UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA AGRONÓMICA, ALIMENTARIA Y DE BIOSISTEMAS

MÁSTER EN BIOLOGÍA COMPUTACIONAL

DEPARTAMENTO DE BIOTECNOLOGÍA - BIOLOGÍA VEGETAL

DEPARTAMENTO DE NEUROLOGÍA

COMPARATIVE ANALYSES OF SPLICING AND ISOFORM VARIANTS BY SHORT AND LONG-READ SEQUENCING

TRABAJO FIN DE MÁSTER

Autor: Javier Guo Shen

Tutor Externo: Silvia Corrochano Sánchez
Tutor Académico: Jesús Israel Pagán Muñoz

July, 2025

INDEX

SUMMARY

FUS mutations cause aggressive, early-onset Amyotrophic lateral sclerosis (ALS) and a form of dementia frontotemporal (FTD) disorders. Although the effect of *FUS* mutations on motor neuron degeneration has been largely investigated, the more global systemic consequences of these mutations, particularly on the RNA-processing landscape in a variety of other tissues, are not fully characterized. In the present work, we investigated transcriptomic alterations driven by the homozygous FUSDelta14 mutation using a knock-in mouse model.

We re-analysed the transcriptomic data from the RNA sequencing (RNA-Seq) by short read (Illumina) in four tissues (brain, spinal cord, liver, and skeletal muscle) of the control and mutant mice and develop the best pipeline for gene expression (different references) and splicing analysis (rMATs Turbo and IsoformSwitchAnalyzer). Further, to overcome inherent short-read limitations, we optimised the sequencing of cDNA using the long read technology from Oxford Nanopore, from library preparation through data analysis, enabling direct comparison between sequencing platforms.

Most changes in gene expression were tissue-dependent, with shared themes around RNA metabolism, lipid transport, mitochondrial activity and glial development. *Fus* and several pseudogenes of ribosomal genes were repeatedly dysregulated across all tissues. Splicing analysis via rMATS uncovered hundreds of genotype-specific events, primarily exon skipping and intron retention, including the skipping of the exon 14, what characterise the mouse model. However, the transcript-level quantification highlighted discrepancies in this reconstruction. As a first approximation, nanopore-based long-read RNA sequencing of the hippocampus seems to allow us to identify full-length isoforms and novel transcripts, demonstrating its effectiveness in dissecting complex splicing landscapes.

In summary, the present study reveals that the FUSDelta14 mutation causes widespread changes in gene expression and RNA splicing in a tissue-restricted manner. These studies indicate that long-read sequencing includes an essential source of resolution in identifying functionally relevant isoform changes and acts as a potent method to dissect transcriptomic dysregulation in ALS-FTD models.

1. INTRODUCTION

1.1 Amyotrophic Lateral Sclerosis–Frontotemporal Dementia

Neurodegenerative disorders are a group of different diseases characterized by a progressive loss of the function of neurons (nerve cells) and death. It is affected mostly in areas of the brain or the spinal cord, often associated to disabling symptoms such as cognitive impairment, motor deficit or behavioral disturbances. In the literature, these diseases are relatively distinct clinically but with many sharing molecular processes including protein misfolding and errors in RNA processing/output.(Soto & Pritzkow, 2018).

Among these disorders, the amyotrophic lateral sclerosis–frontotemporal dementia (ALS–FTD) spectrum has gained increasing attention in the past two decades and, as it brings together two conditions that were previously considered entirely distinct, each leading to different symptoms, considered as part of a disease spectrum. Since there are up to 50% of ALS patients have demonstrated frontotemporal dysfunction (Van Langenhove et al., 2012) and approximately 15% of patients with FTD presents signs of motor neuron involvement (Burrell et al., 2016). ALS-FTD spectrum disorders are highly complex and heterogeneous, affecting multiple tissues and biology systems in the course of life. While our understanding of ALS and FTD has advanced over the years, their underlying pathogenic mechanisms remain elusive.

FTD presents clinically with an early onset progressive cognitive and language impairment and behavioural abnormalities with selective loss of neurons predominantly within the frontal and temporal lobes (Bang et al., 2015). It is the second most common dementia after Alzheimer's disease (Van Langenhove et al., 2012).

ALS was first reported 140 years ago by the French physician Jean-Martin Charcot (Rowland & Shneider, 2001). ALS primarily involves the degeneration of both upper and lower motor neurons at the spinal or bulbar level (Brooks et al., 2000), resulting in a broad spectrum of phenotypes which are region dependent, making it a highly heterogeneous disease. The degeneration of upper motor neurons usually causes the Babinski sign, hyperreflexia and an exaggerated jaw or gag reflex. In contrast, the degeneration of lower motor neurons causes muscle atrophy and weakness (Hardiman et al., 2017,Swinnen & Robberecht, 2014).

A study carried out by DeJesus-Hernandez (DeJesus-Hernandez et al., 2011) reported that in the ALS-FTD spectrum there were common genetic mutations in genes which physiological function is implicated in the regulation of the RNA, being mainly RNA-binding proteins, including *TARDBP, C9ORF72, TBK1, and FUS.*

1.2 Fused in sarcoma (FUS)

Fus (Fused in Sarcoma) has emerged as a critical gene in ALS-FTD pathogenesis, as we mentioned before. *Fus* encodes a multifunctional RNA/DNA-binding protein belonging to the heterogeneous nuclear ribonucleoprotein (HNRNP) family of proteins. These proteins are predominantly nuclear and are involved in several biological processes: including transcriptional regulation, RNA splicing, transport, and DNA repair (Lagier-Tourenne et al., 2012; Ling et al., 2013; Sama et al., 2014). Pathogenic mutations in *FUS*, especially those affecting its nuclear localization signal (NLS), result in cytoplasmic mislocalization and aggregation of the FUS protein, leading to toxic gain-of-function effects (Kwiatkowski et al., 2009). These aggregates are observed in motor neurons and glia of the motor cortex and spinal cord (Vance et al., 2009) but also in the frontal cortex, hippocampus and striatum in the FTD-*FUS* subtype of frontotemporal dementia (FTD) (MacKenzie et al., 2011) Their presence interferes with RNA metabolism and have been shown to alter alternative splicing patterns in neurons, and also in glial cells (Fujioka et al., 2013; Rogelj et al., 2012).

Moreover, in studies of ALS-FTD, splicing alteration has been observed in cells and brain tissues from patients (Luisier et al., 2018; Prudencio et al., 2015). These transcriptomic abnormalities are thought to contribute directly to disease pathogenesis.Furthermore, experiments conducted with both cellular and animal models expressing mutant FUS exhibit motor neuron degeneration and transcriptomic dysregulation, suggesting that aberrant RNA processing may underline common mechanisms of neurodegeneration in neurological disorders (Sharma et al., 2016).

Fus variants can cause rare juvenile, aggressive genetic forms of amyotrophic lateral sclerosis (ALS). Up to 10% familial ALS cases are caused by variants in *Fus* (Kwiatkowski et al., 2009). Over 50 variants in the *Fus* gene have been identified in ALS family, the majority of which are dominant, missense changes clustered in and around the C-terminal nuclear localization signal (NLS) domain (encoded by exon 15) (Zakaryan & Gehring, 2006). Because the NLS is necessary for the import of FUS into the nucleus, mutations in this region can cause mislocalisation of FUS into the cytoplasm (Naumann et al., 2018), resulting in nuclear loss of function (Ishigaki & Sobue, 2018) and cytoplasmic toxic gain of function (Sharma et al., 2016).

To address the consequences of the FUSDelta14 mutation, a FUS mislocalisation variant first described in a single sporadic case associated with highly aggressive juvenile ALS with fast progression(DeJesus-Hernandez et al., 2010), we used a gene-targeted mouse model in which the endogenous *Fus* locus carries the same Δ14 mutation. This variant affects the splice acceptor site of intron 13, resulting in skipping of exon 14 and a subsequent frameshift in exon 15 that disrupts

the nuclear localization signal (NLS). Additionally, it includes a humanised exon 15, faithfully reproducing the aberrant C-terminal sequence of the mutant FUS protein observed in patients (Devoy et al., 2017).Knock-in models like this offer physiological relevance, in particular for the assessment of systemic and early disease manifestations, since the mutant gene is expressed under its native promoter rather than being artificially overexpressed. Heterozygous FUSDelta14 ($Fus^{\Delta 14/+}$) mice develop mild, late-onset neuromuscular and motor symptoms resembling features of ALS, though these symptoms do not progress to terminal disease within the animals' normal lifespan (Devoy et al., 2017). In contrast, Knock-in homozygous FUSDelta14 ($Fus^{\Delta 14/\Delta 14}$) mice were generated to evaluate phenotypic differences, as recently described (Ali et al., 2023). These animals show a more severe phenotype, characterized by an early-onset neurodevelopmental abnormalities, impaired motor function and systemic metabolic alterations. These observed phenotypic variation from heterozygous and homozygous animals propose a dose-sensitive impact of the FUSDelta14 mutation, whereby total depletion of FUS nuclear localization results into developmental defect and systemic homeostasis abnormalities in addition to neurodegeneration.

1.3 RNA sequencing necessity for ALS-FTD

Transcriptomic approaches have become essential tools to investigate and characterize disease mechanisms, given the strong involvement of RNA-binding proteins and RNA metabolism in ALS-FTD. RNA sequencing (RNA-seq) along with other sequencing technologies enables genes expression profiling and detection of splicing variants and novel or aberrant transcripts (Z. Wang et al., 2009). Mis-splicing of mRNA plays a major role in the pathogenesis of this spectrum as it alters the function of proteins, frequently resulting in truncated, non-functional or even toxic proteins. Investigations of this mechanism in the transcriptomic field not only offer a better understanding of disturbed regulatory mechanisms that lead to diseases but also hold promise for the discovery of biomarkers and novel therapeutic strategies, urgently needed for patients under those conditions (Mehta et al., 2023).

1.4 Short read and long read sequences differences

Short-read RNA-seq methods including Illumina have been the method of choice for transcriptome studies, primarily due to their high accuracy and depth of coverage. But there are certain inherent limitations that prevent transcript diversity comprehension. Some of this can include the inability to reconstruct full-length transcripts to resolve complex alternative splicing events (Prudencio et al., 2015). The moment long-read sequencing technologies, developed by Oxford Nanopore Technologies (ONT) has arrived, part of these issues have been solved. ONT allows for direct full-length transcript sequencing and accurate relative quantification, presenting a highly parallel, real-time alternative without the requirement of assembly (Garalde aliet al.,

2018).

Additionally, sequencing with ONT have revealed a higher complexity of isoform usage and have been particularly useful in the analysis of neurodegenerative illnesses when splicing regulation is impaired (Maggi et al., 2024,Clark et al., 2020,Heberle et al., 2023).
It is worth mentioning that Pacific Biosciences (PacBio) is another tool in the long-read field, that also offers high-fidelity (HiFi) sequencing with excellent accuracy, making it another alternative for isoform discovery.

While long-read sequencing was originally associated with higher error rates, recent advances in base-calling algorithms and hybrid sequencing approaches have greatly improved their error profile (Soneson et al., 2019). Comparison of the three tools can be visualized in the table below:

	Illumina	PacBio	Oxford Nanopore
Strength	- Very high throughput (100–1,000× more reads than long-read). - Well-characterized bias and error profiles. - Wide range of established computational workflows, including for degraded RNA	- Captures full-length transcripts (1–50 kb). - Simplifies transcriptome reconstruction (ab initio assembly).	- Captures full-length transcripts (1–50 kb). - Simplifies ab initio transcriptome assembly. - Reduced need for reverse transcription or PCR. - Enables detection of RNA base modifications and estimation of poly(A) tail lengths.
Limitations	- Isoforms must be inferred computationally (no full-length reads). - Limited isoform resolution and alternative splicing detection. - Sample preparation (RT, PCR, size selection) introduces bias.	- Lower throughput (500K–10M reads). - Sample prep biases still present. - Higher sequencing error rate. - Not optimal for degraded RNA.	- Low throughput (500K–1M reads). - Error profiles not fully resolved. - Biases during library prep still under study. - Similar concerns to PacBio for error rates.
Key Applications	- DGEWTA, small RNA, single-cell analysis, spatial omics, de novo RNA assembly, translatome analysis, structural and RNA-protein interaction analysis, etc.	- Isoform discovery, ab initio transcriptome analysis, fusion transcript discovery, MHC, HLA, or other complex transcript analysis.	- Isoform discovery, ab initio transcriptome analysis, fusion transcript discovery, MHC, HLA, or other complex transcript analysis. - Detection of RNA modifications.

