nature portfolio

Peer Review File

Alternative splicing across the *C. elegans* nervous system

Corresponding Author: Dr Marc Hammarlund

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)
This manuscript describes the generation of a differential alternative splicing (DAS) atlas of 55 neuron types in C. elegans using individual neuron class, bulk RNA-seq generated from the CeNGEN project. The main goal of this work was to identify key genes that undergo DAS in specific neuron types. Additionally, the authors identify certain RNA sequence features that correlate with specific DAS events and train a model to generate a regulatory network of putative splicing factors involved in DAS, this can be used to generate hypotheses about novel splicing factors involved in DAS.

The authors highlight the capabilities of the tools they generated by highlighting a few genes that undergo DAS as examples. First, they provide an overview of the three ways they analyze their data using the ric-4 gene as an example. Raw data visualization can be used to view read coverage and junction counts in any genome browser, local quantification can be used to view the percent spliced-in for specific local splicing variation (LSV) events, while transcript-level quantification is used to measure specific transcript levels in each of the 55 neuron types and average transcript levels across all 55 neuron types. The authors then validate their analysis of DAS with previous literature using elp-1 and daf-2, showing that for daf-2 specifically, neuron-specific exon 11.5 splicing events identified in their data mostly agree with Tomioka et al., 2016. The authors then show that their data can uncover novel DAS events such as cassette exons, alt. 1st exons, and alt. splice sites.

The authors attempt in vivo validation for the specific alternative splicing events they highlight (sax-3, ceh-8, unc-40). Additionally, they are careful to not over-interpret their data, cautioning users to manually inspect raw data for any splicing event of interest. Overall, this work provides C. elegans researchers with numerous tools to study differential alternative splicing in 55 neuronal types and highlights the amazing capability of specific neuronal cell-types to possess their own unique transcriptome, in part due to DAS.

Main points
1. For all RT experiments, did the authors perform no reverse transcriptase controls? None are shown or mentioned in the methods. This is essential. Also, were the primers across exons or within each tested exon? Amplification across an expected junction would be important (not the primer spanning the junction but primers on two consecutive exons).
2. The description of “relative” and “absolute” expression levels at the end of page 7 is somewhat unclear. I believe that by “absolute” the authors mean relative across different neuron types (and not just within a neuron type). This is not true absolute quantification (which would require some external normalization such as the addition of spike-ins) and I find it misleading without further explanation.
3. In their global analysis of DAS the authors identify ~1,940 genes that are differentially spliced across neuron classes. They then classify these across different types of alternative splicing. It is very difficult to get from the current analyses how broad or specific these splicing events are, and quantitatively how complete they are. Showing a distribution of some specificity measure for all called DAS events, as well as a distribution of the maximum delta of inclusion/skipping would be very helpful to better understand the degree and possible impact of these events.
4. When describing the different types of splicing events, the authors state that intron retention is rare. This does not look too rare in their figure. Also, as the authors discuss, this number is at odds with the number reported by Wolfe et al. using a different algorithm. From both papers, it is unclear whether the authors think one approach has more false positives or false negatives. Perhaps the authors can intersect both sets of called events and collectively validate some of these events.
5. There is little analysis or discussion of what the consequences of all these splicing events would be at the protein level. For example, for all intron retention events, I imagine it would be relatively straightforward to compare lengths of predicted translation products with or without the retained intron, to broadly understand the impact of these events (i.e. premature stop or addition of extra, intron-encoded sequence).

6. For the microexon analysis, it would be useful to show how differentially spliced these are, perhaps through a table or a heatmap showing maximum delta PSI?
7. The network to identify new splicing factors can be a useful resource but its validation is currently not well documented. For the two cases the authors explore, they should report their observations in terms of true positives, false positives, false negatives (I realize true negatives would require much extensive additional validation). For example, from the last paragraph of page 18 it is very difficult to extract which of the known factors were found, which were not observed and which were found but don't play a role. Some experimental validation of new predicted regulators would be important to further support that this network can make good predictions.
8. Related to the web tool, it is great to have a way to access the data, thank you! I do have two comments though that would improve usability: the filter categories don't match the different types of alternative splicing events listed in the paper, it would be helpful if splicing events were classified in the same way. The second point is that there is currently no way to search by neuron

