The CeNGEN Consortium

List of consortium members ordered by first name:

Abigail Poff, Alec Barrett, Alexis Weinreb, Becca Collings, Berta Vidal, Chen Wang, Cyril Cros, David M. Miller, III, Emily A. Bayer, Erdem Varol, Eviatar Yemini, HaoSheng Sun, Isabel Courtney, John A. Tipps, Manasa Basavaraju, Marc Hammarlund, Maryam Majeed, Oliver Hobert, Rebecca M. McWhirter, Seth R. Taylor, Smita Krishnaswamy

Department of Genetics, Yale University School of Medicine, New Haven, CT, USA

Alec Barrett, Alexis Weinreb, Manasa Basavaraju, Marc Hammarlund

Department of Neuroscience, Yale University School of Medicine, New Haven, CT, USA

Alec Barrett, Alexis Weinreb, Manasa Basavaraju, Marc Hammarlund

Department of Biological Sciences, Howard Hughes Medical Institute Columbia University, New York, NY, USA

Berta Vidal, Chen Wang, Cyril Cros, Emily A. Bayer, Eviatar Yemini, HaoSheng Sun, Maryam Majeed, Oliver Hobert

Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN, USA

Abigail Poff, Becca Collings, David M. Miller, III, Isabel Courtney, John A. Tipps, Rebecca M. McWhirter, Seth R. Taylor