Table 1. Summary of strengths, limitations, and key applications of Illumina, PacBio, and Oxford Nanopore RNA sequencing platforms.

In the present study, we use RNA-seq, using not only short-read but also long-read technology to assess the transcriptomic effects of *Fus* mutations in a mouse model for ALS. Our primary goal is to compare computational pipelines as well as optimization of long-read sequencing protocol, in order to provide a full characterization of gene expression and alternative splicing in the given disease context.

1.5 OBJECTIVES

This study aims to optimize and compare analytical approaches for both short- and long-read RNA sequencing data to better understand gene expression, splicing events and isoforms analysis, and apply this knowledge to the effect of *Fus* mutation in a mouse model.

1. Analyze a short-read RNA sequencing data from a *Fus* mutant mouse model previously published by (Ali et al., 2023; Devoy et al., 2017), using an alternative computational pipeline and an updated *Mus musculus* reference genome to evaluate and identify the most efficient and accurate pipeline for gene expression analysis.
2. Splicing analysis with the previous database with two splicing-detection algorithms, rMATS Turbo (Y. Wang et al., 2024) and IsoformSwitchAnalyzeR to benchmark methods for splicing events, and with the aim of determining the effects of the *Fus* mutation on the splicing landscape.
3. Optimization of the ONT for long-read RNA sequencing, using the SQK-PCB114. 2A kit, to improve the efficiency of library preparation, accuracy of sequencing and analysis of the long-read RNA-seq to further enhance our understanding at isoform level.

2. MATERIAL AND METHODS

2.2 Bioinformatics analysis for short read RNA sequencing

Gene expression analysis

The dataset was obtained from (Ali et al., 2023), available in the Gene Expression Omnibus (GEO) under accession number GSE245520. Paired-end FASTQ files were preprocessed using fastp (v0.24.0+galaxy3) via the Galaxy platform. Specifically, only bases with a minimum Phred quality score of 30 were considered "qualified".

Reads were then aligned to the reference genome using HISAT2 (v2.1.0) employing both mm10 and mm39 gene annotation files from the GENCODE and RefSeq databases to facilitate downstream comparative analyses and gene counts were obtained using featureCounts (v2.1.1+galaxy0). Differential expression analysis was conducted using DESeq2 (v1. 46. 0) in R (v4.4.2). (Normalized expression was calculated in terms of FPKM and genes with adjusted P-value< 0.05 was considered to be significant). Functional enrichment analysis was conducted via Overrepresentation Analysis (ORA), which identified significantly enriched Gene Ontology (GO) terms.
To streamline this workflow, we also developed a modular bash pipeline “ilumina_pipeline.sh” that automates preprocessing, alignment, quantification, and differential analysis for short reads as; this is available in https://github.com/javiersg072/Fus_splicing_pipeline.

rMATS Turbo splicing analysis

Alternative splicing events were analyzed and visualized using rMATS (v4.3.0) (RNA-Mutually Inclusive/Exclusive Transcript Splicing), a computational tool designed for detecting differential alternative splicing from RNA-Seq data. Preprocessed aligned BAM files from HISAT2 and a corresponding gene annotation file (GTF) were used as input. Splicing events, including exon skipping (SE), alternative 5' or 3' splice sites (A5SS, A3SS), mutually exclusive exons (MXE), and retained introns (RI), were identified using the rMATS turbo pipeline under paired-end mode.

Following differential splicing analysis, significant splicing events were selected based on false discovery rate (FDR) < 0.05 and a minimum absolute $\Delta\Psi$ ($|\Delta\Psi|$) threshold of > 0.05. Splicing patterns were visualized using rMATS-plot, which then generates sashimi plots to illustrate exon inclusion or exclusion levels across conditions.

The STRING database (Search Tool for the Retrieval of Interacting Genes/Proteins) was used to

analyze the relative pathway associations of the identified genes and to determine the primary functional enrichment of gene clusters. Protein-protein interaction networks were generated, and functional annotations were assessed based on Gene Ontology (GO) analysis. To identify FUS-bound genes, we used the POSTAR3 database of experimentally validated RNA-binding protein targets (http://111.198.139.65/RBP.html).

Isoform switch analyzer analysis

Transcript abundance was estimated using Salmon (v1.10.1) in quasi-mapping mode. A decoy-aware index was built from the GRCm39 genome and GENCODE vM36 transcripts.

Paired-end FASTQ files were processed with bias correction enabled (sequence and GC bias), and quantification was guided using the GENCODE GTF annotation (--geneMap). Salmon was run with automatic library type detection, fragment length settings (mean: 250 bp, SD: 25 bp), and optimized parameters to improve accuracy. Output quantifications (quant.sf) were generated per sample for downstream isoform analysis.

To investigate the functional consequences of isoform usage changes, we applied the IsoformSwitchAnalyzeR pipeline (Vitting-Seerup et al., 2019). The tool is run also in R (v4.4.2), it integrates isoform-level expression data to identify significant switches and evaluate alterations in coding sequences, protein domains, and other structural features. Moreover, the tool provides enrichment analyses to highlight pathways and biological processes that may be affected by isoform switching, thereby contributing to a more detailed understanding of the functional relevance of the splicing changes detected in the dataset.

2.3 SQK-PCB114.2A kit optimization

RNA was extracted from hippocampus male mice at 9 months of age using the kit “Mini Lipid Tissue RNAeasy” (Qiagen). Then, the concentration was measured with Qubit Fluorometer (Thermo Fisher Scientific), and RNA integrity was assessed with the Agilent TapeStation system, employing the High Sensitivity RNA ScreenTape assay (Agilent Technologies, 2015). The sequencing workflow follows the PromethION cDNA-PCR Barcoding Kit 24 V14 protocol provided by Oxford Nanopore Technologies. Several modifications were introduced, such as during the reverse transcription step, which is the point 3 in the provided protocol, up to 100 ng of RNA was used and the reverse transcription was extended to 90 minutes at 42ºC instead of 30 min. Additionally, the mixing time with the Hula mixer was extended to 10 minutes in all the steps. Concentrations in PCR vary from one sample to another (as shown in **table 1**).

A sample was reamplified using a method to remove amplified cDNA artefact (*A Method to Remove Amplified CDNA Artefact in SQK-PCS111 and SQK-PCB111.24 Libraries*, n.d.)

In the final PCR extension step, where each cycle typically lasts 6 minutes, we reduced the number of cycles from o 6. The sequencing was performed using a flow cell with 5,600 active pores. After the initial loading, the flow cell was washed and reloaded with the same amount of input material at 24 hours and again at 48 hours. The run was automatically stopped at 72 hours.

	Sample	**µl in PCR**	**pcr reamplificacion**
B10	Hippocampus	139.2 ng	10ul=32.22 ng
B15	Hippocampus	110.4 ng	

Table 1. RNA input amounts used in PCR and reamplification reactions for selected samples. Samples B10 and B15 represent whole lysates.

2.4 Gene expression and splicing analysis for long read RNA sequencing

Oxford Nanopore Library Preparation and Sequencing

For the, long read analysis, raw Nanopore reads were demultiplexed according to their corresponding barcodes and merged into single compressed FASTQ files per sample (samples from barcode 10 and 15). The quality of the merged reads was initially assessed with NanoPlot v1.44.1 to evaluate read length and quality score distributions. Quality filtering was subsequently performed using Chopper v0.9.1, applying a minimum read quality threshold of Q6 to remove low-quality reads. The filtered datasets were evaluated again to confirm the improvement in read quality.

A continuation filtered reads were then aligned to the mouse reference genome (GRCm39, Gencode release 107) using Minimap2 v2.28-r1209, employing splice-aware alignment parameters optimized for long-read RNA sequencing (-ax splice -uf -k14). The generated SAM alignment files were converted to BAM format, sorted, and indexed using Samtools v1.13. Finally, alignment quality metrics were generated for each sample using Samtools flagstat. StringTie2 (v2.2.1) was used for transcript assembly, and featureCounts (v2.0.3) was employed for quantification. Gene annotations were downloaded from Ensembl release 113: Mus_musculus.GRCm39.113.gtf.gz (ftp://ftp.ensembl.org/pub/release-113/gtf/mus_musculus/)

3. RESULTS

Gene expression from short read RNA sequencing

Annotation and Assembly Comparison of Detected Transcripts

In our study, it was important to assess the impact of genome annotation source and assembly version on our transcript detection to ensure consistent and reliable results to our posterior analysis. To achieve this, we performed a comparative analysis using DESeq2 in R and then followed by visual inspection through Venn diagrams (Figure 1). Each diagram illustrates the number and proportion of shared and unique genes identified across different combinations of annotation sources (GENCODE and NCBI) and genome assemblies (mm10 and mm39).

We can observe in Figure 1 that if the source remains from the same database, the similarity of genome assembly shares mostly the same genes (between 75%–80%), meaning that the majority of gene annotations are conserved across genome versions, and differences are likely due to improved annotation accuracy (Figure 1A, B). In contrast, if we compare the different, even within the same genome assembly, we can observe a notable decrease in shared gene content (60% approx.), suggesting that discrepancies arise not from the reference genome itself, but from variations in how different databases define and curate gene models. Moreover, there is an increase in the detected genes numbers for the annotation in RefSeq (Figure 1 C, D)

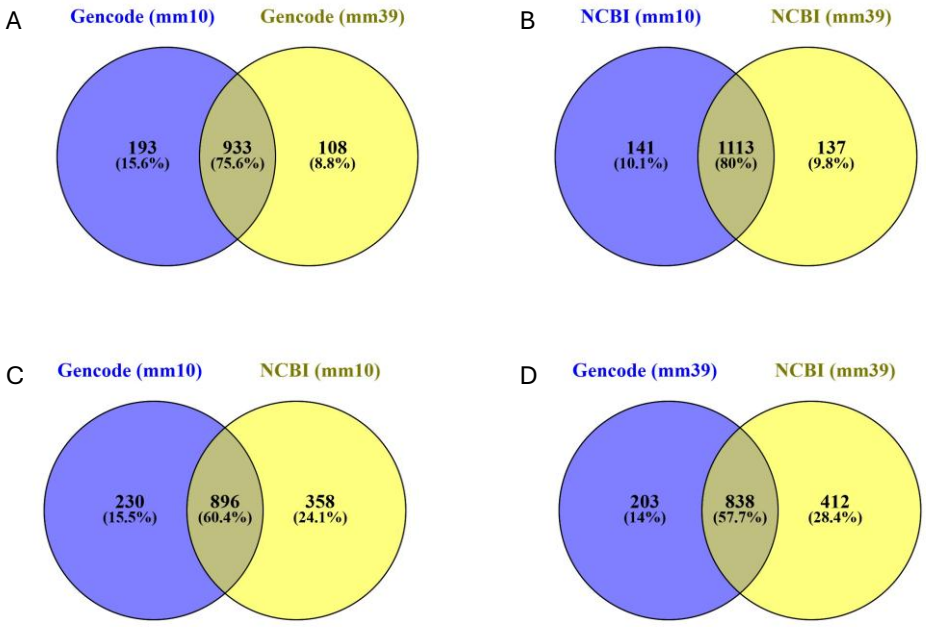


Figure 1. Comparison of genes overlaps between genome annotations and assemblies using Venn diagrams. Gencode version mm10 vs Gencode version mm39 (A). NCBI Refseq version mm10 vs NCBI Refseq version mm39 (B). Gencode version mm10 vs NCBI Refseq version mm10 (C). Gencode version mm39 vs NCBI Refseq version mm39 (D).

As a continuation, we compared the overlap of significant alternative splicing (AS) events detected across mouse genome assemblies and annotation sources using Venn diagrams (Figures X). Filtering was performed based on a false discovery rate (FDR) threshold and an absolute IncLevelDifference cutoff.

