**Distribution Agreement**

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Xianglin Shi Date

Potential biomarkers for ALS/FTD as a consequence of TDP-43 loss of function

By

Xianglin Shi

Master of Public Health

Department of Biostatistics and Bioinformatics

_____

Steve Qin, PhD

(Thesis Advisor)

_____

Jie Jiang, PhD

(Reader)

# Potential biomarkers for ALS/FTD as a consequence of TDP-43 loss of function

By

Xianglin Shi

B.S., China Pharmaceutical University, 2020

Thesis Committee Chair: Steve Qin, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Biostatistics
2023

**Abstract**

Potential biomarkers for ALS/FTD as a consequence of TDP-43 loss of function

**Background:** Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are two distinct yet frequently intersecting neurodegenerative disorders that share common pathological mechanisms. The linkage of ALS and FTD is supported by a shared pathological, the cytoplasmic aggregation and nuclear clearance of a DNA/RNA binding protein called transactive response DNA binding protein 43 kDa (TDP-43, TARDBP) (Ling, S. et al.,2013). TDP-43 has been proven to suppress the incorporation of cryptic exons during splicing. The resulting mRNAs may be subjected to nonsense-mediated decay (NMD) if there is a premature stop codon, leading to loss of regular expression (Brown et al., 2022; Klim et al., 2019; Ma et al., 2022; Melamed et al., 2019). On the other hand, the abnormal mRNAs with cryptic exon inclusions can produce a new stretch of amino acids in the encoding protein only in most ALS/FTD patients with loss of nuclear TDP-43, but not in healthy humans. Such novel peptides represent better biomarker candidates because the signal-to-noise ratio will be high.

**Objectives:** This project aims to identify potential ALS/FTD biomarkers based on our understanding of the recently revealed TDP-43 function in cryptic exon repression.

**Methods:** Alternative splicing analysis and differential gene expression analysis were applied to find the cryptic exons and related genes. We used the IPSCN cell with TDP-43 knockdown model compared with the control group.

**Results:** We identified 67 genes with alternative splicing events in both MAJIQ and LeafCutter, among those genes 33 with detectable gene expression changes (Log2Folder change > 0.2). Confirmed genes can be identified in CSF or as a secreted protein. There were 12 genes that met the conditions, reconfirmed their expression among tissues. Finally, 6 potential genes can be our candidate biomarkers.

**Conclusions:** In this study, we used bioinformatics tools to build a pipeline to find biomarker candidates from RNA-seq data. We found 6 potential genes that can be our candidate biomarkers. This way improves the progress in developing new biomarkers.

# Potential biomarkers for ALS/FTD as a consequence of TDP-43 loss of function

By

Xianglin Shi

B.S., China Pharmaceutical University, 2020

Thesis Committee Chair: Steve Qin, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Biostatistics
2023

**Acknowledgement**

To begin with, I express my gratitude to the Rollins School of Public Health at Emory University for their continuous support throughout my academic journey. Pursuing a two-year study at RSPH enabled me to establish a robust knowledge base in Biostatistics, which has been invaluable to my academic and professional growth. Moreover, this program helped me discover my research interests in bioinformatics and set me on a career path that aligns with my passions and aspirations.

I am deeply appreciative of my thesis advisor, Steve Qin, and my reader Jie Jiang. They provided me the opportunity to conduct research in bioinformatics and shared their experiences and thoughts. Encouraged me when I am facing troubles.

I also want to thank all my friends in Rollins, they give me support and kindness when I am feeling disappointed about my life and study. Especially my friend Siwei Deng, he gave me a lot of guidance in bioinformatics tools, and emotional support. Without him I cannot finish all these works.

Lastly, I would like to express gratitude towards myself. Though I tend to have a negative outlook, worrying about the uncertainties of the future and the possibility of chaos in the world. Despite not always treating my body the best, I appreciate the way it continues to support my life, providing me with the energy to think critically and analyze situations. My effort is very small in the whole of human history, if it can help a little bit in the future, then it doesn't be wasted.

## Table of contents

## 1. Introduction

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are two distinct yet frequently intersecting neurodegenerative disorders that share common pathological mechanisms. ALS predominantly affects motor neurons, leading to gradual muscle weakness, whereas FTD is characterized by alterations in behavior, language, and executive functioning. The linkage of ALS and FTD is supported by a shared pathological feature in patient postmortem tissues, i.e., the cytoplasmic aggregation and nuclear clearance of a DNA/RNA binding protein called transactive response DNA binding protein 43 kDa (TDP-43, TARDBP) (Ling, S. et al.,2013). Approximately 97% of ALS patients and 50% of FTD patients display the deposition of TDP-43 aggregates accompanying nuclear mis-localization (Scotter et al.,2015; Jo. M., et al.,2020).

Diagnosis of ALS and FTD is primarily based on physician observations together with some neurophysiological examinations such MRI scans. However, it is difficult to diagnose patients in the early stage. Nuclear clearance of TDP-43, in the absence of TDP-43 inclusions, has been associated with early neuronal atrophy in ALS/FTD patients (Nana et al., 2019). In addition, ALS/FTD disease severity correlates with the abundance of TDP-43 pathology (Cathcart et al., 2021; Mackenzie et al., 2015). These observations suggest that molecular consequences resulting from TDP-43 loss of function may represent robust disease biomarkers for diagnosing and/or tracking disease progression.

