Improving the efficiency of single-cell genome sequencing based on overlapping pooling strategy and CNV analysis

Jing Tu, Zengyan Yang, Na Lu and Zuhong Lu

Article citation details
R. Soc. open sci. 9: 211330.
http://dx.doi.org/10.1098/rsos.211330

Review timeline
Original submission: 21 August 2021
Revised submission: 7 December 2021
Final acceptance: 4 January 2022

Note: Reports are unedited and appear as submitted by the referee. The review history appears in chronological order.

Review History
RSOS-211330.R0 (Original submission)

Review form: Reviewer 1

Is the manuscript scientifically sound in its present form?
No

Are the interpretations and conclusions justified by the results?
Yes

Is the language acceptable?
Yes

Do you have any ethical concerns with this paper?
No

Have you any concerns about statistical analyses in this paper?
No
Recommendation?
Major revision is needed (please make suggestions in comments)

Comments to the Author(s)
The authors submitted a manuscript reporting a study assorting cells into subclones using the overlapping pooling method without constructing libraries for each cell. The aim of the study is novel, however, it looks the results did not fully support the conclusions.

Here are my comments:
1. In the simulation, the legitimacy of using 80 cells should be stated. Is that the conventional number of cells used in any single cell genome study?
2. Could bootstrap be performed with more repeats in your simulation of the pooling library? It is unclear whether reads in the pooling libraries from the actual experiment are in the same pattern as simulated?
3. Results should be succinct and simple. There is too much jargon. Section 4.2 is like method
4. Legend of figure 2 should be more descriptive.

Review form: Reviewer 2

Is the manuscript scientifically sound in its present form?
Yes

Are the interpretations and conclusions justified by the results?
Yes

Is the language acceptable?
Yes

Do you have any ethical concerns with this paper?
No

Have you any concerns about statistical analyses in this paper?
No

Recommendation?
Accept with minor revision (please list in comments)

Comments to the Author(s)
The authors employed branch and bound (B&B) algorithm to analyze CNV in the in silico overlapping pooling single cell sequencing data. In general, the study is well designed and performed. However, the following concerns need to be addressed.

1. The title of the manuscript is not specific and could mislead readers. The study was centered in CNV analysis of single cell sequencing. Suggest to include CNV analysis in the title.
2. Language need to polished. Some sentences are difficult to read and confusing. Grammar errors present.
3. Pooling parament k seems affect the restoration rate, however it was not discussed.
4. A significant discrepancy of restoration rate has been observed on figure 3 between p=40 and p=27 when k=10%. Readers might be interested in more discussion.
Decision letter (RSOS-211330.R0)

We hope you are keeping well at this difficult and unusual time. We continue to value your support of the journal in these challenging circumstances. If Royal Society Open Science can assist you at all, please don't hesitate to let us know at the email address below.

Dear Dr Tu

On behalf of the Editors, we are pleased to inform you that your Manuscript RSOS-211330 "Improving the efficiency of single cell genome sequencing based on overlapping pooling strategy" has been accepted for publication in Royal Society Open Science subject to minor revision in accordance with the referees' reports. Please find the referees' comments along with any feedback from the Editors below my signature.

We invite you to respond to the comments and revise your manuscript. It will be important to address comprehensively all of the comments of both reviewers. Below the referees’ and Editors’ comments (where applicable) we provide additional requirements. Final acceptance of your manuscript is dependent on these requirements being met. We provide guidance below to help you prepare your revision.

Please submit your revised manuscript and required files (see below) no later than 7 days from today's (ie 29-Nov-2021) date. Note: the ScholarOne system will 'lock' if submission of the revision is attempted 7 or more days after the deadline. If you do not think you will be able to meet this deadline please contact the editorial office immediately.

Please note article processing charges apply to papers accepted for publication in Royal Society Open Science (https://royalsocietypublishing.org/rsos/charges). Charges will also apply to papers transferred to the journal from other Royal Society Publishing journals, as well as papers submitted as part of our collaboration with the Royal Society of Chemistry (https://royalsocietypublishing.org/rsos/chemistry). Fee waivers are available but must be requested when you submit your revision (https://royalsocietypublishing.org/rsos/waivers).