In the comparison between mm39 GENCODE and mm10 GENCODE (Figure 2), we observed a high proportion of shared events across all splicing event types. This suggests that while the genome assembly version has a moderate influence on splicing event detection, it preserves a strong level of consistency between mm10 and mm39. The remaining events (ranging from 7% to 17%) appear to be assembly specific, likely due to shifts in genomic coordinates and updates in transcript annotations.

In contrast, the comparison between mm39 GENCODE and mm39 NCBI RefSeq revealed notably lower overlap rates for splicing events, approximately 50% overall, although skipping exon (SE) events showed a surprisingly high overlap of around 80%. The discrepancies likely arise from differences in transcript structure definitions, isoform diversity, and annotation strategies between different sources GENCODE and RefSeq, and this last one has a greater impact on splicing event detection than the genome assembly itself.

Based on these findings, we ultimately selected the NCBI RefSeq mm39 genome and its corresponding annotation for alignment. All subsequent analyses were performed using this reference.

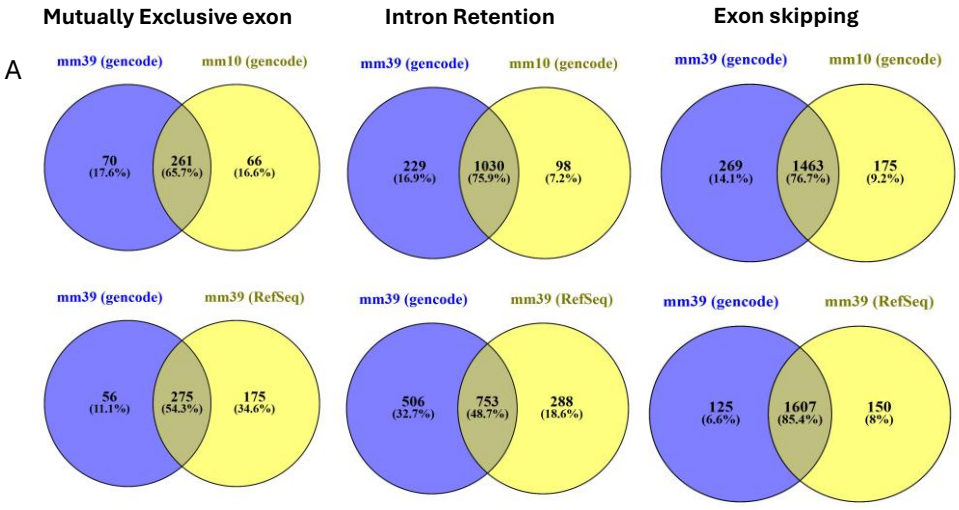


Figure 2. Overlap of alternative splicing events across genome assemblies and annotation sources. Mutually exclusive exon (first column), intron retention (second column) and exon skipping (third column). Gencode version mm39 vs Gencode version mm10 (first row). Gencode version mm39 vs NCBI Refseq version mm39 (second row).

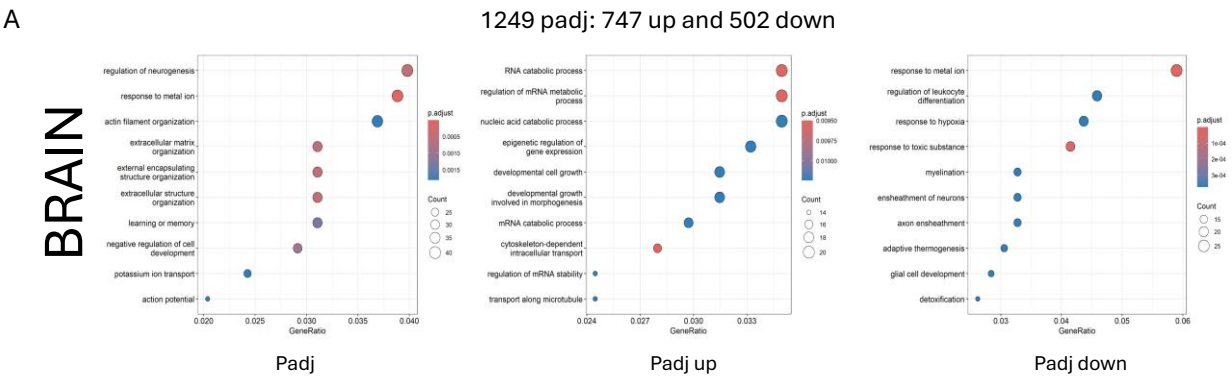
Differential Gene Expression and splicing analysis using a new genome version

To identify biological pathways and functional categories that are statistically enriched within a set of differentially expressed genes (DEGs), Over-Representation Analysis (ORA) was performed. This helps us identify potential biological roles that are altered by the mutation. For this analysis, only the mm39 assembly from the NCBI RefSeq annotation was used to ensure the most up-to-date gene models and a consistent reference for downstream interpretation. Taking a look at Figure 3, in the brain (BR), FUSDelta14 mutants show enrichment in upregulated pathways such as RNA metabolism, RNA and nucleic acid catabolic processes, but downregulation in pathways related to glial cell development, ensheathment of neurons and axon (Figure 3A). In the spinal cord (SPC), the downregulated pathways include oxidative phosphorylation, ATP biosynthesis, and ribonucleoprotein biosynthetic processes (Figure 3B). These changes suggest a tissue-specific in the central nervous system impact from the *Fus* mutation, with a possible consequence of altered energy metabolism and translational regulation being particularly characteristic of the SPC tissue.

Liver samples (LIV) showed an increase in lipid metabolism and mitochondrial function, possibly leading to systemic metabolic disruption (Figure 3C). Meanwhile, in the tibialis anterior (TA), we found altered pathways related to lipid transport and muscle processes, suggesting that skeletal muscle is particularly vulnerable to the *Fus* mutation, potentially contributing to neuromuscular symptoms (Figure 3D).

Doing a further analysis on the shared genes across all tissues, we found out that there were five potential genes: *Fus, Gm16439,* and the pseudogenes *Rps15a-ps3, Rps15a-ps5, and Rps15a-ps7*, with *Fus* and *Rps15a* as common targets of the FUSDelta14 mutation. *Fus*, the mutated gene, shows consistent upregulation, while the presence of ribosomal pseudogenes suggests a shared disruption in translational regulation. *Gm16439*, a predicted gene of unknown function, may be a novel candidate linked to FUS-related pathology.

A



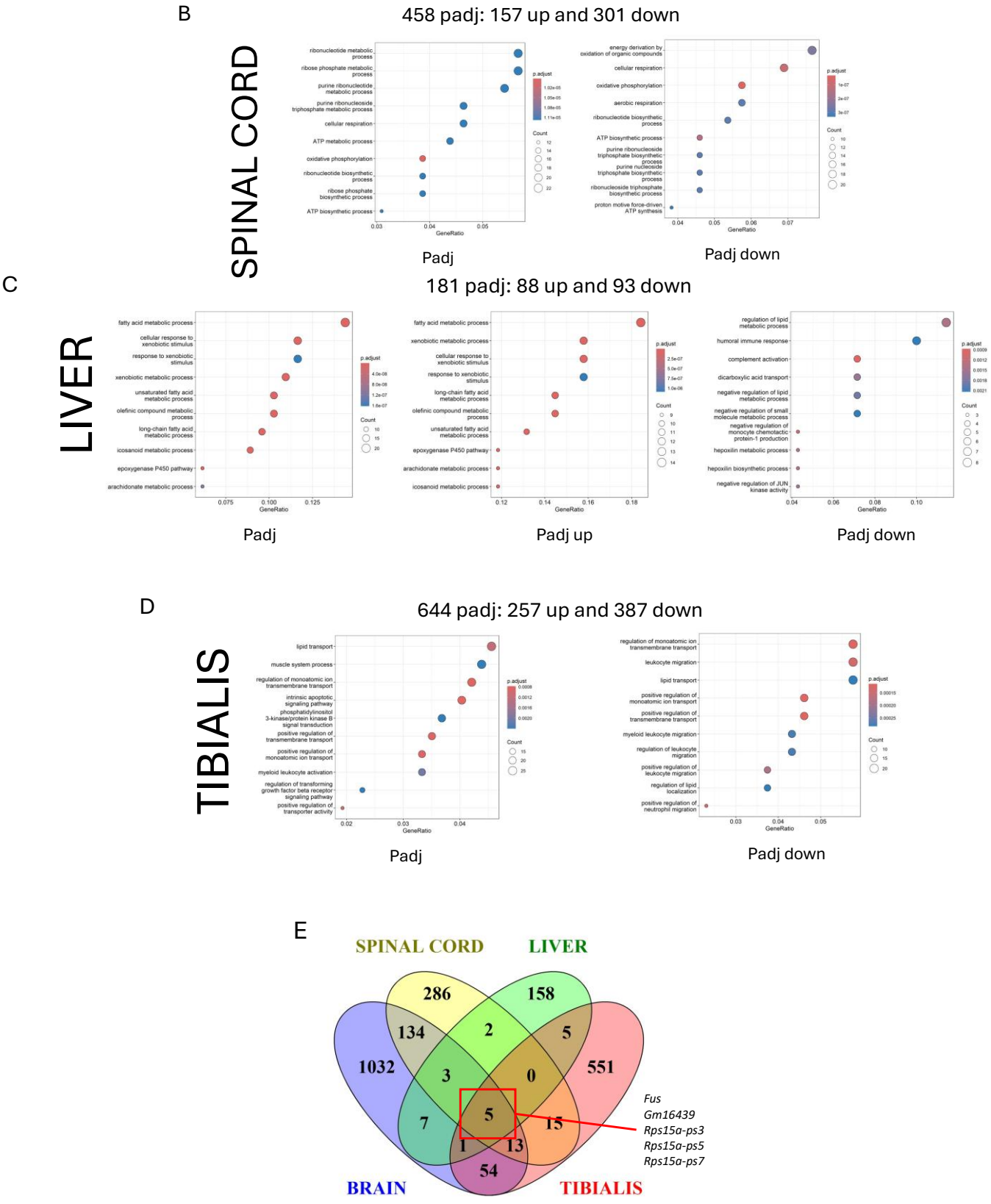


Figure 3. Over-representation analysis (ORA) of differentially expressed genes in FUSΔ14/Δ14 (HOM) versus wild-type (WT) mice across four tissues: brain (BR) (A), spinal cord (SPC) (B), liver (LIV) (C), and tibialis anterior (TA)(D). Dot plots show significantly enriched GO biological

processes for upregulated and downregulated gene sets (FDR < 0.05). Dot size indicates gene count, and color denotes genotype. The Venn diagram illustrates the overlap of significantly differentially expressed genes among the four tissues, with the largest number of unique genes found in the brain (E).

One of the main functions of Fus is RNA regulation, particularly the control of alternative splicing. In this mouse model, *Fus* undergoes a skipping of exon 14, causing a nuclear loss of function, which may further disrupt global splicing homeostasis. To further investigate the impact of Fus dysfunction on transcript processing, we performed a comprehensive analysis on alternative splicing events.

To visualize and validate alternative splicing events detected computationally, we used rMATS Turbo, which systematically classified splicing events into the following categories:

- Exon skipping (SE)
- Retained introns (RI)
- Alternative 5′/3′ splice sites (A5SS/A3SS)
- Mutually exclusive exons (MXE)

Splicing events were filtered using an adjusted p-value (FDR < 0.05) and an absolute ΔΨ (Inclusion Level Difference) > 0.05, revealing a significant number of genotype-specific alterations.

