that enables all of the

270 computation to be run in the web browser without requiring user data to be uploaded to a

271 remote server⁸. In addition to our T-cell reference, we have already included a human glioma

272 myeloid cell reference⁹ and a human bone marrow hematopoiesis reference¹⁰. These

273 references can also be used easily from Python, the command line, or through the starCAT

274 website. This database substantially expands the impact of our work beyond T-cells to all cell

275 types, greatly increasing the utility of our resource.

276 **Reviewer #1 comments:**

277

278 **R1.1. gene expression program GEP needs to be defined and explained in the**
279 **introduction**

280

281 We thank the reviewer for their comment. We have added additional text in the introduction to
282 define gene expression program:

283

284 *Transcriptomes reflect the expression of multiple gene expression programs (GEPs) –*
285 *co-regulated genes that are co-expressed for specific biologic functions such as defining*
286 *a cell type, activation state, lifecycle process, or response to an external stimulus¹¹.*

287

288 **R1.2. The abbreviation of the tool is confusing -- with "*". Not sure how it will be**
289 **searchable on the web etc**

290 **R1.3. Not sure what wildcard character stands for in the name of the tool**

291

292 We thank the reviewer for these comments. We now refer to the tool as starCAT throughout the
293 manuscript, rather than *CAT. We agree that this will improve the searchability of this term and
294 its code. We have added the extra text below to better explain the name:

295

296 *Here, we present star-CellAnnoTator (starCAT), an approach to score cells based on a*
297 *fixed, multidataset catalog of GEPs. The wildcard character "*", written as "star" in the*
298 *name, indicates that the approach can be applied using a reference from any tissue or*
299 *cell-type, whereas our specific instantiation for T-cell GEPs is written as T-CellAnnoTator*
300 *or TCAT.*

301

302 **R1.4. In the results section, the author says that the batch effect could not hinder the**
303 **identification of irreproducible GEP. This needs to be explained in more detail**

304

305 We thank the reviewer for this comment. Batch effects can result in learning GEPs reflecting
306 non-generalizable dataset-specific signals. For example, running cNMF prior to batch-correction
307 identified multiple highly correlated versions of the Translation GEP, where each version was
308 significantly enriched within one batch. However, after batch correction, a single GEP was used
309 consistently across the batches and had higher correlation across datasets. To make this
310 potential pitfall more clear to the readers, we have added the following additional text to the
311 introduction:

312

313 *For our approach, it was essential to amalgamate the GEP spectra from multiple*
314 *datasets. However, we found that batch effects could sometimes cause cNMF to learn*
315 *multiple batch-specific versions of a GEP, resulting in redundant GEPs with less*
316 *correlation across datasets. Most batch correction methods are not compatible with*
317 *cNMF since they create many negative values or correct low-dimensional embeddings*
318 *rather than gene-level data. We therefore modified Harmony¹² to produce batch-*
319 *corrected non-negative values for gene-level data rather than principal components.*

320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363

R1.5. Line 135. Authors claimed that reference datasets shared only 90% of genes. This needs to be explained and explained why this is expected

We thank the reviewer for this comment. Public datasets may share only a subset of the same genes due to differences in alignment pipelines, reference transcriptome versions, and filtering steps taken by the generators of the dataset. We therefore wanted to ensure that our simulation was robust to non-overlapping gene lists. We have added the text below to clarify why we chose this simulation parameter:

To ensure that starCAT is robust to incompletely overlapping genes between the reference and query dataset, we simulated these datasets to share only 90% of genes in common.

R1.6. Line 153. It would be helpful to provide the table with a description of all datasets being used with details about those databases. Also explain what criteria was used to select those public datasets

We thank the reviewer for their comment. We have added a new tab to **Table S1** providing information about the reference and validation datasets included in the paper. We have also added text clarifying the criteria we used to select these datasets:

We analyzed T-cells from 7 diverse datasets including blood and tissues from healthy individuals or individuals with Covid-19, cancer, rheumatoid arthritis, or osteoarthritis (Table S1, Figure S1F). In addition to selecting datasets for a breadth of phenotypes, we also selected for large sample sizes (>70,000 T-cells each), and inclusion of CITE-Seq data, when possible, to aid in GEP interpretation. We also included two Covid-19 PBMC, two healthy PBMC, and two tissue datasets each to help quantify cross-dataset GEP reproducibility.

R1.7. Line 151. Which other tissues in addition to blood were used

We thank the reviewer for their comment. We have included with the new **Table S1** a list of all tissues included for each dataset. This includes the following: joint synovium, lung, lung-draining lymph node, mesenteric lymph node, bone marrow, thymus, skeletal muscle, liver, spleen, omentum, duodenum, jejeunal lamina propria, jejeunum epithelium, Ileum, caecum, transverse colon, sigmoid colon, basal cell carcinoma, B-cell lymphoma, bladder cancer, esophageal cancer, fallopian tube carcinoma, hepatocellular carcinoma, lung cancer, multiple myeloma, ovarian cancer, pancreatic cancer, renal carcinoma, thyroid carcinoma, and uterine carcinoma.

R1.8. Line 160. p-value was not provided

We thank the reviewer for their comment. We have provided the P-values for the correlation statistics in the text ($P < 1 \times 10^{-50}$ for all pairs of GEPs).

364 **R1.9. Line 181. Mentions IGKV gene. The relationship to the T cell needs to be explained**

365

366 We thank the reviewer for their comment and recognize that this reference to an
367 immunoglobulin gene in a T-cell paper may be surprising. Although each of the discovery
368 datasets was subset specifically to T-cells, there are still cases where a non-T-cell may have
369 escaped the filters, likely due to being a doublet with a T-cell (i.e. when a T-cell and non T-cell
370 are captured in the same droplet). Thus, cNMF can learn some gene expression programs
371 (GEPs) that correspond to non-T-cell cell-types. We suspect the cGEP containing *IGKV*
372 reflected a plasmablast doublet GEP and thus labeled it as such. We have clarified this by
373 adding the following additional text:

374

375 *In addition to the 46 T-cell cGEPs, we identified six cGEPs corresponding to non T-cell*
376 *populations, including erythrocytes (HBA2, HBA1, HBB) and plasmablasts (JCHAIN,*
377 *IGKC, IGKV3-20). We suspect these cGEPs were due to residual doublets in the*
378 *datasets and retained them to flag doublet-associated transcriptional signals.*

379

380 **R1.10. Line 222. Mentioned ground truth. It needs to be explained why manual gating of**
381 **surface proteins will provide ground truth. Also, it would be helpful to discuss if such a**
382 **dataset indeed can be considered as ground truth or gold standard See this paper**
383 **discussing the type of gold standard datasets [https://www.nature.com/articles/s41467-](https://www.nature.com/articles/s41467-019-09406-4)**
384 **[019-09406-4](https://www.nature.com/articles/s41467-019-09406-4)**

385

386 We thank the reviewer for this comment. We agree that defining cell subsets based on manual
387 gating of a small number of surface markers is fundamentally limited. However, flow cytometry
388 based gating of canonical surface protein markers is the primary way immunologists have
389 traditionally defined T-cell subsets. Thus, much of immunology knowledge about T-cell subsets
390 relates to populations defined based on these markers. Consequently, this approach is
391 frequently used as a gold standard in computational immunology, e.g. in benchmarking studies
392 of bulk deconvolution¹³ or single-cell RNA-Seq cell type annotation¹⁴.

393

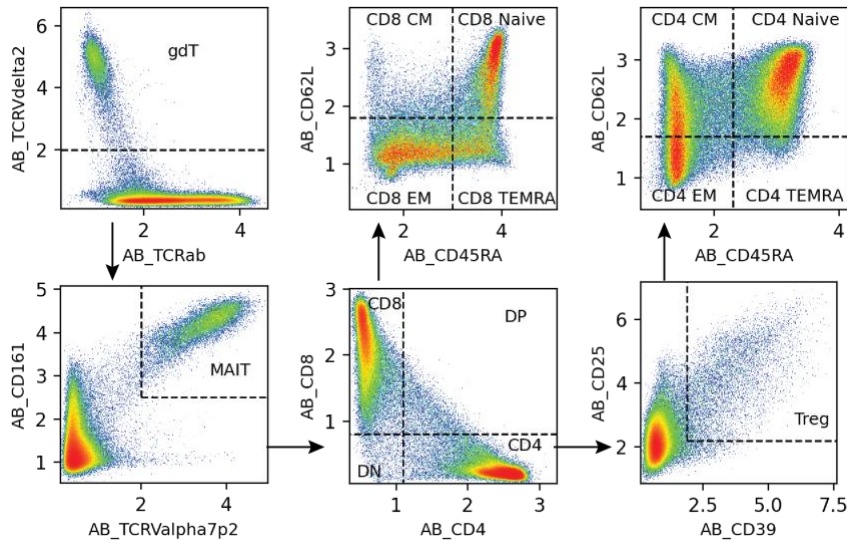
394 While we used gating of CITE-Seq data rather than flow cytometry for our predictive task, the
395 data looked very comparable to standard flow cytometry T-cell gating (**See Figure S3A**
396 **reproduced below for convenience**). Even if the gating is unlikely to be 100% accurate for
397 classifying functionally specific T-cell subsets, we still find it useful as a feature to be predicted
398 in this analysis. First, it is based on protein data and is therefore independent of the
399 transcriptome-features used for prediction. Second, it connects back to the conventional flow
400 cytometry paradigm used in traditional T-cell immunology and can thus ground single-cell
401 analyses in the literature.

402

403 Thus, we have removed the term “ground truth” from this section of the text and instead focus
404 on explaining why this feature is a useful one to predict in the task. We have revised the text as
405 below to explain this:

406

407 We used manual gating of surface proteins to define ten conventional T-cell subsets as
 408 prediction targets (**Figure S3A**). These protein-based annotations are useful for
 409 benchmarking as they are measured independently from the transcriptome-based
 410 features used for prediction and connect directly to the wealth of literature about
 411 canonical flow-cytometry defined T-cell subsets.
 412



413 **Figure S3. Benchmarking starCellAnnoTator on simulated and real datasets.** (A) Manual
 414 gating of Flu-Vaccine dataset analogous to **Figure S2A**.
 415

416 **Reviewer #2 comments:**

417
 418 **R2.1** In this paper by Kotlia et al., the authors effectively apply their previously published
 419 **cNMF** algorithm to create an annotator for T-cell gene expression programs, with
 420 **successful results**. Annotating T cell subsets and unifying prior descriptions of T cell
 421 **functional states** is a very common and important task for many studies involving single-
 422 **cell sequencing analysis** of T cells. Their aggregation and consolidation of 46 GEPs
 423 **describing possible T cell states** into a single easy-to-use package is potentially quite
 424 **useful**. Although this package and method does seem to be generally useful, it is not
 425 **clear to me** that this method is highly differentiated from standard approaches. For the
 426 **examples provided**, there also does not seem to be major new insights gained based on
 427 **the use of this approach**.

428
 429 We thank the reviewer for their comment about the general usefulness of the approach.

430
 431 To address the reviewer’s point about differentiation from standard approaches, we have added
 432 new analyses (**Key revision 2-3, above**) demonstrating how it outperforms reference transfer
 433 algorithms for annotating cell types while also outperforming gene set based approaches for
 434 annotating T-cell activation. See response to **R2.2** for a further discussion about how the
 435 approach is distinct from previous methods.

