A computational framework for mapping isoform landscape and regulatory mechanisms from spatial transcriptomics data

Jiayu Su[1,2,3#], Yiming Qu[2], Megan Schertzer[3], Haochen Yang[1,4], Jiahao Jiang[5], Tenzin Lhakhang[1], Theodore M. Nelson[1], Stella Park[3], Qiliang Lai[6], Xi Fu[1,4], Seung-won Choi[1,2], David A. Knowles[2,3,7#], Raul Rabadan[1,2,4#*]

[1]Program for Mathematical Genomics, Columbia University, New York, NY, USA
[2]Department of Systems Biology, Columbia University, New York, NY, USA
[3]New York Genome Center, New York, NY, USA
[4]Department of Biomedical Informatics, Columbia University, New York, NY, USA
[5]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[6]Department of Computer Science, Rice University, Houston, TX, USA
[7]Department of Computer Science, Columbia University, New York, NY, USA
[#]Correspondence: js5756@cumc.columbia.edu; dak2173@columbia.edu; rr2579@cumc.columbia.edu
[*]Lead Contact

Abstract

Transcript diversity including splicing and alternative 3'end usage is crucial for cellular identity and adaptation, yet its spatial coordination remains poorly understood. Here, we present SPLISOSM (SpatiaL ISOform Statistical Modeling), a computational framework for detecting isoform-resolution patterns from spatial transcriptomics data. SPLISOSM leverages multivariate testing to account for spot- and isoform-level dependencies, demonstrating robust and theoretically grounded performance on sparse data. In the mouse brain, we identify over 1,000 spatially variable transcript diversity events, primarily in synaptic signaling pathways linked to neuropsychiatric disorders, and uncover both known and novel regulatory relationships with region-specific RNA binding proteins. We further show that these patterns are evolutionarily conserved between mouse and human prefrontal cortex. Analysis of human glioblastoma highlights pervasive transcript diversity in antigen presentation and adhesion genes associated with specific microenvironmental conditions. Together, we present a comprehensive spatial splicing analysis in the brain under normal and neoplastic conditions.

Highlights

- Multivariate tests of spatial variability and association for transcript usage
- Conserved isoform variability in synaptic pathways across mouse and human brains
- Coordinated RNA binding protein expression drives region-specific regulation
- Tumor microenvironment shapes spatial transcript landscapes in human glioma

Keywords: SPLISOSM, Spatial transcriptomics, Alternative splicing, Alternative polyadenylation, Transcript 3'end diversity, Spatial variability, Differential isoform usage

Introduction

Almost all human genes undergo orchestrated pre-mRNA processing with precise spatiotemporal regulation[1-3]. The brain exhibits particularly high transcript diversity[4], where region-specific expression of RNA binding proteins (RBPs) drives specialized local transcript landscapes that control neurogenesis[5,6], synaptic plasticity[7,8], and circuit formation[9,10]. Disruption of this delicate regulatory balance contributes to numerous neurological disorders[11,12], including autism spectrum disorders[13], Alzheimer's disease[14], amyotrophic lateral sclerosis[15] and schizophrenia[16], as well as cancer progression and metastasis[17-19]. Despite its critical importance, our understanding of how alternative isoform usage is spatially organized within tissues remains surprisingly limited.

Modern sequencing-based spatial transcriptomics (ST) platforms naturally provide at least some isoform-specific information, whether through 3' fragments in short-read approaches[20-22] or full-length transcripts in long-read sequencing[23,24]. Despite this, most ST analysis pipelines aggregate expression at the gene level, discarding the rich transcript diversity information inherently available in the data. In particular, few computational approaches[25-27] detect spatial patterns in isoform usage and elucidate regulatory mechanisms, creating a significant gap in our understanding of tissue architecture and the molecular complexity underlying spatial functional domains.

Analyzing spatial isoform patterns presents several formidable challenges. First, isoforms from the same gene are intrinsically interdependent, with expression levels constrained by total transcribed pre-mRNAs. This multivariate and compositional nature invalidates standard statistical approaches designed for gene expression analysis. Early attempts either inappropriately treated isoforms as independent genes[24] or relied on oversimplified univariate metrics to represent transcript variation[25,26], significantly compromising statistical power. Second, data sparsity --- already problematic in ST --- is exacerbated at the isoform level when gene counts are subdivided across multiple transcripts, obscuring biological signals and creating artifacts. For example, regional variation in gene expression can masquerade as ratio usage patterns, while zero total UMI counts at many locations make isoform ratios undefined. Even when detected, these ratios frequently appear binary despite representing continuous biological preferences[28,29], complicating operations like log-ratio transformations. Common workarounds such as pseudo-counts introduce additional biases that distort relationships between observations[30]. Finally, spatial autocorrelation violates the independent observation assumptions fundamental to most differential association tests. This generates false positives from spurious associations that merely reflect mutual dependence on spatial coordinates rather than biologically meaningful connections.

To address these challenges, we present SPLISOSM (SPatiaL ISOform Statistical Modeling), a computational framework for analyzing isoform usage in spatial transcriptomics data. SPLISOSM employs theory-driven spatial and isoform kernels to capture complex dependencies between spatial locations and transcripts, maximizing statistical power while maintaining well-calibrated test statistics even with extremely sparse data. We demonstrate SPLISOSM's capabilities in multiple brain ST datasets, identifying thousands of genes with spatially variable processing in both mouse and human tissues. To uncover regulatory mechanisms driving these spatial patterns, we test

for conditional association between isoform ratios and potential regulators like RBPs, both confirming known connections and discovering novel region-specific regulatory relationships. Finally, we extend our analysis to glioblastoma cohorts, revealing how transcript diversity is shaped by both normal brain architecture and disease-specific alterations in the tumor microenvironment.

Results

Designing isoform-level spatial variability and association tests

In ST data, gene variability arises from both cell type distribution patterns and location-specific variation within cell types. While several methods exist for identifying spatially variable genes[31-35], they often prioritize different spatial patterns, resulting in limited consensus among their results[36]. Moreover, these approaches cannot be easily adapted for multivariate isoform analysis due to their reliance on specific model assumptions. SpatialDE[32], for example, builds on a Gaussian linear mixed model with analytically integrated spatial random effects. Extending this framework to multinomial isoform distributions is computationally intractable, yielding low power and inflated false positives in sparse datasets (Supplementary Note).

Drawing on non-parametric statistics' effectiveness modeling total expression from sparse ST data[34], we reformulated both spatial variability detection and differential usage analysis as multivariate independence testing. SPLISOSM transforms isoform compositions, spatial coordinates and other covariates of interest into feature spaces, capturing their complex relationships through the kernel-based Hilbert-Schmidt Independence Criterion (HSIC)[37,38] (Figure 1). Significant patterns are identified by comparing against the null hypothesis that a gene's isoform preference is independent of either spatial locations or a given covariate, e.g. expression of a specific RBP (Methods).

Our approach introduces two key technical innovations: First, we mathematically prove that all low-rank approximations of kernel tests --- including those in SPARK-X[34] --- inevitably sacrifice statistical power (Theorem 1, Supplementary Note). This insight led us to develop enhanced spatial kernels[39] that use spectral graph theory and prioritize signals by their spatial frequency, limiting power losses to high-frequency fluctuations representing noise when processing large-scale data. Second, we established a rigorous framework for sparse isoform data with a new compositional kernel that handles undefined ratios (Theorem 2, Supplementary Note), optimizing efficiency while maintaining test validity through a mean replacement strategy that utilizes all spatial observations.

Based on these theoretical advances, SPLISOSM implements three complementary spatial variability tests tailored to different biological layers: HSIC-GC for total gene expression, HSIC-IR for isoform usage ratios, and HSIC-IC for isoform counts (Methods). For genes showing spatial variability in isoform preference, we apply conditional differential usage tests to identify potential regulatory mechanisms. Intuitively, SPLISOSM controls for spatial confounding by removing location effects from both isoform usage and candidate factors under either non-parametric or parametric frameworks, then tests

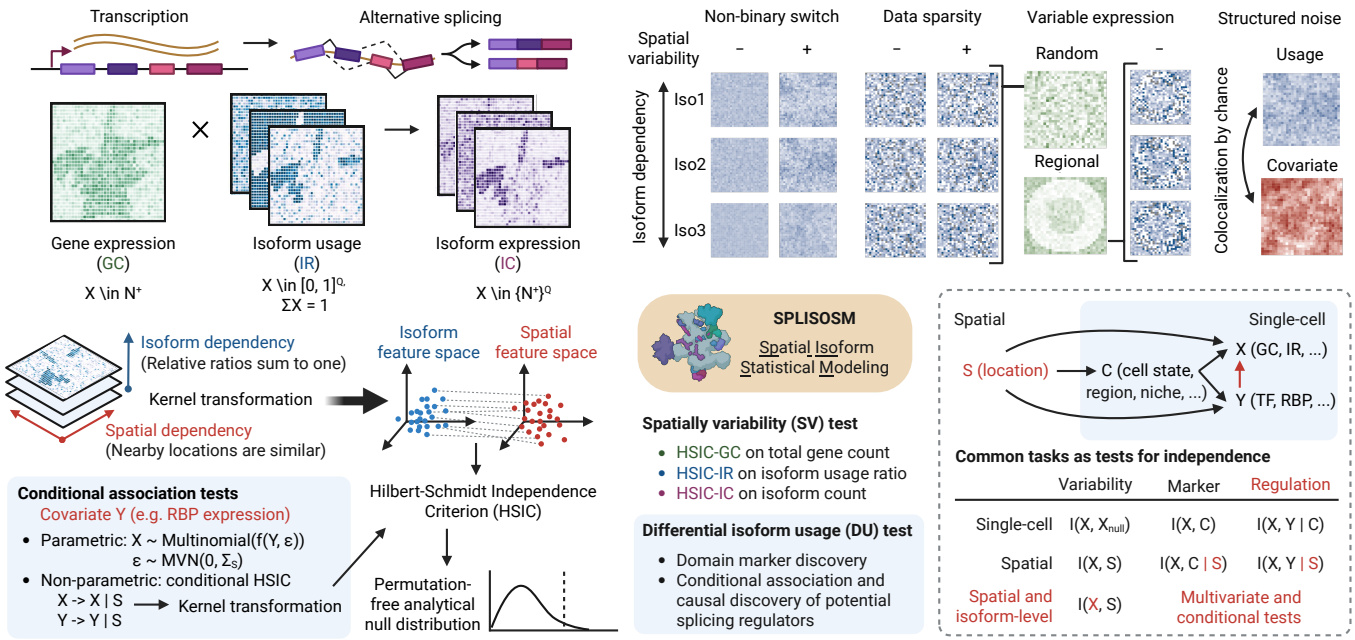

Figure 1: A computational toolbox for spatial isoform pattern discovery
Alternative pre-mRNA processing generates multiple isoforms from the same gene, with final isoform expression determined by both gene expression (controlled by transcriptional factors) and isoform usage (regulated by splicing machinery and RBPs). SPLISOSM addresses challenges in spatial isoform pattern analysis by recasting spatial variability (SV) and differential isoform usage (DU) detections as multivariate independence tests. For SV testing, multivariate isoform usage and spatial coordinates are projected into respective feature spaces to assess their nonlinear dependencies using the Hilbert-Schmidt Independence Criterion (HSIC) with optimized spatial and isoform kernels. For DU testing, we implement a nonparametric conditional HSIC test and a parametric generalized linear mixed models (GLMM) with spatial random effects to control for confounding spatial autocorrelation and reduce spurious associations.

whether the residual variation shows significant association (Methods). By treating spatial coordinates as a continuous cell state representation, our approach enables cell-type agnostic, region-specific discovery, complementing existing single-cell isoform studies[40-42] while offering unique advantages in contexts where discrete cell typing is challenging, such as in tumor microenvironments.

SPLISOSM produces well-calibrated and permutation-free p-values in simulation

In the absence of existing methods for spatial isoform analysis, we first validated SPLISOSM through carefully designed simulations of isoform-level ST datasets with shallow sequencing depth (<5 UMI per gene per spot), which produced nearly binary observed isoform usage patterns reflecting both gene expression and isoform preference variation (Figure 2A, Methods).

Our spatial variability tests target distinct biological signals with specialized statistics (Figure 2B). Among them, HSIC-IR correctly identified spatial patterns in isoform usage while ignoring changes in total gene expression (Scenario 4 vs 5). It employs a modified linear kernel with mean replacement to handle compositional vectors with excessive zeroes and undefined ratios, providing optimal power without introducing artifacts (Figure S1A and B). Complementarily, HSIC-IC detected both isoform usage (Scenario 1 vs 3) and gene expression spatial variation (Scenario 1 vs 4). We also observed improvements in the gene-level HSIC-GC compared to established methods like SPARK-X through the optimized spatial kernel (Scenarios 4-6), validating our theoretical predictions.

For association analysis, we focus on the challenge of correlated noise, which creates patterns that falsely align with covariates (Scenario 1 vs 2). Conventional approaches, including discrete T-tests and continuous multivariate correlation, and generalized linear models (GLMs), all exhibited significant p-value inflation under spatial confounding (Figure 1C, Scenarios 1 vs 2 and 4 vs 5). SPLISOSM overcomes this limitation by introducing spatial conditioning within existing differential analysis frameworks (Methods). Simulations confirmed that our non-parametric conditional HSIC maintains proper null calibration while preserving sensitivity to subtle covariate effects (Figure 1C, Figure S1C-E), and that our parametric generalized linear mixed model (GLMM) partially controlled false positives compared to standard GLMs.

Integrative analysis reveals spatial alternative splicing programs in adult mouse brain enriched for synaptic and membrane trafficking functions

Recent studies[23,24] have generated spatial isoform maps of postnatal and adult mouse brains by combining 10X Visium spatial barcoding with Oxford Nanopore (ONT) sequencing. These emerging long-read ST datasets suffer from limited throughput and measurement noise, providing an ideal testbed to demonstrate SPLISOSM's utility in biological discovery.

We first analyzed the SiT[24] mouse olfactory bulb (MOB) dataset containing 918 spots, with less than 0.5 total ONT UMI per spot on average and more than 75% zero-coverage

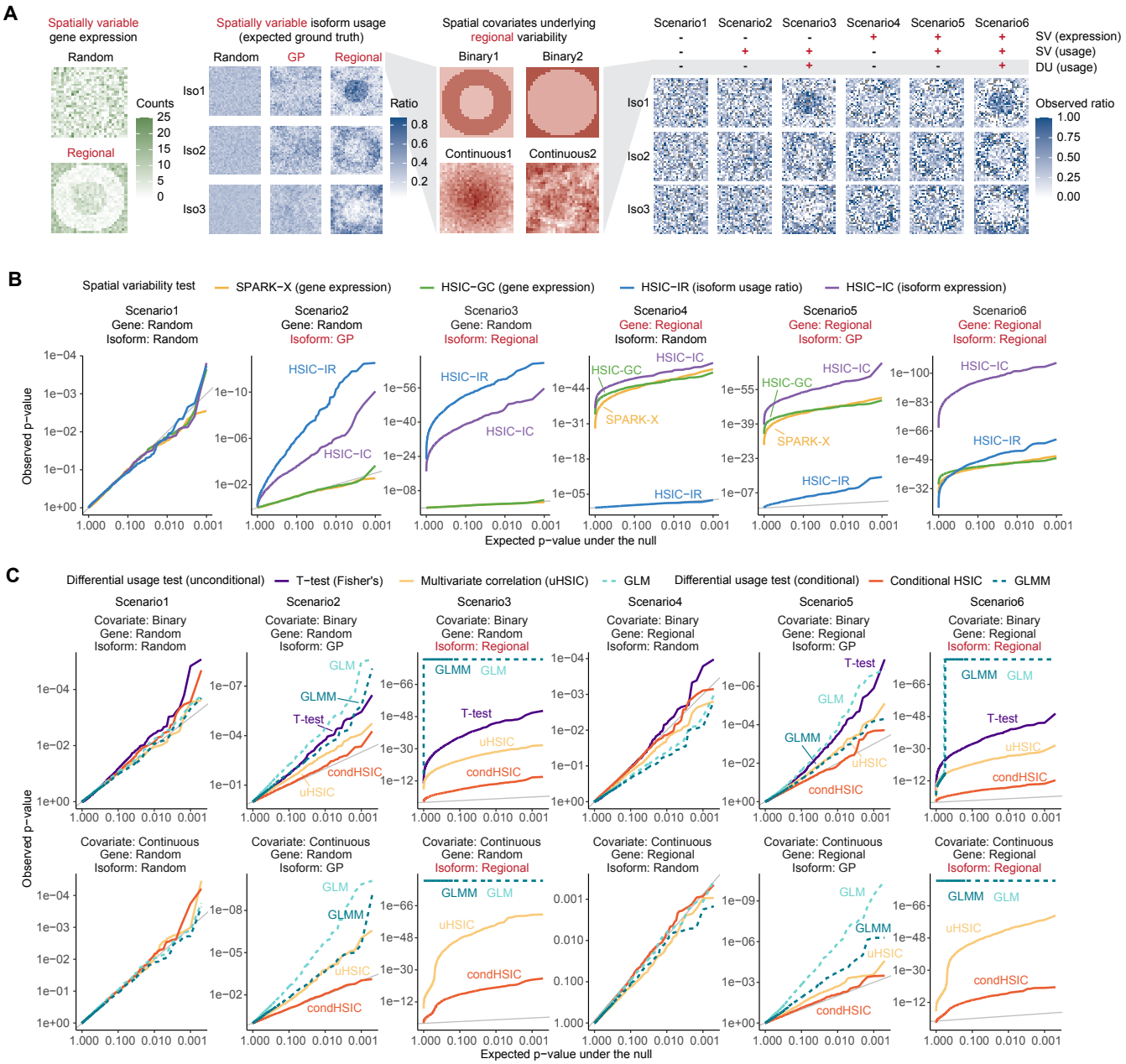

Figure 2: SPLISOSM produces well-calibrated and permutation-free p-values in simulation
(A) Six simulation scenarios with spatial variability at different layers. Regional variability was generated from two binary and two continuous spatial covariates. **(B)** Q-Q plots of SV test p-values across 1,000 simulated genes per scenario. Regional gene expression (Scenarios 4-6) was included to introduce artifacts in the observed ratios. **(C)** Q-Q plots of differential usage test p-values aggregated from 2,000 tests (1,000 genes x 2 covariates) per covariate type. Variable isoform usage independent of covariates (Scenarios 2 and 5) was sampled from a Gaussian process (GP) to mimic spurious differential associations. For binary covariates, T-tests were performed on individual isoforms with p-values combined at gene-level using Fisher's method. For GLM and GLMM, null models with zero effect size were fitted and tested using the score statistic.

spots for a median gene (Figure S2A). Despite this sparsity, SPLISOSM detected over 100 spatially variably spliced (SVS) genes (p-value < 0.05, Figure S2B), including previously undocumented spatial variation in membrane trafficking genes (*Clta, Cltb, Ap2m1, Snap25*) with known tissue specificity and several ribosomal genes (*Rps9, Rps24, Rpl5*). Importantly, HSIC-IR identified spatial transcript variation independent of total expression levels (Figure S2C), revealing functionally distinct signatures enriched in neurodegenerative disease pathways (Figure S2D) that could define major olfactory layers based solely on their usage ratios (Figure S2H). Structurally, SVS genes showed increased exon skipping near the 3'end and motif enrichment of neural RBPs from the RBFOX and ELAVL families (Figure S2E-G). By associating isoform usage with RBP expression, we identified 7 significant RBP-SVS gene pairs, with 2 showing direct RBP binding near regulated events in crosslinking immunoprecipitation (CLIP) data[43] (Figure S2I, Methods). The remaining associations likely represent CLIP false negatives (as with Fmr1-Myl6, Figure S2J) or indirect regulation (as with Rbfox3-Cltb, Figure S2K).

