

# Deep Transcriptomics Reveals Cell-Specific Isoforms of Pan-Neuronal Genes

Zachery Wolfe<sup>1</sup>, David Liska<sup>2</sup>, Adam Norris<sup>3</sup>

1. Department of Biological Sciences, Southern Methodist University, Dallas, TX 75205, United States

2. Office of Information Technology, Southern Methodist University, Dallas, TX 75205, United States

3. Department of Biochemistry, University of California, Riverside, 3401 Watkins Drive, Boyce Hall, Riverside, CA, 92521, United States.

Correspondence: [adamn@ucr.edu](mailto:adamn@ucr.edu)

## ABSTRACT

Profiling gene expression in single neurons using single-cell RNA-Seq is a powerful method for understanding the molecular diversity of the nervous system. Profiling alternative splicing in single neurons using these methods is more challenging, however, due to low capture efficiency and sensitivity. As a result, we know much less about splicing patterns and regulation across neurons than we do about gene expression. Here we leverage unique attributes of the *C. elegans* nervous system to investigate deep cell-specific transcriptomes complete with biological replicates generated by the CeNGEN consortium, enabling high-confidence assessment of splicing across neuron types even for lowly-expressed genes. Global splicing maps reveal several striking observations, including pan-neuronal genes that harbor cell-specific splice variants, abundant differential intron retention across neuron types, and a single neuron highly enriched for upstream alternative 3' splice sites. We develop an algorithm to identify unique cell-specific expression patterns and use it to discover both cell-specific isoforms and potential regulatory RNA binding proteins that establish these isoforms. Genetic interrogation of these RNA binding proteins *in vivo* identifies three distinct regulatory factors employed to establish unique splicing patterns in a single neuron. Finally, we develop a user-friendly platform for spatial transcriptomic visualization of these splicing patterns with single-neuron resolution.

## INTRODUCTION

A neuron's development and function are controlled both by the genes it expresses and the isoforms of those genes it selects. Much recent effort has gone into determining the transcriptional states of individual neurons, but analyzing co-transcriptional and post-transcriptional regulatory events such as alternative splicing is a more challenging task<sup>1-5</sup>. The inherent low capture efficiency and sensitivity of single-cell sequencing approaches causes difficulty in accurately calling alternative splicing, especially for lowly- or moderately-expressed genes<sup>4,6,7</sup>. In addition, the molecular mechanisms by which neurons establish cell-specific isoforms are difficult to identify using current single-cell transcriptomic approaches in isolation.

Although difficult to assess with single-cell resolution, alternative splicing can have profound impacts on neuronal development and function<sup>8,9</sup>. Even relatively small alternatively-spliced exons can have major impacts

on neuronal function, for example conferring switch-like gene regulation<sup>10,11</sup> or generating complex functional diversity from a single gene<sup>12-14</sup>. Regulation of alternative splicing is thus critical for the function of individual neurons, but identifying alternative splicing and defining the underlying regulatory mechanisms with single-cell resolution remain major challenges.

We set out to address these challenges using deep transcriptomes from individual neuron types across the *C. elegans* nervous system. Because the developmental trajectory of the *C. elegans* nervous system is invariant across individuals, thousands of the same neuron can be isolated from thousands of isogenic individuals. The resulting deep single-neuron transcriptomes, complete with multiple biological replicates, allow robust analysis of RNA splicing with uniform coverage throughout the gene body and the ability to assay even lowly expressed genes. Obtaining single-cell splicing patterns in a powerful genetic system like *C. elegans* also allows the next step to be taken: identifying the regulatory mechanisms by which single-cell isoforms are established.

Here we identify neuronal alternative splicing with cell-specific resolution using deep transcriptomes generated by the CeNGEN consortium<sup>15,16</sup>. These global splicing maps reveal several striking observations, including (1) genes that are expressed pan-neuronally but harbor highly cell-specific splice variants, (2) an abundance of differential intron retention across neuronal cell types, and (3) a single neuron highly enriched for selecting upstream alternative 3' splice sites. We develop an algorithm to identify unique cell-specific expression, and use it to identify both cell-specific isoforms and potential regulatory RNA binding proteins that establish these isoforms. We genetically interrogate these RNA binding proteins *in vivo* and identify three distinct regulatory factors employed to establish unique splicing patterns in a single neuron. Finally, we develop a user-friendly platform for spatially visualizing these splicing patterns with cell-specific resolution.

## RESULTS

### Deep transcriptomes reveal cell-specific splicing patterns

To dissect global splicing choices at single-cell level with high confidence, we took advantage of a valuable new resource, the deep cell-specific transcriptomes generated by the CeNGEN consortium<sup>15,16</sup>. The consortium genetically labeled individual *C. elegans* neurons with fluorescent reporters, grew many thousands of isogenic animals, then dissociated and sorted a population of thousands of labeled neurons which were used to generate RNA-Seq libraries (Figure 1A).

The resulting deep-sequenced libraries thus represent relatively large amounts of starting RNA sampled from thousands of genetically-identical copies of a given neuronal cell type. Deep sampling permits analysis of genes that are not highly expressed, uniform transcript coverage enables splicing analysis across the entire gene body, and independent biological replicates increases statistical power (Figure 1A). The consortium has generated deep-sequenced data with multiple biological replicates for 46 of the 118 anatomically-distinct neuron types in *C. elegans*.

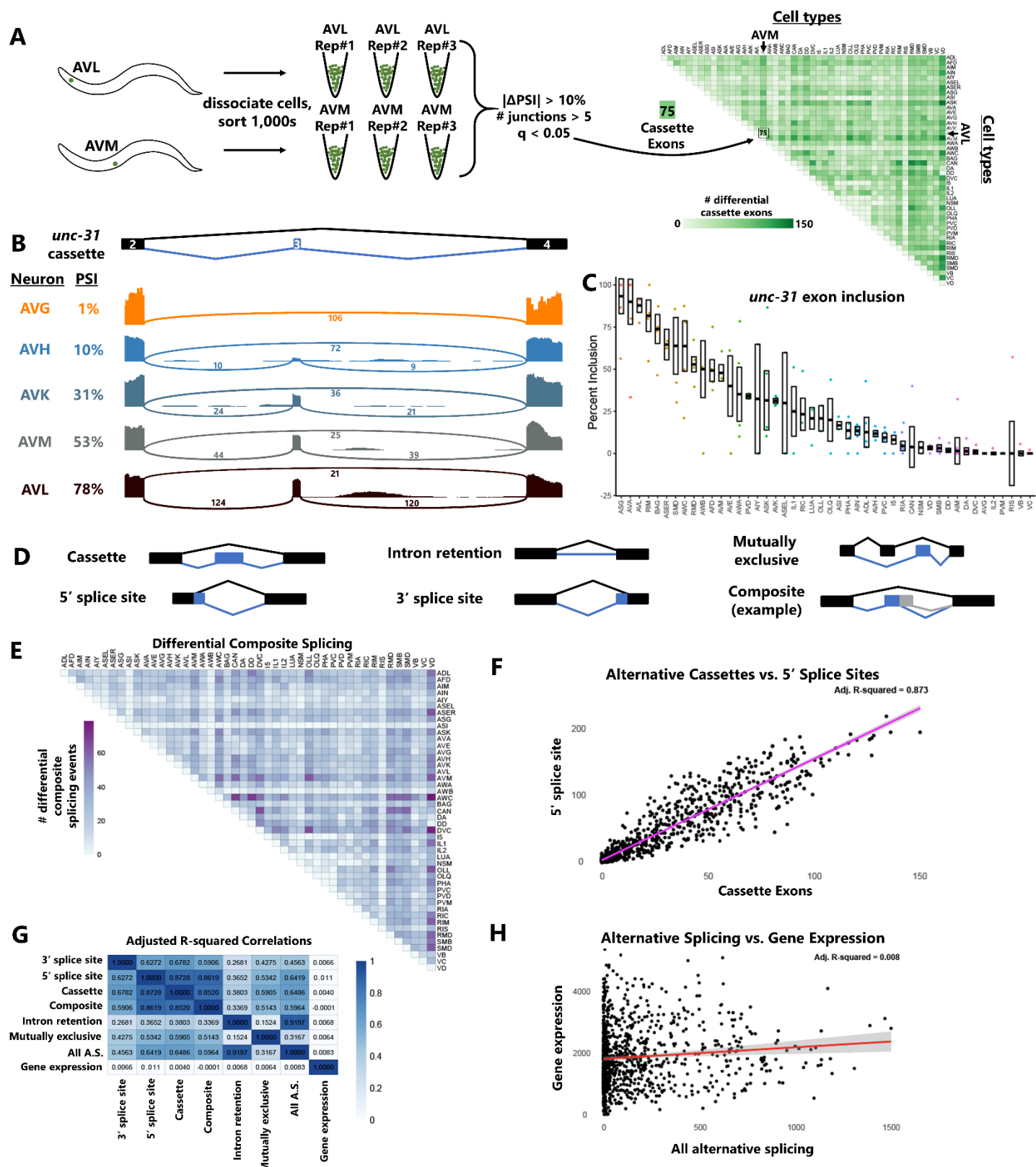
Using this dataset, we performed pairwise comparisons of gene expression (using DESeq2<sup>17</sup>) and alternative splicing (using JUM<sup>18</sup>) between all possible pairs of neuronal cell types. In the example in Figure 1A we compare alternative cassette exons (exon skipping) between the AVM mechanosensory neuron and the AVL multimodal neuron. Using thresholds of  $|\Delta\text{PSI}|$  (change in Percent Spliced In)  $> 10\%$  and FDR-corrected q-value  $< 0.05$ , we identified 75 differential cassette exons. Results for all 2,070 pairwise comparisons are shown for cassette exons in Figure 1A.

Such a systematic global accounting of single-neuron splicing choices enables both gene-centric and cell-centric analysis of alternative splicing. An example of gene-centric analysis is displayed in Figure 1B-C, focusing on exon 3 of the *unc-31/CADPS* gene, which is an essential conserved mediator of neuropeptide release<sup>19</sup>. Splicing levels of this cassette exon vary widely across neuron types, from mostly included (AVL, 78% included) to half included (AVM, 53%) to completely skipped (AVG, 1%). Figure 1C reveals a nearly continuous gradation of splicing levels of this exon across all neuronal cell types. This figure also highlights the power of biological replicates in distinguishing cell-specific measurements that are highly reproducible (*e.g.* AVL) versus those that are highly variable (*e.g.* AIY)

Cell-centric analysis of alternative splicing is displayed in Figure 1A, 1E, and Supplemental Figure S1. JUM is able to interrogate many modes of differential alternative splicing, as shown in Figure 1D, including technically challenging categories that are often ignored, such as "composite" events composed of multiple intertwined splicing choices. Comparisons of differential composite splicing across all cell-type comparisons are displayed in Figure 1E, and data across all cell types for each class of alternative splicing is displayed in Supplemental Figure S1 and Spreadsheet S1.

We asked whether differential cell-type expression of various alternative splicing events are correlated with each other. For example, if AVL-AVM comparison reveals many differential alternative cassette exons, are there likewise many differential alternative 5' splice sites? Indeed, such a comparison across all cell types reveals a strong correlation (Figure 1F). Similar comparisons across all splicing types reveals consistently high correlations (Figure 1G), indicating that splicing differences across cell types are concerted across types of splicing, rather than certain cell types bearing unique splice-type signatures.

In contrast, differences in gene expression across cell types correlate poorly with various types of differential alternative splicing (Figure 1G-H), indicating that gene expression and alternative splicing patterns are globally distinct and are regulated orthogonally across neuronal cell types. Taken together, this deep neuron-specific transcriptomic data provides a powerful resource for interrogating alternative splicing and gene expression across individual neuron types.



