nature portfolio

Peer Review File

Accurate long-read transcript discovery and quantification at single-cell, pseudo-bulk and bulk resolution with Isosceles

Reviewer #1 (Remarks to the Author):

The manuscript by Kazba et al. focuses on transcript isoform discovery and quantitation in single cells.

I think this is a timely contribution to the fast-growing field of long-read sequencing analysis software.
Isosceles focuses on reference-guided de novo detection, accurate quantification, and downstream analysis of
full-length isoforms at either single-cell, pseudo-bulk, or bulk resolution levels.

IMHO, the main competitors of Isosceles for transcriptome isoform and discovery and quantitation are
IsoQuant (PMID: 36593406) and Bambu (PMID: 37308696).

While I appreciate the addition of Isosceles to the toolkit, I would like to see a few points being adressed in a revised version of this manuscript.

(1) I think the manuscript would benefit from an independent benchmark data set as outlined in:
PMID: 37783886
Data set selection is not an issue then.

(2) You are probably aware of the lrgasp preprint:
https://www.biorxiv.org/content/10.1101/2023.07.25.550582v1
Could you please update your manuscript and Figure 2 to include the same information (if applicable) as in
Figure 2: Overview of evaluation for Challenge 1: transcript identification with a reference annotation (from the preprint).
e.g. F1 score is missing.

(3) Since you feature the single cell "level" in your title,
I was a bit dissapointed to see that you "only" use Sicelore
for CBC assignment and UMI detection.
The authors from Sicelore and others (PMID: 33906975) have pointed out some limitations
Maybe, I missed it, but a clear advantage and edge over other solution would be to improve on CBC and UMI detection performance.
While the latter could help with quantitation, the first could is highly relevant for reducing noise in cell type discovery.

(4) Lastly, would your software also work for PacBio data as demonstarted for the competitors.

Reviewer #2 (Remarks to the Author):

Please, see the PDF attached.

Reviewer #2 Attachment on the following page

General comments

The paper entitled “Accurate long-read transcript discovery and quantification at single-cell resolution with Isosceles” describes a new tool for transcriptome reconstruction and quantification of long-read transcriptomic data, of both bulk and single-cell origin. In the view of rapid expansion of RNA sequencing using PacBio and Oxford Nanopore technologies, this tool is a useful instrument and a valuable addition to the scientific community.

I would like to emphasize the quality of Isosceles documentation and its convenient installation. Although I personally do not use R that often, it was still quite easy to install and test Isosceles on the provided toy dataset. Moreover, R implementation might be a benefit for researchers performing further downstream analysis using other R packages. Isosceles has both the short tutorial and the complete user manual. Since software is the main outcome of the project, usability and user-friendliness is a very important aspect. Data used in the paper as well as all the commands used during the benchmarking are provided.

Isosceles yields decent transcript models and superior expression estimations based on simulated data. The authors also provide a nice example of Isosceles being applied in realistic study using mouse brain single-cell data.

Overall, the manuscript is well-structured and easy to follow. Below I summarize a few concerns and suggestions, addressing which, in my opinion, may improve the quality of the current manuscript.

Major comments

- My main concerns are related to Figure 3 (and the following Figure S4), which is one of the key figures supporting conclusions about Isosceles superior quantification accuracy.
    - It is not entirely clear how exactly the 4000 transcripts were selected. The authors state “4000 top highly variable genes/transcripts”. Was the same set used for each tool? Was the selection based on the expression estimates from some particular tool?
    - The authors also provide a similar Supplementary Figure S4b,c, where all expressed transcripts are considered for the analysis (which sounds more relevant than selecting a subset of transcripts). The absolute difference in Spearman correlation (matched vs decoy) for different tools seems to be almost identical when using all transcripts, which does not support the conclusion about Isosceles quantification superiority.
    - One way to support the conclusions made by these plots, would be, for example, to apply the same methodology and to create similar plots using simulated data. This could be useful to understand how these plots (Fig. 3b,c) correlate with actual quantification accuracy measured using ground truth. Simulated data also allows to create reference plots similar to 3b,c using ground truth expression and