Most events are tissue-specific, with relatively few shared across all four tissues. Among the shared genes (Figure 4) are *Tsc22d1, Commd10*, and *Agpat3* (A3SS), *Sirt6, Knop1*, and *Mrpl2* (A5SS), and *Firmc1, 2210408I21Rik*, etc (MXE). The rest of shared genes in RI and SE events are listed in Supplementary Table S1. To further explore the potential regulatory influence of *Fus* on alternative splicing, we cross-referenced the genes undergoing alternative splicing events across tissues with *Fus*-bound targets identified in the POSTAR3 database. As summarized in (Table S1), a substantial proportion of the alternatively spliced genes, particularly in SE (80%) and RI (53%) events, are *Fus* targets. Other event types such as A3SS (100%), A5SS (67%) and MXE (100%) also show considerable overlap with *Fus*-bound targets. Skipped exons (SE) and retained introns (RI) show the highest number of events, while mutually exclusive exons (MXE) are least frequent. Notably, among the retained intron events, we detected *Tardbp*, a key RNA-binding protein also associated with the ALS-FTD spectrum. To explore the functional impact of splicing changes identified by rMATS, we performed GO enrichment analysis using STRING for all splicing event types, and in the case of A3SS, key pathways include RNA processing in brain, endocytosis regulation in spinal cord, locomotor rhythm in liver, and cellular metabolic process in tibialis (Figure S1). To assess functional convergence between differentially expressed genes and those affected by splicing events, we performed GO enrichment analysis using STRING, separately for each tissue. Only the brain showed significant enrichment, with top terms including

"cell projection" and "neuron projection", both closely related to cellular extensions involved in neuronal communication and signaling. Full results are shown in Supplementary Figure S2.

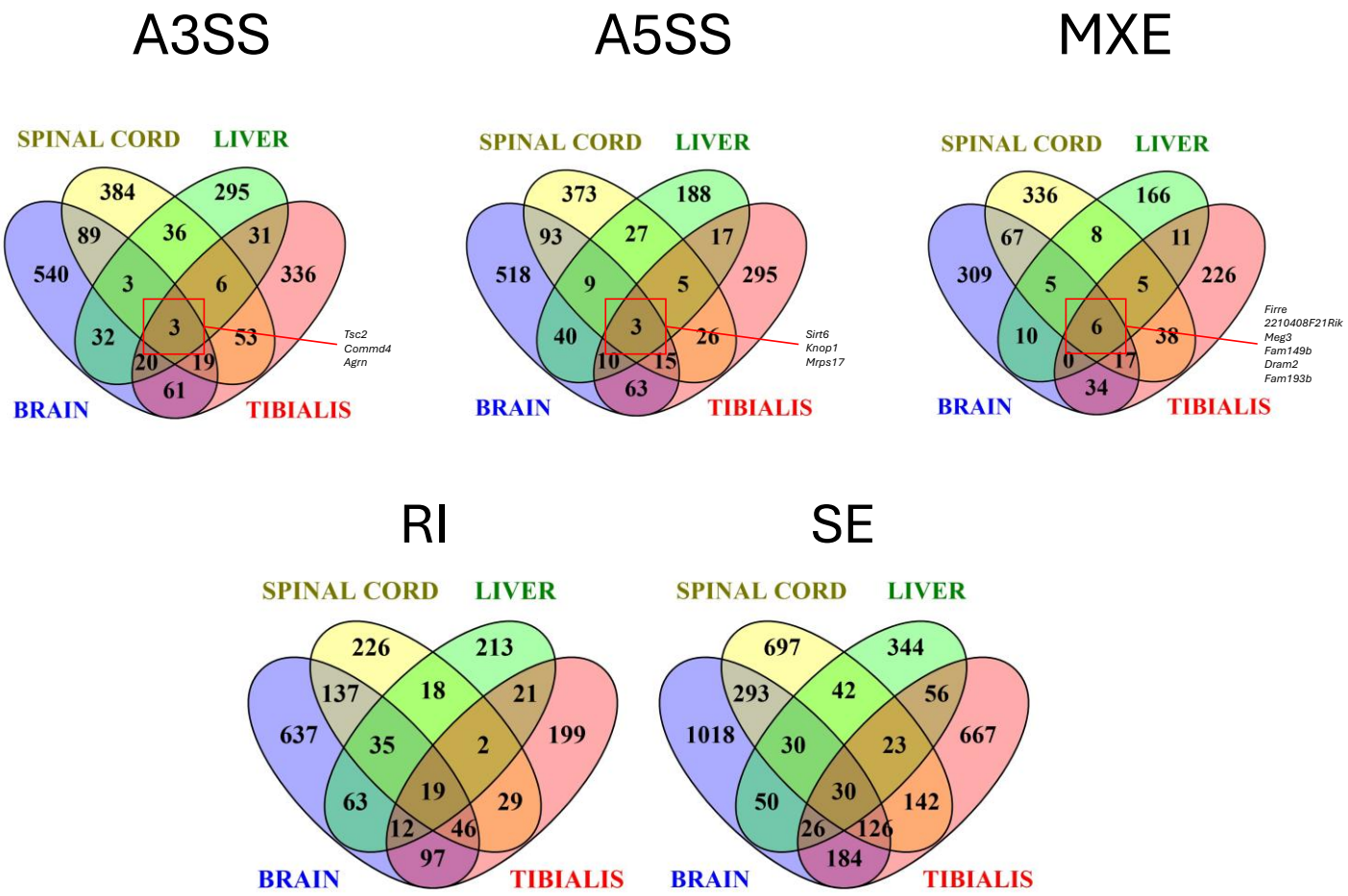


Figure 4. Venn diagrams showing tissue-specific and shared alternative splicing events (A3SS, A5SS, MXE, RI, SE) across brain, spinal cord, liver, and tibialis from rMATS analysis.

As a follow-up, to confirm the splicing mutation consequence of the model we generated Sashimi plots for a representative exon skipping (SE) event within the *Fus* gene, across the four tissues (Figure 5). These plots visualize exon–intron structure and splice junction usage, allowing direct comparison between wild-type (WT) and $Fus^{\Delta 14/\Delta 14}$ (HOM) genotypes. The remaining isoform events are listed in Table S2; however, only the skipped exon (SE) event is common across all tissues.

The results were consistent across all tissues: HOM (group 1) samples exhibited complete exon skipping, with nearly no read coverage over the skipped exon and strong junction reads connecting the flanking exons. Upon checking the coordinates, we identified the affected exons as 13, 14 and 15, with exon 14 being skipped, as expected for the homozygous Δ14 mutation.

This proves and confirms our initial mouse model for these experiments

In contrast, the WT group (group 2) showed full exon inclusion (IncLevel = 1.00), compared to an IncLevel of 0.00 in HOM, confirming a complete differential splicing event between the two genotypes.

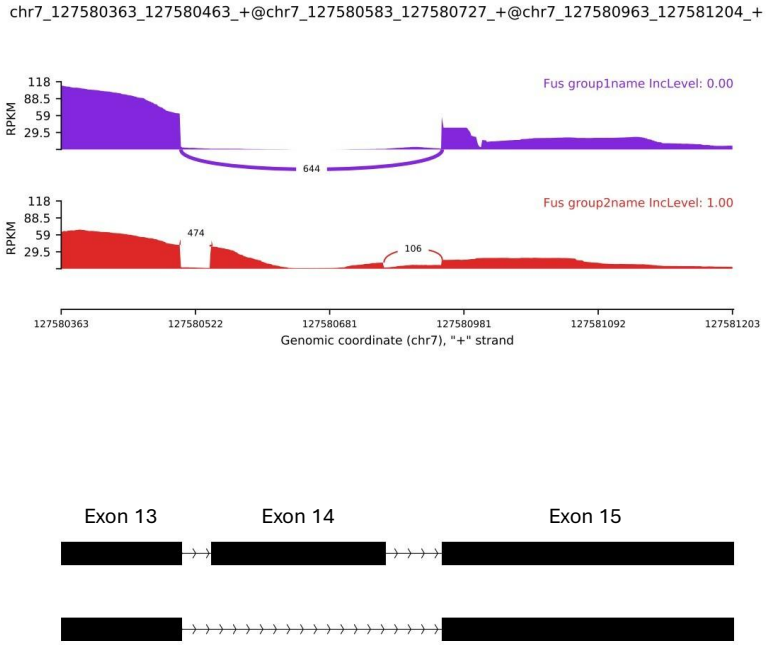


Figure 5. Sashimi plots showing a consistent exon skipping (SE) event in the *Fus* gene. The *Fus*$^{\Delta 14/\Delta 14}$ (HOM) group (purple) and the wild-type (WT) group (red). Arcs represent splice junction read support, and RPKM values indicate transcript abundance. This result highlights a genotype-driven, tissue-wide splicing shift in *Fus* due to the Δ14 mutation.

Comparative Transcript-Level Quantification across Tissues Results

To build on the splicing event analysis, we next examined how these changes manifest at the isoform level, aiming to understand their potential impact on transcript functionality.

Transcript-level expression quantification using Salmon was performed across four tissues: brain frontal cortex (BR), lumbar spinal cord (SPC), liver (LIV), and tibialis anterior (TA). For each tissue, we analyzed:

1. Overall gene expression between homozygous mutants (hom) and wild type (wt). (Figure 6)
2. Isoform-level structures indicating exon inclusion/skipping and coding potential. (Figure 7)
3. Isoform-specific expression levels. (Figure 7)
4. Fractional usage of isoforms, i.e., the proportion of total gene expression contributed by each isoform. (Figure 7)

In figure 6, we observe the overall gene expression levels of *Fus*. Across all tissues, expression tends to be elevated in HOM compared to WT, suggesting genotype-driven transcriptional

upregulation. This is particularly evident in the BR and SPC, where differential expressions appear most pronounced. Additionally, in terms of expression levels, the brain tends to exhibit the highest Fus expression among all tissues, followed by the spinal cord, liver, and tibialis.

This can suggest that the *Fus* mutation is altering the regulation of the *Fus* gene itself, most likely leading to its upregulating either directly (through disrupted RNA binding or splicing) or indirectly (as part of a downstream effect or stress response). Interestingly, as we can observe this gain of expression in all these tissues, this can indicate that FUSDelta14 mutation exerts a global effect on the regulation of *Fus* expression across the entire organism rather than affecting it only in specific cell types.

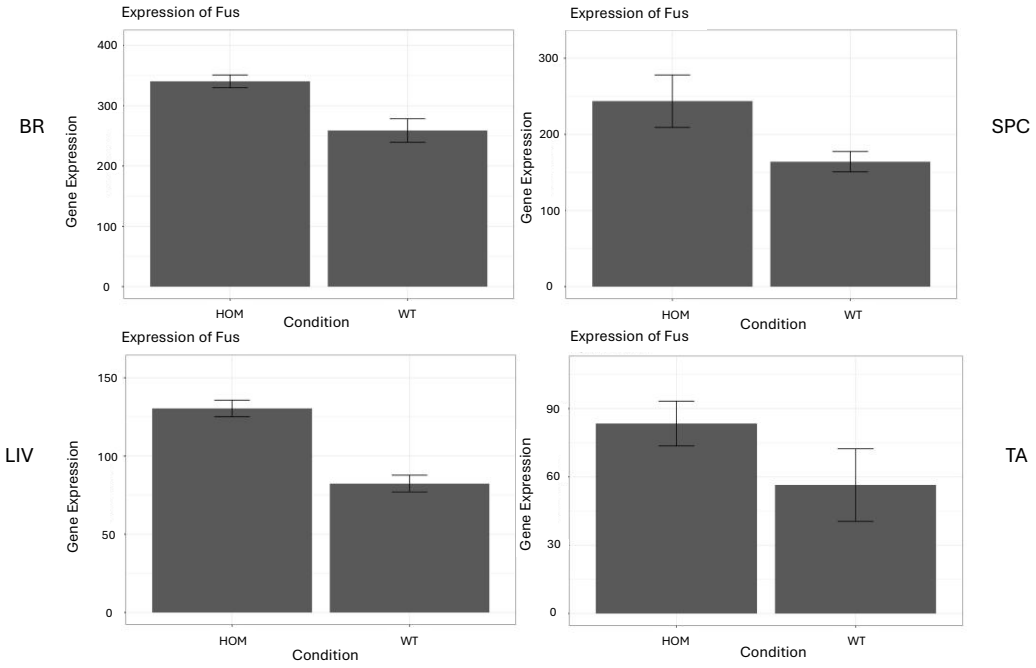


Figure 6. *Fus* gene expression levels in wild-type (WT) and FUSΔ14/Δ14 (HOM) mice across four tissues: brain (BR) (top left), spinal cord (SPC) (top right), liver (LIV) (bottom left), and tibialis anterior muscle (TA) (bottom right).

