CellSpliceNet: Explainable Multimodal Transformers for Splicing in *C. elegans* Neurons

Arman Afrasiyabi[1,2,♡], **Jake Kovalic**[4,♡], **Chen Liu**[1,♡], **Egbert Castro**[3,♡],
Alexis Weinreb[2,♡], Erdem Varol[6,♣], David M. Miller III[5,♠],
Marc Hammarlund[2,♡], Smita Krishnaswamy[1,2,7,♡]

Department of {[1]Computer Science, [2]Genetics, [3]Computational Biology and Biomedical Informatics,
[4]Applied Physics, [5]Cell & Developmental Biology, [6]Computer Science & Engineering,
[7]Applied Math, [8]Biological Sciences}

♣New York University ♠Vanderbilt University ♡Yale University

Abstract

Alternative splicing diversifies the transcriptome, yet its regulation remains difficult to decode. We present CellSpliceNet, an interpretable transformer model that predicts splicing outcomes across *C. elegans* neurons by integrating four modalities: long-range genomic sequence, local RNA regions of interest (ROIs), predicted RNA secondary structure, and cell-type–specific gene expression. Modality-specific encoders—including graph-signal scattering for structure and expression—feed a multimodal multi-head attention module that preserves per-modality signals while enabling expression-informed interactions with sequence and structure. Attention pooling highlights salient biology (e.g., splice boundaries and single-stranded loop regions) and enables deep model interpretability. On held-out data, CellSpliceNet outperforms strong baselines and achieves Spearman $\rho = 0.88$, with robust accuracy across neuron subtypes.

1 Introduction

RNA alternative splicing is a fundamental mechanism that promotes functional diversity in eukaryotic organisms by allowing a single gene to produce multiple mRNA and protein isoforms [1], thereby greatly expanding the functional repertoire encoded by the genome [2]. Recent work has shown the capability of transformers [3] and LLMs [4] to learn motif co-occurrence from large sequence corpora and surpass classic tools like GeneSplicer and MaxEntScan [5, 6, 7, 8, 9], however comprehensive modeling of RNA alternative splicing remains a challenge. These models often miss junction-proximal regulatory motifs, secondary-structure constraints, and cell-specific factors, limiting mechanistic insight into splicing variation. Major obstacles include RNA's four-letter alphabet (weakening standard tokenization), ambiguous splice-related motifs that boost false discoveries, noise from weakly constrained noncoding regions [10, 11], and the difficulty of modeling cell-specific regulation driven by interacting splice factors.

To address these challenges, we propose CellSpliceNet, a deep multimodal transformer-based framework that integrates four key sources of information: the full-length RNA sequence, its predicted secondary structure, regions of interest (ROIs) around splice junctions, and single-cell gene expression profiles. By explicitly modeling junction-proximal sequence context—where regulatory elements often reside—and rendering secondary structures as nucleotide–base-pair graphs, a *novel the employ of geometric scattering* on these graphs (with Dirac node signals) to distill multiscale structural cues. This design enables CellSpliceNet to capture both global and local features critical to splice-site selection while remaining mechanistically interpretable. CellSpliceNet delivers state-of-the-art performance alongside fine-grained interpretability for dissecting splicing regulation.

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

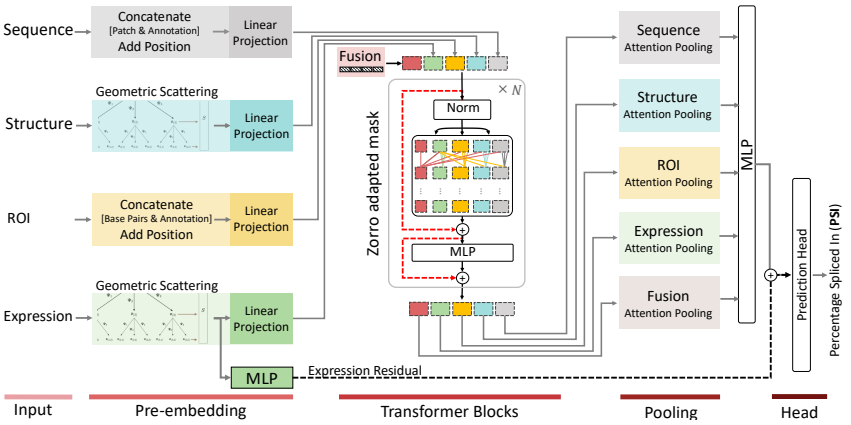

Figure 1: `CellSpliceNet` integrates multimodal inputs including (global) sequence, structure, (local) region of interest and expression to predict the percentage spliced in (PSI) metric.

