Cell-type-specific alternative splicing in the cerebral cortex of a Schinzel-Giedion Syndrome patient variant mouse model

Emma F. Jones[1], Timothy C. Howton[1], Tabea M. Soelter[1], Anthony B. Crumley[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

Email Addresses

Emma F. Jones: efjones@uab.edu, Timothy C. Howton: tchowton@uab.edu, Tabea M. Soelter: tsoelter@uab.edu, Anthony B. Crumley: acrumley@uab.edu, Brittany N. Lasseigne: bnp0001@uab.edu (*corresponding author)

Data Availability Statement

The authors have used publicly available raw sequencing data from GEO at accession number GSE237816. All code, containers, and intermediate data objects can be accessed on GitHub, DockerHub, and Zenodo at the following links:
Github: https://github.com/lasseignelab/230926_EJ_Setbp1_AlternativeSplicing
Dockerhub:
https://hub.docker.com/repository/docker/emmafjones/setbp1_alternative_splicing/general
Zenodo:
GitHub Repo: 10.5281/zenodo.12548384
Docker: 10.5281/zenodo.12534825
Data: 10.5281/zenodo.12535061

Funding Statement

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

Conflict of Interest Statement

The authors declare they have no conflicts of interest.

Ethics Approval Statement

Not applicable

Patient Consent Statement

Not Applicable

Consent for publication

Not Applicable

Clinical Trial Registration

Not Applicable

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 used STARsolo to quantify gene and splice junction (SJ) expression for 51,465 nuclei previously generated from the cerebral cortex of atypical *Setbp1*^{S858R} SGS patient variant mice (n = 3) and wild-type control mice (n = 3). After cell type annotation, we performed pseudobulk differential gene expression and SJ usage (SJU) analyses across cell types and conditions. We identified 34 genes with statistically significant alterations in SJU. Oligodendrocytes had the most genes with changes in SJU, followed by astrocytes, excitatory, and inhibitory neurons. One gene, *Son*, a splicing cofactor known to cause the neurodevelopmental disorder ZTTK Syndrome, had SJU changes in all six non-vascular cell types we measured in *Setbp1*^{S858R} compared to controls. This is the first research to report AS changes in the cerebral cortex of an SGS model and the first study to link SGS to perturbations in *Son*.

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 (TF) 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} 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 controls, 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]. Additionally, a recent study identified 38 alternatively spliced (AS) genes in the peripheral blood of a typical SGS patient using bulk RNA-seq[11]. Moreover, *SETBP1* itself is known to undergo AS in humans[12], and two SETBP1 targets, *LSM2* and *ZMAT2,* are splicesome and pre-splicesome components[13,14]. As cell-type-specific AS is essential for neurodevelopment[15], neurodevelopmental disorders[16], and epilepsy[17], we hypothesized that AS was altered in the cerebral cortex of *Setbp1*^{S858R} mice compared to controls.

Here, we re-processed our previously generated *Setbp1*^{S858R} patient variant mouse snRNA-Seq data[10] to examine the role of AS in atypical SGS (**Figure 1A**). We compared splice junction usage (SJU) in the cerebral cortex of *Setbp1*^{S858R} mice and age-matched controls and identified cell-type-specific genes and pathways with AS between conditions (**Figure 1B-E**). While we did not identify cell-type-specific AS in *Setbp1*, we found that the splicing cofactor *Son,* which causes the Mendelian disease ZTTK syndrome characterized by delayed psychomotor development and developmental delay[18], had common AS alterations between *Setbp1*^{S858R} and controls in excitatory neurons, inhibitory neurons, astrocytes, oligodendrocytes, OPCs, and microglia. Our findings further implicate AS in SGS by identifying cell-type-specific and -shared genes and pathways altered in *Setbp1*^{S858R} cerebral cortex.

