nature portfolio

Peer Review File

Mapping Glioblastoma's Isoform Diversity Using Long-Read Single-Cell Analysis

Corresponding Author: Dr Brian Hon Yin Chung

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In this manuscript, Tang et al. apply single-cell long-read RNA sequencing to map isoform-level transcript diversity in glioblastoma across 27 patient samples. The study identifies differential transcript usage between tumor and non-tumor cells, across tumor subtypes, and discovers novel isoforms with potential therapeutic applications. The work presents a valuable resource for the GBM research community and proposes interesting therapeutic strategies such as dual-targeting approaches and neoantigen-based vaccines. However, several methodological concerns limit the interpretability and novelty of the findings. Importantly, there is insufficient methodological detail to evaluate the robustness of key claims. Increased rigor in statistical analyses, improved clarity in presentation, and more complete methodological documentation would substantially strengthen the manuscript.

Major Comments:

1. The evaluation and comparison of short-read and long-read snRNA-seq is illuminating. However, the presented data is not comprehensive enough to fully understand the benefits of scLR-seq. While gene level expression is highly correlated, it would be helpful to include a Y = X line in Figure 1C to determine if there is a systematic bias in either platform. It appears by glance that scLR-seq tends to under-estimate expression compared to scSR-seq, but it is difficult to tell. Additionally, the authors claim that UMAP-based clustering from long-read data demonstrated a clearer separation of cell states compared to short-read clustering. The evidence for this is not quite clear—it appears that the scSR and scLR-seq data were processed differently. For example, the Methods section suggests that Harmony was used for batch correction in scLR-seq but not in scSR-seq, which could easily explain the differences in UMAP. Authors report that scSR-seq data was integrated with the GBM atlas but do not mention whether scLR-seq was also integrated as well. These differences make the comparisons between scLR-seq and scSR-seq throughout the manuscript hard to interpret. Moreover, visual differences in UMAP structure are notoriously unreliable evidence, as UMAP is a stochastic algorithm sensitive to hyperparameters and the embedding does not preserve global distances or density.

2. The authors focus on isoform level switches in scLR-seq. Because part of the novelty proposed in this study is the comparison between scSR-seq and scLR-seq, a contrastive analysis of isoform level switches in scSR-seq would be helpful to highlight the usefulness of scLR-seq. For example, of the isoform switches detected in scLR-seq, how many could have been detected in scSR-seq? Are there any scSR-seq based switches that could not be corroborated in scLR-seq despite decent coverage of the gene, suggesting that scSR-seq identifies false-postives?

3. The claim that tumor cells frequently express protein-coding isoforms while non-tumor cells express retained-intron isoforms is made in a few parts of the paper. However, there is no systematic evidence presented for this claim. No figure or analysis is cited when this claim is first made on page 8. How many coding to RI switches are there? Are they the most common switch type? Are they more than expected?

4. During the discussion of DTU across tumor cell-types in Figure 2, the authors appear to pool all tumor cells across samples together for the analysis. It would be interesting to see if the identified isoform switches are driven by inter-tumor or intra-tumor heterogeneity. Are isoform switches patient specific? Because a simple Wilcoxon-rank sum test was used to identify DTUs, the authors were unable to correct for sample-specifics effects. It would be helpful to model these sample-specific effects explicitly, e.g. using a method like MAST.

5. Novel transcripts were identified by aligning reads against GENCODE v32 annotations. However, v32 is more than 5 years old (Oct 2019), and according to the GENCODE documentation, the most current version (v49) has more than 2 times as many annotated transcripts (227,462 vs 507,365). The reliance on v32 limits the novelty of this analysis. What percentage of the 6,523 novel isoforms are annotated in recent versions of GENCODE?

6. When analyzing novel isoforms, the authors switch from discussion of AC-like, MES-like, OPC-like and NPC-like tumor cells to differentiated-like, glial-neuronal and stem-like. I do not see these cell types in the UMAP in figure 1D. Were cells re-clustered for this analysis? Why not use the original designation of tumor cell-type?

7. The Methods section is quite bare. There were many moments while reading the paper where I wanted to understand the methods used but could not find any explanation. For example, the authors mention using Wilcoxon rank-sum test for identifying isoform differences between tumor and non-tumor cells. What was the p-value cutoff? Was multiple hypothesis testing controlled for? Was testing done within each patient or across patients? I don't see anything in the methods section besides citation of a scanpy function. Similar remarks apply to other reported terms such as "expression score" and "marker score" which appear not to be defined in Methods. Additionally, things like the identification of "differentiated-like tumor cells" is not explained in Methods and thus it is unclear how the authors identified these cell types.

Minor Comments:

1. When figures are cited, make sure the text matches the content of the figure. For example, it is unclear why Figure 1A is cited at the end of the first paragraph. Figure S1A should be cited after "long-read data provided more uniform coverage across gene bodies", and it is unclear why S1A is cited in the following sentence.

2. Figure S1A shows that the depth of scLR-seq is around ~2-4x that of scSR-seq. This would be important to highlight. Is it due to the biotin enrichment?

3. Figure 1B shows short-read only barcodes but not long-read only barcodes. This seems to imply that there were no long-read read only barcodes. Is this correct? Can the authors clarify why they did not include long-read only barcodes?

4. Figure 2B is an upset plot with only a single intersection and so is uninformative. A Venn diagram may be more suitable.

5. Multiple "stringent criteria" are reported with "p-value < 0.05". Is this reported p-value corrected for multiple hypotheses? If so, how was the correction done? If not, can the authors justify why they used uncorrected p-values.

6. I recommend changing the color intensity on Fig S1D. It is very difficult to see gains/losses by eye.

7. It would be helpful to display the expression of each transcript from figure 2E and 2H in a supplemental figure (e.g. in a violin plot) so that it is easy to verify that these isoform differences are statistically significantly different. The Sankey plots are illuminating but obscure quantitative differences.

Reviewer #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #3

(Remarks to the Author)
The authors present a GBM dataset combining long- and short-read single-cell sequencing. Given the lack of long-read single-cell datasets, this work is a valuable contribution that demonstrates the transcriptomic diversity of tumor cell populations, providing novel insights for personalized therapeutic targeting. The proposed framework for identifying extracellular and intracellular target pairs in patients is conceptually interesting and could be clinically relevant.

Major comments

1. One of the main novelties is the identification of 6,523 high-confidence new isoforms. However, this number includes ISM, NIC, and NNIC categories, which may be misleading, as ISM transcripts can represent truncated alignments to known isoforms rather than true novel transcripts. It would strengthen the manuscript to report the proportions of ISM, NIC, and NNIC, and to characterize novel isoforms in terms of the splicing events such as exon skipping, novel exon inclusion, intron retention. While some of this information is presented in the Supplementary Tables, including a summary in the main text could be helpful for readers.

2. No experimental (proteomic, immunopeptidomic, or targeted RT-PCR) validation is currently presented. To support the therapeutic claims, the authors could validate at least several key isoforms and/or their peptides. Alternatively, they could support their findings by checking whether the same isoforms or novel splice junctions can be detected in another GBM

datasets.

3. The statement “protein-coding isoforms were expressed in tumor cells, whereas non-coding or retained-intron isoforms predominated in non-tumor cells” should be supported by quantitative evidence, such as numbers in the text or a figure. Additional analyses such as comparisons of transcript or UTR lengths between switched isoforms could also be useful. It would also be interesting to check how shifts in transcript biotypes correlate with changes in gene expression.

Minor comments
1. Figure 2B is not particularly informative in its current form.
2. Figures 2E and 2H are difficult to interpret due to the large differences in cell numbers between groups, labelling numbers would improve clarity.
3. The abbreviation “DET” in Figures 2C and 2G is unclear.
4. In addition to the four major categories (FSM, ISM, NIC, NNIC), the list of novel transcripts also includes “genic” isoforms which may represent technical artifacts and should be addressed or excluded.

Reviewer #4

(Remarks to the Author)
The authors present a map of glioblastoma isoform diversity using long-read single-cell analysis. The work is compelling and adds to a growing literature using this approach for isoform analysis in a variety of tumors.

I have a few suggestions that I would like to see addressed to improve the rigor and impact of their findings which I have offered below:

1. Before getting to their long-read single-cell analysis, the authors should start by performing a differential splicing analysis on bulk RNA sequencing performed on these samples and determine whether single-cell replicates findings on the spliting landscape of GBM obtained from bulk seq through something like a Sankey diagram showing the changes in differential splicing events from bulk to scRNA-seq data.
2. The authors introduce their splicing results by summarizing median number of transcript isoforms per protein-coding gene and the portion of total transcripts that exhibited isotope shift - these metrics should be compared between tumor cells vs. non-tumor cells in their samples and across tumor cells of the four molecular subtypes.
3. Can the authors report the distribution of the various types of alternative splicing events (e.g. skipped exon, alternative 3′ splice site, alternative 5′ splice site, and retained intron) in tumor cells of the various subtypes vs. stromal cells?
4. While the authors perform nice analysis of specie variants amongst known single cell molecular GBM subtypes (AC-like, MES-like, etc.), did the authors find that splice variants reveal any intriguing clustering into subtypes that have not yet been described?
5. Protein level corroboration on these samples should be performed for splice variants distinguishable by antibodies
6. The authors should present results on what they found in term of splice variants for secreted proteins, particularly cytokines and ECM proteins which have known splice variants with roles in cancer biology.
7. They should interrogate RNA-binding proteins to identify key regulators of splicing diversity in GBM.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I appreciate the authors' efforts in addressing my previous comments, particularly in clarifying how the scSR-seq and scLR-seq datasets were processed and compared. These clarifications will greatly help readers follow the analytical workflow. I have one remaining concern that I believe can be addressed with a small revision.

Major comments:

In support of the claim that “scLR-seq provides enhanced resolution for cell-state identification compared to scSR-seq”, the authors now use scib-metrics package instead of visual analysis of UMAP structure. This is more principled than evaluating UMAPs, but it is still unclear how the scib-metrics support this claim. For example, the biological conservation metric reflects an integration method’s ability to conserve cell-type information between individual and integrated datasets. This metric largely reflects the capability of the integration method (here harmony) to integrate each individual sequencing type. Thus it is unclear how it could support the claim that one sequencing type provides enhanced resolution compared to another sequencing type. Moreover, the differences (0.59 vs 0.58) are very small, and without some kind of statistical framework it is unclear if they are meaningfully different.

To provide evidence that scLR-seq offers enhanced resolution of cell-states, the authors would need to perform some kind of analysis which shows that known biological variation is better captured within scLR-seq compared to scSR-seq, etc. Perhaps there’s a cell-state which does not cluster together in the scLR-seq but is delineated in the scLR-seq.

However, I don't think the claim that "scLR-seq provides enhanced resolution for cell-state identification compared to scSR-seq" is that important to the originality of this paper, and the data required to support this claim is outside the scope. The main novelty is the enhanced isoform resolution rather than cell-state identification, which is mostly bootstrapped from the GBM atlas anyway.

I therefore suggest that the authors simply remove the claim that scLR-seq provides enhanced resolution for cell-state identification, and instead use the scib-metrics to make the more modest claim that integration quality was comparable across platforms, ensuring downstream isoform-level comparisons are not confounded by differential integration performance.
The paper cited for the scib-metrics discussion is "Shortcomings of silhouette in single-cell integration benchmarking" (https://doi.org/10.1038/s41587-025-02743-4) , a paper which seems to be criticizing Silhouette based metrics for evaluating integration results. Ironically, this paper explicitly criticizes scib-metrics (although Silhouette metrics are only a small part of the total batch correction and bio conservation score, so I do not put much weight on the criticism). The original scib-metrics Nature Methods paper should be cited instead (https://doi.org/10.1038/s41592-021-01336-8).

Minor comments:

The number of reported cells on line 174 is "18,2222", which should presumably be 182,222
"differentials" on line 249 should be "differential"
"the" should be removed before MAST on line 280.

Reviewer #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #3

(Remarks to the Author)
The authors have done an amazing job addressing all my concerns. The additional experiments and analyses have significantly strengthened the manuscript. I have only one very minor comment: I still suggest slightly adjusting the wording regarding novel isoforms. In one paragraph, the authors first refer to "6,524 high-confidence novel isoforms (ISM, NIC, and NNIC)," and later to "the novel isoforms (2,500 NIC and NNIC)", which is a bit confusing. I would suggest reserving "novel isoforms" for NIC and NNIC only, and either using a different term (e.g., "non-FSM isoforms", if no other categories are included in the analysis) or simply listing the categories explicitly (ISM, NIC, and NNIC) when referring to all three.

Reviewer #4

(Remarks to the Author)
Requested revisions have been made and I have no further suggestions.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have done an excellent job addressing all my concerns and have substantially strengthened the manuscript. I appreciate their hard work.

Reviewer #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

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

Dear reviewers,

On behalf of the authors, I want to sincerely thank the reviewers
for your comments and suggestions to improve the article. We have taken your valuable advice and modified the manuscript accordingly. Please find below a point-by-point response to the comments.

Reviewer #1:

In this manuscript, Tang et al. apply single-cell long-read RNA sequencing to map isoform-level transcript diversity in glioblastoma across 27 patient samples. The study identifies differential transcript usage between tumor and non-tumor cells, across tumor subtypes, and discovers novel isoforms with potential therapeutic applications. The work presents a valuable resource for the GBM research community and proposes interesting therapeutic strategies such as dual-targeting approaches and neoantigen-based vaccines. However, several methodological concerns limit the interpretability and novelty of the findings. Importantly, there is insufficient methodological detail to evaluate the robustness of key claims. Increased rigor in statistical analyses, improved clarity in presentation, and more complete methodological documentation would substantially strengthen the manuscript.

