nature portfolio

Peer Review File

Large-scale RNA-seq mining reveals ciclopirox olamine triggers TDP-43 cryptic exons

Corresponding Author: Dr Jonathan Ling

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 article by Irika et al., the authors develop a web-based tool to mine publicly available RNA-seq data for usage of alternative splicing or exon-exon junctional events. They then use this tool to look for inclusion of cryptic exons associated with loss of TDP-43 both in the SRA database with known FTD/ALS cases and in the GTEX database with more healthy donor tissues. From these analyses, they found cryptic exon inclusions in RNA-seq datasets in mouse tissues caused by antifungal drug CPX, previously unconnected to TDP-43 biology.

This study presents a resourceful and innovative strategy to re-analyze existing data that will be of great interest to the field. That they link TDP-43-related cryptic splicing to an anti-fungal drug in datasets unrelated to neurodegeneration is particularly novel and interesting. However, there are a few concerns, listed below, that should be addressed to improve the impact of the study and strengthen the hypotheses presented.

Major Comments:
1. The authors describe their SnapMine website as a user-friendly tool for those lacking a certain level of bioinformatics expertise. However, it does not seem that the current version of the tool addresses this gap. The authors should consider adding the following:
a. Flexibility for input coordinates: the current version requires exact coordinates for both the inclusion and exclusion exons. This greatly limits usability, as different studies even using the same hg38 assembly can report slightly different coordinates. We suggest adjusting the program to return nearest match instead of requiring an exact match.
b. Search by gene name: Often biologists, especially those without bioinformatics expertise do not know the coordinates they are looking for but do have a gene of interest.
c. Tissue type: add a graph showing PSI of each cryptic inclusion for each tissue type

2. There is a missed opportunity to explain how SnapMine differs from existing tools that mine splicing data, such as ASCOT, developed by the authors.

3. For the tissue-specificity studies, uneven representation of tissue samples from undiagnosed individuals with TDP-43 pathology may be a confounding variable. To confirm their findings, the authors should focus on some studies where different tissues were obtained from the same individual.

4. For figures 2A-C, the authors present the PSI per sample, but due to the large number of negative samples, it is unclear what percentage of samples have some cryptic inclusions. This is important for their interpretation of potentially “leaky” splice sites or TDP-43 dysfunction occurring in normal tissues, as undiagnosed individuals in the datasets may be an alternative explanation for rare cryptic inclusion.

5. Figure 4D, the i3N sample western is not convincing since the loading/GAPDH signal is so low. It should be repeated with higher loading for acurate CPX effects on TDP-43 protein levels. This is important since the authors use i3N system later.

6. For RNA-seq data, the finding that microglia and fibroblasts are resistant to CPX is particularly interesting since CPX is an anti-fungal used topically. However, these results are difficult to interpret without additional controls. The authors should test

the effect of CPX on TDP-43 levels as they did in the other cell types and/or perform alternative tests of CPX activity to prove it is active in the cells. Alternatively, it could be useful to look at different times or doses to determine if these tissues simply take longer to show the effects. Also, do they see decreased TARDBP expression levels in any of these tissues?

7. For Supplement 2A, why not analyze the RNA-seq data for cryptic inclusion instead of testing individual targets, as cryptic events could be tissue-specific as described by the authors and abundant targets like STMN2 were not chosen in the tested panel.

8. The authors propose that CPX treatment results in TDP-43 pathology through oxidative stress driven by iron/heavy metal toxicity. This is largely supported by detection of higher ferritin levels detected in the ALS patient samples and differential expression of metallothioneins in cultured cells. However, for the patient samples, there could be alternative causes for this effect, such as increased inflammation. To strengthen their model, the authors should perform additional measurements of ferritin in one of their controlled culture systems, iN3, PBMCs etc.

9. The authors state that observations of leaky STMN2 cryptic exon inclusion in the neuroblastoma cells lines rule out STMN2 cryptic splicing as a measure of TDP-43 depletion in the lines, however, they do not show or discuss analysis of TDP-43 levels in these same cell lines. An alternative explanation could be that TDP-43 dysfunction is occurring in these neuroblastoma cell lines. Emerging reports in fact do connect TDP-43 to various human cancers including neuroblastomas: https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2021.755096/full

Minor Comments:

1. Line 231, Figure 3C appears to be mislabeled, should be 3B.

2. Figure 3B, the authors displayed the top 20 samples in SRA and found only TDP-43 KO/KD/FTD cases. Could they instead look further for samples with high PSI, but lower ranking? The top 20 might just reflect the high number of FTD or TDP-43 centric studies in the SRA database and be masking other interesting results. It might be interesting to report the identity of the first non-TDP-43/FTD sample returned from the analysis.

3. Fig 4A has 2 conditions that are labeled the same with different results. Is this legend a typo?

4. Fig. 4F. It would be helpful to label the CE with arrows as done previously.

5. The authors should consider explaining more about the TDP-43 aptamer especially considering the high signal in control 3 (figure 5A).

Reviewer #2

(Remarks to the Author)
In this study, Sinha and colleagues query a database of RNA sequencing, Snaptron, which combines thousands of human and mouse RNA-seq experiments, to test whether splicing events linked to TDP-43 depletion, cryptic exons, are present in a range of different contexts. They identify the drug Ciclopirox as a modifier of TDP-43 levels, and suggest that it acts through increasing cellular stress to heavy metals through acting on iron. They then use digital neuropathology to show that patients with ALS, the principal TDP-43 pathology disease, have increased abnormal ferritin staining, suggesting an interesting mechanism for iron homeostasis in causing TDP-43 depletion in disease.

Overall the paper is interesting and builds on lots of previous work by the team. I have some suggestions to improve the clarity and rigour of the study, particularly the methods around SnapTron and the particular cryptic splicing events used.

