Tab 1

Cell-type-specific alternative splicing in the cerebral cortex and kidney of a *Setbp1*^{S858R} Schinzel-Giedion Syndrome patient variant mouse

Tabea M. Soelter[1#], Emma F. Jones[1#], Timothy C. Howton[1], Anthony B. Crumley[1], Elizabeth J. Wilk[1], Brittany N. Lasseigne*[1]

Affiliations

[1]Department of Cell, Developmental and Integrative Biology, Heersink School of Medicine, The University of Alabama at Birmingham, Birmingham, Alabama, United States of America

*Corresponding author
#Equal contribution

Email Addresses

Tabea M. Soelter: tsoelter@uab.edu, Emma F. Jones: emma.f.jones@wustl.edu, Timothy C. Howton: tchowton@uab.edu, Anthony B. Crumley: acrumley@uab.edu, Elizabeth J. Wilk: lizzyr@uab.edu, Brittany N. Lasseigne: bnp0001@uab.edu (*corresponding author)

Keywords

rare disease, single-cell RNA-Seq, neurodevelopment, gene regulation, splice junction usage, web application

Abstract

Schinzel-Giedion Syndrome (SGS) is an ultra-rare Mendelian disorder caused by gain-of-function mutations in the *SETBP1* gene. While previous studies determined multiple roles for how *SETBP1* and associated pathways may cause disease manifestation, they have not assessed whether cell-type-specific alternative splicing (AS) plays a role in SGS. We quantified gene and splice junction (SJ) expression from snRNA-seq data we previously generated from the cerebral cortex and the kidney of an atypical *Setbp1*^{S858R} SGS patient variant (n = 3) and wild-type (n = 3) mice. We performed pseudobulk differential gene expression and SJ usage (SJU) analyses across cell types and conditions. We identified 33 and 62 genes with statistically significant alterations in SJU in the brain and the kidney, respectively. Astrocytes and T cells had the most genes with cell-type-specific changes in SJU (n = 6 each) in the brain and kidney, respectively. We identified significant SJU in a member of the heterogeneous nuclear ribonucleoprotein family, *Hnrnpa2b1*. These findings were cell-type-specific for inhibitory neurons in the cerebral cortex and cell-type-agnostic in the kidney, suggesting tissue-specificity of AS in *Setbp1*^{S858R} mice. To broaden the impact of our results for the rare disease community, we developed a point-and-click web application as a resource for users to explore single-cell resolution changes in the presence of *Setbp1*^{S858R} at the gene and splice junction level. Overall, we find that AS may be implicated in a tissue- and cell-type-specific manner in the cerebral cortex and kidney of *Setbp1*^{S858R} mice.

Introduction

Schinzel-Giedion Syndrome (SGS) is an ultra-rare autosomal dominant Mendelian disorder caused by variants in the *SETBP1* gene [1]. *SETBP1* codes for a transcription factor expressed ubiquitously in the human body [2], and patients with SGS manifest clinical phenotypes related to the central nervous system, musculoskeletal system, heart, and kidney/urinary tract [3]. Symptoms of SGS include global neurodevelopmental impairment, progressive neurodegeneration, mild-to-profound intellectual disability, treatment-resistant seizures, distinctive craniofacial structure, muscle hypotonia/spasticity, hydronephrosis, and gastrointestinal problems [3–5]. Gain-of-function variants located in (typical SGS) or near (milder atypical SGS) the 12-base-pair hotspot of the degron region of the SETBP1 protein prevent SETBP1 degradation by proteasomes [6]. When the SETBP1 protein accumulates, it causes aberrant proliferation, deregulation of oncogenes and suppressors, unresolved DNA damage, apoptosis resistance [7], and decreased histone pan-acetylation in neural progenitor cells [8]. Models of SGS are critical for further resolving the molecular etiology and future therapeutics.

Recently, the heterozygous *Setbp1*^{S858R} patient variant mouse model (hereby referred to as *Setbp1*^{S858R}) was developed based on the S867R variant discovered in two atypical SGS patients who experienced seizures, developmental delay, and genital abnormalities [6]. *Setbp1*^{S858R} mice have low female fertility, smaller stature, and reduced brain, liver, and kidney organ weight compared to age and sex-matched wild-type mice [9]. With single-nuclei transcriptomics of the *Setbp1*^{S858R} cerebral cortex and kidney compared to wild-type mice, we previously reported that while *Setbp1* was only differentially expressed in excitatory neurons, many of its targets were differentially expressed and regulated in multiple brain and kidney cell types [10], suggesting a role of gene regulatory processes in SGS. Additionally, a recent study identified 38 alternatively spliced (AS) genes in the peripheral blood of a typical SGS patient using bulk RNA-seq [11], further implicating gene dysregulation in SGS. As an essential gene regulatory mechanism [12], AS increases organisms' functional genomic diversity by creating multiple isoforms/transcripts and proteins from a single gene through the joining of different combinations of exons from a single pre-mRNA. *SETBP1* itself is known to undergo AS in humans [13], and two SETBP1 targets, *LSM2* and *ZMAT2,* are splicesome and pre-splicesome components, respectively [14,15]. As cell-type-specific AS is essential for neurodevelopment [16], neurodevelopmental disorders [17], and epilepsy [18], we hypothesized that AS was altered in the cerebral cortex and the kidney of *Setbp1*^{S858R} mice compared to wild-type mice.

Here, we re-processed our previously generated *Setbp1*^{S858R} mouse snRNA-Seq data [10] to investigate the role of AS in SGS–an ultra-rare disease–and to create a valuable resource for the rare disease community and beyond. To quantify cell-type-specific AS, we used splice junction usage (SJU), which measures AS outputs (i.e., the relative isoform abundance produced from a single gene) from transcriptomic data by counting reads that span splice junctions. We calculated and compared cell-type-specific SJU in the cerebral cortex and the kidney of *Setbp1*^{S858R} and age-matched wild-type mice (**Fig. 1A**). This comparison identified cell-type-specific genes and their function and disease associations with AS between conditions (**Fig. 1B-E**). While we did not identify cell-type-specific AS in *Setbp1*, we found other genes with significant AS, like *Hnrnpa2b1*, a member of the heterogeneous ribonucleoprotein family. Our

findings further implicate AS in SGS by identifying cell-type-specific and -shared genes and pathways altered in the cerebral cortex and the kidney of *Setbp1*^{S858R} mice. To allow researchers to explore our AS analysis results on the *Setbp1*^{S858R} snRNA-seq data, we developed a Graphical User Interface. This resource, available as a Shiny web application [19], allows researchers to easily explore our cerebral cortex and kidney AS results for themselves quickly and without programmatic knowledge.

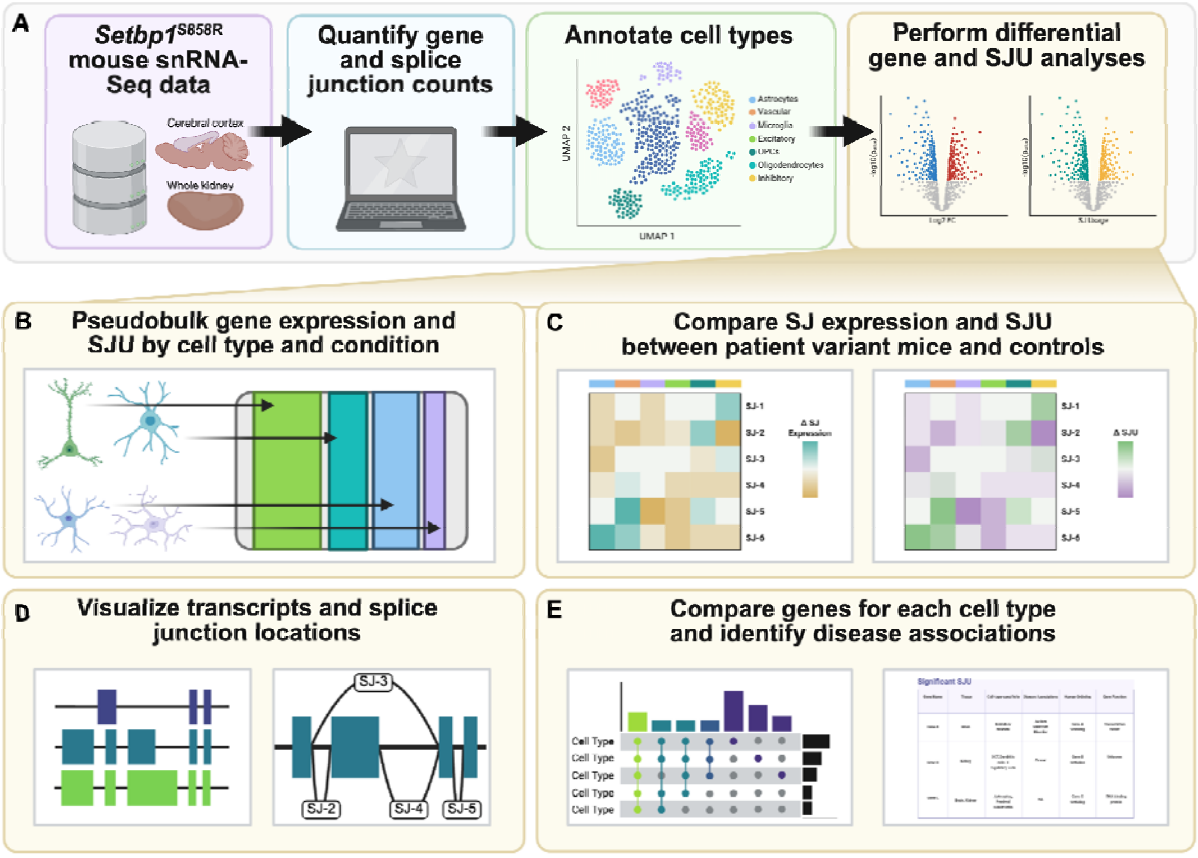


Figure 1. Graphical Abstract. (A) Schematic overview of our processing and analysis pipeline. (B) We analyzed pseudobulk gene expression and calculated SJU for each cell type and condition. (C) We compared SJU values for each cell type using a permutation test to identify cell-type-specific differences in AS between *Setbp1*^{S858R} and wild-type mouse tissues. (D) Next, we visualized all annotated transcripts and splice junction locations for each significant SJU gene. (E) Finally, we compared the genes with significant SJU between cell types and annotated their functions and disease associations to predict their biological relevance.

