nature portfolio

Peer Review File

Splice isoform-perturbation coupled to single cell transcriptome profiling reveals functions of microexons in neurogenesis and autism-linked pathways

Corresponding Author: Professor Benjamin Blencowe

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, Dupas and colleagues present CHyMErA-seq by extending the CHyMErA screening method developed earlier by the group. CHyMErA-seq allows one to delete neural microexons in a high-throughput manner using CRISPR/paired gRNAs and evaluate the functional consequence of the genetic perturbation in individual cells by split-pool single nuclei RNA-seq. This strategy is used to test a set of 37 conserved and neuron-specific microexons in mESCs during in vitro neurogenesis. The study finds neuronal transcriptomic shifts caused by microexon deletion and further examines the functional roles of two microexons (Gfra1 and Clasp1) in neurodevelopment. Deletion of microexons has also been implicated in altering expression of ASD-risk gene networks, providing insights into how dysregulated microexon splicing may contribute to disease. The functional evaluation of dynamic alternative splicing patterns, especially high-throughput technologies, is highly valuable. Indeed, within this class, the biological requirements of microexons are particularly important due to their potential connection with neurodevelopmental disorders.

While the authors have made great strides in the execution of this high throughput strategy, this reviewer has some concerns regarding the approach itself, and the robustness of the downstream gene expression changes, and their relevance to maturation and disease. In particular, the manuscript will benefit from strengthening key aspects, including microexon deletions, as the entire body of work depends on this strategy, characterization of the approach, and phenotyping of neurons with exon deletions.

CHyMErA-seq screening:

1. Microexon deletion: Figure 1, the authors show %del to account for deletions. This is likely an overestimation as smaller PCR amplicon and typically favored. Even if one ignores this bias, the estimated deletion frequency (e.g., top panels in Figure 1d) does not seems to be consistent with the gel images in some cases.

Moreover, the level of deletion in a bulk sample is not an especially useful metric, as what is more relevant is %cells with homozygous deletions. Carefully evaluating (for example with ddPCR), and reporting these numbers will help the audience understand the robustness of this dual Cas approach. The same problem applies to the application of CHyMErA-seq, where it is not clear to what levels homozygous microexon deletions occurs.

2. The representation of guide pairs in the sequenced nuclei library (and potential sources of bias) needs further illustration. The authors have targeted 37 exons with 9 guide pairs each. How many cells were recovered per guide pair for each target that passed QC should be included. How is the correlation of DGEs between guide pairs targeting the same microexon affected by cell#?

3. A major caveat of the CHyMErA-seq approach, when applied to mESCs, is the low editing efficiency, and cells with microexon deletion cannot be isolated directly. The use of QDA to classify cells as KO vs non-perturbed based on gene expression profile is a creative way to deal with the KO vs non-perturbed issue. However, this is an indirect workaround, and the reliability of the approach remains to be determined.

The authors compared CHyMErA-seq derived gene expression changes in response to microexon deletion and that determined by bulk RNA-seq of clonal samples, and the correlation is quite moderate (Figure 4d,e and Extended Data Figure 5d), which is very concerning.

4. It seems that no replicate screening was performed, which adds additional challenge to evaluate the robustness of the screening.

5. In some parts, the description of bioinformatics analysis is not very clear. For example, it is difficult to figure out how perturbation calling was done precisely. The authors might consider using less narratives and include some mathematical descriptions.

Phenotyping:

6. Are the 37 microexons frame preserving? This reviewer assumes so, but it will be helpful to clearly indicate this.

7. The authors have positive control gRNA pairs that target constitutive exons, but do not provide much detail on the transcriptomic characterization of these cells.

8. Pseudotime analyses indicates that the KO cells are broadly enriched in the neuroepithelia clusters (Fig 2d). This suggests there are differentiation defects and KO cells are stuck in the neuroepithelium state? As the microexons are typically not included to significant levels at this stage of differentiation, how do the authors explain KO cells enrichment in neuroepithelia. Moreover, this seems incongruent with the idea that microexons deletions increase maturation.

9. More can be done to solidify/validate gene expression changes observed in the snRNA data. While the authors decided to focus on Gfra1 and Clasp1 for additional analyses and made efforts to generate clonal cell lines with microexon deletion (or skipping/inclusion using dCasRx-RBM25). The evidence provided to support the functional role of each exon is less than compelling in the current form. For example, for Gfra1, the transcritomic changes inferred from the screen essentially has no correlation with that from analysis of clonal cell lines (Extended Data Figure 5d). For Clasp1, the impact of microexon on downstream phosphorylation appears to be quite moderate.

10. The authors used a KS-test to evaluate alterations of gene expression correlation with neuron differentiation rank. It is quite difficult to understand what these significant differences really mean.

11. Ptk2 microexons are known to dramatically elevate autophosphorylation. Do the authors detect signatures of modulated Ras signalling/other downstream signaling activity? Similarly, do the authors observe elevated morphology or migration changes downstream of Ptk2 microexon modulation?

12. GDNF signaling is known to promote survival and growth of neurons. The authors find that Gfra1 microexon deletion modestly increases neurite growth and stimulates expression of mature neuronal genes. Is this a consequence of increased GDNF signalling? i.e., does microexon inclusion decrease Gfra1 signaling? Can the authors supplement neurons with GDNF to amplify the neuritogenesis phenotypes?

13. Camta1 microexon does not appear to be included in mESCs. Nevertheless, the changes in gene expression upon deletion in mESCs appears to enrich downregulated genes that are affected in autism (Fig 5d). this feature appears significant only in the mESCs and not in neurons. Similar results are also obtained for Clasp1 microexon as well. How do the authors explain observations of transcriptomic phenotypes in mESCs that do not express much of the microexon, but not in neurons from the same founders?

Other minor points:

14. Explain hgRNA

15. Line 689, missing “)” after CRISPOR.

16. Line 755, paring --> pairing

17. Regulon scores were z-normalized and p values **and** were calculated for each pairwise comparison by two-tailed unpaired t-test followed by Benjamini-Hochberg correction.

18. Fig. 2b, colors for some of the clusters are too light to be visible

19. Fig. 2e, what are the columns?

20. Fig. 4 legend, “Empirical cumulative distribution functon(ECDF) of differentially expressed genes (DEGs, FDR < 0.05) associated with microexon inclusion compared to **inclusion** for mESCs.”

Reviewer #2

(Remarks to the Author)
This manuscript presents CHyMErA-seq, a platform designed to profile transcriptomic changes at single-cell resolution in response to targeted splice isoform perturbations. Applying this approach to 36 neuronal microexons, the authors identify consequences of these alternative splicing events on gene expression dynamics during neurogenesis. Several microexon-containing genes tested here are implicated in neuron projection development and converge on signaling pathways such as FAK-Src. Functional analyses reveal that specific microexon deletions studied here (similar to the effect of microexons studied elsewhere) lead to increased neurite outgrowth and in general a premature activation of neuronal gene expression programs. The authors also observe parallels between microexon-dependent transcriptomic signatures and gene expression patterns observed in autism spectrum disorder, in line with previous studies.
This study builds on a large body of research from the Blencowe and other labs, which provided the basis for the experimental tools applied here. In the present study, the authors improved editing efficiency and coupled the CHyMErA pipeline with scRNA-seq readout. In addition, most if not all the main results obtained using scRNA-seq here confirm or are in line with previous studies, as the authors themselves point out repeatedly (activation of neuronal gene expression signatures, increased neurite outgrowth, link to ASD,…). On the one hand, this is an encouraging proof-of-concept of the experimental approach, but on the other hand raises questions about the novelty of the findings. In my opinion, it will be necessary to delineate more specifically which results are novel, which are confirmatory, and where the single-cell and isoform-level perspective (which is the main "selling point" of this study) allowed new biological insights into a previously studied biological question.
Specific points and suggested analyses are detailed below.
Especially considering the necessity of pointing out the novelty of the findings compared to previous studies, it would be advisable to be more specific in the wording of the results and avoid vague statements such as the ones below:
Page 12: "Overall, these data confirm transcriptomic signatures detected in the CHyMErA-seq screen and further support important new functions of microexons in neurogenesis"
Page 14: "Collectively, these findings highlight the importance of isoform-specific protein functions of genes in the control of critical and dynamic processes during neurogenesis, and their dysregulation in neurodevelopmental disorders."
Page 15: "Applying this system to neuronal microexons revealed unexpected functional properties [which ones?] of this important class of alternative splicing in the context of neurogenesis, as well as new insights into the multilayered links between microexons and brain disorders [I do not agree that the manuscript contains enough novel insights in this respect to justify this sentence]."

Specific comments:
Page 8/Fig 2: "The Ptk2 microexons (denoted as 'A', 'B' and 'C') display coordinated splicing patterns in vivo…": Their splicing is coordinated, but here the authors delete one of them at a time: Consequently, deletion of microexon alone will create an artificial protein that is not usually produced (because they are included or excluded together). Is there in vivo evidence for the existence of transcripts with two of the three microexons included and one skipped? In addition, the authors state that "notably, we observe transcriptomic signatures for all three microexons targeted in the Ptk2 gene (Fig. 2e)". First of all, this should be rephrased to be less vague and more accurate: (a) …transcriptomic signatures of the DELETION of all three microexons…; (b) …we observe DISTINCT transcriptomic signatures…; (c) these apparent differences between the effect of the deletion of the three microexons should be discussed (in particular, B and C are located in the same domain, but yet deletion of each of them shows very different effects on the transcriptome).

Fig 3a and b: Figure legend and description in the text is very minimalistic. It is not clear to me if the focus here is on gene expression changes between Gfra1 microexon knockout and non-targeting or along the differentiation/pseudotime. It sounds like Fig3a shows only differentially expressed genes (between KO and control – are all of them used for the GSEA in 3b?) and how the expression of these genes changes (only in Gfra1 KO? cells) with pseudotime. Given the drastic changes in expression between pseudotime bins, which population was used for the differential expression testing between KO and control? All cells together? Only neurons?

Page 10: "Collectively, these results support a role for the Gfra1 microexon in suppressing a transcriptomic program to prevent premature activation of neuritogenesis." I am aware that also other microexon deletions have been shown to lead to a more neuron-like phenotype, however, this is still a bit counterintuitive to me. Forcing the non-neuronal splicing outcome (skipping) leads to a more neuron-like gene expression signature and phenotype. I would appreciate if the authors could elaborate on this in the discussion, especially given the fact that this seems to be a general consequence of neuronal nmicroexon deletion. In addition, an analysis of the commonalities and differences in the gene expression signatures of different microexon deletions would be helpful to strengthen the postulated convergence on FAK-Src pathway.

Page 13: "A 'module score' representing expression of genes implicated in ASD…" – Are these genes implicated because of their premature expression (which would be in line with the transcriptomic signature of microexon deletion) or through loss-of-function/mutations? This distinction is relevant for the interpretation of the observed overlap in the list of genes.

Page 13: "We observe that over pseudotime these premature maturation signatures show an early increase in expression of ASD-linked genes…" - Specifically ASD-linked genes? Or general neuronal and the observed association with ASD-linked genes is due to the fact that ASD-linked genes tend to have neuronal functions? What was the background gene list used in the enrichment analyses here?

Minor comments:
Page 3: "These challenges are particularly relevant to a program of evolutionarily conserved, 3-27 nucleotide-long 'microexons', which are predominantly spliced in the nervous system 14, 15." – Should be "predominantly spliced in" or

“predominantly included”.

Page 13: “We also observe enrichment of ASD risk genes identified in three studies in the differential gene expression signatures detected in response to perturbation of Clasp1(A), Ptprf(A), Ptk2(A), Camta1 and Gfra1 microexons in our bulk RNA-seq” – should be clarified, e.g.: “…among the genes differentially expressed in response to…”

Page 14: “Collectively, we observe that numerous neuronal microexons appear to function by ensuring robust and accurate timing of distinct stages of neurogenesis.” – Whether the word “numerous” is appropriate here is debatable. I would suggest to give specific numbers: How many microexons out of the 36 tested? How many microexons out of all known microexons?

