Reviewed Preprint
v1 • April 2, 2026
Not revised

For correspondence:
christoph.d.treiber@gmail.com

Competing interests: No competing interests declared

Funding: See page 17

Reviewing editor: Mohammad M Karimi, King's College London, United Kingdom



# Transposons contribute to splice-isoform diversity in the *Drosophila* brain

**Malak Choucri, Christoph D Treiber**
Department of Biology, University of Oxford, Oxford, United Kingdom

## eLife Assessment

This **valuable** study addresses a timely question regarding the contribution of transposable elements to splice isoform diversity in the Drosophila brain, directly engaging with recent conflicting findings in the field. The work provides **convincing** evidence that TE-gene chimeric transcripts are detectable and that prior discrepancies largely arise from methodological differences in computational pipelines and experimental design. The combination of reanalysis, methodological clarification, and targeted validation represents a technical contribution that will be of interest to researchers studying transcriptome complexity and transposable elements. However, the strength of evidence would be further enhanced by increased methodological transparency, more rigorous experimental controls, and a more cautious interpretation of functional implications.

https://doi.org/10.7554/eLife.110625.1.sa3

## Abstract

The extraordinary complexity of the brain depends in part on the vast diversity of mRNA isoforms it expresses, often in a cell-type specific manner. In a recent study, we found that intronic transposable elements (TEs) are spliced into neural transcripts and diversify the splice isoform repertoire of neurons and glia (Treiber and Waddell, 2020). A recent paper by Azad et al. revisits these findings using their TIDAL analysis pipeline applied to our published data (Azad et al., 2024). Their analysis did not find any of the splicing reads we reported, and although they used RT-PCR to test seven of the 264 TE-gene pairs we had previously reported, they failed to validate TE-gene splicing in any of them. Here, we conduct a quantitative analysis of TE exonisation and show that intronic TE insertions are frequently recruited as alternative exons, with exon usage ranging from rare events to near-complete inclusion in transcripts. We implement this analysis in an improved version of our TEChim software, and present clear support for TE-gene splicing at the seven loci tested by Azad et al. We also identify methodological issues in the experimental and computational design of the Azad et al. study that likely explain their failure to detect TE-gene chimeras, while demonstrating that TE-gene splicing can be detected by RT-PCR under appropriate experimental conditions. Together, our data demonstrates that TE splice isoforms are not rare artefacts but measurable and biologically relevant features of the *Drosophila* brain transcriptome that may contribute to the molecular complexity and functional adaptability of the brain.

## Introduction

Neural transcriptomes are shaped by extensive alternative splicing, which generates a large number of mRNA isoforms that drive the structural and functional diversity within the brain. Among the genomic features that can contribute to this process, transposable elements (TEs) are of particular interest. TEs make up a substantial proportion of eukaryotic genomes and provide a rich and diverse reservoir of potential splice sites. However, their contribution to neural transcript diversity remains poorly understood. In our previous work, we showed that many intronic TEs in

the *Drosophila* genome are spliced into mature transcripts in the brain, giving rise to chimeric mRNAs that incorporate TE-derived sequence (Treiber and Waddell, 2020). By combining high-coverage genomic DNA (gDNA) sequencing (Treiber and Waddell, 2017) with strand-specific RNA sequencing from the same fly strain, we identified 264 TE-gene pairs with clear, breakpoint-spanning read support for splicing into and out of TEs.

TE insertions are not fixed features of genomes but vary extensively between individuals. In *Drosophila*, TE landscapes differ markedly between laboratory strains and even between individuals of an inbred strain (Treiber and Waddell, 2017). Comparable levels of inter-individual TE polymorphisms are well documented in mammals (Akagi et al., 2008), including humans (Bennett et al., 2004; Ewing and Kazazian, 2011). Importantly, recent large-scale transcriptomic and proteogenomic analyses in humans demonstrate that intronic TE insertions are frequently exonised, giving rise to thousands of recurrent, translated splice isoforms with distinct structural and functional properties (Arribas et al., 2024). Hence, variation in TE landscapes has the capacity to generate inter-individual variation of splice isoform repertoires, and potentially protein structure and abundance. Within the nervous system, TE exonisation therefore provides an intriguing mechanism by which genetic variation in TE landscapes could contribute to inter-individual differences in neural function and behaviour.

Here, we introduce a quantitative analysis of TE exonisation using an extended version of the TEChim pipeline, allowing us to measure the extent to which individual intronic TE insertions are used as alternative exons. We show that TE exonisation efficiency varies widely across loci, and is often substantial, and that these values are reproducible across biological replicates. We then examine why a recent re-analysis of the same data by Azad et al. failed to recover these splice junctions, identifying methodological issues that readily account for their negative results. Finally, we revisit our earlier analysis of TE expression modes to place these splice events within the broader context of autonomous vs. non-autonomous TE expression in the brain.

Together, these lines of evidence support a model in which TE-derived splice isoforms are stable and biologically relevant features of the *Drosophila* brain transcriptome, and, in light of growing evidence from other species, part of a conserved strategy by which eukaryotic genomes diversify gene function.

# Results

## Intronic TE insertions show a broad range of exonisation rates

To quantify how frequently individual TE sequences are used as exons, we extended our TEChim pipeline to calculate, for each breakpoint at an endogeneous exon-intron junction, the ratio of TE-derived splice junctions to canonical exon-exon junction (Figure 1A). For each occurrence, TEChim now reports the relative usage of TE-derived splice junctions immediately downstream (for TE-derived splice acceptor sites) or upstream (for donor sites) of the exon-intron junction compared to canonical exon-exon junctions, providing a simple measure of the TE exonisation efficiency for each event. We applied this analysis to our previously published high-coverage RNA sequencing data (Figure 1B, Supplemental table S1). Many of the TE-gene pairs contain more than one chimeric breakpoint, and we therefore obtained splice ratios for 456 individual TE-gene junctions in total.

Across these junctions, splice ratios span almost the full dynamic range from the detection floor to 100%. The median junction has a splice ratio of 3%, and a quarter of junctions exceed 15%, indicating that TE-derived exons contribute substantially to the transcript pool at many sites. Where splice ratios can be estimated, values are highly consistent across biological replicates, suggesting that TE exonisation efficiencies are stable properties of individual TE-gene pairs rather than stochastic noise. Overall, these data show that intronic TE insertions in the *Drosophila* brain are frequently recruited as alternative exons, with usage ranging from rare inclusion events to near-constitutive incorporation at the most strongly exonised loci.

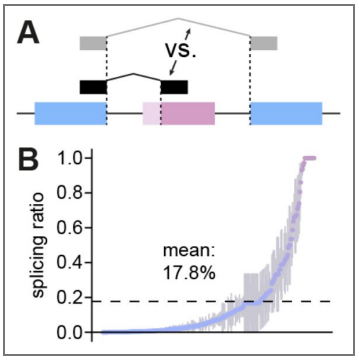



**Figure 1. Exonisation rates at intronic TE insertion sites**

**(A)** Schematic of the approach used to quantify exonisation rates. For each TE-gene breakpoint at an endogeneous exon-intron junction, TEChim computes the ratio of reads supporting TE-derived splice junctions relative to canonical exon-exon junctions. **(B)** Ranked distribution of TE exonisation efficiencies across intronic and exonic TE insertion sites in the Drosophila brain. For each of the 264 TE-gene pairs, the breakpoint with the highest exonisation ratio is shown. Points represent mean exonisation ratios across six biological replicates. Error bars show SEM. The dashed line indicates the mean exonisation rate across all loci. Junction level exonisation ratios, replicate coverage and SEM values are provided in Supplemental Table S1.

