TDP-43 dysregulation of polyadenylation site selection is a defining feature of RNA misprocessing in ALS/FTD and related disorders

Frederick J. Arnold[1] *, Ya Cui[2] *, Sebastian Michels[1,3] *, Michael R. Colwin[1], Cameron Stockford[1], Wenbin Ye[2], Oliver H. Tam[4,5], Sneha Menon[1], Wendy G. Situ[1], Kean C. K. Ehsani[1], Sierra Howard[1], Molly Gale Hammell[4,5], Wei Li[2] & Albert R. La Spada[1,2,6,7,8]

[1]Department of Pathology & Laboratory Medicine, University of California, Irvine; Irvine, CA 92697, USA

[2]Department of Biological Chemistry, University of California, Irvine; Irvine, CA 92697, USA

[3]Department of Neurology, University of Ulm, Ulm, Germany

[4]Institute for Systems Genetics, NYU Langone Health, New York, NY, 10016, USA

[5]Department of Neuroscience & Physiology, NYU Grossman School of Medicine, New York, NY, 10016, USA

[6]Department of Neurology, University of California, Irvine; Irvine, CA 92697, USA

[7]Department of Neurobiology & Behavior, University of California, Irvine; Irvine, CA 92697, USA

[8]UCI Center for Neurotherapeutics, University of California, Irvine; Irvine, CA 92697, USA

* Contributed equally as co-first authors

Corresponding author:
Albert R. La Spada, MD, PhD
Pathology & Laboratory Medicine, Neurology, Biological Chemistry and Neurobiology & Behavior
UCI Center for Neurotherapeutics
University of California Irvine
Irvine, CA, USA 92697
alaspada@uci.edu

Abstract

Nuclear clearance and cytoplasmic aggregation of the RNA-binding protein TDP-43 are observed in many neurodegenerative disorders, including amyotrophic lateral sclerosis (ALS) and fronto-temporal dementia (FTD). Although TDP-43 dysregulation of splicing has emerged as a key event in these diseases, TDP-43 can also regulate polyadenylation; yet, this has not been adequately studied. Here, we applied the dynamic analysis of polyadenylation from RNA-seq (DaPars) tool to ALS/FTD transcriptome datasets, and report extensive alternative polyadenylation (APA) upon TDP-43 alteration in ALS/FTD cell models and postmortem ALS/FTD neuronal nuclei. Importantly, many identified APA genes highlight pathways implicated in ALS/FTD pathogenesis. To determine the functional significance of APA elicited by TDP-43 nuclear depletion, we examined microtubule affinity regulating kinase 3 (MARK3). Nuclear loss of TDP-43 yielded increased expression of MARK3 transcripts with longer 3'UTRs, resulting in greater transcript stability and elevated MARK3 protein levels, which promotes increased neuronal tau S262 phosphorylation. Our findings define changes in polyadenylation site selection as a previously unrecognized feature of TDP-43-driven disease pathology in ALS/FTD and highlight a potentially novel mechanistic link between TDP-43 dysfunction and tau regulation.

While the underlying etiology of many age-related neurodegenerative disorders remains unknown, distinct neuronal subtypes and specific brain regions exhibit a common pathological hallmark of nuclear clearance and cytoplasmic aggregation of TAR DNA-binding protein 43 (TDP-43) [1]. TDP-43 pathology is present in motor neurons in ~97% of patients with amyotrophic lateral sclerosis (ALS) and in the neocortex of ~45% of individuals suffering from frontotemporal dementia (FTD). Furthermore, as many as 57% of Alzheimer's disease (AD) patients display TDP-43 pathology, which correlates with more rapid disease progression and worse cognitive impairment [2,3]. Cytoplasmic aggregates of TDP-43 have also been observed in Huntington's disease [4] and chronic traumatic encephalopathy [5], and are emerging as the defining pathology for limbic-predominant age-related TDP-43 encephalopathy (LATE), a recently described dementia characterized by atrophy in the medial temporal lobes and frontal cortex, though not exclusive to this brain region [6]. Consequently, heterogenous neurodegenerative disorders characterized by TDP-43 pathology may collectively be referred to as 'TDP-43 proteinopathies' [1].

Under physiological conditions, TDP-43 is a predominantly nuclear RNA-binding protein (RBP) that directly binds over 6,000 RNAs [7]. Prior studies of ALS/FTD have identified differentially expressed and alternatively spliced genes associated with TDP-43 nuclear loss of function. This includes a motor neuron growth and repair factor, stathmin-2 (STMN2), in which TDP-43 loss of nuclear function results in de-repression of a cryptic 3' splice site in the *STMN2* gene, favoring inclusion of a cryptic exon in the *STMN2* pre-mRNA, premature transcription termination, and loss of STMN2 protein expression [8-10]. Another gene subject to TDP-43 splicing dysregulation in ALS/FTD is *UNC13A*, as depletion of TDP-43 in the nucleus promotes cryptic exon inclusion within the *UNC13A* pre-mRNA, resulting in loss of UNC13A protein expression [11,12]. Both STMN2 and UNC13A are currently under investigation as candidates for therapeutic intervention and biomarker development in ALS/FTD.

In addition to its role in RNA splicing, TDP-43 regulates alternative polyadenylation (APA) by binding target RNAs near polyadenylation signals (PAS) [13], yet this aspect of TDP-43 function remains relatively unexplored in TDP-43 proteinopathies. Notably, up to 70% of all human pre-mRNAs contain more than one polyadenylation (poly(A)) site, and differential PAS utilization yields transcripts with distinct 3' untranslated regions (3'UTRs) [14,15]. APA is a highly conserved and ubiquitous mechanism of gene regulation that is exceedingly cell-type specific [16]. As it turns out, neurons tend to preferentially utilize distal poly(A) sites [17,18], accounting for the well-established observation that genes expressed in the central nervous system (CNS) encode transcripts containing the longest 3'UTR sequences of all tissues throughout the body. This

3

suggests that APA is likely an important post-transcriptional regulatory mechanism in neurons and other CNS cell types, as the 3'UTR is highly enriched for binding motifs for microRNAs (miRNAs) [19] and RBPs [20]. Hence, by modulating the presence or absence of such cis-regulatory elements, APA can produce transcript isoforms with substantial differences in RNA stability, subcellular localization, nuclear export, and transcription efficiency [21]. These changes in RNA metabolism can markedly impact protein expression, though it is also likely that changes in subcellular localization can modulate gene function without affecting steady-state RNA levels. An important consideration for the study of APA is that standard informatics analysis of RNA-seq data does not fully capture APA events. For this reason, a growing number of computational methods have been developed to infer PAS usage from standard, short-read, bulk RNA-seq transcriptome data, allowing post-hoc analysis of APA from existing datasets [22].

Here we applied the dynamic analysis of polyadenylation from RNA-sequencing (DaPars) tool to transcriptome data sets obtained from neuron cell culture models of ALS/FTD and from ALS/FTD patient postmortem neuron nuclei. We identified hundreds of previously unknown genes regulated by TDP-43, with many genes conserved across multiple cell types and model systems. As a substantial number of these genes function in cellular pathways implicated in ALS/FTD disease pathogenesis, APA analysis can provide insight into how TDP-43 dysfunction alters the transcriptome of neurons and other cell types. To determine the biological significance of TDP-43 APA, we studied its effects on the kinase MARK3, a highly significant hit from APA analysis of ALS/FTD frontal cortex neuron nuclei containing or depleted of TDP-43. We confirmed that upon TDP-43 knock-down, the *MARK3* gene displayed increased RNA stability resulting in elevated MARK3 protein expression in iPSC-derived motor neurons, and we documented that TDP-43 depletion promoted increased phosphorylation of tau at serine 262, a previously established read-out for tau dysfunction in Alzheimer's disease [23]. Our results indicate that TDP-43 dysregulation of polyadenylation site selection is another facet of the RNA processing dysfunction which is a central feature of the disease process in TDP-43 proteinopathies.