Thank you for submitting your manuscript to Royal Society Open Science and we look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Kind regards,
Royal Society Open Science Editorial Office
Royal Society Open Science
openscience@royalsociety.org

on behalf of Dr Paul Schofield (Associate Editor) and Steve Brown (Subject Editor)
openscience@royalsociety.org

Associate Editor Comments to Author (Dr Paul Schofield):
Associate Editor: 1
Comments to the Author:
The referees have suggested several issues that require some further comments or discussion, and while neither indicated that more work needs to be done, these comments should be addressed in the resubmitted manuscript. Of more concern is the comment from both referees that the text is somewhat opaque and difficult to follow in many places. We request that you address these concerns carefully and in particular pay attention to terms, acronyms and algorithms that might be familiar to those working in precisely this area but not more general readers. The paper should
be understandable by a general reader with an interest in the subject without recourse to looking up citations to understand the text of the manuscript. Overall the language in the paper should be looked at carefully for grammar and expression.

Reviewer comments to Author:
Reviewer: 1
Comments to the Author(s)
The authors submitted a manuscript reporting a study assorting cells into subclones using the overlapping pooling method without constructing libraries for each cell. The aim of the study is novel, however, it looks the results did not fully support the conclusions.

Here are my comments:
1. In the simulation, the legitimacy of using 80 cells should be stated. Is that the conventional number of cells used in any single cell genome study?
2. Could bootstrap be performed with more repeats in your simulation of the pooling library? It is unclear whether reads in the pooling libraries from the actual experiment are in the same pattern as simulated?
3. Results should be succinct and simple. There is too much jargon. Section 4.2 is like method
4. Legend of figure 2 should be more descriptive.

Reviewer: 2
Comments to the Author(s)
The authors employed branch and bound (B&B) algorithm to analyze CNV in the in silico overlapping pooling single cell sequencing data. In general, the study is well designed and performed. However, the following concerns need to be addressed.

1. The title of the manuscript is not specific and could mislead readers. The study was centered in CNV analysis of single cell sequencing. Suggest to include CNV analysis in the title.
2. Language need to polished. Some sentences are difficult to read and confusing. Grammar errors present.
3. Pooling parament k seems affect the restoration rate, however it was not discussed.
4. A significant discrepancy of restoration rate has been observed on figure 3 between p=40 and p=27 when k=10%. Readers might be interested in more discussion.

===PREPARING YOUR MANUSCRIPT===

Your revised paper should include the changes requested by the referees and Editors of your manuscript.

You should provide two versions of this manuscript and both versions must be provided in an editable format:
one version should clearly identify all the changes that have been made (for instance, in coloured highlight, in bold text, or tracked changes);
a 'clean' version of the new manuscript that incorporates the changes made, but does not highlight them. This version will be used for typesetting.

Please ensure that any equations included in the paper are editable text and not embedded images.

Please ensure that you include an acknowledgements' section before your reference list/bibliography. This should acknowledge anyone who assisted with your work, but does not qualify as an author per the guidelines at https://royalsociety.org/journals/ethics-policies/openness/.
While not essential, it will speed up the preparation of your manuscript proof if you format your references/bibliography in Vancouver style (please see https://royalsociety.org/journals/authors/author-guidelines/#formatting). You should include DOIs for as many of the references as possible.

If you have been asked to revise the written English in your submission as a condition of publication, you must do so, and you are expected to provide evidence that you have received language editing support. The journal would prefer that you use a professional language editing service and provide a certificate of editing, but a signed letter from a colleague who is a proficient user of English is acceptable. Note the journal has arranged a number of discounts for authors using professional language editing services (https://royalsociety.org/journals/authors/benefits/language-editing/).

---PREPARING YOUR REVISION IN SCHOLARONE---

To revise your manuscript, log into https://mc.manuscriptcentral.com/rsos and enter your Author Centre - this may be accessed by clicking on "Author" in the dark toolbar at the top of the page (just below the journal name). You will find your manuscript listed under "Manuscripts with Decisions". Under "Actions", click on "Create a Revision".