Several prior studies have established that one of the roles of TDP-43 is to suppress the incorporation of cryptic exons during the splicing process. Cryptic exons are located in introns and are normally excluded from the mature mRNAs (Ma, X.R. et al., 2022). In instances where TDP-43 is absent from the nucleus, these cryptic exons are spliced into the messenger RNA. The

**Commented [JJ1]:** Please reinsert these references

resulting mRNAs may be subjected to nonsense-mediated decay (NMD) if there is a premature stop codon, leading to loss of normal expression. Alternatively, they may produce truncated proteins in the absence of premature termination or larger proteins if no frameshifting (Fig. 1.1). Several key RNA targets of TDP-43, including STMN2 and UNC13A, are shown to include cryptic exons and induce NMD in ALS/FTD patients (Brown et al., 2022; Klim et al., 2019; Ma et al., 2022; Melamed et al., 2019). Restoring the expression of normal STMN2 and UNC13A may represent novel therapeutic approaches. However, TDP-43 interacts with and regulates the splicing of thousands of target mRNAs. It remains to be seen whether correcting the expression level of one or two target mRNAs is sufficient to offer tangible clinical benefits while many other downstream pathways remain dysfunctional. On the other hand, the abnormal mRNAs with cryptic exon inclusions can produce a new stretch of amino acids in the encoding protein only in the majority of ALS/FTD patients with loss of nuclear TDP-43, but not in healthy humans. Such novel peptides represent better biomarker candidates because the signal-to-noise ratio will be high.

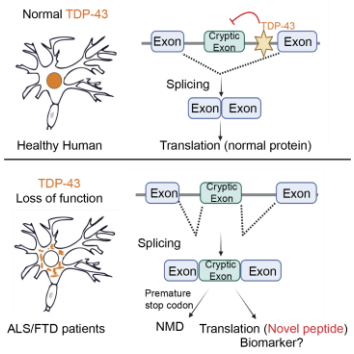



**Figure 1.1 Proteins with new peptides resulting from cryptic exon inclusions when TDP-43 is lost may serve as novel disease biomarkers.** One normal function of TDP-43 is to suppress the splicing of cryptic exons located in introns. Nuclear depletion of TDP-43 results in the inclusion of these cryptic exons in mature mRNAs, which can produce truncated or larger proteins with novel peptides.

This project aims to identify potential ALS/FTD biomarkers based on our understanding of the recently revealed TDP-43 function in cryptic exon repression. Given the commonality of TDP-43 pathology in ALS/FTD, loss of TDP-43 in early disease, and its correlation with disease severity, this study has great promise to fill the gap in the current research of ALS biomarkers.

## 2. Methods

### 2.1 RNA-Seq data

In this study, publicly RNA-sequencing dataset was downloaded from using Gene Expression Omnibus (GEO). This dataset is generated from TDP-43 knockdown model to explore common and unique transcriptional patterns as the result of TDP-43 loss. An IPSC neuronal line was used to detect CE inclusion in TDP-43 knockdown model. Remarkably, TDP-43-deficient human iPSC neuronal models effectively imitate CE events, which were subsequently confirmed in individuals afflicted with frontotemporal dementia/amyotrophic lateral sclerosis (Maize C. Cao et al., 2022).

### 2.2 Preparing of RNA-seq data

FASTQ files were downloaded from NCBI using Sequence Read Archive (SRA) Toolkit (version 2.9.6) and quality control was conducted using FastQC(version 0.12.1)(Mary E. Piper et al., 2022). Samples were assessed based on per base sequence quality, with the threshold set at Phred score of at least 30. Samples were excluded from analysis if they had no reads that survived trimming. If data still contained adapter sequence content were trimmed with Fastp(version 0.23.2)(Shifu Chen et al., 2018) aligned to the GRCh38 genome building using STAR (version 2.7.10a) with gene models from GENCODE V40, samtools (version 1.15.1)( Petr Danecek et al., 2021) was used to sort BAM files.

### 2.3 Differential gene expression analysis

Data was quantified using Feature Counts (version 2.0.3) with gene model from GENCODE V40. Differential expression analysis was performed using DESeq2(version 1.36.0) workflow in R (version 4.2.0) (Michael I Love et al., 2014).

## 2.4 Alternative splicing analysis

We utilized two pipelines for splicing analysis, MAJIQ and Leafcutter.

MAJIQ (version 2.4) (Vaquero-Garcia, J. et al., 2016) was used to identify junction-spanning reads. Distinct local splice variations (LSVs) were identified and estimated the fraction of each junction in each LSV providing precent spliced in (PSI). The PSI changes between TDP-43 negative samples and TPD-43 positive samples, was calculated with delta PSI(ΔPSI)

LeafCutter (version 0.2.9) (Li, Y et al., 2018) reads span exon–exon junctions. Intron clustering was performed using the default settings. Differential excision of the introns between the two conditions (TDP-43-positive neuronal nuclei versus TDP-43-negative neuronal nuclei) were calculated using leafcutter_ds.R. Result graph was generated using Shinyapp.io and prepare_results.R.

## 2.5 Statistical analysis

DESeq2 *p*-values were calculated using Wald tests and corrected for multiple testing using the Benjamini-Hochberg method. Genes with Padj < 0.05 were considered significantly changed.

Log2fold change was calculated to show the ratio of gene expression compared with TDP-43 negative to TDP-43 positive. The value of log2fold > 0.2 was defined as detectable expression changes. Statistical analyses were performed using R (version 4.2.0).

