Rapid and Inducible Mislocalization of Endogenous TDP43 in a Novel Human Model of Amyotrophic Lateral Sclerosis

Johanna Ganssauge[1,2], Sophie Hawkins[*1,3], Seema C. Namboori[*1,3], Szi-Kay Leung[3], Jonathan Mill[3], Akshay Bhinge[1,3#].

1. Living Systems Institute, University of Exeter, EX4 4QD, UK
2. Biosciences, University of Exeter, EX4 4QD, UK
3. Clinical and Biomedical Sciences, University of Exeter, EX4 7BS, UK

* Indicates these authors contributed equally
# indicates corresponding author

Keywords: ALS, TDP43, proteinopathy, iPSC, splicing, microRNA

Abstract

Transactive response DNA binding protein 43 kDa (TDP43) proteinopathy, characterized by the mislocalization and aggregation of TDP43, is a hallmark of several neurodegenerative diseases including Amyotrophic Lateral Sclerosis (ALS). In this study, we describe the development of a new model of TDP43 proteinopathy using human induced pluripotent stem cell (iPSC)-derived neurons. Utilizing a genome engineering approach, we induced the mislocalization of endogenous TDP43 from the nucleus to the cytoplasm without mutating the TDP43 gene or using chemical stressors. Our model successfully recapitulates key pathological features of TDP43 proteinopathy, including neuronal loss, reduced neurite complexity, and cytoplasmic accumulation and aggregation of TDP43. Concurrently, the loss of nuclear TDP43 leads to splicing defects, while its cytoplasmic gain adversely affects microRNA expression. Strikingly, our observations suggest that TDP43 is capable of sustaining its own mislocalization, thereby perpetuating the proteinopathy. The development of this innovative model provides a valuable tool for the in-depth investigation of the consequences of TDP43 proteinopathy. We believe this will accelerate identification of potential therapeutic targets for the treatment of TDP43-associated neurodegenerative diseases including sporadic ALS.

Introduction

Amyotrophic Lateral Sclerosis (ALS) is a devastating neurodegenerative disorder characterised by the relentless degeneration of motor neurons (MNs) in both the brain and spinal cord[1]. This degeneration precipitates a cascade of symptoms including muscle weakness, atrophy, and paralysis, eventually leading to respiratory failure and death typically within three to five years after the onset of symptoms[2].

A hallmark of ALS pathology is the aberrant behaviour of the TAR DNA-binding protein 43 (hereafter TDP43). In ALS patients, TDP43, which normally resides in the nucleus, becomes mislocalized, forming aggregates in the cytoplasm of neurons and glial cells[3]. This phenomenon, termed TDP43 proteinopathy, is implicated in the majority of ALS cases and is considered a central player in the disease's pathogenesis[4]. TDP43 proteinopathy is a common feature in multiple age-associated neurodegenerative diseases including ALS, Limbic-predominant age-related TDP43 encephalopathy (LATE), frontotemporal dementia (FTD) and Alzheimer's disease (AD)[5]. TDP43 is an RNA-binding protein that plays a critical role in RNA metabolism, encompassing RNA splicing, stabilisation, transport, and microRNA processing, thus ensuring proper neuronal function[6]. Recent discoveries have demonstrated alternative splicing defects triggered by loss of nuclear TDP43, particularly via cryptic exon (CE) inclusion in transcripts of important neuronal genes *UNC13A*[7,8] and *STMN2*[9,10]. Notably, the multifaceted functions of TDP43 are vital for neuronal survival and homeostasis[11].

Several methods have been used to model TDP43 mislocalization in neurons. Techniques such as overexpression of mutated TDP43, the use of pharmacological agents to induce stress, and genetic manipulation through RNA interference or CRISPR/Cas9 have been employed in different cellular models to study TDP43 proteinopathy[12,13]. Despite the insights gained from these models, they harbour inherent limitations. Overexpression may not represent physiological conditions, pharmacological induction might induce events unrelated to the disease, and genetic manipulations can be confounded by off-target effects[14]. Furthermore, many models lack the sophisticated cellular environment that characterises human neurons, limiting their applicability in faithfully replicating the disease pathology[15].

In this study, we present an innovative model of TDP43 proteinopathy that allows us to trigger cytoplasmic mislocalization of endogenous non-mutated TDP43 on demand in healthy control iPSC-MNs (Fig.1A). This technological advancement offers an unprecedented level of insight into TDP43 proteinopathy and its role in ALS.

Main

We fused green fluorescent protein (GFP) to the C-terminus of TDP43 in its endogenous locus in healthy control iPSCs using CRISPR-Cas9 genome editing (Fig.1A) and selected iPSC clones with a homozygous knock-in (Supplementary Fig.1A). We initially ascertained that the integration of GFP did not adversely impact the differentiation of iPSCs into motor neurons (MNs). To this end, iPSCs engineered with TDP43-GFP were subjected to differentiation directed towards the MN lineage, employing a set of well-established protocols

[16,17]. Based on the expression of established MN markers - ISLET1 (ISL1) and neurofilament-M (NF-M) - we confirmed that the edited iPSCs (with GFP integration) efficiently differentiated into MNs (Supplementary Fig.1B). Moreover, the TDP43-GFP knock-in MNs did not significantly alter expression levels of TDP43, UNC13A and STMN2, nor cause upregulation of cryptic transcripts in STMN2 or UNC13A (Supplementary Fig.1C). Overall, these results support the premise that the GFP knock-in had no detrimental effects on the iPSCs' ability to differentiate into MNs or TDP43 function.

These engineered iPSCs, hereafter called TDP43-GFP iPSCs, were differentiated into MNs (called TDP43-GFP MNs) where ~80% of the cells in culture immunostained positive for ISL1. We expressed an anti-GFP nanobody (12kDa) using adeno-associated viruses (AAVs) in the TDP43-GFP MNs. Expression of the nanobody (deemed the control nanobody) did not affect TDP43 localization from the nucleus (Fig.1B). We engineered the nanobody with a nuclear export signal (NES). Expression of the NES-nanobody in the TDP43-GFP neurons caused relocation of the fusion protein to the cytoplasm with a concomitant loss in the nucleus (Fig.1B, C). We observed that TDP43 mislocalization leads to reduced dendrite complexity and neuronal soma swelling and an increase in cleaved caspase-3 activation, which is an indicator of apoptosis (Fig.1D,E, Supplementary Fig.1D). We confirmed that expression of the nanobodies in unedited iPSCs did not lead to apoptosis activation or dendrite defects, confirming that the observed phenotypes are due to TDP43 mislocalization (Supplementary Fig.1E). Mislocalization of TDP43 is commonly accompanied by cytoplasmic aggregation in tissue from patients with sporadic ALS. Accordingly, we observed TDP43 aggregates in the cytoplasm homozygous TDP43-GFP MNs that varied in size and number across individual neurons (Fig.1F). Aggregation of TDP43 was accompanied by a reduction in detected protein levels in the soluble fraction of the lysate (Fig.1F, G, Supplementary Fig.1F).

Having demonstrated that our model can recapitulate cellular and biochemical features observed in sporadic ALS, we next explored the molecular consequences triggered by TDP43 proteinopathy. Previous studies have highlighted the prevalence of alternative splicing defects affecting transcripts expressed from *UNC13A* and *STMN2* in ALS[7–10]. These splicing defects are considered hallmarks in the progression of the disease. Given the significance of this phenomenon, we sought to ascertain if MNs expressing the NES nanobody exhibit these characteristic splicing defects. We performed RT-qPCR analysis on motor neurons that expressed the control or NES nanobody. The results strikingly mirrored the commonly observed ALS profile, showing inclusion of cryptic exons in both *UNC13A* and *STMN2* and a reduction in the abundance of canonical transcripts (Supplementary Fig. 1G).

This demonstrated that our model proficiently recapitulates the critical molecular features observed in sporadic ALS.

Next, we ventured to explore the molecular consequences triggered by TDP43 proteinopathy across the whole transcriptome. We performed transcriptomic analysis on day 40 TDP43-GFP MNs to assess changes in the gene expression profiles due to TDP43 mislocalization. Principal component analysis (PCA) demonstrated that the primary axis of variance (90%) is related to the expression of the control and NES nanobody while the much smaller 2nd PC component related to the two clones (Fig.2A). Subsequently, we conducted a differential gene expression analysis using DESeq2 and visualised the results via a volcano plot. Our analysis revealed differential expression of hundreds of genes with a fold change of ≥ 2 and false discovery rate (FDR) < 0.01, suggesting a profound impact of TDP43 dysfunction on the transcriptome of MNs (Fig.2B). Gene ontology (GO) analysis of the differentially expressed genes highlighted a significant enrichment of pathways related to synaptic dysfunction and cytoskeletal defects in axons and dendrites amongst downregulated genes (Fig.2C). This finding is of considerable relevance to ALS pathogenesis, as neurons affected by ALS often exhibit impairments in these functions. In contrast, upregulated genes were enriched for GO terms related to RNA processing including the nonsense mediated decay (Fig. 2C).

Given that alternative splicing (AS) defects are a hallmark of sporadic ALS, we utilized Leafcutter[23] to analyze potential alternative splicing in MNs caused by TDP43 mislocalization. Our analysis identified alterations in the splicing of 231 genes, including *UNC13A* and *STMN2* (FDR < 0.01). To gain functional insights into these splicing defects, we performed pathway enrichment analysis on affected genes, identifying a significant enrichment of terms related to synaptic development (FDR < 0.01). Furthermore, to gain mechanistic understanding of the role of TDP43 in the observed splicing defects, we compared the splicing results with TDP43 eCLIP data. Interestingly, only 35% of the splicing changes were proximal to a TDP43 binding site (Supplementary Fig. 2 and 3). This suggests that the loss of TDP43 from the nucleus may be responsible for a subset of the observed splicing defects but not all, indicating the possibility of additional indirect underlying mechanisms. We compared our splicing results with publicly available transcriptomic data generated from cortical neuronal nuclei with or without TDP43 purified from ALS patient tissue post-mortem. At a stringent significant threshold of 0.01, we detected 106 genes as alternatively spliced in nuclei that showed a loss of TDP43 using Leafcutter. Thirty-five out of these 106 genes were also detected in our iPSC model (overlap p-value = 5.2 x 10e-31) (Fig.2D), indicating strong concordance between the iPSC model and patient data, despite

the differences in sample origin and neuronal subtype. Importantly, 34 of these genes displayed an identical splicing event. To evaluate whether the observed splicing changes were due to nuclear loss of TDP43, we compared our splicing results with transcriptomic data generated in iPSC-derived MNs after TDP43 knockdown (KD) (Fig.2E). Strikingly, almost 60% of the genes that displayed splicing defects due to a global loss of TDP43 also displayed splicing changes in our mislocalization model (1.6 x 10e-40). However, 172 genes that displayed splicing changes due to TDP43 mislocalization were not affected by TDP43 KD. This indicates that nuclear loss of TDP43 alone cannot entirely explain the widespread defects in AS.