Major Comments:

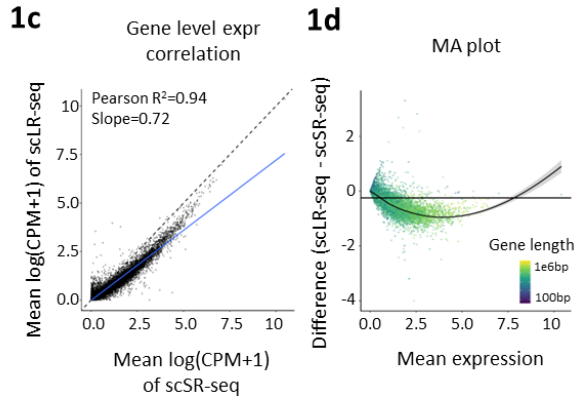
1.1

Reviewer Comment	The evaluation and comparison of short-read and long-read snRNA-seq is illuminating. However, the presented data is not comprehensive enough to fully understand the benefits of scLR-seq. While gene level expression is highly correlated, it would be helpful to include a Y = X line in Figure 1C to determine if there is a systematic bias in either platform. It appears by glance that scLR-seq tends to under-estimate expression compared to scSR-seq, but it is difficult to tell. Additionally, the authors claim that UMAP-based clustering from long-read data demonstrated a clearer separation of cell states compared to short-read clustering. The evidence for this is not quite clear—it appears that the scSR and scLR-seq data were processed differently. For example, the Methods section suggests that Harmony was used for batch correction in scLR-seq but not in scSR-seq, which could easily explain the differences in UMAP. Authors report that scSR-seq data was integrated with the GBM atlas but do not mention whether scLR-seq was also integrated as well. These differences make the comparisons between scLR-seq and scSR-seq throughout the manuscript hard to interpret. Moreover, visual differences in UMAP structure are notoriously unreliable evidence, as UMAP is a stochastic algorithm sensitive to hyperparameters and the embedding does not preserve global distances or density.
Author Response	We thank the reviewer for the positive assessment and are pleased that the study is considered a valuable resource for the GBM research community and that the proposed therapeutic strategies were found to be of interest, as well as for the insightful comments provided. We have updated Figure 1c to include the identity line (Y = X) to facilitate the visualization of systematic deviations between short-read and long-read gene expression estimates. We observe strong concordance between platforms (R^2 = 0.94, β = 0.72; New Fig. 1c),

indicating compression of expression in scLR-seq relative to scSR-seq.

To further investigate this bias, we performed a Bland–Altman (MA) analysis (log expression difference versus mean expression) and modeled expression differences using a multivariable linear regression incorporating both gene length and mean expression. We find that long-read–specific bias is significantly associated with both factors (New Fig. 1d). Mean expression level was the dominant contributor to platform differences ($\beta = -0.26$, P-value $< 2 \times 10^{-16}$). In addition, gene length (defined as the genomic span from transcription start site to transcription end site of the longest annotated transcript) remained a highly significant independent predictor ($\beta = -0.05$ per $\log_{10}$ gene length, P-value $< 2 \times 10^{-16}$).

We have added these analyses to the Results and Methods to clarify gene-level quantification from scLR-seq data and included an MA plot colored by gene length as a new panel to further illustrate the origin of this bias.



We also appreciate the reviewer's feedback regarding the comparison between scSR-seq and scLR-seq clustering. To address these concerns, we have standardized our computational pipeline and performed a quantitative benchmarking framework.

First, we utilized GBMap, a unified glioblastoma atlas encompassing 1.1 million cells across 240 patients, to establish our reference cell-type annotations at both finer and broader resolutions by mapping the scSR-seq data to this extensive atlas. Crucially, since the scSR-seq and scLR-seq datasets were generated from the same physical single cell library, we transferred these high-confidence labels to the scLR-seq data via exact cell barcode matching (Fig. 1e; Supplementary Fig. 2a-c).

We agree with the reviewer that UMAP is primarily a visualization tool and not as the basis for cell-type assignment or biological interpretation. Consequently, we processed both datasets using an identical computational pipeline and then employed the scib-metrics framework[1] to quantify biological conservation and batch integration. These results demonstrate that scLR-seq performs comparably, and slightly superior in batch correction (0.53 vs 0.50) and conservation (0.59 vs 0.58) compared to scSR-seq (New Supplementary Fig. 2f).

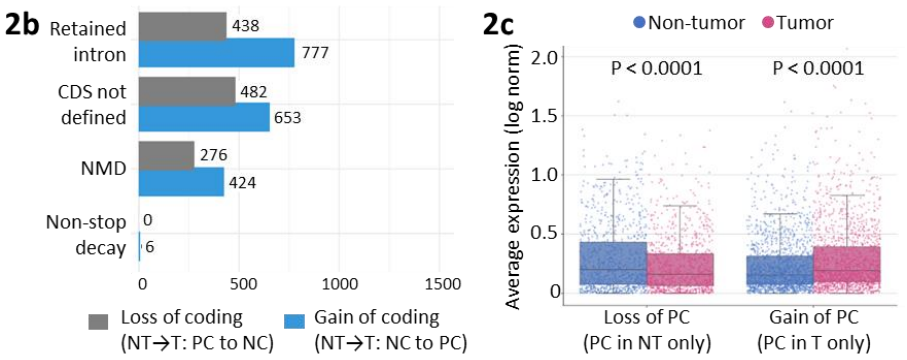
	
Excerpt from revised manuscript	*(P6) 'Gene-level expression estimates derived from short-read and long-read snRNA-seq showed strong concordance across genes, with modest compression of gene counts in long-read relative to short-read data (R^2 = 0.94, β = 0.72; Fig. 1c) ... consistent with the expected expression- and length-dependent deviations in long-read gene-level quantification.* *...* *First, we utilized GBMap, a unified glioblastoma atlas encompassing 1.1 million cells across 240 patients, to establish our reference cell-type annotations at both finer and broader resolutions by mapping the scSR-seq data to this extensive atlas (Supplementary Fig. 2a, 2b). Crucially, since the scSR-seq and scLR-seq datasets were generated from the same physical single cell library, high-confidence cell type labels were transferred to the scLR-seq data via exact cell barcode matching, enabling both fine-grained (Fig. 1e) and broader annotations (Supplementary Fig. 2c). Consistently, the 17 distinct clusters' most differentially expressed markers match with annotated cell identity (Fig. 1f).* *...* *Furthermore, to objectively evaluate whether scLR-seq provides enhanced resolution for cell-state identification compared to scSR-seq, we leveraged the scib-metrics framework to calculate the aggregate scores for biological conservation and batch correction. We observed that scLR-seq achieved slightly higher scores in both biological conservation (0.59 vs. 0.58) and batch correction (0.53 vs. 0.50) compared to scSR-seq (Supplementary Fig. 2f). These results suggest that the long-read data is reliable in distinguishing complex GBM cellular states.'* *Figure Legend:* *(1c) Correlation plot depicting gene expression concordance for 33,666 overlapping genes between short-read and long-read datasets. The dashed diagonal line indicates identity (Y = X).* *(1d) Bland–Altman (MA) plot showing the difference between long-read and short-read gene expression (long-read − short-read) versus mean expression. Points are colored by gene length (bp).* *(S2f) Dot plot showing Biological Conservation Scores and Batch Correction Scores between scSR-seq and scLR-seq modalities.*

1.2

Reviewer Comment	The authors focus on isoform level switches in scLR-seq. Because part of the novelty proposed in this study is the comparison between scSR-seq and scLR-seq, a contrastive analysis of isoform level switches in

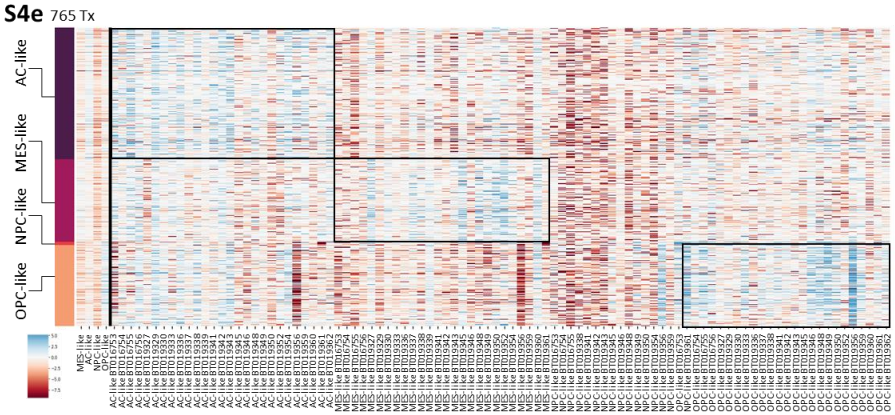
	scSR-seq would be helpful to highlight the usefulness of scLR-seq. For example, of the isoform switches detected in scLR-seq, how many could have been detected in scSR-seq? Are there any scSR-seq based switches that could not be corroborated in scLR-seq despite decent coverage of the gene, suggesting that scSR-seq identifies false-postives?
Author Response	Thank you for the thoughtful suggestion. We agree that while scLR-seq enables direct resolution of full-length transcript isoforms, scSR-seq provide complementary insights for detecting alternative polyadenylation (APA) isoforms. Therefore, we performed isoform quantification analysis on scSR-seq data by scUTRquant[2], which quantifies APA isoforms based on a curated 3' UTRome annotation. This analysis identified 426 genes exhibiting differential transcript usage between tumor and non-tumor cells (New Supplementary Table 4). We compared the tumor and non-tumor isoform markers identified via scLR-seq with those detected by scUTRquant on scSR-seq and observed significant concordance: 128 tumor-specific and 146 non-tumor-specific DTU isoforms detected in scSR-seq were exactly matched by full-length isoforms resolved by scLR-seq, demonstrating cross-platform concordance (New Supplementary Table 4). On the other hand, some differentially used transcripts identified by scLR-seq cannot be detected by scSR-seq (i.e. Fig. 2f), likely due to the inherent 3′ bias of scSR-seq, which limits detection of internal splicing events or alternative 5′ start sites sharing a terminal exon. Therefore, scLR-seq is essential for resolving the full-length architecture of these isoform switches.
Excerpt from revised manuscript	*(P8) 'While scLR-seq enables the direct characterization of full-length transcript isoforms, scSR-seq data provide complementary insights for detecting alternative polyadenylation (APA) isoforms. To assess whether tumor-associated transcript usage identified by scLR-seq is also reflected at the 3 ´ end of transcripts, we analyzed scSR-seq using scUTRquant33, which quantifies APA isoforms based on a curated 3 ´ UTRome annotation. This analysis identified 426 genes exhibiting differential transcript usage between tumor and non-tumor cells (Supplementary Table 4). Analysis of APA-associated structural changes revealed 255 3 ´ UTR lengthening and 171 shortening events, involving 852 transcripts across these 426 genes (tumor relative to non-tumor cells) (Supplementary Fig. 3e), indicating that differential transcript usage is frequently accompanied by gene-specific changes in transcript architecture. Notably, transcripts with shortened 3 ´ UTRs were enriched in tumor-associated pathways, including KRAS signaling and EMT (Supplementary Fig. 3f), consistent with previous reports linking APA-mediated UTR shortening to oncogenic activation34. Moreover, 128 tumor-specific and 146 non-tumor-specific differential transcript usage (DTU) isoforms detected in scSR-seq were exactly matched by full-length isoforms resolved by scLR-seq, demonstrating cross-platform concordance (Supplementary Table 4).'*

1.3

Reviewer Comment	The claim that tumor cells frequently express protein-coding isoforms while non-tumor cells express retained-intron isoforms is made in a few parts of the paper. However, there is no systematic evidence presented for this claim. No figure or analysis is cited when this claim is first made on page 8. How many coding to RI switches are there? Are they the most common switch type? Are they more than expected?
Author Response	Thank you for pointing this out. We have supplemented with numbers for the biotype changed isoforms to make it more clear. In summary, we observed significant biotype switching: 942 genes transitioned from protein-coding in non-tumor cells to non-coding biotypes in tumor cells, primarily through nonsense-mediated decay (NMD) (n=276) and intron retention (n=438). Conversely, higher number of genes (1,530) switched to protein-coding isoforms in tumor cells from non-coding origins in non-tumor cells, involving nonsense-mediated decay (n=424) and retained intron (n=777) biotypes (New Fig. 2b). Consistently, we observed that changes in protein-coding isoform status were significantly associated with expression differences, with gains of protein-coding genes corresponding to higher gene expression levels (New Fig. 2c). These results demonstrated widespread alternative splicing and isoform switching in tumors. 
Excerpt from revised manuscript	*(P8) 'We observed significant biotype switching ... and isoform switching in tumors'* *Figure Legend:* *(2b) Bar plot showing the number of events involving loss of protein-coding annotation (protein-coding in non-tumor cells only) and gain of protein-coding annotation (protein-coding in tumor cells only).* *(2c) Box plot showing gene-level expression associated with protein-coding annotation switches (Loss of PC genes: n=942; gain of PC genes: n=1530) between non-tumor and tumor cells. Mean log-normalized expression per gene is shown for each condition. Statistical significance was assessed using the Mann–Whitney U test. NT: non-tumor; T: tumor. PC: protein-coding; NC: non-coding.*