compute the actual Spearman correlation difference between matched and decoy comparisons.
    - Another way to support this would be to repeat the experiment by using completely different cell lines (possibly not cancer lines), so that the difference between matched and decoy comparisons is extreme.
    - In Fig. 3c,d the authors use specific intervals for the Y axis, i.e. not starting with 0. This provides a somewhat misleading impression on the difference between tools. I suggest using a 0-based Y axis to clearly depict relative differences between bars.
    - Isosceles does have the best median relative difference, Spearman correlation for matched samples and highest difference between matched and decoy comparisons. However, it also shows the most similar expression profiles in decoy comparisons (e.g. the lowest median relative difference). Providing some comments/insights on that matter could be informative.
    - The results provided in Fig. 3d. show FLAMES and Scielore being far worse than Isosceles on simulated data, while the difference seems to be far less dramatic according to real-data experiments (Fig.3c). On the contrary, IsoQuant shows decent performance on simulated data, but the worst mean relative difference for matched samples. I realize that finding the reason for these inconsistencies might be challenging, if not impossible, it would be interesting to hear whether authors have any comments regarding this observation.

Minor comments

- I suggest adding a short caption to Figure 2 stating what kind of data was used to make the figure self-explanatory.
- In Fig.2 (and following supplementary figures) the authors provide FDR only for all detected transcripts. I recommend inserting the same separated plots (annotated and withheld) as for TPR.
- Some plots in Fig.2 look slightly overloaded due to each tool being launched alone and coupled with StringTie. While I appreciate such a fulfilling approach, the authors might want to consider keeping only a few best-performing tools, or only stand-alone performance and move complete benchmarks to the supplementary material (since they are still informative).
- While the authors use both simulated and real data to benchmark Isosceles, it could be also informative to provide some benchmarks using synthetic molecules sequenced on real sequencers, e.g. SIRVs. SIRVs ONT data is publicly available e.g. at https://github.com/nanopore-wgs-consortium/NA12878/blob/master/RNA.md, or in some recent studies, such as LRGASP tool comparison (https://www.biorxiv.org/content/10.1101/2023.07.25.550582v1) or IsoQuant. For benchmarking purposes, in addition to complete SIRV annotation, Lexogen also

provides an annotation with withheld transcripts and an annotation with decoy transcripts.

- It could be informative for users to provide computational performance (CPU time, wall clock time, RAM peak) in the supplementary material.
- Since the authors trained NanoSim models themselves, providing the exact error rates for simulated data in the respective supplementary section might be informative.

Technical comments

- I have tried running Isosceles on some of my SIRV data but encountered an error in the prepare_transcripts function:
  Joining with `by = join_by(transcript_id)`
  Error: length(tx_df$hash_id) not equal to length(unique(tx_df$hash_id))
  I will submit a bug report in the near future.
- According to command lines used for benchmarking, the authors state that they ran IsoQuant twice. I think IsoQuant is capable of outputting both expression values for all reference transcripts and expression values for all discovered transcripts (known and novel) during the same run. I don't know whether it makes a big difference, but the provided benchmarks seem to be reasonable anyway.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The manuscript by Kazba et al. focuses on transcript isoform discovery and quantitation in single cells.

I think this is a timely contribution to the fast-growing field of long-read sequencing analysis software.
Isosceles focuses on reference-guided de novo detection, accurate quantification, and downstream analysis of
full-length isoforms at either single-cell, pseudo-bulk, or bulk resolution levels.

IMHO, the main competitors of Isosceles for transcriptome isoform and discovery and quantitation are
IsoQuant (PMID: 36593406) and Bambu (PMID: 37308696).

While I appreciate the addition of Isosceles to the toolkit, I would like to see a few points being adressed in a revised version of this manuscript.