Results

TDP-43 depletion induces widespread alternative polyadenylation in neuronal cells

To identify alternative polyadenylation (APA) events regulated by TDP-43 in neuronal cells, we employed the dynamic analysis of polyadenylation from RNA-sequencing (DaPars) tool to previously published RNA-sequencing datasets (**Fig. 1a**). While a number of tools have been developed to quantify APA from bulk RNA-seq data [22], we selected DaPars [24], which identifies and quantifies *de novo* polyA site usage by applying a linear regression model to localized changes in RNA-seq read density within the 3'UTR. Given the established role of TDP-43 in regulating the splicing of unannotated 'cryptic' exons, we hypothesized that 'cryptic,' unannotated polyA sites may be similarly utilized upon TDP-43 depletion or mutation. DaPars calculates the percentage of distal polyA site usage index (PDUI) for each transcript, with the results averaged across replicates to determine ΔPDUI in cells where TDP-43 was knocked down or mutated in comparison to control conditions. A positive ΔPDUI denotes increased relative expression of transcripts with a longer 3'UTR, while a negative ΔPDUI signifies increased relative expression of transcripts with a shorter 3'UTR.

Using published RNA-seq data from SH-SY5Y cells, SK-N-BE(2) cells, and i3Neurons in which TDP-43 was depleted by shRNA knock-down [11], we identified hundreds of genes in which TDP-43 knock-down resulted in significant APA (FDR adj $p < 0.05$, $|\Delta PDUI| \geq 0.1$) (**Fig. 1b-g, Supplementary Tables 1-3**). APA genes exhibit both 3'UTR lengthening and shortening upon TDP-43 knock-down; however, we observed increased use of distal polyA sites in 69.9% (460/658) of APA genes in SH-SY5Y cells, 59.6% (429/720) of APA genes in SK-N-BE(2) cells, and 78.1% (250/320) of APA genes in i3Neurons. Nearly 20% of APA events were shared between the closely related SH-SY5Y and SK-N-BE(2) neuroblastoma cell lines (**Fig. 1h**), including 4/10 of the top APA genes by p value (**Fig. 1c, e**). Additionally, 16.3% (52/320) of APA genes in i3Neurons overlapped with at least one of the immortalized neuronal cell lines (**Fig. 1h**), indicating that, while APA is known to be highly cell-type specific [16], many APA events regulated by TDP-43 are conserved across multiple neuronal cell types.

We also applied DaPars to previously published transcriptome data sets in which TDP-43 was depleted by siRNAs in SH-SY5Y cells [10] or in human embryonic stem cell derived motor neurons (hESC-MNs) for 96 hrs [8] (**Supplementary Fig. 1, Supplementary Tables 4, 5**). As expected, we observed fewer APA changes in cells in which TDP-43 was knocked down for a shorter duration or incompletely (~60% in hESC-MNs). Both datasets do, however, exhibit significantly decreased PDUI in *SMC1A*, as previously shown for HEK293 cells upon TDP-43 siRNA knock-down [13].

Given that APA can affect cellular function by regulating mRNA stability and subcellular localization [21], we performed gene ontology (GO) analysis of APA genes in each neuronal cell type (**Supplementary Fig. 2**). Shared terms between datasets for biological pathways known to be highly relevant to ALS/FTD included: 'establishment of protein localization' (SH-SY5Y and i[3]Neurons) and 'cytoskeleton organization' (SH-SY5Y and SK-N-BE(2)), and shared terms for GO molecular functions were: 'kinase activity', 'protein kinase binding', and 'cytoskeletal protein binding' (SH-SY5Y and SK-N-BE(2)). These results indicate that TDP-43 regulation of APA impacts disease-relevant pathways and should be considered, along with differential expression and alternative splicing, as a key aspect of TDP-43 dysfunction.

APA genes directly bound by TDP-43 in the 3'UTR exhibit preferential 3'UTR-lengthening upon TDP-43 depletion

By cross-referencing significant APA genes with TDP-43 eCLIP-seq data from SH-SY5Y cells [25], we found that TDP-43 directly binds either within the 3'UTR or in another region of the transcript to 43.2% (284/658) of APA genes in SH-SY5Y cells, 47.4% (341/720) of APA genes in SK-N-BE(2) cells, and 51.3% (164/320) of APA genes in i[3]Neurons, (**Fig. 2a**, **Supplementary Fig. 3**). In neuronal cells chronically depleted of TDP-43, it is likely that cytotoxicity contributes to gene regulatory changes that do not necessarily reflect direct regulation by TDP-43. Indeed, we noted that in i[3]Neurons, only 36.7% of differentially expressed genes are targets of TDP-43 binding by eCLIP-seq, while a significantly higher proportion of cryptic exon (60.3%) and APA (51.3%) events are observed in genes directly bound by TDP-43 (**Fig. 2b**). In support of our overall finding that TDP-43 depletion preferentially increases use of a distal PAS, we found that APA genes exhibit increased PDUI upon TDP-43 loss – when TDP-43 binds within the 3'UTR, with i[3]Neurons displaying the most significant trend (**Fig. 2c-e;** see also accompanying paper from Gitler and colleagues [26]).

Experimental and functional validation of TDP-43 APA genes in neuronal cells

To experimentally validate TDP-43 APA events, we first utilized the above SH-SY5Y cell model, in which a doxycycline-inducible shRNA against TDP-43 is stably integrated [11]. We selected two APA genes to validate based on their robust and consistent shift in PDUI upon TDP-43 knockdown, and because TDP-43 binds directly within each 3'UTR (**Supplementary Fig. 4**). The Canopy FGF signaling regulator 3 (*CNPY3*) gene was among the 3 greatest PDUI increases in each of our tested neuronal cell lines (**Fig. 2c-e**). *CNPY3* exhibits extensive alternative splicing, with numerous annotated transcript variants. In examining the RNA-seq tracks for *CNPY3*, we found that TDP-43 knock-down significantly increases PAS usage within a specific transcript variant

6

(NM_001318848.2, 'variant two'), defined by an alternative last exon without an upstream splice junction. In line with this data, we confirmed by RT-PCR that TDP-43 depletion results in a significant increase in the use of the variant two 3'UTR relative to variant one (NM_006586.5) (**Fig. 3a**). *CNPY3* APA was similarly described in the accompanying manuscript by Bryce-Smith et al. [27], and was independently detected by Zeng et al. in i3Neurons depleted of TDP-43 in their accompanying study [26]. Increased production of an alternative *CNPY3* isoform upon TDP-43 depletion similarly parallels findings from Zeng et al. in which TDP-43 knock-down resulted in increased expression of an alternative *SFPQ* isoform [26].

In contrast with *CNPY3*, DaPars calculated a significant decrease in distal PAS usage for *SMC1A* (**Supplementary Fig. 4b**), which we experimentally validated by RT-PCR in SH-SY5Y cells (**Fig. 3b**). As noted, APA of *SMC1A* was previously characterized in HEK293 cells upon TDP-43 knockdown, indicating that this is a highly sensitive APA event across diverse cell types. Given that SMC1A is a subunit of the cohesin complex and plays a critical role in chromatin organization, we sought to investigate the biological significance of SMC1A in neuronal cells. Using DIV38 induced pluripotent stem cell-derived motor neurons (iPSC-MNs) in which TDP-43 was knocked down for 10 days (**Supplementary Fig. 5**), we confirmed that TDP-43 loss indeed reduces use of the distal *SMC1A* PAS in human motor neurons (**Fig. 3c**). We further found that APA of *SMC1A* corresponds with a significant increase in SMC1A protein levels in iPSC-MNs (**Fig. 3d**), suggesting that altered regulation of *SMC1A* APA may contribute to impaired chromatin organization upon TDP-43 loss, as previously reported in postmortem neuronal nuclei [28].

Mutant TDP-43 induces APA in genes that function in the oxidative stress response

To explore APA dysregulation that reflects TDP-43 gain-of-function in addition to loss-of-function pathology, we next applied DaPars to RNA-seq data generated from SH-SY5Y cells with homozygous mutation of TDP-43^{N352S} achieved via CRISPR-Cas9 genome editing [10]. Consistent with prior studies, 72% (59/82) of significant APA events corresponded with increased use of a distal PAS (**Fig. 4a, Supplementary Table 6**). Notably, GO analysis revealed significant enrichment of genes that function in the 'response to oxidative stress' pathway (**Fig. 4b**). Numerous studies have implicated oxidative stress in ALS/FTD pathogenesis, including recent evidence that TDP-43 aggregation induces the generation of reactive oxygen species (ROS) [29]. To determine if TDP-43^{N352S} SH-SY5Y cells display an altered oxidative stress response, we measured ROS at 30 minutes and 120 minutes after treating wildtype or TDP-43^{N352S} SH-SY5Y cells with hydrogen peroxide. While there was no initial difference in ROS levels between control and TDP-43^{N352} cells, ROS levels were significantly higher in TDP-43^{N352S} SH-SY5Y cells compared to wildtype cells after 120 minutes of exposure (**Fig. 4c**). This provides proof-of-concept that TDP-

43 APA genes function in cellular pathways implicated in neurodegenerative disease, and that characterization of APA events can highlight disease-relevant phenotypes in cell models of ALS.

