1 **Common ALS/FTD risk variants in *UNC13A* exacerbate its cryptic splicing**
2 **and loss upon TDP-43 mislocalization**

3 **Authors:** Anna-Leigh Brown[1,†], Oscar G. Wilkins[1,2,†], Matthew J. Keuss[1,†], Sarah E. Hill[3,†],
4 Matteo Zanovello[1], Weaverly Colleen Lee[1], Flora C.Y. Lee[1,2], Laura Masino[2], Yue A. Qi[3], Sam
5 Bryce-Smith[1], Alexander Bampton[4,5], Ariana Gatt[4,5], Hemali Phatnani[6], NYGC ALS
6 Consortium[7], Giampietro Schiavo[1,8], Elizabeth M.C. Fisher[1], Towfique Raj[8,9,10,11], Maria
7 Secrier[12], Tammaryn Lashley[4,5], Jernej Ule[1,2,13], Emanuele Buratti[14], Jack Humphrey[8,9,10,11],
8 Michael E. Ward[3,*], Pietro Fratta[1,*]

9 **Affiliations:**

10 [1]Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology,
11 London, UK.

12 [2]The Francis Crick Institute, London, UK.

13 [3]National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD, USA.

14 [4]Queen Square Brain Bank, UCL Queen Square Institute of Neurology, University
15 College London, UK.

16 [5]Queen Square Department of Neurodegenerative Disease, UCL Queen Square Institute
17 of Neurology, University College London, UK.

18 [6]Center for Genomics of Neurodegenerative Disease, New York Genome Center
19 (NYGC), New York, NY, USA.

20 [7]The NYGC ALS Consortium is detailed in supplemental acknowledgments.

21 [8]Nash Family Department of Neuroscience & Friedman Brain Institute, Icahn School of
22 Medicine at Mount Sinai, New York, NY, USA.

23 [9]Ronald M. Loeb Center for Alzheimer's Disease, Icahn School of Medicine at Mount
24 Sinai, New York, NY, USA.

25 [10]Department of Genetics and Genomic Sciences & Icahn Institute for Data Science and
26 Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

27 [11]Estelle and Daniel Maggin Department of Neurology, Icahn School of Medicine at
28 Mount Sinai, New York, NY, USA.

29 [12]Department of Genetics, Evolution and Environment, UCL Genetics Institute,
30 University College London, UK.

31 [13]Department of Molecular Biology and Nanobiotechnology, National Institute of
32 Chemistry, Ljubljana, Slovenia.

33 [14]Molecular Pathology Lab, International Centre for Genetic Engineering and
34 Biotechnology (ICGEB), Trieste, Italy.

35 *Corresponding author. Email: p.fratta@ucl.ac.uk (P.F.); michael.ward4@nih.gov (M.E.W)

36 † These authors contributed equally to this work.

37

38

Abstract:

Variants within the *UNC13A* gene have long been known to increase risk of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD), two related neurodegenerative diseases defined by mislocalization of the RNA-binding protein TDP-43. Here, we show that TDP-43 depletion induces robust inclusion of a cryptic exon (CE) within *UNC13A*, a critical synaptic gene, resulting in nonsense-mediated decay and protein loss. Strikingly, two common polymorphisms strongly associated with ALS/FTD risk directly alter TDP-43 binding within the CE or downstream intron, increasing CE inclusion in cultured cells and in patient brains. Our findings, which are the first to demonstrate a genetic link specifically between loss of TDP-43 nuclear function and disease, reveal both the mechanism by which *UNC13A* variants exacerbate the effects of decreased nuclear TDP-43 function, and provide a promising therapeutic target for TDP-43 proteinopathies.

One-Sentence Summary:

Shared ALS/FTD risk variants increase the sensitivity of a cryptic exon in the synaptic gene *UNC13A* to TDP-43 depletion.

Main Text:

Introduction

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are devastating adult-onset neurodegenerative disorders with shared genetic causes and common pathological aggregates (*1–3*). Genome-wide association studies (GWASs) have repeatedly demonstrated a shared risk locus between ALS and FTD within the crucial synaptic gene *UNC13A,* although the mechanism underlying this association has remained elusive (*4*).

ALS and FTD are pathologically defined by cytoplasmic aggregation and nuclear depletion of TAR DNA-binding protein 43 (TDP-43) in the vast majority (>97%) of ALS cases and in 45% of FTD cases (FTLD-TDP) (*5*). TDP-43, an RNA-binding protein (RBP), primarily resides in the nucleus and plays key regulatory roles in RNA metabolism, including acting as a splicing repressor. Upon TDP-43 nuclear depletion – an early pathological feature in ALS/FTLD-TDP – non-conserved intronic sequences are de-repressed and erroneously included in mature RNAs. These events are referred to as cryptic exons (CEs) and can lead to premature stop-codons/polyadenylation and transcript degradation (*6, 7*). Recently, TDP-43 loss was found to induce a CE in the *Stathmin 2* (*STMN2*) transcript, which can serve as a functional readout for TDP-43 proteinopathy, as it appears selectively in affected patient tissue and its level correlates with TDP-43 phosphorylation (*8-10*).

