SPLICEMAMBA: TISSUE-CONDITIONED STATE-SPACE MODELING OF CASSETTE-EXON INCLUSION

Anonymous authors
Paper under double-blind review

ABSTRACT

Cassette exon inclusion is the predominant form of alternative splicing in mammals and a central mechanism of tissue-specific gene regulation. Predicting cassette exon inclusion remains challenging because cassette exons vary substantially in length, while their splicing outcomes are determined by surrounding *cis*-regulatory sequences, and existing sequence models are largely optimized for classifying individual splice sites rather than quantifying inclusion at the level of the complete splicing event. We introduce SpliceMamba, a selective state-space architecture for tissue-specific prediction of cassette exon Percent Spliced In (PSI). SpliceMamba constructs a unified token representation integrating nucleotide sequence, genomic annotation, and tissue identity; encodes four splice-site-centered windows using a shared selective state-space block; and integrates information across these windows to generate a single tissue-conditioned PSI prediction. On a 56-tissue benchmark, SpliceMamba achieves a Pearson correlation of 0.94 across *event–tissue pairs* using only 0.181M parameters, compared with 0.82–0.85 for splice-site models adapted to the same task. Because pooled correlation can be dominated by between-event variation, we additionally evaluate pairwise tissue-differential inclusion, for which SpliceMamba achieves a correlation of 0.50, compared with 0.23–0.37 for the baselines. *In silico* mutagenesis recovers canonical donor and acceptor motifs and identifies positions whose substitution shifts predicted inclusion between tissues. Ablations within a fixed model scaffold isolate the contributions of genomic features, cross-window integration, and the sequence encoder. Together, these results show that event-level, tissue-conditioned modeling of splice-site-proximal sequence provides an effective and parameter-efficient framework for quantitative PSI prediction.

1 INTRODUCTION

RNA splicing is an important post-transcriptional process that removes introns and ligates exons of precursor messenger RNA (pre-mRNA) into the mature RNA transcript. Approximately 95% of mammalian multi-exon genes undergo RNA alternative splicing (AS) that allows a single gene to produce multiple transcript isoforms (Pan et al., 2009). The spliceosome catalyzes pre-mRNA splicing through precise recognition of conserved *cis*-regulatory sequence motifs at exon-intron boundaries: the 5′ donor splice site (5′ SS), defined by a GU dinucleotide and recognized by U1 small nuclear ribonucleoprotein (snRNP); and the 3′ acceptor splice site (3′ SS), characterized by an AG dinucleotide and recognized by U2 auxiliary factor (U2AF) (Wilkinson et al., 2020; Akinyi & Frilander, 2021).

Cassette exon (also called exon skipping) is the most prevalent AS pattern in mammals and accounts for more than 50% of all AS events in human genome (Kim et al., 2007; Dvinge & Bradley, 2015). Leveraging high-throughput RNA-sequencing (RNA-seq) data, the usage of a cassette exon is quantified by Percent Spliced In (PSI or Ψ), defined as the proportion of RNA-seq reads supporting exon inclusion relative to the total number of reads supporting either inclusion or skipping of the exon (Katz et al., 2010; Shen et al., 2014; Vaquero-Garcia et al., 2016). The splicing regulation of an cassette exon event is determined by the sequences of the cassette exon, flanking introns, and adjacent upstream and downstream exons (Matera & Wang, 2014; Ule & Blencowe, 2019). Moreover, the usage of cassette exons frequently exhibits tissue-specific differences across human tissues (Yeo et al., 2004; Castle et al., 2008; Wang et al., 2008; Tapial et al., 2017).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

AS regulates cell development, differentiation, survival and activities that shape immune responses, infection and inflammation (Kalsotra & Cooper, 2011; Gabut et al., 2011). The cassette exon splicing events are highly associated with diverse human diseases and cancers (Kim et al., 2013; Oltean & Bates, 2014; Danan-Gotthold et al., 2015; Lee & Abdel-Wahab, 2016; McNally et al., 2016). Furthermore, the long-term goal in the RNA splicing field is to build sequence-based models capable of predicting tissue-specific AS patterns (Wang & Burge, 2008). Therefore, we aim to predict tissue-conditioned cassette exon inclusion and pairwise tissue-differential PSI ($\Delta\Psi$) by modeling pre-mRNA sequences.

In summary, this work makes the following contributions:

- We select four splice-site-centered sequence windows together with auxiliary genomic features to represent each cassette exon event.
- We propose SpliceMamba, a tissue-aware Mamba-based framework that efficiently models long-range sequence dependencies across splice-site-centered regions while integrating genomic context to learn tissue-dependent *cis*-regulatory sequence patterns.
- We demonstrate strong performance in predicting tissue-conditioned and tissue-differential cassette exon inclusion levels by benchmarking against established baseline models.

2 Related work

Sequence Modeling for Genomic Sequences. Deep learning approaches for genomic sequence modeling have evolved from convolutional architectures toward increasingly expressive long-range sequence models. Early methods primarily relied on convolutional neural networks (CNNs) to learn local sequence motifs and regulatory patterns from nucleotide sequences (Jha et al., 2017; Bretschneider et al., 2018; Cheng et al., 2021). Transformer-based models subsequently introduced self-attention mechanism, substantially improving the capacity to capture long-range dependencies in genomic sequences (Vaswani et al., 2017). However, the quadratic computational and memory complexity of self-attention can become restrictive for long genomic sequences. More recently, state-space models (SSMs), and in particular the Mamba architecture, have emerged as an efficient alternative for long-range sequence modeling (Gu & Dao, 2023). Mamba combines input-dependent selective state transitions with linear-time sequence processing, allowing the model to selectively retain or discard information for inference.

Deep Learning Models for Tissue-Specific Splicing Prediction. Recent tissue-specific splicing models learn directly from broad genomic sequence contexts to predict the effects of genetic variants on splicing but they are not focused on the specific cassette exon event. SpliceAI (Jaganathan et al., 2019) employs Residual Neural Networks (ResNet) architecture to predict whether the center nucleotide of a 10-kilobase (kb) gene sequence window is donor or acceptor splice site. Subsequently, Pangolin (Zeng & Li, 2022) builds on SpliceAI architecture to predict the tissue-specific splice-site usage by training separate models on four human tissues. Recently, SpliceTransformer (You et al., 2024) combines ResNet with Transformer encoder blocks to predict whether each nucleotide position corresponds to a splice donor, acceptor, or neither across 15 human tissues. Cassette exon prediction presents several challenges. First, the length of cassette exons varies from hundreds of bases to several kilobases, and the mean value is 142.85 bases (Cui et al., 2017). Second, the increased sparsity and limited read coverage from single-cell RNA-seq data (Choi et al., 2020; Jiang et al., 2022) make the accurate estimation of exon inclusion level more difficult. We therefore focus on bulk RNA-seq data, which provides deeper coverage and robust quantification of exon inclusion levels across human tissues.

Our Position. Motivated by the biological importance of alternative splicing, we focus on the cassette exon PSI prediction rather than binary splice-site classification. We propose *SpliceMamba*, a tissue-aware sequence model that integrates four-splice-site-centered sequence windows, selective state-space modeling, efficient parallel scanning, and cross-window integration for tissue-specific PSI predictions. By explicitly incorporating genomic features with sequence contexts, SpliceMamba enables prediction of tissue-conditioned exon usage and tissue-differential exon inclusion levels, which provides a quantitative framework for studying the sequence determinants of differential alternative splicing.

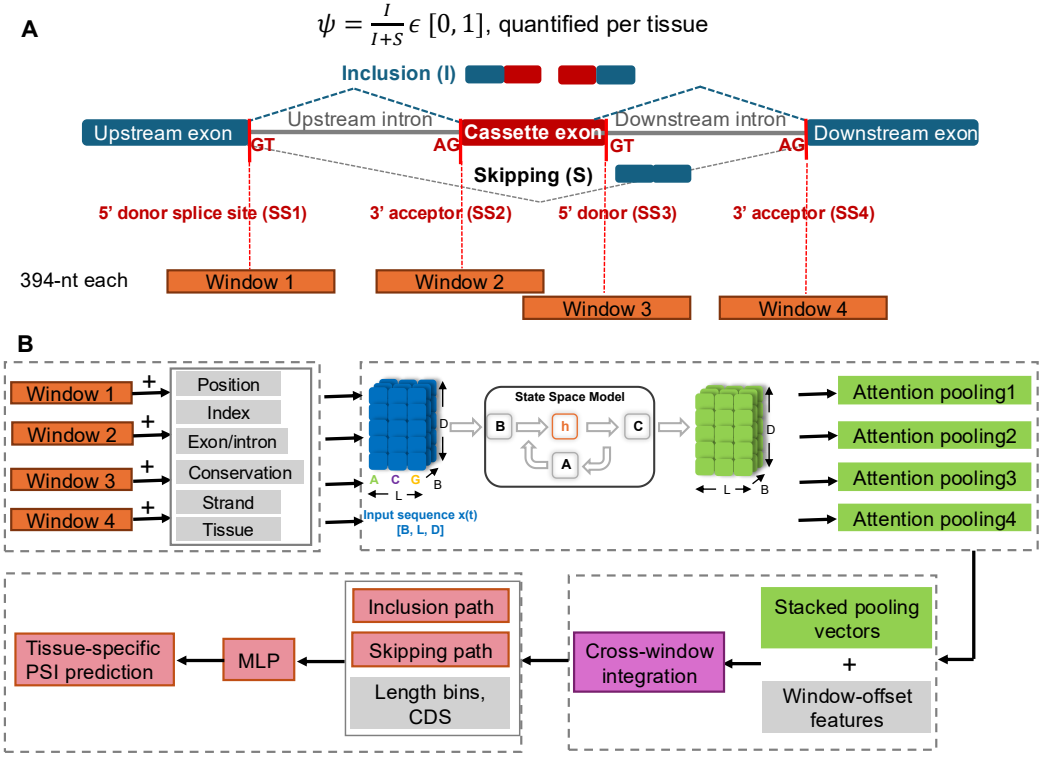


Figure 1: **Study design.** (A) A cassette-exon event and its four sequence windows. The cassette exon is flanked by an upstream and a downstream intron and may be included or skipped from the mature transcript. Inclusion is quantified per tissue as $\Psi = I/(I + S)$, where I counts reads supporting exon inclusion and S counts reads supporting exon skipping. We extract four splice-site-centered 394-nt sequence windows: *window1*, *window2*, *window3*, and *window4*. Adjacent windows may overlap when their splice sites are separated by fewer than 394 nt. (B) The SpliceMamba architecture contains four modules. Module 1 represents each nucleotide of each window by the sum of sequence and six feature embeddings. Module 2 independently encodes each window using a selective SSM with shared learnable parameters across windows, followed by attention pooling to obtain one vector per window. Module 3 incorporates window-offset features and performs cross-window integration. Module 4 combines inclusion- and skipping-path representations with length-bin and CDS features and predicts tissue-specific PSI.

3 METHOD

We propose **SpliceMamba**, a tissue-conditioned sequence modeling framework designed for cassette-exon PSI prediction. Rather than treating the event as a single continuous sequence, each cassette exon event is represented by four splice-site-centered exon–intron junction sequence windows. This design provides a consistent local sequence context for capturing splice-site-proximal sequence patterns involved in cassette-exon inclusion. The SpliceMamba model consists of four components: (1) tissue-specific input representation by incorporating sequence and biological features; (2) a selective SSM encoder with parallel scan; (3) cross-window context integration; and (4) a prediction block. The overall architecture is shown in Figure 1.

3.1 TISSUE-SPECIFIC INPUT REPRESENTATION

Each cassette exon event is represented by four sequence windows $\{\mathbf{X}^{(r)}\}_{r=1}^{4}$, each of length $L = 394$. Appendix A.1, Algorithm 1 summarizes the complete forward pass. Rather than en-

coding nucleotide identity alone, we construct a context-enriched representation by summing seven embedding channels. For position i in window r and target tissue t,

$$\mathbf{e}_i^{(r)} = E_{\mathrm{nt}}\left(X_i^{(r)}\right) + E_{\mathrm{pos}}(i) + E_{\mathrm{reg}}(r) + E_{\mathrm{EI}}(\epsilon_r(i)) + E_{\mathrm{cons}}\left(c_i^{(r)}\right) + E_{\mathrm{str}}(\sigma) + E_{\mathrm{tis}}(t), \quad (1)$$

where all embedding terms have model width d. Here, σ denotes the strand orientation and t denotes the target tissue; both are constant within an event and are broadcast across the sequence window.

The nucleotide embedding E_{nt} represents nucleotide identity within each sequence window. The position embedding E_{pos} encodes the absolute position within each 394-nt window and therefore provides information about the distance to the splice site. The region embedding E_{reg} identifies the window index $r \in \{1, 2, 3, 4\}$. The exon/intron embedding E_{EI} encodes whether a nucleotide belongs to an exonic or intronic context. The conservation embedding E_{cons} represents the discretized per-nucleotide evolutionary conservation level (Siepel et al., 2005). The strand embedding E_{str} represents forward or reverse orientation. Finally, E_{tis} provides tissue identity directly to the sequence encoder, allowing the same nucleotide sequence to be interpreted differently across tissues. Further details are provided in Appendix A.2.

3.2 LocalMamba: Selective SSM models local context within each window

For each window r, the token representations are stacked into

$$\mathbf{H}^{(r)} = \left[\mathbf{e}_1^{(r)}, \ldots, \mathbf{e}_L^{(r)}\right] \in \mathbb{R}^{L \times d}. \quad (2)$$

Each window is processed by K LocalMamba blocks. The detailed block architecture is given in Appendix A.3. Each block uses a pre-normalized residual architecture, projects the input to an expanded inner dimension, separates the signal and gating branches, applies a depthwise convolution and Sigmoid Linear Unit (SiLU) activation to the signal branch, and then performs a selective SSM operation.

For scan position τ, the selective state-space recurrence is

$$\mathbf{s}_\tau = \overline{\mathbf{A}}_\tau \mathbf{s}_{\tau-1} + \overline{\mathbf{B}}_\tau \mathbf{u}'_\tau, \qquad \mathbf{y}_\tau = \mathbf{C}_\tau^\top \mathbf{s}_\tau + \mathbf{D} \odot \mathbf{u}'_\tau, \quad (3)$$

where $\mathbf{s}_\tau$ is the latent SSM state, $\mathbf{u}'_\tau$ is the processed input, and $\mathbf{y}_\tau$ is the SSM output. The discretized transition parameters are

$$\overline{\mathbf{A}}_\tau = \exp(\Delta_\tau \mathbf{A}), \qquad \overline{\mathbf{B}}_\tau = \Delta_\tau \mathbf{B}_\tau, \qquad (\Delta_\tau, \mathbf{B}_\tau, \mathbf{C}_\tau) = g_{\mathrm{sel}}(\mathbf{u}'_\tau). \quad (4)$$

Unlike a fixed SSM, the transition parameters are conditioned on the current input representation, making the state-space dynamics selective. The input-dependent step size Δ_τ controls the rate of state evolution: small values allow information to persist for longer, whereas larger values induce faster decay. The recurrent state is propagated across the entire 394-nt sequence window, allowing the model to capture long-range dependencies within each splice-site-centered context. Further details of the selective SSM are provided in Appendix A.4.