2 Methods

Model Overview and Objective We model exon inclusion by predicting percent spliced in (PSI)—the probability an exon appears in the mature transcript. The model (Fig. 1) is trained end-to-end to minimize mean squared error between predicted and observed PSI. To capture diverse splicing mechanisms, we encode four complementary modalities (detailed below).

Full-length RNA sequence To capture long-range regulatory signals that affect splicing, we tokenize the full-length RNA sequence using a patch-based strategy. Using single nucleotides (A, T, C, and G) as tokens would confine a computational model to a shallow four-symbol vocabulary [12], providing little higher-order context; therefore, we instead group contiguous bases into overlapping k-mer "patches." Each patch is projected into an embedding space through a linear layer, following the Vision Transformer paradigm [13].

RNA secondary structure RNA folding can influence exon inclusion by modulating the accessibility of splice sites and nearby regulatory motifs, a mechanism known as structural masking [14]. To model this, we extract the intronic regions flanking each target exon and predict their secondary structures using ViennaRNA [15]. These structures are converted into graphs, with nodes representing nucleotides and edges representing base pairs. We then cast the structures as nucleotide–base-pair graphs and *pioneer the use of geometric scattering* on these graphs (see supp. mat.), with Dirac node signals, to distill multiscale features relevant to exon inclusion [16].

Regions of interest Regions near splice junctions are enriched for cis-regulatory elements such as enhancers and silencers that modulate exon inclusion. To isolate these signals, we define regions of interest (ROIs) that span 200 base pairs upstream and downstream of each splice junction. These local sequence windows are embedded independently from the full-length transcript to emphasize short-range sequence features critical to splicing.

Gene expression and splicing-factor dynamics Splicing varies across cell types with splicing-factor (SF) expression and activity. We integrate single-cell RNA-seq (imputed with MAGIC [17]) for 243 curated SFs, building neuron-type–stratified expression matrices. To capture co-regulation, we compute pairwise mutual information and discretize it to form SF co-expression graphs (nodes = SFs; edges = significant dependencies). Each graph is encoded with geometric scattering [16] using Dirac signals weighted by average expression (App. A), yielding a compact representation of the regulatory context shaping exon selection.

Fusing heterogeneous modalities Each of the four modalities is processed by a dedicated Transformer encoder. To enable selective exchange while preserving modality-specific learning, we

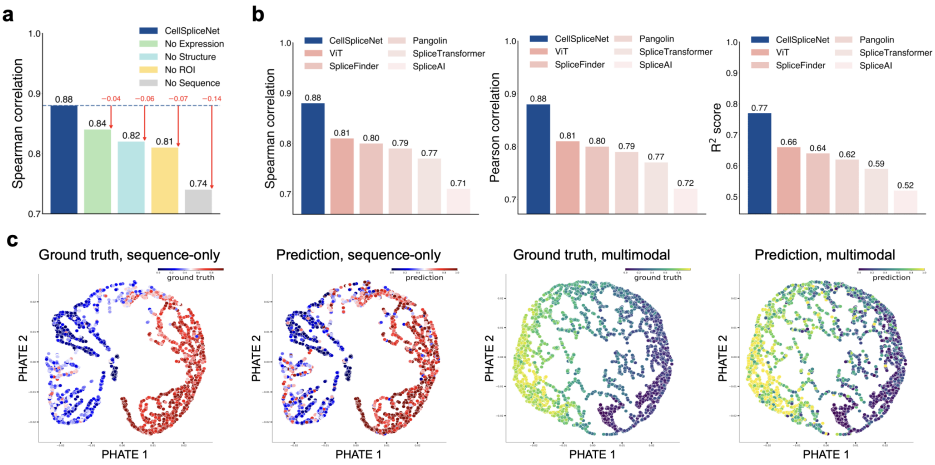

Figure 2: `CellSpliceNet` is highly competitive for PSI prediction. **(a)**. The ablation study **(b)**. Comparisons of the `CellSpliceNet` outperforms 5 different methods on Spearman correlation, Pearson correlation, and R^2. **(c)**. PHATE visualizations of test samples color-coded by PSI values.

use a Zorro-inspired attention mask [18] that blocks cross-modal attention except via designated fusion tokens; expression is allowed to influence all others to reflect SF control over sequence and structure. The resulting embeddings are concatenated and passed to an MLP to predict PSI, trained end-to-end with a squared-error loss (Eqn. (1)).

$$\mathcal{L} = (y_i - \hat{y}_i)^2; \qquad \hat{y}_i = \mathrm{LogSigmoid}\big(\mathrm{MLP}(\mathbf{z}_{\mathrm{cat},i}) + \mathrm{MLP}(\mathbf{z}_{\mathrm{e,i}})\big) \tag{1}$$

Here, $\mathbf{z}_{\mathrm{cat},i}$ represents the concatenated embeddings from all modalities and $\mathbf{z}_{\mathrm{e},i}$ is the residual expression component. This formulation ensures predictions remain within biologically meaningful bounds while penalizing large deviations from ground truth.