Major comments

Intropolis, an earlier effort to assemble splice junctions from SRA that was subsumed into Snaptron, should also be cited in introduction (doi: 10.1186/s13059-016-1118-6).

The authors should state which explicit SnapTron databases they used (with version numbers) and when they were last updated.

Currently the paper is a little confusing to read as they refer to two different sets of cryptic exons throughout. To those outside the admittedly quite niche field of TDP-43-associated cryptic exons, can the authors better justify their two-tier approach?

Furthermore, the authors state that they used cryptic splicing events from references 11, 52-55. How were the data extracted, compared etc? Different papers use different tools, different cutoffs and assumptions. Did they take the union of these results, or apply some kind of filtering on numbers of times a splicing event was reported? This should be expanded upon if possible. One option is that Table 1 could report the number of papers that each event was observed, with citations.

On that same note, is there a relationship between “leakiness” of cryptic exon splicing in non-TDP contexts and the inclusion of these exons in the original TDP-43 knockdown data you used to identify them in? Are these exons less highly included upon knockdown, or more present in the controls? Different TDP knockdown papers used different cutoffs to call a splicing

event “cryptic”. This should be investigated, or at least discussed.

Fig 2D - the STMN2 plot shows large amounts of intronic reads in the cell lines, making it look very different from the TDP knockdown samples. Did the authors attempt to quantify intron retention between exon 1 and the cryptic exon, and compare it to the cryptic exon inclusion?

You state “TDP-43-associated cryptic exons are non-conserved across species and, as such, those identified in mice cannot be targeted in humans” but there is a cryptic exon in UNC13A in both mice and humans. Perhaps you should qualify this sentence? Is that famous cryptic exon not conserved?

Line 336 - for those of us unfamiliar with iron homeostasis, perhaps include a sentence here justifying why you chose to stain for Ferritin.

Was it not possible to construct TDP-43 burden and ferritin stains per cell to see if the two positively correlate in ALS patients? This would strengthen the results considerably.

The finding that SRSF3 knockdown acts synergistically with TDP-43 knockdown to de-repress cryptic splicing is intriguing. As there are other SRSF3 knockdown datasets available, for example in ENCODE, it would be useful to replicate this. Have the authors considered adding the ENCODE RNA-seq knockdowns to SnapMine? Potentially outside of scope but good to discuss.

Minor

SnapMine website has a typo - “Casette” should be Cassette.

Reviewer #3

(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 #4

(Remarks to the Author)
In this manuscript, Sinha et al. developed an R script, and an application called SnapMine that allows splicing junctions to be easily queried. Using this resource, they investigated the background of cryptic exon inclusion and found that HDGFL2 is a relatively specific biomarker for loss of function of TDP-43, whereas the inclusion of other cryptic exons is more common than expected in normal tissues. They searched and found that treatment with an FDA-approved antifungal, ciclopirox olamine (CPX), resulted in similar cryptic exon inclusion, accompanied by reduced level of TDP-43 protein. The link from CPX treatment to TDP-43 depletion and cryptic exon inclusion is not limited to mouse splenocytes but also mouse brain ex vivo, human peripheral blood monocytes, and iPSC derived neurons (i3N). CPX treatment was also studied in combined with proteosome inhibition. RNA seq analysis suggested that CPX treatment may reduce TDP-43 level via heavy metal response and oxidative stress. Overall, I find the study interesting and intriguing. However, I have a series of concerns regarding the conclusions before feeling comfortable recommending publication in Nature Communications.
My main concerns include the following:
(1) Under each cell/tissue type and treatment, a partially overlapping but also differing set of genes displays cryptic exon inclusion. It is not entirely clear whether they are specific to TDP-43 depletion and whether other cryptic exon inclusion events unrelated to TDP-43 depletion also occur in other genes. I think it will help address the first questions if the comprehensive list of genes can be generated along with their expression levels and whether cryptic exon inclusion is observed across all cell types and tissues studied.
(2) Because CPX is an iron chelator, I expect a lot of changes to cellular processes, not just TDP-43 depletion. It is not clear to me whether the link between CPX treatment and cryptic exon inclusion is specific to TDP-43.
(3) Iron dysregulation has been well documented in previous ALS studies, which should be acknowledged.
(4) Similarly to iron dysregulation, stress responses involve diverse cellular changes. It is hard to imagine that the effects are specific to TDP-43.
(5) The quantification of Western blots in Fig. 4D indicates that the abundance of TDP-43 barely decreased upon CPX treatment in i3N cells but cryptic exon inclusion was observed as shown in Fig. 3C. I wonder if there is a problem with quantifying the blots because the TDP-43 blot does show a pronounced change in signal intensity. Also, no replicates and error bars are provided so it is not clear how much variation is expected.
(6) It is concluded that CPX treatment leads to increased degradation of TDP-43 protein (line 302). However, inhibition of proteasomal degradation using MG-132, Carfilzomib, and MLN4924 on top of CPX treatment leads to a further increase in the cryptic exon inclusion. This is apparently inconsistent with prediction.
(7) TDP-43 aggregation is an important aspect of ALS pathology and also linked to oxidative stress. TDP-43 aggregation is not examined at all in this study.

A minor point: Line 292, IGLON5 is not in Fig. 4C.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have made significant revisions following review, which have improved the clarity of the manuscript. In particular, the updated diagrams, explanatory notes and explanation of snap mine in the context of prior splicing tools, are helpful. We appreciate that further efforts to improve usability are beyond the scope of the current tool. We also find that the additional analyses and experiments, including exclusion of putative undiagnosed or presymptomatic patients, repeating TDP-43 immunoblots in different cell types, and exploration of the available neuroblastoma cell line, all improve the rigor of the study. We agree with the decision to move the ferritin measurements and more detailed model to the supplement as additional experiments would be needed to fully test this model, which may be outside the scope of the current study.