**Commented [QZ2]:** 43?

**Commented [QZ3]:** This is not right!

# 3. Results

## 3.1 Alternative splicing events include cryptic exons in TDP-43 knockdown model

To identify potential cryptic exons regulated by depletion of nuclear TDP-43, we utilized IPSC-neuron RNA-seq datasets. This data set includes 3 TDP-43 knockdown samples and 4 control samples. Two pipelines were used to find the cryptic exons, MAJIQ and LeafCutter. We identified 106 alternative splicing events (abs (ΔPSI)>0.2, Confidence >0.95) with MAJIQ. And identified 341 alternative splicing events and 109 events marked as cryptic exons with LeafCutter (abs (ΔPSI)>0.2, P<0.05). There were 67 genes identified in both analyses, including STMN2, and UNC13A that have been validated as subject to regulation by TDP-43 with regards to alternative splicing (Ma, X.R. et al., 2022).

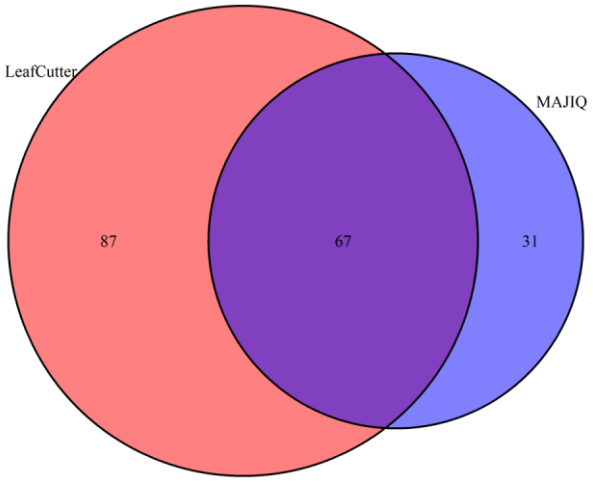



**Figure 3.1.1** The overlap gene number between LeafCutter and MAJIQ

The two pipelines use different definitions for transcript variations and different criteria to control for false positive. The overlap part has at least one region of genes that both pipelines

identified as cryptic exons. The top 25 genes with the most significant p-value were shown in Table 3.1.1.

**Table 3.1.1** Top 25 genes in alternative splicing

| Index | Gene Symbol | Gene Name | Annotation* | p-value** |
|---|---|---|---|---|
| 1 | ACTL6B | Actin like 6B | Cryptic | 2.00E-09 |
| | | | Cryptic | 3.87E-02 |
| 2 | ATG4B | Autophagy related 4B cysteine peptidase | Cryptic | 2.02E-09 |
| 3 | CAMK2B | Calcium/calmodulin dependent protein kinase II beta | Cryptic | 6.74E-10 |
| | | | Cryptic | 4.41E-06 |
| 4 | CBARP | CACN subunit beta associated regulatory protein | Cryptic | 7.29E-09 |
| | | | Annotated | 5.58E-03 |
| 5 | CELF5 | CUGBP Elav-like family member 5 | Cryptic | 9.72E-09 |
| 6 | DDR1 | Discoidin domain receptor tyrosine kinase 1 | Annotated | 2.64E-11 |
| 7 | FAM66C | Family with sequence similarity 66 member C | Cryptic | 4.36E-09 |
| 8 | G2E3 | G2/M-phase specific E3 ubiquitin protein ligase | Cryptic | 3.91E-09 |
| 9 | G3BP1 | G3BP stress granule assembly factor 1 | Annotated | 7.29E-09 |
| 10 | GOLGA7B | Golgin A7 family member B | Cryptic | 2.41E-08 |
| 11 | GRAMD1A | GRAM domain containing 1A | Cryptic | 5.21E-08 |
| 12 | KALRN | Kalirin RhoGEF kinase | Cryptic | 5.49E-10 |
| | | | Annotated | 3.47E-03 |
| | | | Annotated | 3.89E-02 |
| 13 | KLHL18 | Kelch like family member 18 | Annotated | 7.29E-09 |
| 14 | MAST1 | Microtubule associated serine/threonine kinase 1 | Cryptic | 2.95E-09 |
| 15 | MIAT | Myocardial infarction associated transcript | Cryptic | 1.44E-16 |
| | | | Annotated | 3.04E-02 |
| 16 | MYO18A | Myosin XVIIIA | Cryptic | 2.64E-11 |
| 17 | NECAB2 | N-terminal EF-hand calcium binding protein 2 | Cryptic | 8.88E-08 |
| 18 | POLDIP3 | DNA polymerase delta interacting protein 3 | Annotated | 2.31E-07 |

| | | | | |
|---|---|---|---|---|
| 19 | PSD | Pleckstrin and Sec7 domain containing | Cryptic | 1.10E-07 |
| 20 | SEMA4D | Semaphorin 4D | Cryptic<br>Annotated | 3.67E-08<br>4.08E-03 |
| 21 | SETD5 | SET domain containing 5 | Cryptic<br>Annotated | 9.78E-10<br>4.08E-03 |
| 22 | SHANK1 | SH3 and multiple ankyrin repeat domains 1 | Cryptic | 2.42E-07 |
| 23 | SMARCA4 | SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4 | Annotated | 6.96E-11 |
| 24 | SPEG | Striated muscle enriched protein kinase | Cryptic | 2.15E-09 |
| 25 | TRAPPC12 | Trafficking protein particle complex subunit 12 | Cryptic | 1.15E-07 |

*Annotation: shows the alternative splicing event happen in cryptic exon or annotated in normal regulation.