Nuclear clearance of TDP-43 induces APA in ALS/FTD patient neurons

To further determine the significance of TDP-43 APA dysregulation in human patients, we considered APA events in neuron nuclei obtained from seven post-mortem ALS/FTD neocortex samples, where FACS sorting resulted in transcriptome data sets for neurons either containing or depleted of nuclear TDP-43 [28]. We identified 181 APA genes ($|\Delta PDUI| > 0.1$, p<0.05) in neuronal nuclei lacking nuclear TDP-43 (**Fig. 5a, b, Supplementary Table 7**), but unlike in neuronal cell culture models, we observed a preference towards negative $\Delta PDUI$ in postmortem TDP-43 nuclear-depleted neurons, as 72.7% (63/87) of genes exhibited 3'UTR shortening. By correlating the most significant APA events ($|\Delta PDUI| > 0.1$, FDR p<0.05) with eCLIP-seq of TDP-43, we noted that 57.7% of these APA genes are bound by TDP-43 (**Fig. 5c**), suggesting that many APA events are directly regulated by TDP-43. We then performed GO analysis for APA events in coding mRNA with $p < 0.05$ and $[|\Delta PDUI| \geq 0.1]$ and found significant enrichment of genes functioning in the 'histamine response', 'synapse assembly', and 'protein transport' pathways (**Fig. 5d**). These pathways have been previously implicated in ALS disease models, again underscoring potential contributions of APA events to ALS pathobiology [30-32].

The most substantial 3'UTR lengthening in ALS/FTD neuronal nuclei ($\Delta PDUI = +0.363$), was observed in the gene encoding microtubule affinity regulating kinase 3 (*MARK3*) (**Fig. 5b**). Visualizing RNA-seq tracks from this experiment with overlayed eCLIP-seq generated by the ENCODE project [33], we confirmed that TDP-43 binds *MARK3* in its 3'UTR immediately upstream of a canonical ATTAAA polyA signal (hg38: chr14:103503803-103503809), as well as in an upstream intron (**Fig. 5e**). This suggests that nuclear TDP-43 normally represses use of a distal *MARK3* PAS, which then becomes preferentially utilized upon nuclear depletion of TDP-43. Importantly, significant 3'UTR lengthening in *MARK3* was also observed in our APA analysis in SH-SY5Y and SK-N-BE(2) cells (**Fig. 2c, d**), and was independently confirmed in an accompanying paper from Gitler and colleagues [26], indicating that increased utilization of a distal polyA site in *MARK*3 is a prominent effect in neuronal cells upon loss of TDP-43. Given the consistency of this result, we evaluated this phenotype in iPSC-MNs, and confirmed that TDP-43 knock-down can induce markedly increased use of a distal PAS in the *MARK3* gene (**Fig. 5e, f**).

MARK3 APA yields increased mRNA stability and elevated protein levels

As MARK3 is a tau kinase associated with tau S262 phosphorylation in the early stages of Alzheimer's disease pathogenesis [23], and because mutations in closely-related MARK4 can

8

significantly increase AD risk, promote hyperphosphorylation of tau, and induce neuron toxicity [34], we sought to investigate the functional consequences of MARK3 APA in neurons. RT-PCR analysis revealed that TDP-43 knock-down results in increased use of the distal *MARK3* PAS in SH-SY5Y cells (**Fig. 6a**). Although steady-state expression of *MARK3* did not change upon TDP-43 knock-down (**Fig. 6b**), we observed a significantly longer half-life for *MARK3* RNA upon TDP-43 depletion in combination with actinomycin D treatment (**Fig. 6c**), indicating that use of a distal *MARK3* PAS affects transcript turnover. Importantly, increased RNA stability corresponded with elevated MARK3 protein levels in SH-SY5Y cells (**Fig. 6d**). To characterize the cis-regulatory elements within the *MARK3* 3'UTR that may affect its metabolism, we evaluated predicted miRNA binding sites using miRBD [35] and TargetScan [36], and we obtained predicted RBP binding motifs using RBPmap [37] (**Fig. 6e**). Both miRNA prediction tools highlighted a conserved binding motif for miR-142-3p within the *MARK3* 3'UTR (**Supplementary Table 8**). Intriguingly, this miRNA was found to be upregulated in ALS mouse models and in sporadic ALS patients; indeed, serum levels of miR-142-3p have been negatively correlated with ALS clinical outcomes [38]. We also identified two conserved RBP binding motifs within the *MARK3* 3'UTR, which can only be utilized when polyadenylation occurs at a distal MARK3 PAS (**Fig. 6e, Supplementary Fig. 6, Supplementary Table 9**). The first motif is recognized by TDP-43, as well as by other RBPs (e.g. RBM24 and RBM38), while a second T/G-rich motif is predicted to be bound by several other RBPs, such as TIA1 and HuR (**Fig. 6e**). The presence of highly conserved miRNA and RBP binding motifs in the distal region of the *MARK3* 3'UTR suggests that differential recognition of short-versus long-3'UTR isoforms by cis-regulatory elements may account for observed changes in *MARK3* mRNA stability upon TDP-43 knock-down.

Increased MARK3 expression promotes tau S262 hyperphosphorylation in iPSC-MNs

To determine if MARK3 protein levels are sufficient to drive tau S262 hyperphosphorylation in human motor neurons, we transduced iPSC-MNs with lentivirus encoding a shRNA vector against *MARK3*, lentivirus encoding the *MARK3* gene, or control empty vector lentivirus controls (**Supplementary Fig. 7**). In iPSC-MNs subjected to MARK3 over-expression, we detected a ~2.4-fold increase in tau S262 phosphorylation (**Fig. 7a**). Moreover, we observed ~50% reduction in tau S262 phosphorylation in iPSC-MNs subjected to MAP4K3 shRNA knock-down (**Fig. 7a**). When we performed TDP-43 shRNA knock-down in iPSC-MNs, we observed a trend towards increased MARK3 protein expression levels (**Fig. 7b**). Although knock-down of TDP-43 only elicited a trend toward increased MARK3 protein expression, we nonetheless measured a significant increase in tau S262 phosphorylation in iPSC-MNs subjected to TDP-43 shRNA treatment (**Fig. 7c**). These results reveal a potentially novel mechanistic link between TDP-43 and tau biology, and suggest

that TDP-43 dysregulation of neuronal *MARK3* APA may contribute to altered cytoskeletal function in ALS/FTD and related neurodegenerative disorders.

Discussion

The vast majority (>90%) of late-onset neurodegenerative diseases are sporadic, caused by ill-defined interactions between genetic and environmental risk factors. The complex etiology of these disorders has consequently hindered development of broadly effective therapies, and recent advances in the use of biological agents, which hold promise for treating rare familial forms of disease, may only benefit a small fraction of patients. This underscores the need to study convergent neurodegenerative disease mechanisms, such as the aberrant nuclear clearance and cytoplasmic aggregation of TDP-43, which has been implicated in a growing number of neurodegenerative disorders. Indeed, in addition to being a defining histopathological hallmark of ALS/FTD, emerging evidence suggests that TDP-43 dysfunction may play a pivotal role in dementia. Comorbid TDP-43 pathology is correlated with more severe cognitive impairment, more rapid disease progression, and increased brain atrophy in AD [2,3]. Furthermore, TDP-43 has been shown to colocalize with senile plaques and neurofibrillary tangles, and TDP-43 histopathology is strongly associated with hippocampal sclerosis [3]. Hence, delineating the cellular consequences of TDP-43 loss of function in neurons and other CNS cell types has the potential to be broadly relevant across multiple neurodegenerative diseases.

