Open Research Online

The Open University's repository of research publications
and other research outputs

Comparative Analysis of TDP-43-Controlled RNA Processing in Neuronal and Muscle Cells in Mouse and Human

Thesis

How to cite:

Šušnjar, Urša (2022). Comparative Analysis of TDP-43-Controlled RNA Processing in Neuronal and Muscle Cells in Mouse and Human. PhD thesis The Open University.

For guidance on citations see FAQs.

© 2021 Urša Šušnjar

https://creativecommons.org/licenses/by-nc-nd/4.0/

Version: Version of Record

Link(s) to article on publisher's website:
http://dx.doi.org/doi:10.21954/ou.ro.000141fd

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data policy on reuse of materials please consult the policies page.

oro.open.ac.uk

ICGEB
Open University

Urša Šušnjar

Comparative analysis of TDP-43-controlled RNA processing in neuronal and muscle cells in mouse and human

PhD Thesis

advisor: dr. Emanuele Buratti
co-advisor: dr. Philip Van Damme

Trieste, 2021

Contents

List of Figures

List of Tables

Abstract

TDP-43 (TAR DNA-binding protein, encoded by *TARDBP* gene) is a DNA/RNA-binding protein that participates in various steps of RNA metabolism. First identified as a component of cytoplasmic inclusions in motor neurons of patients with amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD), its aberrant aggregation later became recognized in non-neuronal tissues, in particular skeletal muscles of patients with inclusion body myositis (IBM). Despite its ubiquitous expression and pleiotropic functions, it was unclear to what extent TDP-43 participates in basic cellular processes that are uniform across tissues and whether it exhibits tissue-specific functions in cells of different backgrounds.

By silencing TDP-43 in (mouse) neuronal and muscle cell lines we mimicked *loss of TDP-43's function*, a pathomechanism commonly believed to underly disease development. RNA-seq allowed transcriptome-wide detection of transcripts regulated by TDP-43 at the level of their overall abundance (differential expression DEG) or splicing (alternative splicing AS).

In this work we discuss similarities and differences concerning the activity of TDP-43 across cell-types. We identified subsets of unique cell-type-characteristic mRNA targets and those that are commonly mediated by TDP-43 in muscles and neurons, whose tight regulation might underlie functions crucial for cell survival. Based on this, we investigated functional consequences of TDP-43 loss in either cell-type, linking TDP-43 pathology to previously described hallmarks of neuro- and myodegeneration.

We further compared alternative splicing and transcript abundance control as two distinct regulatory mechanisms governed by TDP-43 that are not equally exploited by cells of different types. In addition, we investigated how cell-type-characteristic environments shape TDP-43's function and render it tissue-specific.

Among splicing events that occur in a muscle-specific manner, we started to characterize a TDP-43-dependent switch in *Tbc1d1* splicing, as this gene encodes for a GTPase involved in translocation of glucose transporters and thereby mediates glucose uptake by muscle cells.

With those results we set the ground for future studies investigating the function of TDP-43 in muscles and a putative contribution of TDP-43 redistribution to IBM development. Furthermore, our findings stress the importance to select an appropriate cell model to study tissue-characteristic features.

Finally, we identified two novel TDP-43 targets, *PPFIBP1* and *ASAP2* consistently detected to undergo TDP-43-dependent isoform switch not only in mouse but also in humans. We have shown that the splicing of those is indeed perturbed in brains and skeletal muscles affected by TDP-43 pathology and we therefore believe these two splicing events could make a universal readout of TDP-43 dysfunction across tissues.

List of abbreviations

A3SS	alternative 3' splice site
A5SS	alternative 5' splice site
ALS	amyotrophic lateral sclerosis
AS	alternative splicing / alternatively spliced
CNS	central nervous system
DEG	differentially expressed genes
ER	endoplasmic reticulum
FDR	false discovery rate
FTLD	frontotemporal lobar degeneration
GOF	gain of function
IBM	inclusion body myositis
iCLIP	individual-nucleotide resolution cross-linking and immunoprecipitation
LOF	loss of function
MRF	myogenic regulatory factor
MXE	mutually exclusive exon
NLS	nuclear localization signal
NMD	nonsense-mediated decay
ORF	open reading frame
PSC	premature stop codon
PSI	percent spliced in
PTM	post translational modification
RBP	RNA-binding protein
RI	retained intron
RNA	ribonucleic acid
RNP	ribonucleoprotein
ROS	reactive oxygen species
RRM	RNA-recognition motif
SE	skipped exon
SS	splice site
TARDBP	TAR DNA-binding protein
TDP-43	TAR DNA-binding protein 43

Acknowledgments

When I started my PhD in 2017, I could not imagine, by any means, how my small world and the big world around me would have turned. Now, close to the end, I am looking back with a lot of gratitude that I would like to express here.

I would first thank Emanuele for his *bearable lightness of being* that often eased my impatient soul. Not only he tolerated unusual working hours, he gave me all the freedom in the world and tonnes of support in every endeavor. I hope to have (at least partially) warranted his trust.

I can still remember the very first days in the *new* lab, being welcomed by Sara and Francesca for Cristiana was on her maternity leave and Emanuele was away for a conference. We started to get to know one another by chatting about food (fairly unusual for a non-Italian) and coffee and wedding dresses and makeup (Sara) and these two made me feel comfortable right away. I should thank Sara for insisting on Italian from the very beginning but also later on when everyone else gave up - it is for her I can understand kids in the park now.

Although long days in lab seem somehow distant today, Cristiana's efforts in having us all on track will not be forgotten. She should be thanked for her patience when I asked for the n^{th}-time where the size marker is.

Not only I now know a lot about splicing, I am equipped with other important skills I wouldn't have gained without the teaching experience. Here, I would like to thank Yasmine, who taught me teach. I hope she has forgotten about our weird inexplicable PCRs and my moralizing but she will remember to nurture her curiosity and will bravely accept all scientific challenges to come.

It was my pleasure to share the lab Alice, Matea and Christian, who at their young(er) age undoubtedly believed in (scientific) ideals and often reminded me of my younger self. Their honest enthusiasm would give us hope and drive us further. I am convinced the world will realize these three belong to science (or, maybe, Alice should become a writer).

Ljiljana, sadly, joined our group only when I was about to leave. Despite the little interaction we had, I understood how much one can learn from an experienced person. My research would have benefited from having her aside.

Tahir came like a fresh breeze to our lab. For a month I could enjoy a *second opinion* on everything I was doing and started to miss this as soon as she left. 3 years later, September still reminds me of her.

I should thank Neva for spending lockdown Sundays (on top of all working days) analyzing my data and helping me decipher the secret pipeline of Novogene. That she accepted my caprices - because she needed to change the analysis as I was changing my ideas. She even engaged Robert to help us calculate conservation scores.

Then, I would like to acknowledge another disproportionally gifted person, my dearest Anna-Leigh. With her, I had fun designing mindmaps, running scripts, doubting about results, turning mindmaps around, running scripts again and coming out with even more doubts. Some seeds need to be sent to London once the paper is out.

All this would have not been possible without a strong support I received at home, from mine and Peter's family. At the very beginning, there was my dad, who to me was an example of persistence, dedication and desire to improve. Later on I learned from Peter's dad, who undoubtedly believed in the value of education and appreciated (probably more then anyone else) our efforts.

Then there is Eva, who came *just on time* and lightened up our lockdown life. It is her who sped up the writing (I would not like to imagine how long it would have otherwise taken). She will forever remain our lovely *žuži* and *muc* and we will keep hugging her each time she will be bitten by mosquitoes.

I was lucky enough to have my mom by my side. She adapted a lot in the past year to make our lives easier. I am grateful for days she spent with Eva at Barcola and for teaching her walk. My heart melts when I see Eva scream with excitement as she enters the flat.

And then Peter, the rock on which I've been building my life for the past 15 years. For being my lover, entertainer, cook, comforter, nurse, critic, carpenter, my singer-songwriter, life-coach, news monger, innovator and the sweetest *tata*.

Finally, I would like to virtually hug friends who I met in Trieste; Rose, Kevin, Vedran, Neža and Erick and thank Živa, Erica and Michela for nurturing Slovenian in me.

Cheers to the life full of twists and turns!

1. Introduction

1.1 TDP-43

Transactivation response element (TAR) DNA-binding protein (TDP-43, 43 kDa, encoded by *TARDBP* gene on chromosome 1) was initially described as a transcription repressor of HIV [1] and it gained extensive attention after the discovery that it is involved in neurodegeneration [2]; [3]. Abnormal TDP-43 aggregation was first observed in patients with amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD) but then expanded to the whole spectrum of neurological disease now commonly referred to as TDP-43 proteinopathies. Although TDP-43 happens to be the most studied, there are other proteins belonging to the same family of heterogeneous nuclear ribonucleoproteins (hnRNPs) that are, like TDP-43, mutated or aggregated in diseased brain [4].

TDP-43 is a highly-conserved [5] ubiquitously expressed protein involved in the regulations of several steps of mRNA processing (Figure 1.1). Beside being a transcription factor, TDP-43 regulates alternative splicing and influences transcript stability (including of its own mRNA). As a part of Drosha complex, it plays a role in microRNA biogenesis and furthermore, it modulates mRNA transport, translation and stress granules assembly. Due to the huge number of functions TDP-43 executes in a cell, with its numerous gene targets and protein partners, it makes it very difficult to distinguish between direct TDP-43-regulated events and secondary effects that result from general perturbations of processes summarized in Figure 1.1.

TDP-43 structurally displays a typical hnRNP architecture. hnRNPs are a set of primarily nuclear proteins that transiently cross-link to nascent transcripts produced by RNA polymerase II [6]. Rather then having a single well-defined function, hnRNPs participate in multiple processes involved in RNA metabolism depending on their sub-cellular localization, relative abundance, presence of target RNAs and other cellular components [4].

Despite hnRNPs being versatile and highly diverse proteins, they share some structural characteristics. The most distinguishing feature of hnRNPs is their ability to bind single-stranded nucleic acids in a sequence specific manner (although they also exhibit some unspecific affinity for RNA). They do so through nucleic-acid binding domains. Most of hnRNPs also contain domains with distinctive amino acid composition such as glycine-rich, acidic or proline-rich domains, that are important for mediating protein-protein interactions. Moreover, they undergo

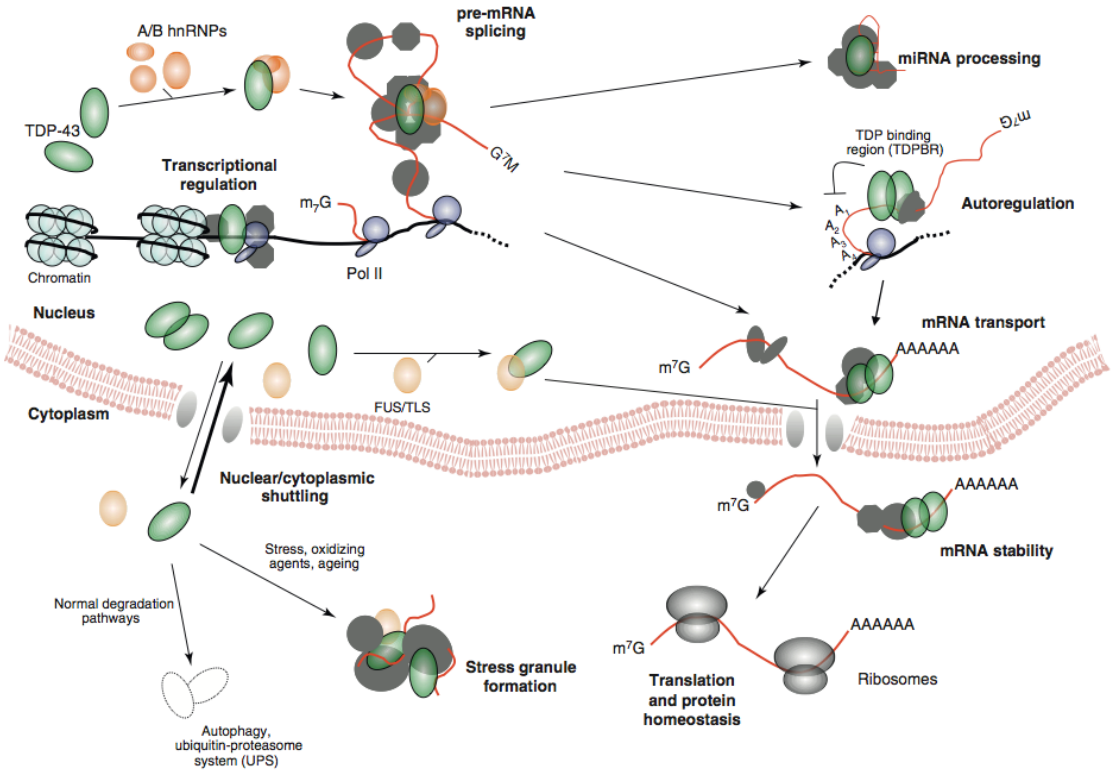


Figure 1.1: **Molecular functions of TDP-43.** TDP-43 binds nucleic acids (DNA in black, RNA in red) and interacts with other hnRNPs (FUS in yellow, hnRNP A/B family in orange, other factors in gray) to govern key steps of mRNA processing both, in the nucleus and in the cytoplasm [4].

post-translational modifications like phosphorylation of serine and threonine residues or arginine methylation and are often subject to alternative splicing giving rise to more than one protein isoform. However, paralogues and isoforms often exhibit parallel functions [4]; [6].

1.1.1 RNA binding

Similar to other hnRNPs, TDP-43 binds RNA in a single-stranded and sequence-specific manner. RNA binding occurs through two RNA recognition motifs (RRM1 and RRM2, Figure 1.2), which are 60 nucleotide long sequences that form a conserved two-dimensional structure. The two sequences are evolutionary well conserved suggesting they might be crucial for the core function of TDP-43 [4]. We still lack structural analysis of TDP-43 bound to RNA due to intrinsic propensity of the protein to aggregate. Initial structural characterization of TDP-43 showed it preferentially binds UG (or TG) tandem repeats, and that the binding affinity increases with the number of repeats. Phenylalanine residues in the RRM1 play a key role in nucleic acid recognition, while RRM2 was proposed to mediate DNA binding and protein structure stabilization [5].

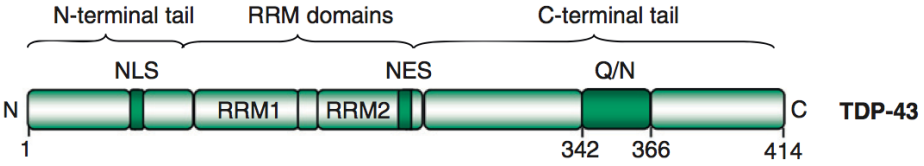


Figure 1.2: **TDP-43 structure.** TDP-43 consists of three main regions: N-terminal tail, RRM regions and C-terminal tail. Important functional domains highlighted herein are NLS (nuclear localization signal), NES (nuclear export signal) and Q/N region (glutamine- and asparagine-rich region) [4].

High-throughput studies using crosslinking and immunoprecipitation (iCLIP) [7] confirmed the preference of TDP-43 to bind UG tandem repeats or long clusters of UG-rich motifs *in vivo*. Single RRM domain usually reorganizes a sequence of four nucleotides, however, the RNA-specificity for long binding sites indicates cooperative binding of TDP-43 to UG-repeats. As a homodimer, TDP-43 recognizes multiple dispersed UG-rich motifs. The number of repeats correlates with the strength of binding, nevertheless, genome-wide analysis revealed UG-rich repeats to be neither necessary nor sufficient to define a TDP-43–binding site [8]. One such example is provided by the *TARDBP* transcript itself, in which TDP-43 binds to non-UG sequences [4]. Almost 40,000 TDP-43 binding clusters were identified in mouse brain, which represents TDP-43 binding sites within 6,304 annotated protein-coding genes, corresponding to approximately one third of the mouse transcriptome [8].

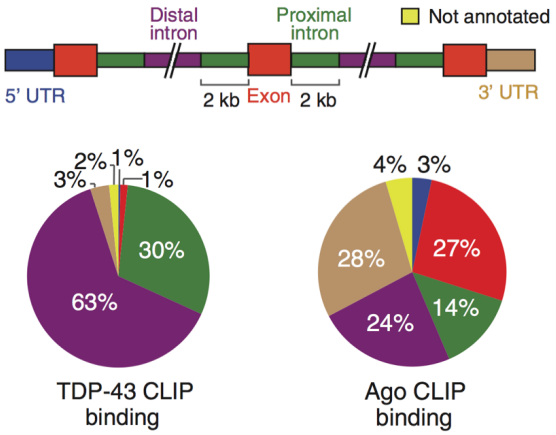


Figure 1.3: **TDP-43 binding.** Binding of TDP-43 within annotated regions of pre-mRNAs shows preferential binding in distal introns. Argonaute is another RNA-binding protein. Distribution of Argonaute biding sites that is different from TDP-43 supports region-specific binding of each protein [8].

iCLIP analysis indicates that the binding position of TDP-43 in a given mRNA determines its splicing function. TDP-43 crosslinking upstream from or within an exon generally leads to

silencing (exclusion) of that particular exon, while TDP-43 binding downstream an exon results in its inclusion. TDP-43 binding further downstream of an exon was associated with silencing of an exon [7]. However, the majority of altered splicing events identified upon depletion of TDP-43 do not have TDP-43 clusters within 2 kb of the splice sites, implicating the possibility of long-distance interactions or indirect effects of TDP-43 through other splicing factors [8]. The majority of TDP-43 binds to distant intronic regions (>2 kb from the neighboring exon) is in contrast with binding maps of some tissue-specific RNA-binding proteins (Figure 1.3). Crosslinking to distal introns would, rather than splicing exclusively, influence other aspects of gene expression, as enrichment of TDP-43 clusters was found in distal introns but not in exons or 5' or 3' UTR of downregulated genes. Introns of genes that were downregulated upon TDP-43 depletion were exceptionally long, which suggests a crucial role of TDP-43 in maintaining the abundance of long intron-containing genes in mouse brain [8].

Moreover, TDP-43 binding was also mapped to lncRNAs and intergenic transcripts, which in fact have the biggest enrichment of UG-repeats, indicating that TDP-43 might connect different aspects of gene expression [7].

1.1.2 Interactions with other proteins

Like RRMs allow binding to nucleic acids, low complexity region at C-terminus (Figure 1.2) enables TDP-43 to interact with other proteins that modulate its RNA-processing abilities. To date, the majority of disease-associated TDP-43 mutations have been described in the low complexity glycine-rich region, implying that the ability of TDP-43 to cooperatively assemble on long RNA binding sites could be involved in disease mechanism [7]. This region has been shown to mediate interaction with several other hnRNPs like hnRNP A1, A2/B1, C1/C2 and A3. Mutational analysis reveals that TDP-43 lacking C-terminus was unable to inhibit splicing of *CTRF* mRNA, probably due to the fact it could not form complexes with other hnRNPs that cooperatively control splicing activity [9]. Collectively, this indicates interactions with other proteins are essential for TDP-43 to remain fully functional.

Proteins interacting with TDP-43 cluster in two distinct networks (Figure1.4). First group is comprised of nuclear proteins taking part in splicing control and other aspects of nuclear RNA metabolism, whereas the second network represents cytoplasmic proteins that regulate mRNA translation. TDP-43-interacting factors include hnRNPs and other splicing factors, RNA helicases, translation regulatory proteins, as well as proteins involved in mRNA transport and stability. Importantly, the interaction of TDP-43 with other proteins sometimes requires prior binding to RNA, or it could be RNA-independent. TDP-43 further interacts with DNA-binding proteins like transcription factors and DNA repair proteins (e.g. Ku70), supporting the view to additionally play a role in DNA metabolism [10].

A big part of the C-terminus (spanning amino acid residues 277-414) was, based on the sequence homology, proposed to bear infectious characteristics and to act as a prion-like domain

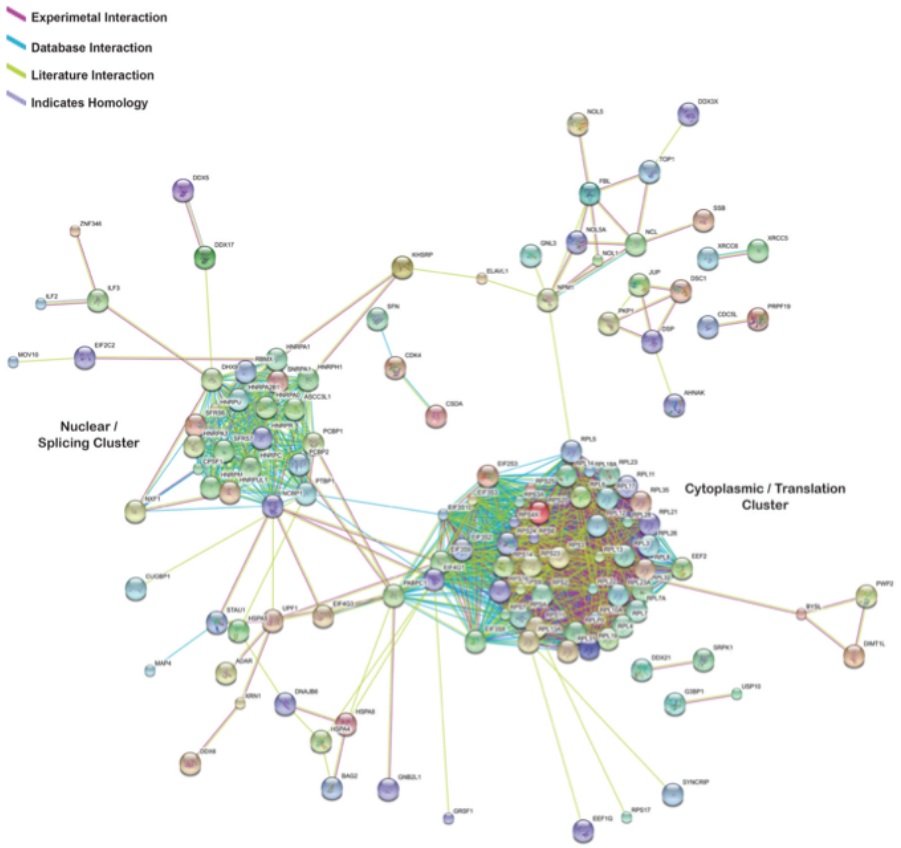


Figure 1.4: **Proteins interacting with TDP-43.** TDP-43-interacting factors clearly fall into two clusters. Nuclear proteins in one cluster take part in mRNA splicing and nuclear RNA processing, whereas cytoplasmic proteins of the other cluster participate in translation and cytoplasmic RNA metabolism [10].

[11]. A shorter segment of that region (residues 342-366) is rich in glutamine/asparagine (Q/N). While on one hand this part mediates protein-protein interactions, it is also responsible for TDP-43 aggregation [12]. The interaction of TDP-43 with its Q/N region is in fact stronger than the interaction of TDP-43 with hnRNP A2 [13].

1.1.3 Sub-cellular localization

TDP-43 displays predominantly nuclear localization, yet it undergoes continuous nucleocytoplasmic shuffling (an ability that is characteristic of most of the hnRNPs [6]), so that low levels of TDP-43 are always present in the cytoplasm under physiological conditions [14]. After discovering pathological TDP-43 inclusions in neurons of ALS/FTLD patients, it became clear that TDP-43 distribution is tightly regulated and any perturbation of that balance would lead to disease.

TDP-43 contains two segments, which allow sub-cellular trafficking; the nuclear localization signal (NLS) located in the N-terminal part and the nuclear export signal (NES) spanning the

RRM2 (Figure 1.2). NLS is a 16 amino acids long sequence containing two clusters of basic residues separated by a stretch of 9-12 amino acids located at the position 82-98 in mouse and human TDP-43. Both basic stretches are required for nuclear targeting. NES, on the other hand, is a leucine-containing sequence spanning the RRM2 domain at the position 239-250 in human TDP-43. Similarly to NLS, NES contains two hydrophobic domains, both of which are required for cytoplasmic translocation. Any disturbance of those sequences blocks TDP-43's exit from the nucleus and leads to formation of nuclear inclusions. Importantly, expression of mutant protein lacking NLS or NES causes TDP-43 aggregation either in the nucleus or in the cytoplasm, and in turn results in sequestration of the endogenous TDP-43 [14]. Most of inclusions observed in TDP-43 pathologies are cytoplasmic and form in the absence of any mutation. This indicates that trafficking of the full-length TDP-43 would be, apart from the NLS and NES, regulated by other mechanisms. Translocation of TDP-43 to the nucleus is an active process carried out by the KPNA/KPNB1 pathway and indeed, depletion of proteins involved in this pathway affects TDP-43 shuffling [15]. Furthermore, C-terminal domain (CTD) fragments generated in some pathological conditions [16] lack NLS, presumably impairing the trafficking of those fragments. 25/35 kDa-long fragments were reported to propagate aggregation of the full-length TDP-43 [17], which might be explained by cytoplasmic trapping. An alternative hypothesis suggests that short fragments interfere with the regular trafficking of TDP-43.

Even in the nucleus TDP-43's distribution is subject to temporal regulation. During embryonic stem cell (ESC) differentiation, TDP-43 is, upon cellular exit from pluripotency, recruited to paraspeckle compartment [18]. Paraspeckles are nuclear formations built around lncRNA *Neat1* (the splicing of which is mediated by TDP-43), which sequesters various proteins, among them also RNA-binding proteins [19]. Partial sequestration in paraspeckles prevents TDP-43 from accessing mRNA, which, besides reduced expression, further decreases its activity [18].

A small portion of cytoplasmic TDP-43 was recently shown to colocalize with mitochondria, in particular with the inner mitochondrial membrane (the cristae) [20]. Since ALS-associated mutations enhance its mitochondrial accumulation [20]; [21], mitochondrial localization was initially considered a rather pathological phenomenon [21]. However, as unmutated TDP-43 is also present in the cristae of non-affected motor neurons, TDP-43 was then proposed to play an essential role in mitochondria, seemingly by interacting with mitochondrial RNAs and helping maintain appropriate transcript levels [20]. Mitochondrial localization signals usually consist of hydrophobic amino acids, and six such stretches exists in TDP-43 [21].

1.2 Alternative splicing

Sequencing of human genome revealed there is a much bigger number of expressed sequences (mRNAs) than there are genes, which raised important questions about genomic complexity and proteomic diversity arising from a limited number of genes (around 20,000) [22]; [23]. Genome-

wide studies estimated about 90-95% of all human genes to undergo alternative splicing, many of which (approximately 37%) give rise to different protein isoforms. Alternative splicing (AS) is a widely-exploited regulatory mechanism that allows generation of transcripts different in their untranslated regions (UTRs) or coding sequences. Splicing events are commonly categorized as exon skipping (SE), alternative splicing of 5' or 3' boundaries between exons and introns (A5SS or A3SS), inclusion of one of two mutually exclusive exons (MXE) or intron retention (RI) (Figure 1.5). However, new types of splicing events (i.e., cryptic exons [24]; [25] and microexons [26]) are starting to emerge due to improved performance of splicing tools.

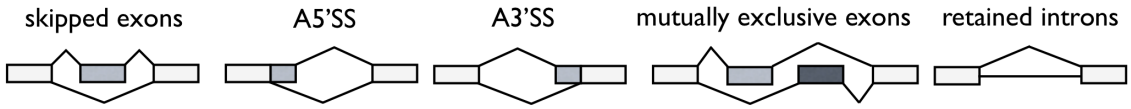


Figure 1.5: **Five categories of AS events.** Splicing events are classified into 5 main categories - exon skipping (SE), selection of a variable 5' or 3' splice site (A5SS and A3SS), inclusion of one of two mutually exclusive exons (MXE) and intron retention (IR). Currently, other types of events (microexons, cryptic exons, mixed events) have become commonly considered along with classical annotated splicing events.

When alternative splicing occurs within UTR, the event would give rise to transcripts that might differ with respect to their mRNA stability, localization and translation. On the other hand, differential splicing within a protein coding region could shift the reading frame or it would, by removal/addition of a sequence, produce an alternative protein isoform. Inclusion of a sequence containing a pre-mature stop codon or causing frame-shifting, would in turn activate non-sense mediated decay (NMD) and reduce transcript level. Furthermore, splicing can affect non-coding transcripts, which can still have a function by modulating protein-coding RNAs.

A big challenge has lately been to understand the physiological significance of alternative splicing. Beyond mapping of individual events, efforts have been made to characterize functional consequences of such alternative events. Differential splicing does not necessarily result in generation of an alternative protein isoform, and even if so, the original biological function of a given protein might remain intact. The role of an individual splice variant (which either gives rise to protein or not) is usually studied by overexpressing one isoform in the absence of others.

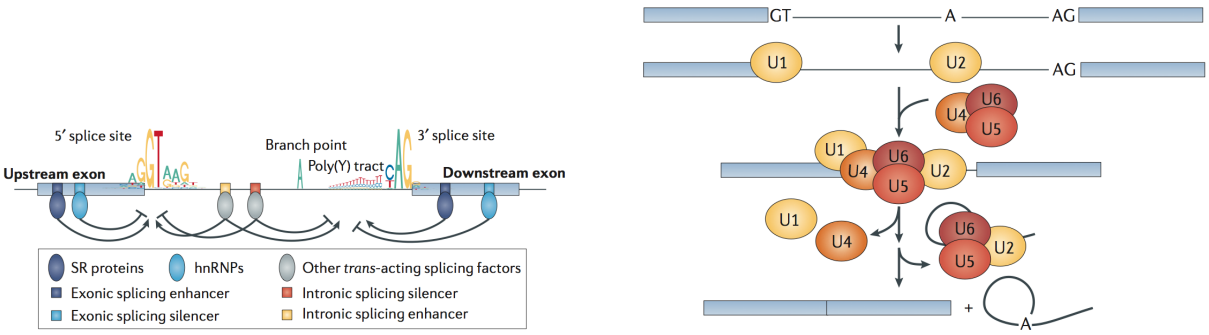
Apart from studying individual splicing events and their biological significance, studies have now progressed to globally explore splicing networks along with possible mechanisms of their coordination. Alternative splicing has been demonstrated to govern various biological processes regulated in space and time such as linage determination, tissue differentiation and organ development [27]; [28]. Moore *et al.* show [28] how certain drugs induce proapoptotic splicing program, i.e., a coordinated upregulation or proapoptotic splice variants that collectively promote apoptosis in arrested cells. Furthermore, time-specific expression of splicing factors during cell differentiation supports the precise temporal coordination of splicing transitions [27]. Thus,

a specific splicing factor would possibly mediate splicing of transcripts from functionally related genes, which participate in the same biological process.

Physiological significance of alternative splicing is further highlighted by the emergence of human disease caused by mutations in *cis*-acting elements (enhancer or silencer sequences, respectively), *trans*-acting splicing factors or components of spliceosome [29]. Beside mutations, upstream signaling pathways could regulate the activity of splicing factors, which would then, by targeting an array of mRNAs, allow quick adjustments in proteome upon internal and external stimuli [30].

1.2.1 Molecular mechanisms of alternative splicing

The primary function of pre-mRNA splicing as a part of mRNA maturation is removal of non-coding introns. The process of splicing is carried out by a large macromolecular machinery called the spliceosome. The major spliceosome, which is responsible for excision of most of introns, consists of five small uridine-rich ribonucleoprotein complexes (U snRNPs; U1, U2, U4, U5 and U6). Each of these snRNPs is built of a stable small nuclear RNA (snRNA) bound by several proteins that additionally associate with multiple splicing factors [31]. The spliceosome has two major tasks: first, to recognize specific sites/sequences associated with exon/intron boundaries known as 5' and 3' splice sites (5' SS and 3' SS) and second, to catalyze the splicing reaction. The end of an upstream exon is demarcated by a consensus sequence motif at its 5' SS (Figure 1.6a), whereas a series of loosely defined sequences (branch point, polypyrimidine tract, 3' SS sequence; Figure 1.6a) mark the start of a downstream exon.



(a) (b)

Figure 1.6: **Molecular mechanisms of alternative splicing. A** A splice site is generally defined by sequences that mark exon/intron boundaries - 5' SS consensus motif, branch point, polypyrimidine tract and 3' SS sequence. The recognition of weak splice sites is facilitated by the presence of auxiliary sequences (*cis*-elements that get bound by *trans*-acting factors). **B** A step-wise recruitment of five U snRNPs to exon-intron boundaries of pre-mRNA. snRNPs together with auxiliary factors (not depicted here) form the major spliceosome and catalyze the intron excision [32].

Assembly of splicing machinery occurs in a step-wise manner (Figure 1.6b). It begins with

recruitment of snRNP U1 to the 5' SS of an upstream exon via base pairing of its snRNA. Next, snRNP U2 interacts with the branch point region, the recognition of which is facilitated by binding of auxiliary factors. Following the recruitment of U4/U5/U6 tri-snRNPs, the spliceosome enters its active conformation and splicing occurs via two transesterification reactions [32]. The first transesterification step is the nucleophilic attack by the 2' OH group of the key adenosine of the branch site to the 5' SS, which produces a structure called intron lariat. In the second transesterification step, the 3' OH group of the 5' SS attacks the 3' SS of the downstream exon, giving rise to spliced mRNA on one hand and excised intron lariat on the other. The intron undergoes subsequent degradation [33].

Splice sites themselves could be strong or weak depending on how closely their sequence resembles the consensus motif. Strong splice sites are, generally speaking, fully used, which leads to constitutive splicing at that particular position. The recognition of weak splice sites, on the other hand, is only partial and depends on the cellular context (resulting in alternative splicing) even with sustained levels of splicing machinery. The recruitment of spliceosomal proteins and the extent to which a certain splice site would be chosen is determined by several factors; 1) the intrinsic strength of the splice site sequence; 2) exon length; 3) the presence of degenerate *cis*-regulating sequences that function either as splicing enhancers (SE) or silencers (SS) and are located within exons or intronic sequences approximately 50 nucleotides from exon/intron boundaries [34]; 4) the secondary structure of mRNA that might either expose or hide regulatory sequences [35]; and 5) the relative abundance of *trans*-acting regulatory proteins and their post-translational modifications. Another level of complexity is added due to the fact that splicing often occurs cotranscriptionally and that the same factors can control both processes [33].

1.2.2 Regulation of alternative splicing by RNA-binding proteins

Classical *trans*-acting splicing regulators are RNA-binding proteins (RBPs) like those from serine/arginine-rich (SR) protein family or hnRNPs that promote and repress splicing in a sequence-specific manner [35]. Originally thought to be important for alternative splicing exclusively, we now understand they also play a crucial role in constitutive exon inclusion [36]. Canonically, SR proteins are viewed as promoters and hnRNPs as splicing repressors [35]; [37]. Very often, however, they behave in both ways either as activators or repressors, depending on the position where they crosslink to the transcript and on their post-translational modifications [35].

RBPs influence splicing through the following modes of action; first, their binding to splicing enhancers might help recruit individual spliceosomal units to the transcript and assist spliceosome formation [38]. Furthermore, binding to exonic sequences could mark exons and thus facilitate splice site recognition. On the other hand, crosslinking of these same proteins to other sequences in a transcript would block functional recognition of a splice site and prevent spliceosomal assembly [37] or repress splicing through sterical hindrance [36]. Alternatively, binding

of splicing factors can induce formation of an RNA loop, which would either prevent usage of a splice site or bring two distant splice sites together [35].

1.2.3 Tissue-specific alternative splicing

Each splicing event is regulated by multiple RNA-binding proteins, therefore, differences in their relative expression and intracellular localization together with the presence of tissue-specific *cis*-regulatory elements underly a "code" for cell-type-specific splicing [35]; [39]. A study comparing transcriptome variation across human tissues demonstrated 50-80% of alternative splicing events to be regulated in a tissue-specific fashion [40], which makes alternative splicing an important driver of phenotypic complexity. In line with that, a multidimensional clustering (PCA) based on exon inclusion levels featuring overall splicing profile in different tissues largely recapitulates tissue type [41].

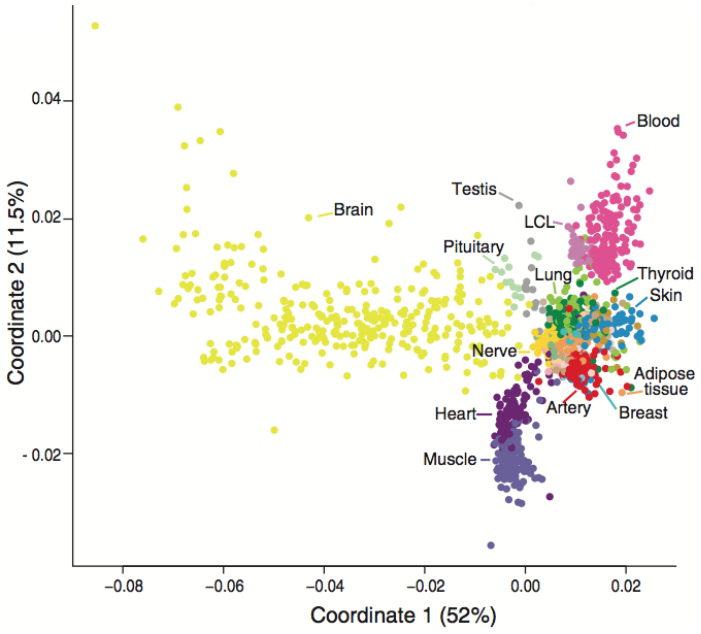


Figure 1.7: **Tissue-specific alternative splicing.** PCA clustering of human tissue samples based on exon inclusion levels (PSI) [41].

Aternative splicing is especially wide-spread in nervous tissue [41]; [42] and neuronal events are better conserved than those occurring elsewhere [43], implying they provide important functions. Splicing alternations in fact underly the transition from undifferentiated precursors to fully differentiated neuronal cells and impact other important processes like neuronal migration, axonal and dendritic outgrowth, establishment of synaptic connections and neuronal plasticity [42]. Several neuron-enriched *trans*-acting factors have been identified, among them NOVA, RBFOX proteins, PTBP proteins and SRRM4 [42]; [44]. Neuronal transcriptome displays a strong enrichment in alternative microexon inclusion (microexons are defined as 3-15 nucleotide

long exons)[26], as well as higher frequency of retained introns. In muscles, for example, intron retention occurs less often [45].

Comparing tissue-specific alternative exons to constitutive exons or other alternative exons, they seem to preserve reading frame more often. They are often enriched in sites where post-translational modifications occur. Exons that show differential usage between tissues are better conserved across species [46]. Furthermore, tissue-regulated exons, especially those that are alternatively spliced in neurons and muscles, often encode for protein segments with unstructured or highly disordered regions, which often mediate protein-protein interactions [42]; [47].

Beside the brain, the most of tissue-specific alternative splicing takes place in striated muscle [29]; [30]. Along with muscle-enriched *trans*-activating regulators like RBM24 [48], RBFOX, CELF, MBNL, and PTB family members, *cis*-regulatory sequences recognized by these factors appear in the proximity of muscle-specific splice sites [30]. Furthermore, the relative expression of muscle-enriched splicing factors changes during muscle differentiation, which would be indicative of splicing contribution to a highly-coordinated muscle differentiation program [30]; [27]. The evidence of precise temporal-coordination of alternative splicing in muscles cames from a study [27], in which they investigated the frequency of specific *cis*-regulatory elements in genes undergoing muscle-characteristic splicing transition at different stages of differentiation. Enrichment of specific regulatory motifs correlated with expression levels of splicing factors crosslinking to those sequences [27]. Genes that undergo muscle-specific alternative splicing are often associated with muscle differentiation (e.g., MEF2 family of transcription factors) or calcium handling (e.g,. mitochondrial calcium uniporter MCU). Moreover, alternative splicing contributes to diversity of contractile proteins, which then give rise to different sacromere types with distinct contractile properties [30].

1.2.4 Conservation of alternative splicing across species

Mouse models (either animal models or murine cell lines) are often used to study disease, in which splicing dysfunction is believed to play a pathological role. Investigating alternative splicing in mouse, it needs to be established to what extent findings from mice can actually be translated into human physiology. Most of human genes have a homologous gene in another species, with only 1% lacking a detectable orthologue. Furthermore, 86% of orthologous genes were found to have the same number of exons in mouse and human [49].

When it comes to alternative splicing, however, the orthology can no longer be investigated at the gene level. As the orthology of a gene-pair cannot guarantee for the orthology of individual exons, one needs to define the extent of the orthology between exons. Alternative Splicing Annotation Project (ASAP) found more than 80,000 mouse and human exons that have at least one orthologue in the other species [50], while later studies [51] found 160,000 exons (92% of all assigned exons) with an orthologous counterpart. A higher level of conservation was observed in the coding region (96%) compared to in 5' UTR (65%) or n 3' UTR (29%), respectively.

The observation that alternative splicing patterns in vertebrates diverged faster than differences in gene expression was mostly attributed to mutations in *cis*-regulatory motifs [52]. Consistent with that, most of the species-specific alternative splicing has been explained by differences in those sequences, followed by differences in *trans*-regulatory factors [53]; [52]; [42]. Conserved alternative exons exhibit higher frequency of *cis*-regulatory elements and higher sequence conservation of exons and surrounding intronic regions. Therefore, conservation of regulatory motifs and mRNA secondary structure [43], rather than exonic sequence similarity exclusively, would be decisive whether the alternative nature of an exon is spared between species. Yet, splicing events that are better conserved since divergence of two species, would more likely be of biological importance. Alternative splicing is more frequent in humans than in mouse [53]; [54], regardless the tissue-type.

The question of splicing conservation has been addressed using different approaches; Thanaraj *et al.* [55] investigated conservation of splice junctions; Yeo et al [43] proposed a model, which integrated multiple exonic features; and Zambelli *et al.* [56] assessed the extent of transcript identity between mouse and human.

Comparing splice junctions, the study reported 15% of alternative and 67% of constitutive human splice junctions to be conserved in mouse [55]. The other, machine learning approach (that integrated multiple exonic features) predicted approximately 2,000 alternative exons to be conserved between mice and human [43]. Predicted conserved alternative exons were more likely to be frame-preserving and less likely to disrupt protein domains than species-specific events. Conserved alternative exons were enriched in genes expressed in the brain and in genes involved in transcriptional regulation, RNA processing, and development. The study comparing orthologous transcripts between the two species [56] found a murine counterpart for 75% of human transcripts.

In conclusion, the usage of alternative exons between human and mouse is partially conserved. Yet, there in no agreement on the level of splicing conservation, in part due to methodological inconsistence, but also due to complexity of regulation and big number of factors influencing exon inclusion. One way to address this question would be to investigate the contribution of individual factors. Comparing TDP-43-mediated splicing in mouse in humans, it has been shown that most of TDP-43-regulated exons in mouse (85% of excluded and 57% of included exons) have an ortholog in humans, for which there is an evidence of alternative inclusion [8]. However, it has not been demonstrated that splicing of those is indeed subject to TDP-43 regulation in humans, as it is in mouse. Importantly, human TDP-43 has previously been expressed in mice, with the aim to investigate the consequences of disease-associated mutations. While human mutations often impaired splicing, expression of wild-type human protein successfully replaced the function of murine TDP-43 and animals remained phenotypically normal [57].

1.2.5 RNA-seq

In the last decade, RNA sequencing (RNA-seq) has emerged as a revolutionary tool for transcriptome studies. It brought great advances to the field of alternative splicing and it rapidly replaced older and less precise expressed sequence tags (EST) methodology or splicing-sensitive microarrays [58]; [59]; [60]. Not only we are now able to detect more splicing events (e.g. find splicing changes of unannotated regions, identify novel splice junctions or detect events that are subject to smaller changes (ΔPSI)), we started to understand that the distinction between constitutive and alternative splice sites is highly arbitrary. Many exons, previously considered as constitutive, were later reported to undergo alternative inclusion (for example in different cell-types or as a response to particular stimuli).

Apart from the mRNA transcribed from protein-coding genes, this same technology allows detection of other RNA species like miRNA, lncRNA, lincRNA [60]. The standard sequencing workflow consists of three stages; extraction of RNA and quality assessment; RNA enrichment, generation of cDNA and preparation of the adapter-ligated sequencing library; and the sequencing performed on a high-throughput platform, most often Ilumina. A given cDNA library is sequenced to a certain read depth, i.e., to obtain a desired number of reads per sample. From this point, the workflow cosists of computational steps which include; sequence alignment (or *de novo* assembly); quantification of the overlapping reads, normalization between samples; and application of statistical models to identify significant changes [61]. Most often, differential exon usage is visualized in the form of Sashimi plot (Figure 1.8), which allows quantitative visualization of reads aligning to a genomic segment [62]. Beside alternative exon usage, it enables a direct assessment of differential gene expression across tissues/conditions.

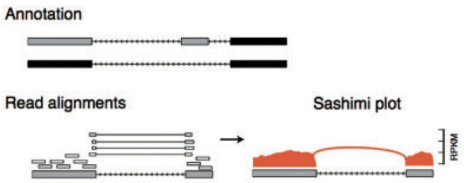


Figure 1.8: **Sashimi plot.** Annotation of RNA-seq reads to a gene that has two isoforms, which differ by the presence of the second exon. Read counts are represented by read densities (reads per kilo base million on y axis) at defined genomic coordinates (x axis). Reads spanning exon junctions are denoted as arcs [62].

Compared to studies focusing on differential gene expression, the experimental design intended to detect splicing alterations requires more attention. Because an alternative exon represents a very short segment within a transcript, the sequencing depth should be higher than that commonly used for the expression analysis [63]. It has been estimated that as few as 10 million reads would suffice to study differential expression compared to at least 50 million reads that would be required for a splicing experiment [64]; [65]. Shen *et al.* [59] highlighted the importance of biological replicates, even when (with a limited budget) at expense of a sequencing

depth. Paired-end sequencing, in which both ends of cDNA fragments are sequenced, produces high-quality data which is easier to align, and was thus recommended to be used for AS detection. It yields longer reads and is more informative than single-end protocol, mostly attributing to the sequences, in which the two reads comprising a pair flank an alternative exon [63].

In addition, splicing analysis faces unique challenges that could be only addressed by the means of powerful statistical tools [63]. One of such challenges is related to tissue-specific alternative splicing. The tissue-characteristic transcriptome is mostly dominated by a small number of genes, so that most of the RNA-reads would align to these highly represented transcripts. Consequently, less reads would be available to estimate the expression of underpresented genes [41], let alone individual exons. The estimation uncertainty of isoform proportion depends on the sequencing coverage of an individual splicing event [59]. Next, some categories of splicing events, namely A5SS, A3SS and microexons, are difficult to investigate since the variable region is often very short [40]. Another issue arises from incompletely spliced transcripts containing intronic fragments, that can be mistaken for retained introns. Therefore, it is difficult to tell an actual retained intron from experimental artifacts [60]. The common analytic approach distinguishes between 5 categories of splicing events (Figure 1.5). However, these basic types of events can be combined to produce more than two isoforms from two neighboring exons. Only analysis specifically-tailored for mixed events would detect such combinations, otherwise, they would be classified as two independent events. Thus, mixed events are often not considered in splicing analysis [59]. It needs to be mentioned here that individual splicing tools perform differently, and the choice of a particular tool should match an underlying biological question. Splicing tools will be further discussed in a latter chapter.

Future studies, however, should incorporate RNA-seq data with iCLIP results and proteomic approaches. The first would give insight on actual binding sites of RNA-binding proteins while the latter would help us understand to what extent alternative transcripts translate to corresponding protein isoforms and how cellular protein levels are mediated by the means of alternative splicing.

1.3 TDP-43 pathologies

TDP-43 was first described as a component of cytoplasmic and nuclear inclusions found in the majority of ALS and approximately half of FTLD patients [2]. Since then, TDP-43 aggregation has been reported as a secondary pathology in Alzheimer disease, Lewy body disorder, hippocampal sclerosis, Perry syndrome and other neurological disease as well as in models of traumatic brain injury [66]; [67]. Furthermore, TDP-43 pathology was later described beyond the spectrum of neuronal disease as in the case of Niemann-Pick C disease [68], a lysosomal storage disease, and in myopathies particularly in inclusion body myositis (IBM) and progressive muscle atrophy (PMA) [66]. This being said, TDP-43 deposition seems to be one of pathological

features pointing to the common mechanisms employed in neuronal and muscle degeneration. Yet, it remains to be established to what extent TDP-43 disturbance contributes to development of different pathologies across tissues. Or, as it has been speculated, formation of TDP-43 aggregates could be a primary event in ALS-FTLD spectrum, while it would represent a secondary event (an epiphenomenon) in other disease [69].

1.3.1 *gain of function* vs. *loss of function* theory

In the context of neurodegeneration and, as recently suggested, muscle wasting, two theories - the *loss of function* and the *gain of function* theory - have been proposed to explain pathologies driven by aberrant TDP-43 behavior. Both theories are widely accepted, nevertheless, there has been an on-going debate [70]; [69]; [71] regarding the contribution of each process to disease development.

The *loss of function* (LOF) theory hypothesizes that upon cytoplasmic accumulation of TDP-43 - which can occur due to missense mutations but also independently of any mutation - lack of TDP-43 in the nucleus leads to disturbance of RNA processing and metabolism. As cytoplasmic aggregates trap more and more protein, this in turn brings about complete nuclear clearance of TDP-43. The *gain of function* (GOF) hypothesis, on the other hand, suggests that the presence of TDP-43 inclusions *per se* is harmful. The two views could not be considered independent, as cytoplasmic aggregates retain the nuclear TDP-43, which would ultimately lead to the loss of its nuclear function. Even though attempts have been made to tell between molecular mechanisms associated with LOF and GOF, these two appear to be rather synergistic and exacerbate the effect of one another [70]; [71]. It has been further suggested that LOF and GOF mechanisms would act at different stages of disease [70].

In addition, more toxic affects could originate from abnormal trapping of RNA and other proteins that usually bind TDP-43 and thus disturb their action elsewhere [72].