Tandon School of Engineering, New York University, New York, NY, USA

Erdem Varol

Department of Computer Science, Yale University, New Haven, CT, USA

Smita Krishnaswamy

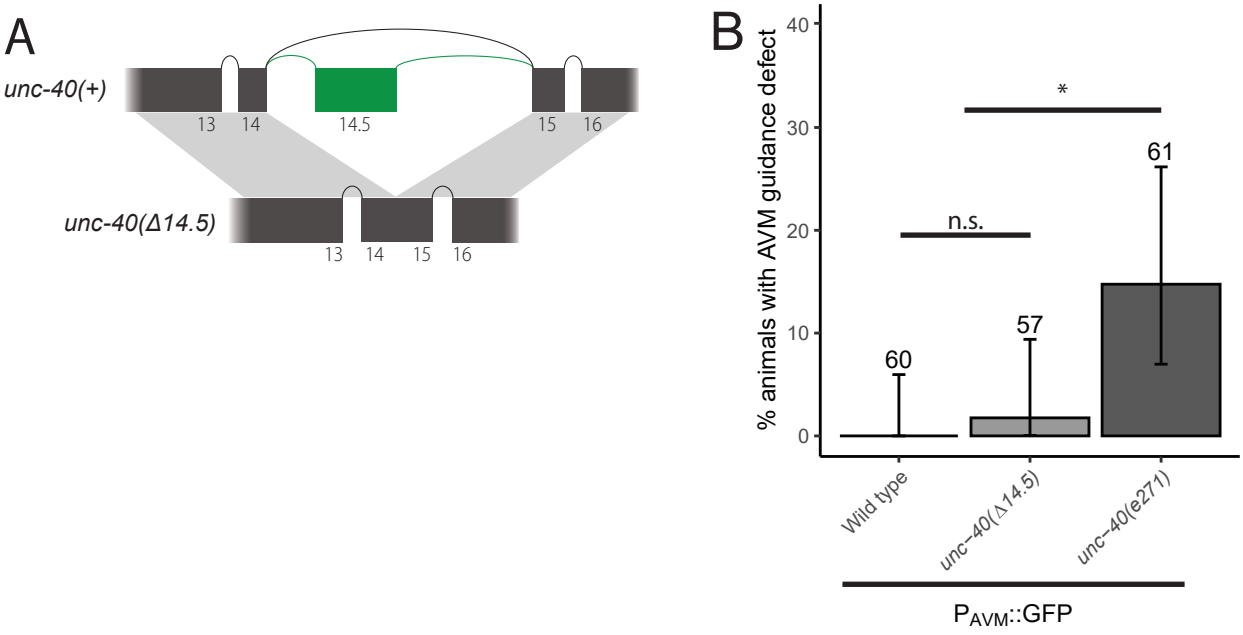

Figure S1: AVM guidance in *unc-40* mutants.
A: Schematic view of the effect of CRISPR deletion of exon 14.5 on the genome. **B:** Proportion of individuals displaying an AVM guidance phenotype, for N2 (wild type), *unc-40(Δ14.5)* mutant lacking exon 14.5, and *unc-40(e271)* nonsense mutant animals. The number of individuals is indicated (collected from 3 biological replicates). Error bars represent 95 % confidence intervals (Clopper and Pearson method). WT vs *e271*, p=0.018; WT vs *Δ14.5*, p=0.979 (proportion tests and Holm correction for multiple comparisons). Source data are provided as a Source Data file.

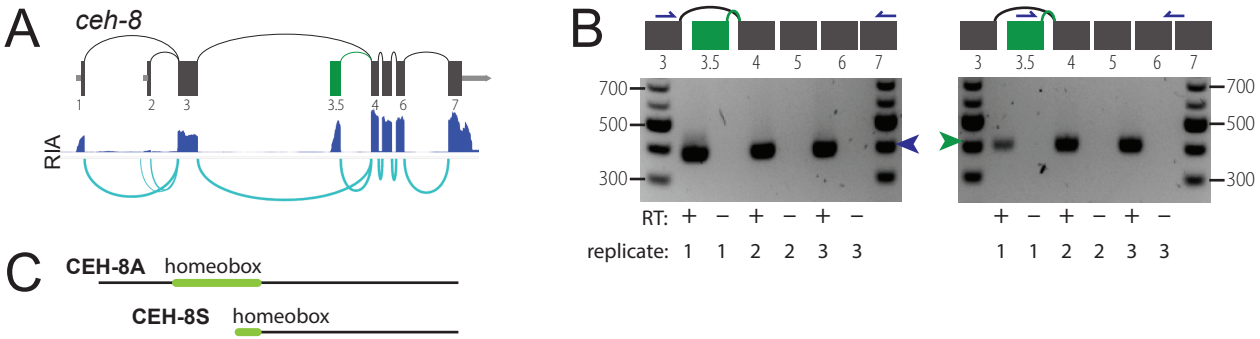

Figure S2: Detection of novel alternative first exon in *ceh-8*.
A: Alternative expression of the novel first exon 3.5 in *ceh-8*. Raw data visualization in neuron RIA, with novel alternative first exon 3.5 (green) vs annotated transcript (grey). **B:** Validation by RT-PCR of the novel exons in cDNA from whole animals (N2 wild type). Top: location of the primers relative to the gene. Left, the primer pair only results in amplification of the annotated transcripts including exon 3; right, the primer pair only results in amplification of the novel transcript expressing alternative first exon 3.5. Arrowheads indicate the expected product sizes (406 bp and 423 bp). For each experiment, the amplification was performed on 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. **C:** Structure of the annotated protein isoform CEH-8A and the novel short isoform starting at N-terminal alternative exon 3.5 (CEH-8S).