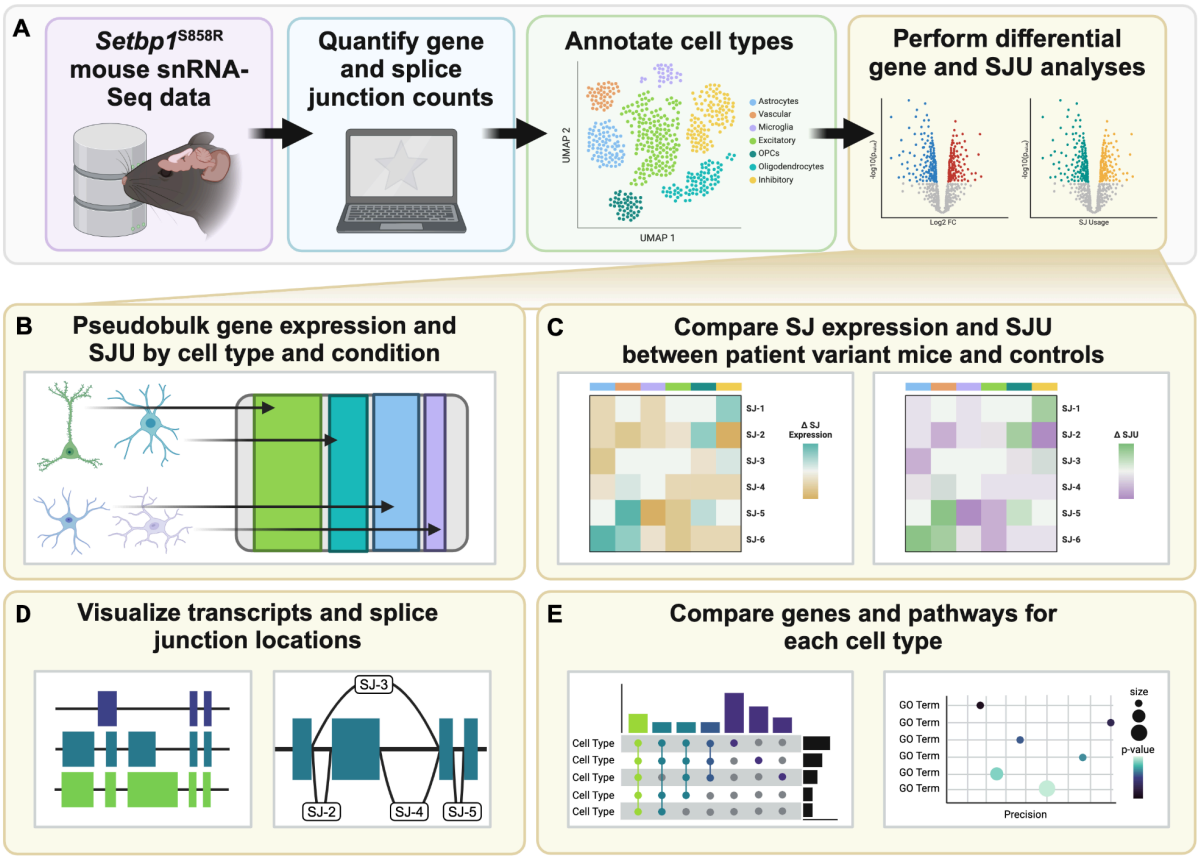

Figure 1. Graphical Abstract. (A) Schematic overview of our processing and analysis pipeline. (B) We analyzed pseudobulk gene expression and calculated SJU for each cell type and condition. (C) We compared SJU values for each cell type using a permutation test (Methods) to identify cell-type-specific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

Materials and Methods

<u>Data acquisition</u>

We obtained raw snRNA-Seq data of six male mouse brain cerebral cortex samples: three 6-week-old C57BL/6J-*Setbp1*^{em2Lutzy}/J mice heterozygous for *Setbp1*^{S858R} (JAX #033235) and three healthy, age-matched controls (JAX #000664) from our recent publication[10], available on GEO from accession number GSE237816.

<u>Data processing</u>

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. In that conda environment, we used STAR version 2.7.10b[19] to build a STAR genome reference from

the GENCODE M31 primary assembly. Next, we ran STARsolo version 2.7.10b [19] 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.

Container reproducibility

For all the analyses following data processing, we used R version 4.3.1 through R Studio version 2023.06.2+561 running in docker containers through Singularity version 3.5.2 on the UAB supercomputer cluster, Cheaha. Docker images for each analysis are specified in each script and available at https://hub.docker.com/repository/docker/emmafjones/setbp1_alternative_splicing/general

Data quality control and filtering

We used the Seurat version 5.0.0 R package for quality control, filtering, and clustering analyses[20]. We imported the gzipped and filtered STARsolo output matrices (barcodes, features, and matrix) into R using the Seurat Read10x function[20]. We created a Seurat object using the CreateSeuratObject function for each sample and condition (*Setbp1*^{S858R} patient variant and wild-type) before merging them into a single Seurat object. 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* since it is frequently over-detected with poly-A capture technologies such as 10x[21]. Then, using default parameters, we performed batch correction using harmony version 1.1.0[22] to preserve biological variation while reducing variation due to technical noise[23]. 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[20] without approximation (approx = FALSE) in order to improve reproducibility. We plotted UMAPs and cell-type proportions to confirm successful integration across conditions.

Clustering and cell type assignment

We also used the Seurat R package (version 5.0.0)[20] for clustering and cell type assignment. We used a clustering resolution of 0.75, identifying 32 clusters using the Leiden option, which relies upon leidenalg v0.10.1[24]. We identified differentially expressed marker genes for each cluster (using the FindAllMarkers from Seurat[20]) 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[25], CellMarker 2.0[26], 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 (**Supporting Information Figure 1**). Due to non-specific cell markers and low cell numbers, we combined pericytes, vascular leptomeningeal, and endothelial cells into one vascular cell type.

We also sub-clustered cluster 22 to differentiate between oligodendrocytes and astrocytes using the FindSubCluster function from Seurat[20] with a resolution of 0.05.

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

To calculate SJU, we divided the total number of SJ counts for a cell type for each SJ and divided by the total number of read counts 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 the R package MARVEL version 2.0.5[27] to integrate SJ and gene expression information into a single R object. We filtered only to include 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. 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 a MARVEL object 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. We compared conditions (patient variant mice and wild-type controls) for each cell type. 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.

Pseudobulk differential gene expression with DESeq2

We also analyzed pseudobulk differential gene expression between patient variant mice and wild-type controls for each cell type using DESeq2 version 1.40.2[28] and apeglm's log2FC shrinkage estimation algorithm[29]. We reported significantly differentially expressed genes as

those with an adjusted p-value of less than 0.05 and a log2 fold change greater than or equal to an absolute value of 0.5.

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 *Setbp1* and *Son.* 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 prime to 3 prime direction of translation.