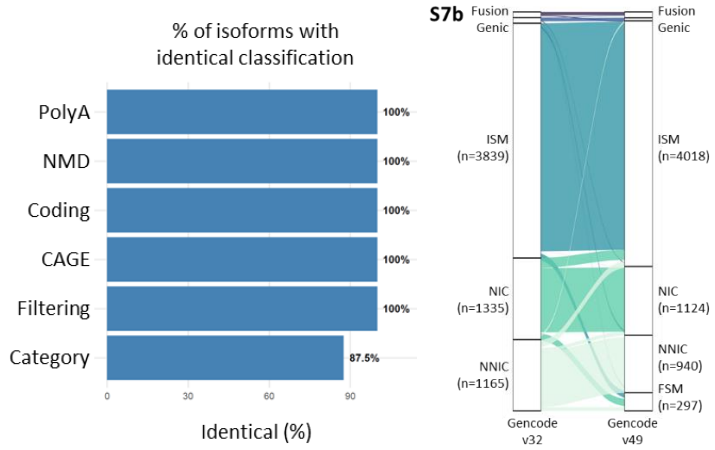
1.4

Reviewer Comment	During the discussion of DTU across tumor cell-types in Figure 2, the authors appear to pool all tumor cells across samples together for the analysis. It would be interesting to see if the identified isoform switches are driven by inter-tumor or intra-tumor heterogeneity. Are

	isoform switches patient specific? Because a simple Wilcoxon-rank sum test was used to identify DTUs, the authors were unable to correct for sample-specifics effects. It would be helpful to model these sample-specific effects explicitly, e.g. using a method like MAST.
Author Response	We thank the reviewer for this insightful comment regarding the influence of inter-tumor versus intra-tumor heterogeneity. In our initial analysis, given the limited co-occurrence of multiple malignant subtypes within individual samples, we pooled tumor cells across the cohort to enable detection of DTU across distinct malignant cell states. To ensure these signals were not driven by a subset of outlier patients, we performed a validation analysis at the individual level by calculating the mean isoform expression score for each of the 765 identified DTU transcripts within malignant cell subtypes on a per-patient basis. This analysis revealed a high degree of concordance between the pooled analysis and patient-level profiles, with cell-type-specific isoform preferences largely preserved across the majority of individuals. Specifically, the malignant subtype displaying the highest mean expression score for a given transcript (averaged across patients) showed strong concordance with the cell group identified in the original pooled analysis (New Supplementary Fig. 4e). While these results indicate that the DTU events reflect shared isoform usage patterns across malignant cell states, we also noted a smaller subset of DTU that were unique to specific individuals (n=207; New Supplementary Table 8).  We also appreciate your suggestion and have conducted further analysis to model sample-specific effects. An additional differential transcript usage analysis using MAST, which accounts for patient-level variability by including sample as a latent variable was done. This analysis was applied to tumor subpopulations. Notably, the results showed strong concordance with our original approach: of the 765 DTUs initially identified, 634 transcripts (81.6%) were also detected as differentially used across tumor subtypes in the MAST analysis (adjusted P-value < 0.05; New Supplementary Table 9), including the representative marker genes highlighted in Figure 2. On the basis of these findings, we have clarified this aspect in the revised Methods and Results sections and incorporated the MAST-based

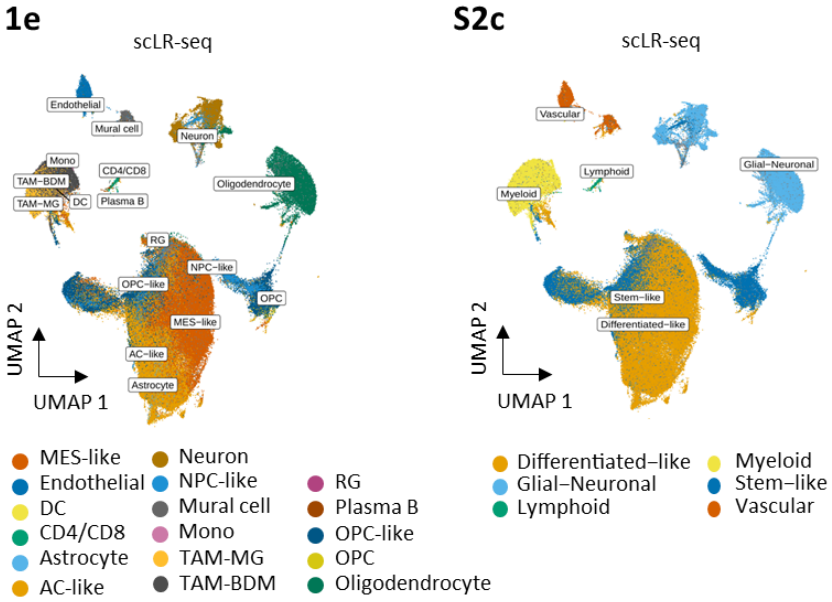
	analysis as an additional validation of our DTU results.
Excerpt from revised manuscript	*(P10) 'To determine whether the identified DTU events were representative of the cohort or driven by inter-patient variability … Collectively, these results suggest that most DTU events represent recurrent isoform usage patterns across malignant cell states.'* *Figure Legend:* *(S4e) Heatmap showing mean expression scores of 765 differentially used transcripts across malignant subtypes for each individual patient.*

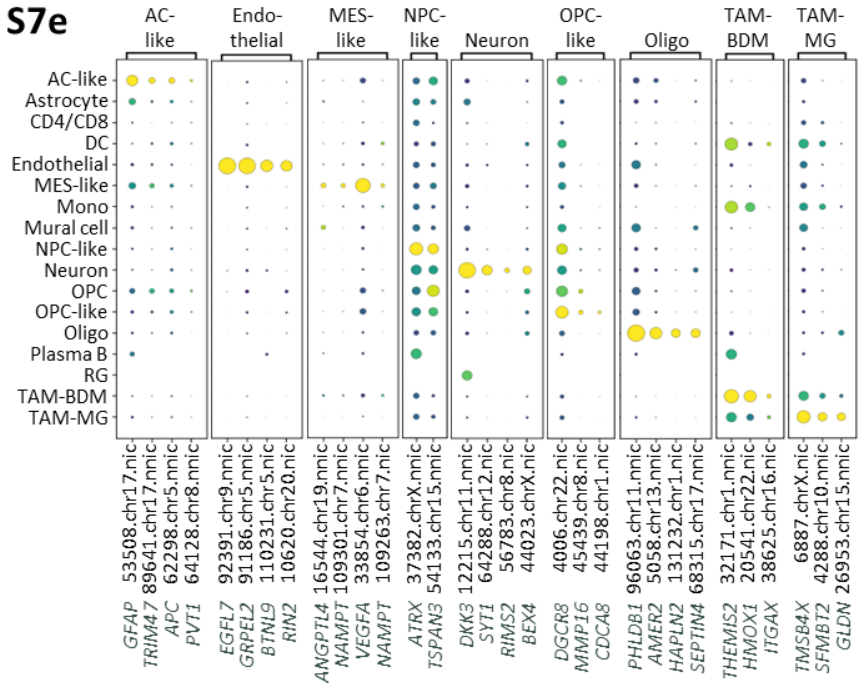
1.5

Reviewer Comment	Novel transcripts were identified by aligning reads against GENCODE v32 annotations. However, v32 is more than 5 years old (Oct 2019), and according to the GENCODE documentation, the most current version (v49) has more than 2 times as many annotated transcripts (227,462 vs 507,365). The reliance on v32 limits the novelty of this analysis. What percentage of the 6,523 novel isoforms are annotated in recent versions of GENCODE?
Author Response	Thank you for the thoughtful comments. We initially used GENCODE v32, corresponding to the 10x Genomics GRCh38-2020-A reference, to ensure consistency across all short- and long-read analyses. We agree that benchmarking against the most recent annotation is critical for a rigorous assessment of transcript novelty. To address this, we re-evaluated all novel isoforms using the latest GENCODE release (v49). Core transcript features including predicted protein-coding potential, lack of NMD classification, poly(A) signal support, and CAGE peak support, remain consistent between annotations. Furthermore, 87.5% of isoforms retained the same SQANTI3 structural category, with category changes observed only in a minority of cases due to newly added reference annotations. Importantly, of the 6,524 isoforms initially classified as novel based on GENCODE v32, 95.4% remain unannotated in GENCODE v49 (New Supplementary Fig. 7b), indicating that the vast majority of isoforms retain as novel under updated reference annotation. 

Excerpt from revised manuscript	*(P13) 'Given the continuous expansion of transcript annotations, we next examined whether the novel isoforms identified here remain unannotated in the most recent GENCODE release (v49). Among 6,524 isoforms initially classified as novel based on GENCODE v32, 6,227 (95.4%) remain unannotated (Supplementary Fig. 7b).'* *Figure Legend:* *(S7b) Sankey plot illustrating changes in transcript structural annotation between GENCODE v32 and v49.*

1.6

Reviewer Comment	When analyzing novel isoforms, the authors switch from discussion of AC-like, MES-like, OPC-like and NPC-like tumor cells to differentiated-like, glial-neuronal and stem-like. I do not see these cell types in the UMAP in figure 1D. Were cells re-clustered for this analysis? Why not use the original designation of tumor cell-type?
Author Response	We thank the reviewer for raising this point and apologize for the confusion caused. The cells were not re-clustered for the novel isoform analyses. Instead, all cells were annotated using a hierarchical labeling scheme that includes both fine-grained subtypes (17 cell types including AC-like, MES-like, OPC-like, and NPC-like; Fig. 1e) and broader, higher-level biological categories (six major cell types: Differentiated−like, Stem−like, Glial−Neuronal, Lymphoid, Myeloid, Vascular), as shown in Supplementary Fig. 2c (formerly Supplementary Fig. 4b).  For the novel isoform differential analyses presented in Fig. 4d and Supplementary Fig. 7f, we focused on the broader cell-type groupings, encompassing stem-like and differentiated-like tumor cells, as well as myeloid, vascular, and glial–neuronal cells. To address this concern and improve clarity, we now additionally present fine-grained subtype–level differential expression of novel isoforms in a new supplementary figure (New Supplementary Fig. 7e). We have also clarified the use of hierarchical cell-type annotations in the revised

	manuscript to avoid further confusion. 
Excerpt from revised manuscript	*(P6) 'First, we utilized GBMap, a unified glioblastoma atlas encompassing 1.1 million cells across 240 patients, to establish our reference cell-type annotations at both finer and broader resolutions by mapping the scSR-seq data to this extensive atlas (Supplementary Fig. 2a, 2b). Crucially, since the scSR-seq and scLR-seq datasets were generated from the same physical single cell library, high-confidence cell type labels were transferred to the scLR-seq data via exact cell barcode matching, enabling both fine-grained (Fig. 1e) and broader (Differentiated−like, Stem−like, Glial−Neuronal, Lymphoid, Myeloid, Vascular) annotations (Supplementary Fig. 2c).'* *Figure Legend:* *(1e) UMAP visualization of 182,222 GBM-derived single cells generated using scLR-seq, colored by 17 fine-grained cell types.* *(S2c) UMAP visualization of single cells generated using scLR-seq, colored by broad cell type categories.* *(S7e) Dot plots highlighting the top differentially expressed novel isoforms across fine-resolution cell states.*

1.7

Reviewer Comment	The Methods section is quite bare. There were many moments while reading the paper where I wanted to understand the methods used but could not find any explanation. For example, the authors mention using Wilcoxon rank-sum test for identifying isoform differences between tumor and non-tumor cells. What was the p-value cutoff? Was multiple hypothesis testing controlled for? Was testing done within each patient or across patients? I don't see anything in the methods section besides citation of a scanpy function. Similar remarks apply to other reported terms such as "expression score" and "marker score"

	which appear not to be defined in Methods. Additionally, things like the identification of "differentiated-like tumor cells" is not explained in Methods and thus it is unclear how the authors identified these cell types.
Author Response	Thank you for highlighting the need for more detailed methodological descriptions. Differential testing was performed across all patients, either between tumor cell clusters and their non-tumor counterparts, or among tumor subtypes (AC-like, MES-like, NPC-like and OPC-like). To control for multiple hypothesis testing, P-values were adjusted using the Benjamini-Hochberg (BH) analysis, and genes or isoforms with an adjusted P-value < 0.05 were considered statistically significant. To improve clarity and standardize terminology, we now refer to the statistical output of differential testing as the "Expression Score" (previously "Marker Score"). The score corresponds to the statistic comparing expression of a gene or isoform in the target group versus all other cells. We have now included score details in the revised Methods section. The identification of higher-level categories such as differentiated-like and stem-like tumor cells is now described in the Methods. Briefly, these are broader groupings for cell type annotation. Full details of the annotation of higher-level categories are provided in Supplementary Fig. 2c and the revised Methods.
Excerpt from revised manuscript	*(P23) 'Differential expression analyses were performed at both the gene and isoform levels using Scanpy (v1.9.3) and Seurat. In Scanpy, differential expression was computed using scanpy.tl.rank_genes_groups(). Expression scores were obtained from the test statistics (scores) returned by Scanpy, which quantify the difference in expression of each gene between the target group and all other cells. Marker identification in Seurat was conducted using FindAllMarkers() with the MAST method, incorporating sample identity as a latent variable (latent.vars) to account for inter-sample variability. Genes and isoforms with an adjusted P-value (Benjamini-Hochberg) < 0.05 and an absolute log fold change > 0.5 were considered statistically significant. To identify genes exhibiting differential transcript usage, we further selected genes for which distinct transcript compositions were observed in more than one group, indicating group-specific isoform usage.'*

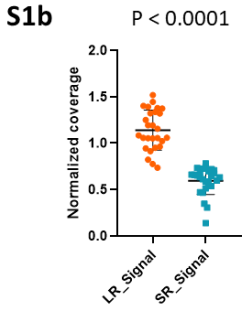
Minor Comments:

1.1

Reviewer Comment	When figures are cited, make sure the text matches the content of the figure. For example, it is unclear why Figure 1A is cited at the end of the first paragraph. Figure S1A should be cited after "long-read data provided more uniform coverage across gene bodies", and it is unclear why S1A is cited in the following sentence.

Author Response	Thank you. We have corrected the citation accordingly.