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In this revision, the authors have clarified some challenging concepts. For instance, adoption of differentiation rank and eCDFs is more reader-friendly and illustrates the nature of gene expression changes upon microexon knockouts, i.e., precocious expression of neural genes. They have also addressed some points about the deletion of constitutive exons from the same set of host genes. Below are additional comments on the revised manuscript the authors might consider:

1. On microexon deletion efficiency in Fig. 1: We do not fully agree with the authors that the emphasis of Fig. 1 is to optimize the CHyMErA system by comparing editing efficiencies between Cas12a constructs. The representative efficiency of the system is a reliability indicator for all downstream efforts. There is still some concern regarding whether the deletion efficiencies listed in Fig. 1c-d reflect clonal ground truth. In Revision Fig. 1b, % edited clones (one assumes) provides a binary measure of deletions (heterozygous and homozygous deletions), and does not provide a clear indication of how frequent homozygous deletions may be, or account for the wider set of possibilities in bulk populations (het. versus hom. deletions). For example, 5/5 clonal lines with heterozygous deletions may be described as 100 % edited clones, but in the bulk population the same clones will have 50% deletions. Moreover, as the length normalized band intensities only corrects for differences in intercalations, it does not correct for the more concerning amplification bias of shorter deletion amplicons. It seems disingenuous to ignore shorter amplicon bias. Perhaps linkage is lost in ddPCRs, but it could still provide a more grounded estimate of bulk editing. Beyond further efforts to resolve true editing efficiency, the authors may consider listing this as a limitation.
2. The notion that heterozygote microexon deletions could have phenotypes is plausible, but in this case the clonal het. lines should be used to demonstrate transcriptomic effects for validation targets.
3. The reliability of QDA still remains to be determined. The authors point to validation of example target exons in this manuscript, but this is unsatisfactory. New analytic pipelines such as this should be benchmarked against published datasets. A good example is single cell CRISPR knock out experiments where it is possible to quantify depletion of target gene.
4. How many KO cells/cluster or subcluster per target for gene expression analyses?
5. The representation of guide pairs in the sequenced nuclei library (and potential sources of bias) still needs further illustration. While we understand some basic ideas about #cells/target a detailed description of what the identities of these cells are will be most informative to the reader. In particular, as this study performs DGE on KO v non-targeting subclusters in pseudo-bulk, it is essential to understand how many cells/target microexon KO within each subcluster. We have noted – with concern – that there is only a moderate correlation between bulk and snRNA comparisons (including Figs. 2g v 3c Gfra1 KO and Ext Fig. 5c) and questioned the DGE outcome of the screen. Understanding whether low Ns could be a contributing weakness is quite important.
6. The microexon KO cells are broadly located in subclusters 0,2,4 and 11, all of which represent neuroepithelial (NE) identities (Fig. 2d). There is a strong, general trend for microexon KO cells to be found in neuroepithelial state, with most non-targeted cells in neuron subclusters (orange points in Fig. 2c). The authors have clarified that the microexons are included in NE stage, but it is not clear if KO cell enrichment in NE has true biological meaning, i.e., differentiation defects, or if this is some unexpected technical issue. As there is no replicate screen or characterization of differentiation using clonal microexon KO cells, the robustness of these findings remains uncertain.
7. Related to Fig. 2g, it would be useful to understand which subclusters enrich precocious neurogenesis signatures.
8. It is true that “estimating the ‘true’ proportion of edited/effected cells in bulk cell populations represents a more general challenge encountered in CRISPR-based screens”, however there are general approaches to mitigate concerns, for example independent replicate screens which have not been performed.
9. On modest correlation between changes in gene expression between snRNA and bulk RNAseq datasets: technical differences between the modalities should be described using comparisons with data generated in this study. One cannot simply take the authors word that there are big differences modalities. Extended Fig. 5c shows an alarmingly poor correlation between bulk RNA-seq from a clonal line and CHyMErA-seq. GO-term overlaps as shown in Fig. 3b could indicate bona fide biology or easy-to-obtain terms within the context of neuronal differentiation. As this one of few validations done in this study, it highlights some critical challenges that remain with demonstrating robustness of screening with CHyMErA system.
10. No experiments were done to further validate transcriptomic or other phenotypes in clonal KO lines.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
The authors have addressed all the points I had raised in the original review and I have no further comments.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I appreciate the authors' efforts to clarify my previous comments. I have no more additional points.

(Remarks on code availability)
N/A

Response to Reviewer Comments:

Reviewer #1

In this manuscript, Dupas and colleagues present CHyMErA-seq by extending the CHyMErA screening method developed earlier by the group. CHyMErA-seq allows one to delete neural microexons in a high-throughput manner using CRISPR/paired gRNAs and evaluate the functional consequence of the genetic perturbation in individual cells by split-pool single nuclei RNA-seq. This strategy is used to test a set of 37 conserved and neuron-specific microexons in mESCs during in vitro neurogenesis. The study finds neuronal transcriptomic shifts caused by microexon deletion and further examines the functional roles of two microexons (Gfra1 and Clasp1) in neurodevelopment. Deletion of microexons has also been implicated in altering expression of ASD-risk gene networks, providing insights into how dysregulated microexon splicing may contribute to disease. The functional evaluation of dynamic alternative splicing patterns, especially high-throughput technologies, is highly valuable. Indeed, within this class, the biological requirements of microexons are particularly important due to their potential connection with neurodevelopmental disorders.

While the authors have made great strides in the execution of this high throughput strategy, this reviewer has some concerns regarding the approach itself, and the robustness of the downstream gene expression changes, and their relevance to maturation and disease. In particular, the manuscript will benefit from strengthening key aspects, including microexon deletions, as the entire body of work depends on this strategy, characterization of the approach, and phenotyping of neurons with exon deletions.

We thank the reviewer for their appreciation of the importance of the large-scale functional analysis of microexons, interest in our strategy and results, and constructive comments, which we have addressed below.

CHyMErA-seq screening:

1. Microexon deletion: Figure 1, the authors show %del to account for deletions. This is likely an overestimation as smaller PCR amplicon and typically favored. Even if one ignores this bias, the estimated deletion frequency (e.g., top panels in Figure 1d) does not seems to be consistent with the gel images in some cases.

Moreover, the level of deletion in a bulk sample is not an especially useful metric, as what is more relevant is %cells with homozygous deletions. Carefully evaluating (for example with ddPCR), and reporting these numbers will help the audience understand the robustness of this dual Cas approach. The same problem applies to the application of CHyMErA-seq, where it is not clear to what levels homozygous microexon deletions occurs.

While we acknowledge that there is some bias in estimating editing efficiency due to more efficient amplification of microexon-deleted amplicons, and that there are inherent challenges in estimating editing efficiencies in bulk cell populations, we would like to emphasize that the main goal of the experiments in Figure 1 was to optimize the CHyMErA system by assessing *relative* editing efficiencies between different Cas12a constructs.

Nevertheless, to address the reviewer's first concern, percent editing efficiencies indicated in Figure 1D do not appear congruent with apparent editing efficiencies from visualizing the gel bands because a length correction was applied (see amendment to the Methods on page 21), which is summarized below) (Gonatopoulos-Pournatzis, Aregger et al. 2020). This correction accounts for weaker fluorescence signal due to reduced intercalation of ethidium bromide into shorter amplicons. Notably, we also compared differences in estimated percent editing efficiencies measured by PCR assays on bulk and 'ground-truth' clonal cells using both corrected and uncorrected measurements (see Revision Fig. 1a below). We observe that the median absolute error after correction is lower ($median_{corrected}$ = 5.85, $median_{uncorrected}$ = 9.22), thus validating the use of this approach.

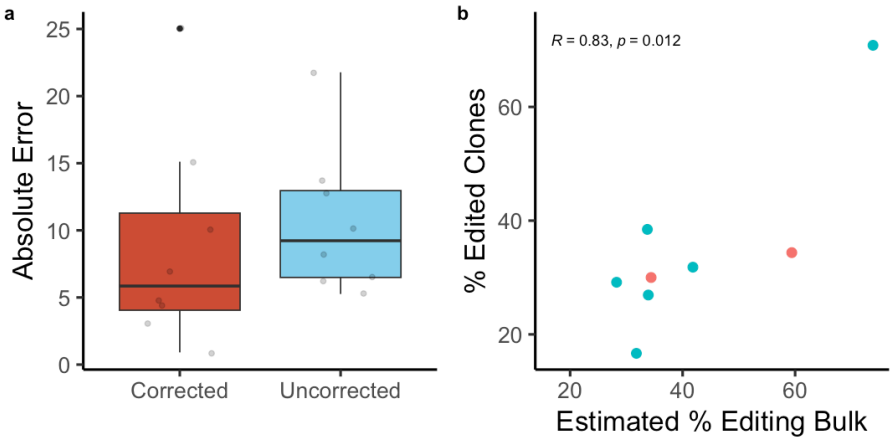
Percent editing was calculated as the quotient of the length-normalized band intensities:

$$\%\ \text{Deletion} = 100 * \frac{\left(\frac{DeletionBand_{intensity}}{DeletionBand_{length}}\right)}{\left(\frac{UncutBand_{intensity}}{UncutBand_{length}} + \frac{DeletionBand_{intensity}}{DeletionBand_{length}}\right)}$$

We have also added text to the figure legend to highlight that this correction has been applied:

Figure 1c and Extended Data Figure 1 a and 1b captions:
"Editing efficiency calculated from length-normalized quantification of ratios of edited and total band intensities (Methods)"

To address the reviewer's second concern, we have assessed our method for estimating editing efficiency from bulk genomic PCR by comparing measurements from 'ground-truth' editing efficiencies from genotyped single-cells derived from the bulk cell populations. Importantly, from these comparisons, we observe a significant correlation between bulk and ground truth editing efficiencies (e.g. see Revision Fig. 1b below; R = 0.83, p = 0.012, Pearson correlation), thereby supporting the validity of our approach.



Revision Figure 1. Assessing the accuracy of bulk cell PCR estimations of genomic editing efficiency. **a)** Comparison of the absolute difference in editing efficiency measured using PCR assays performed on bulk compared to clonal cells, with (red) and without (blue) quantifications correcting for length of PCR products. **b)** Scatterplot comparing percent editing efficiency estimated from genomic PCR assays performed on bulk cell populations, with editing efficiency estimated from the proportions of edited cells observed from the analysis of the single cell clones derived from the CHyMErA deletions tested in this manuscript (red), and deletions used to generate additional clonal lines (blue). R = 0.83, p = 0.012, Pearson correlation. N = 8 guides targeting 7 different exons, median 12 clones per guide.

We thank the reviewer for recommending the use of ddPCR. While this technique could potentially provide more precise measurements of edited and unedited genomes, a standard ddPCR assay would not be informative for distinguishing between homozygous and heterozygous deletions. This is because DNA is randomly partitioned in ddPCR droplets and therefore lacks allelic linkage within single cells. As such, standard ddPCR cannot distinguish between, for example, a population of heterozygous cells and a 50:50 mix of WT and homozygous-edited cells.

To address the reviewer's comment, as mentioned above, we genotyped clonal cell lines derived from populations of cells edited with two independent guides with our enhanced CHyMErA system and observed that 37.5 and 50% of edited clones from these two deletions harbour homozygous deletions. However, it should be noted that based on additional analyses in our submitted manuscript and from the published literature (see below), we would expect that the majority if not all heterozygous deletions of microexons that exhibit robust phenotypic changes when homozygously deleted to also result in measurable phenotypic changes. This conclusion is supported in the present study by the use of our dCasRx-RBM25 targeting system for microexon modulation as an orthogonal assay to measure phenotypes. Specifically, partial skipping of microexons (eg. Clasp1(A), Camta1) directed by dCasRx-RBM25 yielded significant transcriptomic changes that overlap those detected in the CHyMErA-seq screen.

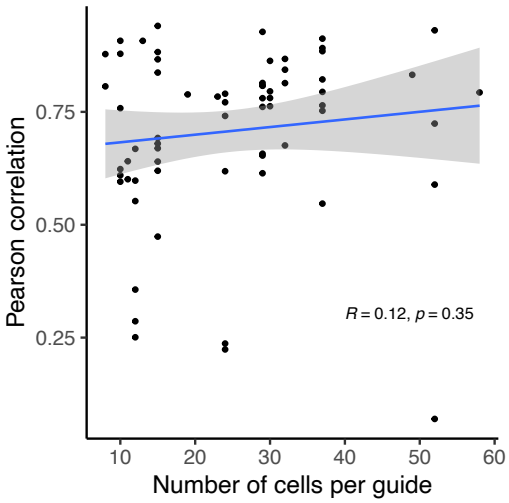
Several published studies have also established that partial disruption of microexons result in significant molecular and behavioural phenotypes. Particularly notable among these are results from analyzing the effects of partial loss of expression of the neuronal microexon-containing isoform of Cpeb4, which results in dysregulation of ASD genes and electrophysiological and behavioural phenotypes in mice (Parras, Anta et al. 2018, Garcia-Cabau, Bartomeu et al. 2024), and heterozygous germline deletion of a neuronal microexon in the Bak1 gene results in increased neuronal apoptosis during development (Lin, Zhang et al. 2020). Furthermore, in a previous study we demonstrated robust yet reduced transcriptomic phenotypes when comparing heterozygous versus homozygous deletions of a neural-differential exon in Puf60 in the same differentiation model as used in the present study (Han, Best et al. 2022).

2. The representation of guide pairs in the sequenced nuclei library (and potential sources of bias) needs further illustration. The authors have targeted 37 exons with 9 guide pairs each. How many cells were recovered per guide pair for each target that passed QC should be included. How is the correlation of DGEs between guide pairs targeting the same microexon affected by cell#?

We refer the reviewer to Extended Data Figures 2a and 2b and the associated text (on page 6) in which we describe the number of cells recovered for individual guides and target: "*Each pair of microexon-targeting gRNAs was represented by a median of 44 cells, and each microexon target was represented by a median of 492 cells (Extended Data Figs. 2a, and 2b)*".

To address the impact of cell number on guide correlation, we plotted this relationship and calculated the correlation (see Revision Fig. 2 below; R = 0.12, p = 0.35, Pearson correlation). This analysis reveals lack of a significant correlation between the numbers of cells expressing different guides targeting the same microexon and changes in transcriptomic signatures, thus indicating that cell number is not a driving factor in the correlations we are observing. To highlight this point, we have added the figure shown below to Extended Data Figure 4e and the following text (on page 7):

"Importantly, cells with guides targeting the same gene exhibited significantly higher gene expression correlation compared to both non-targeting ($p = 7.50 \times 10^{-12}$, Benjamini-Hochberg corrected Wilcoxon rank sum test) and intergenic-targeting controls ($p = 5.85 \times 10^{-9}$, Benjamini-Hochberg corrected Wilcoxon rank sum test) (Extended Data Fig. 4d). *This degree of correlation is to a large extent independent of the number of cells recovered for each guide (Extended Data Fig. 4e)."*



Revision Figure 2. Scatterplot comparing Pearson correlations for guides targeting the same microexon versus the number of cells represented by each guide. Pearson's R and P value are shown.

3. A major caveat of the CHyMErA-seq approach, when applied to mESCs, is the low editing efficiency, and cells with microexon deletion cannot be isolated directly. The use of QDA to classify cells as KO vs non-perturbed based on gene expression profile is a creative way to deal with the KO vs non-perturbed issue. However, this is an indirect workaround, and the reliability of the approach remains to be determined.