In this study, we report a novel CE in *UNC13A* which promotes nonsense-mediated decay, and is present at remarkably high levels in patient neurons. Strikingly, we find that ALS/FTD risk-associated SNPs within *UNC13A* promote increased inclusion of this CE. We thus elucidate the molecular mechanism behind one of the top GWAS hits for ALS/FTD, and provide a promising new therapeutic target for TDP-43 proteinopathies.

Results

TDP-43 knockdown leads to inclusion of a cryptic exon in UNC13A

To discover novel CEs induced by TDP-43 depletion, we performed RNA-seq on human induced pluripotent stem cell (iPSC)-derived cortical-like i[3]Neurons in which we reduced TDP-43 expression through CRISPR inhibition (CRISPRi) (*10–13*). We identified 179 CEs, including several previously reported, such as *AGRN*, *PFKP* and *STMN2* (*6–9*) (**Fig. 1A**; data S1) (**Fig. 1B**; data S2). Interestingly, we observed robust mis-splicing in two members of the UNC13 synaptic protein family, *UNC13A* and *UNC13B* (**Fig. 1C-F***)*. Notably, *UNC13A* polymorphisms modify both disease risk and progression in ALS and FTLD-TDP (*4, 14–21*) pointing towards a potential functional relationship between TDP-43, *UNC13A*, and disease risk.

Inspection of the *UNC13A* gene revealed a previously unreported CE after TDP-43 knockdown (KD), with both a shorter and longer form, between exons 20 and 21 (**Fig. 1C**), and increased IR between exons 31 and 32 (fig. S1B). One ALS/FTLD-TDP risk SNP – *rs12973192* (*17*) – lies 16 bp inside the CE (henceforth referred to as the CE SNP). Another SNP – *rs12608932* (*4*) – is located 534 bp downstream of the donor splice site of the CE inside the same intron (henceforth referred to as the intronic SNP) (**Fig. 1D**). While there are five polymorphisms associated with ALS risk along *UNC13A*, they are all in high linkage disequilibrium (LD) in European populations with both the CE and intronic SNPs, and are present in 35% of individuals (**Fig. 1G**)(*17*). The close proximity of the disease-associated SNPs to the *UNC13A* CE suggests that the SNPs may influence *UNC13A* splicing. In *UNC13B*, TDP-43 KD led to the inclusion of an annotated frame-shift-inducing exon between exons 10 and 11, henceforth referred to as the *UNC13B* frameshift exon (fsE), and increased intron retention (IR) between exon 21 and 22 (**Fig. 1E,F**; fig. S1A).

In support of a direct role for TDP-43 regulation of *UNC13A* and *UNC13B*, we found multiple TDP-43 binding peaks both downstream and within the body of the *UNC13A* CE (**Fig. 1D**) and IR (fig. S1B) (*22*), and *UNC13A* CE inclusion negatively correlated with *TARDBP* RNA levels (rho = -0.43, p=0.077, **Fig. 1H**). Additionally, TDP-43 binding peaks were present near both splice events in *UNC13B* (**Fig. 1F**; fig. S1A) (*22*). We also detected these splicing changes in RNA-seq from TDP-43 depleted SH-SY5Y and SK-N-DZ neuronal lines, as well as in publicly available iPSC-derived motor neurons (MNs) (*8*) and SK-N-DZ datasets (*23*)(**Fig. 1I-L**; fig. S1C), and validated them by PCR in SH-SY5Y and SK-N-DZ cell lines (fig. S1D,E).

UNC13A and UNC13B RNA and protein are downregulated by TDP-43 knockdown

Next, we examined whether incorrect splicing of *UNC13A* and *UNC13B* affected transcript levels in neurons and neuron-like cells. TDP-43 KD significantly reduced *UNC13A* RNA abundance in the three cell types with the highest levels of cryptic splicing (FDR < 0.1; **Fig. 2A**, **Fig. 1I**). Likewise, *UNC13B* RNA was significantly downregulated in four datasets (FDR < 0.1) (**Fig. 2B**). We confirmed these results by qPCR in SH-SY5Y and SK-N-DZ cell lines (fig. S2A). The number of ribosome footprints aligning to *UNC13A* and *UNC13B* was reduced after TDP-43 KD (**Fig. 2C**; fig. S2B, data S3). TDP-43 KD also decreased expression of UNC13A and UNC13B at the protein level, as assessed by quantitative proteomics with liquid chromatography tandem-mass spectrometry and western blot (**Fig. 2D,E**). These data suggest that the mis-splicing in *UNC13A* and *UNC13B* after TDP-43 KD reduces their transcript and protein abundance in neurons.