## TE-Gene splicing is readily detectable with algorithms that span wider junctions

RNA sequencing is a highly sensitive method to uncover TE-gene chimeric mRNAs (Dobin et al., 2013). Crucially, sequencing reads that span TE-gene breakpoints either exist or they don't: software such as our TEChim approach, or Azad et al.'s TIDAL does not infer or predict these reads, but simply finds them in the data. In our previous study, TEChim found read support for splicing events into- and out of TEs for 264 TE-gene pairs. For each of these loci we also detected the corresponding TE insertion in genomic DNA (gDNA) from the same fly strain. For TE families represented in multiple TE-gene pairs, the TE-side breakpoints repeatedly mapped to the same internal position within the TE consensus, rather than being scattered along the TE sequence.

One clear example is an intronic antisense insertion of the retrotransposon *roo* inside the *mustard (mtd)* gene, which encodes an ecdysone-regulated puff protein. In abCherry flies, TEChim identifies 34 reads upstream and 46 reads downstream of the intron-exon junctions in gDNA, confirming the presence and precise location of the insertion (Figure 2A). In strand-specific RNA-seq data, from six independent biological replicates, TEChim then revealed extensive splicing into- and out of *roo*.

We found 42 reads that spanned the upstream intron-TE breakpoint of *mtd* pre-mRNA in the midbrain and 21 reads that spanned the downstream breakpoint, consistent with unspliced pre-mRNA (Figure 2B). In addition, we detected 49 reads that supported splicing between *mtd* exons and internal positions within *roo*. For example, 15 reads spanned a breakpoint that fell precisely at the end of exon 13 and position 5462 within *roo* (Figure 2C). The antisense sequence leading up to position 5462 resembles the end of an intron and contains a cryptic splice acceptor motif. We also found reads from other up-stream exons to the exact same locus within *roo*. Further junctions link position 5190 in *roo* to the next two downstream exons of *mtd*, supported by 28 and 2 reads respectively. These precise and recurrent patterns across biological replicates make it extremely unlikely that the breakpoint-spanning reads arise from amplification artifacts.

Azad et al. applied TIDAL to our RNAseq data and detected some pre-mRNA breakpoint spanning reads at the *mtd-roo* locus (Figure 2B, red bars), but they failed to recover any reads spanning splicing events between roo and neighbouring exons, and reported similar issues for the other TE-gene pairs they examined (see Figure 2 – figure supplement 1-5). This likely reflects a limitation of their method rather than an absence of splicing. In particular, TIDAL clusters split reads within a narrow window (twice the read length) around candidate junctions, whereas many TE-gene splice junctions fall outside this range. As a result, genuine splice junctions that are readily detected by TEChim fall below TIDAL's detection criteria. Taken together with our quantitative exonisation analysis, these locus-level examples show that TE-gene splice isoforms are clearly present in the data and are readily recovered when algorithms are configured to capture the relevant splice junctions.

## Strain-specific TE landscapes require genotype-matched validation

TE insertions vary between different lab strains, a well-established phenomenon also documented by the Lau lab (Rahman et al., 2015). In our original study, we used abCherry flies, which were the F1 offspring of MB008b males crossed to UAS-mCherry virgin females. Azad et al. attempted to validate the splice events we found in these flies, but used w1118 and OreR flies, which differ substantially in their genotype. They examined 22 of our loci and found that some insertions were conserved, while others were specific to our abCherry flies. Although this result is not surprising given known differences in TE content between strains, Azad et al. nonetheless interpret it as evidence that "…the majority (~ 65%) of the predicted TE insertions using RNAseq as inputs could potentially be false positives". This interpretation is problematic and potentially misleading. In summary, conclusions about the reproducibility of TE insertions are not valid when based on genotypically mismatched strains.