436
437 To address the reviewer's point about major new insights, we have added a new analysis using
438 TCAT to identify novel factors that predict tumor responsiveness to immune checkpoint inhibitor
439 (ICI) therapies (**Key revision 1, above**). Briefly, TCAT enabled us to quickly annotate multiple
440 tumor datasets with a consistent set of gene expression features, demonstrating a reproducible
441 finding of increased activation and proliferation gene sets in CD4 and CD8 T-cells across
442 multiple tumor types in non-responders. It also showed consistently increased CD4-naive cGEP
443 usage in responders. We believe this represents a significant new insight that was enabled by
444 TCAT and illustrates the utility of the approach. Accordingly, we have also added the following
445 text to the discussion:

446
447 *TCAT enabled us to identify novel properties of tumor-infiltrating CD4 and CD8 T-cells*
448 *that predict response to immune checkpoint inhibitor (ICI) therapy (**Figure 6**). Across*
449 *three cancer types, non-responsive tumors were enriched for exhaustion, activation, and*
450 *cell cycle cGEPs. This is potentially surprising as, a priori, one might assume that more*
451 *immunologically active tumors would respond better to ICI therapy. Instead, these*
452 *findings suggest that ICI non-responsive tumors contain a higher fraction of activated,*
453 *proliferating, and exhausted CD4 T-cells reflecting an enhanced ability to suppress*
454 *effective cytotoxic responses. By contrast ICI responsive samples were enriched for the*
455 *CD4-Naive cGEP, consistent with previous associations of ICI response with stem-like*
456 *T-cells marked by TCF7². The presence of these cells may suggest the potential of ICI to*
457 *reinvigorate infiltrating T-cells, or alternatively, may suggest the inability of the cancer*
458 *cells to fully exhaust and terminally differentiate all T-cells in the tumor environment.*
459 *These results illustrate how TCAT can aid in characterizing clinically impactful cell states*
460 *in vivo and can help with predicting therapy response.*

461
462 **R2.2 It's a bit unclear why the authors chose to only select two GEPs to turn into a binary**
463 **classification variable; for an annotator, it would be expected and helpful to have binary**
464 **classifications for all programs, otherwise it's unclear how this would be an improvement**
465 **over simply using a calculated gene score with a set of marker genes for a given**
466 **phenotype/differentiation state. The multinomial labeling tool does this somewhat, but**
467 **despite the high number of GEPs here, the level of granularity of T cell phenotypes**
468 **identified by the multinomial labeling tool seems quite limited. As those phenotypes are**
469 **all well-characterized by a few marker genes, it's unclear how this tool is a significant**
470 **improvement over existing methods (e.g. Azimuth).**

471
472 We thank the reviewer for their comment and believe it represents a significant point of
473 confusion about this method which we have sought to improve in the revision. The fundamental
474 advantage of TCAT is the ability to simultaneously annotate a dataset with multiple GEPs that
475 are specifically relevant to T-cells and have been defined in a consistent and reproducible
476 fashion. While it can output some discrete annotations for cases where we had orthogonal data
477 to train a discrete classifier (see below), it primarily outputs continuous GEP scores. While this
478 is similar to calculating scores from a literature-defined gene set, TCAT possesses several
479 advantages over this common approach. First, GEPs used in TCAT provide different weights for

480 each gene, rather than arbitrarily weighting each gene uniformly as in gene set scoring. This
481 can increase accuracy in scoring GEP activity and can increase the ability to resolve related
482 GEPs that may share genes in common but with different weights. In addition, TCAT scores
483 each GEP simultaneously, which allows the relative usage of multiple GEPs to be compared for
484 each cell. This mitigates the risk of confounding multiple related signals, which may happen with
485 gene set scoring. For example, exhausted T-cells may score highly for both an exhaustion and
486 an activation gene set because of overlapping genes between these pathways. TCAT mitigates
487 this by simultaneously modeling the exhaustion and activation GEPs so they can be directly
488 compared. Lastly, there are often multiple gene sets available for a given process which may be
489 derived from distinct experimental contexts, and it can be hard to know which is the most
490 appropriate to use. TCAT provides a practical advantage of expediently scoring a
491 comprehensive set of GEPs that are derived in a consistent manner from *in vivo* data with
492 demonstrated generalizability across multiple scRNA-Seq datasets. We add the new text below
493 to discuss these advantages of TCAT in the introduction:

494
495 *It is possible to compute continuous GEP scores for each cell in a dataset, similar to the*
496 *common practice of scoring cells based on a predetermined gene sets¹⁵. However,*
497 *unlike gene sets, component-based models account for variable gene weights,*
498 *simultaneously model multiple GEPs which reduces confounding of related signals and*
499 *enables comparison of relative GEP activities, and use GEPs derived from a consistent*
500 *in vivo context rather than from potentially diverse experimental systems like microarrays*
501 *or RNA-Seq.*

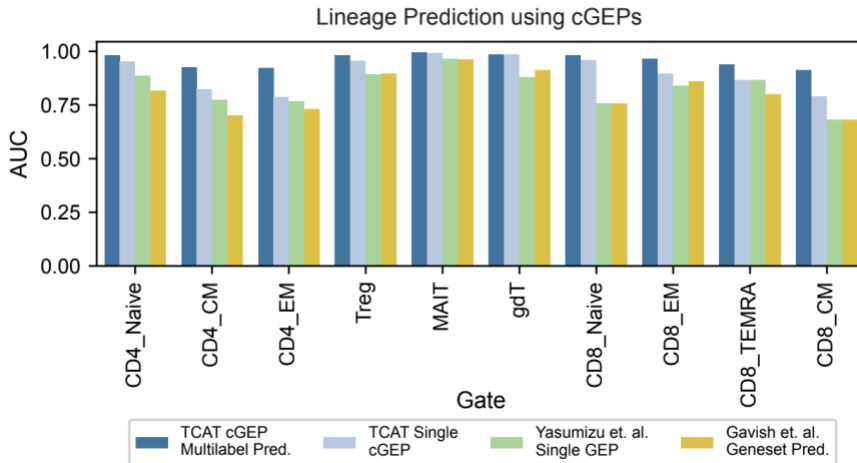
502
503 In addition to these theoretical benefits of TCAT over gene set scoring, we also now perform
504 direct empirical benchmarking (**Key revision 3, above**) showing that TCAT outperforms nine
505 out of nine literature-derived activation gene sets for predicting T-cell activation in an *in vivo*
506 dataset and outperforms seven out of the nine for predicting activation in the AIM-Seq dataset
507 (**New Figure S6F-G, Extended Data 8B-C, above**). We added the following text to the
508 manuscript to highlight this:

509
510 *We benchmarked the ASA score against T-cell activation gene sets for predicting T-cell*
511 *activation as defined above in the AIM-Seq and Flu-Vaccine datasets, two validation*
512 *datasets for this approach. ASA obtained higher accuracy than 9/9 literature-derived T-*
513 *cell activation gene sets in the Flu-Vaccine dataset and 7/9 gene sets in the AIM-Seq*
514 *dataset (**Figure S6F-G**). Thus, using the TCAT-derived ASA score can yield more*
515 *accurate inferences of antigen activation than gene sets both in vivo and ex vivo.*

516
517 We also previously showed that TCAT out-performed a set of scRNA-seq derived T-cell gene
518 sets¹⁶ for annotation of cellular lineage (**Figure S3C, reproduced below for convenience**).

519

C



520
521
522
523
524
525

Figure S3C. Benchmarking CellAnnoTator on simulated and real datasets. Area under the curve (AUC) for prediction of manually gated subset based on a single most associated subset (dark blue), TCAT multilabel prediction (light blue), analogous predictions using the single most associated NMF component published in (Yasumizu et al., 2024), or using gene sets from NMF components in Gavish et al., 202314.

526
527
528
529

These findings show empirically that TCAT provides a significant advantage over gene set annotation for scoring of expression GEPs.

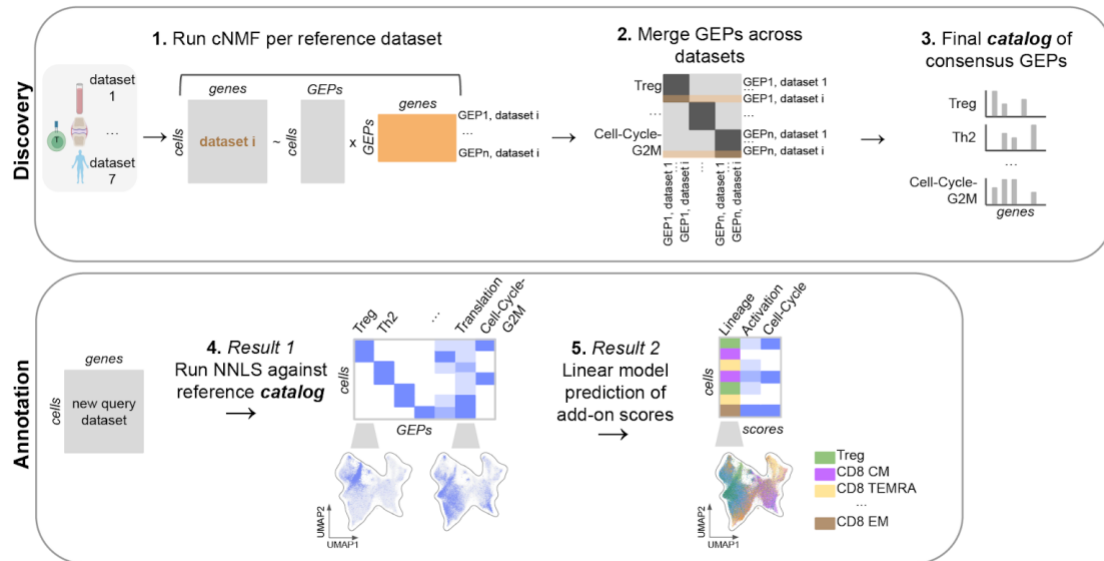
530
531
532
533
534
535
536
537
538
539

We believe that much of the confusion about the method derives from the fact that it also provides the user with several discrete phenotype annotations (subset, antigen-specific activation, and cell cycle) in addition to the continuous GEP scores. We do so because in certain cases, it is useful to have a discrete annotation in addition to a continuous score. For predicting subset and antigen-specific activation (ASA), we could use CITE-Seq data as an orthogonal ground truth label to train a discrete classifier for these features. For most other features, we lack such a ground truth and thus were unable to quantify specificity and sensitivity for specific thresholds. We therefore have opted to not discretize other features with the exception of cell cycle, a metric often discretized in scRNA-seq processing pipelines. We clarified this point in the revised manuscript by adding the following text and modifying our schematic Figure 1A to show these two distinct results (**Revised Figure 1A**).

541
542
543
544
545
546
547

In addition, we utilize these cGEPs to predict and return annotations for additional protein-informed per-cell labels, including lineage, activation score, and cell cycle score.

*We then also utilize the per-cell GEP usages to predict additional per-cell features, including lineage, activation, and cell cycle (**Figure 1A - bottom**).*



548
549
550
551

Revised Figure 1A. Overview of star-CellAnnoTator (starCAT). Schematic of the starCAT pipeline.

552 Lastly, we demonstrated that TCAT outperforms three commonly used reference label transfer
553 softwares (Azimuth, ProjectTILs, and Symphony) for subset prediction (**Key revision 2, above**).
554 While this task is only one small part of what TCAT does, it was useful to demonstrate that
555 TCAT outperforms the state of the art methods for this task.