1.3.2 TDP-43 autoregulatory loop

The amount of TDP-43 in the cell and its nucleo-cytoplasmic partitioning are both tightly regulated. Maintenance of constant protein levels within a cell is achieved by targeting various steps of gene expression and post-transcriptional processing. One possible mechanism often exploited by RNA-binding proteins, which bind their own pre-mRNA, is the activation of an alternative splicing event that triggers a nonsense-mediated decay. Indeed, autoregulatory mechanisms have been described in many hnRNPs [73]. Inability of a cell to maintain the correct protein levels is widely recognized as an age-related phenomenon often associated with disease, especially with regards to neurological disorders [74].

Autoregulation of TDP-43 occurs via its 3' UTR region (Figure 1.1). In the 3' UTR of *Tardbp* mRNA, TDP-43 crosslinks to approximately 700 nucleotides long stretch, which, interestingly, does not contain UG repeats, therefore implying a low affinity binding [8]; [16]. Any

drop in nuclear TDP-43 levels (and reduced TDP-43 binding) allows for the use of the optimal poly-adenylation (poly-A1) site (in contrast to less efficient options of poly-A2 and poly-A4) increasing mRNA stability [16]. Another mechanism that could, to a lesser extent, contribute to TDP-43 autoregulation is inclusion of the alternative exon which introduces a premature stop codon. That shortened *Tardbp* transcript is then subject to nonsense-mediated decay [8]; [16]. Additionally, it has been suggested that TDP-43 binding to its own mRNA might influence other steps of transcription like RNA polymerase II stalling or termination defects [16].

With regards to disease, the negative feedback loop would, after initial insult (i.e., formation of early inclusions), propagate TDP-43 aggregation. As cytoplasmic inclusions trap TDP-43 the level of freely available nuclear protein becomes limited, which in turn yields an elevated pool of stable TDP-43 mRNA. Consequently, more protein is being produced fostering further aggregation and exacerbating the pathological process.

1.3.3 Post-translational modifications of TDP-43

Importantly, it has been suggested that TDP-43 morphology and pathological distribution differ among neurodegenerative and non-neurodegenerative disease [66]. In particular, truncated or aggregated TDP-43 is subject to a variety of post-translational modifications (PTMs) in a tissue-specific fashion. PTMs, either reversible or irreversible, allow proteins to readily change their properties in response to internal or external stimuli, as PTMs influence protein stability and aggregation, but also its sub-cellular localization, functionality, protein-protein or protein-RNA interactions and protein clearance by the proteosome/autophagic response.

It has been shown, for example, that the half-life of full-length TDP-43 differs from that of its truncated form or ubiquitinated/phosphorylated TDP-43 species. Whereas the turnover of full-length TDP-43 was estimated to be 12-34 h, C-terminal domain (CTD) fragments appear to be less stable. On the other hand, phosphorilation and ubiquitination render TDP-43 less prone for degradation possibly by interacting with UPS-mediated degradation or merely by changing protein solubility (Figure 1.9) [69].

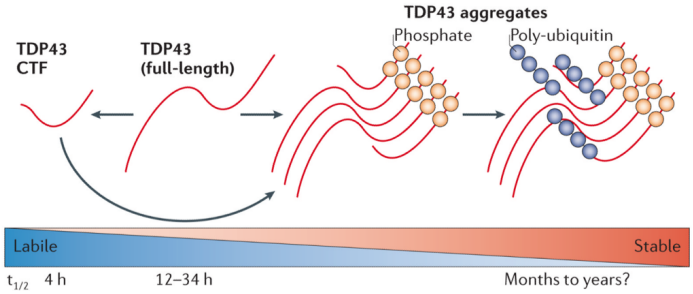


Figure 1.9: **Stability of post-translationally modified TDP-43.** TDP-43 modifications influence TDP-43 stability and turnover [69].

Interestingly enough, sub-cellular distribution and morphology of TDP-43 differ among neu-

rodegenerative and non-neurodegenerative disease, and they do so also across patients diagnosed with the same disease. In part, this could be explained by distinct post-translational processing of this protein. Phosphorylated and ubiquitinated TDP-43 species, which are not readily detected in healthy brain and seem to be a disease-specific phenomenon [69], were detected in ALS/FTLD spectrum. Other PTMs described in ALS/FTLD neurons are generation of CTD fragments, acetylation, cystein oxidation and sumoylation (Figure 1.10). TDP-43 in IBM, however, does not seem to be subject to the same repertoire of PTMs. Also CTD fragments that are commonly present in ALS and play an active role in aggregation [75], have not been observed in IBM. What cannot be excluded here, however, is that to date these PTMs have not been detected in muscles owing to the fact that the presence of TDP-43 pathology in IBM was discovered relatively late compared to ALS and less studies have investigated TDP-43 modification in muscles.

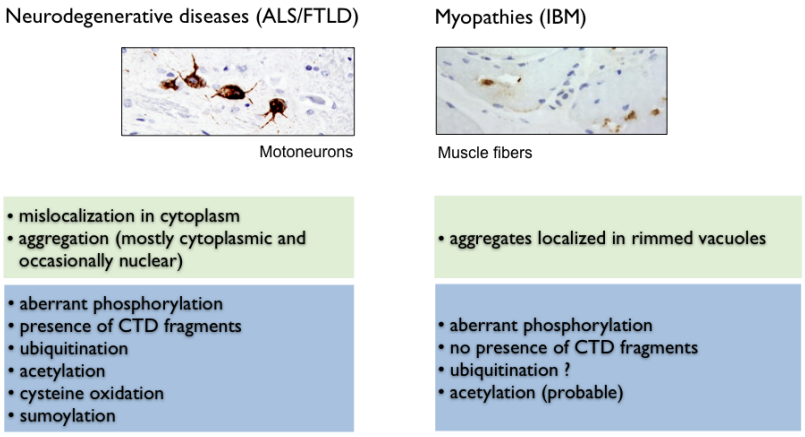


Figure 1.10: **Pathological features of TDP-43 in neurons and muscles.** A characteristic pathological localization of TDP-43 in ALS/FTLD neurons and IBM muscles with the summary of known PTMs affecting each tissue (modified after [66]).

1.3.4 Amyotrophic lateral sclerosis

TDP-43 forms the major component of cytoplasmic and nuclear inclusions found in the brain of the majority of ALS patients and approximately half of FTLD cases [2]; [3]. Excessive effort has been made to reveal the exact role TDP-43 plays in progression of these devastating diseases. Not much could be explained by genetics, since 95% of the patients exhibiting TDP-43 inclusions do not carry mutations in *TARDBP* gene [72].

ALS is a classical neurodegenerative disease characterized by progressive degeneration of motor neurons in the brain and spinal cord, which make direct or indirect connections to innervate muscles. Atrophy of target muscles arises as a consequence of neurodegeneration. Disease onset commonly occurs in mid-adulthood (the mean age being 55 years), although it sometimes begins very early (as early as in twenties) or very late in life. Low prevalence of ALS (about 5

cases in 100,000) reflects rapid mortality. As with other neurodegenerative diseases, ALS starts focally and gradually spreads to neighboring areas. Some motor neurons, for example those that innervate extraocular muscles, are spared until the very late stage. Nevertheless, disease progression gradually leads to paralysis of almost all skeletal muscles and most patients die within 3-5 years after diagnosis [76]; [77].

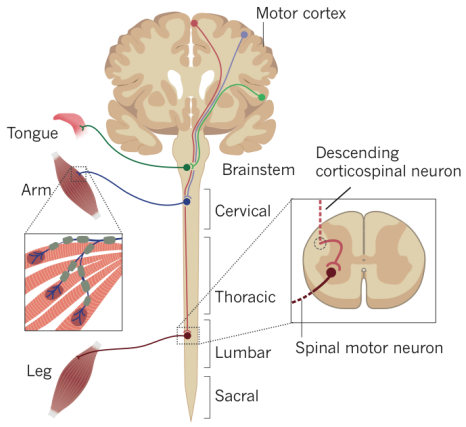


Figure 1.11: **Neuronal populations affected by ALS.** ALS affects the descending upper motor neurons that project from the motor cortex into synapses in the brain stem and spinal cord, and bulbar or spinal motor neurons (lower motor neurons) that project into skeletal muscles [76].

Around 10% of ALS is transmitted in families, mostly as a dominant trait with high penetrance. The first gene recognized to cause ALS was Cu-Zn superoxide dismutase 1 (*SOD1*). Since then, a number of ALS-related mutations have been found in genes including *TARDBP*, *C9orf72*, *FUS*, *VCP* and some others. Genes mutated in ALS encode for proteins that could be by functionality classified in the following categories (Figure 1.12); proteins involved in protein quality control and disposal (*VCP*, *OPTN*); proteins participating in mRNA metabolism (*TARDBP*, *FUS*); and proteins that control cytoskeletal dynamics and axonal architecture (*DCTN1*). This being so, mutations in ALS-associated genes would perturb various cellular processes and contribute to heterogeneity of ALS. The complexity further increases with the list of genetic variants that enhance susceptibility to ALS or that modify the clinical phenotype but are not considered to cause ALS on their own [76]. Lately, it has been even discussed whether ALS should be considered a syndrome rather than a single disease, for the existence of many subtypes that comprise a wide clinical and pathological spectrum [77].

What is common to almost all ALS cases (both, familial and sporadic), however, is mislocalization and aggregation of TDP-43, therefore, TDP-43 inclusions are generally recognized as a hallmark of ALS and some FTLD forms. The importance of TDP-43 disturbance and its direct involvement in ALS was supported by the discovery of ALS-related mutations in *TARDBP* gene. Mutations were, nonetheless, only described in a small subset of patients (namely in 5% of familial and in 1% of sporadic cases). This finding was followed by identification of mutations

in genes encoding other RNA-binding proteins, which drew substantial attention to the role of RNA biology in ALS pathogenesis [76]; [77].

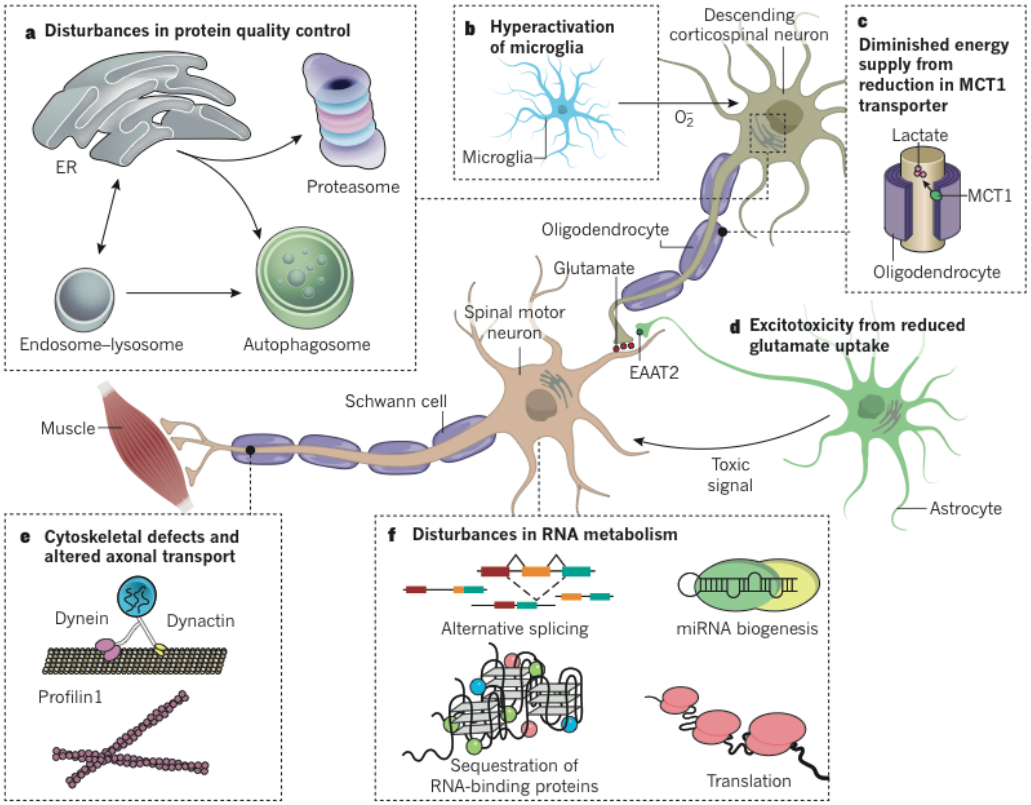


Figure 1.12: **ALS pathology.** Currently proposed picture of ALS pathology involves motor neurons along with neighboring cells that support the motor system. ALS-causing mutations often occur in genes involved in protein quality control **(A)**, mRNA metabolism **(B)**; and axonal transport **(C)**. Neuronal impairments are exacerbated by non-neuronal contribution of glia, astrocytes and oligodendrocytes by the means of neuroinflammation **(D)**; diminished energy supply **(E)** and glutamate exitotoxitcity **(F)** [76].

Studies of ALS have initially focused on neuronal, i.e., cell-autonomous mechanisms, but the interest later expanded to investigate non-cell-autonomous pathology in diseased brain [78]; [76]. It is now becoming clear that neurodegeneration can arise, at least in part, from mechanisms including non-neuronal partners, rather than from neuronal damage alone (Figure 1.12). Other cell populations that appear to be involved in ALS are the following; microglia that are innate immune cells of nervous system; oligodendrocytes, which myelinate axons of upper motor neurons and thereby directly support nuronal cells by providing energy metabolite lactate through monocarboxyalte transporter 1 (MCT1); and astrocytes, another glial cell-type, which carries out recycling of excitatory neurotransmitter glutamate. In support of this view, neuroinflammation, reduction in MCT1 accumulation and glutamate excitotoxcicity were all shown to influence ALS progression [76].

Muscle waisting, which has for long been considered exclusively as a consequence of neuronal loss and muscle dennervation, has lately been questioned as well. Apart from muscle dennervation and atrophy, pathological features like inflammation, fiber necrosis and expression of apoptotic proteins have been reported in ALS muscle [78]. Although early studies failed to detect TDP-43 abnormalities in quadriceps of ALS patients [79], a recent study, in which they extended their examination to other muscle groups (namely axial muscles), identified TDP-43-containing inclusions in muscle fibers of 30% of ALS patients including familial and sporadic cases [78]. ALS myofibers suffer from oxidative stress, mitochondrial dysfunction and bioenergetic disturbances. Yet, it remains unresolved, whether ALS could affect skeletal muscles in a manner that is independent of denervation. Selective degeneration of myofibers of different types depends on their contractile and metabolic features [80].

The evidence of environmental risk factors that might influence ALS progression remains elusive. Frequently studied environmental factors have been chosen arbitrarily and were based on individual observations (exercise, football) or borrowed from other neurodegenerative disease (the effect of smoking) [77].

Energy metabolism in ALS

With our limited knowledge on the role of external environmental factor in ALS development, there is a new interest in the impact the intrinsic environment might have [76]. Results from epidemiological studies imply that people with low BMI are at increased risk to develop ALS, while ALS patients are reported to have higher metabolic rate than control subjects [81].

Despite the evidence coming from epidemiological studies being rather scarce, animal research supports the idea that energy intake can influence disease, as high-energy diet was able to delay ALS onset and slow disease progression [81]. The apparent vulnerability of motor neurons to negative energetic balance contrasts what has been previously observed in other neuronal populations (i.e., those in the hippocampus, cerebral cortex and substantia nigra), which might in fact benefit from energy restriction. Therefore, motor neurons have recently been proposed to have a limited ability to adaptively respond to metabolic stress, for example by increasing production of stress-resistant proteins [81]. These preliminary observations highlight the importance of understanding the interplay between metabolic pathways and disease mechanisms, with the aim to develop strategies to correct metabolic disturbance observed in ALS cases.

1.3.5 Inclusion body myositis

Beyond the spectrum of neurodegenerative disease, aberrant TDP-43 behavior was observed in muscles of patients with inclusion body myositis (IBM). Sporadic IBM (hereafter referred as IBM) is a slowly progressive degenerative and inflammatory disorder of skeletal muscles of unknown cause that begins in middle or late life with no effective treatment available. IBM was not associated with reduced life expectancy, yet, it causes significant disability. The prevalence

widely varies across populations, the highest being reported in the US and Australia while it is considered rare in Asians and African-Americans. A meta-analysis from 2017 assessed a pooled prevalence of 46 patients per 1,000,000 but it also highlighted a steep increase in the prevalence over the last decades, very likely due to increased diagnostic capability [82].

IBM is the most common acquired myopathy with selective involvement of specific muscles (Figure 1.13a), for example finger flexors and quadriceps. Pathologic examination usually reveals extensive endomysial lymphocyte inflammation (mostly cytotoxic T cells) and some less specific features like fibrosis, variation in myofiber size and the presence of necrotic fibers. Diagnosis is generally based on clinical symptoms, supported by laboratory (increase in serum creatine phosphokinase and anti-cN1A) and pathological examination. Due to the lack of a specific marker, diagnosis is often delayed or cases are misdiagnosed [83]; [84].

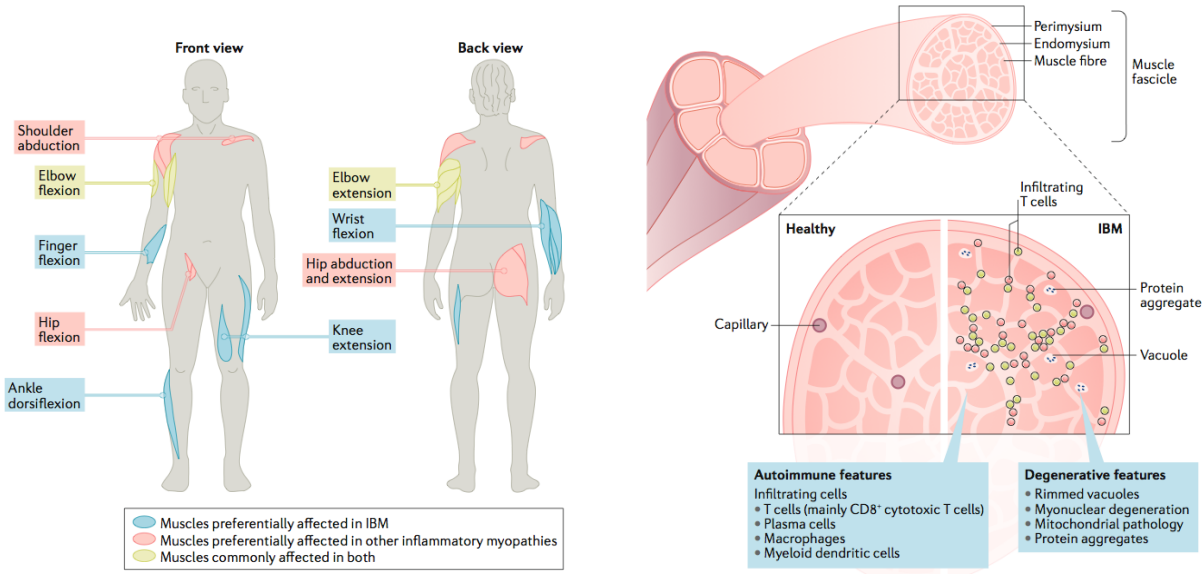


Figure 1.13: **IBM pathology. A** Selective involvement of muscles in IBM with **B** characteristic pathological features [84].

IBM etiology is now understood as an interplay between a degenerative and an inflammatory (autoimmune) component (Figure 1.13a). Some evidence suggests that it is the inflammation which drives degeneration and not *vice versa* [85]. Degenerative features commonly observed in IBM myofibers include; rimmed vacuoles, which were proposed to arise from the breakdown of myonuclei [86]; mitochondrial abnormalities [87]; and cytoplasmic protein inclusions coupled with a burden of ER stress and proteosome dysfunction [84]. The concept of protein aggregation initially focused on amyloid β accumulation and consequently, many IBM mouse models [88] were generated featuring this pathology. However, amyloid β and amyloid β precursor protein (AβPP) were criticized to be non-specifically expressed in degenerating myofibers in all muscle disease. Tau and ubiquitin had similar limitations. Proteins that show higher specificity for IBM

amongst other myopathies are two autophagic markers (p62, LC3) and aggregated TDP-43. The latter was even suggested to have some diagnostic value [89]; [90].

As in neurons, muscle TDP-43 should have predominantly nuclear localization. In IBM, TDP-43 was consistently reported to mislocalize in the cytoplasm (sarcoplasm). In one study, 23% of IBM myofibers were found to contain sarcoplasmic aggregates, which sometimes line around rimmed vacuoles. Sarcoplasmic redistribution was accompanied by nuclear clearance, since TDP-43 was only detected in 12% of the nuclei [90]. Another study [91] reported the presence of deposits in 78% of IBM patients. TDP-43 aggregates were seen in non-necrotic myofibers, typically with minimal morphological abnormalities. Ubiquitination of TDP-43 was reported in some studies [91] but not in others [90].

Taken together, sarcoplasmic deposition of TDP-43 seems to be sensitive and distinguishing feature of IBM among other inflammatory myopathies [90]; [89]. In the need of diagnostic tools, it has been even suggested it could serve as a diagnostic immunohistochemistry marker [92].

Familial form of IBM (sometimes referred to as hIBM, hereditary IBM) has some pathologic similarities to other familial myopathies, but it essentially lacks two main characteristics of sporadic IBM - distinctive muscle involvement and immune cells infiltration [84]. What they do have in common, however, is the presence of TDP-43 inclusions [90]; [91]. One rare form of hIBM occurs together with Paget's disease of bone (PDB) and frontoteporal dementia (FTLD). This form is caused by mutations in valosin-containing protein (*VCP*) gene [93], that participates in protein degradation via UPS and autophagy [94]. As *VCP* mutations were also described in ALS/FTLD, this might be indicative of a genetic overlap which might potentially place IBM in the broader spectrum of degenerative disorders [76] and even imply a mechanistic link. So far, no TDP-43 mutations have been found in IBM patients, but this question could only be addressed in a study with a large patient number.

An intervention study, in which inflammatory stimuli caused cytoplasmic mislocalization of TDP-43 in glia and astrocytes, is in keeping with the view that inflammatory insult acts as a trigger of aggregate formation in muscle [95]. On the other hand, the notion that IBM patients are poorly responsive to anti-inflammatory or immunosuppressive treatments suggests that inflammation *per se* may not be a primary cause of the disease.

How aggregation and mislocalization of TDP-43 impact its function in muscles is currently unknown. Redistribution of TDP-43 together with other hnRNPs [96] sheds some light on putative disturbances in RNA homeostasis. The exact contribution of TDP-43 to govern different steps of mRNA metabolism in skeletal muscles will be addressed in the scope of this thesis.

1.3.6 Aberrant splicing in TDP-43-proteinopathies

Transcriptome-wide studies investigating the consequences of TDP-43 loss, overexpression, or mutations within *TARDBP* gene demonstrated a global effect of TDP-43 on gene expression and alternative splicing. Most of the evidence comes either from cell culture experiments or

from mice models, while analysis of splicing in affected neuronal populations derived from ALS patients or IBM muscle is relatively scarce. However, it is necessary to detect the signature of aberrant TDP-43 behavior *in vivo* and to asses what role altered mRNA processing, in particular splicing dysfunction, plays in development of TDP-43 pathologies. Since TDP-43 levels have been consistently reported to be increased or unchanged in TDP-43 proteinopathies [90]; [91], perturbed splicing might reflect *loss of function* effect caused by cytoplasmic TDP-43 trapping or impairment in protein functionality.

One study reported disturbance of TDP-43-mediated splicing in brain tissues affected by ALS as it was assessed by altered splicing of *POLDIP3* mRNA. In ALS brain, exclusion of exon 3 within *POLDIP3* resembled depletion of TDP-43 in cultured human cells [97]. Another study [98] investigated splicing patterns of 4 TDP-43 target genes - *RWDD1*, *RABGEF1*, *EIF4H* as well as *POLDIP3* - in ALS and healthy brain. Their results suggest that TDP-43 exhibits enhanced splicing function for some exons, and reduced function for some others. Furthermore, altered *BIM* splicing was detected in the brain of FTLD patients positive for TDP-43 inclusions, the pattern of which was consistent with that of TDP-43-silenced neuroblastoma cells [99]. Depletion of TDP-43 was found to cause cryptic splicing (i.e., inclusion of cryptic exons) in mouse and human [100]. Cryptic exons are defined as newly-emerged exons derived from intronic sequences by creation of new splice sites. Under normal condition, TDP-43 appears to repress cryptic exons [24]. In the absence of the right negative splicing regulators (among them also TDP-43), the inclusion of cryptic exons within the host mRNAs would either produce functional or truncated protein, or would alternatively introduce frameshifting followed by NMD. Downregulation of some genes (*ATG4B*, *GPSM2*) which contain cryptic exons was indeed observed in the brain of ALS patients, suggesting that cryptic splicing might play a role in TDP-43 pathologies [24].

A study comparing gene expression and splicing pattern in motor neurons of ALS patients and healthy controls [101] identified immense differences between the two, yet, they failed to demonstrate a direct link between splicing alterations and TDP-43 aggregation. As TDP-43 plays multiple roles in RNA metabolism and interacts with numerous RNA targets and protein partners, this makes it challenging to distinguish between direct effects of TDP-43 depletion from indirect ones.

Beyond neuronal tissues, changes in RNA metabolism have been observed in muscle biopsies from IBM patients. Splicing alterations were attributed to TDP-43 mislocalization, but also to aberrant localization of other hnRNPs [102]. Consistent with results coming from ALS tissues, there is an evidence of altered *POLDIP3* splicing (Cortese *et al.*, unpublished) in IBM muscles. Transcripts containing long introns, which are often targets of TDP-43 [8], appear to be enriched among genes downregulated in IBM [102]. Moreover, altered splicing of *LMNA* was detected in IBM patients [103]. Although that event was not linked to TDP-43 aggregation, it points to the relevance of splicing dysfunction in IBM.

Beside variations in nuclear TDP-43 levels, mutations in TDP-43 could affect splicing activity. Genome-wide splicing changes were investigated in mice expressing human TDP-43 carrying ALS-associated mutations. Mutated protein exhibited normal or enhanced splicing activity for some mRNA targets, but resembled *loss of function* effect for some others [57]. Splicing abnormalities were indeed seen in fibroblasts derived from ALS patients carrying *TARDBP* mutations [104].

1.4 TDP-43 in energy metabolism

Growing evidence points towards impaired energy homeostasis in ALS patients, manifested in hypermetabolism, hyperlipidemia and mitochondrial dysfunction [105]; [106]; [107]. This poses a question whether TDP-43, the aggregation of which is a hallmark of ALS, might be directly involved in development of energetic imbalance. Motor neurons are large, highly polarized and excitable cells with high energetic demand, as they need to maintain the resting membrane potential and the excitability [108]. It is thus possible that these features make motor neurons more vulnerable relative to other cell-types.

In any cell, AMP-activated protein kinase (AMPK) acts as the central energy sensor that gets activated when ATP levels are low or cells get stressed. AMPK is regulated by multiple upstream pathways triggered by environmental stimuli. It has been suggested that abnormal AMPK activation leading to energy disturbance may be relevant for ALS. One study reported that increased AMPK activation enhances nuclear clearance of TDP-43 and its cytoplasmic accumulation. Cytoplasmic deposition of TDP-43 appears to increase reactive oxygen species (ROS) production, which subsequently activates AMPK [109]. Another study conversely demonstrated diminished AMPK activation in spinal cords of transgenic mice, presumably due to upregulation of AMPK phosphatase by mutant TDP-43 [110]. Conflicting results might reflect differences in modulation of AMPK activity by wild-type and mutant TDP-43, the extent of protein overexpression, and different experimental settings.

Furthermore, mitochondrial abnormalities seem to be a pathologic feature of ALS but also IBM. Two studies [111]; [112] found unusual distribution of mitochondria in the brain of mice expressing human TDP-43. Mitochondria are normally evenly distributed around neuronal cell, yet, they accumulated and formed clusters in transgenic mice [111]; [112]. Therefore, TDP-43 was proposed to impact mitochondrial function indirectly by interfering with their intracellular trafficking, which would subsequently lead to abnormal mitochondrial distribution and clustering [111]. Aberrant localization was accompanied by increased fission and reduced fusion markers [112]. Apart from regulating mitochondrial morphology and movement, TDP-43 might additionally affect mitochondrial bioenergetics. A small amount of cytoplasmic TDP-43 localizes in the inner mitochondrial membrane under normal condition, while ALS-associated mutations increase its mitochondrial accumulation. There, TDP-43 interacts with mRNAs of mitochondrial

genes encoding for respiratory complex I subunits, lowers their expression and causes complex I disassembly [21]. Reduced mitochondrial membrane potential, oxygen-consumption rate and cellular ATP levels were observed in HEK293 cells overexpressing wild-type or mutant TDP-43 [21]. Lower cellular ATP levels were also measured in TDP-43-silenced MCF7 cells [20]. In muscles of IBM patients, full-length and truncated TDP-43 colocalized with mitochondrial markers (complexes I-V of respiratory chain) in affected fibers (while these same fibers lacked TDP-43 inclusions), indicating that mitochondrial changes might precede TDP-43 aggregation. Mitochondria were increased in mass, but displayed lower expression of respiratory complexes I and III [113].

Chiang *et al.* [114] generated a *Tardbp* conditional knockout mouse and the postnatal depletion of TDP-43 resulted in rapid fat loss followed by death. Increased fatty acid consumption in those animals was directly linked to downregulation of *Tbc1d1*. *Tbc1d1* encodes for a Rab-GTPase that is involved in translocation of glucose transporter (GLUT4) from perinuclear space to the plasma membrane, therefore, deletion *Tardbp* impaired glucose uptake in skeletal muscle through the mechanisms including Tbc1d1. Consequently, TDP-43 was suggested to be an important regulator of fat metabolism and glucose homeostasis.

2. Materials and methods

2.1 Cell culture

C2C12 immortalized mouse myoblasts (ECACC), NIH-3T3 undifferentiated fibroblasts (available in house), SH-SY5Y human neuroblastoma (ECACC) and RH-30 human rhabdomyosarcoma (kindly donated by Marc-David Ruepp) were maintained in DMEM (Thermo Fisher Scientific), supplemented with 10% FBS (Thermo Fisher Scientific) and antibiotics/antimycotics (Sigma-Aldrich) under standard conditions. NSC34 motorneuron-like mouse hybrid cell line (available in house) was cultured in DMEM (Thermo Fisher Scientific) with 5% FBS (Sigma-Aldrich) and antibiotics/antimycotics (Sigma-Aldrich). All experiments were performed with cells of similar passage number (± 2).

2.1.1 siRNA transfection

To silence TDP-43 in C2C12, NSC34 and NIH-3T3 cells, 40 nM of siTDP or non-targeting siLUC (Table 2.1) were mixed with 54 μl of RNAiMAX (Invitrogen) following the manufacturer's reverse transfection protocol and applied to cells 700,000 seeded in a 10 cm dish. 48 h later, transfected cells were collected for subsequent analysis. The same reagent was used to silence TDP-43 in human SH-SY55Y and RH-30 cells. 400,000 RH-30 were seeded in a 60 mm dish, reversely transfected (human siTDP or siLUC, sequence in ADD) and harvested 48 h later. To deplete TDP-43 in SH-SY5Y cells, 1,000,000 cells were seeded in a 6 mm dish and reversely transfected. After 48 h they were transfected again and harvested 48 later.

Table 2.1: **List of siRNA sequences.**

siRNA	Sequence 5'-3'
siLUC	5'-UAAGGCUAUGAAGAGAUAC-3'
siTDP mouse	5'-CGAUGAACCCAUUGAAAUA-3'
siTDP human	5'-GCAAAGCCAAGAUGAGCCU-3'

2.1.2 C2C12 differentiation

C2C12 myoblasts were seeded into 6-well plates (170,000 cells) and let grow for 24 h hours. Transfection (siLUC/siTDP) was performed on two consecutive days as described above. 72 h after seeding when the confluence reached > 90%, growth medium was replaced by differentiation medium - DMEM supplemented with 2% horse serum (Thermo Fisher Scientific) and antibiotics/antimycotics (Sigma-Aldrich). C2C12 cells were let differentiate for 4 days (or 7 days for untreated cells) changing differentiation medium every 24 h. They were harvested on day 0 (at confluence prior to application of differentiation medium) or at later stages of differentiation.

2.2 ATP assay

Cellular ATP levels were assessed using ATPlite 1step kit (Perkin Elmer) and normalized to cell number. ATP luminescence assay measures production of light caused by the reaction of ATP with added lucifearse and D-luciferin. The emitted light is proportional to the ATP concentration in the suspension.

Herein, we measured ATP levels (intracellular and extracellular) in TDP-43-silenced C2C12 myoblasts. The first round of transfection (siTDP or siLUC) was performed (like described above) on 35 mm plates. Transfected cells were harvested 24 h later and silenced for the second time in tubes such to have 3,000 cells/100 μl of the transfection suspension. 100 μl of the mix was then pipetted into transparent 96-well plate with flat bottom and cells were let attach for 24 h hours. After the incubation, the plate was adapted to room temperature and 100 μl of ATPlite buffer was added to each well containing cells. Culture plates were than shaken for 10 min at 200 rpm for cells to lyse. After the lysis step, the content of each well containing lysed cells was transfered into a white 95-well plates (Perkin Elmer). To reduce the autofluorescence of the plate, white plates were dark-adapted for 10 min prior to the measurement. Luminescence was then measured Perkin Elmer plate reader.

Another 96-well culture plate was prepared in parallel to (indirectly) control for cell number in each well and to account for possible differences in proliferation rate of TDP-43-silenced cells. As ATP-measurement requires cell lysis, cells (nuclei) could not be counted on that same plate. This other 96-well plate was taken from the incubator along with the plate prepared for ATP-measurement 24 h after the second silencing. 50 μl of the medium were removed from each well and substituted with 50 μl of 4% PFA to fix the cells. After 20 min, wells were washed with PBS, nuclei stained with Hoechst stain (Invitrogen) and counted on high-throughput sampler. Each experiment was performed three times with six (technical) repetitions per group. A non-transected control was included in each experiment to account for possible effects of the treatment (transfection was done using RNAiMAX (Invitrogen)) on ATP production. ATP levels were normalized to cell number assessed in parallel.

2.3 Western blot

Whole-cell extracts were resuspended in PBS in the presence of protease inhibitor and sonicated. 15 μg of protein sample were separated on a 10% Bis-Tris gels (Invitrogen) and transferred to the nitrocellulose membrane (Invitrogen). Prior to antibody staining, the membrane was blocked in 4% milk-PBST. The following primary antibodies were routinely used in our experiments: anti-TDP-43 (rabbit, Proteintech, 1:1,000), anti-GAPDH (rabbit, Proteintech, 1:1,000), anti-HSP70 (rat, EnzoLife Science, 1:1,000), anti-tubulin (mouse, available inhouse, 1:10,000) along with HRP-conjugated secondary antibodies: anti-rabbit (goat, Dako, 1:2,000), anti-mouse (goat, Dako, 1:2,000) and anti-rat (rabbit, Dako, 1:2,000).

To detect TBC1D1, 30 μg of the protein sample was separated using a pre-cast gradient gel (4-12%, Invitrogen) and wet blotted onto a PVDF membrane (Merck Millipore). The membrane was then blocked in 4% milk-TBST, stained with anti-TBC1D1 (rabbit, V796, Cell Signaling, 1:1,000, in 2% milk-TBST) and anti-rabbit HRP-conjugated secondary antibody (goat, Dako, 1:2,000).

ECL Star (Euroclone), Luminata (Merck Millipore) or Super Signal Femto (Thermo Fisher Scientific) substrates were used for signal detection according to manufacturer's instructions. Image acquisition and result quantification were conducted using Alliance Q9 Advanced Chemiluminescence Imager (UviTech).

2.4 RNA-extraction and reverse-transcription

Total RNA was isolated using standard phenol-chlorophorm extraction. Briefly, cell pellet harvested from a 10 cm dish was resuspended in 500 μl of Trifast (Euroclone) and let incubate for 5 min. After adding 500 μl of chloroform and 10 min incubation at room temperature, tubes were centrifuged for 20 min, at 4 °C, 16,000 rpm. RNA-containing aqueous phase was transfered to a new tube, mixed with 500 μl of chloroform and centrifuged for 5 min at 4 °C, 16,000 rpm. Next, the aqueous phase was transfered to a new tube and mixed with approximately 500 μl of isopropanol for RNA to precipitate. It was then incubate on ice for 10 min followed by 20 min centrifugation at 4 °C, 16,000 rpm. RNA pellet was washed with 1 ml of 70% ethanol and centrifuged for 10 min at 4°C, 16,000 rpm. Finally, the ethanol was removed and the pellet was resuspended in water. Only undegraded (RIS $>$ 8) RNA of high purity (A260/A230 and A260/A280 $>$ 1.8) was taken for subsequent analysis (PCR experiments and RNA-sequencing).

500 ng of RNA was reversely transcribed using random primers (Eurofins) and Moloney murine leukaemia virus reverse transcriptase (M-MLV, Invitrogen) according to manufacturer's instructions. The reaction was let to occur for 90 min at 37 °C.

2.5 qPCR

For assessment of transcript levels, real-time quantitative PCR was performed using PowerUp SYBR Green master mix (Applied Biosystems) and gene-specific primers (primer sequences in Table 2.2). cDNA was subjected to 45 cycles of the following thermal protocol: 95 °C for 3 min, 95 °C for 10 s, 65 °C for 30 s, 95 °C for 10 s, 65 °C for 1 s. Relative gene expression levels were determined using QuantStudio design and analysis software (Thermofisher Scientific (v1.5.1)) always comparing treated samples (siTDP) with their direct controls (siLUC) normalized against *Gapdh*. p-values were calculated using one-tailed paired t-test as qPCR was conducted to validate expression changes detected by RNA-seq.

Table 2.2: **List of qPCR primers, mouse.**

Gene name	Forward primer 5'-3'	Reverse primer 5'-3'
Tardbp	GCAGTCCAGAAAACATCTGAC	ACACCATCGCCCATCTATCAT
Myog	CAGCCCAGCGAGGGAATTTA	AGAAGCTCCTGAGTTTGCCC
Igf2	CGCTTCAGTTTGTCTGTTCG	AAGCAGCACTCTTCCACGAT
Rfc2	CTGCCGTGGGTTGAAAAATACA	CAGAGGATGCTGGTTGTCTTG
Slmo1	CCACCAATATCACGCTCACGA	CCCCAAGCTAATTCCCTTCACA
Arhgef18	TCAGACAGAAGTGTGGTCCG	GGAGACTGCGAGAGCGAC
Dusp2	ATGGTGGAGATAAGTGCCTGG	GGCTCTGAATCAGGTATGCCA
Tbc1d1	ACAGTGTGGGAAAAGATGCT	AGGTGGAACTGCTCAGCTAG
Myh1	CGGTCGAAGTTGCATCCCTA	TTCTGAGCCTCGATTCGCTC
Myh3	ATGCTTCTCTCTGTCACAGTC	AAGGGCTGGTTCTGAGCTTC
Myolinc short	CTGAATTTGCCTCCCTTCAG	GAAAAGAGGGCATTACACTCGTA
Myolinc long	ACCAGTTGCGAGTGAGCTCCAG	TGGAGGTCCTAAGGTGTTTCCAGT
Myod	GTGGCAGCGAGCACTACA	GCGGTGTCGTAGCCATTC
Myf5	CTGAGGGAACAGGTGGAGAA	CTGTTCTTTCGGGACCAGAC
Myomaker	GCTGGGGTTACGGGGTATAC	GCTGGGTGTAGATGCTCTTGT
Gapdh	AGGTCGGTGTGAACGGATTTG	TGTAGACCATGTAGTTGAGGTCA

2.6 Alternative splicing-sensitive PCR

For detection of alternatively spliced mRNAs, PCR primers were designed complementary to constitutive exonic regions flanking predicted alternatively spliced cassette exon. PCR mix was prepared using gene-specific primers (Sigma, primer sequences in Tables 2.3 and 2.4) and TAQ DNA polymerase (Biolabs or Roche) according to manufacturer's instructions and subjected to 30-45 cycles long thermal protocol optimized for each primer pair. PCR products were separated by capillary electrophoresis (DNA screening cartridge, Qiaxcel) and splicing transitions were quantified using Qiaxcel software (QIAxcel ScreenGel (v1.4.0)). Exon inclusion was calculated by the software. Percentage of the inclusion (Inc. %) reports the area under the curve of the peak representing the longer (inclusion) splicing isoform.

Table 2.3: **List of PCR primers used to detect alternative splicing, mouse.**

Gene name	Exon no.	Transcript	Forward primer 5'-3'	Reverse primer 5'-3'
Poldip3	3	ENSMUST00000058793.13	CATTGGGACTGTAACCCCAG	TGCAAACTTCATCTGCTTGG
Sort1	18	ENSMUST00000135636.5	CAGGAGACAAATGCCAAGGT	TGGCCAGGATAATAGGGACA
Rgp1	3	ENSMUST00000030190.8	TGATCGAAGTGGTAGCTGAGC	AGGGTCTAGCCTCAGGTCAC
Pdp1	2	ENSMUST00000108299.1	TGGTGCTGAGTGAGGGAAGGA	TGCTGGCATGGCATCAGAGAAC
Asap2	24	ENSMUST00000064595.14	TGAGACCTATGGAGCCAT	AGTTCCTGATGTCTTAGCCACA
Tmem2	2	ENSMUST00000237802.1	ATGGAGAGGAGATCTGCA	CATACATGATACCCTGTC
Fam220a	2	ENSMUST00000196487.1	ATCGTGGCTTCCATGATG	CTTAAGGCCACATGCTAG
Dnajc5	5	ENSMUST00000116365.8	TCACCTGCTGCTACTGCT	TGGCAGATGCTGGCTGTAT
Nfya	3	ENSMUST00000046719.13	AGCAATAGTTCCACAGAGC	GACACAGGTACTTGCATGAT
Traf7	4	ENSMUST00000088464.11	AGCTGATGGGACTGGCACAT	AGCACAGCTGGCAACACA
Ppfibp1	4	ENSMUST00000136837.1	TCGCAAAGCCAGGTCCTCA	TCTGGAGATGGTGGAGACA
Sapcd2	2	ENSMUST00000028329.12	TGAGCTGTTATGTGTCCCGGT	CAGGCTGCAGTCCACACCATT
Tbc1d1	12 + 13	ENSMUST00000043893.12	AACTCATGCGGTACCACTCC	TGGCCACTCGAAGGAATATC
Tnik	14	ENSMUST00000159680.8	CACTACGAAGAACAGATGCGTC	TCTGCAGTCTTTCTGCTTGTC
Cipc	3	ENSMUST00000185434.6	ATGAAGAAGCTAGCGCGA	CGCTCTTCATGACGACCAT
Ubr4	23	ENSMUST00000129949.8	AGACTTCCCTTCTGCCTCAG	GCTTTCCAGGGCTTCCAAAG
Lrp8	6	ENSMUST00000106731.4	GTTCTTGCCATCAAACGGTGC	GGTCAGTGCAGATGTGGGAAC
Mef2d	8	ENSMUST00000107559.3	TGATCTGCGGGTCATCACTT	CCCTGGCTGAGTAAACTTGG
Dync1i2	5	ENSMUST00000100028.10	TCCCTCCTCCTATGTCTCCA	CGGGGTCTGAGTTTCCTTTG
Atp6V1H	7	ENSMUST00000044369.13	GCTTGGGGGAAAGAACTGAT	AGGCAAACCGGTACTCATTG
Capzb	9	ENSMUST00000102507.10	GGAAAACAAAATCCGAAGCA	CCTCCACCAGGTCGTTCTTA
Dtna	11	ENSMUST00000115832.4	AAGTCCCTGAGCTGTGCTTC	CGATCAGCCTGTGTTCTTCA
Lrrfip1	5	ENSMUST00000097649.10	GCATCTCTGGGTGGGACTT	GGCGTTGGAGACCATAGCTT
Mef2a	9, 10	ENSMUST00000032776.15	CGAGTTGTCATTCCCCCATCAAGC	GTGTTGTAGGCTGTCGGCATTG
Fxr1	14	ENSMUST00000200392.5	GACTCGACATCAGCGAGACA	GAAGCGCTAGTTGGACCATT
Mtmr3	16	ENSMUST00000128256.2	GTTGGCTACCTGACCACCTG	CTCGACTGGGTTCAAAGAGC
Pkm	9	ENSMUST00000034834.16	CAAGGGGGACTACCCTCTGG	ACACGAAGGTCGACATCCTC

Table 2.4: **List of PCR primers used to detect alternative splicing, human.**

Gene name	Exon no.	Transcript	Forward primer 5'-3'	Reverse primer 5'-3'
POLDIP3	3	ENST00000252115.10	GCTTAATGCCAGACCGGGAGTTGGA	TCATCTTCATCCAGGTCATATAAATT
PPFIBP1	19	ENST00000228425.11	CTCCATGGATGACAACCCCT	TTTTGGCCGAGAAGAAGCAC
ASAP2	23	ENST00000281419.8	AGCATCTTGCAGAATGAGAC	CCTGATCTGTGAGATCCCA
TRAF7	5	ENST00000564067.5	TACAAGCAGCACTGCAGGA	AGCTGACAGCACAGCTTCA
NFYA	3	ENST00000341376.10	AGTTCGACAGAGCAGATTG	GCATGATGGTTTGACCTTGT
TNIK	14	ENST00000284483.12	CAAAGGCGAGAGAAGGAGCTG	CTGATGCTGAAGGGAAACTAAG
TBC1D1	12 + 13	ENST00000508802.5	AGGTATCACTCAGTGAGC	CTTCACAGGAGTCCCACC
SLMO1	5	ENST00000592149.5	ACCGTGCTCACACAAGAAG	TCCTTGCTCGTCCTCTCCTCA
DNAJC5	5	ENST00000470551.1	ACGGAGTTCTACGTGTCC	GCTGTATGACGATCGGCG
RGP1	3	ENST00000378078.5	ATGATTGAAGTGGTAGCAGAG	CTCTCCAGGATCAAGCCT

2.7 mRNA library preparation and RNA-seq

Both polyA cDNA library generation and RNA-seq were performed by Novogene (Beijing, China). cDNA libraries with insert length of 250-300 bp were generated using NEB NextR Ultra RNA Library Prep Kit and the quality of polyA libraries was assessed on the Agilent Bioanalyzer 2100 system. Sequencing was performed on Illumina with paired-end 150 bp (PE

150) sequencing strategy. On average, 135.8 million clean reads were obtained per sample.

2.7.1 Read mapping

* performed by Novogene (Beijing, China)

Raw reads of FASTQ format were then processed through Novogene inhouse scripts to obtain clean reads (131.1 milion on average) by removing reads containing adapters, reads containing poly-N and low quality reads from raw data. Obtained reads were aligned to the mouse genome GRCm38 (mm10) using the Spliced Transcripts Alignment to a Reference (STAR) software (v2.5) [115]. Pair-end clean reads were aligned to the reference genome GRCm38 (mm10). 95% of clean reads (on average) were mapped to the genome, and of these 94.9% of total reads were mapped only once (on average, unique mapping).

2.7.2 Data availability

Datasets generated for this study are deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE171714.

2.8 RNA-seq analysis

2.8.1 Quantification of gene expression level

* conducted by Neva Škrabar

Counting of reads mapped to each gene was performed using HTSeq (v0.6.1) [116]. Raw read counts together with respective gene length were used to calculate Fragments Per Kilobase of transcript per Million mapped reads (FPKM). In contrast to read counts, FPKMs account for sequencing depth and gene length [117] and are frequently used to estimate gene expression levels.

2.8.2 Differential expression

Differential gene expression (DEG) analysis of two conditions was performed using the DESeq2 R package (v2.1.6.3) [118], a tool that utilizes negative binomial distribution model to account for variance-mean dependence in count data and tests for differential expression [119]. Three biological replicates were included per cell-type and condition, in control (siLUC) and TDP-43-silenced (siTDP) cells. Read count matrix was pre-filtered by removing rows with row sum below one and normalized for library size and composition by the regularized log transformation. Multiple testing adjustments were performed using Benjamini and Hochberg's approach to control for the false discovery rate (FDR) ($\alpha = 0.1$, default log fold change threshold (lcfThreshold = 0)). Transcripts with $p_{adj} < 0.05$ were considered as differentially expressed.

DEG identified in both cell lines under different experimental conditions (i.e., the union of differentially expressed genes detected in C2C12 and NSC34 cells) were hierarchically clustered based on their $\log_{10}$(FPKM+1), scaled and visualized with pheatmap R package (v1.0.12) [120].

Further, distance between silenced and control samples of each cell line was illustrated with principal component analysis (PCA), using the R function "prcomp" [121]. For that, raw gene counts were normalized using DESeq2 package (v2.1.6.3) and rlog-transformed values were then plotted as principle component.

Differences in gene expression levels ($\log_{10}$(FPKM+1)) between cell lines were tested for significance using Wilcoxon signed-rank test.

2.8.3 Differential splicing

* rMATS analysis was performed by Novogene (Beijing, China) and MAJIQ analysis was performed by Anna-Leigh Brown

Differential splicing analysis seeks to detect changes in splicing between different conditions or tissues. While advances in RNA-seq have allowed detailed analysis of the transcriptome, the short nature of reads (usually 150-200 bases) which can align to different transcripts produced from the same gene make splicing analysis more challenging compared to detection of expression changes [122].

Nevertheless, a number of computation tools for AS analysis have been developed to date. These are based on two main approaches: isoform-based and count-based methods. Isoform-based methods (e.g., cuffdiff2 [123]) first seek to reconstruct the entire transcript and then estimate the relative abundance of each transcript within a sample. For count-based methods which are further divided into exon-based (e.g., DEXSeq [124]) and event-based methods (e.g., rMATS [59] and MAJIQ [125]), the gene is usually divided into single units (i.e., exons, juctions) and counts are recorded as the number of reads aligned to each counting unit.