**Figure 1: Deep transcriptomes reveal cell-specific alternative splicing patterns.** (A) Schematic of the analysis pipeline. Deep transcriptomes with biological replicates were obtained by the CeNGEN consortium by sorting large isogenic populations of dissociated worms harboring neuron-type-specific fluorescent transgenes. We then used the RNA-Seq data to call differential splicing across all possible pairwise combinations of neuron types. As an example, AVM vs AVL comparison results in 75 differentially-spliced cassette exons according to our  $\Delta$ PSI (10%) and q-value (0.05) cutoffs. (B) Sashimi plots for a cassette exon (exon #3) in *unc-31*, illustrating gradations of inclusion levels across neuron types. Plots are from single biological replicates; PSI values are means across replicates. (C) Median +/- SEM of PSI values across all neuron types, including each biological replicate value, distinguish highly replicable

observations (e.g. AVL) versus variable observations (e.g. AIY). (D) Illustration of the variety of alternative splicing types accommodated by the JUM analysis pipeline. (E) Widespread differential composite splicing, a type of alternative splicing that is historically difficult to analyze. (F-G) Correlation in the magnitude of different gene regulatory events. Each point represents a single cell vs. cell comparison (e.g. AVM vs AVL). (H) Summary of correlation values (adjusted R-square) revealing correlation across different types of splicing, but little correlation between gene expression and alternative splicing.

## Widespread differential intron retention across neuronal cell types

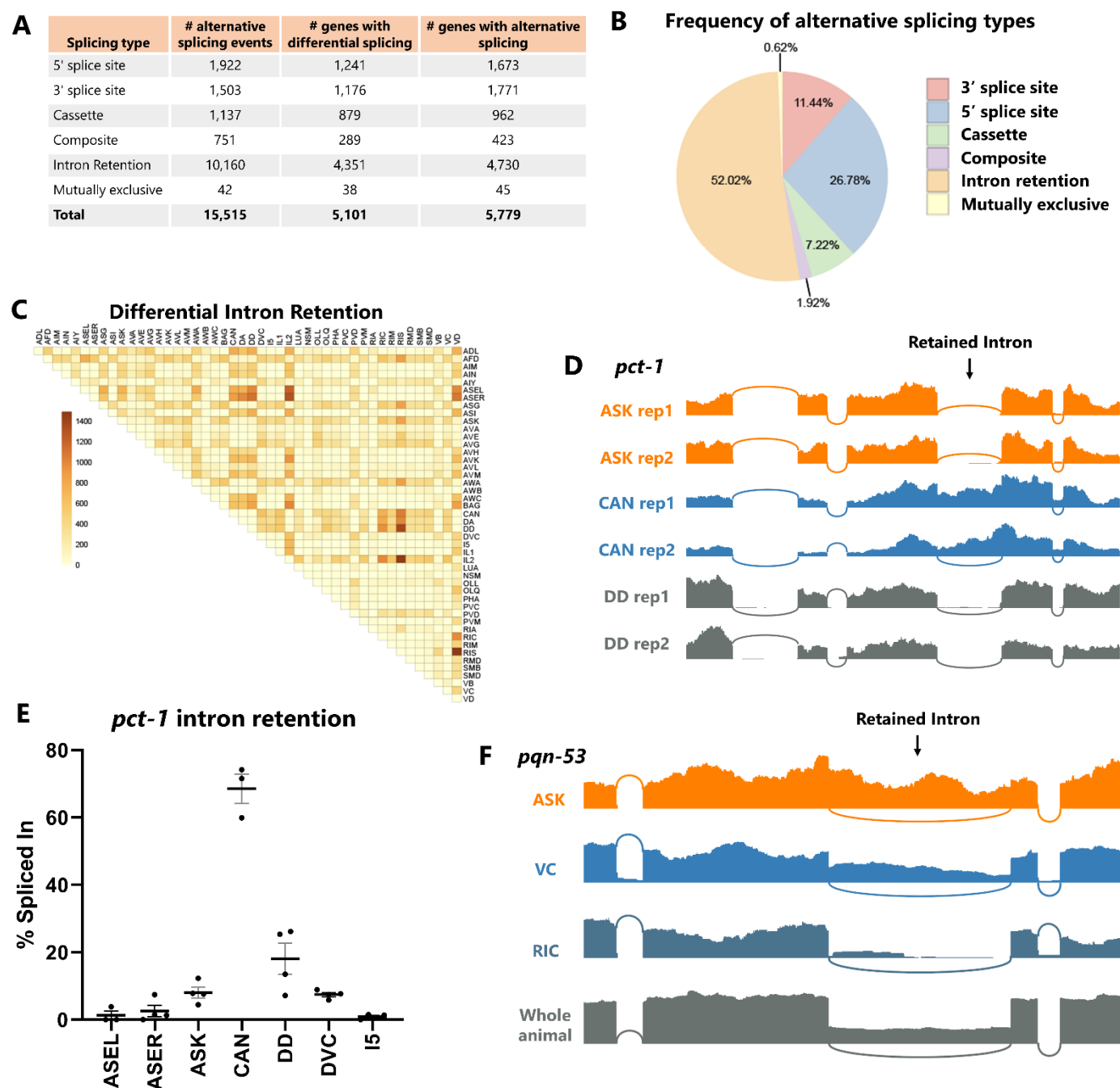
In total we identified 15,515 alternative splicing events in 5,779 distinct genes (Figure 2A). Among these alternative splicing events, we noticed a surprising degree of alternative intron retention across neuronal cell types (Figure 2A-C). Indeed, intron retention accounts for over half (52%) of all differential splicing, followed by alternative 5' splice sites (27%) and 3' splice sites (11%). This is somewhat surprising, as cell-specific intron retention has historically received less attention relative to other splicing types such as cassette exons, although regulated intron retention is a prominent phenomenon in plants<sup>20</sup> and is increasingly appreciated to be important for developmental and stimulus-dependent regulation of neuronal transcripts<sup>10,21-23</sup>. Our data show differential intron retention across neuron types to be very widespread. For example, it is over 7-fold more prevalent than differential cassette exon alternative splicing (Fig 2B).

We sought to test whether the preponderance of intron retention in our data represents ground truth biological observations or might be inflated by technical artifacts. First, we considered whether intron retention signals might falsely arise from stochasticity in lowly-expressed genes (represented by few sequencing reads) or in technical attributes such as PCR artifacts. However, statistically-significant intron retention events tend to be covered by hundreds of reads per biological replicate, and the % included values are reproducible across replicates from a given cell type (Figure 2D-E). For example, an intron in the *pct-1* gene is consistently spliced out in most neurons (~0% retained) but consistently retained in CAN neuron (69%). This argues against a stochastic origin of the intron retention signal.

Second, we considered whether *de novo* mutations in the individual worm strains used to sort specific neuron types might result in splice site mutations causing intron retention specific to that strain. This might occur relatively easily in genes that are not essential for viability or growth. However, using the mapped RNA-Seq data to inspect sequences of the retained introns and surrounding exons, we found no evidence for *de novo* mutations in the intron-retained cell types compared to the intron-spliced cell types (Figure S2).

Finally, we considered whether technical attributes specific to the deep transcriptomic experimental procedure (e.g. cell dissociation, rRNA-depletion) might result in erroneous intron retention signals not found in samples prepared by other methods. We thus compared our data to RNA-Seq data obtained from wild-type whole animal RNA isolated via polyA selection. An example for the gene *pqn-53* is shown in Figure 2F, in which the intron is retained at various levels across neuron types, from mostly retained (ASK, 72%) to mostly spliced (RIC, 10%). Wild-type whole animal RNA-Seq also reveals substantial retention of this intron, thus confirming the observation of alternative splicing of this intron across substantial variations in technical and experimental conditions.

We therefore conclude that the surprising preponderance of differential intron retention across neuronal cell types is likely a real biological phenomenon. If true, cell-specific intron retention would represent an under-appreciated source of gene regulation at the level of individual neurons, representing around half of differential alternative splicing across neuron types.



**Figure 2: Widespread differential intron retention across cell types.** (A) Total number of alternative splicing events (defined as PSIs between 10-90%) and differential alternative splicing ( $|\Delta\text{PSI}| > 10\%$ ) detected in our analysis. Minimum of 5 junction-spanning reads per biological replicate. (B) Fraction of differentially-regulated alternative splicing corresponding to different types of alternative splicing. (C) Differential intron retention across individual neuron types is widespread. (D) RNA-Seq coverage showing an example of a retained intron that is highly retained in CAN but almost completely spliced out in all other neurons. (E) The pattern of *pct-1* intron retention is consistent across multiple biological replicates (individual dots) in various neuron types. (F) An example of a retained intron with retention signal in many neuron types, and confirmation that the retained intron is readily apparent from whole-animal polyA-selected RNA as well.



## Uniqueness index identifies pan-neuronal genes with unique cell-specific isoforms

A global survey of differential alternative splicing in single neurons allowed us to ask whether genes expressed in many or all neurons might have highly-specific isoforms restricted to a single neuron or a small number of neuron types. Such isoforms might be unannotated in current databases obtained from data on whole animals, but could represent highly-specialized isoforms of important neuronal genes with cell-specific function.

To search for such isoforms, we developed an algorithm that aggregates statistically-significant differential alternative splicing across all pairwise cell-type comparisons. This algorithm, which we refer to as a “uniqueness index,” incorporates both statistical significance and effect size to identify splicing events that are robustly different in one cell type compared to all others.

In Figure 3A, an example is shown for the *unc-31* exon #3 cassette exon, in which significant splicing differences in AVL versus all 45 other neurons are summed to obtain a uniqueness index. A positive value indicates that the exon is more included in AVL compared to other cells. The maximum possible value of the uniqueness index is 45, which would represent 100% exon inclusion in AVL and 0% inclusion in all other 45 cells. The minimum value is -45, and a value of 0 indicates no systematic difference in splicing compared to all 45 other neurons. Therefore, the uniqueness index of 23 in Figure 3A indicates a strong exon inclusion signature for the *unc-31* exon in AVL. This is in line with data in Figure 1B-C showing that the exon is included in AVL and a few other neurons, but skipped in most neurons.