To generate a robust list of AS events resulting from TDP43 pathology, we set a stringent p-value threshold of 10e-4 and only retained those genes that were identified in our RNA-seq results and those obtained from ALS/FTD post-mortem tissue. Our analysis identified 12 genes, including *STMN2* and *UNC13A*, that we further investigated using RT-qPCR. This confirmed AS events in all the genes identified, where 10/12 genes displayed a decrease in expression of their canonical transcript levels (Fig.2F). Further, most of these genes also displayed AS changes due to TDP43 KD in iPSC-MNs, although the extent of these changes for a subset of the genes tested was less dramatic (Supplementary Fig.4).

A key drawback with earlier TDP43 models was the inability to characterize early changes after TDP43 mislocalization in neurons. To enable inducible control of TDP43 cytoplasmic localization in our model, we expressed the nanobody under a doxycycline-inducible promoter, and the entire circuit was knocked into the AAVS1 locus in the homozygous TDP43-GFP iPSCs. We called these iPSCs TDP43-GFP-CTRL or TDP43-GFP-NES iPSCs (Fig.3A). We first confirmed that we could induce TDP43 mislocalization in the TDP43-GFP-CTRL/NES iPSC-derived MNs using doxycycline (Dox). Dox (1 µg/ml) triggered significant TDP43 mislocalization in the TDP43-GFP-NES MNs, while the -CTRL line, and the -NES line without Dox, maintained nuclear TDP43 (Fig.3B). Notably, 48 hours of doxycycline treatment was sufficient to cause mislocalization (Supplementary Fig.5A).

To detect early changes in expression and splicing post TDP43 mislocalization, we harvested neurons for RT-qPCR 48 hours after the addition of Dox at day 20 of the differentiation process. The expression levels and splicing patterns for nine genes were compared with control cultures that were continuously treated with Dox until day 40 (Supplementary Fig.5B,C). Strikingly, all nine genes including *UNC13A* and *STMN2* displayed AS events as early as 48 hours post-TDP43 mislocalization. This suggests that

dysfunctional AS could be one of the incipient molecular events in ALS pathogenesis due to TDP43 mislocalization.

A significant area of research in the field of ALS involves pinpointing the upstream triggers responsible for causing the mislocalization of TDP43. The underlying premise of these investigations is that eliminating the trigger may potentially reverse the mislocalization of TDP43. However, it remains to be confirmed whether this hypothesis holds true. We wanted to ascertain whether TDP43, once mislocalized, perpetuates in its mislocalized state even after the initial trigger has been removed. We induced TDP43 mislocalization by applying doxycycline to TDP43-GFP-NES iPSC-derived MNs at day 20. Subsequently, doxycycline treatment was withdrawn after three days (day 23) and neurons were harvested 7 (day 30) and 14 days (day 37) after Dox-withdrawal. As a benchmark for successful induction, positive control cultures were continuously subjected to doxycycline treatment. Conversely, the negative controls comprised neurons that were never exposed to doxycycline. We evaluated TDP43 localization using immunofluorescence microscopy and in parallel assessed nanobody transcript levels using RT-qPCR.

In alignment with our expectations, positive control samples exhibited significant TDP43 mislocalization at both time points, verifying the efficacy of doxycycline in this respect. In contrast, TDP43 was found exclusively in the nucleus of our negative control samples, reiterating the specificity of doxycycline-induced TDP43 mislocalization. Interestingly, we observed persistent TDP43 mislocalization in a subset of neurons even after 14 days of doxycycline withdrawal (Fig.3C), despite near-undetectable levels of nanobody expression (Fig.3D). This was accompanied by the CE inclusion in STMN2 with concomitant decrease in its canonical transcript level (Fig.3E). Approximately 50% of neurons, however, showed partial recovery of TDP43 into the nucleus. This data indicates that removal of the initial trigger is insufficient to completely reverse TDP43 once it has been mislocalized. Previous studies have indicated that mutant mislocalized TDP43 can further trap normal forms of the protein, establishing a positive feedback loop[18,19]. However, these studies were performed by exogenous expression of TDP43 to non-physiological levels. To ascertain whether the non-mutated form of TDP43 was capable of initiating such a feedback, we deployed our heterozygous TDP-GFP iPSC lines and differentiated these into MNs. Expression of the nanobody mislocalized the TDP43-GFP fusion protein as expected. Strikingly, we observed a significant reduction of the untagged TDP43 protein within the nucleus, with a corresponding increase in the cytoplasm (Fig.3F). In a subset of the heterozygous TDP43-GFP neurons, we observed near total loss of nuclear TDP43. To rule out any homozygous TDP43-GFP cell contamination in the heterozygous clones, we screened 30 randomly

selected clones from our heterozygous pool. All of these clearly showed heterozygous GFP knock-in. These results validate our hypothesis that once mislocalized, cytoplasmic TDP43 can sustain a positive feedback that further traps additional TDP43 protein. These observations carry substantial implications for our understanding of TDP43 pathology in ALS. From a therapeutic perspective, this indicates that rectifying the cause of the mislocalization may not be adequate to reverse the ensuing pathology. This understanding underscores the need for therapeutic strategies that can effectively disrupt the sustained loop of TDP43 mislocalization, beyond addressing the initial trigger as well as targeting the downstream consequences due to the mislocalization.

Finally, since TDP43 is involved in microRNA biogenesis, we sought to analyse changes in the microRNA profiles associated with TDP43 mislocalization. For this purpose, small RNA sequencing was carried out on day 40 MNs after triggering TDP43 mislocalization at day 20 by the addition of doxycycline. Our analysis revealed a striking alteration in the landscape of microRNA expression as a result of TDP43 mislocalization. Principal component analysis (PCA) demonstrated distinct separation of the control and mislocalized samples (PC1 77%), emphasising the profound impact of TDP43 mislocalization on microRNA profile (Supplementary Fig. 6A).

Around 150 microRNAs were found to be significantly altered (FDR < 0.01 and fold change ≥ 1.5) upon TDP43 mislocalization. Our data captured downregulation of the microRNAs miR-218, and miR-9, which have previously been shown to be downregulated in ALS MNs[20 21]. Interestingly, these changes were not unidirectional; almost equal numbers of microRNAs were upregulated or downregulated (Supplementary Fig. 6B). The results of this study indicate that TDP43 mislocalization leads to global dysregulation of microRNA expression in iPSC-derived MNs. However, it is noteworthy that our observations contrast with those found in postmortem tissue studies, where all differentially expressed microRNAs were reported to be downregulated. This divergence in findings may suggest that the uniform downregulation observed in postmortem tissues could be attributed to changes that are triggered by end-stage dying neurons, which might not represent the whole spectrum of molecular alterations especially early in the disease progression.

We wanted to evaluate whether alterations in microRNA expression profiles induced by TDP43 mislocalization are congruent with the changes caused by TDP43 knockdown. For this investigation, small RNA sequencing was utilized to assess microRNA expression in motor neurons (MNs) at day 30 following the knockdown of TDP43 using shRNAs. Again, principal component analysis revealed a distinct separation between all control and TDP43

knockdown samples, indicating a clear impact of TDP43 knockdown on microRNA expression profiles (Supplementary Fig.6C). However, a striking contrast was observed in the number of microRNAs affected by TDP43 knockdown as compared to TDP43 mislocalization. Specifically, in the case of TDP43 knockdown, only one microRNA (miR-1249-3p) exhibited a significant change (FDR < 0.01 and a fold change of 1.5). This is in stark contrast to the observations made when TDP43 was mislocalized, where over 150 microRNAs displayed altered expression. To avoid setting thresholds in our comparisons, we first ranked the microRNAs impacted by TDP43 knockdown based on their fold changes. Subsequently, we utilized gene set enrichment analysis (GSEA) to determine if the microRNAs disrupted by TDP43 mislocalization significantly coincided with our knockdown data. We observed a substantial overlap between microRNAs that were downregulated due to TDP43 mislocalization and those downregulated following TDP43 knockdown (Supplementary Fig. 3E). However, we did not observe a significant overlap for microRNAs that were upregulated in both scenarios.

This disparity holds significant implications. Firstly, it suggests that mislocalized TDP43 selectively affects the expression of a subset of microRNAs. Secondly, these observations underscore that the nuclear loss and cytoplasmic gain of TDP43 induce distinct molecular alterations within motor neurons that converge to hasten neuronal dysfunction and demise.

Summary

In summary, our unique iPSC model of TDP43 proteinopathy offers an unprecedented access to investigate cellular and molecular events in human neurons in a temporal fashion downstream of TDP43 mislocalization. We expect this model to not only enhance our understanding of ALS but also other TDP43 proteinopathies and accelerate efforts into developing therapies against these devastating neurodegenerative diseases.

Acknowledgements

We would like to thank Aaron Jeffries and the University of Exeter sequencing centre for the Illumina and Oxford Nanopore Technologies long read sequencing. Thanks to Jessica Board for her help with the TDP43 shRNAs. This study was funded by an MNDA Pilot award and an MRC New Investigator Grant to A.B.. This study was supported by the National Institute for Health and Care Research Exeter Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Materials and Methods

Cell Culture

Motor neurons were generated from iPSCs as described previously[16,17]. Briefly, iPSCs were plated onto matrigel and differentiated by treatment with neuronal differentiation media (DMEM/F12:Neurobasal in a 1:1 ratio, HEPES 10 mM, N2 supplement 1%, B27 supplement 1%, L-glutamine 1%, ascorbic acid 5 µM) supplemented with SB431542 (40 µM), CHIR9921 (3 µM) and LDN98312 (0.2 µM) from day 0 until day 4. Cells were caudalized by treatment with 0.1 µM retinoic acid starting at day 2 and ventralized with 1 µM purmorphamine starting at day 4 and continued till day 10. At day 8, progenitors were replated onto poly-D-lysine/laminin coated wells and differentiated with 10 µM DAPT for 3 days. Non-mitotic cells were removed with a pulse of 10 µg/ml mitomycin-C for 1 hour at day 14. Motor neurons were subsequently maintained in N2B27 media supplemented with 10 ng/ml BDNF and GDNF, with half media changes occurring twice per week.

TDP43-GFP mislocalization was induced via transduction with AAVs expressing anti-GFP nanobodies. In our TDP43-GFP-NES/ CTRL cell line, TDP43-GFP mislocalisation was achieved via addition of 1 µg/mL doxycycline, which was replenished every 48 hours.

Immunocytochemistry

Cells were fixed in 4% paraformaldehyde for 20 minutes at room temperature (RT), followed by permeabilization in ice-cold methanol for 5 minutes. Blocking was performed in 1% BSA (in PBS) for 1 hour (RT) and primary antibodies (Supplementary Table 2) were incubated overnight at 4°C. Next day, wells were washed in PBS. Secondary antibodies (Molecular probes, 1:2,000) were incubated for 1-2 hours (RT), and nuclei were stained with Hoechst 33542 (1:1,000; Molecular Probes). All antibodies were diluted in blocking agent. Plates were imaged using ImageXpress® Pico (Molecular Devices).