3

124   The *UNC13A* CE contains a premature termination codon (PTC) and is thus predicted to
125 promote nonsense-mediated decay (NMD). Cycloheximide (CHX) treatment, which stalls
126 translation and impairs NMD, increased CE inclusion in *UNC13A* after TDP-43 KD. Conversely,
127 CHX did not alter levels of the aberrant *STMN2* transcript, which is not predicted to undergo
128 NMD (**Fig. 2F**). Taken together, our data suggests that TDP-43 is critical for maintaining normal
129 expression of the presynaptic proteins *UNC13A* and *UNC13B* by ensuring their correct pre-
130 mRNA splicing.

131 *UNC13A cryptic exon is highly expressed in TDP-43-depleted patient neurons*

132 To explore whether the *UNC13A* CE could be detected in patient tissues affected by TDP-43
133 pathology, we first analysed RNA-seq from neuronal nuclei sorted from frontal cortices of
134 ALS/FTLD patients (*24*). We compared levels of *UNC13A* CE to levels of a CE in *STMN2*
135 known to be regulated by TDP-43. Both *STMN2* and *UNC13A* CEs were exclusive to TDP-43-
136 depleted nuclei, and, strikingly, in some cases the *UNC13A* CE percent spliced in (PSI) reached
137 100% (**Fig. 3A**). This suggests that in patients there will be a significant loss of UNC13A
138 expression within the subpopulation of neurons with TDP-43 pathology.
139   Next, we quantified *UNC13A* CE inclusion in bulk RNA-seq from the NYGC ALS
140 Consortium, a dataset containing 1,349 brain and spinal cord tissues from a total of 377 ALS,
141 FTLD, and control individuals. The *UNC13A* CE was detected exclusively in FTLD-TDP and
142 ALS-TDP cases (89% and 38% respectively), with no detection in ALS-non-TDP (*SOD1* and
143 *FUS* mutations), FTLD-non-TDP (FTLD-TAU and FTLD-FUS), or control cases. The lower
144 detection rate in ALS versus FTLD is likely due to the lower expression of *UNC13A* in the spinal
145 cord (fig. S3A). Thus, pathological *UNC13A* CEs occur *in vivo* and are specific to
146 neurodegenerative disease subtypes in which mislocalization and nuclear depletion of TDP-43
147 occurs.
148   *UNC13A* CE expression mirrored the known tissue distribution of TDP-43 aggregation
149 and nuclear clearance (*25*): it was specific to ALS-TDP spinal cord and motor cortex, as well as
150 FTLD-TDP frontal and temporal cortices, but absent from the cerebellum in disease and control
151 states (**Fig. 3B**). Despite the CE PSI being diluted by both the presence of unaffected cells and
152 NMD in bulk RNA-seq, we were still able to detect CE above 20% in some samples.
153 Furthermore, although, unlike the STMN2 CE, the *UNC13A* CE induces NMD, it was detected at
154 similar levels to *STMN2* CE in cortical regions, whilst *STMN2* CE was more abundant in the
155 spinal cord (**Fig. 3C**). We next investigated whether *UNC13A* CEs could be visualised by *in situ*
156 hybridisation (ISH) in FTLD patient brains. Using a probe targeting the *UNC13A* CE on frozen
157 frontal cortex tissue, we detected staining significantly above background in 4 out of 5 tested
158 FTLD-TDP cases, but in none of the FTLD-Tau (n=3) or control (n=5) cases (**Fig. 3D**; fig. S3B).
159   To assess if *UNC13A* CE levels in bulk tissue was related to the level of TDP-43
160 proteinopathy, we used *STMN2* CE PSI as a proxy, as it correlates with the burden of
161 phosphorylated TDP-43 in patient samples (*10*). As expected, across the NYGC ALS
162 Consortium samples we observed a significant positive correlation between the level of *STMN2*
163 CE PSI and *UNC13A* CE PSI (rho = 0.55, p = 3.0e-4) (**Fig. 3E**). Combined, our analysis reveals
164 a strong relationship between TDP-43 pathology and *UNC13A* CE levels, supporting a model for
165 direct regulation of *UNC13A* mRNA splicing by TDP-43 in patients.

rs12973192(G) and rs12608932(C) combine to promote cryptic splicing

To test whether the ALS/FTD *UNC13A* risk SNPs promote cryptic splicing, which could explain their link to disease, we assessed *UNC13A* CE levels across different genotypes, and found significantly increased levels in cases homozygous for CE *rs12973192(G)* and intronic *rs12608932(C)* SNPs (fig. S4A-B). To ensure that this was not simply due to more severe TDP-43 pathology in these samples, we normalised by the level of *STMN2* cryptic splicing, and again found a significantly increased level of the *UNC13A* CE in cases with homozygous risk variants (Wilcoxon test, $p < 0.001$) (**Fig. 4A**; fig. S4C,D). Next, we performed targeted RNA-seq on *UNC13A* CE from temporal cortices of ten heterozygous risk allele cases and four controls. We detected significant biases towards reads containing the risk allele ($p < 0.05$, single-tailed binomial test) in six samples, with a seventh sample approaching significance (**Fig. 4B**), suggesting that the two ALS/FTLD-linked variants promote cryptic splicing *in vivo*.