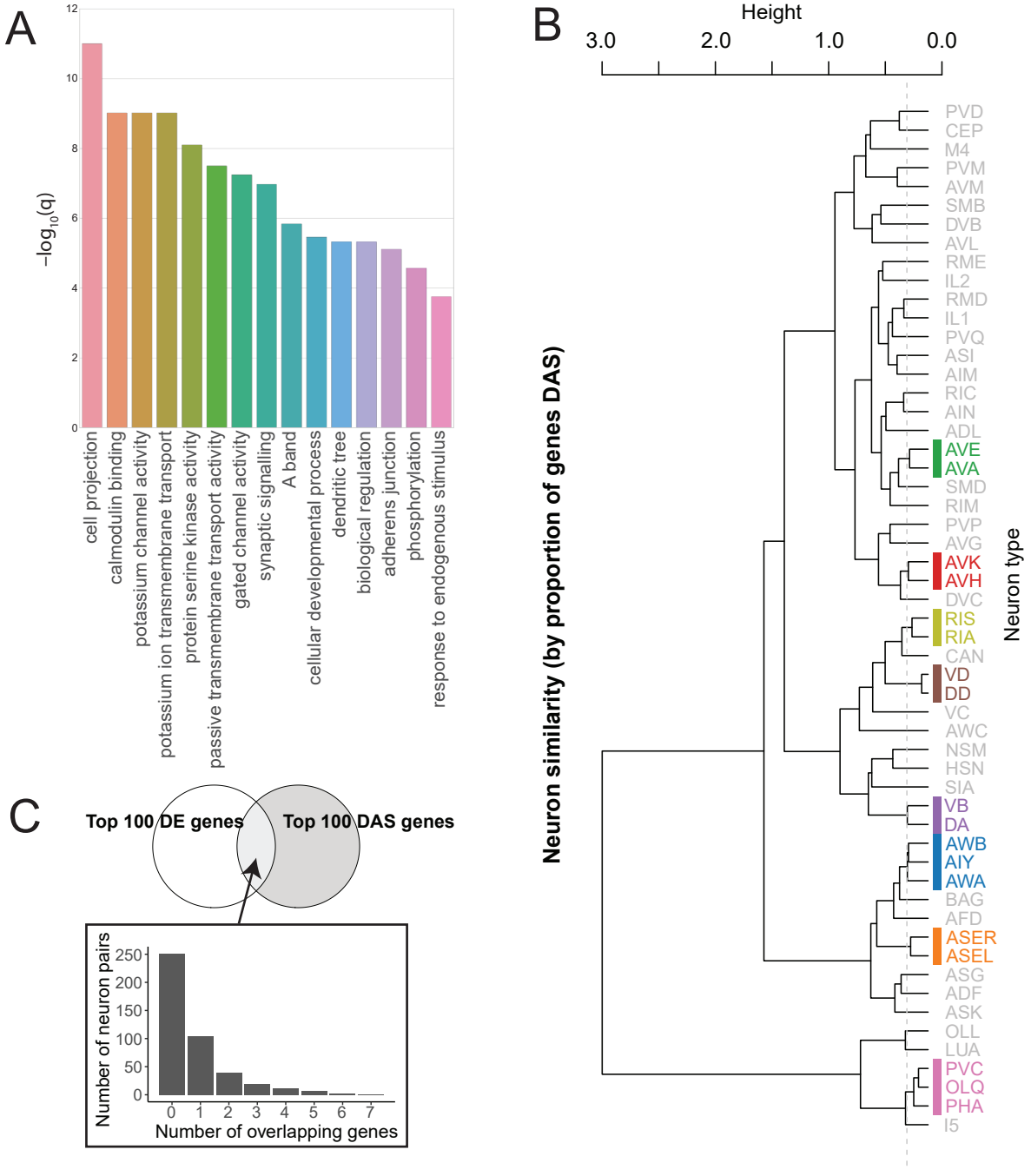

Figure S3: Analysis of DAS genes.
A: Gene Ontology terms enriched among the genes with DAS in neurons. **B:** Pairs of neurons with similar profiles, clustered by proportion of DAS genes among co-expressed genes, corresponding to the heatmap in Fig. 5C. The color blocks indicate clusters identified from a tree cut at a height of 0.2 (grey dashed line). See Fig. 5C for Source Data. **C:** Overlap between the top 100 DAS genes and the top 100 differentially expressed genes for each neuron pair. Source data are provided as a Source Data file.