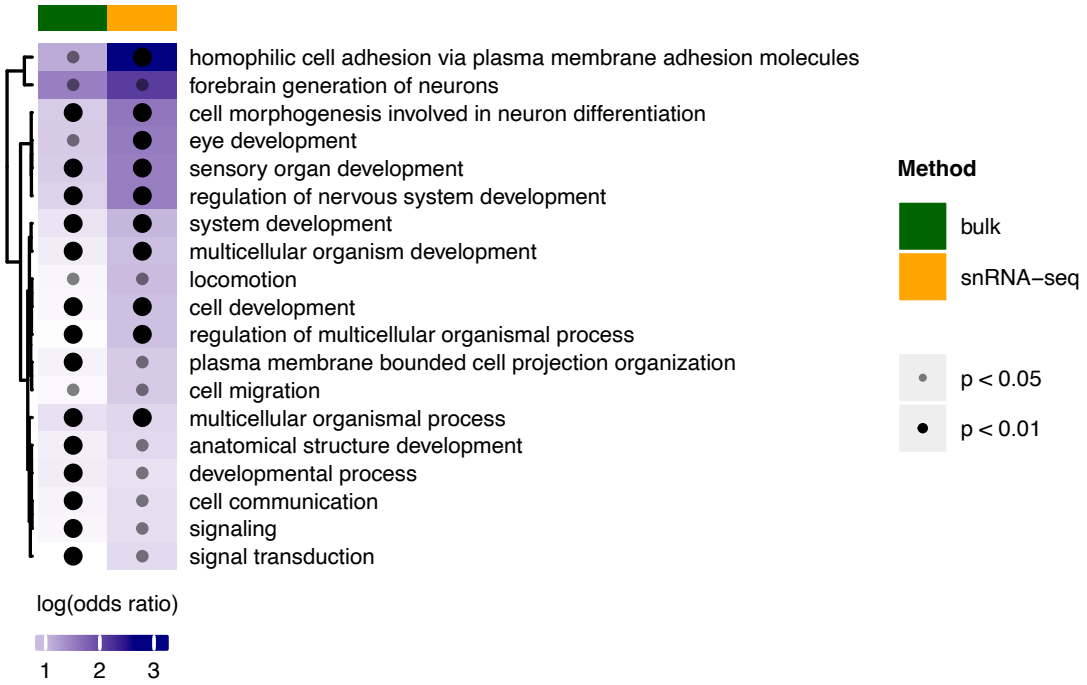
While we acknowledge that a limitation of the CHyMErA-seq strategy is not being able to directly confirm microexon deletion in cells harbouring transcriptomic signatures, the validity of this screening approach is supported by the extensive follow-up experiments provided in our study, including orthogonal assays performed on cell lines harbouring independent, genotyped deletions (Figure 3), and results from independently perturbing microexons using dCasRx-RBM25 (Figure 4). We also note that estimating the 'true' proportion of edited/effected cells in bulk cell populations represents a more general challenge encountered in CRISPR-based screens yet, as in the case of our work, has not hampered providing phenotypic read-outs that serve as a valuable basis of follow-up studies (e.g. (Dixit, Pamas et al. 2016, Papalexi, Mimitou et al. 2021, Santinha, Klingler et al. 2023, Jiang, Dalgarno et al. 2025) to list a few).

The authors compared CHyMErA-seq derived gene expression changes in response to microexon deletion and that determined by bulk RNA-seq of clonal samples, and the correlation is quite moderate (Figure 4d,e and Extended Data Figure 5d), which is very concerning.

We understand the reviewer's concern but given that we are comparing transcriptomic signatures derived from single-nuclei RNA-seq and whole-cell bulk RNA-seq data, we would not necessarily expect to observe strong gene-gene correlations. This is especially the case when considering the considerable heterogeneity of the bulk cell populations and

also the major technical differences involved, for example, in library-preparation and data processing methods. However, despite these significant differences, we observe positive and significant correlations between the bulk and single-nuclear RNA-seq profiles (Fig. 4d and 4e, Extended Data Fig. 5c). Additionally, we observe the same directional-shifts in the expression of genes with a higher neuron differentiation rank in the bulk RNA-seq data (Fig. 3c, Fig, 4b and 4c) as we do in the single nuclear signatures (Fig 2g). These findings therefore support the overall consistency of the detection of altered gene expression programs detected in both the snRNA-seq and the bulk RNA-seq data. This consistency is further reflected in a new analysis of enrichment of Gene Ontology terms associated with upregulated genes by directly comparing the single-nuclear and bulk RNA sequencing data, which is now included as a new panel in Figure 3 (Figure 3b, shown also below). For example, we observe enrichment of multiple common GO terms associated with neuron differentiation and neuronal morphogenesis for genes up-regulated in both snRNA-seq and bulk RNA-seq after Gfra1 microexon deletion.

Page 10:
"Furthermore, we observe a significant overlap between the sets of Gene Ontology (GO) terms enriched for upregulated genes in response to Gfra1 microexon deletion in both bulk RNA-seq and CHyMErA-seq (odds ratio = 20.4, P = 2.2x10^{-16}, Fisher's exact test) (Fig. 3b). We observe that most of the overlapping GO terms are related to processes involved in cellular morphogenesis and neuron differentiation."



Revision Figure 3. Heatmap of log-transformed odds ratios of Gene Ontology (GO) biological process terms significantly enriched (P < 0.05, Fisher's exact test) among up-regulated genes identified by bulk RNA-seq and CHyMErA-seq following Gfra1 microexon deletion. P values were corrected by Benjamini-Hochberg correction.

4. It seems that no replicate screening was performed, which adds additional challenge to evaluate the robustness of the screening.

We performed the CHyMErA-seq screen using multiple distinct hgRNAs targeting each microexon to serve as replicates for edits and to enable assessment of the robustness our transcriptomic measurements. This approach for assessing replication of measurements has the major advantage of avoiding potentially confounding variability due to possible batch effects that one would expect between replicates of entire screen experiments that involve a large number of procedures and months of work-up. We also emphasize again that our study incorporates extensive validation through the application of dCasRx-RBM25 microexon perturbations and other follow-up assays, which served to both confirm and extend results from the CHyMErA-seq data.

5. In some parts, the description of bioinformatics analysis is not very clear. For example, it is difficult to figure out how perturbation calling was done precisely. The authors might consider using less narratives and include some mathematical descriptions.

We thank the reviewer for their comment and have updated the Methods to provide a more detailed explanation of perturbation calling. In addition to including equations describing editing efficiency calculations, we have provided the following new text and associated mathematical equations that describe the perturbation calling strategy:

Page 27:

"We calculated a normalized gene *expression vector for each cell* $(\tilde{x}_i)$ *subtracting the average expression of the neighboring cells* $(\bar{x}_{N_i})$ *for each gene:*

$$\tilde{x}_i = x_i - \bar{x}_{N_i}$$

where:

$$\bar{x}_{N_i} = \frac{\sum_{j \in N_i} x_j}{k}$$

Page 28:

"To classify cells as perturbed or non-targeting we applied quadratic discriminant analysis (QDA), implemented in sci-kit learn (v1.2.2), to calculate classification assignments from multivariate Gaussians fit on normalized expression values $\tilde{x}_i$. QDA models each class as a multivariate Gaussian with its own class-specific covariance matrix Σ_k and mean μ_k and assigns classifications using Bayes' rule:

$$\mathrm{P}(\, y_i = k \mid \tilde{x}_i \,) = \frac{P(\, \tilde{x}_i \mid y_i = k \,) \cdot P(y_i = k)}{P(\tilde{x}_i)}$$

for which we calculate the class conditional likelihood as:

$$\mathrm{P}(\, \tilde{x}_i \mid y_i = k \,) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(\tilde{x} - \mu_k)^{\top}\Sigma_k^{-1}(\tilde{x} - \mu_k)\right)$$

"For each cell, an empirical p-value was calculated by comparing its observed classification accuracy to the null distribution.

$$p_{value_i} = \frac{1 + \sum_{n=1}^{N} I\left[\hat{p}_i \leq p_i^{\mathrm{null},n}\right]}{N + 1}$$

where $\hat{p}_i$ is the observed classification accuracy of cell i, and $p_i{}^{null,n}$ is the accuracy in the nth permutation."

Phenotyping:

6. Are the 37 microexons frame preserving? This reviewer assumes so, but it will be helpful to clearly indicate this.

We have updated the text on page 5 to clarify:

"All of the targeted microexons are frame-preserving and expected to insert 3-9 amino acids into the coding sequence"

7. The authors have positive control gRNA pairs that target constitutive exons, but do not provide much detail on the transcriptomic characterization of these cells.

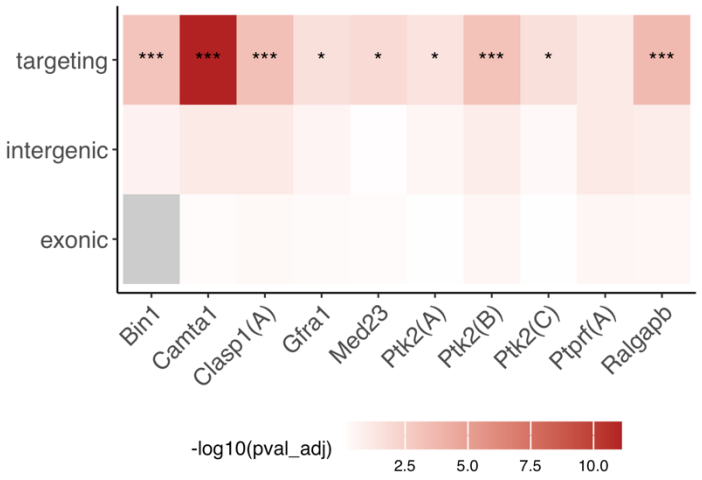
Inclusion of the constitutive exon-targeting guides was intended to serve as positive controls to benchmark transcriptomic responses relative to microexon deletion, intergenic and intronic guide controls. However, assessing transcriptomics signatures from these guides alone was challenging due to the relatively small numbers of cells recovered from these deletions, which was in part due to constraints imposed by a size limitation of our hgRNA library design, as we prioritized the library for targeting microexons over constitutive exons in the same genes. However, in response to the reviewer's comment, we have

performed additional analyses to assess the extent to which transcriptomic signatures detected following microexon deletion are detected in cells expressing guides targeting constitutive exons ('exonic' guides) designed to generate gene knockouts.

For each microexon deletion that resulted in a detectable transcriptomic phenotype, we identified differentially expressed genes (FDR < 0.1) and performed PCA on their normalized expression across single cells (see '*Normalization to non-targeting cells*' in the Methods). We then compared transcriptomic signatures of microexon-targeting, exonic or intergenic negative control guides, with those from non-targeting cells. We calculated Hotelling's T^2 statistic on the top 40 principal components, or fewer if limited by the number of cells in each group, to assess statistical significance.

This analysis reveals that, among all of the targeted genes, no constitutive exon-targeting guides elicited significantly altered the same transcriptomic signatures that were detected in cells targeted for microexon-deletion (i.e. relative to non-targeted cells) (Revision Figure 4, see below). A similar result was observed when comparing with intergenic and non-targeting negative control cells. These results suggest that the transcriptomic signatures induced by microexon deletion are distinct from those generated by constitutive exon-targeting guides, consistent with splicing dysregulation-specific phenotypes. We have added the following text to the manuscript and the accompanying Extended Data Fig. 4g to address the reviewer's comment.

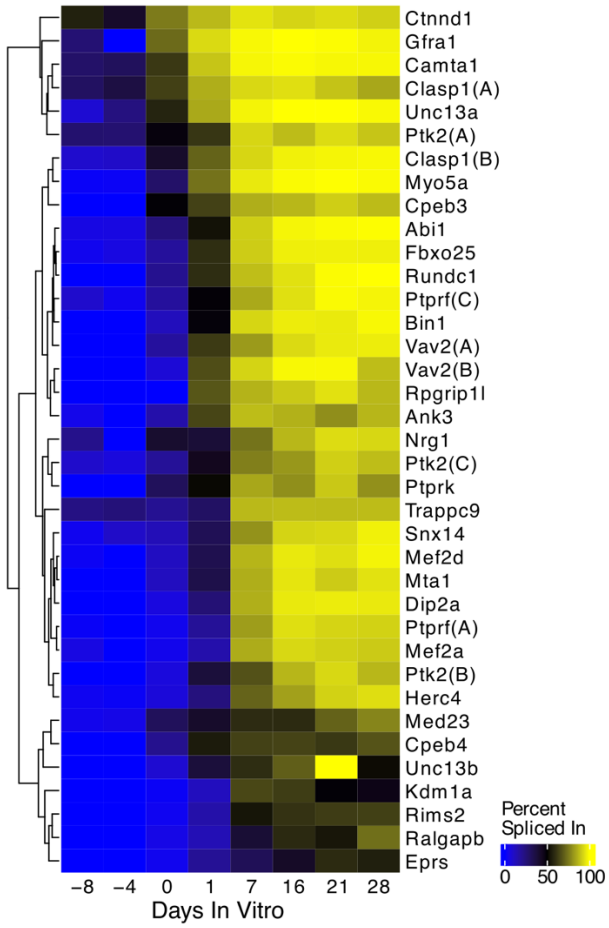
Page 7:
"*To assess the specificity of microexon-deletion phenotypes, we performed pairwise comparisons of microexon deletion-associated transcriptomic signatures between cells expressing microexon-targeting hgRNAs or cells expressing guides targeting constitutive exons in the same gene, with cells expressing non-targeting guides (Methods) (Extended Data Fig. 4g). This analysis reveals that transcriptomic signatures derived from microexon-deletions are distinct from those detected in response to guide-directed perturbation of the corresponding genes."*



Revision Figure 4. Transcriptomic similarity between cells harboring microexon deletions and cells expressing control guides or guides targeting constitutive exons in the same genes. Heatmap showing -log10(corrected p-values) for pairwise comparisons of transcriptomes between non-targeting cells and microexon-deletion cells (targeting), constitutive exon-targeting (exonic) or intergenic negative control guides (grey indicates NA). Cells were embedded using principal component analysis of differentially expressed genes and Hotelling's T^2 test was used to assess group-level differences in multivariate expression space. $P^* > 0.05$, $P^{**} > 0.01$, $P^{***} > 0.001$; Benjamini-Hochberg corrected Hotelling's T^2 test.