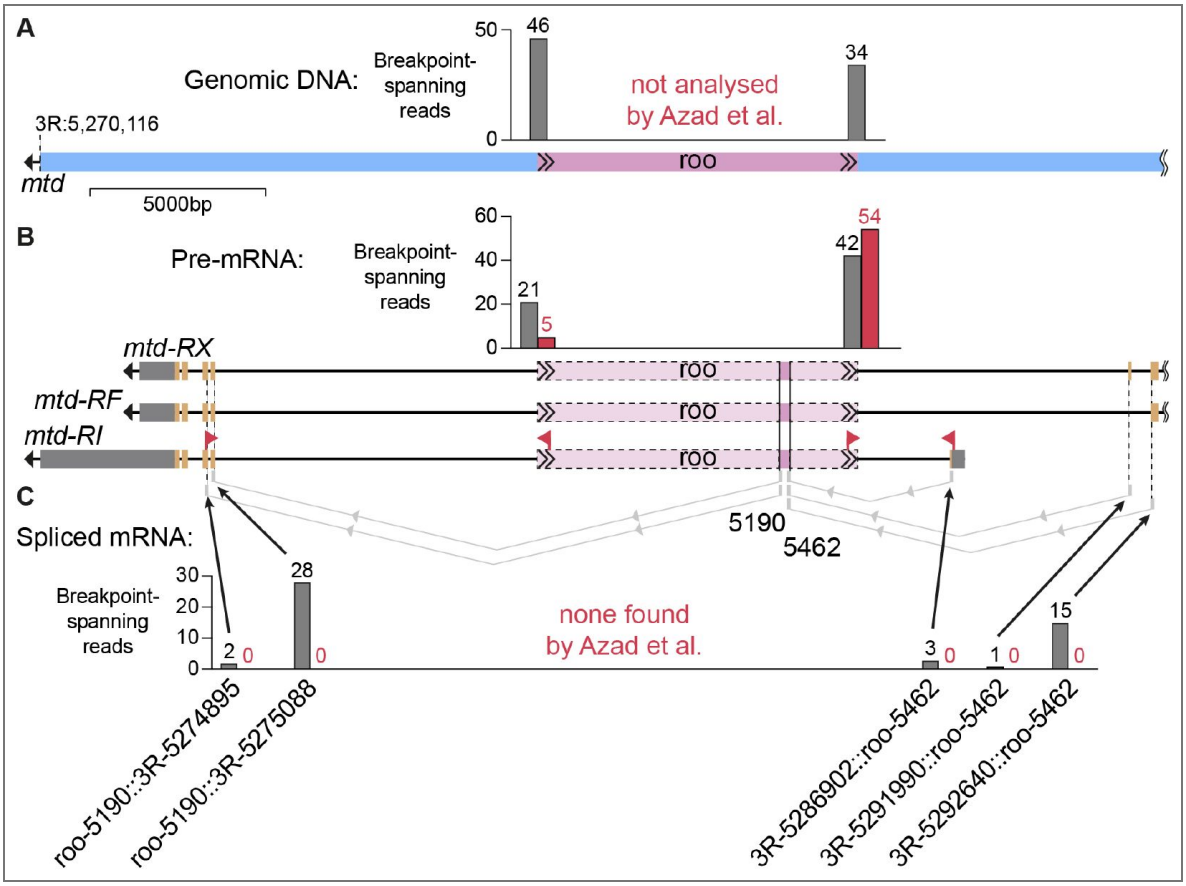



**Figure 2. Antisense insertion of roo in mtd introduces cryptic splice donor- and acceptor sites**

**(A)** Schematic of the genomic locus of the mustard gene, including a roo insertion present in abCherry flies. Bar graphs show the number of breakpoint-spanning reads identified in gDNA-sequencing data from abCherry flies, using TEChim (grey). **(B, C)** Schematic of mustard transcripts -RX, -RF and -RI, with coding sequences (orange) and UTRs (dark grey). The precise splice acceptor- and donor sites on roo are also shown, and sections of roo that are spliced out are shaded in light magenta. Bar graph in B shows the number of breakpoint spanning reads found in mRNA from abCherry midbrains by TEChim (grey) and TIDAL (red). Bar graphs in C show the number of reads that support splicing events (indicated by light grey lines) between up- and downstream exons, and two conserved loci within roo (grey bars). Each bar is labelled with the precise location of the breakpoint on roo and the reference genome. TIDAL did not find any of these reads. In addition, red arrows indicate the precise location of RT-PCR primers used by Azad et al. Note that one primer of each primer pair maps onto a section of the TE that is actually spliced out. Figure is drawn to scale and gDNA and mRNA are aligned to each other.

## Accurate primer design enables robust detection of TE-gene splicing

Strand-specific RNA sequencing enables unbiased, transcriptome-wide detection of splicing events between TEs and genes. Once identified, these events can also be confirmed by a PCR-based approach on cDNA libraries from tissue sample. This requires the careful designing of primers that span both the exon of the gene and the sections of the TE that comprise a given splicing event (Figure 3A).

Azad et al. intended to test a total of nine genomic loci with this method. However, they made a number of methodological errors in their designs and therefore failed to validate any of the tested splicing events. For five of the insertions, *mtd/roo* (Figure 2), *Bx/opus* (Figure 3), *CG17698/tabor* (Figure 2 – figure supplement 1), *Pde1c/I-element* (Figure 2 – figure supplement 2), *Dscam2/blood* (Figure 2 – figure supplement 3), primers were placed within the section of the TE that is spliced out.

In one case, *Dscam2/blood*, additional primers within the spliced region produced bands that the authors interpret as pre-spliced transcripts. However, the band size in the Azad et al. Fig. 5Diii (bottom right panel) is not consistent with this interpretation: an unspliced amplicon would be 2165nt long, whereas a spliced version, as we reported in our study, should be 616nt. The band appears closer to 616nt than 2165nt, although the authors' decision to crop and combine gels complicates this interpretation.

Two loci, *Rh7/micropia* and *CaMKII/pogo*, were tested despite the fact we did not report splicing events, and it is therefore unclear what hypothesis was being evaluated. For *mub/opus* (Figure 1 – figure supplement 4), the tested insertion in w1118 flies is located 31kb away from the insertion we reported in abCherry flies and the forward and reverse genomic PCR primers used by Azad et al. flank the wrong site. For *SelR/mdg3* (Figure 1 – figure supplement 5), splicing occurs with splice isoforms *SelR-RG, -RH and -RJ*, but one of the two primer pairs of Azad et al. was specific to a different splice isoform. Their other primer pair is designed correctly. However, the absence of a band highlights the most important issue with their approach. *Mdg3* is located within the 3'UTR of *SelR*, and this primer pair should produce a PCR fragment even in the absence of splicing. The fact that the authors did not detect a band suggests technical problems with their RT-PCR rather than absence of the underlying transcript.

RT-PCR requires careful optimisation, and while the presence of a band (and its confirmation using Sanger sequencing) provides strong evidence for the existence of a certain genomic locus or splicing event, the absence of a band cannot reliably be taken as evidence for its absence (Bustin et al., 2009).

To address the concerns raised by Azad et al. directly, we genotyped Canton-S flies and found that they were homozygous for the same *opus* insertion in *Beadex* that we had found in abCherry flies. We extracted RNA from two groups of 20 fly brains each, generated cDNA libraries and amplified a fragment consistent with our previous observation, where an upstream exon of *Beadex* is spliced into position 902 of *opus* (Figure 3A, B, Figure 3 – figure supplement 1). We confirmed the band from one of the replicates using Sanger sequencing (Figure 3C), which matched the predicted breakpoints.

In summary, the combination of genotypically mismatched strains and suboptimal primer design is sufficient to explain Azad et al.'s negative RT-PCR results. Consistent with our quantitative exonisation analysis and locus-level RNA-seq evidence, TE-gene chimeric transcripts are readily detected when genotype and primer placement are matched to the documented TE insertion and splice junction.

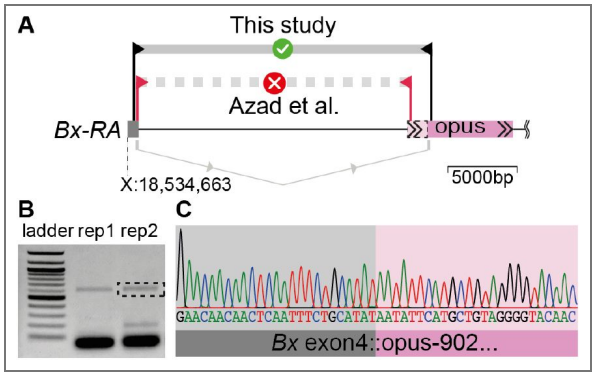



**Figure 3. Accurate primer design generates independent breakpoint-spanning fragments that confirm splicing between Bx and opus**

**(A)** Schematic of the genomic locus of Beadex - splice isoform -RA, drawn to scale. Primer pairs used by Azad et al. are shown in red (at the precise point where they map). The primers used here are drawn in black (forward primer: CCTGATCTCACGGTC TCTGT; reverse primer: CCTTAATGCCTGTCACCACG). The predicted band size for the spliced transcript is 647nt. **(B)** Electrophoresis gel image of PCR product. The ladder is Quick-load 100 bp ladder. The image was edited to remove surface artefacts (e.g. scratches) using global, automated image-processing software with no manual intervention. No bands were added or removed (see the original, unaltered image in Figure 2 – figure supplement 1). Dashed box indicates the fragment that was excised and Sanger sequenced. **(C)** Sanger sequencing result of the band from b. Sequence matches the predicted breakpoint between Beadex exon 4 and position 902 within opus.

### TEs in the brain are predominantly expressed as chimeric transcripts

Our splice-ratio analysis shows that many intronic TE insertions are frequently recruited as altenative exons in neural genes. To place these locus-level observations in a broader context, we next revisit our previous analysis of TE expression modes from our 2020 study (Treiber and Waddell, 2020). In that study, we asked, at the level of TE families, whether TE-derived sequences detected in brain mRNA arise predominantly from autonomous TE expression, or from co-expression with neighbouring genes as chimeric pre-mRNAs and spliced mature mRNAs. Determining the balance between autonomous and non-autonomous TE expression is challenging because TEs are highly repetitive, which complicates the unambiguous assignment of sequencing reads to specific TE copies.

In our 2020 study, we addressed this problem using two complementary approaches. First, we compared the average read depth per nucleotide of individual TE families with the abundance of TE-gene breakpoint-spanning reads for those same families (Figure 4A-C). This analysis indicated that TE expression was most consistent with a non-autonomous origin, arising primarily from co-expression with neighbouring genes. Second, we analysed long terminal repeat (LTR)-containing elements by classifying reads that spanned the LTR and either the internal TE sequence (which can result from both autonomous- and non-autonomous expression), or a flanking genomic region (which can only result from non-autonomous expression). Both approaches supported the conclusion that TEs in the brain are predominantly expressed as part of chimeric mRNAs.

