nature portfolio

Peer Review File

Benchmarking single-cell multi-modal data integrations

Corresponding Author: Professor Qi Liu

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

Version 0:

Decision Letter:

4th Feb 2024

Dear Professor Liu,

Your Registered Report Stage 1 proposal, "Benchmarking single-cell multi-modal data integrations", has now been evaluated by expert reviewers. As you will see from their comments below, although the reviewers find your work of potential interest, they have raised several suggestions and concerns. We are interested in the possibility of proceeding further with your submission at Nature Methods, but would like to consider your response to these concerns before we reach a final decision on acceptance in principle and Stage 2 submission.

I have discussed the reviewer reports in detail with our chief editor and other members of the Nature Methods team, with a view to 1) identifying key priorities that should be addressed in revision, and 2) overruling reviewer requests that we think fall beyond the scope of the study. Please see the attached file which contains the reviewer reports annotated with our editorial recommendations. We invite you to revise your Stage 1 proposal to address these concerns (before you embark on the proposed experiments).

We are committed to providing a fair and constructive peer review process. Please do not hesitate to contact us if there are any reviewer requests that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your paper:
* include a point-by-point response to the reviewers 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
* include a cover letter with the following information

---an anticipated timeline for completing the study (should the study be accepted in principle)
---a statement confirming that, should the study be accepted in principle, you agree to share your raw data and conform to our other open science requirements (see details below)
---a statement confirming that, following acceptance in principle, you agree to register the Stage 1 proposal in a recognized repository (either publicly or under private embargo), until submission of the Stage 2 manuscript (please note we can assist you in uploading your Stage 1 proposal in our Figshare space)
---a statement confirming that if you later decide to withdraw your manuscript from consideration at Nature Methods, you agree to the journal publishing a brief note about the withdrawn proposal

Please resubmit all the necessary files electronically by using the link below to access your home page:

Link Redacted

IMPORTANT: 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 proposal within 4 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.

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 please visit www.springernature.com/orcid.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further. We look forward to seeing the revised manuscript and thank you for the opportunity to consider your work.

Sincerely,

Lin Tang, PhD
Senior Editor
Nature Methods

Reviewers' Comments:
Reviewer #1:
Remarks to the Author:
Fu et al. proposed a comprehensive benchmarking analysis to assess the performance of various single-cell multi-modal data integration methods. With the increasing prevalence of single-cell multi-modal data, numerous integration methods have been developed, posing a challenge for practitioners in choosing the most suitable approach. This paper seeks to address this challenge by conducting an evaluation of 31 integration methods that encompass the integration of DNA, RNA, and protein data. The evaluation criteria include clustering accuracy, batch effect removal, usability, and robustness. The results will provide a thorough assessment of different methods. Additionally, the authors outlined a plan to validate the replicability of the proposed benchmarking analyses. While the paper may offer some new insights into method performance, I have the following major concerns.

1. Significance: Lee et al. (2023 Genome Biol. 2023 Oct 24;24(1):244. doi: 10.1186/s13059-023-03073-x) has already performed a similar evaluation, and this diminished the importance of the novelty and impact of the proposed analyses.

2. Soundness:
a. Limited dataset scope: The proposed evaluation is based on only two datasets. To achieve comprehensive assessment, this limited scope is inadequate. For a publication at the Nature Methods level, a broader evaluation using diverse datasets is imperative. Ideally, this should encompass various tissue types and diseases to provide a thorough understanding of each method's performance across diverse scenarios. In reality, it is unlikely that a single method is uniformly better than the others. Rather, specific methods may excel under particular situations. Thus, I recommend that the authors broaden their dataset selection to enhance the robustness and applicability of their evaluations.
b. Limited dataset scale: The paper proposed the use of two datasets, the CITE-seq dataset (90,261 cells) and the 10x Multiome dataset (69,249 cells). However, both datasets are relatively small in total cell numbers. They may prove inadequate for the purpose of upsampling to simulate large-scale data encompassing around 500,000 cells. The reliance on two small datasets for benchmarking raises a significant concern regarding the paper's ability to generalize findings to larger-scale datasets. Addressing this limitation by incorporating datasets of greater scale would enhance the robustness and applicability of the proposed benchmarking.

3. Potential bias in evaluation: An inherent challenge in benchmarking analyses lies in defining the ground truth. For example, when evaluating clustering results, choosing Seurat's clustering results as the ground truth may favor Seurat. Addressing this potential bias in the evaluation is a crucial aspect that requires careful consideration. The paper should explicitly outline the strategies to mitigate such biases in the benchmarking to ensure a fair and unbiased evaluation.

Other comments:
1. Table 1: ordering by publication time. I suggest the authors to arrange the methods chronologically based on their publication time. This ordering would provide a helpful perspective on the evolution of these methods over time.
2. Table 1: integration capabilities. Additionally, it would be helpful to include information on whether each method is capable of integrating RNA-ATAC, RNA-protein, or both. This added detail would offer a quick reference for readers seeking methods that cater to specific integration requirements.
3. The paper missed two important recent publications: Seurat5 (Nat Biotechnol. 2023 May 25. doi: 10.1038/s41587-023-01767-y) and StabMap (Nat Biotechnol. 2023 May 25. doi: 10.1038/s41587-023-01766-z)

The significance of this paper can be enhanced by incorporating an analysis of multi-modal spatial omics technologies. While these technologies are gaining popularity, it remains uncertain whether methods originally developed for single-cell analysis can effectively analyze such data. Currently, only two methods, SpatialGlue (doi: https://doi.org/10.1101/2023.04.26.538404) and MUSE (Nat Biotechnol. 2022 Aug;40(8):1200-1209), are explicitly designed for multi-modal spatial omics data analysis. However, their comparative performance against single-cell methods is unclear. Including an assessment of spatial data in the paper would not only address this critical gap but also significantly increase the impact and significance of the study.

Reviewer #2:
Remarks to the Author:
In this proposed registered report, Fu et al. proposed a benchmarking of single-cell multi-modal data integration methods. The proposal seeks to be comprehensive in the methods profiled, including 31 different existing integration methods for both paired and unpaired data integration. The results of a systematic benchmarking study for this application, if done well, would be a valuable resource for the community. However, data integration benchmarking is a very difficult thing to get right, especially for multimodal integration as proposed here. The main challenge is there is a lack of ground truth data with regard to cell type labels or embedding space, and so the evaluation of methods needs to be performed extremely carefully to avoid biasing results.

My main concerns are summarized below:

Evaluation of accuracy:

Evaluating the accuracy of data integration methods is the most challenging part of the proposed study, and I have several concerns regarding the proposed methods.

The choice of evaluation metrics needs to be more carefully considered. Some of the suggested metrics do not make sense for evaluation and would only serve to make the results less clear. It's better if the authors can decide on a smaller number of metrics that are well motivated, and minimize dependency on cell type annotations. The LISI metric as described does not take into account differences in cell type abundance between samples, and so the expected diversity of batches in each neighborhood is not taken into account by the metric. For the nearest cell barcode metric, it's unlikely that looking at only one neighboring cell will be informative, looking at a group of k neighboring cells would give a more accurate estimate of the local neighborhood composition. For imputation assessment, the Pearson correlation between raw and imputed values does not work well. All correlations tend to be low due to the sparsity of the raw measurement. Looking at k-nearest-neighbor smoothed gene expression vs model-imputed can give a better measure of model performance.

For assessing the preservation of biological structure, cell type labels are required. The methods used to produce these cell type labels can bias the evaluation of performance of different integration methods. How cell type labels will be determined is not explained in the proposed study. This is a key problem that needs to be addressed.

How the different metrics will be summarized needs to be carefully considered. The current proposal aims to weight bio conservation and batch removal 60/40, but this is arbitrary and different weightings would change overall rankings of methods. The same is proposed for single embedding vs cross embedding. As much as possible, arbitrary numbers like this should be avoided in the evaluation. Separate rankings for bio conservation and batch removal could be reported instead, and left to the reader to determine which is more important for their datasets. The authors also propose to perform min-max scaling of scores across methods. This can hide cases where all methods do well or all do poorly. Choosing a smaller number of metrics to focus on and simply reporting each metric, rather than a scaled value relative to other methods, may be better.

Experimental design:

Here the authors have sought to undertake a comprehensive benchmarking study. However, only one tissue type is proposed here: bone marrow mononuclear cells. This will limit the generalizability of the benchmarking results. The authors should try to be comprehensive in the types of tissues used in the study. Furthermore, there is already a set of rankings and performance evaluations for different multimodal data integrations using the BMMC data here: https://proceedings.mlr.press/v176/lance22a/lance22a.pdf.

Other comments:

Some other methods should be added: Seurat bridge integration (Hao et al. 2023 Nature Biotech), StabMap (Ghazanfar et al. 2023 Nature Biotech)
Seurat integration has multiple dimension reduction methods that can be used: CCA, PCA. These should be evaluated independently. The method can also correct the embedding space or the gene expression space, which can be evaluated separately.
"we will compute the GAM using the CreateGeneActivityMatrix function within the Seurat v3 package" this function has been removed in later versions of Seurat in favor of the more accurate GeneActivity function in the Signac package
It was unclear how upsampling of a dataset would be performed
Some statements too vague, eg " For deep generative models with posterior imputation, we will also assess whether the imputation function can effectively mitigate data dropout and enhance downstream data analysis"
I did not understand the motivation behind this part for the ATAC evaluation: "For the scATAC modality, we will initially involve determining the number of accessible sites (N) within each cell. We will then identify the overlaps between the top N peaks in the algorithm's posterior imputations and the original accessible sites."

Reviewer #3:

Remarks to the Author:
The authors propose to realize a benchmarking of single-cell multi-omics intergation methods considering both paired and unpaired data.

This is a highly relevant topic for the computational biology research community, expecially if concrete guidelines for users could be derived from the work.

The proposed inpiut data, selected tools and evaluation metrics are good. The only modifications I would perform are:

-in the evaluation a set of metrics based on the biological interpretability of the obtained results. In the list of proposed tools there is MOFA that provides a biological interpretation of the obtained latent dimensions, while other methods do not do it (e.g. scVI). An evaluation on the quality of the biological interpretability of the latent space provided by different methods would be highly relevant.

-the proposed input data contain different levels of cell annotation granularity, an evaltation based on differnet levels of granularity would be highly informative.

-including additional data widely used to test single-cell multi-omics integration methods would be beneficial to have a better view on the variability of the obtained performances

-an evaluation on the scalability of the different methods, would finally be highly important (e.g. running them on the full Neurips data)

No bias can be identified in the proposed analysis.

The description is methodologically clear. In the final submission would be important to have a Jupyter notebook or similar reproducing all the performed analysis and helping the research community to evaluate new methods with repsect to those proposed in this work.

****

OPEN SCIENCE REQUIREMENTS

REPORTING SUMMARY AND EDITORIAL POLICY CHECKLISTS
When revising your manuscript, please update your 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.

IMAGE INTEGRITY
When submitting the revised version of your manuscript, please pay close attention to our Digital Image Integrity Guidelines and to the following points below:

-- that unprocessed scans are clearly labelled and match the gels and western blots presented in figures.
-- that control panels for gels and western blots are appropriately described as loading on sample processing controls
-- all images in the paper are checked for duplication of panels and for splicing of gel lanes.