Efficient parallel evaluation. Direct evaluation of Eq. 3 is sequential along the sequence length L. We therefore reformulate each state update as an affine transformation and exploit the associativity of affine composition. This permits all states to be computed using a parallel prefix scan, reducing the computational dependency depth from $O(L)$ to $O(\log L)$. For $L = 394$, the scan requires 9 parallel stages. The complete derivation and implementation details are given in Appendix A.5.

$$N_{\mathrm{stage}} = \lceil \log_2 L \rceil = \lceil \log_2 394 \rceil = 9 \quad (5)$$

3.3 Cross-Window Integration and Prediction

After LocalMamba encoding, each window is reduced to a single vector using attention pooling with padding positions masked:

$$\mathbf{q}^{(r)} = \mathrm{AttnPool}\left(\mathbf{H}^{(r)}, \mathbf{X}^{(r)}\right), \qquad r = 1, \ldots, 4. \tag{6}$$

The four pooled representations are stacked as

$$\mathbf{Q} = \left[\mathbf{q}^{(1)}; \mathbf{q}^{(2)}; \mathbf{q}^{(3)}; \mathbf{q}^{(4)}\right] \in \mathbb{R}^{4\times d}. \tag{7}$$

To explicitly encode the genomic geometry between neighboring splice sites, we introduce six window-offset features, corresponding to two features for each adjacent window pair. Let $\mathbf{o} \in \mathbb{R}^6$ denote these features and $\mathbf{W}_{\mathrm{off}}$ denote their projection matrix. The projected offsets are added to the four window representations before cross-window integration:

$$\mathbf{Q} \leftarrow \mathbf{Q} + \mathrm{Reshape}\left(\mathbf{o}\mathbf{W}_{\mathrm{off}}\right). \tag{8}$$

Because each window has fixed length and is centered on a splice site, two adjacent windows overlap whenever their splice sites are separated by fewer than L nucleotides. The resulting offset features therefore provide an explicit representation of the relative genomic geometry among the four splice-site-centered windows.

During training, we additionally apply dropout at the granularity of whole windows while retaining at least one window. This encourages the model to integrate information across multiple splice-site contexts rather than relying on a single window.

The resulting representation is processed by the cross-window module:

$$\widetilde{\mathbf{Q}} = \mathrm{CrossSite}(\mathbf{Q}). \tag{9}$$

Two fixed read-outs are then constructed. The inclusion-path representation aggregates all four splice-site contexts, whereas the skipping-path representation focuses on the two outer splice sites:

$$\mathbf{q}_{\mathrm{incl}} = \frac{1}{4}\sum_{r=1}^{4} \widetilde{\mathbf{q}}^{(r)}, \qquad \mathbf{q}_{\mathrm{skip}} = \frac{1}{2}\left(\widetilde{\mathbf{q}}^{(1)} + \widetilde{\mathbf{q}}^{(4)}\right). \tag{10}$$

These representations are concatenated with the mean length-bin embedding and the CDS annotation vector. The resulting feature vector is normalized before being passed to the prediction head:

$$\mathbf{z} = \mathrm{RMSNorm}\left(\left[\mathbf{q}_{\mathrm{incl}}; \mathbf{q}_{\mathrm{skip}}; \frac{1}{8}\sum_{j=1}^{8} E_{\mathrm{len}}(\ell_j); \mathbf{f}\right]\right). \tag{11}$$

The predicted PSI is obtained as

$$\hat{y} = \sigma\left(f_{\mathrm{MLP}}(\mathbf{z})\right), \qquad \hat{y} \in (0, 1). \tag{12}$$

Tissue identity is not concatenated at the prediction head because tissue conditioning has already been introduced directly into the nucleotide representation through Eq. 1.

3.4 Training Objective

We train the model using a soft-target binary cross-entropy (BCE) loss computed directly from the output logits. Although PSI is a continuous quantity in $[0, 1]$, its empirical distribution is typically concentrated near the two extremes, corresponding to predominantly included or excluded cassette exons. Soft-target BCE provides an objective that can accommodate continuous targets in $[0, 1]$. We compute BCE from logits rather than from post-sigmoid probabilities to improve numerical stability and avoid saturation under mixed-precision training. The overall training objective is defined as

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{BCE}} + \lambda \mathcal{L}_{\text{Pearson}}, \qquad \mathcal{L}_{\text{Pearson}} = 1 - r \tag{13}$$

where r denotes the Pearson correlation coefficient between the predicted and observed PSI values, and λ controls the contribution of the correlation objective and is selected during model tuning.

During distributed training with *DistributedDataParallel* (DDP), Pearson correlation is computed after differentiable global aggregation of the sufficient statistics across workers. Consequently, the correlation objective is evaluated over the effective global batch rather than independently within each local micro-batch.

4 Experiments

4.1 Common experimental settings

Training hyperparameter settings. We implement the SpliceMamba training and development using PyTorch. All experiments are conducted on four NVIDIA GH200 GPUs. We use AdamW optimization with learning rate 2×10^{-4} and weight decay 5×10^{-4}. The learning rate follows a cosine annealing schedule from 2×10^{-4} to 1×10^{-6} with a warm-up period of 10 epochs. Training is performed for 100 epochs with a batch size of 64 using mixed precision and gradient clipping with maximum norm 1.0. A fixed random seed of 42 is used across all experiments for reproducibility.

Evaluation metrics. We evaluate model performance using the following metrics: within-0.2 accuracy (a), defined as the fraction of predicted PSI values that fall within 0.2 of the corresponding observations; Mean Absolute Error (MAE), which measures the average absolute deviation between predicted and observed PSI values; Root Mean Square Error (RMSE), which measures the square root of the mean squared prediction error and places greater emphasis on larger errors; Pearson correlation (r), which measures the strength of the linear relationship between predicted and observed PSI values; Spearman correlation (ρ), which measures the strength of the monotonic relationship based on the ranks of predicted and observed PSI; and the Coefficient of Determination (R^2), which measures the proportion of variance in observed PSI explained by the model predictions. The same metrics are applied to tissue-differential PSI predictions ($|\Delta\Psi| > 0.15$). The detailed formulas are provided in the Appendix B.1 and Appendix B.2.

4.2 Data and baselines

Tissue-specific splicing datasets. We evaluate model predictive performance on two datasets. The ASCOT resource dataset (Ling et al., 2020) contains $61{,}823$ cassette exons with tissue-specific PSI measurements across 56 human tissues. After filtering out events observed in fewer than two tissues, $61{,}502$ cassette exon events are retained and subsequently partitioned into training and test sets. Cassette exon events located on chromosomes 1, 3, 5, 7, and 9 are assigned to the test set, resulting in $18{,}754$ cassette exon events and $867{,}155$ valid *event–tissue pair* observations. The remaining $42{,}748$ cassette exon events from remaining chromosomes are used for model training.

The GTEx dataset is obtained from GTEx Consortium RNA-seq data (Consortium, 2020), following the cassette exon event generation pipeline (Wu et al., 2026). This dataset contains six human tissues: Heart (Atrial Appendage), Brain (Cerebellum), Liver, Lung, Spleen, and EBV-transformed lymphocytes. Following the same dataset split strategy, a total of $17{,}218$ cassette exon events are partitioned into a test set of $1{,}260$ events and a training set of $15{,}958$ events. After excluding *event–tissue pairs* with missing PSI, the resulting GTEx dataset contains $53{,}992$ training samples and $3{,}758$ test samples. The six GTEx tissues are all represented among the 56 tissues in ASCOT. In

both datasets, the fundamental unit of prediction is an *event–tissue pair*, where the same cassette exon event can contribute multiple observations corresponding to different tissues.

Baselines. We benchmark our method against three established sequence-based splicing models, including SpliceAI, Pangolin, and SpliceTransformer. To enable a consistent event-level PSI comparison, we train these baselines on the same cassette-exon training dataset. For each cassette-exon event, we construct two independent sequence inputs: one centered on the $3'$ acceptor splice site and the other centered on the $5'$ donor splice site. Each input spans $10{,}001$ nt, comprising $5{,}000$ nt upstream of the splice site, the splice-site position itself, and $5{,}000$ nt downstream. This input length is determined by the effective receptive field of these architectures, ensuring that an input of exactly $10{,}001$ nt produces a single output prediction corresponding to the central splice site. Sequences are extracted from GENCODE[1] primary genome assembly (GRCh38.p14) using the splice-site coordinates for each event. For events on the minus strand, the extracted sequences are reverse-complemented to maintain the canonical $5'$–$3'$ orientation. Regions extending beyond chromosome boundaries are padded with nucleotide "N". Because the baselines independently predict values for the acceptor and donor sites, we define the final event-level PSI for a given tissue t as the average of these two predictions:

$$\widehat{\mathrm{PSI}}_{\mathrm{event},t} = \frac{\widehat{\mathrm{PSI}}_{\mathrm{acceptor},t} + \widehat{\mathrm{PSI}}_{\mathrm{donor},t}}{2}. \tag{14}$$

Thus, all three baselines are evaluated using their original splice-site prediction outputs, combined with a deterministic splice-site-to-event aggregation procedure. These adapted models provide an application-level benchmark for cassette-exon PSI prediction under their respective input formulations. Because SpliceMamba uses a richer event-level representation and auxiliary genomic/tissue features, this comparison should not be interpreted as an architecture-only evaluation. Architectural effects are assessed separately through matched-input backbone ablations.

4.3 SpliceMamba shows improvement on tissue-conditioned splicing prediction

We first evaluate SpliceMamba against established splicing models as an application-level benchmark on ASCOT and GTEx. As shown in Figure 2 and Table 1, SpliceMamba achieves the strongest overall predictive performance across the evaluated metrics. On the ASCOT 56-tissue dataset, SpliceMamba achieved an MAE of 0.0771, RMSE of 0.1584, and Pearson correlation of 0.9398, with 90.3% of within-0.2 accuracy. The performance advantage is also maintained on the independent GTEx 6-tissue dataset, where SpliceMamba obtains a Pearson correlation of 0.8137, and an within-0.2 accuracy of 75.2%. In comparison, SpliceAI and Pangolin show higher prediction errors and lower Pearson correlations of 0.5041 and 0.5029, respectively, while SpliceTransformer exhibits worse performance. Together, these results demonstrate that SpliceMamba provides strong PSI prediction across both the large-scale 56-tissue and the independent GTEx evaluation using fewer model parameters. More detailed performance results are shown in Appendix B.3 and B.4.

4.4 Nucleotide-Resolution Interpretability Analysis

To examine whether SpliceMamba learns tissue-dependent sequence effects, we analyzed attribution profiles for events with high predicted inclusion in brain but low predicted inclusion in liver. The upper panel summarizes nucleotide-level attribution across four 394-nt splice-site-centered sequence windows, while the lower panels provide enlarged sequence-logo representations around the splice sites. The ISM heatmap in Figure 3 (Top) reveals that the influence of ISM effects is most pronounced in proximity to splice sites and attenuates progressively with increasing distance from these sites. More ISM effects among pairwise tissue comparisons are shown in Appendix B.6.

The local sequence features exhibit opposite attribution directions between the two tissues in Figure 3 (Bottom). For example, window 3 and window 4 show tissue-dependent reversal of cytosine contributions, with positive contributions in brain and negative contributions in liver. These patterns suggest that the contribution of a local sequence feature is dynamically modulated by tissue context. Thus, SpliceMamba learns context-dependent sequence–splicing relationships, providing an interpretable mechanism for modeling tissue-specific alternative splicing.

[1] https://www.gencodegenes.org/human/

Table 1: Performance comparison of SpliceMamba with three baseline models on ASCOT and GTEx datasets. The reported metric values for each model represent the mean values over pooled *event–tissue pairs*. The best performance for each metric is highlighted in bold, and the second-best performance is underlined.

	ASCOT 56-tissue (n = 867,155)				GTEx 6-tissue (n = 3,758)				
Model	MAE ↓	RMSE ↓	Pearson ↑	a ↑	MAE ↓	RMSE ↓	Pearson ↑	a ↑	Param. (M)
SpliceMamba	**0.0771**	**0.1584**	**0.9398**	**0.903**	**0.1401**	**0.2411**	**0.8137**	**0.752**	0.181
SpliceAI	0.1366	0.2585	0.8350	0.800	<u>0.2458</u>	0.3885	<u>0.5041</u>	<u>0.634</u>	0.697
Pangolin	0.1445	0.2676	0.8232	0.782	0.2586	<u>0.3751</u>	0.5029	0.569	0.697
SpliceTransformer	<u>0.1280</u>	<u>0.2434</u>	<u>0.8543</u>	<u>0.802</u>	0.3591	0.4041	0.1612	0.163	10.936

Note. Pangolin and SpliceAI differ primarily in their original training labels. Pangolin uses tissue-specific splice-site classification and usage labels, whereas SpliceAI uses splice-site classification labels. Both models are adapted for tissue-specific PSI regression prediction. Consequently, their original label designs are not retained in this evaluation, and the two models share the same trainable parameters.

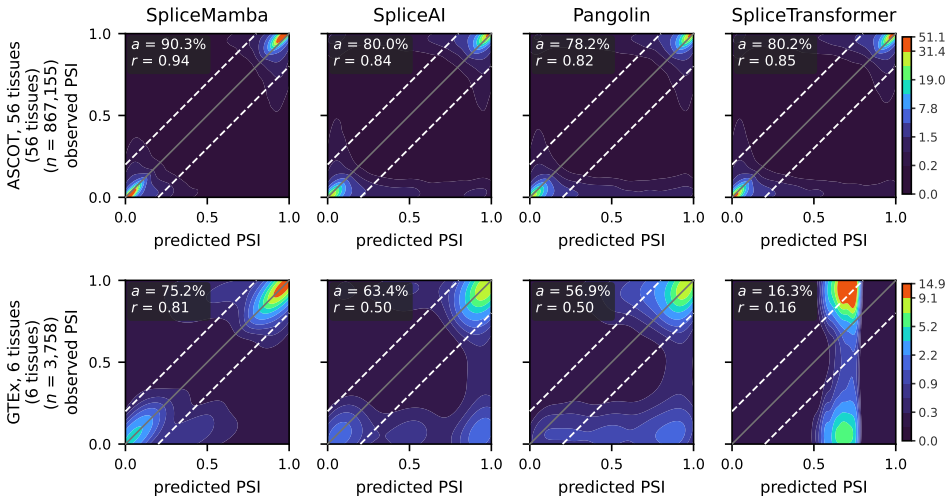


Figure 2: **Heatmaps of predicted against observed PSI on ASCOT and GTEx datasets.** Rows show model predictive performance on 56 tissues and 6 tissues, respectively. Color represents the kernel density estimate of data points. Within each row, all four panels share identical contour levels. The solid grey line is $y = x$ and the dashed lines bound the ± 0.2 band.

4.5 Ablation Study

We conduct an ablation study on the GTEx 6-tissue dataset to assess the contribution of individual components of SpliceMamba (Table 2). Removing auxiliary genomic features reduces prediction performance, with Pearson correlation decreasing from 0.8137 to 0.5230, highlighting the importance of genomic context. Removing sequence information causes a further decline in Pearson correlation to 0.3318, underscoring the important role of sequence features in PSI prediction. Removing the cross-window integration block also reduces performance, indicating that interactions among the four sequence windows contribute to the final splicing prediction.