There are certain advantages and disadvantages associated with each of these methods. For example, some of them detect unannotated novel splicing events (*de novo*) while others only work on previously annotated events; not all support a desired experimental design (i.e, two groups, two groups + a blocking factor, two groups with paired samples, complex designs). Furthemore, above mentioned methods heavily differ in terms of their performance and results they produce [65].

Herein, C2C12 and NSC34 datasets were analyzed using rMATS (replicate multivariate analysis of transcript splicing) software (v3.2.1) [59] as well as MAJIQ (modeling alternative junction inclusion quantification) (v2.1) [125].

rMATS was chosen for its ability to detect five major types of alternative splicing - skipped exons (SE), mutually exclusive exons (MXE), alternative 5' and 3' splice sites (A5SS and A3SS) and intron retention (RI). rMATS quantifies alternative splicing events, each of which can produce exactly two isoforms. Each isoform is adjusted for its length before calculating the ratio

of two isoforms and testing significance of differential splicing between two conditions. Multiple testing is then corrected using Benjamini and Hochberg's method.

Most of the results discussed herein were obtained using rMATS protocol (unless otherwise stated). Depending on the underlying biological question, we set different inclusion criteria, which will be discussed in a later chapter. To investigate global splicing regulation provided by TDP-43, we set a low stringency threshold at which all splicing events at FDR (false discovery rate) < 0.01 were considered significant irrespective of size of the change ΔPSI (percent spliced in) ($\Delta\text{PSI} \neq 0$) or read coverage of the event. For the assessment of common targets in both tissues, splicing events were defined by exact genomic coordinates of alternatively-spliced exons as well as coordinates of the upstream and the downstream exon. Two events were considered the same only if they matched in all coordinates.

Differential splicing analysis of the same data sets was additionally conducted using MAJIQ (v2.1) as previously described elsewhere [25]. A threshold of 0.1 ΔPSI was used to test significant changes between groups (FDR < 0.1) for each of the junctions.

A subset of splicing events was visually validated (qualitatively) using Interactive Genome Viewer (v2.8.0).

2.8.4 Enrichment analysis

Gene Ontology (GO) [126] and Kyoto Encyclopaedia of Genes and Genomes databases (KEGG) [127] are widely used in gene enrichment analysis to classify lists of individual genes based on their expression pattern, or other similar feature, with the aim to predict disturbed biological processes, functions and pathways or any other general trend within a subset of data [128]. Herein, GO enrichment and KEGG analysis were conducted in R using *clusterProfiler* package (v3.14.3) [128] either on the set of differentially expressed genes ($p_{adj} < 0.05$) or alternatively spliced genes (FDR < 0.01). For AS genes detected in C2C12 dataset, we additionally performed GO enrichment analysis using less stringent threshold for inclusion of alternatively spliced genes (where we considered genes at non-corrected p-value < 0.01 instead of FDR < 0.01). Genes of each dataset were assigned Entrez gene identifiers from Bioconductor mouse annotation package *org.Mm.eg.db* (v3.10.0). Enrichment test for GO terms and KEGG pathways were calculated based on hypergeometric distribution. The resulting GO terms/KEGG pathways were considered significant after applying multiple testing corrections with Benjamini-Hochberg method ($p_{adj} < 0.05$). Subsequently, significant GO terms (category: biological process) were functionally grouped or manually edited depending on the underlying biological question.

2.9 Conservation analysis

2.9.1 PhyloP score

* analyzed by Neva Škrabar and Robert Bakarić

Gene/exon conservation analysis within mouse (mm10) was performed by calculating phyloP (phylogenetic p-values) scores, i.e., *per base* conservation scores, generated from aligned genomic sequences of multiple species [129]. For each differentially expressed gene, an average *per gene* phyloP score was computed with bigWigSummary (UCSC). To calculate phyloP score of TDP-43-regulated alternative sequences (hereafter referred to as *per exon* phyloP score), we considered TDP-43-regulated sequences of all event types. Those include A3SS and A5SS (long and short exon), retained introns, MXE (the 1st and the 2nd exon) as well as sequences of skipped exons.

2.9.2 Mouse-human orthology

All genes of the mouse genome (mm10) together with their human orthologues (hg38) were downloaded from Ensembl Biomart database (5. June 2020, [130]).

To asses the orthology at the exon level, we exported coordinates marking the start and the end of each TDP-43-dependent alternative sequence (i.e., cassette exons, the 1st and the 2nd of two mutually exclusive exons, the longer and the shorter variant of alternatively spliced 5' and 3' sequences as well as retained introns as assessed by rMATS at FDR < 0.1) in BED format and converted them from mm10 to corresponding locations of the hg38 build using the UCSC liftOver tool.

2.10 Patient samples

2.10.1 IBM samples

Muscle biopsies (vastus lateralis or biceps) from 4 patients diagnosed with IBM according to the Griggs criteria [131] and 4 healthy controls were obtained from the Bank of Muscle Tissue, Peripheral Nerve, DNA and Cell Culture (Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy) and the Laboratory of Muscle Histopathology and Molecular Biology (IRCCS-Policlinico San Donato, San Donato Milanese, Italy). Participants were investigated for cramps or fatigue, they underwent regular examination, neurophysiology tests and histological examinations. IBM biopsies were taken from moderately affected muscles and routinely investigated for histological and immunohistochemistry features. In case muscle fibrosis was present, it did not compromise a definite pathologic diagnosis. Basic demographic features of all participants are summarized in Table 2.5. Biopsies were stored at 80 °C. Institutional board reviewed the study and ethical approval was obtained.

Table 2.5: **IBM patient characteristics.**

Sample no.	Diagnosis	Sex	Age of biopsy	Muscle
1	control	F	52	vastus lateralis
2	control	F	54	vastus lateralis
3	control	M	48	vastus lateralis
4	control	F	22	biceps
5	IBM	M	65	vastus lateralis
6	IBM	M	n.a.	vastus lateralis
7	IBM	F	68	vastus lateralis
8	IBM	M	77	biceps

2.10.2 ALS/FTLD samples

The New York Genome Center (NYGC) consortium ALS cohort is a large RNA-seq dataset across multiple neuroanatomical regions and diseases. Patient demographics, as well as sample processing and data analysis have been detailed elsewhere [132]; [25]. Herein, we investigated splicing changes in a sub-cohort of ALS and FTLD patients with TDP-43 pathology (ALS-TDP and FTLD-TDP) and healthy controls while excluding ALS patients with SOD1 mutations or FTLD patients without TDP-43 inclusions.

3. Results

3.1 TDP-43's activity in muscle and neuronal cells

TDP-43 has been first found to form pathological aggregates in motor neurons of patients suffering from amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD), nevertheless, recent discoveries have revealed that pathological behavior of this protein is not limited to the brain. IBM, a disease of skeletal muscle, is just one example of non-neuronal TDP-43 proteiopathy. Clinically, ALS and IBM do not exhibit much similarity, yet, some pathological phenomena seem to occur in both disease (abnormal protein accumulation, post-transcriptional modifications (PTMs) of deposited proteins, defects in regulation of protein disposal, mitochondrial abnormalities) [133]; [91]; [90]; [134]. This prompted us to investigate the role of TDP-43 played in cells of muscle and neuronal origin, more particularly, how the absence of TDP-43 would impact cellular processes in each of them. By doing so, I sought to fill the gap and further explicate the activity of TDP-43 in cells other than neurons.

In the last decade, high throughput methodologies have set the ground for a systematic comparison of TDP-43-targeted RNAs across tissues. However, the overlap of TDP-43-controlled events identified by earlier studies (in different cell-types) is rather poor. By studying neuronal cells in parallel to cells of muscle origin, I attempted to establish the extent to which TDP-43 acts in a tissue-specific fashion.

A direct comparison of two cell-types would be needed, if we consider that distinct post-translational modifications and cleavage products of TDP-43 have been reported in muscles and neurons [66]; that Vogler *et al.* [135] have recently described a muscle-characteristic localization of TDP-43 along muscle development; that distinct cell population have been shown to express cell-type-specific milieu of TDP-43-binding partners [41] and other RNA-binding proteins (RBPs) controlling common mRNA targets [136]; [137]. Moreover, all these differences occur in a context of highly variable transcriptome between tissues including coding and non-coding transcripts [138]; [139]. Therefore, TDP-43 is likely to elicit tissue-characteristic functions by targeting unique subsets of transcripts, which encode proteins participating in tissue-specific cellular pathways and provide crucial structural and functional features of a cell. The consequences of TDP-43 dysfunction in muscles could thus possibly differ from those that have so far been described in the central nervous system.

Jeong *et al.* [140] have previously investigated the ability of TDP-43 to repress cryptic exons across various tissues by conditionally deleting TDP-43 in excitatory neurons or skeletal muscle in mouse. Apart from that, however, this is the first time TDP-43 targets would be explored in muscles and neurons in parallel. As I aimed at understanding a cell-autonomous effect of TDP-43 depletion, I sought to employ a simple and robust model. Well-established myoblast and neuroblast cell lines were thus chosen over tissue-specific knockdowns or primary cell cultures, which, in contrast to cell lines, represent a rather heterogeneous cell population. Due to the lack of an appropriate and well-established cell line derived from human skeletal muscle, I employed mouse cells in the first line of experiments, yet I later investigated conservation of TDP-43 targets between mouse and human.

3.1.1 Contributions

- Anna-Leigh Brown performed splicing analysis of C2C12 and NSC34 datasets using MAJIQ

3.1.2 Similar TDP-43 expression in C2C12 and NSC34

To investigate functions of TDP-43 in neuronal and muscle-derived cells, I took advantage of the most commonly used mouse cell lines representing motor neurons (NSC34) and skeletal muscle (C2C12). NSC34 is a hybrid cell line produced by the fusion of motor neurons from spinal cords of mouse embryos and mouse neuroblastoma cells [141] that is used as a *bona fide* cellular model in ALS research and has thus been widely employed to study TDP-43-associated neurodegeneration [72]; [142]. C2C12, on the other hand, is a mouse myoblast cell line derived from satellite cells [143]. It has already provided a model to investigate pathomechanisms contributing to IBM development [144]; [145]; [146]. Recently, C2C12 were chosen as a cellular model to study the role of TDP-43 in muscle development and regeneration [147]; [135]. For simplicity reasons, our experiments were conducted using undifferentiated cells although both cell lines could readily be differentiated - NSC34 by the application of retionic acid [148] and C2C12 upon removal of serum mitogens.

I initially compared protein levels of endogenous TDP-43 in undifferentiated and untreated cells (Figure 3.1a). Expression of housekeeping proteins, which serve as loading controls in western blot, might vary across tissues, thus, TDP-43 expression was normalized against three loading controls (tubulin, HSP70 and P84) that showed constant levels in two cell lines. In contrast, GAPDH displayed variable expression in C2C12 and NSC34, both, at the protein (Figure 3.1a) as well as at the level of the RNA (Figure 3.1b). Accordingly, GAPDH was only used to normalize TDP-43 expression between experiments conducted within the same cell-type but not for the comparison between cell lines.

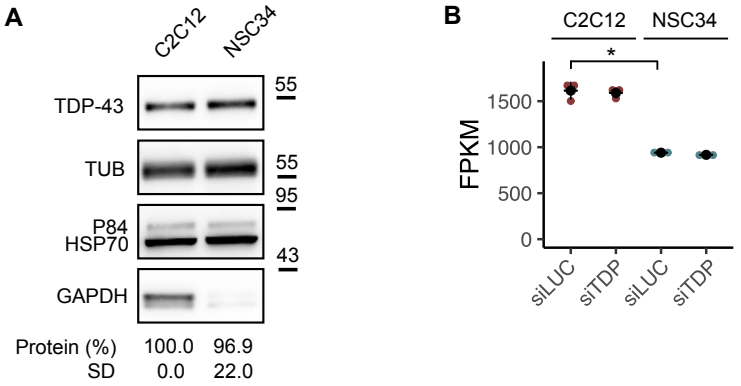


Figure 3.1: **TDP-43 expression in C2C12 and NSC34. A** Western blot shows similar expression of endogenous TDP-43 in C2C12 and NSC34 cells. The amount of TDP-43 was normalized to the sum of peak intensities of three loading controls (tubulin, HSP70 and P84). GAPDH expression varies between C2C12 and NSC34, and was not used for normalization. **B** The plot shows variable expression of *Gapdh* mRNA in siLUC-tranfected C2C12 and NSC34 cells (p-value < 0.05) as assessed by RNA-seq and plotted as FPKM values. Depletion of TDP-43 does not influence *Gapdh* expression. p-value was generated using Kruskal-Wallis test.

TDP-43 is an ubiquitously expressed protein (Protein atlas), the levels of which are tightly controlled by autoregulatoty loop [149]. In mature mouse tissues, TDP-43 was reported to be more abundant in the brain compared to quadriceps muscle [140] and other organs (spinal cord, lung, liver) [150]. Nonetheless, I noted no difference in the amount of total TDP-43 between undifferentiated C2C12 and NSC34 cells (Figure 3.1a) nor in the expression of *Tardbp* at the RNA level of siLUC-transfected cells (Figure 3.2a).

3.1.3 Efficient TDP-43 silencing in C2C12 and NSC34

Given that TDP-43 aggregation in the cytoplasm is accompanied by nuclear depletion, it is still under debate whether aberrant TDP-43 behavior drives disease through mechanisms of *loss of function* or *gain of function*. Our approach, however, models the *loss of function* effect, and would, rather then to investigate pathomechanisms directly, allow to identify mRNAs targeted by TDP-43 in each cell-type.

As complete depletion of TDP-43 (knockout) is detrimental for any cell, I silenced TDP-43 in C2C12 and NSC34 using siRNA. TDP-43 was depleted to a similar extent in both cell lines and reduction of the protein was confirmed by western blot (Figure 3.2a and 3.2b). Loss of TDP-43 functionally reflected in altered splicing of two well characterized TDP-43 targets - *Poldip3* and *Sort1* (Figure 3.2c) [151]; [152]; [153]; [97].

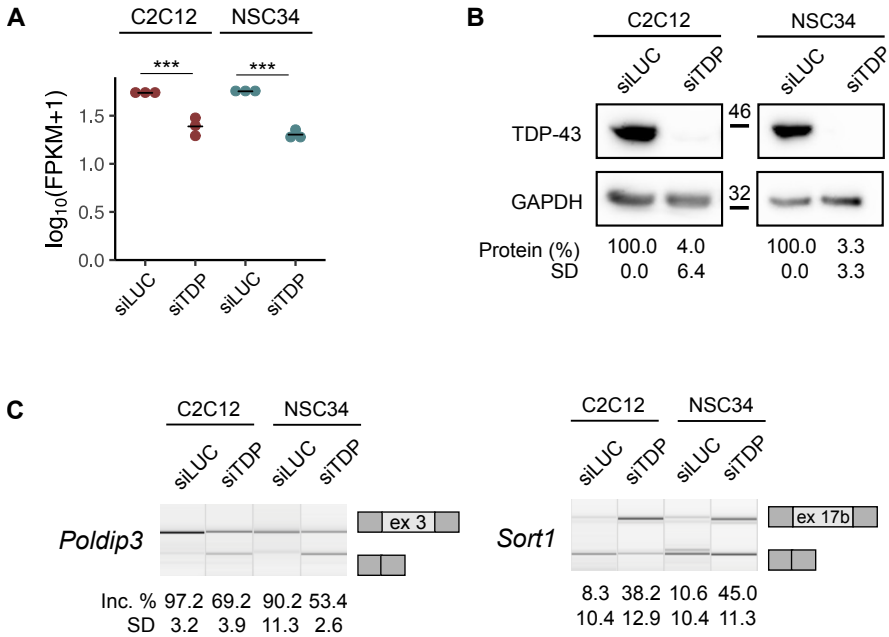


Figure 3.2: **Silencing TDP-43 in C2C12 and NSC34.** **A** Expression levels of *Tardbp* in TDP-43-silenced C2C12 and NSC34 cells along with corresponding controls assessed by RNA-seq and plotted as $\log_{10}$-transformed FPKM values show that TDP-43 was depleted (at the mRNA level) to the same extent in both cell lines. $p_{adj} < 0.001$ for C2C12 and NSC34, respectively. p-values were generated using Wald test and Bejamini-Hochberg multiple testing correction [119]. **B** Western blot shows the reduction of TDP-43 in C2C12 and NSC34 cells upon siTDP transfection. siLUC-transfected cells were used as a control. TDP-43 expression was normalized against GAPDH. **C** TDP-43 depletion led to altered splicing of *Poldip3* and *Sort1*. Semi quantitative RT-PCRs conducted in TDP-43-silenced samples and corresponding controls are shown along with the quantification of splicing changes (% of alternative exon inclusion). The number of the alternative exon is given at the scheme (see the exact transcript numbers in Table 2.3.

Tissues (and cell-types) substantially vary with regards to the expression of individual genes, as these differences define specific biological characteristics and functions. Herein, we studied transcriptome-wide effects of TDP-43 downregulation by performing RNA-seq on polyadenylated mRNA extracted from TDP-43-depleted cells. Compared to the TDP-43-induced changes in gene expression, there is, among all, a considerable variability between C2C12 and NSC34 (siLUC) reflecting a cell-type-characteristic transcriptome as revealed by the first component (PC1) of the principal component analysis plot (PCA) (Figure 3.3a). Nevertheless, TDP-43 silencing accounts for the great majority of the variance between samples within each cell line (PC2). This indicates that depletion of TDP-43 promotes transcription alterations in C2C12 and NSC34, but it does so based on the tissue-characteristic transcriptional profile. Indeed, cluster analysis of differentially expressed genes (DEGs) detected in either cell line and plotted as a heatmap (Figure 3.3b) shows a clear distinction between two cells lines.

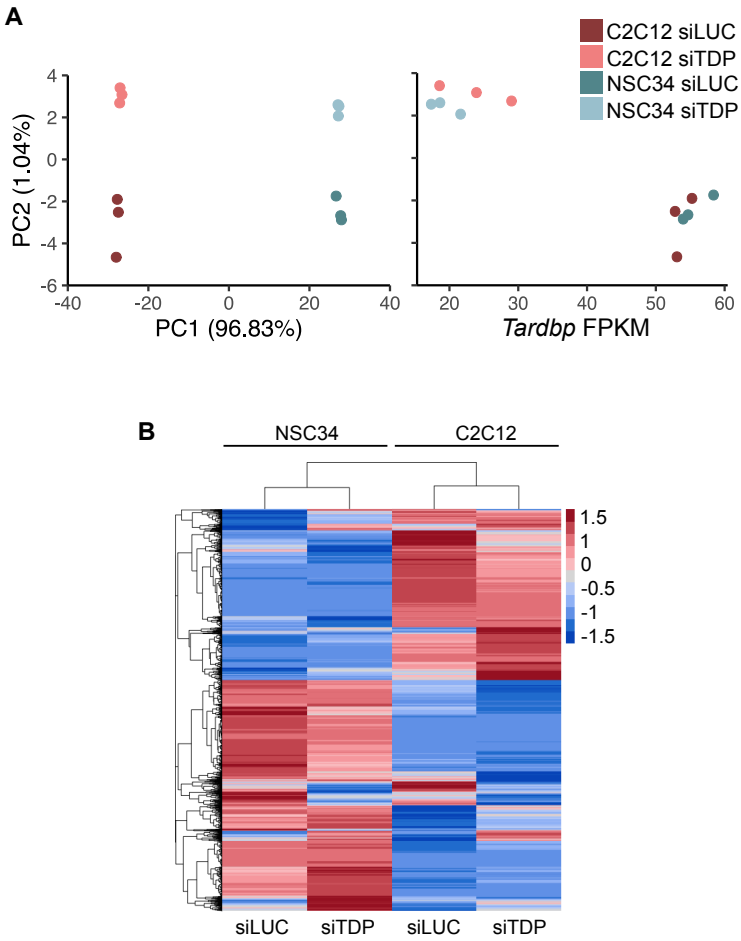


Figure 3.3: **C2C12- and NSC34-characteristic transcriptomes. A** PCA plot visualizes distances between siLUC and siTDP-transfected cells. Raw gene counts were rlog-normalized before plotting. **B** Heatmap featuring differences in gene expression across cell-types and conditions (control siLUC vs. siTDP-transfected cells). Mean FPKM values of 4019 DEGs detected in C2C12 and NSC34 were $\log_{10}$-transformed and scaled before plotting.

3.1.4 TDP-43 targets specific transcripts in C2C12 and NSC34

Considering pleinotropic functions of TDP-43, pathologic loss of this protein affects a big number of RNA-regulating processes, which in turn leads to altered transcript abundance. Initially described as a transcription factor [1], TDP-43 directly regulates transcription rate of target genes. Beyond that, it influences mRNA abundance through a number of posttranscriptional events: by regulation of alternative splicing [5] or by direct binding to mRNAs affecting their stability [154]; [8]. Furthermore, TDP-43 makes part of Drosha and Dicer complexes and is therefore involved in miRNA biogenesis, which would then impact mRNA levels in an indirect manner [155].

To investigate which mRNAs are, directly or indirectly, targeted by TDP-43 in two cell lines, I separately normalized reads of C2C12 and NSC34 datasets and obtained 4019 transcripts, expression of which is subject to TDP-43 regulation. At $p_{adj} < 0.05$, I detected a very similar number of DEGs in C2C12 and NSC34 (2325 and 2324, respectively), with 630 (15.7%) transcripts being commonly disturbed in both cell lines (Figure 3.4a).

Although the majority of bases in the genome is ubiquitously transcribed [156], only transcripts that reach a certain level of expression can be detected by sequencing. TDP-43-dependent changes in transcripts that are abundantly represented in one cell line but have a very low or negligible background expression in the other (i.e., muscle-characteristic genes, which are not transcribed in neuronal cells and *vice versa*), would not be detected by bioinformatic means. To exclude the possibility that some TDP-43-dependent transcription changes appear to be cell-type specific simply due to low abundance of those same RNAs in the other cell line and that the small overlap I see would reflect the fact we miss changes occurring in poorly represented genes, I investigated how big the overlap of TDP-43 targets is considering genes expressed in both cell lines (FPKM in both cell lines > 0.5) (Figure 3.4b). Even with this approach, the proportion of common TDP-43 targets remained rather small (19.3%).

Nonetheless, our data indicates that TDP-43 targets regulated in a cell-type-specific fashion tend to have higher expression in one cell-type than another (Figure3.4c). On average, C2C12-specific TDP-43-regulated mRNAs display higher expression in C2C12 than in NSC34 cells (the median FPKM of 11.6 and 6.6, respectively, p-value $= 1.0 \cdot 10^{-14}$), NSC34-specific targets show higher expression in NSC34 than in C2C12 (the median FPKM of 7.5 and 11.5, respectively, p-value $< 2.2 \cdot 10^{-16}$). For genes commonly regulated in both tissues, expression levels seem to be rather similar (the median FPKM of 15.0 and 16.2 in C2C12 and NSC34, respectively, $p = 0.02$).

In our datasets (Figure 3.4a), the number of downregulated genes slightly outnumbered genes that were upregulated following TDP-43 depletion (Figure 3.4e), however, the overlap was very similar irrespective the direction of the change (14.0% and 15.0% for upregulated and downregulated transcripts, respectively). Comparing the extent of expression changes of commonly regulated transcripts (630) induced by TDP-43 reduction, I saw a positive correlation ($\rho = 0.77$, p-value $< 2.2 \cdot 10^{-16}$) between two cell lines, with a trend towards larger alternations in C2C12 (Figure 3.4d). Of note, there were few mRNAs whose expression was altered in the opposite direction, indicating that TDP-43 loss can elicit contrary effects depending on the cellular environment.

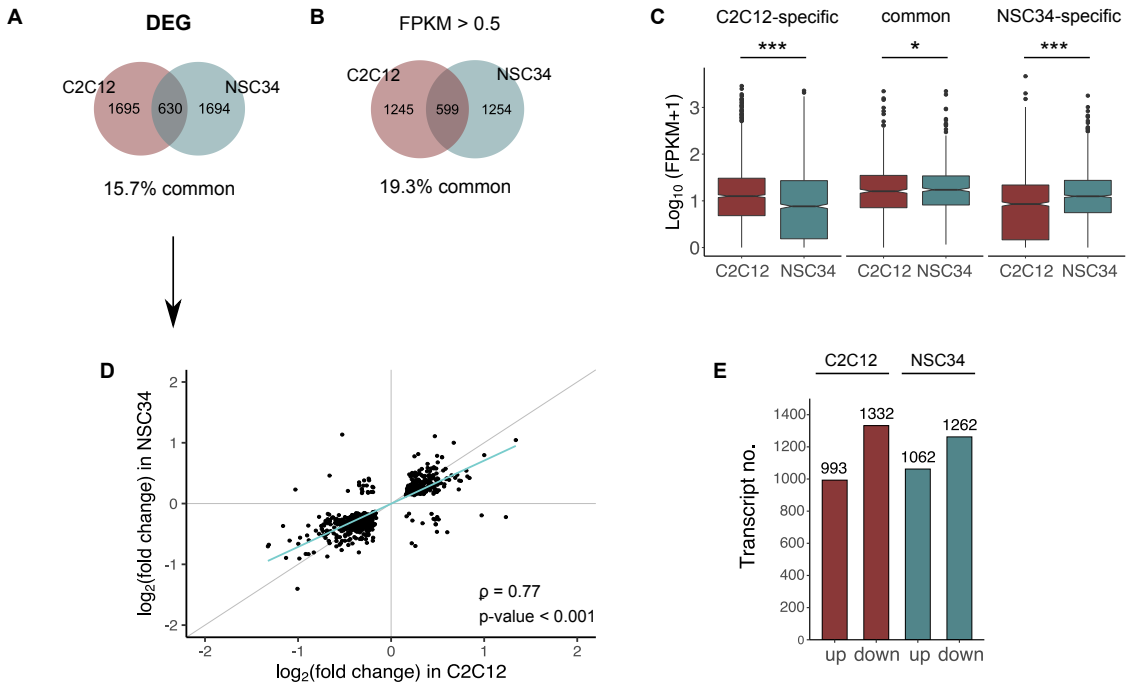


Figure 3.4: **Differentially expressed genes in C2C12 and NSC34. A** Venn diagram shows the number of TDP-43-regulated transcripts identified in C2C12 and NSC34 cells exclusively (1695 and 1694, respectively), along with those that are commonly regulated by TDP-43 in both cell-types (630). **B** Venn diagram shows the overlap (599 transcripts, 19.3%) of TDP-43-regulated DEGs identified in C2C12 and NSC34 cell line (as in **A**) in a subset of transcripts expressed in both cell lines (FPKM in both cell lines > 0.5). **A** and **B** Transcripts with $p_{adj} < 0.05$ were considered as differentially expressed irrespective their $\log_2$ fold change. **C** Box plot shows $\log_{10}$-transformed FPKM values of muscle, neuronal and common TDP-43 targets in siLUC-transfected C2C12 and NSC34 cells. C2C12-specific DEGs exhibit higher expression in C2C12 cells (p-value $< 2.2{\cdot}10^{-16}$), while NSC34-specific DEGs have higher expression in NSC34 (p-value $< 2.2{\cdot}10^{-16}$). Expression levels of common targets is more similar between cell lines (p-value $= 0.02$). Significance was tested using Wilcoxon signed-rank test. **D** Transcription changes of common targets (**A**, 630) are plotted by their $\log_2$ fold change in C2C12 and NSC34 (Spearman's $\rho = 0.77$, p-value $< 2.2{\cdot}10^{-16}$). Grey line represents y = x and the blue line represents the fitted regression. **E** Diagram shows the number of upregulated and downregulated genes detected in C2C12 and NSC34 cells following TDP-43 depletion.

Polymenidou *et al.* [8] detected TDP-43 binding sites within 6,304 annotated protein-coding genes, which makes approximately 30% of the mouse transcriptome. They proposed that TDP-43 binding is crucial to sustain pre-mRNA levels and that mRNA downregulation occurs as a direct consequence of TDP-43 loss, while mRNA upregulation is attributed to indirect effects. In neurons, TDP-43 binding sites were found in protein-coding genes with particularly long introns [8]. Furthermore, Cortese *et al.* [102] additionally demonstrated that transcripts with

long introns are significantly downregulated in sIBM.

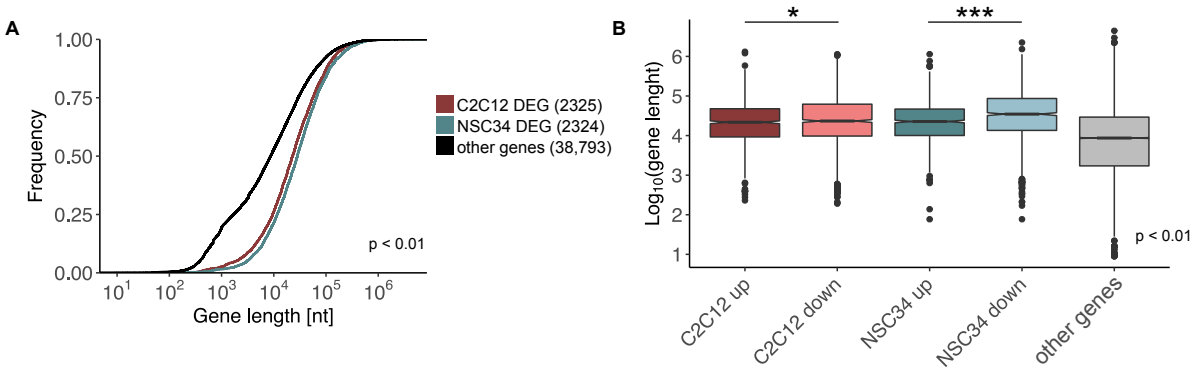


Figure 3.5: **Gene length of TDP-43 targets in C2C12 and NSC34. A** Cumulative distribution plot compares gene lengths of TDP-43-targeted genes (2325 and 2324 in C2C12 and NSC34, respectively) with other genes in mouse genome (38,793). TDP-43-controlled genes are significantly longer (p-value $< 2.2{\cdot}10^{-16}$) than other genes of mouse genome, while DEGs detected in NSC34 appear to be longer than those regulated by TDP-43 in C2C12 (p-value $= 2.1{\cdot}10^{-7}$). **B** Boxplot shows that among TDP-43 targets identified in each cell line, genes downregulated upon TDP-43 knockdown in NSC34 cell line are longer than those with increased expression (p-value $< 2.2{\cdot}10^{-16}$). The difference between the length of upregulated and downregulated genes found in C2C12 is smaller (albeit significant, p-value $=$ 0.01). The difference among all groups was tested with Kruskal-Wallis rank sum test (p-value $< 2.2{\cdot}10^{-16}$), followed by pairwise comparisons using Wilcoxon rank sum test.

Herein, I show that genes targeted by TDP-43 in C2C12 and NSC34 are significantly longer (p-value $< 2.2{\cdot}10^{-16}$) than an average gene in mouse genome (Figure 3.5a). What is more, TDP-43-controlled genes in NSC34 cells appear to be longer than TDP-43 targets in C2C12 (p-value $= 2.1{\cdot}10^{-7}$). The latter did not come as a surprise, for it has been previously demonstrated that genes preferentially expressed in the brain contain longer introns relative to genes transcribed in other tissues [8]. In line with what has been shown by Polymenidou *et al.* [8], downregulated genes appear to be twice longer than upregulated genes in NSC34 (median length of 2,244,856 and 1,139,157 bp for downregulated and upregulated genes, respectively, p-value $< 2.2{\cdot}10^{-16}$). However, the same does not hold true for C2C12 cells, in which downregulated genes are in fact slightly shorter (median length of 1,135,405 and 1,313,497 for downregulated and upregulated genes, respectively, p-value $=$ 0.01). Long genes are expected to contain exceptionally long introns (but not exons) which hold important regulatory elements including those bound by TDP-43 [8]. Why downregulated genes appear to be longer than upregulated genes in NSC34 but not in C2C12 while a similar number of TDP-43 targets was identified in both cell lines, can only be speculated. Possibly, TDP-43-dependent stabilization of mRNAs represents an important regulatory mechanism in neuronal NSC34 cells, which express longer genes containing long introns with multiple TDP-43 biding sites, whereas in muscle-derived C2C12 cells, TDP-43

predominantly acts through processes other than mRNA stabilization.

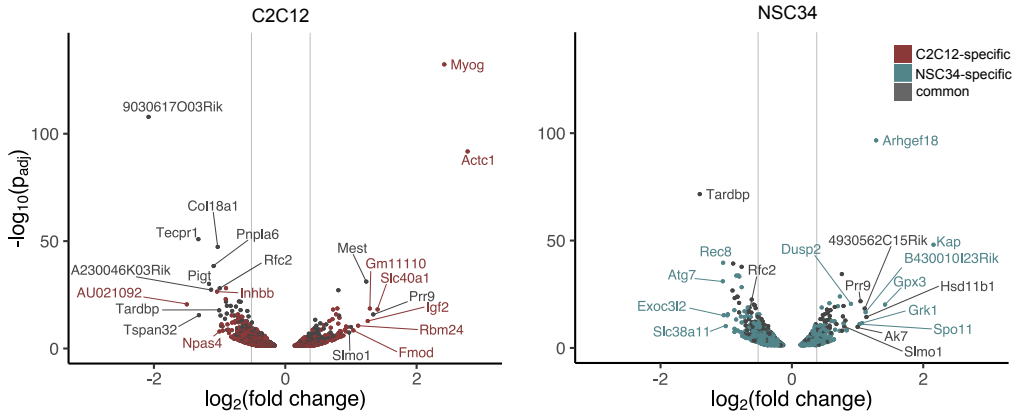


Figure 3.6: **TDP-43-dependent DEGs found in C2C12 and NSC34.** TDP-43-mediated transcription changes in C2C12 and NSC34 represented as volcano plots. C2C12- and NSC34-specific targets are shown in red and blue, respectively, while commonly disturbed transcripts are plotted as grey dots. Vertical lines indicate fold changes of 0.7 (30% increase) and 1.3 (30% decrease). Best hits are labeled with gene names.

Looking at individual targets (Figure 3.6), I hypothesized that the biggest transcriptional changes induced by TDP-43 knockdown occurred in highly expressed genes. Plotting the size of the change ($\log_2$ fold change) against background expression levels (FPKM in siLUC-transfected cells) of all DEGs, however, revealed that there is in fact no correlation between the two (Figure 3.7).

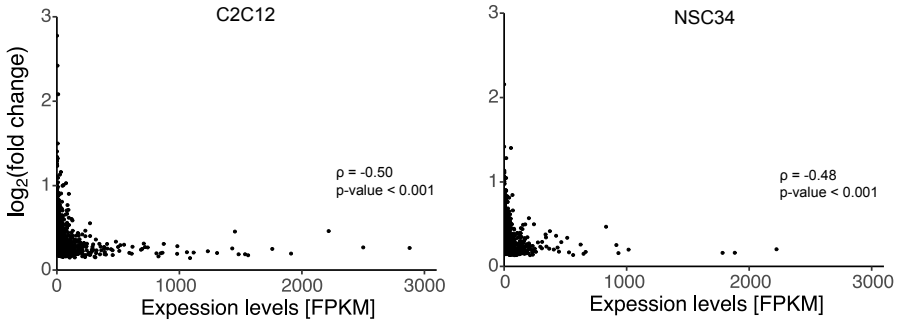


Figure 3.7: **Correlation of expression levels with TDP-43-dependent transcription change.** Scatter plots show there is no correlation between the absolute change in gene expression following TDP-43 knockdown (plotted as $\log_2$-transformed fold change) and the baseline expression of a given transcript (FPKM in siLUC-transfected cells) for DEGs identified in C2C12 (2,325) and NSC34 (2,324) (Spearman's ρ = -0.50, p-value < $2.2{\cdot}10^{-16}$ and Spearman's ρ = -0.48, p-value < $2.2{\cdot}10^{-16}$, respectively).

Even among genes undergoing the largest expression changes, some transcripts appear to be regulated by TDP-43 in a cell-type-specific fashion (Figure 3.6 and 3.8). *Myog* and *Actc1* are two exemplary transcripts that increase upon TDP-43 loss in C2C12 but not in NSC34. *Myog* encodes for myogenin, which acts as a transcriptional activator that promotes transcription of muscle-specific genes and plays a role in muscle differentiation; and *Actc1* encodes for α-actin, which is a muscle-specific actin isoform that builds the contractile apparatus. Since both proteins provide essential muscle functions, they are not be expected to be of high importance for neurons. In fact, expression of *Myog* in NSC34 is so low it could not be detected by qPCR (Figure 3.8b).

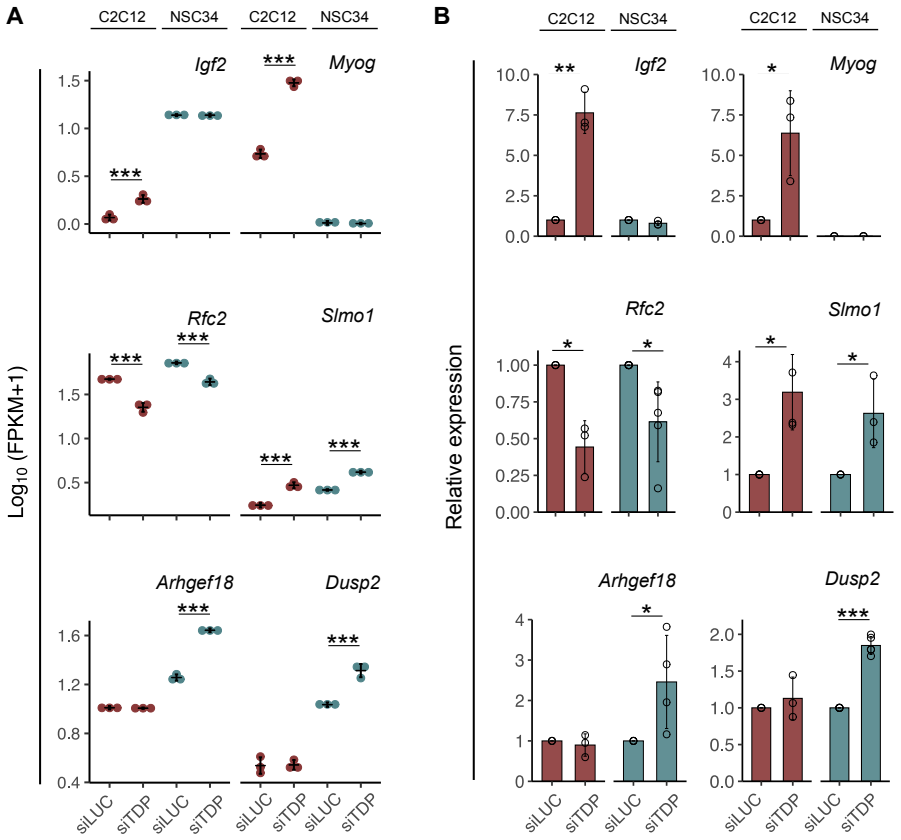


Figure 3.8: **qPCR validation of transcription changes detected by RNA-seq. A** Expression changes of representative DEGs (C2C12-specific vs. common vs. NSC34-specific) as assessed by RNA-seq and plotted as $\log_{10}$-transformed FPKM. p-values were generated using Wald test and Bejamini-Hochberg multiple testing correction [119]. **B** Relative expression changes of DEGs from **A** were validated using qPCR. p-values were generated using Student's t test (paired, one-tailed). * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001.

Actc1, on the other hand, displays a rather similar transcription levels in siLUC-transected C2C12 and NSC34 cells (the mean expression of 0.72 and 0.14 in C2C12 and NSC34, respec-

tively, p-value = 0.03), yet, its expression increases following TDP-43 depletion in C2C12 but not in NSC34 (13.3-fold increase, p-value $< 2.2 \cdot 10^{-16}$) (not shown here). Similarly, *Igf2* undergoes TDP-43-dependent increase in C2C12 cells exclusively, although its levels are higher in NSC34 cells. *Igf2* that encodes for insulin-like growth factor II is required for development and maintenance of the musculoskeletal system and it promotes differentiation of C2C12 myoblasts [157].

Rfc2 and *Slmo1*, two genes commonly regulated by TDP-43, seem to execute more general functions; *Rfc2* (Replication Factor C Subunit 2) encodes a subunit of a replication factor needed during DNA elongation [158] and *Slmo1* (Slowmo Homolog 1) mediates transfer of phosphatidic acid between liposomes [159].

Expression of *Arhgef18* (Rho/Rac Guanine Nucleotide Exchange Factor 18) is controlled by TDP-43 in a NSC34-specific manner. While functions of ARHGEF18 in muscles have not been described so far, it was recently reported to regulate axon branching in cortical neurons [160].

Collectively, our results support the idea that unique sets of transcripts controlled by TDP-43 in each cell-type could be, only to a minor extent, explained by variable expression levels of cell-type-characteristic mRNAs across tissues. Factors other than transcript abundance as such would underly TDP-43's activity and render it cell-type-specific.

3.1.5 TDP-43-dependent mRNA splicing is cell-type-specific

mRNA depletion and splicing changes that arise as a consequence of pathologic TDP-43 behavior are both believed to contribute to neuronal vulnerability [8]; [7]. Previous studies have investigated TDP-43-mediated regulation of alternative splicing in neuronal tissues [8]; [57]; [7]; [99]; [161] or cell lines [162]; [7]. Yet, little is understood about how TDP-43 dysfunction affects pre-mRNA splicing in tissues beyond the central nervous system.

In this study I systematically compared splicing changes that occur following TDP-43 knockdown in the skeletal muscle and neuronal cell line using two independent splicing tools - rMATS and MAJIQ (see the Methods section). Both of these are event-based methods, however, they differ in what they detect as a single splicing event. While rMATS identifies alternative sequences along with their flanking exons and is thus able to distinguish between various event types, MAJIQ seeks for alternatively used junctions, which are considered as individual splicing events. Although rMATS was our primary tool of choice for its ability to categorize splicing events and was used to generate most of the results discussed herein (if not otherwise stated), I additionally conducted splicing analysis by MAJIQ in order to obtain an independent result.

Using rMATS, a considerably higher number of AS events was detected in NSC34 than in C2C12 cells (1,270 and 730, respectively) (Figure 3.9a). As at this point, was interested in global patterns of splicing disturbance that emerge upon TDP-43 loss in each cell-type rather than in finding specific AS events, I set a low stringency threshold and considered all splicing events identified in each dataset regardless the size of the change (ΔPSI different from 0, FDR

< 0.01). However, it needs to be kept in mind that (by not filtering events by the effect size) we are likely to include events, which have little or no biological importance. Considering all significant changes, only a small proportion of AS events (on average 5.2%, which makes 11.4% of alternatively spliced genes) appeared to be commonly regulated by TDP-43 in C2C12 and NSC34 (Figure 3.9a).

Five splicing categories were considered as a part of the rMATS splicing protocol (i.e., SE, MXE, RI, A3SS, A5SS) [59]. Alternative inclusion of cassette exons (hereafter referred to as SE for skipped exons, although cassette exons which display enhanced inclusion fall into this category regardless the direction of change) were the most frequent event detected by rMATS, followed by mutually exclusive exons, intron retention, alternative 3' and 5' splice site selection (Figure 3.9b and 3.13a). The percentage of overlapping events (i.e., commonly detected in both datasets) was small irrespective of the AS category.

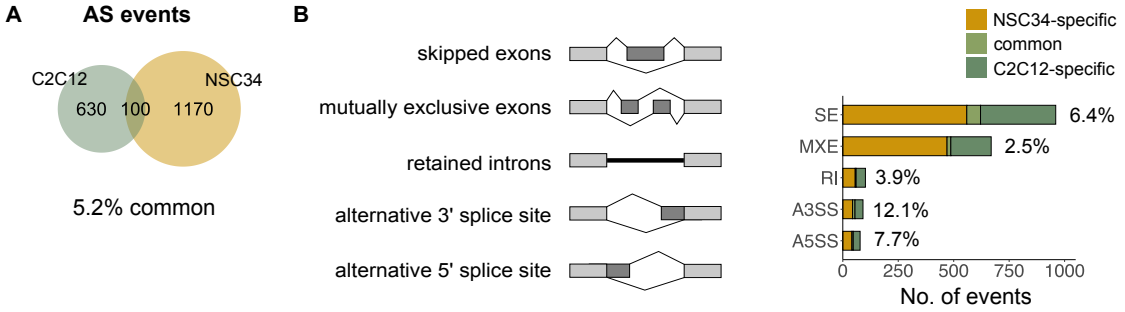


Figure 3.9: **TDP-43-regulated AS events detected by rMATS. A** Venn diagram shows the total number of AS events (detected by rMATS at FDR < 0.01) induced by TDP-43 depletion in C2C12 and NSC34 specifically (630 and 1,170, respectively), together with those commonly identified in both cell lines (100). **B** The number of annotated AS events (from **A**) by event type. SE - exon skipping, MXE - mutually exclusive exons, RI - intron retention, A3SS and A5SS - alternative 3' or 5' splice site. The percentage of overlapping AS events is reported on the plot.

Even more than for differential expression analysis, distinct transcriptomes (Figure 3.3) represent an obstacle when it comes to detection and quantification of splicing changes. Tissue-characteristic transcriptome is usually dominated by a low number of highly-expressed genes. Upon sequencing, more reads would therefore align to alternatively spliced regions of genes with high expression. Since the estimation uncertainty of isoform proportion depends on the sequencing coverage of individual splicing event [59], it is more likely that AS in lowly-expressed genes will not be detected by the statistical means (the probability of false negatives with high p-values). I sought to overcome this (putative) bias using different approaches. First, I only considered AS events that occur in transcripts commonly expressed in both C2C12 and NSC34 (FPKM > 0.5 in both cell lines) and to exclude events within mRNAs that display very low expression in either cell line. Even by doing so, the percentage of overlapping events remained

small (5.8%) (Figure 3.10a). I next assessed the overlap of TDP-43-regulated splicing including all events that showed significant alteration upon TDP-43 silencing in one cell line (FDR $<$ 0.01), while the change was detected at less string threshold (if FDR $>$ 0.01 but p-value $<$ 0.05) in the other one, and *vice versa* (Figure 3.10b) [163]. Using this approach I obtained a higher total number of AS events that entered the analysis along with an increase in the overlap (10.7%). Nonetheless, the proportion of commonly regulated AS events was still very small, suggesting that cell-type-specific activity of TDP-43 I detected could not be sufficiently explained by variable expression levels of those mRNAs.

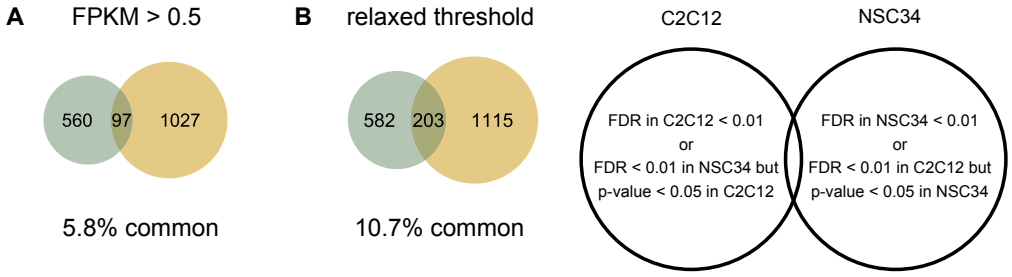


Figure 3.10: **The overlap of TDP-43-regulated splicing found by rMATS.** **A** Venn diagram shows the number of AS events detected by rMATS (as in Figure 3.9a) in C2C12 and NSC34, considering only events that occur within transcripts expressed in both cell lines (FPKM in both cell lines $>$ 0.5). **B** Venn diagram showing the number of AS events detected at relaxed threshold (considering events that were detected at FDR $>$ 0.01 in one dataset and p-value $<$ 0.05 in the other) as explained at the scheme.

Compared to specific TDP-43 targets, commonly spliced transcripts on average showed higher expression in C2C12 and NSC34 cells than those alternatively spliced in a cell-type-specific manner (median expression of of 18.8 and 10.3 FPKM, respectively, p-value $< 2.2{\cdot}10^{-16}$, Figure 3.11a). Furthermore, overlapping events displayed bigger splicing transitions (ΔPSI) (median expression of of 16.8 and 12.4%, respectively, p-value $= 1.0{\cdot}10^{-7}$, Figure 3.11b). We therefore believe that differences between commonly-detected and cell-type-specific events simply reflect the fact that TDP-43-controlled AS events undergoing bigger changes occurring within highly expressed transcripts are more likely to be detected by bioinformatic tools in both cell lines. This is additionaly supported by the observation that for a subset of commonly detected events (100), most of splicing changes occurred in the same direction in both cell lines (83%, $\rho = 0.62$, p-value $= 4.1{\cdot}10^{-12}$, Figure 3.11c), suggesting that TDP-43 exerts a similar function in cells of neuronal and muscle background.