Functional enrichment analysis

For functional enrichment analysis with the R package gprofiler2[30] version 0.2.3, we used the genes for each cell type with differential SJU between patient variant mice and wild-type controls. For background genes, we used all expressed genes in our dataset. We used the gost (gene ontology search term) function, which uses a one-tailed Fisher's exact test, with a g_SCS corrected p-value cutoff of 0.05.

Results

Cerebral cortex 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 of C57BL/6J-*Setbp1*^{em2Lutzy}/J mice heterozygous for *Setbp1*^{S858R} and matched controls and quantified gene and SJ expression using STAR[19]. After processing with STAR, we had 53,175 nuclei (**Methods**). We retained 51,465 nuclei that passed quality control filtering (**Methods**). From there, we annotated cell types (**Supporting Information Figure 2A, B**) using canonical marker gene expression to get seven cell types: astrocytes, excitatory neurons, inhibitory neurons, microglia, oligodendrocytes, oligodendrocyte precursor cells (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 = 33,095; **Supporting Information Figure 2A**). Cell types were evenly distributed across *Setbp1*^{S858R} and wild-type mice (**Supporting Information Figure 2C, D**), underscoring high dataset quality and sample processing consistency.

After annotating cell types, we integrated SJ count information from STAR into our analyses using MARVEL[27]. 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 (**Table S2, Supporting Information Figure 3A, B**). This is consistent with other research, as the fact that neurons express more genes than glia has previously been reported[31]. However, neurons still express more SJs per cell than glial cell types when dividing by the total number of genes expressed (**Supporting Information Figure 3C**), suggesting neuron-specific genes may be more

transcriptionally complex. This also agrees with previous findings that neuron-specific synaptic genes are longer and have more transcript isoforms[32]. The gene we detected with the most SJs 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 mutations can cause multiple Mendelian disorders, including cerebellar ataxia, Emery-Dreifuss muscular dystrophy, and arthrogryposis multiplex congenita[33]. Our results corroborate previous findings that *Syne1* has over 100 exons, resulting in dozens of transcript variants[34]. We saw striking cell-type-specific differences in the number of detected SJs for *Syne1*: we detected 156 SJs in excitatory neurons (the most) but only 29 SJs in microglia (the least). This highlights the cell-type-specificity of SJ expression and, thus, alternative splicing in our single-nuclei dataset.

Setbp1 gene expression, splice junction expression, and splice junction usage are not significantly different in *Setbp1*^{S858R} mice compared to wild-type

Next, we examined *Setbp1* gene expression and SJ usage. *Setbp1* has one annotated transcript in mice, *Setbp1-201* (**Figure 2A**), with six exons and five SJs (**Table 1**). We detected more counts for SJs 1 and 2 at the beginning of the transcript (5′ end) (**Figure 2A**). *Setbp1* was expressed in all cell types and was highest in neurons (mean normalized expression in excitatory and inhibitory neurons was 1.752 and 1.736, and in astrocytes, oligodendrocytes, OPCs, and microglia was 1.175, 0.658, 1.392, and 0.347, respectively; **Figure 2B**), consistent with our previous work[10]. Additionally, *Setbp1* was not significantly differentially expressed (adj p = 0.851 - 0.999, log2FC = -0.04 - 0.0001) between controls and patient variant mice in any cell type. Even with the 3′ bias in our data, we detected reads for all five SJs (**Supporting Information Figure 4**). We observed the largest magnitude of change between *Setbp1*^{S858R} and WT in SJs 2 and 3 (normalized SJ expression was +0.01 and -0.01; **Figure 2C**). None of the changes in SJU for any *Setbp1* SJs or cell type between conditions were significant (permutation test, $p > 0.05$). We identified the largest overall difference (+0.01) in SJU for SJ-2 (i.e., the splice junction between exon 1 and 2), specifically in microglia (+0.3) and oligodendrocytes (-0.2) and interestingly, these changes were in opposite directions for the two cell types (**Figure 2D**).