We further examine how the location of tissue identity affects prediction. Moving tissue embedding to the prediction head decreases performance, indicating that incorporating tissue identity during representation learning is more effective than at the final prediction stage. Replacing the forward-scan of Mamba encoder with a unidirectional LSTM results in a lower Pearson correlation, supporting the effectiveness of the Mamba state-space model for sequence representation. Finally, the bidirectional Mamba variant provides only a modest improvement over the full model, while increasing the training time per epoch. We therefore retain the forward-scan Mamba encoder in the final model, as it provides a favorable trade-off between predictive performance and computational

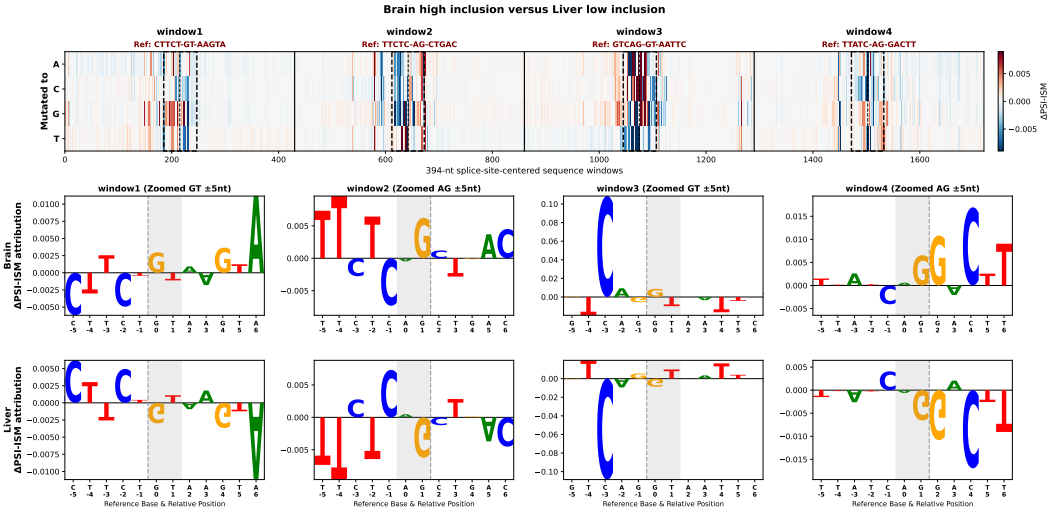


Figure 3: **Nucleotide-level sequence attribution analysis of tissue-specific cassette exon splicing via *in silico* mutagenesis on the GTEx test set.** (Top) The $|\Delta\Psi|$-ISM heatmap illustrates the possible single-nucleotide substitutions across four sequence windows. The y-axis denotes the mutated target nucleotide (A, C, G, and T). Red intensities indicate mutations that increase Brain-specific exon inclusion, whereas blue intensities indicate the reverse. (Bottom) Sequence logos derived from the ISM scores, zoomed to a 12-nt window (±5 nt) centered on the canonical GT and AG dinucleotides. Letter heights reflect the functional contribution of specific nucleotides to the predicted tissue-specific PSI, revealing differential reliance on splice site context between the two tissues.

Table 2: Ablation study of SpliceMamba on the GTEx 6-tissue dataset. **w/o Features**: model using only sequence and tissue information, without auxiliary genomic features. **w/o Sequences**: model using only auxiliary genomic features and tissue information, without sequences. **w/o Cross-window Integration**: model without cross-window integration block. **Tissue at Head**: moving the tissue embedding out of the encoder and into the prediction head. **Local LSTM**: model with a unidirectional LSTM encoder instead of forward-scan Mamba encoder. **Bidirectional LocalMamba**: model with a bidirectional scan of Mamba encoder.

Model Variant	MAE ↓	RMSE ↓	Pearson ↑	Spearman ↑	R^2 ↑	Param. (M)	Time/epoch (min)
Full	0.1401	0.2411	0.8137	0.7143	0.6470	0.181	1.8
w/o Features	0.2448	0.3670	0.5230	0.4702	0.1821	0.170	1.7
w/o Sequences	0.3946	0.4098	0.3318	0.1798	0.0199	0.180	1.7
w/o Cross-window Integration	0.1492	0.2705	0.7751	0.6894	0.5555	0.148	1.8
Tissue at Head	0.1448	0.2683	0.7797	0.6882	0.5629	0.185	2.0
Local LSTM	0.1421	0.2650	0.7859	0.7081	0.5737	0.181	0.8
Bidirectional LocalMamba	0.1383	0.2394	0.8165	0.7118	0.6521	0.181	3.3

efficiency. Overall, these ablations support the integrated design of SpliceMamba for tissue-specific splicing prediction. Additional ablations are shown in Appendix B.8.

5 Conclusion

In conclusion, SpliceMamba accurately predicts tissue-conditioned cassette exon inclusion and tissue-differential PSI by jointly modeling sequence and genomic features. ISM analysis shows that the *cis*-regulatory sequences around the canonical donor and acceptor splice site regions are important for the splicing decision of cassette exon. The ISM-derived sequence-attribution logos identify positions whose substitutions shift predicted inclusion between tissues. The ablations show the effectiveness of modeling both four sequence windows and auxiliary genomic features. Together, this work offers a SSM-based architecture for tissue-specific splicing prediction and achieves stronger predictive performance than existing models.

REFERENCES

Maureen V Akinyi and Mikko J Frilander. At the intersection of major and minor spliceosomes: crosstalk mechanisms and their impact on gene expression. *Frontiers in Genetics*, 12:700744, 2021.

Hannes Bretschneider, Shreshth Gandhi, Amit G Deshwar, Khalid Zuberi, and Brendan J Frey. Cossmo: predicting competitive alternative splice site selection using deep learning. *Bioinformatics*, 34(13):i429–i437, 2018.

John C Castle, Chaolin Zhang, Jyoti K Shah, Amit V Kulkarni, Auinash Kalsotra, Thomas A Cooper, and Jason M Johnson. Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nature genetics*, 40(12):1416–1425, 2008.

Jun Cheng, Muhammed Hasan Çelik, Anshul Kundaje, and Julien Gagneur. Mtsplice predicts effects of genetic variants on tissue-specific splicing. *Genome biology*, 22(1):94, 2021.

Kwangbom Choi, Yang Chen, Daniel A Skelly, and Gary A Churchill. Bayesian model selection reveals biological origins of zero inflation in single-cell transcriptomics. *Genome biology*, 21(1): 183, 2020.

GTEx Consortium. The gtex consortium atlas of genetic regulatory effects across human tissues. *Science*, 369(6509):1318–1330, 2020.

Ying Cui, Meng Cai, and H. Eugene Stanley. Comparative analysis and classification of cassette exons and constitutive exons. *BioMed Research International*, 2017(1):7323508, 2017. doi: https://doi.org/10.1155/2017/7323508. URL `https://onlinelibrary.wiley.com/doi/abs/10.1155/2017/7323508`.

Miri Danan-Gotthold, Regina Golan-Gerstl, Eli Eisenberg, Keren Meir, Rotem Karni, and Erez Y Levanon. Identification of recurrent regulated alternative splicing events across human solid tumors. *Nucleic acids research*, 43(10):5130–5144, 2015.

Heidi Dvinge and Robert K Bradley. Widespread intron retention diversifies most cancer transcriptomes. *Genome medicine*, 7(1):45, 2015.

Mathieu Gabut, Payman Samavarchi-Tehrani, Xinchen Wang, Valentina Slobodeniuc, Dave O'Hanlon, Hoon-Ki Sung, Manuel Alvarez, Shaheynoor Talukder, Qun Pan, Esteban O Mazzoni, et al. An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell*, 147(1):132–146, 2011.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

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.

Anupama Jha, Matthew R Gazzara, and Yoseph Barash. Integrative deep models for alternative splicing. *Bioinformatics*, 33(14):i274–i282, 2017.

Ruochen Jiang, Tianyi Sun, Dongyuan Song, and Jingyi Jessica Li. Statistics or biology: the zero-inflation controversy about scrna-seq data. *Genome biology*, 23(1):31, 2022.

Auinash Kalsotra and Thomas A Cooper. Functional consequences of developmentally regulated alternative splicing. *Nature Reviews Genetics*, 12(10):715–729, 2011.

Yarden Katz, Eric T Wang, Edoardo M Airoldi, and Christopher B Burge. Analysis and design of rna sequencing experiments for identifying isoform regulation. *Nature methods*, 7(12):1009–1015, 2010.

Eddo Kim, Alon Magen, and Gil Ast. Different levels of alternative splicing among eukaryotes. *Nucleic acids research*, 35(1):125–131, 2007.

Yoonjung Kim, Hui Ram Kim, Juwon Kim, Joong-Wook Shin, Hong-Joon Park, Jae Young Choi, Un-Kyung Kim, and Kyung-A Lee. A novel synonymous mutation causing complete skipping of exon 16 in the slc26a4 gene in a korean family with hearing loss. *Biochemical and biophysical research communications*, 430(3):1147–1150, 2013.

Stanley Chun-Wei Lee and Omar Abdel-Wahab. Therapeutic targeting of splicing in cancer. *Nature medicine*, 22(9):976–986, 2016.

Jonathan P Ling, Christopher Wilks, Rone Charles, Patrick J Leavey, Devlina Ghosh, Lizhi Jiang, Clayton P Santiago, Bo Pang, Anand Venkataraman, Brian S Clark, et al. Ascot identifies key regulators of neuronal subtype-specific splicing. *Nature communications*, 11(1):137, 2020.

A Gregory Matera and Zefeng Wang. A day in the life of the spliceosome. *Nature reviews Molecular cell biology*, 15(2):108–121, 2014.

Elizabeth M McNally, Eugene J Wyatt, et al. Welcome to the splice age: antisense oligonucleotide–mediated exon skipping gains wider applicability. *The Journal of Clinical Investigation*, 126(4): 1236–1238, 2016.

Sebastian Oltean and Dave O Bates. Hallmarks of alternative splicing in cancer. *Oncogene*, 33(46): 5311–5318, 2014.

Qun Pan, Ofer Shai, Leo J Lee, Brendan J Frey, and Benjamin J Blencowe. Addendum: Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, 41(6):762–762, 2009.

Shihao Shen, Juw Won Park, Zhi-xiang Lu, Lan Lin, Michael D Henry, Ying Nian Wu, Qing Zhou, and Yi Xing. rmats: robust and flexible detection of differential alternative splicing from replicate rna-seq data. *Proceedings of the national academy of sciences*, 111(51):E5593–E5601, 2014.

Adam Siepel, Gill Bejerano, Jakob S Pedersen, Angie S Hinrichs, Minmei Hou, Kate Rosenbloom, Hiram Clawson, John Spieth, LaDeana W Hillier, Stephen Richards, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome research*, 15(8):1034–1050, 2005.

Javier Tapial, Kevin CH Ha, Timothy Sterne-Weiler, André Gohr, Ulrich Braunschweig, Antonio Hermoso-Pulido, Mathieu Quesnel-Vallières, Jon Permanyer, Reza Sodaei, Yamile Marquez, et al. An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome research*, 27(10): 1759–1768, 2017.

Jernej Ule and Benjamin J Blencowe. Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Molecular cell*, 76(2):329–345, 2019.

Jorge Vaquero-Garcia, Alejandro Barrera, Matthew R Gazzara, Juan Gonzalez-Vallinas, Nicholas F Lahens, John B Hogenesch, Kristen W Lynch, and Yoseph Barash. A new view of transcriptome complexity and regulation through the lens of local splicing variations. *elife*, 5:e11752, 2016.

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.

Eric T Wang, Rickard Sandberg, Shujun Luo, Irina Khrebtukova, Lu Zhang, Christine Mayr, Stephen F Kingsmore, Gary P Schroth, and Christopher B Burge. Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456(7221):470–476, 2008.

Zefeng Wang and Christopher B Burge. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *Rna*, 14(5):802–813, 2008.

Max E Wilkinson, Clément Charenton, and Kiyoshi Nagai. Rna splicing by the spliceosome. *Annual review of biochemistry*, 89(1):359–388, 2020.

Di Wu, Natalie Maus, Anupama Jha, Kevin Yang, Benjamin D Wales-McGrath, San Jewell, Anna Tangiyan, Peter Choi, Jake R Gardner, and Yoseph Barash. Generative modeling for rna splicing prediction and design. *Elife*, 14:RP106043, 2026.

Gene Yeo, Dirk Holste, Gabriel Kreiman, and Christopher B Burge. Variation in alternative splicing across human tissues. *Genome biology*, 5(10):R74, 2004.

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.

Tony Zeng and Yang I Li. Predicting rna splicing from dna sequence using pangolin. *Genome biology*, 23(1):103, 2022.

A Appendix I: Method Details

A.1 SpliceMamba Forward Algorithm

A.2 Input Representation and Additive Embeddings

The input representation contains seven token-level components. For nucleotide position j in window r, the input embedding is

$$\mathbf{h}_{r,j} = \mathbf{e}^{\mathrm{nt}}(x_{r,j}) + \mathbf{e}^{\mathrm{pos}}(j) + \mathbf{e}^{\mathrm{reg}}(r) + \mathbf{e}^{\mathrm{EI}}_{r,j} + \mathbf{e}^{\mathrm{cons}}_{r,j} + \mathbf{e}^{\mathrm{strand}}(\sigma) + \mathbf{e}^{\mathrm{tissue}}(t). \tag{15}$$

This is equivalent to Eq. 1, with the notation $\mathbf{h}_{r,j}$ emphasizing the complete token representation.

1. Nucleotide tokens. Each nucleotide is represented by a learnable categorical embedding. Special padding and masking tokens are included. During training, nucleotide masking is applied with probability $p_{\mathrm{mask}} = 0.25$.

2. Position embedding. The position embedding represents nucleotide positions $1, \ldots, 394$ within each sequence window and therefore provides information about the relative location of a nucleotide within the splice-site-centered context.

3. Region embedding. The Region embedding identifies the four splice-site-centered sequence windows, indexed by $r \in \{1, 2, 3, 4\}$.

4. Exon–intron boundary embedding. The exon/intron indicator $\epsilon_r(i) \in \{0, 1\}$ is deterministically derived from the canonical GT–AG splice-junction boundaries. The corresponding embedding distinguishes exonic and intronic sequence positions.

5. Strand embedding. The strand orientation is represented by a binary categorical embedding corresponding to forward and reverse strand orientation.

6. Nucleotide conservation embedding. Per-nucleotide evolutionary conservation scores are discretized into 11 categorical bins and mapped to learnable embeddings.

7. Tissue embedding. A learnable embedding is assigned to each target tissue and added directly to every nucleotide representation. Tissue information is therefore available throughout the sequence encoder rather than being introduced only at the final prediction head.

8. Window-offset features. Window-offset features are event-level features and present the relative genomic positions of the four splice-site-centered sequence windows. They are projected to the pooled window representations before the cross-window integration block according to Eq. 8. Let

$$\mathbf{o} = \left[o^{\mathrm{overlap}}_{1,2}, o^{\mathrm{dist}}_{1,2}, o^{\mathrm{overlap}}_{2,3}, o^{\mathrm{dist}}_{2,3}, o^{\mathrm{overlap}}_{3,4}, o^{\mathrm{dist}}_{3,4} \right] \in \mathbb{R}^6 \tag{16}$$

Algorithm 1 SpliceMamba: tissue-specific PSI prediction for cassette-exon splicing events

Parameters: d (model dimension); K (number of LocalMamba blocks); N (SSM state size); E (inner expansion factor); $\mathbf{A} \in \mathbb{R}^{Ed \times N}$ (state transition); $\mathbf{D} \in \mathbb{R}^{Ed}$ (skip parameter); $\mathbf{W}_{\text{in}}, \mathbf{W}_{\text{out}}$; g_{sel} (selection projection); $E_{\text{nt}}, E_{\text{pos}}, E_{\text{reg}}, E_{\text{EI}}, E_{\text{cons}}, E_{\text{str}}, E_{\text{tis}}$; $\mathbf{W}_{\text{off}}$; E_{len}; f_{MLP}; p_{reg}.
Input: four sequence windows $\{\mathbf{X}^{(r)}\}_{r=1}^{4}$, $\mathbf{X}^{(r)} \in \{1, \ldots, V_{\text{nt}}\}^{L}$, $L = 394$; per-position conservation $\{\mathbf{c}^{(r)}\}_{r=1}^{4}$; tissue index t; strand σ; window-offset features $\mathbf{o} \in \mathbb{R}^{6}$; length bins $\boldsymbol{\ell} \in \{1, \ldots, V_{\text{len}}\}^{8}$; CDS feature $\mathbf{f} \in \{0, 1\}^{8}$.
Output: predicted PSI $\hat{y} \in (0, 1)$ for event–tissue pair (e, t).