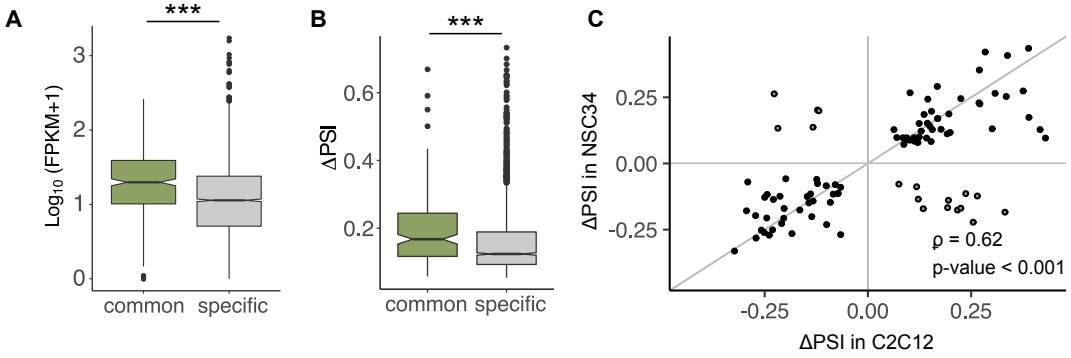


Figure 3.11: **Characteristics of commonly detected AS events. A** Expression levels of transcripts that are commonly spliced in both cell lines (164) or in one cell line exclusively (1268) are plotted as $\log_{10}$-transformed FPKM values (p-value < $2.2{\cdot}10^{-16}$). **B** Absolute changes (ΔPSI) of overlapping splicing events (100) (from Figure 3.10a) compared to those uniquely detected in C2C12 or NSC34 (1800) (p-value = $1.0{\cdot}10^{-7}$). p-values for **A** and **B** were generated by unpaired Wilcoxon rank sum test. **C** The correlation of splicing changes for commonly detected splicing events (100) plotted as ΔPSI in C2C12 and NSC34 (Spearman's correlation coefficient $\rho = 0.62$, p-value = $4.1{\cdot}10^{-12}$).

Splicing analysis was additionally conducted by an independent splicing tool - MAJIQ (ΔPSI > 0.2, FDR < 0.1) (Figure 3.12a). Like with rMATS, a higher number of TDP-43-dependent splicing events was detected in NSC34 cells (305 and 394 in C2C12 and NSC34, respectively). MAJIQ and rMATS quantify in separate ways (junctions or events, respectively) [65], thus comparable results are only produced when the same pipeline is applied. The proportion of commonly regulated annotated splicing events obtained by MAJIQ relative to rMATS (15.5% and 5.2%, respectively) clearly points out to the different performance of two event-based methods on the same dataset [65].

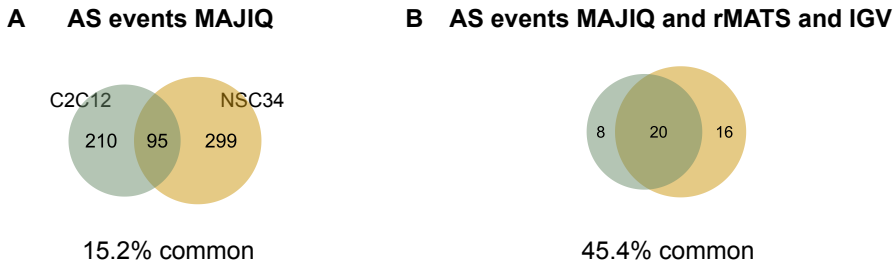


Figure 3.12: **The overlap of TDP-43-regulated splicing detected by MAJIQ. A** Venn diagram shows the number of AS events (junctions) detected by MAJIQ (ΔPSI > 0.2, FDR < 0.1). **B** Venn diagram showing the number of AS events detected by MAJIQ (in **A**) and rMATS (Figure 3.9a), that were additionally validated using Interactive Genome Viewer.

To answer the question how cell-type-specific the activity of TDP-43 really is, we, as our

final approach, estimated the overlap of TDP-43-controlled splicing between C2C12 and NSC34 cell line on a smaller subset of AS events, all of which were detected by both splicing tools (even if at expense of false negatives). Moreover, I visually validated every event found by splicing tools using Interactive Genome Viewer and subsequently excluded those whose changes could not be confirmed (Figure 3.12b).

Among these events, 45% of TDP-43-regulated splicing activity seemed to be shared between NSC34 and C2C12 and I firmly believe this provides a better estimate of TDP-43's cell-type-specificity.

Similar characteristics of TDP-43-regulated events regardless the cell-type

Given a higher number of TDP-43-mediated events detected in neuronal cell line (Figure 3.9a and 3.12a), further explored whether TDP-43 targets in C2C12 and NSC34 would display distinct features. Generally, AS events identified in each cell line did not differ with regards to event type distribution (Figure 3.13a) or length of cassette exons (i.e., cassette exons of SE and MXE, Figure 3.13b). As TDP-43 is known to act both as splicing repressor or enhancer depending on its cross-linking position in a given transcript [7], its depletion would either result in inclusion or exclusion of a target exon. Looking at the ratio of inclusion/skipping events, that was not different between C2C12 and NSC34 targets (Figure 3.13c) neither was the frequency of frame-conserving events (Figure 3.13d). Interestingly enough, alternative sequences regulated by TDP-43 in neuronal cell line seem to be better conserved across species than those in the muscle cell line, which will be further discussed in a later chapter.

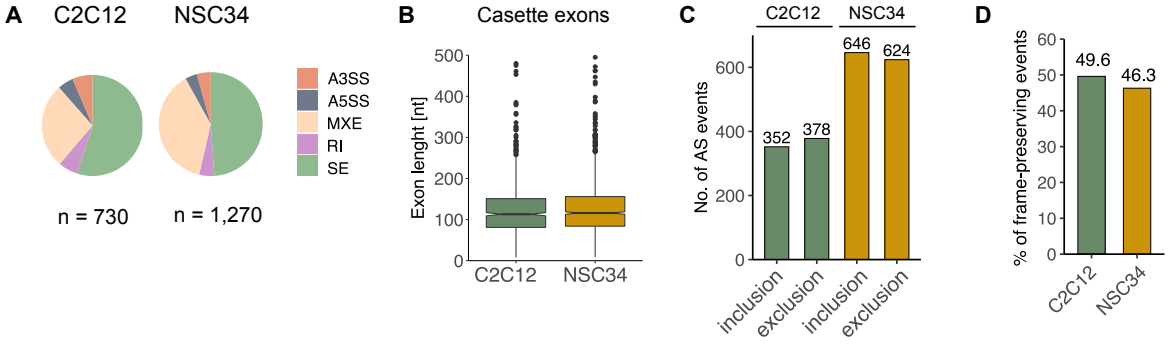


Figure 3.13: **General features of TDP-43-controlled AS events. A** TDP-43-regulated AS events detected in C2C12 and NSC34 cells do not differ in terms of event type distribution (the number below shows the total number of AS events detected in each cell line); **B** the average length of TDP-43-regulated cassette exons (SE and MXE); **C** the ratio between inclusion/exclusion events; **D** the percentage of frame-conserving events.

Common splicing events found in C2C12 and NSC34

Previous studies have already disclosed lists of transcripts, whose splicing is affected by TDP-43 removal or dysfunction [142]; [164]; [165]; [7]. Yet, the reproducibility of target identification is rather poor, most likely due to differences in methodological approaches, low conservation of TDP-43 targets across species [142]; [166]; [164], and, as I have shown, the unique function TDP-43 elicits in each tissue or cell-type. The most consistently reported TDP-43-regulated splicing event across studies and conditions is skipping of exon 3 within *Poldip3*/*POLDIP3* mRNA (both mouse and human) [151]. This being so, inclusion levels (ΔPSI) of *Poldip3* exon 3 often serve as a readout of TDP-43 functionality [167]; [168]; [169]. In search of new splicing events that would, similarly to *Poldip3*/*POLDIP3*, show high reproducibility across experimental settings, I chose mRNAs that underwent the biggest shift in TDP-43-controlled exon inclusion and whose isoform proportion was altered in both C2C12 and NSC34 cells. TDP-43-dependent isoform switch was validated using isoform-sensitive semi-quantitative RT-PCR (Figure 3.14a). In few cases (*Dnajc5*, *Fam220a*), splicing alterations co-occurred with a change in transcript abundance (Figure 3.14b). Herein I describe some of splicing events commonly regulated by TDP-43 in both cell lines.

Skipping of exon 3 (137 bp) within *Rgp1* (RAB6A GEF Complex Partner 1), which occurs upon TDP-43 knockdown is a novel splicing event that has not been previously described since exon 3 seems to be constitutively spliced-in (VastDB [170]). Exon 3 shows high sequence similarity (94%) with its human orthologue and a putative TDP-43 binding site (UG repeats) appear downstream the 5' splice site.

Another target transcript, *Dnajc5* (DnaJ heat shock protein family member C5), encodes for a protein that plays a role in protein folding and membrane trafficking, and has been reported to have anti-neurodegenerative properties [171]. Inclusion of exon 5 (75 bp) leads to ORF disruption and causes NMD in mouse and human (VastDB [170]). TDP-43 depletion enhances exon 5 incorporation, which is accompanied by transcript reduction in C2C12 and NSC34 (24% and 12% decrease, respectively). Multiple UG repeats are found downstream and upstream this alternative exon suggesting TDP-43 binding.

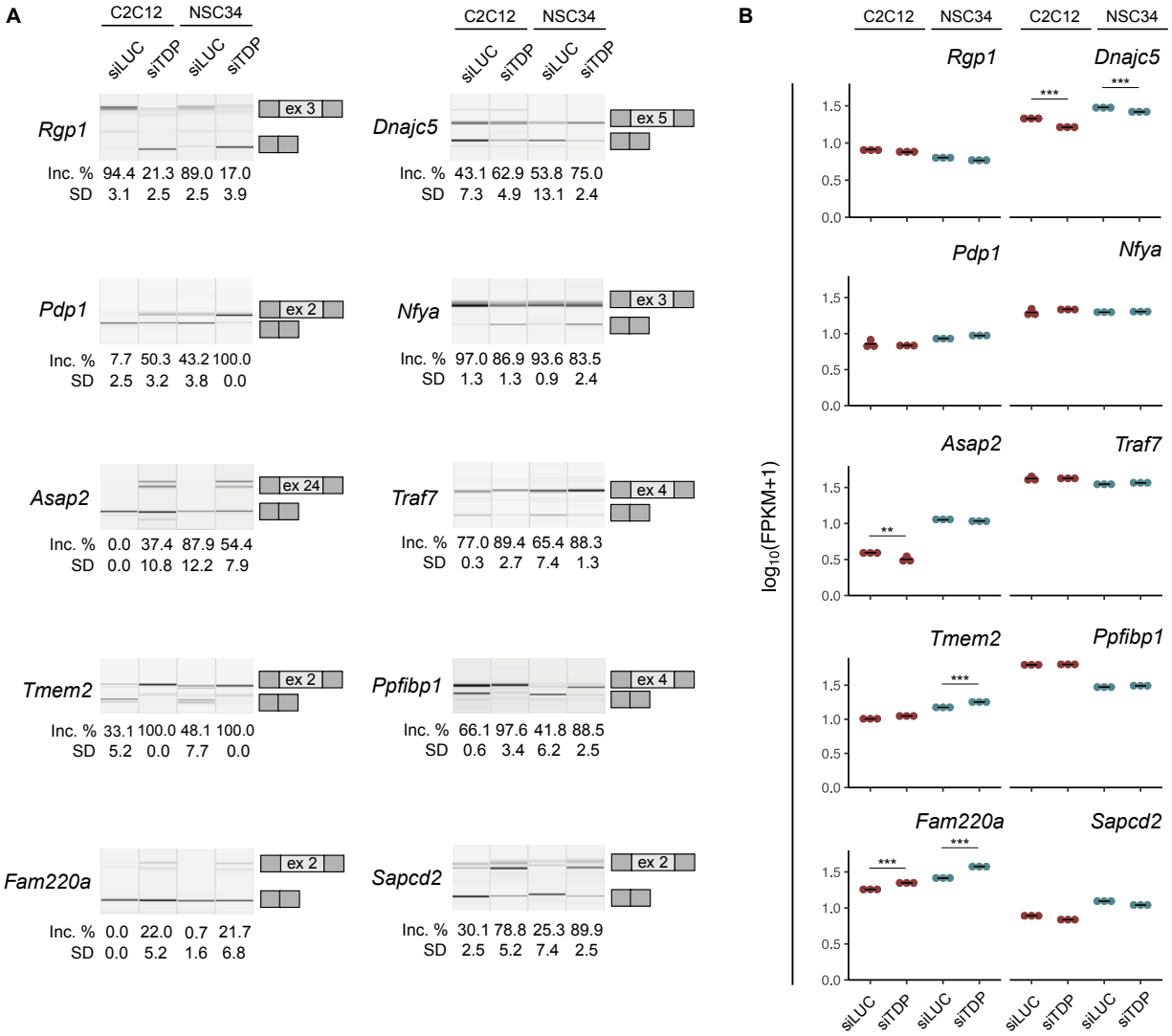


Figure 3.14: **Common TDP-43-dependent splicing events. A** Validation of TDP-43 dependent splicing of 10 representative mRNA targets. Semi quantitative RT-PCR conducted in TDP-43-silenced samples and corresponding controls is shown along with the quantification of splicing changes (% of alternative exon inclusion). The number of the alternative exon is shown in the scheme (see the exact transcript numbers in Table 2.3). **B** Expression changes of AS transcripts (shown in **A**) as assessed by RNA-seq and plotted as $\log_{10}$-transformed FPKM. p-values were generated Kruskal-Wallis test. * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001.

Pdp1 (pyruvate dehyrogenase phosphatase catalytic subunit 1) encodes for one of three components of the pyruvate dehydrogenase complex that participates in glucose metabolism. Exon 2 (35 bp) displays higher inclusion levels in neurons than in muscles (VastDB [170]), however, its incorporation increases upon TDP-43 reduction. Of note, enhanced inclusion of exon 2 was detected in individuals with myotonic dystrophy type 2 [172].

We further validated TDP-43-regulated inclusion of exon 3 (81 bp) within *Nfya* (nuclear transcription factor Y subunit α), which shows high inclusion level in mature *Nfya* mRNA.

Another commonly indetified TDP-43 target is *Asap2* (Arf-GAP with SH3 domain, ANK repeat and PH domain containing protein 2). Exon 2 (135 bp) of *Asap2* has low incorporation levels in muscle cells ($< 20\%$) and some neuronal populations in mouse (VastDB [170]). Its inclusion increases upon TDP-43 silencing in C2C12 and NSC34. As multiple UG repeats were found downstream the 3'SS, the latter might be indicative of TDP-43 binding. Expression of *Asap2* seems to decrease upon TDP-43 depletion, however, it is unclear whether downregulation is associated with this particular TDP-43-controlled splicing event or not.

Exon 4 (154 bp) within *Traf7* (TNF receptor associated factor 7), which shows enhanced inclusion upon TDP-43 loss, has variable extent of inclusion across mouse tissues (0-90%). This splicing event was associated with cell differentiation (VastDB [170]).

Another commonly detected skipping event occurs within *Tmem2* mRNA (transmembrane protein 2). Inclusion of exon 2 (82 bp) leads to frame-shifting, yet it does not introduce a premature stop codon. *Tmem2* encodes for cell surface hyaluronidase that mediates degradation of extracellular high-molecular-weight hyaluronan and therefore regulates angiogenesis and morphogenesis. TDP-43 knockdown results in upregulation (22%) of *Tmem2* in NSC34 cells but not in C2C12. Increased expression has been previously observed following TDP-43 reduction in rat primary neurons [173]. Nevertheless, transcription changes are not necessarily associated with this splicing event.

Furthermore, TDP-43 depletion also led to inclusion of exon 4 (33 bp) within *Ppfibp1* (PPFIA binding protein 1) transcript. Exon 4 displays a variable extent of inclusion across mouse tissues (0-80%) and its incorporation increases with differentiation in muscle and neurons (VastDB [170]). TDP-43 seems to induce skipping of this exon by binding to an upstream UG-rich cluster.

Fam220a is an example of a transcript that undergoes TDP-43-dependent splicing but also increases in abundance upon TDP-43 reduction in C2C12 and NSC34 (23 and 46% increase, respectively). Exon 2 (133 bp) incorporation in fact causes frame-shifting and would be expected to induce NMD. A long strech of UG repeats found in the upstream intron indicates a putative TDP-43 biding site.

Next, we also validated the increase in exon 2 (102 bp) inclusion within *Sapcd2* which occurs as a consequence of TDP-43 silencing.

Importantly, all above-mentioned events were detected in our dataset by both, rMATS and MAJIQ. The assumption that these alternative exons are directly regulated by TDP-43 is supported by the observation that multiple UG repeats or long UG stretches were found in the proximity of all alternative exons (100 bp upstream or downstream). We validated both, inclusion and exclusion events associated with TDP-43 loss. Surprisingly enough, we found an event that induces frame-shifting (within *Fam220a*) and is predicted to cause NMD, however,

the overall transcript levels do not behave as expected.

Cell-type-specific TDP-43-controlled events found in C2C12 and NSC34

While commonly detected TDP-43-dependent splicing events could represent a much needed readout of TDP-43 functionality, finding events that are uniquely controlled by TDP-43 in a given cell population would help us understand how cell-type-specific environments shape TDP-43's function.

In a previous chapter we described different approached to assess the extent of cell-type-characteristic splicing activity of TDP-43. Comparing TDP-43-regulated splicing in two different cell-types we believe the proportion of unique TDP-43's targets is in fact smaller than it was detected by currently available splicing tools (Figure 3.12b). The reasons are as they follow:

- Some splicing events might be detected as significant in one cell line (i.e., in one dataset) but not in another, either due to low expression (e.g., low read coverage of the alternative exon) or due to high variation between samples of one cell line (see the example of *Slmo1*; significance was tested using Fisher's exact test and corrected for multiple testing). Thereby, alternative splicing of *Slmo1* was detected by rMATS in C2C12 but not in NSC34 cell line, although TDP-43-dependent isoform switch can be seen in both cell lines when splice isforms were validated using semi-quantitative PCR (Figure 3.15a).

- Due to differential exon usage across cell-types, the same spicing event (i.e., the same alternative exon as exemplified by *Tesk2*) could be detected as two different AS events, as a given alternative exon is flanked by different neighboring exons in each cell line (Figure 3.15b).

- Some genes might be expressed in one cell line only (Figure 3.15c).

Nonetheless, we were still able to validate some AS events that are indeed regulated by TDP-43 in a cell-type-specific fashion (Figure 3.16). *Tnik*, *Tbc1d1*, *Lrp8* and *Mef2d* are examples of transcripts that undergo TDP-43-dependent splicing in undifferentiated C2C12 but not in NSC34 (Figure 3.16a). On the other hand, splicing of *Dync1I2* and *Ubr4* appears to be subjected to TDP-43 control only in NSC34 cell line. Importantly, these AS events were found within transcripts expressed in both cell lines.

In mouse, exon 14 of *Tnik* mRNA (TRAF2 and NCK interacting kinase) shows high inclusion in mature muscles but not in other tissues (VastDB [170]). In undifferentiated C2C12, however, incorporation of exon 14 increases upon TDP-43 reduction.

TDP-43-dependent inclusion of two alternative exons within *Tbc1d1* that we observed in undifferentiated C2C12 cells will be discussed in a later chapter.

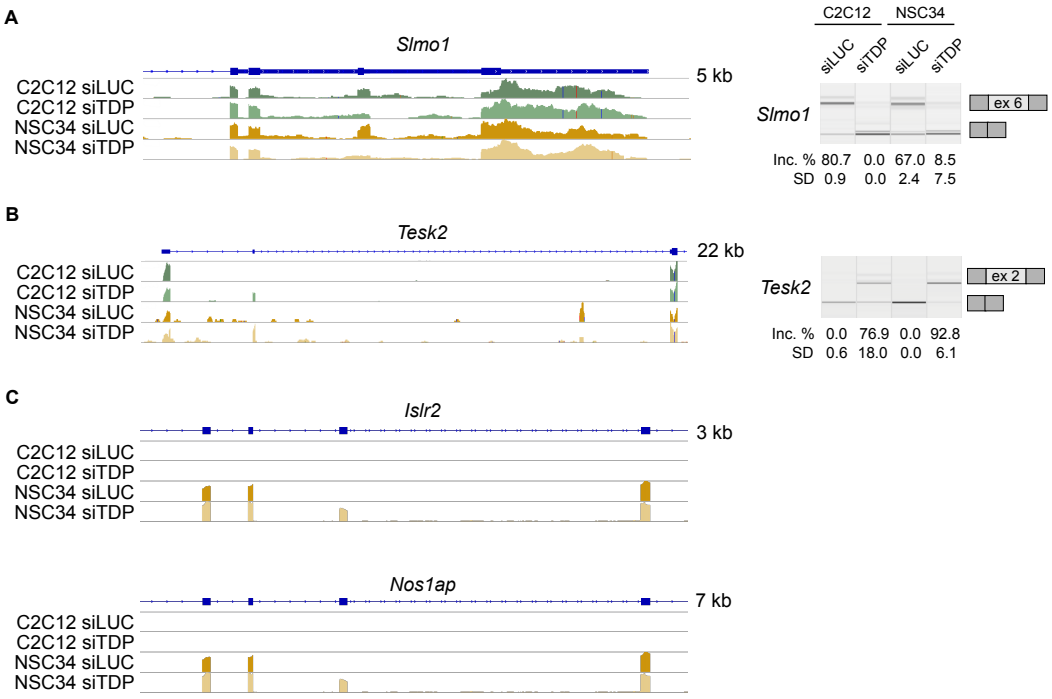


Figure 3.15: **Cell-type-specific detection of AS events by splicing tools. A** TDP-43-dependent splicing of *Slmo1* was only detected in C2C12 dataset (due to higher variation in the NSC34 dataset), while isoform switch was demonstrated in both cell-types using PCR. **B** TDP-43-dependent splicing of *Tesk2* was detected as two different TDP-43-controlled events due to different exon usage (of neighboring exons) in C2C12 and NSC34. Results obtained by PCR demonstrate TDP-43-controlled inclusion of exon 2 in C2C12 and NSC34. **C** Two exemplary transcripts whose expression is NSC34-specific. TDP-43 controlled inclusion of alternative exons could not be detected in C2C12 dataset, as these genes are only expressed in neuronal cells. **A**-**C** Coverage track from Interactive genome viewer (IGV) is shown on the left. **A** and **B** Semi quantitative RT-PCR conducted in TDP-43-silenced samples and corresponding controls is shown along with the quantification of splicing changes (% of alternative exon inclusion). The number of the alternative exon is shown in the scheme (see the exact transcript numbers in Table 2.3).

Two cassette exons - exon 6a (123 bp) and exon 6b (39 bp) - of *Lrp8* (low-density lipoprotein receptor-related protein 8) appear to be constitutively included across different mouse tissues (VastDB [170]), however, in undifferentiated C2C12 and NSC34 they do not make part of the mature *Lrp8* transcript. Incorporation of both exons increases upon TDP-43 reduction in a cell-type-specific manner. Exon 6a was reported to encode for an experimentally validated eukaryotic linear motif (ELM) (VastDB [170]), while exon 6b contains furin cleavage site [174]. *In vivo* phenotype exerted by each splice isoform remains unexplained.

Mef2d encodes for a myocyte enhancer factor 2D, which is a family member of myocyte-specific transcription factors. MEF2D was reported to regulate differentiation of skeletal muscle and neuronal cells. Exon 8 (21 bp), which is regulated by TDP-43 in C2C12 cells but not in NSC34, shows high level of inclusion in mature muscle and neurnal tissues and its inclusion

indeed increases upon cell differentiation. Exon was reported to encode for a *trans*-activation domain [175].

Dync1i2 (dynein cytoplasmic 1 intermediate chain 2) encodes for protein that acts as a motor that allows retrograde transport of organelles and vesicles. Exon 5 (60 bp) that is subject to TDP-43-dependent inclusion in NSC34 cells, is not included in mature muscles. Incorporation of region encoded exon 5 increases the interaction of DYNC1I2 with PAFAH1B1 and DYNC1I1.

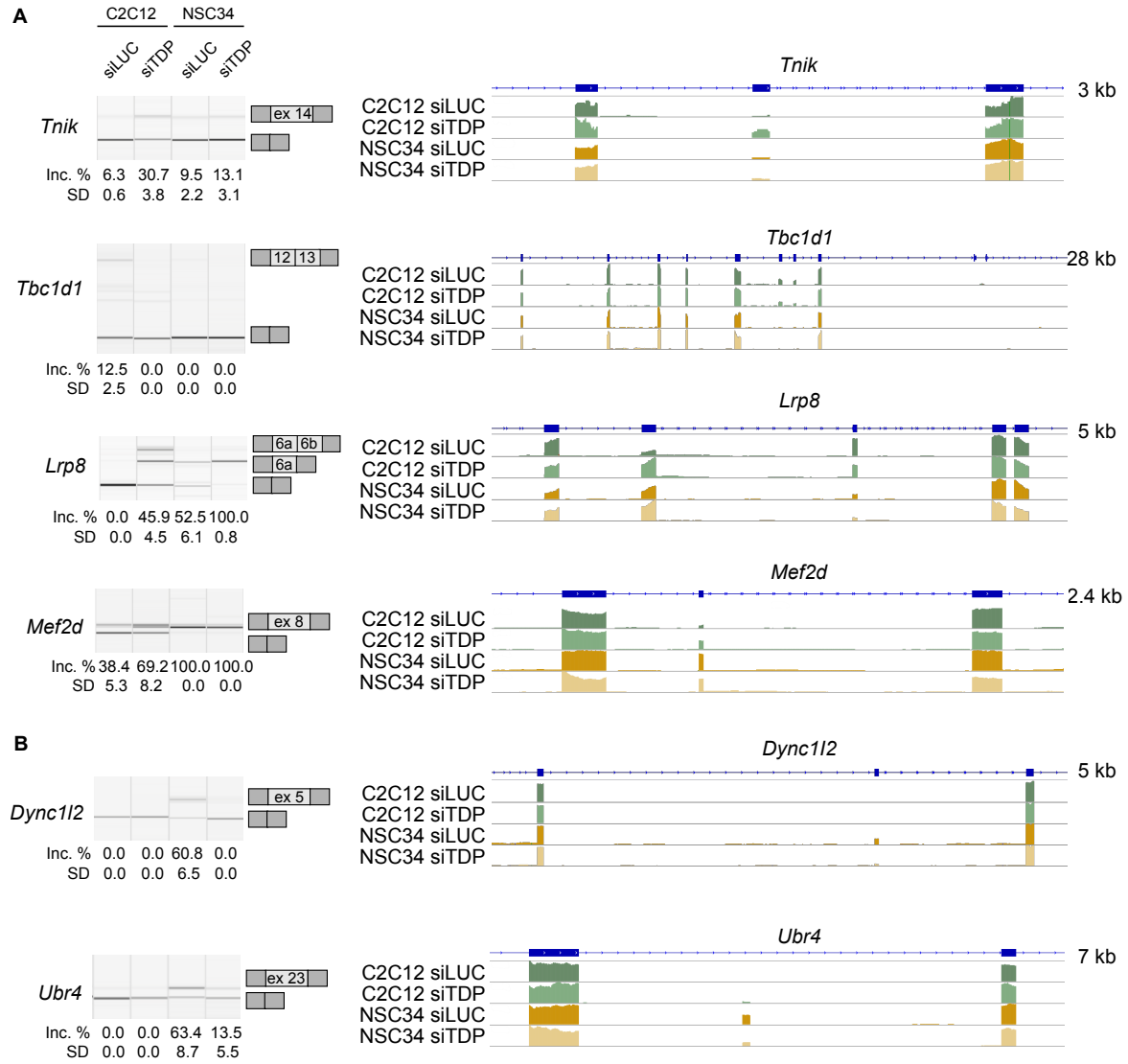


Figure 3.16: **Cell-type-specific TDP-43-dependent splicing events.** Validation of **A** C2C12-specific and **B** NSC34-specific TDP-43-regulated splicing. Semi quantitative RT-PCR conducted in TDP-43-silenced samples and corresponding controls is shown along with the quantification of splicing changes (% of alternative exon inclusion). The number of the alternative exon is shown in the scheme (see the exact transcript numbers in Table 2.3). The coverage track from Interactive genome viewer (IGV) is shown on the right.

Higher expression of RNA-binding proteins in NSC34

The observation that more splicing is subjected to TDP-43 control in NSC34 cells might reflect the importance of alternative splicing as a regulatory mechanism in neurons specifically and support the existence of a characteristic splicing program in neuronal tissues, as already suggested by others [26]; [41]; [176]. Why TDP-43-regulated splicing, but also alternative splicing in general, is more wide-spread in neurons than in muscles can only be speculated. Although TDP-43 is not more abundant in undifferentiated NSC34 relative to undifferentiated C2C12 cells (Figure 3.2a), RNA binding proteins (RBPs) other than TDP-43 might be differentially expressed in these cells (Figure 3.17a). Inspecting expression levels of some RBPs [41], which either directly interact with TDP-43 [10] or influence processing of its target transcripts [137]; [165]; [152], we saw a higher average expression of RBPs in neuronal NSC34 cells (the median expression of 49.5 and 57.5 FPKM in C2C12 and NSC34, respectively, Figure 3.17b) in line with previous observations [41]. Joint functions of multiple RBPs in coordinating mRNA processing might underlie a more complex splicing regulation that is unique for neuronal tissues and explain why TDP-43-regulated splicing is more frequent in NSC34 cells.

Two cell lines employed in our study clearly express a distinct array of RBPs (Figure 3.17a and c), while transcription levels of some are additionally affected by TDP-43 depletion (Figure 3.17d). Binding of tissue-specific splicing factors in the proximity of TDP-43-regulated alternative sequences could influence inclusion/skipping of these same sequences to further boost (the same directionality) or impede (the opposite direction) regulatory activity of TDP-43.

We believe this cooperative effect would explain why some alternative sequences appear be regulated by TDP-43 in a cell-type-specific fashion (Figure 3.16). Having identified and validated some cell-type-specific AS events, we next explored whether TDP-43-dependent expression of specific splicing isoforms could be explained by the co-regulatory activity of other RBPs that are highly expressed in one cell line but are absent (or have negligible levels) in the other. Using siRNA we knocked-down four RBPs expressed in a cell-type-specific manner (i.e., *Elavl4* and *Elavl3* in NSC34; *Celf2* and *Khdrbs3* in C2C12, as these four RNPs displayed the highest expression in one of the cell lines while they were not expressed in the other). Although we only achieved partial reduction of selected RBPs (Figure 3.18a), we could see that expression of tissue-characteristic TDP-43-regulated splice isoforms indeed (Figure 3.16b and 3.16c) depends on the co-regulation by tissue-specific splicing actors. For example, by silencing muscle-characteristic RBPs Celf2 or Khdbrs3 in C2C12 cells we decreased the expression of the long *Tbc1d1* isoform typical of C2C12 cells and obtained a more neuron-like phenotype even in the presence of TDP-43 (Figure 3.18b). By silencing Elavl3 or Elavl4 in NSC34 cells, on the other hand, we shifted the ratio of *Dync1i2* and *Lrp8* splice isoforms expressed in NSC34 cells towards the isoform typically present in C2C12 cells (Figure 3.18c). We therefore show that the tissue-specific splicing activity of TDP-43 arises as a consequence of different cellular environments, in particular characteristic expression of RBPs that modulate inclusion of TDP-43 targeted exons.

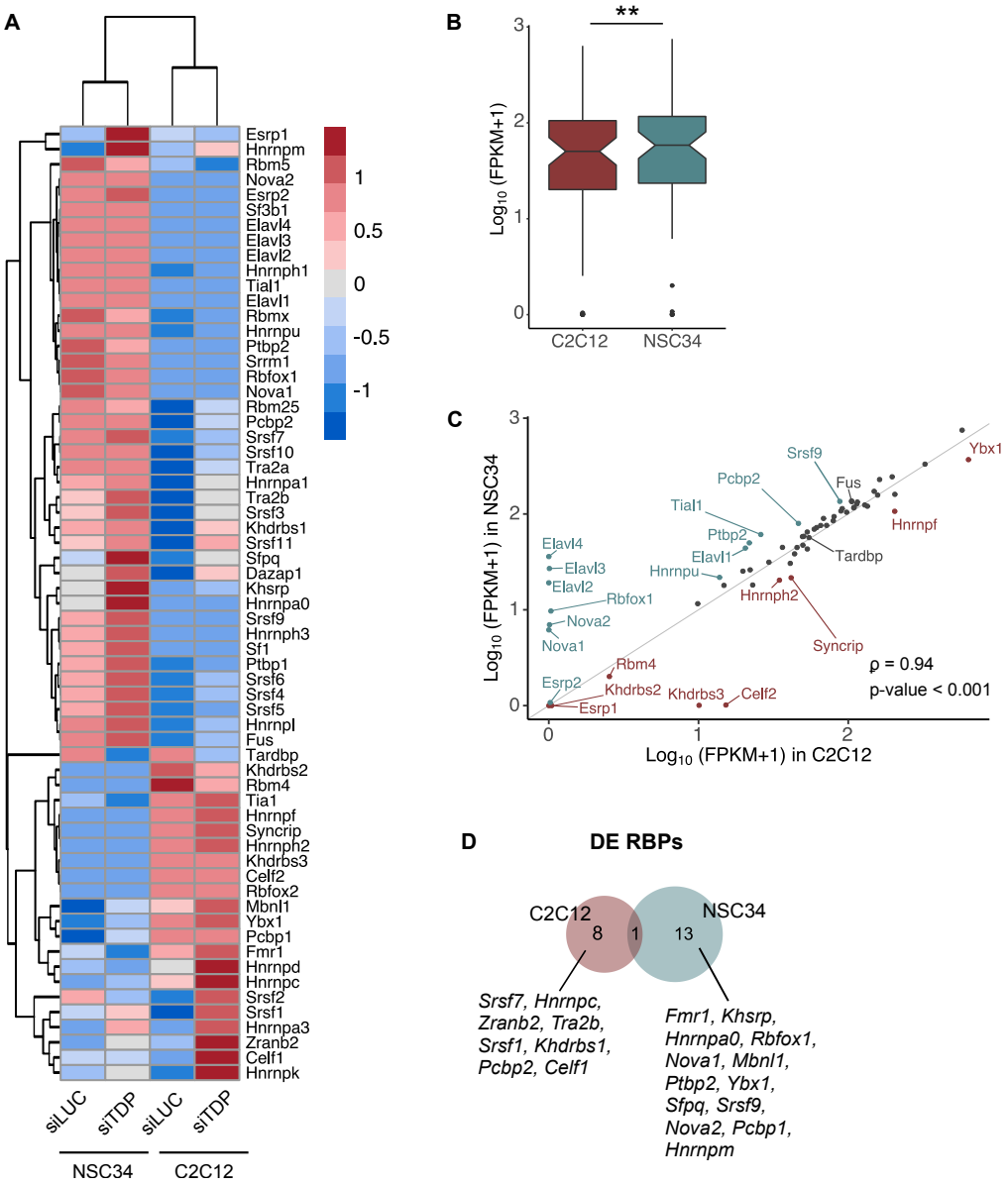


Figure 3.17: **RBP expression in C2C12 and NSC34. A** Heatmap illustrates differential expression of 63 RBPs in C2C12 and NSC34 cell line [41]. z-scores were generated by row. **B** Boxplot shows that NSC34 cells on average display higher expression of 63 RNA-binding proteins [41] compared to C2C12 cells (p-value = 0.0028). Average expression levels of all 63 transcripts are plotted as $\log_{10}$-transformed FPKM values. p-value was generated using Wilcoxon signed-rank test. **C** Expression of 63 RBPs (plotted as $\log_{10}$-transformed FPKM values) in C2C12 and NSC34 cells (Spearman's $\rho = 0.94$; p-value $<2.2{\cdot}10^{-16}$). Those with higher expression in one cell line than another ($> 150\%$) are shown in red (C2C12) or blue (NSC34). Grey line represents y = x. **C** Venn diagram shows RBPs the expression of which changes following TDP-43 reduction. The overlapping event is downregulation of *Tardbp*.

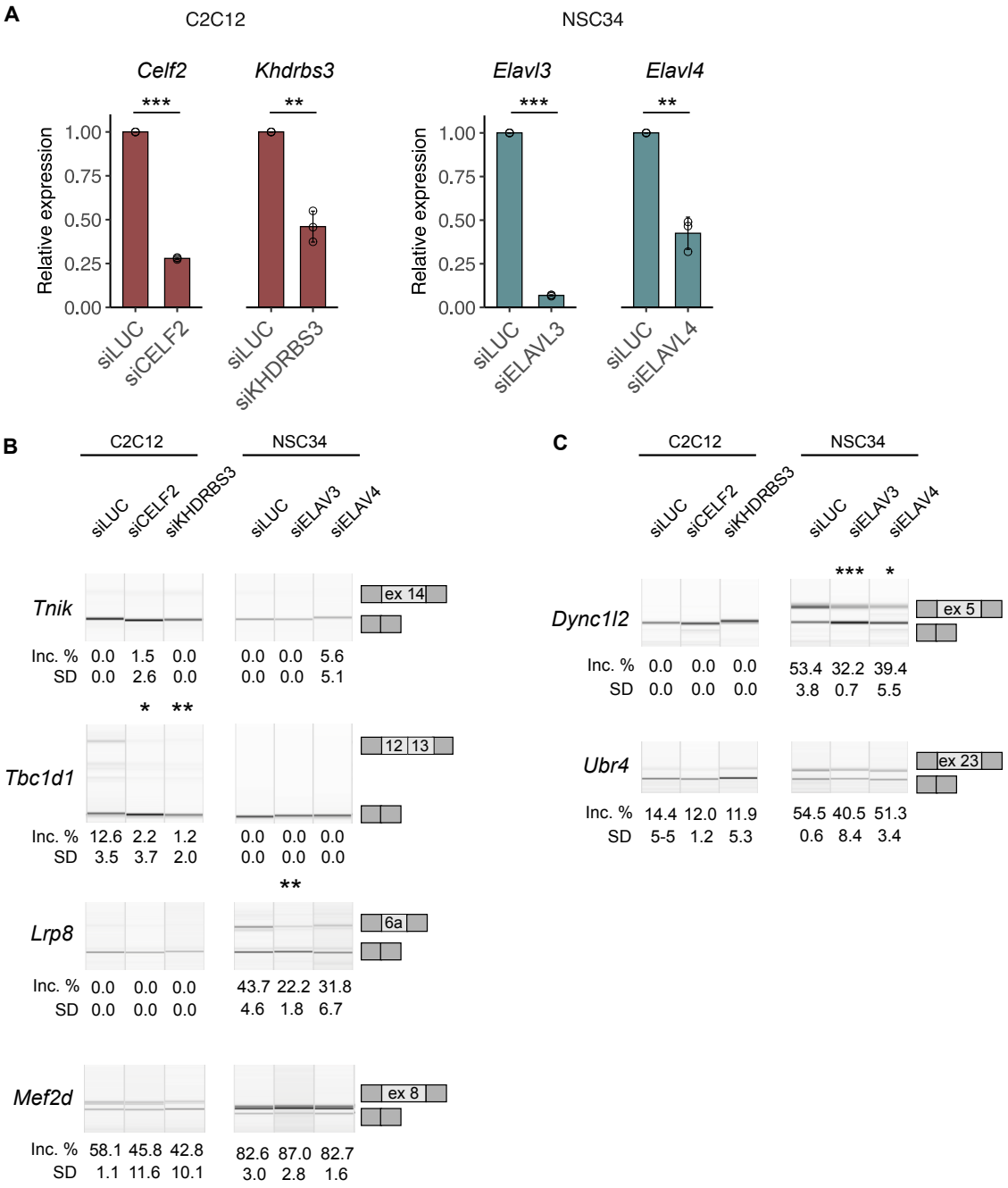


Figure 3.18: **Cell-type-specific RBPs influence TDP-43-dependent splicing.** **A** Relative expression changes of tissue-specific RBPs after siRNA transfection (siELAVL3, siELAVL4, siCELF2, siKHDRBS3 vs. siLUC) were validated using qPCR. p-values were generated using Student's t test (paired, one-tailed, n = 3 per group). **B** C2C12-specific and **C** NSC34-specific TDP-43-regulated splicing events are additionally controlled by tissue-specific RBPs. Semi quantitative RT-PCR conducted in TDP-43-silenced samples and corresponding controls is shown along with the quantification of splicing changes (% of alternative exon inclusion). The number of the alternative exon is shown in the scheme (see the exact transcript numbers in Table 2.3).

3.1.6 TDP-43 depletion leads to disturbance of common and cell-type-specific functions in C2C12 and NSC34

To further classify genes detected as differentially expressed (DEG) or differentially spliced (AS) we conducted functional analysis. The purpose of such analysis was to predict biological processes which might be under TDP-43's control. If TDP-43-mediated transcripts found in C2C12 and NSC34 were functionally linked and thus enriched in particular cellular pathways, that would imply on TDP-43's ability to regulate specific processes in muscles and neurons.

In mouse brain, TDP-43 has been shown crucial for maintenance of mRNAs that encode proteins involved in synaptic activity [8]. In line with the current literature, GO enrichment analysis of DEGs suggested a great portion of neuronal processes like *vesicle-mediated transport* and *regulation of postsynaptic membrane neurotransmitters* to be affected by TDP-43 loss in NSC34 cells (Figure 3.19a). Among C2C12-enriched GO terms, on the other hand, we found those directly associated with muscle characteristic features like *striated muscle development* or *muscle cell migration*, which agrees with results highlighting the importance of TDP-43 in skeletal muscle formation and regeneration [147]; [135].

While the percentage of overlapping DEGs was only 15.7% (Figure 3.4a), almost one third of all biological processes (28%) seemed to be disturbed upon TDP-43 depletion in both cell lines (Figure 3.19a).

Given that currently proposed picture of pathological processes implicated in myopathies bears several similarities with neurodegenerative diseases [177]; [133]; [91], we investigated commonly enriched GO terms to see, whether any of them could detect abnormalities previously described in the above-mentioned diseases. Significant GO terms enriched by DEGs in both C2C12 and NSC34 (Figure 3.19b) suggest that some common TDP-43-mediated mechanisms might contribute to development of TDP-43-proteinopathies in both muscle and neuronal tissues. Pathomechanisms include aberrant protein accumulation (i.e., of ubiquitin, amyloid β, α-synuclein, phosphorylated τ and TDP-43), post-translational modifications of deposited proteins (by phosphorylation, ubiquitination, acetylation, sumoylation), defects in protein disposal (26S proteasome and autophagy) and mitochondrial abnormalities.

However, despite the 28% overlap between predicted biological responses to TDP-43 depletion (GO: biological process), individual differentially expressed transcripts enriched in common GO terms were remarkably different between C2C12 and NSC34 (Figure 3.19c). This supports the idea that TDP-43 influences similar biological processes in both muscles and neurons, but it does so by mediating expression levels of genes encoding for distinct proteins that participate in those pathways. It is important to note, however, that we cannot exclude that two distinct subsets of genes (clustering within same pathways) were detected to be differentially expressed in each dataset (C2C12 and NSC34).

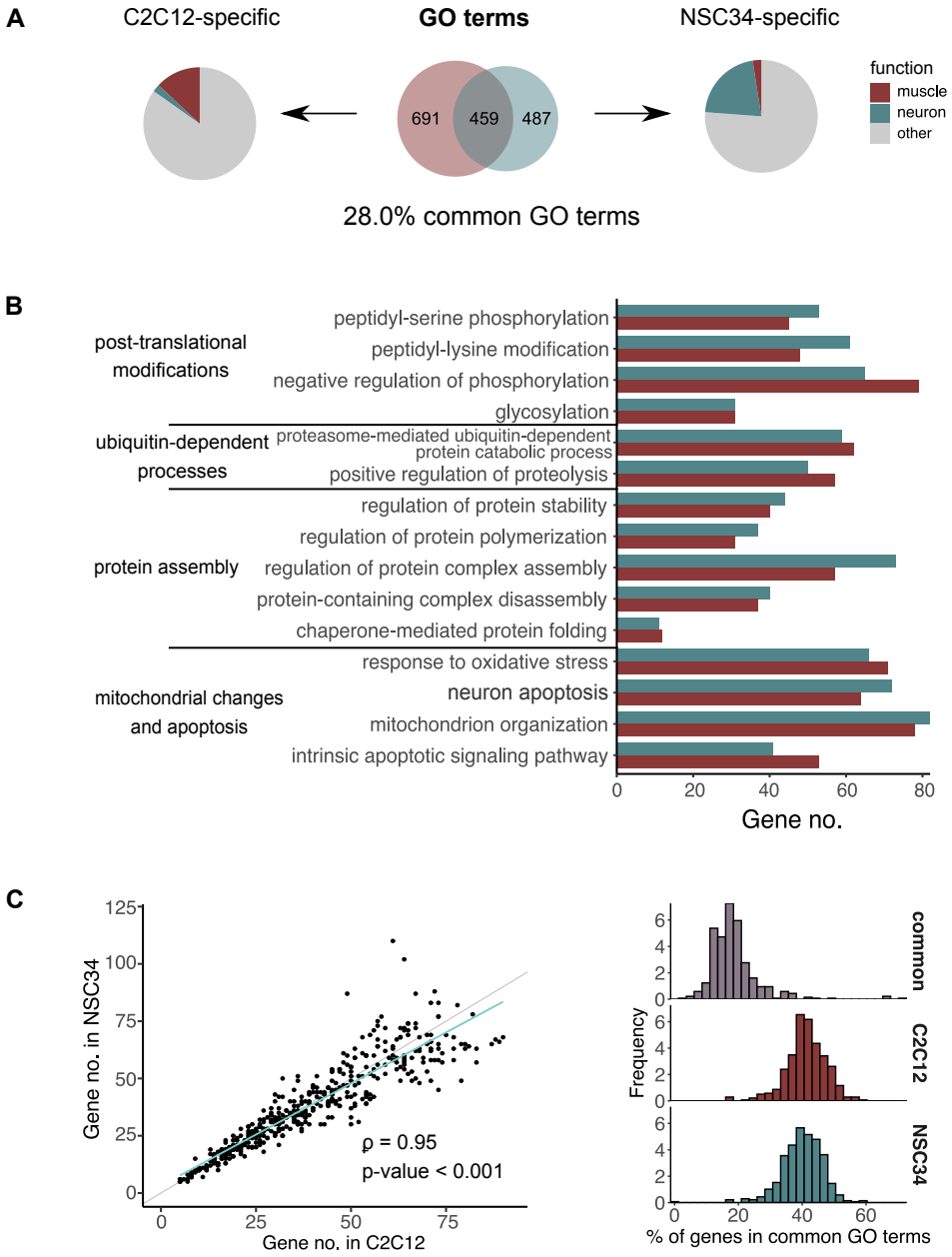


Figure 3.19: **GO enrichment analysis of DEGs.** **A** Venn diagram shows the number of cell-type-specific and overlapping GO terms enriched by DEGs identified in C2C12 or NSC34 dataset. GO terms (category: biological process) were grouped based on their names as those implying muscle- (red) or neuron-related features (blue). **B** Representative GO terms (category: biological process) commonly enriched by DEGs in C2C12 and NSC34 cells suggesting pathological abnormalities described in neurodegenerative and myodegenerative disease (hand curated). **C** The number of DEGs found in commonly enriched GO terms (**A**) is similar between two cell lines (left). Grey line represents y = x and the blue line the fitted regression (Spearman's $\rho = 0.95$, p-value $< 2.2 \cdot 10^{-16}$). Frequency plot shows that commonly regulated terms are highly enriched for cell-type-specific TDP-43-regulated DEGs (right).

We next employed GO analysis on misspliced genes identified in C2C12 and NSC34 dataset. Since the number of C2C12 AS genes entering GO analysis (578) was considerably lower than that of NSC34 genes (1018), the analysis resulted in fewer GO terms found to be enriched in C2C12 compared to many in NSC34 (23 and 203, respectively) (Figure 3.20a). As expected, GO terms enriched in NSC34 cells exclusively suggest that misspliced transcripts predominantly encode for proteins implicated in processes taking place in the nervous system (e.g., *axonogenesis*, *regulation of neuron differentiation*) (Figure 3.20a). This is in line with earlier studies demonstrating that in human neuroblastoma cells SH-SY5Y, TDP-43 targets encode for proteins regulating neuronal development and those involved in neurodegenerative disease [7].

On the contrary, a great portion of GO terms uniquely enriched in C2C12 (56%) suggested involvement of misspliced genes in DNA-related processes (e.g., *covalent chromatin modification* or *regulation of chromosome organization*), while only one implied on a muscle characteristic feature (i.e., *regulation of cardiac muscle cell action potential*) (Figure 3.20c). As we thought this observation might be biased due to the fact we were only looking at very few GO terms (18), we repeated enrichment analysis using a more relaxed threshold (Figure 3.20d). Although less stringent threshold could produce some false positives, it was intentionally chosen to increase the number of splicing events and genes that enter GO analysis. At p-value < 0.01 (in contrast to FDR < 0.01), 1130 events in 869 genes entered the analysis, which gave 45 enriched GO terms. Even so, DNA-related processes comprised more then a third of all GO terms, which was not the case for NSC34.

This is a particularly interesting observation as DNA-related processes play an important role in muscle differentiation. In adult skeletal muscle, DNA and histone modifications participate in adaptive response to environmental stimuli, which challenge structural and metabolic demands and thus make skeletal muscle a very plastic tissue [178]; [179].

During muscle development and regeneration, the early commitment towards myogenic lineage involves epigenetic changes mediated by chromatin remodeling enzymes like histone deacetylases (HDACs), histone acetyltransferases (HATs) and histone methyltransferases (HMTs) [180] which influence the activity of major muscle-specific transcription factors like myogenic regulatory factor (MRF) and myocyte enhancer factor (MEF2) that in turn govern all steps of myogenic commitment, differentiation and proliferation. In keeping with this, *Dnmt3a*, *Dnmt3b*, *Hdac9*, *Hdac7*, *Prdm2* are just few of chromatin-modifying enzymes that underwent splicing changes following TDP-43 knockdown in C2C12 but not in NSC34 cells. Interestingly, telomere shortening was described in primary muscle cultures from sIBM patients suggesting premature senescence [181] and epigenetic changes have been described in congenital myopathies [182]. Therefore, results obtained in C2C12 imply another possible mechanism on how TDP-43 may control gene expression in muscle in an indirect fashion and eventually participate in disease. Recently, loss of TDP-43 was found to be associated with increased genomic instability and R-loop formation [183]; [184], possibly through mechanisms involving *Poldip3*. POLDIP3 has

been shown to play a role in maintaining genome stability and preventing R-loop accumulation at sites of active replication [185].