We would like to thank both reviewers for reading and commenting on our manuscript, and are glad this reviewer found our work to be a 'timely contribution to the fast-growing field of long-read sequencing analysis software'. We appreciate the reviewer's helpful comments and suggestions, addressing which has improved the manuscript. Please find our responses and revisions to each point provided below.

(1) I think the manuscript would benefit from an independent benchmark data set as outlined in: PMID: 37783886
Data set selection is not an issue then.

We appreciate the great suggestion. We now include an accuracy benchmark using this dataset (Mixtures A and B from PMID: 37783886). The results display roughly identical Spearman correlation for the top programs Isosceles, IsoQuant, and Bambu when compared to ground-truth (all 0.97 on Mix A and 0.98 on Mix B), with Isosceles performing better on one of the mixture sets with mean relative difference (0.71 vs. 0.74 on Mix A and equal at 0.78 on Mix B). We have added this data to Fig. S3, and to the text on Page 4 (lines 125-130).

(2) You are probably aware of the lrgasp preprint:
https://www.biorxiv.org/content/10.1101/2023.07.25.550582v1
Could you please update your manuscript and Figure 2 to include the same information (if applicable) as in
Figure 2: Overview of evaluation for Challenge 1: transcript identification with a reference annotation (from the preprint).
e.g. F1 score is missing.

We thank the reviewer for this suggestion. In the revised manuscript, we now cite the LRGASP preprint in the introduction and include the F1-score in Fig. 2 and S2 panels. We also updated the main text to reflect these changes (Page 4, lines 116-118).

(3) Since you feature the single cell "level" in your title,
I was a bit dissapointed to see that you "only" use Sicelore
for CBC assignment and UMI detection.
The authors from Sicelore and others (PMID: 33906975) have pointed out some limitations
Maybe, I missed it, but a clear advantage and edge over other solution would be to improve on CBC and UMI detection performance.
While the latter could help with quantitation, the first could is highly relevant for reducing noise in cell type discovery.

The reviewer makes an important point here: the accuracy of quantification and downstream analysis is reliant on performant upstream processing, including CBC assignment & UMI detection. As mentioned, this is a significant remaining gap in the field– one that is sufficiently large in scope to warrant several entire papers and software suites dedicated to it, as the reviewer points out (PMID: 33906975).

In our manuscript, we used Sicelore for this because it is one of the most conservative methods, but acknowledge that it is not without limitations. Therefore, we have designed Isosceles to be agnostic to the upstream choice of CBC or UMI detection/assignment software, enabling compatibility with future progress in this area. For example, Oxford Nanopore is also developing a CBC/UMI detection suite for single-cell nanopore data (see *wf-single-cell*, https://github.com/epi2me-labs/wf-single-cell) which is also compatible with downstream analysis using Isosceles. In response to this comment, we have updated Isosceles documentation on GitHub to include instructions for running wf-single-cell, or other CBC/UMI detector, and have added a benchmark of Isosceles quantifications starting from either Sicelore or wf-single-cell CBC/UMI calls in the revised manuscript (Fig. S5d). Performing the matched vs decoy cell line benchmark (from Fig. 3), we find that Sicelore barcode calls perform significantly better. Therefore, we provide the Sicelore CBC/UMIs to all programs in the manuscript.

As for featuring the “single-cell level” in the title, in our manuscript we focused on another equally important and major gap in long-read single-cell analysis, which is the flexible applicability of EM for quantification at single-cell and pseudo-bulk resolutions. This is one of the main innovations of Isosceles, which we believe does provide a clear advantage over other methods that do not support such capability. We illustrate this advance in outperformance for benchmarks shown in Figures 2 & 3, as well as in practice, illustrating that these capabilities are enabling for flexible experimental design and biological discovery in our single-cell case study (Figure 4). In the revised manuscript we add “pseudo-bulk, and bulk” resolutions to the title, so as to more clearly highlight this capability across multiple resolution levels.

(4) Lastly, would your software also work for PacBio data as demonstarted for the competitors.