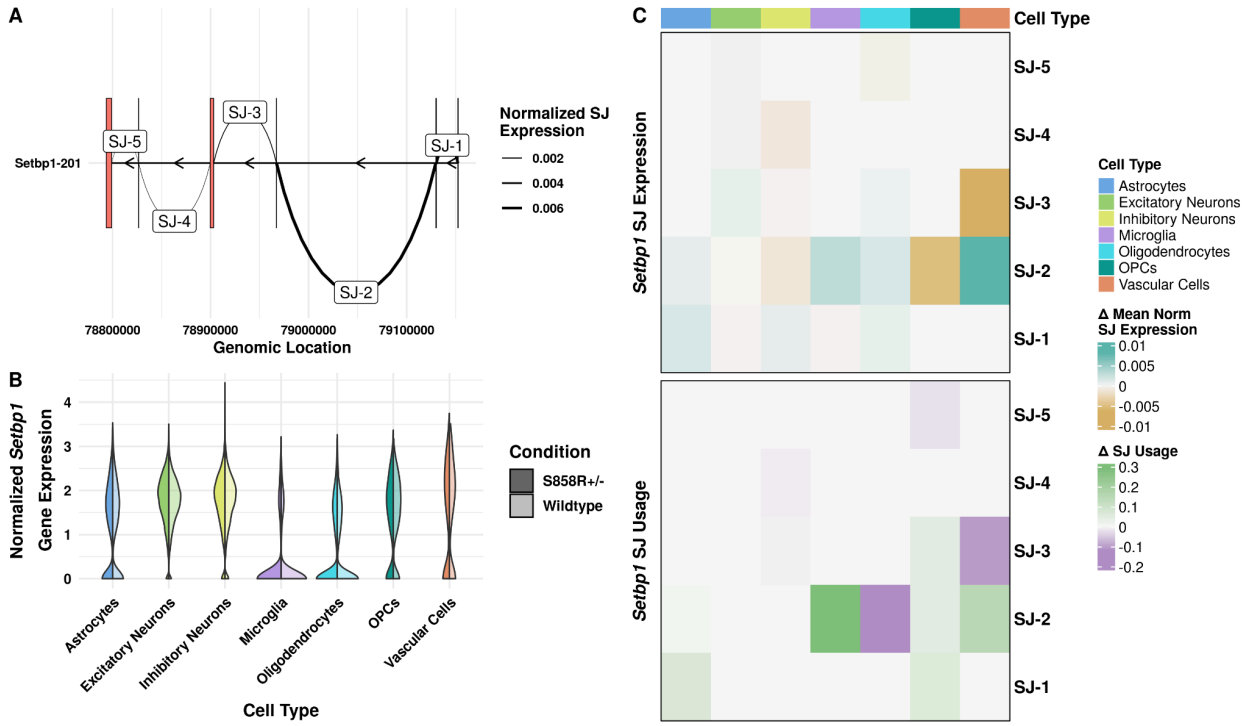

Figure 2. *Setbp1* gene and SJ expression. (A) Transcript diagram showing the transcript structure of *Setbp1* in mouse. SJs are labeled on the curved lines connecting exons. Arrows indicate the direction of transcription (which for *Setbp1* is antisense). Connecting line thickness indicates the mean normalized SJ expression for a given SJ. (B) Split violin plots showing normalized *Setbp1* expression across cell type (x-axis) and condition (darker shade represents patient variant mice, and lighter shade represents wild-type mice). (C) Heatmaps showing the changes in normalized mean SJ expression (top) and usage (bottom) between patient variant mice and wild-type controls for the five different *Setbp1* SJs. The top annotation of columns shows cell type. A positive delta value indicates expression or usage was higher in the variant than in wild-type mouse brain tissue, and a negative indicates expression or usage was higher in controls.

Table 1. SJ locations of *Setbp1.* All genomic locations of mouse *Setbp1* annotated in GENCODE M31 on chromosome 18.

SJ Short Name	SJ Start	SJ End
SJ-1	79130396	79152402
SJ-2	78967258	79129765
SJ-3	78826645	78967203
SJ-4	78826645	78899698
SJ-5	78799041	78826473

Setbp1^{S858R} mice have cell-type-specific splicing in all measured cerebral cortex cell types

To determine significant changes in SJU in other genes, we performed a permutation analysis (**Methods**) using the MARVEL[27] R package for each cell type between *Setbp1*^{S858R} and controls. We detected 34 genes with significant changes in SJU (permutation test, $p < 0.05$ and delta > 1 – see **Methods** for details). The cell type with the most genes with significant SJU was oligodendrocytes (n = 10, permutation test, p-val < 0.05), followed by astrocytes (n = 9), excitatory neurons (n = 9), and inhibitory neurons (n = 9) (**Figure 3**). Our SJU differences were primarily cell-type-specific (n = 26), where astrocytes had the most cell-type-specific genes (n = 8), followed by oligodendrocytes (n = 6) (**Figure 3**). Our 34 significant SJU genes did not overlap with the 37 mouse orthologs of genes previously reported with AS in whole blood of an SGS patient proband compared to their non-SGS parents[11].

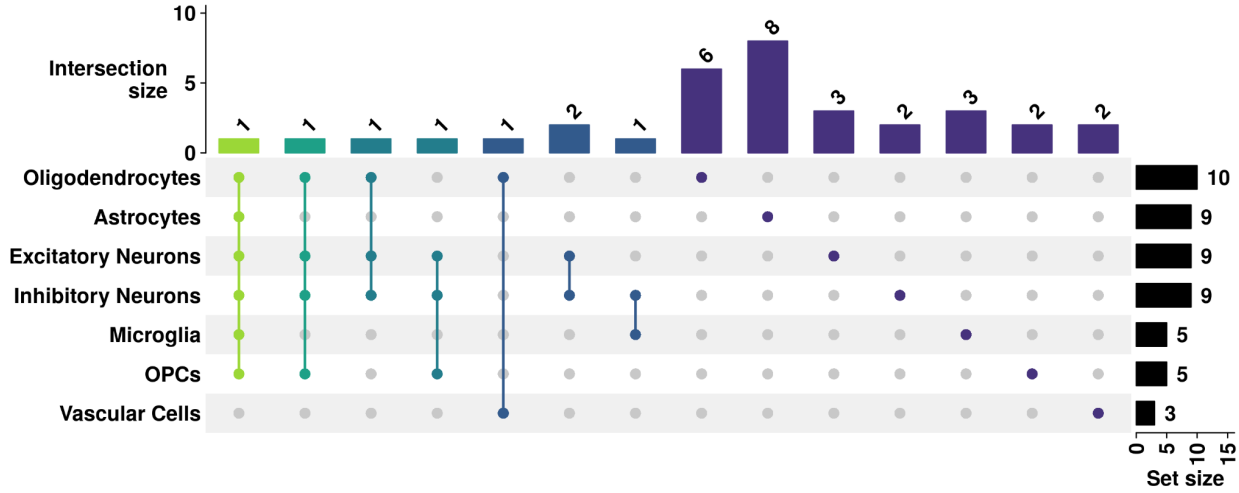

Figure 3. *Setbp1*^{S858R} patient variant mice have cell-type-specific splicing in all cell types. UpSet plot of genes with significant SJU changes between patient variant mice and wild-type controls split by cell type. Brighter colors indicate higher overlap between cell types, with yellow-green denoting genes shared by six cell types to 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.