Module 1: Input token representation
for $r = 1, \ldots, 4$ **do**
  for $i = 1, \ldots, L$ **do**
    $\mathbf{e}_i^{(r)} \leftarrow E_{\text{nt}}(X_i^{(r)}) + E_{\text{pos}}(i) + E_{\text{reg}}(r) + E_{\text{EI}}(\epsilon_r(i)) + E_{\text{cons}}(c_i^{(r)}) + E_{\text{str}}(\sigma) + E_{\text{tis}}(t)$
  end for
  $\mathbf{H}^{(r)} \leftarrow [\mathbf{e}_1^{(r)}, \ldots, \mathbf{e}_L^{(r)}]$
end for

Module 2: LocalMamba encoder
for $r = 1, \ldots, 4$ **do**
  for $k = 1, \ldots, K$ **do**
    $\mathbf{Z} \leftarrow \text{RMSNorm}(\mathbf{H}^{(r)})$
    $[\mathbf{U}, \mathbf{R}] \leftarrow \text{Split}(\mathbf{Z}\mathbf{W}_{\text{in}})$
    $\mathbf{U}' \leftarrow \text{SiLU}(\text{DWConv1D}(\mathbf{U}))$
    $(\boldsymbol{\Delta}, \mathbf{B}, \mathbf{C}) \leftarrow g_{\text{sel}}(\mathbf{U}')$
    $\mathbf{a}_\tau \leftarrow \exp(\Delta_\tau \mathbf{A})$ and $\mathbf{b}_\tau \leftarrow \Delta_\tau \mathbf{B}_\tau \odot \mathbf{u}'_\tau$
    $\{\mathbf{s}_\tau\}_{\tau=1}^{L} \leftarrow \text{ParallelScan}\left(\{(\mathbf{a}_\tau, \mathbf{b}_\tau)\}_{\tau=1}^{L}\right)$
    $\mathbf{y}_\tau \leftarrow \mathbf{C}_\tau^{\top} \mathbf{s}_\tau + \mathbf{D} \odot \mathbf{u}'_\tau$
    $\mathbf{Y} \leftarrow [\mathbf{y}_1, \ldots, \mathbf{y}_L]$
    $\mathbf{H}^{(r)} \leftarrow \mathbf{H}^{(r)} + \text{Dropout}\left([\mathbf{Y} \odot \text{SiLU}(\mathbf{R})]\mathbf{W}_{\text{out}}\right)$
  end for
  $\mathbf{H}^{(r)} \leftarrow \text{RMSNorm}(\mathbf{H}^{(r)})$
  $\mathbf{q}^{(r)} \leftarrow \text{AttnPool}(\mathbf{H}^{(r)}, \mathbf{X}^{(r)})$
end for

Module 3: Cross-window integration
$\mathbf{Q} \leftarrow \text{Stack}(\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(4)}) \in \mathbb{R}^{4 \times d}$
$\mathbf{Q} \leftarrow \mathbf{Q} + \text{Reshape}(\mathbf{o}\mathbf{W}_{\text{off}})$
if training and $p_{\text{reg}} > 0$ **then**
  $\mathbf{Q} \leftarrow \text{WindowDropout}(\mathbf{Q}, p_{\text{reg}})$
end if
$\widetilde{\mathbf{Q}} \leftarrow \text{CrossSite}(\mathbf{Q})$

Module 4: Prediction
$\mathbf{q}_{\text{incl}} \leftarrow \frac{1}{4} \sum_{r=1}^{4} \widetilde{\mathbf{q}}^{(r)}$
$\mathbf{q}_{\text{skip}} \leftarrow \frac{1}{2} \left(\widetilde{\mathbf{q}}^{(1)} + \widetilde{\mathbf{q}}^{(4)}\right)$
$\mathbf{z} \leftarrow \text{RMSNorm}\left(\text{Concat}\left(\mathbf{q}_{\text{incl}}, \mathbf{q}_{\text{skip}}, \frac{1}{8} \sum_{j=1}^{8} E_{\text{len}}(\ell_j), \mathbf{f}\right)\right)$
$\hat{y} \leftarrow \sigma\left(f_{\text{MLP}}(\mathbf{z})\right)$
return $\hat{y}$

denotes the six offset features. Specifically, we consider three splice-site pairs, splice-site 1 to splice-site 2 ($o_{1,2}$), splice-site 2 to splice-site 3 ($o_{2,3}$), and splice-site 3 to splice-site 4 ($o_{3,4}$). $o_{1,2}^{\text{dist}}$ represents the length of upstream intron, $o_{2,3}^{\text{dist}}$ represents the length of cassette exon, and $o_{3,4}^{\text{dist}}$ represents the length of downstream intron. We determine whether two sequence windows overlap in genomic coordinates and record the corresponding overlap length. Thus, $o_{1,2}^{\text{overlap}}$ represents the

overlapped length between *window1* and *window2*, the $o_{2,3}^{\text{overlap}}$ represents the overlapped length between *window2* and *window3*, and $o_{3,4}^{\text{overlap}}$ represents the overlapped length between *window3* and *window4*.

The resulting feature distribution is highly consistent between the training and test sets. The cassette exon overlaps in 98.4% of training events and 98.7% of test events, with median overlap lengths of 109 nt and 102 nt, respectively. In contrast, overlaps between upstream intron and downstream intron occur less frequently, reflecting the greater variation in the intronic sequence context. The upstream intron overlaps occur in 28.1% of training events and 20.6% of test events, with median overlap lengths of 165 nt and 195 nt, respectively. Similarly, the downstream intron overlaps occur in 28.6% of training events and 20.2% of test events, with median overlap lengths of 160 nt and 182 nt. Overall, the window-offset features capture structural characteristics of cassette exon events.

9. Exon–intron length-bin features. Exon and intron length bins are incorporated at the prediction head since they are event-level features that remain constant across positions within a sequence window. Their learned representations are averaged before being concatenated with the inclusion and skipping representations.

10. CDS annotation features. Each splicing event is assigned an 8-dimensional one-hot CDS annotation vector. The categories are *de novo*, coding region, noncoding region, start-coding region, end-coding region, mixed categories, non-coding gene, and frame shift. The resulting vector $\mathbf{f} \in \{0,1\}^8$ is concatenated with the learned representations at the prediction head according to Eq. 11.

A.3 Detailed LocalMamba Architecture

For an input representation $\mathbf{H}$, a LocalMamba block first projects the input into a signal branch and a gating branch:

$$[\mathbf{U}, \mathbf{R}] = \text{Split}\left(\text{RMSNorm}(\mathbf{H})\mathbf{W}_{\text{in}}\right). \tag{17}$$

The signal branch is processed using a depthwise one-dimensional convolution followed by the SiLU activation:

$$\mathbf{U}' = \text{SiLU}\left(\text{DWConv1D}(\mathbf{U})\right). \tag{18}$$

The input-dependent selective parameters are then obtained as

$$(\mathbf{\Delta}, \mathbf{B}, \mathbf{C}) = g_{\text{sel}}(\mathbf{U}'). \tag{19}$$

The selective SSM produces the sequence representation

$$\mathbf{Y} = \text{SelectiveSSM}\left(\mathbf{U}', \mathbf{\Delta}, \mathbf{B}, \mathbf{C}\right), \tag{20}$$

where the position-wise state recurrence is given by Eq. 3. The SSM representation is modulated by the gating branch:

$$\mathbf{Z} = \mathbf{Y} \odot \text{SiLU}(\mathbf{R}). \tag{21}$$

The final residual output of the block is

$$\mathbf{H}_{\text{out}} = \mathbf{H} + \text{Dropout}\left(\mathbf{Z}\mathbf{W}_{\text{out}}\right). \tag{22}$$

The default SpliceMamba model stacks $K = 3$ LocalMamba blocks for each of the four sequence windows.

A.4 Detailed Selective SSM

The continuous-time state-transition matrix is parameterized as

$$\mathbf{A} = -\exp(\mathbf{A}_{\log}), \tag{23}$$

which constrains the state dynamics to be non-positive.

For each sequence position τ, the input-dependent step size is computed as

$$\Delta_\tau = \mathrm{clip}\left(\mathrm{Softplus}(\widetilde{\Delta}_\tau), \Delta_{\min}, \Delta_{\max}\right). \tag{24}$$

For the LocalMamba encoder, we use

$$\Delta_{\min} = 10^{-3}, \qquad \Delta_{\max} = 10^{-1}. \tag{25}$$

The continuous-time state transition is discretized as

$$\overline{\mathbf{A}}_\tau = \exp\left(\Delta_\tau \mathbf{A}\right), \tag{26}$$

and the input contribution is

$$\mathbf{b}_\tau = \Delta_\tau \mathbf{B}_\tau \odot \mathbf{u}'_\tau. \tag{27}$$

Together, these quantities yield the recurrence in Eq. 3.

A.5 Parallel Selective Scan

The selective state update in Eq. 3 can be written as the first-order recurrence

$$\mathbf{s}_\tau = \mathbf{a}_\tau \odot \mathbf{s}_{\tau-1} + \mathbf{b}_\tau, \qquad \mathbf{a}_\tau = \exp(\Delta_\tau \mathbf{A}), \qquad \mathbf{b}_\tau = \Delta_\tau \mathbf{B}_\tau \odot \mathbf{u}'_\tau, \tag{28}$$

with initial state $\mathbf{s}_0 = \mathbf{0}$.

Each recurrence step can be represented as an affine transformation

$$T_\tau(\mathbf{s}) = \mathbf{a}_\tau \odot \mathbf{s} + \mathbf{b}_\tau. \tag{29}$$

Composing two affine transformations gives

$$\begin{aligned}(T_2 \circ T_1)(\mathbf{s}) &= \mathbf{a}_2 \odot (\mathbf{a}_1 \odot \mathbf{s} + \mathbf{b}_1) + \mathbf{b}_2 \\ &= (\mathbf{a}_2 \odot \mathbf{a}_1) \odot \mathbf{s} + (\mathbf{a}_2 \odot \mathbf{b}_1 + \mathbf{b}_2).\end{aligned} \tag{30}$$

Accordingly, define the binary product

$$(\mathbf{a}_2, \mathbf{b}_2) \otimes (\mathbf{a}_1, \mathbf{b}_1) = (\mathbf{a}_2 \odot \mathbf{a}_1,\ \mathbf{a}_2 \odot \mathbf{b}_1 + \mathbf{b}_2). \tag{31}$$

Because $\otimes$ represents function composition, it is associative:

$$(\mathbf{p}_3 \otimes \mathbf{p}_2) \otimes \mathbf{p}_1 = \mathbf{p}_3 \otimes (\mathbf{p}_2 \otimes \mathbf{p}_1), \tag{32}$$

although it is not commutative. Associativity is sufficient to evaluate the prefix compositions using a parallel scan.

Let

Algorithm 2 Parallel prefix scan for the selective recurrence

Input: $\{(\mathbf{a}_\tau, \mathbf{b}_\tau)\}_{\tau=1}^{L}$
Output: $\{\mathbf{s}_\tau\}_{\tau=1}^{L}$
$(\mathbf{A}^*, \mathbf{B}^*) \leftarrow (\{\mathbf{a}_\tau\}, \{\mathbf{b}_\tau\})$
$j \leftarrow 1$
while $j < L$ **do**
    $\mathbf{B}^*_\tau \leftarrow \mathbf{A}^*_\tau \odot \mathbf{B}^*_{\tau-j} + \mathbf{B}^*_\tau$ for all $\tau > j$ in parallel
    $\mathbf{A}^*_\tau \leftarrow \mathbf{A}^*_\tau \odot \mathbf{A}^*_{\tau-j}$ for all $\tau > j$ in parallel
    $j \leftarrow 2j$
end while
return $\mathbf{s}_\tau = \mathbf{B}^*_\tau$

$$T_{1:\tau} = T_\tau \circ \cdots \circ T_2 \circ T_1 = \left(\mathbf{a}^{(1:\tau)}, \mathbf{b}^{(1:\tau)}\right). \tag{33}$$

Because $\mathbf{s}_0 = \mathbf{0}$, the state is obtained directly from the additive component:

$$\mathbf{s}_\tau = T_{1:\tau}(\mathbf{0}) = \mathbf{b}^{(1:\tau)}. \tag{34}$$

The SSM output is subsequently computed as

$$\mathbf{y}_\tau = \mathbf{C}_\tau^\top \mathbf{s}_\tau + \mathbf{D} \odot \mathbf{u}'_\tau. \tag{35}$$

Parallel prefix-scan algorithm. The scan requires $\lceil \log_2 L \rceil$ parallel stages. For the sequence length used in SpliceMamba,

$$\mathrm{Depth}_{\mathrm{scan}} = O(\log L) = O(\log 394), \qquad \mathrm{Depth}_{\mathrm{seq}} = O(L) = O(394). \tag{36}$$

Thus, the sequential dependency depth is reduced from $O(L)$ to $O(\log L)$ while preserving the same recurrence.

Blocked evaluation. The full scan materializes the coefficient tensors for every sequence position. To reduce peak memory, we divide the sequence into blocks of length M. The sequence therefore contains $\lceil L/M \rceil$ segments. Within each segment, Algorithm 2 is applied independently. The final state from one segment is then folded into the first coefficient of the next segment:

$$\mathbf{b}_{\tau_0} \leftarrow \mathbf{a}_{\tau_0} \odot \mathbf{s}_{\tau_0 - 1} + \mathbf{b}_{\tau_0}, \tag{37}$$

where τ_0 denotes the first position of the current segment.

The blocked implementation has peak memory complexity

$$O\left(n_{\mathrm{b}} M E d N\right), \tag{38}$$

where n_{b} is the batch size, Ed is the inner feature dimension, and N is the SSM state size. The default implementation uses $M = 32$.

B Appendix II: Additional Experimental Results

B.1 Detailed Performance Evaluation Metrics

Model performance is evaluated for tissue-specific PSI prediction and pairwise tissue-differential splicing prediction using Pearson correlation (r), Spearman correlation (ρ), mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and within-0.2 accuracy (a).

Notation. Let $\Psi_{e,t} \in [0,1]$ denote the observed PSI for event e in tissue t, and let $\hat{y}_{e,t}$ denote the corresponding model prediction. Let $m_{e,t} \in \{0,1\}$ denote the observation mask.

The set of valid observations is

$$\mathcal{O} = \{(e,t) : m_{e,t} = 1 \ \wedge \ \Psi_{e,t}, \hat{y}_{e,t} \text{ are finite}\}, \qquad N = |\mathcal{O}|. \tag{39}$$

The corresponding sample means are

$$\bar{\Psi} = \frac{1}{N} \sum_{(e,t)\in\mathcal{O}} \Psi_{e,t}, \qquad \bar{y} = \frac{1}{N} \sum_{(e,t)\in\mathcal{O}} \hat{y}_{e,t}. \tag{40}$$

For notational simplicity, the subscript $(e,t) \in \mathcal{O}$ is omitted from the following equations.