RNA Sequencing and analysis

RNA was extracted using the QIAzol and the NEB Monarch RNA extraction kit and ribosomal RNA was depleted using the NEBNext rRNA depletion kit (NEB) . Libraries for Illumina short-read sequencing were prepared using the NEBNext UltraII RNA library kit

(NEB) according to manufacturer instructions. RNA fragment sizes were assessed by RNA pico chip (Agilent Bioanalyzer). Sequencing was performed at the Exeter sequencing centre using the NovaSeq platform. Reads were mapped to the human genome assembly hg19 using STAR[22] and counts for each gene per sample were generated using the R package featureCounts[23]. Differential expression analysis was performed using DESeq2[24]. Alternatively spliced transcripts were detected using Leafcutter[25]. Libraries for long read ONT platform were prepared according to the manufacturer's instructions and sequenced on a PromethION sequencer at the Exeter sequencing centre. Reads were mapped to hg19 using minimap2[26] and transcript isoforms were detected using FICLE[27]. Small RNA sequencing was performed commercially through Macrogen and the fastq data was mapped and counts generated using the miRDeep2 package[28]. Differential expression was performed using DESeq2.

RT-qPCR

Following lysis in QIAzol, total RNA was purified using The RNeasy Micro Kit (Qiagen). A total of 500 ng RNA was used for the generation of cDNA using the High-capacity cDNA kit (ThermoFisher). The RT-qPCR was performed using the Promega® 2x Master Mix and the QuantStudio™ 6 (Applied Biosystems) for 40 cycles. Housekeeping genes GAPDH, HPRT1, and RPL13 were used for normalization. Using the $\Delta\Delta$Ct method, relative fold changes were calculated per sample relative to their appropriate control. Significance testing was performed using the student's T-test. Primer sequences have been included in Supplementary Table 1.

Western Blotting

Motor neuron samples at D42 were lysed in ice-cold RIPA buffer supplemented with HALT™ protease phosphatase inhibitors. Lysates were centrifuged at 15,700 x g, for 30 minutes at 4°C. The supernatant was used as the soluble protein fraction. A total of 3 µg protein was separated by SDS-PAGE gel electrophoresis (Bio-Rad) and transferred to 0.2 µm nitrocellulose membrane (Amersham™ Protran™) for immunoblotting. Membranes were blocked for 1 hour at RT using Intercept-T blocking buffer (Li-Cor). Primary antibodies (Supplementary Table 2) were incubated overnight at 4°C. Secondary antibodies were incubated for 1 hour at RT. All antibodies were diluted in Intercept-T with 0.2% Tween-20 and washed in TBST. Blots were imaged using the Li-Cor Odyssey Fc Imager. ImageJ (FIJI) was used for quantification of band intensity. All bands were normalized to alpha-tubulin.

References

1. Brown, R. H. & Al-Chalabi, A. Amyotrophic Lateral Sclerosis. *N. Engl. J. Med.* **377**, 162–172 (2017).
2. Hardiman, O. *et al.* Amyotrophic lateral sclerosis. *Nat Rev Dis Primers* **3**, 17071 (2017).
3. Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).
4. Ling, S.-C., Polymenidou, M. & Cleveland, D. W. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* **79**, 416–438 (2013).
5. de Boer, E. M. J. *et al.* TDP-43 proteinopathies: a new wave of neurodegenerative diseases. *J. Neurol. Neurosurg. Psychiatry* **92**, 86–95 (2020).
6. Buratti, E. & Baralle, F. E. TDP-43: gumming up neurons through protein-protein and protein-RNA interactions. *Trends Biochem. Sci.* **37**, 237–247 (2012).
7. Brown, A.-L. *et al.* TDP-43 loss and ALS-risk SNPs drive mis-splicing and depletion of UNC13A. *Nature* **603**, 131–137 (2022).
8. Ma, X. R. *et al.* TDP-43 represses cryptic exon inclusion in the FTD-ALS gene UNC13A. *Nature* **603**, 124–130 (2022).
9. 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).
10. Melamed, Z. ’ev *et al.* Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration. *Nat. Neurosci.* **22**, 180–190 (2019).
11. 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).
12. Barmada, S. J. *et al.* Cytoplasmic mislocalization of TDP-43 is toxic to neurons and

enhanced by a mutation associated with familial amyotrophic lateral sclerosis. *J. Neurosci.* **30**, 639–649 (2010).

13. Miguel, L. *et al.* Accumulation of insoluble forms of FUS protein correlates with toxicity in Drosophila. *Neurobiol. Aging* **33**, 1008.e1–15 (2012).
14. Scaramuzzino, C. *et al.* Protein arginine methyltransferase 1 and 8 interact with FUS to modify its sub-cellular distribution and toxicity in vitro and in vivo. *PLoS One* **8**, e61576 (2013).
15. Lenzi, J. *et al.* ALS mutant FUS proteins are recruited into stress granules in induced pluripotent stem cell-derived motoneurons. *Dis. Model. Mech.* **8**, 755–766 (2015).
16. Maury, Y. *et al.* Combinatorial analysis of developmental cues efficiently converts human pluripotent stem cells into multiple neuronal subtypes. *Nat. Biotechnol.* **33**, 89–96 (2015).
17. Bhinge, A., Namboori, S. C., Zhang, X., VanDongen, A. M. J. & Stanton, L. W. Genetic Correction of SOD1 Mutant iPSCs Reveals ERK and JNK Activated AP1 as a Driver of Neurodegeneration in Amyotrophic Lateral Sclerosis. *Stem Cell Reports* **8**, 856–869 (2017).
18. Winton, M. J. *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**, 13302–13309 (2008).
19. Gasset-Rosa, F. *et al.* Cytoplasmic TDP-43 De-mixing Independent of Stress Granules Drives Inhibition of Nuclear Import, Loss of Nuclear TDP-43, and Cell Death. *Neuron* **102**, 339–357.e7 (2019).
20. Reichenstein, I. *et al.* Human genetics and neuropathology suggest a link between miR-218 and amyotrophic lateral sclerosis pathophysiology. *Sci. Transl. Med.* **11**, (2019).
21. Zhang, Z. *et al.* Downregulation of microRNA-9 in iPSC-derived neurons of FTD/ALS patients with TDP-43 mutations. *PLoS One* **8**, e76055 (2013).
22. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

23. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

24. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

25. Li, Y. I. *et al.* Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158 (2018).

26. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

27. Leung, S. K. *et al.* Long-read transcript sequencing identifies differential isoform expression in the entorhinal cortex in a transgenic model of tau pathology. *bioRxiv* 2023.09.20.558220 (2023) doi:10.1101/2023.09.20.558220.

28. Friedländer, M. R. *et al.* Discovering microRNAs from deep sequencing data using miRDeep. *Nat. Biotechnol.* **26**, 407–415 (2008).

Figure 1

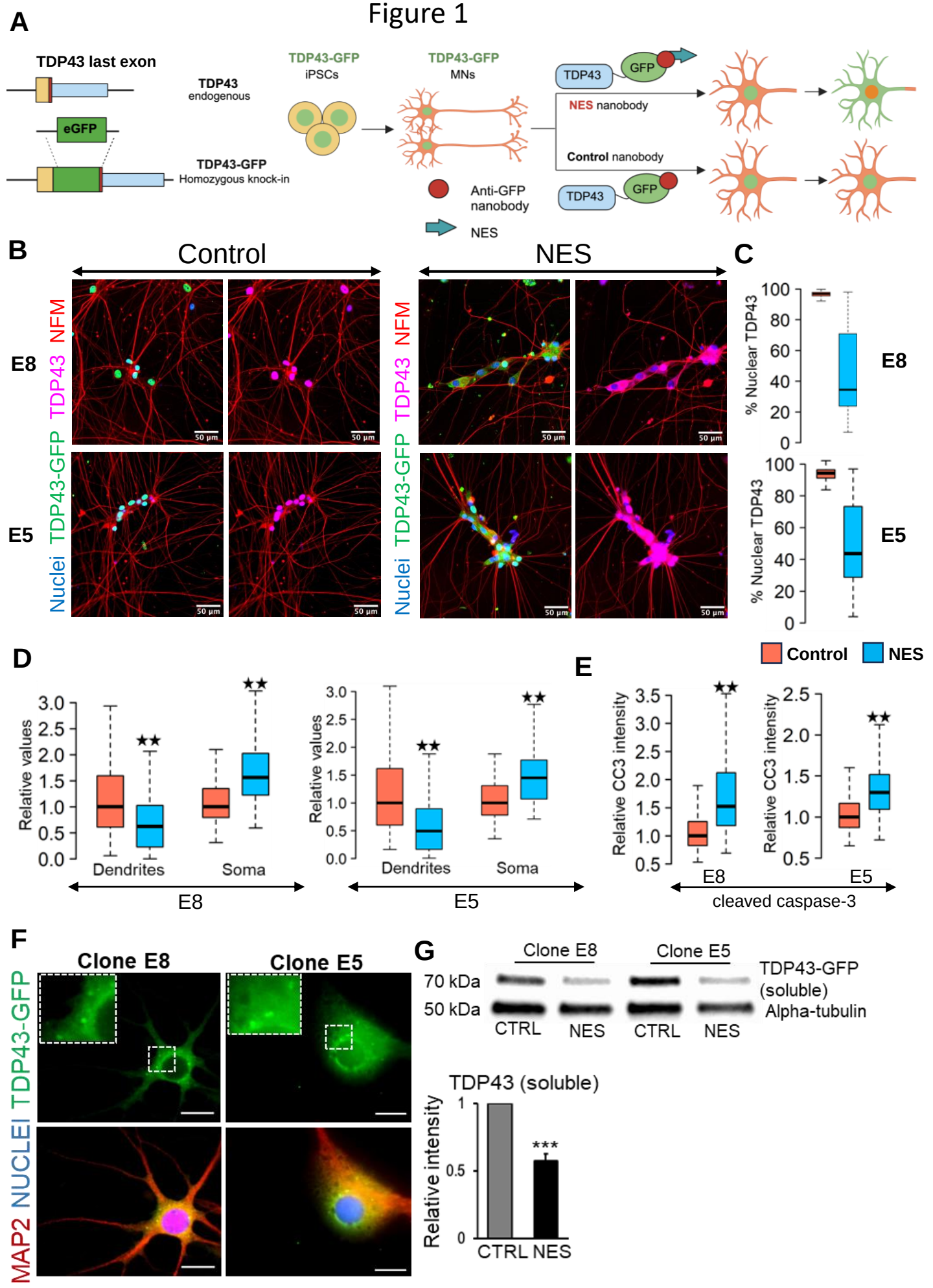

A
TDP43 last exon
TDP43
endogenous
eGFP
TDP43-GFP
Homozygous knock-in
TDP43-GFP
iPSCs
TDP43-GFP
MNs
TDP43
GFP
NES nanobody
Control nanobody
Anti-GFP
nanobody
NES
B
Control
NES
E8
E5
NFM
TDP43
TDP43-GFP
Nuclei
50 µm
C
% Nuclear TDP43
E8
E5
Control
NES
D
Relative values
Dendrites
Soma
E8
E5
E
Relative CC3 intensity
E8
E5
cleaved caspase-3
F
Clone E8
Clone E5
TDP43-GFP
NUCLEI
MAP2
G
Clone E8
Clone E5
70 kDa
50 kDa
TDP43-GFP
(soluble)
Alpha-tubulin
CTRL
NES
CTRL
NES
TDP43 (soluble)
Relative intensity
CTRL NES

Figure 1. A Human iPSC-based model of TDP43 proteinopathy in MNs

(A) Schematic depicting genome editing of healthy iPSCs to knock-in GFP into the C-terminus of endogenous *TARDBP*, encoding TDP43. The resulting TDP43-GFP iPSCs are differentiated into motor neurons (MNs) and transduced with AAVs encoding anti-GFP nanobodies. The “NES” nanobody includes a sequence for a strong nuclear export signal, which transports nuclear TDP43-GFP into the cytoplasm. The control nanobody (lacking NES) has no effect on TDP43-GFP localization.