Attach your point-by-point response to referees and Editors at the 'View and respond to decision letter' step. This document should be uploaded in an editable file type (.doc or .docx are preferred). This is essential, and your manuscript will be returned to you if you do not provide it.

Please ensure that you include a summary of your paper at the 'Type, Title, & Abstract' step. This should be no more than 100 words to explain to a non-scientific audience the key findings of your research. This will be included in a weekly highlights email circulated by the Royal Society press office to national UK, international, and scientific news outlets to promote your work. An effective summary can substantially increase the readership of your paper.

At the 'File upload' step you should include the following files:
-- Your revised manuscript in editable file format (.doc, .docx, or .tex preferred). You should upload two versions:
  1) One version identifying all the changes that have been made (for instance, in coloured highlight, in bold text, or tracked changes);
  2) A 'clean' version of the new manuscript that incorporates the changes made, but does not highlight them.
-- An individual file of each figure (EPS or print-quality PDF preferred [either format should be produced directly from original creation package], or original software format).
-- An editable file of each table (.doc, .docx, .xls, .xlsx, or .csv).
-- An editable file of all figure and table captions.
Note: you may upload the figure, table, and caption files in a single Zip folder.
-- Any electronic supplementary material (ESM).
-- If you are requesting a discretionary waiver for the article processing charge, the waiver form must be included at this step.
-- If you are providing image files for potential cover images, please upload these at this step, and inform the editorial office you have done so. You must hold the copyright to any image provided.
-- A copy of your point-by-point response to referees and Editors. This will expedite the preparation of your proof.

At the 'Details & comments' step, you should review and respond to the queries on the electronic submission form. In particular, we would ask that you do the following:
-- Ensure that your data access statement meets the requirements at https://royalsociety.org/journals/authors/author-guidelines/#data. You should ensure that you cite the dataset in your reference list. If you have deposited data etc in the Dryad repository, please only include the 'For publication' link at this stage. You should remove the 'For review' link.
-- If you are requesting an article processing charge waiver, you must select the relevant waiver option (if requesting a discretionary waiver, the form should have been uploaded, see 'File upload' above).
-- If you have uploaded any electronic supplementary (ESM) files, please ensure you follow the guidance at https://royalsociety.org/journals/authors/author-guidelines/#supplementary-material to include a suitable title and informative caption. An example of appropriate titling and captioning may be found at https://figshare.com/articles/Table_S2_from_Is_there_a_trade-off_between_peak_performance_and_performance_breadth_across_temperatures_for_aerobic_scope_in_teleost_fishes_/3843624.

At the 'Review & submit' step, you must view the PDF proof of the manuscript before you will be able to submit the revision. Note: if any parts of the electronic submission form have not been completed, these will be noted by red message boxes - you will need to resolve these errors before you can submit the revision.

Author's Response to Decision Letter for (RSOS-211330.R0)

See Appendix A.

RSOS-211330.R1 (Revision)

Review form: Reviewer 1

Is the manuscript scientifically sound in its present form?
No

Are the interpretations and conclusions justified by the results?
Yes

Is the language acceptable?
Yes

Do you have any ethical concerns with this paper?
No

Have you any concerns about statistical analyses in this paper?
Yes

Recommendation?
Accept as is

Comments to the Author(s)
The manuscript is improved
Review form: Reviewer 2

Is the manuscript scientifically sound in its present form?
Yes

Are the interpretations and conclusions justified by the results?
Yes

Is the language acceptable?
Yes

Do you have any ethical concerns with this paper?
No

Have you any concerns about statistical analyses in this paper?
No

Recommendation?
Accept as is

Comments to the Author(s)
My comments on the original submission has been addressed. No additional comments. I recommend acceptance of this manuscript.

Decision letter (RSOS-211330.R1)

We hope you are keeping well at this difficult and unusual time. We continue to value your support of the journal in these challenging circumstances. If Royal Society Open Science can assist you at all, please don't hesitate to let us know at the email address below.