1.2

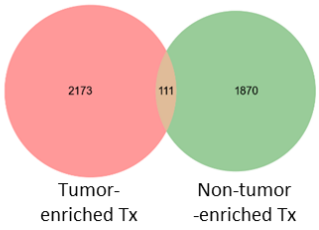
Reviewer Comment	Figure S1A shows that the depth of scLR-seq is around ~2-4x that of scSR-seq. This would be important to highlight. Is it due to the biotin enrichment?
Author Response	We thank the reviewer for this insightful observation regarding the coverage differences between the two modalities. By extracting the normalized signal from 10bp bins across the gene body, we quantified that scLR-seq provides a 1.63-fold higher signal density compared to scSR-seq (New Supplementary Fig. 1b). As you correctly point out, the biotinylated cDNA and pulled down on streptavidin beads is effective in depleting incomplete transcripts, primers, or non-cDNA fragments. The pull-down helps retain long fragments by reducing the presence of shorter nonspecific products, which could otherwise dominate the sequencing output. 
Excerpt from revised manuscript	*(P6) 'Compared to short-read data, which shows a coverage bias toward the 3' end, long-read data provided more uniform and higher coverage across gene bodies (1.63-fold increase; Supplementary Fig. 1a, Supplementary Fig. 1b). This improvement is likely due to the biotinylated cDNA capture and streptavidin bead pull-down, which effectively depletes incomplete transcripts, primers, and non-cDNA fragments.'* *(S1b) Dot plot showing normalized coverage distributions from TSS to TES (per 10bp bin) for scSR-seq and scLR-seq. Statistical significance was assessed using the Mann–Whitney U test.*

1.3

Reviewer Comment	Figure 1B shows short-read only barcodes but not long-read only barcodes. This seems to imply that there were no long-read read only barcodes. Is this correct? Can the authors clarify why they did not include long-read only barcodes?
Author Response	Thank you. Short-read and long-read sequencing were performed on the same 10x Genomics cDNA library derived from an identical pool of single cells. Cell barcodes were initially identified using short-read scRNA-seq data, which remains the standard and most reliable approach for barcode calling[3,4]. In our dataset, we observed a median of 82 barcodes per sample that were detected only in the long-read

	scRNA-seq data (Source Data Table 1B). To ensure consistency with the well-defined cell population established by short-read sequencing and to avoid potential artefacts from low-confidence barcode detection, these long-read-only barcodes were excluded from all downstream analyses. We have now clarified this rationale in the revised manuscript to prevent confusion.
Excerpt from revised manuscript	*(P23) 'Cell barcodes were defined based on matched short-read scRNA-seq data, and long-read reads were restricted to this curated set of validated barcodes derived from the same cDNA library.'*

1.4

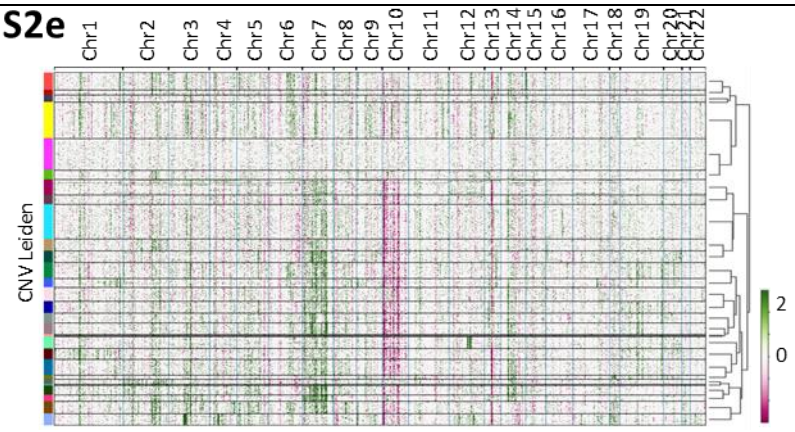
Reviewer Comment	Figure 2B is an upset plot with only a single intersection and so is uninformative. A Venn diagram may be more suitable.
Author Response	Thank you. We modified the upset plot to venn diagram accordingly. **S3c** 
Excerpt from revised manuscript	*(S3c) Venn diagram showing genes exhibiting differential transcript usage between tumor and non-tumor cells. The overlapping region indicates genes for which distinct isoforms are detected as differentially expressed in both tumor and non-tumor populations.*

1.5

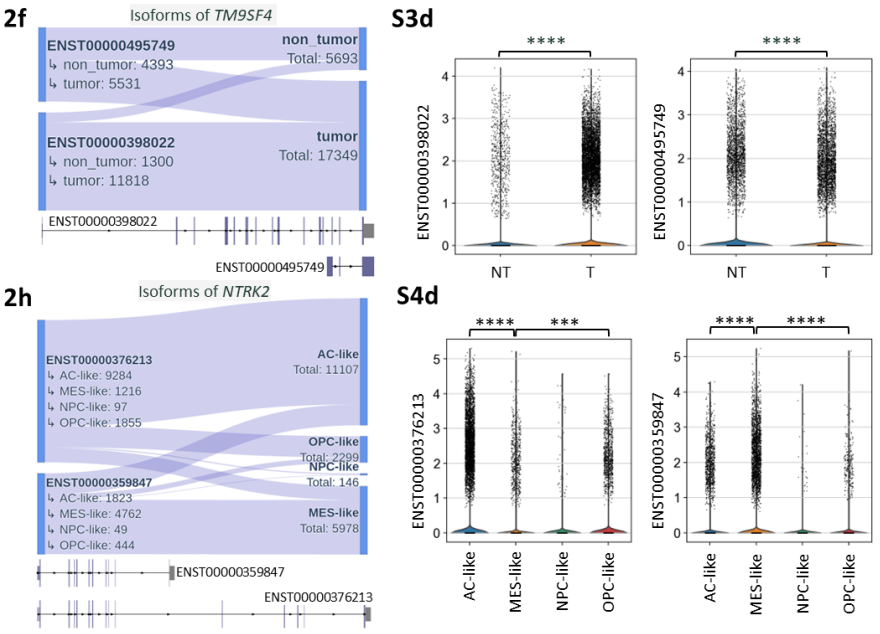
Reviewer Comment	Multiple "stringent criteria" are reported with "p-value < 0.05". Is this reported p-value corrected for multiple hypotheses? If so, how was the correction done? If not, can the authors justify why they used uncorrected p-values.
Author Response	Thank you. All reported p-values were corrected for multiple hypothesis testing using the Benjamini–Hochberg method. We have revised the manuscript to explicitly state that an adjusted P-value threshold of 0.05 was used throughout the analysis.

1.6

Reviewer Comment	I recommend changing the color intensity on Fig S1D. It is very difficult to see gains/losses by eye.
Author Response	We agree with the reviewer. The color scale in Supplementary Fig. 1d (Now Supplementary Fig. 2e) has been adjusted to improve contrast and visibility.

	
Excerpt from revised manuscript	*(S2e) Copy-number variation profiling across individual chromosomes and corresponding UMAP projection of scLR-seq data, with cells classified as normal or malignant based on inferCNV.*

1.7

Reviewer Comment	It would be helpful to display the expression of each transcript from figure 2E and 2H in a supplemental figure (e.g. in a violin plot) so that it is easy to verify that these isoform differences are statistically significantly different. The Sankey plots are illuminating but obscure quantitative differences.
Author Response	We thank the reviewer for the helpful suggestion. We have updated the Sankey plots in Figures 2E and 2H (New Figs. 2f, 2h) to include absolute cell counts and flow values. As suggested, we have added New Supplementary Figs. 3d, 4d, which display the expression distributions for each highlighted transcript. 
Excerpt from revised	*(2f) Representative examples of cell frequency distributions for isoforms differentially expressed between tumor and non-tumor cells.* *(2h) Representative examples of isoform expression patterns across*

manuscript	*tumor subtypes, highlighting subtype-specific isoforms.* *(S3d) Dot plot showing the expression of TM9SF4 isoforms in tumor and non-tumor cells.* *(S4d) Dot plot showing the expression of NTRK2 isoforms across tumor subtypes.*

Reviewer #2, ECR

Reviewer Comment	I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.
Author Response	Thank you very much for taking the time to review our manuscript and for providing constructive comments.

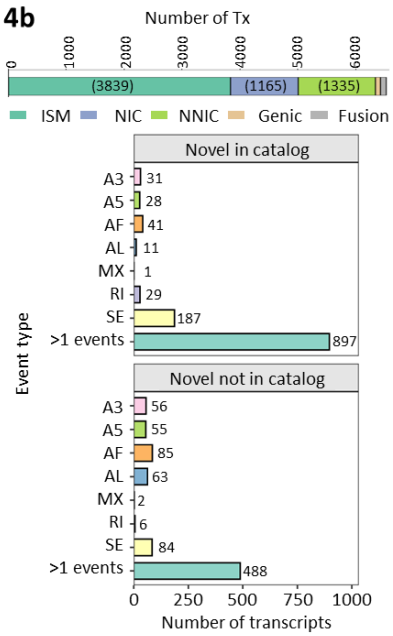
Reviewer #3

The authors present a GBM dataset combining long- and short-read single-cell sequencing. Given the lack of long-read single-cell datasets, this work is a valuable contribution that demonstrates the transcriptomic diversity of tumor cell populations, providing novel insights for personalized therapeutic targeting. The proposed framework for identifying extracellular and intracellular target pairs in patients is conceptually interesting and could be clinically relevant.

Major comments:

3.1

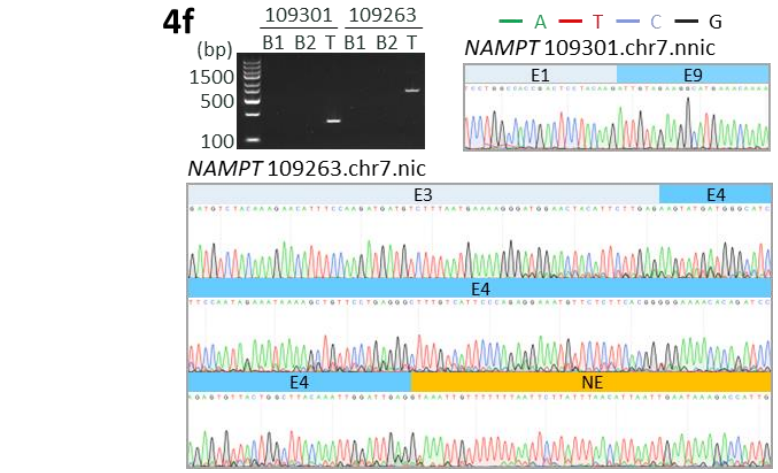
Reviewer Comment	One of the main novelties is the identification of 6,523 high-confidence new isoforms. However, this number includes ISM, NIC, and NNIC categories, which may be misleading, as ISM transcripts can represent truncated alignments to known isoforms rather than true novel transcripts. It would strengthen the manuscript to report the proportions of ISM, NIC, and NNIC, and to characterize novel isoforms in terms of the splicing events such as exon skipping, novel exon inclusion, intron retention. While some of this information is presented in the Supplementary Tables, including a summary in the main text could be helpful for readers.
Author Response	We appreciate the reviewer's recognition of the contribution this work makes to characterizing transcriptomic and isoform diversity, as well as the constructive comments provided. As suggested, we reported the proportions of ISM, NIC, and NNIC and further utilized SUPPA2 to decompose the full-length transcript structures to local splicing events of the novel isoforms (2,500 NIC and NNIC). Among these, 27% (n=679) of novel isoforms are defined by single splicing events (including skipped exon; retained intron; mutually exclusive exons; and alternative 5'/3' splice sites). Notably, the majority of novel isoforms (n = 1,385) featuring two or more concurrent alternative splicing events, highlighting the structural complexity captured by scLR-seq. This summary has been added to the Results

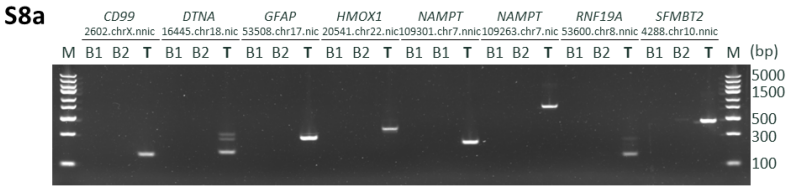
	section, Source Data Table 4B and visualized in New Fig. 4b. 
Excerpt from revised manuscript	*(P14) 'We further utilized SUPPA2 to decompose the full-length transcript structures to local splicing events of the novel isoforms (2,500 NIC and NNIC). Among these, 27% (n=679) of novel isoforms are defined by single splicing events (including skipped exon; retained intron; mutually exclusive exons; and alternative 5'/3' splice sites). Notably, the majority of novel isoforms (n = 1,385) featuring two or more concurrent alternative splicing events, highlighting the structural complexity captured by scLR-seq (Fig. 4b).'* *Figure Legend:* *(4b) Horizontal bars represent the total transcript count within each SQANTI3 structural category (upper), stratified by their constituent splicing events (bottom). SE, skipped exon; RI, retained intron; MX, mutually exclusive exons; A5/A3, alternative 5'/3' splice site; and AF/AL, alternative first/last exon. >1 Events denotes transcripts containing multiple concurrent splicing event types.*

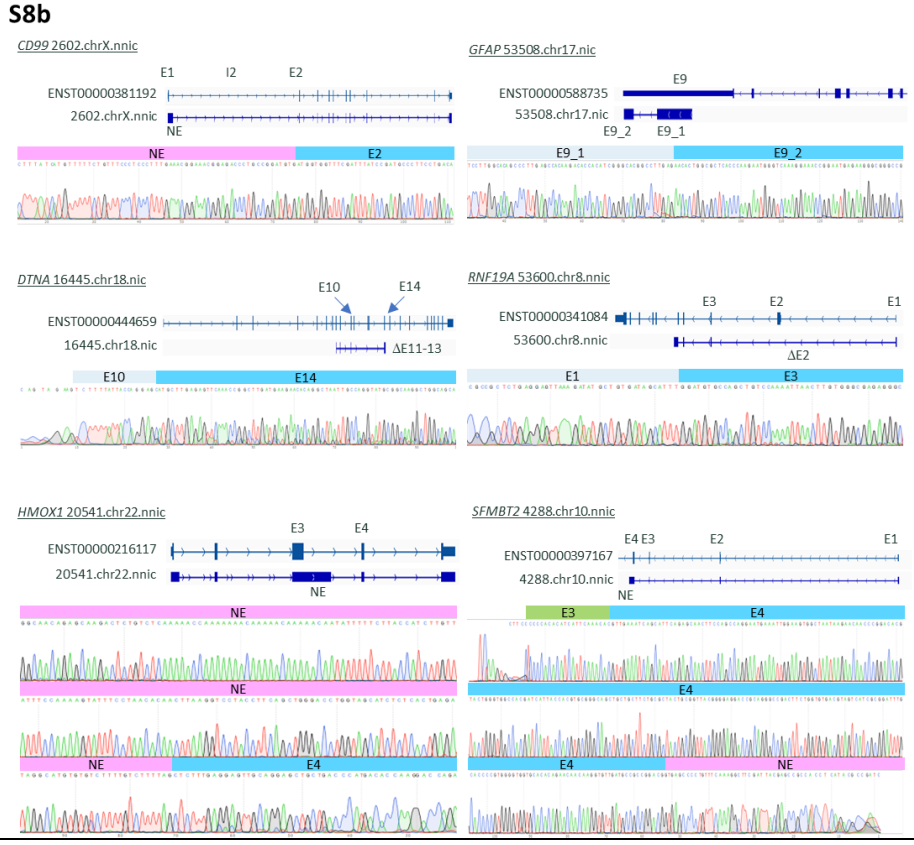
3.2

Reviewer Comment	No experimental (proteomic, immunopeptidomic, or targeted RT-PCR) validation is currently presented. To support the therapeutic claims, the authors could validate at least several key isoforms and/or their peptides. Alternatively, they could support their findings by checking whether the same isoforms or novel splice junctions can be detected in another GBM datasets.
Author Response	We thank the reviewer for this important comment and for the constructive suggestions to strengthen the support for our findings. To experimentally validate the isoforms identified by our scLR-seq pipeline, we performed the following complementary validation analyses, which are now included in the revised manuscript: 1. We designed isoform-specific primers spanning the neo-