556

557 **R2.3 Lines 129-147; Fig S1:**

558 **Removing only 4 GEPs out of 16 for performance testing purposes seems low; what if a**
559 **dataset had only CD4 or CD8 T cells for example? How would *CAT perform on such a**
560 **dataset? Figure S1 is hard to interpret quantitatively; extracted accuracy values would be**
561 **helpful.**

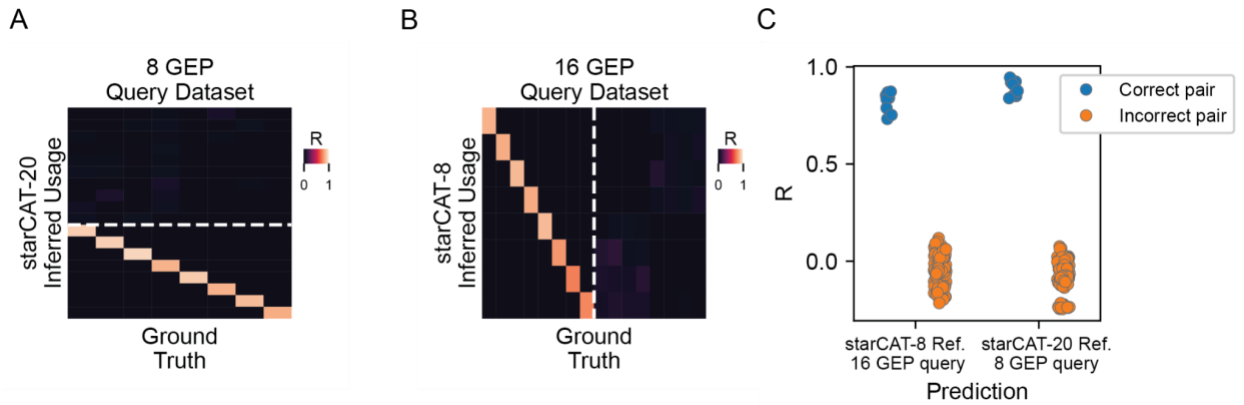
562

563 We thank the reviewer for these comments. To address the first point, we have now tested the
564 accuracy of using a starCAT reference with 20 GEPs in the reference, of which only 8 are in the
565 query (analogous to using a reference trained on both CD4 and CD8 T-cells to predict solely
566 CD8 T-cells). We also tested the inverse where a reference contains only 8 GEPs but the query
567 contains 16. We found generally comparable results to what we presented previously in **Figure**
568 **S1 (New Extended Data Figure 1** reproduced below for convenience). We now include the
569 following additional text in the manuscript describing this finding:

570

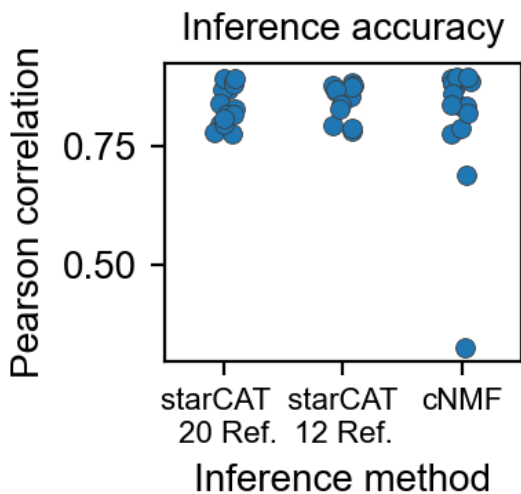
571 *We tested the robustness of starCAT when a lower fraction of GEPs overlapped*
572 *between the reference and query datasets. We observed similar prediction accuracies*
573 *when predicting a simulated query dataset with 8 GEPs using a reference with 20 GEPs,*
574 *and predicting a query with 16 GEPs using a reference with only 8 GEPs (**Extended***
575 ***Data Figure 1**). This shows starCAT can be robust to references and queries that*
576 *contain a substantially smaller fraction of overlapping GEPs.*

577



578
 579 **New Extended Data Figure 1. Additional starCAT simulation analyses** (A) Pearson
 580 correlation of ground truth simulated usages of each GEP (columns) vs inferred usages
 581 (rows) for starCAT using a simulated 20 GEP reference to predict an 8 GEP query
 582 dataset, where the 8 GEPs overlap with the reference. (B) Same as A but using an 8
 583 GEP reference and a 16 GEP query dataset, containing 8 overlapping GEPs. (C)
 584 Extracted pearson correlation values from (A) and (B) indicating whether an inferred
 585 GEP matched the ground truth GEP based on the color.
 586

587 Regarding the second comment, we agree that the correlation heatmap figures in **Figure S1**
 588 could be confusing to the reader. Therefore, to help with interpretation, we have added a
 589 summary of the correlation between the inferred GEP usages and the corresponding simulated
 590 ground truth usages (**Revised Figure S1C**). The stripplot shows that the starCAT reference with
 591 extra or missing cGEPs had similarly good correlations for their overlapping GEPs in the query
 592 dataset, while cNMF had less accurate GEP inferences for 2 out of the 16 GEPs.
 593



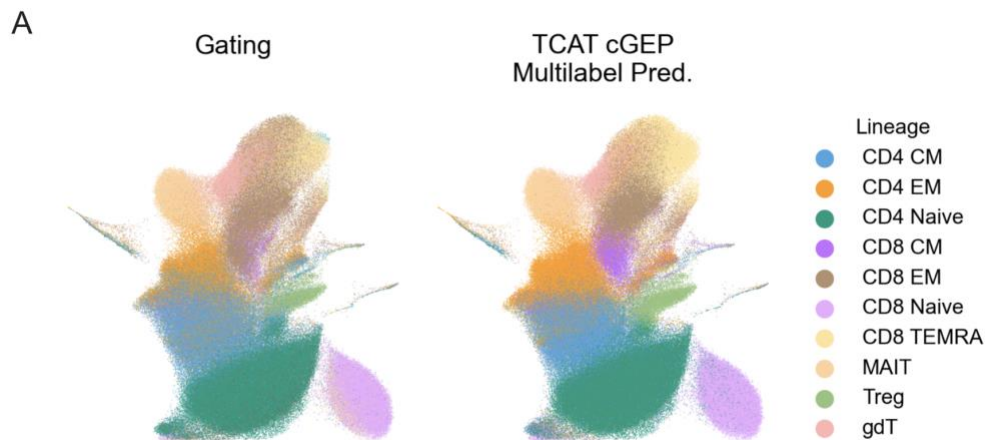
594
 595 **Revised Figure S1C. Characterizing starCAT.** Pearson correlation between inferred
 596 gene expression programs and the corresponding ground truth usages, extracted from
 597 (B).
 598

599 **R2.4 Fig 2A:**
600 **The UMAP in Fig 2A does not appear to effectively separate or cluster the identified cell**
601 **types; since these labels are being used as ground truth this may be concerning and**
602 **should be addressed. A corresponding UMAP showing the multinomial label predictions**
603 **would be helpful here.**
604

605 We thank the reviewer for this comment. We believe this comment suggests a useful way for us
606 to help illustrate an important aspect of our manuscript which we have hoped to clarify. For this
607 annotation (previously called “ground truth”), we used CITE-seq protein markers to separate
608 canonical subsets of T-cells, an approach analogous to traditional isolation of T-cell lineages by
609 immunologists using flow cytometry sorting on protein markers (**See R1.10**). However, single
610 cell transcriptomes reflect additional gene expression programs besides cell lineage, such as
611 functional (ex. cell cycle, activation) or artifact gene expression programs. Thus, cells may not
612 cleanly separate by lineage on a UMAP.
613

614 In addition to adding a new UMAP showing the multinomial label predictions (**Revised Figure**
615 **2A** reproduced below), we have added the following text to the corresponding results section to
616 clarify this:

617
618 *These protein-based annotations are useful for benchmarking as they are measured*
619 *independently from the transcriptome-based features used for prediction and connect*
620 *directly to the wealth of literature about canonical flow-cytometry defined T-cell subsets.*
621 *Protein-derived subsets largely separated on a gene-expression derived UMAP (**Figure***
622 ***2A - left**). However, there was also significant overlap of memory populations, which*
623 *may reflect the impact of non-subset-specific signals influencing the transcriptome,*
624 *which we further explored in the next section.*
625



626
627 **Revised Figure 2A.** Left - UMAP of the Flu-Vaccine dataset colored by the manually
628 gating shown in **Figure S3A**. Right - UMAP colored by TCAT multinomial label
629 prediction.
630
631

632 **R2.5 Fig 2C, Fig S3C:**
633 **The figure description in S3C does not match the legend in the figure; is multilabel**
634 **prediction the light blue color or the dark blue color? If the legend is correct, then for 2C**
635 **and S3C, why does the multilabel prediction perform better than the single cGEP version**
636 **(the explanation for how this was calculated is also somewhat unclear)?**
637

638 We thank the reviewer for this comment. For the previous **Figure 2C and S3B** (which are
639 **Extended Data 5A-B** in the revision), and **Figure S3C** (which remains as is), the dark blue line
640 indicates the multilabel prediction and the light blue line indicates prediction using the single
641 cGEP that was most correlated with the subset. For multilabel prediction, we fit a multinomial
642 logistic regression that predicts the probability a cell belongs to each subset as a function of its
643 usage of all cGEPs. To determine the AUC for a given subset prediction using this model, we
644 obtain a receiver operator characteristic (ROC) by varying the threshold on the probability
645 assigned to that subset between 0 and 1. By contrast, for the single cGEP prediction, we
646 identify the single cGEP that has the best AUC for classifying each subset using a ROC based
647 on varying the thresholds on the usage of that cGEP between 0 and 1. The multilabel prediction
648 can outperform the single cGEP prediction by leveraging the ability to combine multiple cGEPs
649 rather than just using a single cGEP. This is especially helpful for complex subsets such as CD4
650 EM which benefit from considering multiple effector states (e.g. Th17, Th2, cytotoxic, etc.). We
651 have added additional text to help clarify this point for the reader:

652
653 *Since subsets contain heterogeneity not captured in univariate analysis such as multiple*
654 *polarized populations within CD4 effectors, we performed multivariate analysis using all*
655 *cGEPs for simultaneous multi-label prediction (**Methods, Figure 2A - right**). We trained*
656 *the classifier on the COMBAT dataset and evaluated its performance on the Flu-Vaccine*
657 *dataset. Across lineages, the classifier outperformed commonly used categorical*
658 *reference mapping methods, including Azimuth⁵, Symphony⁷, and ProjecTILs⁶, as well*
659 *as transcriptome clustering (Figure 2B-C). For example, TCAT greatly outperformed*
660 *other methods for classification of CD4 effector memory cells in the Flu-Vaccine dataset*
661 *(accuracy 0.80 compared to 0.57, 0.60, and 0.52 for Azimuth, Symphony, and*
662 *ProjecTILs respectively). TCAT also predicted TEMRA lineage (accuracy 0.83), which*
663 *was not predicted by any of these alternative methods. The TCAT classifier was also*
664 *more accurate than de novo clustering across all nine clustering resolutions tested, with*
665 *average balanced accuracy differences ranging from 0.10 to 0.032 (**Figure 2B-C,***
666 ***Extended Data Figure 5A**). By incorporating information from multiple cGEPs, the multi-*
667 *label predictor yielded increased prediction accuracy than the single most associated*
668 *cGEP for eight out of ten lineages.*

669
670 We have also updated the description of **Extended Data 5B** to correctly match the figure
671 legends.