Dear Dr Tu,

It is a pleasure to accept your manuscript entitled "Improving the efficiency of single cell genome sequencing based on overlapping pooling strategy and CNV analysis" in its current form for publication in Royal Society Open Science. The comments of the reviewer(s) who reviewed your manuscript are included at the foot of this letter.

If you have not already done so, please ensure that you send to the editorial office an editable version of your accepted manuscript, and individual files for each figure and table included in your manuscript. You can send these in a zip folder if more convenient. Failure to provide these files may delay the processing of your proof.

Please remember to make any data sets or code libraries 'live' prior to publication, and update any links as needed when you receive a proof to check - for instance, from a private 'for review' URL to a publicly accessible 'for publication' URL. It is good practice to also add data sets, code and other digital materials to your reference list.

Our payments team will be in touch shortly if you are required to pay a fee for the publication of the paper (if you have any queries regarding fees, please see https://royalsocietypublishing.org/rsos/charges or contact authorfees@royalsociety.org).
The proof of your paper will be available for review using the Royal Society online proofing system and you will receive details of how to access this in the near future from our production office (openscience_proofs@royalsociety.org). We aim to maintain rapid times to publication after acceptance of your manuscript and we would ask you to please contact both the production office and editorial office if you are likely to be away from e-mail contact to minimise delays to publication. If you are going to be away, please nominate a co-author (if available) to manage the proofing process, and ensure they are copied into your email to the journal.

Please see the Royal Society Publishing guidance on how you may share your accepted author manuscript at https://royalsociety.org/journals/ethics-policies/media-embargo/. After publication, some additional ways to effectively promote your article can also be found here https://royalsociety.org/blog/2020/07/promoting-your-latest-paper-and-tracking-your-results/.

Thank you for your fine contribution. On behalf of the Editors of Royal Society Open Science, we look forward to your continued contributions to the Journal.

Kind regards,
Royal Society Open Science Editorial Office
Royal Society Open Science
openscience@royalsociety.org

on behalf of Dr Paul Schofield (Associate Editor) and Steve Brown (Subject Editor)
openscience@royalsociety.org

Reviewer comments to Author:
Reviewer: 2
Comments to the Author(s)
My comments on the original submission has been addressed. No additional comments. I recommend acceptance of this manuscript.

Reviewer: 1
Comments to the Author(s)
The manuscript is improved

Follow Royal Society Publishing on Twitter: @RSocPublishing
Follow Royal Society Publishing on Facebook:
https://www.facebook.com/RoyalSocietyPublishing.FanPage/
Read Royal Society Publishing’s blog:
https://royalsociety.org/blog/blogsearchpage/?category=Publishing
Associate Editor Comments to Author (Dr Paul Schofield):
Associate Editor: 1
Comments to the Author:
The referees have suggested several issues that require some further comments or discussion, and while neither indicated that more work needs to be done, these comments should be addressed in the resubmitted manuscript. Of more concern is the comment from both referees that the text is somewhat opaque and difficult to follow in many places. We request that you address these concerns carefully and in particular pay attention to terms, acronyms and algorithms that might be familiar to those working in precisely this area but not more general readers. The paper should be understandable by a general reader with an interest in the subject without recourse to looking up citations to understand the text of the manuscript. Overall the language in the paper should be looked at carefully for grammar and expression.

Response: Thanks for your comments. We revised our manuscript according to your comments and reviewers’ comments. Point-by-point response to the reviewers’ comment is listed below. We have done our best to make the pipeline clear and easy to follow. Meanwhile, the language has been polished carefully for grammar and expression.

Appendix A
Reviewer: 1
Comments to the Author(s)

The authors submitted a manuscript reporting a study assorting cells into subclones using the overlapping pooling method without constructing libraries for each cell. The aim of the study is novel; however, it looks the results did not fully support the conclusions.

Here are my comments:

1. In the simulation, the legitimacy of using 80 cells should be stated. Is that the conventional number of cells used in any single cell genome study?