(B) Representative images depicting TDP43 localization in two homozygous TDP43-GFP lines (E8, E5). MNs were transduced at day 20 of the differentiation process and fixed at day 40 for the immunostaining. TDP43 is expressed in the nucleus in the presence of the control nanobody and relocates to the cytoplasm in the presence of the NES nanobody. TDP43-GFP indicates signal from the anti-GFP antibody while TDP43 indicates signal from the anti-TDP43 antibody. Neurons were also stained with neurofilament-M (red) and Hoechst 33342 (blue nuclear stain). Scale bar = 50 µm.

(C) Quantification of the % of nuclear TDP43 in individual neurons transduced with control or NES nanobodies.

(D) Quantification of morphological defects in NES nanobody treated MNs versus control. Mislocalized TDP43 (NES) causes a reduction in dendritic tree length (Dendrites) and soma swelling (Soma), compared to neurons expressing nuclear TDP43 (control).

(E) Quantification of cleaved caspase-3 (CC3) expression as a marker of apoptosis. Cytoplasmic TDP43 causes significant elevation in apoptosis compared to nuclear TDP43.

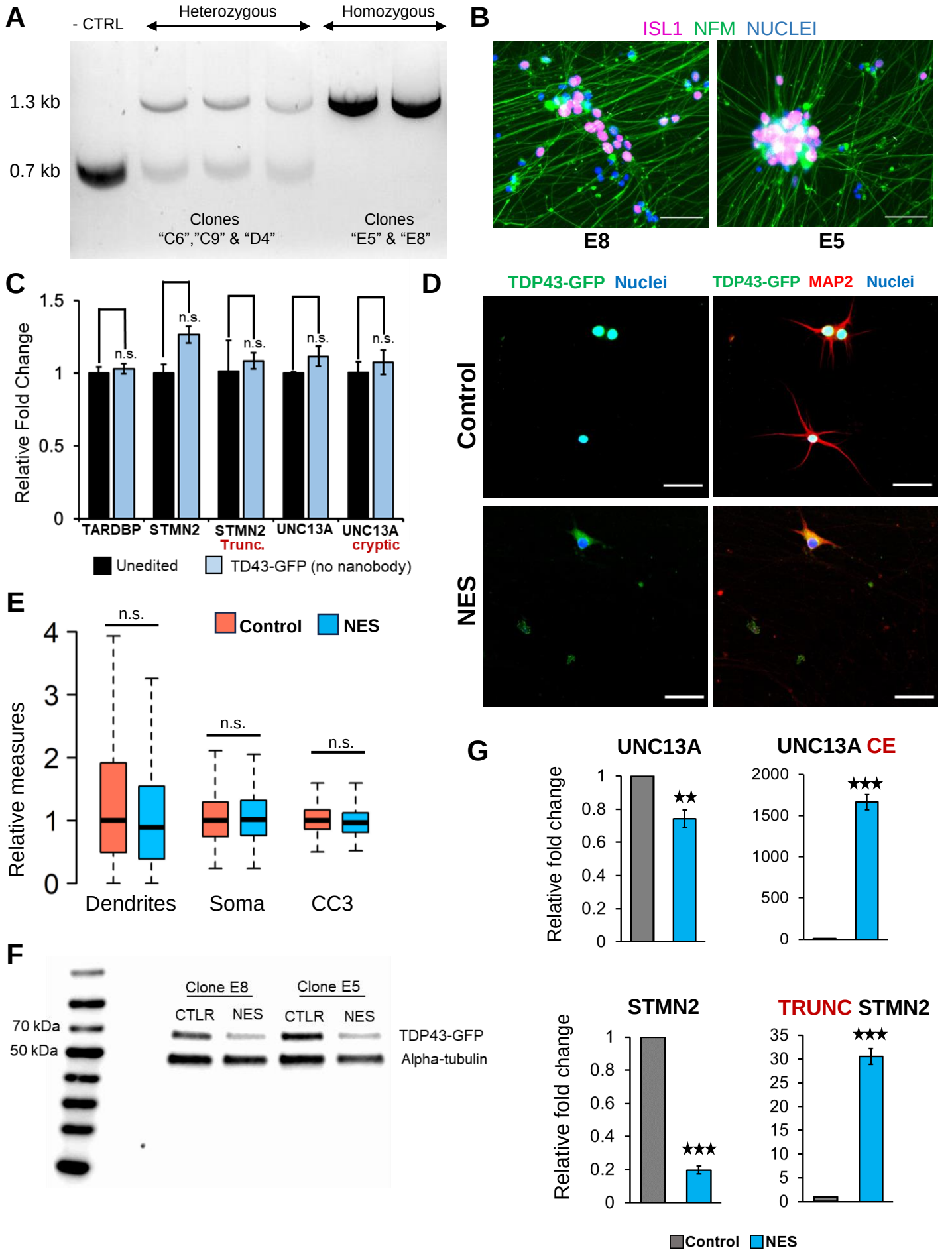
Panels C, D, and E display data for the two homozygous TDP43-GFP lines (E8 and E5), transduced with nanobodies at day 20 and fixed for staining at day 40. N = 3 independent differentiations per clone. At least 100 neurons were included per condition per replicate.

(F) Representative images showing cytoplasmic aggregation of TDP43 in day 40 homozygous TDP43-GFP MNs transduced with the NES nanobody at day 18. Scale bar = 15 µm.

(G) Upper panel: representative western blot image of soluble TDP43-GFP in NES versus control MNs at day 42, 22 days post-transduction with AAV nanobodies. Lower panel: Quantification of soluble TDP43-GFP from western blot. N=3 independent differentiations, two for E5 iPSCs, one for the E8 iPSCs.

** indicates $p < 0.01$,*** indicates $p < 0.001$.

Supplementary Figure 1



Supplementary Figure 1. Related to Figure 1

(A) Confirmation of the three heterozygous and two homozygous TDP43-GFP knock-in clones using PCR-gel electrophoresis. The unedited parent healthy iPSCs served as a negative control.

(B) Representative images of TDP43-GFP homozygous clones E8 and E5 stained for MN markers neurofilament-M (NFM) and ISLET1 (ISL1) at day 40. Nuclei were stained with Hoechst 33342 Scale bar = 50 µm.

(C) RT-qPCR of unedited (healthy) iPSC-MNs and untransduced homozygous TDP43-GFP knock-in MNs at day 30, suggesting no adverse effects of GFP knock-in on TDP43 expression or alternative splicing. P-values were calculated using a one-tailed Student's T-test. Error bars show SEM. None of the differences were significant.

(D) Representative immunofluorescent images of decreased dendrite complexity in NES compared to control MNs. Cells are from the E8 homozygous GFP knock-in clone. Scale bar = 50 µm.

(E) Morphological analysis in unedited iPSC MNs treated with control or NES nanobodies. No significant effect is observed on dendrite morphology, soma swelling or apoptosis (CC3 intensity) a p-value threshold of 0.01. N=2 independent differentiations. At least 150 neurons were analyzed per replicate. CC3 = cleaved caspase-3

(F) Uncropped western blot (from Fig.1G) of soluble TDP43 in control (CTRL) and mislocalized (NES) MNs lysed at D42, 22 days post-transduction with AAVs. Clones represent independent differentiations of E5 and E8 homozygous TDP43-GFP knock-in lines.

(G) RT-qPCR showing abnormal splicing of UNC13A and STMN2 in day 40 homozygous TDP43-GFP MNs, 20 days post-transduction with NES or control nanobodies. Error bars show SEM. CE = cryptic exon, TRUNC = truncated.

n.s. = not significant, ** indicates $p < 0.01$,*** indicates $p < 0.001$.