672
673 *Receiver operator curves (ROCs) for prediction of manually gated subset based on*
674 *TCAT multilabel prediction (dark blue), a single most associated TCAT cGEP (light*
675 *blue), analogous predictions using the single most associated NMF component*

676 *published in Yasumizu et al., 2024¹¹, or using gene sets from NMF components in*
677 *Gavish et al., 2023¹⁴. Individual points show accuracies of discrete predictions based on*
678 *cGEP multilabel regression, or clustering with the Leiden resolution specified in the*
679 *legend.*

680

681 **R2.6 The text in some figures is quite small and almost unreadable (e.g. Fig 1A, S2B)**

682

683 We thank the reviewer for their comment. We have improved the legibility of text in figures,
684 including in Fig 1A and S2B.

685

686 **R2.7 Color schemes in figures are completely mixed throughout; it would be clearer if the**
687 **same color schemes were used for representing the same data types**

688

689 We thank the reviewer for their comment. We have updated the figure color schemes to be
690 consistent throughout the manuscript. For instance, we used a single color scheme for plotting
691 cGEP usages throughout the manuscript (e.g. **Figure 3B-D, Figure S4B-C, Figure 5D, Figure**
692 **6A**) and for coloring cell lineages (**Figure 1A, Figure 2A, Figure 4D**).

693

694 **R2.8 Typo in line 984**

695

696 We thank the reviewer for their comment. We have edited this text.

697 **Reviewer #4 comments:**

698

699 **R4.1**

700 **A. Summary of the key result**

701

702 **This paper proposes a new way to analyse single T cell expression as being the sum of a**
703 **variety of independent gene expression profiles, i.e. instead of an individual T cell being**
704 **described as exhausted or expanding it can be thought of as occupying a state**
705 **somewhere between both (and indeed up to 46 different GEPs).**

706 **To do this the authors applied the recently published cNMF algorithm, but first had to**
707 **design a new batch normalisation strategy to allow the data from 7 previously published**
708 **studies to be amalgamated in such a way as to be compatible with cNMF.**

709 **They find several GEPs to be represented in all the T cell datasets considered, though**
710 **these include trivial GEPs such as cell cell cycle and cytoskeletal associated GEPs that**
711 **should be in all cells which flatters the comparison. GEPs were annotated by manual**
712 **inspection of the top weighted genes, and poorly supported GEPs removed.**

713 **TCAT GEPs were benchmarked against the previous tool NMFproj and clustering**
714 **approaches using an independent CITEseq experiment. TCAT was then applied to the**
715 **COMBAT covid database and shown to be able to deconvolute subsets of proliferating T**
716 **cells which had previously been grouped together, and showed that the previous dataset**
717 **contained many cells of poor quality that had not been filtered out. TCAT also**
718 **discriminates between CD8+ T cells polarisation states.**

719 The authors introduce their AIMseq technique, looking at TCR dependent T cell
720 activation and cGEPs that correlate with antigen reactivity. Using the cGEP enriched in
721 activated T cells they then generate an ASA score by which the antigen reactivity of a cell
722 can be assessed, and use this to analyse T cells derived from various cancer samples (in
723 part by determining a ASA threshold at which to specify a cell as antigen reactive). The
724 authors show that tumor mutation burden and the proportion of antigen-specific T cells
725 correlate as would be expected, and then further analyse the cGEP phenotype of
726 bystander T cells.

727

728 Thank you for your thorough and thoughtful summary of our manuscript.

729

730 **R4.2**

731 **B. Originality and significance: if not novel, please include reference**

732 **Originality: NMF approaches have recently been applied to circulating CD4+ T cells**

733 **(Single-cell transcriptome landscape of circulating CD4+ T cell populations in**

734 **autoimmune diseases - ScienceDirect); this new study seeks to define gene expression**

735 **profiles more broadly and determines a larger number of GEPs in an effort to make a**

736 **shared reference representation of cell states, in particular T cell states.**

737 **The significance is hard to determine: the authors are correct that having a well**

738 **characterised reference panel of GEPs that (in theory) captures all possible T cell GEPs**

739 **would be of value in mapping studies (especially smaller studies) onto a common**

740 **representation of cell states is of value to the field of single cell sequencing. The**

741 **downside is that the cNMF approach is conceptually harder to understand than simpler**

742 **approaches using reference maps such as Azimuth and ProjecTILs, which may limit**

743 **adoption. However the authors have made an admirable effort to make their tool available**

744 **on github, present appropriate reference implementations, and are working on a web**

745 **based tool that can be deployed on remote compute resources, which should facilitate**

746 **implementation.**

747 **As the costs of sequencing continue to fall, and the size and availability of single cell**

748 **datasets increases, this tool will likely be of increasing value (in particular to meta**

749 **analyses)**

750 **While the authors present a huge amount of data, this paper would benefit from less data**

751 **and a more compelling use case – for example the Sakaguchi group paper puts a clear**

752 **focus on autoimmunity, I'd encourage the authors to put more emphasis on the last third**

753 **of their paper looking at cancer to emphasise the novelty or additional value (though I**

754 **appreciate that *CAT has uses beyond TCAT).**

755

756 Thank you for this comment. Based on this, we have benchmarked our method against Azimuth

757 and ProjecTILs to illustrate the significantly improved performance (**Key revision 2, New**

758 **Figure 2B, S3B, above**). We have also further developed our website which can be seen at the

759 following link <https://immunogenomics.io/starcat>. It now hosts additional reference catalogs for

760 glioma-derived myeloid cells and bone-marrow derived CD34+ hematopoietic stem cells (**Key**

761 **revision 4, above**).

762

763 Most significantly, we have greatly expanded our analysis of cancer to show how TCAT reveals
764 features of CD4+ tumor-infiltrating lymphocytes that predict response to immune checkpoint
765 inhibitors (ICIs) in melanomas, colorectal cancer, and non-melanoma skin cancer (**Key revision**
766 **1 and New Figure 6, above**). Surprisingly, tumors with more proliferating and activated CD4 T-
767 cells were more likely to progress despite ICI. By contrast, tumors with more cells expressing
768 the CD4-naive cGEP (and consequently more *TCF7*) were associated with improved ICI
769 responsiveness. Thank you for raising these points. We believe the updated analysis is more
770 compelling and greatly increases the quality of the paper.

771

772 **R4.3**

773 **C. Data & methodology: validity of approach, quality of data, quality of presentation**

774 **As a general point the paper seems undecided as to whether it's a technical paper for a**
775 **specialist audience with a high background level of understanding, or a newly**
776 **generalisable method that might be of interest to less specialist readers who have single**
777 **cell sequencing data.**

778

779 We thank the reviewer for this comment. We agree that certain aspects of the paper are quite
780 technical depending on the background of the reader. Specifically, we envision three target
781 audiences: (1) T-cell biologists who might wish to use TCAT on their data (2) other biologists
782 who may use starCAT for other cell types of interest, and (3) single-cell methods developers
783 who may be interested in the general approach. Currently, the paper is most geared toward (1)
784 and (3) but may be harder for non-T-cell biologists and non-computational researchers to
785 understand. Although it is not possible to equally address all audiences, we have made edits to
786 better sign-post the potential utility of the method to non-T-cell biologists. Specifically, we have
787 added the following concluding sentence to the abstract:

788

789 *Our software package starCAT similarly enables reproducible annotation of other cell*
790 *types and tissues.*

791

792 We also note that the beginning of the discussion summarizes the general method of starCAT
793 rather than the specific T-cell implementation:

794

795 *Here, we introduced starCellAnnoTator (abbreviated starCAT) for annotating scRNA-*
796 *Seq data with predefined GEPs. starCAT exploits the observation that functionally*
797 *informative GEPs learned by cNMF are reproducible across datasets and contexts and*
798 *can thus aid interpretation of new datasets.*

799

800 In addition, we now specifically indicate some of the non T-cell references that are hosted on
801 our website:

802

803 *We demonstrated starCAT with T-cells but it is equally applicable to other cell types and*
804 *tissues. We make the starCAT software publicly available and have created a repository*
805 *to host cGEP catalogs, including a human glioma myeloid⁹ and human bone marrow*
806 *hematopoiesis reference¹⁰, enabling easy application to new datasets. Users studying*

807 other tissues and cell-types can contribute their own catalogs to the repository, akin to
 808 the molecular signatures database (MSigDB)^{17,18}, but hosting GEPs for annotation of
 809 scRNA-Seq data rather than gene-sets for enrichment testing. We hope starCAT will aid
 810 in comprehensive identification of GEPs underlying cell behavior across tissues and
 811 diseases.

812

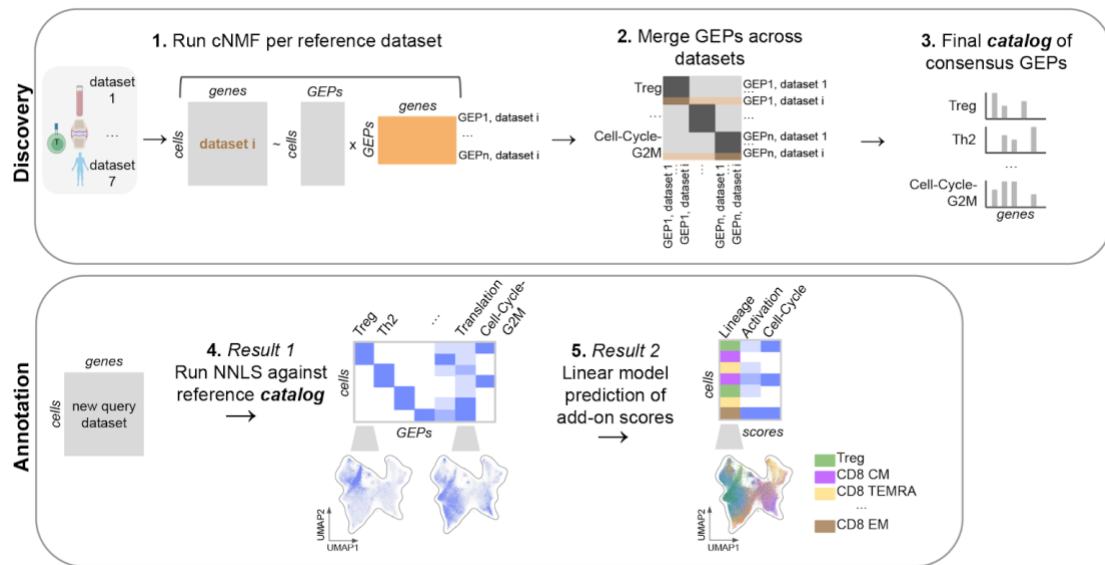
813 **R4.4**

814 Given the journal is Nature Methods the figure legends are excessively terse which can
 815 make the data hard to understand: for example Fig 1A providing the overview of the
 816 whole approach is described in a single line of text. The figure itself could be expanded
 817 upon (or at least made larger) to help the reader grasp the core concept of this paper,
 818 while Fig1B is difficult to read and could be better linked to Fig1A_step2.

819 Compared to the previous NMF paper this paper has a more detailed methods section
 820 and a similar or more detailed supplementary data. The main figures are less crowded
 821 than the supplementals and have been prepared with care commensurate for a top tier
 822 journal.

823

824 Thank you for this comment. We have expanded the size of **Figure 1A** and made the included
 825 text more legible (**Figure 1A**, reproduced below).