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

DATA AVAILABILITY
We strongly encourage you to deposit all new data associated with the paper in a persistent repository where they can be

freely and enduringly accessed. We recommend submitting the data to discipline-specific and community-recognized repositories; a list of repositories is provided here: http://www.nature.com/sdata/policies/repositories

All novel DNA and RNA sequencing data, protein sequences, genetic polymorphisms, linked genotype and phenotype data, gene expression data, macromolecular structures, and proteomics data must be deposited in a publicly accessible database, and accession codes and associated hyperlinks must be provided in the “Data Availability” section.

Refer to our data policies here: https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-data

To further increase transparency, we encourage you to provide, in tabular form, the data underlying the graphical representations used in your figures. This is in addition to our data-deposition policy for specific types of experiments and large datasets. For readers, the source data will be made accessible directly from the figure legend. Spreadsheets can be submitted in .xls, .xlsx or .csv formats. Only one (1) file per figure is permitted: thus if there is a multi-paneled figure the source data for each panel should be clearly labeled in the csv/Excel file; alternately the data for a figure can be included in multiple, clearly labeled sheets in an Excel file. File sizes of up to 30 MB are permitted. When submitting source data files with your manuscript please select the Source Data file type and use the Title field in the File Description tab to indicate which figure the source data pertains to.

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

Version 1:

Decision Letter:

20th May 2024

Dear Professor Liu,

Your Registered Report Stage 1 proposal, "Benchmarking single-cell multi-modal data integrations", has now been evaluated by expert reviewers. As you will see from their comments below, although the reviewers find your work of considerable potential interest, they have raised several suggestions and concerns. We are very interested in the possibility of proceeding further with your submission at Nature Methods, but would like to consider your response to these concerns before we reach a final decision on acceptance in principle and Stage 2 submission.

I have discussed the reviewer reports in detail with our chief editor and other members of the Nature Methods team, with a view to 1) identifying key priorities that should be addressed in revision, and 2) overruling reviewer requests that we think fall beyond the scope of the study. Please see the attached file which contains the reviewer reports annotated with our editorial recommendations. We invite you to revise your Stage 1 proposal to address these concerns (before you embark on the proposed experiments).

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

When revising your paper:
* include a point-by-point response to the reviewers 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
* include a cover letter with the following information

---an anticipated timeline for completing the study (should the study be accepted in principle)
---a statement confirming that, should the study be accepted in principle, you agree to share your raw data and conform to our other open science requirements (see details below)
---a statement confirming that, following acceptance in principle, you agree to register the Stage 1 proposal in a recognized repository (either publicly or under private embargo), until submission of the Stage 2 manuscript (please note we can assist you in uploading your Stage 1 proposal in our Figshare space)
---a statement confirming that if you later decide to withdraw your manuscript from consideration at Nature Methods, you agree to the journal publishing a brief note about the withdrawn proposal

Please resubmit all the necessary files electronically by using the link below to access your home page:

Link Redacted

IMPORTANT: 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 proposal within 4 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.

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 please visit www.springernature.com/orcid.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further. We look forward to seeing the revised manuscript and thank you for the opportunity to consider your work.

Sincerely,

Lin Tang, PhD
Senior Editor
Nature Methods

Reviewers' Comments:
Reviewer #1:
Remarks to the Author:
The authors have appropriately addressed my previous concerns. I don't have any additional comments.

Reviewer #2:
Remarks to the Author:
In their revised proposal, the authors have addressed some of my initial comments. Some comments still require additional consideration:

1. Datasets proposed: this study is still very limited in the types of cells it proposes to use for benchmarking. Most of the datasets listed are immune. Since there is now plenty of published multimodal data available from a variety of tissues, the authors should aim to be more comprehensive in this aspect.

2. Treatment of the scATAC data: the authors propose to use different methods of creating the gene activity matrix (GeneActivity vs CreateGeneActivityMatrix functions) depending on dataset. All the scATAC-seq datasets used should have

fragment files available, and so the GeneActivity function should be used uniformly. If there are cases where a fragment file is not available, this can be constructed from the BAM file. The GeneActivity function is also incorrectly attributed to the Seurat package in the revised text.

3. Weighting of evaluation metrics: I again want to reiterate my previous comment that an overall ranked list that combines bio conservation and batch mixing in an arbitrarily weighted manner may not be the most useful thing to report. I recognize that 60/40 weighting was proposed in the scIB paper, but I disagree that this should be interpreted as an accepted standard in the field. Rather than trying to condense all the evaluation metrics into a single number, the authors can report the performance of each method across each of the different categories (bio conservation, etc.)

4. One previous comment was not addressed: some statements are too vague, eg " For deep generative models with posterior imputation, we will also assess whether the imputation function can effectively mitigate data dropout and enhance downstream data analysis". How will enhanced downstream analysis be evaluated in this case? I also caution against use of the term "dropout" as there is some debate whether scRNA-seq is zero-inflated. Instead terms like "sparsity" can be used.

Version 2:

Decision Letter:

3rd Jul 2024

Dear Professor Liu,

Thank you once again for submitting your revised Stage 1 Registered Report, entitled "Benchmarking single-cell multi-modal data integrations." After consulting again with Reviewer 2, who thinks the points previously raised by Reviewer 3 are now sufficiently addressed, I am delighted to say that we can offer acceptance in principle. You may progress to Stage 2 and complete your study as approved.

As you know, a condition of acceptance-in-principle is that the authors agree to deposit their Stage 1 accepted manuscript in a repository, either publicly or under embargo until Stage 2 manuscript acceptance and publication. We are very keen to showcase our accepted-in-principle manuscripts, so that our readers, reviewers, and potential authors can gain insight into the requirements of the format as well as an idea of the types of projects that are suitable for publication as Registered Reports in Nature Methods. We have set up a space on figshare (https://springernature.figshare.com/registered-reports_nmethods) to host all of our accepted-in-principle manuscripts, which can either be made public or kept under embargo until Stage 2 acceptance (depending on author preference). This gives you the opportunity to have your work publicly associated with Nature Methods, and of course we will be very pleased to showcase your report if you agree to share it publicly.

If you agree with posting your Stage 1 manuscript on our figshare space, we will upload it on your behalf and either set it public or place it under embargo, depending on your choice. Your protocol will be licensed under a CC BY license (Creative Commons Attribution 4.0 International License). The CC BY license allows for maximum dissemination and re-use of open access materials and is preferred by many research funding bodies. Under this license users are free to share (copy, distribute and transmit) and remix (adapt) the contribution including for commercial purposes, providing they attribute the contribution in the manner specified by the author or licensor (read full legal code: http://creativecommons.org/licenses/by/4.0/legalcode) Please note that any use of https://springernature.figshare.com will be subject to the figshare terms of use. Figshare has the right to enforce these terms and conditions where applicable. Use of third party services and sites will be subject to the relevant terms of use and will apply if we act on your behalf in this regard. Do let me know if you would like to take up this option or if you have any questions regarding the manuscript deposition requirement.

Please also note that depositing the work on our figshare space does not preclude deposition of your Stage 1 manuscript on other repositories – your manuscript can also be posted on any other public repository of your choice.

Following completion of your study, we invite you to resubmit your finalized manuscript as a Stage 2 Registered Report. We will send the Stage 2 manuscript to our reviewers for a final check, but they will be instructed that any comments on novelty and/or potential significance of the results will not factor into our final decision.

IMPORTANT: Please note that your manuscript can still be rejected for publication at Stage 2 if the Editors consider any of the following to hold:

-- The authors substantially alter the rationale for the study as approved in the Stage 1 submission (please note that the Introduction should not be significantly modified from the Stage 1 manuscript).
-- The authors fail to adhere closely to the approved experimental plan. (Please contact us as soon as possible for advice if at any point you need to make any changes to your experimental plan.)
-- The authors' conclusions are not justified given the data obtained.
--Any post hoc (unregistered) analyses are not justified or are overly dominant in shaping the authors' conclusions.
-- Our open science requirements (detailed below) are not followed.

Should authors choose to withdraw their Registered Report at any time, we will publish a Withdrawn Registration notice.

When you are ready, please use the following link to access your home page and submit your Stage 2 Registered Report:

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 expect your Stage 2 Registered Report to be submitted by the date specified in your latest cover letter. If unforeseen circumstances prevent submission by that date, please contact us as soon as possible.

Thank you again for submitting your work to Nature Methods and we look forward to receiving your Stage 2 Registered Report! Please do not hesitate to reach out to me at any time if you have questions.

Yours sincerely,

Lin Tang, PhD
Senior Editor
Nature Methods

OPEN SCIENCE REQUIREMENTS

REPORTING SUMMARY AND EDITORIAL POLICY CHECKLISTS
When submitting your Stage 2 manuscript, please include a reporting summary and editorial policy checklists. If you have any questions about these checklist, please see http://www.nature.com/authors/policies/availability.html or contact me.

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 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.

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

-- that unprocessed scans are clearly labelled and match the gels and western blots presented in figures.
-- that control panels for gels and western blots are appropriately described as loading on sample processing controls
-- all images in the paper are checked for duplication of panels and for splicing of gel lanes.

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

DATA AVAILABILITY
We strongly encourage you to deposit all new data associated with the paper in a persistent repository where they can be freely and enduringly accessed. We recommend submitting the data to discipline-specific and community-recognized repositories; a list of repositories is provided here: http://www.nature.com/sdata/policies/repositories

All novel DNA and RNA sequencing data, protein sequences, genetic polymorphisms, linked genotype and phenotype data, gene expression data, macromolecular structures, and proteomics data must be deposited in a publicly accessible database, and accession codes and associated hyperlinks must be provided in the “Data Availability” section.

Refer to our data policies here: https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-data

To further increase transparency, we encourage you to provide, in tabular form, the data underlying the graphical representations used in your figures. This is in addition to our data-deposition policy for specific types of experiments and large datasets. For readers, the source data will be made accessible directly from the figure legend. Spreadsheets can be submitted in .xls, .xlsx or .csv formats. Only one (1) file per figure is permitted: thus if there is a multi-paneled figure the source data for each panel should be clearly labeled in the csv/Excel file; alternately the data for a figure can be included in multiple, clearly labeled sheets in an Excel file. File sizes of up to 30 MB are permitted. When submitting source data files with your manuscript please select the Source Data file type and use the Title field in the File Description tab to indicate which figure the source data pertains to.

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 any 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 here.

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 please visit www.springernature.com/orcid.

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.

Reviewers' Comments:

Reviewer #2:
Remarks to the Author:
The authors have now addressed all of my comments

Version 3:

Decision Letter:

Our ref: NMETH-RR54370C

7th Jan 2025

Dear Dr. Liu,

Thank you for submitting your revised manuscript "Benchmarking single-cell multi-modal data integrations" (NMETH-RR54370C). It has now been seen by the original referees and their comments are below. In light of these review comments, we'll be happy in principle to publish it in Nature Methods, pending minor revisions to comply with our editorial and formatting guidelines.

We think the additional analyses suggested by Reviewer #1 are not needed for publication of this Stage 2 Registered Report manuscript. That said, please briefly discuss these points as future directions in the Discussion section.

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.