We therefore recommend the study for publication. We note two minor things that the authors may want to consider for improving clarity of the figures. Fig 3B and D – perhaps label gray in the figure or legend as not measured (if that is indeed what gray represents). Fig 4F – the control images are too small for the reader to see puncta (or lack of puncta) within the cells, consider including enlarged insets for control as well to improve comparison (or put only insets in the main figure and reserve larger images for supplement).

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
The authors have paid close attention to my suggestions and have strengthened their manuscript. I have no further suggestions.

(Remarks on code availability)

Reviewer #3

(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 #4

(Remarks to the Author)
The authors addressed most of my concerns in the revision. Therefore, I support its publication in Nature Communications. I still have a request regarding Figure 4D. The Western blot of TDP-43 for human PBMCs shows substantial decrease upon CPX treatment. However, the bar graph on the right shows very little reduction. Can the authors check if this apparent discrepancy is an error in quantification? Also regarding the same panel, if biological replicates have been done for the Western blotting analysis, the data should be plotted in the bar graph with error bars indicated.

(Remarks on code availability)

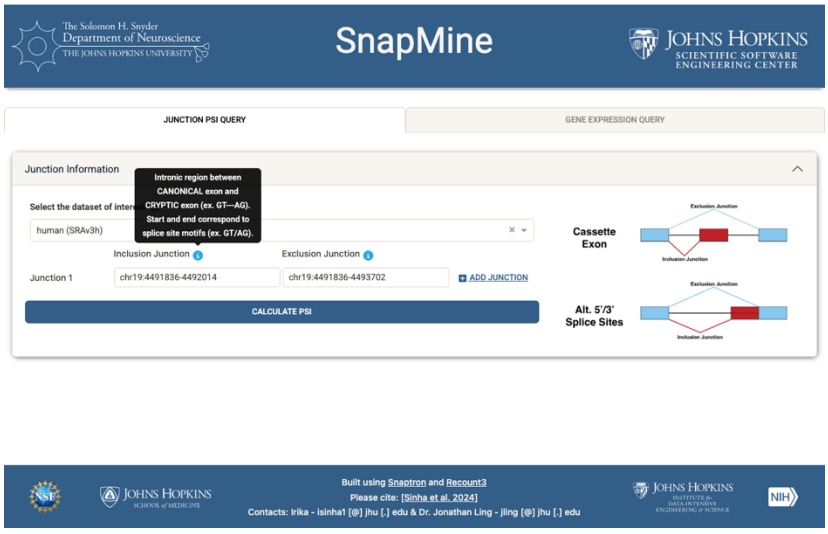
Reviewer #1 (Remarks to the Author):
Major Comments:

1. The authors describe their SnapMine website as a user-friendly tool for those lacking a certain level of bioinformatics expertise. However, it does not seem that the current version of the tool addresses this gap. The authors should consider adding the following:
    a. Flexibility for input coordinates: the current version requires exact coordinates for both the inclusion and exclusion exons. This greatly limits usability, as different studies even using the same hg38 assembly can report slightly different coordinates. We suggest adjusting the program to return nearest match instead of requiring an exact match.

We thank Reviewer #1 for suggesting improvements in flexibility regarding the input coordinates. While using a nearest-match approach could enhance usability, it's challenging to implement reliably due to the proximity of alternative or cryptic splice sites, which often differ by only 1–2 bases, and nearest-match could result in different matches for certain samples. Alternatively, implementing nearest-match logic globally, by taking junction count sums across all sample, risks selecting more abundant canonical junctions (instead of intended rare or cryptic events) and would likely compromise accuracy.

However, we recognize that we can improve SnapMine's usability. We have enhanced the SnapMine website by clearly defining "inclusion" and "exclusion" junctions through updated diagrams and explanatory notes:

- Inclusion junction: "Intronic region between CANONICAL exon and CRYPTIC exon (e.g., GT---AG). Coordinates correspond to splice-site motifs (e.g., GT/AG)."
- Exclusion junction: "Intronic region between TWO CANONICAL exons (e.g., GT---AG). Coordinates correspond to splice-site motifs (e.g., GT/AG)."



Furthermore, we strongly recommend users cross-validate junction coordinates using visualization tools like UCSC Genome Browser or IGV before querying SnapMine, and have added a notice to the application to help ensure mitigate discrepancies in junction coordinates. We are working with our software development team to implement it on the live application.

b. Search by gene name: Often biologists, especially those without bioinformatics expertise do not know the coordinates they are looking for but do have a gene of interest.

Alternative splicing frequently generates multiple isoforms with diverse exon inclusion patterns. Indeed, certain genes, such as *EPB41L4A* and *KIF21A*, contain multiple cryptic exons. Thus, searching by gene name alone would return many splicing events. To assist users in identifying TDP-43 cryptic exons, specifically, we have included Supplemental Tables 1 and 2 to provide genomic coordinates for commonly studied cryptic exons, along with their right and left junctions.

c. Tissue type: add a graph showing PSI of each cryptic inclusion for each tissue type

SnapMine currently utilizes a common graphical representation across all four data compilations. However, only the GTEx and TCGA datasets include tissue-type metadata, making it challenging to integrate tissue-specific graphs into the existing website without significant changes to the underlying infrastructure. To address this request, we are working with our software development team to incorporate the option to generate boxplots split by tissue type for GTEx and TCGA. This improvement is currently planned, but may require some time to implement fully.

2. There is a missed opportunity to explain how SnapMine differs from existing tools that mine splicing data, such as ASCOT, developed by the authors.