826

827 **Revised Figure 1. Overview of star-CellAnnoTator (starCAT).** (A) Schematic of the
 828 starCAT pipeline. The discovery phase (top) identifies consensus GEPs (cGEPs) by
 829 running consensus NMF on multiple datasets and combining the results. The annotation
 830 phase (bottom) uses the cGEPs to annotate new query datasets and to compute
 831 additional scores and discrete classifiers.

832

833 **R4.5**

834 **D. Appropriate use of statistics and treatment of uncertainties**

835 I do not have the requisite background in statistics to comment on the mathematical
 836 suitability of the new normalisation techniques used to combine the 7 datasets. Some

837 **figures lack statistics, and/or the statistical tools to compare these types of highly**
838 **complex single cell datasets are not yet well established.**

839
840 **The authors have made the code underlying every figure available to allow readers to**
841 **repeat their analyses which is admirable and underscores their high level of technical**
842 **ability. If one were to criticise this might benefit from a little more curation - for example**
843 **Analysis/TCAT/Polarization_MarkerHeatmaps.ipynb contains a number of KeyErrors**
844 **which suggest that this might not be the final version used.**

845
846 Thank you for this comment. We have cleaned up extraneous notebook cells that were part of
847 our initial exploratory analysis but not part of the final paper. This removed additional KeyErrors
848 from the notebooks. We have also updated the documentation clarifying in which notebook each
849 figure was produced.

850
851 We have also added additional statistical analysis to figures where they were missing. For
852 example, **Figure 4F** which shows association of activation-associated cGEPs with lineages has
853 been re-made using pseudobulked analysis to enable more rigorous statistical hypothesis
854 testing. See the response to **R4.14** below.

855
856 **R4.6**

857 **E. Conclusions: robustness, validity, reliability**

858 **The authors demonstrate a number of use cases for *CAT using TCAT to analyse diverse**
859 **T cell repertoires to show that they can more accurately deconvolute T cell subsets.**

860 **F. Suggested improvements: experiments, data for possible revision**

861 **In general review figures and legends for clarity.**

862 **Fig S5B not mentioned in text**

863
864 Thank you for this comment. We have added the text reference for this figure.

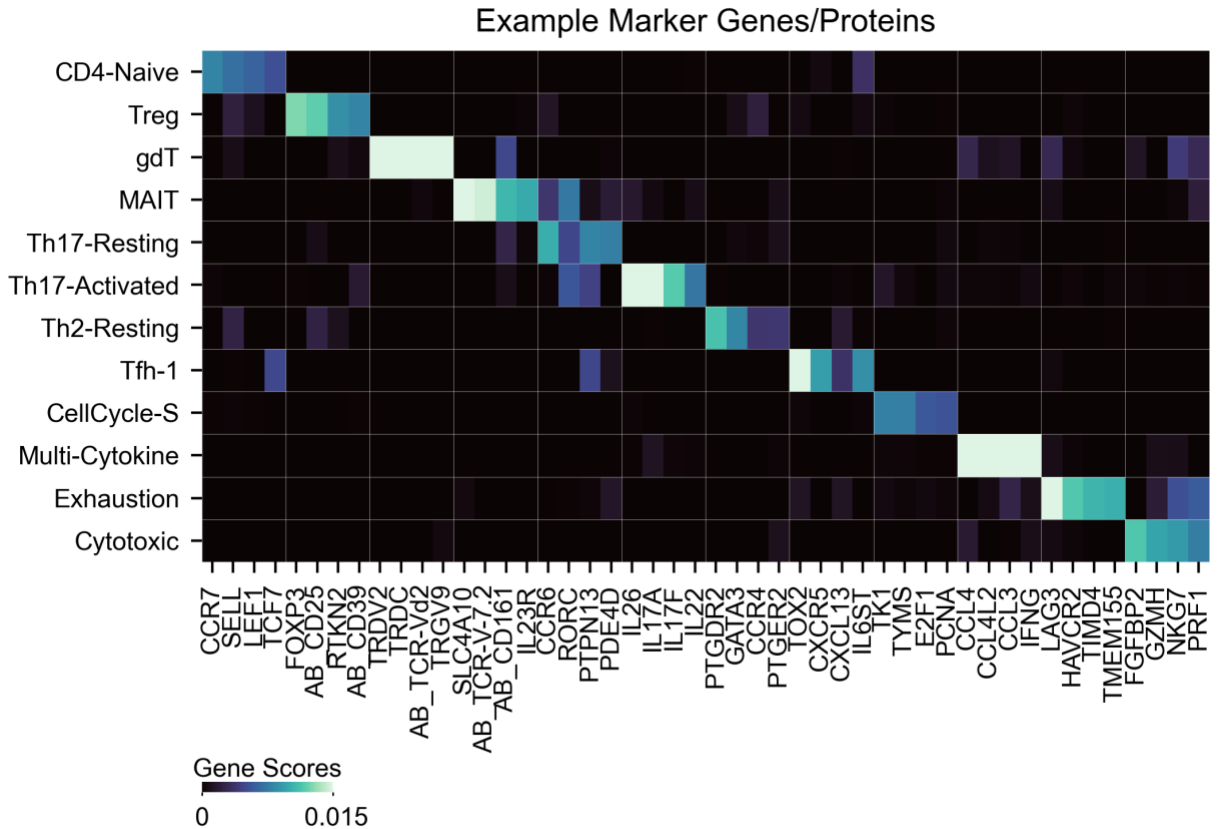
865
866 **R4.7**

867 **Section 1**

868 **I would also suggest tying the results of single cell sequencing/TCAT more closely to**
869 **what the scientific audience know about T cells – for example in Fig1D how are known**
870 **markers like ENTPD1 and CXCL13 that happen not to be among the top genes per GEP**
871 **distributed? Some information here – such as that HLA-DRA and HLA-DRB are top genes**
872 **in the HLA cGEP – doesn't add much information.**

873
874 Thank you for this comment. One of the original goals for this figure was to emphasize the
875 specificity and ease of interpretability of the discovered cGEPs using some well-known example
876 programs with clear defining markers. However, we agree that including non T-cell specific
877 cGEPs such as the HLA adds less information. We have therefore replaced the HLA, ISG, IEG,
878 and cytoskeleton cGEPs with Exhaustion (which includes *ENTPD1*) and Tfh-1 (which includes
879 *CXCL13*) (**Revised Figure 1D**, below).

880



881
882 **Revised Figure 1D.** Marker genes for selected example cGEPs in cNMF gene score
883 units.
884

885 **R4.8**

886 **I was surprised to see CXCL13 as a top weighted gene in T follicular helped cells, as this**
887 **has been reported by many groups (Platten/Green, Wu, Rosenberg) as being a key**
888 **marker for tumor reactive T cells**
889

890 Thank you for this comment. We agree that *CXCL13* has been identified as an important marker
891 of tumor-reactivity. It has also been shown to be a key marker of T follicular helper cells^{19–21} and
892 T peripheral helper cells^{22,23}, which is consistent with the role of these cells in providing B-cell
893 help. We suspect that many chronically antigen stimulated cells within tumors and rheumatoid
894 arthritis synovia become Tph and Tfh. Thus this is not contradictory with *CXCL13* being
895 upregulated in tumor reactive T-cells that have become Tph and Tfh cells. We also have
896 identified a *CXCL13*-expressing activation cGEP, labeled OX40/EBI3, which we believe may be
897 associated with tumor-reactivity (**See R4.16**). We added the following text to the manuscript
898 noting the literature relating *CXCL13* to tumor-reactivity:
899

900 *The OX40/EBI3 cGEP was identified specifically in tissues and was most common in*
901 *tumor-infiltrating T-cells (Figure 1B, Extended Data Figure 7D). It includes CXCL13, a*
902 *marker of tumor-reactivity²⁴, as well as several activation-induced markers used to*
903 *define AIM-positivity, including TNFRSF4, encoding OX40, and IL2RA, encoding CD25.*

904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935

R4.9

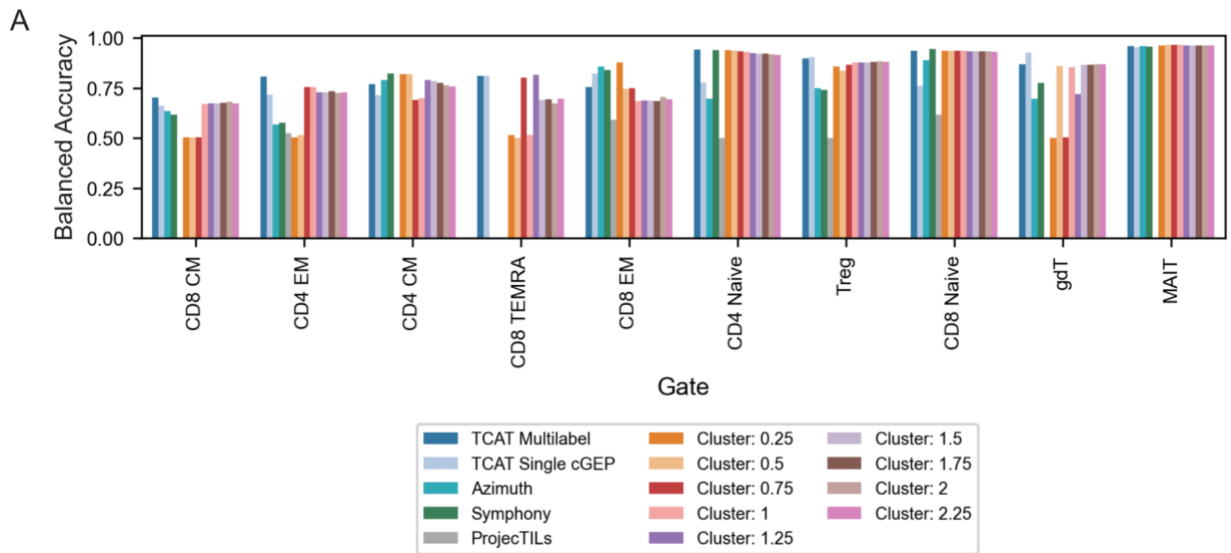
Section 3

Fig 2B – Accuracy does not reflect false positive assessments, a metric like G-Mean might better show the improved performance of the TCAT multilabel approach vs the single cGEP especially for CD8 EM. Is this what the authors mean in Fig2C with “balanced accuracy”? Possibly this is stated in the methods, but would be easier to state this in the legend/results.

We thank the reviewer for their comment and we agree about the importance of choosing a metric that balances both sensitivity and specificity. Throughout this portion of the text, we used a balanced accuracy metric which works well when the data is imbalanced, which is the case for rare cell populations. In the binary case (as in the previous **Figure 2B**, now **Extended Data Figure 5A**), balanced accuracy is calculated as the arithmetic mean of the sensitivity and specificity for a prediction. It thus incorporates both false positive and false negative assessments. Balanced accuracy also generalizes to the multiple class prediction problem, in which case it is defined as the mean sensitivity across classes (used in new **Figure 2B** for simplicity, displaying a single accuracy value per method rather than one per lineage). This approach in effect reflects false positive rates, as a high false positive rate of one class will result in a low true positive rate (sensitivity) of others. While geometric mean would likely also work well, we chose arithmetic mean as we thought it may be more familiar for some readers. To help clarify why we used balanced accuracy, we have added the following text to the manuscript, which is in reference to the multi-class case:

To measure performance of a prediction, we quantified its balanced accuracy, calculated as the mean sensitivity across classes, an approach which gives equal weighting to each class and is appropriate for cases of class imbalance, such as in cases of rare cell populations.