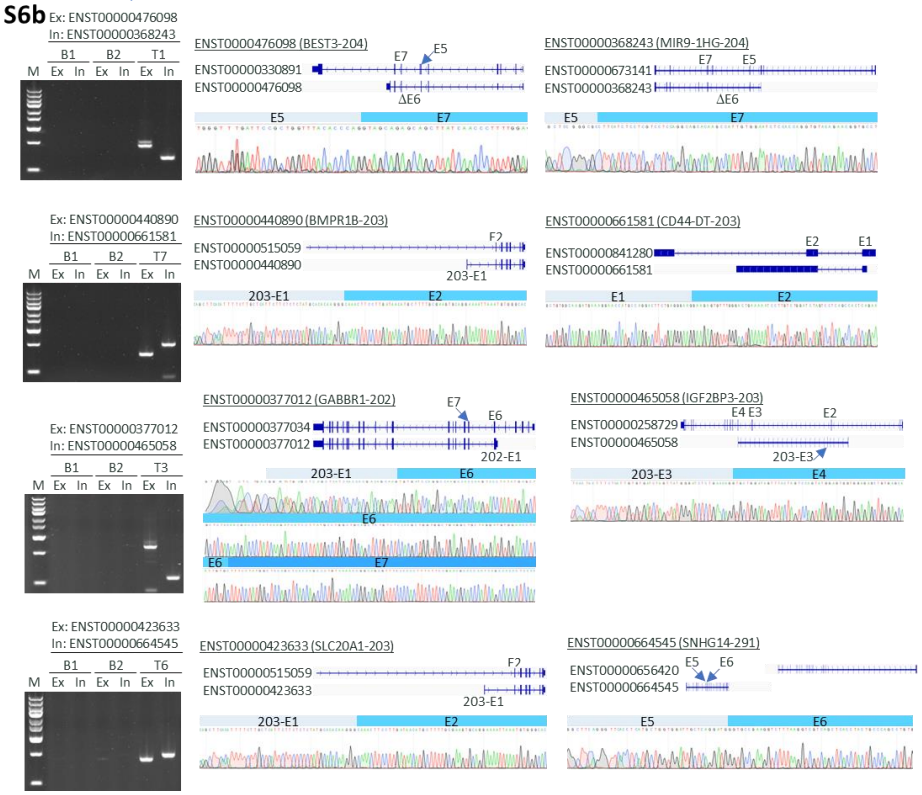
junctions of eight selected novel isoforms including the *NAMPT*. PCR followed by Sanger sequencing was performed using patient samples and control blood. All eight novel isoforms showed expected size (New Fig. 4f, Supplementary Fig. 8a) and exact sequence concordance with the predicted splice junctions (New Supplementary Fig. 8b) in tumor samples but not in control blood, demonstrating the high accuracy of our isoform reconstruction.



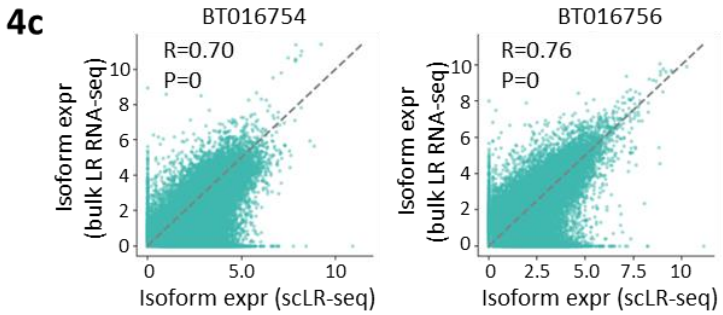




2. Moreover, we also designed primers targeting four pairs of annotated isoforms that form extracellular–intracellular targets highlighted in Fig. 3f and validated their specific expression in tumor samples from those individual (New Supplementary Fig. 6b).



3. To further corroborate isoform expression independently of single-cell analysis, we performed bulk direct RNA long-read sequencing (Oxford Nanopore Technology) on two GBM tumor samples. Isoform-level expression (n = 205,662, including 6,524 novel isoforms) quantified from bulk long-read data showed strong concordance with scLR-seq measurements (Pearson R = 0.70 and 0.76, P-value < 0.05), supporting the robustness and reproducibility of the isoform counts derived from our scLR-seq pipeline (New Fig. 4c).

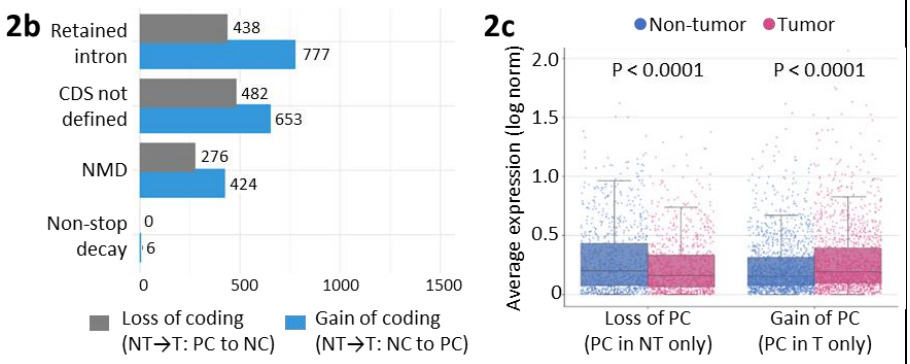


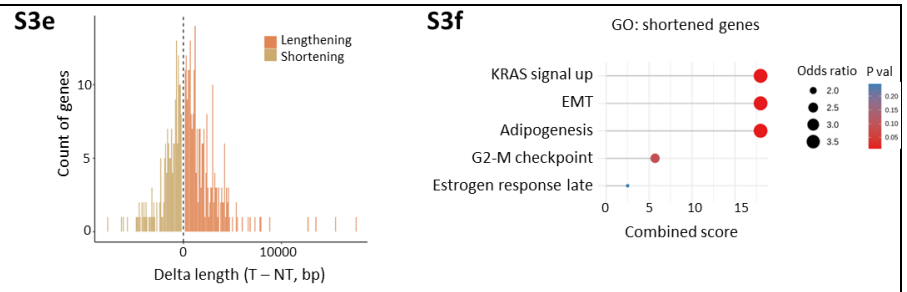
Together, these orthogonal validations, including junction validation and independent long-read bulk direct RNA-seq concordance, provide experimental support for both the existence and quantitative reliability of the identified isoforms. We hope these combined approaches address the reviewer's concern and strengthen confidence in our therapeutic

	prioritization framework.
Excerpt from revised manuscript	*(P12) 'Four selected target pairs were further validated by PCR and Sanger sequencing. All produced clear bands in matched tumor samples, and the sequences matched the expected reference, while no corresponding signals were detected in blood controls (Supplementary Fig. 6b).'* *(P15) 'To experimentally validate novel isoforms identified by scLR-seq, we selected eight novel isoforms for which junction-spanning primers could be designed. PCR followed by Sanger sequencing was performed using patient samples and control blood. All eight novel isoforms showed expected size (Supplementary Fig. 8a) and exact sequence concordance with the predicted splice junctions (Supplementary Fig. 8b) in tumor samples but not in control blood, demonstrating the high accuracy of our isoform reconstruction … The novel NAMPT splice junctions were confirmed by Sanger sequencing (Fig. 4f), illustrating the ability of scLR-seq to uncover previously unrecognized transcript variants.'* *(P14) 'To further validate our scLR-seq results, we performed direct RNA long-read sequencing (ONT) at bulk level on two tumor samples, obtaining approximately 27 million ONT long reads per sample. Isoform-level expression (n = 205,662, including 6,524 novel isoforms) measured by bulk RNA-seq showed high concordance with scLR-seq (Pearson R = 0.70 and 0.76, P-value < 0.05) (Fig. 4c), supporting the reliability of the isoform counts obtained from scLR-seq. Although bulk RNA-seq cannot resolve cell-type-specific expression, this validation confirms that the isoform-level patterns captured by scLR-seq are suitable for downstream differential analyses.'* *Figure Legend:* *(4f) Sanger sequencing validation of the novel splice junction in the newly identified NAMPT isoforms. Representative chromatograms and Sanger images are shown. E, exon; NE: neo exon; B, blood control; T: tumor sample.* *(S8a-b) PCR and Sanger sequencing validation of the novel splice junction in another six newly identified isoforms. Representative chromatograms (A) and Sanger images (B) are shown. E, exon; NE: neo exon; B, blood control; T: tumor sample.* *(S6b) Sanger sequencing validation of selected extracellular–intracellular target pairs in tumor samples and control blood. Representative chromatograms and Sanger images are shown.* *(4c) Scatter plot showing the correlation of log-normalized transcript expression between scLR-seq and bulk long-read RNA-seq.*

3.3

Reviewer Comment	The statement "protein-coding isoforms were expressed in tumor cells, whereas non-coding or retained-intron isoforms predominated in non-tumor cells" should be supported by quantitative evidence, such as numbers in the text or a figure. Additional analyses such as

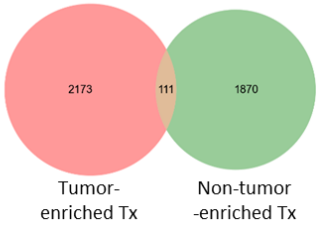
	comparisons of transcript or UTR lengths between switched isoforms could also be useful. It would also be interesting to check how shifts in transcript biotypes correlate with changes in gene expression.
Author Response	Thank you for the valuable comment. We now supplemented with quantitative evidence for biotype switching. In summary, 942 genes transitioned from protein-coding in non-tumor cells to non-coding biotypes in tumor cells, primarily through nonsense-mediated decay (NMD) (n=276) and intron retention (n=438). Conversely, higher number of genes (n=1,530) switched to protein-coding isoforms in tumor cells from non-coding origins in non-tumor cells, involving nonsense-mediated decay (n=424) and retained intron (n=777) biotypes (New Fig. 2b). Consistently, we observed that changes in protein-coding isoform status were significantly associated with expression differences, with gains of protein-coding genes corresponding to higher gene expression levels (New Fig. 2c). These results demonstrated widespread alternative splicing and isoform switching in tumors.  We also appropriate the reviewer's suggestion regarding UTR lengths analysis. We agree that while scLR-seq enables direct resolution of full-length transcript isoforms, scSR-seq provide complementary insights for detecting alternative polyadenylation (APA) isoforms. To assess whether tumor-associated isoform switching identified by scLR-seq is also reflected at the 3′ end of transcripts, we analyzed scSR-seq data using scUTRquant[2], which quantifies APA isoforms based on a curated 3′ UTRome annotation. This analysis identified 426 genes exhibiting differential transcript usage between tumor and non-tumor cells (New Supplementary Table 4). Analysis of APA-associated structural changes revealed 255 3′ UTR lengthening and 171 shortening events, involving 852 transcripts across these 426 genes (tumor relative to non-tumor cells) (New Supplementary Fig. 3e), indicating that differential transcript usage is frequently accompanied by gene-specific changes in transcript architecture. Notably, transcripts with shortened 3′ UTRs were enriched in tumor-associated pathways, including KRAS signaling and EMT (New Supplementary Fig. 3f), consistent with previous reports linking APA-mediated UTR shortening to oncogenic activation[5].

	 Moreover, we compared DTU events detected by scSR-seq with those identified by scLR-seq. Notably, 128 tumor-specific and 146 non-tumor-specific DTU isoforms detected in scSR-seq were exactly matched by full-length isoforms resolved by scLR-seq, demonstrating cross-platform concordance.
Excerpt from revised manuscript	*(P8) 'We observed significant biotype switching ... These results demonstrated widespread alternative splicing and isoform switching in tumors.'* *(P8) 'While scLR-seq enables the direct characterization of full-length transcript isoforms, scSR-seq data provide complementary insights for detecting alternative polyadenylation (APA) isoforms. To assess whether tumor-associated transcript usage identified by scLR-seq is also reflected at the 3′ end of transcripts, we analyzed scSR-seq using scUTRquant34, which quantifies APA isoforms based on a curated 3′ UTRome annotation. This analysis identified 426 genes exhibiting differential transcript usage between tumor and non-tumor cells (Supplementary Table 4). Analysis of APA-associated structural changes revealed 255 3′ UTR lengthening and 171 shortening events, involving 852 transcripts across these 426 genes (tumor relative to non-tumor cells) (Supplementary Fig. 3e), indicating that differential transcript usage is frequently accompanied by gene-specific changes in transcript architecture. Notably, transcripts with shortened 3′ UTRs were enriched in tumor-associated pathways, including KRAS signaling and EMT (Supplementary Fig. 3f), consistent with previous reports linking APA-mediated UTR shortening to oncogenic activation35. Moreover, 128 tumor-specific and 146 non-tumor-specific differential transcript usage (DTU) isoforms detected in scSR-seq were exactly matched by full-length isoforms resolved by scLR-seq, demonstrating cross-platform concordance (Supplementary Table 4).'* *Figure Legend:* *(2b) Bar plot showing the number of events involving loss of protein-coding annotation (protein-coding in non-tumor cells only) and gain of protein-coding annotation (protein-coding in tumor cells only).* *(2c) Box plot showing gene-level expression associated with protein-coding annotation switches (Loss of PC genes: n=942; gain of PC genes: n=1530) between non-tumor and tumor cells. Mean log-normalized expression per gene is shown for each condition. Statistical significance was assessed using the Mann–Whitney U test.* *(S3e) Histogram showing the number of genes and corresponding*

