

Peer Review Information

Journal: Nature Cell Biology

Manuscript Title: Real-time and programmable transcriptome sequencing with PROFIT-seq

Corresponding author name(s): Dr Fangqing Zhao

Editorial Notes:

Reviewer Comments & Decisions:

Decision Letter, initial version:
--

*Please delete the link to your author homepage if you wish to forward this email to co-authors.

Dear Dr Zhao,

Your manuscript, "Real-time and programmable transcriptome sequencing with PROFIT-seq", has now been seen by 2 referees, who are experts in transcriptome sequencing (referee 1) and functional genomics (referee 2). Despite our best efforts, we have been unable to receive the comments of reviewer 3, who is an expert in RNA sequencing. If/when we do receive the missing comments, we will pass them on to you. As you will see from their comments (attached below) they find this work of potential interest, but have raised substantial concerns, which in our view would need to be addressed with considerable revisions before we can consider publication in Nature Cell Biology.

Nature Cell Biology editors discuss the referee reports in detail within the editorial team, including the chief editor, to identify key referee points that should be addressed with priority, and requests that are overruled as being beyond the scope of the current study. To guide the scope of the revisions, I have listed these points below. We are committed to providing a fair and constructive peer-review process, so please feel free to contact me if you would like to discuss any of the referee comments further.

In particular, it would be essential to:

A- Clarify details of the approach, assays and analysis used, as per Reviewer#1 pt 2, 4 and Reviewer#2

B- experimentally address the point about enrichment levels (Reviewer#1 pt 3)

C- Address the potential artefacts generated by the RCA approach, elaborate on any strategies to overcome the issue of artefacts and potential for the method to generate false chimerics (Reviewer#2)

D- Ensure all sequencing data is publicly available (the HRA003930 dataset is not currently publicly accessible) and custom code generated is made accessible. Ensure that all human participant data (with patient identifiers) is diasaggregated in the manuscript and follows our guidelines:

<https://www.nature.com/nature-portfolio/editorial-policies/ethics-and-biosecurity#Research-with-human-participants-their-data-or-biological-material>

E-All other referee concerns pertaining to strengthening existing data, providing controls, methodological details, clarifications and textual changes, [EDITOR DELETE AS APPROPRIATE] should also be addressed.

F-Finally please pay close attention to our guidelines on statistical and methodological reporting (listed below) as failure to do so may delay the reconsideration of the revised manuscript. In particular please provide:

- a Supplementary Figure including unprocessed images of all gels/blots in the form of a multi-page pdf file. Please ensure that blots/gels are labeled and the sections presented in the figures are clearly indicated.

- a Supplementary Table including all numerical source data in Excel format, with data for different figures provided as different sheets within a single Excel file. The file should include source data giving rise to graphical representations and statistical descriptions in the paper and for all instances where the figures present representative experiments of multiple independent repeats, the source data of all repeats should be provided.

We would be happy to consider a revised manuscript that would satisfactorily address these points, unless a similar paper is published elsewhere, or is accepted for publication in Nature Cell Biology in the meantime.

When revising the manuscript please:

- ensure that it conforms to our format instructions and publication policies (see below and www.nature.com/nature/authors/).
- provide a point-by-point rebuttal to the full referee reports verbatim, as provided at the end of this letter.
- provide the completed Editorial Policy Checklist (found here <https://www.nature.com/authors/policies/Policy.pdf>), and Reporting Summary (found here <https://www.nature.com/authors/policies/ReportingSummary.pdf>). This is essential for reconsideration of the manuscript and these documents will be available to editors and referees in the event of peer review. For more information see <http://www.nature.com/authors/policies/availability.html> or contact me.

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

Please submit the revised manuscript files and the point-by-point rebuttal to the referee comments using this link:

[Redacted]

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

We would like to receive a revised submission within six months. We would be happy to consider a revision even after this timeframe, however if the resubmission deadline is missed and the paper is eventually published, the submission date will be the date when the revised manuscript was received.

We hope that you will find our referees' comments, and editorial guidance helpful. Please do not hesitate to contact me if there is anything you would like to discuss.

Best wishes,

Sabrya Carim

Sabrya Carim, PhD
(she/her/hers)
Associate Editor, Nature Cell Biology
Nature Portfolio

Springer Nature
The Campus, 4 Crinan Street, London N1 9XW, UK
sabrya.carim@springernature.com
<https://orcid.org/0000-0001-9485-1938>

Reviewers' Comments:

Reviewer #1:

Remarks to the Author:

In the manuscript entitled “Real-time and programmable transcriptome sequencing with PROFIT-seq” Zhang et al. introduce the PROFIT-seq method which involves the adaptive sampling of transcriptome libraries to enrich genes of interest. The authors use a clever approach to get around the limitations of adaptive sampling and showcase their approach for several applications.

First, they test their web-based implementation of the adaptive sampling API on run simulations. Second they characterize their modified RT priming strategy on HeLa cell RNA. Third, they validate the combined priming and adaptive sampling approach on colorectal polyp samples which they enrich for a panel of cancer relevant genes. Fourth, they establish a custom Salmon-based quantification approach. Fifth, they use a different priming approach on sputum samples to characterize PROFIT-seq for pathogen detection - SARS-CoV-2 and pneumonia. Sixth, they implemented a 100Mb panel of cancer genes which they applied to a group of colorectal polyp samples.

I generally like the approach and I believe the authors achieve some enrichment of target gene transcripts with their adaptive sampling approach. However, as with all adaptive sampling approaches, the enrichment levels are modest compared with probe-based enrichment approaches. On the other hand, they are “free” as in they require only computational, not experimental interventions.

I also think the authors are doing too much here causing the manuscript’s cohesiveness and clarity to suffer. There is the priming approach, the enrichment approach and the EM quantification approach all applied to several scenarios.

I think the pathogen detection approach in particular is an unnecessary detour especially since it uses a different priming strategy altogether. I think that part would be better suited for its own manuscript.

There are other major and minor issues as well that I will outline below:

1) The authors cite but don't name the R2C2 approach which they use throughout the entire manuscript. They refer to an "RCA assay" but use a molecular biology protocol that is identical to R2C2, they use oligo sequences like the splint sequence directly from R2C2, and their analysis approach sounds very much like the C3POa analysis pipeline associated with R2C2.

To be clear, there is nothing wrong with using R2C2 in the way the authors did. I think the way the concatemerization in the R2C2 protocol is used to circumvent adaptive sampling is clever.

I just think that attribution is important. Name the method.

2) If C3POa was used for analysis, the C3POa github should be linked. If not, the analysis code should be posted publicly and results (like output and accuracy) be compared to C3POa.

3) Enrichment levels.

The authors state that "72.89% of genes were successfully enriched". The way "enriched" is used here, without any treatment, you would expect 50% of genes to be successfully "enriched". What this really means in this case is that 37% of genes are covered less with the PROFIT-seq approach compared to the control run. This aligns with the rest of the manuscript - enrichment is just very low.

Additionally, runs were limited to less than 36 hours, even though MinION flow cells can produce data for 48 hours especially with reloads. Were runs reloaded? And is the complete run duration shown in all plots. It is quite possible that after reloads, the control runs would generate more on-target reads than the PROFIT-seq runs.

I think to address this the authors should a) accompany all on-target read plots with total-read plots, and b) show data for full run durations if flow cells were run longer than the plots indicate

4) The combinatorial priming strategy.

A single strand random hexamer primer will bind any RNA molecule, not just circular, even when added after the other primers. Any given round of priming and RT isn't 100% effective. I am therefore surprised by the outcome of the RT reactions. Only 3% of reads are classified as coming from non-polyA transcripts while 40% of reads lack a poly-A tail.