Modality-specific interpretation mechanism `CellSpliceNet` is a multimodal multi-head transformer to integrate the four modalities along with a set of learnable fusion tokens. While conceptually related to a recent work on multimodal protein analysis [19], our approach introduces more advanced techniques for information fusion. As previously described, the total number of tokens for the {sequence, structure, ROI, expression} modalities are $\{N, S, B, K\}$ respectively, where each token is embedded into $\mathbb{R}^D$. Concretely, we denote:

$$\mathbf{Z}^{\mathrm{seq}} = \{\mathbf{z}_n^{\mathrm{seq}}\}_{n=1}^{N}, \quad \mathbf{Z}^{\mathrm{roi}} = \{\mathbf{z}_b^{\mathrm{roi}}\}_{b=1}^{B}, \quad \mathbf{Z}^{\mathrm{stc}} = \{\mathbf{z}_s^{\mathrm{stc}}\}_{s=1}^{S}, \quad \mathbf{Z}^{\mathrm{exp}} = \{\mathbf{z}_s^{\mathrm{exp}}\}_{s=1}^{K},$$

where $\mathbf{z} \in \mathbb{R}^D$. In addition, we introduce P learnable fusion tokens: $\mathbf{Z}^{\mathrm{fus}} = \{\mathbf{z}_p^{\mathrm{fus}}\}_{p=1}^{P}, \quad \mathbf{z}_p^{\mathrm{fus}} \in \mathbb{R}^D$. The multimodal tokens are formed by concatenating the modality-specific tokens.

In a vanilla transformer [3], a standard multi-head self-attention module will be used to compute key $\mathbf{K}$, query $\mathbf{Q}$, and value $\mathbf{V}$ as learned projections of $\mathbf{Z}^{\mathrm{multimodal}}$. To do this, we used a `Zorro` attention mechanism [18], which accommodates specific attention constraints. Specifically, let $\mathsf{modality}(i) \in \{\mathsf{seq}, \mathsf{stc}, \mathsf{roi}, \mathsf{exp}, \mathsf{fus}\}$ denote the modality type of the i-th token. We define a `Zorro` mask $\mathbf{MASK} \in \{0,1\}^{L \times L}$, where $L = N + S + B + K + P$, is:

$$\mathbf{MASK}_{ij} = \begin{cases} 1 & \text{if } \mathsf{modality}(i) = \mathsf{modality}(j) \text{ or } \mathsf{modality}(i) \in \{\mathsf{exp}, \mathsf{fus}\} \\ 0 & \text{otherwise.} \end{cases}$$

Attention-based pooling and PSI prediction Processing the tokens with the multimodal transformer module is a dimension-preserving step, which means $\mathbf{H}^{\mathrm{multimodal}} \in \mathbb{R}^{(N+2S+B+P)\times D}$. We can therefore re-partition the multimodal hidden state by modality. This gives us the sequence hidden state ($\mathbf{H}^{\mathrm{seq}} \in \mathbb{R}^{N\times D}$), structure hidden state ($\mathbf{H}^{\mathrm{stc}} \in \mathbb{R}^{S\times D}$), ROI hidden state ($\mathbf{H}^{\mathrm{roi}} \in \mathbb{R}^{B\times D}$), expression hidden state ($\mathbf{H}^{\mathrm{exp}} \in \mathbb{R}^{S\times D}$), and fusion hidden state ($\mathbf{H}^{\mathrm{fus}} \in \mathbb{R}^{P\times D}$). For each modality, we apply attention pooling to aggregate the hidden state into a single vector, following RELSO [20]. Concretely, for each modality-specific hidden state $\mathbf{H}^{\mathrm{mod}}$, we learn a low-dimensional attention score:

$$\boldsymbol{\beta}^{\mathrm{mod}} = \mathrm{Softmax}(\theta_2^{\mathrm{mod}}\, \sigma(\mathrm{BatchNorm}(\theta_1^{\mathrm{mod}}(\mathbf{H}^{\mathrm{mod}})^\top))),$$