**p-value: Benjamini-Hochberg p-value

As observed in Table 3.1.1, a single gene may undergo multiple alternative splicing events. To effectively narrow down the pool of candidate genes, it is imperative to validate the target genes via gene expression analysis, with the aim of identifying those that exhibit detectable changes between the TDP-43 knockdown group and the control group.

## 3.2 Confirm gene expression changes in TDP-43 knockdown model

To confirm the gene expression differences were determined by TDP-43 knockdown, a principal component analysis was performed. From Figure 3.2.1, the first principal component shown on the x-axis indicates 63.31% variation, which distinguishes the TDP-43 knockdown and control groups.

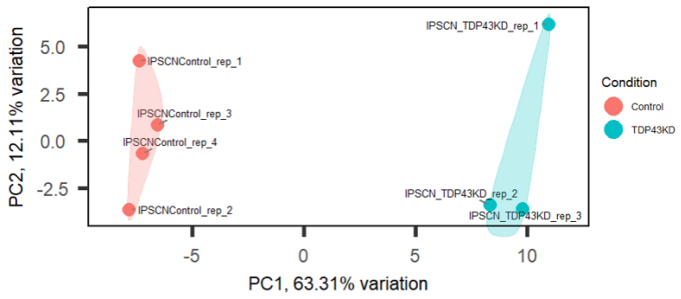



**Figure 3.2.1** Principal component analysis of IPSCN cell

We examined differential gene expression between TDP-43 negative samples and TDP-43 positive samples. From Figure 3.2.2 there were 33 genes with significant and detectable gene expression changes, 19 lower-level expression and 14 genes had high-level expression in TDP43-negative samples.

**Commented [QZ4]:** Should sort this list.

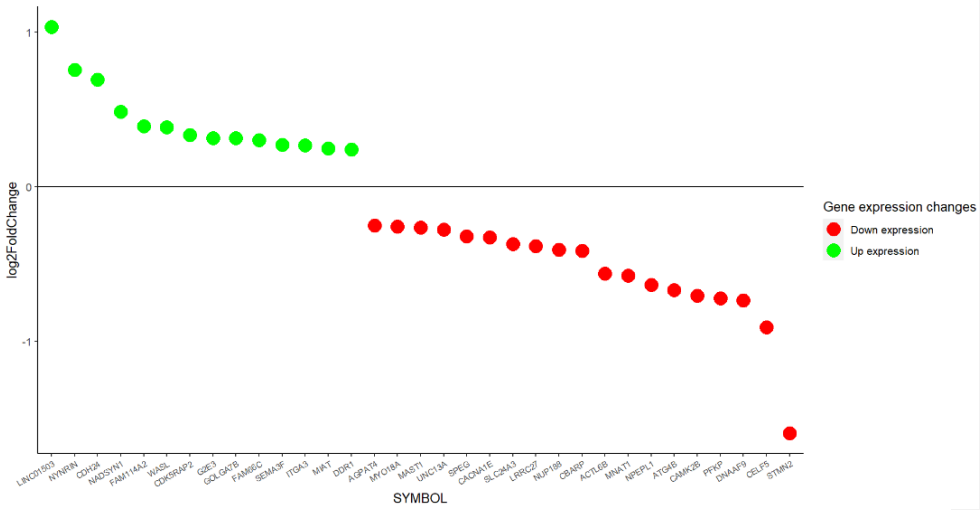



**Figure 3.2.2** Gene expression difference TDP-43 knockdown vs control group

The gene expression levels of STMN2 and LINC01503 demonstrate significant differences between the groups under investigation. The STMN2 gene encodes a protein that exerts regulatory effects on neuronal growth and is also implicated in osteogenesis. Previous studies have linked decreased expression of STMN2 with Down's syndrome and Alzheimer's disease. In contrast, LINC01503 is a gene that has been associated with Squamous Cell Carcinoma and does not generate any peptides. Next, we need to consider whether the gene can transcribe peptides and proteins and whether the proteins can be detected in cerebrospinal fluid or other body fluids.

## 3.3 Confirm peptide in CSF and Secretome datasets

ALS and FTD are characterized by the selective loss of neurons. In the early stages of the disease, the detection of genes in the brain is challenging. This is primarily due to the limited accessibility of most genes, which are exclusively expressed in brain tissue and are not readily detectable while the patient is alive. Therefore, it is imperative to identify candidate genes that can be detected in brain tissue, as well as secreted proteins or those that can be measured in cerebrospinal fluid, to facilitate early diagnosis and disease monitoring.

In order to select candidate genes that are potentially secreted from the cell or detectable in cerebrospinal fluid (CSF), we performed validation of the previously identified candidate genes in two separate datasets (Macron, C. et al.,2018; Chen et al., 2019). Following this validation, a total of 12 genes were selected that were at least identified in one of the datasets.