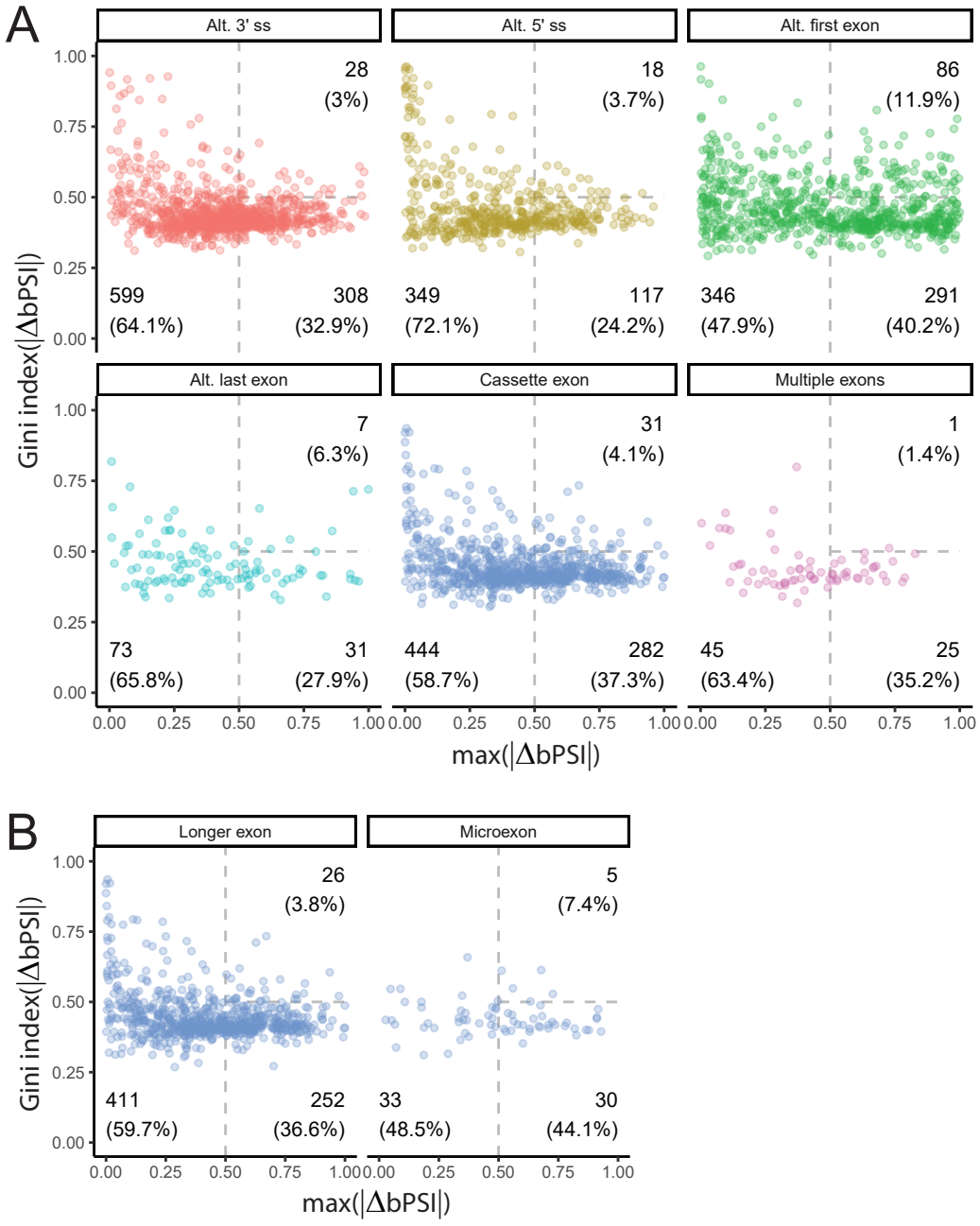

Figure S4: Variability of canonical event types.
A, B: Individual events of each canonical type measured by SUPPA. The horizontal axis indicates the maximal absolute difference in bPSI for each event across neuron types, such that an event with low maximum absolute delta bPSI displays a similar bPSI level in all neuron types, and an event with high absolute delta bPSI displays a large difference for at least one pair of neuron types. The vertical axis represents the Gini coefficient of the absolute differences in bPSI across all neuron pairs, such that an event with low Gini coefficient displays similar values of delta bPSI across neuron pairs, while a high Gini coefficient indicates an event for which few neuron pairs display a large difference in bPSI. The dashed lines delimitate values of 0.5, the number of events in each quadrant are indicated. The events belonging to each canonical type are represented (A), while the cassette exons are further separated (B) between microexons (≤ 27 bp) and larger exons (>27 bp). Source data are provided as a Source Data file.