While the specific mechanisms linking TDP-43 dysfunction to neurodegeneration remain unclear, independent lines of investigation point towards aberrant RNA metabolism as the key driver of disease. Indeed, a number of independent studies have identified disease-relevant differentially expressed transcripts (e.g. *STMN2*) and alternatively spliced transcripts (e.g., *UNC13A*) that are being regulated by TDP-43 [8,10-12]. TDP-43 is also known to regulate APA, and previous work in HEK293 cells defined a position-dependent principle by which binding of TDP-43 within 75 nucleotides of a PAS represses its usage, while binding further downstream of a PAS site enhances its usage [13]. However, prior to this investigation and the two studies accompanying this report [26,27], the role of TDP-43 dysregulation of APA in neurodegenerative disease had not been undertaken. Here we found that TDP-43 loss of nuclear function changes polyadenylation site selection in hundreds of transcripts in neuron cell culture models of TDP-43 depletion or mutation and in postmortem ALS/FTD patient neuron nuclei containing or lacking TDP-43. While the majority of these APA events were cell-type specific, a number of highly reproducible changes in PAS usage were identified across multiple datasets, including significantly increased distal PAS usage in *CNPY3*, significantly reduced distal PAS usage in *SMC1a*, and significantly increased distal PAS usage in *MARK3*. Two reports accompanying this study also evaluated the effect of TDP-43 depletion in postmortem ALS/FTD patient neuron nuclei using an identical transcriptome data set [28], and despite employing distinct APA detection algorithms and bioinformatics analysis pipelines, both investigations identified hundreds of TDP-43 target genes subject to APA [26,27],

11

underscoring the relevance of TDP-43 dysregulation of polyadenylation site selection to the disease process in TDP-43 proteinopathy. To assess the functional significance of TDP-43 APA in ALS/FTD disease pathogenesis, we also performed GO analysis on the APA hits and identified various pathways, including response to oxidative stress, which we validated as impaired in TDP-43^{N352S} SH-SY5Y cells by documenting elevated ROS upon treatment with hydrogen peroxide.

To examine the functional implications of TDP-43 APA, we pursued directed studies on the tau kinase MARK3, our top hit from DaPars analysis of ALS/FTD patient post-mortem frontal cortex neuron nuclei containing or depleted of TDP-43. After confirming that TDP-43 knock-down in SH-SY5Y cells resulted in increased expression of a *MARK3* transcript isoform with a lengthened 3'UTR, we measured *MARK3* RNA expression levels, but did not detect any change in *MARK3* RNA levels upon TDP-43 knock-down. However, when we performed TDP-43 knock-down in combination with actinomycin D treatment to block transcription, we documented a significant increase in *MARK3* RNA expression levels, indicative of increased RNA transcript stability. Immunoblot analysis of MARK3 protein in SH-SY5Y cells subjected to TDP-43 knock-down revealed markedly increased expression of MARK3 protein, confirming that use of a distal PAS upon TDP-43 depletion results in increased MARK3 protein. Inclusions of tau protein accumulate in the brains of patients affected with familial and sporadic FTD [39], and tau is centrally involved in AD where as many as 57% of affected patients display TDP-43 nuclear clearance and cytoplasmic aggregation in addition to the hallmark neurofibrillary tangles of tau protein [2]. To determine if TDP-43 APA of MARK3 could contribute to disease pathogenesis in ALS/FTD and AD, we evaluated the effect of TDP-43 knock-down on MARK3 phosphorylation of tau at S262, a well-established MARK3 phosphorylation site with potential pathophysiological implications, because tau phosphorylation at this residue, which is located within a tau microtubule-binding site [40], can reduce the microtubule-stabilizing properties of tau and antagonize microtubule tracking at the plus end by disrupting interaction between end-binding protein 1 and tubulin [41] [42]. We found that TDP-43 knock-down in iPSC-derived human motor neurons resulted in significantly increased tau S262 phosphorylation, indicating that TDP-43 APA of *MARK3* could be contributing to neurodegeneration. Similarly, in an accompanying report from Zeng et al., TDP-43 APA of the FTLD-TDP risk gene *TMEM106B* was detected and then documented in post-mortem frontal cortex samples from FTLD-TDP patients, and proposed to alter *TMEM106B* translational efficiency, potentially destabilizing TMEM106B dimer formation [26].

Nearly two decades ago, the discovery of TDP-43 aggregates in the brains of patients with ALS and FTD revolutionized the neurodegenerative disease field [43], and set the stage for intense research into the role of TDP-43 dysfunction in the pathogenesis of ALS/FTD and related

disorders. Over the last decade, considerable evidence has accumulated indicating that TDP-43 dysregulation of RNA splicing is likely a key step in the disease process for a broad range of disorders, now known as TDP-43 proteinopathies. However, in addition to regulating splicing, TDP-43 performs another fundamentally important function in the processing of RNA transcripts – selection of the site of polyadenylation, which dictates the length of a transcript's 3'UTR. It is important to emphasize that 3'UTR length is a critical factor in regulating gene function. Within the 3'UTR are binding sites for miRNAs and RBPs, and these interactions dictate the stability of a transcript and hence modulate its expression at the protein level. Here we report that TDP-43 APA of MARK3 can alter its protein expression with consequences for tau function, underscoring the potential importance of TDP-43 dysregulation of polyadenylation site selection in disorders featuring tau pathology and TDP-43 nuclear clearance and cytoplasmic aggregation. Furthermore, cis-acting regulatory elements controlling mRNA subcellular localization exist within 3' UTRs [44], and their presence or absence may determine if a mRNA localizes to the cell soma, the axon, or dendrite processes of a neuron. Subcellular localization of RNAs within neurons and other CNS cell types could profoundly affect cellular function; hence, altered subcellular mRNA transcript localization due to APA could be contributing to neurodegeneration, as reported for *ELK1* in an independent study from Bryce-Smith et al. accompanying this study [27]. Hence, our results, along with independent findings accompanying this report from two studies using different methods of APA analysis [26,27], reveal that TDP-43 dysregulation of polyadenylation site selection could be driving the pathogenesis of various neurodegenerative diseases, likely in combination with altered splicing and cryptic exon inclusion. Defining genes subject to TDP-43 APA should thus be a major focus of future studies, as a number of these genes could be targets for developing biomarkers and engineering novel therapies. As antisense oligonucleotides (ASOs) can regulate polyadenylation site usage by steric hindrance [45,46], delineation of genes subject to APA may yield candidates for a next generation of ASOs and other biological agents for treatment of TDP-43 proteinopathy.

Materials & Methods

DaPars software, described previously [24], allows for the joint analyses of multiple samples based on a two-normal mixture model, which is used to calculate the PDUI value. Briefly, we extract a 3′UTR annotation for each gene using the "DaPars_Extract_Anno.py" script within DaPars2. We then used the "samtools flagstat" command [47] to calculate the sequencing depth for each sample. Finally, we use DaPars2 to calculate the percentage of the distal poly (A) site usage index (PDUI) divided by the total expression level of each transcript across samples.

Cell culture

Human neuroblastoma SH-SY5Y cells (ATCC) were maintained in 50% EMEM (ATCC, 30-2003), 50% Ham's F12 (ThermoFisher, 11765047), supplemented with 10% fetal bovine serum, and 50 U/mL penicillin–streptomycin. Cells were grown at 37 °C and 5% CO_2.

Doxycycline-inducible knockdown of TDP-43

SH-SY5Y cells with stable integration of a doxycycline-inducible shRNA cassette targeting TDP-43 were previously generated [11]. Cells were treated with 1 µg/mL doxycycline (Sigma-Aldrich, D9891-1G) for 7 days to induce TDP-43 knockdown.

iPSC-derived motor neurons (iPSC-MNs)

iPSCs from the KOLF2.1J line were maintained in mTeSR (STEMCELL Technologies, 100-0276). According to manufacturer's instructions, iPSCs were differentiated into motor neurons from day 0-14 (STEMCELL Technologies, 100-0871) and subsequently matured from day 14- (STEMCELL Technologies, 100-0872). iPSC-MNs were transduced using lentiviral vectors at day 28 and collected at day 38. Cells were cultured at 37°C and 5% CO_2.