Having validated SPLSISOSM's performance on the MOB dataset, we expanded our analysis to include two larger SiT samples of mouse coronal brain section (CBS, Figure 3A). Given the brain's complex cellular organization, we asked whether spatial isoform usage contributes to functional heterogeneity across brain regions. SPLISOSM detected 150 SVS genes shared by both replicates (adjusted p-value < 0.05, Figure 3B) --- more than double the originally reported 61 genes --- without relying on region annotations. Most SVS genes also exhibited spatial variability in total expression (Figure 3B). Our spatial variability tests demonstrated strong consistency between replicates (Spearman's rho of test p-values = 0.52, 0.39, 0.43 for HSIC-IR, HSIC-GC, and HSIC-IC, respectively), while SPARK-X biased toward the sample with higher sequencing coverage (Spearman's rho = 0.26). In line with the MOB results, these SVS genes were enriched in membrane trafficking and synaptic signaling (*Clta, Cltb, Snap25, Stxbp1, Ap2a1, Atp6v1e1, Napa*), with many implicated in neurodegenerative diseases (Figure 3C). We also observed widespread isoform variability in ribosomal genes (*Rps24, Rps6, Rpl13a, Rpl5, Rps9*), aligning with previous studies showing cell-type and tissue-specific ribosome composition[44]. Importantly, transcript usage variability alone proved sufficient to define brain regions with comparable spatial resolution to expression-based annotation methods (Figure 3D, Methods). Gene-wise clustering further organized spatially variable isoform preference into region-specific programs, likely capturing shared regulation (Figure 3E). Together, these findings underscore the biological significance of isoform specificity in establishing and maintaining brain function.

At the structural level, SVS genes showed enrichment for exon skipping particularly near the 3'end, as well as alternative first exons (Figure 3F and G). Although RBP expression patterns collectively mirrored brain region specialization (Figure 3H), no single known RBP motif could fully explain the observed spatial exon variability, prompting us to systematically search for all potential splicing modulators --- whether acting through direct binding or as indirect cofactors --- that might drive regional specificity. Our differential analysis identified 746 RBP-SVS pairs significant in both replicates (Figure 3I, Methods). The top regulators included both well-established neural RBPs like QKI, RBFOX, and CELF families and less-studied RBPs like ARPP21. To replicate these relationships, we examined available brain CLIP datasets[43], which independently confirmed 4 of 9

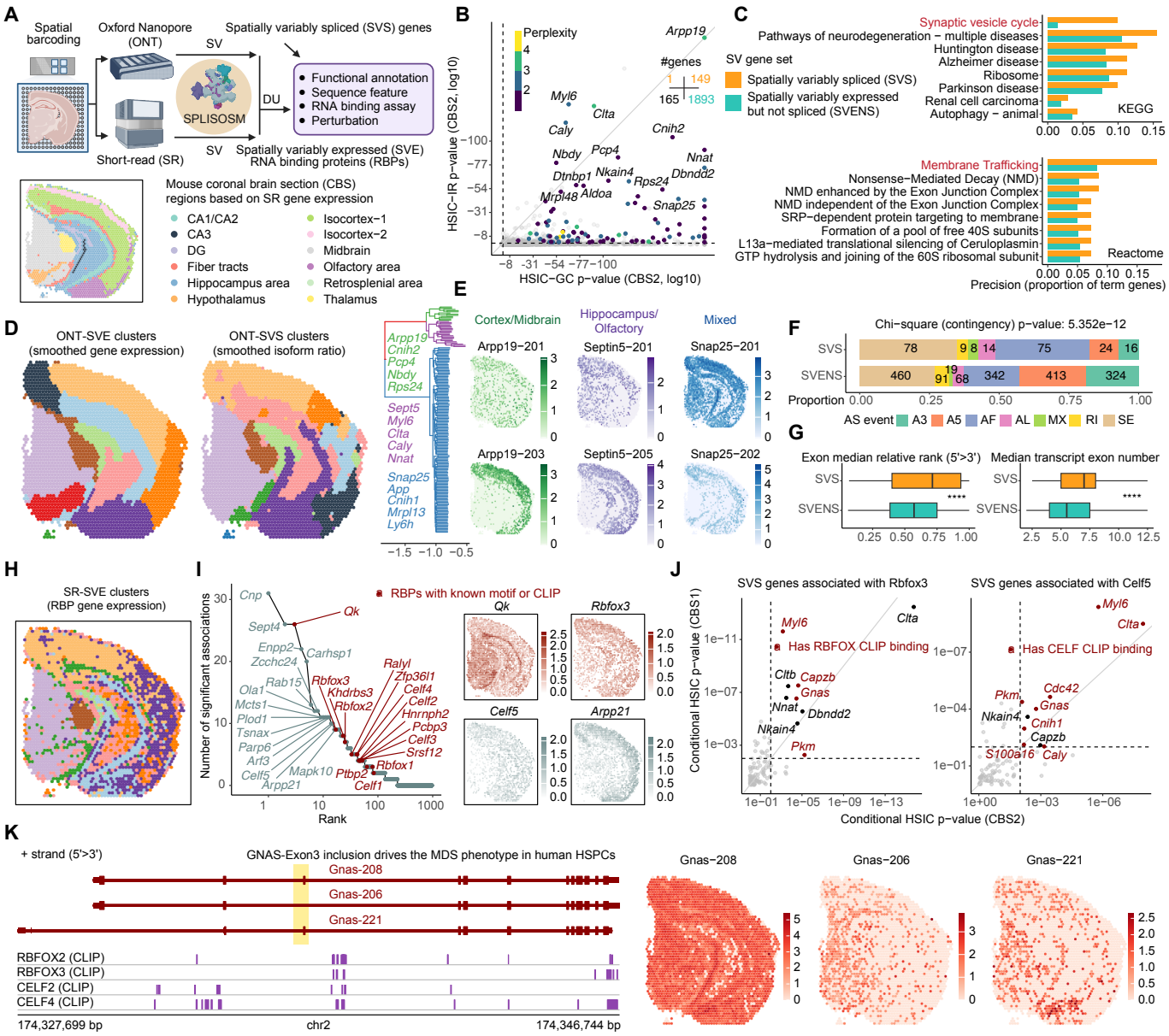

Figure 3: Integrative analysis reveals spatial alternative splicing programs in adult mouse brain enriched for synaptic and membrane trafficking functions

(A) Workflow overview. **(B)** Comparison of spatial variability in gene expression (x-axis) versus isoform usage (y-axis) in the CBS2 sample. Points are colored by expression perplexity (effective number of isoforms per gene), calculated as the exponential of global isoform usage entropy. Inset shows number of events significant in both CBS1 and CBS2 replicates (adjusted p-value < 0.05), categorized as spatially variably spliced (SVS) or spatially variably expressed but not spliced (SVENS) genes. **(C)** Pathway enrichment analysis comparing SVS versus SVENS genes, with terms arranged by precision differences between the two groups. **(D)** Unsupervised clustering based on gene-level expression of SVE genes (left) or relative isoform ratios of SVS genes (right). Isoform counts were locally smoothed to compensate for data sparsity. **(E)** Selected isoform usage programs identified through hierarchical clustering of SVS genes. **(F)** Distributions of alternative splicing event types among SVS and SVE genes. **(G)** Selected Matt exon features of skipping exons (SE) and mutually exclusive exons (MX) in SVS and SVENS genes. **(H)** Unsupervised clustering based on SVE RBP expression from short-read sequencing data. **(I)** Top RBPs ranked by number of significantly associated targets, color-coded by availability of documented motifs (CISBP-RNA) or CLIP data (POSTAR3). Log-normalized expression of selected RBPs shown at right. **(J)** Differential isoform usage p-values for Rbfox3 (left) and Celf5 (right) in two replicates. Validated targets with RBP binding near regulated events in brain-CLIP data are highlighted in red. **(K)** *Gnas* transcript structure and spatial log-normalized expression patterns.

significant *Rbfox3* associations. Although CELF5-specific CLIP data remains unavailable, 8 of 10 of its predicted target genes showed binding by other CELF family members, supporting CELF5's role as a major neural splicing modulator in adult mouse brain (Figure 3J).

Among SVS genes, *Gnas* showed particularly intriguing regulation (Figure 3K). Exon 3, which is spliced out specifically in midbrain and fiber tracts, generates a hyperactive G-alpha protein driving differentiation defects and myeloid malignancies when included due to SRSF2 and U2AF1 mutations[45]. Our analysis reveals that RBFOX and CELF families regulate *Gna*s exon 3 skipping in the brain (Figure 3K), raising the possibility that neural RBP dysregulation in non-neural tissues could contribute to disease pathogenesis[45].

Spatial 3'end transcript diversity in adult mouse brain extends beyond alternative polyadenylation and shows functional convergence on signaling pathways

In addition to alternative splicing, mammalian nervous systems are also marked by active alternative polyadenylation (APA)[46] that generates transcript diversity near the 3'end (TREND[47]). While APA often modulates gene expression by altering 3′ untranslated region (UTR) length, some events directly modify coding sequences through intronic polyadenylation, producing truncated or functionally distinct proteins[48].

Since sequencing-based ST protocols typically capture the 3′ end of polyadenylated mRNA transcripts, we reasoned that SPLISOSM could be applied to study APA regulation across spatial dimensions even with short-read data (Figure 4A, Methods). In a new 10X Visium sample of mouse coronal brain section (CBS), we identified 815 SVS genes with spatially variable TREND usage (adjusted p-value < 0.01, Figure 4B), with 95 shared with SiT ONT CBS samples. This substantial increase in SVS genes mirrored higher sequencing depth (median 1 UMI per gene per spot) and the subsequent ability to test more genes in SR data. We observed similar patterns at near-cellular spatial resolution in a Slide-seqV2 dataset of mouse hippocampus[22], where 42 of 52 SVS genes were also variable in the Visium dataset (adjusted p-value < 0.01, Figure S3A), despite extreme data sparsity (median UMI < 0.1 per gene per spot). These results confirmed the robustness and reproducibility of our approach for studying isoform variation in various ST data.

Although more than 80% of spatially variable TREND events mapped to 3'UTRs as expected, we found significant enrichment at junctions and alternative exons (Figure 4C). Some non-3'UTR events represented internal poly(A) priming sites, supported by the enrichment of A-rich motifs, allowing us to detect upstream diversity deeper into the transcript. The functional importance of TREND became evident as 25% of events overlapped coding regions. Of the 7 previously reported protein-altering poly(A) site shifts between neurons and glia[49], we recapitulated 6 (*Map4, Atp2a2, Cdc42, Klc1, Itsn1, Ptpn2*) while providing new insights into their spatial distribution (Figure 4E and Figure S3B). More generally, SVS genes converged on key neural signaling pathways, including glutamatergic and adrenergic synaptic transmission and Rap1, Ca^{2+}, cAMP, and MAPK cascades (Figure 4D). We observed additional protein-coding changes in genes involved in diverse cellular functions (Figure S3C), including proteostasis (*Hsp90aa1, Hsp90ab1*),

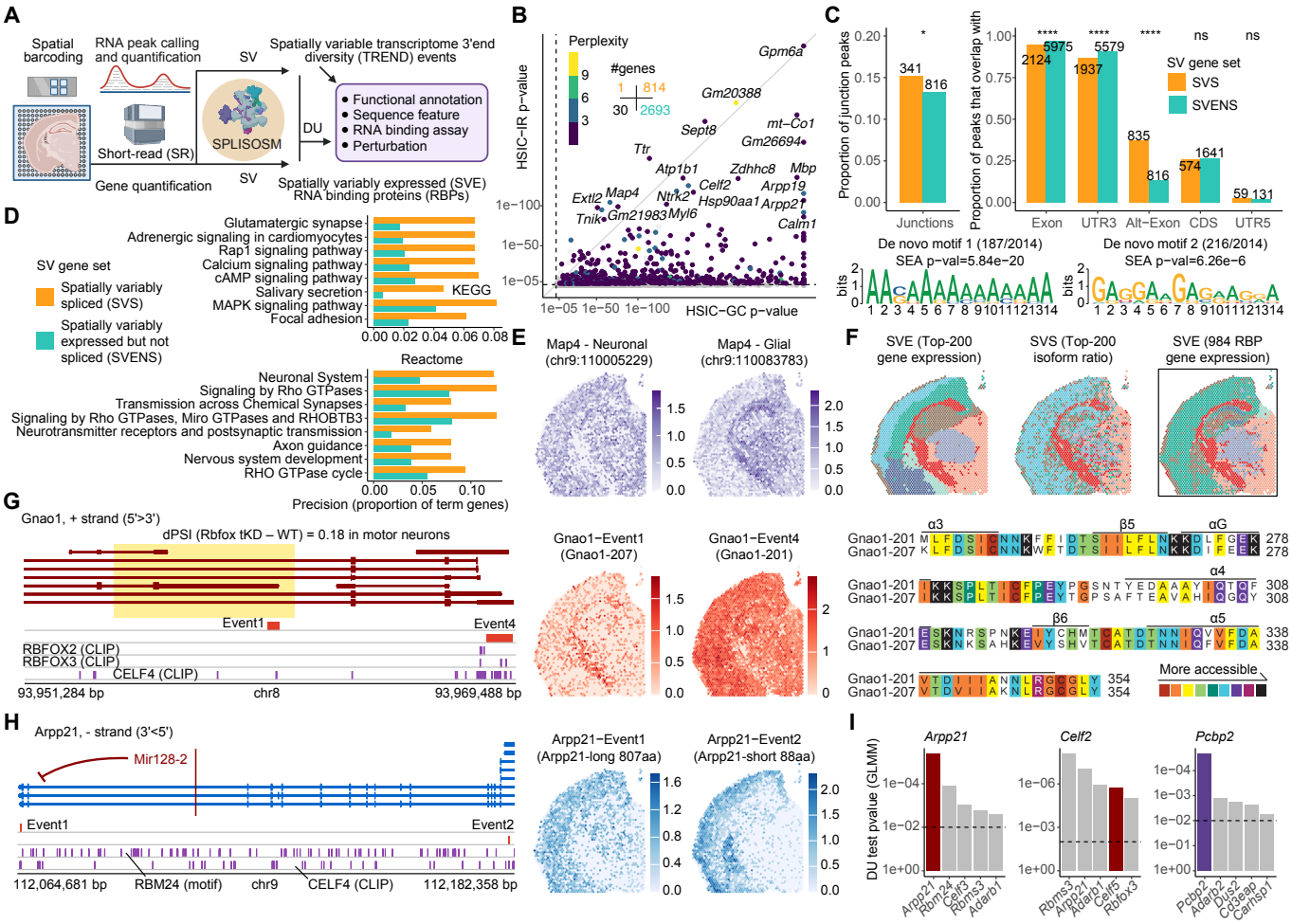

Figure 4: Spatial 3'end transcript diversity in adult mouse brain extends beyond alternative polyadenylation and shows functional convergence on signaling pathways

(A) Workflow overview. **(B)** Comparison of spatial variability in gene expression (x-axis) versus TREND usage (y-axis) in the SR coronal brain section sample. Points are colored by expression perplexity (effective number of isoforms per gene), calculated as the exponential of global isoform usage entropy. Inset quantifies significant events (adjusted p-value < 0.01). **(C)** TREND event annotation and motif enrichment results (bottom). A TREND event may overlap with multiple genomic features. Alternative exons (Alt-Exon) are defined as exons located within introns of other transcript variants. **(D)** Pathway enrichment analysis comparing SVS and SVENS genes. **(E)** Spatial log-normalized expression of the neuronal and glial *Map4* variants. **(F)** Unsupervised clustering based on gene-level expression of top 200 SVE genes (left), relative TREND ratios of top 200 SVS genes (middle), and gene expression of all SVE RBPs (right). **(G)** *Gnao1* transcript structure and spatial log-normalized expression patterns. The resulting protein sequence variation, color-coded by solvent-accessible surface area, is shown on the right. **(H)** *Arpp21* transcript structure and spatial log-normalized expression patterns. **(I)** Top RBPs associated with TREND usage in *Arpp21*, *Celf2* and *Pcbp2*.

membrane trafficking and secretion (*Septin8, Septin11, Ap1s2*, *Olfm1, Lgi3*), and enzymatic activities (*Kalrn, Rexo2, Ppp2r5c*). Finally, major brain regions could be defined using TREND preference alone in agreement with our long-read analysis, with isocortex layers being the exception due to weaker signals (Figure 4F).

To understand the spatial regulation of TREND, we performed association analysis between event usage and RBP gene expression. Contrary to conventional thinking, spatial variability in core alternative polyadenylation machinery components was insufficient to explain the observed 3'end diversity patterns (Figure S3D). Instead, neural splicing regulators identified in our ONT analysis such as QKI, RBFOX, and CELF families also emerged as top regulators for APA, supporting their central role in shaping the adult mouse brain transcriptome[50]. A prominent RBFOX target is *Gnao1*, which encodes a subunit of the heterotrimeric G proteins. We independently validated Rbfox-Gnao1 regulation using CLIP[43] and knockout data[51]. Our spatial analysis uncovered previously unrecognized segregation in *Gnao1* last exon usage between cerebral nuclei and fiber tracts versus other brain regions, generating two protein isoforms with different C-termini (Figure 4G). The regional isoform specificity extends to other G proteins (*Gnai2, Gnas, Gnal, Gnaq*) and G-protein-coupled receptors (*Gabbr1, Grm5, Adgrl1*), indicating a potential role of alternative splicing and polyadenylation in fine-tuning signal transduction for spatially diverse physiological functions and pharmacological responses[52,53].

Our catalog also brought to light self-regulatory mechanisms driving transcript diversity in RBP genes themselves. For instance, the last exon usage of *Arpp21* showed strong correlation with its total gene expression (Figure 4H), consistent with a regulatory model where miR-128-2, produced from the longer isoform's second-to-last intron, inhibits *Arpp21* expression[54]. We also verified self-regulation in the 3'UTR of CELF family[55] (Figure 4I) and discovered additional spatially variable 3'UTR lengths across RBPs (*Elavl2, Elavl3, Qk, Rbfox1, Hnrnpa2b1, Hnrnpk*). One particularly intriguing novel finding involves *Pcbp2*, which showed enriched usage of a minor upstream last exon in isocortex and thalamus. This alternative exon encodes a protein variant with an extended disordered C-terminus (Figure S3E), whose usage inversely correlated with overall *Pcbp2* expression in both mouse brain and across human tissues (Figure S3F). CLIP data[56] revealed stronger human PCBP2 binding to this minor last exon compared to the canonical one, suggesting a selective auto-regulatory feedback loop that contributes to spatial specialization.

In summary, we report extensive 3'end transcript diversity in the mouse brain that extends well beyond alternative polyadenylation. These region-specific isoform patterns are orchestrated by neural RBPs through complex regulation, providing a molecular foundation for understanding the functional specialization of brain regions.

RBFOX and CELF families regulate neural transcript diversity cooperatively

Causal inference from observational studies faces drawbacks from unknown confounding factors[57]. Since SPLISOSM's differential test captures both direct and indirect associations in the general form, it is not immune to such confounding effects. By

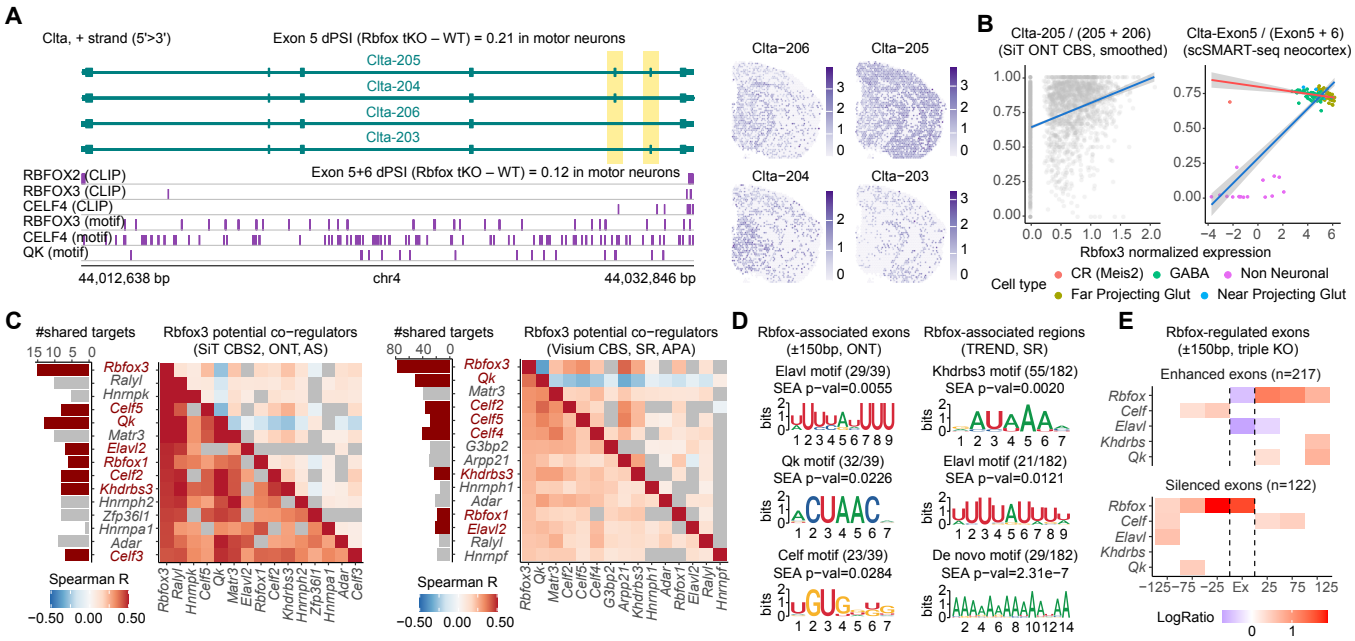

Figure 5: RBFOX regulates neural transcript usage with other splicing factors cooperatively

(A) *Clta* transcript structure and spatial log-normalized expression in the ONT-CBS2 sample. **(B)** Relationship between normalized *Rbfox3* expression and *Clta* isoform ratio (left) or cell-type-aggregated exon PSI values (right). Spatial data was locally smoothed before calculating isoform ratios for visualization. **(C)** Identification of Rbfox3 co-regulators in ONT (left) and SR (right) samples. Heatmaps display pairwise correlations of gene expression (upper triangle) or conditional HSIC test p-values (lower triangle). RBPs with reported RBFOX interaction are highlighted. **(D)** Motif enrichment in Rbfox1/2/3 associated SVS targets using alternative exons and 150bp flanking introns (left) or TREND regions (right). **(E)** Motif enrichment in exons with >=25% PSI changes following RBFOX triple knockout in Day 10 motor neurons, categorized as silenced (increased PSI after KD) or enhanced (decreased PSI). Sequences were binned by distance to exon boundaries (x-axis) to show positional enrichment.

comparing spatial RBP associations with perturbation studies in specific cellular environments, we can nevertheless extract meaningful biological insights.