Minor points
1. In Figs 1C and D, colors indicating ric-4a and ric-4b are swapped, which may lead to confusion, please modify for consistency
2. In Fig 1D at the bottom, the standard deviation is not stated
3. For figures 3B and 4B, shouldn't the Y-axes be labeled the same, as they both represent PSI?
4. For fig 3C, maybe a schematic of primer location would be help to avoid confusion regarding size of amplicons. Also, the strain/specific sample from which cDNA was made from should be stated, and most importantly, no-RT controls should be shown.
5. For figure 5E, maybe the X-axis should be labeled "Number of neuron types sampled" instead
6. In line 123, I presume that authors meant "percent spliced in" instead of "percent selected in"
7. In line 130, "MAJIQ" is spelled incorrectly
8. In line 39, perhaps provide an estimate of the % of alt. spliced human genes?
9. In line 52, perhaps the two sentences can be rewritten to form one sentence
10. In line 220, perhaps the authors could expand on the possible significance of a truncated homeobox domain and if this phenomenon has been seen before in other proteins of the same family?
11. In line 320, could the authors comment on how many microexons are estimated to be in the C. elegans transcriptome?
12. In line 434, "interpret" should be "interpreting"
13. In line 448, "effort" should be "efforts"

(Remarks on code availability)

Reviewer #2

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

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
Weinreb et al analyze data generated by the CeNGEN project to produce an atlas for 55 neuron types in C. elegans and use it to study differential alternative splicing through parallel approaches of raw data or local splicing variation visualization/quantification and transcript-level quantification, illustrated by the analysis of the ric-4 gene and other known alternatively spliced genes from the literature. The authors intend to make these analyses tools available through an interoperable web portal. Detailed studies of transcripts of the unc-40, sax-3 and ceh-8 genes reveal the existence of (potentially biologically relevant) unannotated isoforms expressed only in specific neuron types and thus not detected in bulk RNA studies. The authors generalize this concept using their complete dataset and argue that > 1,300 novel isoforms have been identified as well as nearly 2,000 genes (projected to 3,000) that display differential splicing in different neuron types. This recapitulates and expand previous knowledge in the literature and shows particular prominence in genes related to neuronal excitability. Finally, the authors identify general sequence features in neuron-type differentially spliced transcripts, including an increase in frequency of regulated microexons, and develop a complex computational approach to identify networks of regulation that suggest potential splicing factors involved in the neuron type-specific control of alternative splicing.

Assessing alternative splicing at single cell resolution has been fraught with controversy due to artifacts in amplification of tiny amounts of RNA that can prevent accurate quantification of isoforms. The authors have taken advantage of the highly uniform development of C. elegans to isolate RNA from identical neurons from many animals, thus overcoming the issue of RNA amounts limitation. The paper, along with the co-submission by Wolfe et al. provides a highly valuable resource to identify cell-specific splicing variants and define mechanism of regulation.

In my opinion the following revisions would help to improve the manuscript:

