nature communications

Article

https://doi.org/10.1038/s41467-025-58293-5

Alternative splicing across the *C. elegans* nervous system

Received: 16 May 2024

Accepted: 18 March 2025

Published online: 16 May 2025

Check for updates

Alexis Weinreb[1,2], Erdem Varol[3], Alec Barrett[1,2], Rebecca M. McWhirter[4], Seth R. Taylor[4,6], Isabel Courtney[4], Manasa Basavaraju[1,2], Abigail Poff[4], John A. Tipps[4], Becca Collings[4], The CeNGEN Consortium*, Smita Krishnaswamy[1,5] ✉, David M. Miller III[4] ✉ & Marc Hammarlund[1,2] ✉

Alternative splicing is a key mechanism that shapes transcriptomes, helping to define neuronal identity and modulate function. Here, we present an atlas of alternative splicing across the nervous system of *Caenorhabditis elegans*. Our analysis identifies novel alternative splicing in key neuronal genes such as *unc-40*/DCC and *sax-3*/ROBO. Globally, we delineate patterns of differential alternative splicing in almost 2000 genes, and estimate that a quarter of neuronal genes undergo differential splicing. We introduce a web interface for examination of splicing patterns across neuron types. We explore the relationship between neuron type and splicing, and between splicing and differential gene expression. We identify RNA features that correlate with differential alternative splicing and describe the enrichment of microexons. Finally, we compute a splicing regulatory network that can be used to generate hypotheses on the regulation and targets of alternative splicing in neurons.

Differential alternative splicing is a fundamental mechanism that increases molecular diversity. Splicing involves processing of pre-mRNAs by the spliceosome, resulting in removal of intronic sequences. Most metazoan genes undergo splicing, and splicing is critical not only for producing mature mRNAs but also for nuclear export and therefore translation. Alternative splicing (AS) occurs when a pre-mRNA is processed in more than one way, resulting in removal of different introns and the consequent production of mature mRNAs with different sequences. AS can alter the mRNA coding potential, resulting in expression of different protein isoforms. AS can also affect the stability and other features of the mRNA. The majority of human genes undergo AS[1,2], and defects in AS have been linked to disease[3].

Differential AS occurs when AS is regulated spatially or temporally, so that different cells express separate isoforms (differential AS is often referred to simply as 'AS'; here, the terms are distinct). In the nervous system, where it is most prevalent[4,5], differential AS controls multiple aspects of neuron identity[6], including global AS switches during development[7], isoform differences in neuron type specification[8,9], axon specification, guidance, and synaptogenesis[10,11], and has been linked to brain disorders[3,12,13]. Differential AS has also been studied in cancer; indeed, some forms of cancer are dependent on differential AS[14]. AS, because it is not cell-specific, can be identified by methods such as sequencing bulk cDNA. By contrast, identification of differential AS requires comparison of splicing between different samples, and thus has been less well characterized. One theme that has emerged is that differential AS is often not all-or-none, but rather is characterized by different ratios of splice isoforms across time or space.

How is splicing regulated? For AS, cis-acting factors—sequences or structures within the pre-mRNA—can regulate splicing diversity. The logic of cis regulation has been analyzed to identify features that give rise to AS[15–19]. But for *differential* AS, features within the nascent transcript are not sufficient, since the transcript is the same in all cell types. There must also exist trans-acting factors that regulate AS in a cell type-

[1]Department of Genetics, Yale University School of Medicine, New Haven, CT, USA. [2]Department of Neuroscience, Yale University School of Medicine, New Haven, CT, USA. [3]Tandon School of Engineering, New York University, New York, NY, USA. [4]Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN, USA. [5]Department of Computer Science, Yale University, New Haven, CT, USA. [6]Present address: Department of Cell Biology and Physiology, BYU, Provo, UT, USA. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: smita.krishnaswamy@yale.edu; david.miller@vanderbilt.edu; marc.hammarlund@yale.edu

specific manner, for example by interacting with the spliceosome to promote or inhibit particular splicing events[3]. Several approaches have been developed to establish splicing regulatory networks[5,20–25], or to integrate *trans* features within a framework developed on sequence motifs[26–29]. Nevertheless, our understanding of the 'splicing code'—the regulatory framework involving cis and trans elements that determines differential AS across all transcripts—is incomplete.

The nematode *Caenorhabditis elegans* is a powerful model organism for studies of the nervous system including AS[30]. Although gene expression atlases of developing and aging *C. elegans* cells are now available[31–38], less has been done to systematically establish AS patterns[39–41].

Here, we analyze data generated by the CeNGEN project to produce an atlas of AS for 55 single neuron types in *C. elegans*. We develop analytical tools to make the data available to the research community. We study differential AS between neuron types and show that key neuronal genes display broad patterns of differential AS. Focusing on canonical AS events, we establish overall patterns of differential splicing. Finally, we develop a principled computational approach to extract a regulatory network for differential AS, and use the network to identify candidate factors that regulate differential alternative splicing.

Results

Visualization and analysis of alternative exon usage

The CeNGEN project generated a data set covering 55 individual neuron types suitable for splicing analysis. As described previously[42,43], a series of *C. elegans* strains were used, each with specific promoters that uniquely label individual neuron types. For each neuron type, neurons were recovered by Fluorescence Activated Cell Sorting (FACS) from L4 hermaphrodites, with multiple independent biological replicates (average 3.8 replicate per neuron type, Supp. Data 1). Libraries were prepared using an optimized ribodepletion protocol[43] and sequenced on the Illumina platform, an approach that yielded robust coverage across the gene body (Fig. 1A). Our recent analysis distinguished 128 neuron types in the *C. elegans* nervous system each defined by its unique transcriptome[34]. Here, we isolated 55 neuron types, or 43% of known neuron types. To analyze alternative splicing (AS) in this data set we used three parallel approaches: (1) raw data visualization, (2) local quantification, and (3) transcript-level quantification. We illustrate our approaches on the gene *ric-4*, the homolog of the SNARE protein SNAP25, which displays an alternative first exon expressed differentially between neuron types[44,45]. All approaches are available online at www.splicing.cengen.org.

Raw data visualization to detect alternative splicing

Raw data visualization is a direct approach for splicing analysis that depends on displaying raw read counts in a genome browser. Direct visualization allows inspection of exon and splice usage in the full context of all the data for that gene. Raw data visualization does not use a statistical model and can be applied to individual biological replicates, to pooled data for each individual neuron type, or to all samples grouped together. Our browser is based on JBrowse2[46]. For each individual biological replicate, we generated a pair of browser tracks. The tracks underwent minimal filtering for clarity (see Methods).

First, a density plot indicates the number of reads aligned at a particular genomic position (normalized by the total number of reads in that sample and multiplied by one million, yielding Counts Per Million). Second, a splice junction track indicates the number of junction-spanning reads supporting that junction, without any normalization. We computed similar tracks for each neuron type, using all biological replicates for that type. Here, the density histogram represents the mean coverage across replicates at each genomic position (for each base pair). In addition, the junction-spanning reads (see Methods) are summed for each junction, to give a total junction usage track for that neuron. Finally, to allow rapid examination of a genomic locus across many neurons, we generated an additional set of six "global" tracks: the mean coverage (for each genomic position) across all neuron types, the minimum and maximum coverage at each genomic position across all neuron types, and the sum, minimum, and maximum of junction-spanning reads for each splice junction across neuron types. The mean exonic coverage and sum of splice junction tracks enable convenient visualization of an "average" transcript across neuron types. The minimum and maximum tracks facilitate the identification of rare transcripts: if a single neuron type expresses a given exon, it will be apparent in the maximum coverage track; if a single neuron does not express a given exon, it will not appear in the minimum track (and similarly for splice junctions).

For example, for the gene *ric-4* (Fig. 1B), we apply raw data visualization to the neuron types NSM and PVM. Read coverage shows that the distal first exon (blue box), corresponding to transcript *ric-4a*, is preferentially expressed in NSM. PVM displays weaker preferential usage of the proximal first exon (orange box), corresponding to transcript *ric-4b*. In addition, NSM displays 292 junction-spanning reads connecting the distal first exon, and only 18 junction-spanning reads connecting the proximal first exon. By contrast, in PVM there are 124 junction-spanning reads connecting the distal first exon, and 283 junction-spanning reads connecting the proximal first exon. This analysis indicates that *ric-4* undergoes differential alternative splicing in NSM vs PVM.

Quantification of alternative splicing

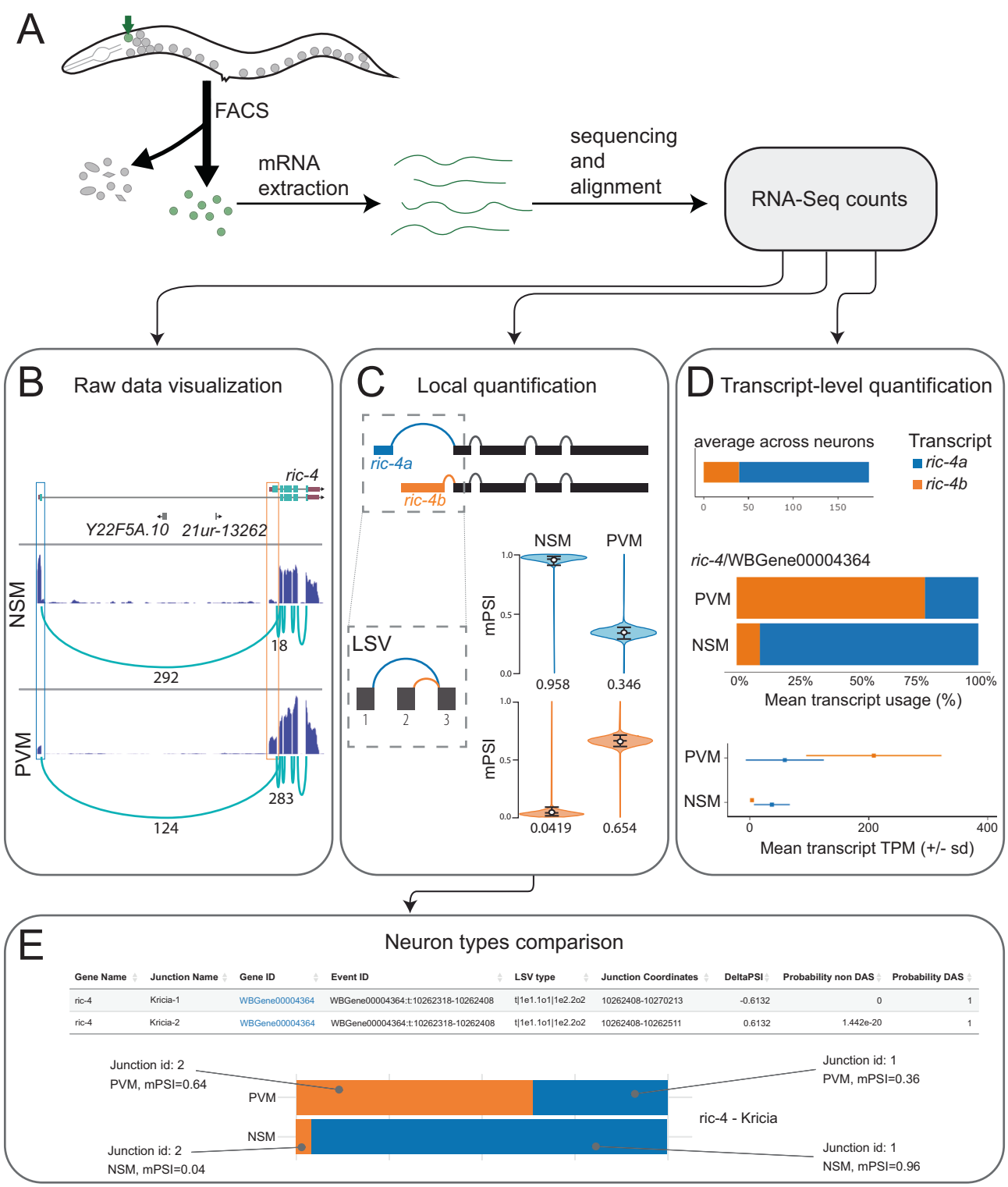
Although inspecting raw data on the genome browser is useful for visualizing differential AS at the single-gene level, additional methods are needed for genome-wide analysis. For this purpose, we quantified splice junction usage with the software package MAJIQ[47]. MAJIQ defines a Local Splicing Variation (LSV) as a set of splice junctions (SJ) starting from the same source exon or ending in the same target exon. For each LSV, the relative usage of each possible SJ is quantified, and a MAJIQ-defined Percent Selected Index (mPSI) is estimated from a Bayesian model. For example, an exon skipping/inclusion event (known as a 'cassette exon') is represented by MAJIQ as two LSVs: one upstream LSV, containing two splice junctions (one SJ that links the upstream exon to the cassette exon, and one SJ that skips the cassette to link to the downstream exon), and a second LSV, also with two SJs (one SJ from the upstream exon into the downstream exon, the other SJ from the cassette exon into the downstream exon). Similarly, an alternative first exon is represented in MAJIQ by a single LSV, immediately downstream of the alternative exons. Quantitative data generated using MAJIQ can be represented using VOILA[47], which displays the mPSI of junctions belonging to an LSV using violin plots.

In the case of *ric-4*, the alternative first exon is quantified as a single LSV containing two splice junctions, and quantification demonstrates the preferential use of the *ric-4a* exon in NSM and the *ric-4b* exon in PVM (Fig. 1C). We use the quantitative data generated by MAJIQ to analyze global splicing patterns across genes and neuron types in the following sections.

Besides quantifying individual splicing events, it is also useful to visualize alternative events in the context of complete transcripts. To address this question, we used the software package StringTie in quantification mode to analyze transcript levels in individual neuron types[48]. Given a set of annotated transcripts (we used all transcripts annotated in WormBase), StringTie uses a maximum flow computational approach to estimate the expression level of each transcript. Thus, StringTie output represents not only the relative abundance of the alternative transcripts of a gene, but also the measured level of each transcript in each neuron type (in Transcript Per Million or TPM units).

For example, in the case of *ric-4* (Fig. 1D), StringTie analysis compares the levels of the transcripts *ric-4a* and *ric-4b* in NSM and

https://doi.org/10.1038/s41467-025-58293-5



Gene Name	Junction Name	Gene ID	Event ID	LSV type	Junction Coordinates	DeltaPSI	Probability non DAS	Probability DAS
ric-4	Kricia-1	WBGene00004364	WBGene00004364:t:10262318-10262408	t|1e1.1o1|1e2.2o2	10262408-10270213	-0.6132	0	1
ric-4	Kricia-2	WBGene00004364	WBGene00004364:t:10262318-10262408	t|1e1.1o1|1e2.2o2	10262408-10262511	0.6132	1.442e-20	1

Fig. 1 | Overview of data collection and splicing analysis. A Schematic of the experimental procedure. **B**–**E** Four methods to analyze alternative splicing, applied to the gene *ric*-4/SNAP25 in the neurons NSM and PVM. **B** Raw data visualization. Top track: Gene model of *ric-4* (along with non-coding RNAs *21ur-13262* and *Y22F5A.10*). Bottom tracks: Read coverage and junction counts. Numbers denote junction-spanning reads for the splice junctions of interest. Blue and orange boxes indicate alternative first exons. **C** Local Splicing Variation (LSV) visualization. Top: Gene model of *ric-4*, highlighting the alternative first exons corresponding to *ric-4a* (blue) and *ric-4b* (orange). Bottom: Posterior mPSI (MAJIQ-defined Percent Selected Index) estimates displayed as violin plots. Numbers indicate the posterior expected mPSI (summing to one in each neuron, as the orange and blue junctions are mutually exclusive). **D** Transcript-level quantification. Top: Average transcript TPM across all sequenced neurons. Middle: Relative transcript usage (in proportion to the total TPM for the gene, in each neuron), in NSM and PVM. Bottom: Transcript quantification, indicating the mean +/- standard deviation of TPM across samples for each transcript in each neuron. **E** Neuron-wise comparison between PVM and NSM. Top: The application returns a list of splice junctions with differential usage between the two neurons, including two junctions belonging to the same LSV in the gene *ric-4*. Bottom: Visualization of junction usage for this LSV in neurons PVM and NSM.