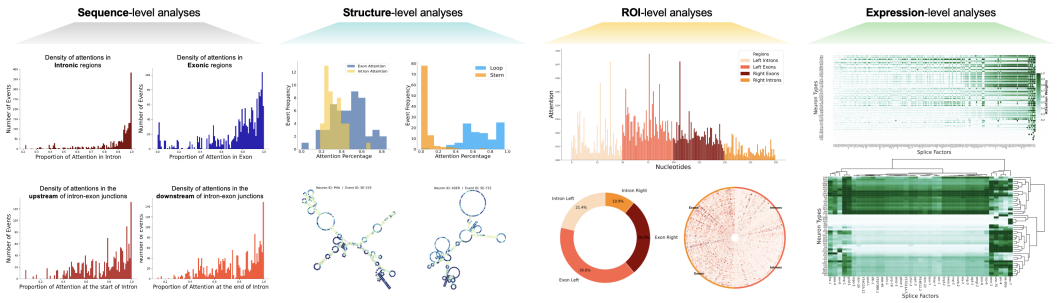

Figure 3: Illustration of modality-specific attention analyses enabled by `CellSpliceNet`.

where β^{mod} is the attention weights across the input for the given modality, $\theta_1^{\text{mod}} \in \mathbb{R}^{d \times D}$ and $\theta_2^{\text{mod}} \in \mathbb{R}^{1 \times d}$ are learned parameters, BatchNorm denotes a batch normalization layer, and $\sigma(\cdot)$ denotes a GELU activation. The resulting attention weights β^{mod} are applied to $\mathbf{H}^{\text{mod}}$ to produce a single hidden state vector: $\mathbf{h}^{\text{mod}} = (\boldsymbol{\beta}^{\text{mod}})^\top \mathbf{H}^{\text{mod}} \in \mathbb{R}^D$. We obtain pooled hidden states for sequence, structure, ROI, expression, and fusion. These are concatenated: $\mathbf{h}_{\text{cat}} = \left[\mathbf{h}^{\text{seq}} || \mathbf{h}^{\text{stc}} || \mathbf{h}^{\text{roi}} || \mathbf{h}^{\text{exp}} || \mathbf{h}^{\text{fus}}\right]$. We then feed this concatenated vector $\mathbf{h}_{\text{cat}}$ into a k-layer MLP, and the output of the MLP is subsequently passed through a linear layer. The final prediction is obtained using a LogSigmoid activation function, effectively mapping the output to the log-probability space.

Results

Performance We ablated the four `CellSpliceNet` inputs to quantify their respective contributions, namely the full-length RNA sequence, ROI, secondary structure, and gene expression. The full multimodal `CellSpliceNet` achieved Spearman 0.88 on PSI prediction; removing any single modality reduced performance to 0.74 (No Sequence), 0.81 (No ROI), 0.82 (No Structure), and 0.84 (No Expression), underscoring their complementarity (Fig. 2a). For benchmarking, we compared `CellSpliceNet` with SpliceTransformer [21], SpliceFinder [22], SpliceAI [23], Pangolin [24], and ViT [13] using Spearman, Pearson, and R^2 (Fig. 2b). `CellSpliceNet` led by +0.07 on both correlations and +0.11 R^2 over ViT, with even larger margins over the four domain-specific baselines. Results are means across three random seeds, confirming multimodal integration yields state-of-the-art PSI prediction. Additionally, PHATE [25] embeddings were generated from the test set sequence modality prior to the final MLP layer (Fig. 2**c**). Embeddings are colored by target and predicted PSI values, revealing a close alignment between the learned latent space and the ground truth. Taken together, these results demonstrate that the multimodal architecture of `CellSpliceNet` advances the state of the art in splicing prediction.

Interpretability We probe the attention pooling layer after multi-head attention in `CellSpliceNet` to locate modality regions most predictive of PSI. Across neuron types, attention concentrates at intron–exon boundaries (Fig. 3, first column). Aggregated histograms show most weight on exonic and immediately flanking intronic patches, consistent with canonical splice sites and nearby auxiliary elements. Mapping the structure stream's attention onto predicted secondary structures for exon-centred windows reveals a strong bias toward unpaired nucleotides: loop residues, especially within exonic cores, receive the highest weights, while adjacent stems are de-emphasised (Fig. 3, second column). Genome-wide, loops significantly outscore stems.