To specifically examine whether the CE or the intronic SNP of *UNC13A* promote CE splicing, we generated four variants of minigenes containing *UNC13A* exon 20, intron 20, and exon 21, featuring both risk alleles (2R), both non-risk alleles (2H), the risk allele within the CE (*rs12973192*) (RE), or the risk allele in the intron (*rs12608932*) (RI) (**Fig. 4C**). We then expressed these minigenes in SH-SY5Y cells with doxycycline-inducible TDP-43 knockdown. We found that both the CE SNP and, to a lesser extent, intronic SNP independently promoted CE inclusion, with the greatest overall levels detected for the 2R minigene (**Fig. 4D,E**).

To explore how these two SNPs might act to enhance CE splicing, we analyzed a dataset of *in vitro* RNA heptamer/RBP binding enrichments, and examined the effect of the SNPs on relative RBP enrichment (*26*). Strikingly, when investigating which RBPs were most impacted in their RNA binding enrichment by the CE-risk SNP, TDP-43 had the third largest decrease of any RBP, with only two non-human RBPs showing a larger decrease (**Fig. 4F,G**; fig. S4E,F). To test whether the CE SNP directly inhibited *in vitro* TDP-43 binding, we performed isothermal titration calorimetry using recombinant TDP-43 and 14-nt RNAs. As predicted, we observed an increased Kd for RNA containing the CE risk SNP (**Fig. 4H;** fig. S4G,H; Data S4). Together these data predict that the *UNC13A* CE SNP may directly inhibit TDP-43 binding.

To directly study the impact of the SNPs on TDP-43 binding to *UNC13A* pre-mRNA, we performed TDP-43 iCLIP with cells expressing either the 2R or 2H minigene. We observed a striking enrichment of crosslinks within the ~800nt UG-rich region containing both SNPs in intron 20 (**Fig. 4I**). When comparing the 2R with the 2H minigene, the peaks with the largest fractional changes were in close proximity of each SNP; similarly, we detected a 21% decrease in total TDP-43 crosslinks centred around the CE SNP and a 73% increase upstream of the intronic SNP (**Fig. 4I, J**, fig. S4I; 50 nucleotide windows). These data demonstrate that these two disease-risk SNPs distort the pattern of TDP-43/RNA interactions, decreasing TDP-43 binding near the CE donor splice site, thus exacerbating *UNC13A* CE inclusion upon nuclear TDP-43 depletion.

Discussion

Our results support a model wherein TDP-43 nuclear depletion and the intronic and CE SNPs in *UNC13A* synergistically reduce expression of *UNC13A*, a gene that is critical for normal neuronal function. In this model, when nuclear TDP-43 levels are normal in healthy individuals, TDP-43 efficiently binds to *UNC13A* pre-mRNA and prevents CE splicing, regardless of

210   *UNC13A* SNPs. Conversely, severe nuclear depletion of TDP-43 in end-stage disease induces
211   CE inclusion in all cases. However, the common intronic and CE SNPs in *UNC13A* alter TDP-43
212   binding to *UNC13A* pre-mRNA and may make *UNC13A* CE more sensitive to partial TDP-43
213   loss that occurs early in degenerating neurons, explaining their associated risk effect. Strikingly,
214   we found that both risk alleles for these SNPs independently and additively promoted cryptic
215   splicing *in vitro*. Intriguingly, when the two variants are not co-inherited, as seen in East Asian
216   individuals with ALS, an attenuated effect is observed (*20*). A similar phenomenon wherein SNP
217   pairs both contribute to risk has been widely studied at the APOE locus in Alzheimer's disease
218   (*27*).

219   UNC13-family proteins are highly conserved across metazoans and are essential for
220   calcium-triggered synaptic vesicle release (*28*). In mice, double knockout of *UNC13A* and
221   *UNC13B* inhibits both excitatory and inhibitory synaptic transmission in hippocampal neurons
222   and greatly impairs transmission at neuromuscular junctions (*29, 30*). In TDP-43-negative
223   neuronal nuclei derived from patients, the *UNC13A* CE is present in up to 100% of transcripts,
224   suggesting that expression of functional *UNC13A* is greatly reduced, which could impact normal
225   synaptic transmission.

226   TDP-43 loss induces hundreds of splicing changes, a number of which have also been
227   detected in patient brains. However, it has remained unclear whether these events – even those
228   that occur in crucial neuronal genes – contribute to disease pathogenesis. That genetic variation
229   influencing the *UNC13A* CE inclusion can lead to changes in ALS/FTD susceptibility and
230   progression strongly supports *UNC13A* downregulation to be one of the critical consequences of
231   TDP-43 loss of function. Excitingly, *UNC13A* provides a generalizable therapeutic target for
232   97% of ALS and approximately half of FTD cases. These findings are also of interest to other
233   neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease and chronic
234   traumatic encephalopathy, in which TDP-43 depletion is also observed in a significant fraction of
235   cases.