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,

Lin Tang, PhD
Senior Editor
Nature Methods

Reviewer #1 (Remarks to the Author):

The authors have conducted a very comprehensive benchmarking evaluation across various data types, algorithms, and scenarios. The results presented in this paper will undoubtedly serve as a valuable resource for researchers interested in multi-omic data integration. I have only one suggestion that, if addressed, could further strengthen the paper.

Spatial Multi-omic Data Integration: While I appreciate the authors' addition of evaluations for spatial multi-omic data integration, the current analysis remains too limited. I recommend the following improvements:
1. Incorporate non-spatial multi-omic data integration algorithms into the evaluation. Comparing algorithms specifically designed for spatial multi-omic data with more generic algorithms would provide a broader perspective on their relative performance and utility.
2. Extend the scope of the evaluation beyond RNA and protein integration. I encourage the authors to benchmark their methods using RNA and epigenetic data, such as those provided in the following paper: https://www.nature.com/articles/s41586-023-05795-1. This would enrich the analysis and demonstrate the applicability of the proposed methods to a wider range of multi-omic data types.

Addressing these points would enhance the breadth and impact of the paper, making it even more beneficial to the research community.

Reviewer #2 (Remarks to the Author):

The authors have conducted a thorough benchmarking study of multimodal single cell data integration methods. They have included a broad range of computational tools, and applied these to a broad range of datasets. Overall, the benchmarking is comprehensive and carried out in an unbiased way, as should be the goal for a benchmarking study. As this is a stage 2 registered report, my review must assess whether the plan described in stage 1 has been faithfully carried out. I believe that it has been. The authors also provide a balanced discussion of the result and highlight key areas that should be focused on in future development. Overall, this benchmarking study will be a useful resource for the field.

Version 4:

Decision Letter:

20th May 2025

Dear Professor Liu,

Thank you very much for sending us the updated version of your Registered Reports, "Benchmarking single-cell multi-modal data integrations". I am pleased to inform you that this paper has now been accepted for publication in Nature Methods. The received and accepted dates will be 12th Nov 2023 and 20th May 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) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including self-archiving policies. Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

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.

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. Thank you very much again for publishing your paper at Nature Methods!

Best regards,

Lin Tang, PhD
Senior Editor
Nature Methods

Review 1

Fu et al. proposed a comprehensive benchmarking analysis to assess the performance of various single-cell multi-modal data integration methods. With the increasing prevalence of single-cell multi-modal data, numerous integration methods have been developed, posing a challenge for practitioners in choosing the most suitable approach. This paper seeks to address this challenge by conducting an evaluation of 31 integration methods that encompass the integration of DNA, RNA, and protein data. The evaluation criteria include clustering accuracy, batch effect removal, usability, and robustness. The results will provide a thorough assessment of different methods. Additionally, the authors outlined a plan to validate the replicability of the proposed benchmarking analyses. While the paper may offer some new insights into method performance, I have the following major concerns.

1.1 Significance: Lee et al. (2023 Genome Biol. 2023 Oct 24;24(1):244. doi: 10.1186/s13059-023-03073-x) has already performed a similar evaluation, and this diminished the importance of the novelty and impact of the proposed analyses.

Response:

Thank you for your comment! As you mentioned, Lee et al. recently conducted a benchmark study on 5 unpaired scRNA and scATAC (diagonal) integration tools and 4 multiome-guided (mosaic) integration tools, identifying Seurat v4 reference mapping as the best method for integrating scRNA, scATAC, and multiome data[1]. We sincerely appreciated such works, however, the benchmark strategy, the systematicness and completeness presented in our study are distinct from existing works like Lee et al.'s study. And our study aims to achieve greater novelty and impact through the following points:

(1) Firstly, our benchmark proposal will encompass 40 multimodal integration algorithms for RNA, ATAC, protein, and spatial modalities. This includes paired multi-modal integration methods, unpaired diagonal integration methods, and unpaired mosaic integration methods, where such benchmark strategy and the greater number of benchmark tools will provide a more systematic and objective evaluation compared to existing works like Lee et al.'s study. Specifically, the paired multi-modal integration methods in our proposed study will include 17 paired scRNA and scATAC integration methods, 13 paired scRNA and ADT integration methods, and 3 spatial multiomics integration methods. Additionally, unpaired integration methods will include 14 scRNA and scATAC diagonal integration methods, 8 CITE-seq mosaic integration methods, and 8 multiome mosaic integration methods. The current benchmark algorithms are summarized in Table 3 and detailed in Tables 1 and 2 in our revised Stage 1 proposed manuscript as follows:

Table 1. Paired multi-modal integration algorithms in this benchmark

software	scRNA+scATAC	scRNA+ADT	Spatial multiomics	journal	year
Citefuse		✓		*Bioinformatics*	*2020*
MOFA+	✓	✓		*Genome Biology*	*2020*

scAI	✓			*Genome Biology*	*2020*
cobolt	✓			*Genome Biology*	*2021*
DCCA	✓			*Bioinformatics*	*2021*
scDEC	✓			*Nature machine intelligence*	*2021*
scMM	✓	✓		*Cell Reports Methods*	*2021*
scMVAE	✓			*Briefings in Bioinformatics*	*2021*
SeuratV4 WNN	✓	✓		*Cell*	*2021*
totalVI		✓		*Nature Methods*	*2021*
MultiMap	✓	✓	✓	*Genome Biology*	*2021*
bindSC		✓		*Genome Biology*	*2022*
Multigrate	✓	✓		*BioRxiv*	*2022*
SAILERX	✓			*Nucleic Acids Research*	*2022*
SCALEX		✓		*Nature Communications*	*2022*
sciPENN		✓		*Nature machine intelligence*	*2022*
scMDC	✓	✓		*Nature Communications*	*2022*
scMVP	✓			*Genome Biology*	*2022*
uniport	✓			*Nature Communications*	*2022*
DeepMAPS	✓	✓		*Nature communications*	*2023*
Maxfuse		✓		*Nature Biotechnology*	*2023*
MultiVI	✓			*Nature Methods*	*2023*
scMCs	✓			*Bioinformatics*	*2023*
SpatialGlue			✓	*BioRxiv*	*2023*

Table 2. Unpaired multi-modal integration algorithms in this benchmark

software	**diagonal scRNA+scATAC**	**mosaic scRNA+scATAC**	**mosaic scRNA+ADT**	**journal**	**year**
Liger iNMF	✓			*Cell*	*2019*
Seuratv3 CCA	✓			*Cell*	*2019*

Seuratv3 RPCA	✓			*Cell*	*2019*
unionCom	✓			*Bioinformatics*	*2020*
cobolt		✓		*Genome Biology*	*2021*
GLUE	✓			*Nature Biotechnology*	*2021*
Liger online iNMF	✓			*Nature Biotechnology*	*2021*
totalVI			✓	*Nature Methods*	*2021*
bindSC	✓			*Genome Biology*	*2022*
Liger UiNMF	✓			*Nature Communications*	*2022*
Multigrate		✓	✓	*bioRxiv*	*2022*
Pamona	✓			*Bioinformatics*	*2022*
SCALEX	✓			*Nature Communications*	*2022*
sciPENN			✓	*Nature machine intelligence*	*2022*
uniport	✓			*Nature Communications*	*2022*
scVAEIT		✓	✓	*PNAS*	*2022*
GCN-SC	✓			*Briefings in Bioinformatics*	*2023*
Maxfuse	✓			*Nature Biotechnology*	*2023*
MultiVI		✓		*Nature Methods*	*2023*
SIMBA	✓			*Nature Methods*	*2023*
SeuratV5 bridge		✓	✓	*Nature Methods*	*2023*
StabMap		✓	✓	*Nature Methods*	*2023*
scMoMaT		✓	✓	*Nature communications*	*2023*
MIDAS		✓	✓	*Nature Biotechnology*	*2024*

Table 3. The summary of tested benchmark algorithm tutorials for each integration type

Integration type	Algorithms	Deep models	Generative imputation	Non-deep models	Non-executable
paired scRNA+scATAC	*17*	*12*	*6*	*5*	*3*
paired scRNA+ADT	*13*	*6*	*4*	*7*	*1*
unpaired diagonal scRNA+scATAC	*14*	*6*	*-*	*8*	*1*

unpaired mosaic scRNA+scATAC	*8*	*6*		*2*	*0*
unpaired mosaic scRNA+scADT	*8*				*0*
Spatial multi-omics	*3*	*1*	-	*2*	*0*

(2) Secondly, our proposed study will address the conclusion bias towards deep learning models which will probably occur in existing works. Since their study evaluated deep learning models using only a CPU platform, it would be unfair to assess the performance of GPU-based multi-modal integration tools. This approach may lead to biased conclusions, especially considering that GLUE and MultiVI were significantly slower (more than 4 times) than other integration tools.

As listed in Table 3 above, we have conducted a pilot evaluation for all benchmark algorithms on both CPU and GPU servers. This ensures a fair benchmark study for both deep learning and non-deep learning tools.

(3) Our benchmark study will also address potential biases identified in existing works, where Seurat tools are applied to annotate the PBMC benchmark dataset and Seurat is found to be the top method in clustering accuracy. However, this may suffer from the exact same issue you mentioned in point 1.3.

To mitigate this potential bias, we have examined the cell annotation methods for all benchmark datasets and selected datasets with cell annotations independent of cell clustering in joint embedding or scRNA embedding. In our revised proposal, we have identified three datasets with cell labels that meet these requirements for unbiased clustering evaluation. We have provided detailed information about the cell annotation methods and the reasons for choosing these datasets in our revised manuscript as follows:

"*BMMC datasets*

The BMMC Multiome and CITE-seq datasets[2] contain nested batches of donors and experimental sites. The cell annotations were first generated for each modality and batch separately, and then manually annotated the cell subclusters inconsistent between the two modalities by evaluating which modality marker features more clearly described the specific cellular subpopulation[2]. The cell type annotations of BMMC datasets do not rely on joint embedding clustering or scRNA modality clustering, which largely ensures the independence of clustering results of all benchmark algorithms.

SHARE-seq dataset

The SHARE-seq mouse skin dataset[3] was annotated for scRNA and scATAC modalities separately, and then marked cell labels inconsistent between the two modalities as "mixed". We will use cells

with consistent cell labels in two modalities as the ground truth for cell clustering, as these annotated cell labels will not favor the clustering results from joint embedding or scRNA embedding.

Spatial multiomics datasets

We will use three spatial CITE-seq datasets[4, 5] for spatial multi-modal integration methods evaluation. Among the three spatial multi-omics datasets, we will use the human lymph node dataset for biological conservation evaluation, as this dataset contains cell annotations by manually annotated tissue structures from histological images, which are irrelevant to joint embedding or scRNA embedding.

"

1.2. Soundness:

a. Limited dataset scope: The proposed evaluation is based on only two datasets. To achieve comprehensive assessment, this limited scope is inadequate. For a publication at the Nature Methods level, a broader evaluation using diverse datasets is imperative. Ideally, this should encompass various tissue types and diseases to provide a thorough understanding of each method's performance across diverse scenarios. In reality, it is unlikely that a single method is uniformly better than the others. Rather, specific methods may excel under particular situations. Thus, I recommend that the authors broaden their dataset selection to enhance the robustness and applicability of their evaluations.

b. Limited dataset scale: The paper proposed the use of two datasets, the CITE-seq dataset (90,261 cells) and the 10x Multiome dataset (69,249 cells). However, both datasets are relatively small in total cell numbers. They may prove inadequate for the purpose of upsampling to simulate large-scale data encompassing around 500,000 cells. The reliance on two small datasets for benchmarking raises a significant concern regarding the paper's ability to generalize findings to larger-scale datasets. Addressing this limitation by incorporating datasets of greater scale would enhance the robustness and applicability of the proposed benchmarking.