Figure 2

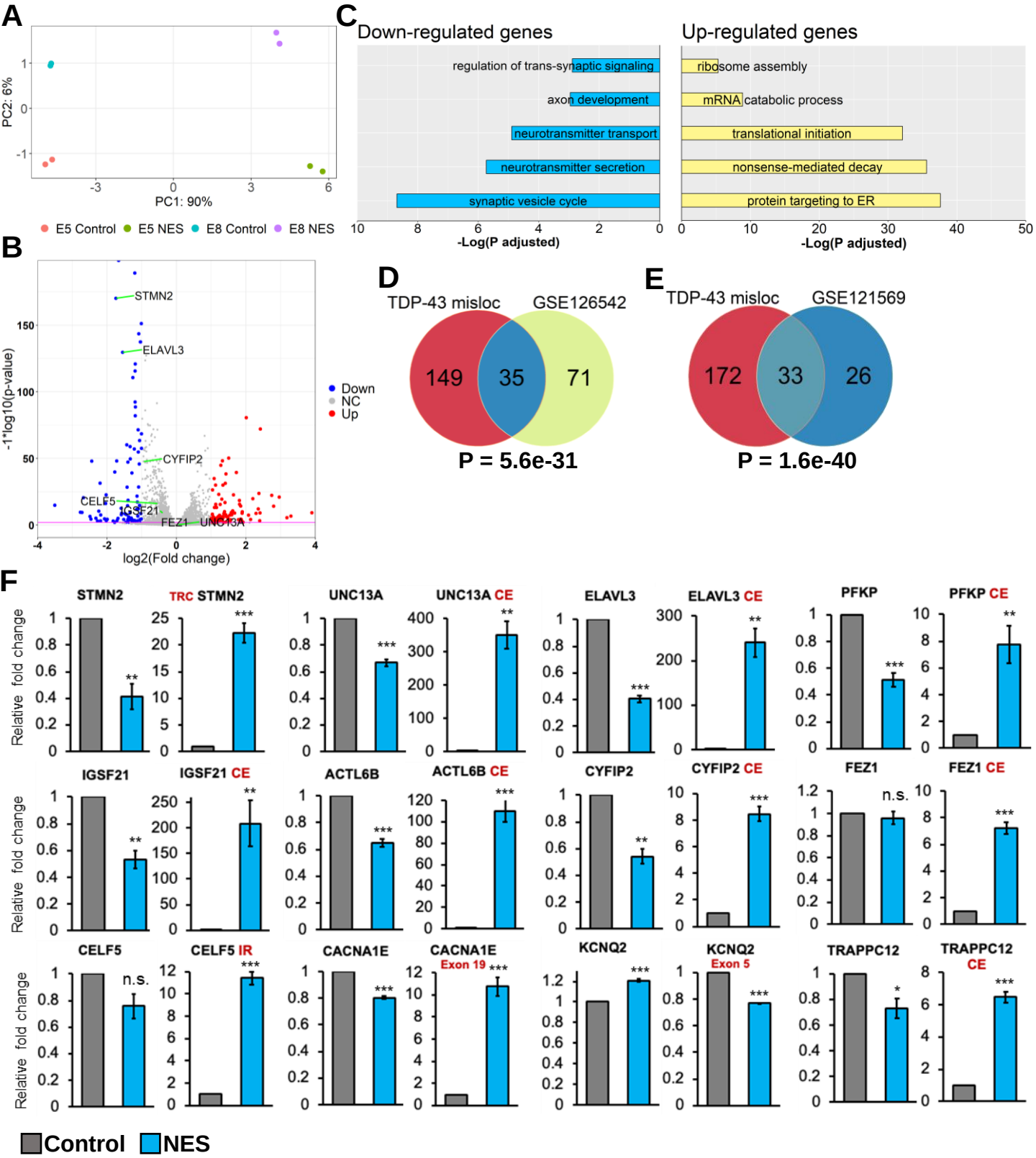

A
PC2: 6%
PC1: 90%
E5 Control
E5 NES
E8 Control
E8 NES
C
Down-regulated genes
regulation of trans-synaptic signaling
axon development
neurotransmitter transport
neurotransmitter secretion
synaptic vesicle cycle
-Log(P adjusted)
Up-regulated genes
ribosome assembly
mRNA catabolic process
translational initiation
nonsense-mediated decay
protein targeting to ER
-Log(P adjusted)
B
STMN2
ELAVL3
CYFIP2
CELF5
IGSF21
FEZ1
UNC13A
Down
NC
Up
-1*log10(p-value)
log2(Fold change)
D
TDP-43 misloc
GSE126542
149
35
71
P = 5.6e-31
E
TDP-43 misloc
GSE121569
172
33
26
P = 1.6e-40
F
STMN2
TRC STMN2
UNC13A
UNC13A CE
ELAVL3
ELAVL3 CE
PFKP
PFKP CE
IGSF21
IGSF21 CE
ACTL6B
ACTL6B CE
CYFIP2
CYFIP2 CE
FEZ1
FEZ1 CE
CELF5
CELF5 IR
CACNA1E
CACNA1E Exon 19
KCNQ2
KCNQ2 Exon 5
TRAPPC12
TRAPPC12 CE
Relative fold change
n.s.
Control
NES

Figure 2. Transcriptional Consequences of TDP43 Mislocalization

A) Principal component analysis of gene counts for E8 and E5 iPSC-MNs.

B) Volcano plot displaying genes differentially expressed due to TDP43 mislocalization. Red: upregulated genes, Green: downregulated genes, NC: No Change. Horizontal line denotes a p-value threshold of 0.01.

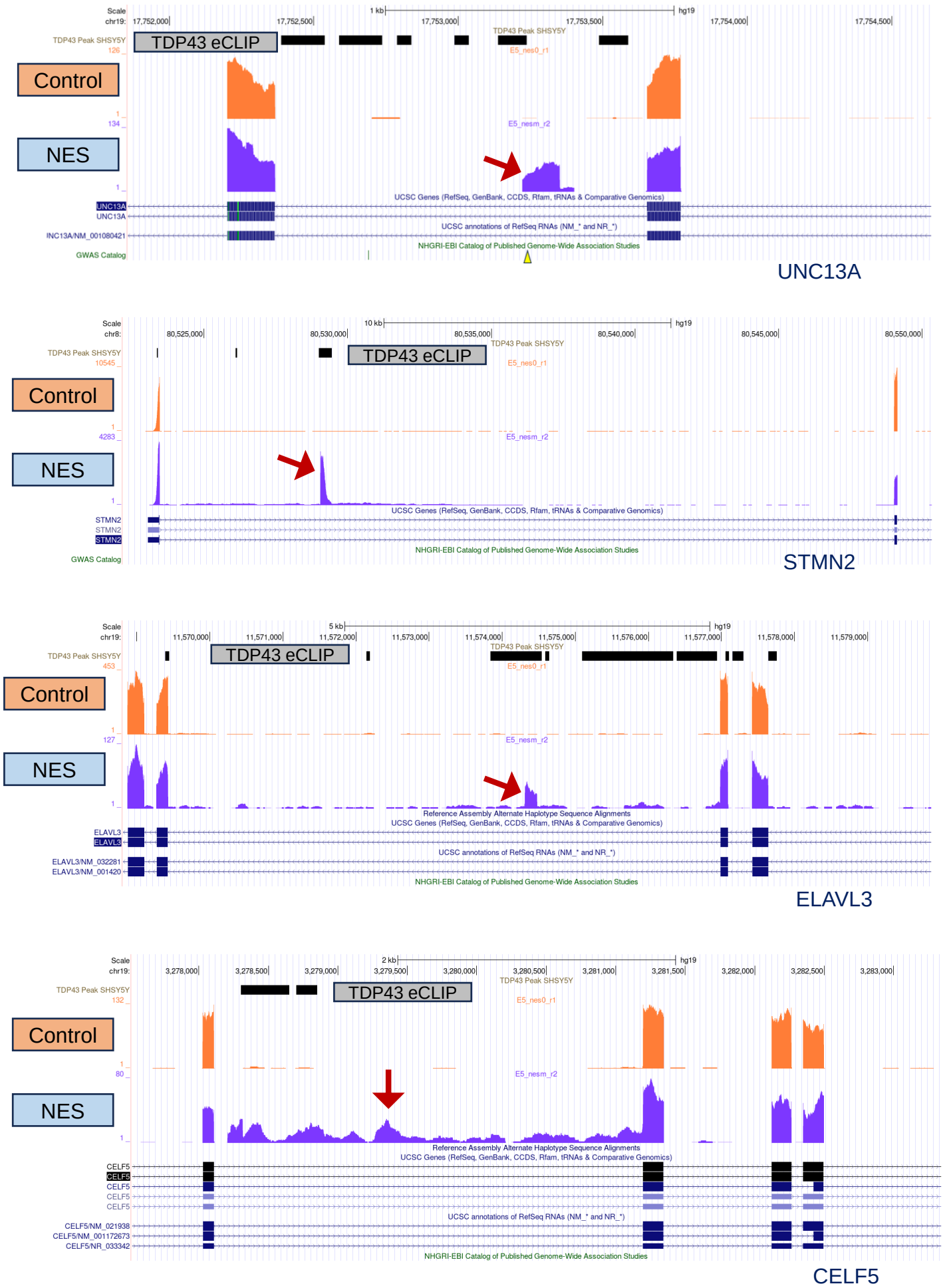
C) Gene ontology enrichment analysis on the differentially expressed genes in the TDP43 model. The top five enriched pathways have been displayed.

D) Venn diagram displaying the overlap between mis-spliced genes identified due to TDP43 mislocalization in the TDP43-GFP iPSC MNs and cortical neuronal nuclei without TDP43 obtained from ALS/FTD patient tissue (GSE126542).

E) Venn diagram displaying the overlap between mis-spliced genes identified due to TDP43 mislocalization in the TDP43-GFP iPSC MNs and TDP43 knockdown using siRNAs in healthy iPSC-MNs (GSE121569).

F) RT-qPCR validating alternative splicing changes resulting from TDP43-GFP mislocalization in iPSC-MNs. MN samples were lysed at D30, 12 days post-transduction with AAVs. Replicates were three independent differentiations of homozygous TDP43-GFP knock-in lines, (two of E5, one of E8). CE = cryptic exon, TRC = truncated, IR = intron retention.
* indicated $p < 0.05$, ** indicates $p < 0.01$. *** indicates $p < 0.001$.

Supplementary Figure 2

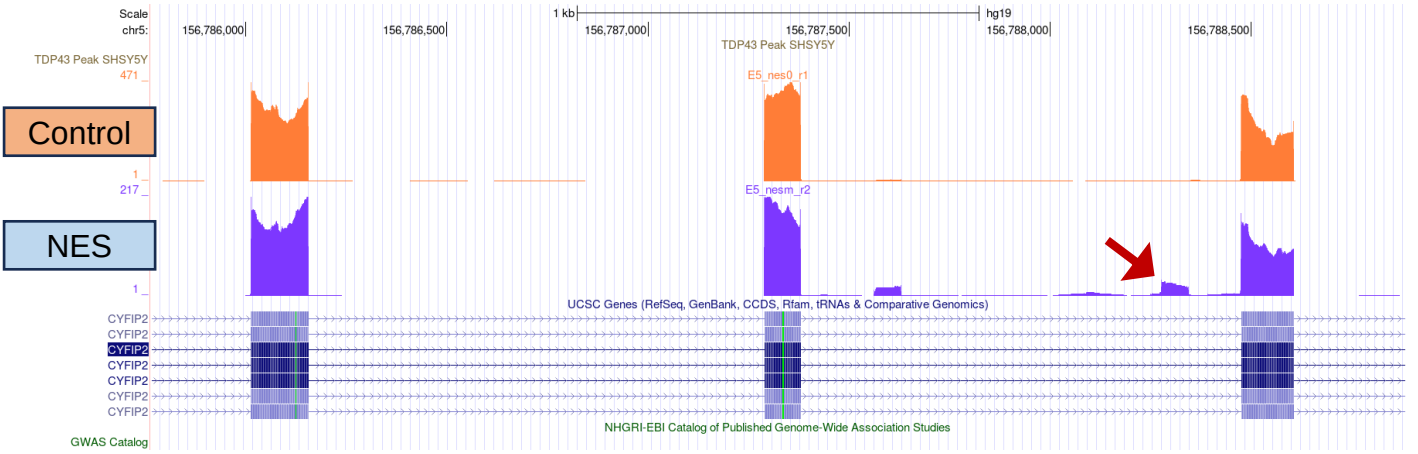


Supplementary Figure 2 (Related to Figure 2)

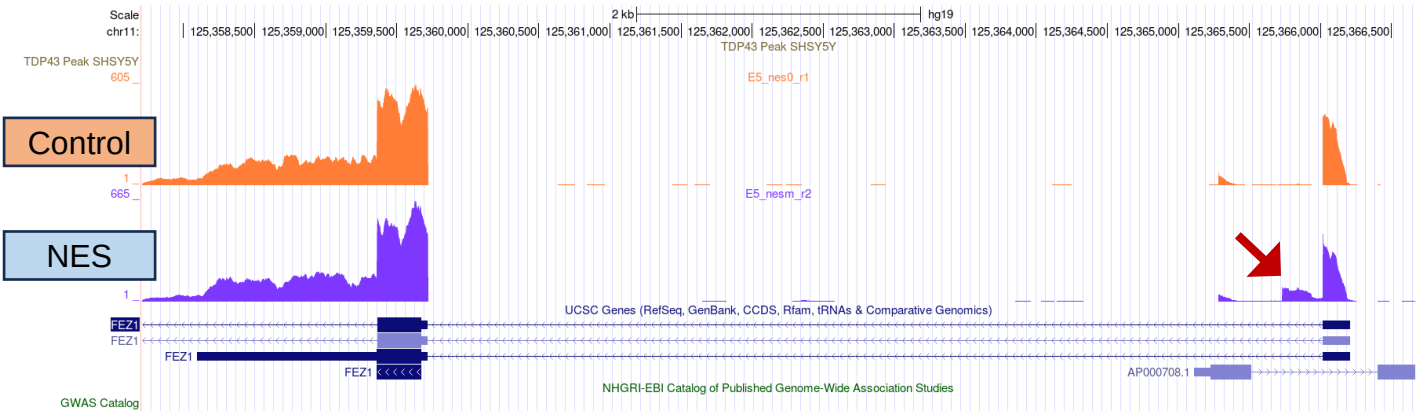
UCSC Genome browser screen shots depicting cryptic exon inclusion (*UNC13A, STMN2, ELAVL3*) and intron retention (*CELF5*) resulting from TDP43 mislocalization. In all cases, there is at least one TDP43 binding site, identified by SH-SY5Y eCLIP data (GSE122648). Screenshots are from the homozygous E5 TDP43-GFP clone, at MN day 40, 20 days post-transduction with AAVs. Red arrows indicate the alternative splice event.