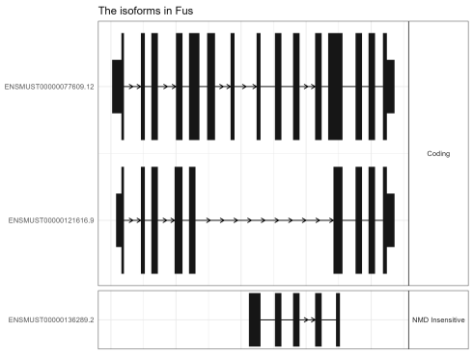
The figure 7 of each tissue slide shows the structural representation of isoforms (1st column). These diagrams illustrate which exons are retained or skipped in each isoform and whether the isoforms are predicted to be coding or insensitive to nonsense-mediated decay (NMD).

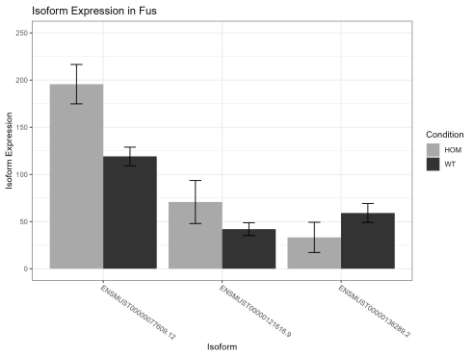
The second column shows the expression levels of each isoform across genotypes. Coding isoforms tend to be more abundant in HOM, especially in BR and SPC, with expression levels reaching approximately 150–200 units, this can suggest an enhanced productive splicing. In contrast, NMD-insensitive isoforms show reduced expression, indicating they are either not preferentially spliced or are more tightly regulated under normal conditions. In the last column, shows the relative usage of each isoform (isoform fraction) of the different tissues. For example, in all tissues, a particular coding isoform ENMUST00000077609.12 dominates the expression

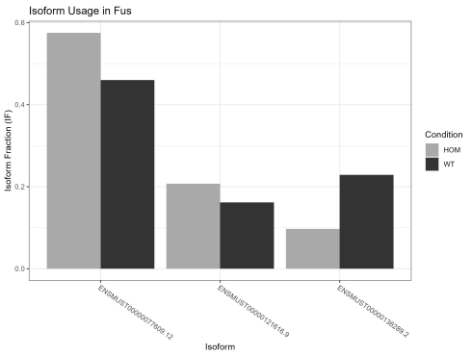
landscape in HOM. Meanwhile, the relative isoform usage of ENMUST00000136289.2 (Fus-207) in HOM from the LIV context, is increased compared to the other tissues. All these results suggest a higher sensitivity to *Fus* perturbation in liver.

We do not fully trust this isoform analysis because, as shown in the plots for the homozygous samples (HOM), exon 14 appears to be included in the different isoforms. For example, the first transcript displayed in the graph across all tissues is ENSMUST00000077609.12 (Fus-201), which includes a total of 15 exons, exon 14 among them. Another transcript, ENSMUST00000121616.9 (Fus-204), also contains exon 14. However, this cannot be correct, as the homozygous mutation in our model causes this exon to be skipped as clearly demonstrated by the rMATS sashimi plots in the previous analysis. This potentially highlights a limitation of short read-based isoform inference. This is one of the reasons why we are implementing long-read sequencing technologies like Nanopore and optimizing the protocol to achieve more accurate isoform resolution.

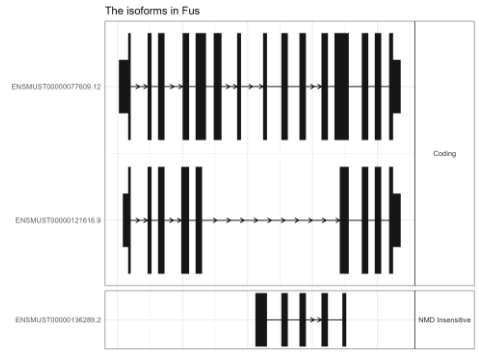
BRAIN

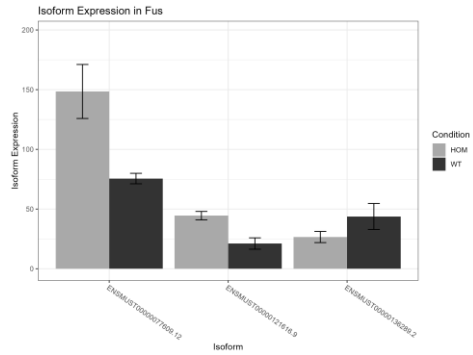


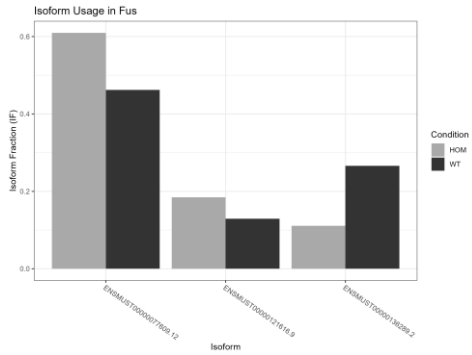




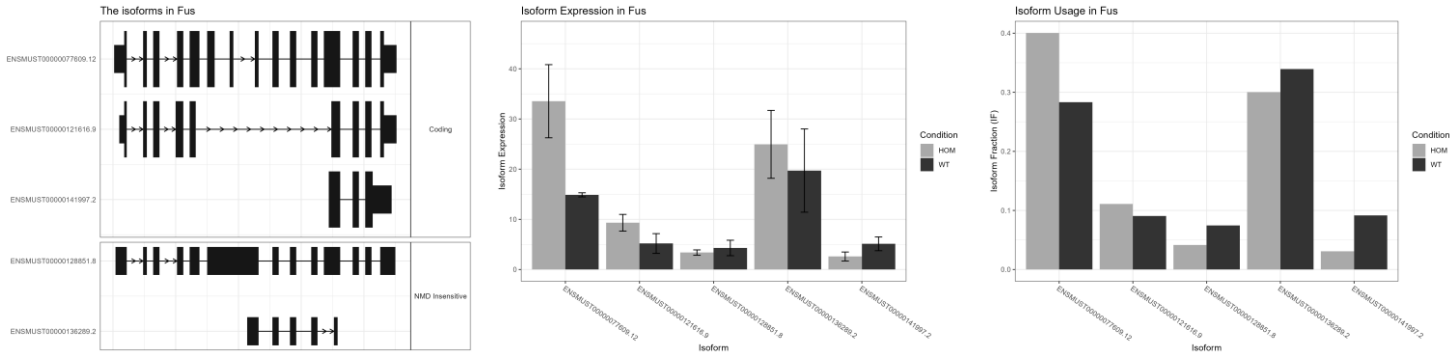
SPINAL CORD







LIVER



TIBIALIS

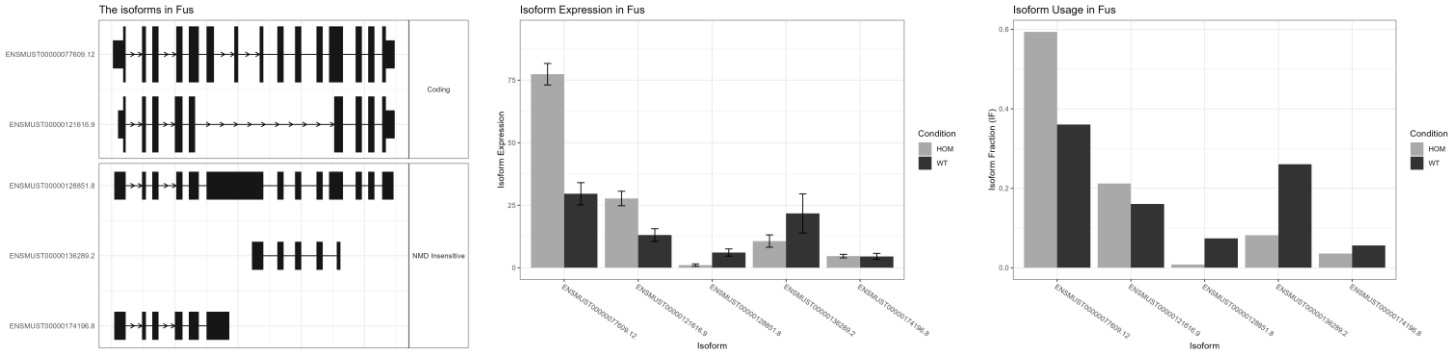


Figure 7. Differential expression and isoform usage of the *Fus* gene across conditions.

SQK-PCB114.2A kit optimization workflow

In the initial results following the original pipeline, we encountered a poor library, the read length distribution was below than expected from the protocol. In our optimization of the SQK-PCB114.2A kit for long-read Nanopore sequencing, we compared the effects of the normal workflow versus including a reamplification step after library preparation. For these purposes, we did the studies on hippocampus.

The reamplified libraries produced greater total reads and better overall alignment capabilities, indicating that reamplification improves the quality and mappability of sequence read data. (Table 2)

However, when looking at read length N50, the length above which 50% of the total bases are found, the N50 values of the B11 reamplified samples (585 bp and 561 bp respectively) were lower than those of the non-reamplified samples (B15), which had values of 679 and 604 bp.

The increased alignment rate of reamplified samples was evident, with 93–94% of reads aligning compared to only 81–82% of non-reamplified samples.

This may suggest that while reamplification boosts sequencing efficiency and alignment success, it introduces a shift in fragment size distribution, favoring shorter molecules. This is due to that the PCR amplification tends to enrich shorter fragments, which amplify more efficiently than longer ones but with more variability. As a result, even though the library complexity and depth increase, the average and N50 read lengths decrease, and this might impact the applications from these long reads in subsequent analysis, such as isoform-level transcriptome analysis.

	HPC reamplified (B11)	**HPC (B15)**
Mean read length	507.7	502.8
Mean read quality	14.8	14.8
Median read length	466.0	388.0
Median read quality	16.6	16.8
Number of reads	13,392,075.0	8,640,084.0
Read Length N50	585.0	679.0
Standard deviation of read length	287.7	395.2
Total bases	6,798,996,470.0	4,344,320,013.0
Mapped reads	97.04%	91.58%
Primary mapped reads	93.39%	82.77%

Table 2. Summary statistics of read quality and length for samples Hippocampus reamplified (B11) and without it (B15). Metrics include mean and median read length and quality, total number of reads, N50 read length, standard deviation of read length, total bases and mapping statistics (expressed as percentages).

StringTie was then used to assemble and quantify transcripts, allowing us to obtain expression profiles and identify isoform structures. The returned output is a GTF file with all the genes and transcript models assembled from the aligned reads, each of these transcripts also include its FPKM and TPM values (expression levels) (Figure 8).

Long-read transcriptome profiling reveals substantial isoform diversity in the *Fus* gene region, including both annotated and novel splicing events. Additionally, it returns more isoform than in the short read method. These data support complex alternative splicing regulation and highlight the advantages of long-read sequencing for comprehensive isoform discovery.