8. Pseudotime analyses indicates that the KO cells are broadly enriched in the neuroepithelia clusters (Fig 2d). This suggests there are differentiation defects and KO cells are stuck in the neuroepithelium state? As the microexons are typically not included to significant levels at this stage of differentiation, how do the authors explain KO cells enrichment in neuroepithelia. Moreover, this seems incongruent with the idea that microexons deletions increase maturation.

In fact, the analyzed microexons are strongly included at this stage (e.g. refer to Day In Vitro 7 timepoint in Figure 1f, shown below). The pseudotime analysis merely indicates that within the trajectory these cells are less mature. Contrary to the reviewer's interpretation that these cells may be arrested at this stage of differentiation, these cells have an upregulated differentiation signature at this early stage and do differentiate. This is illustrated in the validation experiments focusing on the Gfra1 microexon deletion, in which we observe that the microexon deletion does not cause differentiation arrest, but rather an upregulation of differentiation-related transcriptional programs and increased neuritogenesis (refer to Figure 3).



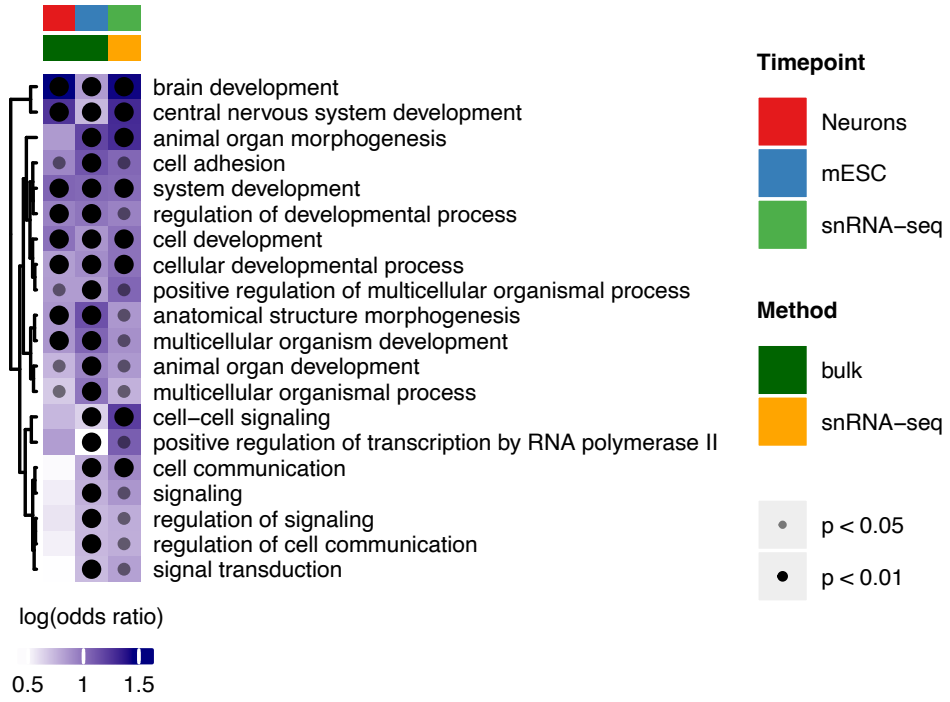
Revision Figure 5. Percent spliced in (PSI) profiles of microexon targets selected for the CHyMErA-seq analysis during *in vitro* neuronal differentiation. From Figure 1f.

9. More can be done to solidify/validate gene expression changes observed in the snRNA data. While the authors decided to focus on Gfra1 and Clasp1 for additional analyses and made efforts to generate clonal cell lines with microexon deletion (or skipping/inclusion using dCasRx-RBM25). The evidence provided to support the functional role of each exon is less than compelling in the current form. For example, for Gfra1, the transcriptomic changes inferred from the screen essentially has no correlation with that from analysis of clonal cell lines (Extended Data Figure 5d). For Clasp1, the impact of microexon on downstream phosphorylation appears to be quite moderate.

To validate CHyMErA-seq-detected gene expression changes, we performed bulk RNA-seq analysis on clonal cell lines harbouring microexon deletions that were generated using gRNAs that are different from those used for the CHyMErA-seq screen, and we also manipulated microexons using dCasRx-RBM25 targeting. As described in response to the comment # 3 above, gene-gene level correlations are not expected to be high between snRNA-seq and bulk RNA-seq data, in large part because of the significant technical differences involved in sample preparation and analysis. Nevertheless, we still observe

significant and positive correlations (see Extended Data Figure 5c, Figure 4d and 4e). Furthermore, consistently, we also observe significant overlaps between GO functional terms that are enriched in the sets of genes with neuronal differentiation-related transcriptional signatures between these datasets (refer to Revision Figure 3). By performing the same analysis on gene expression signatures resulting from Clasp1 dCasRx-RBM25 manipulation, we also observed significant overlap with GO term enrichments derived from CHyMErA-seq and the bulk RNA-seq (odds ratio$_{Neurons}$ = 52.1, $P_{Neurons} < 2.2x10^{-16}$; odds ratio$_{mESC}$ = 29.3, $P_{mESCs} < 2.2x10^{-16}$, Fisher's exact test) (see Revision Fig.6 below).

Page 11:
"*We further observe significant overlap of GO terms enriched among genes that are upregulated in the CHyMErA-seq gene signature, and in mESCs (odds ratio = 29.3, P < 2.2x10^{-16}, Fisher's exact test) and DIV7 neurons (odds ratio = 52.1, P < 2.2x10^{-16}, Fisher's exact test)) (Fig. 4f). These GO terms are associated with functions related to the regulation of nervous system development, morphogenesis and cell-signaling.*"



Revision Figure 6. Heatmap of the log(odds ratio) of representative significantly enriched GO terms in bulk RNA-seq and CHyMErA-seq following Clasp1(A) microexon exclusion compared to inclusion in mESCs and neurons (Methods). P values are FDR adjusted.

10. The authors used a KS-test to evaluate alterations of gene expression correlation with neuron differentiation rank. It is quite difficult to understand what these significant differences really mean.

The neuron differentiation rank for each gene corresponds to how strongly the expression of a gene correlates with the position of a given cell on the differentiation trajectory. These highly ranked genes are enriched in neuronal processes, while lowly ranked genes are enriched in processes involved in cell division and catabolism (see Extended Data Fig. 4i).

We agree with the reviewer that the use of the KS-test here is somewhat indirect and have adapted the analysis with a more straightforward comparison. We now assess the significance in the distribution of neuron differentiation ranks for up- and down-regulated genes with a two-sided Wilcoxon rank sum test comparing the up- or down-regulated distribution of ranks with the distribution derived from genes that are not significantly altered (i.e. FDR ≥ 0.05). Therefore, we can interpret the enrichment of genes based on their regulation towards higher or lower neuron differentiation rank scores. We have added the following statement to the text to clarify this analysis and have accordingly updated Figures 2g, 3c, 4b and 4c to reflect the new statistics:

Page 9:
"To assess how microexon deletion effects neurogenesis, we compared the distributions of scores for up- and down-regulated genes for each perturbation, *or in the case of Ptk2, pooled scores for genes affected across the three microexon perturbations, due to the relatively low cell numbers and sparse data recovered for each individual Ptk2 perturbation. Notably, compared to background genes (i.e. non-differentially expressed, FDR $\geq$0.05), we observe that upregulated genes associated with altered transcriptomic signatures are significantly enriched for higher CellRank2 scores* for microexon deletions in the Bin1, Clasp1, Gfra1, Med23, Ptprf and Ralgapb genes ($P < 0.05$, Benjamini-Hochberg corrected two-sided Wilcoxon rank sum test) (Fig. 2g)."

11. Ptk2 microexons are known to dramatically elevate autophosphorylation. Do the authors detect signatures of modulated Ras signalling/other downstream signaling activity? Similarly, do the authors observe elevated morphology or migration changes downstream of Ptk2 microexon modulation?

We are aware of the excellent work performed on the Ptk2 microexons and have referenced these papers (Burgaya, Toutant et al. 1997, Toutant, Costa et al. 2002). We refer the reviewer to page 13 where we highlight the link between microexon deletions, including Ptk2 microexons, and the Ras/MAPK-signaling pathways (which is downstream of FAK-integrin signaling):

Page 13:
"Of these six dysregulated regulons associated with microexon deletion, three are downstream of MAPK-ERK signaling pathway (Ets2, Erg1 and Elk1), which has been implicated in core processes during neurogenesis (Gantner, Hunt et al. 2021). Notably, in addition to the role of the Clasp1(A) microexon in augmenting MAPK-ERK signaling shown in this study, roles in MAPK signaling have been previously described for Bin1 (Wang, Jia et al. 2017), Gfra1 (Paratcha, Ledda et al. 2003), Med23 (Wang, Huang et al. 2009, Yang, Zhao et

al. 2012, Yang, Li et al. 2023), Ptprf (Wang, Weng et al. 2000, Wang, Kuo et al. 2007) and Ptk2 (Mitra, Hanson et al. 2005) proteins (which also form a protein-protein interaction network, Fig. 2f) suggesting that the microexons in these proteins may convergently regulate this signalling pathway."

We observe significant downregulation of the regulon score associated with Egr1, which is activated downstream of MAPK signaling pathways in response to Ptk2 microexon deletions (Figure 5a). Since Ptk2 microexons have already been characterized in some detail, we did not follow up on these with morphological assays and instead focused on relatively less-well characterized microexons in the Gfra1 and Clasp1 genes.

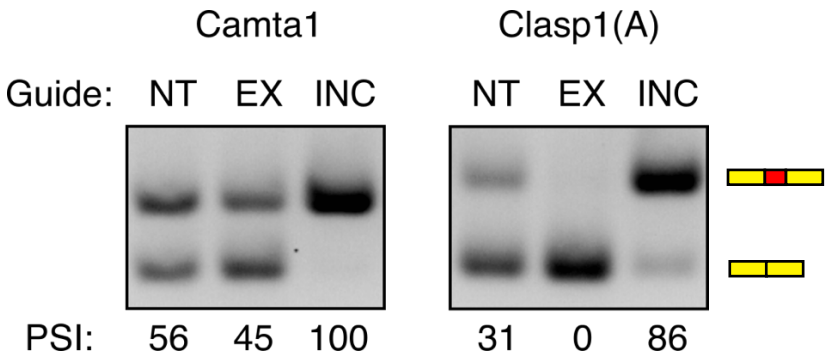
12. GDNF signaling is known to promote survival and growth of neurons. The authors find that Gfra1 microexon deletion modestly increases neurite growth and stimulates expression of mature neuronal genes. Is this a consequence of increased GDNF signalling? i.e., does microexon inclusion decrease Gfra1 signaling? Can the authors supplement neurons with GDNF to amplify the neuritogenesis phenotypes?

We interpret that these changes are due to increased Gfra1 signaling due to previous work showing that inclusion of the microexon in Gfra1 decreases its affinity for Gdnf and Nrtn and thereby downregulates Ret-signaling (Charlet-Berguerand, Le Hir et al. 2004). We inadvertently omitted this citation from the original text and have corrected this error in the revised manuscript. While we agree with the reviewer that it would be interesting to titrate Gdnf and measure effects on neurite outgrowth, we feel that this experiment is beyond the scope of our manuscript, which is primarily focused on introducing the CHyMErA-seq technology, validation of this system, and initial follow-up characterization on selected interesting screen hits.

13. Camta1 microexon does not appear to be included in mESCs. Nevertheless, the changes in gene expression upon deletion in mESCs appears to enrich downregulated genes that are affected in autism (Fig 5d). this feature appears significant only in the mESCs and not in neurons. Similar results are also obtained for Clasp1 microexon as well. How do the authors explain observations of transcriptomic phenotypes in mESCs that do not express much of the microexon, but not in neurons from the same founders?

Actually, the Camta1 and Clasp1 microexons do show basal inclusion in mESCs (approximately 56 PSI and 31 PSI, respectively; see Revision Figure 7 below, and Revision Figure 5 (i.e. Fig 1f)). Furthermore, in these experiments we are comparing dCasRx-RBM25-directed skipping and inclusion in these cells, which affords a larger difference in the inclusion levels of these microexons between the two conditions. Strikingly, we observe that dCasRx-RBM25-directed skipping of the microexon in mESCs is sufficient to induce transcriptomic changes corresponding to a more neuronal signature, as compared to dCasRx-RBM25-directed exon-inclusion mESCs. These results highlight that by forcing splicing level differences of these microexons in mESCs using dCasRx-RBM25, we observe transcriptomic changes similar to those seen in the CHyMErA-seq screen data, which are

enriched in ASD-linked genes that are often implicated in early nervous system development (Satterstrom, Kosmicki et al. 2020) as is indicated in Figure 5d.



Revision Figure 7. RT-PCR quantification of inclusion levels of Camta1 and Clasp1(A) microexons when transduced with non-targeting ('NT') exon skipping ('EX') or inclusion ('INC') guides in dCasRx-RBM25 expressing mESCs.