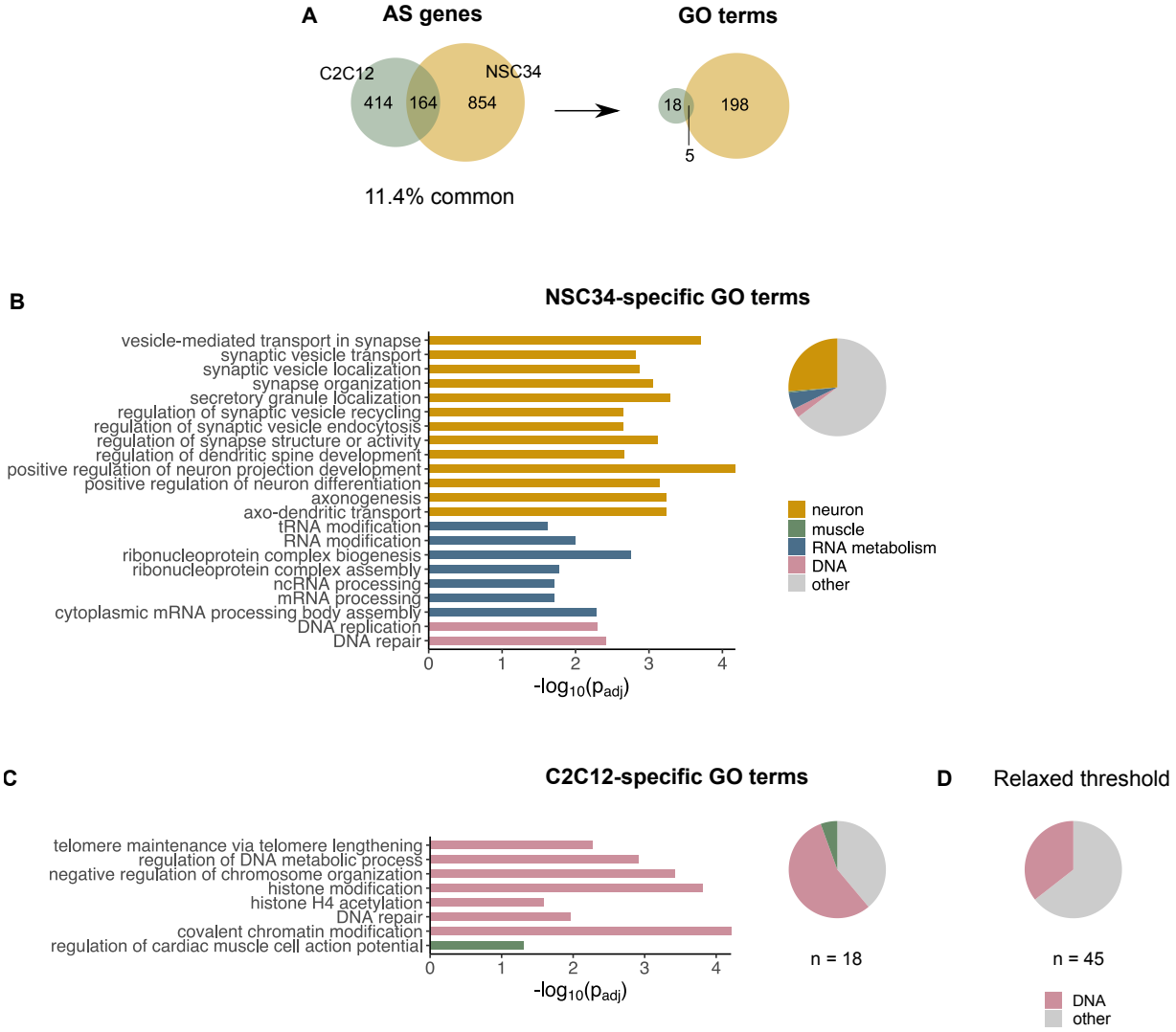


Figure 3.20: **GO enrichment analysis of AS genes.** **A** Venn diagrams show the number of alternatively spliced transcripts (as detected by rMATS, FDR $<$ 0.01) in C2C12 and NSC34 cells together with GO terms (category: biological process, $p_{adj} < 0.05$) enriched in AS genes detected in each cell line. **B** GO terms uniquely enriched in NSC34 (198) imply on deregulation of neuronal processes, mRNA metabolism and DNA biology in NSC34 cells (representative GO terms are shown on the plot). **C** GO terms uniquely enriched in C2C12 (18) suggest involvement of TDP-43-regulated AS genes in DNA-modifying processes (representative GO terms are shown on the plot). **D** GO enrichment analysis (refers to **C**) was additionally conducted on AS genes detected in C2C12 at stringent threshold (p-value $<$ 0.01 instead of FDR $<$ 0.01). Resulting GO terms (45) imply on disturbance of DNA-related biological processes.

3.1.7 TDP-43 regulates alternative splicing and transcript abundance

About 80% of alternative splicing was estimated to occur co-transcriptionally [186] and the two mechanisms have been known to influence one another in a coordinated manner [33]. In our case, however, only a small portion of transcripts undergoing TDP-43-dependent splicing additionally displayed changes in the overall transcript abundance (21.9% and 21.2% in C2C12 and NSC34, respectively) (Figure 3.21a). At least in C2C12 cells, transcripts whose splicing was affected by loss of TDP-43 more often decreased in abundance, which might be indicative of nonsense mediated decay (Figure 3.21b).

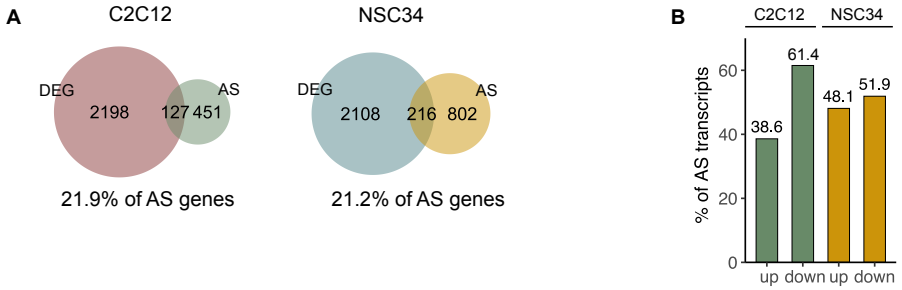


Figure 3.21: **Transcripts subject to TDP-43-dependent splicing and expression level changes. A** Venn diagrams show the percentage of transcripts whose splicing (AS) and abundance levels (DEG) are affected by TDP-43 depletion in each cell line (21.9% of AS genes in C2C12 and 21.0% of AS genes in NSC34, respectively). **B** Barplot shows the percentage of down- and upregulated genes among those that are misspliced following TDP-43 loss.

KEGG analysis performed on sets of differentially expressed or alternatively spliced genes supports the idea that TDP-43 could mediate a particular molecular pathway such as *axon guidance* by the means of regulating the overall transcript abundance (DEG) or by influencing alternative splicing (Figure 3.22) of genes that encode for proteins participating in that pathway.

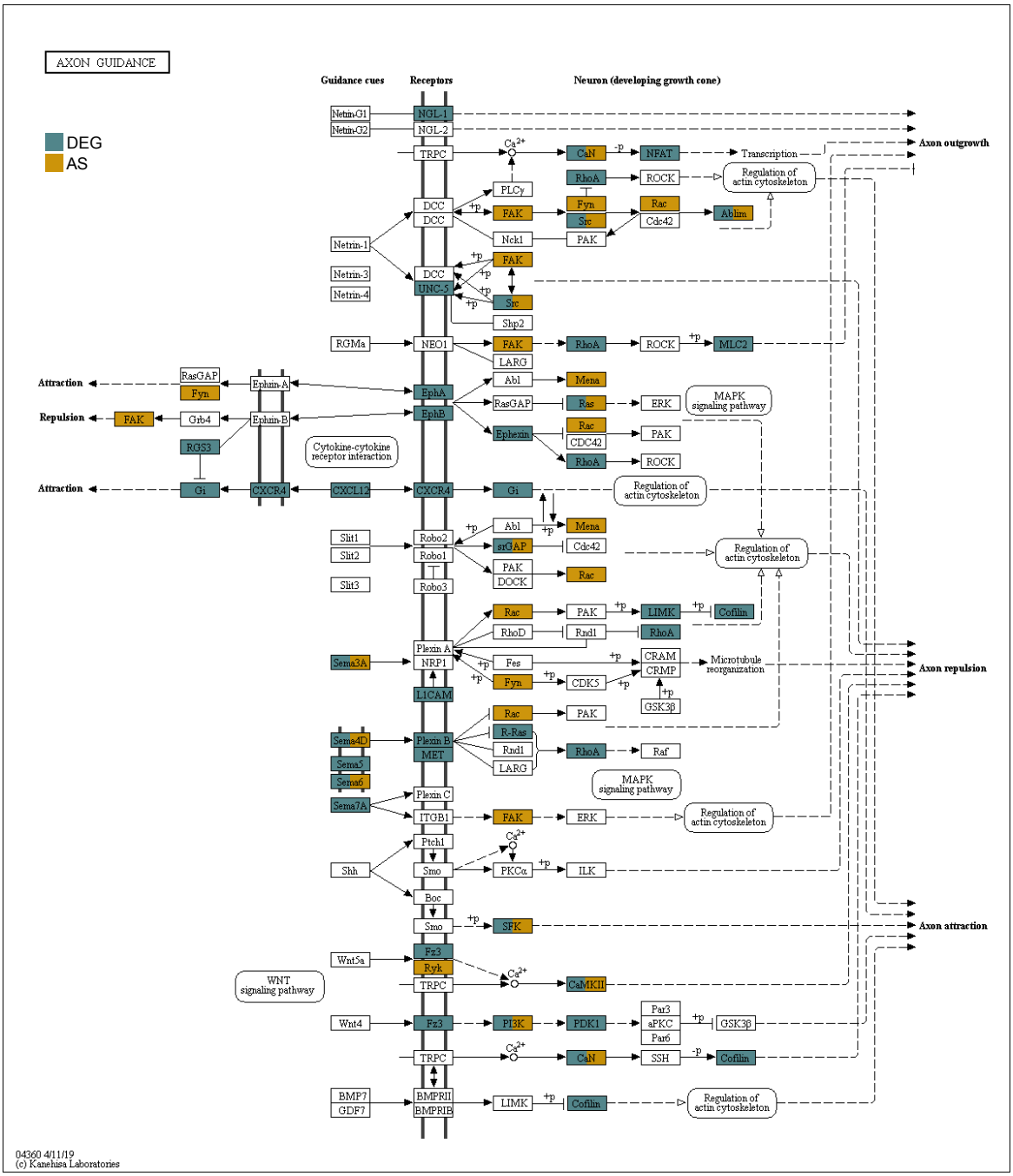


Figure 3.22: **An exemplary KEGG pathway enriched in TDP-43 targets.** A representative KEGG pathway (axon guidance pathway, mmu04360) significantly enriched ($p_{adj} < 0.05$) in AS and DE genes in NSC34 cells demonstrates that TDP-43 might influence axon guidance by regulating AS and expression levels of transcripts encoding proteins that participate this biological process. Proteins encoded by AS transcripts are shown in yellow and those encoded by DEG are shown in blue.

3.1.8 Conclusions

- Expression of TDP-43 is very similar between undifferentiated C2C12 and NSC34 cells.
- Expression changes induced by TDP-43 knockdown occur based on a cell-type-characteristic transcriptome.
- TDP-43-mediated control of transcript abundance is equally important in C2C12 and NSC34 cells.
- TDP-43-dependent splicing is more pronounced in neuronal NSC34 cell line.
- Different arrays of RBPs are expressed in two cell line.
- We estimate that about 50% of splicing events are subject to TDP-43 regulation in both, C2C12 and NSC34.
- Cell-type-specificity of TDP-43's splicing activity can be explained by differential expression of host mRNAs and cooperative regulation provided by other RBPs that are expressed in a cell-type-specific manner.
- TDP-43 mediates splicing and abundance of transcripts that encode for proteins participating in essential muscle-related functions in C2C12 cells or typical neuronal functions in NSC34.
- TDP-43 loss is predicted to impact biological processes, which have been previously described as a part of ALS and IBM aetiology.

3.2 Functional consequences of TDP-43 knockdown in C2C12 cells

In recent years, advent of sequencing techniques fostered global investigation of transcriptional and post-transcriptional dynamics. This, together with the advances in computational tools, provided the means for investigation of transcription (splicing) networks and mechanisms of their regulation.

The Gene Ontology (GO) annotations, which link each gene (i.e., protein-coding or non-coding RNAs encoded by each gene) with a functional information obtained from many different organisms [126], are then used for interpretation of big scale experiments (e.g., RNA-seq) to gain insights into functions potentially affected under given experimental conditions.

We therefore conducted GO enrichment analysis of differentially expressed (DEG) and alternatively spliced (AS) genes detected in our RNA-seq datasets of TDP-43-silenced myoblasts and neuroblasts (see the previous chapter). While computational predictions like GO analysis provide a useful tool for identification of networks that consist of co-regulated transcripts from functionally related genes and can give hints on biological processes that are possibly disturbed under a certain condition, outcomes of such analysis should be validated using molecular experiments. In the context of myodegeneration, we aimed to explore whether TDP-43 loss would (due to disturbance of transcript abundance or splice isoform expression) affect specific physiological functions in muscle-derived cells. Functional studies have previously demonstrated a central role played by TDP-43 in maintaining core neuronal processes: axonogenesis [187]; [188], neuron development [189], synaptic function [173]; [190] and apoptosis [188]. Due to lack of information on features governed by TDP-43 in muscles specifically, we investigated the consequences of TDP-43 depletion in muscle-derived C2C12 cells. GO enrichment analysis performed on genes differentially expressed following TDP-43-knockdown revealed a significant enrichment of GO terms (category: biological process) related to muscle cell differentiation (Figure 3.23), which prompted us to investigate eventual involvement of TDP-43 in myogenesis, since this had not been explored by then. It has to be mentioned, however, that two studies [147]; [135] addressing the same phenomenon were published at the time differentiation experiments were being conducted.

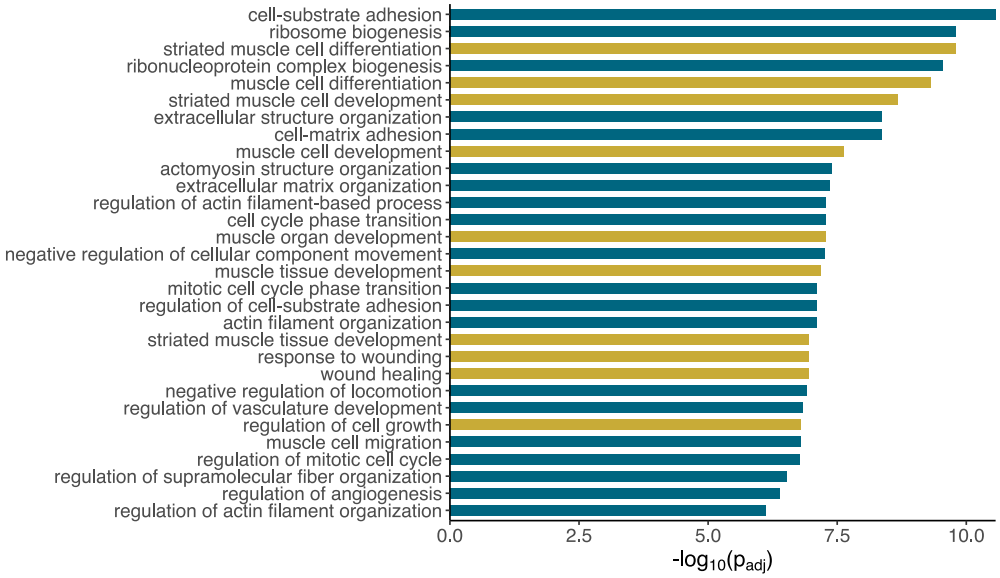


Figure 3.23: **GO enrichment analysis of DEGs detected in TDP-43-silenced C2C12 cells.** Top 30 GO terms (by p_{adj}, category: biological process) enriched by DEGs ($p_{adj} < 0.05$) detected upon TDP-43 depletion in C2C12 myoblasts. GO terms directly implying on disturbance of differentiation-related processes are shown in yellow.

3.2.1 C2C12 differentiation

Myogenesis is a complex process that first takes place during embryonic development and later during muscle growth and regeneration. Myoblasts (or myogenic cells) arise from asymmetric division of mitotically quiescent satellite cells. They proliferate, differentiate and either fuse to pre-existing myofibers or form new myofibers. C2C12 cell line is often used as a model of muscle cells differentiation. In culture, differentiating C2C12 cells contain a sub-population of reserve cells, which are satellite-like cells arrested in G0 phase. They can, anytime, re-enter cell cycle and replenish the pool of differentiating cells [191].

Myogenic differentiation is a precisely orchestrated process that depends on a family of myogenic regulatory factors (MRF) - myogenic factor 5 (MYF5), myogenic differentiation (MYOD), myogenic regulatory factor 4 (MRF4), and myogenin (MYOG) are the core components of the myogenic pathway. Of these, myogenin expression happens later after a cell has been already determined to undergo myogenic fate, and is crucial for terminal differentiation. In addition to MRFs, a number of growth factors, non-coding RNAs, epigenetic regulators and other signaling pathways collectively contribute to muscle formation *in vivo* and *in vitro* [147]; [191]. Furthermore, precisely regulated splicing changes were described to be important in biological transitions from undifferentiated cells to mature myocytes. Bland *et al.* [27] identified 95 alternative splicing events that undergo robust transitions during C2C12 differentiation in a time-specific manner. Some splicing factors, such as those from FOX and Muscleblind-like

families, are enriched in muscles.

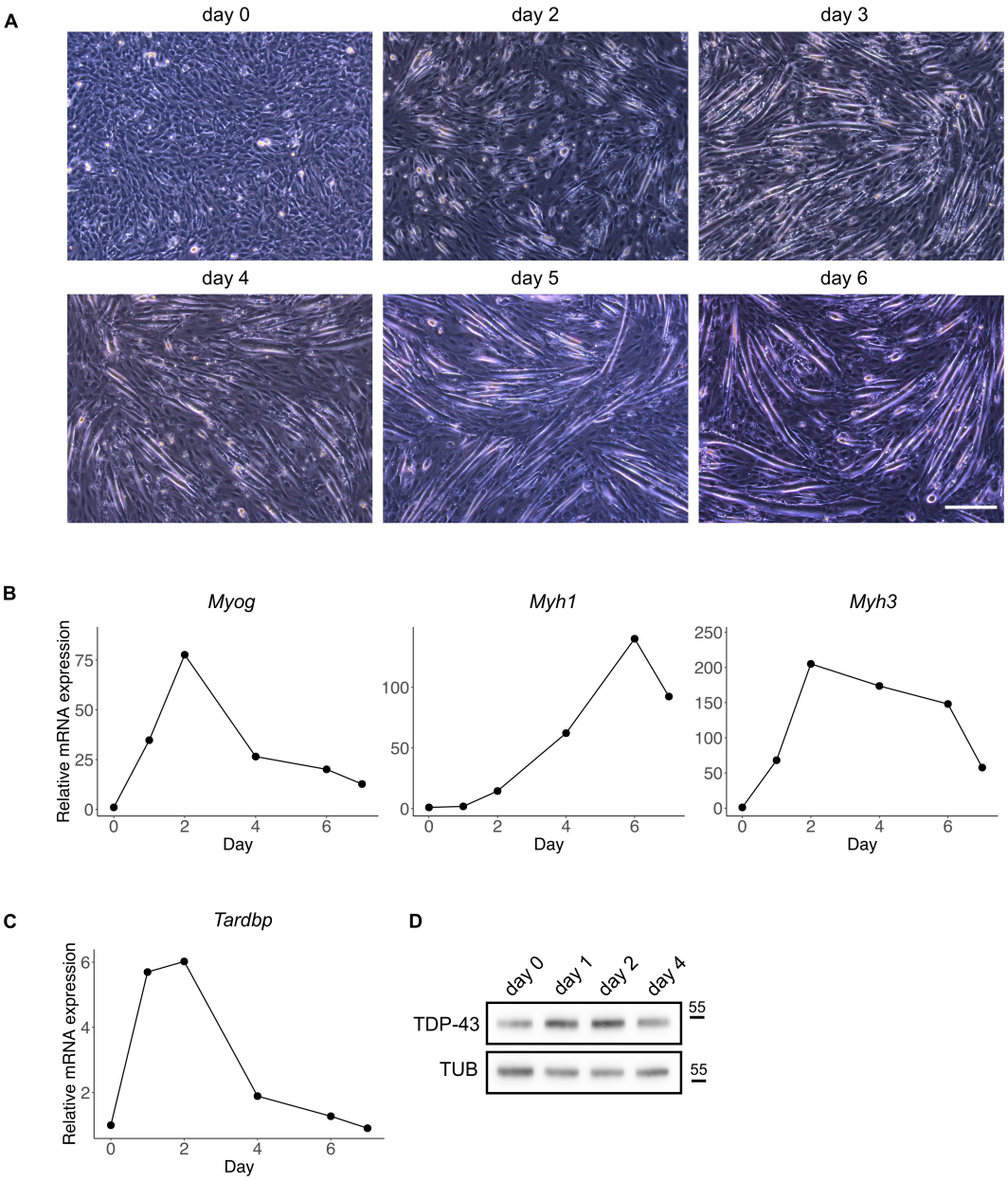


Figure 3.24: **Differentiation of C2C12 myoblasts. A** Morphological changes that occur along myoblast transition into myotubes. White line marks the length of 200 nm. **B** Relative expression changes of *Myog*, *Myh1*, *Myh3* and **C** *Tardbp* 7 days after induction of differentiation as assessed by qPCR. Expression of individual genes was normalized against *Gapdh*. Means of (technical) triplicates from one of three representative experiments are shown on the plot (therefore, no SD and p-values were generated). **D** Western blot shows expression levels of TDP-43 in differentiating C2C12 cells (day 0-4). TDP-43 expression was normalized to α-tubulin.

Before investigating the role of TDP-43 in C2C12 differentiation, we induced C2C12 differentiation in non-treated cells and assessed TDP-43 levels at different time points. Briefly, confluent myoblasts were induced to differentiate by replacing growth medium with differentiation medium, which has low concentrations of serum-derived mitogens. Upon exposure to low serum conditions cells upregulate myogenin, exit cell cycle and form multinucleated myotubes [192] (Figure 3.24a). Beside morphological changes, we measured expression of 3 differentiation markers; *Myog*, a muscle-specific transcription factor; and two myosin heavy chain genes - *Myh1* and *Myh3* (Figure 3.24b). mRNA levels of all three genes were low in undifferentiated myoblasts. Consistently with the literature [193] their expression increased upon differentiation (although in distinct temporal patterns) and then decreased again in the late stage.

The same was found for the expression of *Tardbp* (Figure 3.24c). Its mRNA levels were low in undifferentiated C2C12 cells but a 6-fold increase was seen 24 h hours after induction, which could be observed at the protein level (Figure 3.24d).

In line with other reports [147]; [135] (which were published just after we had conducted differentiation experiments), transient increase in *Tardbp* expression seems to be characteristic of the first part of cell transition, whereas it returns to the levels of undifferentiated cells at later stages. Vogler *et al.*, 2018 [135] recently demonstrated that TDP-43 plays an essential role in muscle formation (and regeneration). In differentiating muscle, TDP-43 was reported to temporarily translocate to the cytoplasm and form cytoplasmic amyloid-like assemblies called myo-granules, which bind mRNAs encoding for sarcomeric proteins. Myo-granules get cleared as myofibers mature.

3.2.2 TDP-43-dependent splicing in C2C12 differentiation

Increase in TDP-43 levels during the first days of C2C12 differentiation prompted us to explore what effect, if any, TDP-43 depletion has on C2C12 differentiation. We thus silenced TDP-43 in C2C12 myoblasts and induced differentiation by exposing confluent cells to low-serum conditions. The differentiation was only performed for 4 days, for we have previously shown (Figure 3.24) that morphological and transcription changes associated with cell transition occur as early as from day 1 on.

Given that TDP-43 acts as a splicing factor, we further explored whether depletion of TDP-43 influences splicing of transcripts involved in muscle development. We selected a set of genes known to undergo differentiation-dependent splicing transitions - *Atp6V1H*, *Capzb*, *Dtna*, *Fxr1*, *Lrrfip1*, *Mef2a*, *Mtmr3*, *Pkm* and *Tbc1d1* [27] (Figure 3.25 and 3.26). Among these, only splicing of two cassette exons within *Tbc1d1* appeared to be TDP-43-dependent and was thus investigated further.

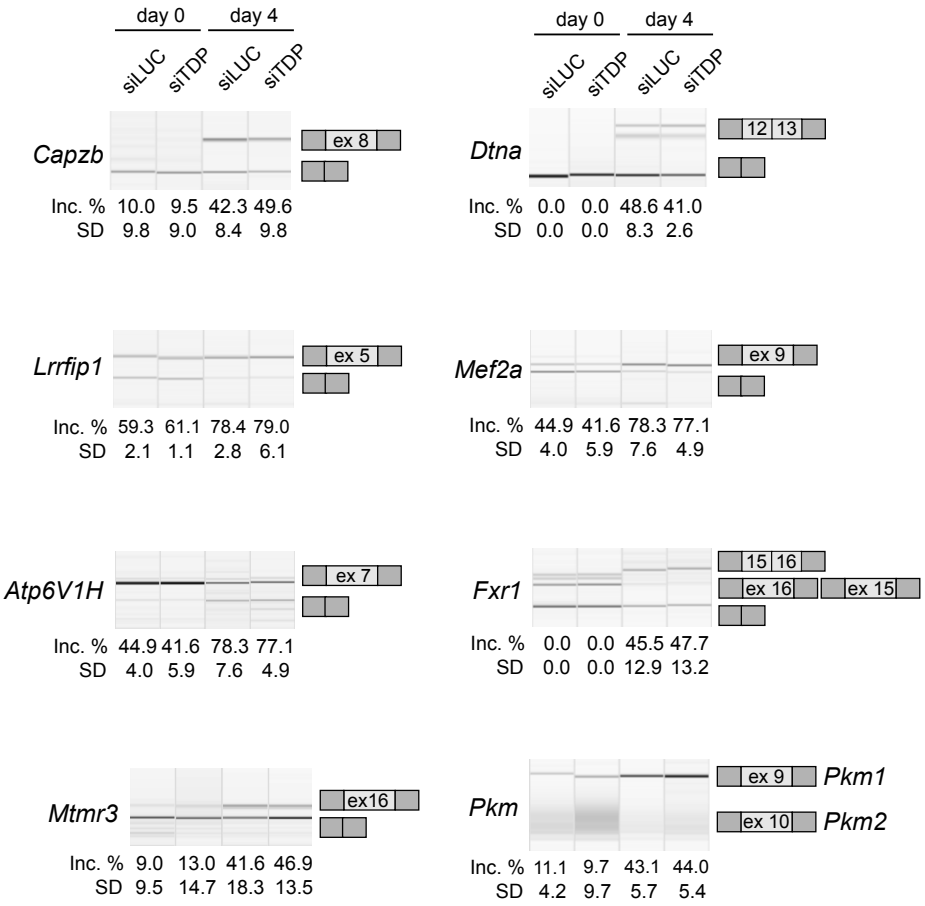


Figure 3.25: **Splicing transitions associated with C2C12 differentiation.** Splicing transitions (day 0 vs. day 4) of eight representative mRNAs found to be associated with C2C12 differentiation by Bland *et al.* [27] were validated using semi-quantitative RT–PCR comparing siTDP-transfected samples with corresponding controls (siLUC). Quantification of splicing changes is given bellow (% of alternative exon inclusion) and the number of the alternative exon is shown in the scheme (see the exact transcript numbers in Table 2.3).

3.2.3 TDP-43 regulates splicing of *Tbc1d1*

TBC1D1 (TBC1 domain family member 1) is believed to regulate insulin-mediated glucose uptake by modulating GLUT4 (glucose transporter 4) translocation from peri-nuclear space to plasma membrane.

Tbc1d1 was previously reported to undergo robust splicing transition upon differentiation [27]. Indeed, the long isoform that is barely present in C2C12 myoblasts at the beginning of differentiation (around 12% of the total transcript) becomes prevalently expressed later as C2C12 differentiate into myotubes (more than 60% at day 7) (Figure 3.26a). Increased expression of long *Tbc1d1* mRNA isoform (containing exons 12 and 13) can also be seen on the protein level

(Figure 3.26b), as both transcripts translate into proteins (with the size of 128 and 138 kDa, respectively). Despite not being quantified, the ratio of two protein isoforms seems to reflect that of mRNAs.

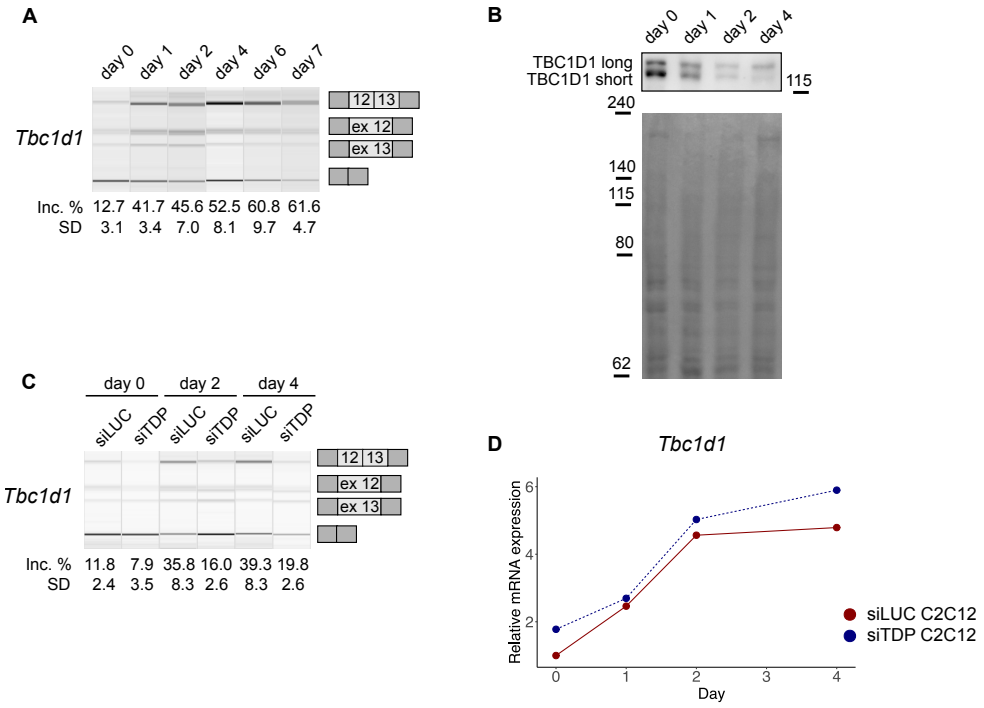


Figure 3.26: **TDP-43-dependent *Tbc1d1* splicing in differentiating C2C12.** **A** *Tbc1d1* isoform switch occurs along C2C12 differentiation in non-treated C2C12 cells. The ratio between splice isoforms was assessed using semi-quantitative RT-PCR. Percentage of the long isoform (with exons 12 and 13) is given bellow. **B** Both *Tbc1d1* mRNA isoforms (128 and 138 kDa, respectively) are translated into protein. Western blot shows increase in the big TBC1D1 protein isoform along C2C12 differentiation in non-treated C2C12 cells. Pierce staining of the membrane is shown bellow. **C** The ratio between splice isoforms of *Tbc1d1* gene at the initial stage of C2C12 differentiation (day 0, day 2, day 4) in TDP-43-silenced cells and corresponding controls (siLUC) was assessed using semi-quantitative RT-PCR. Percentage of the long isoform (with exons 12 and 13) is given bellow. **D** Expression changes (relative to day 0) of *Tbc1d1* in siLUC- (red) or siTDP-transfected (blue) in C2C12 cells 4 days after induction of differentiation as assessed by qPCR. Gene expression was normalized against *Gapdh*, and means of (technical) triplicates from one of three representative experiments are shown on the plot (therefore, no SD and p-values were calculated).

TDP-43 depletion blocked the inclusion of two alternative exons (12 and 13) not only in undifferentiated myoblasts (day 0) but also in differentiating cells (on day 2 and 4) (Figure 3.26c). Inclusion of exons 12 and 13 seems to be linked since inclusion of either exon (on its own) rarely occurs.

Beside splicing, TDP-43 has been previously proposed to regulate *Tbc1d1* expression [114]; [194]. In inducible knock-out cell models and skeletal muscle of mice, TDP-43 loss led to reduced *Tbc1d1* expression [114]. In our model of cell differentiation, however, the total expression of *Tbc1d1* appeared to be higher or unchanged (depending on the time point) in TDP-43-silenced cells (Figure 3.26d, mean values of three technical replicates obtained in one of three representative experiments are plotted).

We have previously shown that the long *Tbc1d1* isoform could be detected in undifferentiated C2C12 but not undifferentiated NSC34 cells (Figure 3.16a). To investigate whether inclusion of exons 12 and 13 is a muscle-characteristic event or the two exons were just spliced out in neurons, we additionally silenced TDP-43 in murine fibroblast cell line NIH-3T3. A very weak upper band was detected in siLUC-transfected cells, which did correspond to the size of the long mRNA isoform, however, it represented less than 5% of the total transcript (Figure 3.27a). Interesting enough, another splice isoform was detected in NIH-3T3 cells, the size of which could not be explained by the length of known (annotated) exons located between exon 11 and 14, in which primers align. Nevertheless, that isoform was not subject to TDP-43-regulated inclusion and was not of further interest.

Even though we did not detect the long *Tbc1d1* in undifferentiated NSC34 and NIH-3T3 cells, we further investigated whether this isoform gets expressed in a muscle-specific manner or alternatively, it is related to differentiation itself. Like in muscle-derived cells, enhanced retention of exons 12 and 13 was observed during neurogenesis of cortical glutamatergic neurons from murine ESCs [195]. Given that long *Tbc1d1* isoform is actually the one prevalently expressed across multiple brain regions and in mature skeletal muscle in mice and human (VastDB [170]), we reasoned that rather than being muscle-specific, retention of exons 12 and 13 within mature *Tbc1d1* is an event associated with differentiation itself, of both muscle and neuronal cells. We believe this notion explains why the long isoform was not detected in undifferentiated NSC34 cells. Since inclusion of exons 12 and 13 depends on the presence of TDP-43 at all stages of myoblast transition and since TDP-43 binding sites was detected by iCLIP downstream exon 13 in mouse brain, it is very likely that TDP-43 in fact controls retention of exons 12 and 13 also in neuronal cells. On the other hand, the long protein isoform is not expressed in adipocytes differentiated from NIH-3T3 fibroblasts [196].

Insertion of two cassette exons within *Tbc1d1* (159 and 120 nucleotides, respectively) does not lead to frame-shifting neither it introduces a premature stop codon, thus, it is not predicted to cause NMD. TDP-43 loss did not significantly impact *Tbc1d1* transcription levels in any of cell lines tested (Figure 3.27b). Of note, all these experiments were conducted using undifferentiated cells, which is in contrast to inducible knock-outs and mature tissues investigated by Chiang *et al.* [114]. It is possible that depletion of TDP-43 induces a different response in undifferentiated cells compared to mature tissue.

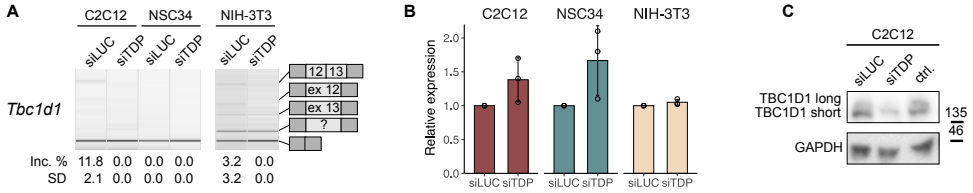


Figure 3.27: ***Tbc1d1* isoforms in undifferentiated C2C12, NSC34 and NIH-3T3 cells.** **A** Proportion of different splice isoforms of *Tbc1d1* gene in undifferentiated C2C12 and NSC34 cells, as well as in mouse fibroblast cell line NIH-3T3 was assessed by semi-quantitative RT-PCR. Percentage of the long isoform (with exons 12 and 13) is given bellow. **B** Expression of *Tbc1d1* does not change upon TDP-43 silencing in C2C12, NSC34 and NIH-3T3 cells as validated by qPCR. Significance was tested using Student's t test (paired, two-tailed). **C** Both protein isoforms of TBC1D1 were detected by western blot in control C2C12 myoblasts (either siLUC-transfected or untreated - ctrl.), while only the smaller protein isoform was expressed in TDP-43-silenced cells.

Importantly, TDP-43-dependent expression of the long protein isoform in C2C12 myoblasts was abolished following TDP-43 knockdown (Figure 3.27c) as this was confirmed by the western blot. Eventual differences in the functionality of two protein isoforms remain unknown. What we know, however, is that there are two phosphorylation sites (S660 and S700) found within the region encoded by exons 12 and 13 (93 aa) of the long TBC1D1 isoform in mouse (1255 aa) (Figure 3.28a), *in vitro* [197] and *in vivo* [198]. Both serine residues were predicted to undergo AMPK (5' AMP-activated protein kinase)-dependent phosphorylation [197]; [198]; [199] as assessed by AICAR stimulation. In human skeletal muscle, the two residues get phosphorylated in response to exercise but not by insulin [200]. Phosphorylation of TBC1D1 along with subsequent binding to 12-3-3 proteins have been shown to not affect its GAP (GTPase-activating) activity towards Rab10, Rab8 and Rab14 [197]. TBC1D1 phosphorylation was instead proposed to impact its recruitment to GLUT4 (glucose transporter type 4) vesicles. TBC1D1 interacts with IRAP (insulin-regulated aminopeptidase), a protein resident in vesicles, and this interaction was reduced upon TBC1D1 phosphorylation by AMPK. Taken together, phosphorylation of TBC1D1 was proposed to prevent recruitment of TBC1D1 to GLUT4 vesicles and activate Rabs located within those vesicles [197] (Figure 3.28b).

In mouse models, depletion of TDP-43 caused TBC1D1 downregulation [114] while TDP-43 overexpression increased TBC1D1 levels [194]. TDP-43 overexpression inhibited translocation of GLUT4 vesicles in the skeletal muscle and impaired glucose uptake [194]. In our cell line models, expression of *TBC1D1* remained unaltered following TDP-43 loss, however, we described an TDP-43-dependent isoform switch of TBC1D1. Given that the long TBC1D1 contains two phosphorylation sites, which have been proposed to influence recruitment of TBC1D1 to GLUT4 vesicles, it would be interesting to explore whether inclusion of TDP-43-regulated exons would alter the activity of TBC1D1 such to affect GLUT4 translocation and glucose uptake in stimulated skeletal muscle.

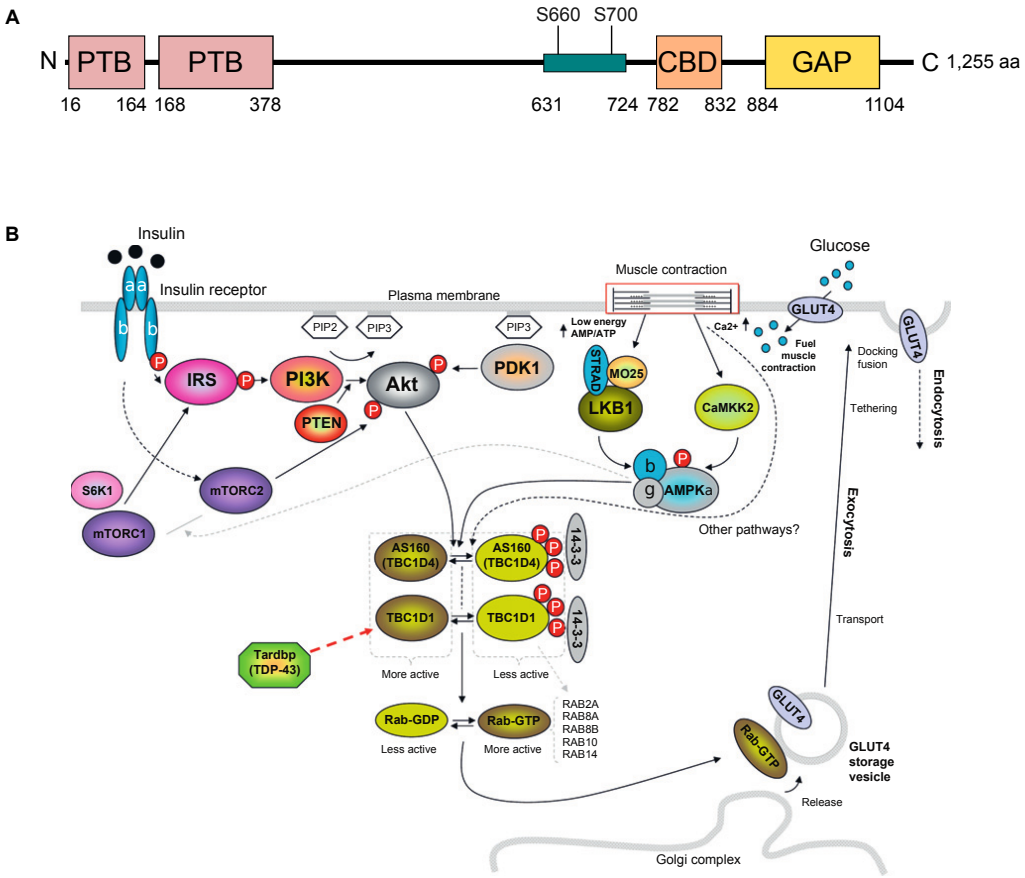


Figure 3.28: **TBC1D1 structure and function.** **A** Annotated functional domains in the long (1255 aa) isoform of TBC1D1 include two protein tyrosine binding (PTB) domains at the N-terminus, a calmodulin-binding (CBD) domain and a GTPase-activating (GAP) domain at the C-terminus. Two reported AMPK-phosphorylation sites (S660 and S700) that lie within the alternative region encoded by alternative exons 12 and 13 (in green) are indicated. Coordinates (aa) of each region are given bellow (modified after [196]; [197]). **B** Schematic representation of the role played by TBC1D1 in regulation GLUT4 translocation in the skeletal muscle [201]. 14-3-3: 14-3-3 protein; AKT: protein kinase B (PKB); AMP: adenosine monophosphate; AMPK: 50 adenosine monophosphate-activated protein kinase; AS160: Akt substrate of 160 kDa; ATP: adenosine-triphosphate; CAMKK2: calcium/calmodulin-dependent protein kinase kinase 2; GLUT4: glucose transporter type 4; GMP: guanosine monophosphate; GTP: guanosine-50-triphosphate; IRS: insulin receptor substrate protein; LKB1: liver kinase B1; MO25: mouse protein-25; mTORc1: mammalian target of rapamycin, in multiprotein complex 1; mTORc2: mammalian target of rapamycin, in multiprotein complex 2; PDK1: pyruvate dehydrogenase lipoamide kinase isozyme 1; PI3K: phosphatidylinositol 3-kinase; PIP2: phosphatidylinositol (3,4,5)-biphosphate; PIP3: phosphatidylinositol (3,4,5)-triphosphate; PTEN: phosphatase and tensin homolog; Rab: rab protein (member of the Ras superfamily of monomeric G proteins); STRAD: STe20-related adaptor; Tardbp: Tat activating regulatory DNA-binding protein; TBC1D1: TBC1 (tre-2/USP6, BUB2, cdc16) domain family, member 1; TBC1D4: TBC1 (tre-2/USP6, BUB2, cdc16) domain family, member 4; TDP-43: TAR DNA-binding protein 43.

Interestingly enough, GO analysis indicated altered GTPase activity upon TDP-43 depletion as associated GO terms were enriched by genes that were either differentially expressed (DEG) or alternatively (AS) spliced (Figure 3.29). It is thus possible that TDP-43 regulates GTPase activity in a coordinated manner, by influencing transcript levels and splice isoform expression of GTPases - TBC1D1 but also others.

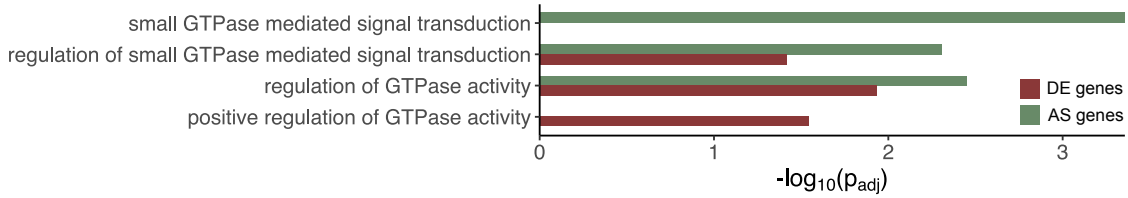


Figure 3.29: **GTPase activity in TDP-43-silenced C2C12 cells.** GO terms (category: biological process) significantly enriched by DE (red) and AS genes (green) genes detected in undifferentiated C2C12 cells suggest disturbance of GTPase activity.

3.2.4 TDP-43 depletion leads to ATP reduction in C1C12 cells

Given that TDP-43 silencing blocked the expression of the long TBC1D1 isoform, which could potentially affect glucose uptake and energetic balance of the cell, we measured ATP production in TDP-43-depleted myoblasts. ATP is generated via glycolysis in the cytoplasm and oxidative phosphorylation in the mitochondria. The relative contribution of each ATP-generating process depends on substrate and oxygen availability, and the energy demand [202]. Herein we demonstrate a 20% reduction in ATP levels of TDP-43-depleted myoblasts as compared to siLUC-transfected or non-treated cells (Figure 3.30). This is consistent with results obtained by Izumikawa *et al.* [20] in MCF-7 human breast cancer cell line, which they claim was due to impaired mitochondrial homeostasis, as they have shown for TDP-43 to regulate processing of mitochondrial transcripts.

Substrate usage and activation of a specific energy-generating pathway can be regulated at the enzymatic level, for example by varying expression of glycolytic regulators or by generation of splicing isoforms exhibiting differential functionality (as in the case of two pyruvate kinase (PK) isoforms, one of which promotes aerobic glycolysis and the other oxidative phosphorilation [203]).

Further investigation would be needed to understand whether lack of the bigger TBC1D1 protein isoform could account for decreased ATP production. Nevertheless, there is some evidence indicating this might be the case. Firstly, TDP-43 has been recently recognized as a regulator of glucose homeostasis *in vivo*, potentially via effects on TBC1D1 [194]; [114]. Secondly, reduced inclusion of alternative exons 12 and 13 within *Tbc1d1* was observed in various type of human cancers (The Cancer Genome Atlas project [204]). In fact, the evidence linking metabolic switch to changes in cell state was first provided by Warburg [205] upon investigation

of cancer cells. However, it later became apparent that metabolic reprogramming plays a role in other processes such as cell differentiation and has been also associated with muscle disorders [206]; [207].

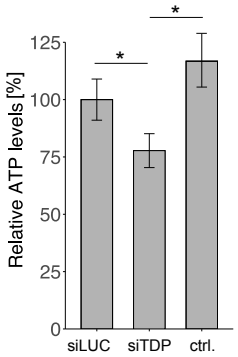


Figure 3.30: **ATP levels in TDP-43-silenced C2C12 myoblasts.** Undifferentiated cells were transfected using siRNA (siLUC or siTDP) on two consecutive days and ATP levels in lysed cells (intracellular and extracellular ATP) were assessed by a luminometric assay 24 h hours after the second transfection. A non-treated control (ctrl.) was included as well to rule-out the effect of Hiperfectamine treatment on cellular ATP production. Fluorescence measured in each well was normalized to the number of cells determined by Hoechst-stained nuclei count. The plot shows the mean and the standard deviation of three independent experiments. p-values were generated by two-way ANOVA followed by Tukey's test.

3.2.5 Conclusions

- Expression of *Tardbp* increases 24 h after differentiation onset in C2C12 cells and returns to pre-differentiation levels at day 4.
- Splicing transitions occur along C2C12 differentiation.
- Increased expression of the long *Tbc1d1* is associated with C2C12 differentiation.
- TDP-43 controls inclusion of two cassette exons (12 and 13) in undifferentiated C2C12 myoblasts and myotubes.
- TDP-43 loss does not impact the overall *Tbc1d1* expression in undifferentiated C2C12 and NSC34 cells.
- Two AMPK-dependent phosphorylation sites (S660 and S700) are found within the sequence encoded by exons 12 and 13 in the long TBC1D1 isoform in mouse.
- TDP-43 knockdown results in 20% reduction in ATP levels in C2C12 cells.

3.3 Conservation of TDP-43 targets in humans

TDP-43 is well conserved across species. High sequence similarity of *TARDBP* gene in 4 taxa - human, mouse, *C. Elegans* and *D. Melanogaster* - implies on essential functions TDP-43 provides in those organisms [208]. As a consequence, several animal models (i.e., *C. Elegans*, *D. Melanogaster*, mouse) [209]; [136] or cellular models derived from different species (mouse, human, yeast [210]) have been employed to study structural features of TDP-43 as well as its function, with the general intention to explain its contribution in the development of TDP-43-proteionopathies in humans.