PVM. First, the transcript *ric-4a* is more common than *ric-4b* when averaging across all neurons (*ric-4a* is detected at 134 TPM, *ric-4b* at 47 TPM on average). Second, examining the relative transcript usage in individual neurons, 90% of the *ric-4* expression in NSM is attributed to *ric-4a*, whereas 80% of the *ric-4* expression in PVM is attributed to *ric-4b*. Comparing the total transcript expression values, *ric-4a* is expressed at 38 + /− 31 TPM in NSM and 60 + /− 67 TPM in PVM, whereas *ric-4b* is expressed at 4 + /− 0.9 TPM in NSM and 212 + /− 116 TPM in PVM, where the interval given corresponds to the standard deviation across biological replicates from the same neuron type.

Finally, a common research need is to determine genes that differ in their splicing patterns between two neurons or sets of neurons. We thus provide an additional tool at www.splicing.cengen.org to interrogate the quantifications performed by MAJIQ. For example, comparing the PVM and NSM neurons yields 49 splice junctions belonging to 27 LSVs in 22 genes with differential splicing. Among these differentially spliced junctions are the two junctions corresponding to the alternative first exon of *ric-4*, identifying this gene as a good candidate for further investigation with the event-centric tools described above (Fig. 1E). This approach provides an important investigative method to select splicing events potentially relevant to a known phenotype.

Comparison to known instances of alternative splicing

Next, we asked if our analysis aligns with existing data on alternative exon usage in *C. elegans* neurons. A two-color splicing reporter previously indicated that *elp-1*/EMAP undergoes differential alternative splicing, with exon 5 skipped in touch neurons[49]. Indeed, our data confirms exon 5 skipping in the AVM touch neurons (Fig. 2A). In another case, a cassette exon (11.5) in *daf-2*/IGFR was previously reported to undergo differential alternative splicing in many neuron classes[50] (Fig. 2B). Using local quantification (Fig. 2C), we also find differential alternative splicing at exon 11.5, with the splicing patterns we observe in good agreement with previous results. A similar pattern is seen by visual exploration of the raw data (Fig. 2D, red box). It is interesting to observe that despite relatively clear data for exon 11.5, our transcript-level analysis shows that the transcript known to contain this exon (*daf-2c)* is only modestly enriched in individual neuron types, likely owing to the relatively large number of alternative transcripts of the *daf-2* gene[44] (Fig. 2E). Together, these results indicate that our data are consistent with existing in vivo observations at single-neuron type resolution.

Web tools for splicing analysis

To facilitate use by the scientific community, these data and analytical methods are available via a web portal at www.splicing.cengen.org. For raw data visualization, the user can select a gene or genomic region, and also choose the data to display: all individual samples are available, as well as the averaged data for each neuron type and the global data showing mean, maximum and minimum. For each data set displayed, the user can select whether to display the read counts, the exon-spanning reads, or both. In addition, due to our use of interoperable formats, our tracks can be imported to other genome browsers (such as WormBase or UCSC), and tracks generated by other projects can be displayed in our genome browser, allowing simultaneous examination of data from separate sources. For local quantification using MAJIQ, we display the results using VOILA. Finally, for transcript-level quantification, we developed a custom web application to display the results. For a single gene, the application displays both relative transcript usage and absolute transcript expression for all annotated transcripts. Multiple genes can be represented as a heatmap of transcript expression. These three tools offer users complementary levels of interpretation: quantification of transcript usage reflects the underlying biology of RNA processing. However, expression levels of complete transcripts are difficult to infer from short-read data and may be inaccurate (see Discussion). By contrast, local LSV-level quantification more directly reflects our measurements and is thus likely more accurate. For complex alternative transcripts, interpretation can be complicated by the need to consider multiple splice junctions simultaneously. Finally, the browser view does not offer rigorous quantification but can be used to examine the full context of a genomic region, including constitutive exons and non-coding RNAs.

Axon guidance receptor gene *unc-40/DCC* is differentially spliced in specific neurons

Most gene models of splicing in *C. elegans* were obtained from sequencing bulk samples. However, if differential alternative splicing occurs in only a small number of cells, these rare splicing patterns might not be detected. To test this idea, we used manual inspection of raw data visualizations to examine well-studied neuronal genes.

Although a single transcript is annotated for the gene *unc-40*/DCC[44], our analysis detected two novel exons, exon 8.5 and exon 14.5 (Fig. 3A). In particular, exon 14.5 is preferentially included in AVM, whereas other neuron types (e.g. AVL, AWA) exclusively express the canonical splice variant, skipping exon 14.5 (Fig. 3B). Using RT-PCR, we validated the presence of exon 14.5-including transcripts in cDNA extracted from whole animals, as well as the conventional skipped transcript (Fig. 3C). Interestingly, the additional exons (8.5, 14.5) do not disrupt the open reading frame, and lead to insertions between known domains of the UNC-40 protein (Fig. 3D). To determine whether the inclusion of exon 14.5 represents a transcript unique to *C. elegans*, we examined the locus of the orthologs of *unc-40* in the closely related nematodes *Caenorhabditis briggsae* and *Caenorhabditis brenneri* (Fig. 3E), and examined bulk RNA-Seq data for those species available from Wormbase[44]. We found that *C. briggsae* displays four annotated transcripts, with cassette exons corresponding to exons 8.5 and 14.5 of *C. elegans*. On the other hand, *C. brenneri* presents evidence of unannotated exons corresponding to exons 8.5 and 14.5. This finding suggests that other nematode species express a similar exon. The additional exons in *Cbr*-UNC-40 and *Cbn*-UNC-40 encode protein sequences with high identity with exons 8.5 and 14.5 of *Cel*-UNC-40 (Fig. 3F).

The expression of *unc-40* is necessary for ventral guidance of the AVM axon[51,52]. Mutation of *unc-40* leads to aberrant anterior growth of the AVM axon with a penetrance of 20-40%. To assess whether the additional exon 14.5 is necessary for the function of UNC-40 in AVM, we performed a CRISPR deletion of this exon (Figure S1A), directly joining exons 14 and 15. We did not observe AVM guidance defects following excision of exon 14.5 (Figure S1B). It is possible that this exon plays a role in a function other than axon guidance in the AVM neuron.

Thus, *C. elegans unc-40* has previously unannotated alternative transcripts, with conserved sequence and potential functional impact. In general, our analysis of splicing in individual neuron types can identify novel mRNA sequences.

sax-3/Robo and the homeobox factor *ceh-8* have novel alternative first exons

We also identified novel splice variants in the Slit receptor gene *sax-3*/ROBO. The *sax-3* annotation shows two alternative transcripts, differing by 13 bp in exon 11 length (Fig. 4A) and 29 bp in the length of the annotated 5′ UTR (not shown in figure for clarity). We found a novel alternative splice site that shortens exon 9 by 15 bp. In addition, we detected a novel alternative first exon 5.5 (positioned between annotated exons 5 and 6). LSV quantification shows that the annotated alternative splice site in exon 11 is not used in the neurons sequenced here. By contrast, the novel alternative splice site in exon 9 and the alternative first exon 5.5 are both differentially expressed in broad subsets of neuron types in our data set (Fig. 4B, showing AVL and AVM as examples). We confirmed the in vivo expression of both alternative first exons by RT-PCR (Fig. 4C, D). Both of these novel events affect

https://doi.org/10.1038/s41467-025-58293-5

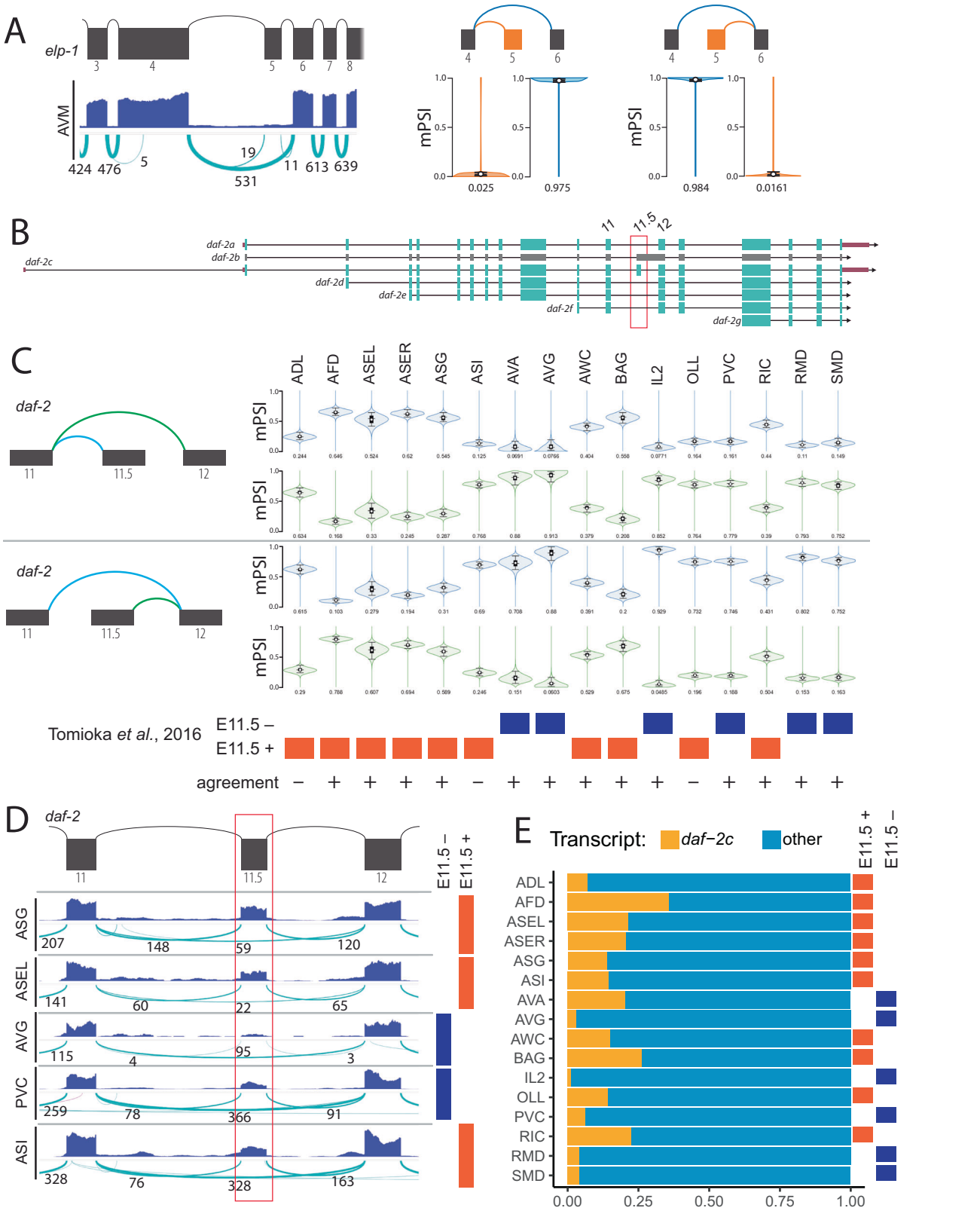


Fig. 2 | Comparison to previous splicing data. A Exon 5 skipping in *elp-1* in AVM neuron. Left: Raw data visualization in the neuron AVM. Right: Local quantification of the two LSVs corresponding to the cassette exon. The violin plots represent the posterior distribution computed by MAJIQ. **B**–**D** Cassette exon 11.5 in *daf-2* in individual neuron types. **B** *daf-2* gene structure. The cassette exon 11.5 (red box) is unique to the transcript *daf-2c*. **C** Local quantification of the upstream (top) and downstream (middle) events flanking the cassette exon 11.5. Left: Schematic representation of splice junctions constituting the local event. Right: Splice junction quantifications in 16 neuron types. Bottom: Inclusion pattern of exon 11.5 from Tomioka et al.[50], and agreement with our data. **D** Raw data visualization of exon 11.5 alternative splicing in 5 neuron types. Top: Gene model. Bottom: Exonic and junction-spanning counts. Right: Exon 11.5 inclusion pattern from Tomioka et al.[50] in the same neurons. **E** Transcript quantification of *daf-2* in the same neurons as (**C**). The transcripts *daf-2a*, *d*, *e*, and *f*, (other) which do not include exon 11.5, were grouped together. Right: Exon 11.5 inclusion pattern from Tomioka et al. [50].

https://doi.org/10.1038/s41467-025-58293-5

coding potential: the alternative splice site in exon 9 alters the amino acid sequence of the intracellular domain of SAX-3, whereas the alternative first exon 5.5 generates a short isoform (SAX-3S) lacking four of the five Ig domains, but encoding its own signal peptide in frame with the remainder of the protein (Fig. 4E). We examined the locus of the *C. briggsae* and *C. brenneri* orthologs in bulk RNA-Seq data, and find a remarkable conservation of the gene structure, including the novel exon 5.5 (Fig. 4F). These orthologous exons 5.5 encode an identical amino acid sequence (Fig. 4G).

Finally, we identified a novel alternative first exon in the homeobox transcription factor *ceh-8* (Figure S2A, B). Interestingly, this transcript would lead to translation of a protein with a truncated homeobox domain (Figure S2C). This finding is consistent with data from recent studies that also show that alternative splicing can lead to partial or complete loss of a DNA-binding domain[8,53]. The *ceh-8* locus is not well conserved in *C. briggsae* and *C. brenneri*, precluding direct comparison of alternative transcripts. Together, these examples demonstrate that our approach can detect novel alternative first