Lentiviral vector transduction

Lentiviral constructs encoding GFP (VectorBuilder, LVM[VB010000-9298rtf]-C), MARK3-V5 (Addgene, 107235), shRNA control (Sigma, SHC002), tardbp shRNA (Sigma, TRCN0000016038), or MARK3 shRNA (Sigma, TRCN0000001564) were packaged into lentivirus by VectorBuilder. iPSC-MNs were transduced with 10 MOI lentivirus at DIV 28 for 10 days.

shRNA target sequences:
Control: CAACAAGATGAAGAGCACCAA
Tardbp: GCTCTAATTCTGGTGCAGCAA
MARK3: TGTGTGTGAAGTGGTGTATAT

RT-PCR analysis

RNA was isolated from cell lysates (Qiagen, 74106) and reverse transcribed to generate cDNA (ThermoFisher, 11756500). PCR was performed using 10-30 ng cDNA template, 0.5 µM forward and reverse primers, and One*Taq* DNA polymerase mastermix (NEB, M0488L). PCR products were separated by agarose gel electrophoresis using 1.5% agarose (VWR, 0710) in 1X TBE buffer at 150V.

RT-PCR primer sequences:

MARK3 fwd: GGAAATGGAAGTGTGCAAGC

MARK3 prox rev: TAGAAGATGCAGACGTTATTGCC

MARK3 dist rev: CACACACAAGCAATGTTCACAAC

CNPY3 var1 3'UTR fwd: CCAGCATCCTCTGTCCTGA

CNPY3 var1 3'UTR rev: GAAGAGGGCACAGCCAAG

CNPY3 var2 3'UTR fwd: GACCACCTGGGATCTTCCT

CNPY3 var2 3'UTR rev: CACATCGTGGATCTTGCTGAG

SMC1A prox 3'UTR fwd: CCTGTCTGGATCCCTAAGCTG

SMC1A prox 3'UTR rev: CTCCAGACCTAACATCACCTCTG

SMC1A dist 3'UTR fwd: GTTAGTCAGTAGCAGTAGGAGGAG

SMC1A dist 3'UTR rev: GCATTCACAGGGAAATAAGGAAGAC

Quantitative reverse transcription PCR

Relative fold change of MARK3 was determined by qRT-PCR using SYBR Green Master Mix (ThermoFisher, A25776) with GAPDH as an endogenous control.

RT-qPCR primer sequences:

MARK3 fwd: CGCTCTCTGCTCCTCCTGTT

MARK3 rev: TGCTGGTCTGACTCCTTTTCGG

GAPDH fwd: CAGTCTCCTCACCACAAAGTG

GAPDH rev: CCATGGTGTCTGAGCGATGT

Immunoblot analysis

Protein extracts were prepared in radioimmunoprecipitation assay lysis buffer (ThermoFisher, 89900) with protease and phosphatase inhibitors (ThermoFisher, 87786). Lysates were sonicated at 4 °C (Bioruptor Pico), denatured in LDS sample buffer (ThermoFisher, NP0007) with a reducing agent (ThermoFisher, NP0009), and heated at 70 °C for 10 min. Lysates were electrophoresed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis using a 4%–20% gradient polyacrylamide gel (Bio-Rad, 5678093) and transferred to 0.45-μm polyvinylidene difluoride membrane (Bio-Rad, 1704157). Blots were incubated with primary antibodies diluted in 5% nonfat milk in TBST (TRIS-buffered saline, 0.05% Tween-20) overnight at 4 °C and with secondary antibodies diluted in 5% nonfat milk in TBST (ThermoFisher, A16078 [goat anti-mouse HRP, 1:3000], A16110 [goat anti-rabbit HRP, 1:3000], A32728 [goat anti-mouse Alexa FluorTM 647, 1:1000], A32733 [goat anti-rabbit Alexa FluorTM 647, 1:1000]) for 60 min at RT. Detection was performed using enhanced chemiluminescence substrate (Genesee Scientific, 20–300S) and imaged on a ChemiDoc MP (Bio-Rad). Primary antibodies were as follows: TDP-43 (Proteintech, 12892-1-AP, 1:3000), MARK3 (Cell Signaling Technology, 9311S, 1:2000), SMC1A (abcam, ab9262, 1:2500), phospho-tau Ser262 (ThermoFisher, 44-750-G, 1:2000), TAU-5 (ThermoFisher, AHB0042, 1:2000), and α-tubulin (ThermoFisher, 62204, 1:3000 or Novus Biologicals, NBP2-80570, 1:3000).

Actinomycin D mRNA stability experiment

SH-SY5Y cells with doxycycline-inducible TDP-43 shRNA expression were treated with 5 μg/mL actinomycin D (Sigma-Aldrich, A1410) for 0, 2, 4, 8, or 12 hours. Cells were subjected to a PBS wash prior to RNA isolation (Qiagen, 74106).

Detection of reactive oxygen species (ROS)

Wildtype and TDP-43^{N352S} SH-SY5Y cells were plated on a 96-well plate and treated for 120 minutes with 100 μM H_2O_2. A fluorometric intracellular ROS detection kit was used to detect ROS levels at 30 and 120 minutes (Sigma, MAK144) using a Varioskan microplate reader (ThermoFisher).

Gene Ontology (GO) analysis

GO analysis was performed using ShinyGO v0.77 [48] for GO Biological Processes and GO Molecular Function pathway databases. GO analysis was performed for APA events in protein-coding genes with $\Delta PDUI \geq 0.1$ and $p < 0.05$. Graphical representations of the top 10 pathways with an FDR adj. $p < 0.05$ cut-off were generated in R version 4.3.1, utilizing the ggplot and sjPlot packages.

Statistical analysis

As indicated, for pairwise comparisons, statistical significance was determined by two-tailed Student's t-test. For multiple comparisons, statistical significance was determined by one-way analysis of variance (ANOVA) with *post hoc* Tukey test. The significance level (α) was set at 0.05 for all experiments. Data were compiled and analyzed using Microsoft Excel or Prism 10 (GraphPad).

References

1.   de Boer, E.M.J.*, et al.* TDP-43 proteinopathies: a new wave of neurodegenerative diseases. *J Neurol Neurosurg Psychiatry* **92**, 86-95 (2020).

2.   Josephs, K.A.*, et al.* TDP-43 is a key player in the clinical features associated with Alzheimer's disease. *Acta Neuropathol* **127**, 811-824 (2014).

3.   Meneses, A.*, et al.* TDP-43 Pathology in Alzheimer's Disease. *Mol Neurodegener* **16**, 84 (2021).

4.   Sanchez, I.I.*, et al.* Huntington's disease mice and human brain tissue exhibit increased G3BP1 granules and TDP43 mislocalization. *J Clin Invest* **131**(2021).

5.   Anderson, E.N.*, et al.* Traumatic injury compromises nucleocytoplasmic transport and leads to TDP-43 pathology. *Elife* **10**(2021).

6.   Nelson, P.T.*, et al.* Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. *Brain* **142**, 1503-1527 (2019).

7.   Polymenidou, M.*, et al.* Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat Neurosci* **14**, 459-468 (2011).

8.   Klim, J.R.*, et al.* ALS-implicated protein TDP-43 sustains levels of STMN2, a mediator of motor neuron growth and repair. *Nat Neurosci* **22**, 167-179 (2019).

9.   Baughn, M.W.*, et al.* Mechanism of STMN2 cryptic splice-polyadenylation and its correction for TDP-43 proteinopathies. *Science* **379**, 1140-1149 (2023).

10.   Melamed, Z.*, et al.* Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration. *Nat Neurosci* **22**, 180-190 (2019).

11.   Brown, A.L.*, et al.* TDP-43 loss and ALS-risk SNPs drive mis-splicing and depletion of UNC13A. *Nature* **603**, 131-137 (2022).

12.   Ma, X.R.*, et al.* TDP-43 represses cryptic exon inclusion in the FTD-ALS gene UNC13A. *Nature* **603**, 124-130 (2022).

13.   Rot, G.*, et al.* High-Resolution RNA Maps Suggest Common Principles of Splicing and Polyadenylation Regulation by TDP-43. *Cell Rep* **19**, 1056-1067 (2017).

14.   Hoque, M.*, et al.* Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat Methods* **10**, 133-139 (2013).

15.   Derti, A.*, et al.* A quantitative atlas of polyadenylation in five mammals. *Genome Res* **22**, 1173-1183 (2012).

16.   Li, L.*, et al.* An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nat Genet* **53**, 994-1005 (2021).

17.   Zhang, H., Lee, J.Y. & Tian, B. Biased alternative polyadenylation in human tissues. *Genome Biol* **6**, R100 (2005).

18. Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J.O. & Lai, E.C. Widespread and extensive lengthening of 3' UTRs in the mammalian brain. *Genome Res* **23**, 812-825 (2013).

19. Gu, S., Jin, L., Zhang, F., Sarnow, P. & Kay, M.A. Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nat Struct Mol Biol* **16**, 144-150 (2009).