Available mouse models mimicking TDP-43-associated neurodegeneration are usually able to recapitulate some of pathological phenotypes commonly observed in ALS/FTLD patients, but not all of them [150]; [211]; [57]. In order to establish good disease models [212], one needs to understand similarities and differences in the performance of TDP-43 across species. One possible approach to address this would be by assessing binding/splicing performance of TDP-43 derived from different species on the same mRNA target constructs (minigenes), or alternatively, by comparing subsets of orthologous mRNA transcripts targeted by this same protein in each organism. Considering evolution of intronic/exonic regulatory sequences (enhancers/silencers) [213]; [214]; [153], it is very likely that unique subsets of mRNAs would be targeted by TDP-43 in each species.

In fact, a study comparing *in vitro* splicing activity of human TDP-43 and that from *D. melanogaster* and *C. elegans* showed that, notwithstanding high sequence similarity of RRMs, there is a variation in splicing regulation provided by TDP-43 derived from different species [149]. Furthemore, exons 17b within mouse and human *Sort1/SORT1* transcripts are repressed to a different exent in human cells, due to presence of additional splicing enhancers in mouse transcript [153], suggesting that evolution of regulatory sequences in target transcripts determines whether a particular alternative exon would be subject to TDP-43 control in a given species. Taken together, those findings suggest that conservation of TDP-43 regulatory function depends on 1) intrinsic properties of the protein in each species, 2) the evolution of regulatory sequences in target transcripts as well as 3) evolution of other *trans*-regulatory factors that influence performance of TDP-43 in a cooperative manner.

Transcriptome-wide, mice expressing human WT TDP-43 (and as a consequence, with reduced levels of endogenous TDP-43 due to autoregulation) displayed altered splicing profiles (824 AS events identified by microarray) [57], demonstrating that human TDP-43 fails to control inclusion levels of some alternative exons, which are otherwise regulated by endogenous (mouse) TDP-43. Nonetheless, those transgenic mice (expressing WT human TDP-43) remained phenotypically normal, suggesting that human TDP-43 is able to provide some essential regulatory activity in mouse. Even though one study reported for a great portion of TDP-43-regulated exons in mouse to have a prior evidence of alternative splicing in humans [8], it has not been

directly assessed whether those alternative exons are subject to TDP-43 regulated inclusion in humans.

We still lack a clear understanding of how well TDP-43 activity in mouse recapitulates that of human TDP-43 in human cells, and to what extent target transcripts (orthologues) would be shared between mouse and human.

To address this issue we first assessed conservation levels of sequences regulated by TDP-43 across species. We further tested, which of TDP-43-controlled events commonly detected in both mouse cell lines (C2C12 and NSC34) appear to be regulated by TDP-43 in human cells representing neurons (SH-SY5Y) and muscle (RH-30). Finally, we compared subsets of TDP-43 targets (RNA-seq) identified in human cell line (SH-SY5Y) with those found in mouse (C2C12, NSC34).

3.3.1 Contributions

- Neva Škrabar and Robert Bakarić performed comparative sequence analysis and calculated PhyloP scores.
- Sara Cappelli silenced TDP-43 in SH-SY5Y cells and extracted RNA. RNA samples were sequenced by Novogene (Beijing, China).

3.3.2 Sequence conservation

Computational techniques that determine the evolution rate of certain sequences allow for detection and characterization of functional elements. Sequence elements, which evolve slower then it would be expected under natural drift (conservation, as opposed to fast evolving sequences - acceleration), are assumed to be of functional importance providing conserved functions. Comparative sequence analysis of aligned mammalian genomes has therefore been employed to identify various functional elements like protein-coding genes or sequences coding for functional RNAs [129].

As it has been previously demonstrated that AS events specific to neural tissues display slower evolution rate than AS events detected in other tissues [215]; [42]; [26], we aimed to compare sequence conservation of TDP-43-regulated transcripts detected in (mouse) muscle and neuronal cells.

PhyloP (phylogenetic p-value) score, a measure that has been employed herein, predicts evolutionary *per base* conservation at individual alignment sites. The score represents the negative logarithmic value of the p-value (-log(p-value)) under a null hypothesis of neutral evolution using a 59-way comparison between mouse and other species. Positive scores indicate conservation while negative scores indicate fast evolution at a certain genomic position [129].

Importantly, we compared conservation of sequences subject to TDP-43 regulation either at the level of overall transcript abundance (DEG) or alternative splicing (AS). Yet we did not

attempt to investigate functional elements that might possibly influence the activity of TDP-43 in each tissue.

TDP-43-dependent DEG are not particularly well conserved

We first tested whether TDP-43-dependent differentially expressed genes (DEG) detected in C2C12 and NSC34 might differ with regard to their evolutionary sequence conservation. This was done by computing average *per gene* PhyloP scores, which average PhyloP values obtained for individual bases within each gene. This approach measures conservation of the entire sequence from gene start site to the end, and does not distinguish between functional elements of the gene (i.e., introns, exons, 5' and 3' UTR).

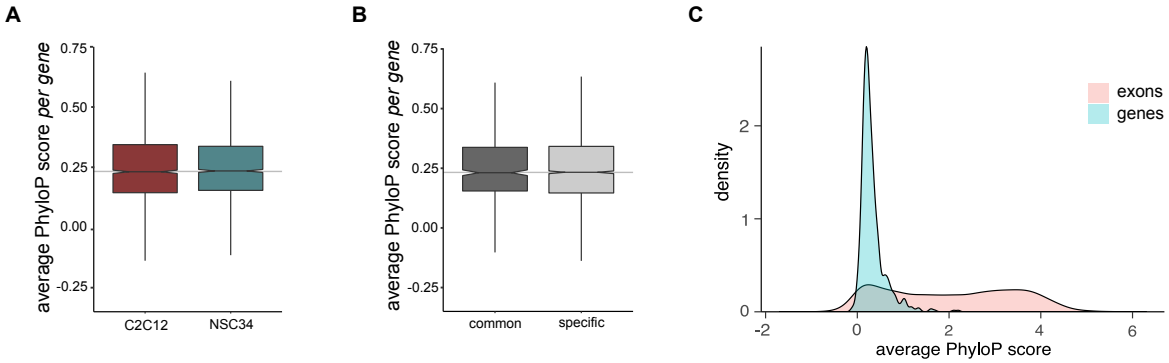


Figure 3.31: **Conservation of DEGs detected in C2C12 and NSC34. A** Average PhyloP conservation scores *per gene* plotted as box plots show TDP-43-regulated DEG detected in C2C12 (2,325) and NSC34 (2,324) are equally well conserved across species (p-value = 0.48). **B** Average PhyloP conservation scores *per gene* plotted as box plots show TDP-43-regulated genes detected in both cell lines (630) are not better conserved across species than those detected in one cell line exclusively (3,389) (p-value = 0.90). p-values was generated using Wilcoxon signed-rank test, boxes represent the median with the first and the third quartile, whiskers represent values that fall within 1.5-times the interquartile range and the grey line represents the average *per gene* PhyloP score of all genes in mouse genome. **C** The density plot shows bimodal and wider distribution in in PhyloP scores considering each exon within DEGs compared to *per gene* PhyloP scores. This means that within each DEG, there are two types of exons - some that are highly conserved and some that show low conservation.

Comparing the median *per gene* PhyloP scores of all DEGs identified in C2C12 and NSC34, no difference was observed. Although gene expression in brain has diverged less than in other organs [216], TDP-43-regulated transcripts in both tissues seem to display a similar evolution rate. The median PhyloP score of TDP-43-controlled transcripts (either in C2C12 or NSC34) was in fact similar to the median PhyloP score of all genes (Figure 3.31a, grey line), meaning that TDP-43-regulated transcripts are not particularly well conserved in respect to an average gene in the mouse genome (Figure 3.31a).

We further compared overlapping TDP-43 targets (detected in both, C2C12 and NSC34) with cell-type-specific DEGs. Commonly regulated genes showed the same level of conservation across species as those regulated by TDP-43 in a tissue-specific fashion (Figure 3.31b).

However, it is important to note PhyloP scores calculated *per exon* are generally higher than those determined *per gene*, as they do not include intronic sequences. Furthermore, there is a high variation in conservation level of individual exons within genes (Fig 3.31c).

TDP-43-dependent AS sequences detected in NSC34 are better conserved across species

Given that transcripts whose overall abundance depends on TDP-43 in mouse are not particularly well conserved across species (Figure 3.31), we next investigated the extent of evolutionary conservation of TDP-43-mediated alternative exons (herein, I will refer to them as *alternative sequences* instead of *alternative exons*, as retained introns, alternative 5'- and 3'-sequences were analyzed along with cassette exons). The average PhyloP score *per exon* was calculated for each alternative sequence controlled by TDP-43.

In fact, TDP-43-controlled alternative sequences show higher conservation than an average exon in mouse genome. Furthermore, TDP-43-regulated sequences in NSC34 are better conserved than those detected in C2C12 cells (p-value < 0.001) (Figure 3.32a), in line with the notion that neuronal AS events in general are more often conserved between vertebrates than AS events specific to any other organ [215]. This would imply on biological importance of alternative splicing characteristic to neuronal tissue and as a consequence, slower evolution rate of alternatively spliced exons in neurons [26]. Most of the difference between conservation of alternative sequences identified in C2C12 and NSC34 (Figure 3.32a) is in fact attributed to PhyloP scores of skipped exons (p-value < 0.001) (Figure 3.32b), as they represent the most frequent AS event regulated by TDP-43 (Figure 3.13a). Compared to alternative exons, TDP-43-dependent introns (i.e., retained introns) are less conserved, as intronic sequences are well-known to evolve faster [129]. However, TDP-43-controlled alternative exons show conservation scores similar to adjacent constitutive exons (i.e., flanking exons).

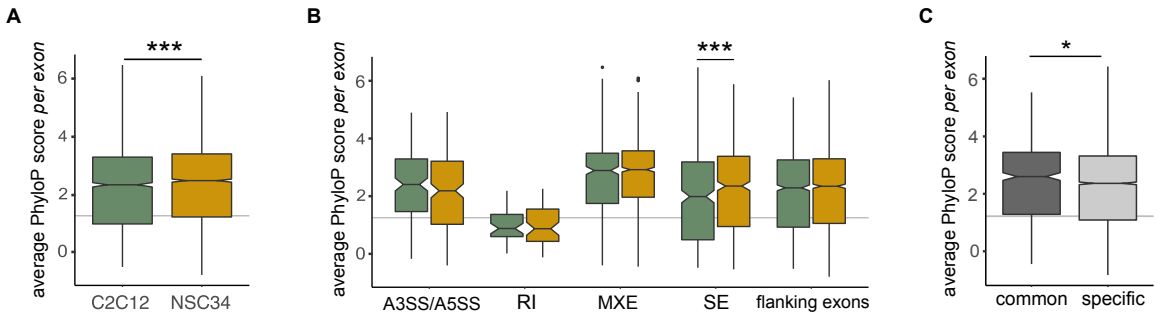


Figure 3.32: **Conservation of alternatively spliced sequences detected in C2C12 and NSC34.** **A** Average PhyloP conservation scores *per exon* plotted as box plots show TDP-43-regulated alternative sequences detected in NSC34 cells (4,281) are better conserved across species than those detected in C2C12 cells (2,372) (p-value < 0.001). p-value was generated using Wilcoxon signed-rank test. **B** Average PhyloP conservation scores *per exon* of TDP-43-regulated alternative sequences stratified by event type (SE, MXE, RI, A3SS, A5SS). The difference (p-value < 0.001) among all groups was tested using Kruskal-Wallis chi-squared test and pair-wise comparison was conducted using Wilcoxon singed-rank test. green - AS events in C2C12, yellow - AS events in NSC34. **C** Average PhyloP conservation scores *per exon* plotted as box plots show TDP-43 regulated alternative sequences detected in both cell lines (634) are better conserved across species than those detected in one cell line exclusively (6,019) (p-value $= 0.02$). p-values was generated using Wilcoxon signed-rank test, boxes represent the median with the first and the third quartile, whiskers represent values that fall within 1.5-times the interquartile range and the grey line represents the average *per exon* PhyloP score of all (annotated) exons in mouse genome.

Comparing common and tissue-specific AS events, alternative sequences subject to TDP-43 regulation in both tissues on average show higher level of conservation than sequences regulated by TDP-43 in a cell-type-specific manner (Figure 3.32c).

3.3.3 Conservation of TDP-43-regulated splicing in human cell lines

We started our comparison of tissue-specific TDP-43's performance using mouse cell lines representing muscles and neurons (see an earlier chapter). Mouse cells were chosen over human ones due to the lack of an appropriate and well-established cell line derived from human skeletal muscle.

In the previous subsection we discussed the evolutionary conservation of TDP-43-regulated transcripts and exons, however, we subsequently tested whether TDP-43-mediated splicing events detected in mouse are subject to TDP-43 regulation also in human cells. We know that sequence similarity *per se* does not guarantee for an orthologous region to be dependent on TDP-43 in both species. Beside on the exonic sequence, incorporation of a particular region into mature mRNA depends on *cis*-regulatory motifs (conservation of those was not investigated herein) and *trans*-acting factors [42]; [52]; [215]; [217]. In fact, most of species-characteristic

alternative splicing was attributed to mutations in *cis*-regulatory sequences and *trans*-acting factors [52], rather than mutations accumulated within alternatively-spliced sequences.

Human cell lines representing neuronal (SH-SY5Y) and muscle cells (RH-30)

We investigated conservation of TDP-43-dependent splicing using two human cell lines, one representing neuronal cells and the other one serving as proxy of skeletal muscles.

SH-SY5Y, a human neuroblastoma cell line [218]; [219] has been widely employed as a neronal cellular model to study general features of TDP-43 (e.g., structure, binding) [136]; [220] as well as its involvement in neurodegeneration [7]; [221]. RH-30, on the other hand, is a rhabdomyosarcoma cell line, that exhibits some structural and electrophysiological characteristics of mammalian skeletal muscle [222]; [223]. Although it has not been established in myodegenerative disease research, it has been previously used to investigate FUS-dependent gene expression of muscle characteristic transcripts [224].

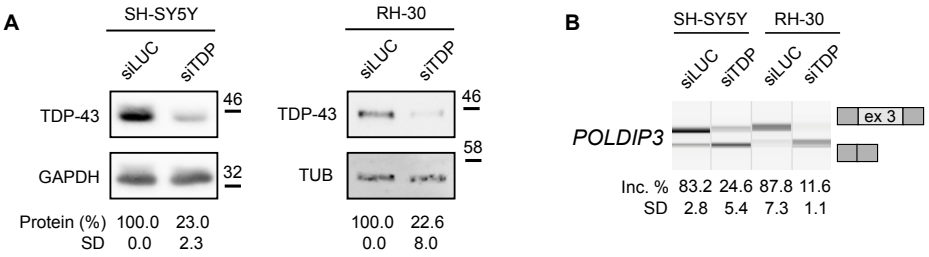


Figure 3.33: **Depletion of TDP-43 in SH-SY5Y and RH-30 cells. A** The western blot shows efficient reduction of TDP-43 in SH-SY5Y and RH-30 cells upon siTDP transfection. siLUC-transfected cells were used as a control. TDP-43 expression was either normalized against GAPDH or tubulin. **B** Altered splicing of *POLDIP3* upon TDP-43 knockdown was validated using semi-quantitative RT-PCR. Inc. % - the abundance of the longer splice isoform.

We transected SH-SY5Y and RH-30 using siTDP and achieved efficient TDP-43 depletion (Figure 3.33a). In both cell lines, lack of TDP-43 resulted in increased skipping of exon 3 within*POLDIP3*, which functionally reflected in TDP-43 loss (Figure 3.33b). TDP-43-dependent processing of *POLDIP3* and its mouse orthologue *Poldip3* has been previously shown to be conserved in mouse and human [151]; [8]; [97].

Morphologically, TDP-43-depleted cells did not differ from siLUC-transfected controls 48 h after transfection.

Conservation of alternative splicing validated in human cell lines

We next investigated possible splicing orthology on a subset of cassette exons, inclusion of which has been found to be TDP-43-mediated in mouse cell lines. To do so, we first referred to the

VastDB database [170] and Ensembl [130], and subsequently validated putative isoform switch in TDP-43-depleted SH-SY5Y and RH-30 cells using semi-quantitative RT-PCR.

There are three possible outcomes regarding the conservation of TDP-43-regulated splicing between mouse and human;

- an alternative exon found in mouse does not even have an orthologoues exon in human (Figure 3.34);
- an orthologous exon does exists in humans, nevertheless, its inclusion is not TDP-43-dependent (Figure 3.35);
- both, an alternative exon identified in mouse as well as its human orthologue are subject to TDP-43-regulation (Figure 3.36).

Fam220a, *Tesk2*, *Tmem2* and *Sapcd2* are four exemplary transcripts, whose splicing is controlled by TDP-43 in mouse but they lack orthologous exons in humans. Orthology at the transcript level does not guarantee for the orthology of each exon. Low PhyloP scores determined for alternative exons found in mouse (0.0334, 0.1070, 0.2206, 0.0679 for *Fam220a* exon 2, *Tesk2* exon 2, *Tmem2* exon 2 and *Sapcd2* exon 2, respectively) imply on poor sequence conservation of those exons across species.

Mouse *Fam220a* has two human orthologues, *SMIM10L2A* and *SMIM10L2B*, none of which has a prior evidence of alternative splicing. Exon 2 of *Fam220a* is repressed by TDP-43 in mouse, yet it absent in human orthologue of this gene.

The same holds true for exon 2 within *Tesk2*, that in repressed by TDP-43 in mouse C2C12 cell line. This exon does not appear in its human orthologue *TESK2*.

Mouse *Tmem2* (gene synonym *Cempi2*) and its human counterpart *TMEM2* show very low level of similarity. Exon 2 of *Tmem2*, the inclusion of which is enhanced upon TDP-43 depletion, is not annotated as an exon in humans, and has evolved to become part of an intron. Nonetheless, TDP-43 binding was detected in the intronic sequence of human *TMEM2* transcript.

Likewise, exon 2 of *Sapcd2* makes part of the first intron in human *SAPCD2* transcript.

Importantly, TDP-43 has recently been described to provide a previously unknown function, i.e., to repress splicing of novel type of alternative exons called cryptic exons [24]; [140]; [100]. In the presence of TDP-43, cryptic exons do not get spliced into the mature transcript (and were thus not recognized as exons), as they would often disrupt reading frame and lead to NMD. Despite the fact that exon 2 of *Tmem2* and exon 2 of *Sapcd2* were not predicted to be conserved in humans, we tested splicing of human *TMEM2* and *SAPCD2* for possible emergence of crypric exons following TDP-43 knockdown (Figure 3.34). To do so, we designed primers in the upstream and downstream constitutive exons flanking the intronic regions, in which a TDP-43-regulated alternative exon was detected in mouse. However, no cryptic exons were detected upon TDP-43 loss in SH-SY5Y cells.

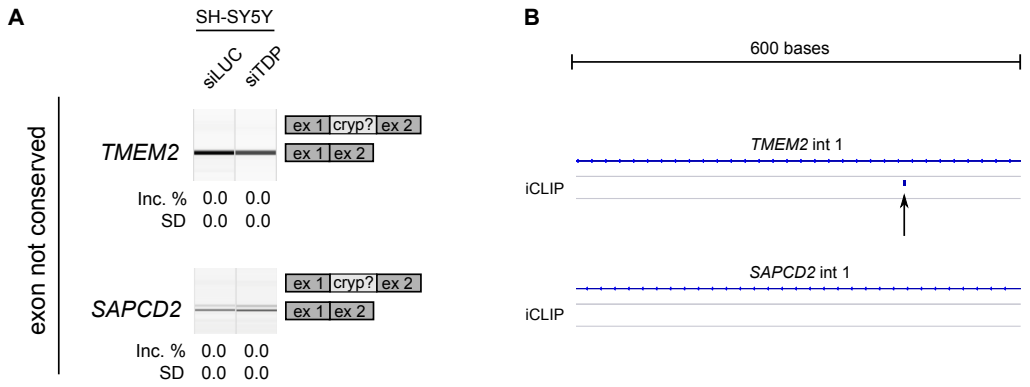


Figure 3.34: **Alternative exons found in mouse *Tmem2* and *Sapcd2* do not have orthologoues in humans. A** Semi-quantitative RT-PCR conducted in TDP-43 silenced SH-SY5Y cells and corresponding controls is shown along with the quantification of splicing changes (% of alternative exon inclusion). Primers were designed in constitutive exons flanking the intronic regions, in which an alternative exon was predicted to emerge (see the exact transcript numbers in Table 2.4). **B** Schemes show TDP-43 binding sites (arrow) identified by iCLIP analysis in SH-SY5Y cells [7] in the vicinity of putative alternative exons represented on panel **A**.

Next, we found some examples of cassette exons, which are conserved between mouse and human, yet their inclusion is regulated by TDP-43 in a species-specific fashion, i.e., only in mouse but not in human cell lines (Figure 3.35).

Among those, exon 3 of *Rgp1* displays almost constitutive inclusion ($> 90\%$) in mouse and also in its human orthologue *RGP1* [170]. Furthermore, exons 3 display high sequence similarity (94.2%). While TDP-43 depletion leads to exon 3 skipping in mouse, its inclusion does not appear to be TDP-43-dependent in human cell lines. In fact, no TDP-43 biding sites were identified in the proximity of exon 3 in SH-SY5Y cells.

Inclusion of exon 3 in mouse *Dnajc5* and its human orthologue *DNAJC5* is predicted to cause NMD. Exon 3 shows variable levels of inclusion across tissue in mouse ($< 60\%$) and human ($< 40\%$). While the absence of TDP-43 in mouse leads to enhanced inclusion (and as a result transcript downregulation), this exon is not controlled by TDP-43 in human cell lines, despite the fact we detected TDP-43 binding within that exon along and saw a downregulation of *DNAJC5* transcript in TDP-43-silenced SH-SY5Y cells.

Exons 5 of mouse *Nfya* and human *NFYA* are alternatively included into mature transcripts. The extent of exon 5 incorporation varies very much between tissues (0-100%), yet its inclusion is not controlled by TDP-43 in human cell lines tested. Consistently, no TDP-43 cross-linking was detected in the proximity of exon 5 in SH-SY5Y cells.

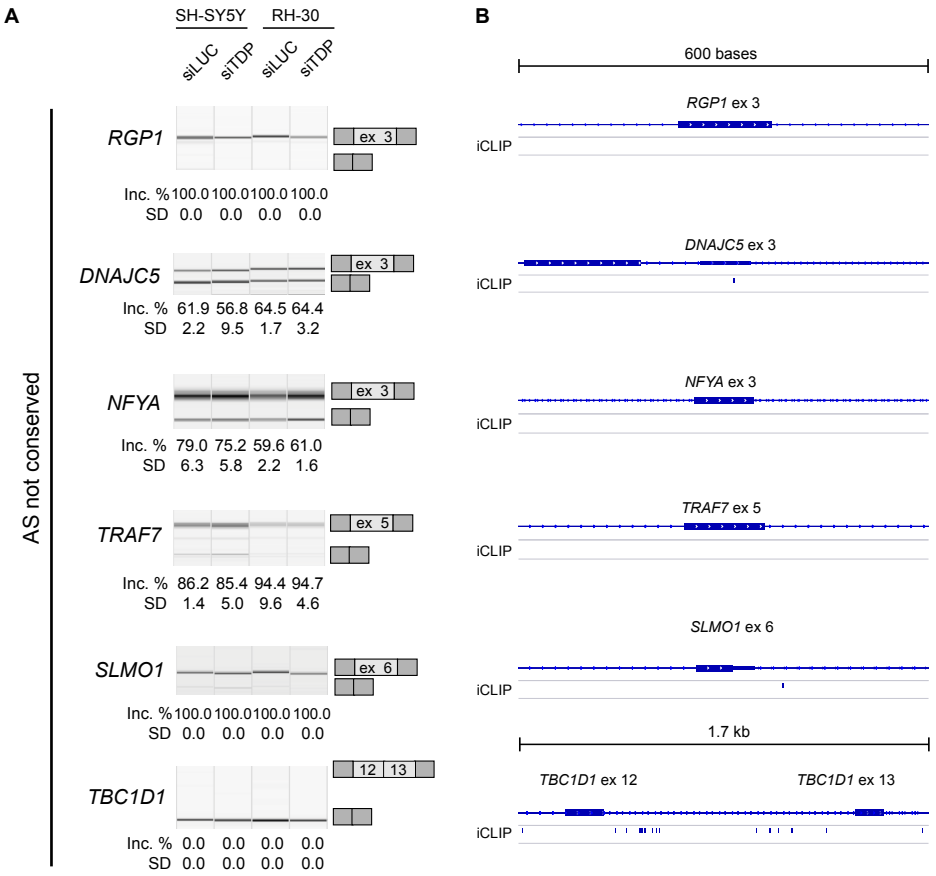


Figure 3.35: **Alternative exons found in mouse are not controlled by TDP-43 in humans. A** Semi-quantitative RT-PCR conducted in TDP-43 silenced SH-SY5Y cells and corresponding controls is shown along with the quantification of splicing changes (% of alternative exon inclusion). Primers were designed in constitutive exons flanking the putative alternative exon (see the exact transcript numbers in Table 2.4). **B** Schemes show TDP-43 binding sites where identified by iCLIP analysis in SH-SY5Y cells [7] in the vicinity of alternative exons represented on panel A.

Exon 5 within mouse *Traf7* displays alternative inclusion ($> 70\%$) across tissues. Whereas in mouse TDP-43 was seen to enhance the inclusion of this exon, it does not control its incorporation in human *TRAF7*. In fact, exon 5 is constitutively included into mature *TRAF7* transcript in all tissues. No TDP-43 binding sites were found in flanking regions in SH-SY5Y cells.

Exon 6 of mouse *Slmo1* (gene synonym *Prelid3a*), which in mouse has some evidence of alternative inclusion in different tissues ($> 60\%$) seems to be constitutively included in human *SLMO1*. Unlike in mouse, depletion of TDP-43 in human cells does not cause exon skipping, although TDP-43 biding was detected downstream exon 6 in SH-SY5Y cells.

Inclusion of exons 12 and 13 within *Tbc1d1* seemed to be a muscle-specific TDP-43-dependent event (as it was discussed in the previous chapter). Importantly, enhanced incorporation of those

two exons has been associated with cell differentiation (either muscle of neuronal) [27], thus it is likely that the long isoform is poorly expressed in undifferentiated cells we used (C2C12 and NSC34). Even though these two alternative exons show strong inclusion in human skeletal muscles and different brain regions (> 60%) as compared to other tissue, we did not detect the long *TBC1D1* isoform (including exons 12 and 13) neither in SH-SY5Y nor in RH-30 cells. Since SH-SY5Y and RH-30 cells were used in their undifferentiated state, they were expected to lack the long *TBC1D1* variant. Importantly, numerous TDP-43 binding sites were identified in flanking introns, thus we speculate that TDP-43 might nevertheless regulate inclusion of these two cassette exons in humans, but it does so only in differentiated cells or mature tissue.

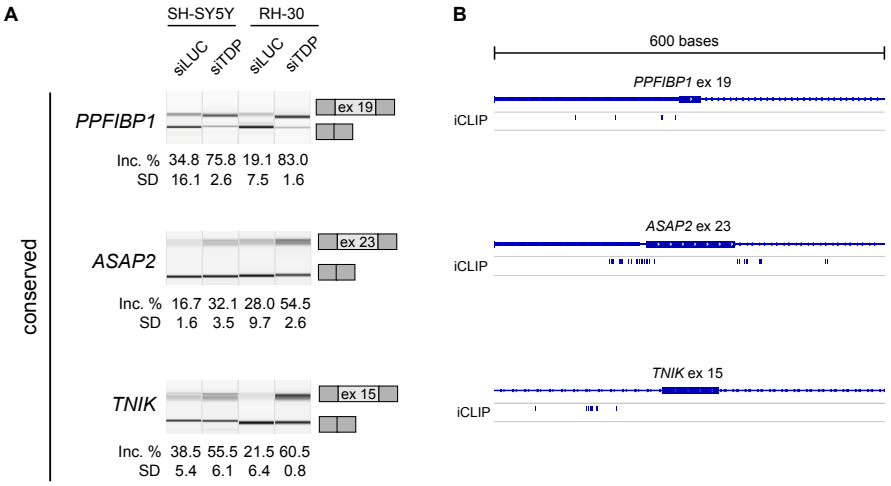


Figure 3.36: **Alternative exons targeted by TDP-43 in mouse and human cell lines. A** Semi quantitative RT-PCR conducted in TDP-43 silenced SH-SY5Y cells and corresponding controls is shown along with the quantification of splicing changes (% of alternative exon inclusion). Primers were designed in constitutive exons flanking the putative alternative exon (see the exact transcript numbers in Table 2.4). **B** Schemes show TDP-43 binding sites that were identified by iCLIP analysis in SH-SY5Y cells [7] in the vicinity of alternative exons represented on panel A.

Finally, we identified splicing events, which are consistently regulated by TDP-43 in mouse and human (Figure 3.36).

Both, exon 4 of *Ppfibp1* as well as its human counterpart exon 19 of *PPFIBP1* gene show tissue-characteristic inclusion levels, and are consistently repressed by TDP-43 in all cell lines tested. Moreover, there is the evidence of TDP-43 biding in the upstream intron in SH-SY5Y cells.

Exon 24 of *Asap2* shows variable extent of inclusion across mouse tissues (< 60%) and the same holds true for its human orthologue within *ASAP2* transcript (0-100%). Since TDP-43 binding sites was detected within the alternative exon itself as well as in flanking introns, that is indicative of direct repression of exon 23 due to TDP-43 binding.

TDP-43-controlled splicing of exon 14 within mouse *Tnik* and exon 15 within human *TNIK* has been previously validated in neuronal cells [7] and HEK-293 [225]. While in mouse we only detected inclusion of exon 14 upon TDP-43 depletion in muscles, exon 15 of *TNIK* transcript appears to be regulated by TDP-43 in both human cell lines tested - SH-SY5Y and RH-30. Many TDP-43 binding sites were identified downstream the alternative exon in SH-SY5Y, suggesting a direct repression of exon 15 inclusion by TDP-43.

We have shown, using a set of exemplary splicing events, that TDP-43 does not necessarily regulate splicing of orthologoues exons, even when an exon orthology exists between mouse and human at the level of the sequence itself. This might be explained by divergence of *cis*-regulatory motifs across species as well as expression of *trans*-acting factors (i.e., RBPs) that either enhance or block TDP-43 binding/ activity.

Generally speaking, multiple TDP-43 binding sites are usually identified in the proximity of TDP-43-dependent alternative exons (*PPFIBP1*, *ASAP2*, *TNIK*), in contrast to none (*RGP1*, *NFYA*, *TRAF7*) or few (*DNAJC5*, *SLMO1*) within/nearby exons not subject to TDP-43 regulation. While TDP-43 binding alone is a good predictor of whether TDP-43 influences the usage of a given splice site, TDP-43 cross-linking is neither necessary (indirect regulation or TDP-43 binding in distal intronic regions, which nevertheless influences splicing) nor sufficient for inclusion/skipping of a particular exon [7]. The probability of TDP-43's binding to a given mRNA increases with the length of UG clusters. This increased affinity for long biding sites makes it different from CELF2, which preferentially recognizes shorter UG repeats. TDP-43 cooperatively binds to multiple bindings sites, as each of the two RRMs recognizes 4 nuecleotides and it usually binds in a homodimer state [7].

3.3.4 Small percentage of TDP-43-regulated events detected in mouse and human cell lines

After having looked at the sequence conservation of TDP-43-regulated transcripts and alternatively spliced regions (Figure 3.31 and 3.32) as well as having investigated the regulation of exemplary orthologous alternative exons in mouse and human (Figure 3.34-3.36) using RT-PCR, we next conducted a transcriptome-wide comparison of events regulated by TDP-43 in mouse (C2C12 and NSC34) and human (SH-SY5Y) cell lines.

For each event (DEG or AS, respectively) found to be regulated by TDP-43 in mouse cell lines, we searched for eventual orthology in humans (either at the gene or exonic level) and looked whether an orthologous sequence is be subject to TDP-43 control in SH-SY5Y cells.

Transcripts regulated by TDP-43 in mouse and human cells

Only 811 genes were detected as DE in TDP-43-silenced SH-SY5Y cells (compared to 2,325 and 2,324 in C2C12 and NSC34, respectively) at a similar sequencing depth (130 million reads,

paired-end), possibly due to lower degree of silencing (around 55% depletion of TDP-43 at the RNA level).

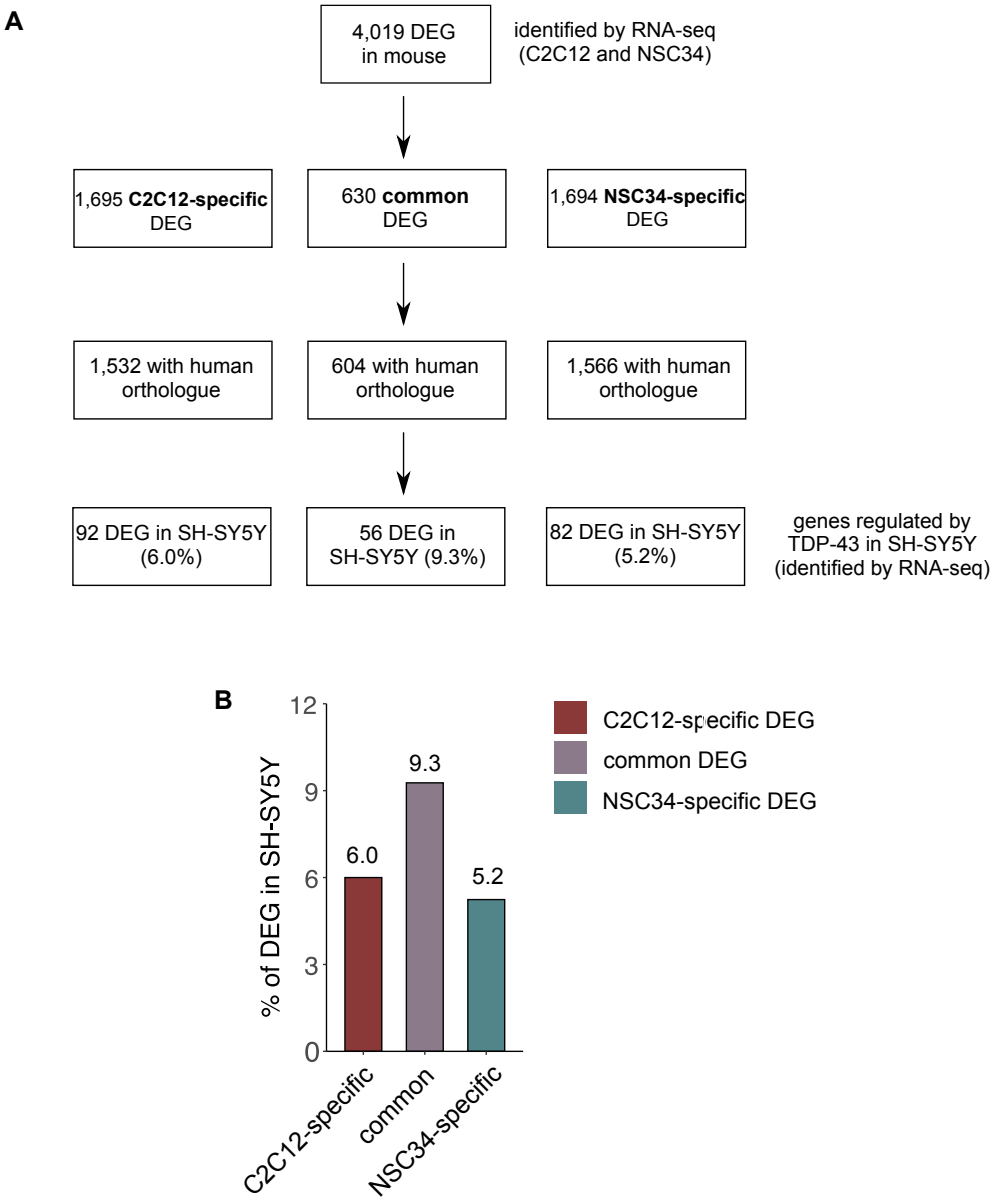


Figure 3.37: **Schematic representation of conservation search for DEGs regulated by TDP-43 in mouse and human cell lines. A** Of 4,019 genes found to be TDP-43-dependent in mouse cell lines, 3,702 have at least one orthologous gene in humans (BioMart database [226]). The proportion of human orthologues, which are indeed regulated by TDP-43 in SH-SY5Y is shown in **B**. In total, 811 genes were detected as DE in SH-SY5Y cells upon depletion of TDP-43 at $p_{adj} < 0.05$.

For each transcript, whose abundance was TDP-43-dependent in mouse cell lines, we first

searched whether an orthologues gene exists in human (one or more) (Figure 3.37a).

Transcripts with human orthologues, whose expression levels were affected by TDP-43 knockdown in both mouse cell lines, appear to be slightly more likely regulated by TDP-43 in human neuroblastoma cell line SH-SY5Y than those undergoing TDP-43-mediated regulation in C2C12 or NSC34 exclusively (6.0%, 9.3%, 5.2% for C2C12-specific, common and NSC34-specific targets, respectively) (Figure 3.37b). Interestingly enough, NSC34 and SH-SY5Y are equally distant in terms of TDP-43 regulation than SH-SY5Y and C2C12.

AS events regulated by TDP-43 in mouse and human cells

Inspecting cross-species conservation of TDP-43-dependent splicing, the same alternative splicing pipeline was applied to identify AS in our mouse and human datasets (rMATS, v4.0.2). In general, alternative splicing is reported to be more frequently exploited in human cells than in mouse [54] irrespective of cell-type, and twice as frequent in all analyzed primate organs relative to equivalent organs of other species [215]. Using the same splicing tool (at FDR < 0.01), a markedly higher number of AS events was detected in SH-SY5Y cells (5,935) compared to C2C12 or NSC34 (730 or 1,270, respectively).

Conversion of mm10 coordinates to the hg38 build using liftOver tool (UCSC) allowed a direct comparison between orthologoues sequences regulated by TDP-43 in mouse and human genome (Figure 3.38a). Some TDP-43-controlled sequences identified in mouse could not be converted to orthologous positions in the human genome as they evolved heavily since divergence of two species (as in the case of alternative exons within *Tmem2* and *Sapcd2* (Figure 3.34)).

Of those for which an orthologous position could be found in the hg38, 22 sequences (19.5%) commonly controlled by TDP-43 in both mouse cell lines were subject to TDP-43 regulation also in SH-SY5Y, compared to 136 (17.0%) and 227 (14.6%) sequences regulated by TDP-43 in a cell-type-specific fashion (Figure 3.38a and 3.39).

TDP-43-dependent events conserved between mouse and human belong to all splicing categories (Figure 3.38b). Interestingly enough, TDP-43 appears to control retention of some orthologoues introns in both species. Results obtained from the comparison of the two RNA-seq datasets (TDP-43-silenced mouse and human cells) are in good agreement with what we have previously demonstrated (Figure 3.34–3.36) on a smaller set of events using VastDB conservation predictions [170] and RT-PCR validation of TDP-43-dependent splicing in human cell lines (SH-SY5Y and RH-30).

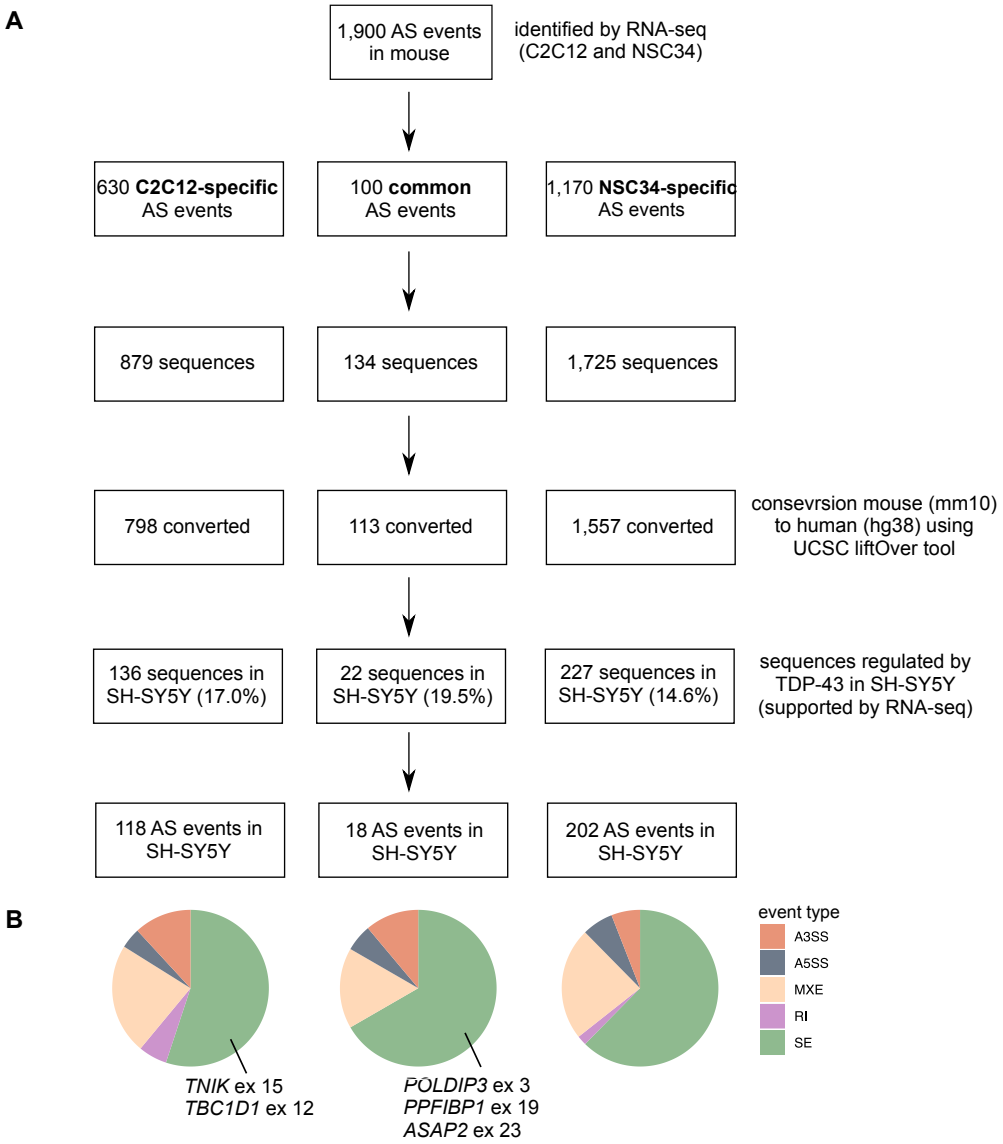


Figure 3.38: **Schematic representation of conservation search for AS sequences regulated by TDP-43 in mouse and human cell lines. A** Coordinates of TDP-43-regulated sequences detected in mouse cell lines (aligned to mm10) were converted to orthologous positions in the hg38 using liftOver tools. **B** Pie charts show distribution of conserved TDP-43-regulated AS events by event type. The total number of events is shown in the box above. *TNIK* exon 15 and *TBC1D1* exon 12 are two C2C12-specific events regulated by TDP-43 in SH-SY5Y cells. *POLDIP3* exon 3, *PPFIBP1* exon 19 and *ASAP2* exon 23 undergo TDP-43-dependent splicing in C2C12 and NSC34 as well as in SH-SY5Y cells.

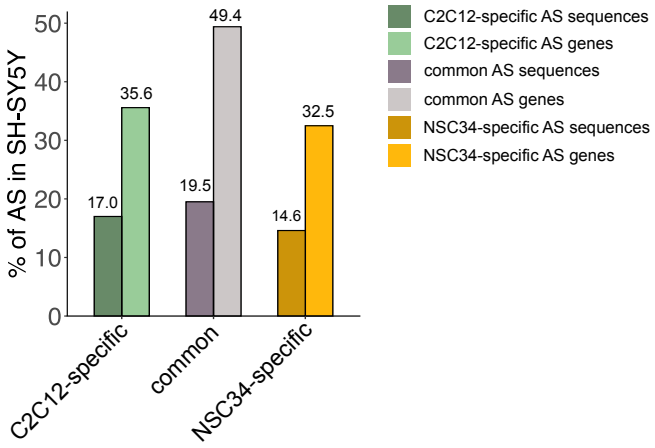


Figure 3.39: **Proportion of TDP-43-controlled AS events identified in mouse that are subject to TDP-43 regulation in human SH-SY5Y cell line.** Bar plot shows the percentage of TDP-43-regulated alternative sequences detected in mouse, orthologues of which are controlled by TDP-43 in SH-SY5Y cells, along with the proportion of AS genes identified in mouse undergoing TDP-43-dependent splicing in SH-SY5Y. The percentage of orthologoues events found to be TDP-43-controlled in the SH-SY5Y is given on the plot.

The observation that a similar portion of TDP-43-regulated sequences identified in muscle and neuronal mouse cell line (17.0% and 14.6%, respectively (Figure 3.38a and 3.39)) is mediated by TDP-43 in human neuroroblastoma cell line came as a surprise, as one would expect a bigger overlap between NSC34 and SH-SY5Y for both cell lines being of neuronal origin.

Although NSC34 and SH-SY5Y represent *bona fide* (mouse and human) cellular models to study pathomechanisms in ALS, one needs to consider differences between the two; SH-SY5Y is a neuroblastoma cell line [218]; [219], while NSC34 was established by fusion of mouse neuroblastoma cells with primary mouse embryonic stem cells enriched in motor neuorns [141]. Both cell models demonstrate neuronal characteristics (synaptophysin expression in SH-SY5Y cells and the ability to accumulate/release dopamine upon potassium challenge; the presence of voltage-gated ion channels, axonal transport and choline acetyltransferase activity in NSC34 cells [227]), nonetheless, SH-SY5Y have been criticized that in their undifferentiated state (likewise used here) they display a tumor-like rather than neuronal-phenotype [227]. This could partially explain why, in terms of TDP-43-controlled splicing, SH-SY5Y are similarly distant from neuronal (NSC34) and from muscle (C2C12) mouse cells.

We have shown in a previous subsection (Figure 3.32c) that alternative sequences identified upon TDP-43 depletion in both mouse cell lines (common) tend to be better evolutionary conserved than those regulated by TDP-43 in a cell-type specific manner (specific). Consistent with that, a slightly higher percentage of TDP-43-controlled sequences common to C2C12 and NSC34 is subject to TDP-43 regulation in human cell line (19.5%) compared to those identified

in one mouse cell line exclusively (17.0% and 14.6% in C2C12 and NSC34, respectively) (Figure 3.39). Nonetheless, it should not be excluded (as it was discussed in an earlier chapter, Figure 3.11 and 3.15) that some AS events, which are in fact commonly controlled by TDP-43 in both mouse cell lines, were not detected in either of the two datasets and therefore appeared to be tissue-specific.

3.3.5 Conclusions

- TDP-43-mediated genes (DEG) are no better conserved across species in respect to other genes in mouse genome.
- TDP-43 targets (DEG) identified in NSC34 and C2C12 cell lines are equally well conserved at the sequence level.
- NSC34-specific TDP-43-dependent alternative sequences show higher conservation rate across species that those regulated by TDP-43 in C2C12 cells.
- Commonly detected alternative sequences are better conserved than those regulated by TDP-43 in a tissue-specific fashion.
- Exon 23 of *ASAP2*, exon 19 of *PPFIBP1* and exon 15 of *TNIK* are subject to TDP-43-dependent inclusion in mouse and human cell lines.
- TDP-43 binding is a good predictor of whether an exon would undergo TDP-43-dependent inclusion.
- We estimate that about 5-10% of TDP-43-targeted transcripts (DEG) identified in mouse undergoes TDP-43-controlled expression changes in human SH-SY5Y cell line.
- 15–20% of all TDP-43-controlled AS events detected in mouse are estimated to undergo TDP-43-dependent splicing in SH-SY5Y cells.

3.4 Aberrant splicing in proteinopathies

Widespread impairment of mRNA metabolism has been consistently reported in diseased tissues affected by ALS and FTLD [97]; [48]; [99]; [24]; [104]; [132] and recently also in skeletal muscles of patients with IBM and myofibrillar myopathy [102]; [167].

The fact that TDP-43 participates at multiple stages of RNA metabolism makes it challenging to distinguish between direct consequences of TDP-43 dysfunction *in vivo* from general perturbation of gene expression and pre-mRNA processing in disease state. Even in the absence of any disease, gene expression levels heavily vary between individuals and are, in our experience, additionally influenced by the experimental procedure itself (e.g., how and when biopsies are taken). Compared to gene expression, the relative abundance of characteristic splice isoforms displays less inter-individual variation and is less sensitive to sample handling. This being so, it would make a more stable readout of TDP-43 functionality and was therefore the subject of our investigation.

In fact, splicing activity remains one of the best characterized features of TDP-43 to date. This includes TDP-43's function to control inclusion of regular (annotated) alternative exons, which was described first, as well as its recently discovered role to repress incorporation of cryptic exons [24]; [140]; [100].

Identification of TDP-43-associated splicing events that are disturbed in disease would explain pathomechanisms driven by aberrant TDP-43 behavior and help us understand whether splicing defects might potentially underlie TDP-43 proteinopathies. One such example is *SORT1* that was observed to be misspliced in the temporal cortex of FTLD-TDP cases. SORT1 a neuronal progranulin (PGRN) receptor that mediates PGRN endocytosis. Depletion of TDP-43 leads to enhanced inclusion of exon 17b in the mature *SORT1* transcript producing truncated SORT1 receptor that can still bind PGRN but cannot internalize it, leading to PGRN imbalance [153]. Another example is inclusion of a cryptic exon within stathmin-2 (*STMN2*) due to TDP-43 dysfunction in the spinal cord of ALS patients. Downregulation of *STMN2* occurs as a direct consequence of TDP-43 loss, which is no longer able to repress cryptic inclusion that leads to NMD. As *STMN2* encodes for a microtubule regulator, reduced STMN2 levels affect axonal outgrowth and regeneration [168]; [228].