A revealing example is the spatial regulation of *Clta* isoforms (Figure 5A). Knockout studies demonstrated that RBFOX proteins suppress *Clta* exons 5 and 6 inclusion in developing motor neurons[51], yet our spatial test revealed a surprising positive correlation between *Rbfox3* expression and the Clta-205 isoform containing these exons. Analysis of exon inclusion in a single-cell neocortex atlas initially confirmed this positive relationship across brain cell types (Figure 5B). The perceived contradiction resolved, however, when examining only neuronal subtypes: within GABAergic and far-projecting glutamatergic neurons, *Rbfox3* negatively correlates with exon inclusion, aligning with the knockout results. This context-dependent relationship implies the presence of cell-type-specific factors, countering RBFOX, that drive up *Clta* exon 5 usage in neurons.

The case of *Clta* echoes recent findings that network inference methods can identify regulatory partners correctly but mischaracterize their regulatory mode[58]. We hypothesized that such conflicts stem from multiple splicing factors working in concert. In our spatial data, we found several additional genes (*Myl6, Cltb, Tln1, Klc1*; Figure S4B-D) where RBFOX knockout effects contradicted spatial associations. Supporting our hypothesis, previous experiments have shown that repressive CELF binding can override RBFOX enhancement at *MYL6* exon 6 in human T cells[59].

To systematically identify potential RBFOX co-regulators, we assessed agreement between RBP differential usage results and spatial co-expression patterns (Figure 5C). The top candidates included RBPs with established RBFOX interactions: CELF family proteins, known to act antagonistically with RBFOX in muscle and heart tissue[59]; ELAVL2, which shares targets with RBFOX1 in human neurons[60]; and QKI, recently predicted to affect RBFOX function[61]. Expression of these collaborating RBPs remained stable after RBFOX depletion (Figure S4A), pointing toward direct cooperation rather than cross-regulation feedback. Sequence motif analysis provided further validation, showing significant enrichment of ELAVL, CELF, and QKI binding motifs near Rbfox-associated variable exons detected in the ONT data, while KHDRBS3 (a QKI homolog) and ELAVL motifs were enriched near Rbfox-associated TREND regions in short-read data (Figure 5D). Consistently, we observed position-dependent overrepresentation of CELF and ELAVL motifs around exons responding to RBFOX triple knockout in the perturbation data[51], supporting the collaborative regulation between multiple RBPs on the same targets (Figure 5E).

Evolutionarily conserved transcript diversity in synaptic genes across the human prefrontal cortex

Given the critical role of splicing and APA in brain functions, we hypothesized that spatial transcript diversity patterns would be highly preserved throughout mammalian evolution. To investigate this, we reanalyzed 12 10X Visium samples[62] from healthy human brain (Figure 6A, Methods), identifying 861 SVS genes in the dorsolateral prefrontal cortex (DLPFC, adjusted p-value < 0.05). Power to detect spatial variability in these human samples has not yet saturated with respect to sequencing depth (Figure 6B and Figure

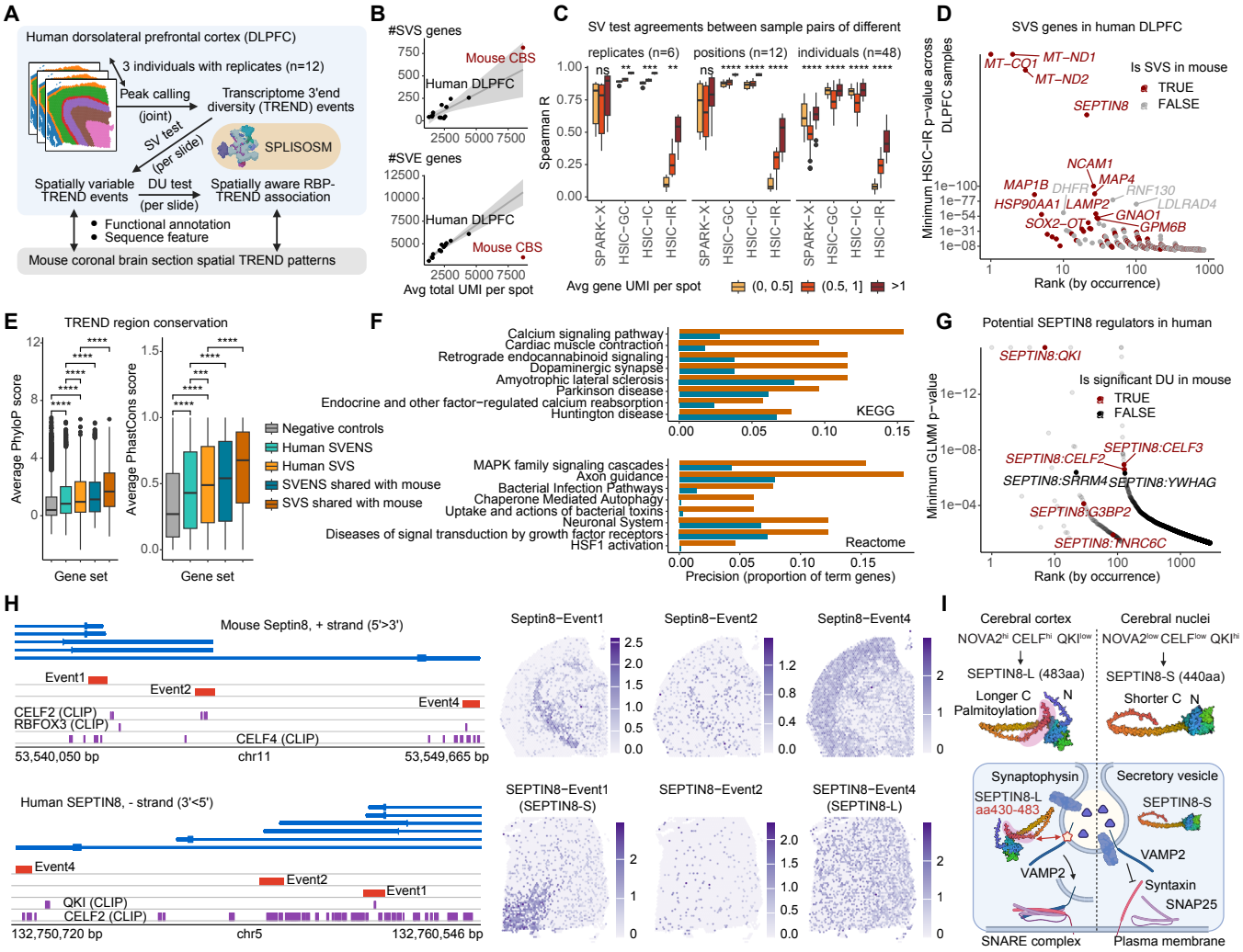

Figure 6: Evolutionarily conserved transcript diversity in synaptic genes across the human prefrontal cortex
(A) Workflow overview. **(B)** Relationship between total read coverage in TREND regions per spot (x-axis) and number of significant spatially variably spliced (SVS, top) or variably expressed (SVE, bottom) genes across datasets. **(C)** P-value correlation distributions between technical replicates (left), different sections from same donor (middle), and different donors (right). Genes are stratified by average TREND read counts per spot. **(D)** SVS genes ranked by recurrence (number of significant samples, x-axis) and the minimal HSIC-IR p-values across samples (y-axis). Highlighted genes have mouse homologs that are also spatially variably spliced. **(E)** Distributions of average conservation scores in TREND regions across gene categories: non-variable controls, human-specific SVENS/SVS, and human-mouse shared SVENS/SVS. **(F)** Pathway enrichment analysis comparing SVS versus SVENS genes conserved between human and mouse. **(G)** RBP-SVS associations ranked by recurrence (x-axis) and minimal GLMM p-values across samples (y-axis). Top potential regulators of *SEPTIN8* are highlighted and colored by whether the association is conserved in mouse. **(H)** *SEPTIN8* transcript structure and spatial log-normalized expression in mouse (top) and human (bottom, sample 151673). **(I)** Comparative illustration of SEPTIN8-long and SEPTIN8-short protein variants.

S5A), so improved RNA capture per spot would likely reveal additional variable events. When evaluating performance across replicates, SPLISOSM's expression-based tests (HSIC-GC and HSIC-IC) demonstrated better power and consistency compared to SPARK-X, which was hampered by suboptimal kernel design (Figure 6C). The HSIC-IR test showed weaker concordance, primarily due to data sparsity and noise, though this improved significantly for genes with higher expression.

Despite these technical limitations, we found strong evidence for cross-species conservation at both individual gene and functional levels. Of the human SVS genes, 178 had mouse homologs that also exhibited spatial variability (Figure 6D and Figure S5B-C). Moreover, conserved TREND regions displayed significantly higher sequence conservation than invariant regions (Figure 6E). Functional enrichment analyses in both species highlighted the importance of isoform-level regulation in synaptic signaling and chaperone-mediated autophagy, processes implicated in various neurodegenerative disorders (Figure 6F, Figure S5D).

The molecular regulation governing spatial isoform diversity also showed remarkable cross-species similarities. MAP4, which organizes microtubules in muscles and neurons, exhibited patterns corresponding to neuronal-vs-glial usage in both species (Figure 4E and Figure S5G). While PTBP2 was previously identified as a MAP4 regulator in embryonic brain cortex[49], its limited spatial variability in adult brain suggested the involvement of additional factors in region-specific regulation (Figure S5E). Our differential usage analyses pointed to RBFOX1, QKI, and CELF families as likely co-factors, with CLIP data revealing conserved binding sites supporting this hypothesis (Figure S5F).

SEPTIN8 provides another compelling example of evolutionarily conserved regulation. This gene encodes a GTP-binding protein that polymerizes into filaments essential for cellular architecture. Our spatial analysis revealed a distinct regional pattern with a shorter SEPTIN8 isoform predominating in mouse fiber tracts and human DLPFC white matter (Figure 6H). This truncated protein lacks the extra C-terminal region that interacts with VAMP2 to facilitate SNARE-complex assembly[63], which also features a filopodia-inducing motif (FIM) that enhances dendritic arborization when palmitoylated[64] (Figure 6I). The alternative last exon of Septin8 is regulated by NOVA2 in mouse motor neurons[64]. Our analysis indicates that NOVA2 likely cooperates with QKI and CELF proteins jointly to achieve the precise spatial patterning in the brain (Figure 6G).

Spatial transcript diversity in human glioma is shaped by immune infiltration and microenvironment composition

The capability to detect patterns without cell-type or region annotations makes SPLISOSM especially valuable for complex tissues with ambiguous cellular states. Glioblastomas are characterized by their infiltrative nature and heterogeneous spatial structures[65]. To explore isoform distributions in this challenging context, we systematically mapped spatial transcript usage across 24 human glioma samples (Figure 7A, Methods), identifying 306 spatially variably processed (SVP) genes in 11 ONT samples[66] and 2,828 SVP genes in 13 SR samples[65] (adjusted p-value < 0.05, Figure 7B). While around 50%

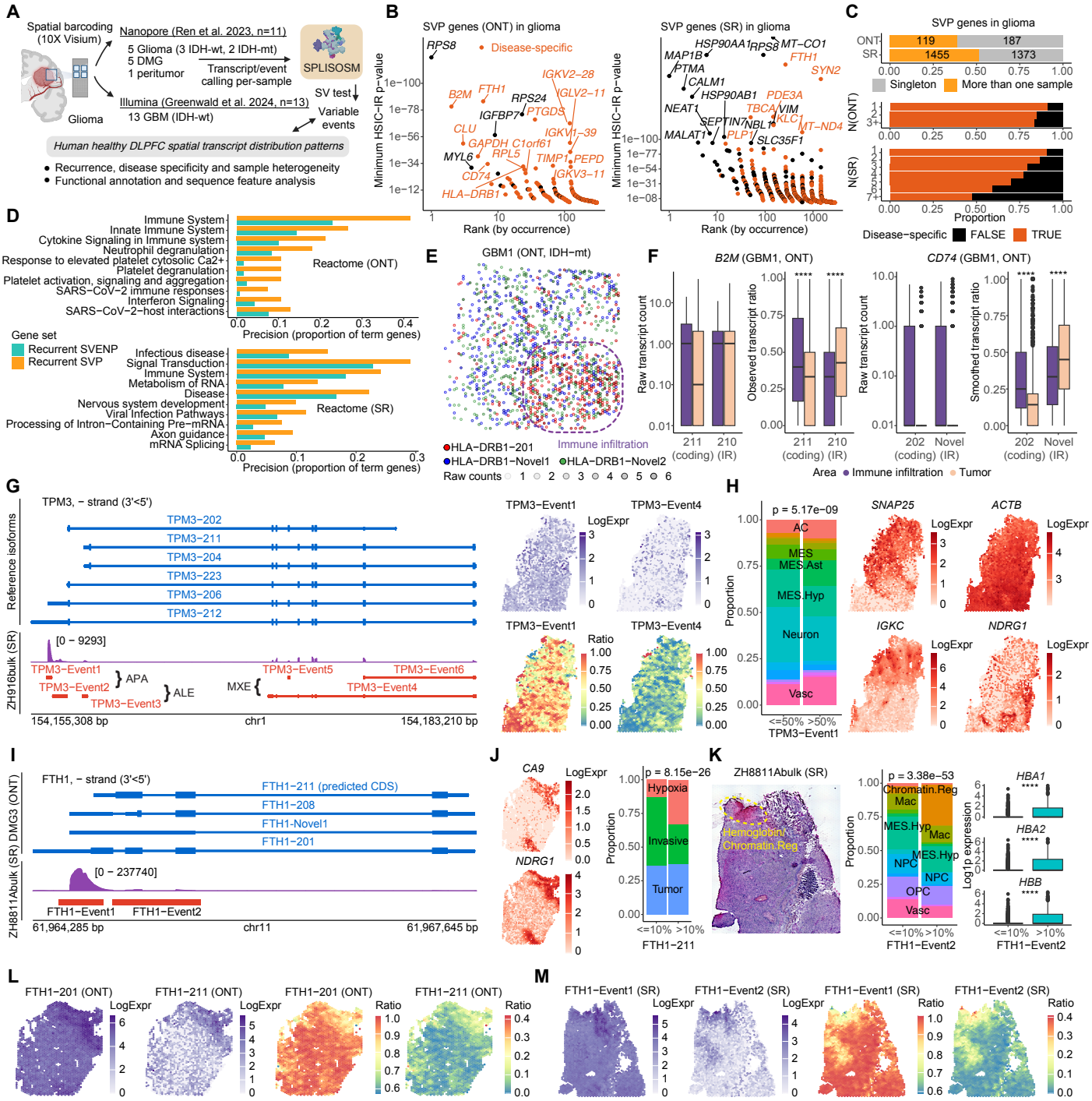

Figure 7: Spatial transcriptomic diversity in human glioma is shaped by microenvironment composition and immune infiltration
(A) Workflow overview. **(B)** Spatially variably processed (SVP) genes ranked by recurrence (x-axis) and the minimal HSIC-IR p-values across samples (y-axis) in ONT (left) and SR (right) cohorts. Genes not variable in healthy DLPFC are highlighted as disease specific. **(C)** Distribution of recurrent and glioma-specific SVP genes. **(D)** Pathway enrichment analysis comparing recurrent SVP (>=2 samples) genes versus recurrent SVENP (>=4 samples) genes. **(E)** Spatial distribution of *HLA-DRB1* isoforms in the ONT sample GBM1. **(F)** *B2M* and *CD74* isoform expression in GBM1, comparing tumor versus immune-infiltrated regions. Three novel *CD74* intron-retention isoforms are aggregated as one for visualization. **(G)** *TPM3* transcript structure, RNA-seq coverage, and spatial patterns in the SR sample ZH916bulk. Event counts were locally smoothed for ratio visualization, same in **(L)** and **(M)**. **(H)** Distribution of glioblastoma metaprograms in regions with high/low TPM3-Event1 usage. Shown on the right are log-normalized expression of selected markers: *SNAP25* (neuron), *ACTB* (glial), *IGKC* (B cell), *NDRG1* (hypoxia). **(I)** *FTH1* transcript structure in the ONT sample DMG3 and its RNA-seq coverage in the SR sample ZH8811Abulk. **(J)** Hypoxia marker expression (left) and correlation with FTH1-211 usage in DMG3. **(K)** Histology (left), metaprogram distribution (middle) and hemoglobin expression (right) in regions with high/low FTH1-Event2 usage in ZH8811Abulk. **(L)** Spatial distribution of *FTH1* isoforms in DMG3. **(M)** Spatial distribution of *FTH1* TREND usage in ZH8811Abulk.

SVP genes were sample-specific and disease-specific, the most frequently recurring SR-SVP genes overlapped substantially with those found in healthy brain tissue, reflecting compositional changes of non-tumor cells (Figure 7C, Figure S6A and B). As in healthy brain, approximately 25% of all variable transcript events in glioma had potential implications for protein structure, with alternating exon accounting for much of this variation (Figure S6D). Beyond RNA processing, glioma-specific transcript diversity also emerged from genomic alterations. In one pediatric sample, we detected variable immunoglobulin transcripts resulting from VDJ recombination, revealing clonality in B-cell infiltration (Figure S6C).

The importance of immune interactions was further underscored by consistent variability in antigen presentation components across samples, often correlated with immune infiltration patterns (Figure 7D, Figure S6E-G). This included HLA class I (*HLA-A, B, C*) and class II (*HLA-DRA, DRB1, DPB1*) genes, whose sequence variation likely stems from both alternative splicing and germline variability[67] (Figure 7E). For non-polymorphic elements such as *B2M* and *CD74*, we found preferential usage of intron-retaining, potentially non-functional transcripts in tumor regions (Figure 7F). While the SR data could not capture identical variability events due to its 3'end bias, it demonstrated similar functional convergence toward immune-related pathways, particularly in infectious disease response and viral infection pathways (Figure 7D, Figure S6G).

Our analysis also revealed dysregulation within the splicing machinery itself. In the SR data, functional enrichment identified extensive transcript variability in the 3'end of spliceosome constituents (Figure S6H). This result agrees with reports on extensive cross-regulation splicing networks in human cells[68] and offers a potential explanation for the widespread expression dysregulation of splicing factors in glioblastoma[69].

Furthermore, the tumor ecosystem shaped transcript diversity through cell-cell communications. We detected spatial variability in signal transduction, focal adhesion and cytoskeleton regulation pathways (Figure S6I), reflecting the intricate cellular interactions and extracellular matrix remodeling within tumors. *TPM3*, a tropomyosin family member with well-documented cancer-associated splicing patterns across tumor types[70], exhibited complex and distinct splicing variation (Figure 7G). We identified an upstream *TPM3* event (Event4) that partially overlapped with neuronal and hypoxia signatures, while the canonical form (Event1) associated with glial programs and increased actin expression while negatively correlating with B-cell infiltration (Figure 7H). These results resonates with recent reports of differential splicing of actin cytoskeleton components between peripheral and core glioblastoma cells[71]. Additionally, we observed spatially variable usage of the last exon of *GFAP*, encoding two variants of an astrocytic intermediate filament protein whose ratio is linked to glioma grade and malignancy[72,73] (Figure S6J). The location-specific distribution of these cytoskeleton gene isoforms indicates dynamic responses to local microenvironmental cues, modulating migration potentials and driving glioblastoma invasion.