**Table 3.3.1** 12 Genes can be detected in CSF or Secretome

| Index | Gene Synbol | CSF Proteome* | Secretome** | Related Diseases or Biological Functions*** |
|---|---|---|---|---|
| 1 | STMN2 | Yes | No | Down's syndrome and Alzheimer's disease |
| 2 | NYNRIN | Yes | No | Hereditary Wilms' Tumor |

| | | | | |
|---|---|---|---|---|
| 3 | CBARP | Yes | No | Spinal Muscular Atrophy, Type Ii and Rapp-Hodgkin Syndrome |
| 4 | WASL | Yes | No | WASL include Wiskott-Aldrich Syndrome and Buruli Ulcer |
| 5 | UNC13A | Yes | No | Amyotrophic Lateral Sclerosis and Autism |
| 6 | DDR1 | Yes | Yes | Spondylometaepiphyseal Dysplasia, Short Limb-Hand Type and Breast Cancer |
| 7 | CACNA1E | Yes | No | Developmental And Epileptic Encephalopathy and Van Der Woude Syndrome |
| 8 | MYO18A | Yes | No | Atypical Chronic Myeloid Leukemia, Bcr-Abl1 Negative and Chronic Tic Disorder |
| 9 | SPEG | Yes | No | Myopathy, Centronuclear and Myopathy, Centronuclear |
| 10 | SEMA3F | Yes | Yes | Neuroma and Megacolon |
| 11 | LRRC27 | Yes | No | An important paralog of this gene is SCRIB |
| 12 | NADSYN1 | Yes | No | Vertebral, Cardiac, Renal, And Limb Defects Syndrome and Congenital Vertebral-Cardiac-Renal Anomalies Syndrome |

*CSF preteome constructed in PRIDE database.

**Secretum in the SPRomeDB searchable database.

*** Related Diseases or biological functions for each gene was gotten from GeneCards websit.

To validate the gene expression of the 12 proposed genes across diverse tissues, we turned to the GTEx portal, TPM>=15 was considered as high expression in this tissue. The neurological features of ALS typically involve the degeneration of both upper and lower motor neurons, as well as the brain stem and corticospinal tracts. In contrast, FTD is characterized by a progressive decline in the temporal and frontal cortex. As illustrated in Figure 3.3.1, the expression profiling revealed that most of the candidate genes were primarily expressed in brain tissue, while MYO18A and SPEG displayed relatively high expression in muscle tissue. Additionally,

MYO18A and NASYN1 were found to exhibit high expression levels in whole blood. Conversely, LRRC27 and SEMA3F were disregarded due to their inadequate expression levels in brain and muscle tissue.

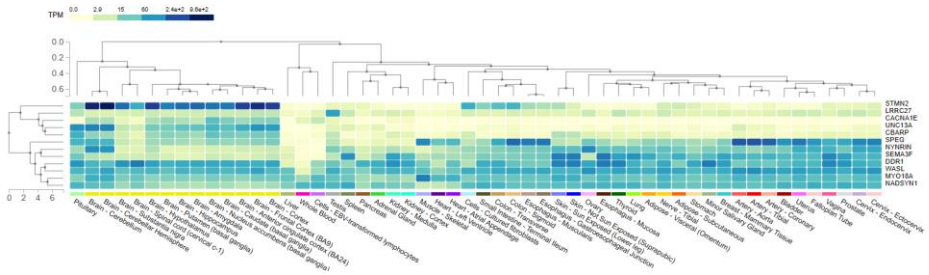
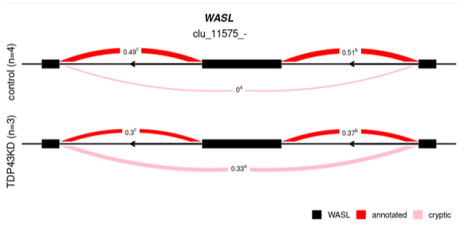



**Figure 3.3.1** Gene expression in different tissue for 12 candidate genes

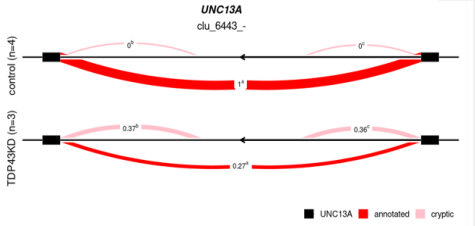
Ideally, in control samples, the occurrence of cryptic exons should either be negligible or completely absent. In this study, we have carefully chosen six specific genes as targets for investigation. As illustrated in Figure 3.3.2, among these genes, namely, WASL, UNC13A, and SPEG, no evidence of alternative splicing events with cryptic exon was observed in the control group but was observed in the TDP-43 knockdown group. In contrast, MYO18A, NADSYN1and STMN2 exhibited a minimal number of alternative splicing events in the control group but showed a significant increase in the TDP-43 knockdown group. These genes can be candidates biomarkers in diagnosing ALS/FTD patients.

a



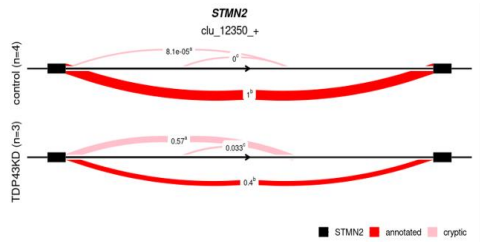

| | chr | start | end | verdict | ΔPSI |
|---|---|---|---|---|---|
| a | chr7 | 123684580 | 123692347 | novel annotated pair | 0.325 |
| b | chr7 | 123689150 | 123692347 | annotated | -0.169 |
| c | chr7 | 123684580 | 123689042 | annotated | -0.156 |

b



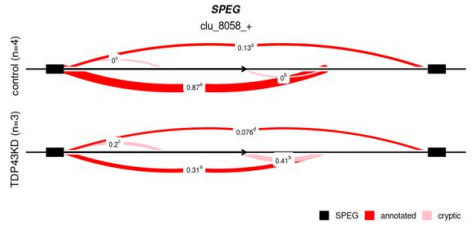

| | chr | start | end | verdict | ΔPSI |
|---|---|---|---|---|---|
| a | chr19 | 17641556 | 17642845 | annotated | -0.735 |
| b | chr19 | 17641556 | 17642414 | cryptic_fiveprime | 0.37 |
| c | chr19 | 17642541 | 17642845 | cryptic_threeprime | 0.364 |