Beside gaining a mechanistic insight, screening for TDP-43-regulated transcripts that are consistently misspliced in disease represents the first step towards biomarker discovery. As it is impossible to assess TDP-43's state (i.e., its nuclear depletion, aggregation) or detect its pathological forms (e.g., its cleavage or PTMs) directly in the tissue as the brain and the spinal cord are not readily accessible, aberrantly spliced transcripts detected in the blood would serve as a proxy to asses eventual TDP-43 dysfunction and could be used for diagnostic purposes but also to monitor TDP-43 activity during disease progression or in response to potential to TDP-43 targeting therapies [132]. Recently, misspliced and truncated form of *STMN2* was identified

Table 3.1: **TDP-43-controlled transcripts that undergo splicing alterations in disease.**

Aberrantly spliced transcripts	Disease	Tissue	Study
POLDIP3	ALS	thalamus, motor cortex, spinal cord	[97]
RWDD1, *RAB-GEF1*, *EIF4H*, *POLDIP3*	ALS	spinal cord	[48]
MAST2, *MEF2D*, *KIF2A*, *KIF1B*, *PFPK*, *SPAG9*, *BCL2L11*, *PDCD6*	ALS, FTLD	middle temporal gyrus	[99]
GPSM2, *ATG4B*	FTLD, ALS	middle temporal gyrus, motor cortex	[24]
STMN2	ALS, FTLD	spinal cord, motor, frontal and temporal cortices	[132]
FUS, *OPTN*, *ATXN2*, *UBQLN2*, *DCTN1*, *GRN*, *ELP3*, *C9ORF72*	ALS	cervical spinal cord	[104]
MAPT	IBM	skeletal muscle	[102]
POLDIP3, *FNIP1*, *BRD8*	myofibrillar myopathy	skeletal muscle	[167]

as a marker of TDP-43 pathology in bulk tissues from ALS and FTLD patients. Importantly, truncated transcript was only detected in brain regions with characterized TDP-43 pathology. Similarly to *STMN2*, nuclear depletion of TDP-43 allows for inclusion of a cryptic exon within *UNC13A*, resulting in NMD and protein loss [25]. Compared to classical annotated TDP-43-dependent splicing events (i.e., SE, MXE, RI, A3SS, A5SS), cryptic exons would probably make a better disease marker owing to the fact that they, by definition, are not expected to appear in mature transcripts in the absence of disease (PSI = 0). However, cryptic splicing often causes downregulation of host transcripts though the NMD pathway [100]. Considering eventual diagnostic potential of TDP-43-controlled cryptic splicing, two patents have recently been filed [24]; [25] that pertain incorporation of cryptics in mature mRNA in TDP-43-proteinopathies as the basis for identification of biomarkers and developement of related therapeutic strategies.

Studies investigating possible splicing impairment in disease with TDP-43-pathology have reported aberrant splicing of different TDP-43-controlled transcripts (Table 3.1). As this could certainly be due to differences in methodological approaches (e.g., microarrays, RNA-seq, RT-PCR), study design (i.e., the subset of transcripts that was being investigated) and sample type (bulk tissue, neurons derived from patient's iPSCs), we hypothesize, based on our previous results, that depletion of nuclear TDP-43 would affect splicing of unique mRNA targets in different cell-types depending on the cellular environment, which clearly shapes the function of TDP-43.

Herein, we investigated the splicing of a set of TDP-43-dependent alternative exons across tissues (i.e., different brain regions and the skeletal muscle) with reported TDP-43 pathology.

3.4.1 Contributions

- Anna-Leigh Brown performed splicing analysis of patient datasets (IBM and the NYGC cohort) as described in [25]
- Cristiana Stuani generated Figure 3.40c

- Sara Cappelli silenced TDP-43 in SH-SY5Y and SK-N-DZ cells and extracted RNA. RNA samples were sequenced by Novogene (Beijing, China)

3.4.2 TDP-43-associated splicing signature

Our splicing signature consisted of six TDP-43-dependent exons, which were chosen for the following reasons.

Exon 3 of *POLDIP3* and exon 15 of *TNIK* have been previously validated as TDP-43 targets in our lab using different cell lines (HEK-293, Hela, SH-SY5Y, SK-N-DZ, NSC34).

Exon 3 within *POLDIP3*, in our experience, shows high sensitivity to reduced TDP-43 levels/activity, therefore, its inclusion levels are routinely screened using RT-PCR for functional validation of TDP-43 depletion/activity *in vitro* (Figure 3.40a).

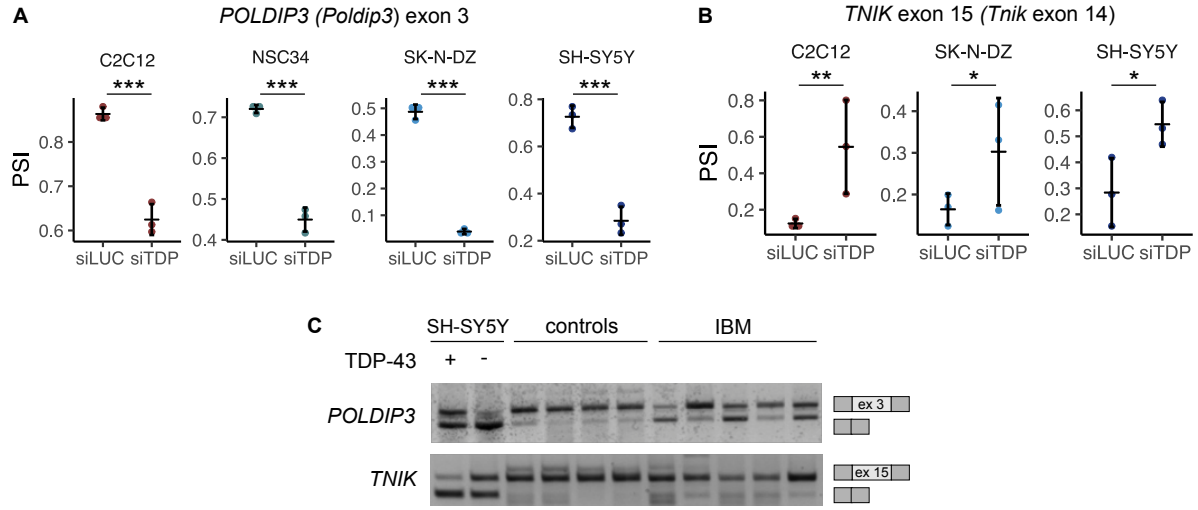


Figure 3.40: **Splicing of *POLDIP3* and *TNIK* in mouse and human cell lines.** **A** Different inclusion levels of *POLDIP3* (*Poldip3*) exon 3 were detected in TDP-43-silenced cells as assessed by RNA-seq in four cell lines. **B** Different inclusion levels of *TNIK* exon 15 (*Tnik* exon 14) detected in three (out of four) cell lines tested. p-values were generated using Fisher's exact test and corrected for multiple testing. * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001. **C** Splicing patterns of *POLDIP3* and *TNIK* in skeletal muscles of IBM patients resemble that of TDP-43-silenced SH-SY5Y cells (Stuani, Buratti, unpublished). The images shows the result obtained by RT-PCR.

Another previously characterized TDP-43-regulated event is found within *TNIK* mRNA. Inclusion of exon 15 (encoding for 29 amino acids) is enhanced upon TDP-43 knockdown and gives rise to a longer protein isoform [225]. The fact that we did not see enhanced exon 14 inclusion in TDP-43-silenced mouse NSC34 cells (Figure 3.16) implies that in mouse, TDP-43 controls exon 14 in a cell-type-specific manner. In humans on the other hand, exon 15 seems to be consistently dependent on TDP-43 irrespective of cell-type (SH-SY5Y, SK-N-DZ, RH-30,

HEK-293) (Figure 3.36; 3.40b) and was therefore considered as a suitable readout of TDP-43 dysfunction in patients. Furthermore, altered splicing of both, *POLDIP3* and *TNIK*, has already been observed in a preliminary study conducted in our lab investigating splicing patterns of some TDP-43 targets in IBM muscles (Figure 3.40c).

Next, our splicing signature also included two novel TDP-43-regulated events within *PPFIBP1* and *ASAP2* identified by this study using mouse cells lines (Figure 3.14), splicing of which is conserved in human cells (SH-SY5Y, RH-30) (Figure 3.36a). Importantly, both alternative exons appear to be targeted by TDP-43 in a direct fashion (Fig 3.36b).

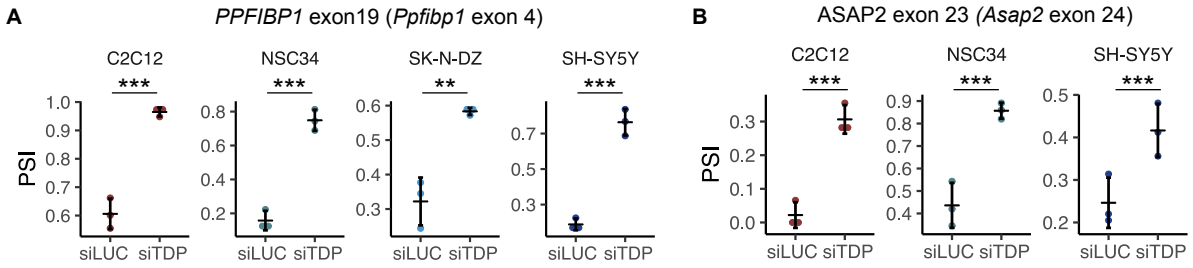


Figure 3.41: **Splicing of *PPFIBP1* and *ASAP2* in mouse and human cell lines.** **A** Different inclusion levels of *PPFIBP1* exon 19 (*Ppfibp1* exon 4) were detected in TDP-43-silenced cells as assessed by RNA-seq in four cell lines. **B** Different inclusion levels of *ASAP2* exon 23 (*Asap2* exon 24) detected in three (out of four) cell lines tested. p-values were generated using Fisher's exact test and corrected for multiple testing. * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001.

Furthermore, we investigated TDP-43-dependent splicing of *TBC1D1* mRNA (Figure 3.42). As all splicing experiments in this work were conducted using undifferentiated cells (mouse and human), the long isoform of *TBC1D1* transcript (with exons 12 and 13) was only detected in undifferentiated C2C12 cells, and even in these cells, it would represent only of about 10% of the total transcript (Figure 3.27a). Given that the long *TBC1D1* is not detected in undifferentiated NSC34, SH-SY5Y, SK-N-DZ and RH-30 cells, we could not demonstrate its dependence on TDP-43 using undifferentiated cells (Figures 3.42b and 3.35a). Nonetheless, there is sufficient evidence to assume that in differentiated cells, which are the ones that actually build mature tissues, inclusion of exons 12 and 13 does occur and is regulated by TDP-43 (despite not exclusively). First, we (Figure 3.26a) and the others [27] have shown that expression of the long *Tbc1d1* isoform increases with C2C12 differentiation, and that inclusion of exons 12 and 13 depends on the presence of TDP-43 at all stages of myoblast differentiation (Figure 3.26c). Like in muscle-derived cells, the expression of the long *Tbc1d1* isoform is enhanced during neurogenesis of cortical glutamatergic neurons from murine ESCs [195]. The fact that TDP-43 binding sites were found within introns flanking the two alternative exons of *TBC1D1* (Figure 3.42b) in SH-SY5Y cells [7] suggests that TDP-43 regulates their inclusion in a direct fashion. As exons

12 and 13 are normally retained in the mature *TBC1D1* transcript in muscle and neuronal tissues (of healthy individuals) (Figure 3.42b), it is reasonable to assume that their splicing might be affected by TDP-43 loss/dysfunction in disease. Given all these, we subsequently inspected eventual disturbance of exon 12/13 inclusion in patient's tissues.

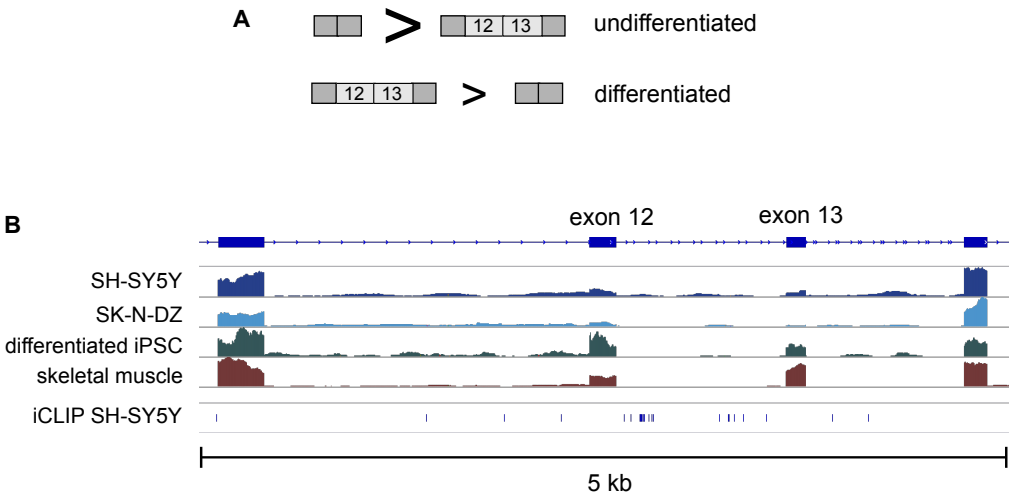


Figure 3.42: **Two splicing isoforms of *TBC1D1* mRNA. A** The scheme represents two isoforms of *TBC1D1* transcript. The short isoform prevails in undifferentiated cells while the long isorom that contains exons 12 and 13, becomes important in differentiated cells. **B** Exons 12 and 13 do not make part of mature *TBC1D1* transcript in undifferentiated human neuronal cells (SH-SY5Y and SK-N-DZ), while they are detected within *TBC1D1* mRNA from differentiated iPSCs and skeletal muscles from healthy individuals. TDP-43 binding sites (lower track) that were identified by iCLIP analysis in SH-SY5Y cells [7] within introns flanking exons 12 and 13.

3.4.3 Splicing of TDP-43 targets is tissue-specific

Alternative splicing has long been known as an important driver of phenotypic diversity of different cell-types [40] with 50-80% of AS events being regulated in a tissue-specific fashion [39]. The tissue-characteristic splicing "code" is determined by the relative expression of RNA-binding proteins and their subcellular localization along with the presence of tissue-specific *cis*-regulatory regions [35]; [39]. We have shown in a previous chapter that splicing regulation executed by TDP-43 highly depends on cell-type and the cellular environment as such. However, before exploring plausible splicing defects in patients, it is worthy to inspect the relative inclusion of TDP-43-controlled alternative exons across different brain regions. As exemplified by *ASAP2* exon 23 and *TBC1D1* exon 13, neuronal tissues significantly vary with regards to the inclusion of the two alternative exons (Figure 3.43a).

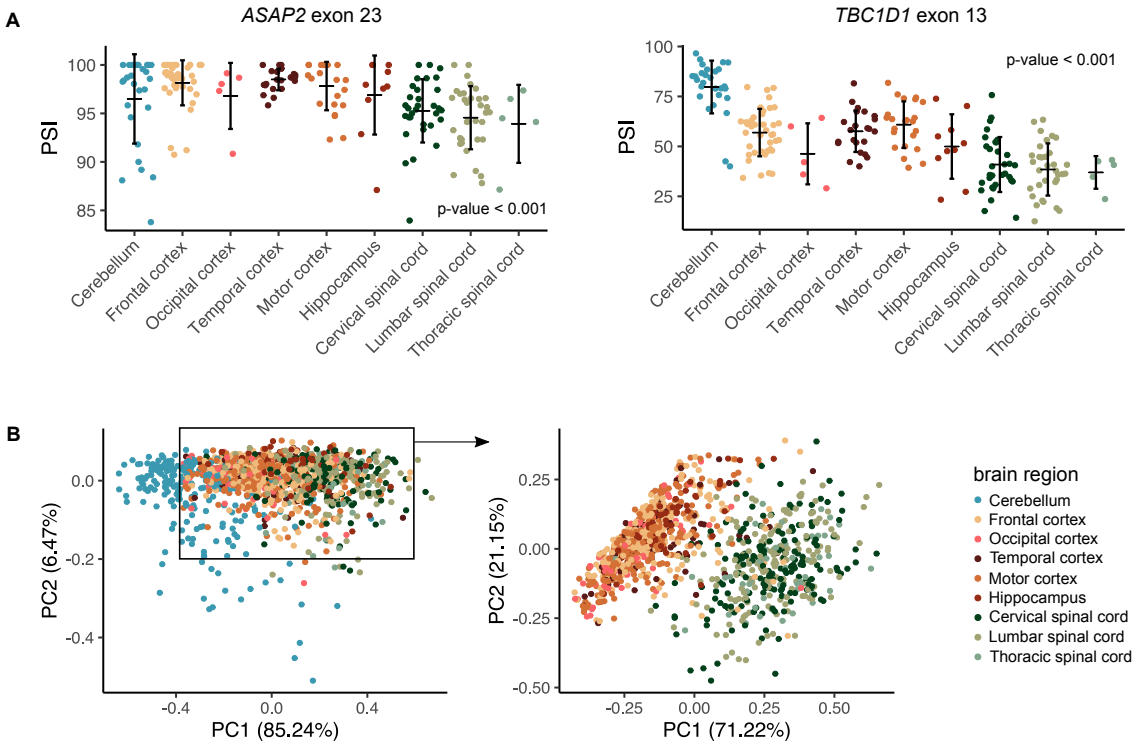


Figure 3.43: **Tissue-specific splicing of six TDP-43-controlled alternative exons. A** Dot plots demonstrate variable inclusion levels of alternative exons across neuroanatomical regions of healthy controls (in the absence of TDP-43 pathology), as exemplified by two alternative exons – exon 23 of *ASAP2* and exon 13 of *TBC1D1*. p-values were generated using Kruskal-Wallis chi-squared test. **B** PCA plots show clustering of different brain regions based on inclusion levels (PSI) of six TDP-43-regulated alternative exons. Cerebellum (blue) is excluded from the PCA on the right.

Indeed, a PCA clustering featuring multiple anatomical regions of the brain from healthy individuals based on inclusion levels (PSI) of six TDP-43-regulated alternative exons (*POLDIP3* exon 3, *TNIK* exon 15, *PPFIBP1* exon 19, *ASAP2* exon 23, *TBC1D1* exons 12 and 13) reveals that their splicing profiles in fact recapitulate different brain areas (Figure 3.43b). Whereas cerebellum appears to form an independent cluster, we see that splicing of six exons in the spinal cord (cervical, lumbar, thoracic) generally differs from that in the brain (cortical brain regions and hippocampus).

On the other hand, splicing of above-mentioned genes is not influenced by age (Figure 3.44a) neither it depends on sex (Figure 3.44b).

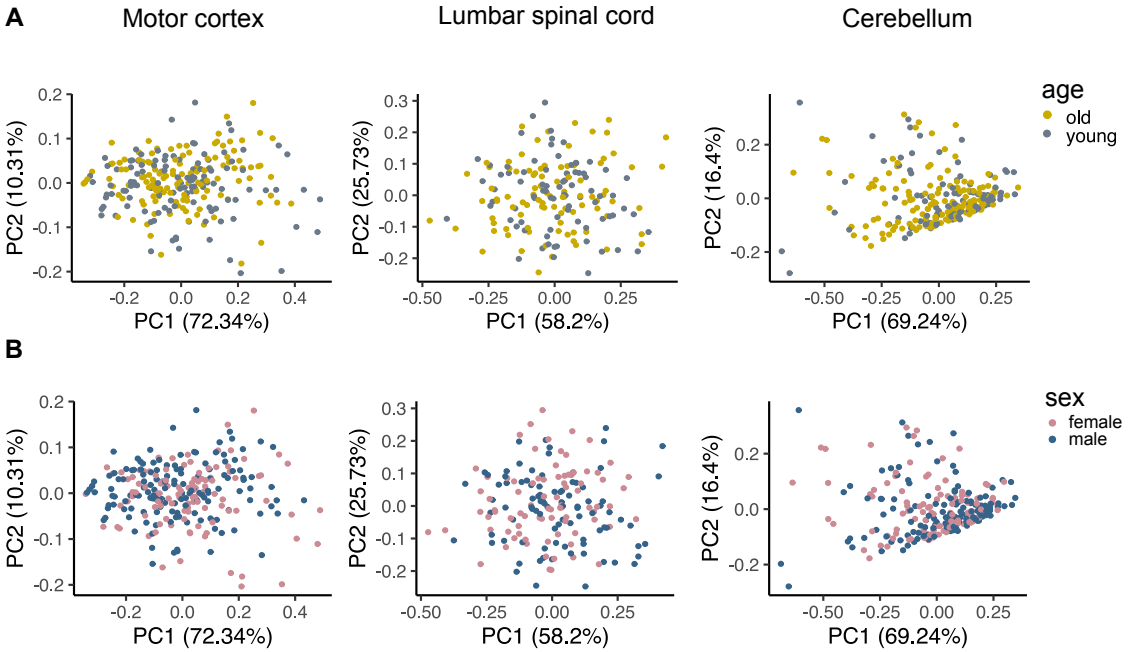


Figure 3.44: **Splicing of six TDP-43-controlled alternative exons by age and sex.** PCA plots feature inclusion levels of six alternative exons in three representative neuroanatomical regions by age (**A**) and sex (**B**). All samples were plotted together irrespective of the diagnosis (ctrl, ALS, FTLD). n motor cortex: 255, n lumbar spinal cord: 173, n cerebellum: 247. Age bellow median age (66 years) was considered young.

Given the tissue-characteristic isoform expression and cell-type-specific activity of TDP-43, we explored splicing of TDP-43 targets in each brain region independently, rather than analyzing them together. Tissue-specific accumulation of truncated *STMN2*, which is believed to be a good readout of TDP-43 dysfunction [228]; [168], in fact occurs in anatomical regions previously known to be affected by TDP-43 pathology [132]; [2]; [3].

3.4.4 Aberrant splicing of TDP-43-dependent exons detected in diseased tissues

To explore if splicing impairment could be detected in TDP-43 proteinopathies, we measured inclusion levels of six TDP-43-mediated alternative exons in IBM muscles as well as in pathological neuroanatomical regions of ALS and FTLD cases with reported TDP-43 pathology (ALS-TDP and FTLD-TDP).

The number of IBM samples available for the analysis (RNA-seq of skeletal muscle biopsies) was rather limited (4 IBM patients vs. 4 healthy controls, Table 2.5) (Figure 3.45). Another limitation concerning IBM samples lies in the fact that sampled muscles were not histologically examined for the presence of sarcoplasmic TDP-43 inclusions. Thus, it could only be assumed,

based on the current literature [90]; [91], that sampled muscles actually exhibited TDP-43 pathology.

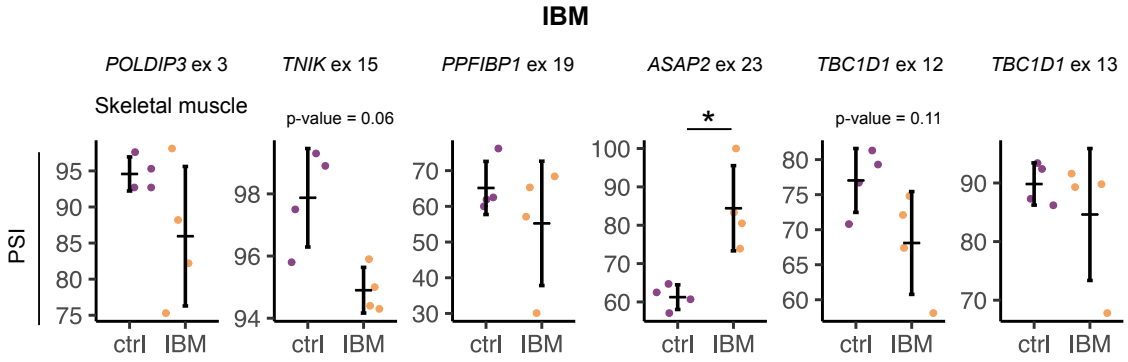


Figure 3.45: **TDP-43-controlled splicing in IBM muscles.** Inclusion levels (PSI) of six alternative exons in skeletal muscle biopsies from IBM patients vs. healthy controls (n = 4 per group). p-values were generated using Wilcoxon rank sum test. * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001.

Comparing IBM muscles with those of controls, we, out of six TDP-43-dependent AS events, observed significantly increased inclusion of exon 24 within *ASAP2*, while two other exons (*TNIK* exon 15 and *TBC1D1* exon 12) displayed a trend towards altered splicing (despite non-significantly). Interestingly enough, splicing of *TNIK* changes in the opposite direction as seen upon TDP-43 depletion in cell lines (Figure 3.40b) yet, this change is consistent with what has been previously observed in IBM muscles (Figure 3.40c).

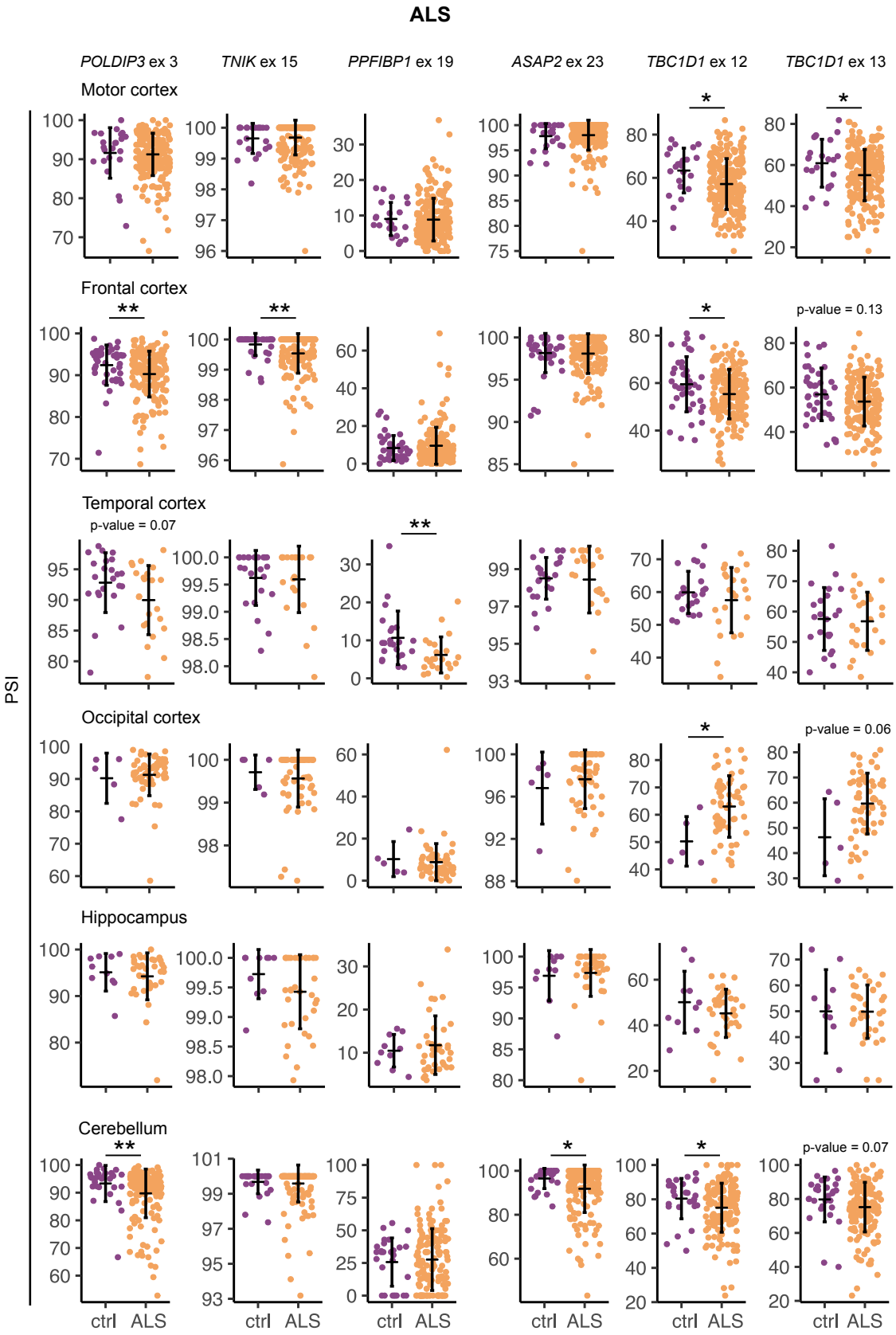
The ALS/FTLD RNA-seq dataset, on the other hand, comes from a large *post-mortem* cohort comprising of samples from multiple neuroanatomical regions and diseases collected by the New York Genome Center (NYGC) ALS Consortium. Donors of interest for our study (a smaller subset of the NYGC cohort) were classified according to their diagnosis (ALS, FTLD, controls without neurological disease). FTLD cases were additionally inspected for the presence of TDP-43 aggregates, while ALS patients (except for those with *SOD1* mutations) were assumed to have TDP-43 pathology although they were not specifically investigated for that.

Samples from multiple neuroanatomical regions were available for ALS-TDP cases including those from different cortical areas (motor cortex, frontal cortex, temporal cortex, occipital cortex), hippocampus, cerebellum and spinal cord (cervical and lumbar spinal cord) (Figure 3.46). Unadjusted analysis (not adjusted for age and gender) revealed that altered splicing, although not necessarily of the same genes, was detected in motor cortex and spinal cord. Surprisingly enough, we also found differences in other regions including cerebellum.

Similarly to other neuroegenerative disease, ALS starts focally and spreads to other brain areas [76]. All ALS cases of the subcohort investigated herein were assumed to have TDP-43 pathology, however, abnormal accumulation of TDP-43 is believed to be more prevalent in

some anatomical regions. One study reports that 96% of all ALS patients exhibit TDP-43 pathology in the inferior olive compared to 56% in prefrontal cortex and only 15% in medial temporal lobe regions [229]. It was suggested that the distribution and severity of TDP-43 pathology recapitulate clinical manifestation of ALS and the stage of the disease [229]. While pTDP-43 (phosphorylation of TDP-43) is restricted to motor cortex, brain stem, motor nuclei of cranial nerves and spinal cord motoneurons at the initial stage of ALS, it gradually spreads (at later stages) to prefrontal neocortex, striatum, temporal lobe and hippocampus with rare cases displaying the involvement of cerebellum as well [230].

The fact that TDP-43-associated alterations in splicing were detected in brain tissues other than motor cortex and spinal cord could possibly reflect the spread of TDP-43 pathology at later stages of the disease. In this same dataset, the expression of truncated *STMN2* was most often observed in the spinal cord, motor cortex and frontal cortex, however, it was not exclusively restricted to those areas. Despite only in few cases, truncated *STMN2* was also detected in the occipital cortex and cerebellum [132].


ALS
POLDIP3 ex 3
TNIK ex 15
PPFIBP1 ex 19
ASAP2 ex 23
TBC1D1 ex 12
TBC1D1 ex 13
Motor cortex
Frontal cortex
Temporal cortex
Occipital cortex
Hippocampus
Cerebellum
p-value = 0.07
p-value = 0.13
p-value = 0.06
p-value = 0.07
PSI
ctrl
ALS

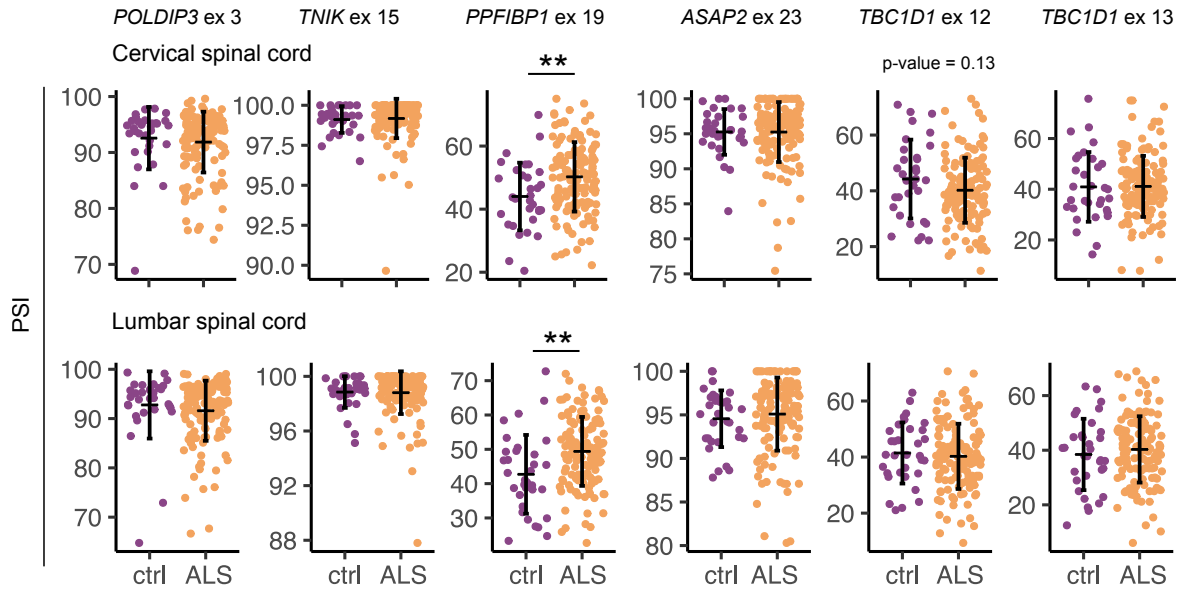


Figure 3.46: **TDP-43-controlled splicing in ALS brain.** PSI of six alternative exons in different brain regions of ALS patients and healthy controls (n motor cortex: 223 ALS and 23 ctrl, n frontal cortex: 173 ALS, 43 ctrl, n temporal cortex: 23 ALS, 23 ctrl, n occipital cortex: 57 ALS, 5 ctrls, n hippocampus: 37 ALS, 10 ctrls, n cerebellum: 155 ALS, 28 ctrls, n cervical spinal cord: 134 ALS and 32 ctrl, n lumbar spinal cord 136 ALS and 33 ctrl). p-values were generated using Wilcoxon rank sum test. * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001.

Compared to ALS cases, splicing of TDP-43-controlled exons seems to be more heavily perturbed in FTLD tissues (frontal cortex, temporal cortex and cerebellum datasets were available) (Figure 3.47). In this same dataset, truncated *STMN2* was detected in frontal and temporal cortices but not in cerebellum [132]. A study comparing regional distribution of pTDP-43 in FTLD-TDP (behavioral variant) found that cerebellum could only be affected in some rare FTLD cases at the very final stage [231] and Neumann *et al.* [2] reported cerebellum is spared in all FTLD cases.

Collectively, our results demonstrate that splicing of TDP-43-targets (i.e., the proportion of TDP-43-dependent splice isoforms of the same transcript) is perturbed in tissues with TDP-43 pathology.

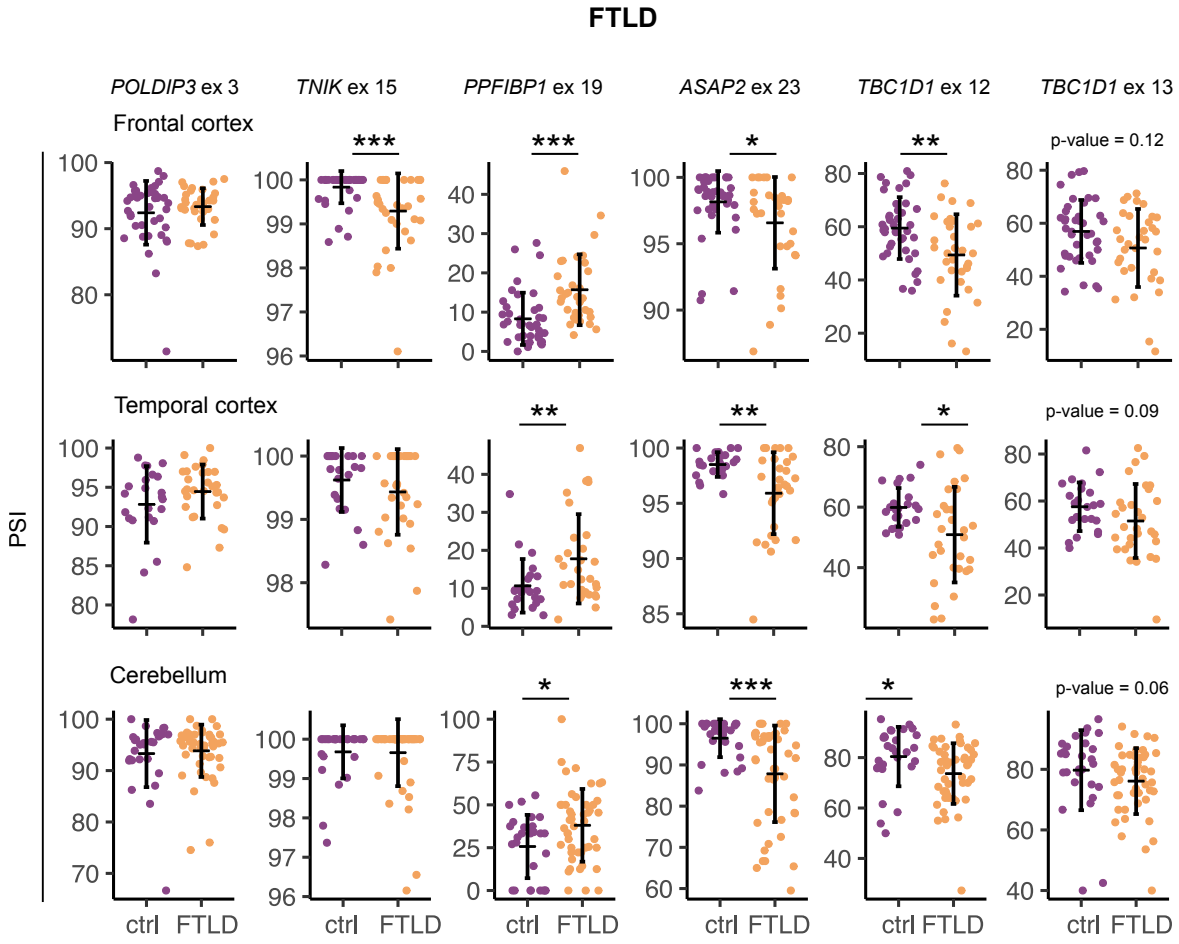


Figure 3.47: **TDP-43-controlled splicing in FTLD brain.** PSI of six alternative exons in frontal and temporal cortices and cerebellum of FTLD patients with reported TDP-43 pathology and healthy controls (n frontal cortex: 33 FTLD and 40 ctrl, n temporal cortex: 30 FTLD and 23 ctrl, n cerebellum: 47 FTLD and 28 ctrl). p-values were generated using Wilcoxon rank sum test. * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001.

3.4.5 Conclusions

- Expression of characteristic splice isoforms of TDP-43 targets differs across anatomical brain regions.
- Splicing of TDP-43 targets that were considered as a part of the splicing signature is not affected by age neither it depends on gender.
- Altered splicing of TDP-43-targeted transcripts was detected across different neuroanatomical regions affected by ALD/FTLD as well as in skeletal muscles of IBM patients.

4. Discussion and outlook

TDP-43 inclusions remain the hallmark of ALS/FTLD and are frequently recognized as a secondary pathology in other neurodegenerative disease. As a growing body of evidence pointed towards pathologic TDP-43 behavior (i.e., misslocalization and aggregation) in tissues beyond CNS, we sought to fill this gap and investigated TDP-43's activity in muscle-derived cells.

Plausible cell-type-specific molecular features of TDP-43 have not been investigated before. We believe, nevertheless, that elucidating the role of TDP-43 in muscles specifically was very much needed to first understand its involvement in IBM pathogenesis (as well as that of other myopathies) but also to find disease makers that could be employed for diagnosis of TDP-43 pathologies in muscles. Tissues could be differentially affected by lack of a protein or any protein modification [232]. We attempted to understand whether that is the case with TDP-43 and we thus conducted a direct and systematic comparison of consequences exerted by TDP-43 depletion in cells of muscle and neuronal origin.

Our experimental setting modeled loss of TDP-43's function. Compared to tissues, cell lines represent a rather robust system that allows investigation of cell-autonomous response devoid of confounding effects due to presence of other cell populations. Molecular analysis of diseased tissues (with compromised TDP-43's function) would, on the other hand, reflect average changes occurring in a heterogeneous population of cells (e.g., myocytes and inflammatory cells that infiltrate diseased muscle) which could hinder identification of primary disease mechanisms [233].

Two cell lines employed herein, C2C12 and NSC34, were used in their undifferentiated state. As we understand that homeostatic levels of TDP-43 are tightly regulated and that tissue-specific levels of splicing factors determine their splicing activity [214], we compared expression of endogenous TDP-43 in both cell-types and noted no difference, neither at the protein nor at the RNA level. We thus excluded that a variable amount of TDP-43 *per se* could be responsible for eventual differences between cell lines. Of note, two cell lines display a markedly distinct transcriptome, which needs to be kept in mind with such experimental design (2 cell lines x 2 experimental conditions).

TDP-43 is a versatile protein participating at different steps of an mRNA life cycle. By sequencing poly-adenylated RNAs, we gained insight into TDP-43-mediated control of transcript abundance and alternative spicing. Gene expression and splicing control are two complemen-

tary regulatory mechanisms, that ultimately tune protein levels and protein structure. While TDP-43 turned out to be equally important in mediating abundance (i.e., expression levels) of individual mRNAs in both cell-types (as a similar number of target transcripts was detected by DE analysis), TDP-43-dependent splicing (assessed by differential splicing analysis) appeared to be of bigger importance in neurons. The latter supports the idea that some functions of TDP-43 (namely pre-mRNA processing) are more critical for one cell-type then another. Notably, alternative splicing in general has been known to be widely exploited by neurons relative to any other cell population [26]; [41], and TDP-43 as a splicing factor makes part of neuronal regulatory network.

Only a small subset of mRNAs whose splicing was affected by TDP-43 knockdown (about 20%) additionally displayed altered transcript abundance. Yet we have observed an enrichment of TDP-43-targeted AS and DE genes in overlapping GO terms. This suggests that TDP-43 controls particular functional processes by the means of mediating expression levels or isoform proportion of mRNAs encoding for proteins that participate a given biological process, like it was exemplified by the *axon guidance pathway*.

Given that TDP-43 is essential for cell survival and embryo development, one would expect, regardless eventual differences in cellular environments, that there should be a subset of genes involved in crucial biological processes which are commonly regulated by TDP-43 in every cell-type [166]. Until now, the overlap between TDP-43 targets identified across studies remains rather small. While this certainly reflects differences in technical approaches and experimental models to be employed (human vs. mouse, tissues vs. cell lines, health vs. disease), it also points out towards cell-type-specific behavior of TDP-43.

A big advantage of this study lies in the fact that TDP-43's function in the two cell lines was investigated in parallel, from the very fist cell culture experiment till the final analysis, ruling out any experimental bias.

Even so, the proportion of commonly detected TDP-43 targets (15.7% overlap of DEG and 5.2% for AS events, respectively) remained small. This observation could only partially be attributed to differences in baseline expression of individual transcripts. Our results clearly demonstrate that unique subsets of genes are regulated by TDP-43 in each cell-type, nonetheless, the exact extent of cell-type specificity is still under debate. Other studies, for example, reported a 1.5% overlap between TDP-43-regulated splicing events detected in SH-SY5Y and Hela cells [97] and a 7.2% overlap between neocortex and spinal cord in mouse (as assessed by microarray) [57].

Cell-type characteristic transcriptome (i.e., variable expression of AS transcripts and cell-type-specific exon usage) makes detection of splicing alterations (in an unbiased manner in both cell-types) rather challenging. It is very likely that a certain TDP-43-controlled event would be detected by a splicing software in one dataset (i.e., from one cell line) but not in another, leaving the impression of cell-type-specific performance of TDP-43, while it actually reflects the

analytical bias.

To estimate the extent of the overlap in a more reliable manner, we analyzed our dataset using two different splicing tools (rMATS and MAJIQ) that are known to perform strikingly differently on the same data [65]. In fact, we obtained 5.2% overlap with one (for splicing events) and 15.2% overlap with the other (for splicing junctions). We further narrowed down the list of putative TDP-43-regulated AS events by selecting those that were detected by both pipelines in each cell line, and were additionally confirmed by visual investigation. We are fully aware that this approach left out many actual splicing events (false negatives), however, we used this smaller subset of TDP-43 targets (which we could rely on) to determine the actual proportion of alternative splicing commonly regulated by TDP-43 in two cell-types. In our estimate, around 50% of AS events subject to TDP-43 control appear to be universally regulated irrespective of cell-type.

Despite the fact we disclosed unique sets of TDP-43 targets in two cell-types, TDP-43-mediated events bear similar features. Herein we showed - what has been previously observed elsewhere [8] - that TDP-43 mediates abundance of mRNAs encoded by long genes (i.e., transcripts containing extremely long introns) not only in neuronal but also in muscle cells. Moreover, splicing events detected in either cell line belong to the same AS categories and displayed the same frequency of frame-shifting. The size of expression or splicing changes ($\log_2$ fold change or ΔPSI, respectively) in common targets showed high correlation between muscle and neuronal cells, suggesting that for commonly detected targets TDP-43 loss exerts a similar effect.

GO analysis suggested a partial functional overlap of TDP-43 in both cell-types, as unique TDP-43 targets (DE or AS genes) were often enriched in overlapping GO terms. Those were in fact associated with basic cellular functions (e.g., *ribosome biogenesis*) highlighting the central role played by TDP-43 in every cell. Among cellular processes predicted to be commonly disturbed upon TDP-43 knockdown, we also found those generally assumed to be involved in development of TDP-43-proteinopathies in muscle and neurons (e.g., aberrant protein accumulation, defects in protein disposal, mitochondrial abnormalities). Although it still needs to be established whether pathomechanisms driven by aberrant TDP-43 behavior are as central in IBM as they are in ALS, our comparison provides the first indication that two seemingly distant disease might be mechanistically linked.

Considering high (morphological and functional) specialization of neuronal and muscle tissue, we also sought for eventual differences in TDP-43's performance in muscle and neurons. It is worthy to note here that two cell lines, despite undifferentiated, display a very much cell-type-characteristic transcriptome and that many genes appear to be transcribed in a cell-type specific manner. Also, it is very unlikely that TDP-43 depletion would induce transcription of muscle-characteristic genes (for example *Myog* that encodes for myogenin) in neuronal cell line where it is usually not expressed. Nevertheless we have shown that even for transcripts with comparable expression levels, some undergo TDP-43-dependent expression changes in one cell line but not

in another. Likewise, inclusion of certain alternative exons appears to be uniquely controlled by TDP-43 in a cell-type-specific fashion.

Recently, the activation of cryptic exons following TDP-43 knockdown was shown to be tissue-specific [140]. Despite not being discussed in this work we believe that emergence of cryptic exons in TDP-43-depleted cells indeed shows more tissue-specificity relative to other categories of annotated TDP-43-controlled alternative splicing (investigated as a part of our rMATS protocol).

Beside at the level of individual target mRNAs, enrichment analysis suggested some functional specificity, i.e., disturbance of neuronal features upon TDP-43 knockdown in neuronal cells and impaired acquisition on muscle-characteristic phenotype in muscle cells. This might be attributed to the fact that TDP-43 mediates unique sets of genes in each cell line (which encode for transcripts participating in neuronal or muscle-characteristic processes). Yet at the same time it probably reflects a more general disturbance of core cellular processes (as for example ribosome biogenesis) that are subject to TDP-43 control in every cell, which would in turn impact expression/processing of cell-type specific transcripts. An important drawback of our experimental design is that we were not able to distinguish between direct and indirect consequences of TDP-43 depletion.

Interestingly enough, one study proposed that motor neurons are more sensitive to overexpression of mutated TDP-43 than other neuronal populations [221] consistently with selective vulnerability of motor neurons in ALS. In this regard, it should not be excluded that muscles and neurons could have different sensitivity to TDP-43 loss/dysfunction.

In fact, differences in TDP-43's state (i.e., its exact localization, its aggregation status, post-translational modifications [234]) have been described across TDP-43-proteinopathies affecting muscle and brain [66]. We thus assume that intrinsic properties of TDP-43 itself together with the characteristic cellular environment may impact TDP-43's regulatory activity.

We further explored how factors other than intrinsic features of TDP-43 could shape its function and render it tissue-specific. As an example, cell-type-characteristic GC methylation patterns were suggested to impact intron retention in a cell-type-specific [235]. Interestingly enough, intron retention was most frequently detected cell-type-specific event we observed in C2C12 and NSC34 datasets.

Notwithstanding a very similar expression of TDP-43 in the two cell lines, we noted a higher abundance of other RNA binding proteins in neuronal cell line. Moreover, we showed that unique sets of splicing factors are expressed in neuronal and muscle cells. As each splicing event depends on coordinated control provided by multiple RBPs and the relative expression of these underlies tissue-specific splicing [35]; [39], we explored whether presence of cell-type-specific RBPs (i.e., ELAVL4, ELAVL3, CELF2, KHDRBS3) impacts inclusion of TDP-43-targeted sequences. Having identified (and validated) a set of cassette exons regulated by TDP-43 in a cell-type-specific fashion, we then silenced tissue-specific splicing factors and demonstrating that their depletion

indeed evens out regulation provided by TDP-43. The observation that more TDP-43-regulated AS events were detected in neuronal dataset is likely related to more complex splicing program and the importance of alternative splcing in neuronal tissues [26]; [41].