236

237

238   **References**

239   1. R. Ferrari, D. Kapogiannis, E. D. Huey, P. Momeni, FTD and ALS: a tale of two diseases.
240   *Curr. Alzheimer Res.* **8**, 273–294 (2011).
241   2. P. Couratier, P. Corcia, G. Lautrette, M. Nicol, B. Marin, ALS and frontotemporal dementia
242   belong to a common disease spectrum. *Rev. Neurol. (Paris).* **173**, 273–279 (2017).
243   3. A.-L. Ji, X. Zhang, W.-W. Chen, W.-J. Huang, Genetics insight into the amyotrophic lateral
244   sclerosis/frontotemporal dementia spectrum. *J. Med. Genet.* **54**, 145–154 (2017).
245   4. M. A. van Es, J. H. Veldink, C. G. J. Saris, H. M. Blauw, P. W. J. van Vught, A. Birve, R.
246   Lemmens, H. J. Schelhaas, E. J. N. Groen, M. H. B. Huisman, A. J. van der Kooi, M. de
247   Visser, C. Dahlberg, K. Estrada, F. Rivadeneira, A. Hofman, M. J. Zwarts, P. T. C. van
248   Doormaal, D. Rujescu, E. Strengman, I. Giegling, P. Muglia, B. Tomik, A. Slowik, A. G.
249   Uitterlinden, C. Hendrich, S. Waibel, T. Meyer, A. C. Ludolph, J. D. Glass, S. Purcell, S.
250   Cichon, M. M. Nöthen, H.-E. Wichmann, S. Schreiber, S. H. H. M. Vermeulen, L. A.
251   Kiemeney, J. H. J. Wokke, S. Cronin, R. L. McLaughlin, O. Hardiman, K. Fumoto, R. J.
252   Pasterkamp, V. Meininger, J. Melki, P. N. Leigh, C. E. Shaw, J. E. Landers, A. Al-Chalabi,
253   R. H. Brown, W. Robberecht, P. M. Andersen, R. A. Ophoff, L. H. van den Berg, Genome-
254   wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for

255   sporadic amyotrophic lateral sclerosis. *Nat. Genet.* **41**, 1083–1087 (2009).

256 5. M. Neumann, D. M. Sampathu, L. K. Kwong, A. C. Truax, M. C. Micsenyi, T. T. Chou, J.
257   Bruce, T. Schuck, M. Grossman, C. M. Clark, L. F. McCluskey, B. L. Miller, E. Masliah, I.
258   R. Mackenzie, H. Feldman, W. Feiden, H. A. Kretzschmar, J. Q. Trojanowski, V. M.-Y. Lee,
259   Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral
260   sclerosis. *Science*. **314**, 130–133 (2006).

261 6. J. Humphrey, W. Emmett, P. Fratta, A. M. Isaacs, V. Plagnol, Quantitative analysis of
262   cryptic splicing associated with TDP-43 depletion. *BMC Med. Genomics*. **10**, 38 (2017).

263 7. J. P. Ling, O. Pletnikova, J. C. Troncoso, P. C. Wong, TDP-43 repression of nonconserved
264   cryptic exons is compromised in ALS-FTD. *Science*. **349**, 650–655 (2015).

265 8. J. R. Klim, L. A. Williams, F. Limone, I. Guerra San Juan, B. N. Davis-Dusenbery, D. A.
266   Mordes, A. Burberry, M. J. Steinbaugh, K. K. Gamage, R. Kirchner, R. Moccia, S. H.
267   Cassel, K. Chen, B. J. Wainger, C. J. Woolf, K. Eggan, ALS-implicated protein TDP-43
268   sustains levels of STMN2, a mediator of motor neuron growth and repair. *Nat. Neurosci.* **22**,
269   167–179 (2019).

270 9. Z. Melamed, J. Lopez-Erauskin, M. W. Baughn, O. Zhang, K. Drenner, Y. Sun, F.
271   Freyermuth, M. A. McMahon, M. S. Beccari, J. Artates, T. Ohkubo, M. Rodriguez, N. Lin,
272   D. Wu, C. F. Bennett, F. Rigo, S. Da Cruz, J. Ravits, C. Lagier-Tourenne, D. W. Cleveland,
273   Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent
274   neurodegeneration. *Nat. Neurosci.* **22**, 180–190 (2019).

275 10. M. Prudencio, J. Humphrey, S. Pickles, A.-L. Brown, S. E. Hill, J. Kachergus, J. Shi, M.
276   Heckman, M. Spiegel, C. Cook, Truncated stathmin-2 is a marker of TDP-43 pathology in
277   frontotemporal dementia. *J. Clin. Invest.* (2020).