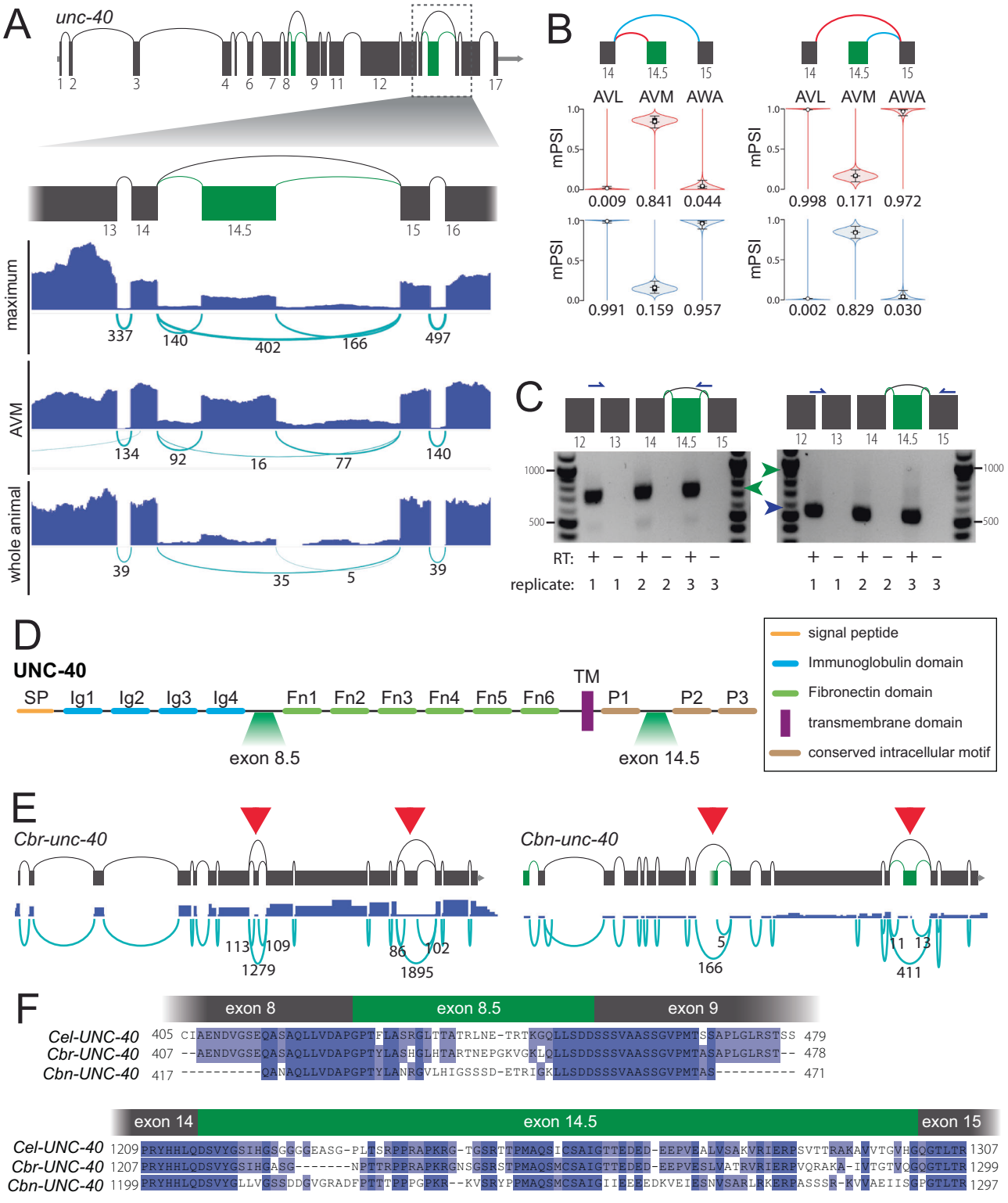


Fig. 3 | Detection of novel cassette exon in unc-40. A Alternative expression of a novel cassette exon in *unc-40*. Top: Schematic of the gene structure, showing the novel alternative exons 8.5 and 14.5 (green) and the annotated constitutive exons (gray). Bottom: Genome browser representation around exon 14.5, displaying the maximum track, the AVM track, and samples collected from whole animals. **B** Local quantification of exon 14.5 inclusion. The violin plots represent the posterior distribution computed by MAJIQ. **C** Validation by RT-PCR of the novel exon 14.5 in cDNA from whole animals (N2 wild type). Top: Location of the primers relative to the alternative exon 14.5. Left, the primer pair only results in amplification from a transcript with exon 14.5 included; right, the primer pair flanking exon 14.5 is able to amplify from both transcripts. Arrowheads indicate the expected product sizes, green arrowhead: transcript with exon 14.5 included (left: 754 bp; right: 837 bp); blue arrowhead: transcript with exon 14.5 skipped (582 bp). For each experiment, the amplification was performed on three biological replicates along with a control without reverse transcriptase (RT). Molecular weights reported in base-pairs. Source data are provided as a Source Data file. **D** Protein structure of UNC-40, indicating the cassette exons positions vs known protein domains. **E** Structure of the *unc-40* orthologous genes in *Caenorhabditis briggsae* (*Cbr-unc-40*) and *C. brenneri* (*Cbn-unc-40*). Top: Annotated gene models (unannotated features in green). Bottom: RNA-Seq data. Red arrowheads denote orthologs of exon 8.5 and 14.5. **F** Alignment of the protein sequences of *C. elegans* UNC-40, *C. briggsae* UNC-40, and *C. brenneri* UNC-40 around exon 8.5 orthologs (top) and around exon 14.5 orthologs (bottom).

exons. Such events also contribute to transcript diversity, but are generated by different biological mechanisms than other forms of alternative splicing. While most alternative splicing is performed by the spliceosome, alternative first exons are the result of alternative promoter usage.

Global detection of novel splicing events across neuron types

Given these examples of novel isoform detection, we sought to identify candidate novel splice junctions across our data set. We used STAR to generate a preliminary list of 1,026,619 unannotated splice junctions[54]. We filtered these junctions using multiple criteria to focus on well-supported novel junctions, for example by requiring high expression relative to neighbor genes (see Methods). We also leveraged the biological replicates in our data set, requiring novel splice junctions to be present in at least half the samples from a single neuron type, minimum of two. This analysis yielded 1722 novel junctions (Supp. Data 2). Attaching novel junctions to annotated genes is a transcript discovery task which is challenging using short reads, and current tools do not reach high accuracy[48]. However, each novel junction must belong to a gene in its immediate neighborhood; we provide a list of all neighbor genes (Supp. Data 2). We also sought to estimate the number of genes containing novel splicing events, without precise knowledge of which gene each junction belongs to. As multiple junctions may belong to the same novel transcript, we expect the number of affected genes to be less than the number of novel splice junctions. We thus estimated the number of genes containing novel splicing events under three assumptions: (1) minimizing the number of novel junction-containing genes that could explain the observed pattern, (2) maximizing this number, or (3) estimating an average number of novel junction-containing genes by selecting random samples, under the assumption that the novel junctions are uniformly distributed among the genes. Following this procedure, we estimate that approximately 1361 genes (at least 1353 and at most 1368 genes) contain novel junctions detected in our data. Overall, this analysis indicates that many novel splice sites and mRNA isoforms remain to be described, and provides a list of reliable candidates for future study.

Detection of differential AS between neuron types identifies genes associated with neuronal excitability

Our analysis of known differential alternative splicing events (*ric-4, elp-1, daf-2*; Figs. 1, 2) and identification of novel events (*unc-40, sax-3, ceh-8*; Figs. 3, 4) demonstrate that our data set can be used to identify instances of differential alternative splicing (DAS) in the *C. elegans* nervous system. To identify candidate DAS events across all genes and neuron types, we quantified differential event usage using MAJIQ[47]. Specifically, within its Local Splicing Variation framework, MAJIQ models the relative usage of splice junctions that share a common source or target exon. Across the 55 neuron types in our data set, we detected 1940 genes displaying DAS. To validate these findings, we compared the genes detected here to a list of 542 genes compiled from the literature (Supp. Data 3, see Methods) and found a large overlap of 461 genes (85%; Fig. 5A). This comparison indicates that the novel instances of DAS we detect are indeed strong candidates.

Next, we sought to explore the function of candidate DAS genes. Gene Ontology analysis[55] showed enrichment of terms related to neuronal function (Figure S3A). To investigate the specific role of DAS, we examined major neuronal functional gene classes[56]. We found that the prevalence of DAS is highly variable by gene class. For example, the majority of potassium channels and voltage-gated calcium channels undergo differential alternative splicing, whereas ribosome subunits and neuropeptide-encoding genes tend to be similarly spliced across neuron types (Fig. 5B). This analysis suggests that DAS is preferentially used to fine tune neuronal excitability, enhancing functional diversity among different neuronal types.

Global patterns and prevalence of DAS

With a list of DAS events in hand (Fig. 5A), we examined global patterns of DAS across neuron types. We performed pairwise comparisons of all neuron types in our data set, and assessed the proportion of DAS genes among the genes co-expressed in both neurons of the pair (Fig. 5C, full table in Source Data file). Clustering this data revealed that that a group of 10 ciliated sensory neurons (ASK, ADF, ASG, ASEL, ASER, AFD, BAG, AWA, AIY, AWB) have similar global patterns of AS. In addition, some pairs of neurons with similar functions have similar splicing profiles (DA-VB, DD-VD, AVM-PVM, AVH-AVK) (Figure S3B). Interestingly, another group of 6 neurons comprising I5, LUA, OLQ, OLL, PHA, and PVC also display similar AS profiles (Fig. 5C), even though they do not share known functional or morphological characteristics. These data suggest that at least in some cases, neurons with similar characteristics adopt similar patterns of DAS, presumably to support specific functional specialization.

Besides DAS, gene expression patterns are highly correlated with neuron type. In fact, gene expression patterns alone can be used to group single cells into clusters corresponding to individual neuron types[34]. Given the very strong association between gene expression and neuron type, we wondered if DAS patterns and gene expression patterns are related. In this case, both gene expression and alternative splicing might encode the same information reflecting the underlying structure of neuronal cell identities. To test this model, we compared (for each neuron pair) the most strongly differentially expressed (DE) genes to the most strongly DAS genes, and found limited overlap (Figure S3C). In addition, for each pair of neurons, we compared the number of DE genes to the number of DAS genes and found only a weak correlation (Fig. 5D). Together, this analysis indicates that DAS and gene expression are two largely independent dimensions of neuron type identity[57].

We then aimed to determine the number of DAS genes in the entire nervous system. As our dataset covers 55 neuron types out of 119 classes (118 canonical classes, plus ASEL/ASER), we used a downsampling approach to estimate the number of detected DAS genes depending on the number of neuron types sequenced (Fig. 5E). By projection, we estimate that 3192 genes are differentially alternatively spliced within the nervous system (see Methods), corresponding to about one quarter of all genes expressed in neurons.

https://doi.org/10.1038/s41467-025-58293-5

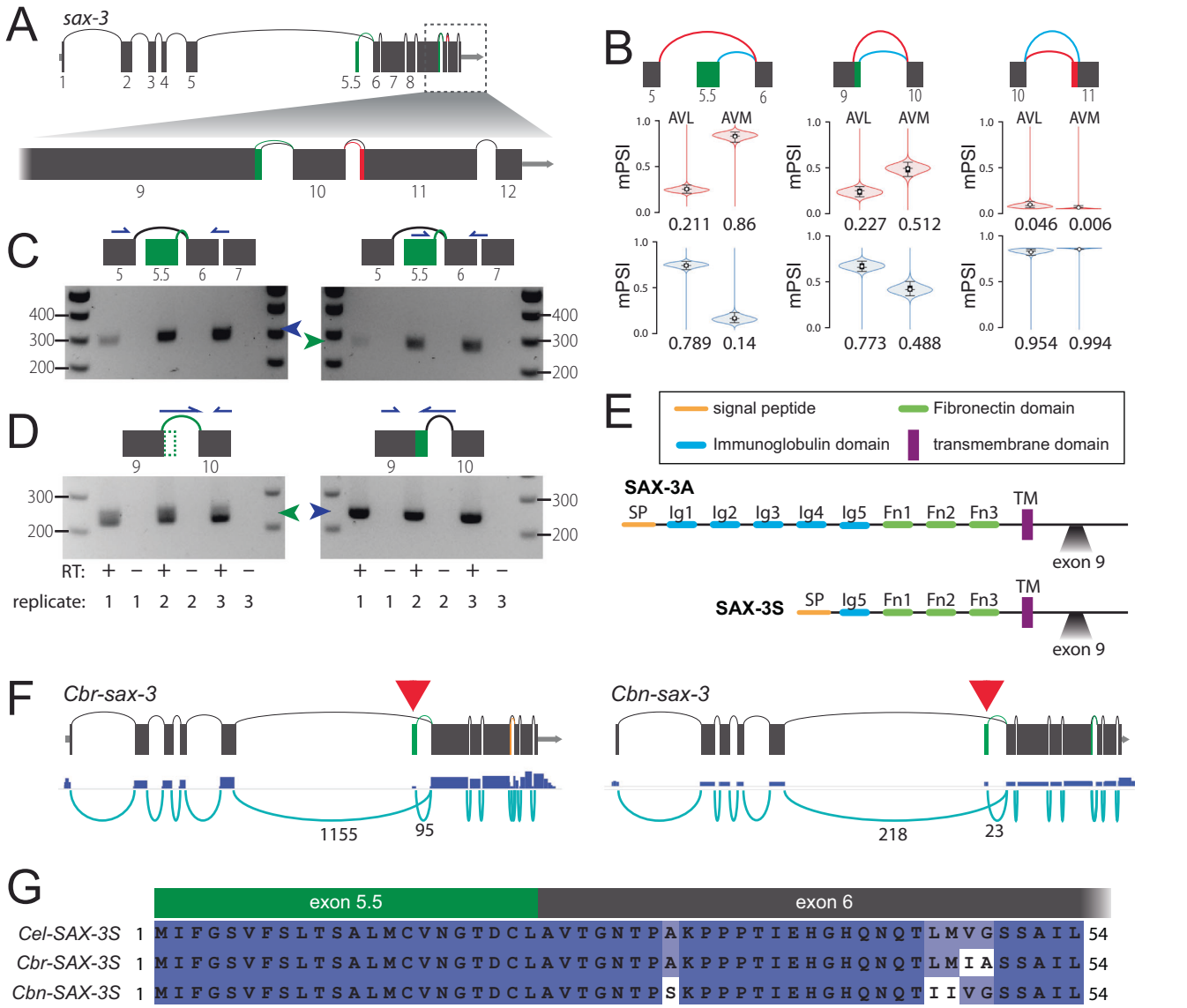


Fig. 4 | Detection of novel alternative first exon in sax-3. A Structure of the gene *sax-3*/ROBO indicating the novel exon 5.5 and alternative 5′ splice site at the end of exon 9 (green), and the annotated alternative 3′ splice site at the start of exon 11 (red). Bottom: Enlarged 3′ end of the locus. **B** Local quantification of LSVs in the representative neurons AVL and AVM for the novel alternative first exon 5.5 (left), the novel alternative splice site (middle), and the annotated alternative splice site (right). Violin plots represent the posterior distribution computed by MAJIQ. **C**, **D** Validation by RT-PCR of the novel exons in cDNA from whole animals (N2 wild type). Alternative first exon 5.5 (**C**), top: Location of primers relative to the exon. Left, the primer pair only amplifies the annotated transcript including exons 1-5; right, the primer pair only amplifies the novel transcript using exon 5.5 as first exon. Arrowheads: the expected product sizes (307 bp, 286 bp). Alternative 5′ splice site in exon 9 (**D**), top: Location of the primers relative to the event. Left, the primer pair only amplifies the novel transcript with shorter exon 9; right, the primer pair only amplifies the annotated transcript with longer exon 9. Arrowheads: expected product sizes (242, 248 bp). Three biological replicates along with a control without reverse transcriptase. Molecular weights reported in base-pairs. Source data are provided as a Source Data file. **E** Protein structure of SAX-3, indicating the impact of the novel splice variants. Top: Overall structure of annotated isoform SAX-3A. Bottom: Overall structure of the novel short isoform starting at novel exon 5.5 (SAX-3S). The position of the novel splice site in exon 9 is indicated. **F** Structure of the orthologs of *sax-3* in *C. briggsae* (*Cbr-sax-3*) and *C. brenneri* (*Cbn-sax-3*). Top: Gene models. Bottom: RNA-Seq data. Red arrowheads denote the orthologs of exon 5.5. **G** Protein sequence alignment of the short isoforms of SAX-3 orthologs from *C. elegans*, *C. briggsae*, and *C. brenneri*.