1. While the authors' approach takes advantage of the unique uniform developmental plan of C elegans, it is not possible to discern, in the cases in which exons/introns/splice sites are not fully used or fully skipped, that individual cells do not have a distribution of PSI values because at the end the analysis is carried out in a population of cells (arguably identical, but that is exactly the point, they may not be so). It would be would to discuss this limitation in the text.
2. Figure 3 and 4: the analyses of transcripts from other species is interesting but raises the question of why, if the isoforms were unannotated in C. elegans and only the analyses of individual neuron types allowed their description, the isoforms are however readily detectable in bulk RNAseq datasets from other species. Is there a difference in the relative abundance of these transcripts between C. elegans and the other species?
3. Figures 3C, 4C, 4H: it is very good that the authors have used RT-PCR for the independent validation of the RNA-seq pipeline analyses. However, these results and associated schemes are difficult to understand. The position of the primers used for RT-PCR amplification of each product should be schematized and the predicted sizes of the amplification products for each isoform should be indicated. In the cases in which amplification products from two isoforms are expected (e.g. exon inclusion and skipping when amplification occurs using primers complementary to the distal exons), the position of the electrophoretic mobility of the two isoforms should be indicated.
4. Figure 5C: the results are very interesting but it may be helpful to further explore the structure of the data by running run PCA analyses to see whether there are clusters of neurons that differ more than others in differential splicing patterns (general or type-specific) and whether such clusters bear any relationship with other developmental, functional, anatomical, etc. criteria. The authors find a biological rationale for the ASK…AWB group (ciliated neurons). Is there a functional commonality for the I5…OLL group as well?
5. Figure 6: in sharp contrast with the results from the co-submission by Wolfe et al, intron retention is not a category of splicing changes particularly enriched in neuron type-specific regulation. The authors speculate in the Discussion that this apparent discrepancy may be due to the use of SUPPA as RNA-seq analysis tool. It would be relevant for the authors (of both papers, in fact) to solve this discrepancy by at least running a subset of samples on another pipeline (e.g. in the case of Weinreb et al using one more adapted to the detection of intron retention) and see whether this is indeed a plausible explanation.
6. Figure 6: the functional enrichment in microexons does not seem very persuasive (64% vs 51%). Is the difference between short and long exons more acute if microexons are compared to internal exons of more standard size (e.g. longer than 50 nucleotides)?
7. Figure 7: the computational approach used to derive regulatory networks is complex and professional but no proof is provided that the pipeline generates new predictions that can be validated. The authors could try to validate some novel predictions, at least bioinformatically. For example, is there an enrichment in putative binding sites for the specific splicing factors predicted to control a subset of alternative splicing event? In particular, the authors find a number of transcription factors as potential regulators of particular differential splicing events. At the same time, they find that the use of alternative promoters (first exons) is a prominent category of differential transcripts in their dataset. Is there a link between these two observations, i.e. could the authors predict binding sites for transcription factors found to regulate particular events and identify binding sites for these factors in the differential promoters? (I understand that the network was built on cassette exons, but they could do the same on differential first exon events to facilitate this analysis). Alternatively (or in addition), the effect of transcription factors could be indirect by inducing (or repressing) the expression of splicing factors, can these indirect effects be predicted through the presence of transcription factor binding sites in the promoters of putative splicing factor regulators?
8. The web portal www.splicing.cengen.org does not seem to be active and it is therefore not possible to provide feedback on this -very likely helpful- resource.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
This study by Weinreb et al uses the wealth of data generated by the CENGEN consortium to examine differential splicing patterns across neuron types. They provide examples of previously undiscovered alternative splicing events, discover types of genes that are most frequently alternatively spliced, and describe the most frequent types of alternative splicing events. Lastly, they describe the development of a splicing regulatory network prediction algorithm that may identify new splicing regulators for genes of interest. Taken together, this is a generally thorough work that will be an excellent resource for the community, and provides a wealth of information that will undoubtedly lay the groundwork for many future studies of alternative splicing in the C. elegans nervous system. There are a few issues that need addressing before the manuscript should be accepted.

Major Concerns:
1) The splicing regulatory network tool fails to predict most of the known splicing regulators of daf-2 (Figure 7E). The fact that it fails to predict what would be considered ground truth examples calls into question the overall reliability of the splicing network algorithm, yet the authors simply mention that "we did not detect" these known regulators. This warrants a much more thorough discussion.
2) The Authors give multiple examples of identifying novel isoforms or instances of alternative splicing (Figures 3 and 4), but it would have been nice to provide some biological significance to these new events. These genetic experiments are fairly straightforward in the worm, and at least one example should have been included.

Minor Concerns:

1) Discussion of previous studies of tissue specific splicing events in C. elegans is somewhat cursory. More in depth discussion would be helpful
2) Line 429 - the authors mention "(ref., Wolfe et al., manuscript)", which is likely a note to cite https://www.biorxiv.org/content/10.1101/2024.05.16.594572v1. The authors should update their references accordingly.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors addressed most of our points. There are a few minor things that we would strongly recommend including in the final version of their paper.

1. Plot S3A: This new plot provides a very good overview. However, this analysis would be more helpful if the table used to make it was included, with all events and the max delta PSI and Gini coefficient for each event, as an excel file.

2. To address the point of what is the consequence of the alternative splicing events on protein production, the authors provide a new analysis shown in 6E. This would be much more informative if it was done only for events in ORFs. Also, for the events that result in insertions that are multiples of 3, it would be helpful to indicate how many insert a stop codon.

3. The presented data and new analyses clearly show that the model is not very powerful (to a large extent because not enough data exist, both to make solid predictions or test them). We still think there is value in having this attempt, but doubt that other researchers will implement this as is. However, we do think that to increase usability of the data, the authors could include the expression across all neurons of the different SF shown in Table S5. At least this will provide a set of candidates for anyone interested in exploring further.

4. Typo in line 507: out/our

(Remarks on code availability)

Reviewer #2

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

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
I am happy with the revisions and I recommend publication in Nature Communications.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
This revised manuscript addressed mine and the majority of the other reviewers suggestions, and is greatly improved. This paper describes an excellent resource to the broad community, and provides new insight into differential alternative splicing.

I recommend the manuscript for acceptance.

(Remarks on code availability)

We would like to thank the editor and the reviewers for their thoughtful consideration of our manuscript. We have addressed all the reviewer comments including new experiments and analyses, which have confirmed and extended our original findings. Our responses are detailed below, and our changes highlighted in the manuscript text.