Pearson correlation. Pearson correlation measures linear agreement between predicted and observed PSI:

$$r = \frac{\sum(\hat{y} - \bar{y})(\Psi - \bar{\Psi})}{\sqrt{\sum(\hat{y} - \bar{y})^2}\sqrt{\sum(\Psi - \bar{\Psi})^2}}. \tag{41}$$

Spearman correlation. Spearman correlation measures monotonic agreement and is defined as Pearson correlation applied to the ranks:

$$\rho = r\left(\mathrm{rank}(\hat{y}), \mathrm{rank}(\Psi)\right). \tag{42}$$

Midranks are assigned to tied observations.

Mean absolute error and root mean square error.

$$\mathrm{MAE} = \frac{1}{N}\sum|\hat{y} - \Psi|, \qquad \mathrm{RMSE} = \sqrt{\frac{1}{N}\sum(\hat{y} - \Psi)^2}. \tag{43}$$

Coefficient of determination. The coefficient of determination is defined as

$$R^2 = 1 - \frac{\sum(\hat{y} - \Psi)^2}{\sum(\Psi - \bar{\Psi})^2}. \tag{44}$$

We report R^2 separately from Pearson correlation because the two metrics capture different aspects of prediction quality. Pearson correlation measures linear association, whereas R^2 additionally depends on prediction calibration relative to the observed mean.

Standard-deviation tracking. The population standard deviations of predicted and observed PSI are

$$s_{\hat{y}} = \sqrt{\frac{1}{N}\sum(\hat{y} - \bar{y})^2}, \qquad s_{\Psi} = \sqrt{\frac{1}{N}\sum(\Psi - \bar{\Psi})^2}. \tag{45}$$

Their ratio,

$$R_{\mathrm{std}} = \frac{s_{\hat{y}}}{s_{\Psi}}, \tag{46}$$

provides an additional diagnostic of the spread of model predictions relative to the observed PSI distribution.

Within-0.2 accuracy. For comparison with published splicing baselines, we also report the fraction of predictions within 0.2 PSI of the observed value:

$$a = \frac{1}{N} \sum \mathbb{1}\left[|\hat{y} - \Psi| < 0.2\right]. \tag{47}$$

B.2 Pairwise Tissue-Differential Splicing Prediction

We additionally evaluate tissue-differential splicing for the three tissue pairs considered in this study: Heart–Brain, Heart–Liver, and Brain–Liver.

For tissue pair (t, t'), the predicted and observed differential PSI values are defined as

$$\widehat{\Delta\Psi}_{e,(t,t')} = \hat{y}_{e,t} - \hat{y}_{e,t'}, \qquad \Delta\Psi_{e,(t,t')} = \Psi_{e,t} - \Psi_{e,t'}, \qquad \Delta\Psi_{e,(t,t')} \in [-1, 1]. \tag{48}$$

Because both tissue-specific predictions are generated by the same model, the differential prediction is obtained directly from the corresponding PSI predictions.

A tissue pair contributes to the differential evaluation only when both tissue observations are available:

$$\mathcal{D} = \{(e, (t, t')) : m_{e,t} m_{e,t'} = 1\}. \tag{49}$$

For the differential predictions, MAE, RMSE, Pearson correlation, Spearman correlation, and R^2 are computed using the same definitions as Eqs. 41–44, with

$$\Psi \to \Delta\Psi, \qquad \hat{y} \to \widehat{\Delta\Psi}, \qquad \mathcal{O} \to \mathcal{D}. \tag{50}$$

The mean used in the correlation and R^2 calculations is therefore

$$\overline{\Delta\Psi} = \frac{1}{|\mathcal{D}|} \sum_{(e,(t,t'))\in\mathcal{D}} \Delta\Psi_{e,(t,t')}. \tag{51}$$

B.3 Tissue-wise prediction performance on ASCOT 56-tissue dataset

Per-tissue Pearson correlation of SpliceMamba Table 3 reports the Pearson correlation coefficient between predicted and observed PSI values for each of the 56 ASCOT tissues. SpliceMamba achieves consistently high correlations across tissues, with Pearson r ranging from 0.930 to 0.955, indicating stable predictive performance across diverse tissue contexts.

Per-tissue within-0.2 accuracy on the 56-tissue ASCOT test set Table 4 reports within-0.2 accuracy a in each of the 56 tissues. SpliceMamba is the most accurate model in every tissue, with $a = 0.903$ over all pooled pairs against 0.800 for SpliceAI, 0.782 for Pangolin and 0.802 for SpliceTransformer. Pooled and macro-averaged within-0.2 accuracy coincide for every model (0.903 and 0.903 for SpliceMamba), indicating that no subset of tissues dominates the aggregate.

We also evaluate the performace on the same six tissues as the GTEx dataset, including Heart (Atrial Appendage), Brain (Cerebellum), Liver, Lung, Spleen, and EBV-transformed lymphocytes. Accuracy on the six tissues is indistinguishable from accuracy over all 56 for every model: 0.902 against 0.903 for SpliceMamba, 0.798 against 0.800 for SpliceAI, 0.781 against 0.782 for Pangolin and 0.801 against 0.802 for SpliceTransformer, differences of at most 0.002 and within twice the standard error of the six-tissue estimate in each case. Therefore, the results derived from the 6-tissue and 56-tissue analyses reflect the same biological phenomenon but they are assessed at differing levels of granularity.

Table 3: Per-tissue Pearson correlation coefficients (r) of SpliceMamba on the ASCOT 56-tissue test set (n = 867,155). The 56 tissues are listed in their original order from ASCOT, with the first 28 tissues shown in the left column and the remaining 28 tissues shown in the right column. The reported r of each tissue type are the mean values. A fixed random seed of 42 is used.

Tissue	**Pearson** r	**Tissue**	**Pearson** r
Retina - Eye	0.938	Sigmoid - Colon	0.941
RPE/Choroid/Sclera - Eye	0.948	Transverse - Colon	0.941
Subcutaneous - Adipose	0.940	Gastroesoph. Junc. - Esophagus	0.942
Visceral (Omentum) - Adipose	0.942	Mucosa - Esophagus	0.942
Adrenal Gland	0.945	Muscularis - Esophagus	0.939
Aorta - Artery	0.942	Fallopian Tube	0.954
Coronary - Artery	0.943	Atrial Appendage - Heart	0.941
Tibial - Artery	0.941	Left Ventricle - Heart	0.942
Bladder	0.951	Cortex - Kidney	0.948
Amygdala - Brain	0.942	Liver	0.945
Anterior cingulate - Brain	0.939	Lung	0.939
Caudate nucleus - Brain	0.940	Minor Salivary Gland	0.946
Cerebellar Hemisphere - Brain	0.932	Skeletal - Muscle	0.937
Cerebellum - Brain	0.932	Tibial - Nerve	0.937
Cortex - Brain	0.937	Ovary	0.941
Frontal Cortex - Brain	0.937	Pancreas	0.943
Hippocampus - Brain	0.941	Pituitary	0.940
Hypothalamus - Brain	0.938	Prostate	0.941
Nucleus accumbens - Brain	0.939	Not Sun Exposed - Skin	0.939
Putamen - Brain	0.940	Sun Exposed (Lower leg) - Skin	0.939
Spinal cord (C1) - Brain	0.941	Ileum - Small Intestine	0.943
Substantia nigra - Brain	0.941	Spleen	0.943
Mammary Tissue - Breast	0.939	Stomach	0.943
EBV-xform lymphocytes - Cells	0.946	Testis	0.930
Leukemia (CML) - Cells	0.945	Thyroid	0.940
Xform. fibroblasts - Cells	0.943	Uterus	0.943
Ectocervix - Cervix	0.955	Vagina	0.943
Endocervix - Cervix	0.955	Whole Blood	0.942

Table 4: **Within-0.2 accuracy on the ASCOT 56-tissue test set.** Within-0.2 accuracy a is the fraction of observed *event–tissue pairs* whose predicted PSI lies within 0.2 of the observed value, and se is its binomial standard error $\sqrt{a(1-a)/n}$. n is the number of *event–tissue pairs*. *Pooled* weights every *event–tissue pair* equally; *macro* averages the 56 per-tissue values, weighting tissues equally, and its se is the standard error of that mean. *6 tissues* denotes the metrics evaluation for the same six tissues as the GTEx six-tissue benchmark, matched by tissue identity. A fixed random seed of 42 is used across all models. The best value in each row is highlighted in **bold**.

		SpliceMamba		SpliceAI		Pangolin		SpliceTransformer	
Tissue	n	$a\uparrow$	$se\downarrow$	$a\uparrow$	$se\downarrow$	$a\uparrow$	$se\downarrow$	$a\uparrow$	$se\downarrow$
Testis	16,967	**0.880**	0.0025	0.762	0.0031	0.750	0.0032	0.769	0.0031
Lung	16,390	**0.898**	0.0024	0.794	0.0030	0.778	0.0031	0.795	0.0030
Sun Exposed (Lower leg) - Skin	16,289	**0.897**	0.0024	0.791	0.0030	0.777	0.0031	0.797	0.0030
Thyroid	16,280	**0.898**	0.0024	0.799	0.0030	0.781	0.0031	0.799	0.0030
Subcutaneous - Adipose	16,224	**0.898**	0.0024	0.794	0.0030	0.779	0.0031	0.798	0.0030
Tibial - Nerve	16,243	**0.891**	0.0024	0.793	0.0030	0.775	0.0031	0.792	0.0031
Not Sun Exposed - Skin	16,105	**0.898**	0.0024	0.793	0.0031	0.779	0.0031	0.798	0.0030
Mammary Tissue - Breast	16,115	**0.897**	0.0024	0.795	0.0030	0.779	0.0031	0.796	0.0030
Mucosa - Esophagus	16,040	**0.903**	0.0023	0.801	0.0030	0.784	0.0031	0.802	0.0030

Continued on next page

Table 4 continued from previous page

		SpliceMamba		SpliceAI		Pangolin		SpliceTransformer	
Tissue	n	$a\uparrow$	$se\downarrow$	$a\uparrow$	$se\downarrow$	$a\uparrow$	$se\downarrow$	$a\uparrow$	$se\downarrow$
Muscularis - Esophagus	15,979	**0.900**	0.0024	0.798	0.0030	0.780	0.0031	0.800	0.0030
Tibial - Artery	15,986	**0.899**	0.0024	0.797	0.0030	0.780	0.0031	0.800	0.0030
Skeletal - Muscle	15,960	**0.896**	0.0024	0.792	0.0031	0.772	0.0032	0.791	0.0031
Pituitary	15,933	**0.900**	0.0024	0.799	0.0030	0.780	0.0031	0.798	0.0030
Stomach	15,892	**0.903**	0.0023	0.802	0.0030	0.783	0.0031	0.802	0.0030
Transverse - Colon	15,906	**0.903**	0.0023	0.803	0.0030	0.782	0.0031	0.801	0.0030
Visceral (Omentum) - Adipose	15,892	**0.901**	0.0024	0.798	0.0031	0.781	0.0031	0.799	0.0030
Pancreas	15,867	**0.904**	0.0023	0.805	0.0030	0.787	0.0031	0.810	0.0030
Prostate	15,842	**0.900**	0.0024	0.800	0.0030	0.781	0.0031	0.800	0.0030
Sigmoid - Colon	15,787	**0.903**	0.0024	0.800	0.0030	0.784	0.0031	0.802	0.0030
Gastroesoph. Junc. - Esophagus	15,799	**0.902**	0.0024	0.802	0.0030	0.784	0.0031	0.804	0.0030
Atrial Appendage - Heart	15,760	**0.902**	0.0024	0.797	0.0031	0.778	0.0032	0.797	0.0031
Ileum - Small Intestine	15,705	**0.905**	0.0023	0.803	0.0030	0.783	0.0031	0.802	0.0030
Hypothalamus - Brain	15,723	**0.898**	0.0024	0.793	0.0031	0.774	0.0032	0.796	0.0031
Frontal Cortex - Brain	15,676	**0.899**	0.0024	0.794	0.0031	0.776	0.0032	0.795	0.0031
Aorta - Artery	15,664	**0.902**	0.0024	0.799	0.0031	0.783	0.0032	0.802	0.0030
Left Ventricle - Heart	15,667	**0.904**	0.0024	0.800	0.0031	0.781	0.0032	0.804	0.0030
Caudate nucleus - Brain	15,639	**0.902**	0.0024	0.797	0.0031	0.778	0.0032	0.798	0.0031
Adrenal Gland	15,664	**0.908**	0.0023	0.810	0.0030	0.788	0.0031	0.809	0.0030
Nucleus accumbens - Brain	15,636	**0.899**	0.0024	0.793	0.0031	0.776	0.0032	0.797	0.0031
Cortex - Brain	15,707	**0.897**	0.0024	0.791	0.0031	0.772	0.0032	0.794	0.0031
Cerebellar Hemisphere - Brain	15,593	**0.890**	0.0025	0.781	0.0032	0.764	0.0033	0.789	0.0031
Whole Blood	15,598	**0.904**	0.0024	0.801	0.0031	0.783	0.0032	0.806	0.0030
Coronary - Artery	15,605	**0.902**	0.0024	0.801	0.0031	0.784	0.0032	0.805	0.0030
Vagina	15,537	**0.906**	0.0023	0.802	0.0031	0.786	0.0031	0.806	0.0030
Hippocampus - Brain	15,460	**0.902**	0.0024	0.797	0.0031	0.778	0.0032	0.801	0.0031
Ovary	15,424	**0.904**	0.0024	0.803	0.0031	0.787	0.0032	0.805	0.0030
Minor Salivary Gland	15,477	**0.910**	0.0023	0.807	0.0030	0.791	0.0031	0.810	0.0030
Xform. fibroblasts - Cells	15,495	**0.904**	0.0024	0.808	0.0030	0.788	0.0031	0.808	0.0030
Cerebellum - Brain	15,395	**0.891**	0.0025	0.784	0.0032	0.765	0.0033	0.787	0.0032
Anterior cingulate - Brain	15,382	**0.901**	0.0024	0.794	0.0031	0.772	0.0032	0.797	0.0031
Spleen	15,381	**0.903**	0.0024	0.802	0.0031	0.784	0.0032	0.805	0.0031
Putamen - Brain	15,261	**0.903**	0.0024	0.799	0.0031	0.778	0.0032	0.799	0.0031
Spinal cord (C1) - Brain	15,296	**0.905**	0.0024	0.801	0.0031	0.782	0.0032	0.801	0.0031
Amygdala - Brain	15,337	**0.903**	0.0024	0.799	0.0031	0.779	0.0032	0.801	0.0031
Uterus	15,267	**0.905**	0.0024	0.803	0.0031	0.786	0.0032	0.806	0.0031
Substantia nigra - Brain	15,217	**0.903**	0.0024	0.800	0.0031	0.781	0.0032	0.803	0.0031
Liver	15,140	**0.908**	0.0024	0.805	0.0031	0.786	0.0032	0.808	0.0031
Cortex - Kidney	15,021	**0.910**	0.0023	0.810	0.0031	0.790	0.0032	0.812	0.0031
EBV-xform lymphocytes - Cells	14,862	**0.909**	0.0024	0.809	0.0031	0.793	0.0032	0.812	0.0031
Retina - Eye	14,810	**0.900**	0.0025	0.794	0.0032	0.776	0.0033	0.798	0.0032
RPE/Choroid/Sclera - Eye	14,711	**0.916**	0.0023	0.816	0.0031	0.798	0.0032	0.822	0.0030
Leukemia (CML) - Cells	13,905	**0.908**	0.0024	0.810	0.0032	0.790	0.0033	0.811	0.0032
Bladder	13,906	**0.920**	0.0023	0.820	0.0031	0.800	0.0032	0.823	0.0031
Fallopian Tube	13,636	**0.922**	0.0023	0.819	0.0032	0.799	0.0033	0.824	0.0031
Ectocervix - Cervix	13,140	**0.925**	0.0023	0.823	0.0032	0.807	0.0033	0.829	0.0031
Endocervix - Cervix	12,962	**0.928**	0.0023	0.826	0.0032	0.809	0.0033	0.833	0.0031
All 56 (pooled)	867,155	**0.903**	0.0003	0.800	0.0004	0.782	0.0004	0.802	0.0004
All 56 (macro)	—	**0.903**	0.0011	0.800	0.0013	0.782	0.0013	0.803	0.0014
6 tissues	92,928	**0.902**	0.0010	0.798	0.0013	0.781	0.0013	0.801	0.0013

B.4 BENCHMARKING RESULTS ON GTEX 6-TISSUE DATASET

Per-tissue within-0.2 accuracy on the GTEx test set To provide a detailed comparison across six tissues, Table 5 provides the number of *event–tissue pairs* (n), within-0.2 accuracy (a), and

Table 5: Tissue-wise predicted within-0.2 accuracy of SpliceMamba and three baseline models on the GTEx test set (n = 3,758). Here, a denotes the proportion over n *event–tissue pair* observations, and se is the corresponding binomial standard error, $se = \sqrt{a(1-a)/n}$. "All 6" aggregates all six tissues, while "3 tissue pairs" denotes the Heart–Brain, Heart–Liver, Brain–Liver tissue pairs.