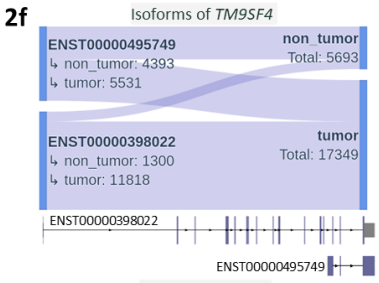
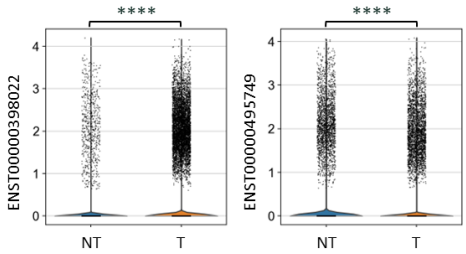
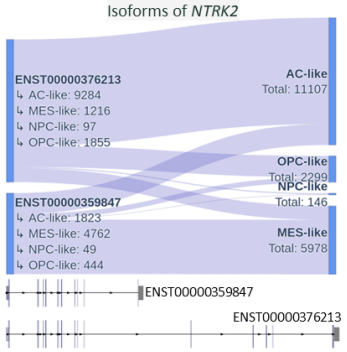
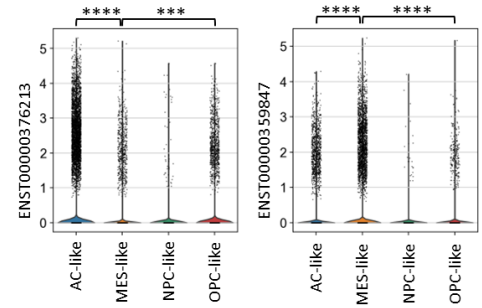
	changes in 3' UTR length between tumor and non-tumor cells. *(S3f) Top enriched gene ontology terms (MSigDB: Hallmark gene sets) for 171 genes with shortened 3' UTRs in tumor cells.*

Minor Comments:

3.1

Reviewer Comment	Figure 2B is not particularly informative in its current form.
Author Response	Thank you. We modified the upset plot to venn diagram accordingly. **S3c** 
Excerpt from revised manuscript	*(S3c) Venn diagram showing genes exhibiting differential transcript usage between tumor and non-tumor cells. The overlapping region indicates genes for which distinct isoforms are detected as differentially expressed in both tumor and non-tumor populations.*

3.2

Reviewer Comment	Figures 2E and 2H are difficult to interpret due to the large differences in cell numbers between groups, labelling numbers would improve clarity.
Author Response	Thank you for the useful guidance. We have updated the Sankey plots in Figures 2e and 2h (New Figs. 2f, 2h) to include absolute cell counts and flow values accordingly. Moreover, we have added New Supplementary Figs. 3d and 4d, which display the expression distributions for each highlighted transcript. **2f**  **S3d**  **2h**  **S4d** 

Excerpt from revised manuscript	*(2f) Representative examples of cell frequency distributions for isoforms differentially expressed between tumor and non-tumor cells.* *(2h) Representative examples of isoform expression patterns across tumor subtypes, highlighting subtype-specific isoforms.* *(S3d) Dot plot showing the expression of TM9SF4 isoforms in tumor and non-tumor cells.* *(S4d) Dot plot showing the expression of NTRK2 isoforms across tumor subtypes.*

3.3

Reviewer Comment	The abbreviation "DET" in Figures 2C and 2G is unclear.
Author Response	Thank you for pointing this out. We have revised the figure legends to clearly define all terms: DET, differentially expressed transcripts; DEG, differentially expressed genes.

3.4

Reviewer Comment	In addition to the four major categories (FSM, ISM, NIC, NNIC), the list of novel transcripts also includes "genic" isoforms which may represent technical artifacts and should be addressed or excluded.
Author Response	We agree with your comments. Our initial summary was subjected to the quality control including CAGE-peak support, PolyA signal support, and NMD-prediction filtering, which didn't specifically include filtering on structure category. We agree that genic isoforms can be difficult to interpret biologically. Therefore, they were excluded from all subsequent analyses of tumor-specific novel isoforms. We have updated the Methods section to explicitly state that genic isoforms are excluded from the final set.
Excerpt from revised manuscript	*(P27) 'Genic isoforms are excluded from the final set.'*

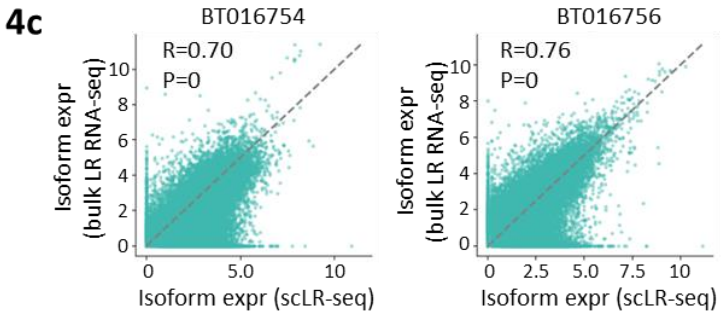
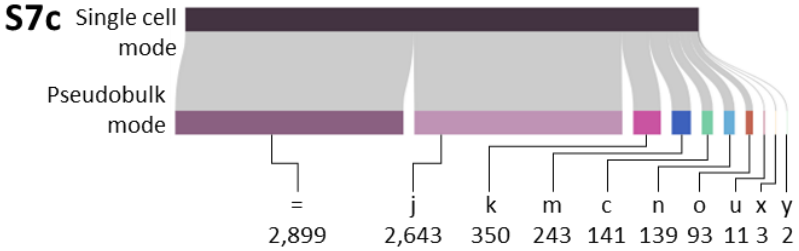
Reviewer #4

The authors present a map of glioblastoma isoform diversity using long-read single-cell analysis. The work is compelling and adds to a growing literature using this approach for isoform analysis in a variety of tumors.

I have a few suggestions that I would like to see addressed to improve the rigor and impact of their findings which I have offered below:

4.1

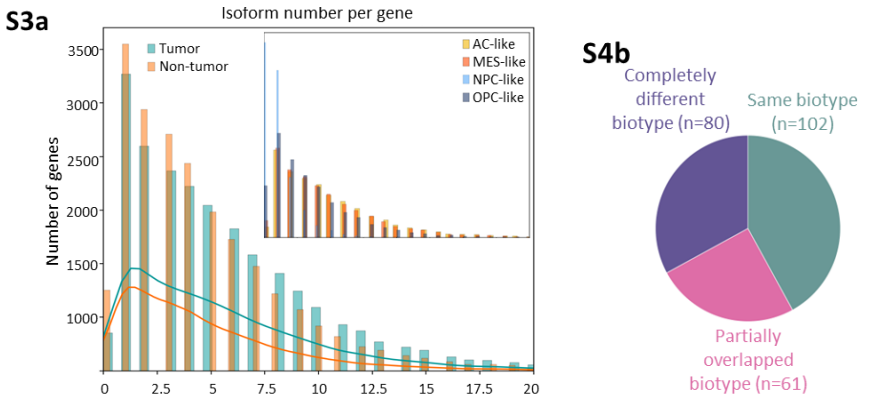
Reviewer Comment	Before getting to their long-read single-cell analysis, the authors should start by performing a differential splicing analysis on bulk RNA sequencing performed on these samples and determine whether single-cell replicates findings on the spliting landscape of GBM obtained from bulk seq through something like a Sankey diagram showing the changes in differential splicing events from bulk to scRNA-seq data.

Author Response	We thank the reviewer for highlighting the value of this work and its potential application of long-read approaches to isoform analysis across cancers, as well as for the valuable suggestions. Despite limitations in sample availability, we were able to perform direct RNA long-read sequencing (ONT) at bulk level on two tumor samples, obtaining approximately 27 million ONT long reads per sample. Briefly, we re-isolated RNA from these tumors and carried out bulk long-read sequencing. The reads were then mapped to a reference transcriptome that included both the annotated reference transcripts and our newly identified novel isoforms (n = 6,524). The bulk RNA-seq results confirmed our scLR-seq findings, showing a high correlation at the isoform level between scLR-seq and bulk RNA-seq (r = 0.70 and r = 0.76; p=0; New Fig. 4c). Although the normal brain samples are not available and bulk RNA-seq does not allow for distinguishing between different cell types or separating tumor from non-tumor cells, the strong isoform-level concordance supports the reliability of the isoform counts obtained from scLR-seq and the downstream differential transcript usage identified.  To address the reviewer's concern regarding bulk versus single-cell comparisons, we used pseudobulk scLR-seq data as a bulk-equivalent reference to evaluate whether splicing patterns observed at single-cell resolution recapitulate bulk-level splicing landscapes. Whereas our initial novel isoform discovery was performed in a barcode-aware, single-cell mode, isoform structures were additionally reconstructed at the sample level by aggregating reads into pseudobulk profiles to perform novel isoform structure discovery in a bulk-equivalent mode. Overall, bulk- and single-cell–derived novel isoforms' structure showed substantial structural concordance as assessed by *gffcompare*, including 2,899 exact matches and 2,643 junction-level matches (New Supplementary Fig. 7c). This analysis has been incorporated into the revised manuscript. 
Excerpt	*(P13) 'Novel isoform structures were firstly detected using a barcode-*

from revised manuscript	*aware mode, in which transcript models are reconstructed for each cell to enable cell-type-specific isoform discovery'* *(P14) 'In addition to barcode-aware reconstruction of novel isoform at cell-level, reads were aggregated into sample-level pseudobulk profiles to perform novel isoform structure discovery in a bulk-equivalent mode. Overall, bulk- and single-cell–derived novel isoforms' structure showed substantial structural concordance as assessed by gffcompare, including 2,899 exact matches and 2,643 junction-level matches (Supplementary Fig. 7c).'* *(P14) 'To further validate our scLR-seq results, we performed direct RNA long-read sequencing (ONT) at bulk level on two tumor samples, obtaining approximately 27 million ONT long reads per sample. Isoform-level expression (n = 205,662, including 6,524 novel isoforms) measured by bulk RNA-seq showed high concordance with scLR-seq (Pearson R = 0.70 and 0.76, P-value < 0.05) (Fig. 4c), supporting the reliability of the isoform counts obtained from scLR-seq. Although bulk RNA-seq cannot resolve cell-type-specific expression, this validation confirms that the isoform-level patterns captured by scLR-seq are suitable for downstream differential analyses.'* *Figure Legend:* *(4c) Scatter plot showing the correlation of log-normalized transcript expression between scLR-seq and bulk long-read RNA-seq.* *(S7c) Sankey plot summarizing classification of isoforms reconstructed in barcode-aware single-cell mode relative to isoforms reconstructed from pseudobulk (sample-level aggregated) reads. Class codes indicate the structural relationship between single-cell–derived isoforms and pseudobulk models: "=" denotes exact intron-chain matches; "j" indicates isoforms sharing at least one splice junction match; "k" denotes containment of reference; "m" indicates retained matched intron(s); "c" represents containment within a larger transcript; "n" denotes retained novel intron structure; "o" indicates generic exon overlap on the same strand; "u" represents unknown/intergenic transcripts; "x" denotes overlap with exons on the opposite strand; and "y" indicates transcripts containing a reference within its intron.*

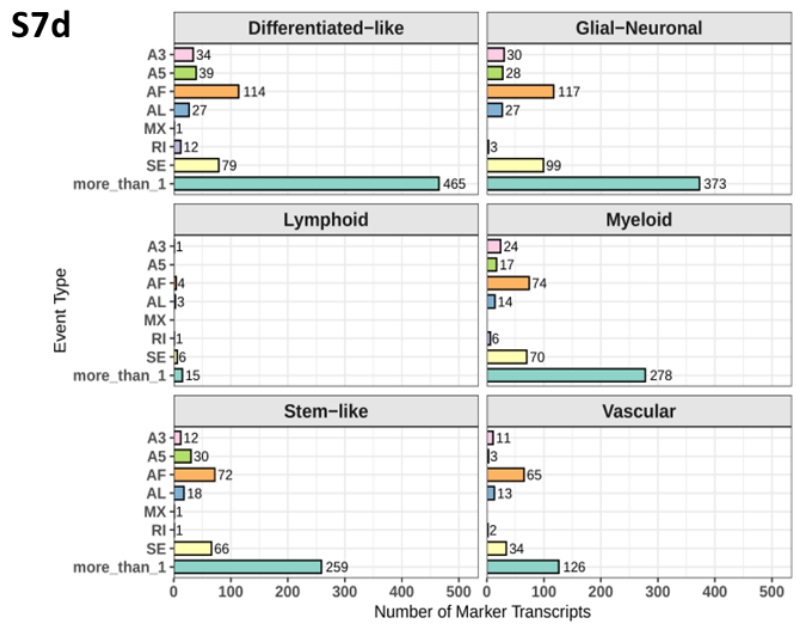
4.2

Reviewer Comment	The authors introduce their splicing results by summarizing median number of transcript isoforms per protein-coding gene and the portion of total transcripts that exhibited isotope shift - these metrics should be compared between tumor cells vs. non-tumor cells in their samples and across tumor cells of the four molecular subtypes.
Author Response	Thank you for the suggestion. Following the recommendation, we compared the median number of transcript isoforms per protein-coding gene across tumor subtypes (New Supplementary Fig. 3a). We observed that AC-like, MES-like, and OPC-like cells exhibited similar isoform number distributions, which were higher than those in NPC-like cells. In addition, we examined biotype switches among tumor subtypes to further characterize subtype-specific transcript diversity (New

	Supplementary Fig. 4b). 
Excerpt from revised manuscript	*Figure Legend:* *(S3a) Bar plot showing the number of expressed transcripts per gene in tumor versus non-tumor cells (isoforms expressed in at least 10 cells), as well as across malignant subtypes.* *(S4b) Pie chart categorizing these genes based on whether they exhibit completely distinct biotypes, partially overlapping biotypes, or the same biotype across malignant subtypes.*

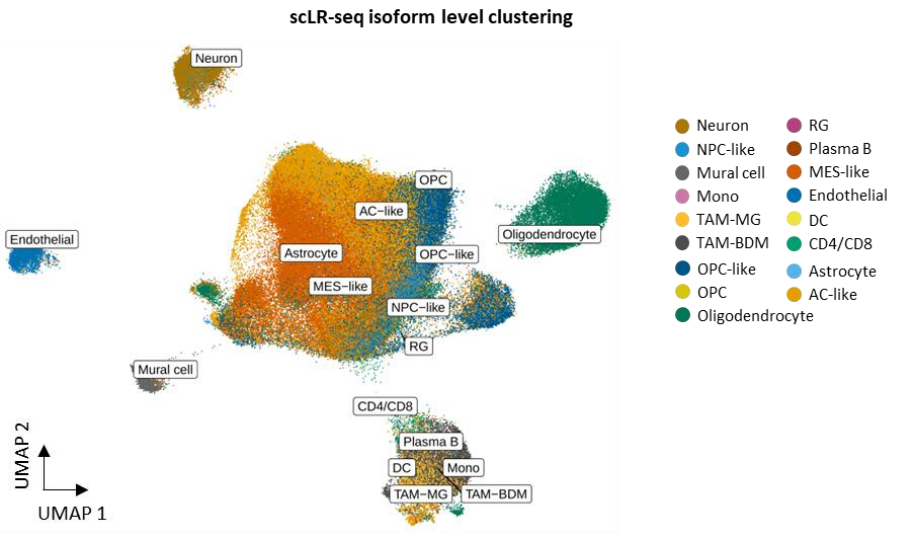
4.3

Reviewer Comment	Can the authors report the distribution of the various types of alternative splicing events (e.g. skipped exon, alternative 3′ splice site, alternative 5′ splice site, and retained intron) in tumor cells of the various subtypes vs. stromal cells.
Author Response	Thank you for the valuable suggestion. We further utilized SUPPA2 to map full-length transcript structures to local splicing events of the novel isoforms (2,500 NIC and NNIC). Among these, 27% (n=679) of novel isoforms are defined by single splicing events (including skipped exon; retained intron; mutually exclusive exons; and alternative 5'/3' splice sites) (New Fig. 4b). Notably, the majority of novel isoforms (n = 1,385) featuring two or more concurrent alternative splicing events, highlighting the structural complexity captured by scLR-seq. As suggested, we further examined the distribution of alternative splicing event types by analyzing significantly differentially expressed isoform markers across major cell types. The comparative distribution of splicing event categories is summarized in New Supplementary Fig. 7d.