We agree. To clearly articulate the distinction between SnapMine and ASCOT, we have revised the manuscript on lines 84-93 and added the following:

“We previously developed ASCOT to perform comprehensive, annotation-free identification of binary splicing events using a computationally intensive junction-walking method. Due to its complexity, ASCOT's analyses are limited to precomputed compilations (GTEx, ENCODE, and two manually curated mouse datasets) and cannot directly query new events or rapidly extend analyses to large-scale archives like TCGA or the entirety of SRA. […] SnapMine employs a complementary approach to ASCOT, where users with a predefined splice junction query can instantly access splicing information across hundreds of thousands of diverse public RNA-seq datasets.”

3. For the tissue-specificity studies, uneven representation of tissue samples from undiagnosed individuals with TDP-43 pathology may be a confounding variable. To confirm their findings, the authors should focus on some studies where different tissues were obtained from the same individual.

(See response to point 4 below)

4. For figures 2A-C, the authors present the PSI per sample, but due to the large number of negative samples, it is unclear what percentage of samples have some cryptic inclusions. This is important for their interpretation of potentially “leaky” splice sites or

TDP-43 dysfunction occurring in normal tissues, as undiagnosed individuals in the datasets may be an alternative explanation for rare cryptic inclusion.

We agree with the reviewer that some individuals profiled by GTEx may have undiagnosed TDP-43 pathology. To address this concern, we performed an additional analysis of GTEx data that introduced exclusion criteria to account for possible undiagnosed cases. Tissues were considered positive for a cryptic exon if the average PSI was greater than or equal to 5% in at least one sample from that tissue. Since TDP-43 nuclear clearance leads to the inclusion of multiple cryptic targets, individuals who had five or more different cryptic exons measured in a single tissue type were excluded from the analysis on the basis that they likely had an undiagnosed TDP-43 pathology. This led to the exclusion of 62 individuals, most of whom had multiple detected cryptic exons within brain tissues. The majority of individuals were between the ages of 50-69, which is potentially consistent with pre-symptomatic onset of TDP-43 pathology.

Figure 2A-C has been updated after removing donors with potential TDP-43 pathology. We have also added Supplementary Figure 2 to describe individuals excluded due to potential TDP-43 pathology (Fig S2).

5. Figure 4D, the i3N sample western is not convincing since the loading/GAPDH signal is so low. It should be repeated with higher loading for accurate CPX effects on TDP-43 protein levels. This is important since the authors use i3N system later.

We agree that the levels of signal for the loading control in i3N are low. Additionally, the signal for the mouse brain blot was similarly weak. Both the i3N and mouse brain western blots in Figure 4D have been re-created with higher levels of loaded protein. The graph was updated accordingly and ratios to GAPDH were re-calculated for all blots.

6. For RNA-seq data, the finding that microglia and fibroblasts are resistant to CPX is particularly interesting since CPX is an anti-fungal used topically. However, these results are difficult to interpret without additional controls. The authors should test the effect of CPX on TDP-43 levels as they did in the other cell types and/or perform alternative tests of CPX activity to prove it is active in the cells. Alternatively, it could be useful to look at different times or doses to determine if these tissues simply take longer to show the effects. Also, do they see decreased TARDBP expression levels in any of these tissues?

We agree and believe that it is valuable to test multiple concentrations of CPX on the resistant cell-types and measure TDP-43 levels.

As mentioned, CPX is a topically used antifungal. Therefore, we performed further experiments on human dermal fibroblasts (HDF). HDFs were purchased and treated with CPX at a variety of times and with a variety of conditions. HDFs were treated with 20µM CPX for four hours at 2-days, 3-days, or 4-days post-seeding to investigate whether growth stage and confluency impacted cryptic exon incorporation. We found that the growth stage of the cells does not change the resilience of the cells to CPX and that TDP-43 was not depleted.

Furthermore, HDFs were treated with 20µM, 50µM, or 100µM of CPX for 24 or 48 hours. We observed that 100 µM treatment led to substantial cell death, while the TDP-43 levels within surviving cells did not differ from controls. However, 24 hour treatment of HDFs with 20µM or 50µM of CPX was sufficient to deplete TDP-43 levels and induce cryptic exons. These data have been added to Figure S3 A-E.

Lines 321-329 have been added to the manuscript to describe these experiments:
"Our results confirmed CPX's effect leading to TDP-43 depletion at a 20µM concentration for four hours in a cell-type specific manner (Fig. S3A-C). Increasing the dosage and treatment time of CPX to 50 µM for 24 or 48 hours in human dermal fibroblasts was also sufficient to induce TDP-43 depletion and cryptic exons (Fig. S3D-E), although 48-hour treatments also greatly increased cell death of the samples. 100 µM treatment appears not to impact TDP-43 protein levels while still inducing cryptic exons. We hypothesize that treatment led to initial cell death of saturated cells while those surviving after 24 or 48 hours had a survivable dosage of the drug and TDP-43 depletion (Fig. S3D-E). Additionally, CPX had the same effect in IMR-32 cells (Fig. S3I-K)."

7. For Supplement 2A, why not analyze the RNA-seq data for cryptic inclusion instead of testing individual targets, as cryptic events could be tissue-specific as described by the authors and abundant targets like STMN2 were not chosen in the tested panel.

While it is true that cryptic splicing events can be tissue-specific, Supplemental Figure 2A (now 3A) considers cryptic inclusion of individual targets that are considered widely expressed and have been detected in multiple tissues. *STMN2* is a primarily neuronally expressed gene and was not selected for testing in fibroblasts or microglia. *PFKP, HDGFL2, EPB41L4A*, and *AGRN* are more widely expressed and are readily detected across multiple tissues. RT-PCR was also used as an orthogonal validation method instead of relying solely on PSI values calculated from RNA-seq data, as it provided a consistent analytical approach across all samples given that sequencing every condition was not feasible.