Other minor points:

14. Explain hgRNA

We have elaborated on the description of hgRNAs in the Introduction:

Page 3:
"CHyMErA employs simultaneous expression of Cas9 and Cas12a nucleases, together with hybrid guide (hg)RNAs, *which are fusions of Cas9 and Cas12a guides expressed from a single U6 promoter and* processed into pairs of individual Cas9 and Cas12a gRNAs that direct genomic fragment deletion."

15. Line 689, missing ")" after CRISPOR.

Corrected.

16. Line 755, paring --> pairing

Corrected.

17. Regulon scores were z-normalized and p values **and** were calculated for each pairwise comparison by two-tailed unpaired t-test followed by Benjamini-Hochberg correction.

Corrected.

18. Fig. 2b, colors for some of the clusters are too light to be visible

We have adjusted the color in the plot to make the points more visible.

19. Fig. 2e, what are the columns?

We have replaced this figure with a revised version that illustrates enrichment of gene ontology terms for the DEGs associated with microexon deletions thereby providing a more biologically interpretable figure.

20. Fig. 4 legend, “Empirical cumulative distribution function (ECDF) of differentially expressed genes (DEGs, FDR < 0.05) associated with microexon inclusion compared to **inclusion** for mESCs.”

Correction made.

Reviewer #2

This manuscript presents CHyMErA-seq, a platform designed to profile transcriptomic changes at single-cell resolution in response to targeted splice isoform perturbations. Applying this approach to 36 neuronal microexons, the authors identify consequences of these alternative splicing events on gene expression dynamics during neurogenesis. Several microexon-containing genes tested here are implicated in neuron projection development and converge on signaling pathways such as FAK-Src. Functional analyses reveal that specific microexon deletions studied here (similar to the effect of microexons studied elsewhere) lead to increased neurite outgrowth and in general a premature activation of neuronal gene expression programs. The authors also observe parallels between microexon-dependent transcriptomic signatures and gene expression patterns observed in autism spectrum disorder, in line with previous studies.

This study builds on a large body of research from the Blencowe and other labs, which provided the basis for the experimental tools applied here. In the present study, the authors improved editing efficiency and coupled the CHyMErA pipeline with scRNA-seq readout. In addition, most if not all the main results obtained using scRNA-seq here confirm or are in line with previous studies, as the authors themselves point out repeatedly (activation of neuronal gene expression signatures, increased neurite outgrowth, link to ASD,...). On the one hand, this is an encouraging proof-of-concept of the experimental approach, but on the other hand raises questions about the novelty of the findings. In my opinion, it will be necessary to delineate more specifically which results are novel, which are confirmatory, and where the single-cell and isoform-level perspective (which is the main “selling point” of this study) allowed new biological insights into a previously studied biological question.

Specific points and suggested analyses are detailed below. Especially considering the necessity of pointing out the novelty of the findings compared to previous studies, it would

be advisable to be more specific in the wording of the results and avoid vague statements such as the ones below:

Page 12: “Overall, these data confirm transcriptomic signatures detected in the CHyMErA-seq screen and further support important new functions of microexons in neurogenesis”

Page 14: “Collectively, these findings highlight the importance of isoform-specific protein functions of genes in the control of critical and dynamic processes during neurogenesis, and their dysregulation in neurodevelopmental disorders.”

Page 15: “Applying this system to neuronal microexons revealed unexpected functional properties [which ones?] of this important class of alternative splicing in the context of neurogenesis, as well as new insights into the multilayered links between microexons and brain disorders [I do not agree that the manuscript contains enough novel insights in this respect to justify this sentence].”

We thank the reviewer for their feedback and recommendation to better highlight the novel and distinct aspects of our study. In our view, one of the most important and novel contributions of our study is that it reports the first scalable system for the systematic perturbation of exons coupled to a single cell transcriptomics read-out. This approach enables for the first time the parallel interrogation of the functions of dozens of exons in a single experiment, which in turn affords direct comparisons of the activities of multiple splice variants in specific biological processes and pathways. In particular, while a small number of previous studies have revealed functions of microexons in the control of transcriptomic outputs and neuritogenesis, these studies focused on analyzing one microexon at a time in bulk cell populations and at single time points. In contrast, by analyzing the functions of multiple microexons at single cell resolution at different stages of development our study reveals convergent functions of microexons in specific signalling pathways and in the timing of neurogenesis-associated transcriptomic signatures. Lastly, while our study purposefully targeted a few microexons that have been previously studied to compare specific biological outcomes (i.e. affording a level of independent validation), most of the microexons interrogated in our study have not been previously analyzed on any level and therefore their functional roles were entirely unknown.

To address the reviewer’s comments relating to sentences on pages 12, 14 and 15, we have modified the text to clarify the most important and novel findings or out study, as well as how our results relate to – but complement and extend – the findings of previous studies:

Page 12:

“Overall, these data confirm the dysregulation of gene programs associated with neuron morphology and development observed in the CHyMErA-seq screen and extend these observations by highlighting disruption of downstream signaling cascades as a result of microexon perturbation.”

Page 14:

"*Taken together with the observations described above indicating functions for these microexons in controlling neuronal development, our results highlight a potential role for their dysregulation in contributing to the etiology of neurodevelopmental disorders.*"

Page 16:

"Applying this system to neuronal microexons uncovered alternative splicing-dependent gene expression signatures associated with the regulation of developmental timing during neurogenesis, and further illuminated an additional connection between microexon dysregulation and neurodevelopmental disorders."

Furthermore, to provide context and contrast our work to previous studies, we have added the following text to the discussion:

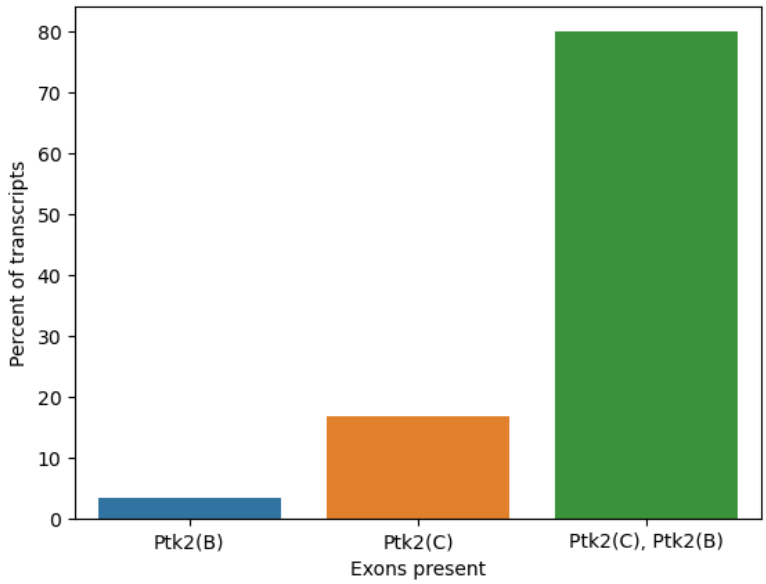
Page 16:
"In contrast to these previous studies, through the development of the CHyMErA-seq system the results of the present study revealed new and unexpected functions of microexons in the control of single cell transcriptomes during neurogenesis."

Specific comments:

Page 8/Fig 2: "The Ptk2 microexons (denoted as 'A', 'B' and 'C') display coordinated splicing patterns in vivo…": Their splicing is coordinated, but here the authors delete one of them at a time: Consequently, deletion of microexon alone will create an artificial protein that is not usually produced (because they are included or excluded together). Is there in vivo evidence for the existence of transcripts with two of the three microexons included and one skipped? In addition, the authors state that "notably, we observe transcriptomic signatures for all three microexons targeted in the Ptk2 gene (Fig. 2e)". First of all, this should be rephrased to be less vague and more accurate: (a) …transcriptomic signatures of the DELETION of all three microexons…; (b) …we observe DISTINCT transcriptomic signatures…; (c) these apparent differences between the effect of the deletion of the three microexons should be discussed (in particular, B and C are located in the same domain, but yet deletion of each of them shows very different effects on the transcriptome).

We thank the reviewer for these important questions. While there are predicted transcripts in Refseq for Ptk2 isoforms matching all possible combinations of alternative splicing exons A, B and C, some of these transcripts appear to represent minor fractions of the total transcript pool for this gene. This is especially true of transcripts containing exons B or C, which show coordinated splicing ('A'-only transcripts are annotated in GENCODE and show an earlier inclusion in neurodevelopment). To directly observe the presence of B and C combinations in neural data, we analysed single-cell isoform sequencing data from a

recent paper that utilized samples from the developing human brain (Patowary, Zhang et al. 2024) (Revision Figure 8 below). This analysis indicates that 20% of transcripts containing 'B' or 'C' harbour one or the other microexon, while 80% harbour both. However, given the frame-preserving lengths of these exons (18 and 21 nucleotides), we expect that the resulting transcripts are still translated into proteins.



Revision Figure 8. Bar plot of Ptk2 exons present in long-read single-cell data from developing human embryos (Patowary, Zhang et al. 2024). Bars represent the percent of transcripts containing Ptk2 microexons 'B' and/or 'C' (n = 60 total transcripts).

Finally, while we individually targeted the Ptk2 A-C microexons, there were relatively few cells harbouring these deletions resulting in sparse data for the individual perturbations. The description of the transcriptomics read-out therefore relates to the average changes in gene expression after pooling data from cells harboring the different perturbations. We have modified the text to clarify this:

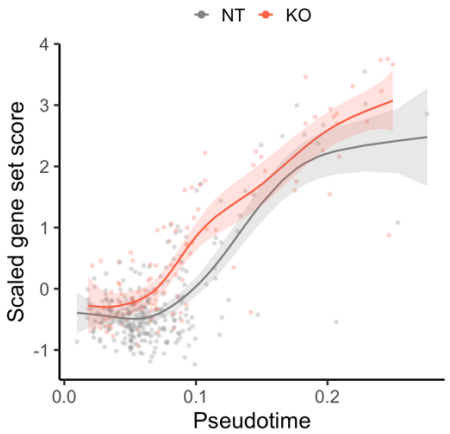
Page 9:
"To assess how microexon deletion effects neurogenesis, we compared the distributions of scores for up- and down-regulated genes for each perturbation *or in the case of Ptk2, pooled scores for genes affected across the three microexon perturbations, due to the relatively low cell numbers and sparse data recovered for each individual Ptk2 perturbation.*"

Fig 3a and b: Figure legend and description in the text is very minimalistic. It is not clear to me if the focus here is on gene expression changes between Gfra1 microexon knockout and non-targeting or along the differentiation/pseudotime. It sounds like Fig3a shows only

differentially expressed genes (between KO and control – are all of them used for the GSEA in 3b?) and how the expression of these genes changes (only in Gfra1 KO? cells) with pseudotime. Given the drastic changes in expression between pseudotime bins, which population was used for the differential expression testing between KO and control? All cells together? Only neurons?

To address the reviewer's question, we have replaced the previous Figure 3a with a simpler version that is also shown below (Revision Figure 9). The figure now shows the scaled aggregate expression of DEGs derived from Gfra1 microexon deletion in the CHyMErA-seq screen over pseudotime for both Gfra1 deletion cells and non-targeting cells. We hope that this panel better illustrates the temporal changes in the expression of upregulated genes as a consequence of the Gfra1 microexon deletion cells through pseudotime, as compared to non-targeting cells. The GSEA was performed using log-fold changes between Gfra1 targeted cells and non-targeting cells. We have moved this figure to Extended Data Figure 5b. The legends for these figures (and others in our manuscript) have been expanded in order to make them easier to follow and the corresponding procedures have been added to the Methods:

"Temporal analysis of differential gene expression
Cell-wise scores computed per cell by summing normalized log-expression values (log1p-transformed) of genes significantly upregulated in Gfra1 microexon deletion (KO) versus non-targeting (NT) cells (logFC > 0, FDR < 0.05). Scores were z-scaled across all cells. A generalized additive model (GAM) was then fitted to the scaled scores as a function of pseudotime using a smooth interaction with genotype. Model fits and 95% confidence intervals were computed using the mgcv package in R(Wood 2011)."



Revision Figure 9. Temporal dynamics of Gfra1 microexon deletion-associated transcriptomic signature. A generalized additive model (GAM) was fitted to scaled expression scores for genes significantly upregulated (logFC > 0, FDR < 0.05) in Gfra1 microexon deletion cells (KO, n = 103; red) and non-targeting controls (NT, n = 278; grey) across pseudotime (adjusted R^2 = 0.72; KO vs. NT estimate = 0.39; p = 6.7×10^{-9}) (Methods). Shaded areas represent the 95% confidence intervals of the fit.