We performed functional enrichment analyses of the genes with significant SJU between conditions by cell type using gprofiler2[30]. The cell types that had the most enriched terms were oligodendrocytes and excitatory neurons (**Figure 4A-B**, n = 5 each, Fisher's exact test, adj. $p < 0.05$). Oligodendrocyte SJU gene-associated terms included the iron-related terms ferroptosis, mineral absorption, and iron homeostasis, as well as membrane trafficking and vesicle-mediated transport (**Figure 4A**). The significant enrichment terms related to SJU genes for excitatory neurons were nuclear speck, adherens junction, and three microRNAs (miR-296-3p, miR-340-5p, and miR-19b-3p) (**Figure 4B**). Additionally, inhibitory neuron SJU genes were also enriched (n = 2, Fisher's exact test, adj. $p < 0.05$) for the microRNA miR-19b-3p term in addition to the Spg33-Rtn complex (**Figure 4C**), suggesting this miRNA may have altered

neuron-specific (i.e., in both excitatory and inhibitory) activity in *Setbp1*^{S858R} compared to control. Previous studies have shown in mouse hippocampal neurons that miR-19b-3p is more highly expressed under chronic restraint stress[35] than controls, and changes in AS with genes associated with this microRNA may indicate a neuronal stress response. We also identified terms from significant SJU genes for pathways and miRNAs in astrocytes, OPCs, and vascular cells (n = 3, 1, 1, Fisher's exact test, adj. $p < 0.05$) (**Figure 4D-F**). We detected that astrocyte SJU genes were enriched for the microRNAs miR-7234-5p, miR-7214-5p, and miR-329-5p, while OPCs were enriched for miR-340-5p, and vascular cells were enriched for the eicosanoid lipid synthesis map term (**Figure 5D-F**). Finally, we detected no pathway enrichment in microglia-significant SJU genes. These findings support that cell-type-specific SJU programs are altered in the cerebral cortex of *Setbp1*^{S858R} mice.