		SpliceMamba		SpliceAI		Pangolin		SpliceTransformer	
Tissue	n	$a\uparrow$	$se\downarrow$	$a\uparrow$	$se\downarrow$	$a\uparrow$	$se\downarrow$	$a\uparrow$	$se\downarrow$
Heart (Atrial Appendage)	920	0.763	0.0140	0.646	0.0158	0.577	0.0163	0.155	0.0119
Brain (Cerebellum)	857	0.698	0.0157	0.574	0.0169	0.523	0.0171	0.314	0.0159
Liver	679	0.776	0.0160	0.677	0.0179	0.588	0.0189	0.137	0.0132
Lung	469	0.763	0.0196	0.652	0.0220	0.548	0.0230	0.072	0.0120
Spleen	418	0.794	0.0198	0.684	0.0227	0.648	0.0234	0.062	0.0118
EBV-transformed lymphocytes	415	0.747	0.0213	0.593	0.0241	0.561	0.0244	0.113	0.0156
All 6	3,758	0.752	0.0070	0.634	0.0079	0.569	0.0081	0.163	0.0060
3 tissue pairs	2,456	0.744	0.0088	0.629	0.0097	0.561	0.0100	0.206	0.0082

binomial standard errors (se) across six GTEx tissues for SpliceMamba, SpliceAI, Pangolin, and SpliceTransformer.

Pearson correlation heatmaps on 3-tissue pairs The heatmaps of predicted and observed PSI distributions further support that SpliceMamba achieves better tissue-specific PSI prediction in Figure 4. Across 2,456 samples from brain, heart, and liver tissues, SpliceMamba achieves the highest correlation ($r = 0.80$) and within-0.2 accuracy ($a = 74.4\%$), compared with $r = 0.49$/$a = 62.9\%$ for SpliceAI, $r = 0.48$/$a = 56.1\%$ for Pangolin, and $r = 0.16$/$a = 20.6\%$ for SpliceTransformer (Figure 4, top).

We further evaluate the models on 526 tissue-differential splicing events identified from pairwise comparisons of Heart–Brain, Heart–Liver, and Brain–Liver (Figure 4, bottom). On this subset, SpliceMamba maintains the highest Pearson correlation of 0.50 and within-0.2 accuracy of 46.2% compared to SpliceAI (0.37/47.3%), Pangolin (0.26/41.8%), and SpliceTransformer (0.23/36.9%). Together, these results show that SpliceMamba provides accurate quantitative PSI prediction and captures tissue-associated variation in cassette-exon inclusion.

B.5 Methodology: In silico mutagenesis and interpretability visualization

To characterize the sequence features underlying tissue-specific cassette exon regulation, we employed an exhaustive *in silico* mutagenesis (ISM) framework at single-nucleotide resolution. For each selected tissue-differential cassette exon event, we systematically perturb every valid nucleotide in four splice-site-centered sequence windows and quantify the resulting changes in the model-predicted PSI. The differential effects between two tissues are then summarized using ΔPSI-ISM and visualized at both global and splice-site-focused resolutions.

Step 1: Directional event selection and reference sequence extraction. We first identify cassette exon events exhibiting substantial tissue-specific splicing differences using three tissue pairs: Heart–Brain, Heart–Liver, and Brain–Liver. Events are required to have valid PSI measurements in both tissues and an observed differential splicing magnitude satisfying

$$|\Delta\mathrm{PSI}_{\mathrm{obs}}| \geq 0.15. \tag{52}$$

For each tissue pair, we further distinguish high-inclusion and high-exclusion events according to their observed PSI values, using

$$\mathrm{PSI}_{\mathrm{high}} > 0.7, \qquad \mathrm{PSI}_{\mathrm{low}} < 0.3. \tag{53}$$

The direction of the tissue comparison is defined according to the observed inclusion pattern. Among candidate events satisfying these directional criteria, we select representative events based on the agreement between model-predicted and observed tissue differences. Specifically, the prediction error is defined as

$$E_{\Delta\mathrm{PSI}} = \left|\Delta\mathrm{PSI}_{\mathrm{pred}} - \Delta\mathrm{PSI}_{\mathrm{obs}}\right|. \tag{54}$$

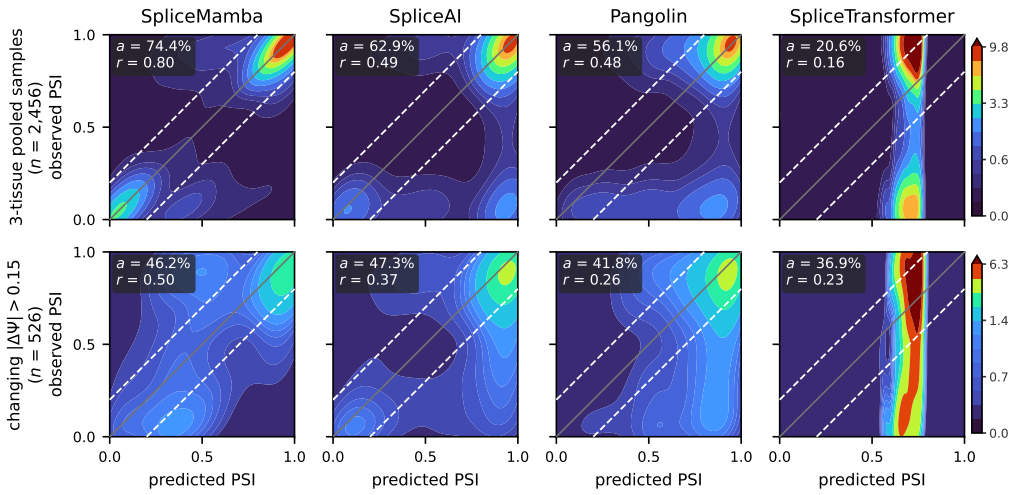


Figure 4: **Heatmaps of predicted against observed PSI on the GTEx test set.** Columns are models. The top row shows 3-tissue (Heart, Brain, Liver) pooled samples (n = 2,456), and the bottom row shows the changing subset with observed $|\Delta\Psi| > 0.15$ (n = 526). Color represents the kernel density estimate of data points. Within each row, all four panels share identical contour levels. The two rows are independently scaled to accommodate their distinct sample sizes and corresponding density ranges. The solid grey line is $y = x$ and the dashed lines bound the ± 0.2 band.

The candidate with the smallest $E_{\Delta\mathrm{PSI}}$ is selected for each directional category, thereby prioritizing events with pronounced tissue-specific splicing differences that are also well captured by the model.

For each selected event, the reference input consists of four 394-nt sequence windows centered on the four splice junctions: *window1*, the upstream exon–upstream intron junction; *window2*, the upstream intron–cassette exon junction; *window3*, the cassette exon–downstream intron junction; and *window4*, the downstream intron–downstream exon junction.

Step 2: Exhaustive single-nucleotide mutagenesis. For each selected event, we perform exhaustive *in silico* mutagenesis across all four 394-nt sequence windows. At every position containing a valid nucleotide (A, C, G, or T), the reference nucleotide is individually replaced by each of the three alternative nucleotides. Formally, for a reference nucleotide r_i at position i, the mutation set is

$$\mathcal{A}_i = \{A, C, G, T\} \setminus \{r_i\}. \tag{55}$$

Thus, each valid nucleotide position is evaluated under all three possible single-nucleotide substitutions. The resulting mutation library is processed in batches through the trained SpliceMamba model. All non-sequence inputs, including strand information, conservation values, exon/intron length features, coding-sequence features, and pairwise window features, are kept fixed during mutagenesis so that the measured change is attributable specifically to the nucleotide perturbation.

Step 3: Tissue-specific model inference and ΔPSI-ISM attribution. For each selected event, the reference sequence is first evaluated independently for each tissue to obtain the corresponding baseline PSI prediction. Each mutated sequence is then evaluated under the same tissue condition. For tissue t, the ISM attribution of substituting the reference nucleotide at position i with alternative nucleotide a is defined as

$$\mathrm{ISM}_t(i, a) = \widehat{\mathrm{PSI}}_t\left(x_{i\to a}\right) - \widehat{\mathrm{PSI}}_t(x), \tag{56}$$

where x denotes the reference sequence and $x_{i\to a}$ denotes the sequence in which nucleotide r_i is replaced by a.

To quantify tissue-specific sequence effects, we calculate the differential ISM score between tissues A and B as

$$\Delta\mathrm{PSI\text{-}ISM}(i, a) = \mathrm{ISM}_A(i, a) - \mathrm{ISM}_B(i, a). \tag{57}$$

Equivalently, this can be expressed directly in terms of the tissue-specific PSI predictions:

$$\Delta\text{PSI-ISM}(i,a) = \left[\widehat{\text{PSI}}_A\left(x_{i\to a}\right) - \widehat{\text{PSI}}_B\left(x_{i\to a}\right)\right] - \left[\widehat{\text{PSI}}_A(x) - \widehat{\text{PSI}}_B(x)\right]. \tag{58}$$

A positive ΔPSI-ISM value indicates that the nucleotide substitution increases the predicted PSI difference between tissues A and B, whereas a negative value indicates a relative shift toward tissue B.

Step 4: Global ΔPSI-ISM heatmap visualization. The resulting attribution scores are organized into a matrix with four rows corresponding to the four possible substituted nucleotides (A, C, G, and T) and columns corresponding to the concatenated positions across the four 394-nt sequence windows. The matrix is visualized using a diverging heatmap centered at zero, allowing positive and negative tissue-specific effects to be distinguished. To enhance the visualization of sequence determinants near splice junctions, the ± 12-nt regions surrounding the core splice-site dinucleotides are assigned increased horizontal width, while the remaining positions are displayed at a compressed scale. The four sequence windows are explicitly separated and labeled as *window1*–*window4*. The canonical splice-site dinucleotide within each window is additionally marked to facilitate localization of mutation-sensitive regions.

Step 5: Splice-site-focused attribution logo generation. To provide a high-resolution representation of the sequence determinants surrounding canonical splice sites, we generate attribution-based sequence logos for the ± 5-nt region surrounding the core splice-site dinucleotide. For *window1* and *window3*, the core motif is identified as the canonical 5′ GT dinucleotide, whereas for *window2* and *window4*, the core motif is identified as the canonical 3′ AG dinucleotide.

For each position i, an attribution score is assigned to the reference nucleotide r_i based on the average effect of its three possible substitutions:

$$S_i^{\text{ref}} = -\frac{1}{3} \sum_{\substack{a\in\{A,C,G,T\} \\ a\neq r_i}} \Delta\text{PSI-ISM}(i,a). \tag{59}$$

The negative sign converts the effect of disrupting the reference nucleotide into an importance-like score: a strongly negative average mutation effect produces a positive score for the reference nucleotide, indicating that the reference base supports the corresponding tissue-specific prediction.

The attribution logo is generated separately for the two tissues. For Tissue A, the reference-base score is represented directly, whereas the attribution for Tissue B is obtained by reversing the sign of the differential score:

$$S_i^{(A)} = S_i^{\text{ref}}, \qquad S_i^{(B)} = -S_i^{\text{ref}}. \tag{60}$$

The resulting logos focus on the local sequence context surrounding the canonical splice-site dinucleotides and provide a compact representation of nucleotide-level model attribution. These logos should be interpreted as model-derived attribution profiles rather than conventional sequence-frequency logos; they describe sequence patterns that contribute to the tissue-conditioned PSI predictions rather than directly establishing biological causality.

Implementation details and required libraries. The ISM pipeline and visualization framework are implemented entirely in Python. The required software libraries and their specific roles are as follows:

- **`torch`** (PyTorch): Handles the tensor operations, data loading, and highly parallelized GPU-accelerated forward-pass inference required to rapidly evaluate the thousands of mutated sequences per splicing event.
- **`numpy`**: Utilized for array manipulations, calculating offset positions for core dinucleotides, computing differential attribution matrices, and general mathematical operations.
- **`pandas`**: Employed for structuring the event metadata, sorting candidate splicing events based on observation/prediction deltas, and exporting the raw ISM attribution scores into structured CSV formats for supplementary analysis.

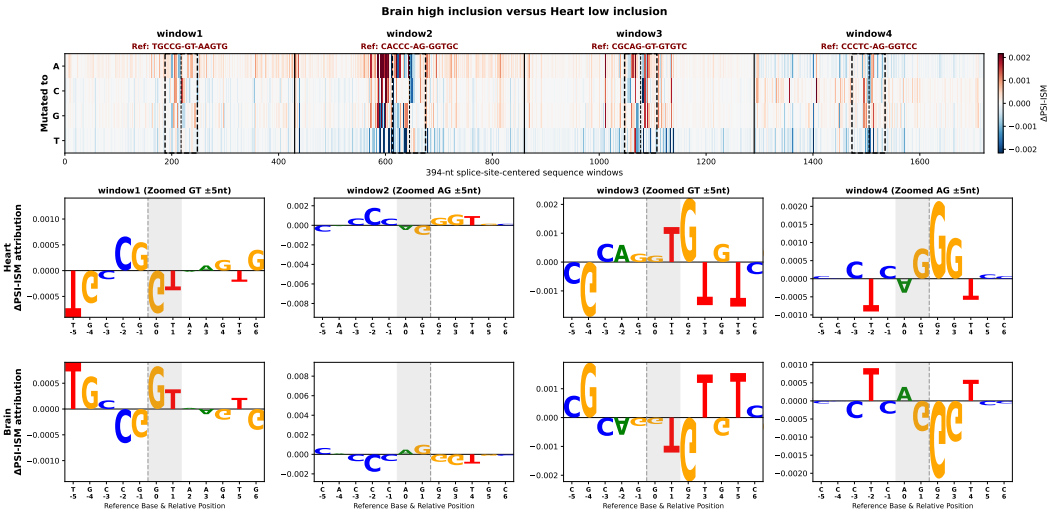


Figure 5: **Nucleotide-level ISM analysis of Brain (high exon inclusion) versus Heart (low inclusion).** (Top) The $|\Delta\Psi|$-ISM heatmap illustrates the model sensitivity to all possible single-nucleotide substitutions across four 394-nt splice-site-centered sequence windows. The y-axis denotes the mutated target nucleotide (A, C, G, and T). Red intensities indicate mutations that promote Brain-specific exon inclusion, whereas blue intensities indicate the reverse. (Bottom) Sequence logos derived from the ISM scores, zoomed to a 12-nt window (±5 nt) centered on the canonical 5' GT and 3' AG dinucleotides (highlighted in gray, with the exon-intron boundary marked by a dashed line). Letter heights reflect the functional contribution of specific nucleotides to the predicted tissue-specific PSI, revealing differential reliance on splice site context between the two tissues.