We thank the reviewer for the helpful suggestion which we address in the revised manuscript by adding a benchmark to compare Isosceles, IsoQuant, and Bambu across long-read sequencing platforms for the same cell line (GM12878). We compare nanopore reads from the Nanopore WGS Consortium (cDNA Pass basecalls from: https://github.com/nanopore-wgs-consortium/NA12878/blob/master/RNA.md, PMID: 31740818) with PacBio reads from the ENCODE Consortium (ENCFF450VAU). As an independent pseudo-'ground-truth', we compare each long-read quantification to the short-read Illumina qualifications downloaded from the ENCODE Consortium for that cell line (ENCFF485OUK). We find that all three programs perform well in ONT vs. Illumina comparisons, however Isosceles does display slightly higher Spearman correlation for PacBio vs. Illumina and PacBio vs. ONT (Fig. S4a). These data suggest that Isosceles is also able to perform well using PacBio data.

Reviewer #2 (Remarks to the Author):

General comments

The paper entitled “Accurate long-read transcript discovery and quantification at single-cell resolution with Isosceles” describes a new tool for transcriptome reconstruction and quantification of long-read transcriptomic data, of both bulk and single-cell origin. In the view of rapid expansion of RNA sequencing using PacBio and Oxford Nanopore technologies, this tool is a useful instrument and a valuable addition to the scientific community.

I would like to emphasize the quality of Isosceles documentation and its convenient installation. Although I personally do not use R that often, it was still quite easy to install and test Isosceles on the provided toy dataset. Moreover, R implementation might be a benefit for researchers performing further downstream analysis using other R packages. Isosceles has both the short tutorial and the complete user manual. Since software is the main outcome of the project, usability and user-friendliness is a very important aspect. Data used in the paper as well as all the commands used during the benchmarking are provided.

Isosceles yields decent transcript models and superior expression estimations based on simulated data. The authors also provide a nice example of Isosceles being applied in realistic study using mouse brain single-cell data.

Overall, the manuscript is well-structured and easy to follow. Below I summarize a few concerns and suggestions, addressing which, in my opinion, may improve the quality of the current Manuscript.

We thank the reviewer for testing our software and reading our manuscript, and are glad they found our manuscript to be “well-structured and easy to follow”, and Isosceles to be well documented, convenient to install, and a “useful instrument and a valuable addition to the scientific community.” We appreciate the reviewer’s insightful and constructive feedback, addressing which has significantly improved the quality of the revised manuscript. Please find our response and description of revisions to address each specific point below:

Major comments

● My main concerns are related to Figure 3 (and the following Figure S4), which is one of the key figures supporting conclusions about Isosceles superior quantification accuracy.

We acknowledge the initial lack of clarity in describing the experimental design and parameter choices for Figure 3 and the accompanying Figure S5 (formerly S4).

In this benchmark we put forth a realistic challenge for cell type discovery, testing whether the methods can correctly differentiate identical samples from similar but different decoy samples. We designed the experiment to give each program an equal and fair chance, opting for

parameter choices with this intent. We have expanded our discussion and rationale for parameter selection, addressing these concerns with significant revisions in the analysis based on this valuable feedback.

○ It is not entirely clear how exactly the 4000 transcripts were selected. The authors state “4000 top highly variable genes/transcripts”. Was the same set used for each tool? Was the selection based on the expression estimates from some particular tool?

We apologize for the lack of clarity on this point, and have improved this section of the results (Page 6, lines 175-176) and methods text (Page 17, lines 536-539) in response.

Each program is used independently from start to finish in the benchmark. So, the highly variable transcripts (HVT) are determined from each program’s expression estimates and those HVTs are subsequently utilized to differentiate matched and decoy cell lines.

Specific code describing how the HVTs were selected using the ‘getTopHVGs’ function from the *scran* package, is documented in Isosceles_Paper repository here: https://timbitz.github.io/Isosceles_Paper/reports/nanopore_bulk_sc_benchmarks_hvts_4000.html

○ The authors also provide a similar Supplementary Figure S4b,c, where all expressed transcripts are considered for the analysis (which sounds more relevant than selecting a subset of transcripts). The absolute difference in Spearman correlation (matched vs decoy) for different tools seems to be almost identical when using all transcripts, which does not support the conclusion about Isosceles quantification superiority.