Materials and Methods

<u>Data acquisition</u>

We obtained raw snRNA-Seq data of six male mouse brain cerebral cortex and kidney samples: three 6-week-old C57BL/6J-*Setbp1*^{em2Lutzy}/J mice heterozygous for *Setbp1*^{S858R} (RRID:IMSR_JAX:033235) and three healthy, wildtype C57BL/6J age-matched wild-type mice

(RRID:IMSR_JAX:000664) from our recent publication[10], available on GEO from accession number GSE237816.

Data processing

We built a conda environment using Anaconda3 version 2023.07-2 to process raw sequencing data and provided the parameters to build this environment in an environment.yml file (available at https://github.com/lasseignelab/230926_EJ_Setbp1_AlternativeSplicing/blob/main/bin/conda/environment.yml. In that conda environment, we used STAR version 2.7.10b[20] to build a STAR genome reference from the GENCODE M31 primary assembly. Next, we ran STARsolo version 2.7.10b[20] on each sample with the following recommended options to best replicate 10x Genomics' CellRanger's filtering protocol and achieve the most recovered cell barcode similarity with our previous work: --soloType CB_UMI_Simple, --soloFeatures GeneFull_Ex50pAS SJ, --soloCellFilter EmptyDrops_CR, --soloUMIlen 12, --clipAdapterType CellRanger4, --outFilterScoreMin 30, --soloCBmatchWLtype 1MM_multi_Nbase_pseudocounts, --soloUMIfiltering MultiGeneUMI_CR, --soloUMIdedup 1MM_CR.

Data quality control and filtering

We removed ambient RNA from our gene counts using SoupX [21] using the raw matrix, barcodes, and feature files generated by STAR Solo. Since splice junction counts already exhibit very high sparsity, we did not repurpose SoupX to apply it to splice junction counts, thereby avoiding the exacerbation of their sparsity. We used the filtered barcodes, features, and matrix TSV files as input for preprocessing in Seurat. We used the Seurat version 5.0.0 R package for quality control, filtering, and clustering analyses [22] in the cerebral cortex and the kidney. We imported the gzipped and filtered STARsolo output matrices (barcodes, features, and matrix) into R using the Seurat Read10x function [22]. We created a Seurat object using the CreateSeuratObject function for each sample and condition (*Setbp1*^{S858R} and wild-type) before merging them into a single Seurat object per condition. We observed all cells were equally distributed across cell cycle phases using Seurat's MergeLayers and CellCycleScoring functions and converting default human gene IDs to mouse IDs using bioMart version 2.56.1. We filtered at the cell level (i.e., mitochondrial ratio < 0.05, between 1,000 and 15,000 genes per cell). We also removed *Malat1* from the cerebral cortex since it is frequently over-detected with poly-A capture technologies such as 10x[23]. Then, using default parameters, we performed batch correction using Harmony version 1.1.0[24] to preserve biological variation while reducing variation due to technical noise[25]. We scaled and normalized expression data using the ScaleData and NormalizeData functions with default settings (i.e., a scale factor equal to 10,000 and natural-log normalization). We then performed Principal Component Analysis (PCA) using the RunPCA function from Seurat[22] without approximation (approx = FALSE) in order to improve reproducibility. We plotted Uniform Manifold Approximation and Projections (UMAPs) and cell-type proportions to confirm successful integration across conditions for cerebral cortex and kidney samples.

Clustering and cell type assignment

We also used the Seurat R package (version 5.0.0)[22] for cell type assignment. We used a clustering resolution of 0.75, identifying 32 clusters in the cerebral cortex and a resolution of 1.25, identifying 50 clusters in the kidney using the Leiden option, which relies upon leidenalg v0.10.1[26]. We identified differentially expressed marker genes for each cluster (using the FindAllMarkers from Seurat[22]) with a log fold change threshold > 0.2 and a Bonferroni-adjusted p-value < 0.05. We assigned cell types using differential expression of cell-type-specific genes identified through PanglaoDB[27], CellMarker 2.0[28], and the Allen Brain Cell Atlas at https://portal.brain-map.org/atlases-and-data/bkp/abc-atlas (**Table S1**). We visually examined the expression of these canonical cell-type markers by making feature and dot plots. Due to non-specific cell markers and low cell numbers, we combined pericytes, vascular leptomeningeal, and endothelial cells into one vascular cell type in the cerebral cortex.

SJ expression normalization

To normalize SJ expression for data visualization, we employed a method similar to Seurat's NormalizeData function, which divides feature counts by the total counts for that cell, multiplying by a scale factor, and applying a natural-log transformation. Here, we divided SJ counts by the total SJ counts for that cell, multiplied by a scale factor of 1000 (determined by the average number of SJs expressed per cell in our dataset), and finally used the base R function log1p to natural-log-transform the resulting values.

SJU calculation with MARVEL

To calculate SJU, we divided the total SJ count for a cell type for each SJ by the total read count for that gene. As SJU is a percentage, we converted SJU values over 100 to a maximum of 100 (which only occurred for low-count genes) and any infinite values (indicating no gene expression) to NaN.

Differential SJ and gene expression analysis with MARVEL

We used MARVEL version 2.0.5 [29], an R package with multiple analytical tools for investigating single-cell resolution SJU, to integrate SJ and gene expression information into a single R object per tissue. We filtered the data to include only splice junctions within a single gene annotated in GENCODE release M31. We imported the filtered gene expression data processed with Seurat and SJ expression data processed with STARsolo [20]. Then, we filtered the SJ data for cells that passed the Seurat quality control metrics described above and combined them into a single sparse count matrix. Finally, using the combined sparse matrix, we created cerebral cortex and kidney MARVEL objects using the CreateMarvelObject.10x function.

We also used the MARVEL package for differential SJ and gene expression analyses. For differential SJU, we employed MARVEL's CompareValues.SJ.10x function, which applies a permutation test on the SJU of individual SJs for a given cell group, as transcripts of splice junctions are rare, creating extreme sparsity, which cannot be overcome with a non-parametric rank-based test [29]. We compared conditions (*Setbp1*^{S858R} and wild-type mice) for each cell type across tissues. We used a p-value cutoff of 0.05 (meaning out of 100, the permutation of random condition assignment would randomly have 5 or fewer absolute values of the delta

greater than actual delta values) and an absolute value of delta SJU of greater than one. We used at least 5% of cells expressing a gene or SJ as gene and SJ expression cutoffs.

To identify genes that are differentially expressed and have SJU, we used MARVEL's CompareValues.Genes.10x function. This function employs a Wilcoxon rank sum test on normalized log2-transformed gene expression values.

Significant SJU disease annotation

We observed a total of 84 unique genes with significant SJU in the cerebral cortex and the kidney, mapping to 74 human orthologs. To annotate genes with significant differential SJU for disease associations, we compared COSMIC cancer gene census (version 100), SFARI Gene (2025 Q3 release), and GenCC (accessed 250811) to our genes following mouse to human ortholog mapping (gprofiler2 version 0.2.3). The COSMIC cancer gene census consists of expert-curated cancer-driving genes, tumors associated with cancer-driver genes, and assignment of genes to Tier 1 (strong functional and mutational support) or Tier 2 (strong support either functionally or mutationally, but not both) [30]. Of note, none of our genes with significant SJU in either organ were associated with germline tumors in COSMIC, so we only reported somatic tumor associations. The SFARI Gene includes autism-associated genes, with gene scoring for the strength of evidence supporting associations [31,32]. Finally, GenCC is an international gene curation coalition with submitters from diagnostic laboratories and established online resources (e.g., Gene2Phenotype, OMIM, Orphanet, ClinGen) [33].