We also added the following text to the legend of **Extended Data Figure 5A** (shown below) to clarify the calculation of balanced accuracy in the binary case:



937

938

939

940

941

942

943

944

945

R4.10

946

Section 4

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

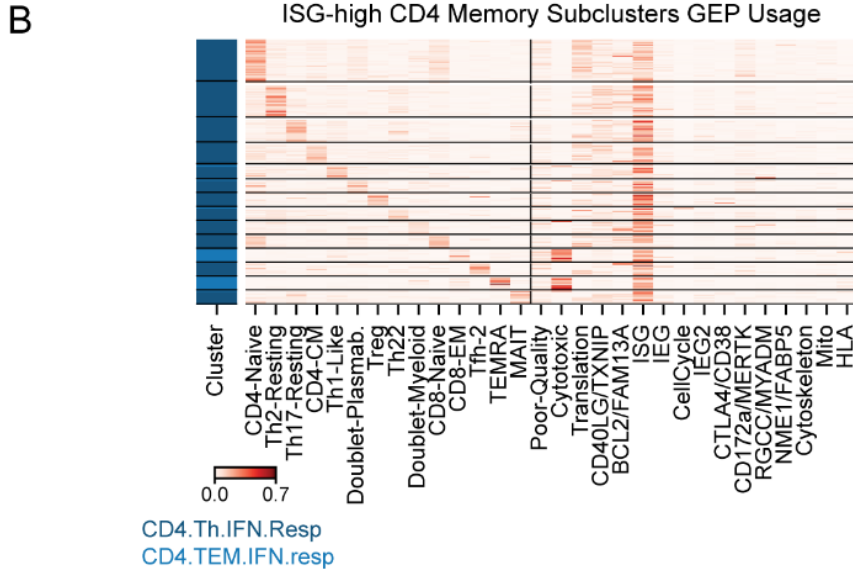
New Extended Data 5A. Balanced accuracy comparisons of TCAT's prediction of manually gated subsets, as compared to Azimuth, Symphony, ProjectTILs, and clustering with multiple Leiden resolution parameters. Balanced accuracy is calculated as the mean sensitivity and specificity of a prediction, thus reflecting both true positive and negative rates and appropriate in cases of class imbalance.

While TCAT does deconvolute the proliferating T cell cluster well, many studies perform regression of cell cycle genes so this is more a limitation of the COMBAT Covid-19 dataset than a feature of TCAT (though it nicely shows how previous data benefits from mapping onto a common reference).

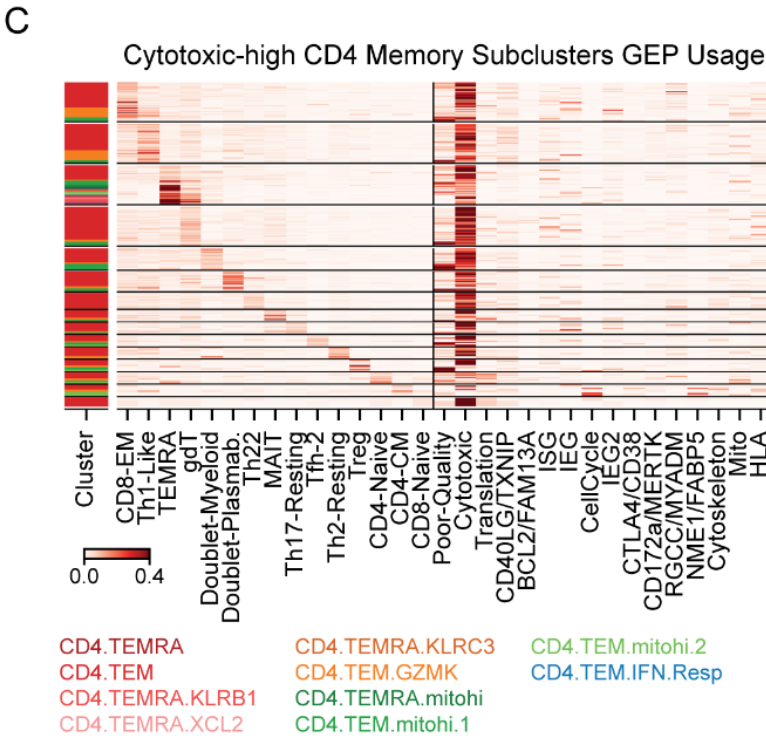
We thank the reviewer for highlighting this point. We agree that some researchers do address cell cycle as a confounder through regressing of cell cycle genes. However, cell cycle is just one example of a confounding non-lineage signal. We show similar effects of other signals such as interferon, cell quality, and cytotoxicity (**Figures S4B-C** reproduced below for convenience). In theory, these could all be regressed out, but we find our approach to be significantly more streamlined, and regressing out other signals is neither widespread nor validated. We like to keep the cell cycle example as a common, important, and easy to understand example of how lineage and non-lineage signals can get confounded. We also find it a biologically important example because scientists may frequently wish to know which lineages and conditions have the most proliferation. We agree that this could be characterized using alternative approaches (regress cell cycle to first cluster the cells, then use cell cycle gene set scoring to infer proliferating cells). However, our approach is significantly more streamlined and thus likely easier for many users. We have added the text below to clarify these points:

966
967
968
969
970
971
972
973

Next, we illustrate how TCAT reveals cellular heterogeneity not visible with clustering. As a first example, we consider cell cycle, a common signal that obscures other aspects of proliferating cells²⁵ in the COMBAT Covid-19 dataset. While it is possible to regress out cell cycle during preprocessing²⁶, this does not always work well, can remove correlated biological signals, and does not address analogous signals like cell quality, ISG, and cytotoxicity that can also obscure cell type. TCAT addresses these limitations by simultaneously modeling the GEPs underlying cell cycle, cell lineage, and more.



974



975

976 **Figure S4B-C. (B)** Usage of selected cGEPs (columns) in cells (rows) grouped by maximum
977 subset cGEP. Cells are drawn from subclusters with high usage of the ISG cGEP, indicated in
978 the colorbar. **(C)** Same as **(B)** but only showing cells from subclusters with high cytotoxicity
979 cGEP usage.

980

981 **R4.11**

982 **Looking at Fig S4 comparing the two covid databases COMBAT with UK-Covid, while**
983 **TCAT can deconvolute proliferating T cells into constituent subsets to show the**
984 **important role of the T peripheral helper cGEP in covid patients vs healthy, there is little**
985 **explanation of the large differences between these two covid datasets. Could the authors**
986 **comment on this – I believe this is important given that a core argument is that datasets**
987 **can (should?) be mapped to a common reference panel of GEPs. Dataset quality seems**
988 **to be part of the answer given that the authors later comment on the high incidence of a**
989 **cGEP indicative of poor cell quality in the COMBAT study. Why is the proportion of**
990 **CD8trm cGEP cells in the proliferative state so different between studies? I would**
991 **speculate that some cGEP are dependent on genes with low expression, such that poor**
992 **sample handling (e.g. long delay between sample acquisition and single cell capture)**
993 **depletes some cGEP more than others? The authors used an overdispersion metric to**
994 **identify key genes for the cNMF analysis, but does this come at a price of robustness? I**
995 **would like to see some discussion of these differences**

996

997 Thank you for this comment. While we are not certain about the precise reason for the
998 differences in rates of proliferation (and activation, as shown in **Figure 5G** and mentioned in
999 **R4.15**) between the two Covid-19 datasets, it is an important topic which we now specifically
1000 address in the text. Following up on the reviewer’s suggestion, we observed a positive
1001 correlation between both proliferation and antigen-specific activation (ASA) rates with several
1002 signals that are proxies for sample quality, including usage of the Poor-Quality and
1003 mitochondrial cGEPs, and the detected number of UMIs per cell. ASA and proliferation cGEPs
1004 were positively correlated with PoorQuality and mitochondrial cGEP usage and negatively
1005 correlated with average library size (**New Extended Data Figure 9A**). This is consistent with the
1006 hypothesis that low quality samples have more activation. We hypothesize that differences in
1007 cell processing pipelines may result in lower quality samples with higher rates of cell death and
1008 non-specific activation and corresponding cell proliferation.

1009

1010 We have added a new paragraph and supplementary figure highlighting the differences in
1011 proliferation and activation between the Covid-19 datasets and their potential explanations. As
1012 all reference mapping algorithms can be influenced by batch artifacts and downstream effects of
1013 processing methods, we also added text indicating the importance of careful interpretation of
1014 cross-dataset comparisons, particularly regarding cGEPs which may be affected by processing
1015 pipelines, such as artifact and activation-associated GEPs:

1016

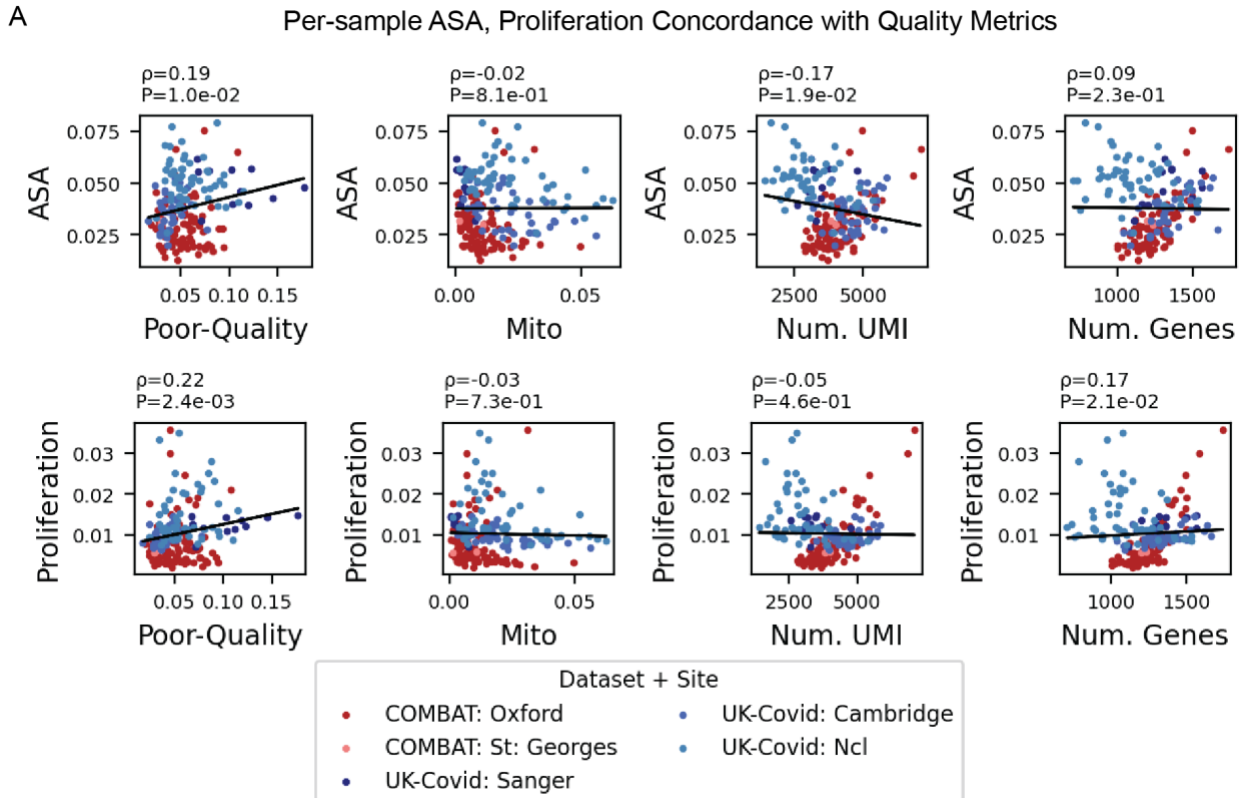
1017

1018 *We quantified higher rates of activation in both healthy and Covid-19 samples in UK-*

1019 *Covid than in the COMBAT dataset. We hypothesized that this could reflect non-specific*

1020
1021
1022
1023
1024
1025
1026

activation related to sample processing. Consistent with this, the two Covid-19 datasets greatly differed in sample quality, as marked by higher usage of the Poor-Quality and Mito cGEPs in the UK-Covid dataset. ASA and proliferation were both correlated with sample quality, as reflected by correlation with the Poor-Quality cGEP and library size (**Extended Data Figure 9**). Thus, it is important to carefully interpret cross-dataset differences in activation as these could reflect differences in sample quality.