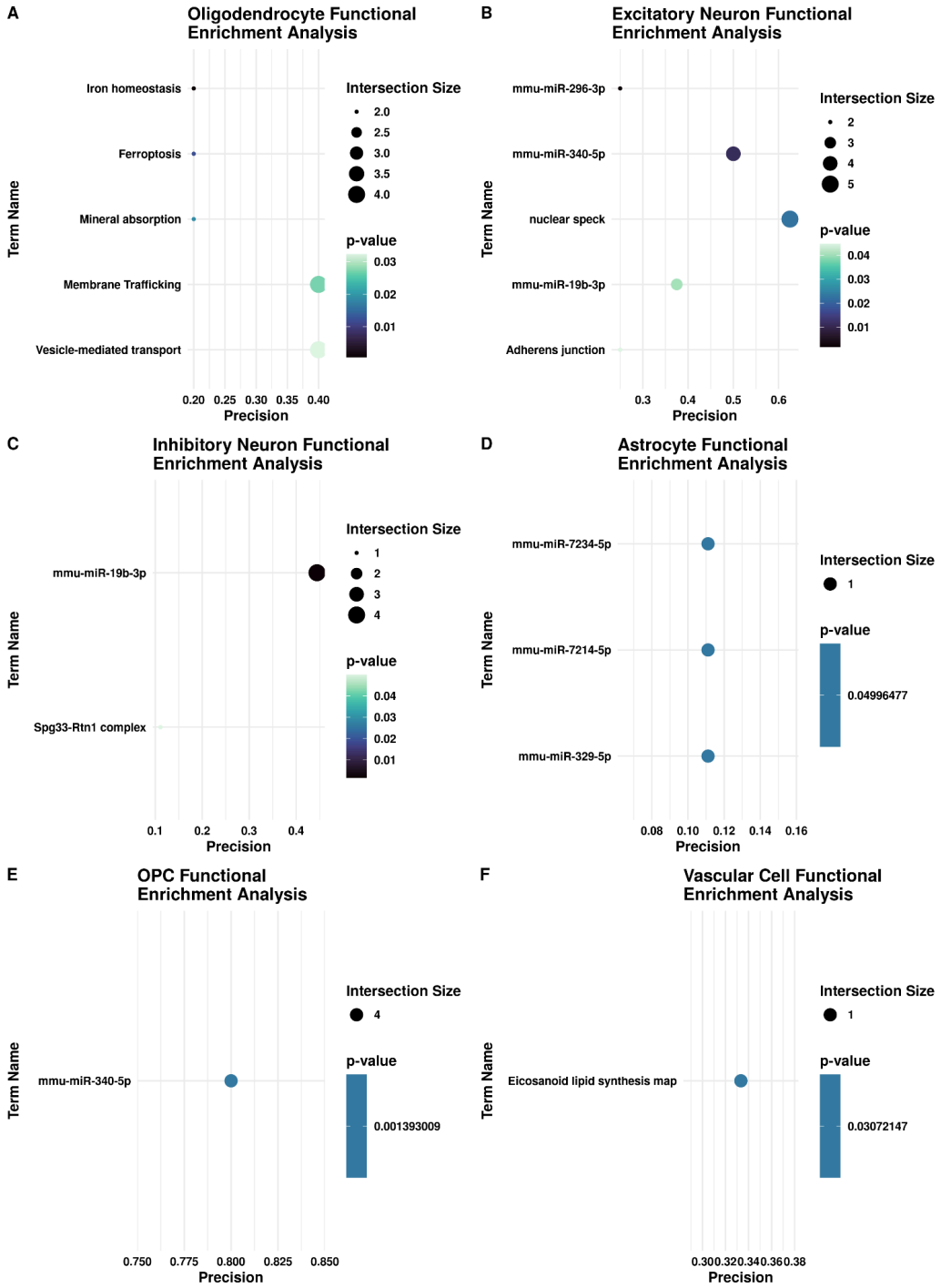

Figure 4. *Setbp1*^{S858R} patient variant mice have cell-type-specific pathway enrichment in six cell types. (A-F) Dot plots of significantly enriched pathways associated with genes with significant SJU changes in (A) oligodendrocytes, (B) excitatory neurons, (C) inhibitory neurons, (D) astrocytes, (E) oligodendrocyte precursor cells, and (F) vascular cells. Dot size indicates a larger intersection of terms, and darker color denotes a lower adjusted p-value.

Setbp1^{S858R} mouse cerebral cortex had significant SJU in *Son* in non-vascular cell types

Son, a splicing cofactor, was the only gene with significant SJU changes between the patient variant and control mice in all of the cerebral cortex non-vascular cell types we measured, i.e., excitatory neurons, inhibitory neurons, astrocytes, oligodendrocytes, OPCs, and microglia (**Figure 3**). *Son* has 14 annotated transcripts in mice (**Figure 5A**), five of which are protein-coding with a defined coding sequence (CDS). While expressed in all cell types, there were no significant differences in *Son* pseudobulk gene expression between the *Setbp1*^{S858R} and controls in any cell type. Still, it did have a higher normalized mean expression in non-neuronal cell types (>2) than in neuronal cell types (<1.5) (**Figure 5C**). Specifically, SJ-2, which includes *Son*'s largest exon, exon 3 (**Table S2,** SJ location visualization available in **Supporting Information Figure 5**), has lower SJU in all *Setbp1*^{S858R} mice cell types compared to control (change in mean normalized expression ranged between -0.305 and -0.103) and is significantly reduced in all 6 of the non-vascular cell types ($p < 0.05$, delta > 1; **Figure 5D**). This suggests several cell types are expressing an alternative isoform in SGS mice, such as *Son-205*, which is missing exon 3 (**Figure 5A, B**). The next most significant SJ, SJ-1 (mean normalized SJ expression = 0.7, **Table S1**), also had significantly increased usage in astrocytes and oligodendrocytes in the *Setbp1*^{S858R} cerebral cortex. Given the significant changes in SJU for all non-vascular cell types in this study, *Son* may be a potential contributor to SGS molecular phenotypes.