Response:

Thank you for your constructive suggestion! We've broadened the scope of our dataset in response to the limitations noted in our initial proposal. Now, we're including nine benchmark datasets that cover various tissue/sample types like BMMC (bone marrow mononuclear cells), HSPC (hematopoietic stem and progenitor cells), PBMC (peripheral blood immune cells) from COVID-19 patients, skin, spleen, thymus, and brain. (as shown below, Table 4 in the revised manuscript, with details provided in Table S2).

Additionally, the COVID-19 CITE-seq dataset contains more than 700,000 cells, which will improve the robustness of the study with a larger dataset scale. For paired or unpaired integration evaluation of scRNA and scATAC, due to the much higher dimensionality in ATAC (>100,000) compared to ADT (10-200), we found that a large proportion of methods failed to integrate datasets with 500,000 cells or 100,000 cells. As a result, we will evaluate the accuracy of CITE-seq methods

on a dataset of approximately 700,000 cells, but only use simulated datasets with 500,000 cells to assess the scalability of scRNA and scATAC integration tools. The table of benchmark datasets in our revised proposal is as follows:

Table 4 Summary of multi-modal benchmarking datasets in this study

Dataset name	Multi-omics type	Tissue/sample type	Batch	Species	Cell number	Tested tasks
BMMC Multiome	*RNA + ATAC*	*BMMC*	*12 donors from 4 sites*	*Human*	*69,249*	*All accuracy metrics and robustness metrics*
BMMC CITE-seq	*RNA + ADT*	*BMMC*	*12 donors from 4 sites*	*Human*	*90,261*	*All accuracy metrics and robustness metrics*
HSPC Multiome	*RNA + ATAC*	*HSPC*	*4 donors of 5 time points*	*Human*	*105,942*	*Batch removal, cell alignment and imputation*
HSPC CITE-seq	*RNA + ADT*	*HSPC*	*4 donors of 5 time points*	*Human*	*70,988*	*Batch removal, cell alignment and imputation*
SHARE-seq skin	*RNA + ATAC*	*Skin*	*-*	*Mouse*	*34,774*	*Biological conservation, cell alignment and imputation*
COVID19 CITE-seq	*RNA+ADT*	*PBMC*	*143 donors*	*Human*	*781,123*	*batch removal and paired and unpaired RNA+ADT Scalability,*
Lymph node spatial	*spatial+RNA+ADT*	*Lymph node*	*2 samples*	*Human*	*6,843*	*Biological conservation and Batch removal*
Thymus spatial	*spatial+RNA+ADT*	*Thymus*	*4 samples*	*Mouse*	*17,824*	*Batch removal*
Spleen SPOTS	*spatial+RNA+ADT*	*Spleen*	*2 samples*	*Mouse*	*5,336*	*Batch removal*

1. 3. Potential bias in evaluation: An inherent challenge in benchmarking analyses lies in defining the ground truth. For example, when evaluating clustering results, choosing Seurat's clustering results as the ground truth may favor Seurat. Addressing this potential bias in the evaluation is a crucial aspect that requires careful consideration. The paper should explicitly outline the strategies to mitigate such biases in the benchmarking to ensure a fair and unbiased evaluation.

Response:

Thank you for your insightful comment! In response to your comment 1.1, to mitigate this potential bias, we present an un-biased evaluation strategy to examine the cell annotation methods for all benchmark datasets and select datasets with cell annotations independent of cell clustering in joint embedding or scRNA embedding. We only use cell labels in these unbiased datasets for the evaluation of biological conservation accuracy. Details are described in response to your comment 1.1.

Other comments:

1.4 Table 1: ordering by publication time. I suggest the authors to arrange the methods chronologically based on their publication time. This ordering would provide a helpful perspective on the evolution of these methods over time.

Response:

Thank you for your suggestion! We now ordered benchmark methods based on their publication time in Tables 1-2 and supplementary Table 2 in our revised manuscript.

1.5 Table 1: integration capabilities. Additionally, it would be helpful to include information on whether each method is capable of integrating RNA-ATAC, RNA-protein, or both. This added detail would offer a quick reference for readers seeking methods that cater to specific integration requirements.

Response:

Thanks for your comment! We showed the capability of all benchmark algorithms for multi-modal integration types and modality in Tables 1-2 and supplementary Table 2 in our revised manuscript.

1.6 The paper missed two important recent publications: Seurat5 (Nat Biotechnol. 2023 May 25. doi: 10.1038/s41587-023-01767-y) and StabMap (Nat Biotechnol. 2023 May 25. doi: 10.1038/s41587-023-01766-z)。

Response:

Thank you for your suggestion! To provide a more comprehensive benchmark for multi-modal integration, we now added two new unpaired mosaic integration tasks to our revised proposal. These tasks integrate paired multi-modal datasets with unpaired ones. We will benchmark eight multiome mosaic integration methods and eight CITE-seq mosaic integration methods, including Seurat v5[6] and StabMap[7]. We completed the testing for these mosaic integration methods during this revision. Additionally, we have designed specific benchmark metrics, datasets, and workflows for evaluating unpaired mosaic multi-modal integrations.

1.7 The significance of this paper can be enhanced by incorporating an analysis of multi-modal spatial omics technologies. While these technologies are gaining popularity, it remains uncertain whether methods originally developed for single-cell analysis can effectively analyze such data. Currently, only two methods, SpatialGlue (doi: https://doi.org/10.1101/2023.04.26.538404) and MUSE (Nat Biotechnol. 2022 Aug;40(8):1200-1209), are explicitly designed for multi-modal spatial omics data analysis. However, their comparative performance against single-cell methods is unclear. Including an assessment of spatial data in the paper would not only address this critical gap but also significantly increase the impact and significance of the study.

Response:

Thank you for your helpful suggestion! This comment is insightful. Following your suggestion, we first investigate algorithms which support spatial multiomics analysis including SpatialGlue[8] and MUSE[9]. MUSE method supports multi-modal integration of spatial-transcriptomic with imaging modality, which is out of the scope of our benchmark study for integrating DNA, RNA and ADT modality. And SpatialGlue method supports data integration and analysis of spatial-RNA-ATAC, spatial-RNA-ADT and spatial-RNA-epigenome, which has been added to benchmark methods in our proposed study.

Besides SpatialGlue, we have found another two algorithms available for spatial multiomics data integration: MultiMAP[10] and MEFISTO[11]. The MultiMAP method is a tool for joint dimension reduction of multiple modalities to the same latent space, which reports to support spatial multiomics integration in their paper. Although there is no official tutorial for spatial multi-omics analysis in their source GitHub, we have successfully applied this tool using spatial position along with single-cell multi-modal data. MEFISTO is the latest updated tool in the MOFA[12] package, which extends multiomics data analysis with function of spatial and temporal information integration. We have adapted codes from official guidelines for spatial-transcriptome analysis and multi-omics analysis, and successfully implemented MEFISTO for spatial multi-omics data. The demo Jupyter notebook for spatial multi-omics data analysis with MEFISTO and MultiMAP has been uploaded to our GitHub folder at https://github.com/bm2-lab/SCMMI_Benchmark/tree/main/scripts/spatial/. Additionally, we will compare batch removal performance of these spatial multi-omics analysis tools with two spatial-transcriptome batch removal tools - Harmony[13] and STAligner[14] - to assess potential data analysis options for novel spatial multi-omics datasets.

Editor recommendations:
Reviewer 1 has raised important concerns on several aspects of this analysis including significance, soundness, potential bias in evaluation, and other comments, some of which are also noted by other reviewers. We think all these points should be well addressed in the revision. Among other revisions/changes, we strongly encourage you to include an analysis of multi-modal spatial omics technologies to enhance the significance of this analysis, as suggested by Reviewer 1.

Response:
Thanks! We have added analysis and datasets for spatial multi-modal technologies, and revised our proposed manuscript to address all raised concerns from reviewer 1 to improve the significance and soundness of our benchmark study.

Reviewer 2
In this proposed registered report, Fu et al. proposed a benchmarking of single-cell multi-modal data integration methods. The proposal seeks to be comprehensive in the methods profiled, including 31 different existing integration methods for both paired and unpaired data integration. The results of a systematic benchmarking study for this application, if done

well, would be a valuable resource for the community. However, data integration benchmarking is a very difficult thing to get right, especially for multimodal integration as proposed here. The main challenge is there is a lack of ground truth data with regard to cell type labels or embedding space, and so the evaluation of methods needs to be performed extremely carefully to avoid biasing results.

My main concerns are summarized below:
Evaluation of accuracy:

2.1 Evaluating the accuracy of data integration methods is the most challenging part of the proposed study, and I have several concerns regarding the proposed methods. The choice of evaluation metrics needs to be more carefully considered. Some of the suggested metrics do not make sense for evaluation and would only serve to make the results less clear. It's better if the authors can decide on a smaller number of metrics that are well motivated, and minimize dependency on cell type annotations. The LISI metric as described does not take into account differences in cell type abundance between samples, and so the expected diversity of batches in each neighborhood is not taken into account by the metric. For the nearest cell barcode metric, it's unlikely that looking at only one neighboring cell will be informative, looking at a group of k neighboring cells would give a more accurate estimate of the local neighborhood composition. For imputation assessment, the Pearson correlation between raw and imputed values does not work well. All correlations tend to be low due to the sparsity of the raw measurement. Looking at k-nearest-neighbor smoothed gene expression vs model-imputed can give a better measure of model performance.

Response:
Thanks for detailed and insightful suggestions! Following your comments, we have revised our proposal to improve our metrics selection as follows:

(1) To minimize redundancy, we have condensed four batch effect removal metrics to three, five biological conservation metrics to four and three cell alignment accuracy metrics to two. Specifically, we merge metrics of Batch ASW_{site} and batch ASW_{sample} to single batch ASW metric. We remove "celltype ASW" from biological conservation evaluation. And we remove seurat alignment score in cell alignment evaluation. To lessen the reliance on cell type annotations, we modified "cell labels" to cell clusters in batch ASW metric computation for each algorithm. We have revised the batch ASW metric in our updated manuscript as follows:

"For batch mixing evaluation, we will initially adjust the batch silhouette width of cell i by equation (1) and calculate the average batch silhouette width of all cells in cell cluster j as batch ASW_j (equation 2). The final batch ASW score (equation 3) will be an average of batch ASW scores across all cell clusters."

(2) For LISI metrics, we agree with reviewer that this metric does not considering the difference in labels abundance diversity. However, we will only compare this metric score of different methods within identical datasets, not across diverse ones in our benchmark study. This

approach prevents bias from variations in cell type abundance or batch size. Thus, we will maintain this metric as a local accuracy measure in our revised manuscript.