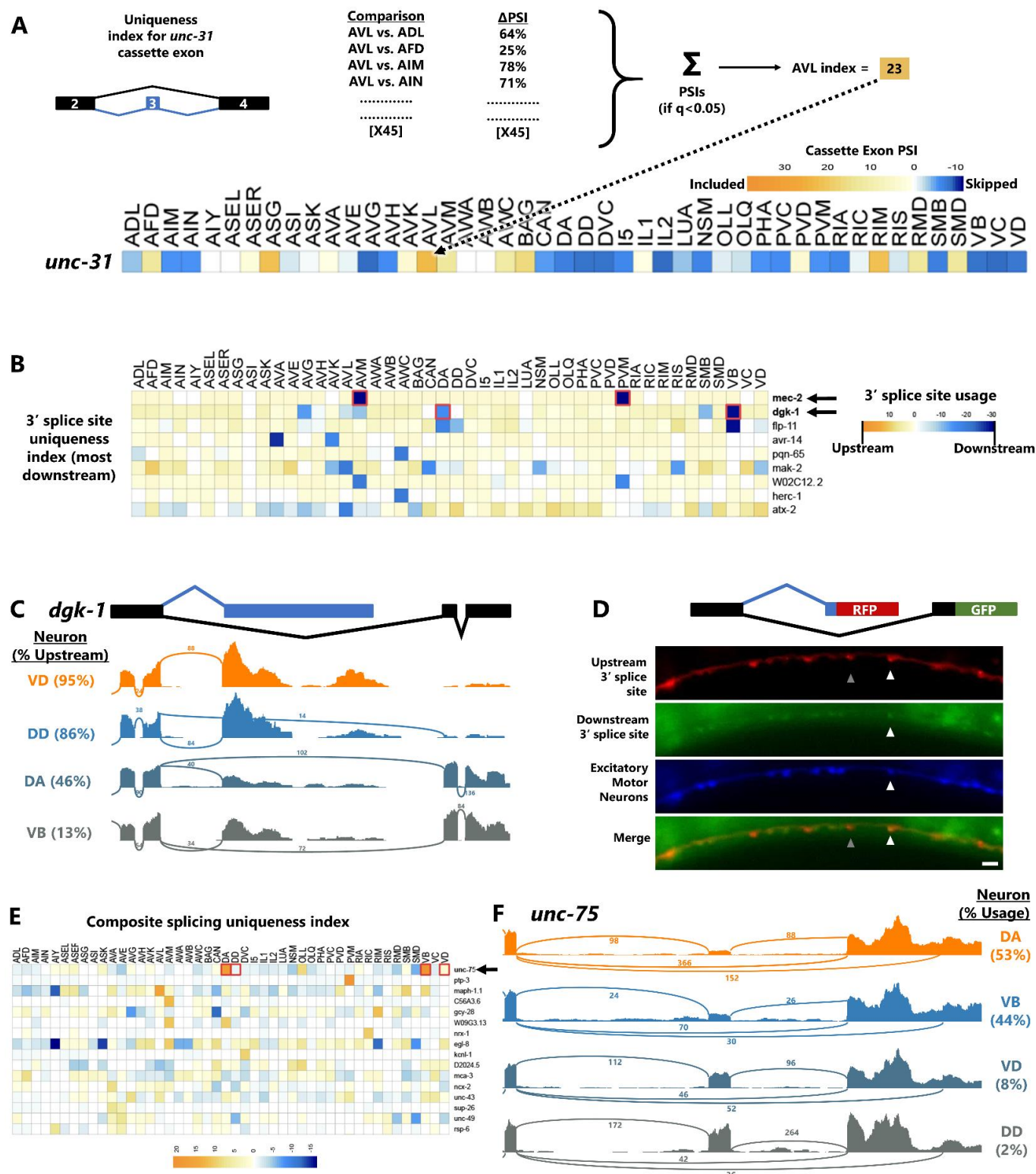
In the case of alternative 3' splice sites (Figure 3B), the highest uniqueness value belongs to the gene *mec-2/Stoml3*, which is predicted to select a downstream 3' splice site in mechanosensory touch neurons (AVM and PVM), while all other neurons select the upstream 3' splice site (Figure 2B, row 1). This is precisely in line with our recent work *in vivo* showing that touch neurons exclusively select the downstream 3' splice site, and that this choice is necessary for the function of the MEC-2 protein in touch neurons<sup>14</sup>.

The second-highest 3' splice site uniqueness value (Figure 3B, row 2) belongs to the gene *dgk-1*, diacylglycerol kinase theta, which exhibits pan-neuronal expression<sup>24</sup>. While most neurons select the upstream 3' splice site, a small number of neurons also select an unannotated downstream 3' splice site. The consequence of this alternative splicing choice is two different mRNA 3' ends (Figure 3C) encoding for either 67 (canonical) or 31 (non-canonical) amino acids. After an initial 6 amino acids of sequence homology, the two alternative isoforms encode distinct amino acid sequences. Interestingly, the cells that uniquely select the downstream splice site (DA and VB neurons) are both excitatory (cholinergic) motor neurons, while the neighboring inhibitory motor neurons (VD and DD) exclusively select the upstream splice site (Figure 3C).

To confirm the *dgk-1* splicing patterns predicted by the uniqueness index *in vivo*, we generated transgenes in which the upstream splice site is RFP-tagged and the downstream splice site is GFP-tagged (Figure 3D). As predicted, most neurons are only RFP+, except for some motoneurons in the ventral nerve cord which are both RFP+ and GFP+. Co-labeling with a BFP expressed in excitatory motoneurons confirms that the neurons expressing both RFP and GFP are excitatory motoneurons, while inhibitory motoneurons only express RFP. Therefore, as predicted by the uniqueness index, a small subset of neurons expresses an unannotated unique isoform of the important pan-neuronal gene *dgk-1*.

The uniqueness index can be applied to any type of alternative splicing (Figure S3, Supplemental Spreadsheet S2), even complex composite splicing (Figure 3E). For example, the neuronal RNA binding protein *unc-75/CELF* gene undergoes composite alternative splicing in which a cassette exon may be included or skipped, and if skipped an alternative 3' splice site choice is available (Fig 3F). These splicing choices affect the third RNA Recognition Motif (RRM) domain of UNC-75, encoding either a full-length RRM (exon included), a truncated

RRM (exon skipped + upstream 3' splice site), or no RRM (exon skipped + downstream 3' splice site). The uniqueness index reveals that in most neurons, the primary splicing choice is cassette exon inclusion, followed by equal amounts of upstream and downstream 3' splice site choice. However, in two specific cell types (again, the DA and VB motor neurons), exon skipping predominates, and the upstream 3' splice site is preferentially selected. Thus, the uniqueness index identifies even complex splicing arrangements with signatures unique to individual neuron types.



**Figure 3: Uniqueness index identifies pan-neuronal genes with cell-specific isoforms.** (A) Schematic of the uniqueness index algorithm- for each alternative splicing event, a single neuron (e.g. AVL) is compared against all other 45 neuron types. The sum of all



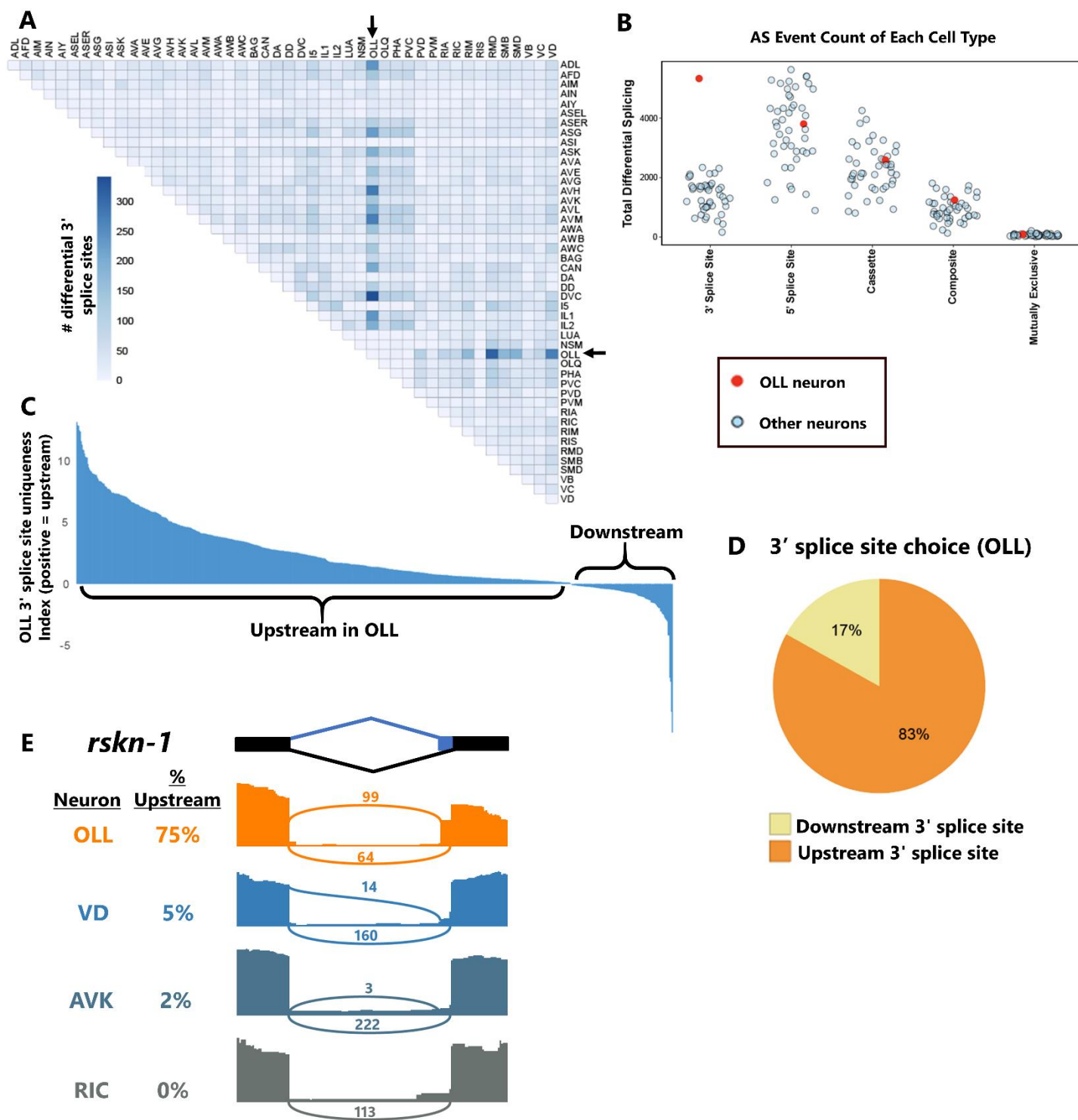
significant ( $q < 0.05$ )  $\Delta$ PSIs is the uniqueness index for that cell/splicing-event pair. A score is computed for every splicing event in every neuron type. (B) Uniqueness index for alternative 3' splice sites, sorted by most "uniquely downstream" splice site choice. (C) Example of *dgk-1* alternative splice site, showing upstream usage in most neurons (e.g. VD and DD) but downstream only in DA and VB. Percentages on left = mean across biological replicates. (D) Transgenic animals harboring pan-neuronal (*rgef-1* promoter) *dgk-1* GFP/RFP splicing reporter as well as cholinergic neuron (*unc-17* promoter) BFP. Scale bar = 10  $\mu$ m. (E) Uniqueness values are obtainable even for complex splicing patterns such as composite splicing. (F) Read coverage illustrating the top uniquely-regulated composite event, in the RNA Binding Protein *unc-75*. DA/VB neurons skip the cassette exon and select the downstream 3' splice site, while other neurons (e.g. VD/DD) include the cassette exon and select the upstream 3' splice site. Plots are from single biological replicates; PSI values are means across replicates.

## Upstream 3' splice site selection in OLL sensory neurons

While most cells have strong correlations between different types of alternative splicing (Fig 1G-H), we found at least one strong exception to this pattern in the OLL sensory neuron. The OLL neuron consistently exhibits a large quantity of alternative 3' splice site selection compared with most other cell types (Figure 4A, S4). On the other hand, no such pattern is evident for other splicing types in OLL (see Figure 1E-F, 2B). Comparing total differential splicing across cell types shows that OLL is a strong outlier for alternative 3' splice sites, but not for other types of alternative splicing (Figure 4B).

We asked whether the alternative 3' splice site usage in OLL follows a consistent pattern (i.e., enriched for upstream vs. downstream splice site usage). Inspecting the uniqueness index values for 3' splice site usage in OLL compared to all other neurons, there exists a strong bias toward positive values (indicating upstream 3' splice site usage) and very few negative values (Figure 4C). Indeed, 83% of alternative 3' splice sites in the OLL show upstream splice site usage (Figure 4D).

The upstream 3' splice site bias in OLL includes both 3' splice sites that are located close to each other (minimum 5 nts, *ZK287.7*) and those that are separated by substantial distance (maximum 2,604 nts, *Y39E4B.10*). An example of the former is shown in Figure 4E, where an alternative 3' splice site (6 nts) in the ribosomal S6 kinase gene *rskn-1* undergoes primarily upstream splice site selection in OLL (75% upstream) but primarily downstream splice site selection in other neurons (0-5% upstream). Together these observations reveal that in at least some cases, individual neurons possess unique and specific splice-type identities. OLL sensory neurons are not only enriched in alternative 3' splicing, they are also specifically enriched in upstream 3' splice site choices.