Microenvironmental influences extended to ribosomal proteins as well. We observed spatial variability in transcripts encoding ribosomal proteins (including *RPS8, RPS24, RPL5, RPLP1, RPS6, RPS12*) associated with tumor infiltration, frequently involving retained introns or truncated 3' ends (Figure S7). While some such as *RPS8* also

exhibited variability in normal brain, the altered read distributions clearly displayed glioblastoma-specific changes (Figure S7A), aligning with prior findings that ribosomal isoform switching occurs in response to microenvironmental cues and contributes to metabolic reprogramming phenotypic plasticity in glioblastoma[74].

FTH1 offers another prominent example of microenvironment-driven transcript usage (Figure 7I). This iron-storage protein, upregulated under hypoxic conditions, protects cells from ferroptosis by reducing reactive iron[75]. We discovered that *FTH1* upregulation in hypoxic tumor regions coincided with increased expression of 3' truncated transcripts lacking the final exon-exon junction (Figure 7J). In the SR data, we identified a minor event spanning *FTH1* exons 2 and 3, likely representing another non-canonical isoform. This variant predominantly appeared in regions with erythrocytes and overlapped with the chromatin regulation metaprogram (Figure 7K). Collectively, our findings suggest there is coordinated regulation between oxygen-sensing, iron metabolism, and stress response pathways through transcript-level modulation of FTH1.

Discussion

Spatial omics techniques are transforming our understanding of tissue organization, yet few studies have explored molecular diversity at the isoform level. Here we present SPLISOSM, a computational framework designed for robust detection of spatially variable transcript processing and its regulation in spatial transcriptomics data. With theory-grounded design of non-parametric kernel tests, our approach achieves high statistical power even in extremely sparse data while delivering well-calibrated, permutation-free p-values.

SPLISOSM has broad applicability across ST platforms and sequencing protocols. We show that standard 3' end sequencing-based methods capture information on isoform diversity, enabling researchers to extract novel insights from existing datasets. While primarily validated on 10X Visium-based spatial barcoding, SPLISOSM performs effectively across various spatial resolutions, including near-single-cell Slide-seqV2 data. To enhance computational efficiency, we implemented low-rank approximations based on spectral graph theory, allowing variability testing for more than 50,000 spots within minutes on a standard laptop with minimal power loss (Supplementary Note). This scalability extends SPLISOSM's potential applications to imaging-based spatial transcriptomics with isoform-specific probes, which could illuminate transcript subcellular localization preferences. As current sequencing depth (few thousands UMI per spot) has not yet saturated the detection of SVP genes, advances in RNA capture efficiency will further expand our understanding of spatial transcript diversity.

Using SPLISOSM, we have uncovered substantial and evolutionarily conserved spatial transcript diversity across mouse and human brain regions. Our findings reveal previously unappreciated spatial isoform heterogeneity in signal transduction pathways, particularly in G proteins (*Gnas, Gnao1, Gnai2, Gnal, Gnaq*) and G protein-coupled receptors (*Gabbr1*, *Grm5*, *Adgrl1*). This molecular diversity may have significant implications for specialized neural function and region-specific pharmacological responses[76]. Moreover, our integrative regulatory analyses shed light on the complex regulatory network of

splicing machinery. Neural RBPs, including RBFOX and CELF family members, orchestrate transcript diversity with spatiotemporal precision[77] through context-specific expression. While differential association patterns and motif enrichment analyses strongly support coordinated regulation, questions remain about whether these splicing factors operate additively or through direct interactions. Future experiments will be crucial to fully elucidate and validate the molecular basis of cooperative splicing regulation between RBFOX and other neuronal RBPs.

In glioblastoma, our comprehensive analysis reveals how spatial transcript diversity is shaped by microenvironmental composition and metabolic conditions. Immune infiltration emerges as a particularly important driver of isoform variation. In particular, the upregulation of truncated and intron-retaining transcripts in antigen presentation genes implies that glioblastomas may employ post-transcriptional alternative processing as an immune evasion mechanism. Targeting these events may enhance immune response and improve immunotherapy benefits. We also observed significant diversity in ribosomal transcripts, suggesting translational regulation could play an underappreciated role in tumor heterogeneity and thus exposing new therapeutic vulnerabilities[74]. Further research is needed to connect these patterns with pathogenesis and treatment outcomes.

In summary, SPLISOSM brings a new isoform perspective to ST data analysis. The spatial organization of transcript diversity we have mapped in healthy and malignant brain tissues provides a foundation for understanding how alternative isoform processing contributes to specialized functions and disease pathology.

Limitations of the study

The spectrum of detectable transcript diversity events differs across sequencing protocols, and we rely on established pipelines for their definition and quantification. For some data, the preprocessing step could be further optimized to improve coverage through better error correction or shifting focus toward local splicing events. Additionally, while our differential usage test accounts for spatial autocorrelation, it still tests each regulatory relationship independently without explicitly controlling for confounding factors such as co-expression among RNA-binding proteins. Future work may extend the spatial conditioning framework to test conditional independence with multiple potential regulators based on causal graphical models.

Acknowledgements

We thank Chaolin Zhang for input on the splicing regulatory analysis and Daniel Moakley for assistance on the preprocessing of single-cell exon splicing data. We thank Junwei Song for access to processed 10X Visium ONT glioblastoma data and transcript annotation. We thank Jun Hou Fung for feedback on the theoretical analysis. We thank Sanja Vickovic, David McKellar, and Karin Isaev for helpful discussions. This work was funded by the National Institutes of Health, National Cancer Institute (R35CA253126, U01CA243073, P01CA285250 to R.R.), the Edward P. Evans Center for MDS at Columbia University (to J.S. and R.R.), and the National Science Foundation (CAREER DBI2146398 to D.A.K.). Any opinions, findings, and conclusions or recommendations

expressed in this material are those of the authors and do not necessarily reflect the views of the NSF or the NIH.

Contributions

J.S. conceived the project, designed and implemented SPLISOSM, conducted the theoretical analysis, benchmark experiments, and applications on spatial transcript diversity in mouse and human samples. Y.Q. contributed to the parallelization and benchmark of the parametric models; J.S., M.S., and S.P. designed and conducted experimental validation; H.Y. applied SPLISOSM to the Slide-seqV2 hippocampus data; J.J. processed the short-read glioblastoma samples; T.L. developed the website; Q.L. contributed to the analysis of 3'end variability in neurodegenerative diseases; T.M.N., X.F. and S.C. contributed to the analysis and interpretation of transcript diversity in glioblastoma samples; D.A.K. and R.R supervised the project. J.S took lead in writing the manuscript with input from all authors. All authors read and approved the final manuscript.

Conflict of interest

R.R. is a founder of Genotwin and a member of the SAB of DiaTech and Flahy. None of these activities are related to the work described in this manuscript.

Data and code availability

This paper analyzes existing, publicly available data. Accession details can be found in the corresponding Methods sections. The SPLISOSM python package is available at https://github.com/JiayuSuPKU/SPLISOSM under BSD-3 License, and code to reproduce analyses included in this paper along with processed data and intermediate results will be made available at https://github.com/JiayuSuPKU/SPLISOSM_paper.

Methods

SPLISOSM Overview

SPLISOSM analyzes isoform-level spatial transcriptomics data through a two-step approach: first detecting spatial variability (SV), then identifying differential usage (DU) by associating isoform patterns with potential regulatory factors. The framework accommodates quantification results of both full-length isoforms from long-read sequencing and local transcript diversity events from short-read sequencing platforms including 10X Visium and Slide-seqV2.

Unlike gene-level analysis, SPLISOSM recognizes that multiple isoforms ($q > 1$) from the same gene are interdependent. Their relative proportions form a multivariate statistical variable across the spatial random field of n locations

$$\{X_i \in [0,1]^q, i = 1, \cdots, n\}.$$

Together with the 2D spatial coordinates $\{Y_i \in \mathbb{R}^2,\ i = 1, \cdots, n\}$ and covariates such as RNA binding protein expression $\{Z_i \in \mathbb{N},\ i = 1, \cdots, n\}$, we can formulate spatial variability detection and differential association analysis as multivariate independence tests $\mathrm{T}(X,\ Y)$ and $\mathrm{T}(X|Y,\ Z|Y)$. The spatial conditioning in DU tests is critical for eliminating spurious associations caused by spatial autocorrelated noise and hidden confounders.

Below we outline SPLISOSM's key components, with comprehensive methodology, theoretical analysis and implementation details available in the Supplementary Note.

Unconditional kernel independence test

A commonly used statistic for multivariate correlation analysis is the RV coefficient, defined as the normalized matrix norm of the q-by-2 cross-covariance matrix $C_{XY} = C_{ij} = \mathrm{Cov}\left(X^i, Y^j\right)$ between X and Y

$$\mathrm{s} \coloneqq |\mathrm{C_{XY}}|_{\mathrm{F}}^2 = \mathrm{tr}\left(\mathrm{C_{XY}C_{XY}^T}\right).$$

To capture complex, non-linear spatial patterns beyond linear correlation, SPLISOSM employs kernel transformations to map variables into reproducing kernel Hilbert spaces (RKHSs). In these potentially infinite-dimensional feature spaces, we measure statistical dependence using the Hilbert-Schmidt Independence Criterion (HSIC), defined as the Hilbert-Schmidt norm of the cross-covariance operator $C_{XY}: \mathcal{G} \to \mathcal{F}$ that connects the respective feature spaces $\mathcal{F}$ and $\mathcal{G}$ of X and Y,

$$\mathrm{HSIC(X, Y)} \coloneqq |C_{XY}|_{HS}^2.$$

Similar to the univariate gene expression SV test SPARK-X[34], we build calibrated, permutation-free testing procedures based on the empirical HSIC estimator,

$$\mathrm{T_{HSIC}}(X,\ Y) = \frac{1}{n}\mathrm{tr}(K_X \mathrm{H} L_Y \mathrm{H}).$$

Here, K_X and L_Y are the n-by-n isoform and spatial kernel matrices that specify feature maps for X and Y, and $\mathrm{H} = \mathrm{I} - \frac{1}{\mathrm{n}}11^T$ is the centering matrix. Under the null hypothesis of statistical independence, T_{HSIC} follows an asymptotic distribution of chi-square mixture, with weights derived from the eigenvalues $\{\lambda_i\}$ and $\{\mu_j\}$ of K_X and L_Y,

$$\mathrm{T_{HSIC}} \xrightarrow{\mathrm{d}} \sum_{i,j=1}^{n} \lambda_i\, \mu_j z_{ij}^2.$$

The statistical power of HSIC-based non-parametric test depends heavily on effective kernel design. To effectively profile spatial isoform patterns in sparse data, SPLISOSM introduces two technical innovations:

First, we prove that any low-rank approximation of the spatial kernel significantly reduces test power (Theorem 1, Supplementary Note), highlighting a key limitation of existing SV tests like SPARK-X. Based on this insight, we developed a new full-rank spatial kernel using the Intrinsic Conditional Autoregressive (ICAR) model[39],

$$L_Y = (D - \rho W)^{-1},$$

where W is the n-by-n binary adjacency weight matrix of the mutual k-nearest-neighbor spatial graph, $D = \mathrm{diag}\{\sum_j W_{ij}\}$ is the degree matrix, and $\rho \in [0,1)$ is the spatial autocorrelation coefficient ensuring invertibility.

In essence, this approach represents spatial patterns as signals on the spatial graph to be decomposed onto the eigenvectors of L_Y through Graph Fourier Transform (GFT)[78], allowing us to prioritize patterns by frequency in the graph spectral domain without manually selecting spatial bandwidths. Additionally, the sparse structure of D and W enhances computational efficiency. For large-scale datasets (e.g. hundreds of thousands of spots) where full-rank spatial kernels become computationally prohibitive, we developed GFT-based low-rank kernels that approximate the kernel inverse L_Y^{-1} through sparse eigen-decomposition, preserving biologically meaningful low-frequency patterns while limiting power loss to less important high-frequency signals.

Second, to handle sparse isoform data where many spots have zero gene coverage, we designed a Zero-Padded Centered (ZPC) kernel for isoform ratios X_n with $n - m$ undefined values,

$$K_{X_n} = \begin{pmatrix} H_m K_{X_m} H_m & 0 \\ 0 & 0 \end{pmatrix},$$

where X_m represents isoform ratios at m spots with non-zero gene coverage, and $K_{X_m} \coloneqq \phi(X_m)\phi(X_m)^T$ is the isoform kernel with transformation ϕ for compositional data (e.g. log-ratio transformations). Our empirical analysis showed that a simple linear kernel $k(x, y) = \langle x, y \rangle$ (with identity transformation) delivers optimal performance.

The ZPC version of a linear kernel is equivalent to a two-step procedure where the $n - m$ undefined ratios in X_n are first replaced with global averages per isoform before constructing the n-by-n isoform kernel for independence test. The NA replacement allows us to reuse the n-by-n spatial kernel across genes with varying sparsity patterns, substantially improving computational efficiency. We mathematically prove that this approach provides a bounded approximation to testing procedures that simply omit spots with undefined ratios (Theorem 2, Supplementary Note).

Detailed implementations of the ICAR kernel and SPLISOSM's three spatial variability tests --- HSIC-GC (gene expression), HSIC-IR (isoform usage ratio), and HSIC-IC (isoform expression) --- are available as Algorithms 1-4 in the Supplementary Note. The GFT-based spatial kernel construction is presented in Algorithm 5.

Conditional kernel independence test

Linking isoform usage X with potential regulators Z requires addressing a fundamental challenge: X and Z may appear related simply because both follow similar spatial patterns irrelevant to splicing regulation. To learn the causal relationship, we need to condition on spatial coordinates Y and all other confounders if available. However, ST data provide

only snapshots: at each location Y_i, we observed only one (X_i, Z_i) pair, making direct estimation of conditional distributions for $X|Y$ and $Z|Y$ impossible.

We address this using the conditional independence framework from Zhang et al[38], which shares information across observations by learning regression functions $f_X: Y \to X$ and $f_Z: Y \to Z$, then testing kernel independence of the residuals.

For linear ridge regression with L2 regularization λ, we can compute the residuals as

$$\mathrm{X} - \widehat{\mathrm{X}} = (\mathrm{I} - \mathrm{Y}(\mathrm{Y}^{\mathrm{T}}\mathrm{Y} + \lambda\mathrm{I})^{-1}\mathrm{Y}^{\mathrm{T}})\mathrm{X} = (\mathrm{I} - \mathrm{Y}\mathrm{Y}^{\mathrm{T}}(\mathrm{Y}\mathrm{Y}^{\mathrm{T}} + \lambda\mathrm{I})^{-1})\mathrm{X}$$

where $Y \in R^{n\times 2}$ represents spatial coordinates and $X \in R^{n\times \mathrm{d}}$ the spatial variable to predict. By substituting the linear kernel YY^T with general spatial kernels K_Y, we extend the solution to non-linear relationships

$$X - \hat{X} = \lambda(\mathrm{K_Y} + \lambda\mathrm{I})^{-1}X \coloneqq R_Y^X X.$$

This allows us to construct a kernel for the conditional variable $X|Y$ as $K_{X|Y} = R_Y^X K_X R_Y^X$ where $R_Y^X = \lambda(K_Y + \lambda I)^{-1}$ is learned from the data. The parameter λ controls conditioning strength: at $\lambda = 0$, X is fully explained by coordinates Y, leaving no information for $X|Y$; as $\lambda \to \infty$, the residuals $\mathrm{X} - \widehat{\mathrm{X}} \to \mathrm{X}$, which gives the unconditional test. Since genes have diverse spatial patterns, we estimate λ and K_Y for each variable X (and Z) individually using Gaussian Process Regression.

Using linear kernels K_X and K_Z, our conditional association test statistic becomes

$$\mathrm{T_{HSIC}}(X|Y, Z|Y) = \frac{1}{n}\mathrm{tr}\left(\left(R_Y^X \tilde{X}\right)^T \left(R_Y^Z \tilde{Z}\right)\right)$$

where $\tilde{X}$ *and* $\tilde{Z}$ are column-wise centered isoform usage ratios and covariates, and R_Y^X and R_Y^Z are the residual operators estimated from Gaussian Process regression.

Detailed implementations of the HSIC-based unconditional and conditional DU tests are provided as Algorithms 6 and 7 in the Supplementary Note. For parametric approaches, we can also introduce spatial conditioning through generalized linear mixed models (GLMMs) with spatial random effects. Specifically, we implemented the score test for differential usage where null GLMM models without covariates are first fitted to account for background spatial patterns. Covariate significance is then assessed by computing the gradient (i.e. score statistic) at zero effect size. Details on model fitting and hypothesis testing are available in Chapter 3 of the Supplementary Note.

General test configurations

For all isoform spatial variability tests, we used raw UMI counts from isoform quantification results (i.e. isoform counts in ONT data and TREND event counts in SR data). Input data varied by method: HSIC-GC and SPARK-X used summed isoform UMI counts (a scalar) per gene per spot, HSIC-IC used the vector of isoform counts per gene per spot, and HSIC-IR used observed isoform ratios. For spots with zero gene coverage (undefined ratios), we substituted with the global mean ratio (justified in Theorem 2).

For differential usage testing with RBP expression, we used log1p-normalized Visium RBP gene expression (library size = 1e-4) as covariates. We first identified spatially variably expressed RBPs (adjusted p-values < 0.01 in both HSIC-GC and SPARK-X tests) to include as covariates, excluding ribosomal proteins and non-canonical RBPs. SVE tests were conducted on Visium-based RBP gene raw counts. Only SVS genes with average total ONT/TREND counts >0.5 per spot were investigated for DU with RBP. For HSIC-based DU tests, we used observed isoform ratios with mean replacement for undefined values. Gaussian process regression was implemented using GaussianProcessRegressor from the scikit-learn[79] package with learnable RBF and white noise kernels. For parametric models (GLM/GLMM), we used isoform count vectors as inputs, grouping genes with identical isoform numbers to enable parallel processing. We calculated p-values using score tests on fitted null models under asymptotic chi-square distribution.

All SPLISOSM SV results in the main paper were computed with the full-rank ICAR kernel with $\rho = 0.99$ and the number of mutual nearest neighbors $k = 4$ for constructing the spatial graph. HSIC-based test p-values were calculated using the Liu's method[80] to approximate the chi-square mixture null distribution. Multiple testing correction was applied using the Benjamini-Hochberg method[81].

External datasets of RNA binding proteins

We obtained mouse and human RBP lists from EuRBPDB[82]. For motif analysis, we obtained RBP binding motifs from CISBP-RNA[83] and refined them using the R package universalmotif[84] to trim low-information peripheral positions before scanning with FIMO from the MEME suite[85] (p-value threshold = 1e-3, background model computed using the default fasta-get-markov). RBP CLIP data was downloaded from the POSTAR3[43] database and the ENCODE[86] portal (for PCBP2 human eCLIP). In our downstream analysis, we prioritized RBPs with validated binding motifs or CLIP evidence as potential splicing regulators. APA-related RBPs were collected based on regulators previously characterized in a high-throughput screening study[87].

Data simulation

We generated *in silico* spatial isoform quantification data using a hierarchical generative model following two steps: first sampling total gene expression and expected isoform usage per spot, then sampling observed isoform counts from spot-specific multinomial distributions.

For total gene expression, we sampled from Poisson distributions with spot-specific means. In Scenarios 1-3 (without spatial gene expression variability), these means were drawn from independent Gaussian distributions (μ=5, σ^2=1). In Scenarios 4-6 (with regional gene expression patterns), means were sampled from multivariate Gaussian distributions with binary covariate-controlled means and covariance matrices combining spatial autocorrelation (ICAR)[39] and white noise.

For expected isoform usage ratios, we sampled in the log-ratio space with $q-1$ degrees of freedom where $q=3$ is the number of isoforms per gene. In Scenarios 1 and 4 (without spatial isoform variability), log-ratios were sampled from independent Gaussian distributions (μ=0, σ^2=0.2) and transformed via softmax to obtain expected ratios. In Scenarios 2 and 5 (with Gaussian process variability), log-ratios were sampled from multivariate Gaussian distributions with ICAR-based spatial covariance. In Scenarios 3 and 6 (with regional isoform usage), log-ratios were determined by binary covariates with added ICAR-based spatial autocorrelated noise.