(3) The nearest cell barcode/cell type metric functions exactly the same but has different names as "cell match score" metric applied in "Match modality" task in the BMMC Kaggle paper[15], which is also recommended in your point 2.4. As we found that "cell match" may better fit this metric than "nearest cell barcode", we renamed this metric to "Cell match/Cell type match" in our revised proposal.
Overall, we have revised the benchmark metrics in Table 5 of our revised manuscript as follows:

"**Table 5. Summary of accuracy metrics**

Metrics	***Group***	***Characteristic***	***Input form***
batch ASW	*batch effect removal*	*cell-cluster-based*	*Latent embedding*
GC	*batch effect removal*	*cell-cluster-based*	*KNN graph*
graph iLISI	*batch effect removal*	*local metric*	*KNN graph*
ARI	*biological conservation*	*global metric*	*Latent embedding*
AMI	*biological conservation*	*global metric*	*Latent embedding*
graph cLISI	*biological conservation*	*local metric*	*KNN graph*
isolated labels ASW	*biological conservation*	*rare cell type*	*Latent embedding*
FOSCTTM	*cell alignment accuracy*	*proportion of false match*	*Latent embedding*
cell match/cell type match	*cell alignment accuracy*	*proportion of true match*	*Latent embedding*
AUPR	*imputation accuracy*	*accuracy of peaks' ranking*	*ATAC profile*
Pearson's r	*imputation accuracy*	*correlation to ground truth*	*RNA or ADT profile*

"

(4) Following your suggestion, we evaluated k-NN smoothing method for scRNA and scATAC raw profiles. As expected, k-NN smoothing can improve the non-zero ratio in the raw profile and enhance the imputation accuracy metric score for algorithms' imputation output. However,

as some algorithms may also directly apply k-NN methods for imputation (such as StabMap), it may be unfair to evaluate these methods. To mitigate the potential bias with k-NN smoothing profile, we will evaluate the imputation results using both raw profiles and k-NN smoothing profiles for accuracy evaluation (see Figure 3 below). We have added a section in our pilot data analysis to evaluate k-NN smoothing measurement, and revised the imputation evaluation methods in our manuscript as follows:

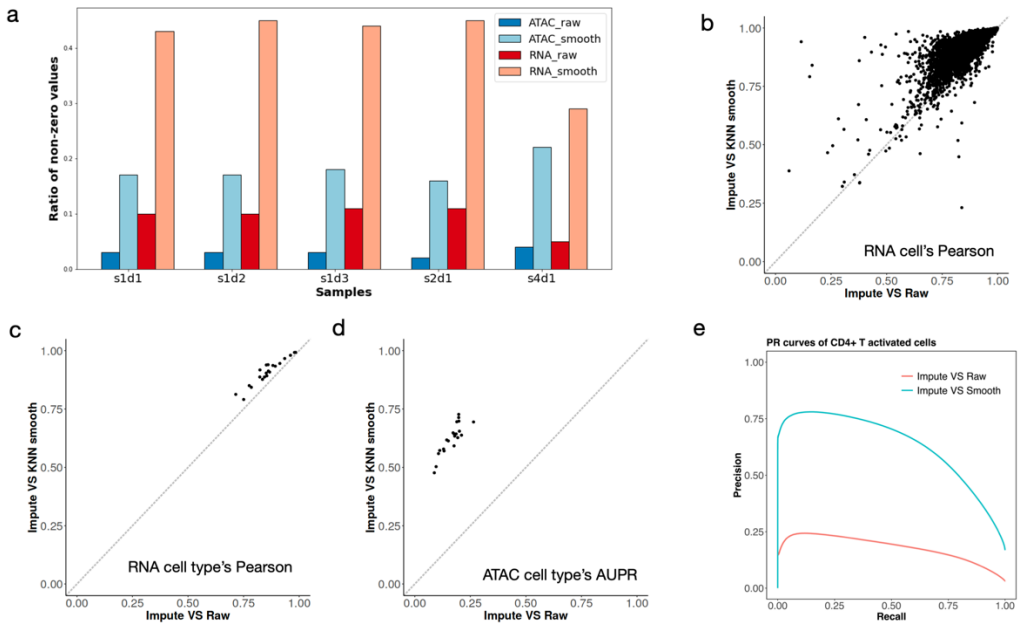

Figure 3. Assessment of k-NN smoothing method for RNA and ATAC imputation evaluation. (a) ratio of non-zero values in raw measurement and k-NN smooth expression of RNA and ATAC in 5 samples of BMMC datasets. (b-c) Pearson correlations of MultiVI RNA imputation with raw measurement (x-axis) and k-NN smoothed expression (y-axis) for each cell (b) and each cell type (c). (d) AUPR values of MultiVI ATAC imputation for each cell type using raw measurement (x-axis) and k-NN smoothed expression (y-axis) as ground truth. (e) PR curve of MultiVI ATAC imputation evaluated with raw measurement (red) and k-NN smoothed expression (green) as ground truth.

"Pilot data analysis

Previous study applied raw measurement of single-cell profile as imputation ground truth[16]. However, the raw measurement itself may suffer from high dropout rate in single-cell RNA-seq and ATAC-seq data, and may not work well for imputation assessment. As k-NN smoothing method may mitigate the potential bias in ground truth, we further investigate whether k-NN smoothing expression would be better than raw measurement for imputation evaluation. Here, we applied both k-NN smoothing expression and raw measurement with 5 samples in BMMC multiome datasets[2] as ground truth to evaluate the accuracy of MultiVI imputation results for both RNA and ATAC modality. Firstly, k-NN smoothing expression mitigated the data sparsity in both RNA and ATAC modality, and contained significant higher fractions of non-zero values in all five samples (Figure

3a). Next, we evaluated MultiVI RNA imputation accuracy for each cell or each cell type using Pearson's r metric (Figure 3b, c). K-NN smoothed RNA expression showed relative higher Pearson correlation value than raw measure to MultiVI imputation results in most of cells (Figure 3b) and all cell types (Figure 3c). Also, we evaluated MultiVI ATAC imputation accuracy for each cell type using AUPR (area under precision recall curve) metric (Figure 3d, e). MultiVI showed much higher AUPR using k-NN smoothed ATAC profile than raw measurement as ground truth (Figure 3d), as more severe data sparse issue exists in scATAC data than scRNA data. For CD4+ T activated cells, we found consistent higher precision with k-NN smoothed expression at all recall level (Figure 3e). These findings indicate the potential improvement of k-NN smoothed expression comparing with raw measurement for raw data dropout, and the k-NN smoothed expression has also been applied as gold standard for imputation evaluation."

"For both paired integration methods and unpaired mosaic integration methods, we will evaluate the accuracy of posterior imputation with gold standards of both raw data profile and KNN imputed profile from raw count data with k-NN smoothing package[17]. For miss modality imputation, we will also compare generative imputation results with KNN impute results in StabMap[7], which impute the miss modality from paired multi-modal data."

2.2 For assessing the preservation of biological structure, cell type labels are required. The methods used to produce these cell type labels can bias the evaluation of performance of different integration methods. How cell type labels will be determined is not explained in the proposed study. This is a key problem that needs to be addressed.

Response:

Thank you for your helpful comments! Following your suggestion, we have inspected the cell annotation methods for all benchmark datasets, and select datasets with cell annotations independent to cell clustering in joint embedding or scRNA embedding. In our revised proposal, we have identified three datasets with cell labels meet these requirements for unbiased clustering evaluation. We added the detailed cell annotation methods and reasons to choose the dataset in our revised manuscript as following:

"*BMMC datasets*

The BMMC Multiome and CITE-seq datasets[2] contains nested batches of donors and experimental sites. And the cell annotations were first generated for each modality and batch separately, and then manually annotated the cell subclusters inconsistent between two modalities by evaluating which modality marker features more clearly described the specific cellular subpopulation[2]. The cell type annotations of BMMC datasets do not rely on joint embedding clustering or scRNA modality clustering, which will be largely ensure the independency to clustering results of all benchmark algorithms.

SHARE-seq dataset

SHARE-seq mouse skin dataset[3] was annotated for scRNA and scATAC modality separately, and then marked cell labels inconsistent between two modalities as "mixed". And we will use cells with consistent cell labels in two modalities as the ground truth for cell clustering, as these annotated cell labels will not favor the clustering results from joint embedding or scRNA embedding.

Spatial multiomics datasets

We will use three spatial CITE-seq datasets[4, 5] for spatial multi-modal integration methods evaluation. Among three spatial multi-omics datasets, we will use the human lymph node dataset for biological conservation evaluation, as this dataset contains cell annotations by manually annotated tissue structures from histological image, which is irrelevant to joint embedding or scRNA embedding."

2.3 How the different metrics will be summarized needs to be carefully considered. The current proposal aims to weight bio conservation and batch removal 60/40, but this is arbitrary and different weightings would change overall rankings of methods. The same is proposed for single embedding vs cross embedding. As much as possible, arbitrary numbers like this should be avoided in the evaluation. Separate rankings for bio conservation and batch removal could be reported instead, and left to the reader to determine which is more important for their datasets. The authors also propose to perform min-max scaling of scores across methods. This can hide cases where all methods do well or all do poorly. Choosing a smaller number of metrics to focus on and simply reporting each metric, rather than a scaled value relative to other methods, may be better.

Response:

Thank you for your insightful comments! Following your suggestions, we have revised our proposal for the metrics summary. We agreed with the reviewer on this point, and we will report the original metric score like biological conservation and batch removal without scaling, and rank distinct types of metrics separately. Nevertheless, we will also keep to apply min-max scaling and weighting or averaging when calculating the summarized ranks for benchmark algorithms in summary tables or figures. The goal of min-max scaling and weighting or averaging applied here is to get a summarized rank from different types of metrics. And it is difficult to fairly summarize different metrics with varying ranges without scaling. Additionally, we will maintain weights of 0.6 and 0.4 for bio-conservation and batch removal respectively in method rankings, as referenced from the scIB benchmark study[18] for the scIB overall score, which is taken as accepted weight strategies in this domain. For other metric types, we will use the averaging method for all other metrics with similar significance; this summarized score will be used solely for ranking purposes. The method of summarizing metrics has been revised in our updated proposal as follows:

"*In the accuracy evaluation, we will first report the original accuracy metric scores for benchmark algorithms, and rank these algorithms by the accuracy metric scores of bio-conservation, batch*

effect removal and cell alignment accuracy in all benchmark datasets. To get the accuracy metrics ranking from metrics of different ranges, we will first apply min-max normalize to each metric for the ith benchmark algorithm as follows:

$$score_i = \frac{metric_i - metric_{min}}{metric_{max} - metric_{min}} \quad (12)$$

The algorithm score will be normalized by scores of the best and worst performing algorithms in each metric, with the best method receiving a score of 1 and the worst a score of 0. Second, in line with scIB study which took co-efficient of 0.4 and 0.6 for batch effect removal and biological conservation[18], we will first report the average score for both biological conservation and batch effect removal metrics in all tested datasets. Then, the final embedding accuracy score will also be reported using the following equation:

$$s_{embed\_acc} = 0.4 * \bar{s}_{batch} + 0.6 * \bar{s}_{bio\_cons} \quad (13)$$