**Figure 4: Widespread upstream 3' splice site selection in OLL neurons.** (A) OLL exhibits abundant alternative 3' splice site choice across cell-type comparisons. (B) OLL (red dot) is an outlier in the total number of differential alternative 3' splicing events called across all cell-cell comparisons, but not for other types of alternative splicing. (C) Uniqueness values reveal that OLL is specifically enriched for upstream 3' splice site selection. (D) Summary of statistically-significant alternative 3' splice site choices in OLL showing that 83% are biased to upstream selection in OLL. (E) Read coverage illustrating an example of OLL selecting upstream 3' splice site in the *rskn-1* gene, while other neurons select the downstream splice site. Plots are from single biological replicates; PSI values are means across replicates.

## Broad applications for uniqueness index and splicing data

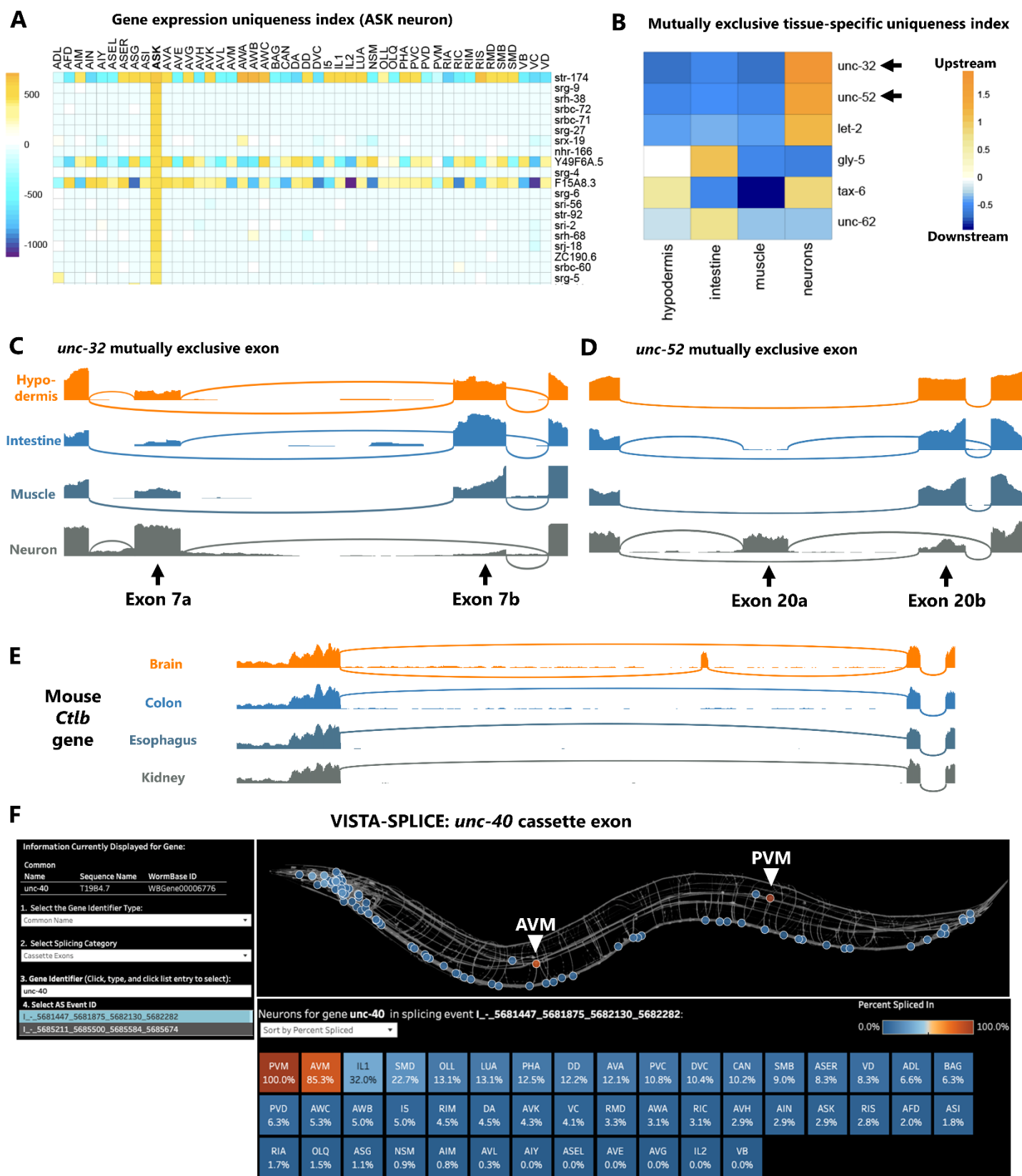
We developed the uniqueness index to identify cell-specific splicing in *C. elegans*, but it is also applicable across various organisms and data types. As one example, we applied the uniqueness index to gene expression data to identify unique expression signatures in single neurons. Figure 5A shows the top twenty unique genes in the sensory neuron ASK. Strikingly, almost all (80%, 16/20) are “serpentine” G-Protein Coupled Receptor genes, a large family of genes (*srg/srbc/srh/str*/etc.) for which specific ligands are mostly unknown, but many of which encode predicted sensory GPCRs<sup>25</sup>. This is consistent with the role of ASK as a chemosensory neuron<sup>26</sup>. Although none of the top ASK-enriched genes have previously been studied, scRNA-Seq datasets from both embryos and L4 larvae<sup>15,27</sup> corroborate the unique expression of these genes in ASK. This highlights the utility of the uniqueness index for working with gene expression data and generating hypotheses about neuron-specific function for unstudied genes.

As a second example of the utility of the uniqueness index, we applied it to tissue-specific RNA-Seq data<sup>28</sup> to identify alternative splicing choices unique to specific tissues (intestine, muscle, hypodermis, neurons) (Figure 5B, S5). Focusing on mutually exclusive exons (Figure 5B), we found that the highest-magnitude tissue-specific splicing event is in a gene for the V-ATPase *unc-32*. In neurons the upstream exon is selected, while in other tissues the downstream exon is selected. This agrees with previous *in vivo* reporters demonstrating selection of the upstream exon specific to the neurons. The second-highest-magnitude splicing event is in the Perlecan *unc-52* gene, which has not been studied *in vivo* but which displays a similar pattern to *unc-32*: the upstream exon is selected in neurons, while the downstream exon is selected in other tissues (Figure 5D).

We also sought to apply the uniqueness index to tissue-specific splicing in a different species. Using RNA-Seq data from 13 different mouse tissues<sup>29</sup>, we indexed brain-specific genes and isoforms. The top brain-specific genes are all associated with synaptic transmission (*Grm5*, *Gabra5*, *Gabra6*, *Cagnag3*) and/or other brain-specific functions (*Mobp*, *Gpr26*, *Lamp5*) (Fig S5). At the level of alternative splicing, we identified many brain-specific isoforms in which one isoform is expressed in the brain, and the other expressed in all other tissues (Fig S5, Spreadsheet S5-6). For example, clathrin light chain B (*Ctlb*) exon 5 is alternatively spliced such that the exon is mostly included in brain but always skipped in other tissues (Figure 5E). Together these observations demonstrate the utility of the uniqueness index across different data sources (cell-specific and tissue-specific) and species (worm and mouse).

To increase the accessibility of our neuron-specific splicing data to a broader audience, we developed a visualization platform we call VISTA-SPLICE. This builds on our visualization platform VISTA, in which gene expression values are displayed as a “spatial heatmap” across the *C. elegans* nervous system<sup>30</sup>. Here we adopt the same visualization scheme, except “Percent Spliced In” values are displayed for splicing events of interest. Many splicing types (e.g. cassettes, retained introns, 3’ splice sites) can be visualized. An example is shown (Figure 5F) for an unannotated cassette exon in the *unc-40/DCC* Netrin receptor. This exon encodes an additional 85 amino acids residing between the P1 and P2 intracellular domains, both of which are important for Netrin-mediated neurite outgrowth<sup>31</sup>. VISTA-SPLICE makes it clear that the exon undergoes cell-specific alternative splicing (Figure 5F). *unc-40* expression is detected in many neurons, most of which skip the exon (~0% inclusion). But two neurons, the AVM and PVM mechanosensory “touch neurons” are clear outliers (85% and 100% inclusion, respectively). Visualizations of such cell-specific splicing patterns should make neuronal splicing analysis accessible to a broader community. The dashboard is freely available at

[https://public.tableau.com/app/profile/smu.oit.data.insights/viz/VISTA-SPLICEVisualizingtheSpatialTranscriptomeofC\\_E/VISTA-SPLICE](https://public.tableau.com/app/profile/smu.oit.data.insights/viz/VISTA-SPLICEVisualizingtheSpatialTranscriptomeofC_E/VISTA-SPLICE)



**Figure 5: Broad applications for uniqueness index and splicing data.** (A) Uniqueness index for gene expression values, as determined by DESeq2, sorted here for most unique to ASK neurons. Of the top 20 unique genes to ASK, 80% are serpentine GPCRs (*str/srg/srh/srb/srx/etc.*) (B) Tissue-specific alternative splicing uniqueness index for mutually exclusive exons, sorted by highest upstream exon selection. The top two mutually exclusive exons are then displayed in (C-D) as read coverage demonstrating upstream exon selection unique to neurons in both cases. (E) *Ctla* exon 5 is mostly included in mouse brain, but is completely skipped in other tissues. (F) Screen shot of the VISTA-SPLICE tool displaying exon inclusion levels for novel cassette exon in the *unc-40* Netrin receptor gene. Note that the two touch neurons AVM and PVM display high inclusion levels, while all other neurons display low exon inclusion.

## Identification of causal regulatory factors establishing cell-specific isoforms

A comprehensive cell-specific catalogue of alternative splicing in a genetically tractable organism such as *C. elegans* holds promise for going beyond descriptive analysis and proceeding to investigate underlying regulatory mechanisms. To this end we focused on two of the cassette exons with the strongest uniqueness values but with opposite splicing patterns. First, the *unc-40* cassette exon (see Figure 5F) which is highly included in touch neurons (AVM/PVM) and skipped in other neurons (Figure 6A). Second, a cassette exon (currently annotated as a constitutively-spliced exon) in the ligand-gated ion channel *lgc-31* which is ~0% included in AVM/PVM, and ~100% included in other neurons.

We generated transgenic bichromatic alternative splicing reporters to confirm *in vivo* the predicted splicing patterns. These transgenic reporters (Figure 6B) yield RFP fluorescence when an alternative cassette exon is included, and GFP when skipped. As predicted, most neurons are GFP+ (exon skipped) for the *unc-40* transgene, and RFP+ (exon included) for the *lgc-31* transgene (Figure S6), while in AVM/PVM the patterns are reversed (Figure 6C). Expressing the splicing reporters pan-neuronally allowed us to interrogate the splicing patterns in additional neuron types not covered by the CeNGEN RNA-Seq data. For example, we predicted the ALM/PLM neurons (the other two touch neuron cell types) would have the same unique splicing patterns as AVM/PVM, and indeed this is the case for both *unc-40* and *lgc-31* (Figure 6C, S6).

The *in vivo* splicing reporters also allowed us to ask whether the splicing events are subject to cell-specific developmental regulation or exhibit variability across individuals. We found that the splicing patterns for both genes are invariant across developmental stages and are also invariant across individuals (Figure S6), indicating that these splicing events are strictly and deterministically regulated at the single-neuron level.