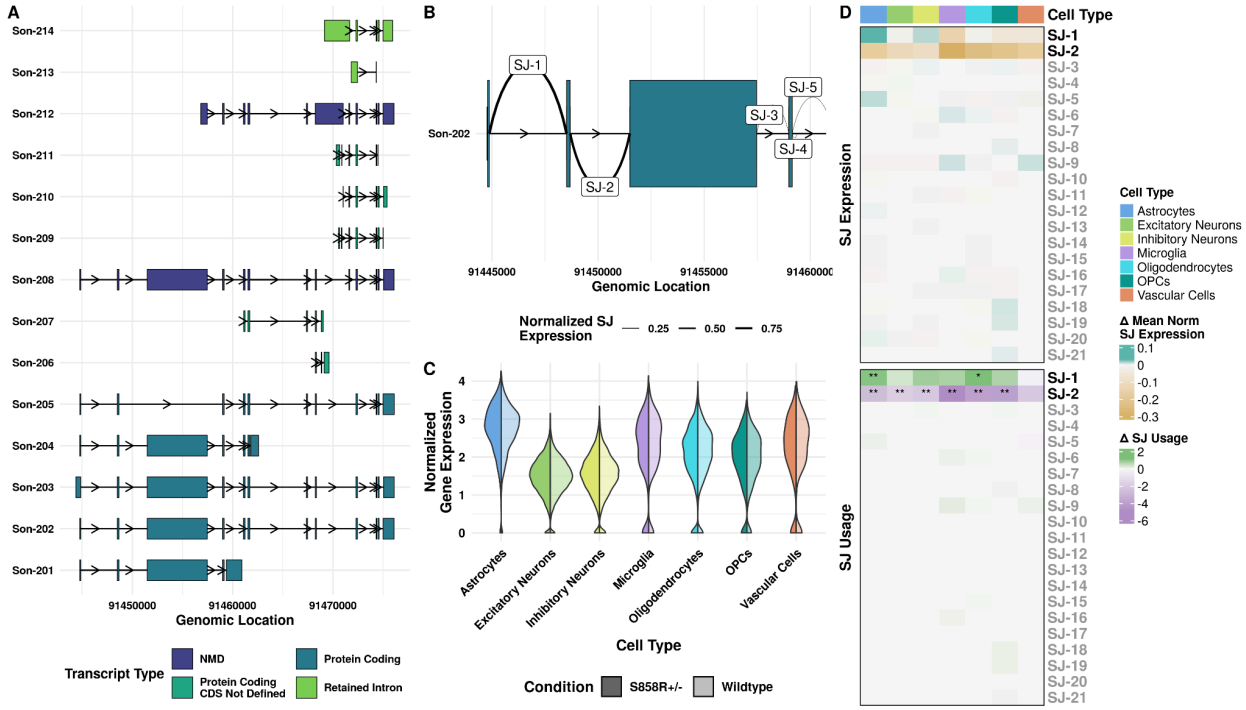

Figure 5. *Son* has AS changes across *Setbp1*^{S858R} patient variant mice and controls in all non-vascular cell types. (A) Transcript of all 14 annotated transcripts of *Son*. The color indicates transcript classification: indigo = transcripts flagged for nonsense-mediated decay (NMD), dark teal = protein-coding transcripts, turquoise = protein-coding transcripts, but coding

sequence (CDS) is not defined, and green = transcripts with retained intron events. Arrows indicate the direction of transcription. (B) Zoomed-in transcript diagram of Son-202, which includes exons 1-4. Connecting line thickness corresponds to the mean normalized SJ expression. (C) Split violin plots showing *Son* gene expression per cell for all cell types, split by condition. (D) Heatmaps of the changes in normalized mean SJ expression (top) and usage (bottom) between patient variant mice and wild-type controls for 21 SJs of *Son*. The top heatmap annotation indicates cell type. A positive delta indicates expression or usage was higher in variant mice than wild-type controls, and a negative indicates expression or usage was higher in controls. Asterisks designate significant changes in SJU; $p < 0.05$ = *, $p < 0.01$ = **.

Discussion

In this study, we analyzed cell-type-specific gene and SJ expression and usage from cerebral cortex single nuclei profiles of *Setbp1*^{S858R} mice compared to age- and sex-matched control mice. We identified SJU differences in the 7 cell types we examined, including 34 genes with significant changes in SJU between the *Setbp1*^{S858R} and controls. We found that 26 of these genes had significant SJU that were cell-type-specific in *Setbp1*^{S858R} samples. Additionally, we found multiple cell-type-specific pathways for significant SJU genes in six cell types by functional enrichment analysis. Given the emphasis of previous research in the field, specifically on neural progenitor cells and neurons[7,8,36], this study builds on our prior work[10], underscoring that SETBP1's role as an epigenetic hub[37] 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, including both excitatory and inhibitory neurons, here we report the most significant SJU differences in *Setbp1*^{S858R} compared to controls were in oligodendrocytes, and the most cell-type-specific SJU was in astrocytes. Our results further implicate that disease-causing patient variants in *SETBP1* induce disease-associated molecular programs (here, differential SJU) in many brain cell types, critical for understanding disease pathogenesis. As the brain has the most complex splicing profile of all tissues[38] and splicing is essential for healthy neurodevelopment[39], this is not surprising but is the first time AS in the cerebral cortex of SETBP1-associated models has been reported.

Also of particular note, we identified that *Son,* which codes for a splicing cofactor[40], has consistent and significant SJU in all non-vascular *Setbp1*^{S858R} cell types, potentially indicating reduced exon 3 usage. Mutations in *SON* are known to cause ZTTK Syndrome, a severe Mendelian neurodevelopmental disorder[18]. As a splicing cofactor, the SON protein promotes AS in many genes essential for neurodevelopment and cell cycle progression[40,41]. Additionally, *Son* knockdown causes defects in neuronal migration in mice, further emphasizing its role in healthy neurodevelopment[42]. SETBP1 accumulation may disrupt splicing machinery, perhaps through one of the many dysfunctional cellular processes associated with it, including chromatin remodeling, DNA damage, cell cycling, or phosphatase activity[43]. Because *SON* is essential for splicing and cell cycle control[40], it is possible that small, sustained changes in *SON* across these cell types could ultimately impact disease phenotype manifestation.

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[27]. Second, while useful, mouse models are not perfect substitutions 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)[12], and two known protein isoforms in UniProt[49]. We did not detect significant SJU changes in the *Setbp1*^{S858R} model, but that does not preclude changes in *SETBP1* isoform expression at a different developmental stage or in patients. Likewise, our genes with significant SJU in cerebral cortex cell types did not overlap with the previously reported genes with changes in AS in a single SGS patient in peripheral blood[11]. This is potentially due to study differences, including species and different sequencing technologies, variant locations (typical vs. atypical), tissue type, etc. Finally, we generated sequencing profiles with the Illumina platform using short-read sequencing. 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 further support that exon 3 has reduced expression in *Setbp1*^{S858R} mouse cerebral cortex but was not covered with our sequencing. Therefore, our analysis was limited to SJ expression and SJU, not full-length transcript expression and usage.