8. The authors propose that CPX treatment results in TDP-43 pathology through oxidative stress driven by iron/heavy metal toxicity. This is largely supported by detection of higher ferritin levels detected in the ALS patient samples and differential expression of metallothioneins in cultured cells. However, for the patient samples, there could be alternative causes for this effect, such as increased inflammation. To strengthen their model, the authors should perform additional measurements of ferritin in one of their controlled culture systems, iN3, PBMCs etc.

We agree that ferritin measurements would be useful to strengthen our model. Unfortunately, levels of ferritin were difficult to quantify in the controlled culture systems, along with other iron regulation associated proteins, likely due to low basal expression. The previous figure has been moved to Supplementary Figure 5 as a hypothesized but unconfirmed pathway through which CPX can modify TDP-43 function, and the current Figure 5 diagram has been simplified.

We have also included additional data regarding the IMR-32 cell line model, in which CPX can induce TDP-43 depletion and cryptic exons (Supplemental Figure 3I-K). Levels of GPX4, a protein involved in inhibiting lipid peroxidation, were measured in the IMR-32 model system and appeared to be slightly reduced upon CPX treatment (Supplementary Figure 3K). Conversely, GPX4 levels were restored to baseline levels with concurrent N-Ac treatment.

9. The authors state that observations of leaky STMN2 cryptic exon inclusion in the neuroblastoma cells lines rule out STMN2 cryptic splicing as a measure of TDP-43 depletion in the lines, however, they do not show or discuss analysis of TDP-43 levels in these same cell lines. An alternative explanation could be that TDP-43 dysfunction is occurring in these neuroblastoma cell lines. Emerging reports in fact do connect TDP-43 to various human cancers including neuroblastomas: https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2021.755096/full

We thank the reviewer for the suggestion. Given that certain neuroblastoma lines have indeed been associated with TDP-43 depletion, we tested one of the referenced neuroblastoma cell lines, IMR-32, with a panel of cryptic exons and found that only cryptic *STMN2* and, to a much lesser extent cryptic *ARHGAP32*, were incorporated under normal culture conditions (without any treatments or experimental manipulations). As both cryptic exons are early polyadenylation events, we believe that the IMR-32 cell line may have leaky cryptic polyadenylation splice sites (Supplementary Figure 3I). Indeed, treatment with shTDP-43 leads to incorporation of the full panel of cryptic exons, confirming that TDP-43 depletion in this cell line should lead to the incorporation of all cryptic exons, not only cryptic polyadenylation sites. As a result, we conclude that TDP-43 is not depleted at baseline in the IMR-32 cell-line. Lines 215-218 have been added to the manuscript to emphasize this:
“This result was validated in the IMR-32 line. Of the tested cryptic exons, only *STMN2* was present at robust levels without any treatment (Fig S3J). There were also low levels of the *ARHGAP32* cryptic exon. However, other cryptic exons associated with TDP-43 depletion in that cell line were not present.”

Finally, we attempted test the other identified neuroblastoma cell lines, but were unable to source them in time for this revision due to issues related to material transfer agreement discussions.

Minor Comments:

1. Line 231, Figure 3C appears to be mislabeled, should be 3B.
This labeling has been corrected.

2. Figure 3B, the authors displayed the top 20 samples in SRA and found only TDP-43 KO/KD/FTD cases. Could they instead look further for samples with high PSI, but lower ranking? The top 20 might just reflect the high number of FTD or TDP-43 centric studies in the

SRA database and be masking other interesting results. It might be interesting to report the identity of the first non-TDP-43/FTD sample returned from the analysis.
We agree that it would be interesting to note the first non-TDP-43/FTD samples with high levels of cryptic exon incorporation. We extracted the top 20 samples in the human SRA database that did not include the words “TDP-43”, “ALS”, or “FTD” in the study title or abstract fields but had the highest average inclusion across the panel of cryptic exons and included it as a supplementary figure. Most of the top 20 samples had incorporation of 2 or 3 cryptic exons at most but none had inclusion of all tested cryptic exons. It is possible this is due to cell-type specificity of cryptic exons or high expression of an individual cryptic exons (Supplementary Figure 1D).

3. Fig 4A has 2 conditions that are labeled the same with different results. Is this legend a typo?
Figure 4A has been corrected by addition of missing detail. While two lanes are from samples both treated with CPX, one was treated with at 5μM and the other at 20μM.

4. Fig. 4F. It would be helpful to label the CE with arrows as done previously.
Figure 4F has been updated with the arrows to identify cryptic exon-associated gel bands as requested. It is now figure 4G due to the addition of 4E.

5. The authors should consider explaining more about the TDP-43 aptamer especially considering the high signal in control 3 (figure 5A).
To further describe the TDP-43 aptamer, we have revised the manuscript and added the following on lines 373-376:
“Aptamers are artificial oligonucleotides with the capability to bind with high specificity to a target, similar to antibodies. This RNA aptamer is highly sensitive and can be used to identify pathological TDP-43 aggregation events, including in pre-symptomatic cases.”

The following line has also been added to the legend of Fig. S5A:
“Due to the high levels of TDP-43 aptamer staining in control 3, it is possible that this patient is an undiagnosed or pre-symptomatic TDP-43 proteinopathy case.”

Reviewer #2 (Remarks to the Author):

Major comments

1. Intropolis, an earlier effort to assemble splice junctions from SRA that was subsumed into Snaptron, should also be cited in introduction (doi: 10.1186/s13059-016-1118-6).

We agree and have added this citation.

2. The authors should state which explicit SnapTron databases they used (with version numbers) and when they were last updated.