Azad et al. attempted to revisit this question using a different approach. Instead of quantifying the number of reads supporting chimeric vs. non-chimeric transcripts, they compare the rank order of TE families in these two categories. This rank-based comparison lacks a clear biological or statistical interpretation and does not provide a quantitative assessment of TE expression modes. As such, their analysis does not alter the conclusions from our original work that, in the Drosophila brain, TE-derived sequences are largely expressed as components of chimeric transcripts, rather than as products of autonomous TE activity.

## Discussion

Here we show that TE-derived splice isoforms are robust and reproducible features of the *Drosophila* brain transcriptome. By quantifying exonisation rates at 456 TE-gene junctions, we find that intronic TE insertions are frequently recruited as alternative exons, with usage ranging from rare events to near-complete incorporation into mature transcripts at a subset of loci. In-depth analyses of candidate TE-gene pairs, together with RT-PCR validation of the *Bx/opus* junction, reinforce the conclusion that chimeric transcripts are readily detectable with carefully-designed experimental and computational approaches. In combination with our previous family-level analysis of TE expression modes, these results support a model in which most TE-derived sequences in the *Drosophila* brain arise from non-autonomous, gene-linked transcripts, rather than widespread autonomous somatic TE expression.

These observations have important implications for how we think about the architecture of neural genes. Intronic TE insertions do not just add passive sequence, but can act as modular exons that are spliced into mature mRNAs (Galit Lev-Maor et al., 2003). Depending on their position within a transcript, TE-derived segments may alter coding sequences, remove protein domains, truncate transcripts, or change untranslated regions that influence RNA stability, localisation or translational control. Because alternative splicing in the nervous system is strongly cell-type specific, TE-driven splice variants provide a straightforward route for generating cell-type specific versions of the same gene product, with distinct TE-derived variants expressed in different neuronal- or glial populations (Brown et al., 2014; Torres-Méndez et al., 2022). Although the present study does not directly address protein output or physiological function, the exonisation rates we report imply that, for many genes, a substantial fraction of transcripts carries TE-derived sequence. This makes it plausible that TE exonisation contributes to the fine-tuning of neural gene function, rather than representing a negligible by-product of noisy splicing.

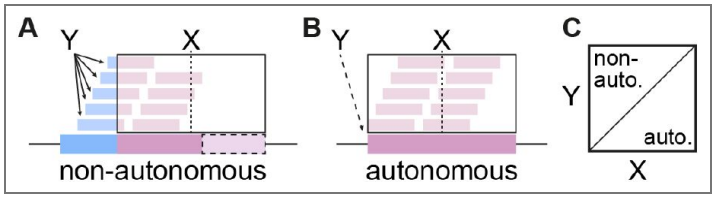



**Figure 4. Schematic highlighting method to assess rate of non-autonomous versus autonomous TE expression**

**(A, B)** The average read depth per nucleotide within a TE (X) is calculated and compared with the number of breakpoint-spanning reads for the same element (Y). **(C)** Primarily non-autonomously expressed TEs sit in the top left half of the correlation plot, while autonomously expressed TEs are on the bottom right side. The data for this analysis is published in Treiber and Waddell, 2020.

Beyond cell-type specific regulation, TE exonisation also provides a simple genetic route to variability in neural transcriptomes between individuals and strains. We and others have shown that TE insertion landscapes differ markedly between *Drosophila* strains, and even between individuals. For example, Azad et al. independently confirmed that several insertions we describe are specific to the abCherry background. In a framework where intronic TEs are frequently exonised, such genetic differences translate directly into inter-individual differences in the splice-isoform repertoire of neural genes. One consequence is that genetically distinct flies may express overlapping, but non-identical TE landscapes, which could contribute to variability in neural circuit properties and behaviour between individuals. In the future it will be important to combine TE-aware transcriptomics with single-cell sequencing to determine how TE-derived isoforms are distributed across neural cell types.

A second set of implications concerns methodology. Our re-analysis of the Azad et al. study, especially of their primer sequences, underscores how algorithmic and experimental choices can mask genuine TE-gene splice isoforms. Algorithms that cluster split reads within narrow genomic windows, or that are not explicitly designed to capture TE-gene architectures, will fail to recover many of the junctions that are visible in the underlying reads. Likewise, RT-PCR validation is only informative when performed in genotype-matched samples using primers that span the reported splice junctions, and when negative results are interpreted in light of assay sensitivity. More generally, TE transcriptomics requires explicit attention to read mapping, breakpoint calling, strain background and primer design. We suggest that the field would benefit from adopting best-practice guidelines similar to those developed for quantitative PCR (Bustin et al., 2009), with clear standards for reporting TE-aware RNA-seq analyses and breakpoint-level validation.

Several limitations of our study point to concrete next steps. First, our quantitative analysis focuses on a curated set of 264 TE-gene pairs, derived from a single *Drosophila* genotype and bulk midbrain RNA-seq data. Many additional TE insertions likely remain below our detection threshold because of limited read depth or cell-type specific expression. Second, splice ratios in bulk RNA-seq data provide a useful summary of exon usage but do not distinguish between different cell types or developmental stages, and they do not directly reveal whether TE-derived isoforms are translated. Addressing these questions will require TE-aware single-cell sequencing, together with targeted proteomics and genetic manipulations that remove or edit individual TE-derived exons. Finally, while our data show that TE-derived isoforms are reproducible and often abundant, their functional consequences for neuronal physiology and behaviour remain to be tested.

In summary, our work strengthens the view that TEs are not merely genomic parasites or sources of noise, but active contributors to neural gene regulation. Intronic TEs in the *Drosophila* brain are frequently exonised, giving rise to stable chimeric transcripts that expand the splice-isoform repertoire of neuronal and glial genes. By clarifying the methodological requirements for detecting these events and by quantifying their prevalence, we provide a framework for future studies to investigate when and how TE-derived isoforms have been co-opted into neural function.

# Materials and Methods

## Fly stocks

Wild-type Canton-S flies (Bloomington stock # 64349) were raised on standard brown cornmeal agar food at room temperature.

## RNA extraction

RNA was extracted using a Promega™ ReliaPrep RNA Tissue Miniprep kit. Midbrains of 10 male and 10 female flies per biological replicate were dissected directly into lysis buffer, and tissue was lysed using a Kinematica AG PT1200 Polytron homogenizer. Samples were processed following the standard protocol, including DNAse1 incubation to reduce genomic DNA contamination. Samples were eluted in 15µl nuclease-free water.

## cDNA library preparation

First-strand cDNA synthesis was performed using a Promega™ GoScript Reverse Transcription Mix Oligo(dT) primer, and 2 x 7µl of total RNA from each sample.

## PCR

cDNA libraries were analysed with a standard PCR protocol, using an APExBT 2X Taq PCR Master Mix and a 1:100 dilution of the stock cDNA libraries.

## Sanger-sequencing

A band with the predicted size was cut out, extracted using a Qiagen QIAquick Gel Extraction Kit, with an elution volume of 30µl. Sanger sequencing was performed by Source Bioscience (Nottingham, UK).