Reviewer #1 (Remarks to the Author):

This manuscript describes the generation of a differential alternative splicing (DAS) atlas of 55 neuron types in C. elegans using individual neuron class, bulk RNA-seq generated from the CeNGEN project. The main goal of this work was to identify key genes that undergo DAS in specific neuron types. Additionally, the authors identify certain RNA sequence features that correlate with specific DAS events and train a model to generate a regulatory network of putative splicing factors involved in DAS, this can be used to generate hypotheses about novel splicing factors involved in DAS.

The authors highlight the capabilities of the tools they generated by highlighting a few genes that undergo DAS as examples. First, they provide an overview of the three ways they analyze their data using the ric-4 gene as an example. Raw data visualization can be used to view read coverage and junction counts in any genome browser, local quantification can be used to view the percent spliced-in for specific local splicing variation (LSV) events, while transcript-level quantification is used to measure specific transcript levels in each of the 55 neuron types and average transcript levels across all 55 neuron types. The authors then validate their analysis of DAS with previous literature using elp-1 and daf-2, showing that for daf-2 specifically, neuron-specific exon 11.5 splicing events identified in their data mostly agree with Tomioka et al., 2016. The authors then show that their data can uncover novel DAS events such as cassette exons, alt. 1st exons, and alt. splice sites.

The authors attempt in vivo validation for the specific alternative splicing events they highlight (sax-3, ceh-8, unc-40). Additionally, they are careful to not over-interpret their data, cautioning users to manually inspect raw data for any splicing event of interest. Overall, this work provides C. elegans researchers with numerous tools to study differential alternative splicing in 55 neuronal types and highlights the amazing capability of specific neuronal cell-types to possess their own unique transcriptome, in part due to DAS.

Main points

1. For all RT experiments, did the authors perform no reverse transcriptase controls? None are shown or mentioned in the methods. This is essential. Also, were the primers across exons or within

each tested exon? Amplification across an expected junction would be important (not the primer spanning the junction but primers on two consecutive exons).

We updated the corresponding figures to show the exact location of the primers relative to the gene and to include no-RT controls.

2. The description of "relative" and "absolute" expression levels at the end of page 7 is somewhat unclear. I believe that by "absolute" the authors mean relative across different neuron types (and not just within a neuron type). This is not true absolute quantification (which would require some external normalization such as the addition of spike-ins) and I find it misleading without further explanation.

This could indeed be misleading, as we are not measuring absolute numbers of mRNA molecules. We reworded this sentence.

3. In their global analysis of DAS the authors identify ~1,940 genes that are differentially spliced across neuron classes. They then classify these across different types of alternative splicing. It is very difficult to get from the current analyses how broad or specific these splicing events are, and quantitatively how complete they are. Showing a distribution of some specificity measure for all called DAS events, as well as a distribution of the maximum delta of inclusion/skipping would be very helpful to better understand the degree and possible impact of these events.

Showing this type of metrics for MAJIQ quantifications is challenging, as LSVs can have variable number of splice junctions. We thus performed this analysis for the SUPPA2 quantifications, which encode all AS as events with binary outcomes, making it convenient to estimate a single metric for each event. We computed the maximal deltaPSI to reflect the strength of differential usage and Gini coefficient to reflect event specificity. We include this as an additional supplementary figure.

4. When describing the different types of splicing events, the authors state that intron retention is rare. This does not look too rare in their figure. Also, as the authors discuss, this number is at odds with the number reported by Wolfe et al. using a different algorithm. From both papers, it is unclear whether the authors think one approach has more false positives or false negatives. Perhaps the authors can intersect both sets of called events and collectively validate some of these events.

We removed the mention of intron retention being rare. As we believe the analysis of intron retention by Wolfe et al. is more thorough, we removed analysis of IR beyond Figure 6A, and refer readers to this paper.

5. There is little analysis or discussion of what the consequences of all these splicing events would be at the protein level. For example, for all intron retention events, I imagine it would be relatively straightforward to compare lengths of predicted translation products with or without the retained intron, to broadly understand the impact of these events (i.e. premature stop or addition of extra, intron-encoded sequence).

We further examined the impact of alternative events in the case of cassette exons and alternative splice sites (Fig 6 E and F). We only considered annotated events. We did not analyze alternative first exons with this approach, as this would require ensuring a correct annotation of start codons, introducing a further level of complexity.

We find that many of these events can impact the protein. However, the difference between the events that are DAS in neurons and those that are not, is relatively small, thus this should be seen as a feature of the *C. elegans* genome rather than unique to our dataset.