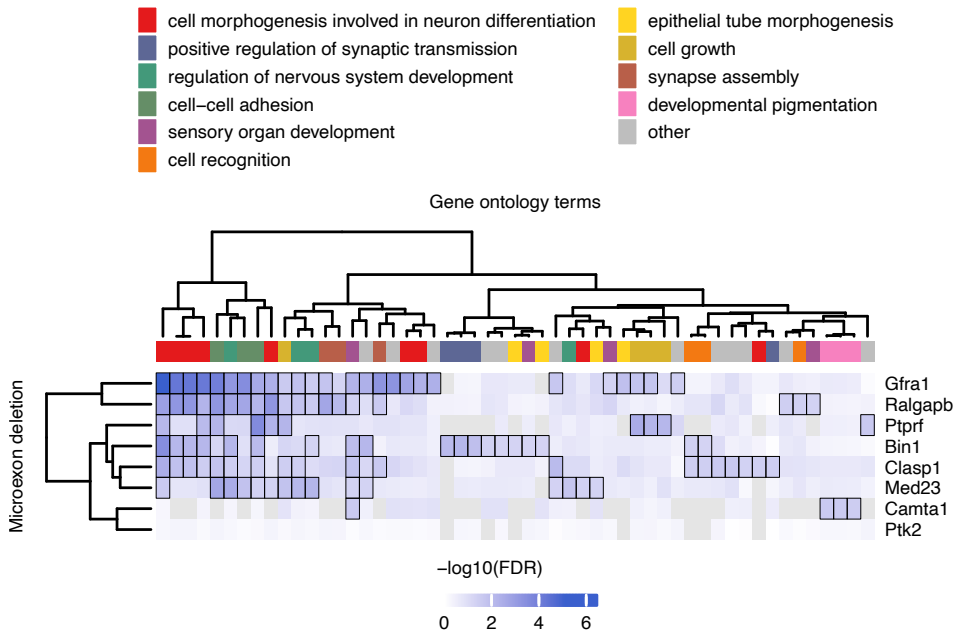
Page 10: “Collectively, these results support a role for the Gfra1 microexon in suppressing a transcriptomic program to prevent premature activation of neuritogenesis.” I am aware that also other microexon deletions have been shown to lead to a more neuron-like phenotype, however, this is still a bit counterintuitive to me. Forcing the non-neuronal splicing outcome (skipping) leads to a more neuron-like gene expression signature and phenotype. I would appreciate if the authors could elaborate on this in the discussion, especially given the fact that this seems to be a general consequence of neuronal nmicroexon deletion.

We have elaborated on these findings in the Discussion by providing the following text:

Page 15:
“Collectively, we observe that numerous neuronal microexons appear to function by ensuring robust and accurate timing of distinct stages of neurogenesis. *In particular, several of the analyzed microexons have convergent effects on transcriptomes that suggest they serve as molecular ‘brakes’ during neurogenesis. This unexpected observation supports the conclusion that specific microexons may have important roles in modulating protein functions to meet context-specific requirements during neurogenesis. For example, the Gfra1 microexon reduces ligand-binding affinity(Charlet-Berguerand, Le Hir et al. 2004), and its deletion in this study leads to the increased expression of genes involved in neuronal differentiation and morphogenesis, along with enhanced neuritogenesis. In contrast, deletion of Ptk2 microexons (which normally enhance phosphorylation activity(Burgaya, Toutant et al. 1997, Toutant, Costa et al. 2002)) leads to downregulation of neurogenesis-related genes (Fig. 2g). This is opposite to the effect seen with deletion of the Clasp1 microexon, which promotes FAK phosphorylation and accelerates neurogenesis.”*

In addition, an analysis of the commonalities and differences in the gene expression signatures of different microexon deletions would be helpful to strengthen the postulated convergence on FAK-Src pathway.

We have updated Figure 2e (Revision Figure 10, below) to highlight the functional terms enriched for each microexon deletion gene expression signature in an effort to enhance the interpretability of these transcriptomic signatures. This analysis shows that several of the microexon deletions result in dysregulation of genes involved in gene sets broadly involved in nervous system development and neuron projection guidance.



Revision Figure 10. Heatmap of top 10 (sorted by FDR adjusted p-value) gene ontology (GO, biological processes) terms for differentially expressed genes (FDR < 0.05) resulting from microexon deletions. GO terms are grouped according to similarity and colored according to parent GO terms. Outline indicates Benjamini-Hochberg corrected P < 0.05. Grey indicates 'NA'.

Page 13: "A 'module score' representing expression of genes implicated in ASD…" – Are these genes implicated because of their premature expression (which would be in line with the transcriptomic signature of microexon deletion) or through loss-of-function/mutations? This distinction is relevant for the interpretation of the observed overlap in the list of genes.

These genes are not implicated in ASD based on expression, but rather their genetic link to ASD. The point we are making here is that ASD-linked genes fall within this program of dysregulated genes upon microexon deletion. This is clarified on page 13:

"A 'module score' representing expression of genes with genetic links to ASD (using ASD association evidence scores 1-3 from Simons Foundation Autism Research Initiative (SFARI) (Abrahams, Arking et al. 2013))".

Page 13: "We observe that over pseudotime these premature maturation signatures show an early increase in expression of ASD-linked genes…" - Specifically ASD-linked genes? Or general neuronal and the observed association with ASD-linked genes is due to the fact that ASD-linked genes tend to have neuronal functions? What was the background gene list used in the enrichment analyses here?

ASD-linked genes represent a subset of the dysregulated genes with neuronal functions. We have added the following text to clarify this point:

Page 14:
"We observe that over pseudotime these premature maturation signatures include an early increase in expression of ASD-linked genes (pseudotime bins 3-6. $p < 0.05$, Benjamini-Hochberg corrected two-sided Wilcoxon rank sum test) followed by a lagging increase in non-targeting cells (pseudotime bin 7, $p < 0.05$, Benjamini-Hochberg corrected Wilcoxon rank sum test), before reaching a plateau (Fig. 5c)."

The background set in the enrichment analysis was all expressed genes used in the differential expression analysis.

Minor comments:

Page 3: "These challenges are particularly relevant to a program of evolutionarily conserved, 3-27 nucleotide-long 'microexons', which are predominantly spliced in the nervous system 14, 15." – Should be "predominantly spliced in" or "predominantly included".

We have corrected this to:

Page 3:
"These challenges are particularly relevant to a program of evolutionarily conserved, 3-27 nucleotide-long 'microexons', which are predominantly *included* in the nervous system".

Page 13: "We also observe enrichment of ASD risk genes identified in three studies in the differential gene expression signatures detected in response to perturbation of Clasp1(A), Ptprf(A), Ptk2(A), Camta1 and Gfra1 microexons in our bulk RNA-seq" – should be clarified, e.g.: "…among the genes differentially expressed in response to…"

We have made this correction

Page 14: "Collectively, we observe that numerous neuronal microexons appear to function by ensuring robust and accurate timing of distinct stages of neurogenesis." – Whether the word "numerous" is appropriate here is debatable. I would suggest to give specific numbers: How many microexons out of the 36 tested? How many microexons out of all known microexons?

We have changed 'numerous' to 'multiple'. However, since 10 out of 36 of the microexon deletions tested that showed robust transcriptomic changes in our analysis displayed altered timing of neurogenesis-associated gene expression patterns, we consider it likely that this functional property of microexons may be more broadly relevant. We have modified the text on page 15 to reflect this distinct observation relating to the function of microexons in our study.

Page 15:

"Collectively, we observe that *multiple* neuronal microexons appear to function by ensuring robust and accurate timing of distinct stages of neurogenesis…"

We thank the reviewers again for their thorough and constructive comments, which have helped us assemble an improved manuscript.

References

Abrahams, B. S., D. E. Arking, D. B. Campbell, H. C. Mefford, E. M. Morrow, L. A. Weiss, I. Menashe, T. Wadkins, S. Banerjee-Basu and A. Packer (2013). "SFARI Gene 2.0: A community-driven knowledgebase for the autism spectrum disorders (ASDs)." Molecular Autism **4**(1): 1-3.

Burgaya, F., M. Toutant, J. M. Studler, A. Costa, M. Le Bert, M. Gelman and J. A. Girault (1997). "Alternatively spliced focal adhesion kinase in rat brain with increased autophosphorylation activity." J Biol Chem **272**(45): 28720-28725.

Charlet-Berguerand, N., H. Le Hir, M. Incoronato, U. di Porzio, Y. Yu, S. Jing, V. de Franciscis and C. Thermes (2004). "Expression of GFRalpha1 receptor splicing variants with different biochemical properties is modulated during kidney development." Cell Signal **16**(12): 1425-1434.

Dixit, A., O. Pamas, B. Y. Li, J. Chen, C. P. Fulco, L. Jerby-Amon, N. D. Marjanovic, D. Dionne, T. Burks, R. Raychowdhury, B. Adamson, T. M. Norman, E. S. Lander, J. S. Weissman, N. Friedman and A. Regev (2016). "Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens." Cell **167**(7): 1853-+.

Gantner, C. W., C. P. J. Hunt, J. C. Niclis, V. Penna, S. J. McDougall, L. H. Thompson and C. L. Parish (2021). "FGF-MAPK signaling regulates human deep-layer corticogenesis." Stem Cell Reports **16**(5): 1262-1275.

Garcia-Cabau, C., A. Bartomeu, G. Tesei, K. C. Cheung, J. Pose-Utrilla, S. Picó, A. Balaceanu, B. Duran-Arqué, M. Fernández-Alfara, J. Martín, C. De Pace, L. Ruiz-Pérez, J. García, G. Battaglia, J. J. Lucas, R. Hervás, K. Lindorff-Larsen, R. Méndez and X. Salvatella (2024). "Mis-splicing of a neuronal microexon promotes CPEB4 aggregation in ASD." Nature 2024: 1-8.

Gonatopoulos-Pournatzis, T., M. Aregger, K. R. Brown, S. Farhangmehr, U. Braunschweig, H. N. Ward, K. C. H. Ha, A. Weiss, M. Billmann, T. Durbic, C. L. Myers, B. J. Blencowe and J. Moffat (2020). "Genetic interaction mapping and exon-resolution functional genomics with a hybrid Cas9–Cas12a platform." Nature Biotechnology **38**(5): 638-648.

Han, H., A. J. Best, U. Braunschweig, N. Mikolajewicz, J. D. Li, J. Roth, F. Chowdhury, F. Mantica, S. Nabeel-Shah, G. Parada, K. R. Brown, D. O'Hanlon, J. Wei, Y. Yao, A. A. Zid, L. C. Comsa, M. Jen, J. Wang, A. Datti, T. Gonatopoulos-Pournatzis, R. J. Weatheritt, J. F. Greenblatt, J. L. Wrana, M. Irimia, A. C. Gingras, J. Moffat and B. J. Blencowe (2022). "Systematic exploration of dynamic splicing networks reveals conserved multistage regulators of neurogenesis." Mol Cell **82**(16): 2982-2999 e2914.

Jiang, L. D., C. Dalgarno, E. Papalexi, I. Mascio, H. H. Wessels, H. Y. Yun, N. Iremadze, G. Lithwick-Yanai, D. Lipson and R. Satija (2025). "Systematic reconstruction of molecular

pathway signatures using scalable single-cell perturbation screens." Nature Cell Biology **27**(3).
Lin, L., M. Zhang, P. Stoilov, L. Chen and S. Zheng (2020). "Developmental Attenuation of Neuronal Apoptosis by Neural-Specific Splicing of Bak1 Microexon." Neuron **107**(6): 1180-1196 e1188.
Mitra, S. K., D. A. Hanson and D. D. Schlaepfer (2005). "Focal adhesion kinase: in command and control of cell motility." Nature Reviews Molecular Cell Biology 2005 6:1 **6**(1): 56-68.
Papalexi, E., E. P. Mimitou, A. W. Butler, S. Foster, B. Bracken, W. M. Mauck, H. H. Wessels, Y. H. Hao, B. Z. Yeung, P. Smibert and R. Satija (2021). "Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens." Nature Genetics **53**(3): 322-+.
Paratcha, G., F. Ledda and C. F. Ibáñez (2003). "The neural cell adhesion molecule NCAM is an alternative signaling receptor for GDNF family ligands." Cell **113**(7): 867-879.
Parras, A., H. Anta, M. Santos-Galindo, V. Swarup, A. Elorza, J. L. Nieto-Gonzalez, S. Pico, I. H. Hernandez, J. I. Diaz-Hernandez, E. Belloc, A. Rodolosse, N. N. Parikshak, O. Penagarikano, R. Fernandez-Chacon, M. Irimia, P. Navarro, D. H. Geschwind, R. Mendez and J. J. Lucas (2018). "Autism-like phenotype and risk gene mRNA deadenylation by CPEB4 mis-splicing." Nature **560**(7719): 441-446.
Patowary, A., P. Zhang, C. Jops, C. K. Vuong, X. Z. Ge, K. C. Hou, M. Kim, N. H. Gong, M. Margolis, D. Vo, X. S. Wang, C. Y. Liu, B. Pasaniuc, J. J. Li, M. J. Gandal and L. de la Torre-Ubieta (2024). "Developmental isoform diversity in the human neocortex informs neuropsychiatric risk mechanisms." Science **384**(6698).
Santinha, A. J., E. Klingler, M. Kuhn, R. Farouni, S. Lagler, G. Kalamakis, U. Lischetti, D. Jabaudon and R. J. Platt (2023). "Transcriptional linkage analysis with in vivo AAV-Perturb-seq." Nature 2023 622:7982 **622**(7982): 367-375.
Satterstrom, F. K., J. A. Kosmicki, J. Wang, M. S. Breen, S. De Rubeis, J. Y. An, M. Peng, R. Collins, J. Grove, L. Klei, C. Stevens, J. Reichert, M. S. Mulhern, M. Artomov, S. Gerges, B. Sheppard, X. Xu, A. Bhaduri, U. Norman, H. Brand, G. Schwartz, R. Nguyen, E. E. Guerrero, C. Dias, C. Autism Sequencing, P.-B. C. i, C. Betancur, E. H. Cook, L. Gallagher, M. Gill, J. S. Sutcliffe, A. Thurm, M. E. Zwick, A. D. Borglum, M. W. State, A. E. Cicek, M. E. Talkowski, D. J. Cutler, B. Devlin, S. J. Sanders, K. Roeder, M. J. Daly and J. D. Buxbaum (2020). "Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism." Cell **180**(3): 568-584 e523.
Toutant, M., A. Costa, J. M. Studler, G. Kadare, M. Carnaud and J. A. Girault (2002). "Alternative splicing controls the mechanisms of FAK autophosphorylation." Mol Cell Biol **22**(22): 7731-7743.
Wang, J., Y. Jia, S. Zhao, X. Zhang, X. Wang, X. Han, Y. Wang, M. Ma, J. Shi and L. Liu (2017). "BIN1 reverses PD-L1-mediated immune escape by inactivating the c-MYC and EGFR/MAPK signaling pathways in non-small cell lung cancer." Oncogene 2017 36:45 **36**(45): 6235-6243.
Wang, W., L. Huang, Y. Huang, J. w. Yin, A. J. Berk, J. M. Friedman and G. Wang (2009). "Mediator MED23 links insulin signaling to the adipogenesis transcription cascade." Developmental cell **16**(5): 764-771.