In addition, it should not be overlooked that the two mouse cell lines we used to model the skeletal muscle and neurons actually differ in their genome (in contrast to tissues derived from the same organism). It would be interesting to know whether eventual SNPs found within regulatory elements might influence TDP-43-dependent splicing, as it was recently the case with *UNC13A* gene [25].

The majority of the work presented herein focuses on transcriptome-wide comparison of two cell-types, which provides no mechanistic evidence on functional significance of individual splicing events. However, we started to explore whether, and how, TDP-43-dependent splicing of *Tbc1d1* mRNA impacts its function.

Stermann *et al.* [236] recently reported that the longer isoform is typically expressed in skeletal muscle of mice, while the shorter one prevails in heart and pancreatic islets implying on the distinct functions these two might have. As we show here, depletion of TDP-43 changes the ratio between the long and the short *Tbc1d1* isoform. TBC1D1 has been previously demonstrated to mediate translocation of glucose transporters from perinuclear space to plasma membrane, thereby influencing glucose uptake in muscle cells. TDP-43 has been recently recognized as a regulator of glucose homeostasis [194]; [114], potentially via effects on TBC1D1. As reduced ATP levels were observed in TDP-43-silenced myoblasts, we speculate this might indicate disturbance of glucose uptake due imbalance of two splice isoforms, but the hypothesis needs further investigation. To assess the role of each *Tbc1d1* isoform in mediating glucose uptake one could block the expression of either of them (i.e., the long or the short *Tbc1d1* isoform) and carry out metabolic assessment of cellular activity using Seahorse analyzer. If reduced ATP-generating capacity of TDP-43-silenced myoblasts could modulate cell transition, remains to be established. By modulating the amount of ATP available for myoblasts to perform their basic and transition-related cellular functions, depletion of TDP-43 could induce metabolic perturbations that also affect muscle differentiation. In fact, there is some evidence linking glucose restriction and myoblasts differentiation. Metformin, a caloric-restriction mimicking drug, reduced myogenic potential of C2C12 cells at higher doses, while it promoted differentiation at lower doses [191]. Culturing C2C12 in glucose-free medium impaired differentiation process as well, as myoblasts at the early stage of differentiation meet 60% of their energy demand by lactate production from glucose [237].

Mouse cell lines were chosen as our cell model (due to lack of an established cell derived from human skeletal muscle) to compare the function of TDP-43 in cells of different backgrounds. While we believe they make a suitable setting to explore the phenomenon of tissue-specific TDP-43 behavior and that similar results would be obtained if one compared two human cell lines, mouse cells could not be used to identify TDP-43-regulated events relevant for human physiology.

A clearer understanding of the extent to which TDP-43-mediated events are conserved between mouse and human is still lacking, yet it is a crucial point that should be addressed in future as it will allow better comparisons of human and mouse models of disease.

Herein we investigated the conservation of TDP-43-mediated transcripts (DEG) and exons (AS). Interestingly enough, alternative sequences regulated by TDP-43 in the neuronal cell line seem to be more conserved across species than TDP-43-regulated sequences in muscles. It has been previously reported that neuronal AS events in general display higher extent of conservation than those occurring elsewhere [43] implying that they provide important functions.

However, as exon orthology (as assessed at the sequence level) could not be predictive of AS conservation since exon incorporation into mature mRNA depends on the exonic sequence but also on *cis*-regulatory motifs and *trans*-acting factors, we silenced TDP-43 in two human cell lines and searched for orthologoues events subject to TDP-43 regulation in both, mouse and human.

We identified two novel targets, *ASAP2* and *PPFIBP1* that undergo TDP-43-dependent splicing in all (mouse and human) cell lines tested. These findings could be of interest to identify common endpoints of mouse and human disease models that could then be used to monitor the efficiency of eventual novel therapeutic approaches or to follow disease course/onset.

Finally, as a proof-of-principle, we showed that splicing alterations of TDP-43-dependent transcripts indeed do take place in different tissues (i.e., skeletal muscle and certain neuroanatomical regions) affected by TDP-43 pathology. While expression levels of a given transcript heavily vary between individuals and, in our experience, seem to be influenced by experimental procedure itself (how and when biopsies are taken), the relative abundance of characteristic isoforms appears to be a more reliable readout. Considering cell-type-specific activity of TDP-43, it is reasonable to deduce that splicing of other TDP-43-controlled transcripts would be affected in the skeletal muscle and in neurons. In conclusion, we show that splicing changes as such indeed represent an indication of pathological conditions both in the skeletal muscle of IBM patients and in the brain of individuals affected with ALS and FTLD.

Bibliography

[1] S. H. Ou et al. "Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs". *J Virol* 69.6 (June 1995), pp. 3584–3596.

[2] M. Neumann et al. "Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis". *Science* 314.5796 (Oct. 2006), pp. 130–133.

[3] T. Arai et al. "TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis". *Biochemical and Biophysical Research Communications* 351.3 (Dec. 2006), pp. 602–611.

[4] E. Buratti and F. E. Baralle. "TDP-43: gumming up neurons through protein–protein and protein–RNA interactions". *Trends in Biochemical Sciences* 37.6 (June 2012), pp. 237–247.

[5] E. Buratti and F. E. Baralle. "Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel dplicing regulator of *CFTR* exon 9". *J. Biol. Chem.* 276.39 (Sept. 2001), pp. 36337–36343.

[6] S. P. Han, Y. H. Tang, and R. Smith. "Functional diversity of the hnRNPs: past, present and perspectives". *Biochemical Journal* 430.3 (Sept. 2010), pp. 379–392.

[7] J. R. Tollervey et al. "Characterizing the RNA targets and position-dependent splicing regulation by TDP-43". *Nat Neurosci* 14.4 (Apr. 2011), pp. 452–458.

[8] M. Polymenidou et al. "Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43". *Nat Neurosci* 14.4 (Apr. 2011), pp. 459–468.

[9] E. Buratti et al. "TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through Its C-terminal tail: an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing". *J. Biol. Chem.* 280.45 (Nov. 2005), pp. 37572–37584.

[10] B. D. Freibaum et al. "Global analysis of TDP-43 interacting proteins reveals strong association with RNA splicing and translation machinery". *J. Proteome Res.* 9.2 (Feb. 2010), pp. 1104–1120.

[11] O. D. King, A. D. Gitler, and J. Shorter. “The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease”. *Brain Research* 1462 (June 2012), pp. 61–80.

[12] M. Budini and E. Buratti. “TDP-43 autoregulation: implications for disease”. *J Mol Neurosci* 45.3 (Nov. 2011), pp. 473–479.

[13] A. D’Ambrogio et al. “Functional mapping of the interaction between TDP-43 and hnRNP A2 in vivo”. *Nucleic Acids Research* 37.12 (July 2009), pp. 4116–4126.

[14] M. J. Winton et al. “Disturbance of nuclear and cytoplasmic TAR DNA-binding protein (TDP-43) induces disease-like redistribution, sequestration, and aggregate formation”. *J. Biol. Chem.* 283.19 (May 2008), pp. 13302–13309.

[15] S. Prpar Mihevc et al. “TDP-43 aggregation mirrors TDP-43 knockdown, affecting the expression levels of a common set of proteins”. *Sci Rep* 6.1 (Dec. 2016), p. 33996.

[16] Y. M. Ayala et al. “TDP-43 regulates its mRNA levels through a negative feedback loop”. *The EMBO journal* 30.2 (2011), pp. 277–288.

[17] X. Zhang et al. “Rapamycin treatment augments motor neuron degeneration in SOD1 G93A mouse model of amyotrophic lateral sclerosis”. *Autophagy* 7.4 (Apr. 2011), pp. 412–425.

[18] M. Modic et al. “Cross-regulation between TDP-43 and paraspeckles promotes pluripotency-ifferentiation transition”. *Molecular Cell* 74.5 (June 2019), 951–965.e13.

[19] J. A. West et al. “Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization”. *Journal of Cell Biology* 214.7 (Sept. 2016), pp. 817–830.

[20] K. Izumikawa et al. “TDP-43 stabilises the processing intermediates of mitochondrial transcripts”. *Sci Rep* 7.1 (Dec. 2017), p. 7709.

[21] W. Wang et al. “The inhibition of TDP-43 mitochondrial localization blocks its neuronal toxicity”. *Nat Med* 22.8 (Aug. 2016), pp. 869–878.

[22] B. Modrek and C. Lee. “A genomic view of alternative splicing”. *Nat Genet* 30.1 (Jan. 2002), pp. 13–19.

[23] H. K. Kim et al. “Alternative splicing isoforms in health and disease”. *Pflugers Arch - Eur J Physiol* 470.7 (July 2018), pp. 995–1016.

[24] J. P. Ling et al. “TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD”. *Science* 349.6248 (Aug. 2015), pp. 650–655.

[25] A. L. Brown et al. *Common ALS/FTD risk variants in UNC13A exacerbate its cryptic splicing and loss upon TDP-43 mislocalization.* preprint. Neuroscience, Apr. 2021.

[26] M. Irimia et al. “A highly conserved program of neuronal microexons is misregulated in autistic brains”. *Cell* 159.7 (Dec. 2014), pp. 1511–1523.

[27] C. S. Bland et al. “Global regulation of alternative splicing during myogenic differentiation”. *Nucleic Acids Research* 38.21 (Nov. 2010), pp. 7651–7664.

[28] M. J. Moore et al. “An alternative splicing network links cell-cycle control to apoptosis”. *Cell* 142.4 (Aug. 2010), pp. 625–636.

[29] D. Baralle and E. Buratti. “RNA splicing in human disease and in the clinic”. *Clinical Science* 131.5 (Mar. 2017), pp. 355–368.

[30] K. Nakka et al. “Diversification of the muscle proteome through alternative splicing”. *Skeletal Muscle* 8.1 (Dec. 2018), p. 8.

[31] M. C. Wahl, C. L. Will, and R. Lührmann. “The spliceosome: design principles of a dynamic RNP machine”. *Cell* 136.4 (Feb. 2009), pp. 701–718.

[32] H. Dvinge et al. “RNA splicing factors as oncoproteins and tumour suppressors”. *Nature Reviews Cancer* 16.7 (2016), p. 413.

[33] A. R. Kornblihtt et al. “Alternative splicing: a pivotal step between eukaryotic transcription and translation”. *Nat Rev Mol Cell Biol* 14.3 (Mar. 2013), pp. 153–165.

[34] X. H.-F. Zhang. “Dichotomous splicing signals in exon flanks”. *Genome Research* 15.6 (May 2005), pp. 768–779.

[35] R. Singh and J. Valcárcel. “Building specificity with nonspecific RNA-binding proteins”. *Nature structural & molecular biology* 12.8 (2005), p. 645.

[36] L. De Conti, M. Baralle, and E. Buratti. “Exon and intron definition in pre-mRNA splicing”. *Wiley Interdisciplinary Reviews: RNA* 4.1 (2013), pp. 49–60.

[37] X.-D. Fu and M. Ares Jr. “Context-dependent control of alternative splicing by RNA-binding proteins”. *Nature Reviews Genetics* 15.10 (2014), p. 689.

[38] J. P. Venables. “Downstream intronic splicing enhancers”. *FEBS letters* 581.22 (2007), pp. 4127–4131.

[39] Y. Barash et al. “Deciphering the splicing code”. *Nature* 465.7294 (2010), p. 53.

[40] E. T. Wang et al. “Alternative isoform regulation in human tissue transcriptomes”. *Nature* 456.7221 (2008), p. 470.

[41] M. Mele et al. “The human transcriptome across tissues and individuals”. *Science* 348.6235 (May 2015), pp. 660–665.

[42] B. Raj and B. J. Blencowe. “Alternative splicing in the mammalian nervous system: recent insights into mechanisms and functional roles”. *Neuron* 87.1 (July 2015), pp. 14–27.

[43] G. W. Yeo et al. “Identification and analysis of alternative splicing events conserved in human and mouse”. *Proceedings of the National Academy of Sciences* 102.8 (2005), pp. 2850–2855.

[44] J. Ule et al. “Nova regulates brain-specific splicing to shape the synapse”. *Nature genetics* 37.8 (2005), p. 844.

[45] U. Braunschweig et al. “Widespread intron retention in mammals functionally tunes transcriptomes”. *Genome research* 24.11 (2014), pp. 1774–1786.

[46] A. Reyes et al. “Drift and conservation of differential exon usage across tissues in primate species”. *Proceedings of the National Academy of Sciences* 110.38 (Sept. 2013), pp. 15377–15382.

[47] M. Buljan et al. “Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks”. *Molecular cell* 46.6 (2012), pp. 871–883.

[48] C. Yang et al. “Partial loss of TDP-43 function causes phenotypes of amyotrophic lateral sclerosis”. *Medical Sciences* (2014), p. 9.

[49] Mouse Genome Sequencing Consortium. “Initial sequencing and comparative analysis of the mouse genome”. *Nature* 420.6915 (Dec. 2002), pp. 520–562.

[50] N. Kim et al. “The ASAP II database: analysis and comparative genomics of alternative splicing in 15 animal species”. *Nucleic Acids Research* 35 (2006), pp. 93–98.

[51] G. C. Fu and W. C. Lin. “Identification of gene-oriented exon orthology between human and mouse”. *BMC genomics*. Vol. 13. 2012, pp. 1–10.

[52] S. Gueroussov et al. “An alternative splicing event amplifies evolutionary differences between vertebrates”. *Science* 349.6250 (Aug. 2015), pp. 868–873.

[53] N. L. Barbosa-Morais et al. “The evolutionary landscape of alternative splicing in vertebrate species”. *Science* 338.6114 (2012), pp. 1587–1593.

[54] Q. Pan et al. “Alternative splicing of conserved exons is frequently species-specific in human and mouse”. *Trends in Genetics* 21.2 (2005), pp. 73–77.

[55] T. Thanaraj, F. Clark, and J. Muilu. “Conservation of human alternative splice events in mouse”. *Nucleic Acids Research* 31.10 (2003), pp. 2544–2552.

[56] F. Zambelli et al. “Assessment of orthologous splicing isoforms in human and mouse orthologous genes”. *BMC genomics* 11.1 (2010), p. 534.

[57] E. S. Arnold et al. “ALS-linked TDP-43 mutations produce aberrant RNA splicing and adult-onset motor neuron disease without aggregation or loss of nuclear TDP-43”. *Proceedings of the National Academy of Sciences* 110.8 (Feb. 2013), pp. 736–745.

[58] A. Breschi, T. R. Gingeras, and R. Guigó. “Comparative transcriptomics in human and mouse”. *Nature Reviews Genetics* 18.7 (2017), p. 425.

[59] S. Shen et al. “rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data”. *Proc Natl Acad Sci USA* 111.51 (Dec. 2014), pp. 5593–5601.

[60] P. K. Thakur et al. *Bioinformatics approaches for studying alternative splicing.* Academic Press, 2019.

[61] R. Stark, M. Grzelak, and J. Hadfield. "RNA sequencing: the teenage years." *Nature reviews. Genetics* (2019).

[62] Y. Katz et al. "Quantitative visualization of alternative exon expression from RNA-seq data". *Bioinformatics* 31.14 (2015), pp. 2400–2402.

[63] Y. Katz et al. "Analysis and design of RNA sequencing experiments for identifying isoform regulation". *Nature methods* 7.12 (2010), p. 1009.

[64] Y. Liu, J. Zhou, and K. P. White. "RNA-seq differential expression studies: more sequence or more replication?" *Bioinformatics* 30.3 (2013), pp. 301–304.

[65] A. Mehmood et al. "Systematic evaluation of differential splicing tools for RNA-seq studies". *Briefings in Bioinformatics* 21.6 (Dec. 2020), pp. 2052–2065.

[66] E. Buratti. "TDP-43 post-translational modifications in health and disease". *Expert Opinion on Therapeutic Targets* 22.3 (Mar. 2018), pp. 279–293.

[67] Y. Chornenkyy, D. W. Fardo, and P. T. Nelson. "Tau and TDP-43 proteinopathies: kindred pathologic cascades and genetic pleiotropy". *Lab Invest* 99.7 (July 2019), pp. 993–1007.

[68] F. Paron, A. Dardis, and E. Buratti. "Pre-mRNA splicing defects and RNA binding protein involvement in Niemann Pick type C disease". *Journal of Biotechnology* 318 (July 2020), pp. 20–30.

[69] E. B. Lee, V. M.-Y. Lee, and J. Q. Trojanowski. "Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration". *Nature Reviews Neuroscience* 13.1 (2012), p. 38.

[70] P. Fratta et al. "Mice with endogenous TDP-43 mutations exhibit gain of splicing function and characteristics of amyotrophic lateral sclerosis". *EMBO J* 37.11 (June 2018).

[71] R. Cascella et al. "Quantification of the relative contributions of loss-of-function and gain-of-function mechanisms in TAR DNA-binding protein 43 (TDP-43) proteinopathies". *J. Biol. Chem.* 291.37 (Sept. 2016), pp. 19437–19448.

[72] M. Budini et al. "TDP-43 loss of cellular function through aggregation requires additional structural determinants beyond its C-terminal Q/N prion-like domain". *Human Molecular Genetics* 24.1 (Jan. 2015), pp. 9–20.

[73] M. C. Wollerton et al. "Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay". *Molecular cell* 13.1 (2004), pp. 91–100.

[74] E. Buratti and F. E. Baralle. "TDP-43: new aspects of autoregulation mechanisms in RNA binding proteins and their connection with human disease". *The FEBS journal* 278.19 (2011), pp. 3530–3538.

[75] L. M. Igaz et al. "Enrichment of C-terminal fragments in TAR DNA-binding protein-43 cytoplasmic inclusions in brain but not in spinal cord of frontotemporal lobar degeneration and amyotrophic lateral sclerosis". *The American journal of pathology* 173.1 (2008), pp. 182–194.

[76] J. P. Taylor, R. H. Brown Jr, and D. W. Cleveland. "Decoding ALS: from genes to mechanism". *Nature* 539.7628 (2016), p. 197.

[77] A. Al-Chalabi and O. Hardiman. "The epidemiology of ALS: a conspiracy of genes, environment and time". *Nat Rev Neurol* 9.11 (Nov. 2013), pp. 617–628.

[78] M. D. Cykowski et al. "Phosphorylated TDP-43 (pTDP-43) aggregates in the axial skeletal muscle of patients with sporadic and familial amyotrophic lateral sclerosis". *Acta neuropathologica communications* 6.1 (2018), p. 28.

[79] G. Sorarú et al. "TDP-43 in skeletal muscle of patients affected with amyotrophic lateral sclerosis". *Amyotrophic Lateral Sclerosis* 11.1-2 (2010), pp. 240–243.

[80] J.-P. Loeffler et al. "The role of skeletal muscle in amyotrophic lateral sclerosis". *Brain Pathology* 26.2 (2016), pp. 227–236.

[81] M. P. Mattson, R. G. Cutler, and S. Camandola. "Energy intake and amyotrophic lateral sclerosis". *NMM* 9.1 (2007), pp. 17–20.

[82] A. Callan et al. "A systematic review and meta-analysis of prevalence studies of sporadic inclusion body myositis". *Journal of neuromuscular diseases* 4.2 (2017), pp. 127–137.

[83] S. A. Greenberg. "Inclusion body myositis". *Current opinion in rheumatology* 23.6 (2011), pp. 574–578.

[84] S. A. Greenberg. "Inclusion body myositis: clinical features and pathogenesis". *Nature Reviews Rheumatology* (2019), p. 1.

[85] C. W. Keller, J. Schmidt, and J. D. Lünemann. "Immune and myodegenerative pathomechanisms in inclusion body myositis". *Annals of clinical and translational neurology* 4.6 (2017), pp. 422–445.

[86] S. Nakano et al. "Histone H1 is released from myonuclei and present in rimmed vacuoles with DNA in inclusion body myositis". *Neuromuscular Disorders* 18.1 (2008), pp. 27–33.

[87] K. A. Rygiel et al. "Mitochondrial and inflammatory changes in sporadic inclusion body myositis". *Neuropathology and applied neurobiology* 41.3 (2015), pp. 288–303.

[88] A. M. Afzali et al. "Animal models in idiopathic inflammatory myopathies: How to overcome a translational roadblock?" *Autoimmunity reviews* 16.5 (2017), pp. 478–494.

[89] A. Hiniker et al. “Comparative utility of LC3, p62 and TDP-43 immunohistochemistry in differentiation of inclusion body myositis from polymyositis and related inflammatory myopathies”. *Acta neuropathologica communications* 1.1 (2013), p. 29.

[90] M. Salajegheh et al. “Sarcoplasmic redistribution of nuclear TDP-43 in inclusion body myositis: redistribution of TDP-43”. *Muscle Nerve* 40.1 (July 2009), pp. 19–31.

[91] C. C. Weihl et al. “TDP-43 accumulation in inclusion body myopathy muscle suggests a common pathogenic mechanism with frontotemporal dementia”. *Journal of Neurology, Neurosurgery & Psychiatry* 79.10 (Oct. 2008), pp. 1186–1189.

[92] A. Verma and R. Tandan. “TDP-43: A reliable immunohistochemistry marker for inclusion body myositis?” *Muscle & nerve* 40.1 (2009), pp. 8–9.

[93] G. D. Watts et al. “Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein”. *Nature genetics* 36.4 (2004), p. 377.

[94] C. C. Weihl and A. Pestronk. “Sporadic inclusion body myositis: possible pathogenesis inferred from biomarkers”. *Current opinion in neurology* 23.5 (2010), p. 482.

[95] A. S. Correia et al. “Inflammation induces TDP-43 mislocalization and aggregation”. *PLoS One* 10.10 (2015), p. 0140248.

[96] J. L. Pinkus et al. “Abnormal distribution of heterogeneous nuclear ribonucleoproteins in sporadic inclusion body myositis”. *Neuromuscular Disorders* 24.7 (2014), pp. 611–616.

[97] A. Shiga et al. “Alteration of POLDIP3 splicing associated with loss of function of TDP-43 in tissues affected with ALS”. *PLoS ONE* 7.8 (Aug. 2012), p. 43120.

[98] C. Yang et al. “Partial loss of TDP-43 function causes phenotypes of amyotrophic lateral sclerosis”. *Proceedings of the National Academy of Sciences* 111.12 (2014), pp. 1121–1129.

[99] J. R. Tollervey et al. “Analysis of alternative splicing associated with aging and neurodegeneration in the human brain”. *Genome Research* 21.10 (Oct. 2011), pp. 1572–1582.

[100] J. Humphrey et al. “Quantitative analysis of cryptic splicing associated with TDP-43 depletion”. *BMC Med Genomics* 10.1 (Dec. 2017), p. 38.

[101] S. J. Rabin et al. “Sporadic ALS has compartment-specific aberrant exon splicing and altered cell–matrix adhesion biology”. *Human molecular genetics* 19.2 (2009), pp. 313–328.

[102] A. Cortese et al. “Widespread RNA metabolism impairment in sporadic inclusion body myositis TDP43-proteinopathy”. *Neurobiology of Aging* 35.6 (June 2014), pp. 1491–1498.

[103] Y. B. Luo et al. “Investigation of splicing changes and post-translational processing of LMNA in sporadic inclusion body myositis”. *International journal of clinical and experimental pathology* 6.9 (2013), p. 1723.

[104] J. R. Highley et al. “Loss of nuclear TDP-43 in amyotrophic lateral sclerosis (ALS) causes altered expression of splicing machinery and widespread dysregulation of RNA splicing in motor neurones: Amyotrophic lateral sclerosis, TDP-43 and mRNA splicing”. *Neuropathol Appl Neurobiol* 40.6 (Oct. 2014), pp. 670–685.

[105] L. Dupuis et al. “Energy metabolism in amyotrophic lateral sclerosis”. *The Lancet Neurology* 10.1 (2011), pp. 75–82.

[106] S. E. Kirk et al. “Biomarkers of metabolism in amyotrophic lateral sclerosis”. *Frontiers in neurology* 10 (2019).

[107] M. Szelechowski et al. “Metabolic reprogramming in amyotrophic lateral sclerosis”. *Scientific reports* 8.1 (2018), p. 3953.

[108] N. D. Perera and B. J. Turner. “AMPK signalling and defective energy metabolism in amyotrophic lateral sclerosis”. *Neurochemical research* 41.3 (2016), pp. 544–553.

[109] Y. J. Liu et al. “Activation of AMP-activated protein kinase α1 mediates mislocalization of TDP-43 in amyotrophic lateral sclerosis”. *Human Molecular Genetics* 24.3 (Feb. 2015), pp. 787–801.

[110] N. D. Perera et al. “Mutant TDP-43 deregulates AMPK activation by PP2A in ALS models”. *PloS one* 9.3 (2014), p. 90449.

[111] X. Shan et al. “Altered distributions of Gemini of coiled bodies and mitochondria in motor neurons of TDP-43 transgenic mice”. *Proceedings of the National Academy of Sciences* 107.37 (Sept. 2010), pp. 16325–16330.

[112] Y. F. Xu et al. “Wild-type human TDP-43 expression causes TDP-43 phosphorylation, mitochondrial aggregation, motor deficits, and early mortality in transgenic mice”. *Journal of Neuroscience* 30.32 (2010), pp. 10851–10859.

[113] M. L. Huntley et al. “Association between TDP-43 and mitochondria in inclusion body myositis”. *Laboratory Investigation* (2019), p. 1.

[114] P. M. Chiang et al. “Deletion of TDP-43 down-regulates Tbc1d1, a gene linked to obesity, and alters body fat metabolism”. *Proceedings of the National Academy of Sciences* 107.37 (2010), pp. 16320–16324.

[115] A. Dobin et al. “STAR: ultrafast universal RNA-seq aligner”. *Bioinformatics* 29.1 (Jan. 2013), pp. 15–21.

[116] S. Anders, P. T. Pyl, and W. Huber. “HTSeq python framework to work with high-throughput sequencing data”. *Bioinformatics* 31.2 (Jan. 2015), pp. 166–169.

[117] A. Mortazavi et al. “Mapping and quantifying mammalian transcriptomes by RNA-seq”. *Nat Methods* 5.7 (July 2008), pp. 621–628.

[118] S. Anders and W. Huber. "Differential expression analysis for sequence count data". *Genome Biol* 11.10 (Oct. 2010), R106.

[119] M. I. Love, W. Huber, and S. Anders. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". *Genome Biol* 15.12 (Dec. 2014), p. 550.

[120] R. Kolde. *pheatmap: Pretty Heatmaps*. Jan. 2019.

[121] R. C. Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2019.

[122] Y. Hu et al. "DiffSplice: the genome-wide detection of differential splicing events with RNA-seq". *Nucleic Acids Research* 41.2 (Jan. 2013), e39–e39.

[123] C. Trapnell et al. "Differential analysis of gene regulation at transcript resolution with RNA-seq". *Nat Biotechnol* 31.1 (Jan. 2013), pp. 46–53.

[124] S. Anders, A. Reyes, and W. Huber. "Detecting differential usage of exons from RNA-seq data". *Genome Research* 22.10 (Oct. 2012), pp. 2008–2017.

[125] C. J. Green, M. R. Gazzara, and Y. Barash. "MAJIQ-SPEL: web-tool to interrogate classical and complex splicing variations from RNA-Seq data". *Bioinformatics* 34.2 (Jan. 2018), pp. 300–302.

[126] M. Ashburner et al. "Gene ontology: tool for the unification of biology". *Nat Genet* 25.1 (May 2000), pp. 25–29.

[127] M. Kanehisa. "KEGG: Kyoto encyclopedia of genes and genomes". *Nucleic Acids Research* 28.1 (Jan. 2000), pp. 27–30.

[128] G. Yu et al. "ClusterProfiler: an R package for comparing biological themes among gene clusters". *OMICS: A Journal of Integrative Biology* 16.5 (May 2012), pp. 284–287.

[129] K. S. Pollard et al. "Detection of nonneutral substitution rates on mammalian phylogenies". *Genome Research* 20.1 (Jan. 2010), pp. 110–121.

[130] A. D. Yates et al. "Ensembl 2020". *Nucleic Acids Research* (Nov. 2019).

[131] R. C. Griggs et al. "Inclusion body myositis and myopathies". *Ann Neurol.* 38.5 (Nov. 1995), pp. 705–713.

[132] M. Prudencio et al. "Truncated stathmin-2 is a marker of TDP-43 pathology in frontotemporal dementia". *Journal of Clinical Investigation* 130.11 (Oct. 2020), pp. 6080–6092.

[133] V. Askanas, W. K. Engel, and A. Nogalska. "Pathogenic considerations in sporadic inclusion-body myositis, a degenerative muscle disease associated with aging and abnormalities of myoproteostasis". *J Neuropathol Exp Neurol* 71.8 (Aug. 2012), pp. 680–693.

[134] S. Yamashita et al. "CYLD dysregulation in pathogenesis of sporadic inclusion body myositis". *Sci Rep* 9.1 (Dec. 2019), p. 11606.

[135] T. O. Vogler et al. "TDP-43 and RNA form amyloid-like myo-granules in regenerating muscle". *Nature* 563.7732 (Nov. 2018), pp. 508–513.

[136] C. Appocher et al. "Major hnRNP proteins act as general TDP-43 functional modifiers both in Drosophila and human neuronal cells". *Nucleic Acids Research* 45.13 (July 2017), pp. 8026–8045.

[137] S. Cappelli, M. Romano, and E. Buratti. "Systematic analysis of gene expression profiles controlled by hnRNP Q and hnRNP R, two closely related human RNA binding proteins implicated in mRNA processing mechanisms". *Front. Mol. Biosci.* 5 (Aug. 2018), p. 79.

[138] M. N. Cabili et al. "Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses". *Genes & Development* 25.18 (Sept. 2011), pp. 1915–1927.

[139] C. Jiang et al. "Identifying and functionally characterizing tissue-specific and ubiquitously expressed human lncRNAs". *Oncotarget* 7.6 (Feb. 2016), pp. 7120–7133.

[140] Y. H. Jeong et al. "Tdp-43 cryptic exons are highly variable between cell types". *Mol Neurodegeneration* 12.1 (Dec. 2017), p. 13.

[141] N. R. Cashman et al. "Neuroblastoma × spinal cord (NSC) hybrid cell lines resemble developing motor neurons". *Dev. Dyn.* 194.3 (July 1992), pp. 209–221.

[142] C. Colombrita et al. "TDP-43 is recruited to stress granules in conditions of oxidative insult". *Journal of Neurochemistry* 111.4 (Nov. 2009), pp. 1051–1061.

[143] M. B. Yaffe et al. "The structural basis for 14-3-3:phosphopeptide binding specificity". *Cell* 91.7 (Dec. 1997), pp. 961–971.

[144] M. Kitazawa, D. N. Trinh, and F. M. LaFerla. "Inflammation induces tau pathology in inclusion body myositis model via glycogen synthase kinase-3β". *Ann Neurol.* 64.1 (July 2008), pp. 15–24.

[145] H. Pluk et al. "Autoantibodies to cytosolic 5'-nucleotidase 1A in inclusion body myositis: anti-cN1A autoantibodies". *Ann Neurol.* 73.3 (Mar. 2013), pp. 397–407.

[146] N. Tawara et al. "Muscle-dominant wild-type TDP-43 expression induces myopathological changes featuring tubular aggregates and TDP-43-positive inclusions". *Experimental Neurology* 309 (Nov. 2018), pp. 169–180.

[147] G. Militello et al. "A novel long non-coding RNA Myolinc regulates myogenesis through TDP-43 and Filip1". *Journal of Molecular Cell Biology* 10.2 (Apr. 2018), pp. 102–117.

[148] P. C. Chen, J. S. Ruan, and S. N. Wu. "Evidence of decreased activity in intermediate-conductance calcium-activated potassium channels during retinoic acid–induced differentiation in motor neuron–like NSC-34 caells". *Cellular Physiology and Biochemistry* 48.6 (2018), pp. 2374–2388.

[149] Y. M. Ayala et al. "Human, Drosophila, and C. elegans TDP43: nucleic acid binding properties and splicing regulatory function". *Journal of Molecular Biology* 348.3 (May 2005), pp. 575–588.

[150] S. L. Huang et al. "A robust TDP-43 knock-in mouse model of ALS". *Acta Neuropathol Commun* 8.1 (Dec. 2020), p. 3.

[151] F. C. Fiesel et al. "TDP-43 regulates global translational yield by splicing of exon junction complex component SKAR". *Nucleic Acids Research* 40.6 (Mar. 2012), pp. 2668–2682.

[152] F. Mohagheghi et al. "TDP-43 functions within a network of hnRNP proteins to inhibit the production of a truncated human SORT1 receptor". *Hum. Mol. Genet.* 25.3 (Feb. 2016), pp. 534–545.

[153] M. Prudencio et al. "Misregulation of human sortilin splicing leads to the generation of a nonfunctional progranulin receptor". *Proceedings of the National Academy of Sciences* 109.52 (Dec. 2012), pp. 21510–21515.

[154] F. C. Fiesel et al. "Knockdown of transactive response DNA-binding protein (TDP-43) downregulates histone deacetylase 6". *EMBO J* 29.1 (Jan. 2010), pp. 209–221.

[155] Y. Kawahara and A. Mieda-Sato. "TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes". *Proceedings of the National Academy of Sciences* 109.9 (Feb. 2012), pp. 3347–3352.

[156] E. Lundberg et al. "Defining the transcriptome and proteome in three functionally different human cell lines". *Mol Syst Biol* 6.1 (Jan. 2010), p. 450.

[157] E. M. Wilson and P. Rotwein. "Control of MyoD function during initiation of muscle differentiation by an autocrine signaling pathway activated by insulin-like growth factor-II". *Journal of Biological Chemistry* 281.40 (Oct. 2006), pp. 29962–29971.

[158] J. Tomida et al. "DNA damage-induced ubiquitylation of RFC2 subunit of replication factor C complex". *Journal of Biological Chemistry* 283.14 (Apr. 2008), pp. 9071–9079.

[159] X. Miliara et al. "Structural insight into the TRIAP1/PRELI-like domain family of mitochondrial phospholipid transfer complexes". *EMBO Rep* 16.7 (July 2015), pp. 824–835.

[160] K. Sasaki et al. "Rho guanine nucleotide exchange factors regulate horizontal axon branching of cortical upper layer neurons". *Cerebral Cortex* 30.4 (Apr. 2020), pp. 2506–2518.

[161] S. J. Rabin et al. “Sporadic ALS has compartment-specific aberrant exon splicing and altered cell–matrix adhesion biology”. *Human Molecular Genetics* 19.2 (Jan. 2010), pp. 313–328.

[162] C. Colombrita et al. “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”. *Journal of Biological Chemistry* 287.19 (May 2012), pp. 15635–15647.

[163] J. Humphrey et al. *FUS ALS-causative mutations impact FUS autoregulation and the processing of RNA-binding proteins through intron retention.* preprint. Neuroscience, Mar. 2019.

[164] L. De Conti et al. “TDP-43 affects splicing profiles and isoform production of genes involved in the apoptotic and mitotic cellular pathways”. *Nucleic Acids Res* 43.18 (Oct. 2015), pp. 8990–9005.

[165] C. Lagier-Tourenne et al. “Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs”. *Nat Neurosci* 15.11 (Nov. 2012), pp. 1488–1497.

[166] E. Buratti, M. Romano, and F. E. Baralle. “TDP-43 high throughput screening analyses in neurodegeneration: advantages and pitfalls”. *Molecular and Cellular Neuroscience* 56 (Sept. 2013), pp. 465–474.

[167] A. Cortese et al. “Altered TDP-43-dependent splicing in HSPB8-related distal hereditary motor neuropathy and myofibrillar myopathy”. *Eur J Neurol* 25.1 (Jan. 2018), pp. 154–163.

[168] J. R. Klim et al. “ALS-implicated protein TDP-43 sustains levels of STMN2, a mediator of motor neuron growth and repair”. *Nat Neurosci* 22.2 (Feb. 2019), pp. 167–179.

[169] A. Roczniak-Ferguson and S. M. Ferguson. “Pleiotropic requirements for human TDP-43 in the regulation of cell and organelle homeostasis”. *Life Science Alliance* 2.5 (2019).

[170] J. Tapial et al. “An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms”. *Genome Res.* 27.10 (Oct. 2017), pp. 1759–1768.

[171] R. D. Burgoyne and A. Morgan. “Cysteine string protein (CSP) and its role in preventing neurodegeneration”. *Seminars in Cell & Developmental Biology* 40 (Apr. 2015), pp. 153–159.

[172] A. Perfetti et al. “Genome-wide identification of aberrant alternative splicing events in myotonic dystrophy type 2”. *PLoS ONE* 9.4 (Apr. 2014), p. 93983.

[173] B. M. Schwenk et al. “TDP-43 loss of function inhibits endosomal trafficking and alters trophic signaling in neurons”. *EMBO J* 35.21 (Nov. 2016), pp. 2350–2370.

[174] C. M. Gallo, A. Ho, and U. Beffert. "ApoER2: functional tuning through splicing". *Front. Mol. Neurosci.* 13 (July 2020), p. 144.

[175] R. K. Singh et al. "Rbfox2-coordinated alternative splicing of Mef2d and Rock2 controls myoblast fusion during myogenesis". *Molecular Cell* 55.4 (Aug. 2014), pp. 592–603.

[176] G. Yeo et al. "Variation in alternative splicing across human tissues". *Genome Biology* (2004), p. 15.

[177] V. Askanas, W. K. Engel, and A. Nogalska. "Sporadic inclusion-body myositis: a degenerative muscle disease associated with aging, impaired muscle protein homeostasis and abnormal mitophagy". *Biochimica et Biophysica Acta - Molecular Basis of Disease* 1852.4 (Apr. 2015), pp. 633–643.

[178] R. Barrès et al. "Acute exercise remodels promoter methylation in human skeletal muscle". *Cell Metabolism* 15.3 (Mar. 2012), pp. 405–411.

[179] S. L. McGee and M. Hargreaves. "Histone modifications and exercise adaptations". *Journal of Applied Physiology* 110.1 (Jan. 2011), pp. 258–263.

[180] V. Guasconi and P. L. Puri. "Chromatin: the interface between extrinsic cues and the epigenetic regulation of muscle regeneration". *Trends in Cell Biology* 19.6 (June 2009), pp. 286–294.

[181] R. Morosetti et al. "Increased aging in primary muscle cultures of sporadic inclusion-body myositis". *Neurobiology of Aging* 31.7 (July 2010), pp. 1205–1214.

[182] O. Rokach et al. "Epigenetic changes as a common trigger of muscle weakness in congenital myopathies". *Hum. Mol. Genet.* 24.16 (Aug. 2015), pp. 4636–4647.

[183] M. Giannini et al. "TDP-43 mutations link amyotrophic lateral sclerosis with R-loop homeostasis and R loop-mediated DNA damage". *PLoS Genet* 16.12 (Dec. 2020). Ed. by D. A. Gordenin, e1009260.

[184] M. Wood et al. "TDP-43 dysfunction results in R-loop accumulation and DNA replication defects". *J Cell Sci* 133.20 (Oct. 2020), jcs244129.

[185] A. Björkman et al. "Human RTEL1 associates with Poldip3 to facilitate responses to replication stress and R-loop resolution". *Genes Dev.* 34.15-16 (Aug. 2020), pp. 1065–1074.

[186] C. Girard et al. "Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion". *Nat Commun* 3.1 (Jan. 2012), p. 994.

[187] C. Fallini, G. J. Bassell, and W. Rossoll. "The ALS disease protein TDP-43 is actively transported in motor neuron axons and regulates axon outgrowth". *Human Molecular Genetics* 21.16 (Aug. 2012), pp. 3703–3718.

[188] J.-H. Han et al. "ALS/FTLD-linked TDP-43 regulates neurite morphology and cell survival in differentiated neurons". *Experimental Cell Research* 319.13 (Aug. 2013), pp. 1998–2005.

[189] V. Di Carlo et al. "TDP-43 regulates the microprocessor complex activity during in vitro neuronal differentiation". *Mol Neurobiol* 48.3 (Dec. 2013), pp. 952–963.

[190] J. N. Sleigh et al. "Mice carrying ALS mutant TDP-43, but not mutant FUS, display in vivo defects in axonal transport of signaling endosomes". *Cell Reports* 30.11 (Mar. 2020), pp. 3655–3662.

[191] T. Pavlidou et al. "Regulation of myoblast differentiation by metabolic perturbations induced by metformin". *PloS one* 12.8 (2017), e0182475.

[192] V. Andrés and K. Walsh. "Myogenin expression, cell cycle withdrawal, and phenotypic differentiation are temporally separable events that precede cell fusion upon myogenesis." *Journal of Cell Biology* 132.4 (Feb. 1996), pp. 657–666.

[193] I. Hatfield et al. "The role of TORC1 in muscle development in Drosophila". *Scientific reports* 5.1 (2015), pp. 1–11.

[194] N. R. Stallings et al. "TDP-43, an ALS linked protein, regulates fat deposition and glucose homeostasis". *PLoS ONE* 8.8 (Aug. 2013), p. 71793.

[195] K. S. Hubbard et al. "Longitudinal RNA sequencing of the deep transcriptome during neurogenesis of cortical glutamatergic neurons from murine ESCs". *F1000Res* 2 (2013), p. 35.

[196] A. Chadt et al. "Tbc1d1 mutation in lean mouse strain confers leanness and protects from diet-induced obesity". *Nat Genet* 40.11 (Nov. 2008), pp. 1354–1359.

[197] S. Mafakheri et al. "AKT and AMP-activated protein kinase regulate TBC1D1 through phosphorylation and its interaction with the cytosolic tail of insulin-regulated aminopeptidase IRAP". *Journal of Biological Chemistry* 293.46 (Nov. 2018), pp. 17853–17862.

[198] E. B. Taylor et al. "Discovery of TBC1D1 as an insulin-, AICAR-, and contraction-stimulated signaling nexus in mouse skeletal muscle". *Journal of Biological Chemistry* 283.15 (Apr. 2008), pp. 9787–9796.

[199] K. Vichaiwong et al. "Contraction regulates site-specific phosphorylation of TBC1D1 in skeletal muscle". *Biochemical Journal* 431.2 (Oct. 2010), pp. 311–320.

[200] J. T. Treebak et al. "Acute exercise and physiological insulin induce distinct phosphorylation signatures on TBC1D1 and TBC1D4 proteins in human skeletal muscle: Phosphorylation signatures of TBC1D1 and TBC1D4 in skeletal muscle". *The Journal of Physiology* 592.2 (Jan. 2014), pp. 351–375.

[201] L. Fontanesi and F. Bertolini. "The TBC1D1 gene". In: *Vitamins & Hormones*. Vol. 91. Elsevier, 2013, pp. 77–95.

[202] R. Koopman, C. H. Ly, and J. G. Ryall. "A metabolic link to skeletal muscle wasting and regeneration". *Front. Physiol.* 5 (2014).

[203] C. J. David et al. "HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer". *Nature* 463.7279 (Jan. 2010), pp. 364–368.

[204] The Cancer Genome Atlas Research Network et al. "The cancer genome atlas pan-cancer analysis project". *Nat Genet* 45.10 (Oct. 2013), pp. 1113–1120.

[205] O. Warburg. "On the origin of cancer cells". *Science* 123.3191 (Feb. 1956), pp. 309–314.

[206] J. G. Ryall. "Metabolic reprogramming as a novel regulator of skeletal muscle development and regeneration". *FEBS J* 280.17 (Sept. 2013), pp. 4004–4013.

[207] J. G. Ryall, J. D. Schertzer, and G. S. Lynch. "Cellular and molecular mechanisms underlying age-related skeletal muscle wasting and weakness". *Biogerontology* 9.4 (Aug. 2008), pp. 213–228.

[208] H. Y. Wang et al. "Structural diversity and functional implications of the eukaryotic TDP gene family". *Genomics* 83.1 (Jan. 2004), pp. 130–139.

[209] N. Strah et al. "TDP-43 promotes the formation of neuromuscular synapses through the regulation of Disc-large expression in Drosophila skeletal muscles". *BMC Biol* 18.1 (Dec. 2020), p. 34.

[210] B. S. Johnson et al. "A yeast TDP-43 proteinopathy model: exploring the molecular determinants of TDP-43 aggregation and cellular toxicity". *Proceedings of the National Academy of Sciences* 105.17 (Apr. 2008), pp. 6439–6444.

[211] C. Stribl et al. "Mitochondrial dysfunction and decrease in body weight of a transgenic knock-in mouse model for TDP-43". *J. Biol. Chem.* 289.15 (Apr. 2014), pp. 10769–10784.

[212] M. Cardoso-Moreira et al. "Developmental gene expression differences between humans and mammalian models". *Cell Reports* 33.4 (Oct. 2020), p. 108308.

[213] P. A. Mercado. "Depletion of TDP 43 overrides the need for exonic and intronic splicing enhancers in the human apoA-II gene". *Nucleic Acids Research* 33.18 (Oct. 2005), pp. 6000–6010.

[214] F. Pagani et al. "Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element". *J. Biol. Chem.* 275.28 (July 2000), pp. 21041–21047.

[215] N. L. Barbosa-Morais et al. "The evolutionary landscape of alternative splicing in vertebrate species". *Science* 338.6114 (Dec. 2012), pp. 1587–1593.

[216] P. Khaitovich et al. "Evolution of primate gene expression". *Nat Rev Genet* 7.9 (Sept. 2006), pp. 693–702.

[217] M. Wainberg, B. Alipanahi, and B. Frey. "Does conservation account for splicing patterns?" *BMC Genomics* 17.1 (Dec. 2016), p. 787.

[218] J. Kovalevich and D. Langford. "Considerations for the use of SH-SY5Y neuroblastoma cells in neurobiology". In: *Neuronal Cell Culture*. Ed. by S. Amini and M. K. White. Vol. 1078. Series Title: Methods in Molecular Biology. Totowa, NJ: Humana Press, 2013, pp. 9–21.

[219] J. L. Biedler, L. Helson, and B. A. Spengler. "Morphology and growth, tumorigenicity, and cytogenetics of human neuroblastoma cells in continuous culture". *Cancer Research* 33.11 (1973), pp. 2643–2652.

[220] A. Russo et al. "Increased cytoplasmic TDP-43 reduces global protein synthesis by interacting with RACK1 on polyribosomes". *Human Molecular Genetics* 26.8 (Apr. 2017), pp. 1407–1418.

[221] L. S. Wu, W. C. Cheng, and C. K. Shen. "Similar dose-dependence of motor neuron cell death caused by wild type human TDP-43 and mutants with ALS-associated amino acid substitutions". *J Biomed Sci* 20.1 (2013), p. 33.

[222] D. P. Gitterman, J. Wilson, and A. D. Randall. "Functional properties and pharmacological inhibition of ASIC channels in the human SJ-RH30 skeletal muscle cell line: ASIC channels in a skeletal muscle cell line". *The Journal of Physiology* 562.3 (Feb. 2005), pp. 759–769.

[223] R. Saab, S. L. Spunt, and S. X. Skapek. "Myogenesis and rhabdomyosarcoma". In: *Current Topics in Developmental Biology*. Vol. 94. Elsevier, 2011, pp. 197–234.

[224] S. Reber et al. "Minor intron splicing is regulated by FUS and affected by ALS-associated FUS mutants". *EMBO J* 35.14 (July 2016), pp. 1504–1521.

[225] C. Colombrita et al. "From transcriptomic to protein level changes in TDP-43 and FUS loss-of-function cell models". *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1849.12 (Dec. 2015), pp. 1398–1410.

[226] D. Smedley et al. "The BioMart community portal: an innovative alternative to large, centralized data repositories". *Nucleic Acids Res* 43.W1 (July 2015), pp. 589–598.

[227] R. B. Parsons. "Comment on: "Cytotoxicity of oxycodone and morphine in human neuroblastoma and mouse motoneuronal cells: a comparative approach"". *Drugs R D* 16.3 (Sept. 2016), pp. 285–286.

[228] Z. Melamed et al. "Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration". *Nat Neurosci* 22.2 (Feb. 2019), pp. 180–190.

[229] R. H. Tan et al. "TDP-43 proteinopathies: pathological identification of brain regions differentiating clinical phenotypes". *Brain* 138.10 (Oct. 2015), pp. 3110–3122.

[230] J. Brettschneider et al. “Stages of pTDP-43 pathology in amyotrophic lateral sclerosis: ALS Stages”. *Ann Neurol.* 74.1 (July 2013), pp. 20–38.

[231] J. Brettschneider et al. “Sequential distribution of pTDP-43 pathology in behavioral variant frontotemporal dementia (bvFTD)”. *Acta Neuropathol* 127.3 (Mar. 2014), pp. 423–439.

[232] A. Feiglin et al. “Comprehensive analysis of tissue-wide gene expression and phenotype data reveals tissues affected in rare genetic disorders”. *Cell Systems* 5.2 (Aug. 2017), pp. 140–148.

[233] E. Y. Liu et al. “Loss of nuclear TDP-43 is associated with decondensation of LINE retrotransposons”. *Cell Reports* 27.5 (Apr. 2019), 1409–1421.e6.

[234] T. J. Cohen et al. “An acetylation switch controls TDP-43 function and aggregation propensity”. *Nature Communications* 6.1 (2015), pp. 1–13.

[235] G. Monteuuis et al. “The changing paradigm of intron retention: regulation, ramifications and recipes”. *Nucleic Acids Research* 47.22 (2019), pp. 11497–11513.

[236] T. Stermann et al. “Deletion of the RabGAP TBC1D1 leads to enhanced insulin secretion and fatty acid oxidation in islets from male mice”. *Endocrinology* 159.4 (Apr. 2018), pp. 1748–1761.

[237] M. Elkalaf, M. Anděl, and J. Trnka. “Low glucose but not galactose enhances oxidative mitochondrial metabolism in C2C12 myoblasts and myotubes”. *PLoS ONE* 8.8 (Aug. 2013), p. 70772.