This points to the ds primer not really doing much while the ss primer largely primes polyA transcripts internally. However, S7B shows that the combinatorial priming approach is largely indistinguishable

from oligodT priming in terms of transcript coverage. Ultimately, this requires clarification. At the very least, polyA and non-poly reads should be separated for an analysis like S7B and for read length plots. It would have been very useful to delineate the products of the individual primers by adding index sequences to the primers themselves.

5) Sequencing coverage plots are too small to show differences.

Minor

1) Throughout the manuscript percent and fold-change are used to quantify enrichment. I recommend sticking with fold-change throughout.

2) “probe design that cannot capture different types of targets simultaneously.” I think this statement needs further explanation.

3) “However, after the adjustment by our quantification algorithm, the correlation was significantly improved ($R=0.733-0.717$)” The authors should probably lead with the smaller number.

4) What does the dashed line in 5C mean? I don’t think it’s explained in the figure legend.

Reviewer #2:

Remarks to the Author:

This study addresses the challenge of detecting specific transcripts within the diverse and complex eukaryotic transcriptome by introducing PROFIT-seq (programmable full-length isoform transcriptome sequencing).

Unlike current targeted RNA-seq methods, which involve complex pre-sequencing enrichment steps that limit the comprehensive characterisation of the transcriptome, PROFIT-seq enriches target transcripts while providing unbiased quantification of the entire transcriptome. The method utilises combinatorial reverse transcription for whole transcriptome capture and a programmable control system for selective enrichment during sequencing.

PROFIT-seq demonstrated significant improvements, achieving over a 300% increase in the yield of effective data and reducing the time required for detecting specific pathogens or key mutations by 75%. The method's application to colorectal polyp development revealed intricate associations between immune responses and bacterial infection, highlighting its potential in revealing complex biological

processes.

The authors present PROFIT-seq as an advanced method that ensures accurate and efficient sequencing of target transcripts, alongside unbiased quantification of both target and non-target transcripts. While I enjoyed reading the manuscript and the data was presented well, I have the following major comments:

Major issues:

- No UMIs and this is a major issue if you are using RCA for amplification. How quantitative is this assay given the known errors associated with ONT sequencing.
- Need to describe the potential artefacts generated by this approach - RCA can generate a lot of false chimeric artefacts which makes the analysis challenging, especially when investigating oncology samples.
- Have you implemented computational strategies to overcome the potential issues of artefacts? It doesn't seem to be described within your methods.
- To demonstrate that your method does not generate too many false chimerics, you need to do species mixing experiments and sequence human and mouse (or another species) and evaluate the effect of chimeric on the data analysis.

Submit your sequencing data to ncbi or ENA because the data is currently behind a controlled access and is not open to the wider research community and it is hosted on CNCB, a platform that is very challenging to work with as a computational biologist. I was not able to download your data because of this, therefore I was only able to perform a limited review of the data. Your data contains HeLa cells, so as a minimum this data does not need to be behind a controlled access agreement.

Reviewer #3:

None

GUIDELINES FOR MANUSCRIPT SUBMISSION TO NATURE CELL BIOLOGY

READABILITY OF MANUSCRIPTS – Nature Cell Biology is read by cell biologists from diverse backgrounds, many of whom are not native English speakers. Authors should aim to communicate their findings

clearly, explaining technical jargon that might be unfamiliar to non-specialists, and avoiding non-standard abbreviations. Titles and abstracts should concisely communicate the main findings of the study, and the background, rationale, results and conclusions should be clearly explained in the manuscript in a manner accessible to a broad cell biology audience. Nature Cell Biology uses British spelling.

MANUSCRIPT FORMAT – please follow the guidelines listed in our Guide to Authors regarding manuscript formats at Nature Cell Biology.

TITLE – should be no more than 100 characters including spaces, without punctuation and avoiding technical terms, abbreviations, and active verbs..

AUTHOR NAMES – should be given in full.

AUTHOR AFFILIATIONS – should be denoted with numerical superscripts (not symbols) preceding the names. Full addresses should be included, with US states in full and providing zip/post codes. The corresponding author is denoted by: "Correspondence should be addressed to [initials]."

ABSTRACT AND MAIN TEXT – please follow the guidelines that are specific to the format of your manuscript, as listed in our Guide to Authors (http://www.nature.com/ncb/pdf/ncb_gta.pdf) Briefly, Nature Cell Biology Articles, Resources and Technical Reports have 3500 words, including a 150 word abstract, and the main text is subdivided in Introduction, Results, and Discussion sections. Nature Cell Biology Letters have up to 2500 words, including a 180 word introductory paragraph (abstract), and the text is not subdivided in sections.

ACKNOWLEDGEMENTS – should be kept brief. Professional titles and affiliations are unnecessary. Grant numbers can be listed.

AUTHOR CONTRIBUTIONS – must be included after the Acknowledgements, detailing the contributions of each author to the paper (e.g. experimental work, project planning, data analysis etc.). Each author should be listed by his/her initials.

FINANCIAL AND NON-FINANCIAL COMPETING INTERESTS – the authors must include one of three declarations: (1) that they have no financial and non-financial competing interests; (2) that they have financial and non-financial competing interests; or (3) that they decline to respond, after the Author Contributions section. This statement will be published with the article, and in cases where financial and non-financial competing interests are declared, these will be itemized in a web supplement to the article. For further details please see <https://www.nature.com/licenceforms/nrg/competing->

interests.pdf.

REFERENCES – are limited to a total of 70 for Articles, Resources, Technical Reports; and 40 for Letters. This includes references in the main text and Methods combined. References must be numbered sequentially as they appear in the main text, tables and figure legends and Methods and must follow the precise style of Nature Cell Biology references. References only cited in the Methods should be numbered consecutively following the last reference cited in the main text. References only associated with Supplementary Information (e.g. in supplementary legends) do not count toward the total reference limit and do not need to be cited in numerical continuity with references in the main text. Only published papers can be cited, and each publication cited should be included in the numbered reference list, which should include the manuscript titles. Footnotes are not permitted.

METHODS – Nature Cell Biology publishes methods online. The methods section should be provided as a separate Word document, which will be copyedited and appended to the manuscript PDF, and incorporated within the HTML format of the paper.

Methods should be written concisely, but should contain all elements necessary to allow interpretation and replication of the results. As a guideline, Methods sections typically do not exceed 3,000 words. The Methods should be divided into subsections listing reagents and techniques. When citing previous methods, accurate references should be provided and any alterations should be noted. Information must be provided about: antibody dilutions, company names, catalogue numbers and clone numbers for monoclonal antibodies; sequences of RNAi and cDNA probes/primers or company names and catalogue numbers if reagents are commercial; cell line names, sources and information on cell line identity and authentication. Animal studies and experiments involving human subjects must be reported in detail, identifying the committees approving the protocols. For studies involving human subjects/samples, a statement must be included confirming that informed consent was obtained. Statistical analyses and information on the reproducibility of experimental results should be provided in a section titled “Statistics and Reproducibility”.

All Nature Cell Biology manuscripts submitted on or after March 21 2016 must include a Data availability statement at the end of the Methods section. For Springer Nature policies on data availability see <http://www.nature.com/authors/policies/availability.html>; for more information on this particular policy see <http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf>. The Data availability statement should include:

- Accession codes for primary datasets (generated during the study under consideration and designated as "primary accessions") and secondary datasets (published datasets reanalysed during the study under consideration, designated as "referenced accessions"). For primary accessions data should be made public to coincide with publication of the manuscript. A list of data types for which submission to

community-endorsed public repositories is mandated (including sequence, structure, microarray, deep sequencing data) can be found here <http://www.nature.com/authors/policies/availability.html#data>.