To evaluate statistical power, we created intermediate scenarios by interpolating between key configurations. For the SV test HSIC-IR, we varied the proportion of spatially autocorrelated noise between Scenarios 4 (regional expression, no isoform usage variability) and Scenarios 5 (regional expression, 50% spatially autocorrelated noise + 50% white noise in the Gaussian process). For DU tests, we adjusted covariate effect sizes between Scenarios 5 (regional expression, no isoform usage association) to Scenarios 6 (regional expression, covariate effect size=0.5).

SiT data of adult mouse brain

We downloaded processed spatial isoform transcriptomics (SiT)[24] isoform quantification results as well as Visium gene expression data of mouse olfactory bulb (MOB) and two coronal brain section samples (CBS1, CBS2) from Gene Expression Omnibus (GSE153859). The SiT ONT datasets contain only reference isoforms mapped to Mouse Gencode vM24. We removed isoforms expressed in <1% of spots and genes with fewer than 2 detected isoforms.

To reduce sparsity for visualization and clustering, we applied local smoothing by borrowing 20% of UMIs from six nearest neighbors per spot, then calculated isoform ratios with mean-padding for remaining undefined values. For spatial clustering, we first identified significant SVE and SVS genes using unsmoothed raw data, then performed dimensionality reduction (PCA, 50 components) and Leiden clustering on either smoothed gene expression or isoform ratios. RBP-based clustering used log1p-normalized Visium expression of SVE RBPs. Clustering resolutions were individually chosen such that the resulting cluster number were similar. For SVS usage program hierarchical clustering, we computed pairwise gene similarity using RV coefficients between multivariate smoothed isoform usage ratios.

For splicing type annotation, we used SUPPA[88] to split isoforms into local alternative splicing events. Skipping exons (SE), mutually exclusive exons (MX), and their flanking introns were analyzed for motif enrichment using MEME suite[85] tools (XSTREME and SEA), with additional exon sequence features extracted using Matt[89].

Visium and Slide-seqV2 data of adult mouse brain

We downloaded RNA-seq bam files of the 10X Visium data of a fresh frozen adult mouse coronal brain section sample from https://www.10xgenomics.com/datasets/adult-mouse-

brain-coronal-section-fresh-frozen-1-standard, and the Slide-seqV2[22] data of adult mouse hippocampus from https://singlecell.broadinstitute.org/single_cell/study/SCP815/sensitive-spatial-genome-wide-expression-profiling-at-cellular-resolution#study-summary.

Transcriptome 3'end diversity (TREND) was quantified using Sierra[90] with Mouse Gencode vM10 reference annotation, with minimal peak cutoffs in FindPeaks set to zero to recover minor events in the quantification stage. TREND events expressed in <1% of spots and genes with fewer than 2 TREND regions were later filtered out. Events were annotated using Sierra's AnnotatePeaksFromGTF with annotation_correction=FALSE to allow multiple annotations per region. Alternative exons (Alt-Exon) were defined as exons located within introns of other transcript variants. Motif enrichment analysis was performed on exons of TREND regions using MEME suite[85] tools (XSTREME and SEA).

For spatial variability tests on gene expression (HSIC-GC and SPARK-X), we used the sum of TREND reads per gene per spot rather than Visium's Space Ranger quantification (except for the RBP SVE test which we always used the latter) to maintain consistency with HSIC-IC/IR tests and ONT data analysis. Slide-seqV2 data was analyzed using HSIC-GC/IC/IR with full-rank spatial kernels. See Supplementary Note for low-rank test performance.

For spatial clustering, we use unsmoothed TREND-based gene expression and isoform usage from top 200 SVE and SVS genes for respective clusters, and Visium-based expression of all SVE RBPs for RBP clusters. We performed dimensionality reduction (PCA, 50 components) and Leiden clustering following the same approach as in our SiT ONT data analysis.

Single cell splicing quantification and RBFOX triple knockout data

We analyzed processed exon and splicing quantification data from the adult mouse neocortex single-cell atlas[91] using results from Moakley et al. 2024[92]. Cassette exon PSI was quantified at cell-type level using Quantas[5] by aggregating cells of identical types. Average RBP expression per cell type was log normalized and standardized to zero means.

For RBFOX triple knockout analysis, we used processed Quantas-based splicing quantification of mouse Day 10 motor neurons from Jacko et al. 2018[51]. Raw RNA-seq data was downloaded from the Sequence Read Archive (SRP128054) to run RBP map and positional motif enrichment analysis using Matt[89].

Visium data of human DLPFC

We downloaded Visium RNA-seq bam files of 12 human dorsolateral pre-frontal cortex (DLPFC) samples[62] through Globus (endpoint: jhpce#HumanPilot10x). Transcriptome 3'end diversity (TREND) was quantified using Sierra[90] with 10X's GRCh38-2024-A reference annotation and default parameters. To create a shared set of TREND events across samples, we merged per-sample peaks using MergePeakCoordinates with default

settings and quantified them with CountPeaks. We obtained conservation scores (phyloP100way and phastCons20way) from the UCSC Genome Browser and calculated average conservation scores specifically for exons within the TREND regions.

ONT and Visium data of human glioma samples

We analyzed two independent cohorts of human glioma samples using long-read and short-read spatial transcriptomics platforms.

For the Ren et al. 2023 cohort[66], we obtained processed ONT isoform quantification results, matched Visium gene expression data, and region annotations for 13 human glioma samples directly from the study authors, as the raw data was restricted by human genetic resources regulations. ONT transcripts were error-corrected, collapsed by shared exon-exon junctions, and quantified using DEMINERS[93] by the original study. For spatial analysis, we included only reference isoforms matching Gencode v32 (hg38) and novel isoforms from known genes. We further filtered out isoforms expressed in <5% of spots and genes with fewer than 2 detected isoforms. SUPPA[88] was used to categorize alternative isoform usage into local alternative splicing events.

For the Greenwald et al. 2024 cohort[65], we processed Visium data from 13 IDH wild-type glioblastoma samples downloaded from SRA (PRJNA994130). Raw RNA-seq reads were aligned to the human reference genome (hg38) using Space Ranger v3.0.1. Metaprograms and spatial annotations were obtained from the authors' GitHub repository (github.com/tiroshlab/Spatial_Glioma). Transcriptome 3'end diversity (TREND) was quantified using Sierra[90] with 10X's GRCh38-2024-A reference annotation and default parameters. Due to sample heterogeneity, we processed each sample independently with sample-specific TREND regions rather than merging across samples. For spatial analysis, we filtered out TREND events expressed in <1% of spots, genes with fewer than 2 TREND regions, and novel genes with symbols beginning with 'ENSG'.

Reference:

1. Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol *10*, 741-754. 10.1038/nrm2777.
2. Lee, Y., and Rio, D.C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu Rev Biochem *84*, 291-323. 10.1146/annurev-biochem-060614-034316.
3. Tian, B., and Manley, J.L. (2017). Alternative polyadenylation of mRNA precursors. Nat Rev Mol Cell Biol *18*, 18-30. 10.1038/nrm.2016.116.
4. Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470-476. 10.1038/nature07509.
5. Weyn-Vanhentenryck, S.M., Feng, H., Ustianenko, D., Duffie, R., Yan, Q., Jacko, M., Martinez, J.C., Goodwin, M., Zhang, X., Hengst, U., et al. (2018). Precise temporal regulation of alternative splicing during neural development. Nat Commun *9*, 2189. 10.1038/s41467-018-04559-0.
6. Zhang, X., Chen, M.H., Wu, X., Kodani, A., Fan, J., Doan, R., Ozawa, M., Ma, J., Yoshida, N., Reiter, J.F., et al. (2016). Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. Cell *166*, 1147-1162 e1115. 10.1016/j.cell.2016.07.025.
7. Ule, J., Ule, A., Spencer, J., Williams, A., Hu, J.S., Cline, M., Wang, H., Clark, T., Fraser, C., Ruggiu, M., et al. (2005). Nova regulates brain-specific splicing to shape the synapse. Nat Genet *37*, 844-852. 10.1038/ng1610.
8. Beffert, U., Weeber, E.J., Durudas, A., Qiu, S., Masiulis, I., Sweatt, J.D., Li, W.P., Adelmann, G., Frotscher, M., Hammer, R.E., and Herz, J. (2005). Modulation of synaptic plasticity and memory by Reelin involves differential splicing of the lipoprotein receptor Apoer2. Neuron *47*, 567-579. 10.1016/j.neuron.2005.07.007.
9. Sanes, J.R., and Zipursky, S.L. (2020). Synaptic Specificity, Recognition Molecules, and Assembly of Neural Circuits. Cell *181*, 1434-1435. 10.1016/j.cell.2020.05.046.
10. Wang, S., DeLeon, C., Sun, W., Quake, S.R., Roth, B.L., and Sudhof, T.C. (2024). Alternative splicing of latrophilin-3 controls synapse formation. Nature *626*, 128-135. 10.1038/s41586-023-06913-9.
11. Li, Y.I., van de Geijn, B., Raj, A., Knowles, D.A., Petti, A.A., Golan, D., Gilad, Y., and Pritchard, J.K. (2016). RNA splicing is a primary link between genetic variation and disease. Science *352*, 600-604. 10.1126/science.aad9417.
12. Walker, R.L., Ramaswami, G., Hartl, C., Mancuso, N., Gandal, M.J., De La Torre-Ubieta, L., Pasaniuc, B., Stein, J.L., and Geschwind, D.H. (2019). Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. Cell *179*, 750-+. 10.1016/j.cell.2019.09.021.
13. Gandal, M.J., Zhang, P., Hadjimichael, E., Walker, R.L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y., et al. (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science *362*. 10.1126/science.aat8127.

14. Raj, T., Li, Y.I., Wong, G., Humphrey, J., Wang, M., Ramdhani, S., Wang, Y.C., Ng, B., Gupta, I., Haroutunian, V., et al. (2018). Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. Nat Genet *50*, 1584-1592. 10.1038/s41588-018-0238-1.
15. Highley, J.R., Kirby, J., Jansweijer, J.A., Webb, P.S., Hewamadduma, C.A., Heath, P.R., Higginbottom, A., Raman, R., Ferraiuolo, L., Cooper-Knock, J., et al. (2014). Loss of nuclear TDP-43 in amyotrophic lateral sclerosis (ALS) causes altered expression of splicing machinery and widespread dysregulation of RNA splicing in motor neurones. Neuropathol Appl Neurobiol *40*, 670-685. 10.1111/nan.12148.
16. Takata, A., Matsumoto, N., and Kato, T. (2017). Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. Nat Commun *8*, 14519. 10.1038/ncomms14519.
17. Liu, Z., Yoshimi, A., Wang, J., Cho, H., Chun-Wei Lee, S., Ki, M., Bitner, L., Chu, T., Shah, H., Liu, B., et al. (2020). Mutations in the RNA Splicing Factor SF3B1 Promote Tumorigenesis through MYC Stabilization. Cancer Discov *10*, 806-821. 10.1158/2159-8290.CD-19-1330.
18. Zhang, Y., Qian, J., Gu, C., and Yang, Y. (2021). Alternative splicing and cancer: a systematic review. Signal Transduct Target Ther *6*, 78. 10.1038/s41392-021-00486-7.
19. Jbara, A., Lin, K.T., Stossel, C., Siegfried, Z., Shqerat, H., Amar-Schwartz, A., Elyada, E., Mogilevsky, M., Raitses-Gurevich, M., Johnson, J.L., et al. (2023). RBFOX2 modulates a metastatic signature of alternative splicing in pancreatic cancer. Nature *617*, 147-153. 10.1038/s41586-023-05820-3.
20. Stahl, P.L., Salmen, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science *353*, 78-82. 10.1126/science.aaf2403.
21. Liu, Y., Yang, M., Deng, Y., Su, G., Enninful, A., Guo, C.C., Tebaldi, T., Zhang, D., Kim, D., Bai, Z., et al. (2020). High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. Cell *183*, 1665-1681 e1618. 10.1016/j.cell.2020.10.026.
22. Stickels, R.R., Murray, E., Kumar, P., Li, J., Marshall, J.L., Di Bella, D.J., Arlotta, P., Macosko, E.Z., and Chen, F. (2021). Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. Nat Biotechnol *39*, 313-319. 10.1038/s41587-020-0739-1.
23. Joglekar, A., Prjibelski, A., Mahfouz, A., Collier, P., Lin, S., Schlusche, A.K., Marrocco, J., Williams, S.R., Haase, B., Hayes, A., et al. (2021). A spatially resolved brain region- and cell type-specific isoform atlas of the postnatal mouse brain. Nat Commun *12*, 463. 10.1038/s41467-020-20343-5.
24. Lebrigand, K., Bergenstrahle, J., Thrane, K., Mollbrink, A., Meletis, K., Barbry, P., Waldmann, R., and Lundeberg, J. (2023). The spatial landscape of gene expression isoforms in tissue sections. Nucleic Acids Res *51*, e47. 10.1093/nar/gkad169.
25. Ji, G., Tang, Q., Zhu, S., Zhu, J., Ye, P., Xia, S., and Wu, X. (2023). stAPAminer: Mining Spatial Patterns of Alternative Polyadenylation for Spatially Resolved Transcriptomic

Studies. Genomics Proteomics Bioinformatics *21*, 601-618. 10.1016/j.gpb.2023.01.003.

26. Olivieri, J., and Salzman, J. (2023). Analysis of RNA processing directly from spatial transcriptomics data reveals previously unknown regulation. bioRxiv. 10.1101/2023.03.13.532412.
27. Fu, Y., Kim, H., Adams, J.I., Grimes, S.M., Huang, S., Lau, B.T., Sathe, A., Hess, P., Ji, H.P., and Zhang, N.R. (2023). Single cell and spatial alternative splicing analysis with long read sequencing. Res Sq. 10.21203/rs.3.rs-2674892/v1.
28. Buen Abad Najar, C.F., Yosef, N., and Lareau, L.F. (2020). Coverage-dependent bias creates the appearance of binary splicing in single cells. Elife *9*. 10.7554/eLife.54603.
29. Isaev, K., and Knowles, D.A. (2024). Investigating RNA splicing as a source of cellular diversity using a binomial mixture model. (PMLR), pp. 163-175.
30. Rich, J.M., Moses, L., Einarsson, P.H., Jackson, K., Luebbert, L., Booeshaghi, A.S., Antonsson, S., Sullivan, D.K., Bray, N., and Melsted, P. (2024). The impact of package selection and versioning on single-cell RNA-seq analysis. bioRxiv.
31. Edsgard, D., Johnsson, P., and Sandberg, R. (2018). Identification of spatial expression trends in single-cell gene expression data. Nat Methods *15*, 339-342. 10.1038/nmeth.4634.
32. Svensson, V., Teichmann, S.A., and Stegle, O. (2018). SpatialDE: identification of spatially variable genes. Nat Methods *15*, 343-346. 10.1038/nmeth.4636.
33. Sun, S., Zhu, J., and Zhou, X. (2020). Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. Nat Methods *17*, 193-200. 10.1038/s41592-019-0701-7.
34. Zhu, J., Sun, S., and Zhou, X. (2021). SPARK-X: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. Genome Biol *22*, 184. 10.1186/s13059-021-02404-0.
35. Dries, R., Zhu, Q., Dong, R., Eng, C.L., Li, H., Liu, K., Fu, Y., Zhao, T., Sarkar, A., Bao, F., et al. (2021). Giotto: a toolbox for integrative analysis and visualization of spatial expression data. Genome Biol *22*, 78. 10.1186/s13059-021-02286-2.
36. Chen, C., Kim, H.J., and Yang, P. (2024). Evaluating spatially variable gene detection methods for spatial transcriptomics data. Genome Biol *25*, 18. 10.1186/s13059-023-03145-y.
37. Gretton, A., Fukumizu, K., Teo, C., Song, L., Schölkopf, B., and Smola, A. (2007). A kernel statistical test of independence. Advances in neural information processing systems *20*.
38. Zhang, K., Peters, J., Janzing, D., and Schölkopf, B. (2012). Kernel-based conditional independence test and application in causal discovery. arXiv preprint arXiv:1202.3775.
39. Su, J., Reynier, J.B., Fu, X., Zhong, G., Jiang, J., Escalante, R.S., Wang, Y., Aparicio, L., Izar, B., Knowles, D.A., and Rabadan, R. (2023). Smoother: a unified and modular framework for incorporating structural dependency in spatial omics data. Genome Biol *24*, 291. 10.1186/s13059-023-03138-x.

40. Booeshaghi, A.S., Yao, Z., van Velthoven, C., Smith, K., Tasic, B., Zeng, H., and Pachter, L. (2021). Isoform cell-type specificity in the mouse primary motor cortex. Nature *598*, 195-199. 10.1038/s41586-021-03969-3.
41. Joglekar, A., Hu, W., Zhang, B., Narykov, O., Diekhans, M., Marrocco, J., Balacco, J., Ndhlovu, L.C., Milner, T.A., Fedrigo, O., et al. (2024). Single-cell long-read sequencing-based mapping reveals specialized splicing patterns in developing and adult mouse and human brain. Nat Neurosci *27*, 1051-1063. 10.1038/s41593-024-01616-4.
42. Patowary, A., Zhang, P., Jops, C., Vuong, C.K., Ge, X., Hou, K., Kim, M., Gong, N., Margolis, M., Vo, D., et al. (2024). Developmental isoform diversity in the human neocortex informs neuropsychiatric risk mechanisms. Science *384*, eadh7688. 10.1126/science.adh7688.
43. Zhao, W., Zhang, S., Zhu, Y., Xi, X., Bao, P., Ma, Z., Kapral, T.H., Chen, S., Zagrovic, B., Yang, Y.T., and Lu, Z.J. (2022). POSTAR3: an updated platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins. Nucleic Acids Res *50*, D287-D294. 10.1093/nar/gkab702.
44. Olivieri, J.E., Dehghannasiri, R., Wang, P.L., Jang, S., de Morree, A., Tan, S.Y., Ming, J., Ruohao Wu, A., Tabula Sapiens, C., Quake, S.R., et al. (2021). RNA splicing programs define tissue compartments and cell types at single-cell resolution. Elife *10*. 10.7554/eLife.70692.
45. Wheeler, E.C., Vora, S., Mayer, D., Kotini, A.G., Olszewska, M., Park, S.S., Guccione, E., Teruya-Feldstein, J., Silverman, L., Sunahara, R.K., et al. (2022). Integrative RNA-omics Discovers GNAS Alternative Splicing as a Phenotypic Driver of Splicing Factor-Mutant Neoplasms. Cancer Discov *12*, 836-855. 10.1158/2159-8290.CD-21-0508.
46. Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J.O., and Lai, E.C. (2013). Widespread and extensive lengthening of 3' UTRs in the mammalian brain. Genome Res *23*, 812-825. 10.1101/gr.146886.112.
47. Ogorodnikov, A., Levin, M., Tattikota, S., Tokalov, S., Hoque, M., Scherzinger, D., Marini, F., Poetsch, A., Binder, H., Macher-Goppinger, S., et al. (2018). Transcriptome 3'end organization by PCF11 links alternative polyadenylation to formation and neuronal differentiation of neuroblastoma. Nat Commun *9*, 5331. 10.1038/s41467-018-07580-5.
48. Tian, B., Pan, Z., and Lee, J.Y. (2007). Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. Genome Res *17*, 156-165. 10.1101/gr.5532707.
49. Hwang, H.W., Saito, Y., Park, C.Y., Blachere, N.E., Tajima, Y., Fak, J.J., Zucker-Scharff, I., and Darnell, R.B. (2017). cTag-PAPERCLIP Reveals Alternative Polyadenylation Promotes Cell-Type Specific Protein Diversity and Shifts Araf Isoforms with Microglia Activation. Neuron *95*, 1334-1349 e1335. 10.1016/j.neuron.2017.08.024.
50. Fisher, E., and Feng, J. (2022). RNA splicing regulators play critical roles in neurogenesis. Wiley Interdiscip Rev RNA *13*, e1728. 10.1002/wrna.1728.
51. Jacko, M., Weyn-Vanhentenryck, S.M., Smerdon, J.W., Yan, R., Feng, H., Williams, D.J., Pai, J., Xu, K., Wichterle, H., and Zhang, C. (2018). Rbfox Splicing Factors Promote

Neuronal Maturation and Axon Initial Segment Assembly. Neuron *97*, 853-868 e856. 10.1016/j.neuron.2018.01.020.