20. Ramakrishnan, A. & Janga, S.C. Human protein-RNA interaction network is highly stable across mammals. *BMC Genomics* **20**, 1004 (2019).

21. Mayr, C. Regulation by 3'-Untranslated Regions. *Annu Rev Genet* **51**, 171-194 (2017).

22. Bryce-Smith, S.*, et al.* Extensible benchmarking of methods that identify and quantify polyadenylation sites from RNA-seq data. *RNA* **29**, 1839-1855 (2023).

23. Lund, H.*, et al.* MARK4 and MARK3 associate with early tau phosphorylation in Alzheimer's disease granulovacuolar degeneration bodies. *Acta Neuropathol Commun* **2**, 22 (2014).

24. Xia, Z.*, et al.* Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat Commun* **5**, 5274 (2014).

25. Tam, O.H.*, et al.* Postmortem Cortex Samples Identify Distinct Molecular Subtypes of ALS: Retrotransposon Activation, Oxidative Stress, and Activated Glia. *Cell Rep* **29**, 1164-1177.e1165 (2019).

26. Zeng, Y.*, et al.* TDP-43 nuclear loss in FTD-ALS causes widespread alternative polyadenylation. [submitted] (2024).

27. Bryce-Smith, S.*, et al.* TDP-43 loss induces cryptic polyadenylation and increased protein synthesis in ALS. [submitted] (2024).

28. Liu, E.Y.*, et al.* Loss of Nuclear TDP-43 Is Associated with Decondensation of LINE Retrotransposons. *Cell Rep* **27**, 1409-1421.e1406 (2019).

29. Bharathi, V., Girdhar, A. & Patel, B.K. Role of CNC1 gene in TDP-43 aggregation-induced oxidative stress-mediated cell death in S. cerevisiae model of ALS. *Biochim Biophys Acta Mol Cell Res* **1868**, 118993 (2021).

30. Burk, K. & Pasterkamp, R.J. Disrupted neuronal trafficking in amyotrophic lateral sclerosis. *Acta Neuropathol* **137**, 859-877 (2019).

31. Apolloni, S.*, et al.* Histaminergic transmission slows progression of amyotrophic lateral sclerosis. *J Cachexia Sarcopenia Muscle* **10**, 872-893 (2019).

32. Gelon, P.A., Dutchak, P.A. & Sephton, C.F. Synaptic dysfunction in ALS and FTD: anatomical and molecular changes provide insights into mechanisms of disease. *Front Mol Neurosci* **15**, 1000183 (2022).

33. Van Nostrand, E.L.*, et al.* A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711-719 (2020).

34. Rovelet-Lecrux, A.*, et al.* De novo deleterious genetic variations target a biological network centered on Aβ peptide in early-onset Alzheimer disease. *Mol Psychiatry* **20**, 1046-1056 (2015).

35. Chen, Y. & Wang, X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res* **48**, D127-D131 (2020).

36. Agarwal, V., Bell, G.W., Nam, J.W. & Bartel, D.P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**(2015).

37. Paz, I., Kosti, I., Ares, M., Cline, M. & Mandel-Gutfreund, Y. RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res* **42**, W361-367 (2014).

38. Matamala, J.M.*, et al.* Genome-wide circulating microRNA expression profiling reveals potential biomarkers for amyotrophic lateral sclerosis. *Neurobiol Aging* **64**, 123-138 (2018).

39. Dickson, D.W., Kouri, N., Murray, M.E. & Josephs, K.A. Neuropathology of frontotemporal lobar degeneration-tau (FTLD-tau). *J Mol Neurosci* **45**, 384-389 (2011).

40. Kellogg, E.H.*, et al.* Near-atomic model of microtubule-tau interactions. *Science* **360**, 1242-1246 (2018).

41. Biernat, J., Gustke, N., Drewes, G., Mandelkow, E.M. & Mandelkow, E. Phosphorylation of Ser262 strongly reduces binding of tau to microtubules: distinction between PHF-like immunoreactivity and microtubule binding. *Neuron* **11**, 153-163 (1993).

42. Ramirez-Rios, S.*, et al.* Tau antagonizes end-binding protein tracking at microtubule ends through a phosphorylation-dependent mechanism. *Mol Biol Cell* **27**, 2924-2934 (2016).

43. Neumann, M.*, et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130-133 (2006).

44. Mendonsa, S.*, et al.* Massively parallel identification of mRNA localization elements in primary cortical neurons. *Nat Neurosci* **26**, 394-405 (2023).

45. Vickers, T.A., Wyatt, J.R., Burckin, T., Bennett, C.F. & Freier, S.M. Fully modified 2' MOE oligonucleotides redirect polyadenylation. *Nucleic Acids Res* **29**, 1293-1299 (2001).

46. Vorlova, S.*, et al.* Induction of antagonistic soluble decoy receptor tyrosine kinases by intronic polyA activation. *Mol Cell* **43**, 927-939 (2011).

47. Li, H.*, et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

48. Ge, S.X., Jung, D. & Yao, R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**, 2628-2629 (2020).

ACKNOWLEDGEMENTS

Author Contributions

F.J.A., Y.C., S.M. and A.R.L.S. provided the conceptual framework for the study. F.J.A., Y.C., S.M. O.H.T., M.G.H., W.L. and A.R.L.S. designed the experiments. F.J.A., Y.C., S.M., M.R.C., C.S., W.Y., S.M., W.G.S., K.C.K.E., and S.H. performed the experiments. F.J.A., Y.C., S.M. O.H.T., M.G.H., W.L. and A.R.L.S. analyzed the data. F.J.A., Y.C., S.M. and A.R.L.S. wrote the manuscript.

Declarations

M.G.H. is on the scientific advisory board of Transposon Therapeutics.

Figure Legends

Figure 1: TDP-43 depletion induces widespread alternative polyadenylation

(**a**) Overview of the study. We applied the dynamic analysis of polyadenylation from RNA-seq (DaPars) tool to published RNA-seq datasets to detect APA events resulting from TDP-43 depletion or mutation in immortalized cell lines, stem cell-derived neurons, and postmortem ALS/FTD neocortex. We then performed experimental and functional validation of select targets. (**b, d, f**) Volcano plots depicting APA events in SH-SY5Y cells (**b**), SK-N-BE(2) cells (**d**), and i3Neurons (**f**) in which TDP-43 was knocked down via shRNA treatment. APA genes with FDR adj. p value < 0.05 and ΔPDUI $\geq$ 0.1 are depicted in red, and APA genes with FDR adj. p value < 0.05 and ΔPDUI $\leq$ 0.1 are depicted in blue. (**c, e, g**) Table of top 10 APA events in SH-SY5Y cells (**c**), SK-N-BE(2) cells (**e**), and i3Neurons (**g**) by FDR adj. p-value with lΔPDUIl $\geq$ 0.1, along with evidence of TDP-43 binding sites in each gene by eCLIP-seq (red). (**h**) Venn diagram illustrating the intersection of APA events across datasets.

Figure 2: APA genes directly bound by TDP-43 in the 3'UTR exhibit preferential 3'UTR-lengthening upon TDP-43 depletion

(**a**) Venn diagram illustrating the proportion of APA genes for which there is published evidence of TDP-43 binding within the 3'UTR. (**b**) Percentage of differentially expressed genes (DEGs), cryptic exon (CE)-containing genes, and APA genes that contain TDP-43 binding sites from RNA-seq of i3Neurons +/- TDP-43 knock-down. (**c, d, e**) Graphical representation of ΔPDUI for all APA genes with TDP-43 binding sites in the 3'UTR.

Figure 3: Experimental and functional validation of TDP-43 APA genes in neuronal cells

(**a**) PCR validation of increased usage of the *CNPY3* variant two 3'UTR upon TDP-43 knock-down in SH-SY5Y cells. (**b**) PCR validation of increased proximal PAS usage of SMC1A upon TDP-43 knock-down in SH-SY5Y cells. (**c**) Validation of SMC1A APA in iPSC-derived motor neurons (iPSC-MNs) in which TDP-43 was knocked down with shRNA for 10 days.
(**d**) Immunoblot analysis of SMC1A in iPSC-MNs reveals that *SMC1A* APA corresponds with an increase in SMC1A protein levels. *p<0.05, **p<0.01, ***p<0.001; two-tailed t-test; n = 3 biological replicates. Data are presented as mean values ± SEM.