- Unique identifiers (accession codes, DOIs or other unique persistent identifier) and hyperlinks for datasets deposited in an approved repository, but for which data deposition is not mandated (see here for details <http://www.nature.com/sdata/data-policies/repositories>).
- At a minimum, please include a statement confirming that all relevant data are available from the authors, and/or are included with the manuscript (e.g. as source data or supplementary information), listing which data are included (e.g. by figure panels and data types) and mentioning any restrictions on availability.
- If a dataset has a Digital Object Identifier (DOI) as its unique identifier, we strongly encourage including this in the Reference list and citing the dataset in the Methods.

We recommend that you upload the step-by-step protocols used in this manuscript to [protocols.io](https://www.protocols.io). More details can be found at <https://www.protocols.io/help/publish-articles>.

DISPLAY ITEMS – main display items are limited to 6-8 main figures and/or main tables for Articles, Resources, Technical Reports; and 5 main figures and/or main tables for Letters. For Supplementary Information see below.

FIGURES – Colour figure publication costs \$600 for the first, and \$300 for each subsequent colour figure. All panels of a multi-panel figure must be logically connected and arranged as they would appear in the final version. Unnecessary figures and figure panels should be avoided (e.g. data presented in small tables could be stated briefly in the text instead).

All imaging data should be accompanied by scale bars, which should be defined in the legend. Cropped images of gels/blots are acceptable, but need to be accompanied by size markers, and to retain visible background signal within the linear range (i.e. should not be saturated). The boundaries of panels with low background have to be demarked with black lines. Splicing of panels should only be considered if unavoidable, and must be clearly marked on the figure, and noted in the legend with a statement on whether the samples were obtained and processed simultaneously. Quantitative comparisons between samples on different gels/blots are discouraged; if this is unavoidable, it should only be performed for samples derived from the same experiment with gels/blots were processed in parallel, which needs to be stated in the legend.

Figures should be provided at approximately the size that they are to be printed at (single column is 86

mm, double column is 170 mm) and should not exceed an A4 page (8.5 x 11"). Reduction to the scale that will be used on the page is not necessary, but multi-panel figures should be sized so that the whole figure can be reduced by the same amount at the smallest size at which essential details in each panel are visible. In the interest of our colour-blind readers we ask that you avoid using red and green for contrast in figures. Replacing red with magenta and green with turquoise are two possible colour-safe alternatives. Lines with widths of less than 1 point should be avoided. Sans serif typefaces, such as Helvetica (preferred) or Arial should be used. All text that forms part of a figure should be rewritable and removable.

We accept files from the following graphics packages in either PC or Macintosh format:

- For line art, graphs, charts and schematics we prefer Adobe Illustrator (.AI), Encapsulated PostScript (.EPS) or Portable Document Format (.PDF). Files should be saved or exported as such directly from the application in which they were made, to allow us to restyle them according to our journal house style.
- We accept PowerPoint (.PPT) files if they are fully editable. However, please refrain from adding PowerPoint graphical effects to objects, as this results in them outputting poor quality raster art. Text used for PowerPoint figures should be Helvetica (preferred) or Arial.
- We do not recommend using Adobe Photoshop for designing figures, but we can accept Photoshop generated (.PSD or .TIFF) files only if each element included in the figure (text, labels, pictures, graphs, arrows and scale bars) are on separate layers. All text should be editable in 'type layers' and line-art such as graphs and other simple schematics should be preserved and embedded within 'vector smart objects' - not flattened raster/bitmap graphics.
- Some programs can generate Postscript by 'printing to file' (found in the Print dialogue). If using an application not listed above, save the file in PostScript format or email our Art Editor, Allen Beattie for advice (a.beattie@nature.com).

Regardless of format, all figures must be vector graphic compatible files, not supplied in a flattened raster/bitmap graphics format, but should be fully editable, allowing us to highlight/copy/paste all text and move individual parts of the figures (i.e. arrows, lines, x and y axes, graphs, tick marks, scale bars etc.). The only parts of the figure that should be in pixel raster/bitmap format are photographic images or 3D rendered graphics/complex technical illustrations.

All placed images (i.e. a photo incorporated into a figure) should be on a separate layer and independent from any superimposed scale bars or text. Individual photographic images must be a minimum of 300+ DPI (at actual size) or kept constant from the original picture acquisition and not decreased in resolution post image acquisition. All colour artwork should be RGB format.

FIGURE LEGENDS – must not exceed 350 words for each figure to allow fit on a single printed NCB page together with the figure. They must include a brief title for the whole figure, and short descriptions of each panel with definitions of the symbols used, but without detailing methodology.

TABLES – main tables should be provided as individual Word files, together with a brief title and legend. For supplementary tables see below.

SUPPLEMENTARY INFORMATION – Supplementary information is material directly relevant to the conclusion of a paper, but which cannot be included in the printed version in order to keep the manuscript concise and accessible to the general reader. Supplementary information is an integral part of a Nature Cell Biology publication and should be prepared and presented with as much care as the main display item, but it must not include non-essential data or text, which may be removed at the editor's discretion. All supplementary material is fully peer-reviewed and published online as part of the HTML version of the manuscript. Supplementary Figures and Supplementary Notes are appended at the end of the main PDF of the published manuscript.

Supplementary items should relate to a main text figure, wherever possible, and should be mentioned sequentially in the main manuscript, designated as Supplementary Figure, Table, Video, or Note, and numbered continuously (e.g. Supplementary Figure 1, Supplementary Figure 2, Supplementary Table 1, Supplementary Table 2 etc.).

Unprocessed scans of all key data generated through electrophoretic separation techniques need to be presented in a supplementary figure that should be labelled and numbered as the final supplementary figure, and should be mentioned in every relevant figure legend. This figure does not count towards the total number of figures and is the only figure that can be displayed over multiple pages, but should be provided as a single file, in PDF or TIFF format. Data in this figure can be displayed in a relatively informal style, but size markers and the figures panels corresponding to the presented data must be indicated.

The total number of Supplementary Figures (not including the “unprocessed scans” Supplementary Figure) should not exceed the number of main display items (figures and/or tables (see our Guide to Authors and March 2012 editorial <http://www.nature.com/ncb/authors/submit/index.html#suppinfo>; <http://www.nature.com/ncb/journal/v14/n3/index.html#ed>). No restrictions apply to Supplementary Tables or Videos, but we advise authors to be selective in including supplemental data.

Each Supplementary Figure should be provided as a single page and as an individual file in one of our accepted figure formats and should be presented according to our figure guidelines (see above).

Supplementary Tables should be provided as individual Excel files. Supplementary Videos should be provided as .avi or .mov files up to 50 MB in size. Supplementary Figures, Tables and Videos must be accompanied by a separate Word document including titles and legends.

GUIDELINES FOR EXPERIMENTAL AND STATISTICAL REPORTING

REPORTING REQUIREMENTS – To improve the quality of methods and statistics reporting in our papers we have recently revised the reporting checklist we introduced in 2013. We are now asking all life sciences authors to complete two items: an Editorial Policy Checklist (found here <https://www.nature.com/authors/policies/Policy.pdf>) that verifies compliance with all required editorial policies and a reporting summary (found here <https://www.nature.com/authors/policies/ReportingSummary.pdf>) that collects information on experimental design and reagents. These documents are available to referees to aid the evaluation of the manuscript. 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>.