1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041

New Extended Data Figure 9A. Per-sample scatterplots showing concordance between mean antigen-specific activation (ASA) score (top) or proliferation (bottom) and various quality metrics, across datasets and sites. Spearman correlation and P-values are shown.

The reviewer also highlighted a difference in proportions of proliferating CD8 TRMs between the two Covid-19 datasets. This is again an important point where we believe there may be a handful of factors at play. As both Covid-19 datasets are derived from blood, the number of CD8 TRMs, or cells with high usage of the CD8 TRM-Like cGEP, are relatively low. However, the presence of circulating TRM cells is somewhat surprising to see and may reflect memory cells migrating to the periphery. Different abundances of proliferating TRMs may reflect differences in clinical properties of the samples, treatments, or other dataset-specific features. Although the rates of proliferation of CD8 TRMs in Covid-19 samples is discordant between these two datasets, we did note a high degree of correlation in rates of proliferation across lineages

1042 between the datasets ($R=0.80$ in Covid-19 samples, $R=0.56$ in healthy). We have added the
1043 following text indicating the concordance of proliferation rates between the two datasets:

1044
1045 *Proliferation rates across lineages were largely concordant between both Covid-19*
1046 *datasets ($R = 0.80$ and $P= 0.00021$ in Covid-19 samples; $R = 0.56$ and $P = 0.025$ in*
1047 *healthy samples).*

1048
1049

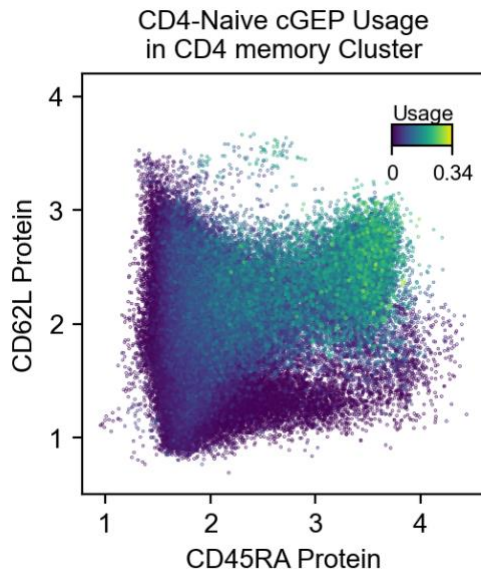
1050 **R4.12**

1051 **When the authors state “Cells with high usage of the CD4-Naive cGEP expressed CD4**
1052 **naive markers including CD45RA protein and SELL RNA, confirming that clustering**
1053 **misclassified them as memory T-cells (Extended Data 4)” some memory T cells also**
1054 **express CD45RA and SELL (e.g. TEMRA or Tcm)**

1055

1056 We thank the reviewer for this point. Our understanding is that TEMRAs are definitionally
1057 CD45RA positive and *SELL* (L-selectin) negative, whereas Tcm are definitionally *SELL* positive
1058 and CD45RA negative. The reviewer is correct that by looking at one marker at a time, rather
1059 than both markers together, we may be incorrectly annotating some TEMRAs or Tcms as CD4-
1060 Naive. To address this point, we have added a biaxial plot of AB_CD45RA vs. AB_CD62L for
1061 the selected population, colored by the usage of CD4-Naive cGEP, which shows that a handful
1062 of the cells in the memory cluster that use the CD4-Naive GEP in fact do express the
1063 combination of markers that label them as naive.

1064



1065

1066 ***New Extended Data Figure 6E. Expression of CD45RA and CD62L colored by usage***
1067 ***of the CD4-Naive cGEP. Protein expression is smoothed using MAGIC²⁷ and color***
1068 ***intensities are averaged over 20 nearest neighbors to reduce overplotting.***

1069

1070 **R4.13**

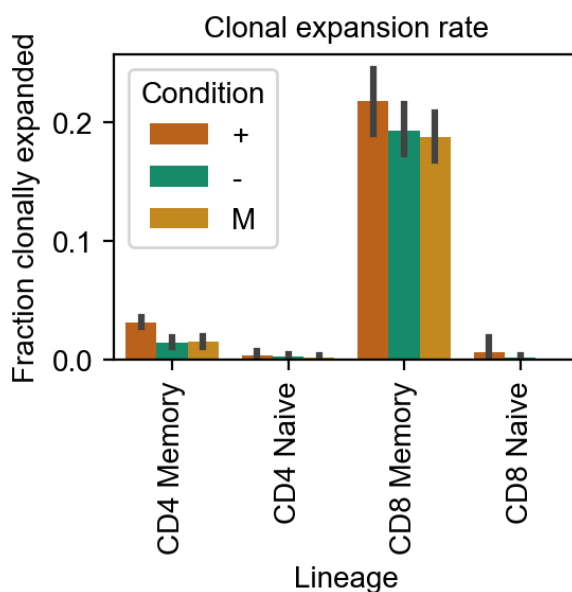
1071 **Section 5 & 6**

1072 **By definition naïve cells should not have encountered their cognate antigen. Clearly the**
1073 **immune system is constantly generating new TCR clonotypes, so indeed some CEFX**
1074 **reactive T cells could be naïve, however can the authors rule out a false positive result?**
1075 **By sequencing TCR clonotypes (as in section 6) one would expect that clonotypes in the**
1076 **naïve population should be less clonal/more diverse, and that T cells sharing the same**
1077 **clonotypes should not be found in both groups. Some analysis of TCR clonality would**
1078 **support this analysis and give additional confidence in the scoring (some of which is**
1079 **described in Fig S6).**

1080
1081 We thank the reviewer for this point and have carried out the analyses they suggested resulting
1082 in the new manuscript text and **New Extended Data Figure 7B**, reproduced below:
1083

1084 *Naive T-cell clones were significantly less likely to be clonally expanded (i.e. have 2+*
1085 *cells with the same beta CDR3 sequence) than memory clones ($P=2.4 \times 10^{-40}$ and*
1086 *$P=9.5 \times 10^{-257}$ for CD4 and CD8 T-cells, respectively). Furthermore, there were*
1087 *significantly more expanded clones in AIM-positive than AIM-negative CD4 memory T-*
1088 *cells and a trend in this direction for CD8 memory T-cells ($P=2.1 \times 10^{-7}$ and 0.14 for CD8*
1089 *and CD8 memory T-cells, respectively, **Extended Data Figure 7B**). Overall, there was*
1090 *significant concordance of AIM responsiveness within a clone: two cells with the same*
1091 *beta CDR 3 chain had a 77% chance of being concordantly AIM positive or AIM*
1092 *negative, compared to only a 51% probability based on chance alone (binomial test*
1093 *$P < 1 \times 10^{-200}$).*

1094



1095
1096
1097 **New Extended Data Figure 7B.** *Fraction of clonally expanded T-cells stratified by*
1098 *memory vs. naive and by treatment condition. Clonal expansion is defined as having 2+*
1099 *T-cells in the same donor with the same CDR3 beta chain. Error bars reflect 95 percent*
1100 *bootstrap confidence intervals.*

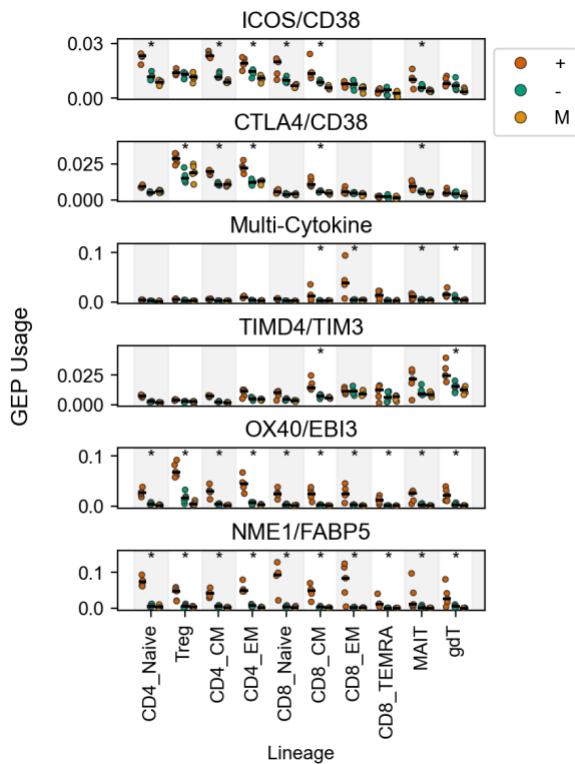
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121

R4.14

While I would agree that proliferation is a core response of TCR activation, using it to validate cGEP choice seems less useful – I believe Fig 4F is one of the least convincing in the paper, with very large error bars and no stats. This should be better explained.

We thank the reviewer for this comment. Figure 4F attempts to demonstrate the lineage-specificity of some of the activation-associated cGEPs. We agree that it would be a significantly stronger figure with clear statistical hypothesis testing. To address this, instead of plotting GEP usage across cells from all samples combined, we now plot the average cGEP usage per pseudobulk sample for each gated lineage (**Revised Figure 4F**). This allows us to do more rigorous statistical hypothesis testing. This shows that while NME1/FABP and OX40/EBI3 are upregulated with antigen recognition in all lineages, TIMD4/TIM3 was specific to CD8_CM and MAIT and multi-cytokine is specific to CD8 memory, MAIT, and gDT.

We also agree with the reviewer that proliferation is just one response to TCR activation and thus is not sufficient validation. We hope the AIM-Seq assay, the correspondence of the cGEPs with activation marker proteins, the clonal abundance analysis, and the association between activation rates and disease states (e.g. activation being higher in Covid-19 than in healthy donors), provide additional validation for the activation-associated cGEPs.