Transcript structure and SJ visualization

We used the R package ggtranscript version 0.99.9 to visualize splice junctions and the GENCODE release M31 gtf to annotate known transcripts of genes with significant SJU. We labeled SJs with the letters "SJ" and a number (e.g.,"-1") indicating the SJ's genomic location on a transcript. If a gene is transcribed in the forward direction, we assign the SJ numbers in ascending order and vice versa. This way, we labeled SJs consistently from the 5 ' to 3' direction of translation.

PPI Network Investigation

We accessed the StringDB (Version 12.0) graphical user interface at https://string-db.org/ to identify and visualize known interactions between the TF Setbp1 and the RNA-binding proteins Hnrnpa1 and Hnrnpa2b1. We included all three proteins in our query under the "Multiple proteins" tab and selected Mus musculus as the organism of interest. To investigate the interactions between the proteins we input, we chose to include 10 additional nodes in the PPI by selecting the "more" button twice (total nodes = 13). In the "Settings" tab, we opted to have edge thickness represent the strength of the data supporting the interactions. The confidence scores reported in the manuscript were extracted by selecting the respective edges, which delineate the combined StringDB score and the evidence that supports it.

Web application development

We built our web application using the R Shiny[34] framework version 1.8.1.1. It is hosted with shinyapps.io at https://lasseignelab.shinyapps.io/setbp1_as.

Reproducibility

For all the analyses following data processing, we used R version 4.3.1 through RStudio version 2023.06.2+561 running in docker containers through Singularity version 3.5.2 on the UAB supercomputer, Cheaha. Docker images for each analysis are specified in each script and available on Docker Hub (https://hub.docker.com/repository/docker/emmafjones/setbp1_alternative_splicing/general and https://hub.docker.com/repository/docker/tsoelter/setbp1_alternative_splicing/general) and Zenodo (https://zenodo.org/records/12534825). All code was independently reproduced by two other lab members and is available on GitHub (https://github.com/lasseignelab/230926_EJ_Setbp1_AlternativeSplicing) and Zenodo (10.5281/zenodo.17601458).

Results

Cerebral cortex and kidney snRNA-Seq from *Setbp1*^{S858R} mice shows cell-type-specific gene and splice junction expression

We processed our previously generated [10] snRNA-seq data set from the cerebral cortex and the kidney of C57BL/6J-*Setbp1*^{em2Lutzy}/J mice heterozygous for *Setbp1*^{S858R} and matched wild-type mice, and quantified gene and SJ expression using STAR [20]. We retained 51,318 cerebral cortex and 75,889 kidney nuclei that passed quality control filtering. From there, we annotated cerebral cortex cell types (**Fig. 2A, B**) using canonical marker gene expression to get seven cell types: astrocytes, excitatory neurons, inhibitory neurons, microglia, oligodendrocytes, OPCs, and vascular cells. As expected for cerebral cortex tissue and in agreement with our previous publication[10], most of the nuclei we annotated were neurons, with the largest group being excitatory neurons (n = 33,569; **Fig. 2A**). Similarly, we performed kidney cell type annotation using canonical marker gene expression to identify 18 cell types (**Fig. S1A, B**). Brain and kidney cell types were evenly distributed across *Setbp1*^{S858R} and wild-type mice (**Fig. 2C, D & Fig. S1C, D**), underscoring the high quality of the dataset and sample processing consistency.

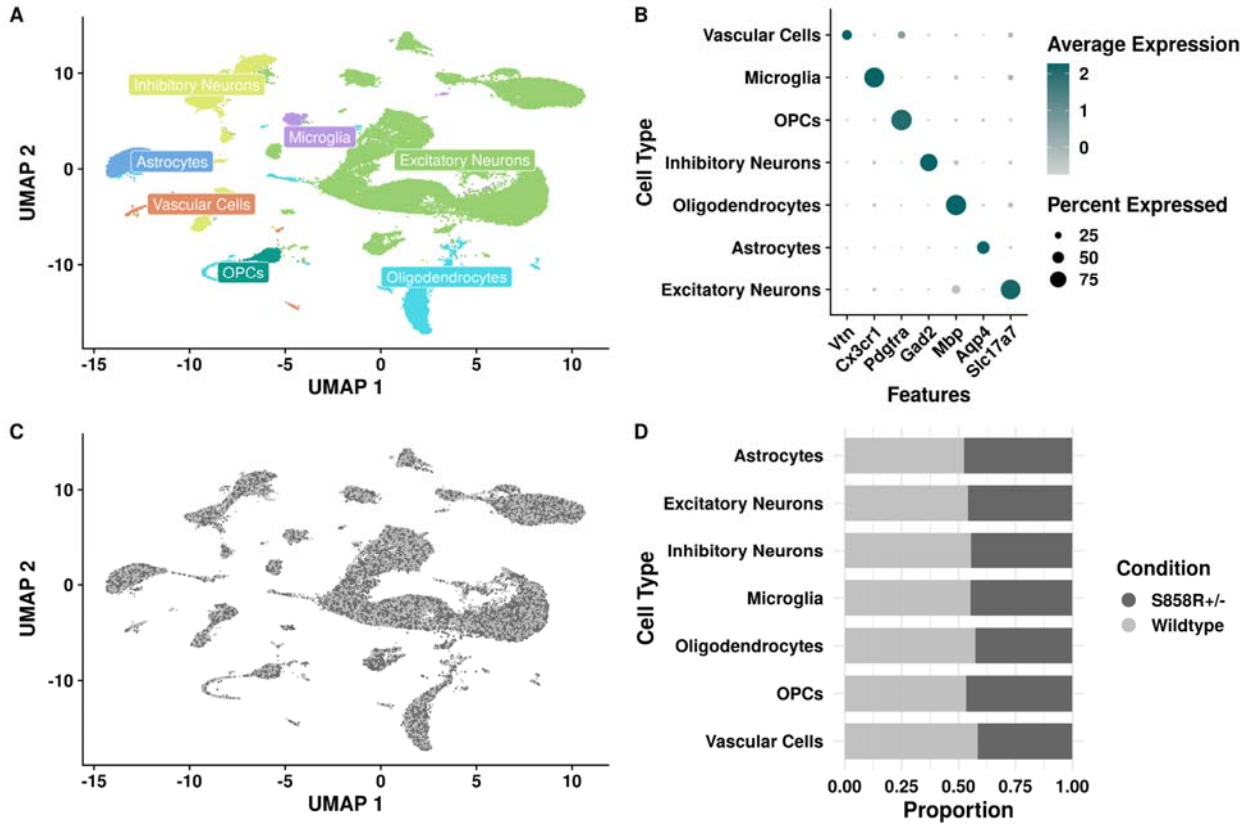


Figure 2. Cell types are evenly distributed across *Setbp1*^{S858R} and wild-type mice. (A) Representative uniform manifold approximation and projection (UMAP) colored by cell types annotated in our integrated dataset. (B) A dot plot of representative cell type markers for cell types is shown in panel A. Size represents the percent of cells expressing a given gene, and color saturation represents average expression. (C) Representative UMAP colored by condition. (D) Bar plot of cell type proportions based on condition. Dark gray denotes *Setbp1*^{S858R} mice, and light gray denotes wild-type mice.

After annotating cell types, we integrated SJ count information from STAR into our analyses using MARVEL, an R package for single-cell splicing analysis[29]. We identified more SJs in cells with more expressed genes (linear regression, $R^2 = 0.8035$). For example, excitatory and inhibitory neurons had both a higher total number of genes expressed and SJs detected per cell than glial cell types in the cerebral cortex (**Table S2, Fig. S2A, B**). This is consistent with other research, as the fact that neurons express more genes than glia has previously been reported[35]. However, neurons still express more SJs per cell than glial cell types when dividing by the total number of genes expressed (**Fig. S2C**), suggesting neuron-specific genes may be more transcriptionally complex. This finding corroborates a recent study that identified distinct neuron-specific AS complexity across multiple RNA-binding proteins and in genes involved in synapse, cell projection, and ion channel activity between neuronal subpopulations using deep scRNA-seq of the mouse cortex [36]. Additional work in murine brains has highlighted the differential splicing of neurexin genes in pre- and postsynaptic neurons [37]. Finally, previous findings indicate that neuron-specific synaptic genes are longer and have more transcript isoforms[38]. In contrast, the number of SJs per cell, normalized for the number of genes per cell across kidney

cell types, is uniformly distributed in the kidney compared to the cerebral cortex cell types (**Fig. S3**). The gene we detected with the most SJs in the cerebral cortex was *Syne1*, which had 156 different SJs in excitatory neurons (**Table S2**). *SYNE1* encodes for spectrin repeat-containing nuclear envelope protein 1, or nesprin-1, in which human variants can cause multiple Mendelian disorders, including cerebellar ataxia, Emery-Dreifuss muscular dystrophy, and arthrogryposis multiplex congenita[39]. Our results corroborate previous findings that *Syne1* has over 100 exons, resulting in dozens of transcript variations[40]. We saw cell-type-specific differences in the number of detected SJs for *Syne1*: we detected 156 SJs in excitatory neurons (the most), but only 25 SJs in microglia (the least). In the kidney, either *Syne1* or *Syne2*, a prognostic marker for clear cell renal cell carcinoma [41], had the most SJs in all 18 cell types (**Table S3**). This highlights the cell-type-specificity of SJ expression and, thus, AS in our single-nuclei dataset.