# Supplemental Figures

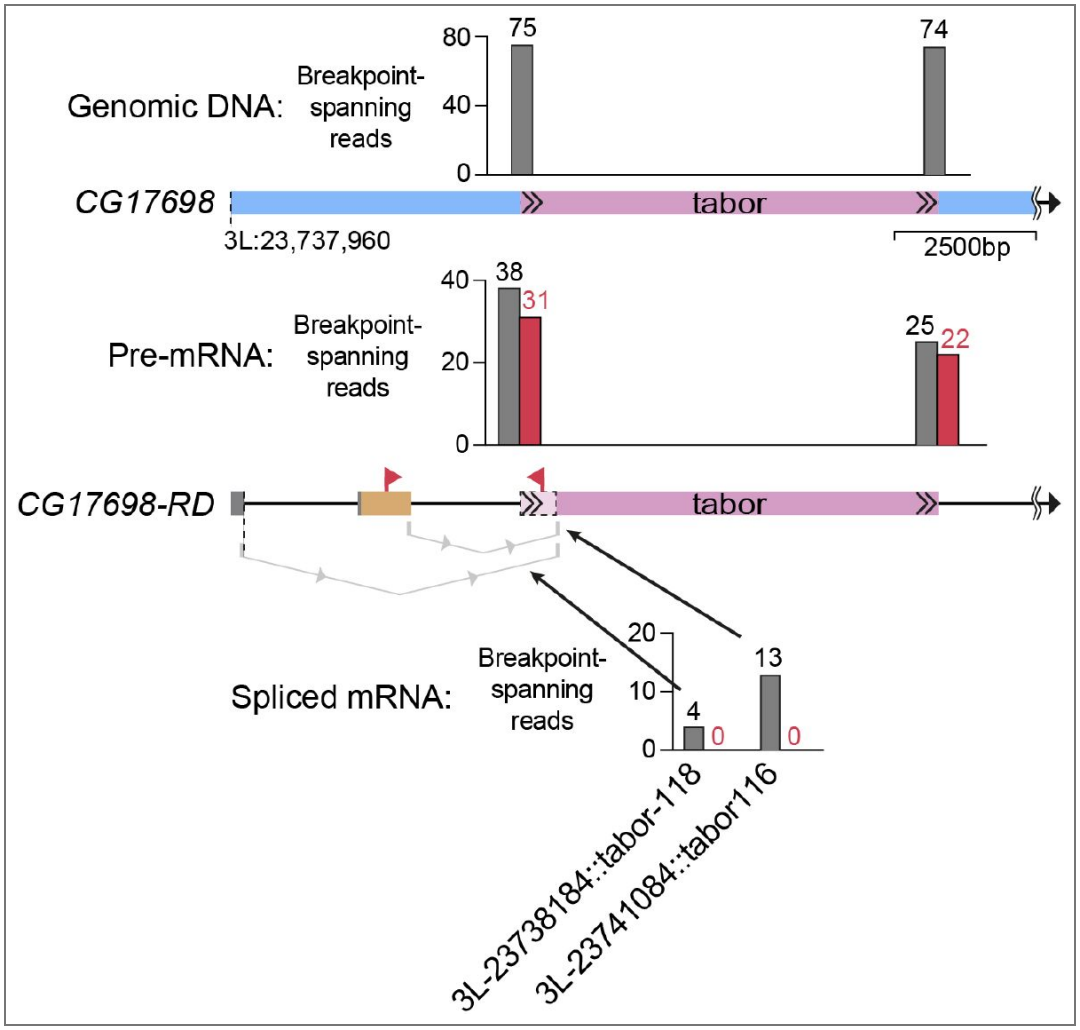



**Figure 1 – figure supplement 1.** ***tabor*** **insertion in** ***CG17698*** Upper panel shows a schematic of the genomic locus of the *CG17698* gene, including a *tabor* insertion present in abCherry flies. Bar graphs at the top show the number of breakpoint-spanning reads identified in gDNA-sequencing data from abCherry flies, using TEChim. Lower panel shows a schematic of the *CG17698-RD* transcript, with coding sequences (orange) and UTRs (dark grey). The splice acceptor site on tabor is also shown, and the section of *tabor* that is spliced out is shaded in light magenta. Bar graphs show the number of breakpoint spanning reads found in mRNA from abCherry midbrains using TEChim (grey) and TIDAL (red). Bar graphs at the bottom show the number of reads that support splicing events

(indicated by light grey lines) between two downstream exons and a conserved locus within tabor (grey bars). Each bar is labelled with the precise location of the breakpoint on tabor and the reference genome. TIDAL did not find any of these reads. In addition, red arrows show precise location of RT-PCR primers used by Azad et al. Note that one primer maps onto a section of the TE that is actually spliced out. Figure is drawn to scale and gDNA and mRNA are aligned to each other.