The rationale for selecting highly variable transcripts (HVT) for cell type discrimination follows the standard in the field for single-cell workflows (eg. PMID: 31217225) and the popular packages *scran*, *seurat*, and *scanpy* all suggest subsetting only ~2k highly variable genes [HVG] for this purpose. We acknowledge that any single choice of HVT number is fairly arbitrary, and as a result, in the revised manuscript we investigate a range of 500, 1k, 2k, 4k, 6k, and 10k HVTs. Here, we observe significantly greater performance for all programs with fewer HVTs utilized, supporting that this approach is also the most effective (Fig. 3c). In the expanded benchmark, Isosceles provides a significant and robust improvement over other methods across the range of HVTs chosen.

Regarding all expressed transcripts, we note that each program actually quantifies and outputs different numbers of total transcripts (with mean pseudobulk expression >= 1 TPM). For example, FLAMES reported 11,765, Sicelore 12,446, IsoQuant 15,478, and Isosceles reported 21,760. As a result, utilizing all transcripts for each program results in less directly comparable metrics between programs. Therefore, in the revised manuscript we use 10,000 HVTs as the

maximum when comparing across programs, which is still sufficient to recapitulate the similarity of results from using all transcripts.

○ One way to support the conclusions made by these plots, would be, for example, to apply the same methodology and to create similar plots using simulated data. This could be useful to understand how these plots (Fig. 3b,c) correlate with actual quantification accuracy measured using ground truth. Simulated data also allows to create reference plots similar to 3b,c using ground truth expression and compute the actual Spearman correlation difference between matched and decoy Comparisons.

We thank the reviewer for this insightful and intriguing idea, addressing which has improved the manuscript. In revision, we implemented this and observed comparable results in the simulated benchmark, with Isosceles consistently outperforming other methods compared to ground-truth. However, one notable discrepancy exists, where IsoQuant performs at equally high accuracy to Isosceles specifically in simulated data between 500-4000 HVTs for Spearman correlation only (with no statistically significant difference between them). This benchmark is now added to Fig. S5c.

○ Another way to support this would be to repeat the experiment by using completely different cell lines (possibly not cancer lines), so that the difference between matched and decoy comparisons is extreme.

We appreciate the suggestion to use extremely different cell lines, which could effectively showcase all software in a clear-cut scenario. However, our experiment was specifically designed to capture the nuances of single-cell analyses, where technical noise and data sparsity often obscure biological signals. In this context, the ability to differentiate subtle variations, which are often as biologically relevant as more pronounced differences, presents a greater analytical challenge.

In response to earlier feedback on Figure 3, we have expanded our benchmark to more precisely evaluate detection of such subtle differences using a range of highly variable transcripts (HVTs). These changes have significantly improved the robustness and reproducibility of our results, which we believe strongly support the conclusions presented in our manuscript.

○ In Fig. 3c,d the authors use specific intervals for the Y axis, i.e. not starting with 0. This provides a somewhat misleading impression on the difference between tools. I suggest using a 0-based Y axis to clearly depict relative differences between bars.

We apologize for the oversight. In the revised manuscript, these bar charts have been replaced with line graphs (with standard error shaded) in order to enable visualization of the difference in

matched and decoy metrics as a function of 500-10k HVTs (Fig. 3c and Fig. S5c). The original bar plots are still available on the individual reports in the 'Isosceles_Paper' repository for each HVT number and both the bar plots and their line graph replacements use a 0-based Y-axis as suggested.

○ Isosceles does have the best median relative difference, Spearman correlation for matched samples and highest difference between matched and decoy comparisons. However, it also shows the most similar expression profiles in decoy comparisons (e.g. the lowest median relative difference). Providing some comments/insights on that matter could be informative.