Wang, W. J., J. C. Kuo, W. Ku, Y. R. Lee, F. C. Lin, Y. L. Chang, Y. M. Lin, C. H. Chen, Y. P. Huang, M. J. Chiang, S. W. Yeh, P. R. Wu, C. H. Shen, C. T. Wu and R. H. Chen (2007). "The tumor suppressor DAPK is reciprocally regulated by tyrosine kinase Src and phosphatase LAR." Molecular cell **27**(5): 701-716.
Wang, X., L. P. Weng and Q. Yu (2000). "Specific inhibition of FGF-induced MAPK activation by the receptor-like protein tyrosine phosphatase LAR." Oncogene 2000 19:19 **19**(19): 2346-2353.
Wood, S. N. (2011). "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models." Journal of the Royal Statistical Society Series B-Statistical Methodology **73**: 3-36.
Yang, X., M. Zhao, M. Xia, Y. Liu, J. Yan, H. Ji and G. Wang (2012). "Selective requirement for Mediator MED23 in Ras-active lung cancer." Proceedings of the National Academy of Sciences of the United States of America **109**(41).
Yang, Y., C. Li, Z. Chen, Y. Zhang, Q. Tian, M. Sun, S. Zhang, M. Yu and G. Wang (2023). "An intellectual disability-related MED23 mutation dysregulates gene expression by altering chromatin conformation and enhancer activities." Nucleic acids research **51**(5): 2137-2150.

Reviewer #1 (Remarks to the Author):

In this revision, the authors have clarified some challenging concepts. For instance, adoption of differentiation rank and eCDFs is more reader-friendly and illustrates the nature of gene expression changes upon microexon knockouts, i.e., precocious expression of neural genes. They have also addressed some points about the deletion of constitutive exons from the same set of host genes. Below are additional comments on the revised manuscript the authors might consider:

We thank the reviewer for their additional helpful comments on our manuscript. Our responses to these are provided below, along with a summary of the corresponding changes we have made to our manuscript (with all new text changes highlighted in yellow).

1. On microexon deletion efficiency in Fig. 1: We do not fully agree with the authors that the emphasis of Fig. 1 is to optimize the CHyMErA system by comparing editing efficiencies between Cas12a constructs. The representative efficiency of the system is a reliability indicator for all downstream efforts. There is still some concern regarding whether the deletion efficiencies listed in Fig. 1c-d reflect clonal ground truth. In Revision Fig. 1b, % edited clones (one assumes) provides a binary measure of deletions (heterozygous and homozygous deletions), and does not provide a clear indication of how frequent homozygous deletions may be, or account for the wider set of possibilities in bulk populations (het. versus hom. deletions). For example, 5/5 clonal lines with heterozygous deletions may be described as 100 % edited clones, but in the bulk population the same clones will have 50% deletions.

We apologize for the confusion. In our previous Rebuttal Fig. 1b, we were in fact showing the percent of edited alleles across clonal cell lines. Therefore, '% Edited Clones' indicates the percentage of editing while considering the zygosity of the deletion; i.e. 5/5 homozygous deletions would be considered 100% edited, whereas 5/5 heterozygous deletions would be considered 50% edited.

Moreover, as the length normalized band intensities only corrects for differences in intercalations, it does not correct for the more concerning amplification bias of shorter deletion amplicons. It seems disingenuous to ignore shorter amplicon bias. Perhaps linkage is lost in ddPCRs, but it could still provide a more grounded estimate of bulk editing. Beyond further efforts to resolve true editing efficiency, the authors may consider listing this as a limitation.

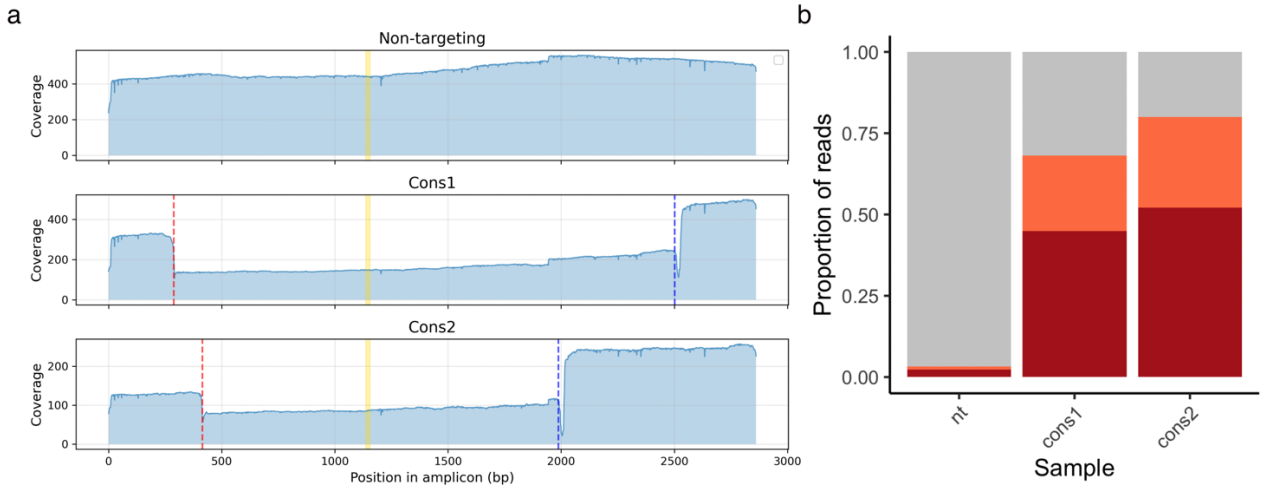
We agree with the reviewer that there could be potential amplification bias favouring detection of edited versus unedited bands. However, as previously shown in our original rebuttal, the editing efficiencies we observe are consistent with those observed in clonal populations, which serve as a 'ground-truth' for estimating editing efficiency.

We appreciate the reviewer's repeated suggestion to apply ddPCR to address this issue. However, in addition to our previous response regarding the inability of this approach to address the stated issues (i.e. zygosity of deletions, which we nevertheless have addressed by analysing clonal lines), we would also like to point out that this approach is far from straightforward and would require the design and optimization of distinct sets of primers with comparable efficiencies to capture edited and unedited genomes.

In order to further address the reviewer's concern, we have now provided an independent measure of the editing efficiency of our enhanced CHyMErA system using Oxford nanopore sequencing of amplified genomic regions targeted for deletion. This approach allows an assessment of editing efficiency independently of possible length-dependent measurement biases associated with agarose gel densitometry, such as ethidium bromide intercalation. We observe significantly reduced sequencing coverage over the expected deletion region for analyzed constructs, relative to the non-targeting control (Rebuttal Fig. 1a). We calculated the editing efficiency in each sample based on reads spanning Cas9/Cas12a deletion sites and the intervening region and observed 45% and 52% editing efficiency for constructs 1 and 2, respectively (Rebuttal Fig. 1b), whereas only 2% of non-targeting reads were classified as 'edited' (although likely due to misclassified, truncated reads). The values for editing efficiency determined using Nanopore sequencing are comparable (i.e. within 10%) to estimates of editing efficiencies calculated using semi-quantitative measurements of agarose gel band intensities. This validation approach is now mentioned in the Methods.

Page 21:

"To independently determine approximate editing efficiencies of the eCHyMErA method, amplified PCR products shown in Figure 1c were analyzed by Oxford Nanopore long-read sequencing (performed by Plasmidsaurus Inc.). Raw reads were aligned to sequences corresponding to the amplified regions. Editing efficiency was quantified by comparing the relative proportions of reads mapping to Cas9/Cas12a deleted regions versus adjacent unedited sites. Editing efficiencies estimated by quantification of gel band intensities and long-read sequencing were within 10% of each other."



Rebuttal Figure 1. Quantifying CHyMErA-directed deletions by Oxford Nanopore long-read sequencing. **a)** Read coverage surrounding the Ptk2 microexon 'A' targeted for CHyMErA deletion. Cut sites for Cas9 and Cas12a are indicated by red and blue dashed lines, respectively; position of microexon is indicated by yellow solid lines. **b)** Quantification of reads representing unedited (grey), repair (orange) or deletion (red) of the intervening genomic regions. Overall read counts analyzed: non-targeting (n=719 reads), Ptk2 deletion construct 1 (n = 624 reads) and Ptk2 deletion construct 2 (n=344 reads).

In summary, by determining ground truth editing efficiencies through the time-consuming generation and genotyping of clonal cell lines, and by using two independent methods for quantification of CHyMErA editing efficiencies in these bulk cell populations, we observe similar efficiencies. Furthermore, it is worth noting that we and others have extensively characterized and validated the CHyMErA system in previous publications directed at measuring phenotypes in response to deletions of targeted exons[1,2]. This includes a recent publication[2] that validates CHyMErA-detected exon deletion growth phenotypes using base-editing-induced skipping. To document how we have quantified editing efficiency using two independent types of measurements, and to acknowledge limitations in the accuracy of estimating editing efficiencies, we have revised the Figure 1 legend:

Figure 1c.

"Editing efficiencies were estimated from length-normalized quantification of ratios of edited and total band intensities, and are consistent (i.e. within 10%) of estimates determined by quantification of Oxford Nanopore sequencing reads mapping specifically to unedited and edited genomic fragments (Methods)."

2. The notion that heterozygote microexon deletions could have phenotypes is plausible, but

in this case the clonal het. lines should be used to demonstrate transcriptomic effects for validation targets.

We have already addressed this point by including validation experiments in our manuscript in which dCasRx-RBM25 was used to directed modulation of microexon splicing. Specifically, targeting the Clasp1(A) microexon results in partial (~50%) exon skipping, yet this level is sufficient to recapitulate transcriptomic phenotypes observed in the screen. As already mentioned in our response, previous focused analyses comparing heterozygous and homozygous deletions of specific neural-differential exons have similarly resulted in significant yet reduced transcriptomic and additional phenotypes for the heterozygous deletions[3, 4]. We have clarified this point in the Discussion.

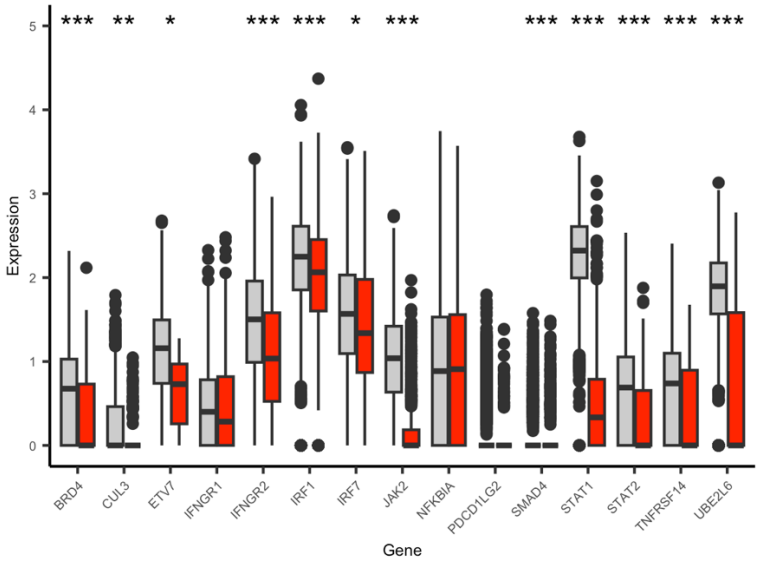
Page 16:

"Moreover, consistent with previous studies demonstrating that heterozygous deletion of alternative exons can have a significant impact on transcriptomic regulation[3, 4], the results of our CHyMErA-seq screen and follow-up experiments using dCasRx-RBM25-directed exon manipulation further demonstrate that induced partial skipping of microexons can lead to significant phenotypes, an observation that may also have relevance for interpreting the significance of partial exon skipping levels detected in patient brain samples."

3. The reliability of QDA still remains to be determined. The authors point to validation of example target exons in this manuscript, but this is unsatisfactory. New analytic pipelines such as this should be benchmarked against published datasets. A good example is single cell CRISPR knock out experiments where it is possible to quantify depletion of target gene.

We thank the reviewer for this excellent suggestion. We have performed the requested benchmarking using a previously published single-cell perturbation screen that utilized Cas9-mediated knockout of target genes[5]. As requested, we scored all cells according to their QDA-derived perturbation signature and quantified differences in the expression level of the target gene in cells classified as 'perturbed', relative to control cells expressing a non-targeting guide. We observe that 12/15 perturbations for which we detect a transcriptomic signature also have significant reductions in expression levels of the targeted gene ($P < 0.05$, Bonferroni-corrected Wilcoxon rank-sum test) (Rebuttal Fig. 2). This benchmarking comparison is now mentioned on page 7:

"Supporting the reliability of these results, benchmarking of our analysis pipeline against a published single-cell CRISPR screen dataset[5] reveals that 12/15 (80%) of perturbations called by our method have significant levels of reduction in expression of the target gene ($P < 0.05$, Bonferroni-corrected Wilcoxon rank-sum test)."



Rebuttal Figure 2. Expression of target gene for non-targeting cells (grey) and cells classified as perturbed by QDA pipeline (red). Data from Papalexi et al 2021. *P < 0.05, **P < 0.01, ***P < 0.001. Bonferroni-corrected Wilcoxon rank-sum test.