STATISTICS – Wherever statistics have been derived the legend needs to provide the n number (i.e. the sample size used to derive statistics) as a precise value (not a range), and define what this value represents. Error bars need to be defined in the legends (e.g. SD, SEM) together with a measure of centre (e.g. mean, median). Box plots need to be defined in terms of minima, maxima, centre, and percentiles. Ranges are more appropriate than standard errors for small data sets. Wherever statistical significance has been derived, precise p values need to be provided and the statistical test used needs to be stated in the legend. Statistics such as error bars must not be derived from $n < 3$. For sample sizes of $n < 5$ please plot the individual data points rather than providing bar graphs. Deriving statistics from technical replicate samples, rather than biological replicates is strongly discouraged. Wherever statistical significance has been derived, precise p values need to be provided and the statistical test stated in the legend.

Information on how many times each experiment was repeated independently with similar results needs to be provided in the legends and/or Methods for all experiments, and in particular wherever representative experiments are shown.

We strongly recommend the presentation of source data for graphical and statistical analyses as a separate Supplementary Table, and request that source data for all independent repeats are provided when representative experiments of multiple independent repeats, or averages of two independent experiments are presented. This supplementary table should be in Excel format, with data for different

figures provided as different sheets within a single Excel file. It should be labelled and numbered as one of the supplementary tables, titled “Statistics Source Data”, and mentioned in all relevant figure legends.

----- Please don't hesitate to contact NCB@nature.com should you have queries about any of the above requirements -----

Author Rebuttal to Initial comments
--

Reviewers' Comments:

Reviewer #1:

Remarks to the Author:

In the manuscript entitled “Real-time and programmable transcriptome sequencing with PROFIT-seq” Zhang et al. introduce the PROFIT-seq method which involves the adaptive sampling of transcriptome libraries to enrich genes of interest. The authors use a clever approach to get around the limitations of adaptive sampling and showcase their approach for several applications.

First, they test their web-based implementation of the adaptive sampling API on run simulations. Second they characterize their modified RT priming strategy on HeLa cell RNA. Third, they validate the combined priming and adaptive sampling approach on colorectal polyp samples which they enrich for a panel of cancer relevant genes. Fourth, they establish a custom Salmon-based quantification approach. Fifth, they use a different priming approach on sputum samples to characterize PROFIT-seq for pathogen detection - SARS-CoV-2 and pneumonia. Sixth, they implemented a 100Mb panel of cancer genes which they applied to a group of colorectal polyp samples.

Response: We greatly appreciate the reviewer’s comments on the novelty and significance of our study.

I generally like the approach and I believe the authors achieve some enrichment of target gene transcripts with their adaptive sampling approach. However, as with all adaptive sampling approaches, the enrichment levels are modest compared with probe-based enrichment approaches. On the other hand, they are “free” as in they require only computational, not experimental interventions.

Response: Thank you for your comments. We agree with reviewer that the enrichment effect of adaptive sampling is modest compared to probe-based enrichment approaches. PROFIT-seq offers the advantages of flexible enrichment for any type of transcripts while preserving the complete composition of the whole transcriptome. In contrast, other probe-based enrichment methods primarily target specific transcripts of interest and may not capture the global expression level effectively.

I also think the authors are doing too much here causing the manuscript's cohesiveness and clarity to suffer. There is the priming approach, the enrichment approach and the EM quantification approach all applied to several scenarios.

I think the pathogen detection approach in particular is an unnecessary detour especially since it uses a different priming strategy altogether. I think that part would be better suited for its own manuscript.

Response: Thank you. While utilizing a distinct priming strategy, our nCov analysis (Fig. 5) highlights how the PROFIT-seq method excels in rapid pathogen detection and accurate variant identification. Additionally, the polyp analysis (Fig. 6) illustrates the clinical utility of PROFIT-seq in exploring intricate host-pathogen interactions. We think it's important to include both case studies to comprehensively showcase the performance and versatility of our PROFIT-seq approach. Therefore, we respectfully request their inclusion in the revised manuscript.

There are other major and minor issues as well that I will outline below:

1) The authors cite but don't name the R2C2 approach which they use throughout the entire manuscript. They refer to an "RCA assay" but use a molecular biology protocol that is identical to R2C2, they use oligo sequences like the splint sequence directly from R2C2, and their analysis approach sounds very much like the C3POa analysis pipeline associated with R2C2.

To be clear, there is nothing wrong with using R2C2 in the way the authors did. I think the way the concatemerization in the R2C2 protocol is used to circumvent adaptive sampling is clever.

I just think that attribution is important. Name the method.

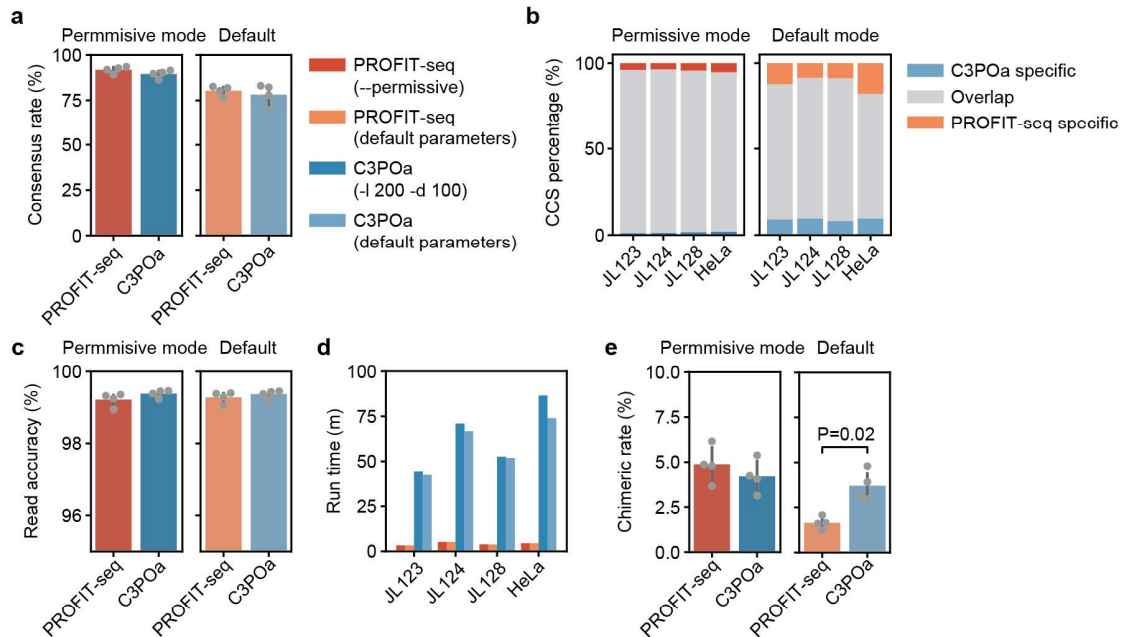
Response: Thanks for the comments. We have revised the manuscript as suggested to clearly name the R2C2 protocol in the main text and methods section.

2) If C3POa was used for analysis, the C3POa github should be linked. If not, the analysis code should be posted publicly and results (like output and accuracy) be compared to C3POa.

Response: Thank you for your valuable suggestion. In the PROFIT-seq protocol, we have introduced a Python script (https://github.com/bioinfo-biols/PROFIT-seq/blob/master/step2_consensus.py) to detect and generate consensus reads from R2C2 concatemers. Briefly, the Parasail¹ library was used to align splint adapters and determine the boundaries of RNA sequences, whereas sequences with intact 5' and 3' flanking adapters were classified as full-length RNAs. The consensus sequence of the captured RNA was computed using adaptive banded partial order alignment². To mitigate chimeric RCA artefacts, all subreads' lengths were required to fall within 80% to 120% of the consensus sequence, with subread similarity exceeding 50%.