Setbp1^{S858R} mice have cell-type-specific splicing in all measured cerebral cortex and most kidney cell types, but not in *Setbp1*

To determine cell-type-specific SJU changes in *Setbp1*^{S858R} compared to wild-type mice, we performed a permutation analysis using the MARVEL[29] R package for each cell type between conditions. We detected 33 and 62 genes with significant changes in SJU in the cerebral cortex and the kidney, respectively (**Table S4;** permutation test, $p < 0.05$ and delta > 1 – see **Methods** section "Differential SJ and gene expression analysis with MARVEL" for details). The cerebral cortex cell type with the most genes with significant SJU was astrocytes (n = 6, permutation test, p-val < 0.05), while the least was OPCs (n = 2) (**Fig. 3A**). In contrast to the cerebral cortex, we did not observe any cell-type-specific AS genes in 3 of 18 kidney cell types (Collecting duct principal cells (CDPCs), thin ascending limb of the loop on henle cells, and dendritic cells; **Fig. 3B**). T cells had the most genes with significant SJU (n = 6) and 6 cell types only had one gene with significant SJU (**Fig. 3B**). Interestingly, almost 82% (27 of 33 genes) of significant SJU genes in the cerebral cortex were cell-type-specific, while only 53% (33 of 62 genes) of significant SJU genes were cell-type-specific in the kidney of *Setbp1*^{S858R} compared to wildtype mice (**Fig. 3**). Additionally, we found that 13 genes with significant SJU that overlapped between the cerebral cortex and the kidney in our *Setbp1*^{S858R} mice (**Table S4**). Out of 13 cross-tissue significant SJU genes in our data, 7 encode RNA-binding proteins and other splicing regulators, including *Pnisr*, *Hnrnpa2b1*, *Srrm2*, *Zcchc7*, *Son*, *Mbnl1*, and *Srek1* (**Table S4**). While *Hnrnpa2b1*, an RNA-binding protein, was cell-type-specific in the cerebral cortex, its AS pattern was broader in the kidney, and *Son*, a splicing co-factor with significant changes in SJU for all non-vascular cell types in the cerebral cortex, was cell-type-agnostic in both the cerebral cortex and the kidney (**Table S4**), highlighting the cell-type- and tissue-specific diversity of AS patterns in *Setbp1*^{S858R} compared to wildtype mice. Consistent with our previous work [10], *Setbp1* was expressed in all cerebral cortex and kidney cell types. However, we did not observe significantly different *Setbp1* gene expression, splice junction expression (SJE), or SJU in *Setbp1*^{S858R} compared to wild-type mice (**Figs. S4 & S5**). While *Setbp1* is not significantly differentially expressed or spliced, our previous work suggests that the S858R variant in *Setbp1*, which encodes a transcription factor, affects downstream gene regulation, as evidenced by the significant transcriptional changes observed in multiple Setbp1 targets. Therefore, the lack of AS in *Setbp1* in our study is unsurprising, as the Setbp1 S858R protein disrupts Setbp1

degradation, allowing it to continue accumulating in the absence of gene expression and regulatory changes [10]. While we observed cerebral cortex-specific changes in expression and usage (i.e., decreased in inhibitory neurons and vascular cells, and increased in excitatory neurons and oligodendrocytes) of SJ-3, which incorporates the Setbp1S858R variant located on exon 4, these differences were not statistically significant during permutation testing (**Fig. S4C**). Additionally, only one *Setbp1* isoform has been identified in mice (*Setbp1-201;* **Fig. S4A**). While we did not identify significant changes in AS of *Setbp1*, our findings indicate altered splicing patterns in *Setbp1*^{S858R} mice with varying degrees of cell-type-specificity across the cerebral cortex and the kidney.

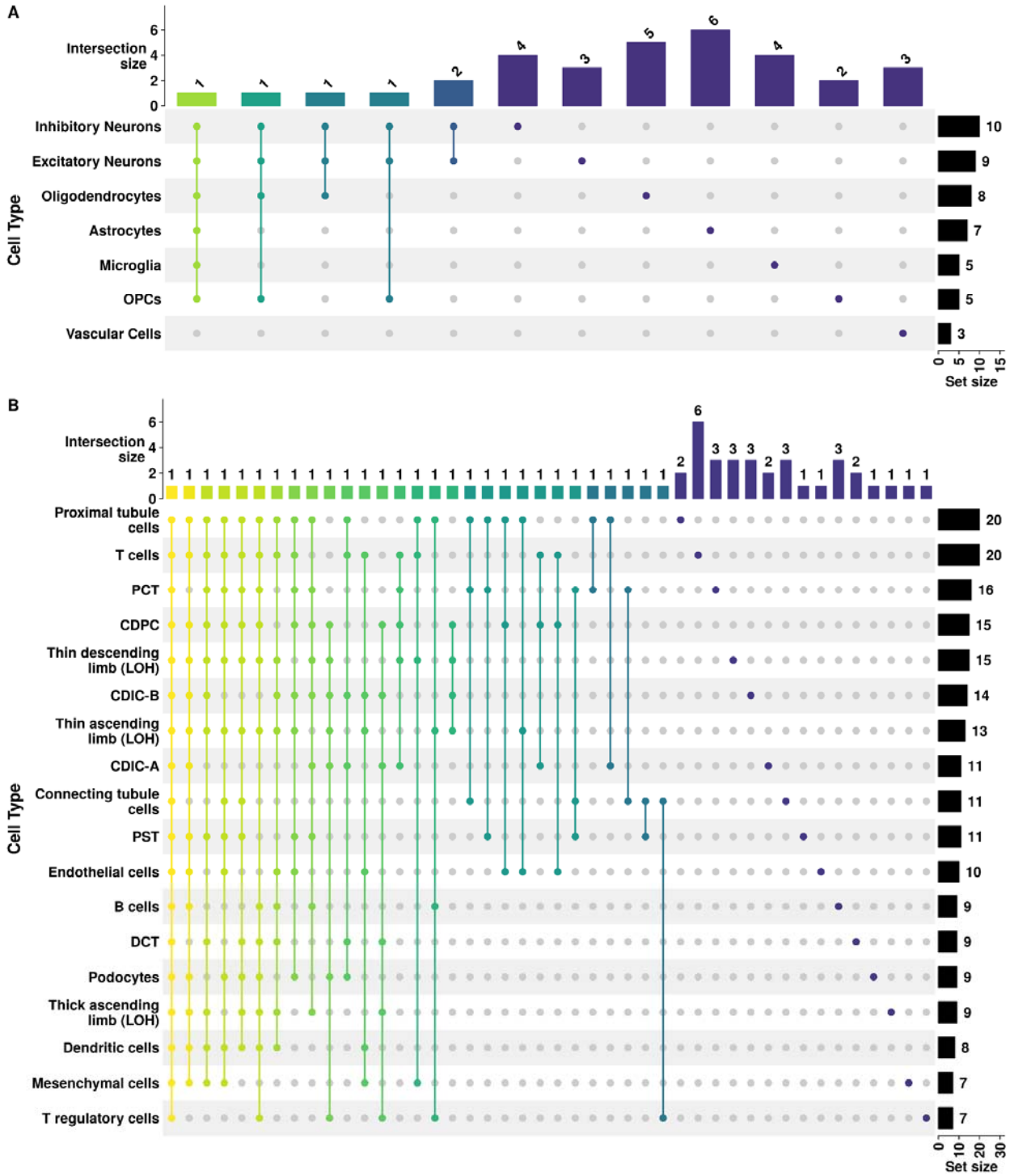


Figure 3. *Setbp1*^{S858R} mice have cell-type-specific splicing in all cerebral cortex and most kidney cell types compared to wild-type mice. UpSet plot of genes with significant SJU changes between *Setbp1*^{S858R} and wild-type mice, split by cell type in (**A**) cerebral cortex and (**B**) kidney. Brighter colors indicate higher overlap between cell types, with purple showing genes unique to each cell type. Bar graphs on the far right represent the total set size for each cell type, while the top bar graphs denote the intersection size.