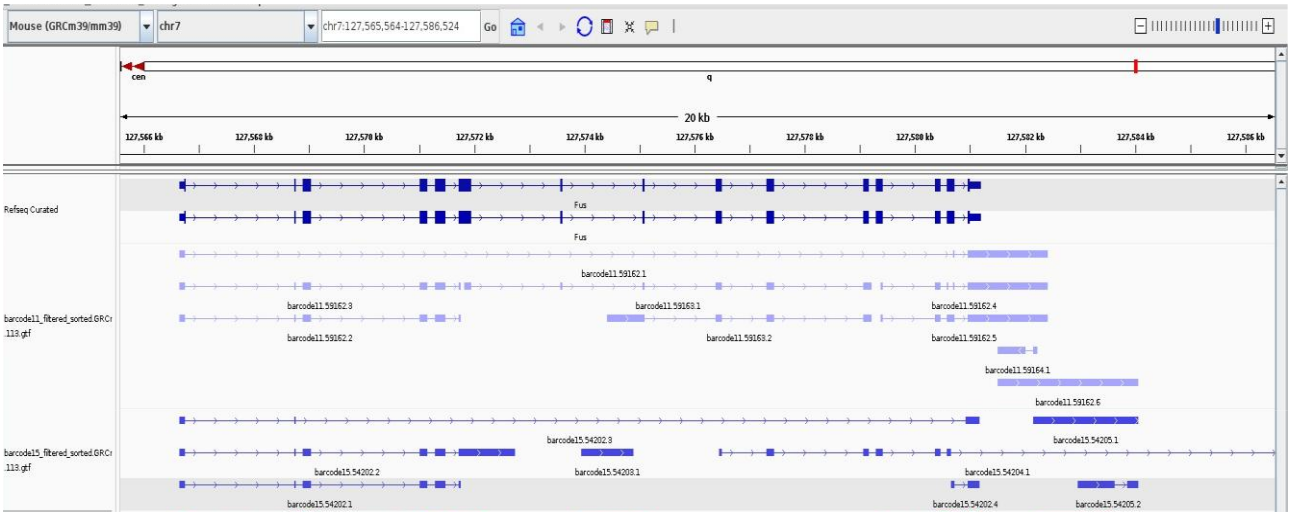


Figure 8. IGV screenshot displaying transcript isoforms of the *Fus* gene region (chr7:127,565,564–127,586,524) in the mouse genome (GRCm39/mm39).

4. DISCUSSION

Our study provides a thorough overview of the impact of the *FusΔ14* mutation on RNA processing in different tissues, using a knock-in mouse model that exhibits characteristics linked to juvenile ALS. Through the use of short- and long-read sequencing, we have developed an approach to untangle the transcriptomic consequences associated with aberrant *Fus* splicing to gain insight into the molecular consequences induced.

We also conducted an analysis comparing the differentially expressed genes (DEGs) with significant adjusted p-values across two different mouse genome annotations and emsambling: mm10 (from 2010) and mm39 (from 2019), gencode and refseq. The goal was to assess whether the choice of genome annotation and ensambling impacts the results, particularly regarding alignment and downstream analyses.

Although the updated annotation (mm39) includes additional genomic features, our results were largely consistent with those obtained using mm10. Differential expression analysis revealed a similar number of significantly altered genes across genome versions, as confirmed by the number of genes deregulated appeared in this article (Ali et al., 2023). Moreover, over-representation analysis (ORA) further supported the consistency of functional enrichment and overlapping biological pathways. Notably, the pathways identified in our analysis overlapped substantially with those reported in the referenced study (Ali et al., 2023). However, discrepancies emerged when comparing annotations from different sources. While the mm39 vs. mm10 comparison

within GENCODE showed high concordance (>75% overlap in DEGs), the comparison between mm39 GENCODE and mm39 RefSeq revealed much lower overlap (~60%), same goes when trying to compare splicing events. Thus, our findings underscore that annotation source has a greater impact on splicing event detection than genome assembly version, highlighting the importance of using a consistent annotation framework in splicing analyses. This is in line with previous human studies showing that conservative annotations like RefSeq yield more precise RNA-seq quantification than broader ones like Ensembl/GENCODE (Chisanga et al., 2022).

Consistent with prior findings (Devoy et al., 2017, Ali et al., 2023), we observed that homozygous *FusΔ14* mice exhibit strong changes in gene expression and alternative splicing compared to wild-type controls, across brain, spinal cord, liver and skeletal muscle.

The tissue-specificity of the splicing alterations, despite a common mutation background, is consistent with previous reports suggesting that *Fus*-mediated splicing is context-dependent. This is in line with findings by (Colombrita et al., 2012), who showed that *Fus* regulates distinct RNA targets in motoneuron-like cells depending on cellular context and RNA structure. Our findings show that although *Fus* expression levels differ between tissues, the disruption of its autoregulation caused by the mutation is maintained. Although exon skipping (SE) and retained introns (RI) were the most frequent splicing modifications, we also observed common alternative splice site usage (A3SS/A5SS) and mutually exclusive exons (MXE) in some genes among tissues. Genes such as *Fus*, a well-known RNA-binding protein and confirmed Fus target, were identified in the reference dataset (POSTAR3), along with additional targets such as *Tsc2, Commd4, Agrn*, while others such as *Sirt6, Knop1, Yipf2* that don't appear as Fus bound gene. But may still represent indirect downstream candidates given their shared splicing behavior across the four tissues suggest they warrant further investigation.

Transcript-level analysis added further resolution by showing isoform-specific changes in expression and usage. Interestingly, the short-read-based quantification did not reflect the expected exon 14 skipping, likely due to the limitations of current isoform reconstruction methods. This discrepancy emphasizes the need for more accurate long-read–based analysis.

To overcome the limitations of short-read RNA-seq, we optimized and applied Nanopore long-read sequencing on hippocampal RNA. Our results demonstrate that reamplification improves read yield and alignment rates, though at the cost of shorter fragment length. The resulting isoform landscape in the hippocampus revealed novel and annotated transcripts for the *Fus* gene, supporting a much greater diversity than detected via short-read methods.

These observations are consistent with results reported by (Sessegolo et al., 2019), who performed a systematic comparison of ONT RNA-seq, ONT cDNA-seq, and Illumina-based sequencing. They concluded that ONT direct RNA reached the most accurate transcript quantification. Specifically, Sessegolo et al. highlighted the use of ONT direct RNA-seq for measuring the abundance of transcripts and for identification of splicing isoforms. However, due to its lack of multiplexing kits and higher cost per sample, direct RNA-seq remains less scalable for large studies. Similarly, a study on Saccharomyces cerevisiae reported that both direct RNA and cDNA sequencing obtained comparable differential expression results (Chen et al., 2025) and (Mock et al., 2023) showed that shallow ONT cDNA-seq provides a viable alternative for fast cancer transcriptome profiling. Although limited in sequencing depth, this approach achieved high concordance with Illumina´s sequencing. Importantly, it offered a fast, low-cost, and portable setup, making it suitable for clinical and translational applications.

Taken together, these comparative studies evaluating each sequencing platforms support the use of ONT cDNA-seq as a robust and scalable method, in transcriptome complexity resolution.

Thus, our nanopore long-read approaches represent a crucial complement to short-read sequencing but still requires careful experimental design (an increasing the sample size (n) and incorporating biological replicates) and an optimized analysis pipelines to fully capture isoform-level regulation in disease models, and in this case of neurodegeneration pathologies like ALS-FTD.

In summary, our study emphasizes the role of *FUS* in preserving splicing fidelity and transcriptome integrity. Combining sequencing platforms and tools enabled discovery of novel splicing events and *FUS* targets. These findings also highlight the requirement for enhancements in technology and library prep protocols which may capture greater reads and full-length transcripts and allow the development of targeted RNA-based therapies.

5. CONCLUSIONS

1. Gene Expression Analysis with Short-Read Data

By reanalyzing the *Fus* mutant mouse dataset from (Ali et al., 2023; Devoy et al., 2017) using an updated reference genome (mm39) and an alternative computational pipeline, we demonstrated that modernized workflows can produce results largely consistent with those obtained from mm10, while offering enhanced genomic resolution. The updated annotations validated previously reported deregulated genes and uncovering subtle changes missed by earlier approaches.

2. Splicing Event Detection and Method Benchmarking

Using rMATS Turbo and IsoformSwitchAnalyzeR, we systematically benchmarked splicing detection tools and identified widespread alterations in splicing associated with *Fus* mutation. Among the most frequent events were skipped exons (SE) and retained introns (RI), with notable overlap between tissues. Importantly, splicing events involving ALS/FTD-relevant genes such as *Tardbp* were detected, strengthening the link between Fus dysfunction and RNA processing defects. Comparative analysis confirmed that combining tools offers complementary insights into splicing patterns and functional impact.

3. Optimization of ONT Long-Read RNA Sequencing

We have further adapted long-read cDNA sequencing via ONT's SQK-PCB114. 2A kit, resulting in enhanced library construction efficiency and sequencing accuracy. This enabled the revealing of complex transcript structures and novel isoforms not detectable with short-read data. Our results affirm the utility of ONT cDNA-seq for transcriptome characterization in neurodegenerative models.

6. REFERENCES

A method to remove amplified cDNA artefact in SQK-PCS111 and SQK-PCB111.24 libraries. (n.d.). Retrieved June 1, 2025, from https://nanoporetech.com/document/requirements/PCS_PCB_artefact

Ali, Z., Godoy-Corchuelo, J. M., Martins-Bach, A. B., Garcia-Toledo, I., Fernández-Beltrán, L. C., Nair, R. R., Spring, S., Nieman, B. J., Jimenez-Coca, I., Bains, R. S., Forrest, H., Lerch, J. P., Miller, K. L., Fisher, E. M. C., Cunningham, T. J., & Corrochano, S. (2023). Mutation in the FUS nuclear localisation signal domain causes neurodevelopmental and systemic metabolic alterations. *Disease Models & Mechanisms*, *16*(10), dmm050200. https://doi.org/10.1242/DMM.050200

Bang, J., Spina, S., & Miller, B. L. (2015). Frontotemporal dementia. *Lancet (London, England)*, *386*(10004), 1672–1682. https://doi.org/10.1016/S0140-6736(15)00461-4

Brooks, B. R., Miller, R. G., Swash, M., & Munsat, T. L. (2000). El Escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis*, *1*(5), 293–299. https://doi.org/10.1080/146608200300079536/ASSET//CMS/ASSET/7C2443A8-BC60-4A13-AC9C-B9A49B669E28/146608200300079536.FP.PNG

Burrell, J. R., Halliday, G. M., Kril, J. J., Ittner, L. M., Götz, J., Kiernan, M. C., & Hodges, J. R. (2016). The frontotemporal dementia-motor neuron disease continuum. *Lancet (London, England)*, *388*(10047), 919–931. https://doi.org/10.1016/S0140-6736(16)00737-6

Chen, Y., Davidson, N. M., Wan, Y. K., Yao, F., Su, Y., Gamaarachchi, H., Sim, A., Patel, H., Low, H. M., Hendra, C., Wratten, L., Hakkaart, C., Sawyer, C., Iakovleva, V., Lee, P. L., Xin, L., Ng, H. E. V., Loo, J. M., Ong, X., … Göke, J. (2025). A systematic benchmark of Nanopore long-read RNA sequencing for transcript-level analysis in human cell lines. *Nature Methods 2025 22:4*, *22*(4), 801–812. https://doi.org/10.1038/s41592-025-02623-4

Chisanga, D., Liao, Y., & Shi, W. (2022). Impact of gene annotation choice on the quantification of RNA-seq data. *BMC Bioinformatics*, *23*(1), 1–21. https://doi.org/10.1186/S12859-022-04644-8/FIGURES/7

Clark, M. B., Wrzesinski, T., Garcia, A. B., Hall, N. A. L., Kleinman, J. E., Hyde, T., Weinberger, D. R., Harrison, P. J., Haerty, W., & Tunbridge, E. M. (2020). Long-read sequencing reveals the complex splicing profile of the psychiatric risk gene CACNA1C in human brain. *Molecular Psychiatry*, *25*(1), 37–47. https://doi.org/10.1038/S41380-019-0583-1;TECHMETA=38,39,77,91;SUBJMETA=337,378,631;KWRD=MOLECULAR+BIOLOGY,NEUROSCIENCE

Colombrita, C., Onesto, E., Megiorni, F., Pizzuti, A., Baralle, F. E., Buratti, E., Silani, V., & Ratti, A. (2012). TDP-43 and FUS RNA-binding Proteins Bind Distinct Sets of Cytoplasmic Messenger RNAs and Differently Regulate Their Post-transcriptional Fate in Motoneuron-like Cells. *The Journal of Biological Chemistry*, *287*(19), 15635. https://doi.org/10.1074/JBC.M111.333450

DeJesus-Hernandez, M., Kocerha, J., Finch, N., Crook, R., Baker, M., Desaro, P., Johnston, A., Rutherford, N., Wojtas, A., Kennelly, K., Wszolek, Z. K., Graff-Radford, N., Boylan, K., & Rademakers, R. (2010). De novo truncating FUS gene mutation as a cause of sporadic amyotrophic lateral sclerosis. *Human Mutation*, *31*(5). https://doi.org/10.1002/HUMU.21241,

DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., Nicholson, A. M., Finch, N. C. A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G. Y. R., Karydas, A., Seeley, W. W., Josephs, K. A., Coppola, G., Geschwind, D. H., … Rademakers, R. (2011). Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron*, *72*(2), 245–256. https://doi.org/10.1016/J.NEURON.2011.09.011

Devoy, A., Kalmar, B., Stewart, M., Park, H., Burke, B., Noy, S. J., Redhead, Y., Humphrey, J., Lo, K., Jaeger, J., Mejia Maza, A., Sivakumar, P., Bertolin, C., Soraru, G., Plagnol, V., Greensmith, L., Acevedo Arozena, A., Isaacs, A. M., Davies, B., … Fisher, E. M. C. (2017). Humanized mutant FUS drives progressive motor neuron degeneration without aggregation in "FUSDelta14" knockin mice. *Brain : A Journal of Neurology*, *140*(11), 2797–2805. https://doi.org/10.1093/BRAIN/AWX248

Fujioka, Y., Ishigaki, S., Masuda, A., Iguchi, Y., Udagawa, T., Watanabe, H., Katsuno, M., Ohno, K., & Sobue, G. (2013). FUS-regulated region- and cell-type-specific transcriptome is associated with cell selectivity in ALS/FTLD. *Scientific Reports 2013 3:1*, *3*(1), 1–13. https://doi.org/10.1038/srep02388

Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipos, B., Lloyd, J. H., Bruce, M., Pantic, N., Admassu, T., James, P., Warland, A., Jordan, M., Ciccone, J., Serra, S., Keenan, J., Martin, S., McNeill, L., Wallace, E. J., Jayasinghe, L., Wright, C., … Turner, D. J. (2018). Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods*, *15*(3), 201–206. https://doi.org/10.1038/NMETH.4577

Hardiman, O., Al-Chalabi, A., Chio, A., Corr, E. M., Logroscino, G., Robberecht, W., Shaw, P. J., Simmons, Z., & Van Den Berg, L. H. (2017). Amyotrophic lateral sclerosis. *Nature Reviews. Disease Primers*, *3*. https://doi.org/10.1038/NRDP.2017.71

Heberle, B. A., Brandon, J. A., Page, M. L., Nations, K. A., Dikobe, K. I., White, B. J., Gordon, L. A., Fox, G. A., Wadsworth, M. E., Doyle, P. H., Williams, B. A., Fox, E. J., Shantaraman, A., Ryten, M., Goodwin, S., Ghiban, E., Wappel, R., Mavruk-Eskipehlivan, S., Miller, J. B., … Ebbert, M. T. W. (2023). *Using deep long-read RNAseq in Alzheimer's disease brain to assess medical relevance of RNA isoform diversity*. https://doi.org/10.1101/2023.08.06.552162

Ishigaki, S., & Sobue, G. (2018). Importance of functional loss of FUS in FTLD/ALS. *Frontiers in Molecular Biosciences*, *5*(MAY). https://doi.org/10.3389/FMOLB.2018.00044,

Kwiatkowski, T. J., Bosco, D. A., LeClerc, A. L., Tamrazian, E., Vanderburg, C. R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E. J., Munsat, T., Valdmanis, P., Rouleau, G. A., Hosler, B. A., Cortelli, P., De Jong, P. J., Yoshinaga, Y., Haines, J. L., Pericak-Vance, M. A., Yan, J., … Brown, R. H. (2009). Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science (New York, N.Y.)*, *323*(5918), 1205–1208. https://doi.org/10.1126/SCIENCE.1166066

Lagier-Tourenne, C., Polymenidou, M., Hutt, K. R., Vu, A. Q., Baughn, M., Huelga, S. C., Clutario, K. M., Ling, S. C., Liang, T. Y., Mazur, C., Wancewicz, E., Kim, A. S., Watt, A., Freier, S., Hicks, G. G., Donohue, J. P., Shiue, L., Bennett, C. F., Ravits, J., … Yeo, G. W. (2012). Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nature Neuroscience*, *15*(11), 1488–1497. https://doi.org/10.1038/NN.3230

Ling, S. C., Polymenidou, M., & Cleveland, D. W. (2013). Converging mechanisms in ALS and FTD: Disrupted RNA and protein homeostasis. *Neuron*, *79*(3), 416. https://doi.org/10.1016/J.NEURON.2013.07.033

Luisier, R., Tyzack, G. E., Hall, C. E., Mitchell, J. S., Devine, H., Taha, D. M., Malik, B., Meyer, I., Greensmith, L., Newcombe, J., Ule, J., Luscombe, N. M., & Patani, R. (2018). Intron retention and nuclear loss of SFPQ are molecular hallmarks of ALS. *Nature Communications 2018 9:1*, *9*(1), 1–15. https://doi.org/10.1038/s41467-018-04373-8

Maggi, J., Feil, S., Gloggnitzer, J., Maggi, K., Bachmann-Gagescu, R., Gerth-Kahlert, C., Koller, S., & Berger, W. (2024). Nanopore Deep Sequencing as a Tool to Characterize and Quantify Aberrant Splicing Caused by Variants in Inherited Retinal Dystrophy Genes. *International Journal of Molecular Sciences*, *25*(17). https://doi.org/10.3390/IJMS25179569,

Mehta, P. R., Brown, A. L., Ward, M. E., & Fratta, P. (2023). The era of cryptic exons: implications for ALS-FTD. *Molecular Neurodegeneration*, *18*(1), 1–9. https://doi.org/10.1186/S13024-023-00608-5/FIGURES/1

Mock, A., Braun, M., Scholl, C., Fröhling, S., & Erkut, C. (2023). Transcriptome profiling for

precision cancer medicine using shallow nanopore cDNA sequencing. *Scientific Reports 2023 13:1*, *13*(1), 1–11. https://doi.org/10.1038/s41598-023-29550-8

Naumann, M., Pal, A., Goswami, A., Lojewski, X., Japtok, J., Vehlow, A., Naujock, M., Günther, R., Jin, M., Stanslowsky, N., Reinhardt, P., Sterneckert, J., Frickenhaus, M., Pan-Montojo, F., Storkebaum, E., Poser, I., Freischmidt, A., Weishaupt, J. H., Holzmann, K., … Hermann, A. (2018). Impaired DNA damage response signaling by FUS-NLS mutations leads to neurodegeneration and FUS aggregate formation. *Nature Communications*, *9*(1). https://doi.org/10.1038/S41467-017-02299-1,

Prudencio, M., Belzil, V. V., Batra, R., Ross, C. A., Gendron, T. F., Pregent, L. J., Murray, M. E., Overstreet, K. K., Piazza-Johnston, A. E., Desaro, P., Bieniek, K. F., DeTure, M., Lee, W. C., Biendarra, S. M., Davis, M. D., Baker, M. C., Perkerson, R. B., Van Blitterswijk, M., Stetler, C. T., … Petrucelli, L. (2015). Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nature Neuroscience*, *18*(8), 1175–1182. https://doi.org/10.1038/NN.4065

Rogelj, B., Easton, L. E., Bogu, G. K., Stanton, L. W., Rot, G., Curk, T., Zupan, B., Sugimoto, Y., Modic, M., Haberman, N., Tollervey, J., Fujii, R., Takumi, T., Shaw, C. E., & Ule, J. (2012). Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. *Scientific Reports 2012 2:1*, *2*(1), 1–10. https://doi.org/10.1038/srep00603

Rowland, L. P., & Shneider, N. A. (2001). Amyotrophic lateral sclerosis. *The New England Journal of Medicine*, *344*(22), 1688–1700. https://doi.org/10.1056/NEJM200105313442207

Sama, R. R. anjith K., Ward, C. L., & Bosco, D. A. (2014). Functions of FUS/TLS from DNA repair to stress response: implications for ALS. *ASN Neuro*, *6*(4). https://doi.org/10.1177/1759091414544472,

Sessegolo, C., Cruaud, C., Da Silva, C., Cologne, A., Dubarry, M., Derrien, T., Lacroix, V., & Aury, J. M. (2019). Transcriptome profiling of mouse samples using nanopore sequencing of cDNA and RNA molecules. *Scientific Reports*, *9*(1). https://doi.org/10.1038/S41598-019-51470-9,

Sharma, A., Lyashchenko, A. K., Lu, L., Nasrabady, S. E., Elmaleh, M., Mendelsohn, M., Nemes, A., Tapia, J. C., Mentis, G. Z., & Shneider, N. A. (2016). ALS-associated mutant FUS induces selective motor neuron degeneration through toxic gain of function. *Nature Communications*, *7*. https://doi.org/10.1038/NCOMMS10465

Soneson, C., Yao, Y., Bratus-Neuenschwander, A., Patrignani, A., Robinson, M. D., & Hussain, S. (2019). A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. *Nature Communications 2019 10:1*, *10*(1), 1–14. https://doi.org/10.1038/s41467-019-11272-z

Soto, C., & Pritzkow, S. (2018). Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases. *Nature Neuroscience 2018 21:10*, *21*(10), 1332–1340. https://doi.org/10.1038/s41593-018-0235-9

Swinnen, B., & Robberecht, W. (2014). The phenotypic variability of amyotrophic lateral sclerosis. *Nature Reviews Neurology 2014 10:11*, *10*(11), 661–670. https://doi.org/10.1038/nrneurol.2014.184

Van Langenhove, T., Van Der Zee, J., & Van Broeckhoven, C. (2012). The molecular basis of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum. *Annals of Medicine*, *44*(8), 817–828. https://doi.org/10.3109/07853890.2012.665471,

Vitting-Seerup, K., Sandelin, A., & Berger, B. (2019). IsoformSwitchAnalyzeR: Analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics*, *35*(21), 4469–4471. https://doi.org/10.1093/BIOINFORMATICS/BTZ247,

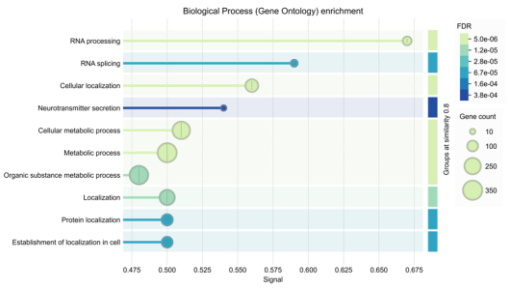
Wang, Y., Xie, Z., Kutschera, E., Adams, J. I., Kadash-Edmondson, K. E., & Xing, Y. (2024). rMATS-turbo: an efficient and flexible computational tool for alternative splicing analysis of large-scale RNA-seq data. *Nature Protocols 2024 19:4*, *19*(4), 1083–1104. https://doi.org/10.1038/s41596-023-00944-2

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57–63. https://doi.org/10.1038/NRG2484;KWRD=BIOMEDICINE

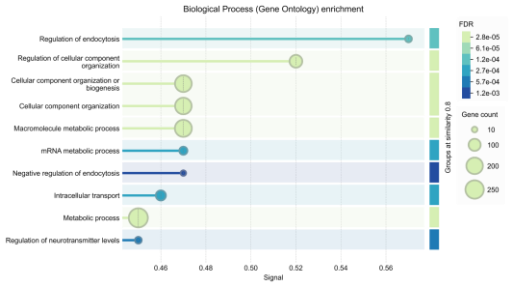
Zakaryan, R. P., & Gehring, H. (2006). Identification and Characterization of the Nuclear Localization/Retention Signal in the EWS Proto-oncoprotein. *Journal of Molecular Biology*, *363*(1), 27–38. https://doi.org/10.1016/j.jmb.2006.08.018