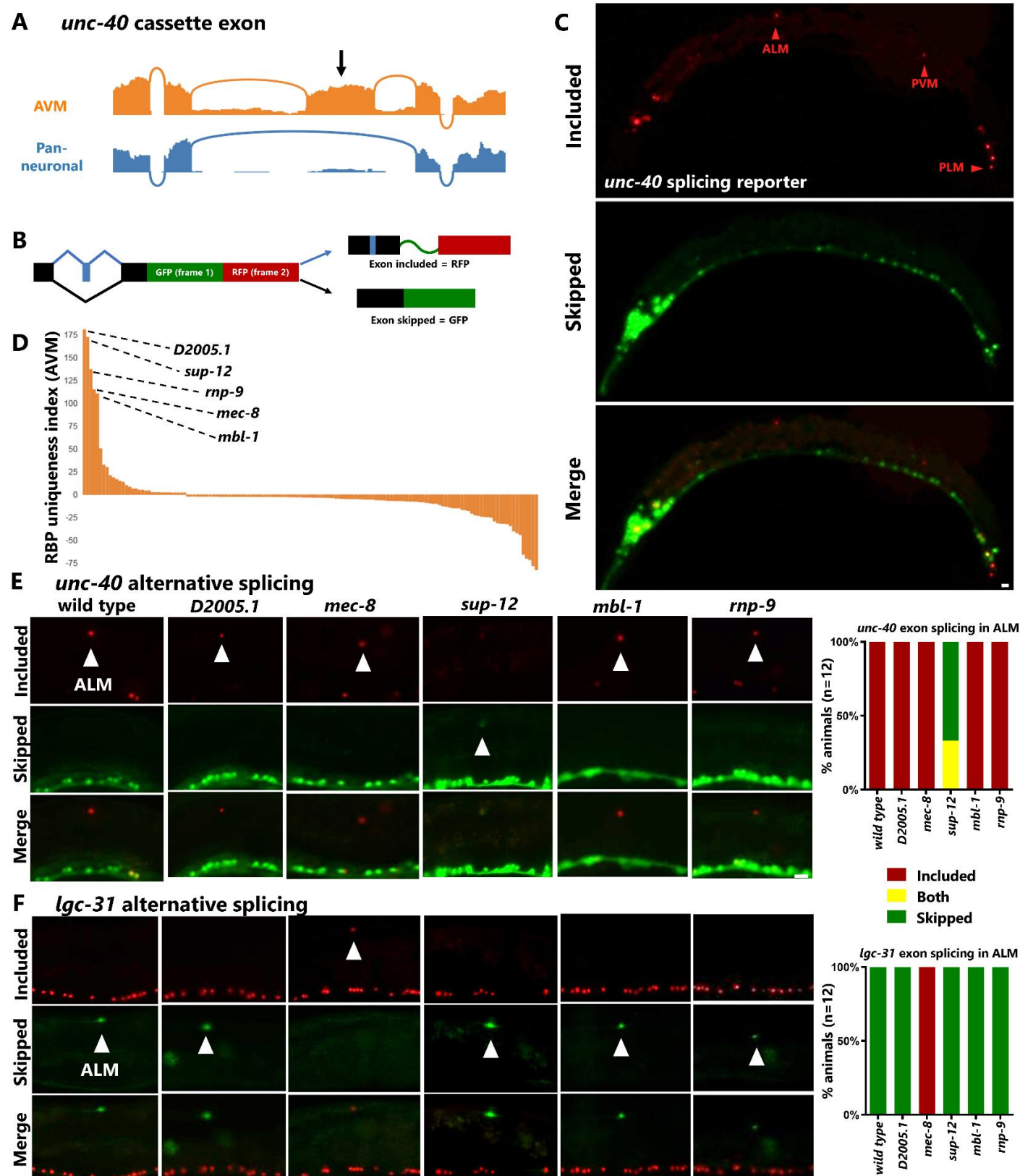
The reporters also allowed us to genetically test *in vivo* for potential *trans*-acting factors that establish the striking *unc-40* and *lgc-31* splicing patterns. To obtain candidate *trans*-acting factors we returned to our gene expression uniqueness index, this time applying it to differential expression of RNA Binding Protein (RBP) genes. Sorting for uniqueness index values in AVM/PVM should reveal RBPs highly expressed in AVM/PVM compared to all other neurons. The RBPs with the highest uniqueness values were similar when sorted by AVM or PVM (Figure 6D, S6), suggesting that these closely-related cell types also express similar complements of RBPs. We selected the top five RBPs (Figure 6D) for further genetic investigation. We previously found two of these RBPs to affect alternative splicing in touch neurons (*mec-8/RBPMS* and *mb1-1/Muscleblind*)<sup>32</sup>, while the other three (*sup-12*, *rnp-9*, *D2005.1*) have not previously been reported to be expressed or function in touch neurons.

We systematically crossed both splicing reporters with mutants for all five RBP genes and tested for changes in cell-specific splicing patterns (Figure 6E-F). For each splicing pattern we found a single RBP responsible for establishing the cell-specific splicing pattern. The *unc-40* cassette exon requires SUP-12/RBM24 (Figure 6E), while the *lgc-31* cassette exon requires MEC-8/RBPMS (Figure 6F). The remaining three RBPs have no effect on either of the cassette exons. In each case, loss of the regulatory RBP affects all four touch neuron types (ALM/AVM/PLM/PVM) resulting in a splicing switch from the unique touch neuron isoform to the default neuronal isoform, and is fully penetrant across individual animals.

Identification of SUP-12/RBM24 as a specific regulator of touch neuron splicing is surprising, as previous work characterized SUP-12 as a muscle-specific splicing factor in *C. elegans*<sup>33,34</sup>. However, our work now identifies additional *sup-12* expression in a small number of neurons (primarily AVM/PVM) and this observation is corroborated by independently-generated scRNA-Seq data<sup>15,27</sup>. This indicates that SUP-12 both directs muscle-specific splicing and also directs touch-neuron-specific splicing within the nervous system. SUP-12 thus joins



the RBPs MBL-1 and MEC-8 as a trio of factors experimentally demonstrated to provide unique splicing attributes to the touch neurons. Therefore, touch neurons possess at least three non-redundant regulatory mechanisms that each achieve the goal of establishing isoforms in touch neurons that differ from those in the rest of the nervous system.



**Figure 6: Identification of causal regulatory factors establishing cell-specific isoforms.** (A) Read coverage for the *unc-40* cassette exon visualized in 5F, showing nearly complete inclusion in AVM neuron, but complete skipping in pan-neuronal RNA-Seq. (B)

Schematic for two-color alternative splicing reporter in which inclusion of a cassette exon is engineered to alter the reading frame, resulting in RFP translation when exon is included and GFP translation when exon is skipped. (C) Splicing reporter for *unc-40* expressed pan-neuronally (*rgef-1* promoter) agrees with prediction from 5F that AVM and PVM express included isoform, and also reveals inclusion in other touch neurons (ALM/PLM), while most other neurons express only the skipped isoform. (D) Uniqueness index for RBP gene expression in AVM neuron reveals small handful of enriched RBPs. (E) Visualizing exon inclusion in the ALM neuron reveals that SUP-12, but none of the other RBPs, is required for cell-specific *unc-40* inclusion. (F) For the *lgc-31* cassette exon, MEC-8, but none of the other RBPs, is required for cell-specific skipping. Scale bar = 10  $\mu$ m.

## DISCUSSION

### Cell-type-specific isoforms of pan-neuronal genes

A useful concept in neuronal gene expression is that of the pan-neuronal versus the cell-type-specific gene. The former constitutes a gene required for fundamental shared processes across all neurons (*e.g.* synaptic vesicle machinery) while the latter constitutes a gene required for specific properties of a given neuronal type (*e.g.* neurotransmitter synthesis)<sup>35,36</sup>. Our work highlights additional layers of complexity in this concept, most notably the pan-neuronal gene with cell-type-specific isoforms.

An interesting example of this is in the gene *dgk-1*, which encodes a diacyl glycerol kinase that exhibits pan-neuronal expression<sup>24</sup> and is required for the normal function of various neuronal types<sup>37,38</sup>. We identify a novel *dgk-1* isoform that is expressed only in excitatory (cholinergic) motor neurons of the ventral nerve cord, while all other neurons express only the canonical isoform. Therefore *dgk-1* is simultaneously a pan-neuronal gene but is cell-type-specific with respect to individual isoforms. This type of regulatory arrangement might be functionally useful, allowing a single gene to simultaneously perform ubiquitous pan-neuronal function while also providing bespoke features unique to individual neuron types.

### Deep transcriptomes reveal complex and novel neuron-specific alternative splicing

Obtaining deep-sequenced populations of single neurons with multiple biological replicates, coupled with a flexible computational pipeline, allowed us to examine many splicing types that are not commonly assayed at the single-neuron level. For example, we identify widespread composite splicing events regulated across individual neurons. Regulated composite splicing requires complex coordinated selection of multiple nearby alternative splice sites. Therefore, the observation of widespread composite alternative splicing raises interesting mechanistic questions of how individual neurons simultaneously regulate the selection of multiple splice sites in a single transcript.

We find that intron retention is by far the most common type of differential splicing across neuronal cell types, although intron retention has not historically been a focus of single-cell transcriptomics analyses. The most commonly-studied splicing type in single-neuron genomics analyses, including our own, is cassette exons, for both technical and historical reasons<sup>3,5,7,39,40</sup>. Nevertheless, our data here indicate that cassette exons make up only a minor fraction of alternative splicing in the nervous system. With future technical and conceptual innovations, we anticipate an expanding appreciation of the diverse alternative splicing landscape in single cells.

We note that complementary work by the CeNGEN consortium (Weinreb *et al.*, co-submitted) using different analysis pipelines largely agrees with the conclusions presented here (*e.g.* highlighting the *unc-40* cassette exon

in AVM, and lack of relationship between cell-specific gene expression and splicing). On the topic of introns, however, they find less differential splicing (using SUPPA) than we do (using JUM). One likely explanation is differences in analysis algorithms. For example, SUPPA depends on transcriptome annotations, while JUM does not. Indeed, intron retention identification has been shown to be particularly sensitive to choice of computational approaches<sup>41</sup>. If differential intron retention is indeed abundant across neurons, this raises interesting questions about both the regulatory and functional implications for individual neurons. Neuronal intron retention is increasingly appreciated as a mechanism for developmental and stimulus-dependent gene regulation<sup>21-23</sup>, and our results suggest that such phenomena are likely even more widespread than previously anticipated.

## Identifying causal regulatory factors that establish individual neuron splicing choices

A strength of studying single-cell transcriptomes in a strong genetic system such as *C. elegans* is the promise of proceeding from descriptive catalogs of cell-specific isoforms to interrogating the regulatory factors that are responsible. We achieved this goal for two different neuron-subtype-specific splicing events in the touch neurons. RBPs expressed uniquely in these neurons were identified using the uniqueness index, then existing null mutants for the RBP genes in question were crossed with *in vivo* splicing reporters to test for changes in cell-specific splicing patterns.

One conclusion from the observations on cell-specific splicing regulation in touch neurons is that there are multiple regulatory routes to the same cell-specific splicing outcome. Here we identify two different RBPs (MEC-8 and SUP-12) that independently establish unique touch-neuron isoforms for different genes. Together with our previous work demonstrating a case of combinatorial regulation by MEC-8 and MBL-1 resulting in cell-specific splicing of the *sad-1* kinase<sup>32</sup>, at least three different unique strategies are deployed to achieve the same outcome (*i.e.* touch-neuron-specific isoforms of broadly-expressed genes). As such, there does not seem to be a single conserved master regulator of touch neuron isoforms. Rather, various factors are deployed that are each readily amenable to the task due to their unique expression in this cell type. It will be interesting to test whether such distributed regulatory networks are common across other neuronal cell types as well.

We anticipate that the experimental framework established here will broadly apply to questions of cell-specific gene regulation. For example, the uniqueness index pipeline we used to identify splicing regulators in touch neurons should also be applicable to identifying transcriptional regulators in specific neurons or tissues, or for nominating candidate chemosensory GPCRs in specific neurons (Figure 5A). We also hope that visualizing the spatial transcriptome with VISTA-SPLICE will lower the barrier to analyzing cell-specific splice isoforms, and the global alternative splicing values reported here will enable researchers to study cell-specific splicing of their genes of interest with ease.