Alternative splicing of the heterogeneous nuclear ribonucleoprotein family A2/B1 (*Hnrnpa2b1*) gene is cell-type- and tissue-specific in *Setbp1*^{S858R} compared to wildtype mice

Since SGS is an ultra-rare Mendelian disorder with fewer than 100 reported cases worldwide [3], we sought to compare our genes with significant SJU to previously identified AS genes from a bulk RNA-seq study in the peripheral blood of a typical SGS patient compared to their healthy parents [11]. None of our predicted SJU genes from the cerebral cortex or the kidney of this atypical SGS mouse model directly overlapped with the 37 orthologs of the 38 AS genes identified in human peripheral blood. However, we predicted significant SJU in the cerebral cortex and the kidney of the *Hnrnpa2b1* gene, encoding an RNA-binding protein that is closely related to the heterogeneous nuclear ribonucleoprotein A1 (*Hnrnpa1*) [42] identified as an AS gene in the peripheral bulk RNA-seq blood study.[11] While the functions of Hnrnpa1 and Hnrnpa2b1 proteins are not identical, the members of the heterogeneous nuclear ribonucleoprotein family act compensatory in disease [43], show structural homology [44], and play essential roles in neurodevelopment [45]. Previous protein-protein interaction (PPI) data compiled in StringDB indicate a strong, high-confidence interaction between the Hnrnpa1 and Hnrnpa2b1 proteins (StringDB confidence score = 0.982; **Fig. S6**). Additionally, StringDB data suggest that Setbp1 interacts with both heterogeneous nuclear ribonucleoproteins through its PPI with Set (StringDB confidence score = 0.791; **Fig. S6**), which is known to be bound by Setbp1 [46]. SETBP1 protein accumulation causes DNA damage in neural progenitor cells derived from SGS patients, as accumulated SETBP1 binds and stabilizes SET, thereby blocking P53 signaling and disrupting cell proliferation and DNA repair [7]. In our snRNA-seq data, we detected 13 transcripts of *Hnrnpa2b1* (**Fig. 4A & S7A**) with 17 and 18 SJs in the cerebral cortex and the kidney, respectively (**Fig. 4C & S7C**). Significant SJU in *Hnrnpa2b1* was specific to inhibitory neurons in the cerebral cortex (**Fig. 4C bottom & Table S4)** and shared across multiple cell types in the kidney (**Fig. S7C bottom & Table S4**) in *Setbp1*^{S858R} compared to wildtype mice. Interestingly, normalized *Hnrnpa2b1* gene expression was highest in astrocytes and microglia in our dataset, rather than inhibitory neurons, in the cerebral cortex (**Fig. 4B**). In contrast, *Hnrnpa2b1* had significant AS in 12 of 18 kidney cell types (i.e., CDPC, connecting tubule cells, distal convoluted tubule (DCT) cells, dendritic cells, proximal convoluted tubule (PCT) cells, podocytes, proximal tubule cells, proximal straight tubule (PST) cells, T cells, Thick ascending limb of the loop of henle (LOH), Thin ascending limb of the LOH, Thin descending limb of the LOH; **Fig. S7C**), indicating a tissue-specific AS pattern in *Setbp1*^{S858R} mice. Overall, our findings of SJU in *Hnrnpa2b1* suggest that AS of genes encoding heterogeneous nuclear ribonucleoproteins are altered across species (human and mouse) and tissues (cerebral cortex, kidney, peripheral blood) and have a cell-type-specific pattern in the cerebral cortex in SGS.

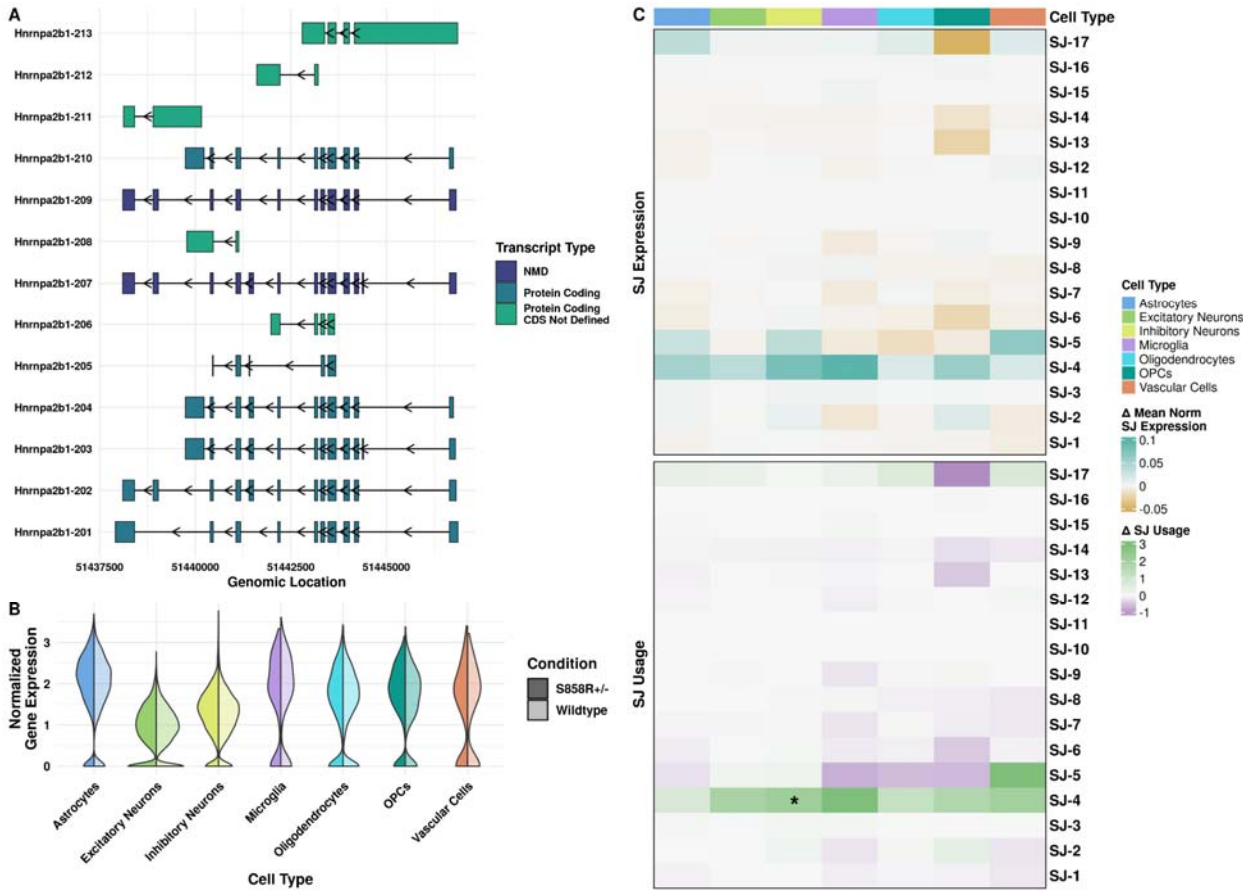


Figure 4. *Hnrnpa2b1* has inhibitory neuron-specific AS changes in the cerebral cortex of *Setbp1*^{S858R} compared to wild-type mice. (A) Transcript models of all 13 annotated transcripts of *Hnrnpa2b1*. The color indicates transcript classification: indigo = transcripts flagged for nonsense-mediated decay (NMD), dark teal = protein-coding transcripts, turquoise = protein-coding transcripts, but coding sequence (CDS) is not defined. Arrows indicate the direction of transcription. (B) Split violin plots showing *Hnrnpa2b1* normalized gene expression per cell for all cell types, split by condition. (C) Heatmaps of the changes in normalized mean SJ expression (top) and usage (bottom) between *Setbp1*^{S858R} and wild-type mice for 17 SJs of *Hnrnpa2b1*. The top heatmap annotation indicates cell type. A positive delta indicates expression or usage was higher in *Setbp1*^{S858R} than in wild-type mice, and a negative delta indicates expression or usage was higher in wild-type mice. The asterisk indicates the significant SJU in inhibitory neurons.

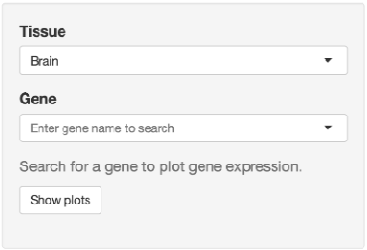
<u>We developed a web application for visualizing gene expression, SJE, and SJU in our SGS model snRNA-seq data</u>

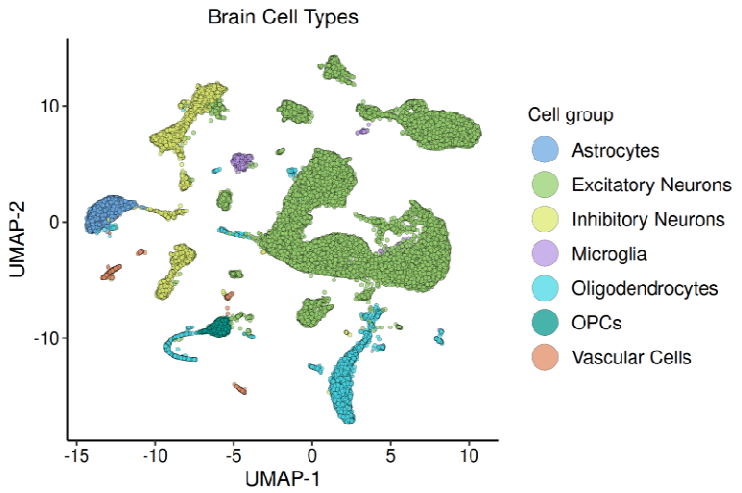
Finally, to expand the accessibility of our analyses to researchers focused on ultra-rare diseases and the broader scientific community, we built an R Shiny[34] web application featuring the data presented here. This application is particularly valuable for the ultra-rare disease community, such as those studying SGS, where data are often scarce, and patient numbers are limited. This application offers several key benefits and use cases:

1. Custom Visualization: Users can create custom gene expression and SJU UMAPs (**Fig. 5A-C**) for any chosen gene in C57BL/6J wildtype and *Setbp1*^{S858R} mouse cerebral cortex and kidney. This flexibility enables researchers to visualize specific genes of interest using pre-processed, normalized, and scaled gene expression values, as well as pre-calculated SJUs.
2. Exploration of Significant Genes: Researchers can further investigate the 33 cerebral cortex and 62 kidney genes we identified with significant SJU and generate AS gene summary plots (as shown in **Fig. 4**).
3. Comparative Analysis: The resource supports exploration within and across conditions (wildtype and *Setbp1*^{S858R}), cell types, tissues, genes, and splice junctions. Users can quickly search for changes in SJ expression and usage between wildtype and *Setbp1*^{S858R}.
4. Gene Isoform Information: The application provides detailed information on gene isoforms, including genomic locations and annotated regions (e.g., NMD, protein coding, and retained introns). This feature helps researchers understand the functional implications of different isoforms.
5. Data Download: Researchers can easily download plots from their custom analyses, making it convenient to use these visualizations for preliminary data analysis, hypothesis generation, or as supporting evidence in their own research.
6. Additional Resources: The web application includes helpful links to external resources (e.g., genome browsers) and a frequently asked questions (FAQ) section to assist users in navigating the tool and understanding the data.