In the ROI stream focusing on the two splice junctions, nucleotide-level attention peaks over the exon core and decays into flanking introns (Fig.3, third column). An "attention wheel" summary shows a systematic upstream bias; quantitatively, upstream exon, downstream exon, upstream intron, and downstream intron account for 39.6%, 28.0%, 21.4%, and 10.9% of ROI attention, respectively (Fig.3third column), despite architectural symmetry.

Finally, `CellSpliceNet` exposes heterogeneous splicing-factor usage across neurons (Fig.3, last column). Clustering factors in the top 85th percentile of attention (Fig.3, last column bottom row) highlights influential regulators such as *smu-1*, *unc-75*, *hrp-1*, *exc-7*, and *grld-1*.

References

[1] Walter Gilbert. Why genes in pieces? *Nature*, 271(5645):501–501, 1978.

[2] Timothy W Nilsen and Brenton R Graveley. Expansion of the eukaryotic proteome by alternative splicing. *Nature*, 463(7280):457–463, 2010.

[3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. *Advances in neural information processing systems*, 30, 2017.

[4] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving Language Understanding by Generative Pre-Training. *OpenAI*, 2018.

[5] Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Jason Liu, Demi Guo, Myle Ott, C Lawrence Zitnick, Jerry Ma, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, 118(15):e2016239118, 2021.

[6] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

[7] Roshan M Rao, Jason Liu, Robert Verkuil, Joshua Meier, John Canny, Pieter Abbeel, Tom Sercu, and Alexander Rives. MSA transformer. In *International Conference on Machine Learning*, pages 8844–8856. PMLR, 2021.

[8] Mihaela Pertea, Xiaoying Lin, and Steven L Salzberg. Genesplicer: a new computational method for splice site prediction. *Nucleic acids research*, 29(5):1185–1190, 2001.

[9] Gene Yeo and Christopher B Burge. Maximum entropy modeling of short sequence motifs with applications to rna splicing signals. In *Proceedings of the seventh annual international conference on Research in computational molecular biology*, pages 322–331, 2003.

[10] Aurélie Kapusta and Cédric Feschotte. Volatile evolution of long noncoding rna repertoires: mechanisms and biological implications. *Trends in Genetics*, 30(10):439–452, 2014.

[11] Chris P Ponting and Wilfried Haerty. Genome-wide analysis of human long noncoding rnas: a provocative review. *Annual review of genomics and human genetics*, 23(1):153–172, 2022.

[12] Roshan Rao, Nicholas Bhattacharya, Neil Thomas, Yan Duan, Peter Chen, John Canny, Pieter Abbeel, and Yun Song. Evaluating protein transfer learning with tape. In *Advances in Neural Information Processing Systems*, pages 9689–9701, 2019.

[13] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2020.

[14] Suzan Ruijtenberg, Stijn Sonneveld, Tao Ju Cui, Ive Logister, Dion De Steenwinkel, Yao Xiao, Ian J MacRae, Chirlmin Joo, and Marvin E Tanenbaum. mRNA structural dynamics shape Argonaute-target interactions. *Nature structural & molecular biology*, 27(9):790–801, 2020.

[15] Ronny Lorenz, Stephan H Bernhart, Christian Höner zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F Stadler, and Ivo L Hofacker. ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, 6:1–14, 2011.

[16] Feng Gao, Guy Wolf, and Matthew Hirn. Geometric Scattering for Graph Data Analysis. In *International Conference on Machine Learning*, pages 2122–2131. PMLR, 2019.

[17] David Van Dijk, Roshan Sharma, Juozas Nainys, Kristina Yim, Pooja Kathail, Ambrose J Carr, Cassandra Burdziak, Kevin R Moon, Christine L Chaffer, Diwakar Pattabiraman, et al. Recovering gene interactions from single-cell data using data diffusion. *Cell*, 174(3):716–729, 2018.

[18] Adrià Recasens, Jason Lin, João Carreira, Drew Jaegle, Luyu Wang, Jean-baptiste Alayrac, Pauline Luc, Antoine Miech, Lucas Smaira, Ross Hemsley, et al. Zorro: the masked multimodal transformer. *arXiv preprint arXiv:2301.09595*, 2023.

[19] Kevin Bijan Givechian, Joao Felipe Rocha, Edward Yang, Chen Liu, Kerrie Greene, Rex Ying, Etienne Caron, Akiko Iwasaki, and Smita Krishnaswamy. ImmunoStruct: a multimodal neural network framework for immunogenicity prediction from peptide-MHC sequence, structure, and biochemical properties. *bioRxiv*, pages 2024–11, 2024.