Additionally, the average from all cell alignment accuracy metrics will be used to compute the cell alignment accuracy score. Third, for paired integration accuracy, we will employ the embedding accuracy score alongside the biological conservation score calculated from both single-batch and mixed batches datasets. For unpaired integration accuracy, we will report the score and ranking of embedding accuracy and cell alignment accuracy for each method, and then rank these methods by average of these two ranks as follows:

$$rank_{unpaired} = \frac{rank_{embedding} + rank_{alignment}}{2} \quad (14)$$

In the robustness evaluation, we will apply the same equations as accuracy evaluation section to compute the accuracy metrics and scores for the downsampled datasets. However, paired multi-modal integration evaluation will only use the appropriate metrics filtered by the methods mentioned in the robustness assessment section. The final robustness score will be computed from the accuracy score at the lowest sequencing depth (s_{acc_10}), which evaluates the method's accuracy under the worst data quality conditions, and the average multi-gradients AUC score from selected metrics at all downsampled sequencing depths (s_{grad_auc}). We will report the robustness score (s_{acc_10} and s_{grad_auc}) and ranks of selected metrics for each benchmark method, and rank robustness of benchmark methods by average ranks of two robustness score as follows:

$$rank_{robustness} = \frac{rank_{acc\_10} + rank_{grad\_auc}}{2} \quad (15)$$

For unpaired mosaic integration, we will further assess the robustness of tested methods to paired multi-modal dataset size using the average multi-gradients AUC score (s_{size_auc}). And robustness of unpaired mosaic integration methods will be finally ranked by average ranks of three robustness score as follows:

$$rank_{robustness} = \frac{rank_{acc\_10} + rank_{grad\_auc} + rank_{size\_auc}}{3} \quad (16)$$

”

Experimental design:

2.4 Here the authors have sought to undertake a comprehensive benchmarking study.

However, only one tissue type is proposed here: bone marrow mononuclear cells. This will limit the generalizability of the benchmarking results. The authors should try to be comprehensive in the types of tissues used in the study. Furthermore, there is already a set of rankings and performance evaluations for different multimodal data integrations using the BMMC data here: https://proceedings.mlr.press/v176/lance22a/lance22a.pdf.

Response:

Thank you for your helpful comments! In response to your point 2.1, we have already used the exact same metric in BMMC Kaggle Paper[15] that you mentioned, and revised the metric name to match this paper in our revised manuscript. Additionally, we now included more tissue types along with nine datasets for different tasks in our revised benchmark proposal. These benchmark datasets are listed as follows in Table 4 of our revised manuscript:

Table 4 Summary of multi-modal benchmarking datasets in this study

Dataset name	Multi-omics type	Tissue/sample type	Batch	Species	Cell number	Tested tasks
BMMC Multiome	*RNA + ATAC*	*BMMC*	*12 donors from 4 sites*	*Human*	*69,249*	*All accuracy metrics and robustness metrics*
BMMC CITE-seq	*RNA + ADT*	*BMMC*	*12 donors from 4 sites*	*Human*	*90,261*	*All accuracy metrics and robustness metrics*
HSPC Multiome	*RNA + ATAC*	*HSPC*	*4 donors of 5 time points*	*Human*	*105,942*	*Batch removal, cell alignment and imputation*
HSPC CITE-seq	*RNA + ADT*	*HSPC*	*4 donors of 5 time points*	*Human*	*70,988*	*Batch removal, cell alignment and imputation*
SHARE-seq skin	*RNA + ATAC*	*Skin*	*-*	*Mouse*	*34,774*	*Biological conservation, cell alignment and imputation*
COVID19 CITE-seq	*RNA+ADT*	*PBMC*	*143 donors*	*Human*	*781,123*	*batch removal and paired and unpaired RNA+ADT Scalability,*
Lymph node spatial	*spatial+RNA+ADT*	*Lymph node*	*2 samples*	*Human*	*6,843*	*Biological conservation and Batch removal*
Thymus spatial	*spatial+RNA+ADT*	*Thymus*	*4 samples*	*Mouse*	*17,824*	*Batch removal*
Spleen SPOTS	*spatial+RNA+ADT*	*Spleen*	*2 samples*	*Mouse*	*5,336*	*Batch removal*

Other comments:

2.5 Some other methods should be added: Seurat bridge integration (Hao et al. 2023 Nature Biotech), StabMap (Ghazanfar et al. 2023 Nature Biotech)

Seurat integration has multiple dimension reduction methods that can be used: CCA, PCA. These should be evaluated independently. The method can also correct the embedding space

or the gene expression space, which can be evaluated separately.

Response:

Thanks for your suggestion! To provide a more comprehensive benchmark for multi-modal integration, we added two new unpaired mosaic integration tasks in our revised proposal, which integrates paired multi-modal datasets with unpaired datasets. We will benchmark 8 multiome mosaic integration methods and 8 CITE-seq mosaic integration methods including Seurat v5[6] and StabMap[7], and completed the testing for these mosaic integration methods during this revision. We have also designed specific benchmark metrics, datasets and workflows for unpaired mosaic multi-modal integration evaluation. Also, we have revised our proposal to test Seurat v3 RPCA and CCA separately. All unpaired integration methods in our benchmark study are listed in Table 2 as following:

Table 2. Unpaired multi-modal integration algorithms in this benchmark

software	diagonal scRNA+scATAC	mosaic scRNA+scATAC	mosaic scRNA+ADT	journal	year
Liger iNMF	✓			*Cell*	*2019*
Seuratv3 CCA	✓			*Cell*	*2019*
Seuratv3 RPCA	✓			*Cell*	*2019*
unionCom	✓			*Bioinformatics*	*2020*
cobolt		✓		*Genome Biology*	*2021*
GLUE	✓			*Nature Biotechnology*	*2021*
Liger online iNMF	✓			*Nature Biotechnology*	*2021*
totalVI			✓	*Nature Methods*	*2021*
bindSC	✓			*Genome Biology*	*2022*
Liger UiNMF	✓			*Nature Communications*	*2022*
Multigrate		✓	✓	*BioRxiv*	*2022*
Pamona	✓			*Bioinformatics*	*2022*
SCALEX	✓			*Nature Communications*	*2022*
sciPENN			✓	*Nature machine intelligence*	*2022*
uniport	✓			*Nature Communications*	*2022*
scVAEIT		✓	✓	*PNAS*	*2022*
GCN-SC	✓			*Briefings in Bioinformatics*	*2023*

Maxfuse	✓			*Nature Biotechnology*	*2023*
multivi		✓		*Nature Methods*	*2023*
SIMBA	✓			*Nature Methods*	*2023*
SeuratV5 bridge		✓	✓	*Nature Methods*	*2023*
StabMap		✓	✓	*Nature Methods*	*2023*
scMoMaT		✓	✓	*Nature communications*	*2023*
MIDAS		✓	✓	*Nature Biotechnology*	*2024*

For algorithms that correct gene space and output graphs rather than latent embeddings, we will apply the same metrics, excluding all ASW metrics for evaluation, as ASW metrics are not available for graph input. We have added the description of a specific evaluation method to the gene expression space correction method in the revised manuscript as follows:

"*For algorithms correcting gene space and output graph rather than latent embedding, we will take graph output of algorithm as input for all metrics evaluation except for ASW metrics.*"

2.6 "we will compute the GAM using the CreateGeneActivityMatrix function within the Seurat v3 package" this function has been removed in later versions of Seurat in favor of the more accurate GeneActivity function in the Signac package.

Response:

Thank you for your suggestion! The "GeneActivity" function in the Signac package takes a scATAC fragment information file as input to compute the signal density across genes as GAM. For datasets without a "scATAC fragment information file", we can only use the "CreateGeneActivityMatrix" function within the Seurat v3 package to compute GAM from scATAC peaks located in gene promoters. We have added a detail description of GAM computation in our revised manuscript as following:

"*For unpaired diagonal integration algorithms, which often involve transforming ATAC chromatin accessibility matrices into gene activity matrices (GAM), we will compute the GAM using the GeneActivity function in Seurat v4 package*[19] *using the ATAC per fragment information file. For certain scATAC datasets without ATAC per fragment information file, we will use CreateGeneActivityMatrix function within the Seurat v3 package*[20] *to compute GAM from the scATAC accessibility profile.* "

2.6 It was unclear how upsampling of a dataset would be performed.

Response:
Thank you for your comment! The upsampling method is only applied for algorithm's scalability evaluation, which would not be influenced by the proportion of cell type labels or cell batch information. We have appended the details of upsampling in our revised manuscript as follows:

"For scalability evaluation, we will generate new simulation datasets (Supplementary Table 2) with a specific number of cells by either downsampling or upsampling from the BMMC Multiome and COVID-19 CITE-seq benchmark dataset. Specifically, for simulation datasets smaller than size of origin dataset, we will randomly select specific number of cells from origin benchmark data as downsampling simulation data. For simulation datasets larger than size of origin data, we will randomly sample from origin datasets for several times and merge these datasets as upsampling simulation dataset."

2.7 Some statements too vague, eg " For deep generative models with posterior imputation, we will also assess whether the imputation function can effectively mitigate data dropout and enhance downstream data analysis". I did not understand the motivation behind this part for the ATAC evaluation: "For the scATAC modality, we will initially involve determining the number of accessible sites (N) within each cell. We will then identify the overlaps between the top N peaks in the algorithm's posterior imputations and the original accessible sites."

Response:
Thank you for your comments! We apologized for the unclear description here. Now we replaced this metric with an AUC-based metric, which was referenced from the scATAC imputation evaluation metric applied in the recent published MIDAS paper[16]. To calculate the AUC, the prediction is the imputation of the algorithm, and the label represents the state of chromatin sites in a single cell or cell type, as 1 or 0. In the evaluation of scATAC data imputation, due to the sparsity of scATAC data, positive labels for accessible sites in each cell are far less than negative labels. Therefore, we further adapted the AUPR metric instead of AUROC metric for scATAC imputation evaluation, which is more appropriate in a data imbalance scenario. To address this point, we have revised our manuscript as follows:

"To calculate the imputation accuracy score, we will take different metrics for scRNA, scATAC and ADT based on their nature of the modality. Specifically, for scRNA and ADT modality representing relative abundance, we will calculate the Pearson correlation between raw expression and algorithm's posterior imputation. For the scATAC modality representing binary states of chromatin sites, the severe data sparsity in the scATAC modality exists and the positive chromatin accessible sites would be far less than negative inaccessible sites. Therefore, we will adapt AUPR (area under precision recall curve) for imputation accuracy evaluation instead of the ATAC AUROC metric applied in the study of MIDAS paper[16], which will be less biased for unbalanced labels."

Editor:

We agree with Reviewer 2 that their concerns on evaluation of accuracy, experimental design and other points are important and should be all addressed.

Response:
Thanks! We have addressed all concerns raised from Reviewer 2 in our revised manuscript.

Reviewer 3
The authors propose to realize a benchmarking of single-cell multi-omics integration methods considering both paired and unpaired data. This is a highly relevant topic for the computational biology research community, especially if concrete guidelines for users could be derived from the work. The proposed input data, selected tools and evaluation metrics are good. The only modifications I would perform are:

3.1 in the evaluation a set of metrics based on the biological interpretability of the obtained results. In the list of proposed tools there is MOFA that provides a biological interpretation of the obtained latent dimensions, while other methods do not do it (e.g. scVI). An evaluation on the quality of the biological interpretability of the latent space provided by different methods would be highly relevant.