Our Shiny web application serves as an excellent resource for the scientific community, facilitating the exploration and analysis of our *Setbp1*^{S858R} snRNA-Seq dataset. The application is publicly available at https://lasseignelab.shinyapps.io/setbp1_as, providing a user-friendly platform for researchers to explore the data and generate new insights for their own research.

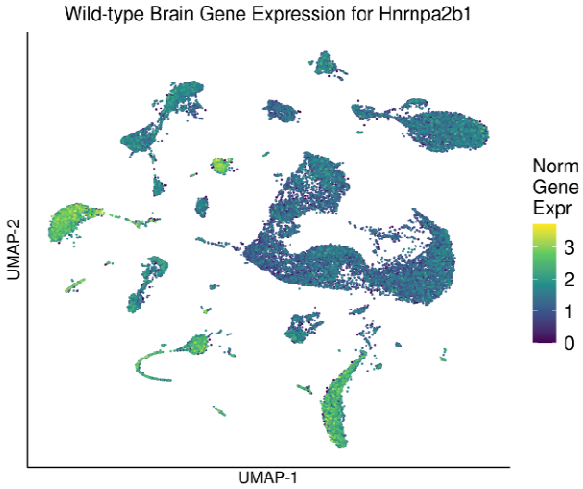
A



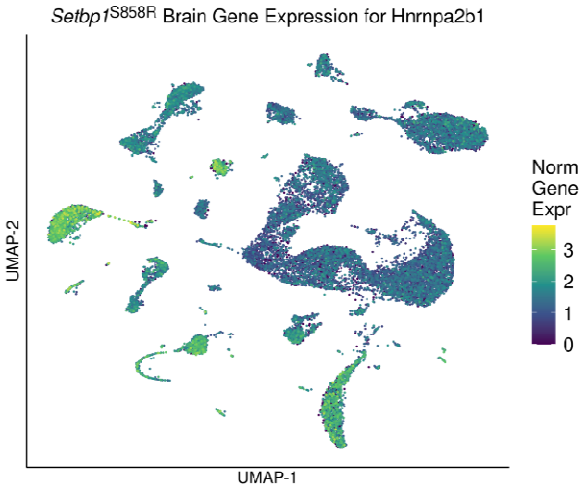


This UMAP displays the assigned cell types of all cells in our dataset. Cell color indicates cell type, following the colors in the figure legend.

B

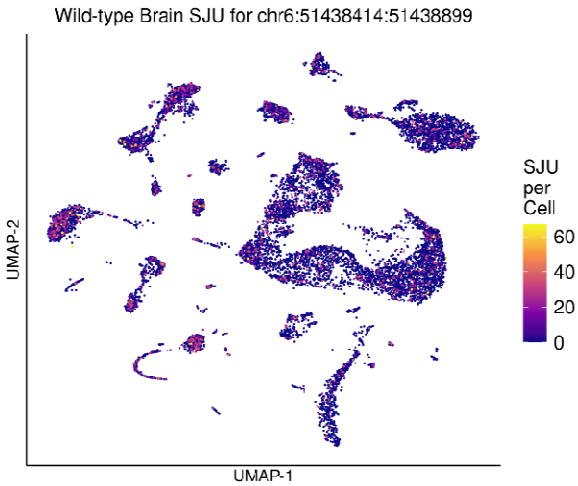


This UMAP displays the normalized and scaled gene expression values for *Hnrnpa2b1* in wild-type mouse brain tissue cells. A brighter color indicates a higher expression level. Download

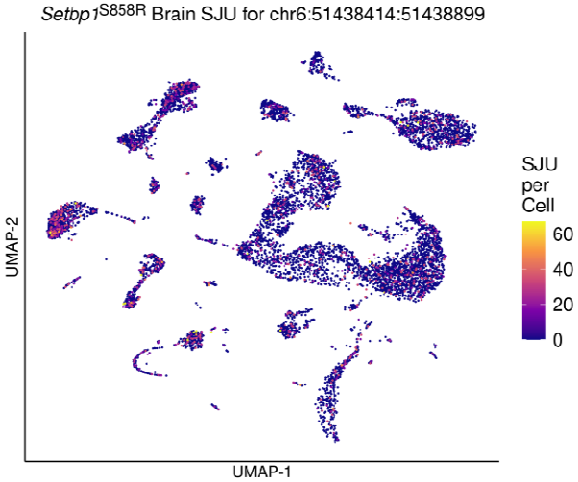


This UMAP displays the normalized and scaled gene expression values for *Hnrnpa2b1* in *Setbp1* ^{S858R} mouse brain tissue cells. A brighter color indicates a higher expression level. Download

C



This UMAP displays the splice junction usage (SJU) values for splice junction chr6:51438414:51438899 from *Hnrnpa2b1* in wild-type mouse brain tissue cells. A brighter color indicates a higher usage level. Please note that our manuscript does not use SJU values per cell, and SJU is a single number calculated for an entire population of cells, such as patient variant cells of a specific cell type. Download



This UMAP displays the splice junction usage (SJU) values for splice junction chr6:51438414:51438899 from *Hnrnpa2b1* in *Setbp1* ^{S858R} mouse brain tissue cells. A brighter color indicates a higher usage level. Please note that our manuscript does not use SJU values per cell, and SJU is a single number calculated for an entire population of cells, such as patient variant cells of a specific cell type. Download

Figure 5. Our web application enables users to access and visualize gene expression and SJU data for the cerebral cortex and the kidney. Representative screenshots of our web application. The “Gene Expression and SJU UMAPs” tab allows users to examine and download the UMAPs annotated by (A) cell types, (B) gene expression, and/or (C) SJU of any gene(s) of interest in our dataset in the cerebral cortex and/or kidney.

Discussion

In this study, we analyzed cell-type-specific gene and SJ expression and usage from cerebral cortex and kidney single nuclei profiles of *Setbp1*^{S858R} mice compared to age- and sex-matched control mice. We identified SJU differences in all cell types in the cerebral cortex and the kidney. We identified 33 and 62 genes with significant changes in SJU between the *Setbp1*^{S858R} and wild-type mice in the cerebral cortex and the kidney, respectively. We found that 82% of significant SJUs were cell-type-specific in the cerebral cortex, while only 53% of significant SJUs in the kidney were specific to a cell type. Additionally, 36 out of 84 total significant SJU genes across tissues have already been shown to be associated with various diseases, cancer, and autism spectrum disorder (Table S4). Given the emphasis of previous research in the field, specifically on neural progenitor cells and neurons[7,8,47], this study builds on our prior work[10], underscoring that SETBP1's role as an epigenetic hub[48] leads to cell-type-specific signatures in atypical SGS cell types. For example, while we reported SJU changes in all cerebral cortex cell types we measured, the excitatory and inhibitory neurons had the most significant SJUs, while astrocytes and oligodendrocytes had the most cell-type-specific SJU genes in *Setbp1*^{S858R} compared to wild-type mice. While *Setbp1* was broadly expressed across all cerebral cortex and kidney cell types, we did not identify significant SJU in any cell type across tissues, indicating that the S858R variant does not lead to changes in AS in *Setbp1* itself. Although the changes did not reach significance, we observed alterations in SJE and SJU for all cerebral cortex and kidney cell types across multiple SJs of *Setbp1,* which may be due to the limitations of short-read RNA-seq data for AS analyses and the current knowledge of Setbp1 isoforms, as only one is currently annotated in the mouse. Interestingly, we observed the greatest SJU increase in SJ-3, the junction spanning exon 3 to the variant-impacted exon 4 of Setbp1, in vascular cells – a cell type not previously implicated in SGS. Our results further implicate that disease-causing patient variants in *SETBP1* induce disease-associated molecular programs (here, differential SJU) in many cerebral cortex and kidney cell types, which are critical for understanding disease pathogenesis. As the brain has the most complex splicing profile of all tissues [49], and splicing is essential for healthy neurodevelopment [50], this finding is not surprising. However, to our knowledge, this is the first report of AS in the cerebral cortex and the kidney of a SETBP1-associated model.