278 11. M. S. Fernandopulle, R. Prestil, C. Grunseich, C. Wang, L. Gan, M. E. Ward, Transcription
279   Factor–Mediated Differentiation of Human iPSCs into Neurons. *Curr. Protoc. Cell Biol.* **79**,
280   e51 (2018).

281 12. R. Tian, M. A. Gachechiladze, C. H. Ludwig, M. T. Laurie, J. Y. Hong, D. Nathaniel, A. V.
282   Prabhu, M. S. Fernandopulle, R. Patel, M. Abshari, M. E. Ward, M. Kampmann, CRISPR
283   Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived
284   Neurons. *Neuron*. **104**, 239-255.e12 (2019).

285 13. C. Wang, M. E. Ward, R. Chen, K. Liu, T. E. Tracy, X. Chen, M. Xie, P. D. Sohn, C.
286   Ludwig, A. Meyer-Franke, C. M. Karch, S. Ding, L. Gan, Scalable Production of iPSC-
287   Derived Human Neurons to Identify Tau-Lowering Compounds by High-Content Screening.
288   *Stem Cell Rep.* **9**, 1221–1233 (2017).

289 14. F. P. Diekstra, P. W. J. van Vught, W. van Rheenen, M. Koppers, R. J. Pasterkamp, M. A.
290   van Es, H. J. Schelhaas, M. de Visser, W. Robberecht, P. Van Damme, P. M. Andersen, L.
291   H. van den Berg, J. H. Veldink, *Neurobiol. Aging*, in press,
292   doi:10.1016/j.neurobiolaging.2011.10.029.

293 15. F. P. Diekstra, V. M. Van Deerlin, J. C. van Swieten, A. Al-Chalabi, A. C. Ludolph, J. H.
294   Weishaupt, O. Hardiman, J. E. Landers, R. H. Brown, M. A. van Es, R. J. Pasterkamp, M.
295   Koppers, P. M. Andersen, K. Estrada, F. Rivadeneira, A. Hofman, A. G. Uitterlinden, P. van
296   Damme, J. Melki, V. Meininger, A. Shatunov, C. E. Shaw, P. N. Leigh, P. J. Shaw, K. E.
297   Morrison, I. Fogh, A. Chiò, B. J. Traynor, D. Czell, M. Weber, P. Heutink, P. I. W. de
298   Bakker, V. Silani, W. Robberecht, L. H. van den Berg, J. H. Veldink, C9orf72 and UNC13A
299   are shared risk loci for ALS and FTD: a genome-wide meta-analysis. *Ann. Neurol.* **76**, 120–
300   133 (2014).