4. How many KO cells/cluster or subcluster per target for gene expression analyses?

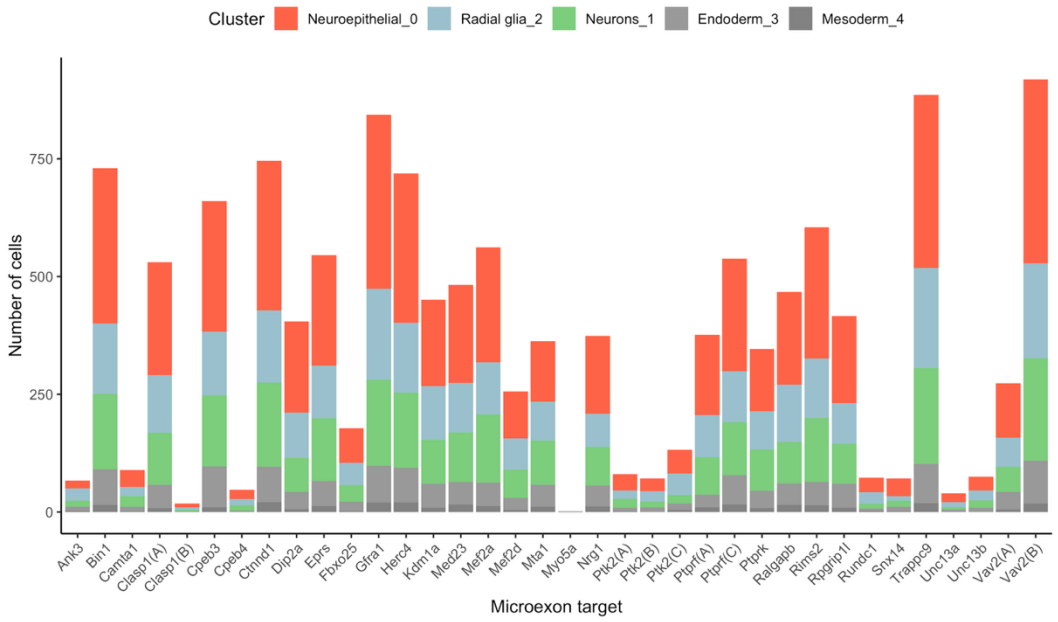
This information is provided in the legend for Figure 2d:

"**d)** Enrichment of 'knockout' cells compared to non-targeting cells within each subcluster, corresponding to deletion of Ralgapb (n = 105), Ptprf (microexon A; n = 158), Ptk2 (microexons A, B and C; n = 112), Med23 (n = 91), Gfra1 (n = 103), Clasp1 (microexon A; n = 130), Camta1 (n = 34) and Bin1 (n = 180) microexons (bottom)."

5. The representation of guide pairs in the sequenced nuclei library (and potential sources of bias) still needs further illustration. While we understand some basic ideas about #cells/target a detailed description of what the identities of these cells are will be most informative to the reader. In particular, as this study performs DGE on KO v non-targeting subclusters in pseudo-bulk, it is essential to understand how many cells/target microexon KO within each subcluster.

To address this concern, we have plotted the number of cells represented by each microexon-targeting guides for each cluster and added this to Extended Data Figure 3c. This new plot is referenced on page 7 of the revised text.

"We clustered cells and analysed marker gene expression within these trajectories to reveal the presence of expected neuroepithelial, radial glial, and neuronal clusters within the neuronal lineage (Fig. 2b, Extended Data Figs. 3a and 3b), with cells expressing microexon targeting guides for 24/36 detected microexon targets represented by at least 25 cells in each cluster within the neuronal lineage (Extended Data Fig. 3c)."



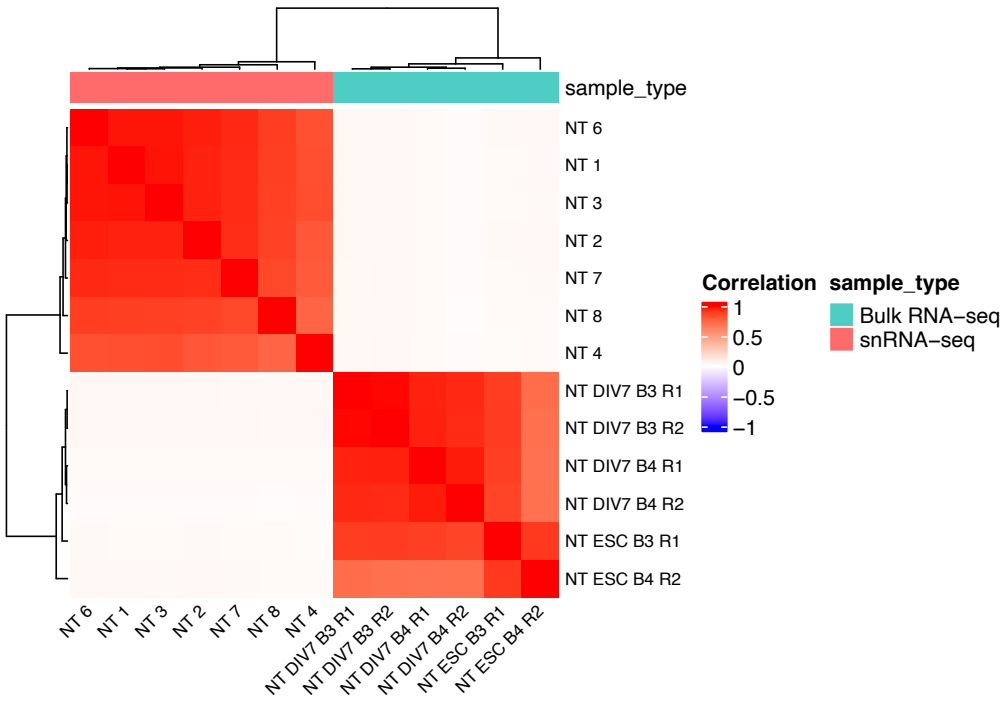
Rebuttal Figure 3. Number of cells assigned to each microexon target according to cluster.

We have noted – with concern – that there is only a moderate correlation between bulk and snRNA comparisons (including Figs. 2g v 3c Gfra1 KO and Ext Fig. 5c) and questioned the DGE outcome of the screen. Understanding whether low Ns could be a contributing weakness is quite important.

We acknowledge the reviewer's repeated concerns regarding the low correlation between the single-nuclei and bulk RNA-seq signatures. However, we would like to again emphasize that there are major inherent differences between these methods that will confound deriving direct correlations. Nevertheless, we have observed and reported positive and statistically significant correlations between the two methods, despite the manifold possible confounding factors.

To further illustrate the challenges, we demonstrate below the substantial differences between the transcriptomic data collected from bulk, whole-cell RNA-seq and single-nuclear RNA seq, when calculating correlations between gene expression patterns from

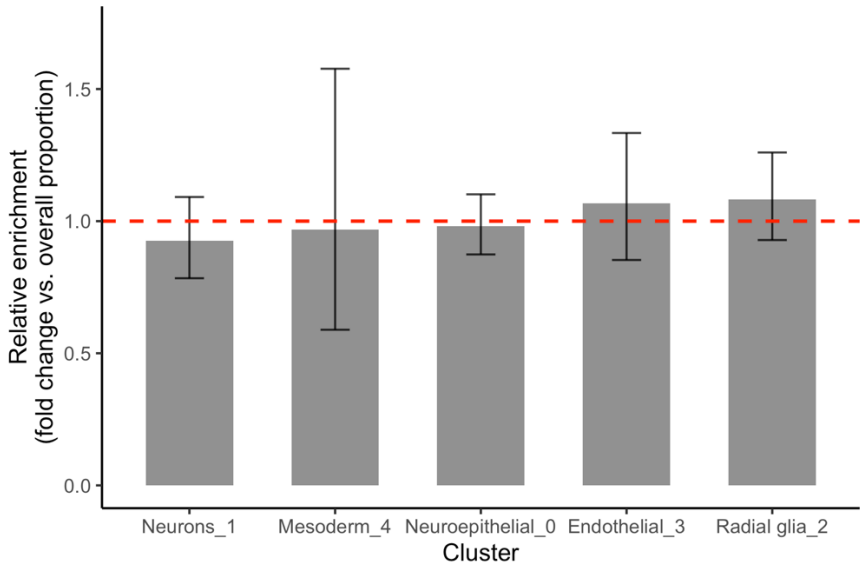
these data. As expected, there are no OVERALL significant correlations between expression levels measured using single-nuclear RNA-seq and the bulk RNA-seq, yet, despite this, we still observe positive and significant correlations for microexon perturbation-dependent gene expression signatures. (Rebuttal Fig. 4).



Rebuttal Figure 4. Visualizing the correlation between gene expression for non-targeting cells in single-nuclear RNA-seq (pink) and whole cell bulk RNA-seq (blue).

6. The microexon KO cells are broadly located in subclusters 0,2,4 and 11, all of which represent neuroepithelial (NE) identities (Fig. 2d). There is a strong, general trend for microexon KO cells to be found in neuroepithelial state, with most non-targeted cells in neuron subclusters (orange points in Fig. 2c). The authors have clarified that the microexons are included in NE stage, but it is not clear if KO cell enrichment in NE has true biological meaning, i.e., differentiation defects, or if this is some unexpected technical issue. As there is no replicate screen or characterization of differentiation using clonal microexon KO cells, the robustness of these findings remains uncertain.

Contrary to the reviewer's assertion, most of the non-targeting cells are in fact not located in the neuronal subclusters. Rather, they are dispersed throughout the clusters (see Rebuttal Figure 5).



Rebuttal Figure 5. Bar plot of relative enrichment of non-targeting cells in each of the clusters relative to the overall proportion of cells. Error bars indicate 95% Wilson score interval.

We are puzzled by the reviewer's suggestion that we have not further characterized this phenotype, since we have indeed characterized clonal cell lines for Gfra1 microexon deletion, as well phenotypes using multiple dCasRx-RBM25 cell lines designed to validate the effects of induced microexon skipping on differentiating cells (please see Figures 3 and 4 and the corresponding text).

7. Related to Fig. 2g, it would be useful to understand which subclusters enrich precocious neurogenesis signatures.

Figure 2g illustrates the distribution of differentially expressed genes resulting from microexon deletion with respect to the neuron differentiation score. As previously described, these scores represent the correlation between gene expression and neuron lineage progression. The information requested by the reviewer is in Figure 2d, where we show which subclusters are enriched in cells classified as 'perturbed'.

8. It is true that "estimating the 'true' proportion of edited/effected cells in bulk cell populations represents a more general challenge encountered in CRISPR-based screens", however there are general approaches to mitigate concerns, for example independent replicate screens which have not been performed.

It is unclear to us how a replicate screen would elucidate the 'true' portion of edited cells. Furthermore, as described in the previous rebuttal, we are performing extensive internal

replication by applying nine distinct guide pairs to target the same microexons. This strategy is significantly more cost efficient than performing multiple replicate screens with fewer guides per target (which would be required to maintain coverage while sequencing the same total number of cells) and avoids encountering batch effects that might otherwise compromise making direct comparisons between replicates.

9. On modest correlation between changes in gene expression between snRNA and bulk RNAseq datasets: technical differences between the modalities should be described using comparisons with data generated in this study. One cannot simply take the authors word that there are big differences modalities.

Please refer above to rebuttal Figure 4 and the associated description of this analysis.

Extended Fig. 5c shows an alarmingly poor correlation between bulk RNA-seq from a clonal line and CHyMErA-seq. GO-term overlaps as shown in Fig. 3b could indicate bona fide biology or easy-to-obtain terms within the context of neuronal differentiation. As this one of few validations done in this study, it highlights some critical challenges that remain with demonstrating robustness of screening with CHyMErA system.

We hope that by illustrating the major differences between these approaches (Rebuttal Fig. 4) the reviewer can appreciate the challenges when making direct comparisons between single-nuclear RNA-seq and whole cell bulk RNA-seq. However, as pointed out above and on pages 10 and 11 of our manuscript, despite these differences, positive and significant correlations were observed between altered gene expression signatures detected using these methods, in addition to the significant overlap in GO terms. Together we believe these results strongly support the validity of the phenotypic effects we observe in the CHyMErA-seq screen. Furthermore, the GO terms are all scored against a background of genes that are expressed in our *in vitro* differentiated neurons and therefore do not represent 'easy-to-obtain' enrichments. Even so, we followed-up on these GO-term enrichments, and in the case of Gfra1, when using clonal lines, we observe that in addition to displaying an enrichment of gene expression signatures that correspond to pathways involved in cell morphogenesis and neuron projection formation (Fig. 3b), we observe that that these neurons display morphological features that reflect the transcriptional changes (i.e. longer neurites) (Figs. 3d and 3e).

10. No experiments were done to further validate transcriptomic or other phenotypes in clonal KO lines.

This statement is incorrect. As clarified above and in our manuscript, we have validated some or our observations using clonal lines in which microexons are targeted at the RNA-level using dCasRx-RBM25 (please refer to Figures 3 and 4 and pages 9-12).

References

1. Gonatopoulos-Pournatzis, T. et al. Genetic interaction mapping and exon-resolution functional genomics with a hybrid Cas9-Cas12a platform. *Nat Biotechnol* **38**, 638-648 (2020).
2. Xiao, M.S. et al. Genome-scale exon perturbation screens uncover exons critical for cell fitness. *Mol Cell* (2024).
3. Gueroussov, S. et al. An alternative splicing event amplifies evolutionary differences between vertebrates. *Science* **349**, 868-873 (2015).
4. Han, H. et al. Systematic exploration of dynamic splicing networks reveals conserved multistage regulators of neurogenesis. *Mol Cell* **82**, 2982-2999 e2914 (2022).
5. Papalexi, E. et al. Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. *Nat Genet* **53**, 322-331 (2021).