Response: Thanks. Actually, the number of single cells used in single cell genome study is uncertain, for example, 47 single cells (10.1093/bib/bbx004), 100 single cells (10.1038/nprot.2012.039), 900 single cells (10.1016/j.cell.2018.03.041) and over 10 thousand single cells (10.1038/nmeth.4154). In addition, we downloaded all single cell sequencing data of the 7 tumor patients from this work (10.1038/ng.3641). Then, 40 cells, 80 cells and 140 cells were randomly selected from the downloaded single cells. Considering the analysis efficiency and the accuracy of result, finally, we used 80 single cells for showing in this paper.

2. Could bootstrap be performed with more repeats in your simulation of the pooling library? It is unclear whether reads in the pooling libraries from the actual experiment are in the same pattern as simulated?

Response: Thanks. First, all the single cell sequencing data used in this paper were downloaded from the articles published on Nature Genetics. The sequencing data is reliable, and the number of reads in the sequencing data of each single cell is different. Second, we have also made settings for the number of input single cell and have done pool simulation for different single cell number. This is a proof-of-concept study, so we did not perform specific experiments to verify this method. But our simulation intersects specific experiments, which can get the real authenticity and accuracy of the results.

3. Results should be succinct and simple. There is too much jargon. Section 4.2 is like method

Response: Thanks for your reminding, we moved the part of “Pooling pattern matrix designation” to method section. And the part of “Pooling pattern matrix solution” belongs to the result part, so we stayed in place.

4. Legend of figure 2 should be more descriptive.

Response: Thanks for your reminding, we added more descriptive and clearer explanation of the Figure 2 as following. 2D-tSNE representation of all single cells included in the study (n = 80). Hierarchical clustering of all single cells on the basis of their CNV patterns. Cell clusters are differentially colored and identified as distinct TNBC patients. All single cells were divided into 7 distinct TNBC patients by hierarchical clustering.
Reviewer: 2
Comments to the Author(s)
The authors employed branch and bound (B&B) algorithm to analyze CNV in the in silico overlapping pooling single cell sequencing data. In general, the study is well designed and performed. However, the following concerns need to be addressed.

1. The title of the manuscript is not specific and could mislead readers. The study was centered in CNV analysis of single cell sequencing. Suggest to include CNV analysis in the title.
   **Response:** Thanks for your reminding, we have added CNV analysis to our title which is modified to “Improving the efficiency of single cell genome sequencing based on overlapping pooling strategy and CNV analysis” now.

2. Language need to polished. Some sentences are difficult to read and confusing. Grammar errors present.
   **Response:** Thanks. We polished our language and made some modifications to the sentences.

3. Pooling parameter $k$ seems affect the restoration rate, however it was not discussed.

4. A significant discrepancy of restoration rate has been observed on figure 3 between $p=40$ and $p=27$ when $k=10\%$. Readers might be interested in more discussion.
   **Response:** Thanks for your comments. As the two comments are all concern with the cell percent $k$, we response them together. We added the results and discussion of the pooling parameter $k$. When the pool number was large ($p = 40$), the tests of smaller percentage of cell showed better restoration results. It is probably due to that more cells in a pool gets highly complex CNV results which might get more bias in calculating the approximate solution. However, when the pool number was 27, the tests of $k = 10\%$ got the worst restoration accuracy. The reason is probably due to that the information gathered in the tests of 27 pools $\times$ 10\% cells were not enough to support fully reconstructed for the initial matrix. It seems that enough information was gathered in the tests of 27 pools $\times$ 15\% cells as the restoration accuracy was much better. Meanwhile, an overall higher restoration rate be found in $p=40$ pools as there are more information provided. As for different $k$ (from 0.1 to 0.25) within the same $p=40$, we could see a negative correlation between restoration rate of cells and $k$. In our opinion, with the number of cells increased in a pool, CNV patterns among the cells conflict with each other so that the restoration is obscured. But when $k=30\%$, the preserved interesting biological structure trumps the influence of the chaos. Due to the feasibility in practical applications, we did not make more attempts for larger $k$. 