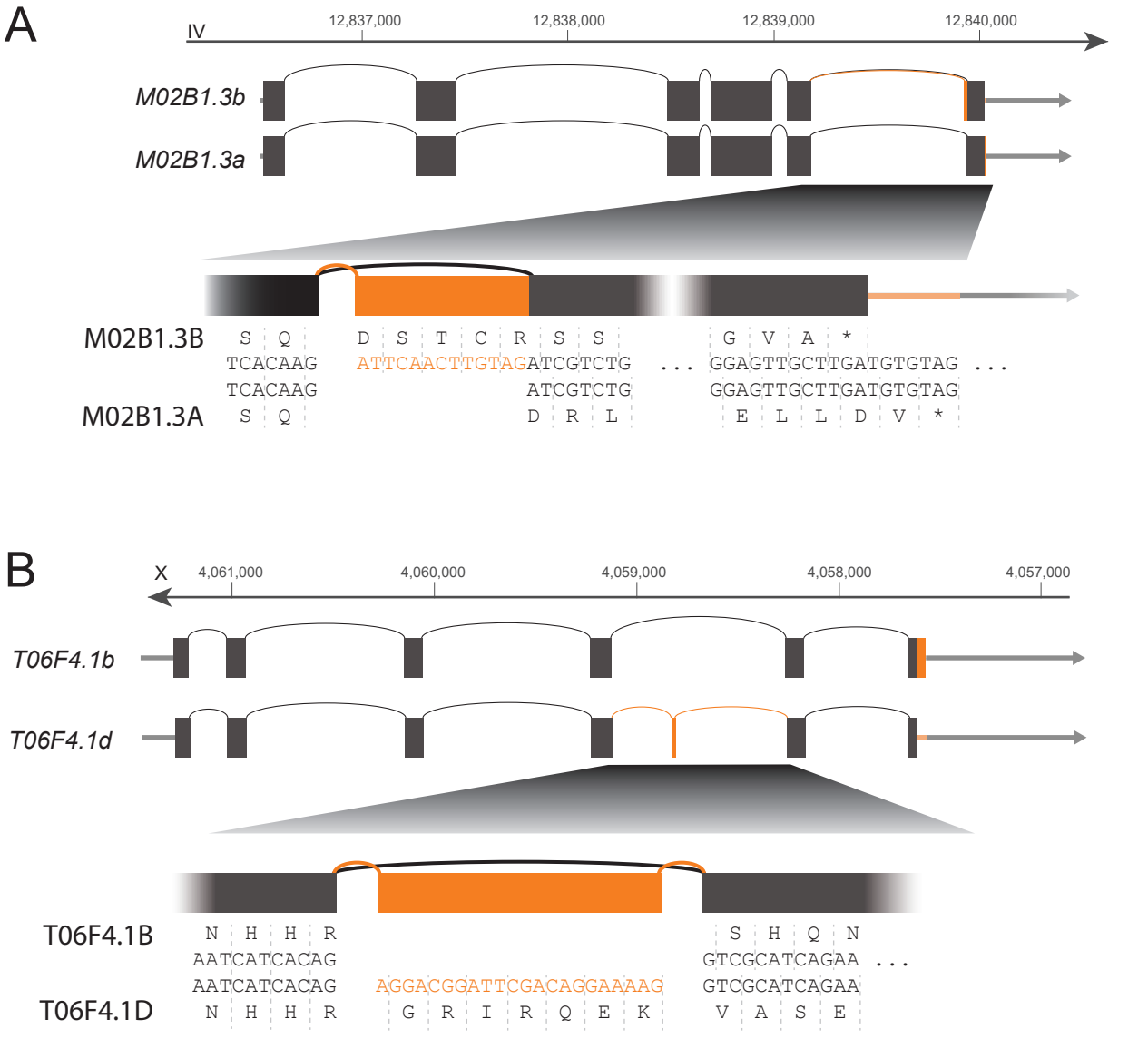

Figure S5: Illustrative examples of alternative events leading to a frame shift.
A: Alternative 3' splice site in the gene *M02B1.3* leading to inclusion of 13 bp. The induced frame shift results in two protein isoforms with differing C terminus. **B:** Alternative exon in *T06F4.1* leading to the inclusion of 22 bp. The induced frame shift results in protein isoforms with differing C terminus. This gene additionally contains alternative first exons encoding different N terminus (not shown).