6. For the microexon analysis, it would be useful to show how differentially spliced these are, perhaps through a table or a heatmap showing maximum delta PSI?

We separately represented max DeltaPSI and Gini coefficient for microexons vs longer exons (see response to point 3) in Fig. S3B.

7. The network to identify new splicing factors can be a useful resource but its validation is currently not well documented. For the two cases the authors explore, they should report their observations in terms of true positives, false positives, false negatives (I realize true negatives would require much extensive additional validation). For example, from the last paragraph of page 18 it is very difficult to extract which of the known factors were found, which were not observed and which were found but don't play a role. Some experimental validation of new predicted regulators would be important to further support that this network can make good predictions.

We have used the known regulatory interactions to compute True and False Positive Rates during the training of the model, which are reported as a ratio in Fig. 7C. We now report the values in the text. However, it should be noted that these known regulatory interactions are established from whole-animal sequencing data, and may not be fully captured in our neuron-only dataset. As such, there is no available ground truth that can be used for a direct evaluation of model performance, we expanded the discussion on this.

Experimental validation in this context is extremely challenging, as it would require testing a large number of interactions to make any meaningful statement. We tried two approaches that could be scaleable to large numbers: HCR (similar to FISH, but with increased sensitivity that can in principle detect single exons) , and single-cell RT-PCR (we dissociated animals with the neurons of interest labeled, pipetted single cells into microwells, and performed RT-PCR). Unfortunately, neither technique gave robust results. Thus, we do not currently have techniques that would allow us to validate large numbers of splicing events in individual neurons.

8. Related to the web tool, it is great to have a way to access the data, thank you! I do have two comments though that would improve usability: the filter categories don't match the different types of alternative splicing events listed in the paper, it would be helpful if splicing events were classified in the same way. The second point is that there is currently no way to search by neuron

For the first point, we performed most of our analyses with MAJIQ, which in our experience best reflects the changes at the transcript level, and, while not easy to use in downstream analyses,

minimizes the risk of misinterpretation. While MAJIQ/VOILA does provide a "modulizer" tool to group LSVs into canonical events, we didn't use it here as it introduces a further level of abstraction. For examining individual event types in the paper, we used SUPPA, which categorizes AS in binary canonical events, making downstream analyses convenient, but requires extra care during interpretation. We thus decided to include the SUPPA quantification as a supplementary material, but not to highlight it on our website.

For the second point, we added a new module to compare neurons, that we introduce in Figure 1.

Minor points

1. In Figs 1C and D, colors indicating ric-4a and ric-4b are swapped, which may lead to confusion, please modify for consistency

We modified Fig 1C to match the color of Fig 1D.

2. In Fig 1D at the bottom, the standard deviation is not stated

Added in main text.

3. For figures 3B and 4B, shouldn't the Y-axes be labeled the same, as they both represent PSI?

Corrected in Fig 4B.

4. For fig 3C, maybe a schematic of primer location would be help to avoid confusion regarding size of amplicons. Also, the strain/specific sample from which cDNA was made from should be stated, and most importantly, no-RT controls should be shown.

We updated the three RT-PCR figures to make the position of the primers more clear, and display the no-RT controls. We added the genotype in the figure legends.

5. For figure 5E, maybe the X-axis should be labeled "Number of neuron types sampled" instead

Corrected.

6. In line 123, I presume that authors meant "percent spliced in" instead of "percent selected in"

This is the phrase used by the authors of the MAJIQ paper. While the authors do not appear to explain their rationale, this may be a more appropriate formulation in two cases: first, an LSV can reflect alternative promoters rather than splicing; second, in a given LSV, some junctions may correspond to an exon being spliced in, while others correspond to an exon being spliced out. Thus a formulation in terms of junction selected may avoid confusion.

In this paper, we work with 3 different definitions under the broad label of “PSI”: the Percent Selected In reported by MAJIQ, the binary PSI generated by SUPPA, and the exonic, binary PSI that we used as input for our network analysis. To avoid confusion, we renamed these three variables as “mPSI”, “bPSI”, and “ePSI” to make their different definitions more explicit.

7. In line 130, “MAJIQ” is spelled incorrectly

Corrected.

8. In line 39, perhaps provide an estimate of the % of alt. spliced human genes?

We now cite Wang 2008 and Pan 2008. We prefer not to give hard numbers as this is an evolving field.

9. In line 52, perhaps the two sentences can be rewritten to form one sentence

We prefer to have the standalone question “How is splicing regulated?” remain, with the goal of highlighting a major focus of the manuscript.