## METHODS

**Strains used:** *adnEx10[rgef-1::dgc-1 splicing reporter + unc-17::BFP]*, *adnEx11[rgef-1::unc-40 splicing reporter]*, *adnEx12[rgef-1::lgc-31 splicing reporter]*, *D2005.1(ok2689)* I, *mb1-1(csb31)* X, *sup-12(ok1843)* X, *rnp-9(ok144)* X, *mec-8(csb22)* I. Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440).

**Imaging:** Fluorescent Images were taken at 10X using a Zeiss Axio Imager.Z1 AX10 microscope and Zeiss ZEN2.5 (blue edition) software. Images were processed using ImageJ 1.54d (NIH, USA. <http://imagej.org>).

**Splicing reporters:** Two-color splicing reporter backbones were obtained as described previously<sup>39,42</sup>. For the inserts, *lgc-31* a 431 bp fragment containing the alternative exon and flanking introns/exons was generated (WBcel235 coordinates chrV:15,513,325-15,513,755), and likewise for *unc-40*, which was an 890 bp fragment (coordinates chrI:5,681,422-5,682,311). For the *dgc-1* 3' splice site selection two-color reporter, two different splicing reporter were generated. In each case, the fluorescent protein was fused immediately upstream of the respective stop codon.

**VISTA-SPLICE:** As with VISTA, upon which VISTA-SPLICE was built, VISTA-SPLICE was created with Tableau and is freely available via Tableau Public. Original anatomical and cell-position information was derived from OpenWorm ([openworm.org](http://openworm.org)) and WormAtlas ([wormatlas.org](http://wormatlas.org)).

### Cell and tissue types:

SRA run selector<sup>43</sup> was used to search for and download SRR files associated with PRJNA952691<sup>16</sup>. A total of 180 replicates from 46 unique cell types were used; each cell type represents a different neuron in *C. elegans*. SRA run selector was also used to search for and download SRR files associated with PRJNA400796<sup>28</sup>. A total of 22 replicates from 4 unique tissue types were used; paired-end reads were not analyzed due to their downstream incompatibility with the remaining single-end replicates.

### DESeq2/HTSeq:

Gene-specific counts were tabulated for each sample using HTSeq<sup>44</sup> and statistically-significant differentially expressed transcripts were identified with DESeq2<sup>17</sup>. Identification of uniquely dysregulated genes was determined by filtration of genes using strict and conservative thresholds. For gene expression, including the expression of RNA binding proteins, this threshold was p-value <0.01 and  $|\log_2FC| > 2$  as determined by DESeq2. All scripts used to filter and quantify DESeq results can be viewed at <https://github.com/xcwolfe/Differential-Expression-in-C-elegans>

### STAR:

Libraries were sequenced and aggregated by Barrett *et. al* (PRJNA952691) and Kaletsky *et. al* (PRJNA400796), then aligned to the worm genome (version WBcel235) using STAR<sup>45</sup>. STAR parameters include a read overhang minimum of 15 bases and a multimap maximum of 20 locations.

### JUM:

The Junction Usage Model (2.0.2) (JUM) was used to identify differentially spliced isoforms in different cell types and quantify their expression levels by computing the  $\Delta PSI$  (difference of Percent Spliced Isoform)<sup>11</sup>. JUM reports six different categories of alternative splicing events: cassette exons, mutually exclusive exons (MXE),

alternative 5' splice sites (A5S), alternative 3' splice sites (A3S), intron retention events, and composite events (a combination of two or more of the previous five categories).

The parameters for JUM A include: --JuncThreshold 5 --Condition1\_fileNum\_threshold (n-1) --Condition2\_fileNum\_threshold (n-1) --IRthreshold 5 --Readlength 100 --Thread 3 (in which n = the number of replicates for a given cell type). The parameters for JUM B include: --Test pvalue --Cutoff 1 --TotalFileNum 8 --Condition1\_fileNum\_threshold (n-1) --Condition2\_fileNum\_threshold (n-1) (in which n = the number of replicates for a given cell type). The parameters for JUM C include: --Test pvalue --Cutoff 1 --TotalCondition1FileNum (n-1) --TotalCondition2FileNum (n-1) --REF refFlat.txt (in which n = the number of replicates for a given cell type). All scripts used to filter and quantify JUM results can be viewed at <https://github.com/xcwolfe/Differential-Expression-in-C-elegans>

**Summation/quantitation of JUM results:** Identification of unique alternative splicing events was determined by filtering of events using strict and conservative thresholds. This includes a junction threshold of 5, p-value <0.05, q-value <0.05 and  $|\Delta\text{PSI}| > 10\%$  as determined by JUM. For our downstream analysis of alternative 3' splice site (A3S) and alternative 5' splice site (A5S) events, we excluded events that contained more than two potential 3' and 5' coordinates, respectively.

**Global Quantification of Differential Alternative Splicing Types:** The parameters used to determine the total counts of unique alternative splicing events included (in addition to the significance thresholds listed above) a minimum threshold of 5 counts of alternative splicing events as detected by JUM. This threshold is also applied for the total counts of uniquely spliced genes. Individual cell type vs. cell type comparisons and tissue type vs. tissue type comparisons (i.e., JUM heatmaps Supplemental Figure S1) retained the minimum threshold of a single (1) count since it is not possible for the same alternative splicing event to be detected more than once in a comparison.

**Sum of first  $\Delta\text{PSIs}$  ("uniqueness" index):** Since JUM is meant to determine and represent alternative splicing (AS) events between a treatment and control group (i.e., 2 different cell types), JUM was run 2082 times – once for each cell type compared to all other cell types ( $46 \times 45 = 2070$ ) and once for each tissue type compared to all other tissue types ( $4 \times 3 = 12$ ). The results of all JUM cell type-cell type comparisons for each AS type were merged into a set of large data frames composed of every single AS event detected in any of the 2070 runs. Additionally, a second set of data frames was composed for every single AS event detected in any of the 12 tissue type vs. tissue type runs. For each alternative splicing event detected in each comparison, the first  $\Delta\text{PSI}$  (the most "upstream" coordinate that was spliced in) of that event was used to represent a particular splicing choice for a given cell/tissue type. When an alternative splicing event was detected in more than one comparison, the first  $\Delta\text{PSI}$  value of every identical AS event in all comparisons was added to one another – this number was referred to as the "sum of all first  $\Delta\text{PSIs}$ " for a given cell/tissue type and given AS event ID. To ensure that  $\Delta\text{PSI}$  represented the most upstream coordinate/exon choice for all AS events, the  $\Delta\text{PSI}$  of A3S events on the positive (+) strand as well as the  $\Delta\text{PSI}$  of A5S events on the negative (-) strand were reversed during the summation of results.

**Usage of alternative splicing junctions by individual cell types:** To determine the splicing junctions used by individual cell types in a non-comparative fashion, we analyzed the detailed output files provided by JUM. Each detailed output file provides the percentage usage of each AS junction for each individual replicate – percentage usage values are consistent between JUM comparisons regardless of the cell types being compared. The percentage usage values of each AS event ID were averaged between replicates of the same cell



type and aggregated into their own data frames. These percentage usage data frames were merged together by common AS event ID to create a “mega” matrix for each AS type (6 mega matrices in total, Supplemental Spreadsheet S3-4). Junctions representing percent inclusion levels (for cassettes or introns) or upstream splicing levels (for 5', 3', and mutually exclusive splicing) were selected for each splicing event ID.

## FUNDING

National Institute of General Medical Sciences of the National Institutes of Health [R35GM133461]; National Institute of Neurological Disorders and Stroke of the National Institutes of Health [R01NS111055].

## REFERENCES

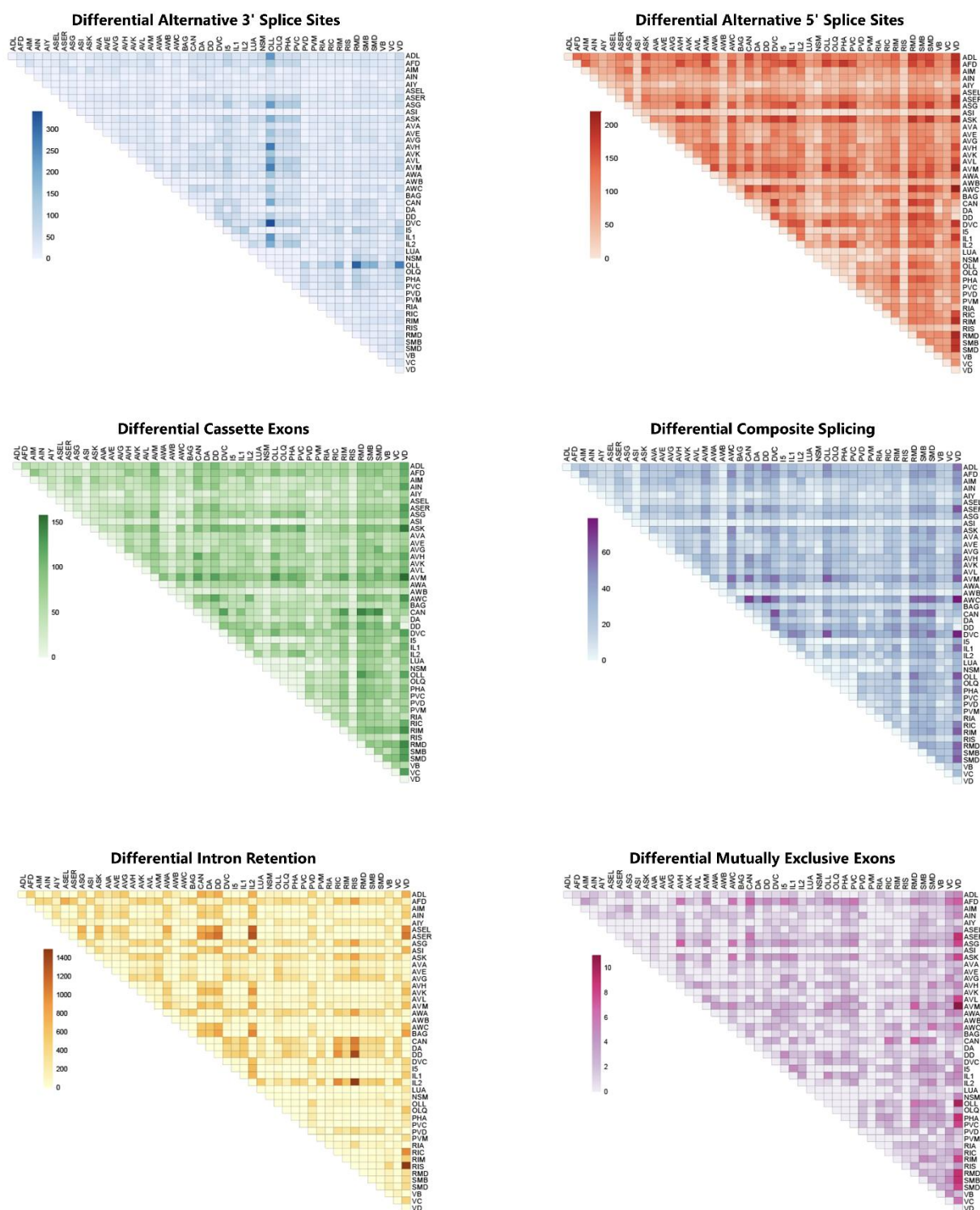
1. Zeng, H. & Sanes, J. R. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nature Reviews Neuroscience* **18**, 530–546 (2017).
2. Tanay, A. & Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338 (2017).
3. Feng, H. *et al.* Complexity and graded regulation of neuronal cell-type-specific alternative splicing revealed by single-cell RNA sequencing. *Proc Natl Acad Sci U S A* **118**, e2013056118 (2021).
4. Westoby, J., Artemov, P., Hemberg, M. & Ferguson-Smith, A. Obstacles to detecting isoforms using full-length scRNA-seq data. *Genome Biol.* **21**, 74 (2020).
5. Song, Y. *et al.* Single-Cell Alternative Splicing Analysis with Expedition Reveals Splicing Dynamics during Neuron Differentiation. *Molecular Cell* **67**, 148-161.e5 (2017).
6. Buen Abad Najjar, C. F., Yosef, N. & Lareau, L. F. Coverage-dependent bias creates the appearance of binary splicing in single cells. *Elife* **9**, e54603 (2020).
7. Buen Abad Najjar, C. F., Burra, P., Yosef, N. & Lareau, L. F. Identifying cell state-associated alternative splicing events and their coregulation. *Genome Res* **32**, 1385–1397 (2022).
8. Norris, A. D. & Calarco, J. A. Emerging Roles of Alternative Pre-mRNA Splicing Regulation in Neuronal Development and Function. *Front Neurosci* **6**, 122 (2012).
9. Marasco, L. E. & Kornblihtt, A. R. The physiology of alternative splicing. *Nat Rev Mol Cell Biol* **24**, 242–254 (2023).
10. Braunschweig, U. *et al.* Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res* **24**, 1774–1786 (2014).
11. Raj, B. *et al.* Cross-regulation between an alternative splicing activator and a transcription repressor controls neurogenesis. *Mol Cell* **43**, 843–850 (2011).