We then compared the performance between the PROFIT-seq analysis script and C3POa (v3.1) using three exemplar polyp samples and the HeLa transcriptome on identical hardware (Intel Xeon W-2191B, 18 CPU cores, 256G DDR4 ECC memory). Both PROFIT-seq and C3POa were run under two different parameters sets: PROFIT-seq in default and permissive modes (without the RCA chimeric filter), and C3POa in default mode and with '-l 200 -d 100' for a less stringent filter. Each tool utilized 18 threads.

As shown in **Rebuttal Fig. 1a**, PROFIT-seq and C3POa showed comparable performance in generating consensus reads (PROFIT-seq: 91.86% / 80.31%; C3POa: 89.51% / 78.21% in permissive and default mode, respectively). The called consensus reads largely overlapped (permissive mode 94.16% / default mode 79.06%, **Rebuttal Fig. 1b**), with similar read accuracy (**Rebuttal Fig. 1c**), demonstrating the high consistency between C3POa and PROFIT-seq results. However, the PROFIT-seq script demonstrated significantly faster processing speeds (Wilcoxon rank sum test, $P=0.02$, **Rebuttal Fig. 1d**), and fewer RCA artefacts (Wilcoxon rank sum test, $P=0.02$) compared to the latest version of C3POa (**Rebuttal Fig. 1e**).



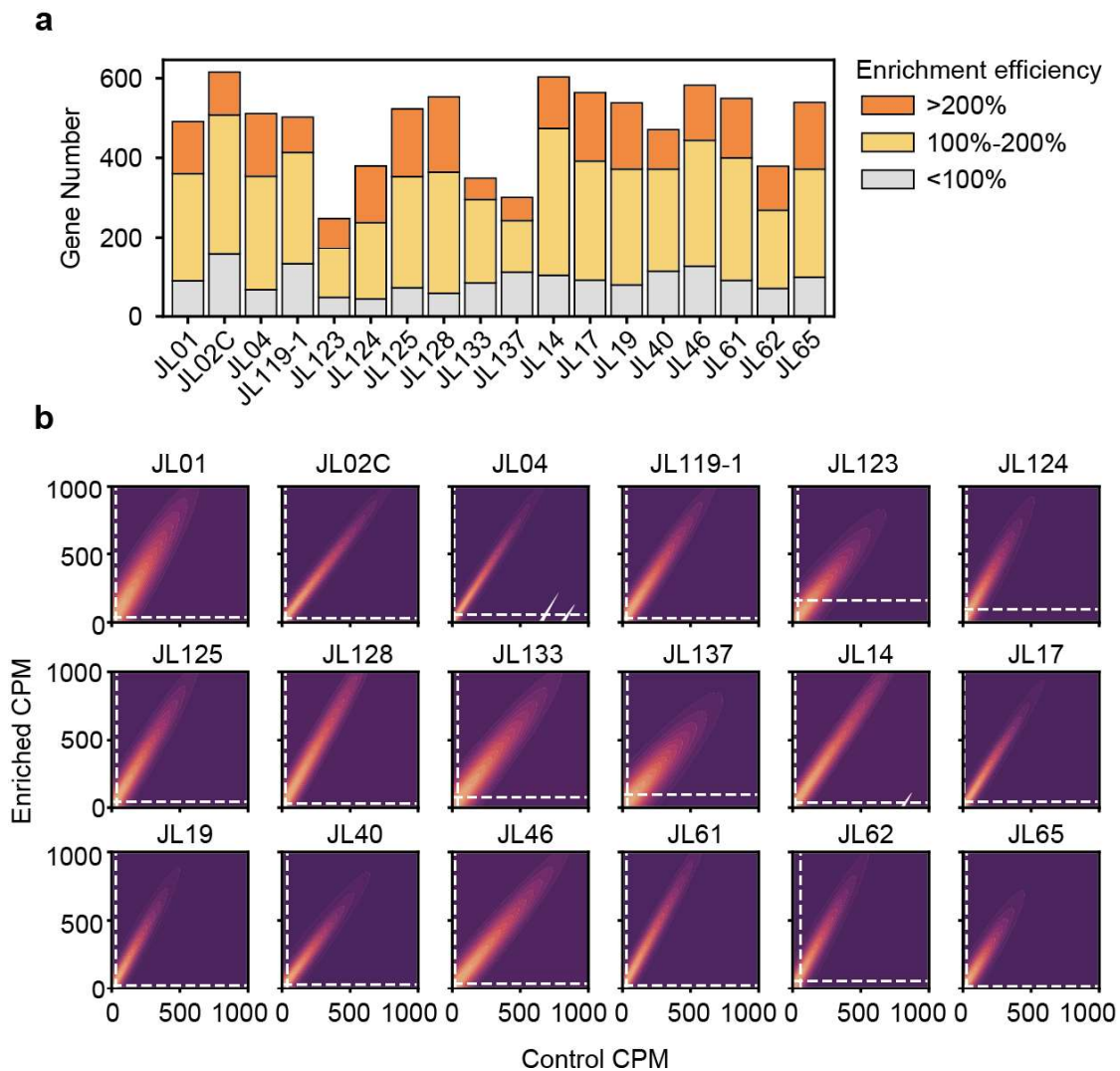
Rebuttal Figure 1. Comparison of PROFIT-seq and C3POa output. (a) Percentage of consensus reads generated by the PROFIT-seq script and C3POa (v3.1) using different parameter sets. (b) Overlap between the results of PROFIT-seq script and C3POa. (c) Per read accuracy of the PROFIT-seq script and C3POa results. Consensus reads were aligned to the reference genome using minimap2 with the option ‘-x splice --cs’, and the per read accuracy was calculated from the reported cs string. (d) Run time for the PROFIT-seq script and C3POa. Colors indicate different tools. (e) The percentage of RCA chimeric reads in PROFIT-seq and C3POa results. All subreads were extracted and aligned to the reference genome using minimap2 with option ‘-x splice’. The RCA chimeric reads were determined if subreads from the same RCA concatemers were aligned to different genomic regions with a distance larger than 1 kb.

1. Daily, J. Parasail: SIMD C library for global, semi-global, and local pairwise sequence alignments. *BMC Bioinformatics* **17**, 81 (2016).
2. Gao, Y. *et al.* abPOA: an SIMD-based C library for fast partial order alignment using adaptive band. *Bioinformatics* **37**, 2209-2211 (2020).

3) Enrichment levels.

The authors state that “72.89% of genes were successfully enriched”. The way “enriched” is used here, without any treatment, you would expect 50% of genes to be successfully “enriched”. What this really means in this case is that 37% of genes are covered less with the PROFIT-seq approach compared to the control run. This aligns with the rest of the manuscript - enrichment is just very low.

Response: Thanks for the comments. The “enriched” genes were defined as those with higher expression levels (measured in CPM) in the adaptive sampling runs compared to the unmanipulated controls. Due to sequencing depth limitations, a large number of lowly expressed genes may have been spuriously affected the enrichment effect. For highly expressed target genes with more than 5 supporting reads in both runs, 80.79% of them were successfully enriched, with 26.43% of the genes having over 2-fold enrichment (**Rebuttal Fig. 2a**). Overall, the target genes panel was significantly enriched (Wilcoxon rank sum test, $P < 0.001$) in all 18 PROFIT-seq runs (**Rebuttal Fig. 2b**). In addition, PROFIT-seq demonstrated over a 3-fold increase in the numbers of target reads in sputum samples (Fig. 5e), indicating its effectiveness in enriching target transcripts.