[20] Egbert Castro, Abhinav Godavarthi, Julian Rubinfien, Kevin Givechian, Dhananjay Bhaskar, and Smita Krishnaswamy. Transformer-based protein generation with regularized latent space optimization. *Nature Machine Intelligence*, 4(10):840–851, 2022.

[21] Ningyuan You, Chang Liu, Yuxin Gu, Rong Wang, Hanying Jia, Tianyun Zhang, Song Jiang, Jinsong Shi, Ming Chen, Min-Xin Guan, et al. SpliceTransformer predicts tissue-specific splicing linked to human diseases. *Nature Communications*, 15(1):9129, 2024.

[22] Ruohan Wang, Zishuai Wang, Jianping Wang, and Shuaicheng Li. SpliceFinder: ab initio prediction of splice sites using convolutional neural network. *BMC bioinformatics*, 20:1–13, 2019.

[23] Kishore Jaganathan, Sofia Kyriazopoulou Panagiotopoulou, Jeremy F McRae, Siavash Fazel Darbandi, David Knowles, Yang I Li, Jack A Kosmicki, Juan Arbelaez, Wenwu Cui, Grace B Schwartz, et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell*, 176(3):535–548, 2019.

[24] Tony Zeng and Yang I Li. Predicting RNA splicing from DNA sequence using Pangolin. *Genome Biology*, 23(1):103, 2022.

[25] Kevin R Moon, David Van Dijk, Zheng Wang, Scott Gigante, Daniel B Burkhardt, William S Chen, Kristina Yim, Antonia van den Elzen, Matthew J Hirn, Ronald R Coifman, et al. Visualizing structure and transitions in high-dimensional biological data. *Nature biotechnology*, 37(12):1482–1492, 2019.

[26] Scott W Emmons. The beginning of connectomics: a commentary on white et al.(1986)'the structure of the nervous system of the nematode caenorhabditis elegans'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1666):20140309, 2015.

[27] Seth R. Taylor, Gabriel Santpere, Alexis Weinreb, Alec Barrett, Molly B. Reilly, Chuan Xu, Erdem Varol, Panos Oikonomou, Lori Glenwinkel, Rebecca McWhirter, Abigail Poff, Manasa Basavaraju, Ibnul Rafi, Eviatar Yemini, Steven J. Cook, Alexander Abrams, Berta Vidal, Cyril Cros, Saeed Tavazoie, Nenad Sestan, Marc Hammarlund, Oliver Hobert, and David M. Miller. Molecular topography of an entire nervous system. *Cell*, 184(16):4329–4347.e23, August 2021.

[28] Zachery Wolfe, David Liska, and Adam Norris. Deep transcriptomics reveals cell-specific isoforms of pan-neuronal genes. *Nature Communications*, 16(1):1–14, 2025.

[29] Paul Davis, Magdalena Zarowiecki, Valerio Arnaboldi, Andrés Becerra, Scott Cain, Juancarlos Chan, Wen J Chen, Jaehyoung Cho, Eduardo da Veiga Beltrame, Stavros Diamantakis, Sibyl Gao, Dionysis Grigoriadis, Christian A Grove, Todd W Harris, Ranjana Kishore, Tuan Le, Raymond Y N Lee, Manuel Luypaert, Hans-Michael Müller, Cecilia Nakamura, Paulo Nuin, Michael Paulini, Mark Quinton-Tulloch, Daniela Raciti, Faye H Rodgers, Matthew Russell, Gary Schindelman, Archana Singh, Tim Stickland, Kimberly Van Auken, Qinghua Wang, Gary Williams, Adam J Wright, Karen Yook, Matt Berriman, Kevin L Howe, Tim Schedl, Lincoln Stein, and Paul W Sternberg. WormBase in 2022—data, processes, and tools for analyzing Caenorhabditis elegans. *Genetics*, 220(4):iyac003, April 2022.

[30] Alexis Weinreb, Erdem Varol, Alec Barrett, Rebecca M McWhirter, Seth R Taylor, Isabel Courtney, Manasa Basavaraju, Abigail Poff, John A Tipps, Becca Collings, et al. Alternative splicing across the c. elegans nervous system. *Nature communications*, 16(1):1–21, 2025.

[31] Eric Nguyen, Michael Poli, Marjan Faizi, Armin Thomas, Michael Wornow, Callum Birch-Sykes, Stefano Massaroli, Aman Patel, Clayton Rabideau, Yoshua Bengio, et al. HyenaDNA: Long-Range Genomic Sequence Modeling at Single Nucleotide Resolution. *Advances in Neural Information Processing Systems*, 36:43177–43201, 2023.