c



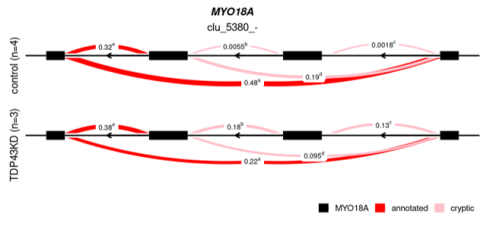

| | chr | start | end | verdict | ΔPSI |
|---|---|---|---|---|---|
| a | chr8 | 79611214 | 79616822 | cryptic_threeprime | 0.568 |
| b | chr8 | 79611214 | 79636802 | annotated | -0.562 |
| c | chr8 | 79613937 | 79616822 | cryptic_unanchored | -0.006 |

d



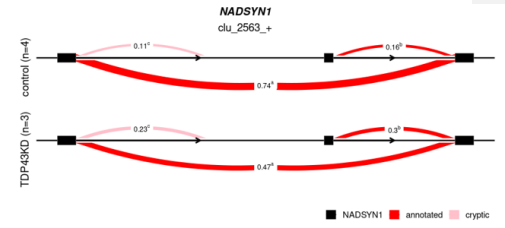

| | chr | start | end | verdict | ΔPSI |
|---|---|---|---|---|---|
| a | chr2 | 219464608 | 219466057 | annotated | -0.566 |
| b | chr2 | 219465786 | 219466057 | cryptic_fiveprime | 0.406 |
| c | chr2 | 219464608 | 219465596 | cryptic_threeprime | 0.208 |
| d | chr2 | 219464608 | 219467174 | annotated | -0.048 |

e




| | chr | start | end | verdict | ΔPSI |
|---|---|---|---|---|---|
| a | chr17 | 29122253 | 29165942 | annotated | -0.255 |
| b | chr17 | 29124688 | 29131383 | novel annotated pair | 0.174 |
| c | chr17 | 29131442 | 29165942 | cryptic_threeprime | 0.123 |
| d | chr17 | 29124688 | 29165942 | novel annotated pair | -0.098 |
| e | chr17 | 29122253 | 29124653 | annotated | 0.057 |

f




| | chr | start | end | verdict | ΔPSI |
|---|---|---|---|---|---|
| a | chr11 | 71474526 | 71478395 | annotated | -0.285 |
| b | chr11 | 71477438 | 71478395 | annotated | 0.154 |
| c | chr11 | 71474526 | 71476201 | novel annotated pair | 0.132 |

**Figure 3.3.2** Six genes with alternative splicing events including cryptic exons. a) WASL, b) UNC13A, c) STMN2, d) SPEG, e) MYO18A, f) NADSYN1

## 4. Discussion

Loss of nuclear TDP-43 and cytoplasmic aggregation is a common feature in ALS/FTD and several other neurodegenerative dise1ses (De Boer et al., 2020). Increasing evidence confirmed loss of nuclear TDP-43affects gene expression and splicing events, especially including cryptic exons. Nevertheless, the precise mechanisms underlying how the repertoire of TDP-43 mRNA targets contributes to pathogenesis and phenotypic changes remain poorly understood. Further investigations are warranted to identify differences in TDP-43 mRNA targets across different cell types and model systems, with a specific focus on the validation of candidate genes, peptides, and splicing events in human neurons affected by ALS/FTD.

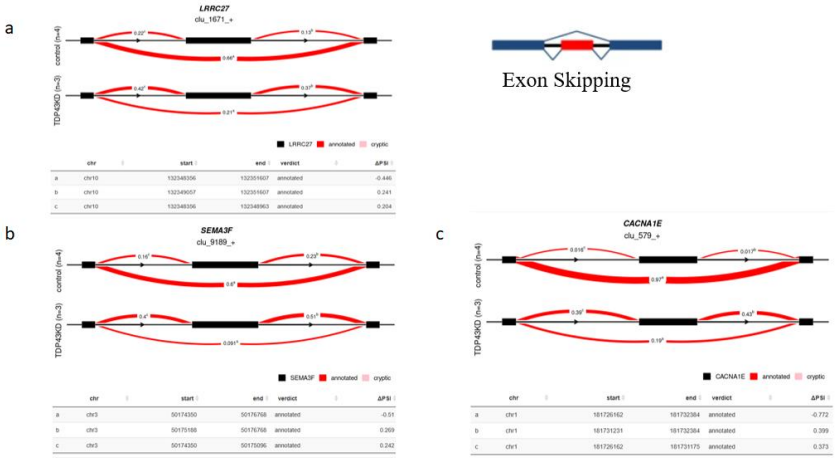
### 4.1 Peptides stability in patients

In order to ensure the stability of biomarkers, it is imperative to reliably detect peptides originating from genes that are associated with the particular disease in patients. Both amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are characterized by a progressive decline in cognitive function and motor control, which may be linked to the accumulation of TDP-43 protein. While it is often assumed that more advanced stages of pathology are associated with greater diagnostic accuracy, the earliest stages of disease offer the greatest opportunity for timely intervention. Specifically, the presence of TDP-43 deposition follows a four-stage progression in ALS and FTD, with initial deposition occurring in the prefrontal cortex (Brettschneider, J. et al., 2013). Genes that are expressed in this region may therefore serve as early diagnostic markers for these conditions, like STMN2, UNC13A.