10. In line 220, perhaps the authors could expand on the possible significance of a truncated homeobox domain and if this phenomenon has been seen before in other proteins of the same family?

We added a sentence with two references to a similar phenomenon.

11. In line 320, could the authors comment on how many microexons are estimated to be in the C. elegans transcriptome?

We added the number (75) to the text. Further, we ensured that the “source data” for figure 6D contains the list of all cassette exons along with their length, enabling examination of this data by readers using only Excel.

12. In line 434, “interpret” should be “interpreting”

Done

13. In line 448, “effort” should be “efforts”

Done

Reviewer #2 (Remarks to the Author):

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

Reviewer #3 (Remarks to the Author):

Weinreb et al analyze data generated by the CeNGEN project to produce an atlas for 55 neuron types in C. elegans and use it to study differential alternative splicing through parallel approaches of raw data or local splicing variation visualization/quantification and transcript-level quantification, illustrated by the analysis of the ric-4 gene and other known alternatively spliced genes from the literature. The authors intend to make these analyses tools available through an interoperable web portal. Detailed studies of transcripts of the unc-40, sax-3 and ceh-8 genes reveal the existence of (potentially biologically relevant) unannotated isoforms expressed only in specific neuron types and thus not detected in bulk RNA studies. The authors generalize this concept using their complete dataset and argue that > 1,300 novel isoforms have been identified as well as nearly 2,000 genes (projected to 3,000) that display differential splicing in different neuron types. This recapitulates and expand previous knowledge in the literature and shows particular prominence in genes related to neuronal excitability. Finally, the authors identify general sequence features in neuron-type differentially spliced transcripts, including an increase in frequency of regulated microexons, and develop a complex computational approach to identify networks of regulation that suggest potential splicing factors involved in the neuron type-specific control of alternative splicing.

Assessing alternative splicing at single cell resolution has been fraught with controversy due to artifacts in amplification of tiny amounts of RNA that can prevent accurate quantification of isoforms. The authors have taken advantage of the highly uniform development of C. elegans to isolate RNA from identical neurons from many animals, thus overcoming the issue of RNA amounts limitation. The paper, along with the co-submission by Wolfe et al. provides a highly valuable resource to identify cell-specific splicing variants and define mechanism of regulation.

In my opinion the following revisions would help to improve the manuscript:

1. While the authors' approach takes advantage of the unique uniform developmental plan of C elegans, it is not possible to discern, in the cases in which exons/introns/splice sites are not fully

used or fully skipped, that individual cells do not have a distribution of PSI values because at the end the analysis is carried out in a population of cells (arguably identical, but that is exactly the point, they may not be so). It would be would to discuss this limitation in the text.

We agree, we now mention this heterogeneity in the Discussion.

2. Figure 3 and 4: the analyses of transcripts from other species is interesting but raises the question of why, if the isoforms were unannotated in C. elegans and only the analyses of individual neuron types allowed their description, the isoforms are however readily detectable in bulk RNAseq datasets from other species. Is there a difference in the relative abundance of these transcripts between C. elegans and the other species?

It doesn't appear that these transcripts are more abundant in *C. briggsae* and *brenneri* than *elegans*. If anything, Wormbase's aggregated data from SRA suggests that it is more highly expressed in *C. elegans* than *briggsae*. Gene (and transcript) annotation is a noisy process, whose outcome depends not only on the quality of the data, but also on the particularities of the algorithms used. Interpreting the differences between the annotations of different species would be a significant undertaking, and would be far beyond the scope of this study. We added a paragraph in the Discussion on this aspect.

3. Figures 3C, 4C, 4H: it is very good that the authors have used RT-PCR for the independent validation of the RNA-seq pipeline analyses. However, these results and associated schemes are difficult to understand. The position of the primers used for RT-PCR amplification of each product should be schematized and the predicted sizes of the amplification products for each isoform should be indicated. In the cases in which amplification products from two isoforms are expected (e.g. exon inclusion and skipping when amplification occurs using primers complementary to the distal exons), the position of the electrophoretic mobility of the two isoforms should be indicated.

We updated the figures to incorporate these recommendations.

4. Figure 5C: the results are very interesting but it may be helpful to further explore the structure of the data by running run PCA analyses to see whether there are clusters of neurons that differ more than others in differential splicing patterns (general or type-specific) and whether such clusters bear any relationship with other developmental, functional, anatomical, etc. criteria. The authors find a biological rationale for the ASK…AWB group (ciliated neurons). Is there a functional commonality for the I5…OLL group as well?