Figure 4: Mutant TDP-43 induces APA in genes functioning in the oxidative stress response

(**a**) Volcano plot depicting APA genes in SH-SY5Y cells expressing homozygous TDP-43[N352S] via CRISPR-Cas9 mediated genome editing. APA genes with FDR adj. p value < 0.05 and ΔPDUI $\geq$ 0.1

22

are depicted in red, and APA genes with FDR adj. p value < 0.05 and ΔPDUI $\leq$ 0.1 are depicted in blue. (**b**) GO BP analysis of APA events in TDP-43^{N352S} cells reveals impaired response to hydrogen peroxide (H_2O_2). (**c**) Wildtype and TDP-43^{N352S} cells were treated with 100 μM H_2O_2. ROS was detected at 30 minutes and 120 minutes using a fluorometric intracellular ROS detection kit. *p<0.05. ****p<0.0001; one-way ANOVA; n = 3 biological replicates. Data are presented as mean values ± SEM.

Figure 5: Nuclear clearance of TDP-43 induces APA in ALS/FTD patient neurons

(**a**) Volcano plot depicting APA genes in neuronal nuclei from 7 post-mortem ALS/FTD neocortex samples sorted by FACS for the presence or absence of nuclear TDP-43. APA genes with p < 0.05 and ΔPDUI $\leq$ -0.1 are depicted in blue and APA genes with p < 0.05 and ΔPDUI $\geq$ 0.1 are in red. Please note that *IKBKAP* is also known as *ELP1*, which is how this gene is referred to in the accompanying paper from Zeng et al. [26] (**b**) Top 10 APA events by FDR adj. p-value with a lΔPDUIl $\geq$ 0.1 along with evidence of TDP-43 binding sites by eCLIP-seq (red). (**c**) Percentage of cryptic exon (CE)-containing genes and APA genes that contain TDP-43 binding sites in ALS/FTD neuronal nuclei depleted of TDP-43. (**d**) GO analysis of biological processes (BP) enriched for coding APA events with p < 0.05 and lΔPDUIl $\geq$ 0.1 highlights genes that function in nucleocytoplasmic transport and synapse assembly. (**e**) ENCODE CLIP-seq data showing the location of TDP-43 binding within the MARK3 3'UTR, as well as in an upstream intronic region (green). MARK3 binding within the 3'UTR is immediately upstream of the distal shift in 3'UTR usage observed in TDP-43 negative neurons (blue vs. red RNA-seq tracks). (**f**) RT-PCR of distal 3'UTR (top band) and proximal 3'UTR (bottom band) of MARK3 in iPSC-MNs in which TDP-43 was knocked down with shRNA for 10 days. *p < 0.05; two-tailed t-test; n = 3 biological replicates and n = 3 technical replicates. Data are presented as mean values ± SEM.

Figure 6: MARK3 alternative polyadenylation results in increased mRNA stability and protein levels in SH-SY5Y cells
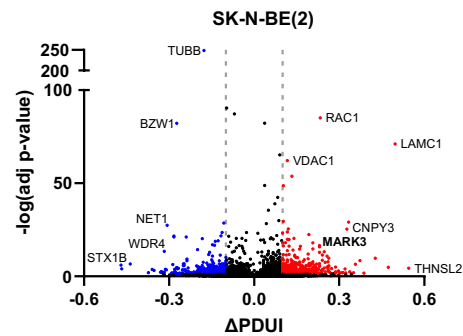
(**a**) PCR validation of increased distal PAS usage of MARK3 upon TDP-43 KD in SH-SY5Y cells.
(**b**) qRT-PCR of MARK3 in SH-SY5Y cells +/- TDP-43 knockdown.
(**c**) RNA stability of MARK3 over time after actinomycin D (5 μg/mL) transcriptional arrest in SH-SY5Y cells upon TDP-43 depletion (red) or untreated (teal).
(**d**) Immunoblot showing increased MARK3 protein levels upon TDP-43 KD in SH-SY5Y cells.
(**e**) Graphical representation of select conserved cis-regulatory elements in the MARK3 3'UTR. *p<0.05, ***p<0.001; two-tailed t-test; n = 3 - 4 biological replicates. Data are presented as mean values ± SEM.

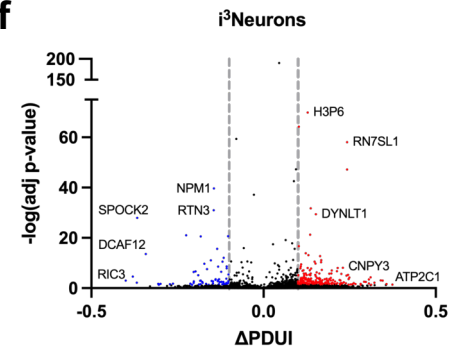**Figure 7: Increased MARK3 expression results in hyperphosphorylation of tau S262 in iPSC-MNs**

(**a**) Immunoblot of S262 phosphorylated tau and total tau in iPSC-MNs transduced with lentivirus encoding control shRNA, shRNA MARK3, GFP (empty vector), or MARK3-V5 for 10 days. *p<0.05, **p<0.01; one-way ANOVA with post-hoc Tukey test; n = 3 biological replicates. Data are presented as mean values ± SEM. (**b**) Immunoblot of MARK3 in DIV 38 iPSC-MNs depleted of TDP-43 with shRNA for 10 days. *p<0.05, ***p<0.001; two-tailed t-test; n = 3 biological replicates. Data are presented as mean values ± SEM. (**c**) Immunoblot of S262 phosphorylated tau and total tau in iPSC-MNs transduced with lentivirus encoding control shRNA or shRNA TARDBP for 10 days. *p<0.05, ***p<0.001; two-tailed t-test; n = 3 biological replicates. Data are presented as mean values ± SEM.

a

Neocortex postmortem ALS/FTD patients
SH-SY5Y: TDP-43^WT, TDP-43^N352S
i3 Neurons
SH-SY5Y & SK-N-BE(2)

nuclear TDP-43 Source: Liu et al. 2019 Cell Reports
Melamed et al. 2019 Nat Neurosci
TDP-43 Source: WT knockdown, WT knockdown — Brown et al. Nature 2022

APA Analysis (DaPars2)
poly(A) usage — Wild Type / TDP-43 KD / mutant — RNA-seq

Calculate the percentage of distal poly(A) site usage index (PDUI)

Gene Ontology (GO)

Experimental and functional validation of APA events

b SH-SY5Y — volcano plot, −log(adj p-value) vs ΔPDUI; labeled genes: TUBB, LAMC1, VDAC1, KCNH2, ETF1, MARK3, CNPY3, USP13, ZNF24, AP2B1

c SH-SY5Y

Gene	PDUI CTL	PDUI TDP-43 KD	ΔPDUI	Adj. p-value	TDP-43 binding
TUBB	0.503	0.346667	-0.15667	1.09 E-248	Exon 3'UTR
LAMC1	0.420	0.986667	0.566667	1.31 E-116	Intron 3'UTR
VDAC1	0.840	0.953333	0.113333	1.98 E-72	Intron
KCNH2	0.677	0.92	0.243333	2.64 E-56	Intron
GATA2	0.767	0.886667	0.12	7.92 E-54	Intron
ETF1	0.433	0.76	0.326667	4.68 E-45	Intron
RAF1	0.870	0.996667	0.126667	1.33 E-38	Intron
CNPY3	0.213	0.67	0.456667	1.55 E-36	Exon Intron 3'UTR
MARK3	0.673	0.903333	0.23	6.29 E-35	Intron 3'UTR
AP2M1	0.773	0.92	0.146667	3.33 E-33	N/A

d SK-N-BE(2) — volcano plot, −log(adj p-value) vs ΔPDUI; labeled genes: TUBB, BZW1, NET1, WDR4, STX1B, RAC1, LAMC1, VDAC1, CNPY3, MARK3, THNSL2

e SK-N-BE(2)