12. Wojtowicz, W. M. *et al.* A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. *Cell* **130**, 1134–1145 (2007).
13. Gracheva, E. O. *et al.* Ganglion-specific splicing of TRPV1 underlies infrared sensation in vampire bats. *Nature* **476**, 88–91 (2011).
14. Liang, X., Calovich-Benne, C. & Norris, A. Sensory neuron transcriptomes reveal complex neuron-specific function and regulation of mec-2/Stomatin splicing. *Nucleic Acids Res* **50**, 2401–2416 (2022).
15. Taylor, S. R. *et al.* Molecular topography of an entire nervous system. *Cell* **184**, 4329–4347.e23 (2021).
16. Barrett, A. *et al.* Integrating Bulk and Single Cell RNA-Seq Refines Transcriptomic Profiles of Specific *C. Elegans* Neurons. <http://biorxiv.org/lookup/doi/10.1101/2022.04.05.487209> (2022) doi:10.1101/2022.04.05.487209.
17. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, (2014).
18. Wang, Q. & Rio, D. C. JUM is a computational method for comprehensive annotation-free analysis of alternative pre-mRNA splicing patterns. *Proceedings of the National Academy of Sciences* 201806018 (2018) doi:10.1073/pnas.1806018115.
19. Speese, S. *et al.* UNC-31 (CAPS) is required for dense-core vesicle but not synaptic vesicle exocytosis in *Caenorhabditis elegans*. *J Neurosci* **27**, 6150–6162 (2007).
20. Wang, B.-B. & Brendel, V. Genomewide comparative analysis of alternative splicing in plants. *Proc Natl Acad Sci U S A* **103**, 7175–7180 (2006).
21. Mazille, M., Buczak, K., Scheiffele, P. & Mauger, O. Stimulus-specific remodeling of the neuronal transcriptome through nuclear intron-retaining transcripts. *EMBO J* **41**, e110192 (2022).
22. Mauger, O., Lemoine, F. & Scheiffele, P. Targeted Intron Retention and Excision for Rapid Gene Regulation in Response to Neuronal Activity. *Neuron* **92**, 1266–1278 (2016).
23. Petrić Howe, M. *et al.* Physiological intron retaining transcripts in the cytoplasm abound during human motor neurogenesis. *Genome Res* **32**, 1808–1825 (2022).
24. Nurrish, S., Ségalat, L. & Kaplan, J. M. Serotonin inhibition of synaptic transmission: Galpha(0) decreases the abundance of UNC-13 at release sites. *Neuron* **24**, 231–242 (1999).
25. Vidal, B. *et al.* An atlas of *Caenorhabditis elegans* chemoreceptor expression. *PLoS Biol* **16**, e2004218 (2018).
26. Bargmann, C. I. & Horvitz, H. R. Chemosensory neurons with overlapping functions direct chemotaxis to multiple chemicals in *C. elegans*. *Neuron* **7**, 729–742 (1991).
27. Packer, J. S. *et al.* A lineage-resolved molecular atlas of *C. elegans* embryogenesis at single-cell resolution. *Science* **365**, eaax1971 (2019).

28. Kaletsky, R. *et al.* Transcriptome analysis of adult *Caenorhabditis elegans* cells reveals tissue-specific gene and isoform expression. *PLoS Genet* **14**, e1007559 (2018).
29. Söllner, J. F. *et al.* An RNA-Seq atlas of gene expression in mouse and rat normal tissues. *Sci Data* **4**, 170185 (2017).
30. Liska, D., Wolfe, Z. & Norris, A. VISTA: visualizing the spatial transcriptome of the *C.elegans* nervous system. *Bioinform Adv* **3**, vbad127 (2023).
31. Gitai, Z., Yu, T. W., Lundquist, E. A., Tessier-Lavigne, M. & Bargmann, C. I. The netrin receptor UNC-40/DCC stimulates axon attraction and outgrowth through enabled and, in parallel, Rac and UNC-115/AbLIM. *Neuron* **37**, 53–65 (2003).
32. Thompson, M. *et al.* Splicing in a single neuron is coordinately controlled by RNA binding proteins and transcription factors. *Elife* **8**, (2019).
33. Anyanful, A. *et al.* The RNA-binding protein SUP-12 controls muscle-specific splicing of the ADF/cofilin pre-mRNA in *C. elegans*. *J Cell Biol* **167**, 639–647 (2004).
34. Kuroyanagi, H., Ohno, G., Mitani, S. & Hagiwara, M. The Fox-1 family and SUP-12 coordinately regulate tissue-specific alternative splicing in vivo. *Mol Cell Biol* **27**, 8612–8621 (2007).
35. Leyva-Díaz, E. & Hobert, O. Robust regulatory architecture of pan-neuronal gene expression. *Curr Biol* **32**, 1715–1727.e8 (2022).
36. Stefanakis, N., Carrera, I. & Hobert, O. Regulatory Logic of Pan-Neuronal Gene Expression in *C. elegans*. *Neuron* **87**, 733–750 (2015).
37. Wiese, M., Antebi, A. & Zheng, H. Intracellular trafficking and synaptic function of APL-1 in *Caenorhabditis elegans*. *PLoS One* **5**, e12790 (2010).
38. Wang, J. *et al.* G protein-coupled receptor kinase-2 (GRK-2) regulates serotonin metabolism through the monoamine oxidase AMX-2 in *Caenorhabditis elegans*. *J Biol Chem* **292**, 5943–5956 (2017).
39. Norris, A. D. *et al.* A Pair of RNA-Binding Proteins Controls Networks of Splicing Events Contributing to Specialization of Neural Cell Types. *Molecular Cell* **54**, 946–959 (2014).
40. Taylor, M., Marx, O. & Norris, A. TDP-1 and FUST-1 co-inhibit exon inclusion and control fertility together with transcriptional regulation. *Nucleic Acids Res* **51**, 9610–9628 (2023).
41. Zheng, J., Lin, C., Wu, Z. & Li, H.-D. A comparison of computational approaches for intron retention detection. *Big Data Min. Anal.* **5**, 15–31 (2022).
42. Choudhary, B., Marx, O. & Norris, A. D. Spliceosomal component PRP-40 is a central regulator of microexon splicing. *Cell Rep* **36**, 109464 (2021).
43. Leinonen, R., Sugawara, H., Shumway, M., & International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res* **39**, D19–21 (2011).

44. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
45. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

## SUPPLEMENTAL FIGURES



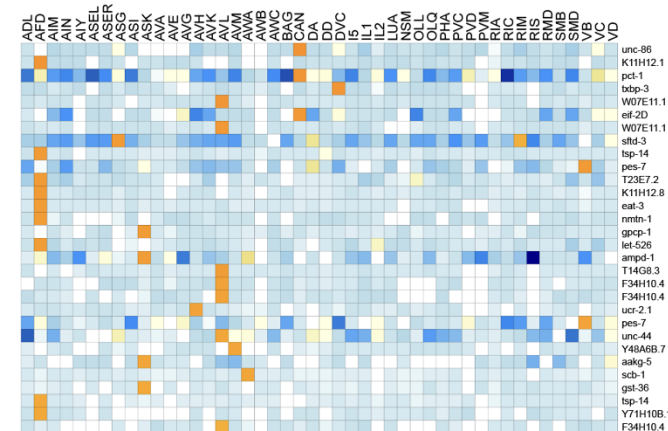
**Figure S1: Alternative splicing across neuronal cell types revealed by deep transcriptomes.** Heatmaps, as in Figure 1E, showing number of differential alternative splicing events across all pairwise comparisons for all classes of alternative splicing detectable by JUM (alternative 3' splice site, 5' splice site, cassette exons, intron retention, mutually exclusive exons, and composite splicing).