Rebuttal Figure 2. Evaluation of the enrichment efficiency. (a) Number of target genes that are significantly enriched (>2-fold) (orange), minorly enriched (1- to 2-fold) (yellow), or not enriched (gray). Only genes with more than 5 supporting reads in the control samples were included. **(b)** Expression level of target genes in adaptive sampling runs and unmanipulated controls. The x and y axis represent the expression level of target genes, the dashed lines represent the threshold of five supporting reads in control and adaptive sampling runs, respectively.

Additionally, runs were limited to less than 36 hours, even though MinION flow cells can produce data for 48 hours especially with reloads. Were runs reloaded? And is the complete run duration shown in all plots. It is quite possible that after reloads, the control runs would generate more on-target reads than the PROFIT-seq runs.

Response: Thanks for the comments. The MinION runs were limited to 36 hours without reloading. Both adaptive sampling runs and controls approached saturation after this period (Figure. 3d), indicating that additional sequencing would not yield significantly more on-target reads. Previous studies have suggested that a nuclease flush and reloading the sample could enhance throughput by recovering active pores, beneficial for both adaptive sampling runs and controls. In particular, adaptive sampling runs tend to benefit more from this process because the nuclease flush restores pores clogged during the ejection of sequencing molecules (Extended Data Fig.4 of Kovaka et al, PMID: 33257863). Therefore, additional reloading would not likely affect the enrichment efficiency.

I think to address this the authors should a) accompany all on-target read plots with total-read plots, and b) show data for full run durations if flow cells were run longer than the plots indicate

Response: Thanks for the comments. We have reorganized Fig. 3b&c as suggested. The Figure.5 already shows the full run duration, and the runs in Fig.3 were truncated to 36 hours because saturation has already been reached as described in the previous response.

4) The combinatorial priming strategy.

A single strand random hexamer primer will bind any RNA molecule, not just circular, even when added after the other primers. Any given round of priming and RT isn't 100% effective. I am therefore surprised by the outcome of the RT reactions. Only 3% of reads are classified as coming from non-polyA transcripts while 40% of reads lack a poly-A tail.

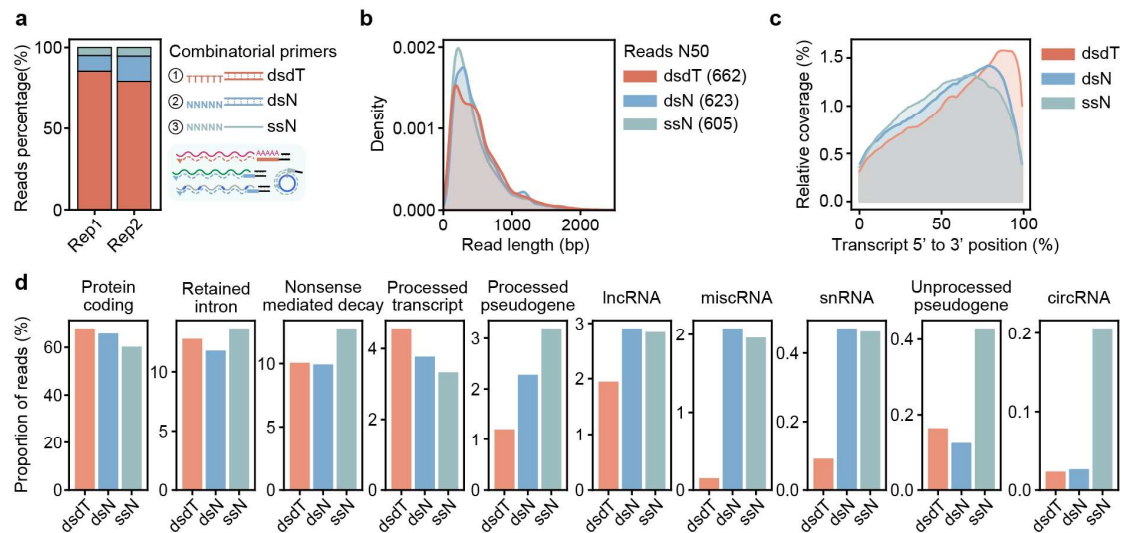
This points to the ds primer not really doing much while the ss primer largely primes polyA transcripts internally. However, S7B shows that the combinatorial priming approach is largely indistinguishable from oligodT priming in terms of transcript coverage. Ultimately, this requires clarification. At the very least, polyA and non-poly reads should be separated for an analysis like S7B and for read length plots.

It would have been very useful to delineate the products of the individual primers by adding index sequences to the primers themselves.

Response: Thanks for your constructive comments. As suggested, we have incorporated index sequences into individual primers, and conducted experiments similar to those presented in Fig.2 to evaluate the effect of combinatorial reverse transcription. The products of each primer were distinguished by aligning their index sequences to the first and last 150 bp of the sequences. As shown in **Rebuttal Fig. 3a**, in two replicates, 79.2% and 85.5% of sequenced molecules originated from double-stranded oligo-dT (dsdT) primers, while double-stranded random hexamers (dsN) and single-stranded random hexamers (ssN) accounted for 9.6% to 15.5%, and 4.9% to 5.2%, respectively. The prevalence of dsdT primers also explains the 3'-biased coverage observed in combinatorial RT compared to random primed libraries in S7B of the original manuscript.

To further evaluate dsN and ssN primers, we analyzed read length and transcript coverage individually. Random hexamers generated reads of similar length (**Rebuttal Fig. 3b**) and displayed a more even distribution across the 5' end of transcripts compared to dsdT-primed reads (**Rebuttal Fig. 3c**). Notably, an increased proportion of reads derived from non-polyadenylated transcripts (e.g. pseudogene, lncRNA, miscRNA, snRNA) was observed in these randomly primed reads. Specifically, ssN primers captured more circRNAs, aligning with our original intent of our combinatorial RT strategy.

We have updated Fig.3 with indexed combinatorial RT results.



Rebuttal Figure 3. Performance of individual primers in combinatorial RT protocol. (a) Percentage of reads derived from different primers. Colors indicate double-stranded oligo-dT (dsdT), double-stranded random hexamer (dsN) and single-stranded hexamer (ssN), respectively. (b) Length of different primed reads in combinatorial RT libraries. (c) Coverage across top 1000 highly expressed transcripts in different primed reads. The x-axis represents relative position along transcript, y-axis is the percent coverage of combined reads. (d) Proportion of reads aligned to transcripts of different biotypes according to the GENCODE annotation. Colors represent different primers. The proportion of circRNAs reads are calculated using CIRI-long.

5) Sequencing coverage plots are too small to show differences.

Response: Thanks for the comments. We have reorganized the coverage plots as suggested.

Minor

1) Throughout the manuscript percent and fold-change are used to quantify enrichment. I recommend sticking with fold-change throughout.

Response: Thanks for the comments. We have revised the manuscript as suggested.

2) “probe design that cannot capture different types of targets simultaneously.” I think this statement needs further explanation.

Response: Thanks for the comments. We have revised the sentence as follows “*Although targeted RNA-seq can improve the sensitivity of detecting target transcripts, the performance of these hybridization capture and amplicon-based methods relies on the design of highly efficient probes or primers for the target regions*”.

3) “However, after the adjustment by our quantification algorithm, the correlation was significantly improved ($P=0.733-0.717$)” The authors should probably lead with the smaller number.

Response: Thanks for the comments. We have revised the sentence as suggested.

4) What does the dashed line in 5C mean? I don’t think it’s explained in the figure legend.

Response: Thanks for the comments. The dashed lines indicate the time taken for adaptive sampling runs to achieve equivalent coverage as 24 hours of control runs. We have revised the figure legend to include the description.