We thank the reviewer for pointing out this observation. We speculate that the greater correlation overall (in both matched and decoys) could be due to the handling of ambiguous reads by the EM algorithm. Here, reads that might be discarded by other methods are still optimally apportioned by Isosceles, providing slightly higher apparent read-depths, and more consistent quantifications for genes with transcript-level ambiguity as a result. We now mention this in our rationale for matched vs. decoy comparisons in the results section relating to Fig. 3c (on Pages 5-6, lines 170-173).

○ The results provided in Fig. 3d. show FLAMES and Scielore being far worse than Isosceles on simulated data, while the difference seems to be far less dramatic according to real-data experiments (Fig.3c). On the contrary, IsoQuant shows decent performance on simulated data, but the worst mean relative difference for matched samples. I realize that finding the reason for these inconsistencies might be challenging, if not impossible, it would be interesting to hear whether authors have any comments regarding this observation.

This is a fair observation. Indeed, we also observe this trend in the revised manuscript's matched vs. decoy comparisons using simulated data (as suggested in the reviewer's earlier comment). Here, IsoQuant also performs better on simulated data than in the exact same benchmark using biological data. That said, as the reviewer correctly points out, it may be very difficult if not impossible to identify the cause of this discrepancy, which is likely related to inherent differences between simulated vs. true biological data. We now mention this observation in the discussion section on Page 8 (lines 261-264).

Minor comments

● I suggest adding a short caption to Figure 2 stating what kind of data was used to make the figure self-explanatory.

We now add a short caption to the Figure 2 legend as suggested.

● In Fig.2 (and following supplementary figures) the authors provide FDR only for all detected transcripts. I recommend inserting the same separated plots (annotated and withheld) as for TPR.

While we appreciate the reviewer's suggestion to separate FDR plots for annotated and withheld transcripts, it isn't feasible due to the experimental design. In our process, we simulate reads from a full transcript annotation file and provide a downsampled annotation file to each program. For TPR (sensitivity analysis), we differentiate between given and withheld transcripts to show disparities between programs. However, for FDR, the result is based on program outputs not aligning with either given or withheld transcripts, resulting in a single FDR for all detected transcripts, rather than segregated ones for given and withheld categories.

In the updated manuscript, we have also introduced the F1-score, encompassing the FDR and therefore also present only this total F1-score for all transcripts (Fig. 2, and Fig. S2).

● Some plots in Fig.2 look slightly overloaded due to each tool being launched alone and coupled with StringTie. While I appreciate such a fulfilling approach, the authors might want to consider keeping only a few best-performing tools, or only stand-alone performance and move complete benchmarks to the supplementary material (since they are still informative).

In the revised manuscript, we keep only the best performing value for each benchmark tool in the main figure (combination or stand-alone), except for Isosceles, which we show with all combinations since it is the focus of the paper. Fig. S2 still retains the full results as suggested.

● While the authors use both simulated and real data to benchmark Isosceles, it could be also informative to provide some benchmarks using synthetic molecules sequenced on real sequencers, e.g. SIRVs. SIRVs ONT data is publicly available e.g. at https://github.com/nanopore-wgs-consortium/NA12878/blob/master/RNA.md, or in some recent studies, such as LRGASP tool comparison (https://www.biorxiv.org/content/10.1101/2023.07.25.550582v1) or IsoQuant. For benchmarking purposes, in addition to complete SIRV annotation, Lexogen also provides an annotation with withheld transcripts and an annotation with decoy Transcripts.

We thank both reviewers for their helpful suggestions to include benchmarks using synthetic molecules sequenced on real sequencers. In response, we have included two additional benchmarks in our revised manuscript (Fig. S3), focusing on comparisons between Isosceles, Bambu, and IsoQuant.

The first benchmark utilizes Sequins, comparing all program quantifications to the ground truths derived from two distinct Sequin mixes. As detailed in our response to Reviewer #1's first comment, Isosceles demonstrated favorable performance across all evaluated metrics.