301   16. B. Gaastra, A. Shatunov, S. Pulit, A. R. Jones, W. Sproviero, A. Gillett, Z. Chen, J. Kirby, I.
302   Fogh, J. F. Powell, P. N. Leigh, K. E. Morrison, P. J. Shaw, C. E. Shaw, L. H. van den Berg,
303   J. H. Veldink, C. M. Lewis, A. Al-Chalabi, Rare genetic variation in UNC13A may modify
304   survival in amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler. Front. Degener.* **17**,
305   593–599 (2016).
306   17. A. Nicolas, K. P. Kenna, A. E. Renton, N. Ticozzi, F. Faghri, R. Chia, J. A. Dominov, B. J.
307   Kenna, M. A. Nalls, P. Keagle, A. M. Rivera, W. van Rheenen, N. A. Murphy, J. J. F. A. van
308   Vugt, J. T. Geiger, R. A. Van der Spek, H. A. Pliner, Shankaracharya, B. N. Smith, G.
309   Marangi, S. D. Topp, Y. Abramzon, A. S. Gkazi, J. D. Eicher, A. Kenna, G. Mora, A. Calvo,
310   L. Mazzini, N. Riva, J. Mandrioli, C. Caponnetto, S. Battistini, P. Volanti, V. La Bella, F. L.
311   Conforti, G. Borghero, S. Messina, I. L. Simone, F. Trojsi, F. Salvi, F. O. Logullo, S.
312   D'Alfonso, L. Corrado, M. Capasso, L. Ferrucci, C. de A. M. Moreno, S. Kamalakaran, D.
313   B. Goldstein, A. D. Gitler, T. Harris, R. M. Myers, H. Phatnani, R. L. Musunuri, U. S. Evani,
314   A. Abhyankar, M. C. Zody, J. Kaye, S. Finkbeiner, S. K. Wyman, A. LeNail, L. Lima, E.
315   Fraenkel, C. N. Svendsen, L. M. Thompson, J. E. Van Eyk, J. D. Berry, T. M. Miller, S. J.
316   Kolb, M. Cudkowicz, E. Baxi, M. Benatar, J. P. Taylor, E. Rampersaud, G. Wu, J. Wuu, G.
317   Lauria, F. Verde, I. Fogh, C. Tiloca, G. P. Comi, G. Sorarù, C. Cereda, P. Corcia, H.
318   Laaksovirta, L. Myllykangas, L. Jansson, M. Valori, J. Ealing, H. Hamdalla, S. Rollinson, S.
319   Pickering-Brown, R. W. Orrell, K. C. Sidle, A. Malaspina, J. Hardy, A. B. Singleton, J. O.
320   Johnson, S. Arepalli, P. C. Sapp, D. McKenna-Yasek, M. Polak, S. Asress, S. Al-Sarraj, A.
321   King, C. Troakes, C. Vance, J. de Belleroche, F. Baas, A. L. M. A. ten Asbroek, J. L.
322   Muñoz-Blanco, D. G. Hernandez, J. Ding, J. R. Gibbs, S. W. Scholz, M. K. Floeter, R. H.
323   Campbell, F. Landi, R. Bowser, S. M. Pulst, J. M. Ravits, D. J. L. MacGowan, J. Kirby, E. P.
324   Pioro, R. Pamphlett, J. Broach, G. Gerhard, T. L. Dunckley, C. B. Brady, N. W. Kowall, J.
325   C. Troncoso, I. Le Ber, K. Mouzat, S. Lumbroso, T. D. Heiman-Patterson, F. Kamel, L. Van
326   Den Bosch, R. H. Baloh, T. M. Strom, T. Meitinger, A. Shatunov, K. R. Van Eijk, M. de
327   Carvalho, M. Kooyman, B. Middelkoop, M. Moisse, R. L. McLaughlin, M. A. Van Es, M.
328   Weber, K. B. Boylan, M. Van Blitterswijk, R. Rademakers, K. E. Morrison, A. N. Basak, J.
329   S. Mora, V. E. Drory, P. J. Shaw, M. R. Turner, K. Talbot, O. Hardiman, K. L. Williams, J.
330   A. Fifita, G. A. Nicholson, I. P. Blair, G. A. Rouleau, J. Esteban-Pérez, A. García-Redondo,
331   A. Al-Chalabi, E. Rogaeva, L. Zinman, L. W. Ostrow, N. J. Maragakis, J. D. Rothstein, Z.
332   Simmons, J. Cooper-Knock, A. Brice, S. A. Goutman, E. L. Feldman, S. B. Gibson, F.
333   Taroni, A. Ratti, C. Gellera, P. Van Damme, W. Robberecht, P. Fratta, M. Sabatelli, C.
334   Lunetta, A. C. Ludolph, P. M. Andersen, J. H. Weishaupt, W. Camu, J. Q. Trojanowski, V.
335   M. Van Deerlin, R. H. Brown, L. H. van den Berg, J. H. Veldink, M. B. Harms, J. D. Glass,
336   D. J. Stone, P. Tienari, V. Silani, A. Chiò, C. E. Shaw, B. J. Traynor, J. E. Landers, F. O.
337   Logullo, I. Simone, G. Logroscino, F. Salvi, I. Bartolomei, G. Borghero, M. R. Murru, E.
338   Costantino, C. Pani, R. Puddu, C. Caredda, V. Piras, S. Tranquilli, S. Cuccu, D. Corongiu,
339   M. Melis, A. Milia, F. Marrosu, M. G. Marrosu, G. Floris, A. Cannas, S. Tranquilli, M.
340   Capasso, C. Caponnetto, G. Mancardi, P. Origone, P. Mandich, F. L. Conforti, S. Cavallaro,
341   G. Mora, K. Marinou, R. Sideri, S. Penco, L. Mosca, C. Lunetta, G. L. Pinter, M. Corbo, N.
342   Riva, P. Carrera, P. Volanti, J. Mandrioli, N. Fini, A. Fasano, L. Tremolizzo, A. Arosio, C.
343   Ferrarese, F. Trojsi, G. Tedeschi, M. R. Monsurrò, G. Piccirillo, C. Femiano, A. Ticca, E.
344   Ortu, V. La Bella, R. Spataro, T. Colletti, M. Sabatelli, M. Zollino, A. Conte, M. Luigetti, S.
345   Lattante, G. Marangi, M. Santarelli, A. Petrucci, M. Pugliatti, A. Pirisi, L. D. Parish, P.
346   Occhineri, F. Giannini, S. Battistini, C. Ricci, M. Benigni, T. B. Cau, D. Loi, A. Calvo, C.