	
Excerpt from revised manuscript	*(P14) 'Differential expression analysis identified many novel isoforms with distinct transcript structures, over half displaying combinations of multiple splicing events (Supplementary Fig. 7d), which were highly cell-type-specific and linked to key functional pathways (Supplementary Fig. 7e; Supplementary Table 12).'* *Figure Legend:* *(S7d) Bar plots showing the distribution of splicing event types stratified by major cell types.*

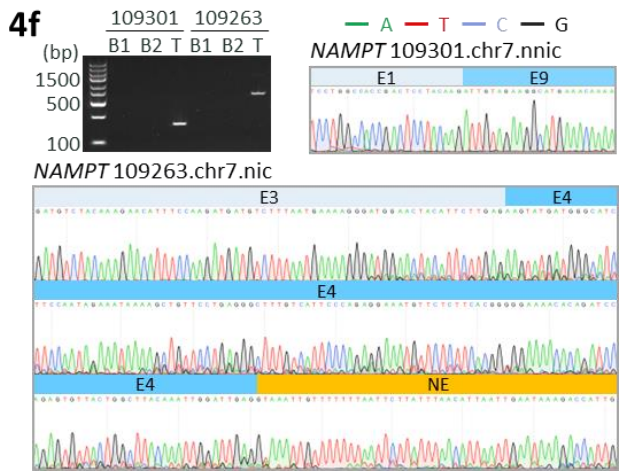
4.4

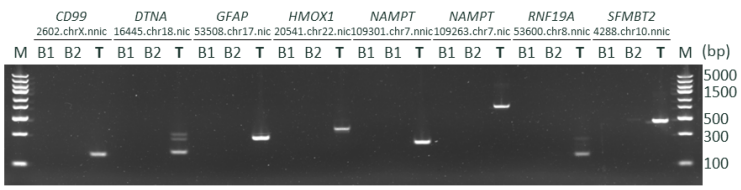
Reviewer Comment	While the authors perform nice analysis of specie variants amongst known single cell molecular GBM subtypes (AC-like, MES-like, etc.), did the authors find that splice variants reveal any intriguing clustering into subtypes that have not yet been described.
Author Response	We thank the reviewer for this thoughtful comments. We performed unsupervised clustering of malignant cells based on transcript-level expression profiles derived from scLR-seq data and compared the resulting clusters with established GBM molecular state annotations. Overall, isoform-based clustering largely recapitulated the known AC-like, MES-like, OPC-like, and NPC-like states, and we did not observe robust evidence for additional, previously unrecognized cell types defined solely by alternative splicing. This result is biologically expected, as GBM molecular states were originally defined by strong lineage-associated gene expression programs reflecting developmental origin and core transcriptional identity. Consequently, while splice variant usage contributes substantially to regulatory and functional diversity within malignant cells, it does not override lineage-defining transcriptional signatures. We have clarified this point in the Results and Discussion and emphasize that the strength of scLR-seq lies in resolving splicing programs within established GBM states.

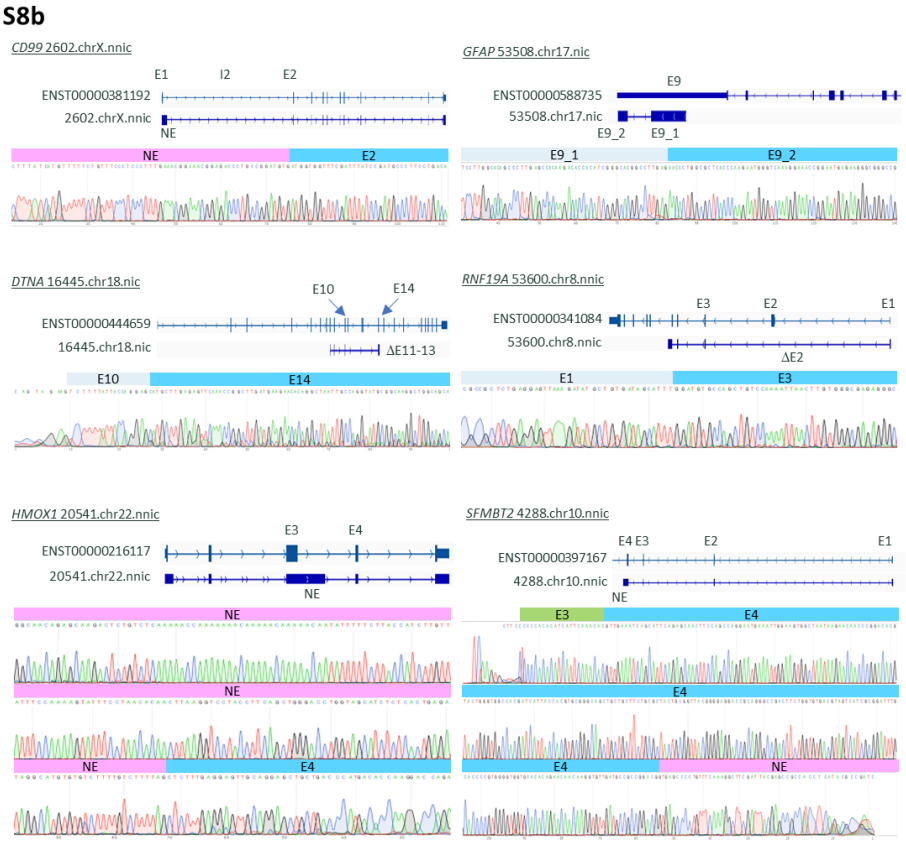
	
Excerpt from revised manuscript	*(P17) 'Furthermore, isoform-level clustering was highly consistent with gene-level clustering and recapitulated established GBM molecular states, reflecting the dominance of lineage-associated transcriptional programs while revealing additional regulatory diversity through alternative splicing.'*

4.5

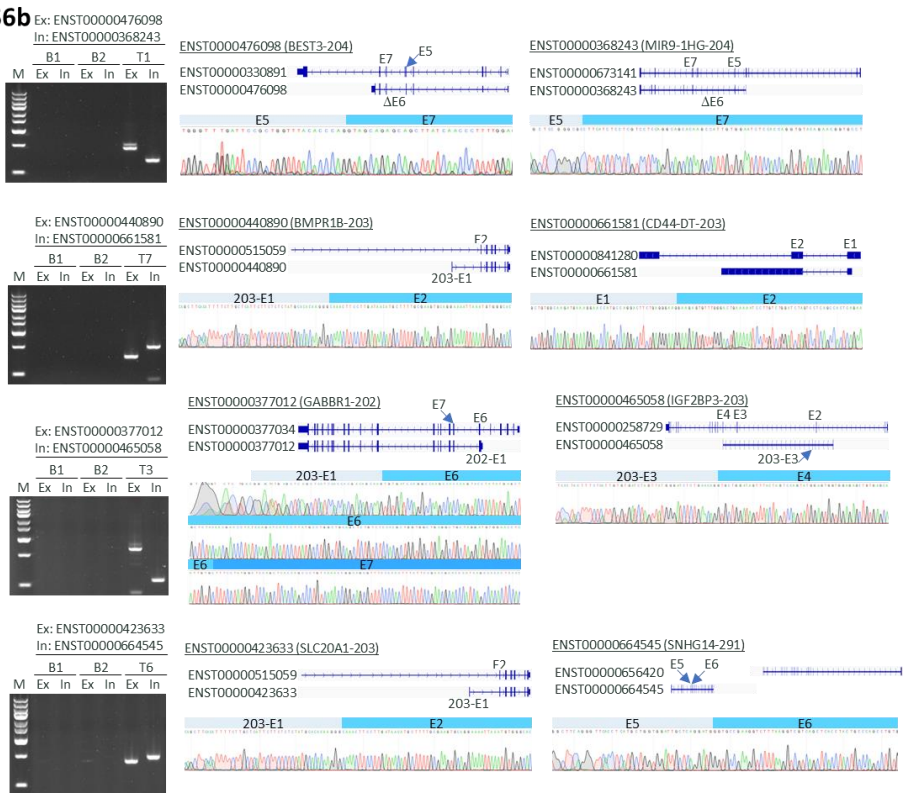
Reviewer Comment	Protein level corroboration on these samples should be performed for splice variants distinguishable by antibodies.
Author Response	We thank the reviewer for this important suggestion. This study is based on precious glioblastoma clinical specimens with limited tissue availability. In addition, isoform-distinguishing antibodies are not available for most splice variants identified here, particularly those differing only by internal exon usage or untranslated regions. As an alternative validation strategy, we performed Sanger sequencing to confirm splice junctions of representative isoforms, thereby providing direct validation of the transcript models resolved by scLR-seq. We designed isoform-specific primers spanning the neo-junctions of eight selected novel isoforms including the *NAMPT*. PCR followed by Sanger sequencing was performed using patient samples and control blood. All eight novel isoforms showed expected size (New Figs. 4f, Supplementary Fig. 8a) and exact sequence concordance with the predicted splice junctions (New Supplementary Fig. 8b) in tumor samples but not in control blood, demonstrating the high accuracy of our isoform reconstruction.







Moreover, we also designed primers targeting four pairs of annotated isoforms that form extracellular–intracellular targets highlighted in Figure 3F and validated their specific expression in tumor samples from those individual (New Supplementary Fig. 6b).



We hope these efforts address the concern raised. We believe that this data establishes a robust framework and reliable reference resource for systematic isoform discovery in GBM using clinically derived samples.

Excerpt from revised manuscript

(P12) 'Four selected target pairs were further validated by PCR and Sanger sequencing. All produced clear bands in matched tumor samples, and the sequences matched the expected reference, while no corresponding signals were detected in blood controls (Supplementary Fig. 6b).'

(P15) 'To experimentally validate novel isoforms identified by scLR-seq, we selected eight novel isoforms for which junction-spanning primers could be designed. PCR followed by Sanger sequencing was performed using patient samples and control blood. All eight novel isoforms showed expected size (Supplementary Fig. 8a) and exact sequence concordance with the predicted splice junctions (Supplementary Fig. 8b) in tumor samples but not in control blood, demonstrating the high accuracy of our isoform reconstruction.

The novel NAMPT splice junctions were confirmed by Sanger sequencing (Fig. 4f), illustrating the ability of scLR-seq to uncover previously unrecognized transcript variants.'

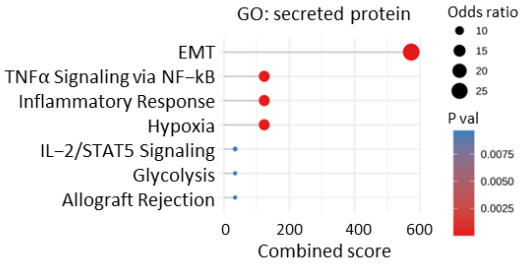
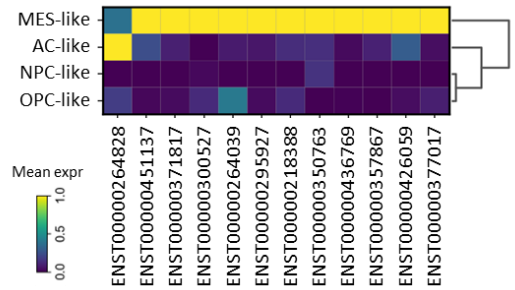
Figure Legend:

(4f) Sanger sequencing validation of the novel splice junction in the newly identified NAMPT isoforms. Representative chromatograms and Sanger images are shown. E, exon; NE: neo exon; B, blood control; T: tumor sample.