Sequence features of alternative splicing

Alternative splicing can take many forms, with implications for both regulatory mechanisms and biological effects. To assess the representation of different alternative splice types in the *C. elegans* nervous system, we grouped events into canonical event types (Fig. 6A). We found that alternative splice sites, cassette exons, and alternative first exons are well represented in the genome. By contrast, alternative last exons and coordinated multiple exon skipping are relatively rare. Using the software SUPPA to estimate a binary Percent Spliced-In (bPSI) for each event, we assessed the prevalence of differential alternative splicing within each event type and found that DAS is well-represented among all forms of alternative splicing. To compare the prevalence of neuron-specific versus tissue-specific event regulation, we retrieved published tissue-specific TRAP-Seq data[39] and processed it using our pipeline. This analysis showed a significantly smaller number of tissue-regulated compared to neuron-regulated events (Fig. 6A), which may reflect a biologically meaningful increase in the amount of AS in the nervous system[4,5], the higher statistical power resulting from the greater number of neuronal samples, or both (see Discussion). Our approach, which

https://doi.org/10.1038/s41467-025-58293-5

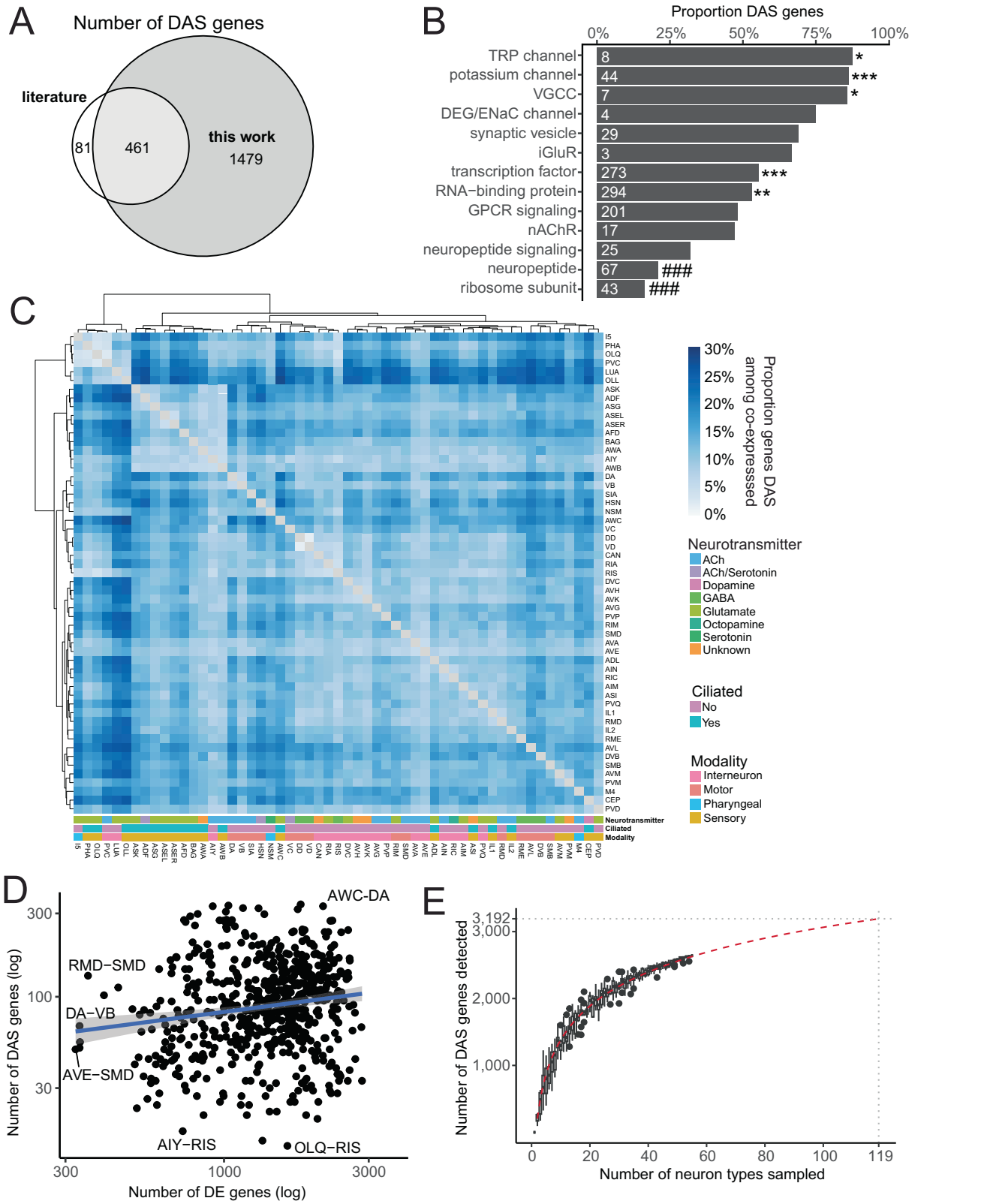
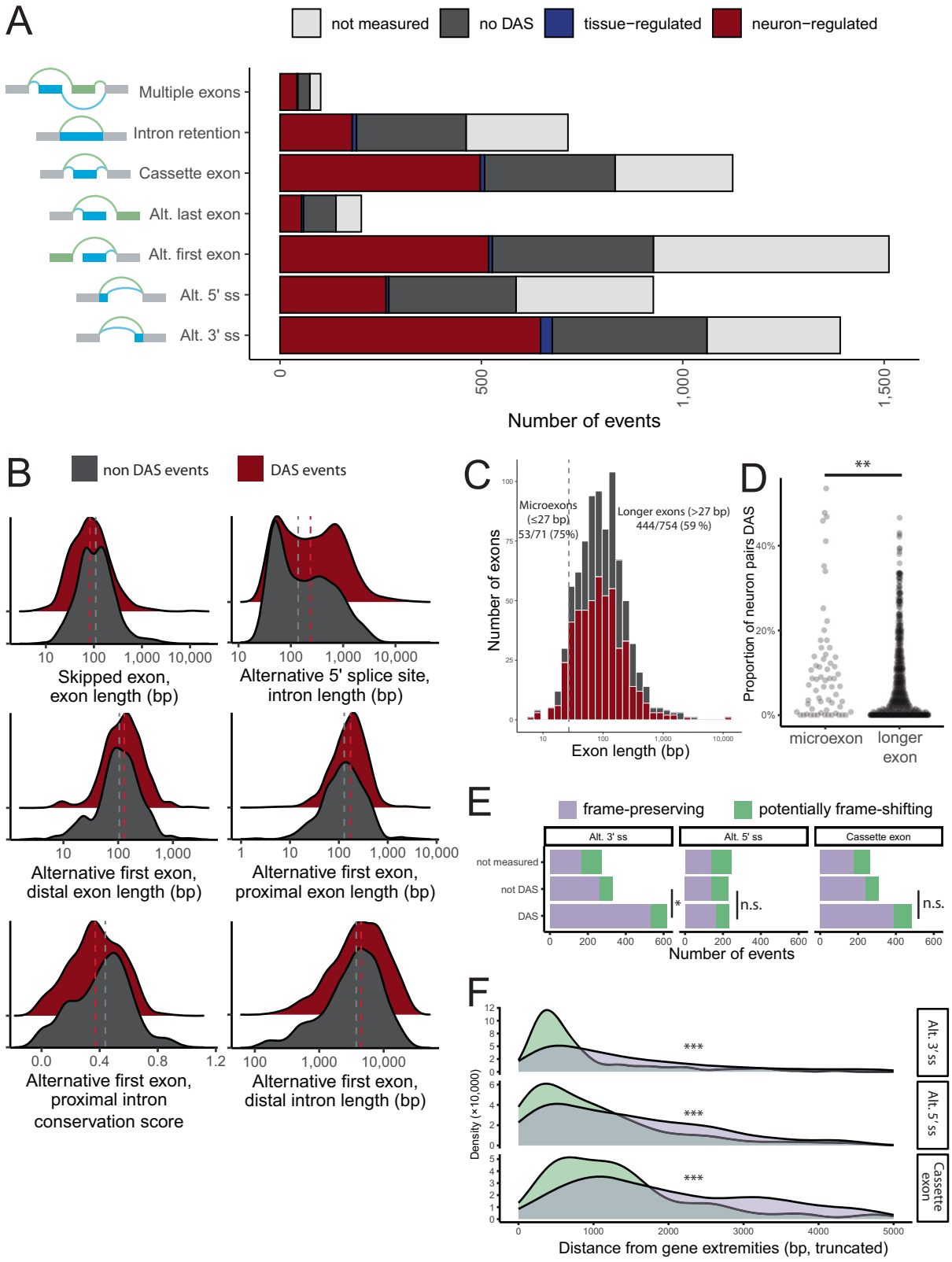


Fig. 5 | Differential AS between neuron types. A Euler plot of the number of genes presenting differential alternative splicing (DAS) in this analysis, and from a literature review (Supp. Data 3). Source data are provided as a Source Data file. **B** Proportion of genes DAS within families of neuronally significant genes[56]. The number of genes in the family is indicated inside each bar. Hypergeometric test: * Significantly enriched, # significantly depleted. One-sided hypergeometric test and Benjamini-Hochberg correction for multiple comparisons. Exact *p*-values reported in Source Data file. **C** Proportion of DAS genes per neuron pair, among genes co-expressed in the neurons of this pair. **D** Number of DAS genes plotted against number of differentially expressed genes for each neuron pair. The line corresponds to a linear fit, the shaded area to the 95% confidence interval. **E** Estimate of the number of genes in which DAS can be detected relative to the number of neuron types sequenced. Each subsampling was repeated 10 times; the box plots are centered on the median and between the first and third quartiles, the whiskers extend to the furthest value no further than 1.5 times the interquartile range. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-025-58293-5



relies on existing genome annotations, detected fewer instances of intron retention compared to an independent analysis of the same dataset[58]. As intron retention cannot be identified through junction-spanning reads and requires careful consideration, we do not further discuss this event type. Overall our analysis indicates that neurons use all available mechanisms to increase their molecular diversity.

Next, we explored the usage and specificity of event types and by analyzing the distribution of absolute differences in bPSI ($|\Delta bPSI|$) across neuron types (Figure S4A, full table in Source Data file). For all canonical event types, one-third to two-thirds of the events displayed a large ($|\Delta bPSI|>0.5$) change in at least one pair of neurons. The Gini index is a measure of specificity, reflecting the number of neuron pairs displaying a substantial $|\Delta bPSI|$. Most events with large max($|\Delta bPSI|$)

Fig. 6 | Landscape of alternative splicing event types. A Number of events by types. Bar length indicates the number of canonical events of each type predicted from genome annotation. Dark gray: events that did not display DAS between neuron types in our dataset; dark red: events DAS in at least one pair of neurons; dark blue: events DAS between tissues, based on published tissue-specific TRAP-Seq data, but not between individual neuron types in our data; light gray: events that could not be measured in either dataset. **B** Distribution (density plot) of length or conservation score of particular sequences within splicing events, for DAS events (red) vs non-DAS events (gray). Only features with a statistically significant difference between DAS and non-DAS events are represented here. Vertical dashed lines represent the median of each group. Lengths in basepairs (bp). **C** Histogram of cassette exon lengths, separating DAS exons (red) and non-DAS exons (gray). The vertical dashed bar at 27 bp delimits microexons. **D** For each exon (each dot), proportion of neuron pairs where the exon is differentially AS between the two neuron types (proportion calculated among the neuron pairs in which the exon-containing gene is expressed in both neurons of the pair). $p = 0.0035$ (two-sided Wilcoxon test). **E** Potential impact of nucleotides addition on open reading frame. For alternative splice sites and cassette exons overlapping a CDS, the color indicates whether inclusion of the alternative event results in a nucleotide addition that is a multiple of 3 (lilac) or is not a multiple of 3 (green), separating events DAS between neuron types. Alt. 3′ ss, $p = 0.023$; alt. 5′ ss, $p = 0.053$; cassette exon, $p = 0.41$ (Chi-square test, comparison of the proportion of PFS events between DAS and non-DAS events). **F** Location of events within the gene body. For alternative splice sites and cassette exons, the color indicates non-PFS (lilac) and PFS events (green). The density curve indicates the distance from the event to the closest extremity (start or end) of the gene. ***: $p < 0.001$ (two-sided Wilcoxon test, comparison of distance from extremity between PFS and non-PFS events, test statistics reported in Source Data). Source data are provided as a Source Data file.

did not exhibit a large Gini index, suggesting broad usage of both splice variants.

Do sequence features also affect *differential* AS?

Previous work has found a role of sequence features in distinguishing AS events from constitutive splicing[15–19]. Do sequence features also affect *differential* AS? To address this question, we examined the association of broad sequence features with differential AS. For each event type, we delineated the genomic regions composing the event locus. For example, for cassette exons, we considered the alternative exon as well as the two flanking introns. For each of these genomic regions, we measured its length, GC content, and conservation score. We compared the resulting values between differentially spliced events vs. events where we did not detect differential AS. In total, this resulted in 72 comparisons (Supp. Data 4), of which 6 were statistically significant (Fig. 6B, Table 1). Strikingly, four of the significant differences were for alternative first exons. Alternative first exons that display differential AS between neuron types have longer exons, a less conserved distal intron, and a longer distal intron. Alternative first exons are features of transcriptional regulation rather than post-transcriptional splicing. These data indicate that sequence features surrounding first exons may be highly variable to support fine-tuning expression of alternative isoforms at the level of transcription.

Alternative splicing of microexons

By contrast, cassette exons presenting differential AS in the nervous system appeared shorter on average. This observation is reminiscent of recent findings that microexons, with a length of 27 bp or less, are differentially spliced between *C. elegans* tissues[39] and frequently display neuron-specific inclusion[59]. We thus focused more closely on the differential splicing of microexons and found that, of the 75 microexons in the annotated genome, 71 are measurable in our dataset. Of those 71, we found that 53 (75%) displayed differential splicing in the nervous system, as opposed to 59% of the larger alternative exons (Fig. 6C, full list of annotated cassette exons available in Source Data file) ($p = 0.0136$ by proportion test). Furthermore, we asked if microexons showed more differential AS than other exons (Fig. 6D). Indeed, microexons were on average DAS in 11% of the neuron pairs tested, as opposed to 7% for longer exons ($p = 0.0035$, Wilcoxon test). Correspondingly, these microexons displayed a modest increase in specificity (Figure S4B). Overall, this analysis supports the previously reported importance of microexon inclusion in the nervous system.

Impact of alternative splicing on coding potential

We then sought to examine the potential impact of alternative splicing on the corresponding protein isoforms. We focused on alternative splice sites and cassette exons, restricting our analysis to events overlapping the coding sequence. Among the events present in the genome annotation and intersecting a coding sequence, 21% of alternative 3′ splice sites, 37% of alternative 5′ splice sites, and 24% of cassette exons result in addition of a number of nucleotides that is not a multiple of 3 (potentially frame-shifting, PFS). Further examination of these PFS events reveals their ability to disrupt the coding sequence of the gene, leading to protein isoforms with alternative N- and C-termini (Figure S5). We examined the proportion of PFS events that were also DAS. Alternative splice sites displayed a weak depletion of PFS events among DAS events (Fig. 6E, Table 2), while cassette exons did not show any meaningful difference in proportion of PFS events. We further hypothesized that PFS events resulting in substantial changes to the protein sequence would be subject to negative selection. Indeed, PFS events are concentrated near the 5′ and 3′ ends of the gene, primarily affecting the sequences of the N- or C-terminus of the protein (Fig. 6F).

Overall, our analysis indicates that differential alternative splicing is associated with specific features of the pre-mRNA. For alternative first exons, which are associated with an increase in intron and exon length, these features likely include transcriptional start sites. Cassette exons displaying DAS, by contrast, tend to be shorter than constitutive exons. In this case, shortness may be a result of selection for protein integrity, with only minor insertions or deletions well-tolerated.

Splicing regulatory network

Differential alternative splicing between neuron types is likely regulated in part by differential expression of splice factor (SF) genes. Our data enables the concurrent measurement of DAS and of gene expression in the same samples across 55 neuron types. We reasoned that these measurements might enable elucidation of a splicing regulatory network that links splice factor expression to DAS. To perform this analysis, we first sought a single measurement that could quantify DAS across genes and neuron types. For this purpose, we restricted our analysis to cassette exons—exons which are either included or excluded from the final transcript (e.g., *unc-40* cassette exon in Fig. 3). For each cassette exon, we computed an exonic Percent Spliced-In measure (ePSI) which captures the extent to which the exon is included in each individual neuron type (see Methods). We also compiled a list of 239 putative SF genes (Supp. Data 5) and quantified their expression (Supp. Data 6). Our list includes well-studied splice factors, as well as genes with only a speculative role in splicing regulation. Since SF genes themselves are heavily AS, we separately quantified expression of each SF transcript. These two measurements—ePSI and SF expression—constitute the input data for our model.

We constructed a covariance matrix which assesses, for each cassette exon ePSI and each SF transcript, how they covary over all 55 neuronal types in our data. In principle, the inverse of this covariance matrix is a precision matrix corresponding to the splicing regulatory network. However, the covariance matrix cannot be directly inverted due to the underdetermined system of equations that comprises many more covariates (SFs and ePSIs) than observations (neuron samples). Thus, we sought to estimate the precision matrix (Fig. 7A).