Response:
Thanks for your helpful suggestion! The biological interpretability in latent space is an important characteristic for summarizing multi-modal integration methods. Accordingly, we will evaluate biological interpretability of all benchmark methods in their latent feature space. We add a description for biological interpretability evaluation in usability section of our revised manuscript as follows:

"We will also evaluate the biological interpretability of latent space output for benchmarked algorithms."

3.2 the proposed input data contain different levels of cell annotation granularity, an evaluation based on different levels of granularity would be highly informative.

Response:
Thank you for your advice! We agree with reviewer that evaluations based on different levels of granularity would be highly informative. Following your advice, we will perform evaluation based on different levels of cell annotations, such as cell batch annotations of samples and sites in BMMC datasets, and cell batch annotations of donors and timepoint in HSPC datasets. We added description for cell annotation granularity evaluation in our revised manuscript as follows:

"In this study, we will consider the granularity of batch annotation for assessing batch effects, including batches from different samples, sites, and time points. For unpaired mosaic integration methods, we will also use this metric to evaluate the batch removal effect between paired and unpaired multi-modal datasets, which will treat each input dataset as a single batch. "

3.3 including additional data widely used to test single-cell multi-omics integration methods would be beneficial to have a better view on the variability of the obtained performances

Response:
Thanks for your suggestion! We have included additional tissue types with 9 datasets for different tasks to improve the variability of the obtained performances. These benchmark datasets are listed in Table 4 in our revised manuscript as follows:

Table 4 Summary of multi-modal benchmarking datasets in this study

Dataset name	Multi-omics type	Tissue/sample type	Batch	Species	Cell number	Tested tasks
BMMC Multiome	*RNA + ATAC*	*BMMC*	*12 donors from 4 sites*	*Human*	*69,249*	*All accuracy metrics and robustness metrics*
BMMC CITE-seq	*RNA + ADT*	*BMMC*	*12 donors from 4 sites*	*Human*	*90,261*	*All accuracy metrics and robustness metrics*
HSPC Multiome	*RNA + ATAC*	*HSPC*	*4 donors of 5 time points*	*Human*	*105,942*	*Batch removal, cell alignment and imputation*
HSPC CITE-seq	*RNA + ADT*	*HSPC*	*4 donors of 5 time points*	*Human*	*70,988*	*Batch removal, cell alignment and imputation*
SHARE-seq skin	*RNA + ATAC*	*Skin*	*-*	*Mouse*	*34,774*	*Biological conservation, cell alignment and imputation*
COVID19 CITE-seq	*RNA+ADT*	*PBMC*	*143 donors*	*Human*	*781,123*	*batch removal and paired and unpaired RNA+ADT Scalability,*
Lymph node spatial	*spatial+RNA+ADT*	*Lymph node*	*2 samples*	*Human*	*6843*	*Biological conservation and Batch removal*
Thymus spatial	*spatial+RNA+ADT*	*thymus*	*4 samples*	*Mouse*	*17,824*	*Batch removal*
Spleen SPOTS	*spatial+RNA+ADT*	*spleen*	*2 samples*	*Mouse*	*5,336*	*Batch removal*

3.4 an evaluation on the scalability of the different methods, would finally be highly important (e.g. running them on the full Neurips data)

Response:
Thanks for your suggestion! Our proposed study has included scalability analysis for all benchmark methods on different datasets, including the simulation datasets generated from NeurIPS data you mentioned. For all paired and unpaired integration methods, we will assess running time, memory consumption and GPU consumption. The scalability analysis is included in the usability section of our benchmark workflow. The detail of scalability analysis method is as following:

"The scalability metrics will evaluate whether the method is executable for distinct dataset sizes (Supplementary Table 2) with limited computation hardware resource. We will also evaluate the biological interpretability of latent space output for benchmarked methods. The scalability of all multi-modal integration tools will be assessed with CPU computation time, maximum memory use and GPU memory consumption. We will use Linux GNU time command to collect the computation time and memory usage, and apply nvitop package in python to collect maximum GPU usage for all. The scalability metrics will evaluate the availability of all methods with the limits of 500GB RAM, 24 hours running time and 24 GB GPU."

3.5 No bias can be identified in the proposed analysis.
The description is methodologically clear. In the final submission would be important to have a Jupyter notebook or similar reproducing all the performed analysis and helping the research community to evaluate new methods with respect to those proposed in this work.

Response:
Thanks again for your positive comment! For reproducibility, we will provide WDL workflow scripts for the execution of all benchmark methods, as well as Jupyter and RMarkdown notebook for reproducing all evaluation and analysis steps. We will also establish an interactive visualization website to access all benchmark results. We have described the reproducible proposal in our manuscript as follows:

"*Benchmark pipeline and interactive website construction*

The benchmark pipeline will be constructed using the Workflow Description Language (WDL; https://github.com/openwdl/wdl). All benchmark algorithms will be encapsulated as modular tasks, enhancing both the extensibility and reproducibility of the pipeline. The entire WDL pipeline and the necessary environment dependencies for execution will be available at GitHub repository (https://github.com/bm2-lab/SCMMI_Benchmark). We will provide a detailed manual to ensure users can apply the benchmarking process to new datasets or integrate new methods seamlessly.

To improve the accessibility of our benchmark study, we will establish an interactive visualization website. Static figures will be created with ggplot2 (v.3.4.0) and interactive figures with plotly (v.4.10.2) in R (v.4.3.1). The HTML web pages will be generated through the rmarkdown package (v.2.3) in R (v.4.3.1).
"

Editor:
Please address all the concerns raised by Reviewer 3, including adding the new evaluation/analysis suggested by this reviewer.

Response:
Thanks! We have addressed all concerns from Reviewer 3 in our revised manuscript.

Reference

1. Lee MYY, Kaestner KH, Li M. Benchmarking algorithms for joint integration of unpaired and paired single-cell RNA-seq and ATAC-seq data. *Genome Biology* **24**, 244 (2023).

2. Luecken MD*, et al.* A sandbox for prediction and integration of DNA, RNA, and proteins in single cells. In: *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*) (2021).

3. Ma S*, et al.* Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell* **183**, 1103-1116 e1120 (2020).

4. Ben-Chetrit N*, et al.* Integration of whole transcriptome spatial profiling with protein markers. *Nat Biotechnol* **41**, 788-793 (2023).

5. Long Y*, et al.* Integrated analysis of spatial multi-omics with SpatialGlue. *bioRxiv*, 2023.2004.2026.538404 (2023).

6. Hao Y*, et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* **42**, 293-304 (2024).

7. Ghazanfar S, Guibentif C, Marioni JC. Stabilized mosaic single-cell data integration using unshared features. *Nat Biotechnol* **42**, 284-292 (2024).

8. Long Y*, et al.* Deciphering spatial domains from spatial multi-omics with SpatialGlue.). Research Square (2023).

9. Bao F*, et al.* Integrative spatial analysis of cell morphologies and transcriptional states with MUSE. *Nat Biotechnol* **40**, 1200-1209 (2022).

10. Jain MS*, et al.* MultiMAP: dimensionality reduction and integration of multimodal data. *Genome Biol* **22**, 346 (2021).

11. Velten B*, et al.* Identifying temporal and spatial patterns of variation from multimodal data using MEFISTO. *Nat Methods* **19**, 179-186 (2022).

12. Argelaguet R*, et al.* Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol* **14**, e8124 (2018).

13. Korsunsky I*, et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289-1296 (2019).

14. STAligner enables the integration and alignment of multiple spatial transcriptomics datasets. *Nat Comput Sci* **3**, 831-832 (2023).

15. Lance C*, et al.* Multimodal single cell data integration challenge: Results and lessons learned. In: *Proceedings of the NeurIPS 2021 Competitions and Demonstrations Track* (eds Douwe K, Marco C, Barbara C). PMLR (2022).

16. He Z*, et al.* Mosaic integration and knowledge transfer of single-cell multimodal data with MIDAS. *Nat Biotechnol*, (2024).

17. Wagner F, Yan Y, Yanai I. K-nearest neighbor smoothing for high-throughput single-cell RNA-Seq data. *bioRxiv*, 217737 (2018).

18. Luecken MD*, et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nat Methods* **19**, 41-50 (2022).

19. Hao Y*, et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021).

20. Stuart T*, et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
The authors have appropriately addressed my previous concerns. I don't have any additional comments.

Response:
Thanks for your positive response! We really appreciate your comments and efforts to review our manuscript.

Reviewer #2:
Remarks to the Author:
In their revised proposal, the authors have addressed some of my initial comments. Some comments still require additional consideration:

1. Datasets proposed: this study is still very limited in the types of cells it proposes to use for benchmarking. Most of the datasets listed are immune. Since there is now plenty of published multimodal data available from a variety of tissues, the authors should aim to be more comprehensive in this aspect.

Response:
Thanks for your helpful comments!

To address the limitation of tissue types in our proposed benchmark, we schedule to append more high-quality single-cell multimodal benchmark datasets. **Of note, not all the available multimodal datasets are applicable for our benchmark and only those datasets with well-curated cell annotations unbiased to existing benchmark integration tools for bio-conservation evaluation or multiple batches for batch effect removal evaluations can be selected.** Specifically, we first collected publicly available single-cell multimodal datasets with modality of scRNA, scATAC or protein as many as possible, including 12 multimodal datasets with modality of scRNA and scATAC and 15 datasets with scRNA and ADT (listed in Table 1 and Table 2 below). Next, we selected datasets with well-curated and unbiased cell annotations or multiple batches (donors or replicates), and remained one dataset for each sample type. In total, we added two scRNA and scATAC multimodal datasets to our revised benchmark proposal, including tissue types of **immune, skin and brain** in all benchmark datasets. We also added three scRNA and ADT datasets to our revised proposal, which would involve tissue types of **immune, spleen, thymus, lung and kidney** in all benchmark datasets. Now we believed that the current curated datasets can serve as the most comprehensive multimodal datasets applicable for our benchmarking that includes different tissues types. The summary tables of collected public multimodal datasets and newly added datasets are listed as below:

Table 1. Collections of publicly available scRNA and scATAC multimodal datasets.

Id	Dataset name	Curated and unbiased cell annotation	Multiple batches	Other info	Source	Included in our study
1	Kaggle human BMMC	Yes	Yes	-	NeuraIPS 2021	Yes
2	Kaggle human HSPC	No	Yes	-	NeuraIPS 2022	Yes
3	SHARE-seq mouse skin	Yes	No	-	[1]	Yes
4	10X Genomics human PBMC	No	Yes	Multiple datasets	https://www.10xgenomics.com/datasets	Newly added
5	Wild type and Alzheimer's Disease Mouse Model Brain	No	Yes	-	https://www.10xgenomics.com/datasets	Newly added
6	Paired-seq adult and fetal mouse cerebral cortex	No	Yes	Duplicated sample types	[2]	No
7	sciCAR mouse kidney	No	No	-	[3]	No
8	SNARE-seq mouse cerebral cortex	No	Yes	Duplicated sample types	[4]	No
9	Fresh embryonic E18 mouse brain	No	Yes	Duplicated sample types	https://www.10xgenomics.com/datasets	No
10	Flash frozen human brain	No	Yes	Duplicated sample types	https://www.10xgenomics.com/datasets	No
11	Human kidney cancer	No	No	-	https://www.10xgenomic	No

					s.com/datasets	
12	Human lymph node	No	No	-	https://www.10xgenomics.com/datasets	No

Table 2 Collections of publicly available scRNA and ADT multimodal datasets.