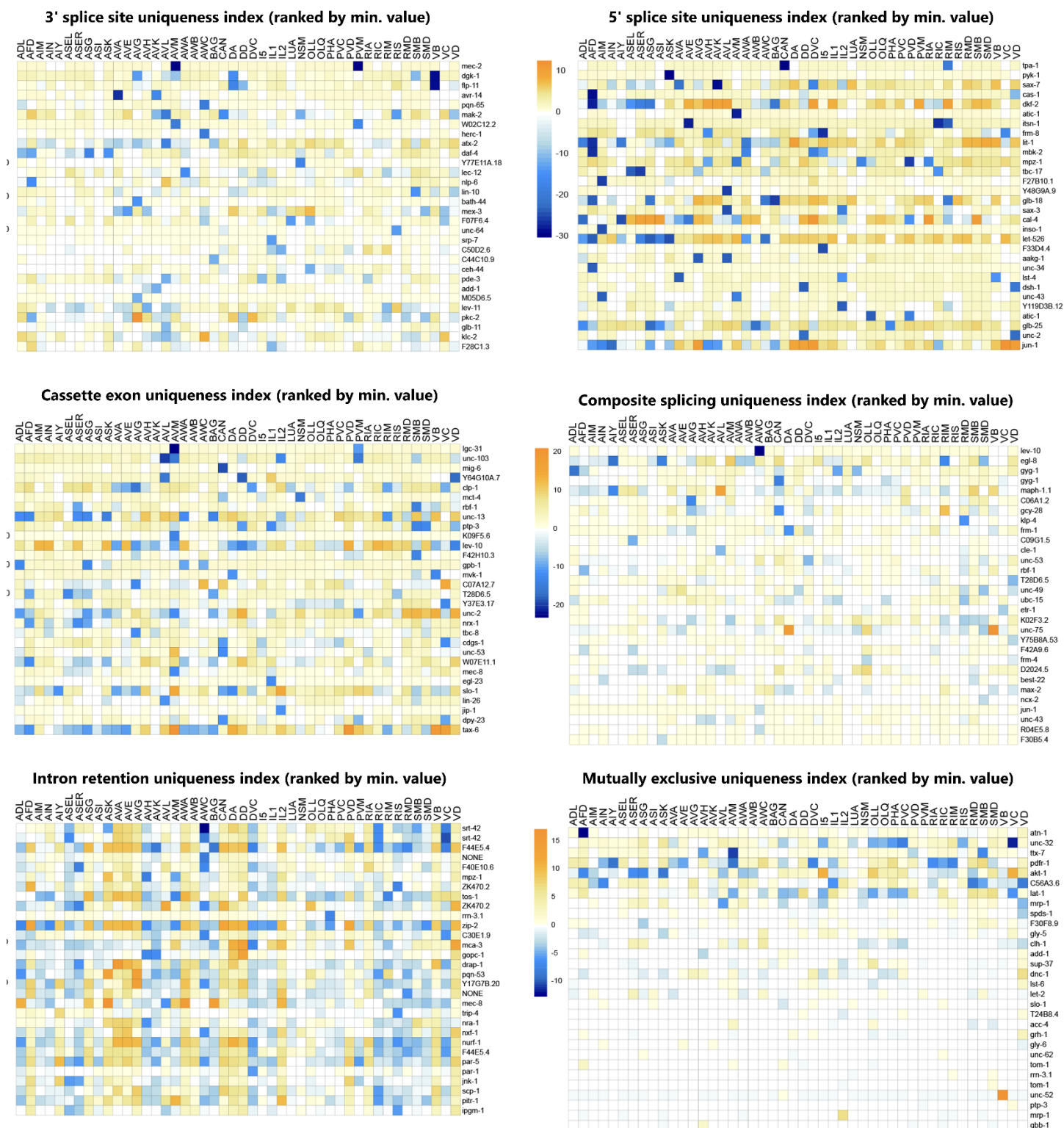




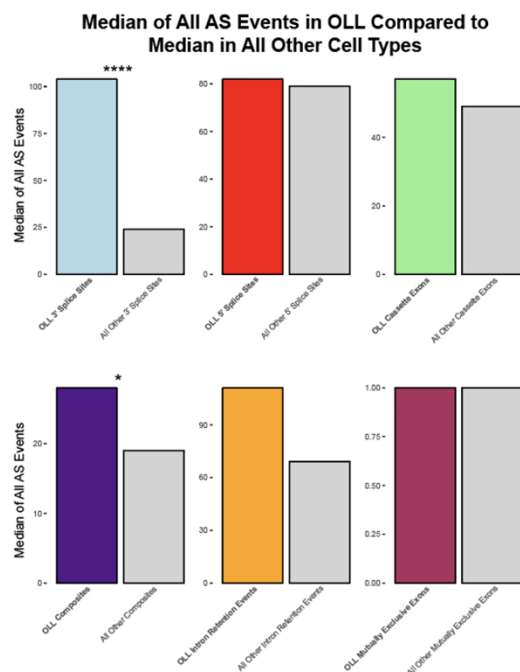
**Figure S2: Intron retention of *pct-1* in CAN neuron is not a consequence of strain-specific mutations.**

Visualization of sequencing reads via IGV showing no mutations in either CAN or ASK neuron RNA Seq (lower tracks) compared to each other, or to reference sequence (upper track). Bottom panel is a sequence alignment between the annotated nucleotides and intron-retained sequenced nucleotides showing no mutations observed.

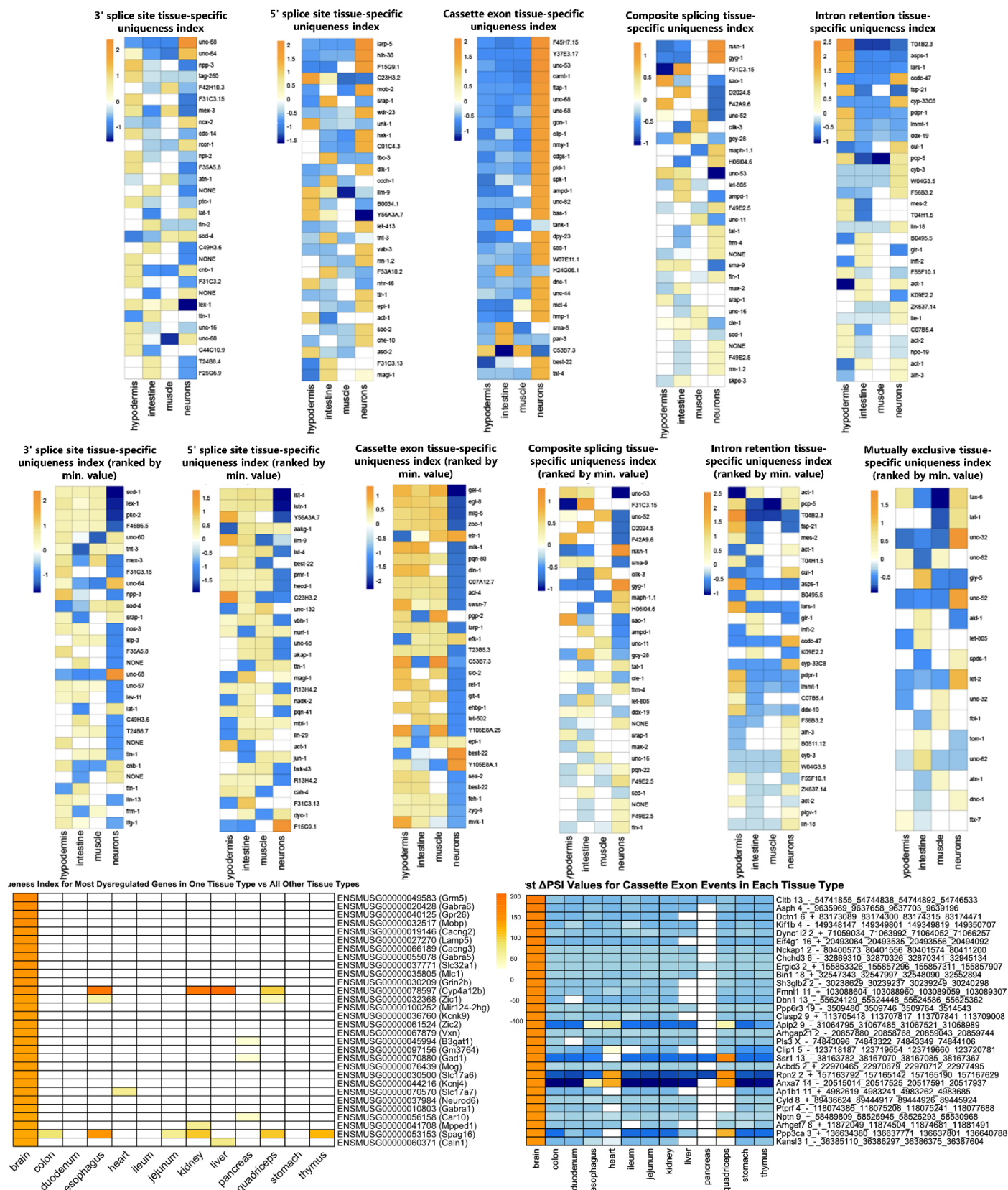




**Figure S3: Uniqueness index values for all alternative splicing types.** Top 30 uniqueness values, upper 6 panels are sorted for highest positive values (corresponding to high PSIs and/or upstream splice site selection), lower 6 panels for most negative values (corresponding to low PSIs and/or downstream splice site selection).



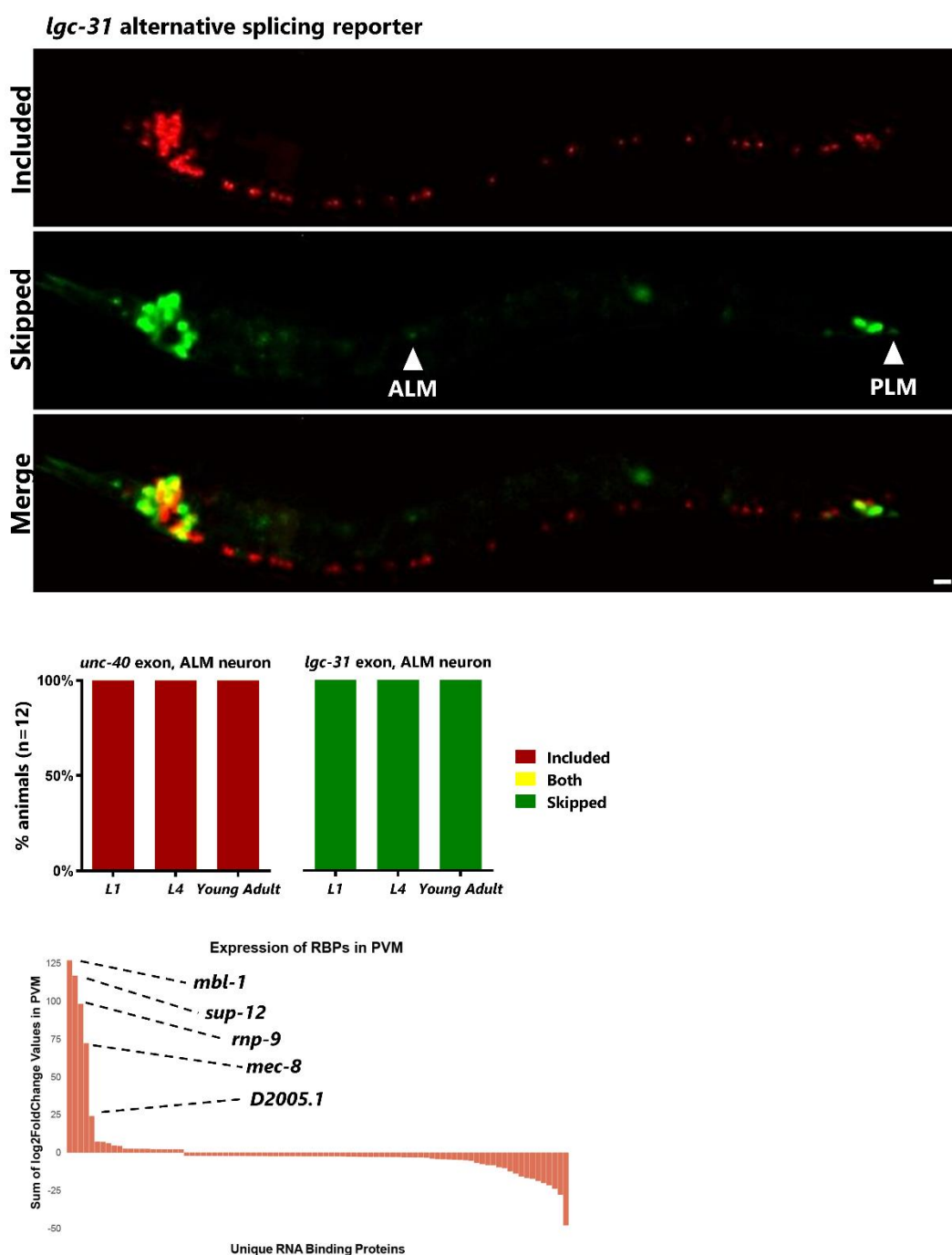
**Figure S4: OLL is uniquely enriched for differential 3' splice site selection.** Median values of number of differential alternative splicing events between OLL and all other neurons (left, colored bars) versus all other neuron pairwise comparisons (right, gray bars). OLL has much more alternative 3' splice site selection compared to other neurons, but this is not the case for other types of alternative splicing. \*\*\* = t-test,  $p < 0.001$ .



**Figure S5: Uniqueness index values for tissue-specific alternative splicing in worm and mouse.** Upper panels, Heatmaps as in Figure 5B but expanded to top 30 splicing events. Upper panels are sorted for highest positive values (corresponding to high PSIs and/or upstream splice site selection), lower panels for most negative values (corresponding to low PSIs and/or downstream splice site selection). Mutually exclusive exons are presented only once because, due to their small number, all of the top values (both positive and negative)



appear on a single heatmap. Lower panels, uniqueness index values for tissue-specific mouse data. Left side, brain-specific gene expression profiles. Right side, brain-specific cassette exons uniquely included in brain compared to all other tissues.



**Figures S6: Unique alternative splicing and RNA Binding Protein expression in touch neurons.** Upper panel, *lgc-31* splicing reporter, as in 6C for the *unc-40* splicing reporter, shows alternative splicing unique to touch neurons (exon skipping, GFP) compared to many other neurons (exon inclusion, RFP). Scale bar

represents 10  $\mu\text{m}$ . Middle panel, splicing phenotypes for *unc-40* and *lgc-31* cassette exons are invariant across individuals and across developmental stages (ALM neuron selected due to ease of unambiguous identification). Lower panel, RNA binding protein uniqueness expression levels, as in Figure 6D for AVM, but shown here for PVM.