## 4.2 Alternative splicing event also change in annotated exons

The TDP-43 knockdown involves not only the cryptic exons but also alterations in annotated exons, as evidenced by the expression patterns of certain genes, depicted in the Figure 4.2.1. Specifically, in the control group, these four genes exhibit small splicing changes within annotated exons, whereas in the TDP-43 knockdown group, splicing events are more frequent in

Commented [JJ5]: Which figure?




exons. Such changes have the potential to result in downstream peptide loss, as the knockdown group exhibits additional exons within the affected genes.

**Figure 4.2.1** Genes with alternative splicing event in exon skipping parrtons but not including cryptic exon. a) LRRC27, b)SEMA3F, c)CACNA1E

LRRC27,CACNA1E and SEMA3F show the same alternative splic subtype, exon skipping which was the most frequent subtypes in all tissues(Ito, D. et al., 2020). This finding suggests that TDP-43 deposition may have a specific impact on the regulation of mRNA splicing events.

## 4.3 Limitations

Our study has several limitations that should be acknowledged. Specifically, the IPSC neuronal model utilized in our experiments resulted in a substantial knockdown of TDP-43, potentially

leading to differential gene expression changes that are not reflective of the pathophysiological conditions occurring in ALS/FTD patients, especially at the early disease stage. Additionally, our selection of a log2-fold change not equal to zero as a criterion for gene differential expression analysis may have included genes with minimal expression changes. Lastly, while the use of CSF database searching provided some evidence for peptide identification, further investigations are required to establish its utility as a stable and detectable biomarker in patients. As such, additional studies utilizing various methods and approaches are necessary to further elucidate the clinical implications of our findings.

## Reference


[1] Brettschneider, J., Del Tredici, K., Irwin, D. J., Grossman, M., Robinson, J., Toledo, J. B., Fang, L., Van Deerlin, V. M., Ludolph, A. C., Lee, V. M., Braak, H., & Trojanowski, J. Q. (2014). Sequential distribution of pTDP-43 pathology in behavioral variant frontotemporal dementia (bvFTD). Acta Neuropathologica, 127(3), 423–439. https://doi.org/10.1007/s00401-013-1238-y

[2] Brown, A.L., O.G. Wilkins, M.J. Keuss, S.E. Hill, M. Zanovello, W.C. Lee, A. Bampton, F.C.Y. Lee, L. Masino, Y.A. Qi, S. Bryce-Smith, A. Gatt, M. Hallegger, D. Fagegaltier, H. Phatnani, J. Newcombe, E.K. Gustavsson, S. Seddighi, J.F. Reyes, S.L. Coon, D. Ramos, G. Schiavo, E.M.C. Fisher, T. Raj, M. Secrier, T. Lashley, J. Ule, E. Buratti, J. Humphrey, M.E. Ward, and P. Fratta. 2022. TDP-43 loss and ALS-risk SNPs drive mis-splicing and depletion of UNC13A. Nature. 603:131-137.

[3] Cathcart, S.J., S.H. Appel, L.E. Peterson, E.P. Greene, S.Z. Powell, A.S. Arumanayagam, A.L. Rivera, and M.D. Cykowski. 2021. Fast Progression in Amyotrophic Lateral Sclerosis Is Associated with Greater TDP-43 Burden in Spinal Cord. Journal of neuropathology and experimental neurology. 80:754-763.

[4] Chen, G., Chen, J., Liu, H., Chen, S., Zhang, Y., Li, P., Thierry-Mieg, D., Thierry-Mieg, J., Mattes, W. B., Ning, B., & Shi, T. (2019). Comprehensive Identification and Characterization of Human Secretome Based on Integrative Proteomic and Transcriptomic Data. Frontiers in Cell and Developmental Biology, 7. https://doi.org/10.3389/fcell.2019.00299

[5] De Boer, E. M., Orie, V. K., Williams, T. J., Baker, M., De Oliveira, H. G., Polvikoski, T., Silsby, M., Menon, P., Van Den Bos, M., Halliday, G. M., Van Den Berg, L. H., Van Den Bosch, L., Van Damme, P., Kiernan, M. C., Van Es, M. A., & Vucic, S. (2020). TDP-43 proteinopathies: a new wave of neurodegenerative diseases. Journal of Neurology, Neurosurgery, and Psychiatry, 92(1), 86–95. https://doi.org/10.1136/jnnp-2020-322983

[6] Scotter, E. L., Chen, H., & Shaw, C. (2015). TDP-43 Proteinopathy and ALS: Insights into Disease Mechanisms and Therapeutic Targets. Neurotherapeutics, 12(2), 352–363. https://doi.org/10.1007/s13311-015-0338-x

[7] Jeong, Y.H., Ling, J.P., Lin, S.Z. et al. (2017). Tdp-43 cryptic exons are highly variable between cell types. Mol Neurodegeneration 12, 13.https://doi.org/10.1186/s13024-016-0144-x

[8] Jo, M., Lee, S., Jeon, Y., Kim, S., Kwon, Y., & Kim, H. S. (2020). The role of TDP-43 propagation in neurodegenerative diseases: integrating insights from clinical and experimental studies. Experimental and Molecular Medicine, 52(10), 1652–1662. https://doi.org/10.1038/s12276-020-00513-7

[9] Ji A, Zhang X, Chen W, et al (2017). Genetics insight into the amyotrophic lateral sclerosis/frontotemporal dementia spectrum Journal of Medical Genetics; 54:145-154. http://dx.doi.org/10.1136/jmedgenet-2016-104271

[10] Klim, J.R., L.A. Williams, F. Limone, I. Guerra San Juan, B.N. Davis-Dusenbery, D.A. Mordes, A. Burberry, M.J. Steinbaugh, K.K. Gamage, R. Kirchner, R. Moccia, S.H. Cassel, K. Chen, B.J. Wainger, C.J. Woolf, and K. Eggan. (2019). ALS-implicated protein TDP-43 sustains levels of STMN2, a mediator of motor neuron growth and repair. Nature neuroscience. 22:167-179.