1122

1123 **Revised Figure 4F. Identifying cGEPs associated with TCR-dependent activation.**
1124 *Average usage of selected Aim-associated cGEPs in +, -, and U cells from different*
1125 *gated subsets, per sample. Lines show the median. * indicates $P < .05$ and average*
1126 *usage in the + cells of greater than 0.01.*
1127

1128 **R4.15**

1129 **The ASA data is very interesting, but again large differences between COMBAT and UK-**
1130 **covid data in terms of absolute percentage are not explained. I would like to see more on**
1131 **this part of the paper – for example why are some cGEPs specific to particular cancers,**
1132 **presumably due to cancer microenvironment factors or similar? Fig 5J starts to get at**
1133 **this with a correlation between tumor mutational load and exhaustion, though seeing**
1134 **how how ASA correlates with survival of individual patients would be more convincing**
1135 **(though such datasets are only just becoming available).**

1136
1137 Thank you for this comment. We have discussed the difference between the COMBAT and UK-
1138 Covid dataset per the comment above (**See R4.11**). We agree that further discussion of the link
1139 between ASA and the tumor microenvironment would be impactful. To this end, we have added
1140 a new analysis characterizing the link between ASA and individual cGEPs and response to
1141 immune-checkpoint inhibitor (ICI) therapy (**Key revision 1 and New figure 6, above**). We find
1142 this to be a more impressive analysis than **Figure S7B** (originally **Figure 6B**) which looks at
1143 cGEP association with particular cancers and thus have substituted it in the revised manuscript.
1144

1145 **R4.16**

1146 **A large number of recent papers have shown CXCL13 to be key marker for tumor-**
1147 **reactive T cells (i.e. not just T peripheral helper cells), while virus reactive T cells seem to**
1148 **show a slightly difference activation pattern. See previous points on CXCL13, I'd like to**
1149 **see this explored more: could it be that the use of a viral peptide pool in AIMseq gives**
1150 **slightly different cGEP enrichment than true tumor reactivity?**

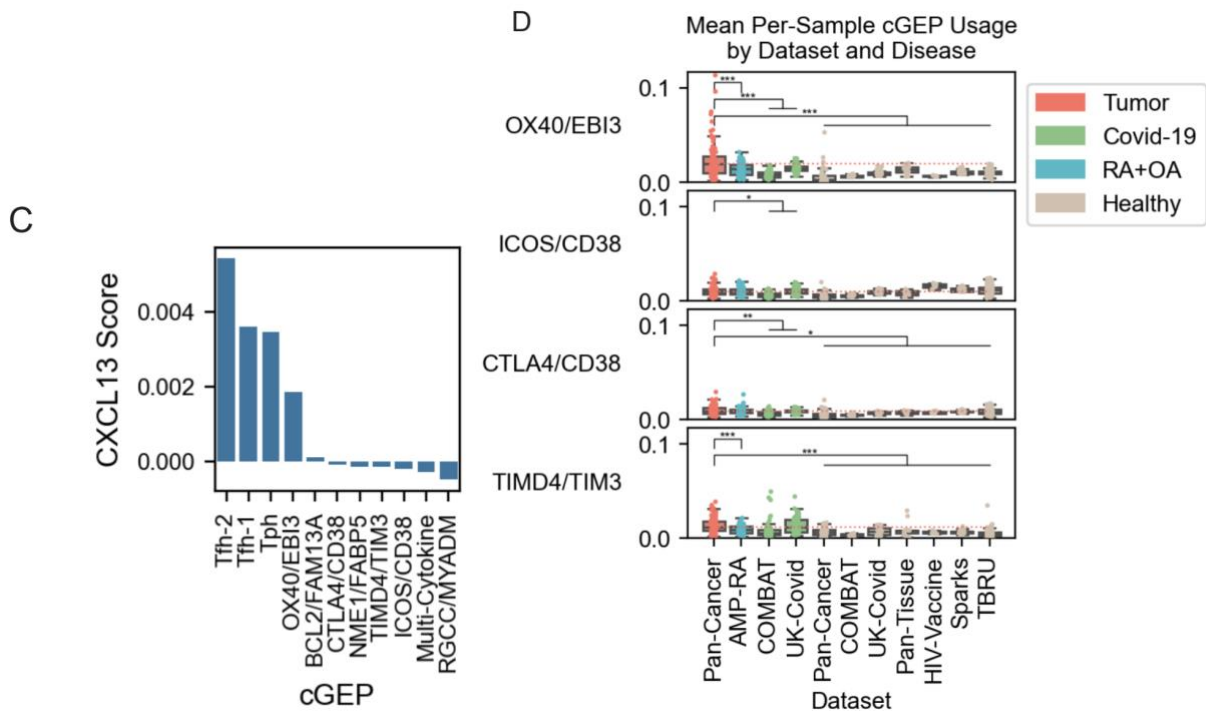
1151
1152 We thank the reviewer for this comment. We agree that there are likely different patterns of
1153 activation in different disease contexts and thus the CEFX stimulation may be biased in the
1154 responses it elicits. To address this, we have added text describing this limitation to the
1155 discussion section of the manuscript:
1156

1157 *A limitation of the AIM-Seq experiment is that it only uses a single costimulation signal,*
1158 *CD28 and CD49d costimulation, and not the full diversity of what is possible in vivo.*
1159 *Furthermore, it only uses a single set of peptides derived from common pathogens. It*
1160 *therefore could bias the patterns of cGEP enrichments that are discovered. For example,*
1161 *a tumor microenvironment may induce exhaustion more than an acute viral infection*
1162 *microenvironment.*

1163
1164 However, we note that the cGEPs we found to be activation-associated and used to define the
1165 ASA score were learned from both tissue and blood datasets from multiple disease contexts,
1166 including cancer. To the reviewer's point, CXCL13 is included among the top associated genes
1167 for the activation-associated OX40/EBI3 cGEP (**New Extended Data Figure 7C**). It is also high

1168 in the Tph, Tfh-1, and Tfh-2 cGEPs. The reviewer raises an interesting point that we may be
 1169 able to characterize different patterns of activation cGEP usages across different disease
 1170 contexts. We note that the OX40/EBI3 cGEP was learned from tumor and normal tissue.
 1171 Beyond this, we conducted an additional analysis showing that OX40/EBI3 usage is significantly
 1172 enriched in tumor tissue with respect to healthy tissue or blood, or other disease contexts (**New**
 1173 **Extended Data Figure 7D**). We describe this finding in the the new text reproduced below:
 1174

The OX40/EBI3 cGEP was identified specifically in tissues and was most common in tumor-infiltrating T-cells (Figure 1B, Extended Data Figure 7D). It includes CXCL13, a marker of tumor-reactivity⁴², as well as several activation-induced markers used to define AIM-positivity, including TNFRSF4, encoding OX40, and IL2RA, encoding CD25.



1181 **New Extended Data Figure 7C-D. (C)** Gene weights for CXCL13 in various cGEPs,
 1182 including Tfh-1, Tfh-2, Tph (positive controls) and all AIM-associated cGEPs. **(D)**
 1183 Boxplots of mean per-sample usage of the AIM-associated cGEPs used to define ASA,
 1184 by dataset and by disease. T-tests were performed for each cGEP to compare
 1185 differences between tumor samples and other disease types (***) indicates $P < 0.001$, **
 1186 indicates $P < 0.01$, * indicates $P < 0.05$). Boxes represent the interquartile range and
 1187 whiskers represent 95% quantile range.
 1188
 1189

1190 **R4.17**

1191 **G. References: appropriate credit to previous work?**

1192 **Yes**

1193

1194 **H. Clarity and context: lucidity of abstract/summary, appropriateness of abstract,**
1195 **introduction and conclusions**

1196 **The abstract, introduction and discussion sections are very well written.**

1197

1198 We thank the reviewer for these comments.

1199

1200

1201

1202 Bibliography

1203 1. Parikh, R. B. *et al.* Uptake and Survival Outcomes Following Immune Checkpoint Inhibitor
1204 Therapy Among Trial-Ineligible Patients With Advanced Solid Cancers. *JAMA Oncol* **7**,
1205 1843–1850 (2021).

1206 2. Sade-Feldman, M. *et al.* Defining T Cell States Associated with Response to Checkpoint
1207 Immunotherapy in Melanoma. *Cell* **176**, 404 (2019).

1208 3. Yost, K. E. *et al.* Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat*
1209 *Med* **25**, 1251–1259 (2019).

1210 4. Li, J. *et al.* Remodeling of the immune and stromal cell compartment by PD-1 blockade in
1211 mismatch repair-deficient colorectal cancer. *Cancer Cell* **41**, 1152–1169.e7 (2023).

1212 5. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29
1213 (2021).

1214 6. Andreatta, M. *et al.* Interpretation of T cell states from single-cell transcriptomics data using
1215 reference atlases. *Nature Communications* **12**, 1–19 (2021).

1216 7. Kang, J. B. *et al.* Efficient and precise single-cell reference atlas mapping with Symphony.
1217 *Nat Commun* **12**, 5890 (2021).

1218 8. Perkel, J. M. No installation required: how WebAssembly is changing scientific computing.
1219 *Nature* **627**, 455–456 (2024).

1220 9. Miller, T. E. *et al.* Programs, Origins, and Niches of Immunomodulatory Myeloid Cells in
1221 Gliomas. *bioRxiv* (2023) doi:10.1101/2023.10.24.563466.

- 1222 10. Li, H. *et al.* The dynamics of hematopoiesis over the human lifespan. *Nature Methods* 1–13
1223 (2024).
- 1224 11. Wagner, A., Regev, A. & Yosef, N. Revealing the vectors of cellular identity with single-cell
1225 genomics. *Nat. Biotechnol.* **34**, 1145–1160 (2016).
- 1226 12. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with
1227 Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
- 1228 13. Avila Cobos, F., Alquicira-Hernandez, J., Powell, J. E., Mestdagh, P. & De Preter, K.
1229 Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat Commun*
1230 **11**, 5650 (2020).
- 1231 14. Colino-Sanguino, Y. *et al.* Performance comparison of high throughput single-cell RNA-Seq
1232 platforms in complex tissues. *Heliyon* **10**, e37185 (2024).
- 1233 15. Cheng, C., Chen, W., Jin, H. & Chen, X. A Review of Single-Cell RNA-Seq Annotation,
1234 Integration, and Cell-Cell Communication. *Cells* **12**, (2023).
- 1235 16. Gavish, A. *et al.* Hallmarks of transcriptional intratumour heterogeneity across a thousand
1236 tumours. *Nature* **618**, 598–606 (2023).
- 1237 17. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for
1238 interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–
1239 15550 (2005).
- 1240 18. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–
1241 1740 (2011).
- 1242 19. Gu-Trantien, C. *et al.* CXCL13-producing TFH cells link immune suppression and adaptive
1243 memory in human breast cancer. *JCI Insight* **2**, (2017).
- 1244 20. Lin, X. *et al.* Follicular Helper T Cells Remodel the Immune Microenvironment of Pancreatic
1245 Cancer via Secreting CXCL13 and IL-21. *Cancers* **13**, 3678 (2021).
- 1246 21. Havenar-Daughton, C. *et al.* CXCL13 is a plasma biomarker of germinal center activity.
1247 *Proc Natl Acad Sci U S A* **113**, 2702–2707 (2016).

- 1248 22. Rao, D. A. *et al.* Pathologically expanded peripheral T helper cell subset drives B cells in
1249 rheumatoid arthritis. *Nature* **542**, 110–114 (2017).
- 1250 23. Kobayashi, S. *et al.* A distinct human CD4+ T cell subset that secretes CXCL13 in
1251 rheumatoid synovium. *Arthritis Rheum* **65**, 3063–3072 (2013).
- 1252 24. Tan, C. L. *et al.* Prediction of tumor-reactive T cell receptors from scRNA-seq data for
1253 personalized T cell therapy. *Nature Biotechnology* 1–9 (2024).
- 1254 25. Buettner, F. *et al.* Computational analysis of cell-to-cell heterogeneity in single-cell RNA-
1255 sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* **33**, 155–160
1256 (2015).
- 1257 26. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell
1258 transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*
1259 **36**, 411–420 (2018).
- 1260 27. van Dijk, D. *et al.* Recovering gene interactions from single-cell data using data diffusion.
1261 *Cell* **174**, 716–729.e27 (2018).

1262