We thank the reviewer for the suggestion. The following details have been added to the methods section on lines 530-534:
"These data were processed during the creation of recount3 and include sequencing data from TCGA v3.0, GTEx v8, SRA human datasets as of 10/06/2019, and SRA mouse datasets as of 01/08/2020 (115,116). The ENCODE shRNA-seq data were processed by Snaptron and uploaded 06/2018 (52)."

3. Currently the paper is a little confusing to read as they refer to two different sets of cryptic exons throughout. To those outside the admittedly quite niche field of TDP-43-associated cryptic exons, can the authors better justify their two-tier approach?

The reviewer makes a good point regarding the different sets of TDP-43-associated cryptic exons. Unfortunately, cryptic exons reside within intronic regions of genes that are highly variable across species. Thus, cryptic exons found in mice are entirely different from those found in human, hence the two sets of cryptic exons. We have attempted to clarify by including the following lines in the manuscript:

"TDP-43-associated cryptic exons are nonconserved across species. Therefore, cryptic exons identified in mice cannot be explored in human data." [Lines 260-261]

"Due to the nonconserved nature of cryptic exons, they differ between human and mouse queries." [Lines 493-494]

Supplementary Tables 1 and 2 are also available for readers to view the mouse and human cryptic exon coordinates used in this study.

4. Furthermore, the authors state that they used cryptic splicing events from references 11, 52-55. How were the data extracted, compared etc? Different papers use different tools, different cutoffs and assumptions. Did they take the union of these results, or apply some kind of filtering on numbers of times a splicing event was reported? This should be expanded upon if possible. One option is that Table 1 could report the number of papers that each event was observed, with citations.

As the reviewer mentions, the authors used splicing events from references 11, 58-62 (updated numbers). However, the authors did not manually analyze the associated data. Instead, cryptic exon events were identified in the following ways:

- 11 - [Ling JP *et al*, PMID 26250685] there is a data table which includes coordinates of cryptic exons identified in HeLa cells
- 58 - [Klim *et al*, PMID: 30643292] identifies 9 transcripts as TDP-43 targets for alternative splicing regulation in Figure 1E using human ESC-derived motor neurons
- 59 - [Melamed *et al*, PMID: 30643298] identifies *STMN2* specifically as a target of TDP-43 that is misspliced after TDP-43 depletion in SH-SY5Y
- 60 - [Ma *et al*, PMID: 35197626] identifies *UNC13A* specifically as a target of TDP-43 that is misspliced after TDP-43 depletion in human post-mortem samples. The authors used MAJIQ and LeafCutter pathways for splicing analysis, but did not verify other novel targets.
- 61 - [Brown *et al*, PMID: 35197628] identifies *UNC13A* and *UNC13B* specifically as targets of TDP-43 that is misspliced after TDP-43 depletion in human iPSC-derived cortical-like i3Neurons
- 62 - [Jeong YH *et al*, PMID 28153034] includes a data table with coordinates of mouse cryptic exons in stem cells, neurons, and myocytes.

The majority of cryptic exons in this analysis are derived from the tables in [Ling JP *et al*, PMID 26250685] and [Jeong YH *et al*, PMID 28153034]. We have modified lines 122-124 in the manuscript regarding cryptic events to:

“We queried the previously described known TDP-43-associated cryptic exon targets identified from human and mouse TDP-43 knockdown datasets (11,58–62)”

5. On that same note, is there a relationship between “leakiness” of cryptic exon splicing in non-TDP contexts and the inclusion of these exons in the original TDP-43 knockdown data you used to identify them in? Are these exons less highly included upon knockdown, or more present in the controls? Different TDP knockdown papers used different cutoffs to call a splicing event “cryptic”. This should be investigated, or at least discussed.

We agree, the reviewer identifies an important point regarding the inherent subjectivity and variability in the criteria used across our field to define cryptic splicing events. In this paper, we use the term “leaky” to describe the cryptic exons for which at least 1% of all samples in the SRA have PSI values greater than or equal to 5%. To fully clarify, cryptic exons investigated in this paper were identified by prior papers and we do not attempt explore novel, previously unreported cryptic exons.

We have added the following lines 201-203 to the manuscript: “Samples were considered “positive” for a cryptic exon if the PSI average for cryptic junctions were greater than or equal to 5%. We considered cryptic exons as “leaky” if least 1% of all samples in the SRA were positive for the cryptic exon.”

We also have added the following lines 219 to the manuscript: “Overall, 4.6% of human SRA samples were positive for the cryptic splice site in STMN2.”

6. Fig 2D - the STMN2 plot shows large amounts of intronic reads in the cell lines, making it look very different from the TDP knockdown samples. Did the authors attempt to quantify intron retention between exon 1 and the cryptic exon, and compare it to the cryptic exon inclusion?

The reviewer is correct that increased levels of intron retention are also observed in the identified cell-lines, likely due to lower levels of transcription as *STMN2* is much more abundant in mature neurons than neuroblastoma cell lines. In this paper, we focused on the cryptic exon junction itself rather than the more complicated calculation of intron retention (only half the intron is retained, which potentially suggests an alternative polyadenylation site instead of complete intron retention). With deeper sequencing, it may be possible to disambiguate between intron retention and alternative polyA, but for the purposes of this manuscript we did not consider this analysis necessary.

7. You state “TDP-43-associated cryptic exons are non-conserved across species and, as such, those identified in mice cannot be targeted in humans” but there is a cryptic exon in UNC13A in both mice and humans. Perhaps you should qualify this sentence? Is that famous cryptic exon not conserved?

The reviewer is correct that there are cryptic exons in both mouse and human *UNC13A*. It is a rare coincidence that both the human (*UNC13A*) and mouse (*Unc13a*) homologues (*UNC13A* and *Unc13a*) have TDP-43-associated cryptic exons. However, they occur at entirely different locations within the gene (intron 20 in human, intron 1 in mouse) and are more of a coincidence rather than conservation.