[32] Michael Poli, Stefano Massaroli, Eric Nguyen, Daniel Y Fu, Tri Dao, Stephen Baccus, Yoshua Bengio, Stefano Ermon, and Christopher Ré. Hyena hierarchy: Towards larger convolutional language models. In *International Conference on Machine Learning*, pages 28043–28078. PMLR, 2023.

[33] Yanrong Ji, Zhihan Zhou, Han Liu, and Ramana V Davuluri. DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome. *Bioinformatics*, 37(15):2112–2120, 2021.

[34] Maxim Zvyagin, Alexander Brace, Kyle Hippe, Yuntian Deng, Bin Zhang, Cindy Orozco Bohorquez, Austin Clyde, Bharat Kale, Danilo Perez-Rivera, Heng Ma, et al. GenSLMs: Genome-scale language models reveal SARS-CoV-2 evolutionary dynamics. *The International Journal of High Performance Computing Applications*, 37(6):683–705, 2023.

[35] Hugo Dalla-Torre, Liam Gonzalez, Javier Mendoza-Revilla, Nicolas Lopez Carranza, Adam Henryk Grzywaczewski, Francesco Oteri, Christian Dallago, Evan Trop, Bernardo P de Almeida, Hassan Sirelkhatim, et al. Nucleotide Transformer: building and evaluating robust foundation models for human genomics. *Nature Methods*, 22(2):287–297, 2025.

[36] Ken Chen, Yue Zhou, Maolin Ding, Yu Wang, Zhixiang Ren, and Yuedong Yang. Self-supervised learning on millions of primary RNA sequences from 72 vertebrates improves sequence-based RNA splicing prediction. *Briefings in bioinformatics*, 25(3):bbae163, 2024.

[37] Jun Cheng, Muhammed Hasan Çelik, Anshul Kundaje, and Julien Gagneur. MTSplice predicts effects of genetic variants on tissue-specific splicing. *Genome Biology*, 22:1–19, 2021.

[38] Jun Cheng, Thi Yen Duong Nguyen, Kamil J Cygan, Muhammed Hasan Çelik, William G Fairbrother, Julien Gagneur, et al. MMSplice: modular modeling improves the predictions of genetic variant effects on splicing. *Genome Biology*, 20(1):1–15, 2019.

[39] Alec Barrett, Rebecca McWhirter, Seth R Taylor, Alexis Weinreb, David M Miller III, and Marc Hammarlund. A head-to-head comparison of ribodepletion and polya selection approaches for caenorhabditis elegans low input rna-sequencing libraries. *G3*, 11(7):jkab121, 2021.

[40] Alec Barrett, Erdem Varol, Alexis Weinreb, Seth R. Taylor, Rebecca M. McWhirter, Cyril Cros, Berta Vidal, Manasa Basaravaju, Abigail Poff, John A. Tipps, Maryam Majeed, Chen Wang, Emily A. Bayer, Molly Reilly, Eviatar Yemini, HaoSheng Sun, Oliver Hobert, David M. Miller, and Marc Hammarlund. Integrating bulk and single cell rna-seq refines transcriptomic profiles of individual c. elegans neurons. *bioRxiv*, 2025.

[41] Diederik P Kingma. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 2015.

Supplementary Materials

A multimodal dataset for *C. elegans* splicing prediction

C. elegans provides a compact, well-annotated nervous system (302 neurons/118 classes) with a complete connectome [26], neuron-resolved transcriptomes [27, 28], and curated gene models [29]. We assemble an *exon-centered, coordinate-aligned, modality-synchronized* resource that, for each target exon, co-registers four views: whole-gene sequence, local secondary structure, junction ROIs, and neuron-type–resolved expression/co-expression. Exon-usage quantifications follow Weinreb et al. [30].

RNA sequence To expose transcript-scale dependencies while retaining base-level detail, we provide *whole-gene* sequences (max 25 kb). Sequences are tiled with a 64 bp sliding window (chosen by validation grid search) to yield patches $\{x_i\}_{i=1}^{N}$ augmented with aligned genomic-context indicators (exon/intron/junction). This design supports long-context modeling [3, 31, 32, 33, 34, 35, 36, 37, 38] without discarding distal signals.

RNA secondary structure For each exon, we extract a symmetric 500 bp window (upstream intron $\|$ exon $\|$ downstream intron) and compute secondary structure with ViennaRNA [15]. We release dot–bracket strings and base–pair graphs ($A \in \{0,1\}^{S_{\text{len}} \times S_{\text{len}}}$). Multiscale, permutation-stable features are obtained via geometric scattering [16] from lazy diffusion P and wavelets $\{\Psi_j\}$, producing $K{=}11$ channels per nucleotide aligned to transcript coordinates.