Table 1 | Sequence features presenting a statistically significant difference between DAS events and non-DAS events. Two-sided Wilcoxon tests with Benjamini-Hochberg correction for multiple comparisons

Event type	Feature	Median when DAS	Median when not DAS	*p*-value (adjusted)	
Alt. first exon	Proximal exon length	174 bp	129 bp	0.0021	**
Cassette exon	Exon length	83 bp	110 bp	0.0030	**
Alt. first exon	Distal intron length	4464 bp	3754 bp	0.0220	*
Alt. 5′ splice site	Intron length	238 bp	138 bp	0.0253	*
Alt. first exon	Proximal intron conservation	0.4040	0.4728	0.0260	*
Alt. first exon	Distal exon length	129 bp	105 bp	0.0262	*
Alt. 5′ splice site	Overhang conservation	0.9927	0.9521	0.0663	#
Alt. last exon	Distal exon conservation	0.6122	0.6641	0.0663	#
Alt. last exon	Distal intron length	2482 bp	1231 bp	0.0663	#
Mutually exclusive exons	First upstream intron conservation	0.5005	0.3695	0.0663	#
Mutually exclusive exons	Second exon length	113 bp	137 bp	0.0663	#
Cassette exon	Exon conservation	0.9826	0.9772	0.0663	#
Cassette exon	Upstream intron GC content	31.06%	31.72%	0.0663	#

Table 2 | Proportion of PFS events among DAS events. Chi-square tests with Holm correction for multiple comparisons

Alt. 5′ ss	Events in frame	Total events	% in frame	*p*-value	p adjusted
non-measurable	141	246	57		
not DAS	138	229	60		
DAS	165	234	71	0.0263	0.0527

Alt. 3′ ss	Events in frame	Total events	% in frame	*p*-value	p adjusted
non-measurable	166	271	61		
not DAS	261	329	79		
DAS	534	619	86	0.0076	0.0227
Cassette exon	**Events in frame**	**Total events**	**% in frame**	***p*-value**	**p adjusted**
non-measurable	178	265	67		
not DAS	240	309	78		
DAS	389	484	80	0.4086	0.4086

To select our method of precision matrix estimation, and optimize the hyperparameters, we used 5-fold cross-validation and computed four metrics (Figure S6A, see Methods). First, the Frobenius norm loss reflects the ability of the method to capture correlations. Second, the Fraction Explained Variance (FEV) of the PSI reconstruction using regression coefficients[60] reflects the ability to capture relationships between SF expression and ePSI. Third, we compiled a ground truth dataset of splicing regulatory interactions for comparison to the network structure (Supp. Data 7). Finally, we used a scale-free criterion on the structure of the network[61]. Using these metrics, we compared the following combinations of approaches. For precision matrix estimation, we considered the glasso[62], QUIC[63], CLIME[64], and SCIO[65] algorithms. For input to these algorithms, we considered either the measured ePSI or reconstructed counts. For normalizing the range of these inputs, we considered applying a Z-score or nonparanormal transformation[66]. Lastly, to handle missing values we considered k-nearest neighbors interpolation or median interpolation. Based on our evaluation metrics in a cross-validation, we chose to use the following methods: glasso as the precision matrix estimator algorithm, ePSI as input, nonparanormal truncated transformation, and k-nearest neighbors imputation (Fig. 7B, arrow, values in Supp. Data 8 and Source Data file), selecting 4-nearest neighbors (Figure S6B).

We then used all the training and validation data, together with the selected methods, to train a final model. We selected an optimal penalty for the glasso algorithm and assessed the model's performance on the as yet unseen test data using a permutation test approach. First, we permuted the model by randomizing the input cassette exon quantifications between samples in the training and validation data. For each permutated dataset, we tested a range of glasso penalty values. We then recalculated the model based on these unstructured data. We compared the performance metrics of these randomized models to that of the model trained on the unaltered data, at each glasso penalty value. For most metrics, the model fails to appropriately reconstruct the validation set when the input data is unstructured, indicating that the model captures relevant relationships (Figure S6C). We also calculated a permutation test *p*-value that reflects the relevance of the model for each metric and selected a glasso penalty value that optimized performance (Fig. 7C, Supp. Data 9). Our final model establishes a virtual network relating putative splice factors to splicing of cassette exons (Supp. Data 10), with a sparsity of 97% and 80% of the variance of ePSI explained. We compiled a "ground truth" dataset of genes whose splicing has been shown to change upon mutation of Splice Factors (see Methods). Using this dataset, we find that our splicing model reaches a True Positive Rate of 8.1% and a False Positive Rate of 4.6%. However, these evaluative metrics are limited by the scarcity of ground truth information on AS (see Discussion).

One application of our splicing network is discovery of potential splice factors. To this end, we examined the putative splice factors with the most central position in the network, ordered by number of connections to cassette exons. We found that many of the most central genes are indeed splice factors known to act in *C. elegans* neurons (Table 3). For example, *mbl-1*/Muscleblind[67], the hnRNP *hrpa-1*[50], and the CELF family gene *unc-75*[68,69] are all known splice factors in *C. elegans* neurons. Our analysis indicates that these key factors likely regulate differential alternative splicing across genes and neuron types, at least of cassette exons. Other central nodes are genes not previously known to function in neuronal splicing. For example, we identified the CELF family gene *etr-1* as a key splice factor. *etr-1* has known roles as a splice factor in muscle[70,71], and our recent data from single-cell RNA-Seq suggests expression in a restricted set of neurons[34]. Together, these data suggest that *etr-1* might play a novel role in regulating neuronal DAS, similar to its known role in muscle. Our analysis also placed the transcription factor *sma-9* and *melo-1*/periphilin 1 in a central position. Neither *sma-9* nor *melo-1* has an established role in splicing regulation. Thus, central nodes in our network, besides known splice factors, may indicate novel regulators of alternative splicing in neurons.

https://doi.org/10.1038/s41467-025-58293-5

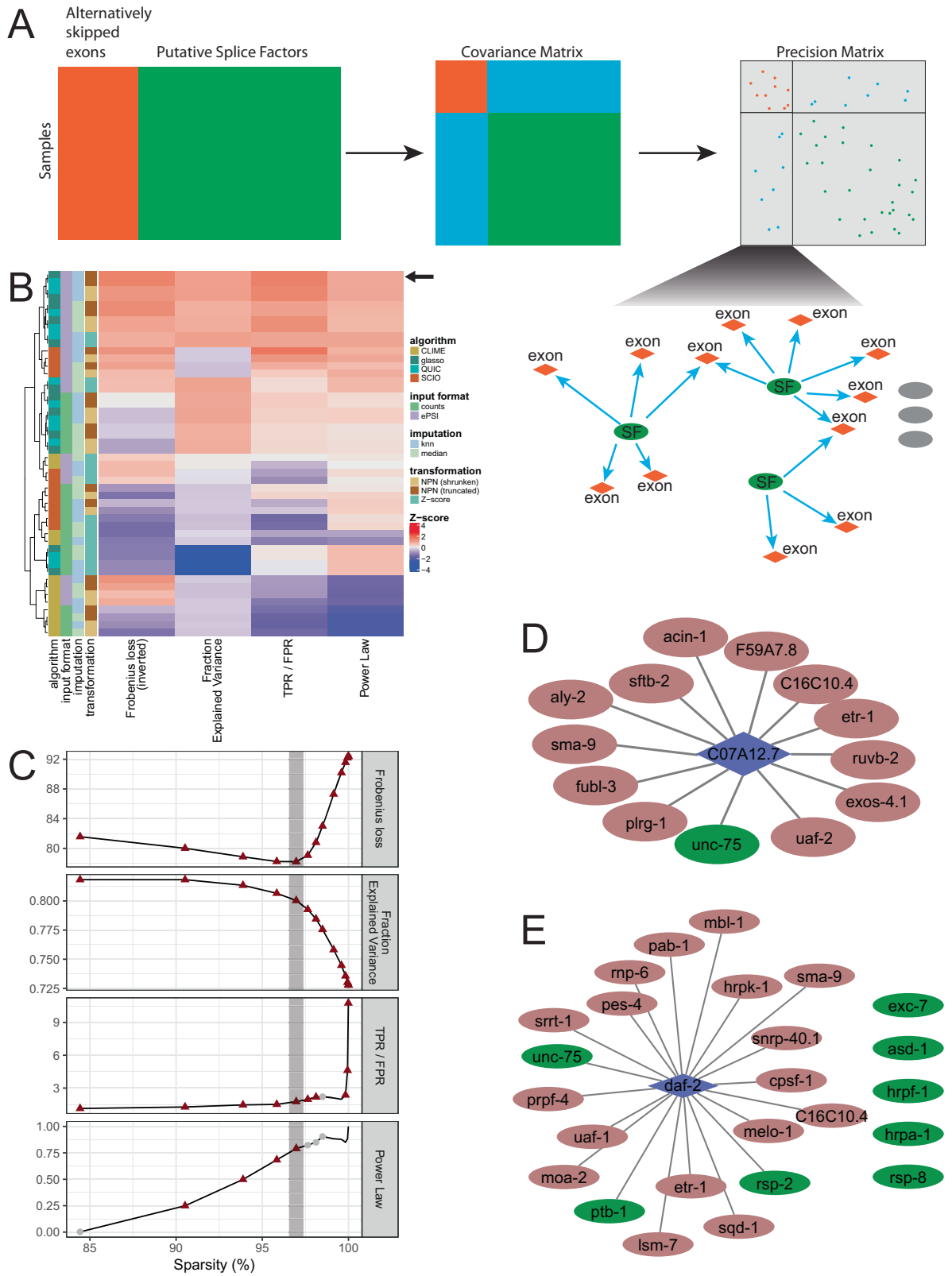


Fig. 7 | Splicing Regulatory Network. A Schematic representation of the approach: The splicing and expression measurements in each samples are grouped in a single data matrix, to compute a covariance matrix. The precision matrix is estimated, whose side quadrants represent the weights of a splicing regulatory network linking Splice Factors (SF) to cassette exons. **B** Comparison of metrics for selected parameters. The four metrics are represented across all methods. The Frobenius loss was inverted such that for all metrics, higher is better. The arrow represents the selected "best" method. **C** The four metrics plotted against sparsity for a range of penalties. The shaded area highlights the selected penalty of 0.25, corresponding to a sparsity of 97%. The red triangles correspond to values significantly different from the permuted data. **D**, **E** Subnetworks centered on exon 5 of *C07A12.7* (**D**) and exon 11.5 of *daf-2* (**E**). Blue diamond: cassette exon, red ellipse: putative splice factor displaying a non-zero weight in the network, green ellipse: splice factor identified from the literature to regulate this exon. Source data are provided as a Source Data file.

Table 3 | Top 20 putative splice factors with highest network degree

Gene id	Gene symbol	Degree	References
WBGene00019347	mbl-1	150	Norris et al.[67], Thompson et al.[105]
WBGene00001340	etr-1	141	
WBGene00004862	sma-9	76	
WBGene00001999	hrpa-1	75	Tomioka et al.[50]
WBGene00013307	melo-1	74	
WBGene00001484	fox-1	68	Kuroyanagi et al.[91], Kuroyanagi et al.[68], Tan and Fraser [99]
WBGene00006807	unc-75	67	Norris et al.[74], Tomioka et al. [50]; Kuroyanagi et al.[68]
WBGene00011279	asd-1	60	Kuroyanagi et al. [91], Kuroyanagi et al. [68], Tan and Fraser[99]; Tomioka et al. [50]
WBGene00006423	asd-2	57	
WBGene00012769	hrpu-1	53	
WBGene00004207	ptb-1	52	Tomioka et al.[50]
WBGene00017816	hrpk-1	51	
WBGene00008362	cfim-2	46	
WBGene00004698	rsp-1	44	
WBGene00006321	sup-12	44	
WBGene00003903	pab-2	44	
WBGene00004702	rsp-5	43	
WBGene00001368	exc-7	42	Norris et al.[74], Tomioka et al. [50], Tan and Fraser[99], Norris et al. [67], Choudhary et al.[106]
WBGene00010908	imph-1	40	
WBGene00007135	cdk-12	39	

A second application of our network is to analyze the regulation of specific cassette exons. We examined the subnetwork of putative SFs connected to *C07A12.7* exon 5 (Fig. 7D) and *daf-2* exon 11.5 (Fig. 7E). For *C07A12.7*, the known regulation by *unc-75*[69] was correctly detected. In addition, 12 other putative interactors appeared in the network: *acin-1*, *aly-2*, *C16C10.4*, *etr-1*, *exos-4.1*, *F59A7.8*, *fubl-3*, *plrg-1*, *ruvb-2*, *sftb-2*, *sma-9*, and *uaf-2*. Similarly, for *daf-2* exon 11.5, we detected the known regulation by *ptb-1*, *rsp-2*, and *unc-75* (we did not detect regulation by *asd-1*, *exc-7*, *hrpa-1*, *hrpf-1*, and *rsp-8*)[50,72]. We obtained an additional 17 predicted interactors: *C16C10.4*, *cpsf-1*, *etr-1*, *hrpk-1*, *lsm-7*, *mbl-1*, *melo-1*, *moa-2*, *pab-1*, *pes-4*, *prpf-4*, *rnp-6*, *sma-9*, *snrp-40.1*, *sqd-1*, *srrt-1*, and *uaf-1*. In addition, we examined the subnetwork of putative splice factors connected to *unc-16*. We quantified 6 separate events, corresponding to 4 exons (two pairs of events correspond to the same exon with differing flanking introns) in the gene *unc-16*, represented as individual nodes in the subnetwork (Figure S6D). We identified the known regulation by *unc-75* and *exc-7*, but did not detect regulation by *prp-40*[69,73,74]. In addition, we found 58 network connections that do not correspond to known regulatory interactions. Thus, our splicing network can generate detailed hypotheses about the regulation of specific splicing events.

Discussion

In this study, we present an atlas of alternative splicing (AS) in the *C. elegans* nervous system, at single neuron type resolution, across 55 neuron types. We develop a toolset of analytic approaches and a website to facilitate their use by the research community. We show that our approach yields results in agreement with existing data, and identifies multiple new examples of alternative splicing in genes with key roles in neuronal function. Systematic quantification reveals broad patterns of differential AS, particularly affecting ion channels, shaping the landscape of neuronal mRNAs separately from gene expression. We provide a broad description of the *C. elegans* neuronal alternative transcriptome, and observe that microexons are notably differentially alternatively spliced (DAS) in the nervous system. Finally, we compute a splicing regulatory network to formulate new hypotheses on splice factor regulation of differential alternative splicing, focused on cassette exons.

With the wider availability of single-cell RNA-Seq methods, large efforts have been made in the recent years to establish atlases of gene expression in *C. elegans*[31–37]. However, these sequencing approaches often present a strong 3′ end bias which make them unsuitable for AS analysis. A more restricted set of studies have focused on AS. In *C. elegans*, these studies have been performed mostly at the level of whole animals, or through comparison between tissues[39–41]. Our data are broadly consistent with these previous analyzes, for example confirming the reported rarity of alternative last exons in the *C. elegans* transcriptome (Fig. 6A), the orthogonality of gene expression and alternative splicing (Fig. 5E, Figure S3C), and an increased differential splicing of shorter cassette exons (Fig. 6C, D)[39].

While our analysis largely aligns with other work, two apparent discrepancies are our global counts of intron retention and alternative first exons. For intron retention, we found relatively fewer instances than a parallel analysis of the same data[58]. A key underlying factor is our use of the tool SUPPA, which only measures intron retention relative to annotated transcripts, and does not detect unannotated instances. For alternative first exons, we found more instances than an analysis of a separate data set[39]. In this instance, SUPPA treats multiple alternative first exons in the same gene as distinct events, rather than grouping them as a single event. Overall, these discrepancies highlight the difficulty of computationally interpreting alternative splicing at genome scale. For this reason, we encourage systematic inspection of the raw data visualization we provide, to guide investigation of splicing at any specific locus.