Reviewer #2:

Remarks to the Author:

This study addresses the challenge of detecting specific transcripts within the diverse and complex eukaryotic transcriptome by introducing PROFIT-seq (programmable full-length isoform transcriptome sequencing).

Unlike current targeted RNA-seq methods, which involve complex pre-sequencing enrichment steps that limit the comprehensive characterisation of the transcriptome, PROFIT-seq enriches target transcripts while providing unbiased quantification of the entire transcriptome. The method utilises combinatorial reverse transcription for whole transcriptome capture and a programmable control system for selective enrichment during sequencing.

PROFIT-seq demonstrated significant improvements, achieving over a 300% increase in the yield of effective data and reducing the time required for detecting specific pathogens or key mutations by 75%. The method's application to colorectal polyp development revealed intricate associations between immune responses and bacterial infection, highlighting its potential in revealing complex biological processes.

The authors present PROFIT-seq as an advanced method that ensures accurate and efficient sequencing of target transcripts, alongside unbiased quantification of both target and non-target transcripts. While I enjoyed reading the manuscript and the data was presented well, I have the following major comments:

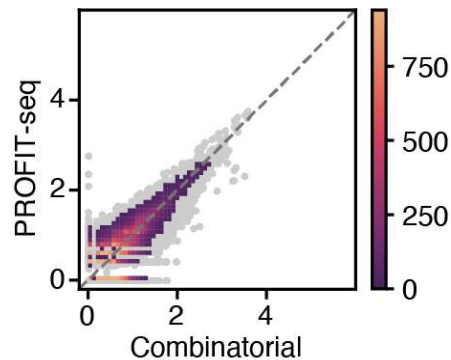
Response: We greatly appreciate the reviewer's comments on the novelty and significance of our study. The comments are carefully addressed in the revised manuscript. Please refer to the point-to-point response for detailed information.

Major issues:

- No UMIs and this is a major issue if you are using RCA for amplification. How quantitative is this assay given the known errors associated with ONT sequencing.

Response: Thanks for the suggestion. The UMIs have been included in the circularization splint sequences (5'-TGAGGCTGATGAGTTCCATANNNTATA
TNNNNATCACTACTTAGTTTTTGATAGCTTCAAGCCAGAGTTGTCTTTTTCTCTTT
GCTGGCAGTAAAAGTATTGTGTACCTTTTGCTGGGTCAGGTTGTTCTTTAGGAGGAG
TAAAAGGATCAAATGCACTAANNNTATATNNNNNGCGATCGAAAATATCCCTTT-
3'). We have updated the analysis script to identify UMI sequences from the consensus splint sequence and perform deduplication to remove RCA duplicates with over >80% sequence overlaps and more than 80% UMI similarity. The description for removing RCA duplicates have been included in lines 971-974 of the revised manuscript.

For the quantitative analysis, we compared the expression levels of the HeLa transcriptome sequenced using the complete PROFIT-seq protocol with those sequenced using only combinatorial RT without R2C2 amplification. As demonstrated in **Rebuttal Fig. 4**, the transcript expression levels from combinatorial RT and R2C2 amplified samples were significantly correlated ($R=0.87$, $P<10^{-31}$), indicating the reliable quantitative capability of our PROFIT-seq protocol.



Rebuttal Figure 4. Gene expression levels in combinatorial RT and PROFIT-seq libraries. The x axis represents gene expression level in combinatorial RT libraries without RCA amplification. The y axis represents the gene expression level in whole PROFIT-seq library.

- Need to describe the potential artefacts generated by this approach - RCA can generate a lot of false chimeric artefacts which makes the analysis challenging, especially when investigating oncology samples.

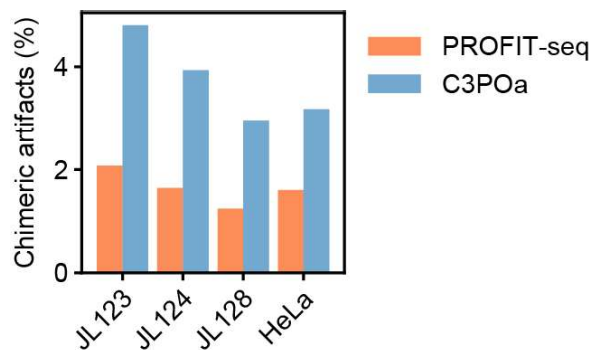
Response: Thanks for the comments. We agree with the reviewer that RCA could produce false chimeric artefacts. Our PROFIT-seq protocol incorporates filters on subread lengths and similarities to mitigate potential artifacts, and we evaluated the chimeric artifacts in the consensus sequences at various aspects. We have included a discussion of this limitation in lines 583 to 586 of the revised manuscript.

- Have you implemented computational strategies to overcome the potential issues of artefacts? It doesn't seem to be described within your methods.

Response: Thanks for the kind suggestion. In the PROFIT-seq protocol, we provided a Python script to detect and generate consensus reads from the R2C2 concatemers. Briefly, the Parasail library was used to align the splint adapters and determine the boundaries of the RNA sequences, whereas sequences with intact 5' and 3' flanking adapters were classified as full-length RNAs. The consensus sequence of the captured RNA was then calculated using adaptive banded partial order alignment. To filter out chimeric RCA artefacts, the length of all subreads must range from 80% to 120% of the consensus sequence, and the similarity between all subreads must exceed 50%. We have included a description of chimeric filtering in line 966-968 of the revised manuscript.

- To demonstrate that your method does not generate too many false chimerics, you need to do species mixing experiments and sequence human and mouse (or another species) and evaluate the effect of chimeric on the data analysis.

Response: Thanks for the comments. We evaluated the impact of false chimeras based on three random polyp samples and the HeLa transcriptome. Briefly, subreads for called consensus reads were extracted from the PROFIT-seq results, and RCA products with subreads aligned to different genomic regions (distance larger than 1kb) are considered as false chimeras. At the same time, similar analysis was also conducted for C3POa (v3.1), which also provides consensus calling on R2C2 nanopore data. As shown in **Rebuttal Fig. 5**, PROFIT-seq generated an average of 1.64% RCA chimeras in called consensus sequences, indicating a significant lower error rate compared to the 3.72% of RCA chimeras in C3POa data (Wilcoxon rank sum test, $P=0.02$).



Rebuttal Figure 5. The percentage of RCA chimeric reads in PROFIT-seq and C3POa results. All subreads were extracted and aligned to the reference genome using minimap2 with option '-x splice'. The RCA chimeric reads were determined if subreads from the same RCA concatemers were aligned to different genomic region with a distance larger than 1 kb.

Submit your sequencing data to ncbi or ENA because the data is currently behind a controlled access and is not open to the wider research community and it is hosted on CNCB, a platform that is very challenging to work with as a computational biologist. I was not able to download your data because of this, therefore I was only able to perform a limited review of the data. Your data contains HeLa cells, so as a minimum this data does not need to be behind a controlled access agreement.

Response: Thanks for the comments. The submission of raw sequencing data of human genetic materials to CNCB are mandated. The polyp and sputum sequencing data can be downloaded from the following link (<https://ngdc.cnecb.ac.cn/gsa-human/s/808cZ72j>). The HeLa RNA-seq data have been deposited in the NCBI SRA database under accession number PRJNA1133093.

Decision Letter, first revision:

Our ref: NCB-A53969A

29th July 2024

Dear Dr. Zhao,

Thank you for submitting your revised manuscript "Real-time and programmable transcriptome sequencing with PROFIT-seq" (NCB-A53969A). It has now been seen by the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Cell Biology, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

Please address the remaining 2 points of reviewer#1 in a revised version

If the current version of your manuscript is in a PDF format, please email us a copy of the file in an editable format (Microsoft Word or LaTeX)-- we can not proceed with PDFs at this stage.