The second benchmark leverages SIRVs, utilizing annotated, withheld, and decoy transcript annotations to evaluate transcript detection. Across the three methods, fairly comparable performance was observed for transcript detection overall (eg. F1-scores ranging from 0.76 to

0.78). Despite identifying fewer withheld transcripts, Isosceles, with zero false positives, slightly outperformed IsoQuant, which had four false positives and missed five annotated structures (Fig. S3c-d).

We updated the text with these results (on Pages 4-5, lines 125-139) and the discussion (on Page 8, lines 264-268) to reflect these additions.

● It could be informative for users to provide computational performance (CPU time, wall clock time, RAM peak) in the supplementary material.

We now add benchmarks of computational performance (both CPU time and RAM peak) as Figure S4b.

● Since the authors trained NanoSim models themselves, providing the exact error rates for simulated data in the respective supplementary section might be informative.

We now include the following exact error rates for simulated data in the methods section on Page 14 (lines 432-433) and can also be found in the read_models directories in Isosceles_Paper: https://github.com/timbitz/Isosceles_Paper/tree/devel/input_data/read_models

Technical comments

● I have tried running Isosceles on some of my SIRV data but encountered an error in the prepare_transcripts function:
Joining with `by = join_by(transcript_id)`
Error: length(tx_df$hash_id) not equal to length(unique(tx_df$hash_id))
I will submit a bug report in the near future.

We appreciate the reviewer running our code and reporting this error. We have now fixed this with pull request #4 (https://github.com/timbitz/Isosceles/pull/4) and linked to the fix in the related issue #2 (https://github.com/timbitz/Isosceles/issues/2).

● According to command lines used for benchmarking, the authors state that they ran IsoQuant twice. I think IsoQuant is capable of outputting both expression values for all reference transcripts and expression values for all discovered transcripts (known and novel) during the same run. I don't know whether it makes a big difference, but the provided benchmarks seem to be reasonable anyway.

This is a reasonable point. In the original submission, we ran IsoQuant twice to maintain consistency with the other programs which were also receiving an IsoQuant-produced GTF file (for Figure 1). However, to respond to this, we went back and compared IsoQuant's performance in *de novo* mode in either (a) one single run or (b) consecutively, with the first run producing the GTF-file that is provided to the second run for quantification. What we found was interesting– IsoQuant performs significantly better overall with the two-run process than it does

in single-run mode (F1-score 84.8% vs. 74.5%; Sensitivity 74.2% vs. 59.9%; FDR 0.9% vs. 1.3% respectively). Therefore we present these two-run quantifications in the paper rather than the single-run results, and mention this in the methods section on Page 13 (lines 415-422).

While we didn't include the single-run results in our main or supplemental figures, we provide a link to this report in the Isosceles_Paper repository in the methods and for the reviewers here: https://github.com/timbitz/Isosceles_Paper/blob/devel/reports_static/simulated_bulk_benchmarks_isoquant.ipynb

---

Lastly, all the changes we have made to Isosceles software and paper repositories are listed in the most recent releases for both:
Isosceles_Paper: https://github.com/timbitz/Isosceles_Paper/releases/tag/0.2.0
Isosceles: https://github.com/timbitz/Isosceles/releases/tag/0.2.0

Reviewer #1 (Remarks to the Author):

I would like to congratulate the authors to this felicitous revision.
The authors have adressed all of my concerns and the manuscript can be IMHO published as is.

Reviewer #2 (Remarks to the Author):

The authors have successfully addressed my main concern regarding the key figure of this manuscript, and patiently clarified all points regarding performed experiments. I sincerely thank the authors for this work. The authors also resolved all minor issues and provided sufficient technical details. Overall, it feels like the revised version has improved in terms of clarity and justification of Isosceles performance.

I have no further questions and wish the authors the best of luck in their further research.

Kind regards
Andrey Prjibelski

Reviewer #2 (Remarks on code availability):

I downloaded and installed Isosceles. I also tested it on my own data and it seems to be working without issues.