.

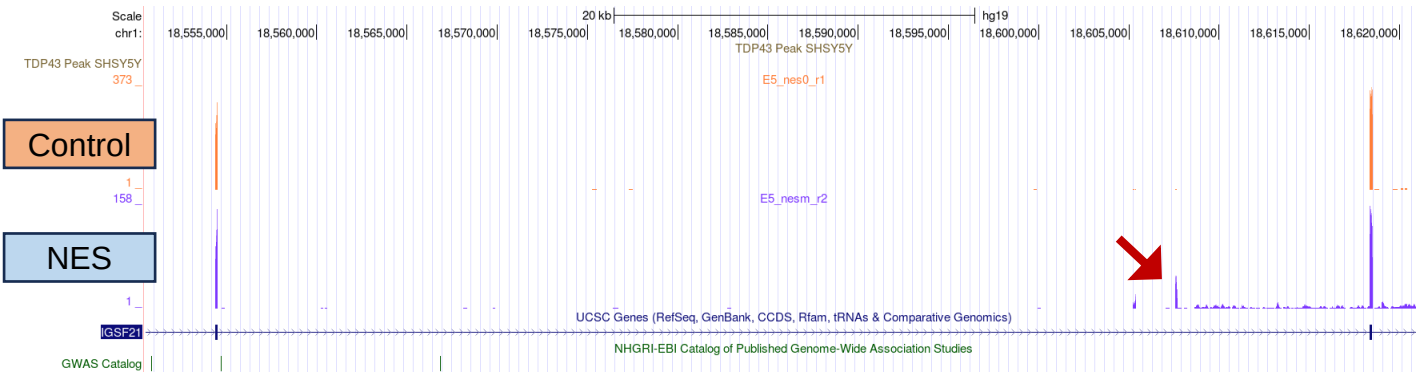
Supplementary Figure 3



CYFIP2



FEZ1



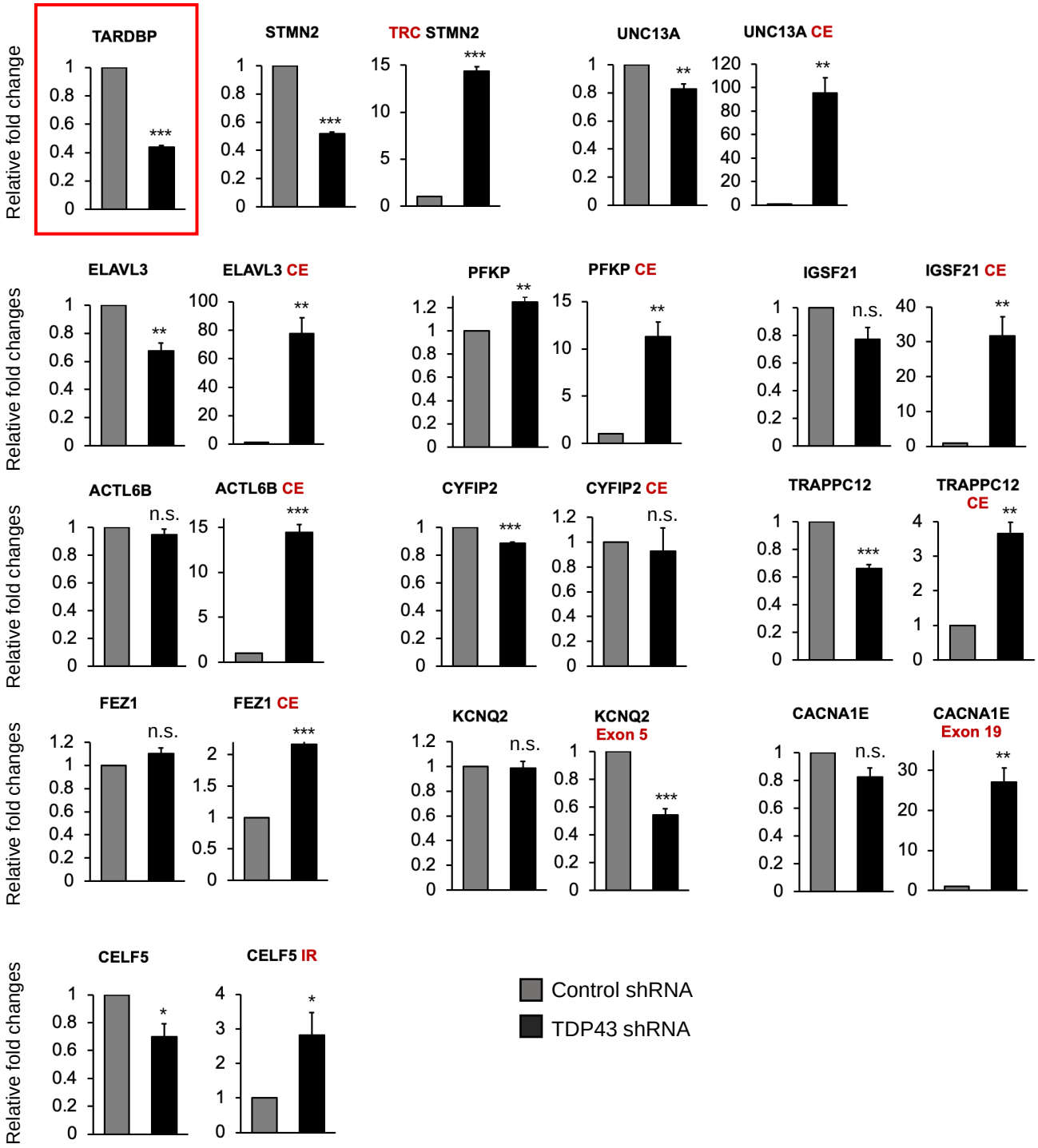
IGSF21

Supplementary Figure 3. (Related to Figure 2)

UCSC Genome browser shots depicting cryptic exon inclusion in *CYFIP2, FEZ1*, and *IGSF21* resulting from TDP43 mislocalization. There are no TDP43 binding sites identified by SH-SY5Y eCLIP data (GSE122648). Screenshots are from the homozygous E5 TDP43-GFP clone, at MN day 40, 20 days post-transduction with AAVs. Red arrows indicate the alternative splice event

.

Supplementary Figure 4



Supplementary Figure 4. (Related to Figure 2)

RT-qPCR showing alternative splicing errors in unedited iPSC MNs, 10 days post-addition of shRNAs targeting TDP43, or controls. The chart showing successful TDP43 knockdown is highlighted in red. Error bars show SEM. CE = cryptic exon, TRC = truncated, IR = intron retention. N = 3 independent differentiations.

* indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p < 0.001$.

.

Figure 3

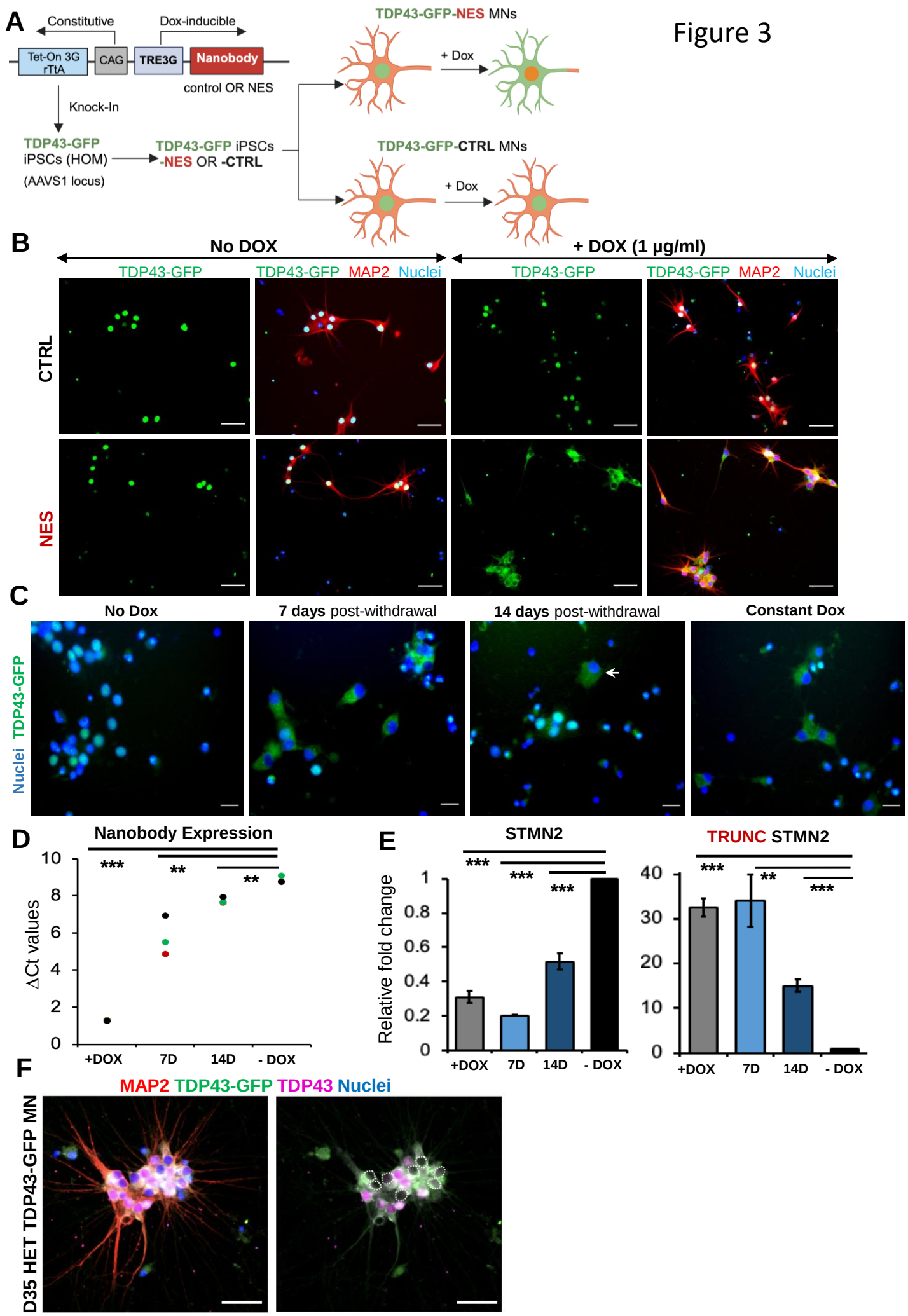

A
Constitutive
Dox-inducible
Tet-On 3G rTtA
CAG
TRE3G
Nanobody
control OR NES
Knock-In
TDP43-GFP iPSCs (HOM)
(AAVS1 locus)
TDP43-GFP iPSCs -NES OR -CTRL
TDP43-GFP-NES MNs
+ Dox
TDP43-GFP-CTRL MNs
+ Dox
B
No DOX
+ DOX (1 µg/ml)
TDP43-GFP
TDP43-GFP MAP2 Nuclei
CTRL
NES
C
No Dox
7 days post-withdrawal
14 days post-withdrawal
Constant Dox
Nuclei TDP43-GFP
D
Nanobody Expression
ΔCt values
+DOX
7D
14D
- DOX
E
STMN2
TRUNC STMN2
Relative fold change
F
MAP2 TDP43-GFP TDP43 Nuclei
D35 HET TDP43-GFP MN

Figure 3. TDP43 Sustains its own Mislocalization

(A) Schematic depicting the knock-in of the nanobody (control or NES) into the human AAVS1 safe harbour locus of our E8 homozygous TDP43-GFP cell line. Nanobody expression was under the control of a doxycycline (Dox)-inducible promoter. Addition of Dox is expected to induce TDP43 mislocalization in the TDP43-GFP-NES MNs but not in the TDP43-GFP-Control MNs.