Junction ROIs To capture motif-scale cis-signals, we provide donor/acceptor-centered ROIs of fixed length $S_{\text{roi}}{=}200$ bp encoded at single-nucleotide resolution ($\{A, C, G, T\}$ one-hot with optional exon/intron tags) plus an index vector mapping each ROI base to transcript coordinates. Tokens use $k{=}1$ (no patching) and are linearly projected with positional encoding.

Expression & co-expression Splicing is cell-type–dependent; we therefore integrate neuron-type–resolved expression from CeNGEN [39, 40] and single-cell atlases [27]. Raw counts undergo QC, library-size normalization, and MAGIC imputation [17]. We curate $f{=}243$ splicing factors and, for each of $K{=}50$ neuron types, construct $\mathbf{E}_k \in \mathbb{R}^{f \times C_k}$. Pairwise mutual information yields weighted co-expression $\mathbf{M}_k$ (optionally binarized to $\mathbf{A}_k$); node signals are mean expression $\mathbf{a}_k$. Geometric scattering encodes $(\mathbf{M}_k, \mathbf{a}_k)$ into topology-aware coefficients, stacked across types for direct alignment with the other modalities.

Intended use. The dataset supports: (i) principled benchmarking of long-context genomic models; (ii) mechanistic attribution linking motifs, structural masking, and distal regulation; and (iii) cell-type–aware prediction grounded in *C. elegans* neurobiology [26, 27, 28, 29].

A Geometric Scattering

Our `CellSpliceNet` employs a novel Geometric Scattering Transform [16] to extract multiscale features from graph-structured modalities (RNA secondary structure and cell-specific co-expression). Given a graph $G = (V, E)$ with adjacency A and degree matrix D, define the lazy random-walk diffusion:

$$P \;=\; \tfrac{1}{2}\big(I + D^{-1}A\big).$$

Graph wavelets at scale $j \in \{1, \ldots, J\}$ are:

$$\Psi_j \;=\; P^{2^{j-1}} - P^{2^j}.$$

For a node signal $x \in \mathbb{R}^{|V|}$, the scattering moments are:

$$S_0 \;=\; \frac{1}{|V|}\sum_{v \in V} x(v),$$

$$S_1(j) \;=\; \frac{1}{|V|}\sum_{v\in V}\Big|\big(\Psi_j x\big)(v)\Big|,$$

$$S_2(j_1,j_2) \;=\; \frac{1}{|V|}\sum_{v\in V}\Big|\big(\Psi_{j_2}\,|\Psi_{j_1}x|\big)(v)\Big|, \qquad j_1<j_2.$$

We concatenate all moments up to $J=4$ to form a feature tensor

$$\mathbf{X}\in\mathbb{R}^{|V|\times S\times d},$$

which serves as the multiscale representation for the secondary-structure (and co-expression) modalities.

B Implementation details

All models were implemented in `PyTorch` and trained on a single NVIDIA A100 GPU. Data loading/preprocessing used standard libraries. We ensured reproducibility with fixed random seeds and a pinned software environment. Preprocessing and end-to-end train/inference scripts, plus pretrained checkpoints, are available in the public repository.

Splits & repeats. We performed a row-level IID split into train/val/test at $65\%/15\%/20\%$: $8{,}605$ / $1{,}969$ / $2{,}667$ samples from $13{,}241$ total, preserved in each fold. To assess robustness we ran k-fold cross-validation and repeated the full train/test procedure $10\times$ with different seeds. All preprocessing and partitioning scripts reside under `preprocessing` (p/).

Leakage control. Normalizers/tokenizers were fit on training only; genomic windows/ROIs were generated once and forbidden to cross splits; augmentation was applied to training only; early stopping and hyperparameters were chosen on validation, and the test set was revealed once at the end.

Training. We optimized CELLSPLICENET with Adam [41] at learning rate 2×10^{-5} and an exponential scheduler (approximately 1% decay per epoch when stepped once per epoch). Early stopping used validation performance; the best checkpoint was retained per fold. RNA sequences were tokenized with a $64\,\mathrm{bp}$ sliding window and truncated/padded to $25{,}000\,\mathrm{bp}$ max; ROI and structure modalities used $200\,\mathrm{bp}$ and $500\,\mathrm{bp}$ windows, respectively. We used batch size 40. The backbone is a 12-layer transformer with d_{model}=256, 8-head self-attention, a 512-width feed-forward sublayer, and dropout 0.2 throughout.