We could not find a functional commonality for this group. As this clustering reflects the number of DAS genes between neuron types, this cluster can be due to an actively regulated splicing program, but can also reflect a lack of differential splicing (i.e. that these neurons are in a common "default state"). Distinguishing between these two possibilities would require further development of advanced analytical tools. In addition, due to the non-binary nature of MAJIQ's LSV quantifications, it is challenging to formulate a PCA or WGCNA-based approach to identifying splicing programs in this context. Thus, we believe this question is best left for future works.

One simple approach is to perform GO enrichment analysis for the list of genes DAS between this cluster of neurons and the others; we do not find any significant enrichment (compared with an appropriate background list).

5. Figure 6: in sharp contrast with the results from the co-submission by Wolfe et al, intron retention is not a category of splicing changes particularly enriched in neuron type-specific regulation. The authors speculate in the Discussion that this apparent discrepancy may be due to the use of SUPPA as RNA-seq analysis tool. It would be relevant for the authors (of both papers, in fact) to solve this discrepancy by at least running a subset of samples on another pipeline (e.g. in the case of Weinreb et al using one more adapted to the detection of intron retention) and see whether this is indeed a plausible explanation.

We believe Wolfe et al.'s examination of intron retention is more thorough than our own and their additional RT-PCR addresses the concerns about validity of these events. We thus do not analyze intron retention beyond figure 6A and refer readers to Wolfe et al.'s paper.

6. Figure 6: the functional enrichment in microexons does not seem very persuasive (64% vs 51%). Is the difference between short and long exons more acute if microexons are compared to internal exons of more standard size (e.g. longer than 50 nucleotides)?

Indeed, increasing the required size of internal exons does lead to a slightly bigger difference. However, as choosing a threshold based on the available data could be a form of "p-hacking" increasing the risk of misinterpretation, we prefer to retain the generally accepted threshold of 27 bp.

We strengthened the filtering step before SUPPA quantification (requiring 5 TPMs for an event to be considered), this leads to a few changes, in particular the enrichment of microexons is now 75% vs 59%.

Overall, this difference does not point to a major difference between microexons and longer exons, however as this class of exon has garnered particular attention in the literature, we believe it might be of interest to some readers to report the difference that can be observed in our dataset.

7. Figure 7: the computational approach used to derive regulatory networks is complex and professional but no proof is provided that the pipeline generates new predictions that can be validated. The authors could try to validate some novel predictions, at least bioinformatically. For example, is there an enrichment in putative binding sites for the specific splicing factors predicted to control a subset of alternative splicing event? In particular, the authors find a number of transcription factors as potential regulators of particular differential splicing events. At the same time, they find that the use of alternative promoters (first exons) is a prominent category of differential transcripts in their dataset. Is there a link between these two observations, i.e. could the authors predict binding sites for transcription factors found to regulate particular events and identify binding sites for these factors in the differential promoters? (I understand that the network

was built on cassette exons, but they could do the same on differential first exon events to facilitate this analysis). Alternatively (or in addition), the effect of transcription factors could be indirect by inducing (or repressing) the expression of splicing factors, can these indirect effects be predicted through the presence of transcription factor binding sites in the promoters of putative splicing factor regulators?

We did look for an enrichment of SF binding sites, but this did not offer strong support either for or against our model's predictions. However, the strength of this analysis is limited by the low number of known SF binding sites, and the relatively low number of predicted targets for each SF.

Because the statistical properties of alternative first exons and alternative cassette exons are quite different, a network approach including these would have to be tailored to the new problem; network performance on identifying Transcription Factors would be a poor predictor of the network performance on identifying Splicing Factors. However, applying a network approach to alternative first exons to identify alternative promoters is indeed a very interesting idea, and studying the interaction of alternative promoters and alternative splicing is a question of great interest, that we hope to include in future works.

We expanded our Discussion to make the known limitations of our model more explicit.

8. The web portal www.splicing.cengen.org does not seem to be active and it is therefore not possible to provide feedback on this -very likely helpful- resource.

We apologize for providing an incorrect address in the text (the "www." should not have been included). We made the appropriate changes to ensure the portal can be reached both with and without "www." to match standard expectations.

Reviewer #4 (Remarks to the Author):

Major Concerns:

1) The splicing regulatory network tool fails to predict most of the known splicing regulators of daf-2 (Figure 7E). The fact that it fails to predict what would be considered ground truth examples calls into question the overall reliability of the splicing network algorithm, yet the authors simply mention that "we did not detect" these known regulators. This warrants a much more thorough discussion.

We expanded our Discussion to make the known limitations of our model more explicit.