For further comparison, we processed the previously published TRAP-Seq dataset[39] using SUPPA and examined the events detected as DAS between serotonergic neurons and the rest of the nervous system both in our dataset and in the TRAP-Seq dataset (not shown). We found little overall agreement between the two datasets, attributable to several causes. First, as our dataset did not measure the MI serotonergic neuron, restricted but strong event inclusion in this neuron (or exclusion) may be evident in the TRAP-Seq measurements and not in our data. Conversely, not all serotonergic neurons share the same splicing patterns, thus the increased resolution of our dataset can detect additional DAS events. For example, in the case of *mvk-1* exon 12, the TRAP-Seq data correctly indicates inclusion in most neurons, but our higher-resolution data detects exclusion in HSN and DA. Finally, only 2 TRAP-Seq replicates are available per tissue, and the sequencing data displays a 3′ bias. This limits statistical power and can preclude correct detection of events far from the 3′ end of the gene. For example, *daf-2* exon 13 appears DAS from our data, but, being located 12.5 kb away from the gene 3′ end, the TRAP-Seq data does not provide sufficient coverage to analyze its usage. Thus, while the two datasets complement each other by enabling measurements inside and outside the nervous system, our data allows deeper, higher-resolution interrogation of events within the nervous system. Studies in other organisms have examined AS at the tissue[21] or single-cell level in the nervous system[5,75–79]. A feature of our dataset is its genome-wide scope and its high-resolution analysis of AS across many individual neuronal types. Thus, this analysis complements and extends previous studies.

Our analysis discovered novel, previously unannotated, exons in at least 1353 genes. Interestingly, in the case of *unc-40*, orthologous exons had been previously annotated in *C. briggsae*, despite evidence suggesting that *Cbr-unc-40* is expressed at a lower level than its *C. elegans* ortholog[44,80]. This highlights the noisy nature of gene and transcript annotation, which is highly dependent both on the available data and the algorithms used for transcript reconstruction.

Consequently, researchers interested in the study of a particular gene should always consult available sequencing data in *C. elegans* and related species to identify potential unannotated isoforms.

We used a computational approach to define a Splicing Regulatory network. A challenge in evaluating the performance of such an approach is the availability of a ground truth, which is instrumental both for selecting hyperparameters during model training, and for evaluating the accuracy of predictions. Although we used a compiled list of interactions between splice factors and splice events as one metric (Supp. Data 7), this approach has strong limitations. First, some of the data are not obtained from neurons, and even the neuronal data does not reach the resolution of individual neuron types. Second, we aggregated this data at the level of entire genes rather than individual events. Third, a substantial portion of this data corresponds to genes whose splicing pattern changes as a consequence of SF (Splice Factor) mutation; these changes may reflect a direct regulatory interaction or be an indirect effect of the SF knock out. In contrast to previous attempts that did not evaluate alternative approaches, and either did not justify the initial selection of hyperparameters[21,24], or used a single criterion[5,22], we adopted an innovative approach to select the hyperparameters of the model using four separate criteria. Our principled tuning method may be beneficial for other models with limited access to ground truth. However, the final step of evaluating the model remains challenging. The True Positive Rate of our final model indicates that only 8% of the regulatory interactions present in the ground truth data are captured by the model. While the ground truth for the selected examples of C07A12.*7, daf-2*, and *unc-16* is more reliable, coming from low throughput studies, these examples also highlight the model's imperfections. Nevertheless, due to its high sparsity, our model can strongly narrow the candidate space, making it a uniquely valuable tool for hypothesis generation. Overall, our final network captures many known interactions between putative splice factors and splice events, and constitutes a powerful tool for the exploration of novel splicing regulatory mechanisms.

Most previous efforts to model the role of splice factors on DAS were performed with RNA-Seq data set either derived from whole animals[24] or from tissue-specific samples[21,22,26,29,81,82]. Our approach enables profiling the transcriptome at single cell type resolution. However, each sample is composed of thousands of individual cells, at a single time point. Previous work reports that 18%[83] or 30%[40] of AS events are differentially regulated during development, which cannot be evaluated in our dataset. The recent progress in single-cell RNA-Seq facilitates higher resolution studies in multiple conditions[5,20]; in particular, current progress in single-cell long-reads sequencing offers great opportunities to extend this analysis[84–86]. These single cell approaches allow for more specificity, capture the heterogeneity within a population of single cells of the same type, and potentially allow for whole-transcript analysis, promising more comprehensive, higher-quality atlases in the near future.

Our approach has cataloged DAS in single neuron types for almost half the canonical neuron types in the *C. elegans* nervous system. Our data indicate that alternative splicing affects the function of key neuronal genes, and reveals substantial novel splicing diversity. These splicing events might control subtle, cell-specific alterations of neuronal form and function that are not accessible by broader forms of genome regulation. Thus, we expect studies of gene function to be informed by these data about differential alternative splicing in specific neuron types. In addition, from the perspective of splicing itself, we use the diversity we have discovered to model regulatory mechanisms that mediate the control of differential alternative splicing.

Methods

FACS isolation and sequencing

For single-cell type bulk RNA-Seq, *C. elegans* strains expressing a fluorescent protein or combination of fluorescent proteins in a single neuron type were dissociated and sorted into Trizol as described previously[34,42,87]. RNA was extracted and sequencing libraries were prepared using the Ovation® SoLo® RNA-Seq library preparation kit, yielding even coverage along the gene body, as described previously[43]. Libraries were sequenced with an Illumina HiSeq 2500 or NovaSeq 6000 (Supp. Data 1). The dataset covers 211 samples corresponding to 55 neuron types, and an additional 8 control samples from whole animal sorts. All neuron types were sequenced in 3-6 replicates, except ADF, M4 (1 replicate each), OLL and PVD (2 replicates). Four samples failed quality control and were excluded from subsequent analyzes.

Following trimming, the RNA-Seq reads were aligned to the *C. elegans* genome (Wormbase WS289) using STAR 2.7.7a[54] with option `--outFilterMatchNminOverLread 0.3` (all other settings left to default). Deduplication was performed using UMI-tools[88]. The pipeline code is available at: https://github.com/cei/bulk_align.

RT-PCR

For RT-PCR, mixed stage N2 *C. elegans* were grown following standard methods[89]. Plates were washed with M9 buffer, and 100 μL of worm suspension was added to 400 μL of Trizol and immediately frozen in liquid nitrogen. Samples were stored at −80 °C. RNA extraction was performed with chloroform in Phase Lock Gel Heavy tubes (Quanta-Bio), treated with DNase I (Thermo Fisher) and purified with the Macherey-Nagel Nucleospin kit following manufacturer's instructions. Finally, cDNA was synthetized using the Affinity Script Multiple Temperature cDNA Synthesis kit (Agilent), following manufacturer's instructions with oligo-dT primers. For each sample, an additional tube was prepared with identical composition, adding water in place of reverse transcriptase. The resulting cDNA was stored at −20 °C. RT-PCR reactions were performed with Phusion polymerase (New England Biolabs), following the manufacturer's protocol. The primers used are listed in Table S1.

CRISPR excision of *unc-40* exon 14.5

To knock-out exon 14.5 of *unc-40*, we injected wild type N2 animals with a Cas9 mix using two crRNAs to create two double-strand breaks near the splice sites of exons 14 and 15, and a single-stranded repair template. The corresponding sequences are listed in Table S1. The deletion was confirmed by sequencing, and crossed with the fluorescent marker for touch receptor neurons. Young adult individuals were examined under a Zeiss Axioplan 2 microscope, imaging approximately 20 individuals of each genotype on 3 separate days. The strains used are listed in Table S2.

Discovery of novel splice junctions

STAR produces junction files, providing a list of splice junctions detected in the processed sample, along with the measured count, annotation status, and other information[54]. We only considered novel junctions (not present in the annotation), that were flanked by canonical splice site motifs. In addition, we only considered splice junctions supported by reads with 12 bp overhang on each side of the junction (STAR's default value for `--outSJfilterOverhangMin`). We defined the neighborhood of a splice junction as the set of genes within 60 bp of either splice site, regardless of the strand. We then filtered the junctions in each sample fulfilling the following criteria:

- Junction no longer than 1000 bp
- At least 2 supporting reads (uniquely mapped, with at least 12 bp overhang following STAR default) supporting that junction
- Not in the neighborhood of an rRNA gene
- In the neighborhood of a protein-coding gene, long-non-coding RNA, or pseudogene
- Has at least 20% as many reads as the most highly detected splice junction from the neighbor genes

After processing each sample with the above filters, we aggregated the junctions across samples and conducted a second round of filtering. We kept novel splice junctions that were detected in at least half the samples from a single neuron type (with a minimum of two samples from a single neuron type). This analysis identified 1722 novel junctions robustly expressed in our dataset. Reliably attributing each junction to a novel transcript is challenging with available methods[48]. Instead, we only attempted to estimate the total number of genes containing novel junctions, without determining their precise identity. To this end, we used three approaches under different assumptions.

First, we determined the minimal number of novel junction-containing genes that could explain the observed pattern by formulating an integer programming problem. For $S = 1722$ novel junctions and $G = 1430$ genes that are neighbors to one of these junctions, we denote the neighbors of junction s as the set of genes N^s. Then, we define the binary variables $(x_{s,g})_{s\in[1,S],g\in[1,G]}$ for each combination of a splice junction s and a gene g, with $x_{s,g} = 1$ if the splice junction s is *not* part of the gene g, and $x_{s,g} = 0$ if the splice junction is in the gene. Thus, since each splice junction is part of a single gene, we can define the following constraint for each splice junction:

$$\forall s \in [1,S],\ \sum_{g\in N^s}(1 - x_{s,g}) = 1 \tag{1}$$

Since $(x_{s,g})$ are binary, this is equivalent to having $x_{s,g} = 1$ for all genes neighbor of s except one, i.e. a junction only belongs to a single gene. Further, we define the objective function as the total number of genes that do not have any novel splice junction. To this end, we define a new set of variables $y_g = \prod_s x_{s,g}\{s|g \in N^s\}$ such that gene g contains a splice junction iff $y_g = 0$. The objective function is thus the sum of all y_g variables. Since the y_g are products, the objective function can be linearized into a sum by introducing additional constraints[90]:

$$\begin{aligned} y_g &\le x_{s,g},\ \forall s|g \in N^s \\ y_g &\ge \textstyle\sum_s x_{s,g} - (n-1) \end{aligned} \tag{2}$$

Where n is the number of $x_{s,g}$ variables in the product y_g. Having formulated this integer programming problem, we used the R package lpSolve to maximize the objective function, thus minimizing the number of novel junction-containing genes while ensuring that each junction is counted in a single gene. Conversely, we also used lpSolve to minimize the objective function, thus maximizing the number of junction-containing genes while ensuring that each junction is counted in a single gene. This provides bounds on the total number of novel junction-containing genes.

To obtain a single, average, number of novel junction-containing genes, we further used a resampling approach. We consider the set of genes neighboring each novel junction. Under the assumption that the novel junctions are uniformly distributed among the genes, we randomly selected a single gene for each junction, and evaluated the total number of genes with novel junctions for this sample. We repeated this procedure 1000 times to estimate an average number of novel junction-containing genes. The complete code, along with an illustrative toy example, is available at: https://gicengenproject/cei/novel_junctions.

DAS with MAJIQ

Local quantification of AS and the analysis of differential AS were performed with MAJIQ 2.3[47]. We built a configuration file using the reference annotation from Wormbase (WS289), strandness forward, and one experiment per neuron type (grouping the biological replicates by experiment). We subsequently ran `majiq build` with parameter `--min-experiments 2` (i.e. a splice junction is retained if it passes filters in at least two replicates from the same neuron), keeping the other options to their default values. We performed mPSI quantification, and delta mPSI quantification between each pair of neuron types with the default parameters.

For the subsequent analysis, we grouped the resulting delta mPSI files from all neuron pairs, obtaining 16,379,082 individual comparisons (for a given LSV in a given pair of neurons). We filtered comparisons to retain only those where the LSV-containing gene was expressed in both neurons of the pair, using the threshold "3" we previously defined based on single-cell RNA-Seq[34], resulting in 8,783,582 comparisons (corresponding to 3787 measurable genes). We define a comparison as DAS if the "probability of not changing" (computed by MAJIQ deltapsi) is lower than 0.05, and the "probability of changing" is higher than 0.5, corresponding to 928,985 comparisons. Finally, we define a gene as DAS if it contains at least one DAS comparison, yielding 1940 DAS genes.

To predict genes expected to display DAS in the nervous system, we compiled a list of 11 studies reporting individual genes[91–97] or performing transcriptomic analyzes in mutant backgrounds disrupting AS in neurons[67,74,98,99]. This resulted in a list of 759 genes (Supp. Data 3), of which 542 are measurable in our dataset (expressed in the neurons sequenced and quantified by MAJIQ). Gene Ontology analysis was performed using a background list of 10,312 genes expressed in at least two neurons sequenced here (as per the threshold above).

To compare the differentially AS to differentially expressed (DE) genes, the DE genes were obtained from integration of single-cell RNA-Seq data with this dataset, as described in ref. 42. For each neuron pair, the DE genes were ordered by absolute fold change and the genes with the 100 highest values were selected. The DAS genes were ordered by absolute delta mPSI of their highest LSV, and the 100 genes containing the highest values were selected. Out of 595 neuron pairs, we could select a top 100 DAS genes for 432 pairs; 15 had 101 genes (because of a tie in the highest delta mPSI), 148 pairs did not have 100 DAS genes. We only represented pairs where we could select 100 top DAS genes.

To predict the total number of DAS genes in the nervous system, we randomly selected between 2 and 55 neuron types among those sequenced, and estimated the number of DAS genes that could be detected. We repeated the procedure 10 times for each number of neurons. We then performed a linear regression of the number of DAS genes detected vs the logarithm-transformed number of neurons subsampled, yielding the relationship $N_{genes} = -302 + 731 \cdot \log(N_{neurons})$. The total number of DAS genes for 119 neuron types is then 3192. We applied the same procedure to estimate the number of genes expressed in the subsampled neurons (above threshold "3" as above). We obtained the relationship $N_{genes} = 4756 + 2127 \cdot \log(N_{neurons})$, and estimate a total number of to 14,920 genes expressed in the *C. elegans* nervous system.

Transcript quantification with StringTie

For the transcript-level quantification, we used StringTie 2.2.1 using the annotation from Wormbase (WS289), without novel transcript discovery. We applied it to the aligned reads from STAR, following deduplication. Code available at: https://github.com/cengenproject/stringtie_quantif.

Website

The output of STAR was used to generate browser tracks. First, the junction counts generated by STAR were processed, and the number of uniquely mapped reads was kept as junction count. Junctions longer than 25,000 bp, and junctions with a count lower than 3 reads were filtered out. Junction counts for individual samples were combined into neuron type average, and global sum, minimum, and maximum. The global tracks underwent a second filtering, requiring 13 and 21 reads for the maximum and sum tracks respectively. The tracks were exported to bed format using the R package rtracklayer[100]. Second, the bam files generated by STAR were used to generate the coverage tracks using custom code, and exported to bigWig format with rtracklayer. The individual tracks can be used in JBrowse2[46] or

downloaded from the website. All code is available at: https://github.com/cengenproject/splicing_browser.

To explore the local splicing quantification in an event-centric manner, the results of the MAJIQ analysis (see above) were loaded in VOILA according to author's instructions (see https://majiq.biociphers.org). To enable exploration in a neuron-centric manner, we further developed a custom R Shiny application operating on the tsv files generated by MAJIQ deltapsi. For comparison of a pair of neurons, the user can directly apply a threshold on the estimates of "probability_changing" and "probability_non_changing" reported by MAJIQ. For comparison of two sets of neurons, the application selects the estimated mPSI of each event in each neuron of both sets, performs a t-test between these two sets, and reports the *p*-value corrected for multiple comparisons by the Benjamini-Hochberg method. The assumptions of the t-test may or may not be met depending on the exact sets of neurons chosen, thus offering no guarantee that the FDR is appropriately controlled. We observed that this approach robustly selects events of interest in the majority of use cases, however we caution against using this approach to draw conclusions about arbitrary sets of neurons, and added a warning to this effect in the application. The source code of this application is available at: https://github.com/cengenproject/das_by_neuron.