[11] Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12). https://doi.org/10.1186/s13059-014-0550-8

[12] Ling, S., Polymenidou, M., & Page, D. C. (2013). Converging Mechanisms in ALS and FTD: Disrupted RNA and Protein Homeostasis. Neuron, 79(3), 416–438. https://doi.org/10.1016/j.neuron.2013.07.033

[13] Li, Y., Knowles, D. A., Humphrey, J., Barbeira, A. N., Dickinson, S. P., Im, H. K., & Pritchard, J. K. (2018). Annotation-free quantification of RNA splicing using LeafCutter. Nature Genetics, 50(1), 151–158. https://doi.org/10.1038/s41588-017-0004-9

[14] Ma, X.R., Prudencio, M., Koike, Y. et al. (2022) TDP-43 represses cryptic exon inclusion in the FTD–ALS gene UNC13A. Nature 603, 124–130. https://doi.org/10.1038/s41586-022-04424-7

[15] Maize C. Cao, Emma L. Scotter (2022); Transcriptional targets of amyotrophic lateral sclerosis/frontotemporal dementia protein TDP-43 – meta-analysis and interactive graphical database. 15 (9): dmm049418. https://doi.org/10.1242/dmm.049418

[16] Mary E. Piper, Meeta Mistry, Jihe Liu, William J. Gammerdinger, & Radhika S. Khetani. (2022). hbctraining/Intro-to-rnaseq-hpc-salmon-flipped: Introduction to RNA-seq using Salmon Lessons from HCBC (first release). Zenodo. https://doi.org/10.5281/zenodo.5833880

[17] Macron, C., Lane, L., Galindo, A. N., & Dayon, L. (2018). Deep Dive on the Proteome of Human Cerebrospinal Fluid: A Valuable Data Resource for Biomarker Discovery and Missing Protein Identification. Journal of Proteome Research, 17(12), 4113–4126. https://doi.org/10.1021/acs.jproteome.8b00300

[18] Mackenzie, I.R., P. Frick, F.A. Grasser, T.F. Gendron, L. Petrucelli, N.R. Cashman, D. Edbauer, E. Kremmer, J. Prudlo, D. Troost, and M. Neumann. (2015). Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. Acta neuropathologica. 130:845-861.

[19] Melamed, Z., J. Lopez-Erauskin, M.W. Baughn, O. Zhang, K. Drenner, Y. Sun, F. Freyermuth, M.A. McMahon, M.S. Beccari, J.W. Artates, T. Ohkubo, M. Rodriguez, N. Lin, D. Wu, C.F. Bennett, F. Rigo, S. Da Cruz, J. Ravits, C. Lagier-Tourenne, and D.W. Cleveland. (2019). Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration. Nature neuroscience. 22:180-190.

[20] Nana, A.L., Sidhu, M., Gaus, S.E. et al. (2019). Neurons selectively targeted in frontotemporal dementia reveal early stage TDP-43 pathobiology. Acta Neuropathol 137, 27–46 https://doi.org/10.1007/s00401-018-1942-8

[21] Petr Danecek, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard, Andrew Whitwham, Thomas Keane, Shane A McCarthy, Robert M Davies, Heng Li, (2021). Twelve years of SAMtools and BCFtools, GigaScience, 10:2 https://doi.org/10.1093/gigascience/giab008

[22] Seddighi, S., Qi, Y., Brown, A., Wilkins, O. G., Bereda, C., Belair, C., Zhang, Y., Prudencio, M., Keuss, M. J., Khandeshi, A., Pickles, S. R., Hill, S. E., Hawrot, J., Ramos, D., Yuan, H., Roberts, J., Sacramento, E. K., Shah, S., Nalls, M. A., . . . Ward, M. E. (2023). Mis-spliced transcripts generatede novoproteins in TDP-43-related ALS/FTD. BioRxiv (Cold Spring Harbor Laboratory). https://doi.org/10.1101/2023.01.23.525149

[23] Shifu Chen, Yanqing Zhou, Yaru Chen, Jia Gu(2018); fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics, Volume 34, Issue 17, 1 Pages i884-i890, https://doi.org/10.1093/bioinformatics/bty560

[24] Tan, R.H., Ke, Y.D., Ittner, L.M. et al. (2017). ALS/FTLD: experimental models and reality. Acta Neuropathol 133, 177–196. https://doi.org/10.1007/s00401-016-1666-6

[25] Vaquero-Garcia, J., Barrera, A., Gazzara, M.R., González-Vallinas, J., Lahens, N.F., Hogenesch, J.B., Lynch, K.W., & Barash, Y. (2016). A new view of transcriptome complexity and regulation through the lens of local splicing variations. eLife, http://dx.doi.org/5. 10.7554/eLife.11752

[26] Ito, D., Taguchi, R., Deguchi, M., Ogasawara, H., & Inoue, E. (2020). Extensive splicing changes in an ALS/FTD transgenic mouse model overexpressing cytoplasmic fused in sarcoma. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-61676-x