52. Markovic, D., and Challiss, R.A. (2009). Alternative splicing of G protein-coupled receptors: physiology and pathophysiology. Cell Mol Life Sci *66*, 3337-3352. 10.1007/s00018-009-0093-4.
53. Sanchez-Reyes, O.B., Zilberg, G., McCorvy, J.D., and Wacker, D. (2023). Molecular insights into GPCR mechanisms for drugs of abuse. J Biol Chem *299*, 105176. 10.1016/j.jbc.2023.105176.
54. Rehfeld, F., Maticzka, D., Grosser, S., Knauff, P., Eravci, M., Vida, I., Backofen, R., and Wulczyn, F.G. (2018). The RNA-binding protein ARPP21 controls dendritic branching by functionally opposing the miRNA it hosts. Nat Commun *9*, 1235. 10.1038/s41467-018-03681-3.
55. Chatrikhi, R., Mallory, M.J., Gazzara, M.R., Agosto, L.M., Zhu, W.S., Litterman, A.J., Ansel, K.M., and Lynch, K.W. (2019). RNA Binding Protein CELF2 Regulates Signal-Induced Alternative Polyadenylation by Competing with Enhancers of the Polyadenylation Machinery. Cell Rep *28*, 2795-2806 e2793. 10.1016/j.celrep.2019.08.022.
56. Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.Y., Cody, N.A.L., Dominguez, D., et al. (2020). A large-scale binding and functional map of human RNA-binding proteins. Nature *583*, 711-719. 10.1038/s41586-020-2077-3.
57. Greenland, S., Pearl, J., and Robins, J.M. (1999). Confounding and collapsibility in causal inference. Statistical science *14*, 29-46.
58. Anglada-Girotto, M., Moakley, D.F., Zhang, C., Miravet-Verde, S., Califano, A., and Serrano, L. (2025). Exon inclusion signatures enable accurate estimation of splicing factor activity. bioRxiv. 10.1101/2024.06.21.600051.
59. Gazzara, M.R., Mallory, M.J., Roytenberg, R., Lindberg, J.P., Jha, A., Lynch, K.W., and Barash, Y. (2017). Ancient antagonism between CELF and RBFOX families tunes mRNA splicing outcomes. Genome Res *27*, 1360-1370. 10.1101/gr.220517.117.
60. Berto, S., Usui, N., Konopka, G., and Fogel, B.L. (2016). ELAVL2-regulated transcriptional and splicing networks in human neurons link neurodevelopment and autism. Hum Mol Genet *25*, 2451-2464. 10.1093/hmg/ddw110.
61. McGurk, M.P., McWatters, D.C., and Burge, C.B. (2024). KATMAP: Inferring splicing factor activity and regulatory targets from knockdown data. bioRxiv, 2024.2006. 2025.600605.
62. Maynard, K.R., Collado-Torres, L., Weber, L.M., Uytingco, C., Barry, B.K., Williams, S.R., Catallini, J.L., 2nd, Tran, M.N., Besich, Z., Tippani, M., et al. (2021). Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat Neurosci *24*, 425-436. 10.1038/s41593-020-00787-0.
63. Ito, H., Atsuzawa, K., Morishita, R., Usuda, N., Sudo, K., Iwamoto, I., Mizutani, K., Katoh-Semba, R., Nozawa, Y., Asano, T., and Nagata, K. (2009). Sept8 controls the binding of vesicle-associated membrane protein 2 to synaptophysin. J Neurochem *108*, 867-880. 10.1111/j.1471-4159.2008.05849.x.

64. Yuan, Y., Xie, S., Darnell, J.C., Darnell, A.J., Saito, Y., Phatnani, H., Murphy, E.A., Zhang, C., Maniatis, T., and Darnell, R.B. (2018). Cell type-specific CLIP reveals that NOVA regulates cytoskeleton interactions in motoneurons. Genome Biol *19*, 117. 10.1186/s13059-018-1493-2.
65. Greenwald, A.C., Darnell, N.G., Hoefflin, R., Simkin, D., Mount, C.W., Gonzalez Castro, L.N., Harnik, Y., Dumont, S., Hirsch, D., Nomura, M., et al. (2024). Integrative spatial analysis reveals a multi-layered organization of glioblastoma. Cell *187*, 2485-2501 e2426. 10.1016/j.cell.2024.03.029.
66. Ren, Y., Huang, Z., Zhou, L., Xiao, P., Song, J., He, P., Xie, C., Zhou, R., Li, M., Dong, X., et al. (2023). Spatial transcriptomics reveals niche-specific enrichment and vulnerabilities of radial glial stem-like cells in malignant gliomas. Nat Commun *14*, 1028. 10.1038/s41467-023-36707-6.
67. Vandiedonck, C., Taylor, M.S., Lockstone, H.E., Plant, K., Taylor, J.M., Durrant, C., Broxholme, J., Fairfax, B.P., and Knight, J.C. (2011). Pervasive haplotypic variation in the spliceo-transcriptome of the human major histocompatibility complex. Genome Res *21*, 1042-1054. 10.1101/gr.116681.110.
68. Rogalska, M.E., Mancini, E., Bonnal, S., Gohr, A., Dunyak, B.M., Arecco, N., Smith, P.G., Vaillancourt, F.H., and Valcarcel, J. (2024). Transcriptome-wide splicing network reveals specialized regulatory functions of the core spliceosome. Science *386*, 551-560. 10.1126/science.adn8105.
69. Fuentes-Fayos, A.C., Vazquez-Borrego, M.C., Jimenez-Vacas, J.M., Bejarano, L., Pedraza-Arevalo, S., F, L.L., Blanco-Acevedo, C., Sanchez-Sanchez, R., Reyes, O., Ventura, S., et al. (2020). Splicing machinery dysregulation drives glioblastoma development/aggressiveness: oncogenic role of SRSF3. Brain *143*, 3273-3293. 10.1093/brain/awaa273.
70. Kahles, A., Lehmann, K.V., Toussaint, N.C., Huser, M., Stark, S.G., Sachsenberg, T., Stegle, O., Kohlbacher, O., Sander, C., Cancer Genome Atlas Research, N., and Ratsch, G. (2018). Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. Cancer Cell *34*, 211-224 e216. 10.1016/j.ccell.2018.07.001.
71. Song, X., Tiek, D., Lu, M., Yu, X., Wu, R., Walker, M., He, Q., Sisbarro, D., Hu, B., and Cheng, S.-Y. (2025). A Single-Cell Atlas of RNA Alternative Splicing in the Glioma-Immune Ecosystem. bioRxiv, 2025.2003. 2026.645511.
72. Stassen, O., van Bodegraven, E.J., Giuliani, F., Moeton, M., Kanski, R., Sluijs, J.A., van Strien, M.E., Kamphuis, W., Robe, P.A.J., and Hol, E.M. (2017). GFAPdelta/GFAPalpha ratio directs astrocytoma gene expression towards a more malignant profile. Oncotarget *8*, 88104-88121. 10.18632/oncotarget.21540.
73. Uceda-Castro, R., van Asperen, J.V., Vennin, C., Sluijs, J.A., van Bodegraven, E.J., Margarido, A.S., Robe, P.A.J., van Rheenen, J., and Hol, E.M. (2022). GFAP splice variants fine-tune glioma cell invasion and tumour dynamics by modulating migration persistence. Sci Rep *12*, 424. 10.1038/s41598-021-04127-5.
74. Larionova, T.D., Bastola, S., Aksinina, T.E., Anufrieva, K.S., Wang, J., Shender, V.O., Andreev, D.E., Kovalenko, T.F., Arapidi, G.P., Shnaider, P.V., et al. (2022). Alternative RNA splicing modulates ribosomal composition and determines the spatial

phenotype of glioblastoma cells. Nat Cell Biol *24*, 1541-1557. 10.1038/s41556-022-00994-w.
75. Fuhrmann, D.C., Mondorf, A., Beifuss, J., Jung, M., and Brune, B. (2020). Hypoxia inhibits ferritinophagy, increases mitochondrial ferritin, and protects from ferroptosis. Redox Biol *36*, 101670. 10.1016/j.redox.2020.101670.
76. Hauser, A.S., Chavali, S., Masuho, I., Jahn, L.J., Martemyanov, K.A., Gloriam, D.E., and Babu, M.M. (2018). Pharmacogenomics of GPCR Drug Targets. Cell *172*, 41-54 e19. 10.1016/j.cell.2017.11.033.
77. Iijima, T., Hidaka, C., and Iijima, Y. (2016). Spatio-temporal regulations and functions of neuronal alternative RNA splicing in developing and adult brains. Neurosci Res *109*, 1-8. 10.1016/j.neures.2016.01.010.
78. Chang, Y., Liu, J., Jiang, Y., Ma, A., Yeo, Y.Y., Guo, Q., McNutt, M., Krull, J.E., Rodig, S.J., Barouch, D.H., et al. (2024). Graph Fourier transform for spatial omics representation and analyses of complex organs. Nat Commun *15*, 7467. 10.1038/s41467-024-51590-5.
79. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research *12*, 2825-2830.
80. Liu, H., Tang, Y., and Zhang, H.H. (2009). A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. Computational Statistics & Data Analysis *53*, 853-856.
81. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J R Stat Soc B *57*, 289-300. DOI 10.1111/j.2517-6161.1995.tb02031.x.
82. Liao, J.Y., Yang, B., Zhang, Y.C., Wang, X.J., Ye, Y., Peng, J.W., Yang, Z.Z., He, J.H., Zhang, Y., Hu, K., et al. (2020). EuRBPDB: a comprehensive resource for annotation, functional and oncological investigation of eukaryotic RNA binding proteins (RBPs). Nucleic Acids Res *48*, D307-D313. 10.1093/nar/gkz823.
83. Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. Nature *499*, 172-177. 10.1038/nature12311.
84. Tremblay, B.J.-M. (2024). universalmotif: An R package for biological motif analysis. Journal of Open Source Software *9*, 7012.
85. Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME Suite. Nucleic Acids Res *43*, W39-49. 10.1093/nar/gkv416.
86. Luo, Y., Hitz, B.C., Gabdank, I., Hilton, J.A., Kagda, M.S., Lam, B., Myers, Z., Sud, P., Jou, J., Lin, K., et al. (2020). New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic Acids Res *48*, D882-D889. 10.1093/nar/gkz1062.
87. Kowalski, M.H., Wessels, H.H., Linder, J., Dalgarno, C., Mascio, I., Choudhary, S., Hartman, A., Hao, Y., Kundaje, A., and Satija, R. (2024). Multiplexed single-cell characterization of alternative polyadenylation regulators. Cell *187*, 4408-4425 e4423. 10.1016/j.cell.2024.06.005.
88. Trincado, J.L., Entizne, J.C., Hysenaj, G., Singh, B., Skalic, M., Elliott, D.J., and Eyras, E. (2018). SUPPA2: fast, accurate, and uncertainty-aware differential splicing

analysis across multiple conditions. Genome Biol *19*, 40. 10.1186/s13059-018-1417-1.

89. Gohr, A., and Irimia, M. (2019). Matt: Unix tools for alternative splicing analysis. Bioinformatics *35*, 130-132. 10.1093/bioinformatics/bty606.
90. Patrick, R., Humphreys, D.T., Janbandhu, V., Oshlack, A., Ho, J.W.K., Harvey, R.P., and Lo, K.K. (2020). Sierra: discovery of differential transcript usage from polyA-captured single-cell RNA-seq data. Genome Biol *21*, 167. 10.1186/s13059-020-02071-7.
91. Tasic, B., Yao, Z., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., et al. (2018). Shared and distinct transcriptomic cell types across neocortical areas. Nature *563*, 72-78. 10.1038/s41586-018-0654-5.
92. Moakley, D.F., Campbell, M., Anglada-Girotto, M., Feng, H., Califano, A., Au, E., and Zhang, C. (2024). Reverse engineering neuron type-specific and type-orthogonal splicing-regulatory networks using single-cell transcriptomes. bioRxiv. 10.1101/2024.06.13.597128.
93. Song, J., Lin, L.A., Tang, C., Chen, C., Yang, Q., Zhang, D., Zhao, Y., Wei, H.C., Linghu, K., Xu, Z., et al. (2025). DEMINERS enables clinical metagenomics and comparative transcriptomic analysis by increasing throughput and accuracy of nanopore direct RNA sequencing. Genome Biol *26*, 76. 10.1186/s13059-025-03536-3.

Main Figure Legends

Figure 1: A computational toolbox for spatial isoform pattern discovery

Alternative pre-mRNA processing generates multiple isoforms from the same gene, with final isoform expression determined by both gene expression (controlled by transcriptional factors) and isoform usage (regulated by splicing machinery and RBPs). SPLISOSM addresses challenges in spatial isoform pattern analysis by recasting spatial variability (SV) and differential isoform usage (DU) detections as multivariate independence tests. For SV testing, multivariate isoform usage and spatial coordinates are projected into respective feature spaces to assess their nonlinear dependencies using the Hilbert-Schmidt Independence Criterion (HSIC) with optimized spatial and isoform kernels. For DU testing, we implement a nonparametric conditional HSIC test and a parametric generalized linear mixed models (GLMM) with spatial random effects to control for confounding spatial autocorrelation and reduce spurious associations.

Figure 2: SPLISOSM produces well-calibrated and permutation-free p-values in simulation

(A) Six simulation scenarios with spatial variability at different layers. Regional variability was generated from two binary and two continuous spatial covariates.

(B) Q-Q plots of SV test p-values across 1,000 simulated genes per scenario. Regional gene expression (Scenarios 4-6) was included to introduce artifacts in the observed ratios.

(C) Q-Q plots of differential usage test p-values aggregated from 2,000 tests (1,000 genes x 2 covariates) per covariate type. Variable isoform usage independent of covariates (Scenarios 2 and 5) was sampled from a Gaussian process (GP) to mimic spurious differential associations. For binary covariates, T-tests were performed on individual isoforms with p-values combined at gene-level using Fisher's method. For GLM and GLMM, null models with zero effect size were fitted and tested using the score statistic.

Figure 3: Integrative analysis reveals spatial alternative splicing programs in adult mouse brain enriched for synaptic and membrane trafficking functions

(A) Workflow overview.

(B) Comparison of spatial variability in gene expression (x-axis) versus isoform usage (y-axis) in the CBS2 sample. Points are colored by expression perplexity (effective number of isoforms per gene), calculated as the exponential of global isoform usage entropy. Inset shows number of events significant in both CBS1 and CBS2 replicates (adjusted p-value < 0.05), categorized as spatially variably spliced (SVS) or spatially variably expressed but not spliced (SVENS) genes.

(C) Pathway enrichment analysis comparing SVS versus SVENS genes, with terms arranged by precision differences between the two groups.

(D) Unsupervised clustering based on gene-level expression of SVE genes (left) or relative isoform ratios of SVS genes (right). Isoform counts were locally smoothed to compensate for data sparsity.

(E) Selected isoform usage programs identified through hierarchical clustering of SVS genes.

(F) Distributions of alternative splicing event types among SVS and SVE genes.

(G) Selected Matt exon features of skipping exons (SE) and mutually exclusive exons (MX) in SVS and SVENS genes.

(H) Unsupervised clustering based on SVE RBP expression from short-read sequencing data.

(I) Top RBPs ranked by number of significantly associated targets, color-coded by availability of documented motifs (CISBP-RNA) or CLIP data (POSTAR3). Log-normalized expression of selected RBPs shown at right.

(J) Differential isoform usage p-values for Rbfox3 (left) and Celf5 (right) in two replicates. Validated targets with RBP binding near regulated events in brain-CLIP data are highlighted in red.

(K) *Gnas* transcript structure and spatial log-normalized expression patterns.

Figure 4: Spatial 3'end transcript diversity in adult mouse brain extends beyond alternative polyadenylation and shows functional convergence on signaling pathways

(A) Workflow overview.

(B) Comparison of spatial variability in gene expression (x-axis) versus TREND usage (y-axis) in the SR coronal brain section sample. Points are colored by expression perplexity (effective number of isoforms per gene), calculated as the exponential of global isoform usage entropy. Inset quantifies significant events (adjusted p-value < 0.01).

(C) TREND event annotation and motif enrichment results (bottom). A TREND event may overlap with multiple genomic features. Alternative exons (Alt-Exon) are defined as exons located within introns of other transcript variants.

(D) Pathway enrichment analysis comparing SVS and SVENS genes.

(E) Spatial log-normalized expression of the neuronal and glial *Map4* variants.

(F) Unsupervised clustering based on gene-level expression of top 200 SVE genes (left), relative TREND ratios of top 200 SVS genes (middle), and gene expression of all SVE RBPs (right).

(G) *Gnao1* transcript structure and spatial log-normalized expression patterns. The resulting protein sequence variation, color-coded by solvent-accessible surface area, is shown on the right.

(H) *Arpp21* transcript structure and spatial log-normalized expression patterns.

(I) Top RBPs associated with TREND usage in *Arpp21*, *Celf2* and *Pcbp2*.

Figure 5: RBFOX regulates neural transcript usage with other splicing factors cooperatively

(A) *Clta* transcript structure and spatial log-normalized expression in the ONT-CBS2 sample.

(B) Relationship between normalized *Rbfox3* expression and *Clta* isoform ratio (left) or cell-type-aggregated exon PSI values (right). Spatial data was locally smoothed before calculating isoform ratios for visualization.

(C) Identification of Rbfox3 co-regulators in ONT (left) and SR (right) samples. Heatmaps display pairwise correlations of gene expression (upper triangle) or conditional HSIC test p-values (lower triangle). RBPs with reported RBFOX interaction are highlighted.

(D) Motif enrichment in Rbfox1/2/3 associated SVS targets using alternative exons and 150bp flanking introns (left) or TREND regions (right).

(E) Motif enrichment in exons with >=25% PSI changes following RBFOX triple knockout in Day 10 motor neurons, categorized as silenced (increased PSI after KD) or enhanced (decreased PSI). Sequences were binned by distance to exon boundaries (x-axis) to show positional enrichment.

Figure 6: Evolutionarily conserved transcript diversity in synaptic genes across the human prefrontal cortex

(A) Workflow overview.

(B) Relationship between total read coverage in TREND regions per spot (x-axis) and number of significant spatially variably spliced (SVS, top) or variably expressed (SVE, bottom) genes across datasets.

(C) P-value correlation distributions between technical replicates (left), different sections from same donor (middle), and different donors (right). Genes are stratified by average TREND read counts per spot.

(D) SVS genes ranked by recurrence (number of significant samples, x-axis) and the minimal HSIC-IR p-values across samples (y-axis). Highlighted genes have mouse homologs that are also spatially variably spliced.

(E) Distributions of average conservation scores in TREND regions across gene categories: non-variable controls, human-specific SVENS/SVS, and human-mouse shared SVENS/SVS.

(F) Pathway enrichment analysis comparing SVS versus SVENS genes conserved between human and mouse.

(G) RBP-SVS associations ranked by recurrence (x-axis) and minimal GLMM p-values across samples (y-axis). Top potential regulators of *SEPTIN8* are highlighted and colored by whether the association is conserved in mouse.

(H) *SEPTIN8* transcript structure and spatial log-normalized expression in mouse (top) and human (bottom, sample 151673).

(I) Comparative illustration of SEPTIN8-long and SEPTIN8-short protein variants.

Figure 7: Spatial transcriptomic diversity in human glioma is shaped by microenvironment composition and immune infiltration

(A) Workflow overview.

(B) Spatially variably processed (SVP) genes ranked by recurrence (x-axis) and the minimal HSIC-IR p-values across samples (y-axis) in ONT (left) and SR (right) cohorts. Genes not variable in healthy DLPFC are highlighted as disease specific.

(C) Distribution of recurrent and glioma-specific SVP genes.

(D) Pathway enrichment analysis comparing recurrent SVP (>=2 samples) genes versus recurrent SVENP (>=4 samples) genes.

(E) Spatial distribution of *HLA-DRB1* isoforms in the ONT sample GBM1.

(F) *B2M* and *CD74* isoform expression in GBM1, comparing tumor versus immune-infiltrated regions. Three novel *CD74* intron-retention isoforms are aggregated as one for visualization.

(G) *TPM3* transcript structure, RNA-seq coverage, and spatial patterns in the SR sample ZH916bulk. Event counts were locally smoothed for ratio visualization, same in (L) and (M).

(H) Distribution of glioblastoma metaprograms in regions with high/low TPM3-Event1 usage. Shown on the right are log-normalized expression of selected markers: SNAP25 (neuron), *ACTB* (glial), *IGKC* (B cell), *NDRG1* (hypoxia).

(I) *FTH1* transcript structure in the ONT sample DMG3 and its RNA-seq coverage in the SR sample ZH8811Abulk.