Furthermore, we predicted significant SJU in the cerebral cortex and the kidney of the *Hnrnpa2b1* gene, encoding an RNA-binding protein of the heterogeneous nuclear ribonucleoprotein class. This class comprises a large family of RNA-binding proteins that contribute to alternative splicing, mRNA stabilization, and transcriptional and translational regulation [51]. Additionally, these proteins act compensatorily in disease, exhibiting high amino acid homology [43,44]. While genes encoding heterogeneous nuclear ribonucleoproteins are

ubiquitously expressed, genetic variants that alter their sequence lead to various neurodevelopmental and neurodegenerative diseases [45]. A human peripheral blood study from an SGS proband compared to their healthy parents identified 38 AS genes, including *Hnrnpa1*, the gene encoding a member of the heterogeneous nuclear ribonucleoprotein class. None of our genes with significant SJU in the cerebral cortex or the kidney directly overlapped with the 37 orthologs of the 38 AS genes identified in the human peripheral blood SGS study [11], which may be due to differences in species (human vs. mouse), sequencing technologies (bulk vs. snRNA-seq), variant locations (typical vs. atypical SGS), and tissue type (peripheral blood vs. cerebral cortex and kidney). However, we identified significant SJU in *Hnrnpa2b1*, another member of the heterogeneous nuclear ribonucleoprotein class, closely related to *Hnrnpa1* [42]. Intriguingly, we identified significant cell-type-specific SJU of *Hnrnpa2b1* in the cerebral cortex. Despite inhibitory neurons possessing the second-lowest gene expression of *Hnrnpa2b1* in the cerebral cortex cell types, they were the only cell type with significant SJU in this gene. Neurons, being non-dividing cells, require tight regulation of mRNA homeostasis, indicating their vulnerability to dysfunction of RNA-binding proteins [51]. Our findings further suggest that AS in members of the heterogeneous nuclear ribonucleoprotein class, like *Hnrnpa2b1,* may play a role in SGS's neurodegenerative phenotype. Interestingly, we also identified significant SJU of *Hnrnpa2b1* in the kidney of our *Setbp1*^{S858R} mice, although we observed AS in the kidney across multiple cell types, indicating tissue-specific AS effects in this protein family. Altogether, this suggests that AS in members of the heterogeneous nuclear ribonucleoprotein class may play a crucial role in SGS, both at the cellular and tissue levels.

To overcome challenges for researchers to quickly explore our data and analyses for themselves, we created a user-friendly point-and-click Shiny web application. Our web application can be used for custom visualization, gene exploration, comparative analysis (across conditions, tissues, cell types, genes, and SJs), and detailed isoform annotations. This resource includes the scaled and normalized gene counts for pre-processed *Setbp1*^{S858R} and wild-type cerebral cortex and kidney samples for 7 and 18 cell types, respectively. It also features pre-calculated SJU across more than 15,000 genes, easy plot downloads, and reference links for gene searches. By integrating these tools into rare disease projects, we can make these data sets more accessible to the broader community. This enhanced accessibility allows researchers and clinicians to interact with and understand complex genetic data more easily, fostering collaboration and accelerating discoveries in rare disease research.

Although our study provides additional insight into dysregulated regulatory mechanisms, specifically AS through SJU, there are several limitations to the current study. First, single-cell/nuclei and SJ read count data are sparse by nature. While we attempted to counteract this with high sequencing depth (~100,000 reads per nucleus), our SJU values were low compared to other publications that used the same computational framework[29]. Additionally, these prior works used cells, rather than nuclei, which may contribute to the differences we observed. Second, while useful, mouse models are not perfect substitutes for patient profiles because of species differences. For example, *Setbp1* has only one annotated transcript in mice, but there are seven annotated protein-coding transcripts in humans (Ensembl, GRCh38.p14)[13] and two known protein isoforms in UniProt[52]. We did not detect significant SJU changes for *Setbp1* in

the *Setbp1*^{S858R} model, but that does not preclude changes in *SETBP1* isoform expression at a different developmental stage or in patients. Finally, we generated sequencing profiles with the Illumina platform using short-read sequencing. While splicing can be measured in short-read RNA-seq by split reads (i.e., reads spanning SJs[53]), this approach does not typically measure full-length transcripts because the average mammalian transcript length is 2-3 kilobase pairs. For example, an increase in splice junction expression or reads between exons 2 and 4 of *Son* would suggest reduced expression of exon 3 in *Setbp1*^{S858R} mouse cerebral cortex. This would help identify potential isoforms expressed, but cannot be used to detect full-length transcript- or isoform- expression or usage. Therefore, our analysis was limited to SJ expression and SJU, rather than full-length transcript expression and usage, which may have generated false positives and/or missed splicing events.

To our knowledge, this work is the first to identify cell-type-specific AS patterns resulting from *Setbp1* patient variants in the cerebral cortex and kidney. Future studies incorporating additional SGS models and leveraging new technologies, such as long-read single-cell profiling capable of capturing full-length transcripts in an individual cell, are critical for further establishing the potential impact of AS in SGS. Additionally, while there are no documented sex differences in SGS, which is an ultra-rare Mendelian disorder, AS is known to change between sexes[54–58], and male and female SGS splicing profiles could be differentially affected by the accumulation of SETBP1; therefore, changes in additional time points and sexes should be explored. Additionally, although outside the scope of this study, future research is needed to determine how splicing may be impacted in a cell-specific manner, particularly to investigate the role of differences in *Hnrnpa2b1* SJU on disease pathogenesis in SETBP1-associated diseases. In conclusion, our findings further implicate the impact of *Setbp1* patient variants on a diverse range of cell types and molecular mechanisms in the cerebral cortex and the kidney, underscoring their importance in contributing to SGS.

Abbreviations

AS	Alternative splicing
CDS	Coding sequence
CDPC	Collecting duct principal cells
DCT	Distal convoluted tubule
FAQ	Frequently asked questions
LOH	Loop of Henle
OPCs	Oligodendrocyte precursor cells
NMD	Nonsense-mediated decay
PCA	Principal component analysis
PCT	Proximal convoluted tubule
PPI	Protein-protein interaction
PST	Proximal straight tubule
snRNA-Seq	single-nuclei RNA sequencing
SGS	Schinzel-Giedion Syndrome

SJ	Splice junction
SJE	Splice junction expression
SJU	Splice junction usage
UMAP	Uniform manifold approximation and projection

Acknowledgments

We would like to acknowledge all current and former members of the Lasseigne Lab for their thoughtful feedback, especially Amanda D. Clark and Vishal H. Oza. The graphical abstract was created using BioRender.

Competing Interests Statement

The authors declare they have no conflicts of interest.

Funding

This work was funded by the UAB Lasseigne Lab funds and the UAB Pilot Center for Precision Animal Modeling (C-PAM) (1U54OD030167).

Data and Resource Availability

The authors have used publicly available raw sequencing data from GEO at accession number GSE237816. Publicly available databases used for analyses include PanglaoDB[27], CellMarker 2.0[28], and the Allen Brain Cell Atlas at https://portal.brain-map.org/atlases-and-data/bkp/abc-atlas. All code, containers, and intermediate data objects can be accessed on GitHub, DockerHub, and Zenodo at the following links:
Github (analysis): https://github.com/lasseignelab/230926_EJ_Setbp1_AlternativeSplicing
Github (web application): https://github.com/lasseignelab/Setbp1_Alternative_Splicing_Shiny
Dockerhub:
https://hub.docker.com/repository/docker/emmafjones/setbp1_alternative_splicing/general
https://hub.docker.com/repository/docker/tsoelter/setbp1_alternative_splicing/general
https://hub.docker.com/repository/docker/tsoelter/sn-ml-drug-repurposing/general
Zenodo: GitHub (analysis): https://zenodo.org/records/17601458, GitHub (web application): https://zenodo.org/records/17653325, Docker: https://zenodo.org/records/12534825, Data: https://zenodo.org/records/17594386

Author Contributions Statement

Tabea M. Soelter: Conceptualization (supporting); data curation (equal); formal analysis (equal); investigation (equal); software – analysis (equal); software – web application (supporting); visualization (equal); validation (equal); writing – editing (equal); writing – original draft (equal). **Emma F. Jones:** Conceptualization (supporting); data curation (equal); formal analysis (equal); investigation (equal); methodology (lead); software – analysis (equal); software – web application (supporting); visualization (equal); writing – original draft (equal); writing – editing (equal). **Timothy C. Howton:** Supervision (supporting); validation (equal); writing - editing (equal). **Anthony B. Crumley:** Software – web application (lead); writing – editing (equal); validation (equal). **Elizabeth J. Wilk:** validation (equal); writing – editing (equal).

Brittany N. Lasseigne: Conceptualization (lead); funding acquisition (lead); project administration (lead); resources (lead); supervision (lead); writing – editing (equal).