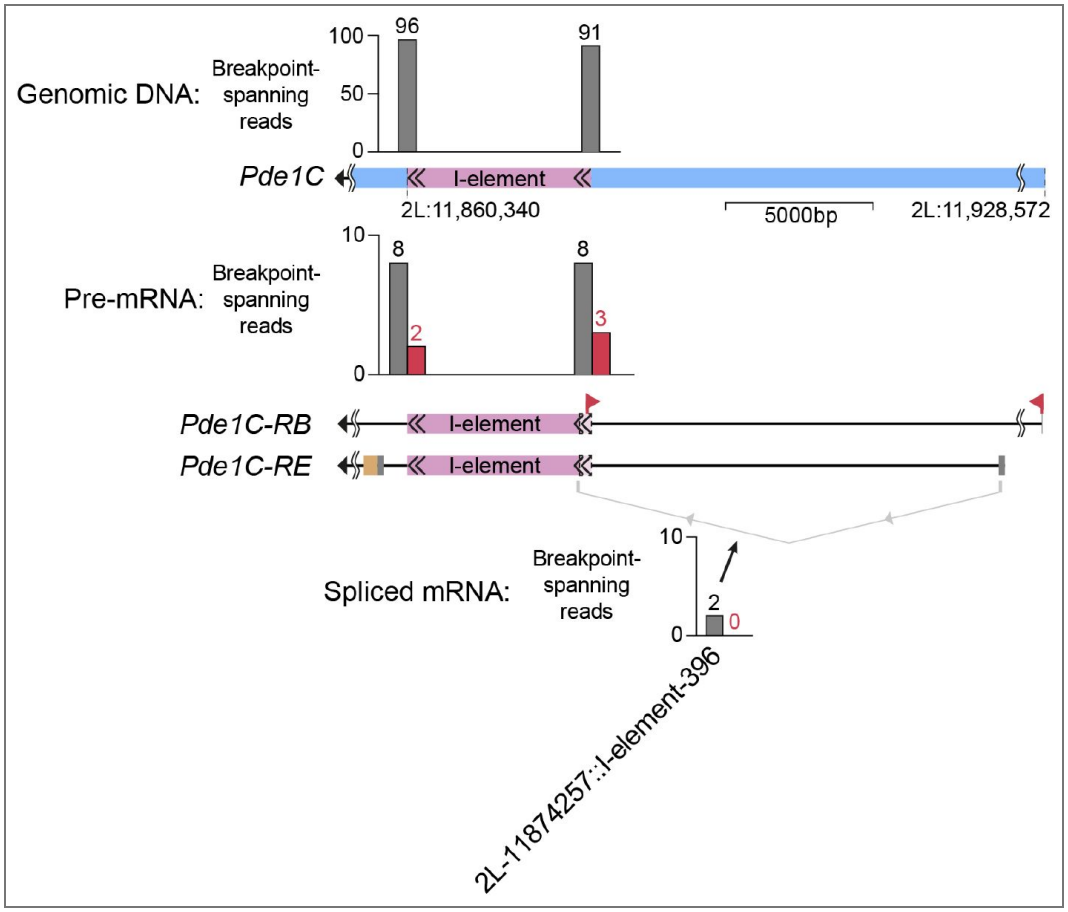



**Figure 1 – figure supplement 2.** ***I-element*** **insertion in** ***Pde1C***

Upper panel shows a schematic of the genomic locus of the *Pde1C* gene, including an *I-element* insertion present in abCherry flies. Bar graphs at the top show the number of breakpoint-spanning reads identified in gDNA-sequencing data from abCherry flies, using TEChim. Lower panel shows a schematic of the Pde1C-RB and -RD transcripts, with coding sequences (orange) and UTRs (dark grey). The splice acceptor site on the *I-element* is also shown, and the section of the *I-element* that is spliced out is shaded in light magenta. Bar graphs show the number of breakpoint spanning reads found in mRNA from abCherry midbrains using TEChim (grey) and TIDAL (red). Bar graph at the bottom shows the number of reads that support splicing events (indicated by light grey line) between the upstream exon and a conserved locus within the *I*-element (grey bar). The bar is labelled with the precise location of the breakpoint on tabor and the reference genome. TIDAL did not find any of these reads. In addition, red arrows show precise location of RT-PCR primers used by Azad et al. Note that one primer maps onto a section of the TE that is actually spliced out and the other primer to an exon of another splice isoform than reported in abCherry flies. Figure is drawn to scale and gDNA and mRNA are aligned to each other.

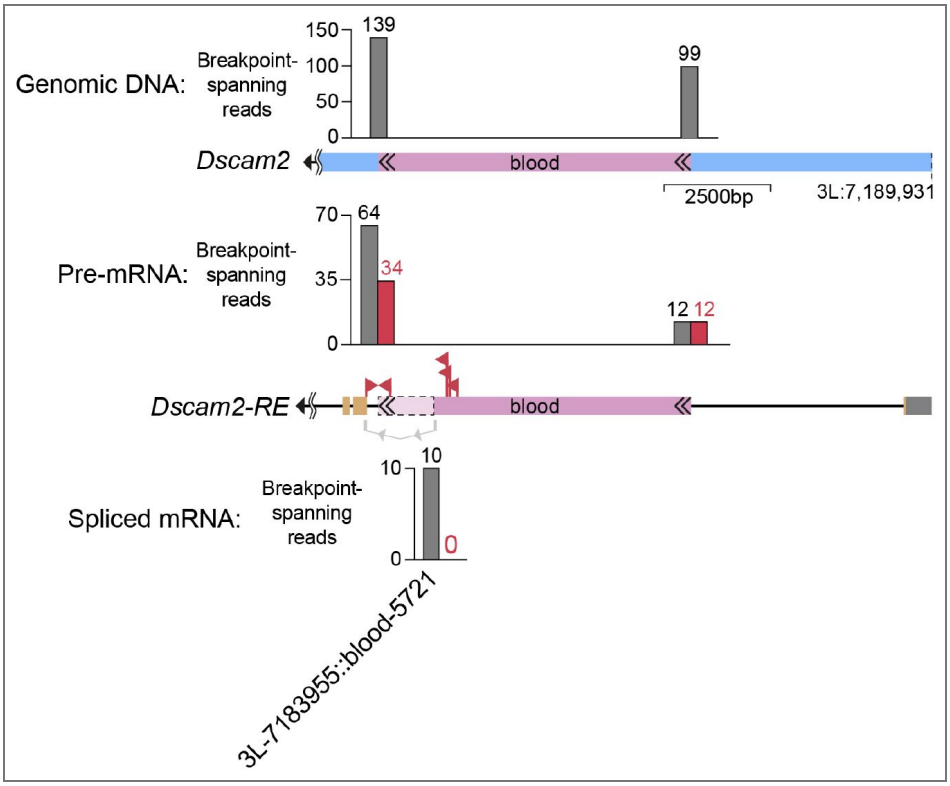



**Figure 1 – figure supplement 3. *blood* insertion in *Dscam2***

Upper panel shows a schematic of the genomic locus of the *Dscam2* gene, including a *blood* insertion present in abCherry flies. Bar graphs at the top show the number of breakpoint-spanning reads identified in gDNA-sequencing data from abCherry flies, using TEChim. Lower panel shows a schematic of the *Dscam2-RE* transcript, with coding sequences (orange) and UTRs (dark grey). The splice donor site on *blood* is also shown, and the section of *blood* that is spliced out is shaded in light magenta. Bar graphs show the number of breakpoint spanning reads found in mRNA from abCherry midbrains using TEChim (grey) and TIDAL (red). Bar graph at the bottom shows the number of reads that support splicing events (indicated by light grey line) between the downstream exons and a conserved locus within *blood* (grey bar). The bar is labelled with the precise location of the breakpoint on *blood* and the reference genome. TIDAL did not find any of these reads. In addition, red arrows show precise location of RT-PCR primers used by Azad et al. Note that one primer maps onto a section of the TE that is actually spliced out. Figure is drawn to scale and gDNA and mRNA are aligned to each other.