(J) Hypoxia marker expression (left) and correlation with FTH1-211 usage in DMG3.

(K) Histology (left), metaprogram distribution (middle) and hemoglobin expression (right) in regions with high/low FTH1-Event2 usage in ZH8811Abulk.

(L) Spatial distribution of *FTH1* isoforms in DMG3.

(M) Spatial distribution of *FTH1* TREND usage in ZH8811Abulk.

Supplementary Figure Legends

Figure S1: Power analysis for spatial variability and differential usage tests, related to Figure 2

(A) Expected and observed isoform ratios in simulations with varying degree of spatial variability, from Scenario 4 (no spatial variance) to Scenario 5 (50% spatial variance) in Figure 2.

(B) Q-Q plots of HSIC-IR p-values with different ratio transformations and undefined ratio replacement strategies. None means spots with NA ratios are removed and ignored for SV testing.

(C) Expected and observed isoform ratios in simulations with varying degree of covariate influence, from Scenario 5 (no association) to Scenario 6 (strong association) in Figure 2.

(D) Q-Q plots of non-parametric DU test p-values. Per-isoform T-test p-values were combined using either Fisher's or Tippett's methods. Conditional HSIC with fixed spatial or KNN kernels are included as approximations of the learnable Gaussian process kernel (default).

(E) Q-Q plots of parametric DU test p-values.

Figure S2: SPLISOSM identifies spatially variable splicing programs in a sparse mouse olfactory bulb dataset, related to Figure 3

(A) Sequencing depth and data sparsity of the SiT-ONT-MOB sample. Red lines indicate the median per-spot coverage (middle) and zero-coverage spots (right) across genes. The ONT-QC preprocessing step filters out isoforms expressed in less than 1% of spots and genes with less than 2 isoforms. Genes passed the filter have median per-spot total UMI of 0.25 in ONT data and 1 in SR data.

(B) Comparison of spatial variability in gene expression (x-axis) versus isoform usage (y-axis) in the MOB sample. Points are colored by expression perplexity (effective number of isoforms per gene), calculated as the exponential of global isoform usage entropy. Inset shows number of significant events (p-value < 0.05), categorized as spatially variably spliced (SVS, HSIC-IR p-value < 0.05) or spatially variably expressed but not spliced (SVENS, HSIC-GC p-value < 0.05 & HSIC-IR p-value > 0.05) genes.

(C) Sequencing depth and perplexity of SVS and SVE genes.

(D) Pathway enrichment analysis comparing SVS versus SVENS genes, with terms arranged by precision differences between the two groups.

(E) Distributions of alternative splicing event types among SVS and SVE genes.

(F) Selected Matt exon features of skipping exons (SE) and mutually exclusive exons (MX) in SVS and SVENS genes.

(G) Motif enrichment in SVS alternative exons (left) and exons with 300bp upstream/downstream flanking introns (right).

(H) Unsupervised clustering based on gene-level expression of SVE genes (middle) or relative isoform ratios of SVS genes (right). Isoform counts were locally smoothed to compensate for data sparsity.

(I) Top RBPs ranked by number of significantly associated targets, color-coded by availability of documented motifs (CISBP-RNA) or CLIP data (POSTAR3). Differential isoform usage p-values for selected RBPs are shown on the right, with validated. Validated RBP-SVS pairs with RBP binding near regulated events in CLIP data are highlighted in red.

(J) Myl6 transcript structure (left), smoothed spatial isoform usage ratios (middle), and its relationship with Rbfox3 log-normalized expression (right).

(K) Cltb transcript structure (left), smoothed spatial isoform usage ratios (middle), and its relationship with Rbfox3 log-normalized expression (right).

Figure S3: Spatially variable TREND events and their regulation in adult mouse brain, related to Figure 4

(A) HSIC-IR test agreements between Visium CBS (y-axis) and Slide-seqV2 hippocampus (x-axis) datasets. Points are colored by expression perplexity.

(B) Differential protein-altering alternative polyadenylation (APA) events between neurons and glial cells. Related to Figure 4E.

(C) Examples of additional protein-altering TREND events and their spatial log-normalized expression.

(D) Top RBPs ranked by number of significantly associated targets, color-coded by available motifs (CISBP-RNA) or CLIP data (POSTAR3), core APA machinery involvement and motif enrichment near SVS TREND events. Spatial log-normalized expression of selected RBPs are shown on the right.

(E) PCBP2 transcript structure in mouse and human, along with eCLIP data from ENCODE.

(F) PCBP2 isoform expression in the spatial mouse CBS data (top) and across GTEx human tissues (bottom). The negative association between PCBP2 gene expression and usage of the upstream exon is visualized on the right.

Figure S4: Discrepancy between knockout and observational data indicates cooperativity in RBFOX regulation, related to Figure 5

(A) Expression changes of selected RBP genes after RBFOX triple knockout where Rbfox genes were disrupted by in-frame stop codons.

(B) Heatmaps displaying DU test p-values between selected RBPs and target SVS genes associated with multiple RBPs. Only significant pairs (p-value < 0.01) are colored.

(C) Examples of two additional SVS genes whose observed isoform usage association with Rbfox3 contradicts the triple knockout data.

(D) Examples of two additional SVS genes whose observed last exon usage association with Rbfox3 contradicts the triple knockout data.

Figure S5: Additional examples of evolutionarily conserved SVS events in human DLPFC, related to Figure 6

(A) Relationship between total read coverage in TREND regions per spot (x-axis) and proportion of significant spatially variably spliced (SVS, top) or variably expressed (SVE, bottom) genes in all tested genes. Related to Figure 6B.

(B) Number and proportion of human DLPFC SV genes that have spatially variable mouse homologs, grouped by recurrence (number of significant DLPFC samples, x-axis).

(C) Sequencing depth of SVS and SVE genes, grouped by conservation and recurrence.

(D) Pathway enrichment analysis comparing human SVS versus human SVENS genes. Related to Figure 6F.

(E) Spatial log-normalized expression of selected RBPs in sample 151673.

(F) RBP-SVS associations ranked by recurrence (x-axis) and minimal GLMM p-values across samples (y-axis). Top potential regulators of MAP4 are highlighted and colored by whether the association is conserved in mouse.

(G) MAP4 transcript structure and spatial log-normalized expression in human sample 151673.

Figure S6: Additional examples of transcript diversity in human glioma, related to Figure 7

(A) Distribution of recurrent SVP genes (left) and relationships between total isoform read coverage per spot (x-axis) and number of SVS and SVE genes (y-axis) across ONT datasets.

(B) Distribution of recurrent SVP genes (left) and relationships between total TREND read coverage per spot (x-axis) and number of SVS and SVE genes (y-axis) across SR datasets.

(C) IG gene isoform diversity in the ONT sample DMG2.

(D) Distributions of AS types in ONT-SV genes and TREND annotations in SR-SV genes.

(E) Pathway enrichment analysis comparing recurrent SVP versus recurrent SVENP genes. Related to Figure 7D.

(F-I) SVP genes involved in selected KEGG pathways.

(J) GFAP transcript structure in the ONT sample DMG2, read coverage in the SR sample ZH916bulk, and the respective isoform or TREND spatial distribution in each sample (right).

Figure S7: Spatial variable usage of reference and novel ribosomal transcripts in human glioma, related to Figure 7

(A) RPS8 transcript structure in the ONT sample GBM3, read coverage in the SR sample ZH8811Abulk and the SR healthy DLPFC sample 151675, and the respective isoform or TREND spatial distribution in each sample. Event counts were locally smoothed for ratio visualization, same in (B).

(B) RPL5 transcript structure in the ONT sample GBM3, read coverage in the SR sample ZH8811Abulk and, and the respective isoform or TREND spatial distribution in each sample.

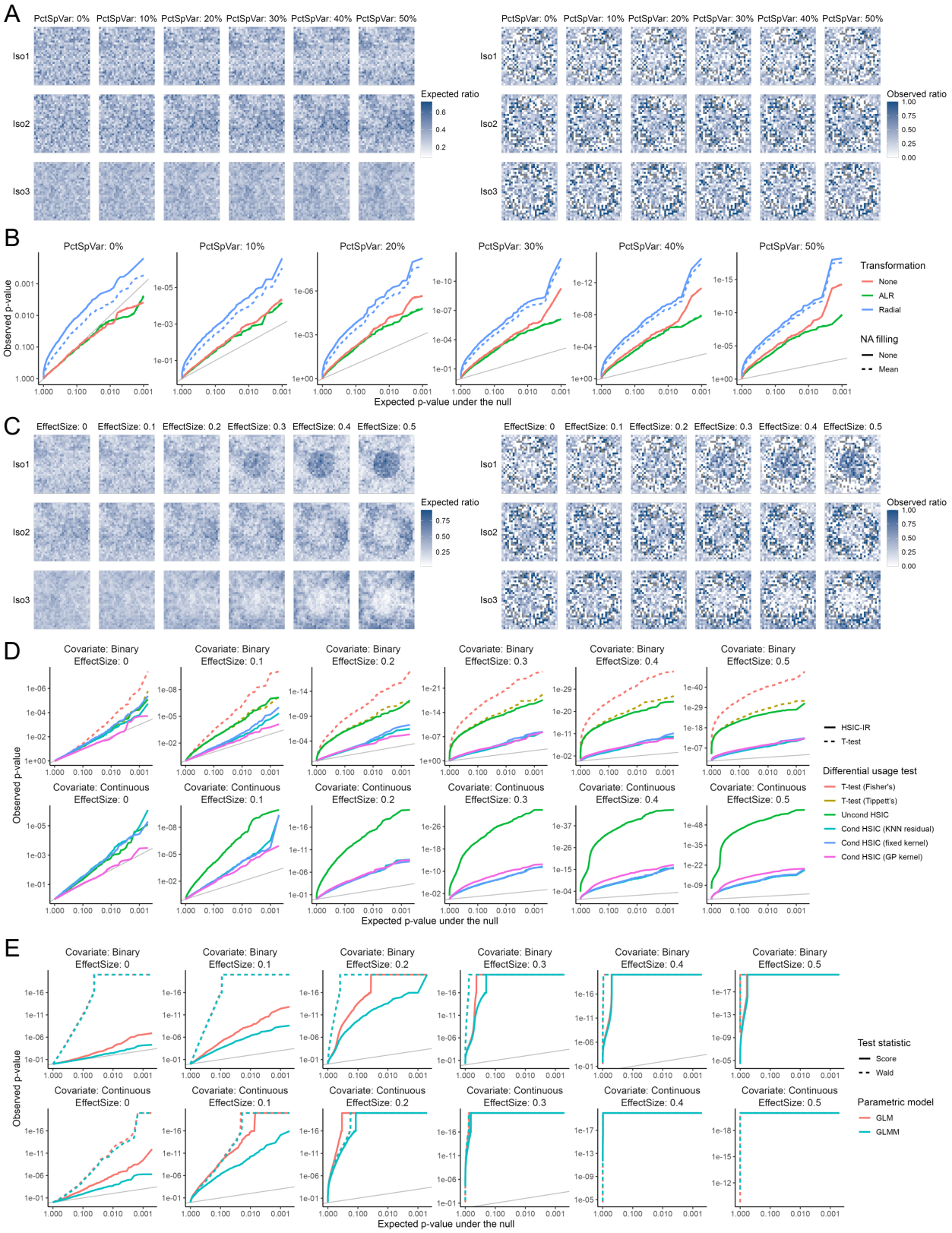

Figure S1: Power analysis for spatial variability and differential usage tests, related to Figure 2

(A) Expected and observed isoform ratios in simulations with varying degree of spatial variability, from Scenario 4 (no spatial variance) to Scenario 5 (50% spatial variance) in Figure 2.

(B) Q-Q plots of HSIC-IR p-values with different ratio transformations and undefined ratio replacement strategies. None means spots with NA ratios are removed and ignored for SV testing.

(C) Expected and observed isoform ratios in simulations with varying degree of covariate influence, from Scenario 5 (no association) to Scenario 6 (strong association) in Figure 2.

(D) Q-Q plots of non-parametric DU test p-values. Per-isoform T-test p-values were combined using either Fisher's or Tippett's methods. Conditional HSIC with fixed spatial or KNN kernels are included as approximations of the learnable Gaussian process kernel (default).

(E) Q-Q plots of parametric DU test p-values.

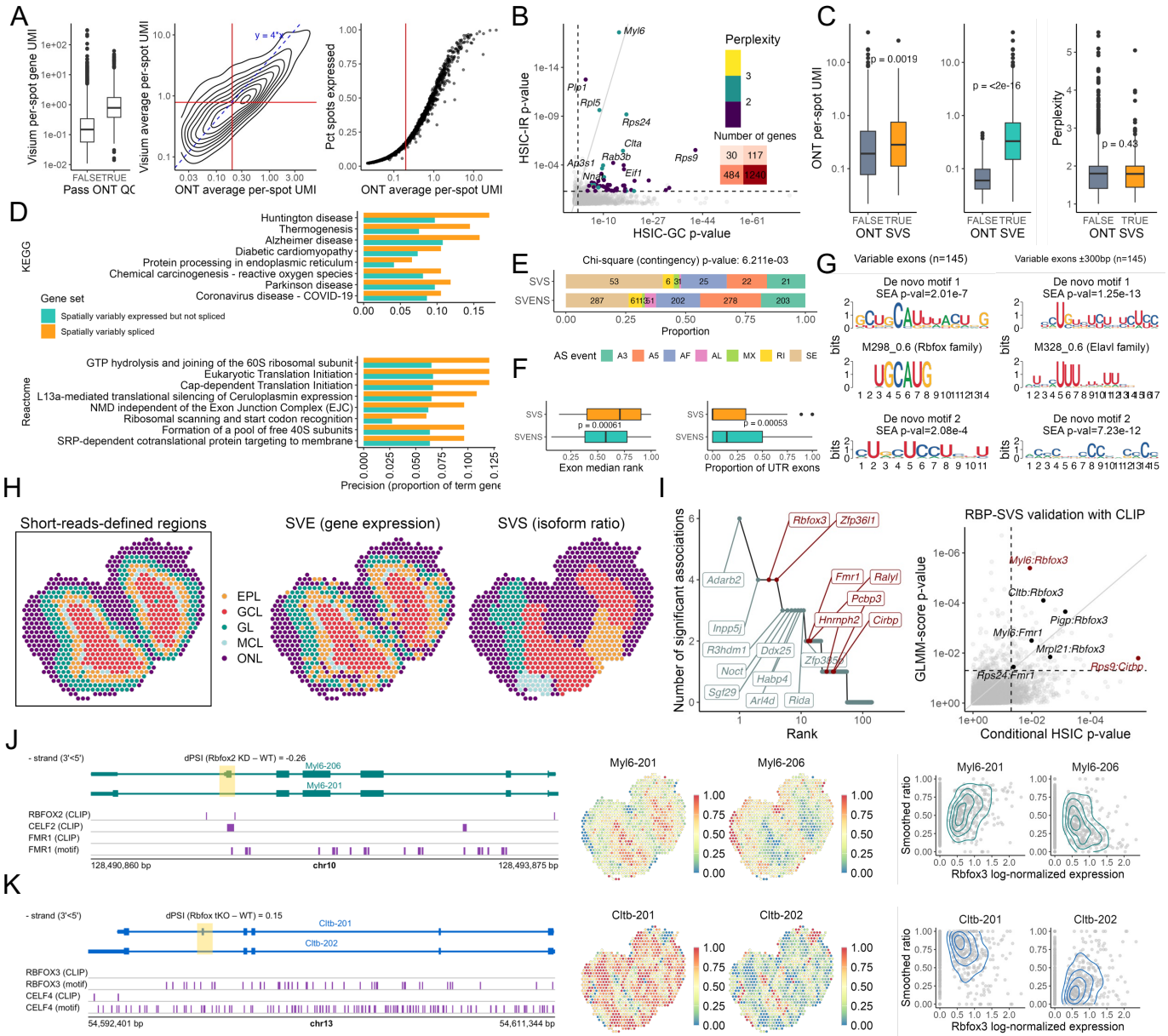

Figure S2: SPLISOSM identifies spatially variable splicing programs in a sparse mouse olfactory bulb dataset, related to Figure 3

(A) Sequencing depth and data sparsity of the SiT-ONT-MOB sample. Red lines indicate the median per-spot coverage (middle) and zero-coverage spots (right) across genes. The ONT-QC preprocessing step filters out isoforms expressed in less than 1% of spots and genes with less than 2 isoforms. Genes passed the filter have median per-spot total UMI of 0.25 in ONT data and 1 in SR data.

(B) Comparison of spatial variability in gene expression (x-axis) versus isoform usage (y-axis) in the MOB sample. Points are colored by expression perplexity (effective number of isoforms per gene), calculated as the exponential of global isoform usage entropy. Inset shows number of significant events (p-value < 0.05), categorized as spatially variably spliced (SVS, HSIC-IR p-value < 0.05) or spatially variably expressed but not spliced (SVENS, HSIC-GC p-value < 0.05 & HSIC-IR p-value > 0.05) genes.

(C) Sequencing depth and perplexity of SVS and SVE genes.

(D) Pathway enrichment analysis comparing SVS versus SVENS genes, with terms arranged by precision differences between the two groups.

(E) Distributions of alternative splicing event types among SVS and SVE genes.

(F) Selected Matt exon features of skipping exons (SE) and mutually exclusive exons (MX) in SVS and SVENS genes.

(G) Motif enrichment in SVS alternative exons (left) and exons with 300bp upstream/downstream flanking introns (right).

(H) Unsupervised clustering based on gene-level expression of SVE genes (middle) or relative isoform ratios of SVS genes (right). Isoform counts were locally smoothed to compensate for data sparsity.

(I) Top RBPs ranked by number of significantly associated targets, color-coded by availability of documented motifs (CISBP-RNA) or CLIP data (POSTAR3). Differential isoform usage p-values for selected RBPs are shown on the right, with validated. Validated RBP-SVS pairs with RBP binding near regulated events in CLIP data are highlighted in red.

(J) Myl6 transcript structure (left), smoothed spatial isoform usage ratios (middle), and its relationship with Rbfox3 log-normalized expression (right).

(K) Cltb transcript structure (left), smoothed spatial isoform usage ratios (middle), and its relationship with Rbfox3 log-normalized expression (right).

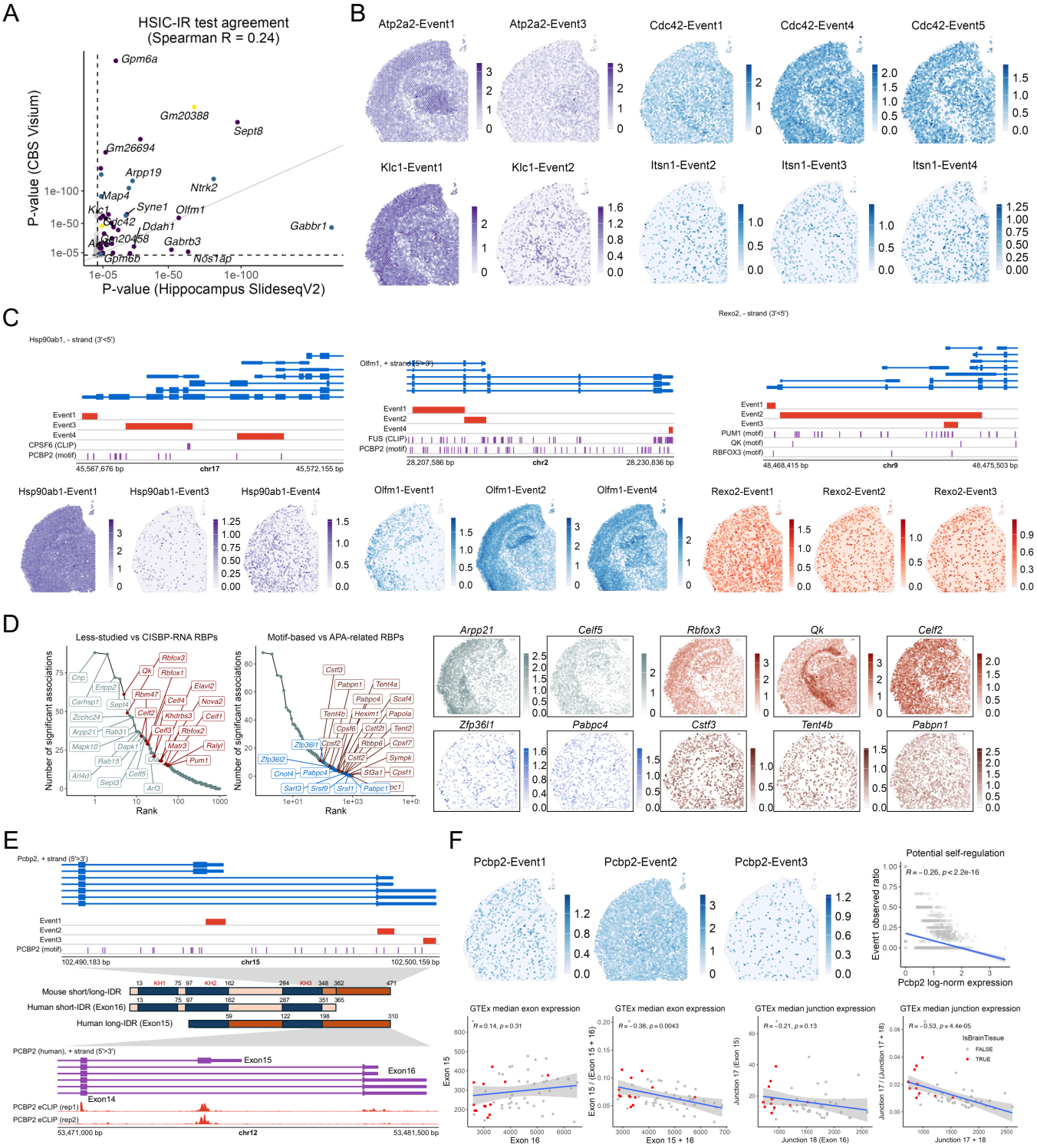

Figure S3: Spatially variable TREND events and their regulation in adult mouse brain, related to Figure 4

(A) HSIC-IR test agreements between Visium CBS (y-axis) and Slide-seqV2 hippocampus (x-axis) datasets. Points are colored by expression perplexity.