- **`matplotlib`**: Serves as the primary plotting backend for rendering the global ΔPSI-ISM pcolormesh heatmaps, managing custom grid layouts (via `GridSpec`), and drawing localized splice-site bounding boxes.
- **`logomaker`**: A specialized Python library used to directly convert the computed $N \times 4$ positional scoring matrices into standard biological sequence logos, rendering the variable-height character stacks mapped to the sequence attributions.

B.6 ISM ANALYSIS ACROSS MULTIPLE TISSUE COMPARISONS

To demonstrate that the regulatory logic captured by our model generalizes across various biological contexts and is not an artifact of a single tissue pair, we expanded our ISM analysis to five additional highly tissue-differential spliced events. These events span multiple tissue comparisons and encompass both directions of differential splicing, including both "Tissue A high inclusion" and "Tissue A high exclusion" events.

Consistent with the representative example in the main text based on GTEx 6-tissue dataset, the supplementary ΔPSI-ISM heatmaps (Figures 5, 6, 7, 8, and 9, Top) confirm that the model heavily prioritizes the proximal splice site regions over distal intronic sequences to determine tissue-specific splicing outcomes. Notably, the zoomed-in sequence logos reveal tissue-specific motif dependencies. The diversity of attribution patterns across these five supplementary events highlights the model capacity to internalize a highly dynamic, tissue-aware splicing code rather than relying on a static consensus motif.

B.7 EFFECTIVENESS OF 394-NT MODELING SEQUENCE WINDOW

We draw the ISM decay curves to provide strong empirical justification for our chosen modeling window size of 394-nt across human tissues.

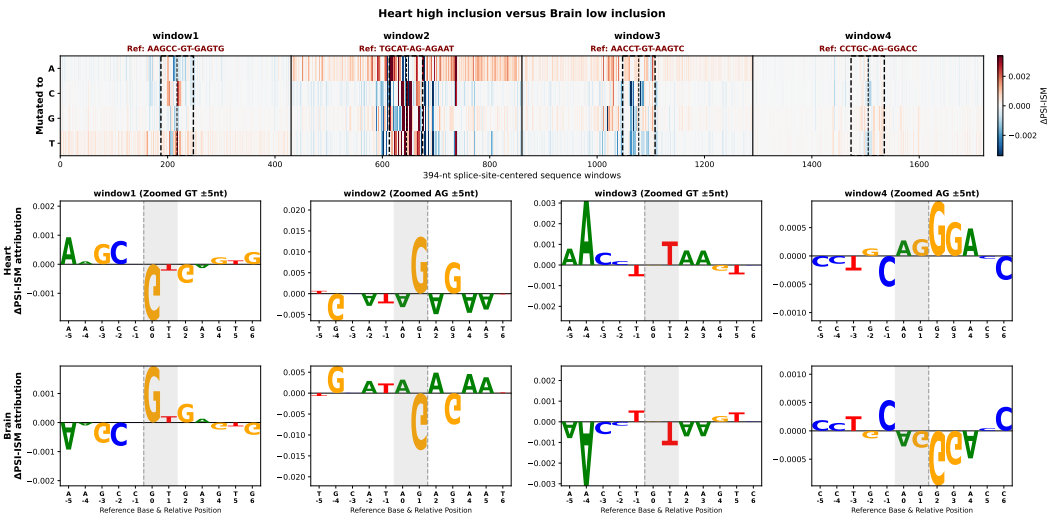


Figure 6: **Nucleotide-level ISM analysis of Heart (high exon inclusion) versus Brain (low inclusion).**

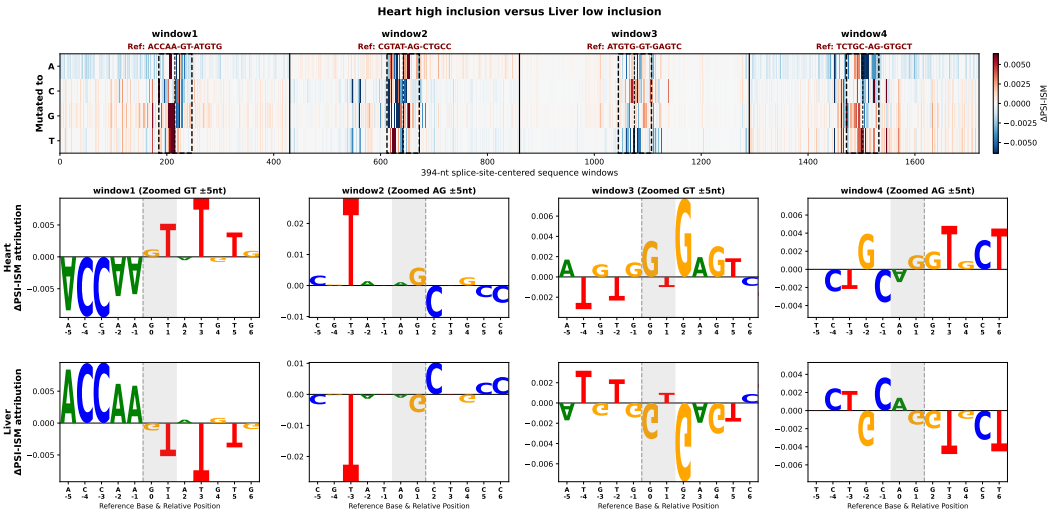


Figure 7: **Nucleotide-level ISM analysis of Heart (high exon inclusion) versus Liver (low inclusion).**

B.7.1 ISM Decay Curve Methodology

1. Unbiased Event Sampling. To accurately estimate the true spatial receptive field without introducing performance bias, the cassette exon events are uniformly sampled from the GTEx 6-tissue test set. The sampling pool is strictly restricted to events with valid, observed Percent Spliced In (PSI) measurements in the target tissue to ensure the saliency profile reflects supervised biological signals. A fixed cohort of 150 events is randomly selected for each tissue type.

2. ISM Execution. For each sampled cassette exon event, a comprehensive *in silico* mutagenesis sweep is performed across all four predefined sequence windows surrounding the splice junctions. Every nucleotide position within these windows is systematically mutated to all three alternative bases. The model performs a forward pass for each substitution to quantify the resulting absolute change in the predicted splicing inclusion rate.

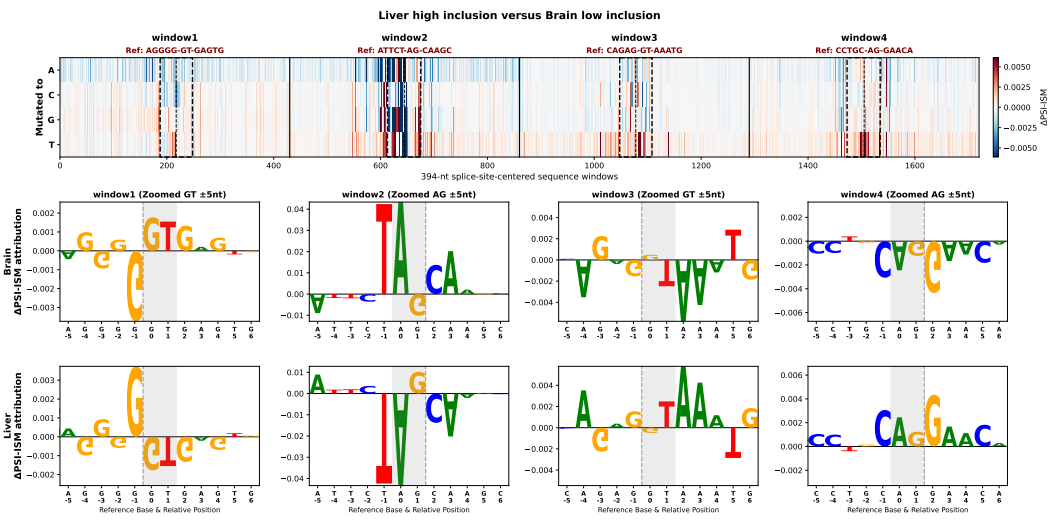


Figure 8: **Nucleotide-level ISM analysis of Liver (high exon inclusion) versus Brain (low inclusion).**

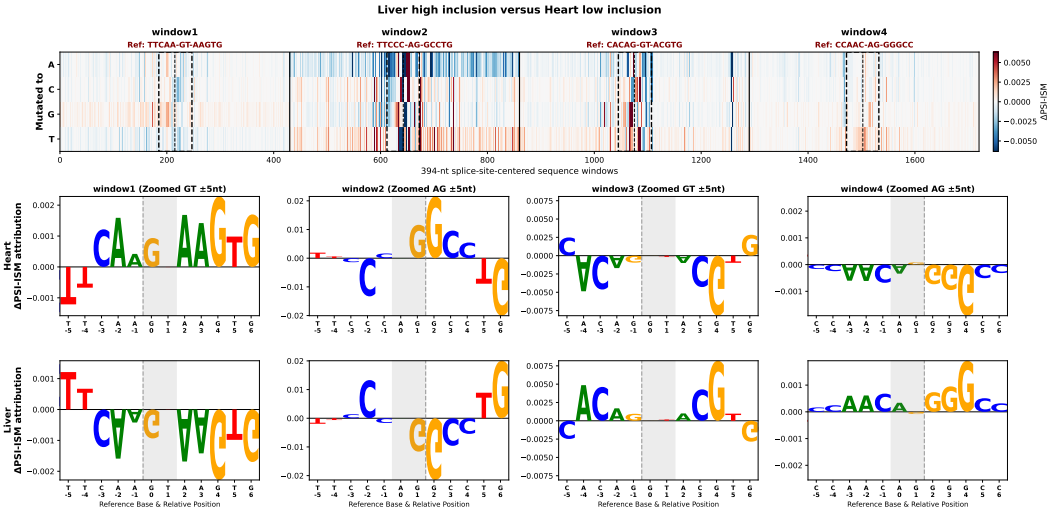


Figure 9: **Nucleotide-level ISM analysis of Liver (high exon inclusion) versus Heart (low inclusion).**

3. Data Aggregation and Noise Minimization. The mutational impacts are averaged across all specific substitutions to yield a single saliency score per nucleotide position. To ensure stable convergence when fitting the exponential decay curve, positions are pooled across the 150 events. When aggregated into 5-nt distance bins relative to the splice site, this cohort size provides 3,000 distinct mutated positions per bin.

B.7.2 ISM DECAY RESULT INTERPRETATION

As illustrated in Figure 10, the disruptive effect of single nucleotide substitutions on cassette exon splicing from brain tissue, exhibits a strong exponential decay as the increase of distance from the splice site. Mutations in the proximity of splice site drastically alter splicing outcomes, whereas this effect rapidly drops to background noise levels at more distal positions. The analysis result shows a asymmetry in the regulatory footprint between exonic and intronic regions. The intronic

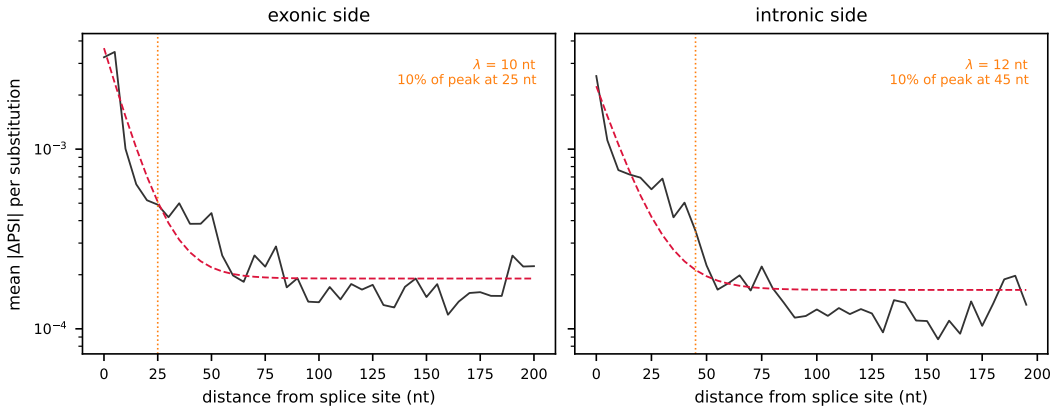


Figure 10: **Distance-dependent impact of single nucleotide substitutions on cassette exon splicing event of brain tissue.** The mean absolute change in PSI (mean $|\Delta PSI|$) per substitution is plotted against the distance from the splice site for exonic (left) and intronic (right) regions. The y-axis is displayed on a logarithmic scale. Solid black lines represent the observed empirical data, while red dashed lines indicate the exponential decay fits. The characteristic decay constant (λ) and the critical distance at which the mutational effect drops to 10% of the peak value are marked by vertical orange dotted lines.

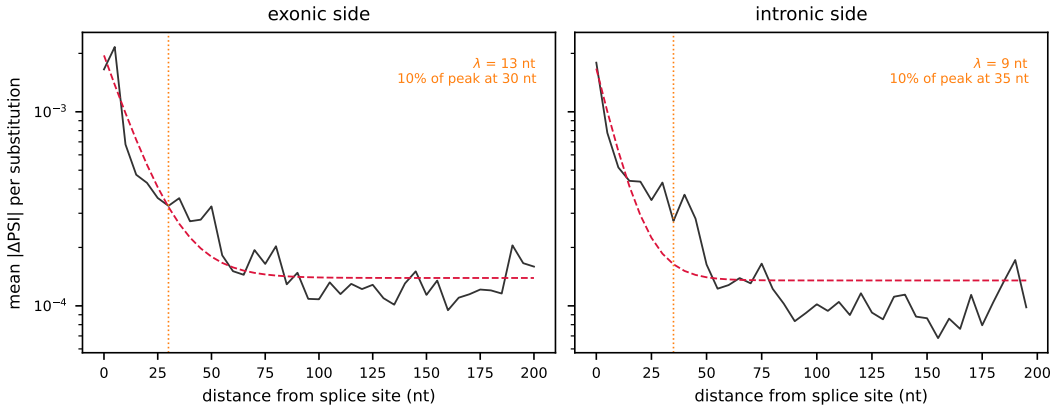


Figure 11: **Distance-dependent impact of single nucleotide substitutions on cassette exon splicing event of heart tissue.** The mean absolute change in PSI (mean $|\Delta PSI|$) per substitution is plotted against the distance from the splice site for exonic (left) and intronic (right) regions. The mutational impact drops to 10% of the peak value at a distance of 30 nt within exonic regions and 35 nt within intronic regions.

side demonstrates a broader region of vulnerability, characterized by a larger decay constant ($\lambda = 12$ nt) and maintaining regulatory sensitivity over a longer distance (decaying to 10% of the peak effect at 45 nt). In comparison, the exonic side decays more sharply and reaches the 10% threshold at just 25 nt. This extended footprint within the intronic region likely reflects the widespread distribution and structural requirements of essential *cis*-regulatory elements necessary for spliceosome assembly, such as the branch point sequence and the polypyrimidine tract.

Figures 11, 12, 13, 14, and 15 also show the distance-dependent decay of mutation effects across heart, liver, lung, spleen, and EBV-transformed lymphocyte cells. Therefore, a 394-nt sequence window spans these critical regulatory sequence regions and captures the vast majority of functional cis-regulatory elements governing splice site recognition.