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

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

Sincerely,

Sabrya Carim, PhD
(she/her/hers)
Associate Editor, Nature Cell Biology
Nature Portfolio

Springer Nature
The Campus, 4 Crinan Street, London N1 9XW, UK
sabrya.carim@springernature.com
<https://orcid.org/0000-0001-9485-1938>

Reviewer #1 (Remarks to the Author):

In their revised manuscript, the authors now acknowledge the use of the R2C2 method and compare their own consensus read implementation to C3POa. As part of this comparison, I think the y-axis in Supplemental Fig. 3e was chosen poorly. It should be 98%-100%. Better yet, it should be logarithmic. But that's minor.

The authors also perform additional experiments to delineate the primer product composition of their approach. Indeed, it looks like the oligodT primer generates most of the product and the ss primer very little. I think sharing this info is important as it contrasts the complexity of the priming strategy with the still-mostly-oligodT outcome.

Generally, my major concerns have been addressed.

Reviewer #2 (Remarks to the Author):

All comments addressed and I don't have any further remarks.

Decision Letter, second revision:

Our ref: NCB-A53969A

2nd August 2024

Dear Dr. Zhao,

Thank you for your patience as we've prepared the guidelines for final submission of your Nature Cell Biology manuscript, "Real-time and programmable transcriptome sequencing with PROFIT-seq" (NCB-A53969A). Please carefully follow the step-by-step instructions provided in the attached file, and add a response in each row of the table to indicate the changes that you have made. Ensuring that each point is addressed will help to ensure that your revised manuscript can be swiftly handed over to our production team.

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

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

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

In recognition of the time and expertise our reviewers provide to Nature Cell Biology's editorial process, we would like to formally acknowledge their contribution to the external peer review of your manuscript entitled "Real-time and programmable transcriptome sequencing with PROFIT-seq". For those reviewers who give their assent, we will be publishing their names alongside the published article.

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

Cover suggestions

COVER ARTWORK: We welcome submissions of artwork for consideration for our cover. For more information, please see our [guide for cover artwork](#).

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

Please note that *Nature Cell Biology* is a Transformative Journal (TJ). Authors may publish their research

with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. [Find out more about Transformative Journals](#)

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.

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

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

Please use the following link for uploading these materials:

[Redacted]

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

Best regards,

Kendra Donahue
Staff
Nature Cell Biology

On behalf of

Sabrya Carim, PhD
(she/her/hers)
Associate Editor, Nature Cell Biology

Nature Portfolio

Springer Nature

The Campus, 4 Crinan Street, London N1 9XW, UK

sabrya.carim@springernature.com

<https://orcid.org/0000-0001-9485-1938>

Reviewer #1:

Remarks to the Author:

I their revised manuscript, the authors now acknowledge the use of the R2C2 method and compare their own consensus read implementation to C3POa. As part of this comparison, I think the y-axis in Supplemental Fig. 3e was chosen poorly. It should be 98%-100%. Better yet, it should be logarithmic. But that's minor.

The authors also perform additional experiments to delineate the primer product composition of their approach. Indeed, it looks like the oligodT primer generates most of the product and the ss primer very little. I think sharing this info is important as it contrasts the complexity of the priming strategy with the still-mostly-oligodT outcome.

Generally, my major concerns have been addressed.

Reviewer #2:

Remarks to the Author:

All comments addressed and I dont have any further remarks.

Author Rebuttal, first revision:

Reviewers' Comments:

Reviewer #1:

Remarks to the Author:

I their revised manuscript, the authors now acknowledge the use of the R2C2 method and compare their own consensus read implementation to C3POa. As part of this comparison, I think the y-axis in Supplemental Fig. 3e was chosen poorly. It should be 98%-100%. Better yet, it should be logarithmic. But that's minor.

Response: Thanks for the reviewer's comments. We have updated the Extended Fig. 3e to 98%-100% as suggested.

The authors also perform additional experiments to delineate the primer product composition of their approach. Indeed, it looks like the oligodT primer generates most of the product and the ss primer very little. I think sharing this info is important as it contrasts the complexity of the priming strategy with the still-mostly-oligo(dT) outcome.

Response: Thanks for the comments. We have included this information in the main figure, and discussed the benefits of keeping mostly oligo(dT) outcome in Discussion as follows:

"PROFIT-seq adapts a combinatorial RT strategy using double-stranded oligo(dT) and random primers as well as single-stranded random hexamers, which enable the simultaneous detection of non-polyadenylated and circular transcripts without skewing the expression level of dominant mRNAs."

Generally, my major concerns have been addressed.

Response: Thanks for the reviewer's comment on helping us improve the manuscript.

Reviewer #2:

Remarks to the Author:

All comments addressed and I don't have any further remarks.

Response: We greatly appreciate the reviewer's comments on the novelty and significance of our study.

Final Decision Letter:

Dear Dr Zhao,

I am pleased to inform you that your manuscript, "Real-time and programmable transcriptome sequencing with PROFIT-seq", has now been accepted for publication in Nature Cell Biology. Congratulations!

Thank you for sending us the final manuscript files to be processed for print and online production, and for returning the manuscript checklists and other forms. Your manuscript will now be passed to our production team who will be in contact with you if there are any questions with the production quality of supplied figures and text.

Over the next few weeks, your paper will be copyedited to ensure that it conforms to Nature Cell Biology 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.

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.

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

Due to the importance of these deadlines, we ask that you please let us know now whether you will be difficult to contact over the next month. If this is the case, we ask you provide us with the contact information (email, phone and fax) of someone who will be able to check the proofs on your behalf, and who will be available to address any last-minute problems.

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

Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details. An online order form for reprints of your paper is available at <https://www.nature.com/reprints/author-reprints.html>. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

Publication is conditional on the manuscript not being published elsewhere and on there being no announcement of this work to any media outlet until the online publication date in *Nature Cell Biology*.

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>

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

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.

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.

If your paper includes color figures, please be aware that in order to help cover some of the additional cost of four-color reproduction, Nature Portfolio charges our authors a fee for the printing of their color figures. Please contact our offices for exact pricing and details.

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

If you have not already done so, we strongly recommend that you upload the step-by-step protocols used in this manuscript to protocols.io (<https://protocols.io>), an open online resource that allows researchers to share their detailed experimental know-how. All uploaded protocols are made freely

available and are assigned DOIs for ease of citation. Protocols and Nature Portfolio journal papers in which they are used can be linked to one another, and this link is clearly and prominently visible in the online versions of both. Authors who performed the specific experiments can act as primary authors for the Protocol as they will be best placed to share the methodology details, but the Corresponding Author of the present research paper should be included as one of the authors. By uploading your Protocols onto protocols.io, you are enabling researchers to more readily reproduce or adapt the methodology you use, as well as increasing the visibility of your protocols and papers. You can also establish a dedicated workspace to collect your lab Protocols. Further information can be found at <https://www.protocols.io/help/publish-articles>.

You can 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 Portfolio.

Please feel free to contact us if you have any questions.

With kind regards,

Sabrya Carim, PhD
(she/her/hers)
Associate Editor, Nature Cell Biology
Nature Portfolio

Springer Nature
The Campus, 4 Crinan Street, London N1 9XW, UK
sabrya.carim@springernature.com
<https://orcid.org/0000-0001-9485-1938>

Click here if you would like to recommend Nature Cell Biology to your librarian
<http://www.nature.com/subscriptions/recommend.html#forms>

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