(S8a,b) PCR and Sanger sequencing validation of the novel splice junction in another six newly identified isoforms. Representative

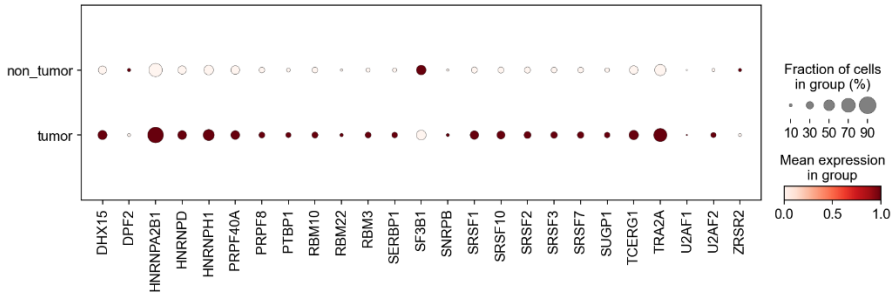
	chromatograms (a) and Sanger images (b) are shown. E, exon; NE: neo exon; B, blood control; T: tumor sample. *(S6b) Sanger sequencing validation of selected extracellular–intracellular target pairs in tumor samples and control blood. Representative chromatograms and Sanger images are shown.*

4.6

Reviewer Comment	The authors should present results on what they found in term of splice variants for secreted proteins, particularly cytokines and ECM proteins which have known splice variants with roles in cancer biology.
Author Response	We appreciate your suggestion. To address this point, we analyzed isoforms encoding secreted proteins among the 740 differentially expressed isoforms (DEIs) enriched in malignant cells. We identified 61 secreted isoforms, including 36 enriched in MES-like cells, 17 in AC-like cells, and 8 in OPC-like cells (New Supplementary Table 10). Notably, secreted isoforms associated with extracellular matrix and EMT-related pathways were predominantly enriched in the MES-like subtype (New Supplementary Figs. 5b, 5c), consistent with the established mesenchymal characteristics in GBM, which include enhanced cell-matrix interactions, tissue remodeling, and invasive behavior. **S5b**  **S5c** 
Excerpt from revised manuscript	*(P11) 'In total, 61 DEIs encoded secreted proteins, including 36 enriched in MES-like cells, 17 in AC-like cells, and 8 in OPC-like cells (Supplementary Table 10). Notably, secreted isoforms involved in extracellular matrix organization and EMT-related pathways were predominantly enriched in the MES-like subtype (Supplementary Fig. 5b, 5c), consistent with the established mesenchymal characteristics in GBM, which include enhanced cell-matrix interactions, tissue remodeling, and invasive behavior.'* *Figure Legend:* *(S5b) Top enriched gene ontology terms (MSigDB Hallmark gene sets)*

	for the 61 identified secreted proteins. *(S5c) Heatmap showing expression of EMT-associated secreted proteins across the four malignant subtypes.*

4.7

Reviewer Comment	They should interrogate RNA-binding proteins to identify key regulators of splicing diversity in GBM.
Author Response	We thank the reviewer for the suggestion, to explore potential regulators of splicing diversity in GBM, we examined our in-house whole-genome sequencing data and identified one patient (BT019346) harboring mutations in RBP gene *TCERG1* (p.R989* and p.R232*) which have been implicated in transcriptional regulation and RNA processing[6]. We further assessed the expression of RBPs in our dataset and observed that many are upregulated in tumor cells, consistent with prior reports[6,7].  While these observations suggest that dysregulation of RBPs may contribute to the altered splicing landscape in GBM, our current data does not allow us to assign specific RBPs to individual splicing events or transcript usage changes. We have therefore added a discussion of these findings and their potential implications for splicing regulation in GBM to the revised manuscript, while acknowledging this as an important direction for future investigation.
Excerpt from revised manuscript	*(P20) 'RNA-binding proteins (RBPs) are known to play central roles in regulating alternative splicing and transcript diversity in cancer. In our cohort, whole-genome sequencing identified mutations in RBP-related gene TCERG1 in one individual, which might play a role in transcriptional regulation and RNA processing. In addition, multiple RBPs showed elevated expression in tumor cells compared with non-tumor cells, which may contribute to the extensive isoform diversity observed in GBM … future studies integrating perturbation-based experiments or joint modeling of RBP binding and splicing outcomes will be necessary to delineate the causal regulators of tumor-specific splicing programs.'*

Once again, thank you very much for your time and valuable comments. We look forward to your favourable response.

Yours sincerely,

Dr. Brian H.Y. Chung

Clinical Associate Professor; School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong

Chief Medical and Scientific Officer; Hong Kong Genome Institute, Hong Kong

Email: bhychung@hku.hk bhychung@genomics.org.hk

Dr. Aya El Helali

Clinical Assistant Professor; Department of Clinical Oncology, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong

E-mail: ahelali@hku.hk

Reference

1. Rautenstrauch, P. & Ohler, U. Shortcomings of silhouette in single-cell integration benchmarking. *Nat. Biotechnol.* 1–5 (2025) doi:10.1038/s41587-025-02743-4.

2. Fansler, M. M., Mitschka, S. & Mayr, C. Quantifying 3′UTR length from scRNA-seq data reveals changes independent of gene expression. *Nat. Commun.* **15**, 4050 (2024).

3. You, Y. *et al.* Identification of cell barcodes from long-read single-cell RNA-seq with BLAZE. *Genome Biol.* **24**, 66 (2023).

4. Lebrigand, K., Magnone, V., Barbry, P. & Waldmann, R. High throughput error corrected Nanopore single cell transcriptome sequencing. *Nat. Commun.* **11**, 4025 (2020).

5. Mayr, C. & Bartel, D. P. Widespread Shortening of 3′UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell* **138**, 673–684 (2009).

6. Mrid, R. B., Guendouzi, S. E., Mineo, M. & Fatimy, R. E. The emerging roles of aberrant alternative splicing in glioma. *Cell Death Discov.* **11**, 50 (2025).

7. Zhang, Q., Ai, Y. & Abdel-Wahab, O. Molecular impact of mutations in RNA splicing factors in cancer. *Mol. Cell* **84**, 3667–3680 (2024).

Dear reviewers,

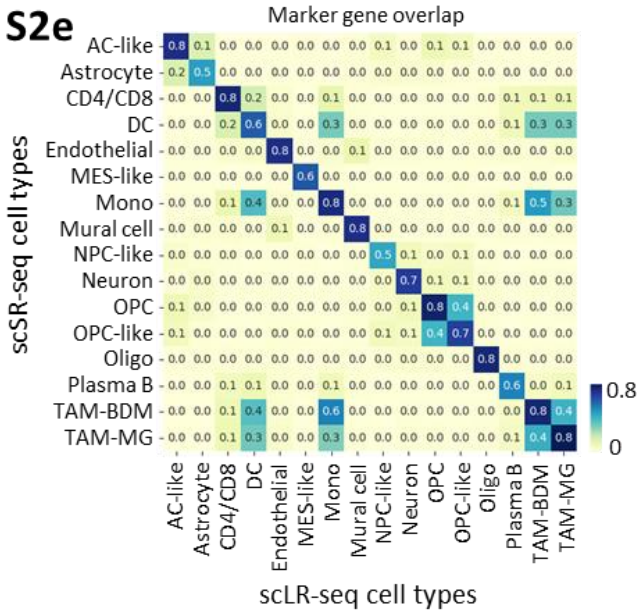
We appreciate the opportunity to further refine our work based on your constructive comments. We have taken your valuable advice and modified the manuscript accordingly. Please find below a point-by-point response to the comments.

Reviewer #1:

I appreciate the authors' efforts in addressing my previous comments, particularly in clarifying how the scSR-seq and scLR-seq datasets were processed and compared. These clarifications will greatly help readers follow the analytical workflow. I have one remaining concern that I believe can be addressed with a small revision.

Major Comments:

Reviewer Comment	In support of the claim that "scLR-seq provides enhanced resolution for cell-state identification compared to scSR-seq", the authors now use scib-metrics package instead of visual analysis of UMAP structure. This is more principled than evaluating UMAPs, but it is still unclear how the scib-metrics support this claim. For example, the biological conservation metric reflects an integration method's ability to conserve cell-type information between individual and integrated datasets. This metric largely reflects the capability of the integration method (here harmony) to integrate each individual sequencing type. Thus it is unclear how it could support the claim that one sequencing type provides enhanced resolution compared to another sequencing type. Moreover, the differences (0.59 vs 0.58) are very small, and without some kind of statistical framework it is unclear if they are meaningfully different. To provide evidence that scLR-seq offers enhanced resolution of cell-states, the authors would need to perform some kind of analysis which shows that known biological variation is better captured within scLR-seq compared to scSR-seq, etc. Perhaps there's a cell-state which does not cluster together in the scLR-seq but is delineated in the scLR-seq. However, I don't think the claim that "scLR-seq provides enhanced resolution for cell-state identification compared to scSR-seq" is that important to the originality of this paper, and the data required to support this claim is outside the scope. The main novelty is the enhanced isoform resolution rather than cell-state identification, which is mostly bootstrapped from the GBM atlas anyway. I therefore suggest that the authors simply remove the claim that scLR-seq provides enhanced resolution for cell-state identification, and instead use the scib-metrics to make the more modest claim that integration quality was comparable across platforms, ensuring downstream isoform-level comparisons are not confounded by differential integration performance. The paper cited for the scib-metrics discussion is "Shortcomings of silhouette in single-cell integration benchmarking" (https://doi.org/10.1038/s41587-025-02743-4) , a paper which seems to be criticizing Silhouette based metrics for evaluating integration

	results. Ironically, this paper explicitly criticizes scib-metrics (although Silhouette metrics are only a small part of the total batch correction and bio conservation score, so I do not put much weight on the criticism). The original scib-metrics Nature Methods paper should be cited instead (https://doi.org/10.1038/s41592-021-01336-8).
Author Response	We thank the reviewer for this important and insightful comment. As the reviewer correctly points out, the scib-metrics framework was developed to evaluate a method's ability to preserve cell-type information during integration. Given the marginal differences observed, we agree that the appropriate interpretation is that, under identical analytical conditions, both platforms demonstrate comparable levels of biological conservation and batch-effect control, rather than providing evidence for enhanced resolution of one modality over the other. To provide a principled validation of the biological variation captured by scLR-seq relative to scSR-seq, we calculated the Jaccard Similarity Index for the marker genes identified in each cell cluster across both datasets. This analysis reveals a high degree of transcriptional consistency, with median Jaccard scores of 0.80 along the diagonal (New Supplementary Fig. 2e). These results indicate that scLR-seq preserves established GBM cell-state identities with an integration quality comparable to the scSR-seq gold standard.  Accordingly, we have revised the manuscript to remove the statement that "scLR-seq provides enhanced resolution for cell-state identification compared to scSR-seq." As suggested, we now use the scib-metrics results and the high cross-platform marker concordance to demonstrate that scLR-seq exhibits comparable performance to scSR-seq, ensuring that the downstream identification of cell-type–specific isoforms is not confounded by differences in integration performance.

	We have also updated the citation with the original scIB benchmarking study (Luecken et al., Nature Methods 2022) to appropriately reference the evaluation framework.
Excerpt from revised manuscript	*(P7) 'Using a Jaccard similarity analysis of the marker genes for each cluster, we observed a high concordance between the scSR-seq and scLR-seq datasets, with a median Jaccard index of 0.80 along the diagonal (Supplementary Fig. 2e).* *Furthermore, we leveraged the scib-metrics framework to calculate the integration scores for biological conservation and batch correction. We observed that scLR-seq achieved comparable scores in both biological conservation (0.59 vs. 0.58) and batch correction (0.53 vs. 0.50) compared to scSR-seq (Supplementary Fig. 2g). These results demonstrates that scLR-seq preserves the established cell-state identities of GBM with integration quality comparable to scSR-seq, supporting its suitability for downstream identification of cell-type–specific isoforms.'* *Figure Legend:* *(2e) Heatmap showing Jaccard Similarity Indices for marker genes identified in scSR-seq versus scLR-seq. Scores were calculated based on the top 100 of significantly enriched markers for each cell state.*

Minor comments:

Reviewer Comment	The number of reported cells on line 174 is "18,2222", which should presumably be 182,222 "differentials" on line 249 should be "differential" "the" should be removed before MAST on line 280.
Author Response	We thank the reviewer for carefully identifying these errors. The corrections have been made accordingly in the revised manuscript.

Reviewer #3:

Reviewer Comment	The authors have done an amazing job addressing all my concerns. The additional experiments and analyses have significantly strengthened the manuscript. I have only one very minor comment: I still suggest slightly adjusting the wording regarding novel isoforms. In one paragraph, the authors first refer to "6,524 high-confidence novel isoforms (ISM, NIC, and NNIC)," and later to "the novel isoforms (2,500 NIC and NNIC)", which is a bit confusing. I would suggest reserving "novel isoforms" for NIC and NNIC only, and either using a different term (e.g., "non-FSM isoforms", if no other categories are included in the analysis) or simply listing the categories explicitly (ISM, NIC, and NNIC) when referring to all three.
Author Response	We sincerely thank the reviewer for the positive feedback and for the helpful suggestions. To improve clarity, we now refer to NIC and NNIC transcripts specifically as "novel isoforms," while using "non-FSM isoforms" when describing the ISM, NIC, and NNIC. These

	revisions have been made consistently throughout the revised manuscript.
Excerpt from revised manuscript	*(P14) 'Applying these quality filters, we retained 6,524 high-confidence non-FSM isoforms (ISM, NIC, and NNIC) … Among the 6,524 isoforms that were not initially classified as full-splice match based on GENCODE v32 … Overall, bulk- and single-cell–derived non-FSM isoforms' structure showed substantial structural concordance as assessed by gffcompare …* *Isoform-level expression (n = 205,662, including 6,524 ISM, NIC and NNIC isoforms) measured by bulk RNA-seq showed high concordance with scLR-seq.* *Differential expression analysis identified many non-FSM isoforms with distinct transcript structures, over half displaying combinations of multiple splicing events.'* *(P15) 'we analyzed the 6,524 high-confidence non-FSM isoforms (ISM, NIC and NNIC) identified in GBM'* *(P28) 'Predicted ORFs for non-FSM isoforms (ISM, NIC, and NNIC) were obtained using SQANTI3 (v5.2.2) by translating the corresponding amino acid sequences'*

Once again, thank you very much for your time and valuable comments. We look forward to your favourable response.

Yours sincerely,

Dr. Brian H.Y. Chung

Clinical Associate Professor, Department of Paediatrics and Adolescent Medicine, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong;

Chief Medical and Scientific Officer; Hong Kong Genome Institute, Hong Kong

Email: bhychung@hku.hk bhychung@genomics.org.hk

Dr. Aya El Helali

Clinical Assistant Professor, Department of Clinical Oncology, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong

E-mail: ahelali@hku.hk