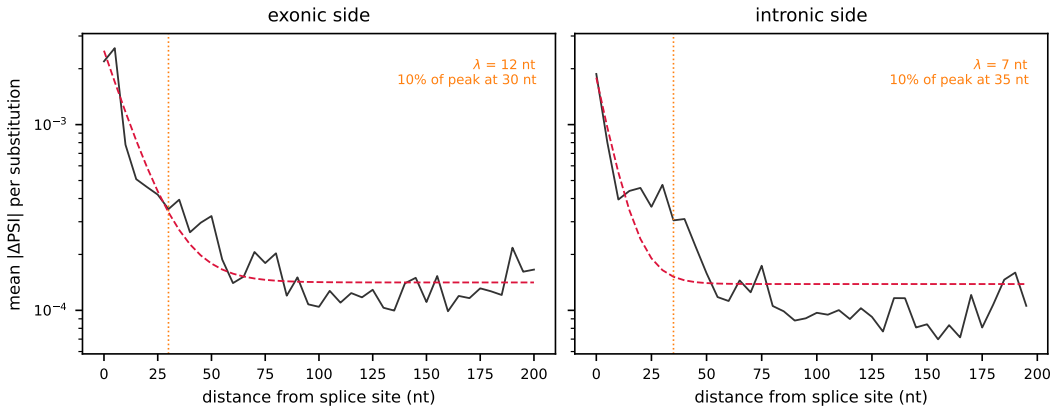


Figure 12: **Distance-dependent impact of single nucleotide substitutions on cassette exon splicing event of liver tissue.** The mean absolute change in PSI (mean $|\Delta PSI|$) per substitution is plotted against the distance from the splice site for exonic (left) and intronic (right) regions. The mutational impact drops to 10% of the peak value at a distance of 30 nt within exonic regions and 35 nt within intronic regions.

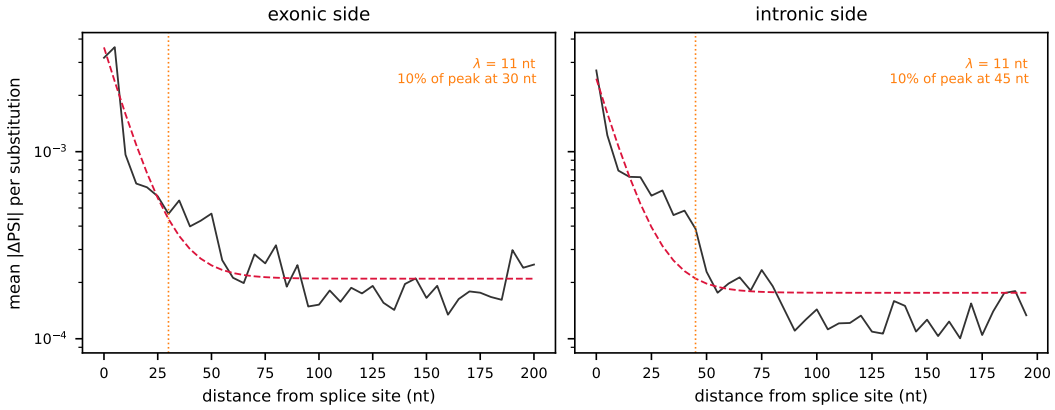


Figure 13: **Distance-dependent impact of single nucleotide substitutions on cassette exon splicing event of lung tissue.** The mean absolute change in PSI (mean $|\Delta PSI|$) per substitution is plotted against the distance from the splice site for exonic (left) and intronic (right) regions. The mutational effect drops to 10% of the peak value at the distance of 30 nt and 45 nt on the exonic and intronic regions, respectively.

B.8 MORE DETAILS IN ABLATION STUDY

Ablation of components. To further investigate the contribution of auxiliary genomic features and training components, we reports theleave-one-out ablation study based on GTEx 6-tissue dataset in Table 6. We evaluate the contribution of mask augmentation, window dropout, Pearson correlation loss, evolutionary conservation, exon/intron boundaries, geometric window offsets, CDS annotations, and length bins.

For the ablation of training components, the removal of mask augmentation results in the Pearson correlation decreasing from 0.8137 in the full model to 0.7609, suggesting that sequence-level augmentation contributes to model robustness. Removing the Pearson loss primarily affects correlation-based metrics, while removing window dropout produces relatively smaller changes in overall performance.

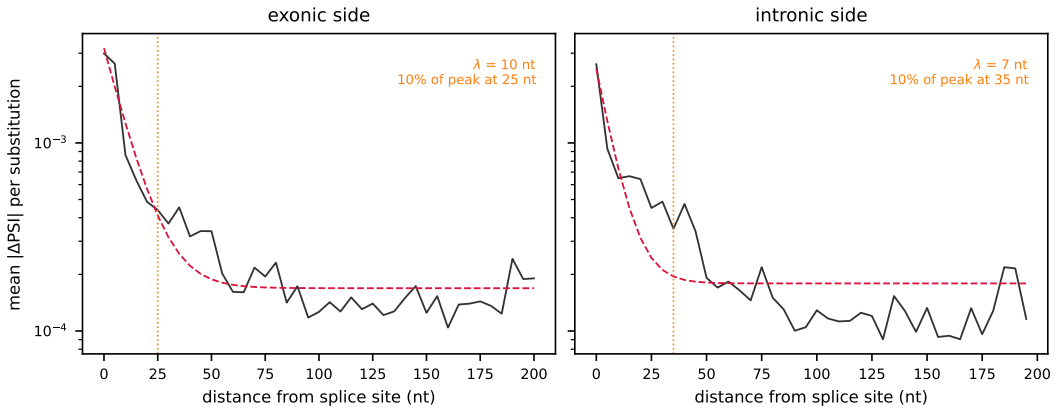


Figure 14: **Distance-dependent impact of single nucleotide substitutions on cassette exon splicing event of spleen tissue.** The mean absolute change in PSI (mean $|\Delta PSI|$) per substitution is plotted against the distance from the splice site for exonic (left) and intronic (right) regions. The mutational effect drops to 10% of the peak value at the distance of 25 nt and 35 nt on the exonic and intronic regions, respectively.

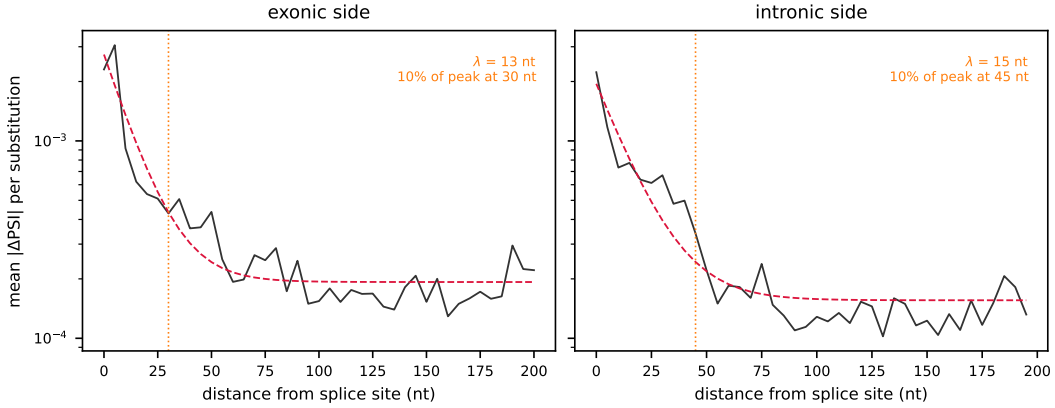


Figure 15: **Distance-dependent impact of single nucleotide substitutions on cassette exon splicing event of the EBV-transformed lymphocyte cells.** The mean absolute change in PSI (mean $|\Delta PSI|$) per substitution is plotted against the distance from the splice site for exonic (left) and intronic (right) regions. The mutational effect drops to 10% of the peak value at the distance of 30 nt and 45 nt on the exonic and intronic regions, respectively.

We also perform the ablation study of individual genomic features one by one. For the input token embeddings, Removing the region embedding decreases performance from 0.8137 to 0.7982 in Pearson correlation, which shows that distinguishing the four sequence windows helps the model capture the distinct spatial organization of cassette exon events. Removing the conservation score embedding decreases the predictive performance. The removal of exon–intron boundary embedding, window-offset features, and CDS annotations lead to moderate performance decrease, indicating that these features provide complementary information about exon–intron boundaries and event-level genomic structure. The removal of exon–intron length bins also produces relatively smaller changes in overall performance. Together, these results demonstrate that the auxiliary features and training strategies in SpliceMamba provide complementary contributions to tissue-specific PSI prediction.

Directionality. By default, we employ a unidirectional scanning strategy. Although we also evaluate a bidirectional encoder variant, its performance shows only marginal improvement over the

Table 6: Detailed ablation study of SpliceMamba on GTEx 6-tissue dataset. Each variant removes one feature or training component from the full model. Lower is better for MAE and RMSE, while higher is better for Pearson, Spearman, and R^2. Parameter count and training time per epoch are reported for reference.

Model Variant	MAE ↓	RMSE ↓	Pearson ↑	Spearman ↑	R^2 ↑	Param. (M)	Time/epoch (min)
SpliceMamba (Full Model)	**0.1401**	**0.2411**	**0.8137**	**0.7143**	**0.6470**	0.181	1.8
w/o Mask augmentation	0.1512	0.2831	0.7609	0.6729	0.5135	0.181	1.9
w/o Window dropout	0.1404	0.2647	0.7902	0.6973	0.5745	0.181	1.9
w/o Pearson loss	0.1445	0.2443	0.8070	0.6927	0.6376	0.181	1.9
w/o Region embedding	0.1476	0.2491	0.7982	0.6945	0.6234	0.181	1.9
w/o Conservation score	0.2342	0.3617	0.5561	0.4805	0.2055	0.180	1.9
w/o EI boundary	0.1431	0.2678	0.7815	0.7002	0.5645	0.181	1.9
w/o Window-offset	0.1405	0.2632	0.7912	0.6934	0.5793	0.179	1.9
w/o CDS flags	0.1466	0.2528	0.7952	0.6959	0.6121	0.180	2.0
w/o EI length bins	0.1394	0.2592	0.7964	0.6947	0.5921	0.173	1.8

Table 7: Hyperparameter sensitivity analysis during model tuning and finally selected configurations for the SpliceMamba model in the main experiments. All runs use the random seed of 42 and are compared on PSI validation performance.

Hyperparameter	**Tested values**	**Selected**
Model dimension (d)	$\{32,\ 64,\ 128\}$	64
Encoder blocks per window (K)	$\{2,\ 3,\ 4\}$	3
SSM state size (N)	$\{8,\ 16,\ 32\}$	16
Dropout rate	$\{0.10,\ 0.15,\ 0.20\}$	0.20
Time step size range (Δ_τ)	$\{(10^{-3}, 10^{-1}),\ (10^{-5}, 10^{-2})\}$	$(10^{-5}, 10^{-2})$
Encoder direction	{unidirectional, bidirectional}	unidirectional
Window size (L)	$\{196,\ 394,\ 786\}$	394
Batch size	$\{16,\ 32,\ 64\}$	64
Learning rate	$\{10^{-5},\ 10^{-4},\ 2 \times 10^{-4}\}$	2×10^{-4}
Weight decay	$\{10^{-4},\ 5 \times 10^{-4},\ 10^{-3}\}$	5×10^{-4}
Pearson weight (λ)	$\{0.25,\ 0.5,\ 1.0\}$	1.0
Loss function	{L1, Huber, soft BCE}	soft BCE

unidirectional full model based on the benchmarking metrics, while per-epoch training time increases from 1.8 to 3.3 minutes for modeling implementation on small-scale GTEx 6-tissue dataset. Consequently, for computational efficiency, we retain the unidirectional forward-scan configuration and report the corresponding performance comparison.

B.9 Hyperparameter Tuning and Sensitivity Analysis

This section summarizes our hyperparameter tuning strategy and the final configurations used in all main experiments. We conduct extensive sensitivity analyses over the search space described in Table 7. Final hyperparameters for the prediction task are chosen based on validation performance.

B.10 Baseline training settings

All four models, including SpliceMamba, SpliceAI, Pangolin, and SpliceTransformer, were trained from random initialization using identical train and test splits of the cassette exon datasets. None incorporated pre-trained or publicly released model weights, ensuring that performance differences reflect architectural distinctions rather than disparities in pretraining data or optimization history. This experimental design deliberately controls for data budget and training protocol to enable a fair and architecture-centric comparison.

The baseline model training settings are summarized in Table 8.

Table 8: Adaptation and training of the baselines. All baselines are trained on the same dataset split as SpliceMamba and evaluated with the same metrics.

	SpliceAI / Pangolin	**SpliceTransformer**
Architecture		
Trunk	16 dilated residual blocks, 32 channels; kernel widths and dilation rates identical in the two published models	two CNN encoders (128 and 64 channels), then a 6-layer Sinkhorn transformer at width 64 over 8,192 positions in the published model
Published output	SpliceAI: 3-class (acceptor, donor, or neither) site identification only. Pangolin: binary classification (splice site or not) head and corresponding usage head per tissue. This give the eight output heads across four tissues (heart, liver, brain and testis).	3-class (acceptor, donor, neither) site and usage over 15 tissues
Adapted output	1×1 convolution to $T = 56$ tissue logits	usage head widened to $T = 56$; splice-site-class heads retained, receive no gradient
Receptive field / crop	10,000 nt; the published crop leaves the center position	context length raised from $4{,}000$ to $5{,}000$ so the crop leaves the center position
Other changes	none	attention evaluated inside the mixed-precision step in the Sinkhorn transformer
Initialization	random	random; end-to-end encoders training
Parameters	0.697M	10.936M
Input and target		
Input	two 10,001 nt windows per cassette exon event, centered on the $3'$ acceptor and $5'$ donor splice site of cassette exon	
Target	per-tissue PSI is shared by both windows; the two center predictions are averaged as final prediction	
Training		
Loss	soft binary cross-entropy loss computed on logits, same as SpliceMamba	
Optimiser	AdamW with cosine annealing, gradient clipping at norm 1.0, mixed precision, same as SpliceMamba	
Learning rate	10^{-4}	10^{-4}
Weight decay	10^{-5}	10^{-5}
Epochs	100	100
Layer dropout	—	0.1 (published value)
Model selection	best epoch by PSI Pearson correlation on the test split	

C Limitations and Future Work

While our work demonstrates the significant biological application of the SpliceMamba for tissue-specific cassette exon prediction, we acknowledge several limitations that also highlight exciting avenues for future research.

The computational demands for pre-training models on large-scale genomic datasets become a critical bottleneck. Specifically, GPU memory limitations and extended training times restrict the feasibility of processing ultra-large genomic datasets.

In addition, we are collecting and integrating more splicing event data from human tissues, which will allow the SpliceMamba model to capture a more comprehensive landscape of *cis*-regulatory elements and improve its generalization to unseen biological contexts.

D LLM Usage Statement

In this work, large language models (LLMs) were used solely as general-purpose assistive tools for grammar correction and language polishing. All technical content, research ideas, experiments, and analyses were conceived, designed, and executed solely by the authors. The LLM did not contribute to the scientific content, experimental design, or interpretation of the results.

E REPRODUCIBILITY STATEMENT

We have taken deliberate steps to ensure the reproducibility of our work. Detailed descriptions of the model architecture, training process, and hyperparameter settings are presented in the main text and appendix. All scripts and source code developed for the experiments in this study, including model implementation, training, evaluation, and benchmarking scripts, will be made publicly available on GitHub upon acceptance.