Supplementary figures

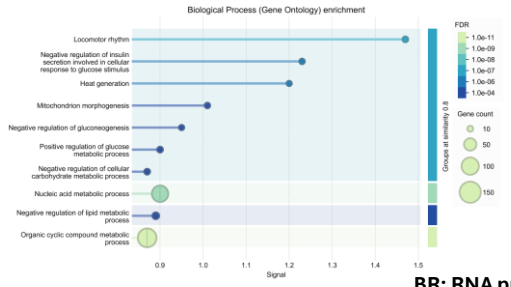
A3SS



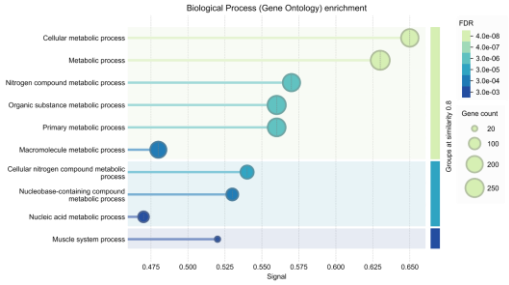
BRAIN



SPINAL CORD



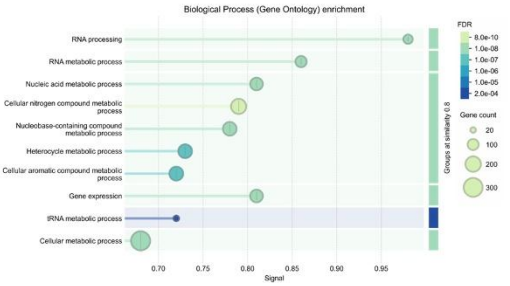
LIVER



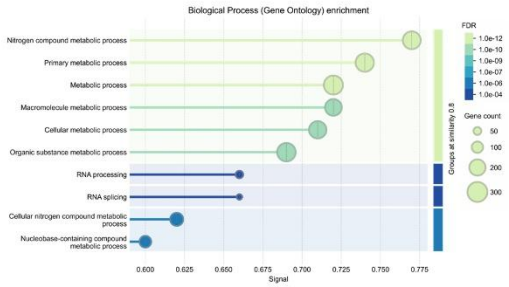
TIBIALIS

BR: RNA processing
SPC: Regulation of endocytosis
LIV: Locomotor rhythm
TA: Cellular metabolic process

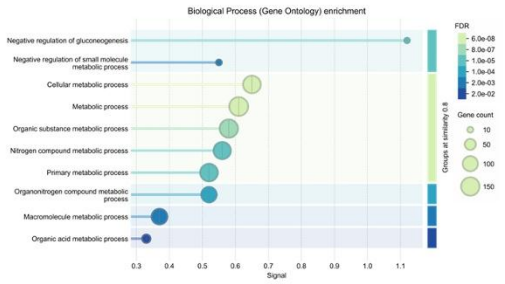
A5SS



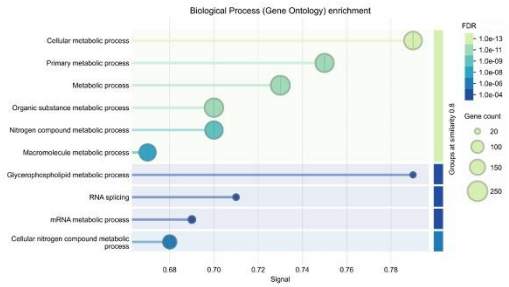
BRAIN



SPINAL CORD

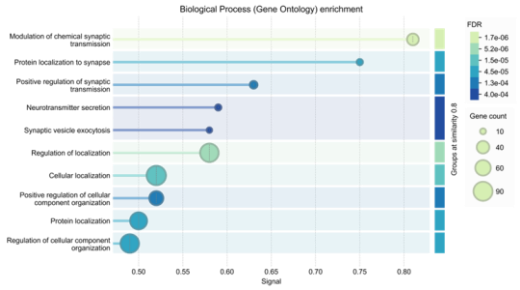


LIVER

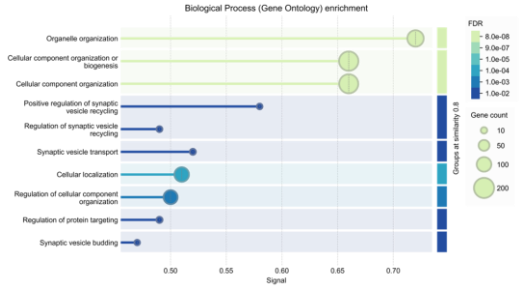


BR: RNA processing
SPC: Metabolic process of nitrogen compounds
LIV: Negative regulation of gluconeogenesis
TA: Cellular metabolic process

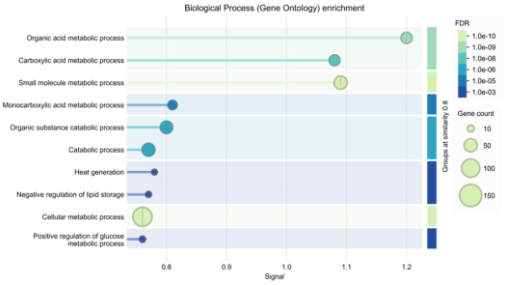
MXE



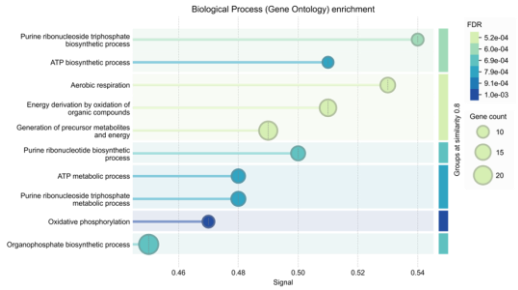
BRAIN



SPINAL CORD



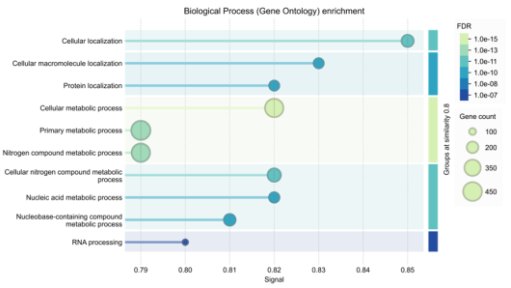
LIVER



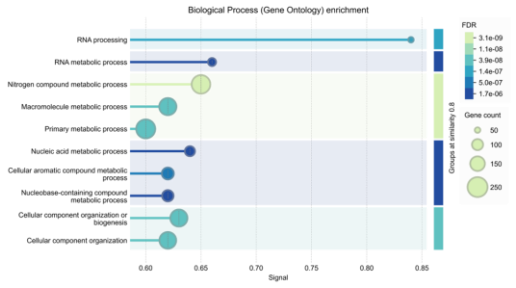
TIBIALIS

BR: Modulation of chemical synaptic transmission
SPC: Organelle organization
LIV: Organic acid metabolic process
TA: Purine ribonucleoside triphosphate biosynthetic process

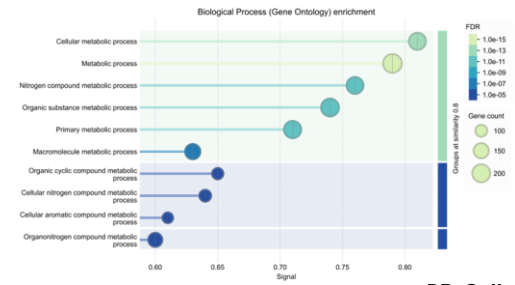
RI



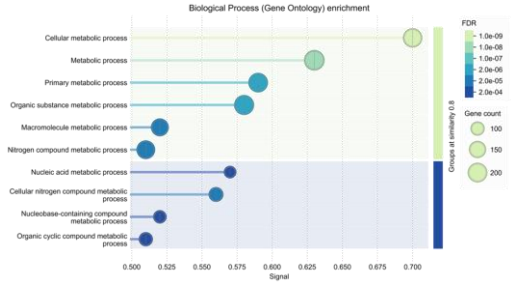
BRAIN



SPINAL CORD



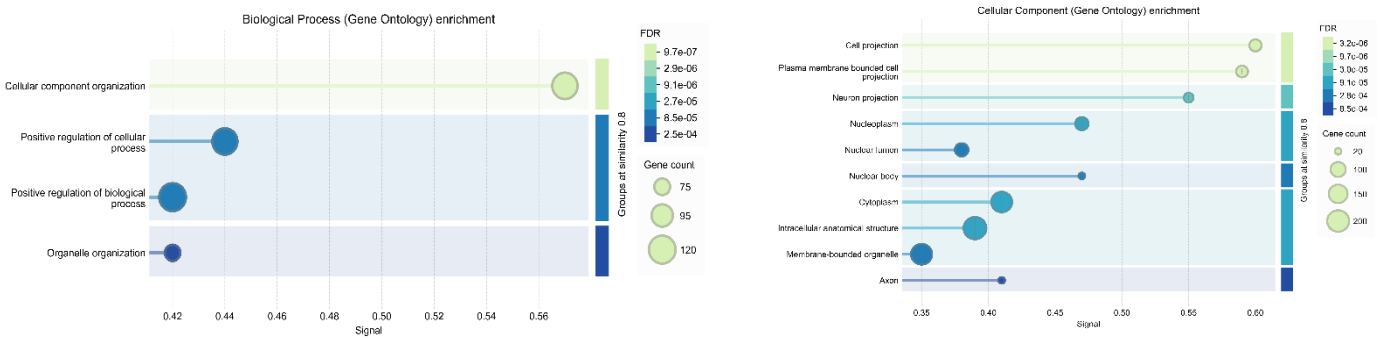
LIVER



TIBIALIS

BR: Cellular localization
SPC: RNA processing
LIV: Cellular metabolic process
TA: Cellular metabolic process

Figure S1. GO Biological Process enrichment (STRING) of genes with significant alternative splicing events detected by rMATS (FDR < 0.05, $|\Delta\Psi| > 0.05$) in brain (top left), spinal cord (top right), liver (bottom left), and tibialis anterior (bottom right). Dot size indicates gene count; color represents FDR significance.



Supplementary Figure S2. GO enrichment in brain for genes overlapping splicing and expression changes. Enriched terms include *cell projection* and *neuron projection*. Dot size reflects gene count; color indicates FDR.

Splicing Event	Total Genes	Fus Targets	Fus Target Genes	Non-targets	Non-target Genes
A3SS	3	3	Commd4, Tsc2, Agrn	0	
A5SS	3	2	Mrps17	1	Knop1, Sirt6
MXE	6	6	2210408F21Rik, Dram2, Fam149b, Fam193b, Meg3, Firre	0	
RI	19	10	Dgat1, Fubp1, Gigyf1, Kifc2, Med1, Ndor1, Ppp1r12c, Prpf4b, Scrib, Tardbp	9	A230072C01Rik, Arfip2, Ccdc163, Cpsf1, Farsa, Foxred1, Rdh5, Slc4a2, Yipf2
SE	30	24	2210408F21Rik, Ap3b1, Atf2, C2cd5, Cops7b, Crem, Dap3, Dram2, Fam193b, Firre, Fus, Hnrnpd, Klhl17, Mtg2, Ncor2, Nbr1, Pank2, Sh3glb2, Snrk, Snx10, Srsf11, Timm9, Mmaa, Trit1	6	Hpf1, Tle2, Ntan1, Scly, Sumf2, Timm10b

Table S1. Summary of alternatively spliced genes by event type and FUS target status. Genes are grouped by splicing event type and classified as FUS targets or non-targets based on POSTAR3 annotation.

tissue	splicing event			strand	PValue	Padj	PSI	Prom HOM	Prom WT
BR	A5SS	Fus	chr7	+	0	0	-0,883	0.036,0.236,0.013,0.184	1.0,1.0,1.0,1.0
	RI	Fus	chr7	+	0	0	0,949	1.0,1.0,1.0,1.0	0.051,0.012,0.087,0.056
SPC									
LIV	SE	Fus	chr7	+	0	0	0,084	1.0,0.969,1.0,0.992	0.969,0.976,0.922,0.756
	SE	Fus	chr7	+	1,3279E-07	5,71288E-05	0,226	1.0,0.888,1.0,0.959	0.9,0.909,0.752,0.384
TA	A5SS	Fus	chr7	+	0	0	-0,846	0.136,0.189,0.182,0.11	1.0,1.0,1.0,1.0
	RI	Fus	chr7	+	0	0	0,935	1.0,1.0,1.0,1.0	0.079,0.026,0.052,0.104

Table S2. Summary of *Fus* of found events across tissues (BR, SPC, LIV, TA) from short-read RNA-seq data. Shown are p-values, PSI, and promoter expression in homozygous mutant (HOM) vs. wild-type (WT).