(B) Immunofluorescent staining showing TDP43 localization in our TDP43-GFP-CTRL or -NES line in response to Dox. Mislocalization only occurs in the TDP43-GFP-NES line with Dox treatment, while all other conditions maintain nuclear TDP43. MNs were fixed and stained at day 35, 15 days post-Dox addition. Scale bar = 50 µm.

(C) TDP43 localization in the TDP43-GFP-NES cell line upon withdrawal of Dox. TDP43 fully mislocalizes following Dox treatment, which persists at 7 days post-withdrawal. After 14 days without Dox, a subset of cells show persistent TDP43 mislocalization in the cytoplasm (white arrow). Some cells show partial recovery of nuclear TDP43. All conditions (except the negative control) received a minimum of 72 hours in Dox prior to withdrawal. Scale bar = 20 µm.

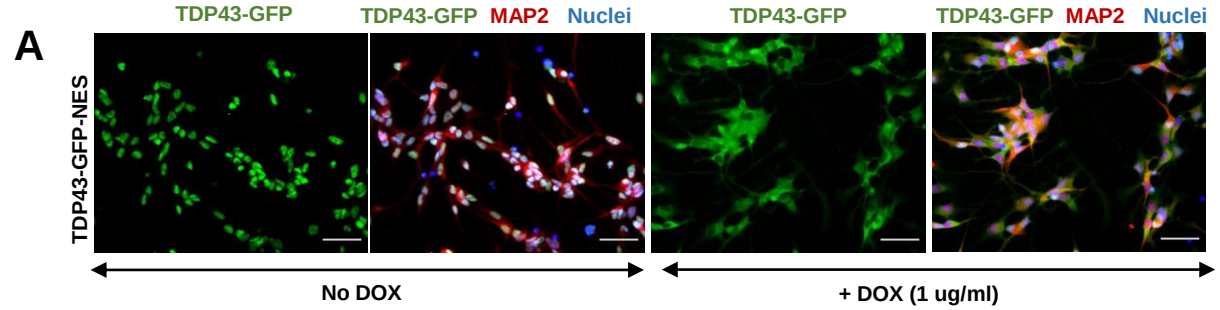
(D) RT-qPCR depicting expression of the NES nanobody. At 14 days without Dox, nanobody levels are reduced to levels similar to the No Dox condition. The three colours indicate different biological replicates. 7D = 7 days post-Dox withdrawal, 14D = 14 days post-Dox withdrawal. +DOX = continuous Dox treatment, - DOX = never received Dox

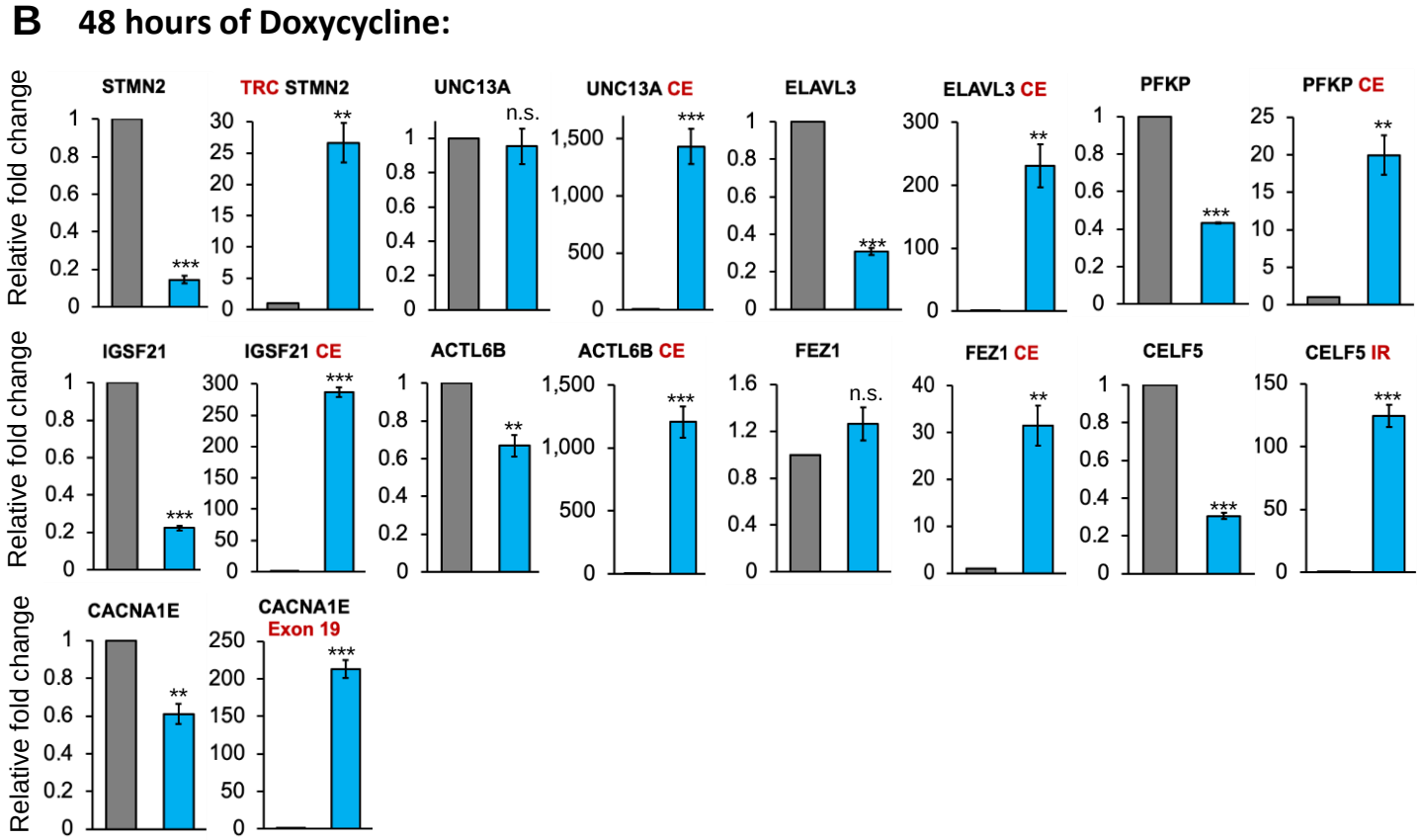
(E) RT-qPCR showing mis-splicing of STMN2 at various stages of Dox-withdrawal. Despite some nuclear TDP43 recovery at day 14, canonical STMN2 transcript expression is reduced by 50%, with associated significant expression of the truncated STMN2 transcript. TRUNC = truncated.

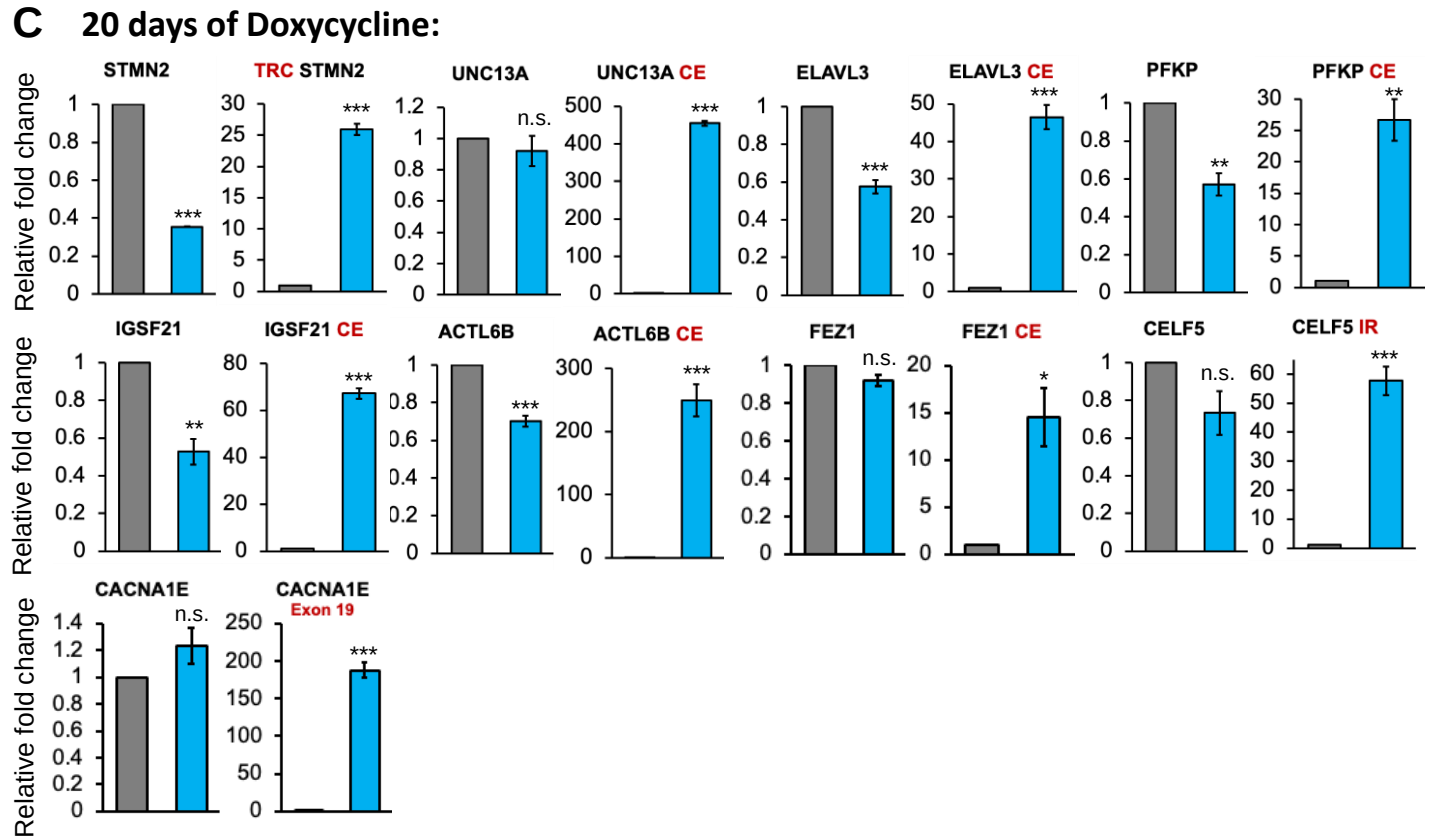
(F) TDP43 localization in heterozygous TDP43-GFP knock-in lines. MNs were transduced at day 20 and fixed for immunostaining at day 35. TDP43-GFP (green) is almost exclusively cytoplasmic, while untagged TDP43 (magenta) is found in both the nucleus and cytoplasm. Approximately 50% of cells show near total nuclear clearance of endogenous TDP43. Scale bar = 50 µm.