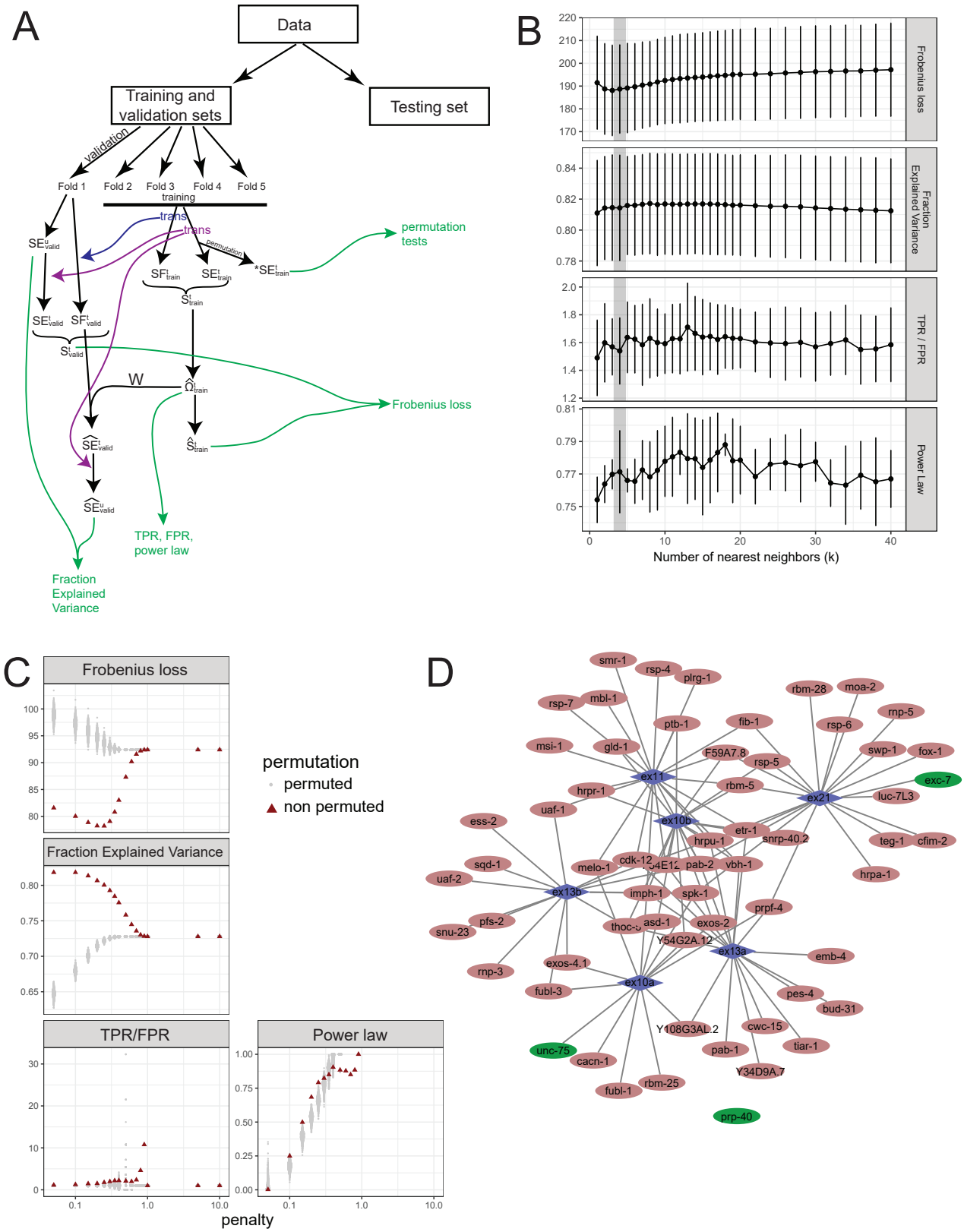

Figure S6: Network computation.
A: Schematic description of the full procedure. *SE*: skipped exon splicing measurements, *SF*: splice factor expressions, *S*: covariance matrix, *Ω*: precision matrix. The superscript *t* or *u* indicates whether the matrix is transformed or untransformed, the subscript *train* or *valid* indicates whether it contains data from the training or validation folds. The green arrows and text correspond to the metrics being computed. **B:** Selection of the optimal number of nearest neighbors. The four metrics are represented at a fixed penalty of 0.2 and plotted against the number of nearest neighbors. The grey line highlights the chosen value of *k*=4. C: Metrics plotted against penalty, the red triangles indicate the computed value, the grey dots were computed after permutation of the input. **D:** Subnetwork centered on the alternative exons of *unc-16*. Symbols as defined in Fig. 5D. Source data are provided as a Source Data file.

Supplementary table 1: list of oligonucleotides used in this work