The following has been added to the manuscript on lines 270-272:
“We note that the cryptic exon in mouse *Unc13a* is different from the human *UNC13A* cryptic exon. They occur at different regions within the gene, are of different lengths, and have different sequences.”

8. Line 336 - for those of us unfamiliar with iron homeostasis, perhaps include a sentence here justifying why you chose to stain for Ferritin.

(Response to point 8 grouped with response for point 9)
Lines 377-378 have been added to the manuscript:
“For grey matter, histological iron burden detection is well represented by immunostaining with a ferritin antibody.”

9. Was it not possible to construct TDP-43 burden and ferritin stains per cell to see if the two positively correlate in ALS patients? This would strengthen the results considerably.

For grey matter, histological iron burden detection is well represented by immunostaining with a ferritin antibody. Other methods of measuring iron directly in tissues (e.g. Perl's stain) rely on histochemical reactions which can be impacted by pre-treatment protocols, etc. and therefore lack reproducibility compared to antibodies and lack ability to capture ferritin-stored, as well as free iron (Lours et al., 2022, Duijn et al., 2013). The spatial staining pattern of ferritin is thought to arise due to upregulation in glial cells surrounding neurons which contain high free iron in a protective response (Kwan et al., 2012). This means that a regional quantification, rather than cell specific comparison, is the most accurate representation of iron stores.

The finding that SRSF3 knockdown acts synergistically with TDP-43 knockdown to de-repress cryptic splicing is intriguing. As there are other SRSF3 knockdown datasets available, for example in ENCODE, it would be useful to replicate this. Have the authors considered adding the ENCODE RNA-seq knockdowns to SnapMine? Potentially outside of scope but good to discuss.

The reviewer is correct that adding the ENCODE RNA-seq shRNA dataset to SnapMine could be a useful resource to look at splicing modifications. Since only a subset of the ENCODE datasets are available in Snaptron and they are not a public part of Recount3, we did not originally include them in SnapMine. Additionally, the ENCODE datasets pre-processed by Snaptron as a part of the SRAv3h compilation does not include the datasets with shRNA or CRISPR-induced SRSF3 knockdown.

The authors have updated the application so that SnapMine now is able to access the ENCODE shRNA datasets that are available among the Snaptron data compilations. These are mainly in K562 and HEPG2 cells and can be queried by selecting "human (ENCODE shRNA)" as the dataset of interest.

Minor

SnapMine website has a typo - "Casette" should be Cassette..
This error has been corrected

Reviewer #4 (Remarks to the Author):

My main concerns include the following:
(1) Under each cell/tissue type and treatment, a partially overlapping but also differing set of genes displays cryptic exon inclusion. It is not entirely clear whether they are specific to TDP-43 depletion and whether other cryptic exon inclusion events unrelated to TDP-43 depletion also occur in other genes. I think it will help address the first questions if the comprehensive list of genes can be generated along with their expression levels and whether cryptic exon inclusion is observed across all cell types and tissues studied.

The reviewer is correct that a comprehensive list of TDP-43 cryptic exon inclusion events would help clarify whether all the measured events are TDP-43-associated. To create such a database, a large array of cell types would need to be queried across both mouse and human since many cryptic exons are cell-type specific and are not conserved between species. Several other labs, with extensive expertise in differentiating iPSCs into various cell types, are actively working on this problem and we are currently waiting on their efforts. Thus, we believe creating a TDP-43 cryptic exon “compendium” is an important goal, but outside of the scope of this paper.

To address the reviewer’s concern of whether the cryptic exons measured are specific to TDP-43, we would like to emphasize that the events mentioned in this paper have been identified and validated by previous studies [Ling, Jonathan P. *et al.* 2015 PMID: 26250685], [Klim *et al.* 2019 PMID: 30643292], [Melamed *et al.* 2019 PMID: 30643298], [Ma *et al.* 2022 PMID: 35197626], [Brown *et al.* 2022 PMID: 35197628].

We have also modified lines 122-124 in the manuscript regarding cryptic events to:
“We queried the previously described known TDP-43-associated cryptic exon targets identified from human and mouse TDP-43 knockdown datasets (11,58–62).”

(2) Because CPX is an iron chelator, I expect a lot of changes to cellular processes, not just TDP-43 depletion. It is not clear to me whether the link between CPX treatment and cryptic exon inclusion is specific to TDP-43.

The reviewer is correct that CPX is an iron chelator and treatment could change more than one process. However, one of its effects is certainly upstream of TDP-43, as a wide variety of TDP-43 cryptic exons are detected. To date, genome-wide induction of cryptic exons is not possible without loss of TDP-43 function. Furthermore, we find that TDP-43 protein levels are depleted upon treatment with CPX (Fig 4D). In analyzing hundreds of thousands of datasets, we find very few datasets that exhibit TDP-43 cryptic exons, despite a variety of experimental manipulations. Thus, while we agree that CPX is inducing many changes to cellular processes, it belongs to a very small set of compounds that can robustly induce TDP-43 cryptic exons. We therefore believe that the pathways activated by CPX warrants further investigation.

(3) Iron dysregulation has been well documented in previous ALS studies, which should be acknowledged.
We agree and have added the following sentence to line 463: “Furthermore, past studies have similarly identified iron dysregulation in ALS cases.” along with the following citations: [Oba *et al*. 1993 PMID: 8234713] [Imon *et al*. 1995 PMID: 8847542] [Kasarskis *et al.* 1995 PMID: 8586987] [Berg *et al*. 2002 PMID: 12709302] [Goodall *et al* 2008 PMID: 18677636] [Jeong *et al*. 2009 PMID: 19158288] [Kwan *et al* 2012 PMID: 22529995] [Veyrat-Durebex *et al*. 2014 PMID: 25101285] [Zheng *et al*. 2017 PMID: 27804118] [Mitani *et al.* 2021 PMID: PMID: 34219295] [Gao *et al*. 2024 PMID: 39001793]

(4) Similarly to iron dysregulation, stress responses involve diverse cellular changes. It is hard to imagine that the effects are specific to TDP-43.
We have grouped our response with the response to point #2.