** indicates $p < 0.01$, *** indicates $p < 0.001$.

Supplementary Figure 5







Supplementary figure 5: (Related to Figure 3)

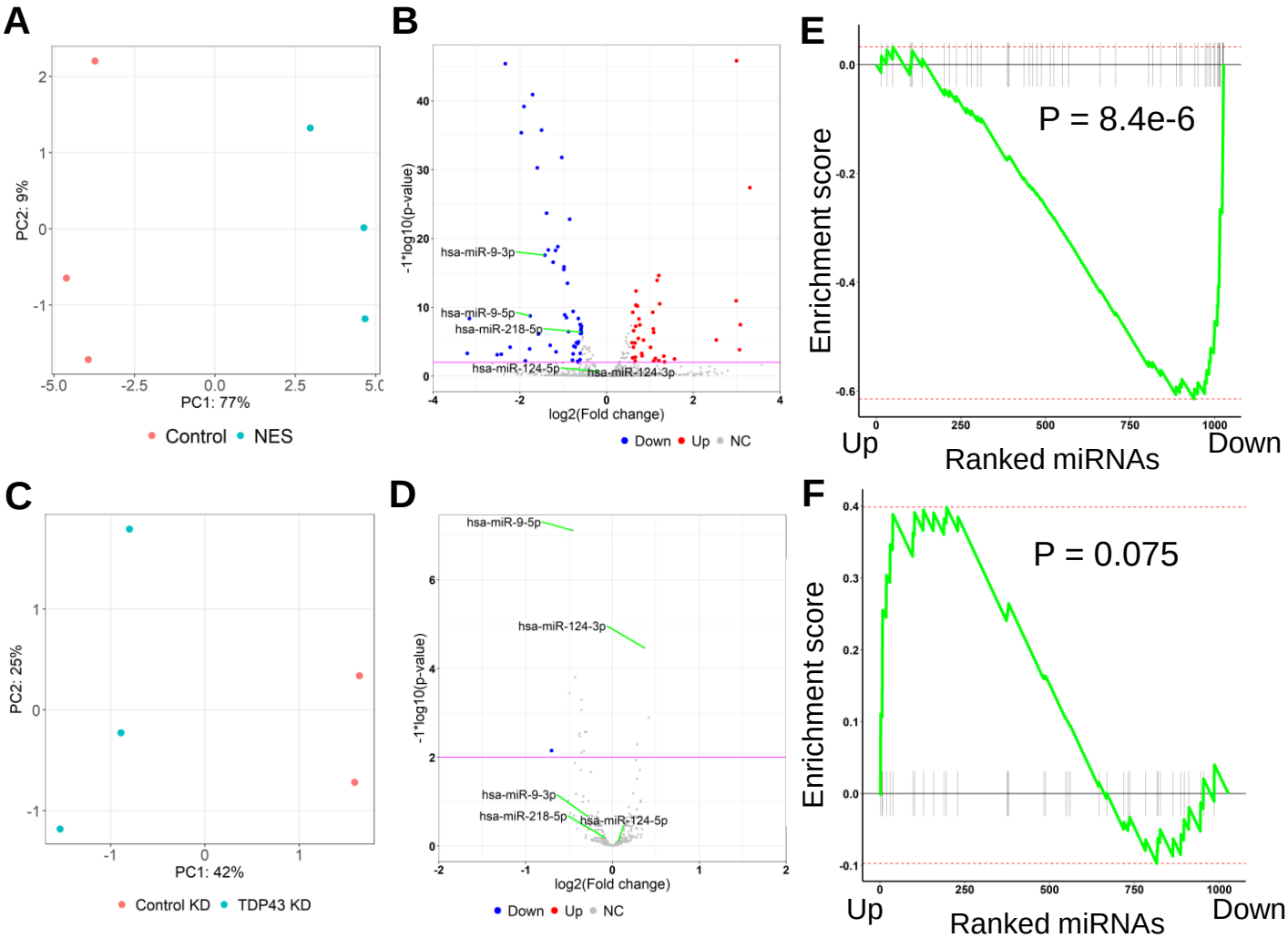
(A) Immunofluorescent staining of day 22 TDP43-GFP-NES MNs showing TDP43 mislocalization at 48 hours post-treatment with 1 µg/ml doxycycline. Scale bar = 50 µm.

(B,C) RT-qPCRs showing alternative splicing errors in Dox-inducible TDP43-GFP-NES versus -CTRL cell lines at **(B)** 48 hours and **(C)** 20 days post-treatment with 1 µg/ml doxycycline. MNs were treated with Dox at day 20 of the differentiation process. CE = cryptic exon, TRC = truncated, IR = intron retention.

* indicates $p < 0.05$, ** indicates $p < 0.01$,*** indicates $p < 0.001$.

.

Supplementary Figure 6



Supplementary figure 6: MicroRNA dysregulation in MNs due to TDP43 Mislocalization or Knockdown.

(A) Principal component analysis of microRNA (miR) counts for the TDP43-GFP-Control and TDP43-GFP-NES MNs.

(B) Volcano plot displaying miRs differentially expressed due to TDP43 mislocalization. Red: upregulated miRs, Green: downregulated miRs, NC: No Change. Horizontal line denotes a p-value threshold of 0.01.

(C) Principal component analysis of miR counts for unedited healthy iPSC MNs transduced with TDP43 or control shRNAs.

(D) Volcano plot displaying miRs differentially expressed due to TDP43 knockdown. Red: upregulated miRs, Green: downregulated miRs, NC: No Change. Horizontal line denotes a p-value threshold of 0.01.

(E, F) Gene-set enrichment analysis plots displaying overlap of differentially expressed miRs due to TDP43 mislocalization or knockdown. miRs differentially expressed in the TDP43 knockdown samples were ranked from most upregulated to the most downregulated. **E**: Gene set used for the GSEA include miRs downregulated due to TDP43 mislocalization, **F:** Gene set used for the GSEA include miRs upregulated due to TDP43 mislocalization**.**

Supplementary Table 1

Target	AS event	AS event loc	Forward sequence	Reverse sequence
UNC13A	canonical transcript		GGACGTGTGGTACAACCTGG	GTGTACTGGACATGGTACGGG
UNC13A-CE	cryptic exon	intron 19	TGGATGGAGAGATGGAACCT	GGGCTGTCTCATCGTAGTAAAC
STMN2	canonical transcript		AGCTGTCCATGCTGTCACTG	GGTGGCTTCAAGATCAGCTC
STMN2 trunc	Cryptic exon	intron 1	GGACTCGGCAGAAGACCTTC	GCAGGCTGTCTGTCTCTCTC
ELAVL3	canonical transcript		TGCAGACAAAGCCATCAACACCC	GCTGACGTACAGGTTAGCATCC
ELAVL3-CE	cryptic exon	intron 3	CCTGCTCTGAGGGATTGAGT	GTACAGGTTAGCATCCCGGA
PFKP	canonical transcript		AGGCAGTCATCGCCTTGCTAGA	ATCGCCTTCTGCACATCCTGAG
PFKP-CE	cryptic exon	intron 3	CTACCAGGGCATGGTGGA	GGAGAGTGTCTCCAGCATCC
ACTL6B	canonical transcript		CTTCCACATCGACACCAATGCC	CAGGTTTGGCTCAGACTTGACG
ACTL6B-CE	cryptic exon	intron 4	CCTGGATCACACCTACAGCA	ACCCAGGAGTTCGAAACTAGC
IGSF21	canonical transcript		GTCTGGAGGAAAACCAGCAC	TCTTGGTGTCATCCAGGTCA
IGSF21-CE	cryptic exon	intron 2	GCCTGCAGGAGGTGTTTATG	CTCTCGCTTGCGGTAGTTCT
CYFIP2	canonical transcript		ATGCCCTGGATTCTAACGGACC	CTTGGTCAGAGCATAGTAGGCG
CYFIP2-CE	cryptic exon	intron 25	CTCCAAGGAACCATTCTCCA	TGAGATTTCTTTCAGGGCTCA
FEZ1	canonical transcript		CCACTGGTGAGTCTGGATGA	TGGATCCCTCCAGTCTTCTG
FEZ1-CE	cryptic exon	intron 1	TGGAGATGTGGGATGATGG	AGGGTCGAAGGTCCTCAAAC
CELF5	canonical transcript		CACCTACTGTGCCAGGGATT	ACTGTCCGCAGGCTTCAC
CELF5-IR	intron retention	intron 5	CGTGAAGTTCTCCTCCCACA	CATCTGCACACACTCACACG
CACNA1E	canonical transcript		CCATTGTCCATCACAACCAG	CCATGCCATACATCTTCAGG
CACNA1E-E19	exon 15 inclusion	exon 19	GAGGCCTTCAACCAGAAACA	CGACATGTGGTGTCTTCTCC
KCNQ2	canonical transcript		AGTACCCCCAGACCTGGAAC	TTCCGCCTCTTCTCAAAGTG
KCNQ2-E5	exon 5 exclusion	exon 5	ACGTCTTTGCCACATCTGC	CCCCTTCTCTGCCAAGTACA
TRAPPC12	canonical transcript		TCGTGGACAAGGAGAACCTCAC	GGACTTGTCTCTTTGCTGCAGC
TRAPPC12-CE	cryptic exon	intron 7	ATGTGCAGCCCAAGTCAAG	TTGCCATGGAGTACATCACC

Supplementary table 1: Table of qPCR primers

Supplementary Table 2

Immunofluorescence antibodies:

Antibody	Species	Dilution	Source
GFP	Goat	1:1,000	NB100-1770 (Novus/Bio-techne)
MAP2	Chicken	1:10,000-16,000	GTX82661 (GeneTex)
TDP43	Rabbit	1:800	10782-2-AP (Proteintech)
Cleaved Caspase 3	Rabbit	1:400	9664L (Cell Signalling)
Neurofilament-M	mouse	1:1,000	Mab1621 (Merck)
ISLET-1	rabbit	1:500	Ab109517 (Abcam)

Western blotting Antibodies:

Antibody	Species	dilution	Source
TDP-43	Rabbit	1:1,000	10782-2-AP (ProteinTech)
Alpha-tubulin	Mouse	1:1,000	AB7291 (Abcam)
IRDye® 680RD anti-mouse	Goat	1:5,000	926-68070 (Li-Cor)
IRDye® 800CW anti-rabbit	Goat	1:5,000	926-32211 (Li-Cor)

Supplementary table 2: antibodies used for immunofluorescence and western blotting