To our knowledge, this work is the first to identify brain cell-type-specific AS patterns resulting from *Setbp1* patient variants, including changes implicating *Son*. Future studies with additional SGS models and leveraging new technologies like 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 extremely rare, AS is known to change between sexes[44–48], 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. Also, while outside the scope of this study, future research to determine how splicing may be impacted in a cell-specific manner and, particularly, what role differences in *Son* SJU may have on disease pathogenesis in *SETBP1*-associated diseases are needed. In conclusion, our findings further implicate the impact of *Setbp1* patient variants on a diverse range of cell types and molecular mechanisms, underscoring their importance in contributing to SGS.

Abbreviations

AS - alternative splicing
CDS - coding sequence
OPCs - oligodendrocyte precursor cells
PCA - principal component analysis
scRNA-Seq - single-cell RNA sequencing
snRNA-Seq - single-nuclei RNA sequencing
SETBP1-HD - *SETBP1* haploinsufficiency disorder
SGS - Schinzel-Giedeion Syndorme

SJ - splice junction
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 Elizabeth J. Wilk, Amanda D. Clark, and Vishal H. Oza. The graphical abstract was created using BioRender.

Author Contributions

Emma F. Jones: Conceptualization (supporting); data curation (lead); formal analysis (lead); investigation (lead); methodology (lead); software (lead); visualization (lead); writing – original draft (lead); writing – editing (equal). **Timothy C. Howton:** Supervision (supporting); validation (equal); writing - editing (equal). **Tabea M. Soelter:** Validation (equal); writing – editing (equal). **Anthony B. Crumley:** Software (supporting); writing – editing (equal). **Brittany N. Lasseigne:** Conceptualization (lead); funding acquisition (lead); project administration (lead); resources (lead); supervision (lead); writing – editing (equal).

Keywords

Schinzel-Giedion Syndrome, *SETBP1*, rare disease, single-cell RNA-Seq, alternative splicing, neurodevelopment, patient variant, splice junction usage, Mendelian Disease, mouse

Appendices

Jones2024_SupportingInformation.docx

Supporting Information

Table_S1.csv

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. 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.
13. 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).
14. Tanis, S. E. J. *et al.* Splicing and Chromatin Factors Jointly Regulate Epidermal Differentiation. *Cell Rep.* **25**, 1292–1303.e5 (2018).
15. Zhang, X. *et al.* Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. *Cell* **166**, 1147–1162.e15 (2016).
16. Patowary, A. *et al.* Developmental isoform diversity in the human neocortex informs neuropsychiatric risk mechanisms. *Science* **384**, eadh7688 (2024).
17. 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).
18. Dingemans, A. J. M. *et al.* Establishing the phenotypic spectrum of ZTTK syndrome by analysis of 52 individuals with variants in SON. *Eur. J. Hum. Genet.* **30**, 271–281 (2022).
19. 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.
20. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* (2023) doi:10.1038/s41587-023-01767-y.
21. 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.

22. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
23. Luecken, M. D. *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50 (2022).
24. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
25. 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).
26. 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).
27. 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).
28. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
29. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2019).
30. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res.* **9**, (2020).
31. 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).
32. Koopmans, F. *et al.* SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron* **103**, 217–234.e4 (2019).
33. Indelicato, E. *et al.* SYNE1-ataxia: Novel genotypic and phenotypic findings. *Parkinsonism*

Relat. Disord. **62**, 210–214 (2019).

34. Gene: SYNE1 (ENSG00000131018) - Summary - Homo_sapiens - Ensembl genome browser 112. https://useast.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000131018;r=6:152121687-152637801.
35. Chen, J., Liu, C., Xu, M., Zhu, J. & Xia, Z. Upregulation of miR-19b-3p exacerbates chronic stress-induced changes in synaptic plasticity and cognition by targeting Drebrin. *Neuropharmacology* **207**, 108951 (2022).
36. 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).
37. Piazza, R. *et al.* SETBP1 induces transcription of a network of development genes by acting as an epigenetic hub. *Nat. Commun.* **9**, 2192 (2018).
38. 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).
39. Sanders, S. J., Schwartz, G. B. & Farh, K. K.-H. Clinical impact of splicing in neurodevelopmental disorders. *Genome Med.* **12**, 36 (2020).
40. Ahn, E.-Y. *et al.* SON controls cell-cycle progression by coordinated regulation of RNA splicing. *Mol. Cell* **42**, 185–198 (2011).
41. Sharma, A. *et al.* Son maintains accurate splicing for a subset of human pre-mRNAs. *J. Cell Sci.* **124**, 4286–4298 (2011).
42. Ueda, M. *et al.* Knockdown of Son, a mouse homologue of the ZTTK syndrome gene, causes neuronal migration defects and dendritic spine abnormalities. *Mol. Brain* **13**, 80 (2020).
43. Antonyan, L. & Ernst, C. Putative Roles of SETBP1 Dosage on the SET Oncogene to Affect Brain Development. *Front. Neurosci.* **16**, 813430 (2022).
44. Trabzuni, D. *et al.* Widespread sex differences in gene expression and splicing in the adult

human brain. *Nat. Commun.* **4**, 2771 (2013).

45. 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).
46. Karlebach, G. *et al.* The impact of biological sex on alternative splicing. *bioRxiv* 490904 (2020) doi:10.1101/490904.
47. 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).
48. 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).
49. UniProt. https://www.uniprot.org/uniprotkb/Q9Y6X0/entry.