(5) The quantification of Western blots in Fig. 4D indicates that the abundance of TDP-43 barely decreased upon CPX treatment in i3N cells but cryptic exon inclusion was observed as shown in Fig. 3C. I wonder if there is a problem with quantifying the blots because the TDP-43 blot does show a pronounced change in signal intensity. Also, no replicates and error bars are provided so it is not clear how much variation is expected.

We agree with the reviewer that the TDP-43 blotting in i3N and mouse brain could be greatly improved. Both the i3N and mouse brain western blots in Figure 4D have been recreated with higher levels of protein loaded per lane. The graph was also updated accordingly and ratios to GAPDH were re-calculated for all blots.

Additionally, replicates for cryptic exon inclusion in i3N are included as a part of Supplementary Figure 3H.

(6) It is concluded that CPX treatment leads to increased degradation of TDP-43 protein (line 302). However, inhibition of proteasomal degradation using MG-132, Carfilzomib, and MLN4924 on top of CPX treatment leads to a further increase in the cryptic exon inclusion. This is apparently inconsistent with prediction.

The reviewer is correct, inhibition of proteasomal degradation does not lead to rescue of the TDP-43 depletion phenotype, suggesting that other degradation pathways may be playing a role. Recent work suggests that various post-translational modifications can regulate the stability of TDP-43 [Suk et al. 2025, PMID: 40149017] and that these processes interact with the proteosome in unexpected ways [Zhang et al. 2024, doi:10.1101/2024.09.04.611121]. It remains to be seen how CPX, iron chelation, and oxidative stress relate to these findings, and ongoing studies in our lab aim to address this question.

(7) TDP-43 aggregation is an important aspect of ALS pathology and also linked to oxidative stress. TDP-43 aggregation is not examined at all in this study.

We agree with the reviewer that aggregated TDP-43 is an important aspect of ALS-FTD pathology. Aggregated, pathogenetic TDP-43 can be isolated in urea-soluble protein fractions [Neumann et al. 2006, PMID: 17023659] [Scialo et al. 2025, PMID: 40157355] [Rummens et al. 2025, PMID: 40157356], thus we compared TDP-43 levels in RIPA-soluble and urea-soluble protein fractions from i3Neurons. Aggregated TDP-43 was not detected in the urea-soluble fraction, indicating that CPX treatment does not lead to insoluble aggregation of the protein.

We have added this data as Figure 4E with the following caption: “Measurement of TDP-43 by immunoblot in RIPA-soluble and urea-soluble fractions indicate no aggregated, urea-soluble TDP-43 in i3Neurons after four-hour treatment with 20µM CPX.”

We have also added the following to the manuscript on lines 334-336:
“However, TDP-43 is not found in the urea-soluble fraction of protein lysate after CPX treatment, suggesting aggregation is not occurring (Fig. 4E).”

A minor point: Line 292, IGLON5 is not in Fig. 4C.
This error has been corrected and IGLON5 has been removed from the sentence.

Reviewer #1 (Remarks to the Author):
The authors have made significant revisions following review, which have improved the clarity of the manuscript. In particular, the updated diagrams, explanatory notes and explanation of snap mine in the context of prior splicing tools, are helpful. We appreciate that further efforts to improve usability are beyond the scope of the current tool. We also find that the additional analyses and experiments, including exclusion of putative undiagnosed or presymptomatic patients, repeating TDP-43 immunoblots in different cell types, and exploration of the available neuroblastoma cell line, all improve the rigor of the study. We agree with the decision to move the ferritin measurements and more detailed model to the supplement as additional experiments would be needed to fully test this model, which may be outside the scope of the current study.

We therefore recommend the study for publication. We note two minor things that the authors may want to consider for improving clarity of the figures. Fig 3B and D – perhaps label gray in the figure or legend as not measured (if that is indeed what gray represents). Fig 4F – the control images are too small for the reader to see puncta (or lack of puncta) within the cells, consider including enlarged insets for control as well to improve comparison (or put only insets in the main figure and reserve larger images for supplement).

We thank Reviewer #1 for their comments. For Fig 3b and 3d, the meaning of a gray tile has been included in the figure legend. For Fig 4F, an enlarged inset has been included in the figure. The full original images have been included in the Source Data.

Reviewer #2 (Remarks to the Author):
The authors have paid close attention to my suggestions and have strengthened their manuscript. I have no further suggestions.

We thank Reviewer #2 for their constructive feedback and prior suggestions.

Reviewer #3 (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.

We thank Reviewer #3 for their constructive feedback and prior suggestions.

Reviewer #4 (Remarks to the Author):

The authors addressed most of my concerns in the revision. Therefore, I support its publication in Nature Communications.
I still have a request regarding Figure 4D. The Western blot of TDP-43 for human PBMCs shows substantial decrease upon CPX treatment. However, the bar graph on

the right shows very little reduction. Can the authors check if this apparent discrepancy is an error in quantification?

We thank Reviewer #4 for their comments. The Western blot of TDP-43 for human PBMCs has been re-quantified using a more representative measure of the background. As suggested, the decrease is more substantial than previously measured.

Also regarding the same panel, if biological replicates have been done for the Western blotting analysis, the data should be plotted in the bar graph with error bars indicated.

Biological replicates showed similar results but were not quantified and therefore not included in the plot.