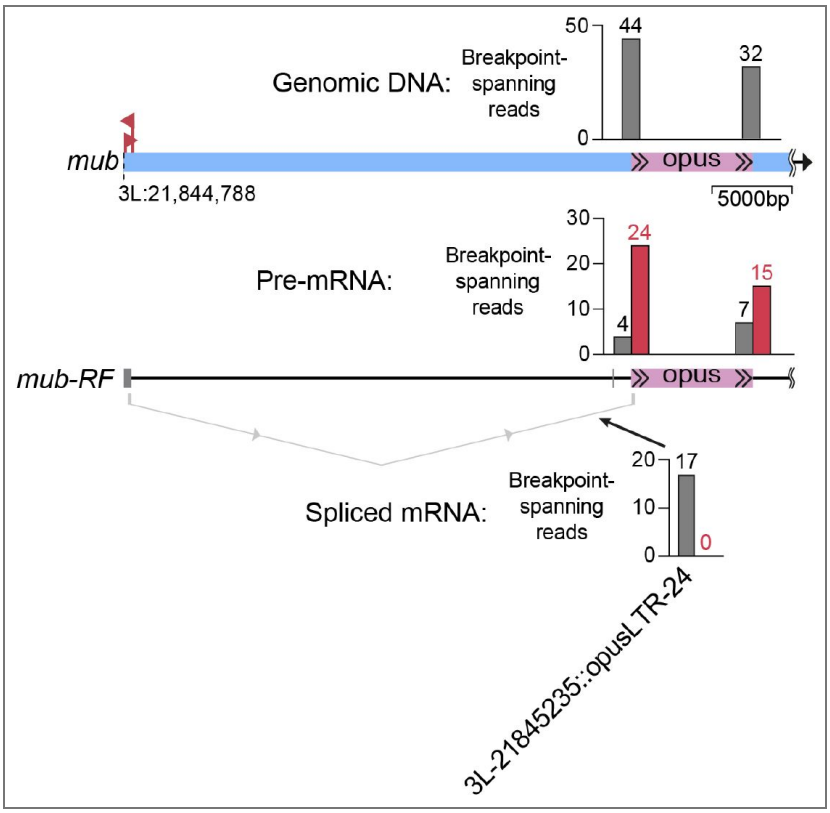



**Figure 1 – figure supplement 4. *opus* insertion in *mub***

Upper panel shows a schematic of the genomic locus of the *mub* gene, including an *opus* insertion present in abCherry flies. Bar graphs at the top show the number of breakpoint-spanning reads identified in gDNA-sequencing data from abCherry flies, using TEChim. Lower panel shows a schematic of the *mub-RF* transcript, with UTRs (dark grey). The splice donor site on *opus* is also shown. The first 24 nucleotides of *opus* are spliced out, but this cannot be visualised at the current scale. Bar graphs show the number of breakpoint spanning reads found in mRNA from abCherry midbrains using TEChim (grey) and TIDAL (red). Bar graph at the bottom shows the number of reads that support splicing events (indicated by light grey line) between the downstream exons and a conserved locus within *opus* (grey bar). The bar is labelled with the precise location of the breakpoint on *opus* and the reference genome. TIDAL did not find any of these reads. In addition, red arrows show precise location of PCR primers used by Azad et al. to genotype their w1118 flies. Note that these primers flank a region ~31k nucleotides upstream of the *opus* insertion in abCherry flies. Figure is drawn to scale and gDNA and mRNA are aligned to each other.

**Figure 1 – figure supplement 5. *mdg3* insertion in *SelR***

Upper panel shows a schematic of the genomic locus of the *SelR* gene, including an *mdg3* insertion present in abCherry flies. Bar graphs at the top show the number of breakpoint-spanning reads identified in gDNA-sequencing data from abCherry flies, using TEChim. Lower panel shows a schematic of the *SelR-RE and -RJ* transcripts, with coding sequences (orange) and UTRs (dark grey). The splice acceptor site on *mdg3* is also shown, and the section of *mdg3* that is spliced out is shaded in light magenta. Bar graphs show the number of breakpoint spanning reads found in mRNA from abCherry midbrains using TEChim (grey) and TIDAL (red). Bar graph at the bottom shows the number of reads that support splicing events (indicated by light grey line) between the upstream exon and a conserved locus within *mdg3* (grey bar). The bar is labelled with the precise location of the breakpoint on *mdg3* and the reference genome. TIDAL did not find any of these reads. In addition, red arrows show precise location of RT-PCR primers used by Azad et al. Note that one primer maps onto a section of another splice isoform then reported in abCherry flies. Figure is drawn to scale and gDNA and mRNA are aligned to each other.

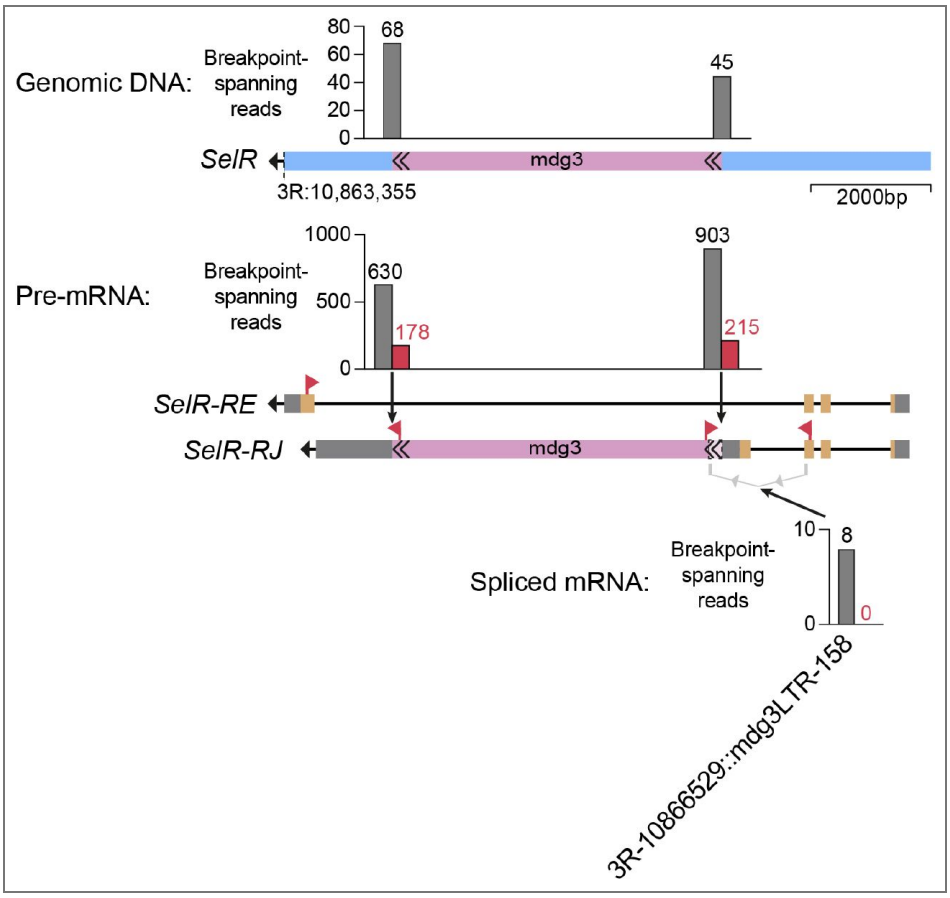
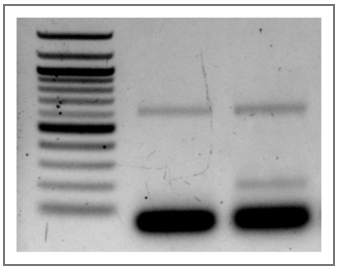



**Figure 2 – figure supplement 1. Original electrophoresis gel image**



# Data availability

Data has been made publicly available as part of a previous study (Treiber and Waddell, 2020).

# Additional information

## Declaration of generative AI and AI-assisted technologies in the writing process

Generative AI (ChatGPT, OpenAI) was used during the drafting process to help refine sentence structure, improve clarity, and explore alternative phrasings. All content, data interpretation, and scientific arguments were conceived, written, and verified by the authors.

## Funding

| Funder | Grant reference number | Author |
|---|---|---|
| UK Research and Innovation (UKRI) | EP/X038882/1 'InnovaTEbehaviour' | Malak Choucri<br>Christoph Daniel Treiber |

## Author ORCID iDs

**Christoph D Treiber**: https://orcid.org/0000-0002-6994-091X

# Additional files

***Supplemental table S1***

# References

**Akagi K,** Li J, Stephens RM, Volfovsky N, Symer DE (2008) Extensive variation between inbred mouse strains due to endogenous L1 retrotransposition. *Genome Research* **18**:869-880 https://doi.org/10.1101/gr.075770.107 | PubMed

**Arribas YA,** Baudon B, Rotival M, Suárez G, Bonté PE, Casas V, Roubert A, Klein P, Bonnin E, Mchich B, *et al.* (2024) Transposable element exonization generates a reservoir of evolving and functional protein isoforms. *Cell* **187**:7603-7620.e22. https://doi.org/10.1016/j.cell.2024.11.011 | PubMed

**Azad MF,** Tong T, Lau NC (2024) Transposable Element (TE) insertion predictions from RNAseq inputs and TE impact on RNA splicing and gene expression in Drosophila brain transcriptomes. *Mobile DNA* **15** https://doi.org/10.1186/s13100-024-00330-z

**Bennett EA,** Coleman LE, Tsui C, Pittard WS, Devine SE (2004) Natural genetic variation caused by transposable elements in humans. *Genetics* **168**:933-951 https://doi.org/10.1534/genetics.104.031757 | PubMed