Supplementary Information

SoelterJones2025_SupplementaryInformation.docx
SoelterJones2025_Supplementary_Table_S2.xls
SoelterJones2025_Supplementary_Table_S3.xls
SoelterJones2025_Supplementary_Table_S4.xls

References

1. Hoischen, A. *et al.* De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat. Genet.* **42**, 483–485 (2010).

2. Whitlock, J. H., Wilk, E. J., Howton, T. C., Clark, A. D. & Lasseigne, B. N. The landscape of SETBP1 gene expression and transcription factor activity across human tissues. *PLoS One* **19**, e0296328 (2024).

3. Duis, J. & van Bon, B. W. M. Schinzel-Giedion Syndrome. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, Seattle (WA), 2024).

4. Schinzel, A. & Giedion, A. A syndrome of severe midface retraction, multiple skull anomalies, clubfeet, and cardiac and renal malformations in sibs. *Am. J. Med. Genet.* **1**, 361–375 (1978).

5. Liu, W.-L., He, Z.-X., Li, F., Ai, R. & Ma, H.-W. Schinzel–Giedion syndrome: a novel case, review and revised diagnostic criteria. *J. Genet.* **97**, 35–46 (2018).

6. Acuna-Hidalgo, R. *et al.* Overlapping SETBP1 gain-of-function mutations in Schinzel-Giedion syndrome and hematologic malignancies. *PLoS Genet.* **13**, e1006683 (2017).

7. Banfi, F. *et al.* SETBP1 accumulation induces P53 inhibition and genotoxic stress in neural progenitors underlying neurodegeneration in Schinzel-Giedion syndrome. *Nat. Commun.* **12**, 4050 (2021).

8. Zaghi, M. *et al.* Balanced SET levels favor the correct enhancer repertoire during cell fate acquisition. *Nat. Commun.* **14**, 3212 (2023).

9. 033235 - Setbp1[S858R] Strain Details. https://www.jax.org/strain/033235.

10. Whitlock, J. H. *et al.* Cell-type-specific gene expression and regulation in the cerebral cortex and kidney of atypical Setbp1S858R Schinzel Giedion Syndrome mice. *J. Cell. Mol. Med.* (2023) doi:10.1111/jcmm.18001.

11. Liu, L. *et al.* Whole-genome sequencing combined RNA-sequencing analysis of patients with mutations in SET binding protein 1. *Front. Neurosci.* **16**, 980000 (2022).

12. Chen, M. & Manley, J. L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* **10**, 741–754 (2009).

13. Gene: SETBP1 (ENSG00000152217) - Summary - Homo_sapiens - Ensembl genome browser 112. https://useast.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000152217;r=18:44680173-45068510.

14. Montemayor, E. J. *et al.* Molecular basis for the distinct cellular functions of the Lsm1-7 and Lsm2-8 complexes. *RNA* **26**, 1400–1413 (2020).

15. Tanis, S. E. J. *et al.* Splicing and Chromatin Factors Jointly Regulate Epidermal Differentiation. *Cell Rep.* **25**, 1292–1303.e5 (2018).

16. Zhang, X. *et al.* Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. *Cell* **166**, 1147–1162.e15 (2016).

17. Patowary, A. *et al.* Developmental isoform diversity in the human neocortex informs neuropsychiatric risk mechanisms. *Science* **384**, eadh7688 (2024).

18. Huang, D.-F. *et al.* Neuronal splicing regulator RBFOX3 mediates seizures via regulating *Vamp1* expression preferentially in NPY-expressing GABAergic neurons. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2203632119 (2022).

19. Chang, W. *et al. Web Application Framework for R [R Package Shiny Version 1.10.0]*. (Comprehensive R Archive Network (CRAN), 2024).

20. Kaminow, B., Yunusov, D. & Dobin, A. STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. *bioRxiv* 2021.05.05.442755 (2021) doi:10.1101/2021.05.05.442755.

21. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, (2020).

22. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* (2023) doi:10.1038/s41587-023-01767-y.

23. Why do I see high levels of Malat1 in my gene expression data? *10X Genomics* https://kb.10xgenomics.com/hc/en-us/articles/360004729092-Why-do-I-see-high-levels-of-Malat1-in-my-gene-expression-data.

24. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

25. Luecken, M. D. *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50 (2022).

26. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).

27. Franzén, O., Gan, L.-M. & Björkegren, J. L. M. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database* **2019**, (2019).

28. Hu, C. *et al.* CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res.* **51**, D870–D876 (2023).

29. Wen, W. X., Mead, A. J. & Thongjuea, S. MARVEL: an integrated alternative splicing analysis platform for single-cell RNA sequencing data. *Nucleic Acids Res.* **51**, e29 (2023).

30. Sondka, Z. *et al.* The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* **18**, 696–705 (2018).

31. Banerjee-Basu, S. & Packer, A. SFARI Gene: an evolving database for the autism research community. *Dis. Model. Mech.* **3**, 133–135 (2010).

32. Abrahams, B. S. *et al.* SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* **4**, 36 (2013).

33. DiStefano, M. T. *et al.* The Gene Curation Coalition: A global effort to harmonize gene-disease evidence resources. *Genet. Med.* **24**, 1732–1742 (2022).

34. Chang, W. *et al.* shiny: Web Application Framework for R. Preprint at https://CRAN.R-project.org/package=shiny (2022).

35. Dopp, J. *et al.* Single-cell transcriptomics reveals that glial cells integrate homeostatic and circadian processes to drive sleep-wake cycles. *Nat. Neurosci.* **27**, 359–372 (2024).

36. Feng, H. *et al.* Complexity and graded regulation of neuronal cell-type-specific alternative splicing revealed by single-cell RNA sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

37. Huntley, M. A. *et al.* Genome-wide analysis of differential gene expression and splicing in excitatory neurons and interneuron subtypes. *J. Neurosci.* **40**, 958–973 (2020).

38. Koopmans, F. *et al.* SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron* **103**, 217–234.e4 (2019).

39. Indelicato, E. *et al.* SYNE1-ataxia: Novel genotypic and phenotypic findings. *Parkinsonism Relat. Disord.* **62**, 210–214 (2019).

40. Gene: SYNE1 (ENSG00000131018) - Summary - Homo_sapiens - Ensembl genome browser 112. https://useast.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000131018;r=6:152121687-152637801.

41. SYNE2 protein expression summary - The Human Protein Atlas. https://www.proteinatlas.org/ENSG00000054654-SYNE2.

42. He, Y. & Smith, R. Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. *Cell. Mol. Life Sci.* **66**, 1239–1256 (2009).

43. Lu, Y. *et al.* Heterogeneous nuclear ribonucleoprotein A/B: an emerging group of cancer biomarkers and therapeutic targets. *Cell Death Discov.* **8**, 337 (2022).

44. Mayeda, A., Munroe, S. H., Cáceres, J. F. & Krainer, A. R. Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J.* **13**, 5483–5495 (1994).

45. Tilliole, P., Fix, S. & Godin, J. D. hnRNPs: roles in neurodevelopment and implication for brain disorders. *Front. Mol. Neurosci.* **17**, 1411639 (2024).

46. Wang, H. *et al.* Identification of a novel de novo mutation of SETBP1 and new findings of SETBP1 in tumorgenesis. *Orphanet J. Rare Dis.* **18**, 107 (2023).

47. Cardo, L. F., de la Fuente, D. C. & Li, M. Impaired neurogenesis and neural progenitor fate choice in a human stem cell model of SETBP1 disorder. *Mol. Autism* **14**, 8 (2023).

48. Piazza, R. *et al.* SETBP1 induces transcription of a network of development genes by acting as an epigenetic hub. *Nat. Commun.* **9**, 2192 (2018).

49. Xu, Q., Modrek, B. & Lee, C. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res.* **30**, 3754–3766 (2002).

50. Sanders, S. J., Schwartz, G. B. & Farh, K. K.-H. Clinical impact of splicing in neurodevelopmental disorders. *Genome Med.* **12**, 36 (2020).

51. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Hum. Genet.* **135**, 851–867 (2016).

52. UniProt. https://www.uniprot.org/uniprotkb/Q9Y6X0/entry.

53. García-Ruiz, S. *et al.* Splicing accuracy varies across human introns, tissues, age and disease. *Nat. Commun.* **16**, 1068 (2025).

54. Trabzuni, D. *et al.* Widespread sex differences in gene expression and splicing in the adult human brain. *Nat. Commun.* **4**, 2771 (2013).

55. Blekhman, R., Marioni, J. C., Zumbo, P., Stephens, M. & Gilad, Y. Sex-specific and lineage-specific alternative splicing in primates. *Genome Res.* **20**, 180–189 (2010).

56. Karlebach, G. *et al.* The impact of biological sex on alternative splicing. *bioRxiv* 490904 (2020) doi:10.1101/490904.

57. Sosnowski, B. A., Belote, J. M. & McKeown, M. Sex-specific alternative splicing of RNA from the transformer gene results from sequence-dependent splice site blockage.

Cell **58**, 449–459 (1989).

58. Han, Y. *et al.* Proteogenomics reveals sex-biased aging genes and coordinated splicing in cardiac aging. *Am. J. Physiol. Heart Circ. Physiol.* **323**, H538–H558 (2022).