For the transcript-level quantification, the quantifications from StringTie (see above) were loaded in a custom R Shiny application, source code available at: https://github.com/cengenproject/isoform_compare/.

Binary DAS with SUPPA2

We used SUPPA 2.3[101] according to the documentation. We generated all local event types (SE, SS, MX, RI, FL) with `--boundary S` based on the genome annotation (Wormbase WS289). We then quantified the event bPSI by running `psiPerEvent` using the StringTie quantifications (see above) as expression file and a threshold of 5 TPMs. Finally, we split the resulting TPM and bPSI files by neuron type, before computing delta bPSI for each pair of neuron, using the settings `–method empirical –combination –gene-correction` and a threshold of 0.3 bPSI. We considered an event differentially AS in a neuron pair if it is displayed a *p*-value lower than 0.1 (*p*-values corrected for multiple comparison by SUPPA with the -gc option), and displays a delta bPSI higher than 0.3.

To explore sequence features of DAS events, we used the R package GenomicFeatures[102] to delineate the genomic regions of interest, the package Biostrings to extract the sequence and calculate its GC content, and the PhastCons conservation score track downloaded from UCSC[103]. The features were compared using a Wilcoxon test followed by Benjamini-Hochberg correction for multiple comparisons, and we applied a threshold of 0.05 to consider a comparison significant. To explore the impact of AS events on protein sequence, we only considered cassette exons and alternative 3' and 5' splice sites. We used the function `join_overlap_inner_directed` in the plyranges R package to determine if the added sequence (the cassette exon, or the overhang in case of alternative splice sites) overlaps with the CDS from the Wormbase WS289 annotation. We then computed the length of the added sequence and determined if it is a multiple of 3.

For microexons, we focused on cassette exons. We compared the number of exons with DAS using a two-proportion test with Yates' continuity correction. To compare the proportion of neuron pairs with DAS, we only analyzed exons detectable in at least 10 neuron pairs (787 out of 913 cassette exons). We used a Wilcoxon test to compare the proportions. Code available at https://github.com/cengenproject/suppa_events.

Reanalysis of the TRAP-Seq dataset

We downloaded the published TRAP-Seq data[39] from the Short Read Archive database (GSE106374; we processed the 10 samples SRR6238092-SRR6238101, we did not re-process the input controls) and aligned them to the *C. elegans* genome (Wormbase WS289) using STAR 2.7.11a with option `--outFilterMatchNminOverLread 0.3`.

Tissue-level coverage and junction genome browser tracks for each of the 5 tissue types analyzed were generated as described above by aggregating the corresponding 2 biological replicates. Transcript-level quantification with StringTie performed as described above and used as input for SUPPA splice event quantification as described above.

All code is available at: https://github.com/cengenproject/reanalysis_tissue

Splicing regulatory network

Quantification of ePSI. Cassette events were extracted from the genome annotation using `suppa generateEvents`[101]. With an approach adapted from[104], we then used `bedtools` and `grep` on the STAR output to count the number of inclusion reads N_i covering the alternative exon, and the number of exclusion reads N_e spanning the alternative splice junction. We then computed exonic Percent Spliced-In (ePSI) from normalized read counts $\overline{N_i}$ and $\overline{N_e}$ based on exon length L_{exon} and read length l_{read}:

$$\begin{aligned}\overline{N_i} &= \frac{N_i}{L_{exon}+l_{read}-1}\\ \overline{N_e} &= \frac{N_e}{l_{read}-1}\\ \mathrm{PSI} &= \frac{\overline{N_i}}{\overline{N_i}+\overline{N_e}}\end{aligned} \tag{3}$$

All relevant code is stored in the Github repository https://github.com/cengenproject/quantif_exon_skipping.

Before use in the model, we removed measurements of a cassette exon in a neuron type if the exon-containing gene is not expressed in that neuron type, based on a thresholding integrating this dataset with single-cell RNA-Seq[42]. Additionally, we only considered neuron types for which we had 3 or more biological replicates. We also filtered the cassette exons, keeping only events covered by more than 20 reads, measured in more than 70 samples from 23 neuron types, and presenting differential splicing between neuron types (standard deviation above 0.3).

Quantification of putative splice factor transcripts. We compiled a list of putative splice factors, available in Supp. Data 5. The transcripts are quantified using StringTie[48], without novel transcript discovery (see above). StringTie gives quantifications in Transcripts Per Million (TPM), which undergo a log10 transformation with a pseudocount of 1 before further processing.

Precision matrix estimation. Here we describe the procedure to select our network construction method (Figure S6A). We build a data matrix where the 127 rows correspond to samples, the 730 columns correspond to cassette exon ePSI (172 events) or splice factor log-TPM (558 transcripts). We perform a first split: 30% of the rows (39 samples) are kept as testing set. The other 70% of samples undergo 5-fold cross-validation: each fold contains 17 or 18 samples, the training is performed on 4 folds, the validation on the held-out fold. As the splicing of a cassette exon in a neuron type can only be meaningfully measured if the gene containing that exon is expressed in that neuron, the ePSI matrix contains missing values that are first imputed (using the column median or k-nearest neighbors). The training matrices $\mathrm{SE}_{\mathrm{train}}$ (containing the skipped exons) and $\mathrm{SF}_{\mathrm{train}}$ (containing the splice factors) are then transformed (using Z-score or NPN transformation), and a covariance matrix S_{train} is computed from the transformed values. We store the variables used for transformation (e.g. the mean and standard deviation for Z-scoring and distribution quantiles for the NPN method) for inverting the transformations later to yield predictions in the original data range. For permutation tests, the $\mathrm{SE}_{\mathrm{train}}$ values randomized within an event (i.e. within a column) after transformation, but before computing the covariance matrix. The covariance matrix is then used to estimate the precision matrix $\widehat{\Omega}_{\mathrm{train}}$ (using glasso, QUIC,

CLIME, or SCIO), which is inverted to recover the estimated covariance matrix $\hat{S}_{\text{train}}$.

Separately, the validation fold matrices SE_{valid} and SF_{valid} are transformed re-using the same parameters as the training folds, to compute the covariance matrix S_{valid}. From the estimated precision matrix, following[60], we extract the quadrants $\widehat{\Omega}_{21}$ (with the splice factors as rows and the cassette exons as columns), and $\widehat{\Omega}_{11}$ (with the cassette exons as rows and columns) and compute $W = \widehat{\Omega}_{21}\widehat{\Omega}_{11}^{-1}$, the matrix of regression coefficients. The splicing measurements in the validation set can then be estimated from the splice factors in the validation set and the precision matrix learned from the training set following:

$$\widehat{\text{SE}}_{\text{valid}} = W^t \cdot \text{SF}_{\text{valid}} \tag{4}$$

Finally, we invert the transformation of $\widehat{\text{SE}}_{\text{valid}}$ using the stored transformation variables (e.g. the mean and standard deviation for Z-scoring and distribution quantiles for the NPN method) to get back to the initial scale.

Model components. We tried several approaches to develop an optimal model. The data matrix SE_{train} can be constructed from ePSIs, a ratio of inclusion and exclusion counts. We reasoned that a model may have a better performance when directly estimating the inclusion and exclusion counts rather than the ratio. Thus we reconstructed counts by multiplying the ePSI with the total count for that exon, and used either ePSIs or reconstructed counts as the columns of SE_{train}.

As our downstream methods are incompatible with the presence of missing values, we need to remove them from SE_{train}. We used median imputation, where the missing value is replaced by the median of the column (i.e. the median of the cassette exon across samples). Alternatively, we used a k-nearest neighbors imputation implemented by the R package impute.

ePSIs (or reconstructed counts) and log-TPMs follow very different distributions, which would distort the covariance computed by simply concatenating them. In addition, they do not follow a normal distribution, making them inappropriate for the downstream algorithms. We thus standardized the training data either with a Z-score transformation, or with nonparanormal transformations[66]. As our cross-validation procedure requires that we transform the validation set using the parameters from the training set, and that we invert the transformation to obtain values in the original scale, we implemented these transformations in the R package projectNPN, available at https://github.com/cengenproject/projectNPN.

Finally, the estimation of the precision matrix can also be performed with several implementations. We used the R packages glasso, QUIC, FLARE (implementing CLIME) and SCIO.

Metrics definitions. First, we compare the covariance matrix measured in the validation set S_{valid} to the covariance matrix estimated from the training set $\hat{S}_{train}$, obtained by inverting the estimated precision matrix. we focus on the quadrant containing the covariance between the skipped exons and the splice factors, as we are interested in the ability of our model to capture this relationship. We then compute the Frobenius loss as $||S_{\text{valid},12} - \hat{S}_{\text{train},12}||_F$ where $||\ ||_F$ represents the Frobenius norm.

Further, we compare the skipping values measured in the validation set SE_{valid} to the values $\widehat{\text{SE}}_{\text{valid}}$ estimated from the precision matrix. To this end, we first compute the residuals: $resid = \widehat{\text{SE}}_{valid} - \text{SE}_{\text{valid}}$. We then estimate, for each event e, the sum of squared residuals: $\text{SS}_{\text{err},e} = \sum \text{resid}_e^2$ and the total sum of squares $\text{SS}_{\text{tot,e}} = \sum \left(\text{SE}_{\text{valid,e}} - \overline{\text{SE}_{\text{valid,e}}}\right)^2$ and define the fraction explained variance as $\text{FEV}_e = 1 - \frac{\text{SS}_{\text{err,e}}}{\text{SS}_{\text{tot,e}}}$. We truncate this value at 0 and average it across events to obtain the mean fraction of explained variance.

To evaluate the biological relevance of the network edges, we extract the quadrant $\hat{\Omega}_{21}$ of the precision matrix (i.e. the adjacency matrix), binarize it (taking an edge for any non-zero entry), and compare it to a ground truth dataset (Supp. Data 7). We compiled this dataset by a review of the literature, considering that a regulatory interaction between a splice factor and a cassette exon is "true" if a change in splicing was detected upon mutation of the splice factor. Note that this dataset suffers from several limitations, notably these interactions do not necessarily take place in the neurons we sequenced here, and these interactions may correspond to different splicing events within the same target gene. Thus, while we expect a better model to obtain a better match with this data, we do not expect a perfect match. We compute the True Positive Rate (TPR) as the fraction of interactions present in the ground truth that are captured by the model, and the False Positive Rate (FPR) as the fraction of interactions that are absent from the ground truth but reported by the model. As both TPR and FPR decrease with increased sparsity, we report the ratio TPR/FPR.

Finally, we seek to constrain the structure of the network. A very sparse network, where each splice factor has at most a single target, or a very dense network, where many splice factors have many targets, would be hard to interpret and likely not capture biologically meaningful interactions. As proposed by ref. 61, we use an approximate scale-free topology criterion. For each splice factor (node in the network), we compute the degree of the node k, and count the number of nodes with the same degree $p(k)$. We then fit a linear regression between $\log(k)$ and $\log(p(k))$, and use the coefficient of determination R^2 as a criterion. A high R^2 suggests that a power law can describe the node degrees, and that the network is scale-free.

All code related to the network modeling is available at: https://github.com/cengenproject/regression_exon_skipping.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw sequencing data is available on GEO (accession GSE229078). The published TRAP-Seq data was downloaded from GEO (accession GSE106374). Source data are provided with this paper.

Code availability

All code is available at https://github.com/cengenproject/ under the corresponding repository names listed in Methods. In addition, a frozen snapshot of the code, along with any dataset necessary to reproduce this work, is deposited on Zenodo, with https://doi.org/10.5281/zenodo.14206313. Full list of the individual Github repositories: • https://github.com/cengenproject/bulk_align. • https://github.com/cengenproject/novel_junctions. • https://github.com/cengenproject/stringtie_quantif. • https://github.com/cengenproject/splicing_browser. • https://github.com/cengenproject/das_by_neuron. • https://github.com/cengenproject/isoform_compare. • https://github.com/cengenproject/suppa_events. • https://github.com/cengenproject/reanalysis_tissue. • https://github.com/cengenproject/quantif_exon_skipping. • https://github.com/cengenproject/projectNPN. • https://github.com/cengenproject/regression_exon_skipping.

References

1. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet* **40**, 1413–1415 (2008).
2. Wang, E. T. et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
3. Daguenet, E., Dujardin, G. & Valcarcel, J. The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing

inform novel therapeutic approaches. *EMBO Rep.* **16**, 1640–1655 (2015).
4. Yeo, G., Holste, D., Kreiman, G. & Burge, C. B. Variation in alternative splicing across human tissues. *Genome Biol.* **5**, R74 (2004).
5. Benegas, G., Fischer, J. & Song, Y. S. Robust and annotation-free analysis of alternative splicing across diverse cell types in mice. *Elife* **11**, e73520 (2022).
6. Furlanis, E. & Scheiffele, P. Regulation of neuronal differentiation, function, and plasticity by alternative splicing. *Annu Rev. Cell Dev. Biol.* **34**, 451–469 (2018).
7. Weyn-Vanhentenryck, S. M. et al. Precise temporal regulation of alternative splicing during neural development. *Nat. Commun.* **9**, 2189 (2018).
8. Lambourne, L. et al. Widespread variation in molecular interactions and regulatory properties among transcription factor isoforms. *bioRxiv* https://doi.org/10.1101/2024.03.12.584681 (2024).
9. Sousa, E. & Flames, N. Transcriptional regulation of neuronal identity. *Eur. J. Neurosci.* **55**, 645–660 (2022).
10. Zheng, S. Alternative splicing programming of axon formation. *Wiley Interdiscip. Rev. RNA* **11**, e1585 (2020).
11. Fisher, E. & Feng, J. RNA splicing regulators play critical roles in neurogenesis. *Wiley Interdiscip. Rev. RNA* **13**, e1728 (2022).
12. Park, C. Y. et al. Genome-wide landscape of RNA-binding protein target site dysregulation reveals a major impact on psychiatric disorder risk. *Nat. Genet* **53**, 166–173 (2021).
13. Quesnel-Vallieres, M., Weatheritt, R. J., Cordes, S. P. & Blencowe, B. J. Autism spectrum disorder: insights into convergent mechanisms from transcriptomics. *Nat. Rev. Genet* **20**, 51–63 (2019).
14. Bonnal, S. C., Lopez-Oreja, I. & Valcarcel, J. Roles and mechanisms of alternative splicing in cancer - implications for care. *Nat. Rev. Clin. Oncol.* **17**, 457–474 (2020).
15. Cheng, J. et al. MMSplice: modular modeling improves the predictions of genetic variant effects on splicing. *Genome Biol.* **20**, 48 (2019).
16. Zeng, T. & Li, Y. I. Predicting RNA splicing from DNA sequence using Pangolin. *Genome Biol.* **23**, 103 (2022).
17. Jaganathan, K. et al. Predicting splicing from primary sequence with deep learning. *Cell* **176**, 535–548 e524 (2019).
18. Rosenberg, A. B., Patwardhan, R. P., Shendure, J. & Seelig, G. Learning the sequence determinants of alternative splicing from millions of random sequences. *Cell* **163**, 698–711 (2015).
19. Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput Biol.* **11**, 377–394 (2004).
20. Arzalluz-Luque, A., Salguero, P., Tarazona, S. & Conesa, A. acorde unravels functionally interpretable networks of isoform co-usage from single cell data. *Nat. Commun.* **13**, 1828 (2022).
21. Saha, A. et al. Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome Res* **27**, 1843–1858 (2017).
22. Iancu, O. D. et al. Cosplicing network analysis of mammalian brain RNA-Seq data utilizing WGCNA and Mantel correlations. *Front Genet* **6**, 174 (2015).
23. Li, W., Dai, C., Liu, C. C. & Zhou, X. J. Algorithm to identify frequent coupled modules from two-layered network series: application to study transcription and splicing coupling. *J. Comput Biol.* **19**, 710–730 (2012).
24. Papasaikas, P., Rao, A., Huggins, P., Valcarcel, J. & Lopez, A. Reconstruction of composite regulator-target splicing networks from high-throughput transcriptome data. *BMC Genomics* **16**, S7 (2015).
25. Michielsen, L. et al. Predicting cell-type-specific exon inclusion in the human brain reveals more complex splicing mechanisms in neurons than glia. *bioRxiv*, https://doi.org/10.1101/2024.03.18.585465 (2024).
26. Cheng, J., Celik, M. H., Kundaje, A. & Gagneur, J. MTSplice predicts effects of genetic variants on tissue-specific splicing. *Genome Biol.* **22**, 94 (2021).
27. Xiong, H. Y. et al. RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806 (2015).
28. Barash, Y. et al. Deciphering the splicing code. *Nature* **465**, 53–59 (2010).
29. Jha, A., Gazzara, M. R. & Barash, Y. Integrative deep models for alternative splicing. *Bioinformatics* **33**, i274–i282 (2017).
30. Gracida, X., Norris, A. D. & Calarco, J. A. Regulation of tissue-specific alternative splicing: C. elegans as a model system. *Adv. Exp. Med Biol.* **907**, 229–261 (2016).
31. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
32. Tintori, S. C., Osborne Nishimura, E., Golden, P., Lieb, J. D. & Goldstein, B. A transcriptional lineage of the early C. elegans embryo. *Dev. Cell* **38**, 430–444 (2016).
33. Packer, J. S. et al. A lineage-resolved molecular atlas of C. elegans embryogenesis at single-cell resolution. *Science* **365**, eaax1971 (2019).
34. Taylor, S. R. et al. Molecular topography of an entire nervous system. *Cell* **184**, 4329–4347 e4323 (2021).
35. Roux, A. E. et al. Individual cell types in C. elegans age differently and activate distinct cell-protective responses. *Cell Rep.* **42**, 112902 (2023).
36. Ghaddar, A. et al. Whole-body gene expression atlas of an adult metazoan. *Sci. Adv.* **9**, eadg0506 (2023).
37. Smith, J. J. et al. A molecular atlas of adult C. elegans motor neurons reveals ancient diversity delineated by conserved transcription factor codes. *Cell Rep.* **43**, 113857 (2024).
38. Large C. R. L. et al. Lineage-resolved analysis of embryonic gene expression evolution in *C. elegans* and *C. briggsae*. *Preprint at* bioRxiv, https://doi.org/10.1101/2024.02.03.578695 (2024).
39. Koterniak, B. et al. Global regulatory features of alternative splicing across tissues and within the nervous system of C. elegans. *Genome Res* **30**, 1766–1780 (2020).
40. Ramani, A. K. et al. Genome-wide analysis of alternative splicing in Caenorhabditis elegans. *Genome Res* **21**, 342–348 (2011).
41. Kaletsky, R. et al. Transcriptome analysis of adult Caenorhabditis elegans cells reveals tissue-specific gene and isoform expression. *PLoS Genet* **14**, e1007559 (2018).
42. Barrett, A. et al. Integrating bulk and single cell RNA-seq refines transcriptomic profiles of individual *C. elegans* neurons. *Preprint* bioRxiv, https://doi.org/10.1101/2025.01.26.63495 (2022).
43. Barrett, A. et al. A head-to-head comparison of ribodepletion and polyA selection approaches for Caenorhabditis elegans low input RNA-sequencing libraries. *G3 (Bethesda)* **11**, jkab121 (2021).
44. Davis, P. et al. WormBase in 2022-data, processes, and tools for analyzing Caenorhabditis elegans. *Genetics* **220**, iyac003 (2022).
45. Hwang, S. B. & Lee, J. Neuron cell type-specific SNAP-25 expression driven by multiple regulatory elements in the nematode Caenorhabditis elegans. *J. Mol. Biol.* **333**, 237–247 (2003).
46. Diesh, C. et al. JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol.* **24**, 74 (2023).
47. Vaquero-Garcia, J. et al. A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife* **5**, e11752 (2016).
48. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
49. 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 (2021).

50. Tomioka, M., Naito, Y., Kuroyanagi, H. & Iino, Y. Splicing factors control C. elegans behavioural learning in a single neuron by producing DAF-2c receptor. *Nat. Commun.* **7**, 11645 (2016).
51. Yu, T. W., Hao, J. C., Lim, W., Tessier-Lavigne, M. & Bargmann, C. I. Shared receptors in axon guidance: SAX-3/Robo signals via UNC-34/Enabled and a Netrin-independent UNC-40/DCC function. *Nat. Neurosci.* **5**, 1147–1154 (2002).
52. Chan, S. S. et al. UNC-40, a C. elegans homolog of DCC (Deleted in Colorectal Cancer), is required in motile cells responding to UNC-6 netrin cues. *Cell* **87**, 187–195 (1996).
53. Leyva-Díaz, E. et al. Alternative splicing controls pan-neuronal homeobox gene expression. *Genes & Dev.* **39**, 209–220 (2025).
54. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
55. Angeles-Albores, D., N Lee, R. Y., Chan, J. & Sternberg, P. W. Tissue enrichment analysis for C. elegans genomics. *BMC Bioinforma.* **17**, 366 (2016).
56. Hobert, O. The neuronal genome of Caenorhabditis elegans. *WormBook* **13**, 1–106 (2013).
57. Dam, S. H., Olsen, L. R. & Vitting-Seerup, K. Expression and splicing mediate distinct biological signals. *BMC Biol.* **21**, 220 (2023).
58. Wolfe, Z., Liska, D. & Norris, A. Deep transcriptomics reveals cell-specific isoforms of pan-neuronal genes. *Nat. Commun* https://doi.org/10.1038/s41467-025-58296-2 (2025).
59. Irimia, M. et al. A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159**, 1511–1523 (2014).
60. Tong, J., Yang, J. Y., Xi, J. T., Yu, Y. G. & Ogunbona, P. O. Tuning the parameters for precision matrix estimation using regression analysis. *Ieee Access* **7**, 90585–90596 (2019).
61. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl Genet Mol. Biol.* **4**, Article17 (2005).
62. Friedman, J., Hastie, T. & Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**, 432–441 (2008).
63. Hsieh, C.-J., Dhillon, I., Ravikumar, P. & Sustik, M. Sparse inverse covariance matrix estimation using quadratic approximation. *Advances in Neural Information Processing Systems* **24**, 2330–2338 (2011).
64. Cai, T., Liu, W. D. & Luo, X. A. Constrained minimization approach to sparse precision matrix estimation. *J. Am. Stat. Assoc.* **106**, 594–607 (2011).
65. Liu, W. D. & Luo, X. Fast and adaptive sparse precision matrix estimation in high dimensions. *J. Multivar. Anal.* **135**, 153–162 (2015).
66. Liu, H., Lafferty, J. & Wasserman, L. The nonparanormal: semiparametric estimation of high dimensional undirected graphs. *J. Mach. Learn Res* **10**, 2295–2328 (2009).
67. Norris, A. D., Gracida, X. & Calarco, J. A. CRISPR-mediated genetic interaction profiling identifies RNA binding proteins controlling metazoan fitness. *Elife* **6**, e28129 (2017).
68. Kuroyanagi, H., Watanabe, Y. & Hagiwara, M. CELF family RNA-binding protein UNC-75 regulates two sets of mutually exclusive exons of the unc-32 gene in neuron-specific manners in Caenorhabditis elegans. *PLoS Genet* **9**, e1003337 (2013).
69. Kuroyanagi, H., Watanabe, Y., Suzuki, Y. & Hagiwara, M. Position-dependent and neuron-specific splicing regulation by the CELF family RNA-binding protein UNC-75 in Caenorhabditis elegans. *Nucleic Acids Res* **41**, 4015–4025 (2013).
70. Milne, C. A. & Hodgkin, J. ETR-1, a homologue of a protein linked to myotonic dystrophy, is essential for muscle development in Caenorhabditis elegans. *Curr. Biol.* **9**, 1243–1246 (1999).
71. Ochs, M. E., McWhirter, R. M., Unckless, R. L., Miller, D. M. 3rd & Lundquist, E. A. Caenorhabditis elegans ETR-1/CELF has broad effects on the muscle cell transcriptome, including genes that regulate translation and neuroblast migration. *BMC Genomics* **23**, 13 (2022).
72. Martinez, B. A. & Gill, M. S. The SR protein RSP-2 influences expression of the truncated insulin receptor DAF-2B in Caenorhabditis elegans. *G3 (Bethesda)* **13**, jkad064 (2023).
73. Choudhary, B., Marx, O. & Norris, A. D. Spliceosomal component PRP-40 is a central regulator of microexon splicing. *Cell Rep.* **36**, 109464 (2021).
74. Norris, A. D. et al. A pair of RNA-binding proteins controls networks of splicing events contributing to specialization of neural cell types. *Mol. Cell* **54**, 946–959 (2014).
75. Booeshaghi, A. S. et al. Isoform cell-type specificity in the mouse primary motor cortex. *Nature* **598**, 195–199 (2021).
76. Karlsson, K. & Linnarsson, S. Single-cell mRNA isoform diversity in the mouse brain. *BMC Genomics* **18**, 126 (2017).
77. Sugino, K. et al. Mapping the transcriptional diversity of genetically and anatomically defined cell populations in the mouse brain. *Elife* **8**, e38619 (2019).
78. Kawasawa, Y. I. et al. RNA-seq analysis of developing olfactory bulb projection neurons. *Mol. Cell Neurosci.* **74**, 78–86 (2016).
79. Zhang, Y. et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34**, 11929–11947 (2014).
80. Levin, M., Hashimshony, T., Wagner, F. & Yanai, I. Developmental milestones punctuate gene expression in the Caenorhabditis embryo. *Dev. Cell* **22**, 1101–1108 (2012).
81. Chen, L. & Zheng, S. Studying alternative splicing regulatory networks through partial correlation analysis. *Genome Biol.* **10**, R3 (2009).
82. Dai, C., Li, W., Liu, J. & Zhou, X. J. Integrating many co-splicing networks to reconstruct splicing regulatory modules. *BMC Syst. Biol.* **6**, S17 (2012).
83. Barberan-Soler, S. & Zahler, A. M. Alternative splicing regulation during C. elegans development: splicing factors as regulated targets. *PLoS Genet* **4**, e1000001 (2008).
84. Gupta, I. et al. Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat. Biotechnol.* **36**,1197–1202 (2018).
85. Joglekar, A. et al. Single-cell long-read sequencing-based mapping reveals specialized splicing patterns in developing and adult mouse and human brain. *Nat. Neurosci.* **27**, 1051–1063 (2024).
86. Hardwick, S. A. et al. Targeted, high-resolution RNA sequencing of non-coding genomic regions associated with neuropsychiatric functions. *Front Genet* **10**, 309 (2019).
87. Spencer, W. C. et al. Isolation of specific neurons from C. elegans larvae for gene expression profiling. *PLoS One* **9**, e112102 (2014).
88. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**, 491–499 (2017).
89. Brenner, S. The genetics of Caenorhabditis elegans. *Genetics* **77**, 71–94 (1974).
90. uit het Broek M., How to linearize the product of two binary variables? *Operations Research Stack Exchange* https://or.stackexchange.com/q/38 (2019).
91. Kuroyanagi, H., Kobayashi, T., Mitani, S. & Hagiwara, M. Transgenic alternative-splicing reporters reveal tissue-specific expression profiles and regulation mechanisms in vivo. *Nat. Methods* **3**, 909–915 (2006).
92. Kabat, J. L., Barberan-Soler, S. & Zahler, A. M. HRP-2, the Caenorhabditis elegans homolog of mammalian heterogeneous nuclear ribonucleoproteins Q and R, is an alternative splicing factor that binds to UCUAUC splicing regulatory elements. *J. Biol. Chem.* **284**, 28490–28497 (2009).
93. Calixto, A., Ma, C. & Chalfie, M. Conditional gene expression and RNAi using MEC-8-dependent splicing in C. elegans. *Nat. Methods* **7**, 407–411 (2010).

94. Galvin, B. D., Denning, D. P. & Horvitz, H. R. SPK-1, an SR protein kinase, inhibits programmed cell death in Caenorhabditis elegans. *Proc. Natl Acad. Sci. USA* **108**, 1998–2003 (2011).
95. Ohno, G. et al. Muscle-specific splicing factors ASD-2 and SUP-12 cooperatively switch alternative pre-mRNA processing patterns of the ADF/cofilin gene in Caenorhabditis elegans. *PLoS Genet* **8**, e1002991 (2012).
96. Ohno, G., Hagiwara, M. & Kuroyanagi, H. STAR family RNA-binding protein ASD-2 regulates developmental switching of mutually exclusive alternative splicing in vivo. *Genes Dev.* **22**, 360–374 (2008).
97. Wang, G. J. et al. GRLD-1 regulates cell-wide abundance of glutamate receptor through post-transcriptional regulation. *Nat. Neurosci.* **13**, 1489–1495 (2010).
98. Barberan-Soler, S., Medina, P., Estella, J., Williams, J. & Zahler, A. M. Co-regulation of alternative splicing by diverse splicing factors in Caenorhabditis elegans. *Nucleic Acids Res* **39**, 666–674 (2011).
99. Tan, J. H. & Fraser, A. G. The combinatorial control of alternative splicing in C. elegans. *PLoS Genet* **13**, e1007033 (2017).
100. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).
101. Trincado, J. L. et al. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol.* **19**, 40 (2018).
102. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol.* **9**, e1003118 (2013).
103. Siepel, A. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
104. Schafer, S. et al. Alternative splicing signatures in RNA-seq data: percent spliced in (PSI). *Curr. Protoc. Hum. Genet* **87**, 11 16 11–11 16 14 (2015).
105. Thompson, M. et al. Splicing in a single neuron is coordinately controlled by RNA binding proteins and transcription factors. *eLife* **8**, e46726 (2019).
106. Choudhary, B., Marx, O. & Norris, A. D. Spliceosomal component PRP-40 is a central regulator of microexon splicing. *Cell Reports* **36**, 109464 (2021).

Acknowledgements

We would like to thank members of the Marc Hammarlund, Shaul Yogev, and Karla Neugebauer labs for helpful discussion and advice. The CeNGEN project is supported by NIH grant R01NS100547. A.W. is supported by Surdna Foundation and the Yale Genetics Venture Fund. E.V. is supported by grant R00MH128772. Sequencing was done with YCGA: research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number 1S10OD030363-01A1. Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440).

Author contributions

A.W. contributed to the acquisition, analysis, and interpretation of data and writing of the manuscript. E.V. and A.B. contributed to the analysis of data. A.B., R.M.M., S.R.T., I.C., M.B., A.P., J.A.T., B.C. contributed to the acquisition of data. S.K., D.M.M., and M.H. contributed to the conception of the work and to writing the manuscript. The CeNGEN Consortium authors were involved in establishing the bulk sequencing dataset on which this splicing analysis is based.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-58293-5.

Correspondence and requests for materials should be addressed to Smita Krishnaswamy, David M. Miller III or Marc Hammarlund.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025

The CeNGEN Consortium

Cyril Cros[7], Berta Vidal[7], Maryam Majeed[7], Chen Wang[7], Emily A. Bayer[7], Molly Reilly[7], Eviatar Yemini[7,8], HaoSheng Sun[7,9] & Oliver Hobert[7]

[7]Department of Biological Sciences, Howard Hughes Medical Institute Columbia University, New York, NY, USA. [8]Present address: Department of Neurobiology, University of Massachusetts Chan Medical School, Worcester, MA, USA. [9]Present address: Department of Cell, Developmental, and Integrative Biology, University of Alabama at Birmingham, Birmingham, AL, USA. A full list of members and their affiliations appears in the Supplementary Information.