**Brown JB,** Boley N, Eisman R, May GE, Stoiber MH, Duff MO, Booth BW, Wen J, Park S, Suzuki AM, *et al.* (2014) Diversity and dynamics of the Drosophila transcriptome. *Nature* **512**:393-399 https://doi.org/10.1038/nature12962 | PubMed

**Bustin SA,** Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, *et al.* (2009) The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* **55**:611-622 https://doi.org/10.1373/clinchem.2008.112797 | PubMed

**Dobin A,** Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15-21 https://doi.org/10.1093/bioinformatics/bts635 | PubMed

**Ewing AD,** Kazazian HH (2011) Whole-genome resequencing allows detection of many rare LINE-1 insertion alleles in humans. *Genome Research* **21**:985-990 https://doi.org/10.1101/gr.114777.110 | PubMed

**Lev-Maor Galit,** Sorek Rotem, Shomron N, Ast G. (2003) The Birth of an Alternatively Spliced Exon: 3' Splice-Site Selection in Alu Exons. *Science* **300**:1288-1291

**Rahman R,** Chirn GW, Kanodia A, Sytnikova YA, Brembs B, Bergman CM, Lau NC (2015) Unique transposon landscapes are pervasive across Drosophila melanogaster genomes. *Nucleic Acids Research* **43**:10655-10672 https://doi.org/10.1093/nar/gkv1193 | PubMed

**Torres-Méndez A,** Pop S, Bonnal S, Almudi I, Avola A, V Roberts RJ, Paolantoni C, Alcaina-Caro A, Martín-Anduaga A, Haussmann IU, *et al.* (2022) Parallel evolution of a splicing program controlling neuronal excitability in flies and mammals. *Sci. Adv*

**Treiber CD,** Waddell S. (2020) Transposon expression in the Drosophila brain is driven by neighboring genes and diversifies the neural transcriptome. *Genome Research* **30**:1559-1569 https://doi.org/10.1101/gr.259200.119 | PubMed

**Treiber CD,** Waddell S. (2017) Resolving the prevalence of somatic transposition in Drosophila. *eLife* **6**:e28297 https://doi.org/10.7554/eLife.28297 | PubMed

**Treiber CD,** Waddell S (2020) Transposon expression in the Drosophila brain is driven by neighboring genes and diversifies the neural transcriptome. NCBI BioProject. ID PRJNA588978 https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA588978

# Peer reviews

## Reviewer #1 (Public review):

Summary:

Choucri and Treiber have reassessed their previous study on TE-gene chimeric transcripts in neural genes in response to Azad et al (2024). Azad and colleagues argued that, contrary to Choucri and Treiber's findings, chimeric TE-mRNAs are relatively infrequent, and they cautioned that further optimization of bioinformatics pipelines is needed to detect TE insertions from RNAseq accurately. In this short response, Choucri and Treiber clearly demonstrate that differences in the tools used between their study and that of Azad et al. likely account for the contrasting results, along with RT-PCR failure in designing primers that would match the chimeric transcript, and the use of different Drosophila lines. The authors emphasize the need for uniform, standardized criteria in such analysis, which would ultimately strengthen and advance the field.

Strengths:

The addition of a ratio to compute the number of splice reads specific to the chimeric transcript and compare to the exon-exon splice reads is really interesting because it opens the door to finally quantify the contribution of chimeric TEs to the overall gene expression, although this is not the scope of the present article. The clear dissection of chimeric transcripts, along with the results from Azad et al, allows us to understand the differences between the two studies confidently. Finally, the discussion on Drosophila lines is indeed essential, given that the lines and even individuals have high TE polymorphism.

Weaknesses:

I think it is necessary to add more detail to this article, for instance, the differences between TEchim and Tidal could be laid out more precisely. Regarding the roo example, one of the caveats of this family, along with others, is the presence of simple repeats. It would be important to show that the simple repeats are not interfering with the read mapping. Regarding the experiments, if we are looking for a standardized protocol, then we should

have a detailed material and methods section, with every experiment, replicate, and PCR temperature clearly defined. Finally, and in my opinion, more importantly, the use of RT negative controls on the RT PCRs, along with DNA PCRs to show insertion presence, is mandatory for testing the presence of chimeric genes. Of course, water negative PCR controls are also needed, and unfortunately, absent from Figure 3.

https://doi.org/10.7554/eLife.110625.1.sa2

## Reviewer #2 (Public review):

Summary:

This study by Choucri and Treiber aims to directly address a recent critique regarding the role of transposable elements (TEs) in diversifying the neural transcriptome of Drosophila. The authors seek to demonstrate that TEs are not merely genomic "noise" but are frequently and reliably "exonized" into brain-specific mRNA. By introducing an upgraded computational pipeline, TEChim, and conducting precise experimental validations, the authors set out to show that TE-mediated splicing represents a genuine biological phenomenon that expands the molecular repertoire of the nervous system.

Strengths:

The study's primary strength lies in its rigorous technical "forensic" analysis of previous failed replication attempts. The authors convincingly demonstrate that the lack of signal in the opposing study stemmed from a fundamental methodological mismatch: the software used by the critics (TIDAL) is logically incapable of detecting splice sites located within TE sequences. Importantly, the authors complement this computational clarification with definitive experimental evidence through an effective "experimental rescue." By employing correctly designed primers and matching the genetic backgrounds of the fly strains, thereby accounting for genomic polymorphisms, they successfully validated all seven loci that were previously reported as undetectable. This dual-pronged strategy, addressing both algorithmic bias and experimental design, establishes a more robust technical benchmark for the detection and validation of TE-derived exons in neural tissues.

Weaknesses:

While the technical rebuttal is highly convincing, the scope of the study remains primarily defensive. As a response to a prior critique, the work focuses on establishing the existence and detectability of chimeric TE-derived transcripts rather than exploring their broader functional consequences. As a result, there is limited new insight into how these TE-modified isoforms influence neural circuit function or organismal behavior. In addition, the detection and validation of these events remain technically demanding, requiring deep sequencing and specialized bioinformatic expertise, which may limit broader adoption by laboratories without dedicated computational resources.

https://doi.org/10.7554/eLife.110625.1.sa1

## Reviewer #3 (Public review):

Summary:

This manuscript by Choucri and Treiber responds to a recent paper by Azad et al., which responds to a paper by Treiber and Wadell (Genome Research, 2020). The controversy relates to the detection of transcripts with transposable elements (TEs) spliced into them in the Drosophila brain.

Strengths:

The authors now argue convincingly that these transcripts exist using an improved, updated version of their pipeline. They also validate some of their findings using RT-PCR and explain why Azad et al. failed to detect these transcripts due to methodological errors. Overall, I am convinced that these transcripts exist and that the TE-derived transcripts described by Choucri and Treiber are real.

Weaknesses:

The authors should mention that combining PCR-amplified cDNA generation with short-read sequencing is suboptimal for detecting TE-fusion transcripts. Recently, direct long-read ONT RNA sequencing, which does not require amplification and spans the entire transcript, has been used to detect similar transcripts in human stem cells and the human brain (PMID: 40848716 & Garza et al, BioRxiv). Had the authors used this technology to validate their findings, there would be no question about these transcripts. If not doing such experiments, then they should at least discuss the possibility and the advantage of the approach.

https://doi.org/10.7554/eLife.110625.1.sa0