(B) Differential protein-altering alternative polyadenylation (APA) events between neurons and glial cells. Related to Figure 4E.

(C) Examples of additional protein-altering TREND events and their spatial log-normalized expression.

(D) Top RBPs ranked by number of significantly associated targets, color-coded by available motifs (CISBP-RNA) or CLIP data (POSTAR3), core APA machinery involvement and motif enrichment near SVS TREND events. Spatial log-normalized expression of selected RBPs are shown on the right.

(E) PCBP2 transcript structure in mouse and human, along with eCLIP data from ENCODE.

(F) PCBP2 isoform expression in the spatial mouse CBS data (top) and across GTEx human tissues (bottom). The negative association between PCBP2 gene expression and usage of the upstream exon is visualized on the right.

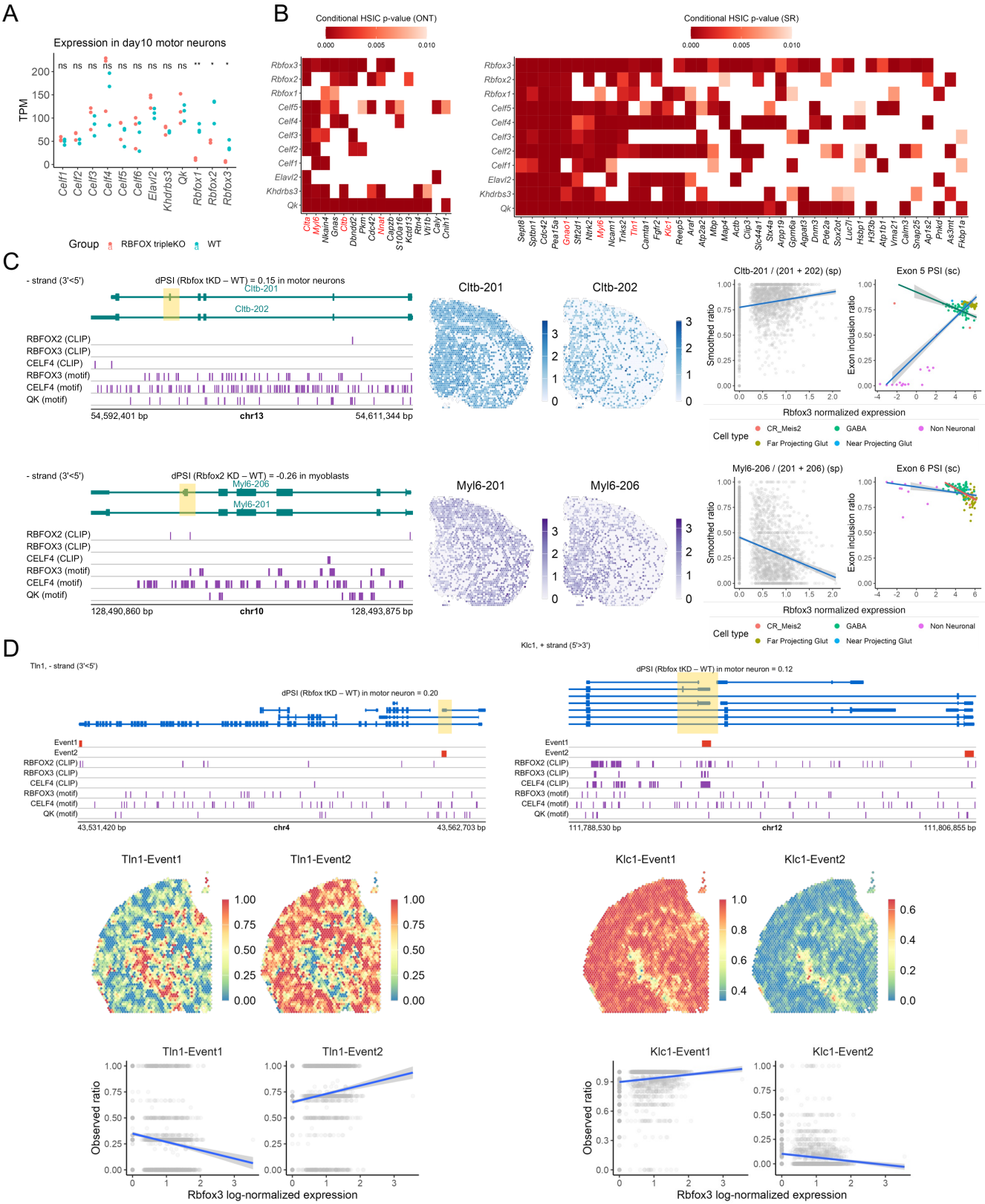

Figure S4: Discrepancy between knockout and observational data indicates cooperativity in RBFOX regulation, related to Figure 5
(A) Expression changes of selected RBP genes after RBFOX triple knockout where Rbfox genes were disrupted by in-frame stop codons.
(B) Heatmaps displaying DU test p-values between selected RBPs and target SVS genes associated with multiple RBPs. Only significant pairs (p-value < 0.01) are colored.
(C) Examples of two additional SVS genes whose observed isoform usage association with Rbfox3 contradicts the triple knockout data.
(D) Examples of two additional SVS genes whose observed last exon usage association with Rbfox3 contradicts the triple knockout data.

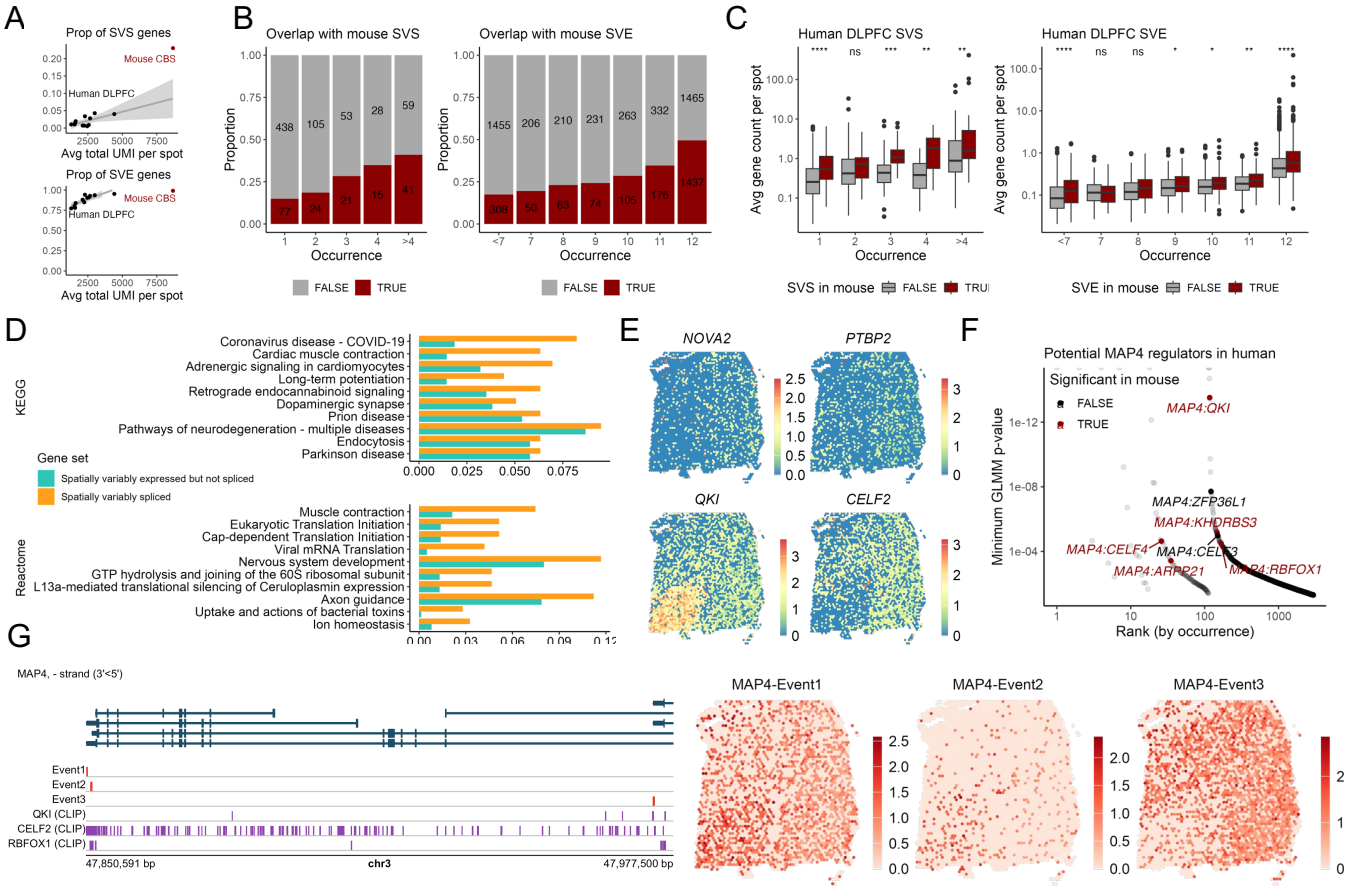

Figure S5: Additional examples of evolutionarily conserved SVS events in human DLPFC, related to Figure 6

(A) Relationship between total read coverage in TREND regions per spot (x-axis) and proportion of significant spatially variably spliced (SVS, top) or variably expressed (SVE, bottom) genes in all tested genes. Related to Figure 6B.

(B) Number and proportion of human DLPFC SV genes that have spatially variable mouse homologs, grouped by recurrence (number of significant DLPFC samples, x-axis).

(C) Sequencing depth of SVS and SVE genes, grouped by conservation and recurrence.

(D) Pathway enrichment analysis comparing human SVS versus human SVENS genes. Related to Figure 6F.

(E) Spatial log-normalized expression of selected RBPs in sample 151673.

(F) RBP-SVS associations ranked by recurrence (x-axis) and minimal GLMM p-values across samples (y-axis). Top potential regulators of MAP4 are highlighted and colored by whether the association is conserved in mouse.

(G) MAP4 transcript structure and spatial log-normalized expression in human sample 151673.

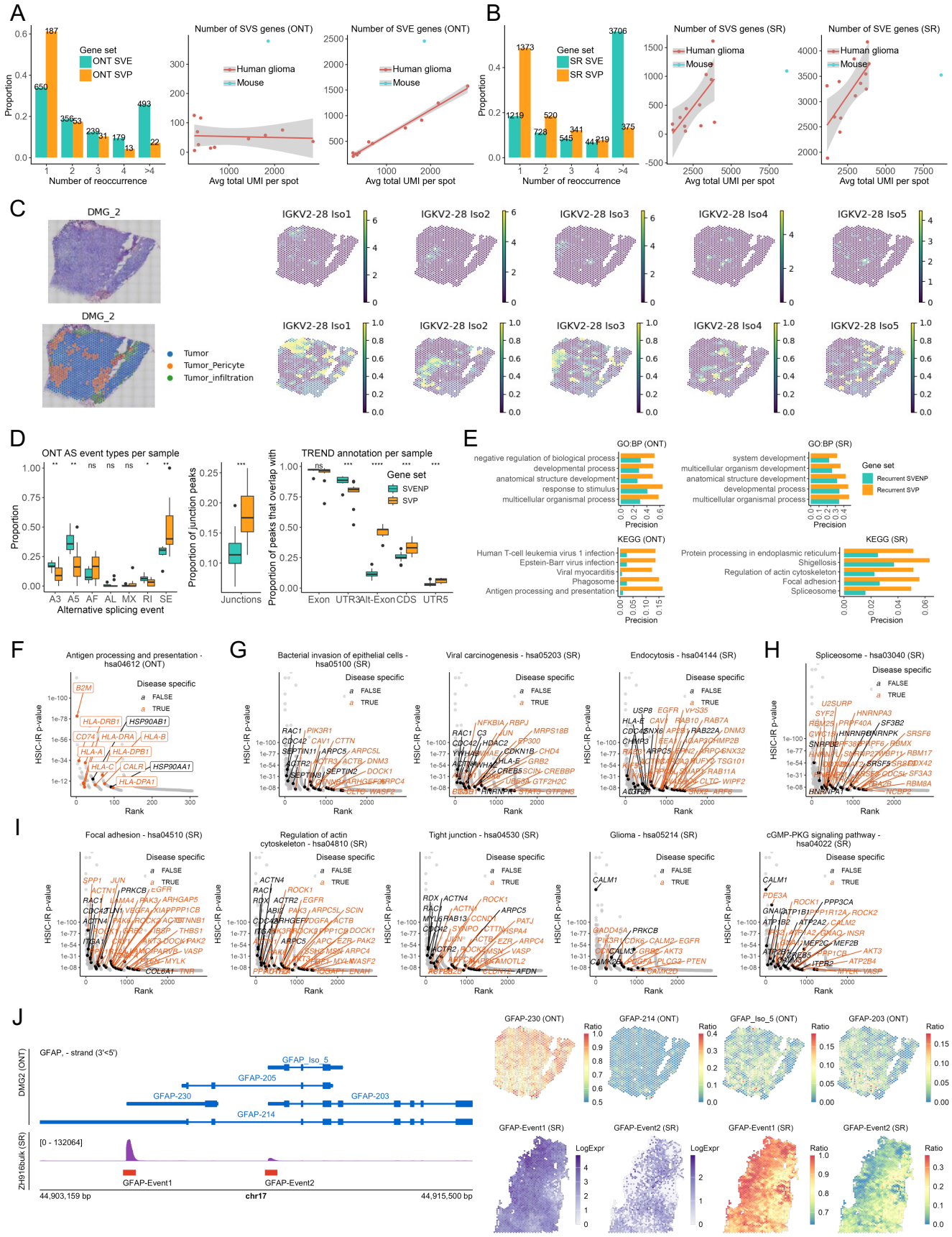

Figure S6: Additional examples of transcript diversity in human glioma, related to Figure 7
(A) Distribution of recurrent SVP genes (left) and relationships between total isoform read coverage per spot (x-axis) and number of SVS and SVE genes (y-axis) across ONT datasets. **(B)** Distribution of recurrent SVP genes (left) and relationships between total TREND read coverage per spot (x-axis) and number of SVS and SVE genes (y-axis) across SR datasets. **(C)** IG gene isoform diversity in the ONT sample DMG2. **(D)** Distributions of AS types in ONT-SV genes and TREND annotations in SR-SV genes. **(E)** Pathway enrichment analysis comparing recurrent SVP versus recurrent SVENP genes. Related to Figure 7D. **(F-I)** SVP genes involved in selected KEGG pathways. **(J)** GFAP transcript structure in the ONT sample DMG2, read coverage in the SR sample ZH916bulk, and the respective isoform or TREND spatial distribution in each sample (right).

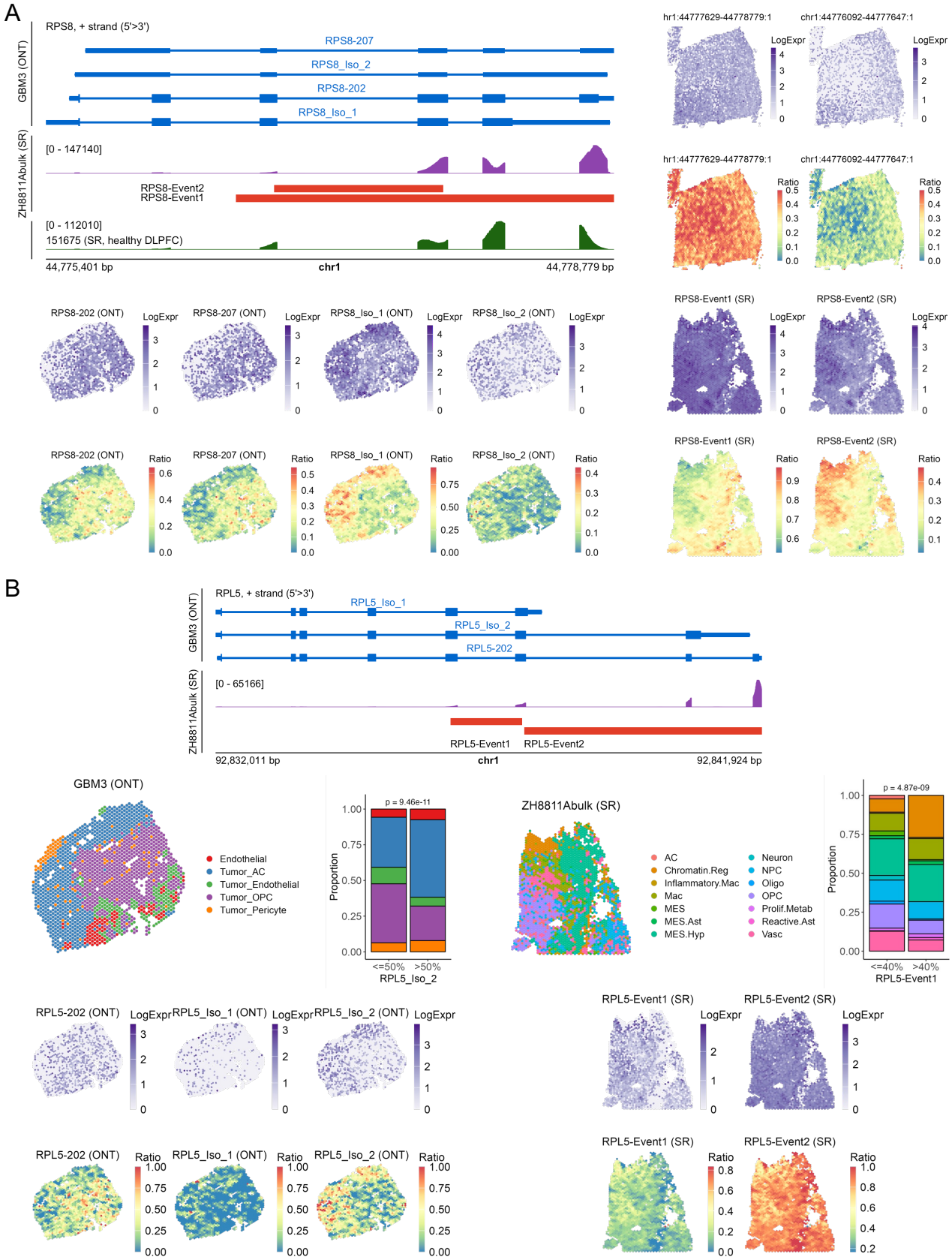

Figure S7: Spatial variable usage of reference and novel ribosomal transcripts in human glioma, related to Figure 7
(A) RPS8 transcript structure in the ONT sample GBM3, read coverage in the SR sample ZH8811Abulk and the SR healthy DLPFC sample 151675, and the respective isoform or TREND spatial distribution in each sample. Event counts were locally smoothed for ratio visualization, same in **(B)**.
(B) RPL5 transcript structure in the ONT sample GBM3, read coverage in the SR sample ZH8811Abulk and, and the respective isoform or TREND spatial distribution in each sample.