2) The Authors give multiple examples of identifying novel isoforms or instances of alternative splicing (Figures 3 and 4), but it would have been nice to provide some biological significance to

these new events. These genetic experiments are fairly straightforward in the worm, and at least one example should have been included.

We generated a deletion of the novel exon 14.5 of unc-40, but did not observe any phenotype (added in Figure S2). It would be interesting to systematically test a larger number of mutants. However our experience as well as the previous literature (e.g. the CRISPR-SGI screen from Norris, Gracida and Calarco, eLife, 2017) suggest that disruption of most alternative events only induce mild phenotypes. Thus this approach would require extensive testing. This is a significant undertaking which is beyond the scope of this paper.

Minor Concerns:

1) Discussion of previous studies of tissue specific splicing events in C. elegans is somewhat cursory. More in depth discussion would be helpful

We expanded the Discussion to be more specific. In addition, we reprocessed the published dataset from Koterniak et al., created additional tracks on our genome browser, added tissue-regulated events to our Figure 6A, and added a paragraph in the Discussion mentioning our direct attempt at comparing serotonergic neurons in the two datasets.

2) Line 429 - the authors mention "(ref., Wolfe et al., manuscript)", which is likely a note to cite https://www.biorxiv.org/content/10.1101/2024.05.16.594572v1. The authors should update their references accordingly.

We added the reference to the preprint.

Reviewer #1 (Remarks to the Author):

1. Plot S3A: This new plot provides a very good overview. However, this analysis would be more helpful if the table used to make it was included, with all events and the max delta PSI and Gini coefficient for each event, as an excel file.

This data is available in the Source Data file. We added explicit mentions in the main text and figure legends.

2. To address the point of what is the consequence of the alternative splicing events on protein production, the authors provide a new analysis shown in 6E. This would be much more informative if it was done only for events in ORFs. Also, for the events that result in insertions that are multiples of 3, it would be helpful to indicate how many insert a stop codon.

We adapted our analysis to only include events intersecting an annotated CDS. This led to a small increase in the proportion of frame-preserving events, both for DAS and non-DAS events, and weakly reduced the difference between DAS and non-DAS events.

For the second question, here are counts of events that intersect a start or stop codon:

Alt. 3' ss				
category	length	no end	stop/start	%
not meas	PFS	144	21	12.7
not meas	%3	188	9	4.6
not DAS	PFS	92	14	13.2
not DAS	%3	268	8	2.9
DAS	PFS	73	30	29.1
DAS	%3	527	17	3.1

Alt. 5' ss				
category	length	no end	stop/start	%
not meas	PFS	166	30	15.3
not meas	%3	160	14	8.0
not DAS	PFS	121	19	13.6
not DAS	%3	145	9	5.8
DAS	PFS	60	30	33.3
DAS	%3	159	14	8.1

Cassette exon				
category	length	no end	stop/start	%
not meas	PFS	86	22	20.4
not meas	%3	178	13	6.8
not DAS	PFS	56	26	31.7
not DAS	%3	221	25	10.2

DAS	PFS	57	44	43.6
DAS	%3	357	39	9.8

Table legend:
"not meas": not measured
"PFS": the length of overhang or cassette exon is not a multiple of 3
"%3": the length of overhang or cassette exon is a multiple of 3
"no end": the overhang or cassette exon does not intersect a start or stop codon
"stop/start": the overhang or cassette exon does intersect a start or stop codon
"%": the proportion of events that intersect a start or stop codon.
Note this is not filtered for events intersecting a CDS, so the total numbers of events are slightly higher than table 2 of the manuscript.

It appears a substantial fraction of events in all categories intersect a start or stop codon. PFS events are more likely to intersect start/stop codons than frame-preserving events, which seems unsurprising (in that case the PFS is likely frame-preserving too, and the added length that is not a multiple of three may be in the UTR). We do not see obvious differences between DAS and non-DAS events. Altogether, it does not seem like this analysis brings new insights and would be of interest to most readers.

3. The presented data and new analyses clearly show that the model is not very powerful (to a large extent because not enough data exist, both to make solid predictions or test them). We still think there is value in having this attempt, but doubt that other researchers will implement this as is. However, we do think that to increase usability of the data, the authors could include the expression across all neurons of the different SF shown in Table S5. At least this will provide a set of candidates for anyone interested in exploring further.

We added a new Supplementary Data File 6 with SF expressions used here. In addition, we added a sentence in the Source availability section to improve the visibility of the Zenodo repository containing the data necessary to reproduce and build on our results.

4. Typo in line 507: out/our

Corrected, thank you.