Gene	PDUI CTL	PDUI TDP-43 KD	ΔPDUI	Adj. p-value	TDP-43 binding
TUBB	0.763	0.587	-0.177	1.93 E-249	Exon 3'UTR
RAC1	0.517	0.75	0.233	7.09 E-86	Intron
BZW1	0.973	0.7	-0.273	6.76 E-83	N/A
LAMC1	0.363	0.86	0.497	7.33 E-72	Intron 3'UTR
VDAC1	0.86	0.977	0.117	6.60 E-63	Intron
CHCHD3	0.803	0.937	0.133	1.52 E-54	Intron
VGF	0.82	0.923	0.103	2.02 E-49	3'UTR
NUCKS1	0.35	0.453	0.103	2.43 E-30	Intron 3'UTR
ETF1	0.38	0.713	0.333	7.42 E-30	Intron
NUDT21	1	0.893	-0.107	2.05 E-29	N/A

f i3Neurons — volcano plot, −log(adj p-value) vs ΔPDUI; labeled genes: H3P6, RN7SL1, NPM1, DYNLT1, SPOCK2, RTN3, DCAF12, RIC3, CNPY3, ATP2C1

g i3Neurons

Gene	PDUI CTL	PDUI TDP-43 KD	ΔPDUI	Adj. p-value	TDP-43 binding
ACTB	0.808	0.91	0.103	6.20 E-65	Exon Intron
NPM1	0.898	0.753	-0.144	2.12 E-40	Intron
EIF4A2	0.290	0.427	0.137	1.73 E-32	Exon Intron 3'UTR
RTN3	0.845	0.7	-0.145	9.52 E-32	N/A
DYNLT1	0.765	0.917	0.152	3.78 E-30	3'UTR
SPOCK2	1	0.633	-0.367	1.21 E-28	3'UTR
CCT3	0.815	0.95	0.135	5.30 E-22	Intron
FKBP1B	0.935	0.71	-0.225	8.80 E-22	Intron
CPLX2	0.990	0.887	-0.103	2.43 E-21	Intron 3'UTR
MSMO1	0.973	0.79	-0.183	2.73 E-21	N/A

h Venn diagram

SH-SY5Y: 506 — SK-N-BE(2): 555 — overlap SH-SY5Y ∩ SK-N-BE(2): 130 — center: 5 — SH-SY5Y ∩ i3: 17 — SK-N-BE(2) ∩ i3: 30 — i3 Neurons: 268

Figure 1

Figure 2

a

SK-N-BE(2)
SH-SY5Y
i3 Neurons
TDP-43 CLIP-seq (3'UTR)

503
247
115
27
21
469
3
3
740
14
2
52
3
15
37
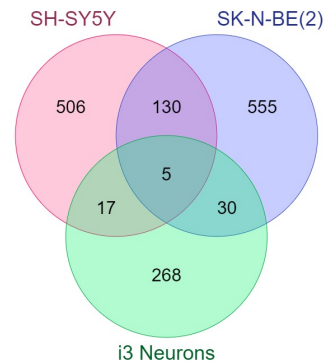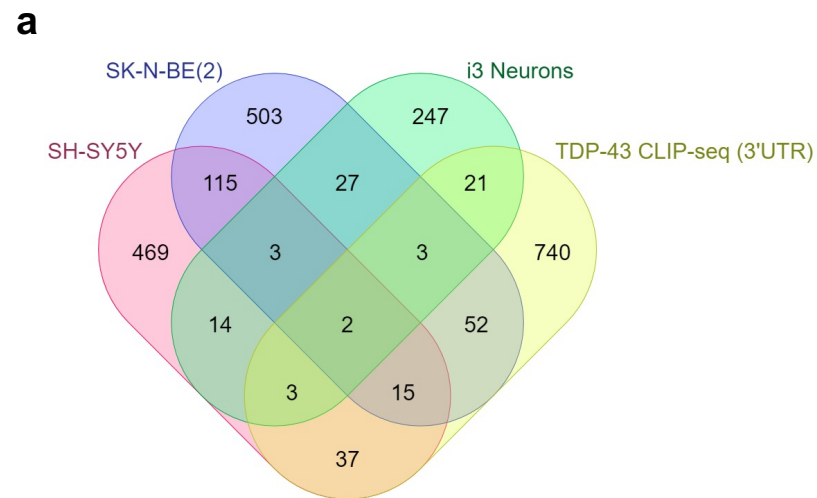
b

i3Neuron RNA-seq analysis	Overlap with TDP-43 eCLIP-seq
Differential expression (DEGs)	1763/4804 **36.7%**
Cryptic exons (CEs)	108/179 **60.3%**
Alternative polyadenylation (APA)	164/320 **51.3%**

c SH-SY5Y

d SK-N-BE(2)

e i³Neurons

ΔPDUI

Figure 3

a

CNPY3

Var2 3'UTR
Var1 3'UTR

200 bp
100 bp

b

SMC1A

Dist 3'UTR
Prox 3'UTR

200 bp
100 bp

c

SMC1A

Dist 3'UTR
Prox 3'UTR

200 bp
100 bp

SMC1A

d

CTL   KD

SMC1A

TDP-43

Total protein

a

(Volcano plot) x-axis: ΔPDUI, y-axis: -log(adjusted p-value). Labeled genes: GNAS, TUBA1A, ACTG1, H2AFZ, LAPTM4B, PSMG2, TBC1D10B

b

SH-SY5Y TDP-43^{N352S} GO Biological Process
Top 10 terms ordered by FDR corrected p-value

GO terms (top to bottom): Response to hydrogen peroxide, Response to reactive oxygen species, Protein-containing complex disassembly, Response to oxidative stress, Cellular component disassembly, Establishment of organelle localization, Actin filament-based proc., Reg. of organelle organization, Intracellular transport, Protein transport

x-axis: Enrichment score. Fold Enrichment: 2, 3, 4, 5. -log10(Enrichment FDR): 2, 3, 4

Cut-off lines drawn at equivalents of p=0.01, p=0.001, p=0.0001

c

30 min — y-axis: A.U. Groups: WT −, N352S −, WT +, N352S +. Significance: ns, ****, ****, ns

120 min — y-axis: A.U. Groups: WT −, N352S −, WT +, N352S +. Significance: ns, ****, ****, *

Figure 4

Figure 5

a

ALS/FTD neocortex

(Volcano plot with x-axis ΔPDUI ranging from -0.6 to 0.6 and y-axis -log (p-value) from 0 to 20. Labeled points: MARK3, HPCAL4, STX3, NFASC, NUP93, EIF4G3, IKBKAP)

b

ALS/FTD neocortex

Gene	PDUI TDP-43+	PDUI TDP-43-	ΔPDUI	Adj. p-value	TDP-43 binding
MARK3	0.544	0.907	0.363	9.41E-16	Intron 3'UTR
HPCAL4	0.978	0.821	-0.157	5.09E-10	N/A
STX3	0.897	0.315	-0.582	3.95E-09	N/A
NFASC	0.981	0.836	-0.146	7.06E-05	Intron
NUP93	0.934	0.734	-0.2	7.06E-05	Intron
EIF4G3	0.524	0.309	-0.216	0.001	3'UTR
UFD1L	1	0.778	-0.222	0.003	N/A
IKBKAP	0.192	0.450	0.257	0.005	N/A
SCCPDH	0.917	0.760	-0.157	0.008	Intron 3'UTR
INTS3	0.857	0.599	-0.259	0.009	Intron

c

ALS/FTD RNA-seq analysis	Overlap with TDP-43 eCLIP-seq
Cryptic exons (CEs)	45/66 **68.2%**
Alternative polyadenylation (APA)	15/26 **57.7%**

d

ALS/FTD neocortex GO Biological Process
Top 10 terms ordered by FDR corrected p-value

(Dot plot showing GO terms: Cellular response to histamine, Inhibitory synapse assembly, Response to histamine, Gamma-aminobutyric acid signaling pathway, Protein import into nucleus, Protein import, Synapse assembly, Nucleocytoplasmic transport, Chemical synaptic transmission, Intracellular protein transport. X-axis: Enrichment score 1.0–3.5. Legend: Fold Enrichment 30/60/90; -log10(Enrichment FDR) 2.50–3.25)

Cut-off lines drawn at equivalents of p=0.01 and p=0.001

e

(Genome browser tracks for MARK3, with 2 kb scale bar. Tracks: MARK3 gene model, CLIP-seq TDP-43 rep1, CLIP-seq TDP-43 rep2, TDP-43 negative rep1 (×3), TDP-43 positive rep1 (×3))

f

MARK3

(Gel image with CTL and KD lanes showing Dist. 3'UTR and Prox. 3'UTR bands, 500 bp marker. Bar graph: Dist 3'UTR / Total, CTL vs KD, with * significance)

Figure 6

Figure 7