Id	**Dataset name**	**Curated and unbiased cell annotation**	**Multiple batches**	**Other info**	**Source**	**Included in our study**
1	Kaggle human BMMC	Yes	Yes	-	NeuraIPS 2021	Yes
2	Kaggle human HSPC	No	Yes	-	NeuraIPS 2022	Yes
3	COVID-19 PBMC	No	Yes	-	[5]	Yes
4	Human lymph node spatial	Partial	Yes	-	[6]	Yes
5	Mouse thymus spatial	No	Yes	-	[6]	Yes
6	Mouse spleen spatial	No	Yes	-	[6]	Yes
7	Human white blood cells	No	Yes	-	[7]	Newly added
8	20k Mixture of NSCLC DTCs from 7 donors	No	Yes	-	https://www.10xgenomics.com/datasets	Newly added
9	Kidney cancer in 40k Mixture of	No	Yes	-	https://www.10xgenomics.com/datasets	Newly added

	Dissociated Tumor Cells					
10	Human PBMCs	No	No	-	[7]	No
11	Human PBMCs	No	No	-	[8]	No
12	ASAP CITE-seq PBMC	No	Yes	Duplicated in sample types	[9]	No
13	human bone marrow	No	No	-	[10]	No
14	30k Cells from 4 dissociated lung tumors	No	No	-	https://www.10xgenomics.com/datasets	No
15	10k Cells from a MALT tumor	No	No	-	https://www.10xgenomics.com/datasets	No

Also, we have appended detailed descriptions for newly added datasets in our revised manuscript as follows:

"10X PBMC dataset

The 10X PBMC dataset is merged from 10k PBMC Multiome RNA + ATAC dataset (https://www.10xgenomics.com/datasets/pbmc-from-a-healthy-donor-no-cell-sorting-10-k-1-standard-2-0-0) and 3k PBMC Multiome RNA + ATAC dataset (https://www.10xgenomics.com/datasets/pbmc-from-a-healthy-donor-no-cell-sorting-3-k-1-standard-2-0-0), including 15,021 cells in total. This dataset will be available for evaluating batch removal between two datasets, as well as cell alignment accuracy evaluation.

10X mouse brain dataset

10X mouse brain dataset is a single cell Multiome RNA + ATAC dataset for brain coronal sections from one hemisphere of Alzheimer's disease mouse and wild type mouse from 10X genomics website (https://www.10xgenomics.com/datasets/multiomic-integration-neuroscience-application-note-single-cell-multiome-rna-atac-alzheimers-disease-mouse-model-brain-coronal-sections-from-one-hemisphere-over-a-time-course-1-standard). We will use two replicates for both wild type mouse and Alzheimer's disease mouse at time point of 2.5 months as our benchmark dataset. This dataset will be available for batch removal evaluation and cell alignment accuracy evaluation.

Human WBC dataset

Human WBC dataset [7] is a human white blood cell (WBC) CITE-seq dataset with 24 samples from 8 donors and 3 time points, which will be available for batch effect removal evaluation. This dataset is annotated by Seurat v4 package [7], which will not apply for evaluation with biological conservation metrics.

10X NSCLC dataset

10X NSCLC dataset is a CITE-seq dataset of non-small cell lung cancer (NSCLC) dissociated tumor cells from 10X Genomics website (https://www.10xgenomics.com/datasets/20k-mixture-of-nsclc-dtcs-from-7-donors-3-v3-1-with-intronic-reads-3-1-standard). This dataset includes 7 donors, and will be available for batch removal evaluation.

10X kidney cancer dataset

10X kidney cancer dataset is a subset of dissociate tumor cells mixture dataset from 10X Genomics website (https://www.10xgenomics.com/datasets/40k-mixture-of-dissociated-tumor-cells-from-3-donors-stained-with-totalseqc-antibodies). We will use two replicates of kidney cancer samples for batch removal evaluation.

“

2. Treatment of the scATAC data: the authors propose to use different methods of creating the gene activity matrix (GeneActivity vs CreateGeneActivityMatrix functions) depending on dataset. All the scATAC-seq datasets used should have fragment files available, and so the GeneActivity function should be used uniformly. If there are cases where a fragment file is not available, this can be constructed from the BAM file. The GeneActivity function is also incorrectly attributed to the Seurat package in the revised text.

Response:
Thank you for your comments!

Among proposed benchmark datasets, HSPC and BMMC Multiome datasets did not have scATAC fragment files in their public links or websites. We have requested the scATAC fragment files of both HSPC and BMMC datasets from the contacting author of the datasets, but still wait for response. If we are not able to obtain BMMC scATAC fragment files, we will generate fragments files from raw SRA data, including using NCBI SRA-toolkit [11] to generate fastq files and cell ranger pipeline [12] to generate the scATAC BAM files as well as fragment files. However, for HSPC datasets, due to the absence of both raw sequencing files and processed scATAC fragment files, we could only generate the GAM with CreateGeneActivityMatrix rather than GeneActivity in Signac, considering that we might not be available for the scATAC fragment files. Furthermore, as we found high resource computational consuming during testing of the above two functions, we will also evaluate the scalability (time and memory) for both the GeneActivity and CreateGeneActivityMatrix functions to report the potential

computational resource requirements in different GAM preparation steps for single-cell multimodal integration.

Also, two newly added scRNA and scATAC multimodal datasets are available with scATAC fragment files, and would be available to calculate GAM with GeneActivity functions in Signac package [13].

Additionally, the incorrect attributions of the Seurat package would been revised to Signac package [13] in our revised stage 1 manuscript.

The scATAC data GAM calculation and evaluation methods for unpaired diagonal integration task are revised in our stage 1 manuscript as follows:

"For unpaired diagonal integration algorithms, which often involve transforming ATAC chromatin accessibility matrices into gene activity matrices (GAM), we will compute the GAM using the GeneActivity function in Signac package [13] using the ATAC per fragment information file. For datasets with raw SRA sequencing data only, we will first apply NCBI SRA-toolkit to generate fastq files and cell ranger ARC pipeline (10x Genomics; v2.0.2) pipeline to generate the ATAC fragment information file. Then we will compute the GAM using the GeneActivity function with the ATAC per fragment information file. For certain scATAC datasets with neither ATAC per fragment information file nor raw sequencing file, we will use CreateGeneActivityMatrix function within the Seurat v3 package [7] to compute GAM from the scATAC accessibility profile."

***"**For unpaired diagonal integration evaluation, we will also evaluate the scalability for the GAM preparation step with GeneActivity [13] and CreateGeneActivityMatrix [7] function, as the generation of GAM will also consume long time and high memory for large scRNA and scATAC multimodal datasets."*

3. Weighting of evaluation metrics: I again want to reiterate my previous comment that an overall ranked list that combines bio conservation and batch mixing in an arbitrarily weighted manner may not be the most useful thing to report. I recognize that 60/40 weighting was proposed in the scIB paper, but I disagree that this should be interpreted as an accepted standard in the field. Rather than trying to condense all the evaluation metrics into a single number, the authors can report the performance of each method across each of the different categories (bio conservation, etc.)

Response:
Thank you for your points! We agree that 60/40 should not be interpreted as an accepted standard in the field of single cell benchmarking, and aggregation of all evaluation metrics would lead to vague interpretation of benchmark results. Following your suggestion, we will report the rank/performance of each method for all metrics in

each category of metrics separately. Then for a comparison purpose, we will also take the average of metrics within one category (bio conservation, for etc.) to report the overall evaluation in terms of a specific metric category. We will not combine the rank among different categories in our proposed benchmark study. We have removed methods descriptions related to 60/40 weighting and revised our manuscript as follows:

"Finally, we will report the original metric scores and average rankings of biological conservation, batch effect removal and cell alignment accuracy for each method respectively."

4. One previous comment was not addressed: some statements are too vague, eg " For deep generative models with posterior imputation, we will also assess whether the imputation function can effectively mitigate data dropout and enhance downstream data analysis". How will enhanced downstream analysis be evaluated in this case? I also caution against use of the term "dropout" as there is some debate whether scRNA-seq is zero-inflated. Instead terms like "sparsity" can be used.

Response:
Thanks for your point! You are correct. Actually, we will use different data dropout rates to simulate the various conditions of data sparsity. We have revised the term "data dropout" to "data sparsity" in our revised proposal. Also, this vague sentence has been deleted in the current version. For the imputation analysis, we will only evaluate the effectiveness in mitigating data sparsity using the scRNA and scATAC profiles as ground truth. We will not assess the imputation function in the downstream data analysis. We did not apply the imputation downstream analysis with the following considerations. Firstly, the mitigation of the data sparsity would subsequently lead to the improvement in imputation downstream data analysis, indicating the potential redundancy for imputation downstream analysis evaluation after data sparsity mitigation evaluation. Secondly, the ground truth for imputation downstream evaluation would have lower confidence than data sparsity mitigation evaluation. The downstream analysis of differential expression would use cell type dependent markers as ground truth, which would be influence single-cell cell cluster and cell type annotation. The imputation downstream analysis would also use bulk sequencing data as ground truth, which would neglect the high resolution in single-cell datasets.

The statements of data sparsity and imputation analysis in the revised stage 1 manuscript are as follows:

"*For paired and mosaic integration methods with imputation function, we will also assess whether the posterior imputation can effectively mitigate data sparsity in simulated high dropout datasets or impute missing modalities in pseudo mosaic*

datasets. And the accuracy of posterior imputation will be evaluated by the raw profile or KNN imputed profile of the modality. "

Reference

1. Ma S, Zhang B, LaFave LM, et al: Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell* 2020, 183:1103-1116 e1120.
2. Zhu C, Yu M, Huang H, Juric I, Abnousi A, Hu R, Lucero J, Behrens MM, Hu M, Ren B: An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. *Nature Structural & Molecular Biology* 2019, 26:1063-1070.
3. Cao J, Cusanovich DA, Ramani V, et al: Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* 2018, 361:1380-1385.
4. Chen S, Lake BB, Zhang K: High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nature Biotechnology* 2019, 37:1452-1457.
5. Stephenson E, Reynolds G, Botting RA, et al: Single-cell multi-omics analysis of the immune response in COVID-19. *Nat Med* 2021, 27:904-916.
6. Long Y, Ang KS, Liao S, et al: Integrated analysis of spatial multi-omics with SpatialGlue. *bioRxiv* 2023:2023.2004.2026.538404.
7. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y, Stoeckius M, Smibert P, Satija R: Comprehensive Integration of Single-Cell Data. *Cell* 2019, 177:1888-1902 e1821.
8. Hao Y, Hao S, Andersen-Nissen E, et al: Integrated analysis of multimodal single-cell data. *Cell* 2021, 184:3573-3587 e3529.
9. Mimitou EP, Lareau CA, Chen KY, et al: Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat Biotechnol* 2021, 39:1246-1258.
10. Triana S, Vonficht D, Jopp-Saile L, et al: Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. *Nat Immunol* 2021, 22:1577-1589.
11. NCBI SRA-Tools. https://github.com/ncbi/sra-tools
12. Cell Ranger pipeline. https://www.10xgenomics.com/cn/support/software/cell-ranger/latest
13. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nature Methods* 2021 18, 1333-1341.