application	target gene	binding region	orientation (mRNA/cDNA)	primer name	primer sequence
RT-PCR	unc-40	exons 12-13	>>	oAW695	TGAGAAAGGTCATGGCCCAT
RT-PCR	unc-40	exons 14.5-15	<<	oAW699	TGTTAGAGTTCCTTGCCCGT
RT-PCR	unc-40	exon 15	<<	oAW675	TGATGCCTTCCAGTTCCTGT
RT-PCR	sax-3	exon 5	>>	oAW692	GCTTATGTGTGCGCTGGAAT
RT-PCR	sax-3	exon 5.5	>>	oAW690	GTCTTCAGCTTGACTTCGGC
RT-PCR	sax-3	exons 6-7	<<	oAW702	CTCCGGTGTCAGGTTTCTTT
RT-PCR	ceh-8	exon 3	>>	oAW627	TGCAAGAGAAACATTGGCTGC
RT-PCR	ceh-8	exon 3.5	>>	oAW701	TTGGTTTTCTAATGGATGGTGC
RT-PCR	ceh-8	exons 6-7	<<	oAW700	TGGTGTTTGAGGCGATTCTTT
CRISPR	unc-40			crAW11	TAAAAATACCTTGTAGATGA.TGG
CRISPR	unc-40			crAW12	GAAAATTTCCGTTAAAAAAA.AGG
CRISPR	unc-40			oAW678	GCTTTGATGATAACTTCGTGTTAGAGTTCCTTGCC.CTTGTAGATGATGaTATCTAGGTGTGGGTGATTCA

Supplementary table 2: strains used in this work

strain	genotype	description
CZ10175	zdIs5 I	GFP in 6 touch receptor neurons (including AVM), otherwise Wild Type.
XE3354	unc-40(wp221) I; zdIs5 I	Deletion of exon 14.5 and its flanking introns, crossed with zdIs5. Designated as unc-40(Δ14.5) in the text.
XE3377	unx-40(e271) I;zdIs5 I	Nonsense mutation in unc-40. CB271 crossed with zdIs5.