Moglia, M. Brunetti, M. Barberis, G. Restagno, F. Casale, G. Marrali, G. Fuda, I. Ossola, S. Cammarosano, A. Canosa, A. Ilardi, U. Manera, M. Grassano, R. Tanel, F. Pisano, M. B. Harms, D. B. Goldstein, N. A. Shneider, S. Goutman, Z. Simmons, T. M. Miller, S. Chandran, S. Pal, G. Manousakis, S. H. Appel, E. Simpson, L. Wang, R. H. Baloh, S. Gibson, R. Bedlack, D. Lacomis, D. Sareen, A. Sherman, L. Bruijn, M. Penny, A. S. Allen, S. Appel, R. H. Baloh, R. S. Bedlack, B. E. Boone, R. Brown, J. P. Carulli, A. Chesi, W. K. Chung, E. T. Cirulli, G. M. Cooper, J. Couthouis, A. G. Day-Williams, P. A. Dion, S. Gibson, A. D. Gitler, J. D. Glass, D. B. Goldstein, Y. Han, M. B. Harms, T. Harris, S. D. Hayes, A. L. Jones, J. Keebler, B. J. Krueger, B. N. Lasseigne, S. E. Levy, Y.-F. Lu, T. Maniatis, D. McKenna-Yasek, T. M. Miller, R. M. Myers, S. Petrovski, S. M. Pulst, A. R. Raphael, J. M. Ravits, Z. Ren, G. A. Rouleau, P. C. Sapp, N. A. Shneider, E. Simpson, K. B. Sims, J. F. Staropoli, L. L. Waite, Q. Wang, J. R. Wimbish, W. W. Xin, H. Phatnani, J. Kwan, D. Sareen, J. R. Broach, Z. Simmons, X. Arcila-Londono, E. B. Lee, V. M. Van Deerlin, N. A. Shneider, E. Fraenkel, L. W. Ostrow, F. Baas, N. Zaitlen, J. D. Berry, A. Malaspina, P. Fratta, G. A. Cox, L. M. Thompson, S. Finkbeiner, E. Dardiotis, T. M. Miller, S. Chandran, S. Pal, E. Hornstein, D. J. MacGowan, T. Heiman-Patterson, M. G. Hammell, N. A. Patsopoulos, J. Dubnau, A. Nath, J. Kaye, S. Finkbeiner, S. Wyman, A. LeNail, L. Lima, E. Fraenkel, J. D. Rothstein, C. N. Svendsen, L. M. Thompson, J. Van Eyk, N. J. Maragakis, J. D. Berry, J. D. Glass, T. M. Miller, S. J. Kolb, R. H. Baloh, M. Cudkowicz, E. Baxi, M. Benatar, J. P. Taylor, G. Wu, E. Rampersaud, J. Wuu, R. Rademakers, S. Züchner, R. Schule, J. McCauley, S. Hussain, A. Cooley, M. Wallace, C. Clayman, R. Barohn, J. Statland, J. Ravits, A. Swenson, C. Jackson, J. Trivedi, S. Khan, J. Katz, L. Jenkins, T. Burns, K. Gwathmey, J. Caress, C. McMillan, L. Elman, E. Pioro, J. Heckmann, Y. So, D. Walk, S. Maiser, J. Zhang, V. Silani, N. Ticozzi, C. Gellera, A. Ratti, F. Taroni, G. Lauria, F. Verde, I. Fogh, C. Tiloca, G. P. Comi, G. Sorarù, C. Cereda, S. D'Alfonso, L. Corrado, F. De Marchi, S. Corti, M. Ceroni, L. Mazzini, G. Siciliano, M. Filosto, M. Inghilleri, S. Peverelli, C. Colombrita, B. Poletti, L. Maderna, R. Del Bo, S. Gagliardi, G. Querin, C. Bertolin, V. Pensato, B. Castellotti, W. Camu, K. Mouzat, S. Lumbroso, P. Corcia, V. Meininger, G. Besson, E. Lagrange, P. Clavelou, N. Guy, P. Couratier, P. Vourch, V. Danel, E. Bernard, G. Lemasson, A. Al Kheifat, A. Al-Chalabi, P. Andersen, A. N. Basak, I. P. Blair, A. Chio, J. Cooper-Knock, P. Corcia, P. Couratier, M. de Carvalho, A. Dekker, V. Drory, A. G. Redondo, M. Gotkine, O. Hardiman, W. Hide, A. Iacoangeli, J. Glass, K. Kenna, M. Kiernan, M. Kooyman, J. Landers, R. McLaughlin, B. Middelkoop, J. Mill, M. M. Neto, M. Moisse, J. M. Pardina, K. Morrison, S. Newhouse, S. Pinto, S. Pulit, W. Robberecht, A. Shatunov, P. Shaw, C. Shaw, V. Silani, W. Sproviero, G. Tazelaar, N. Ticozzi, P. van Damme, L. van den Berg, R. van der Spek, K. van Eijk, M. van Es, W. van Rheenen, J. van Vugt, J. Veldink, M. Weber, K. L. Williams, M. Zatz, D. C. Bauer, N. A. Twine, Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. *Neuron*. **97**, 1268-1283.e6 (2018).

18. K. Placek, G. M. Baer, L. Elman, L. McCluskey, L. Hennessy, P. M. Ferraro, E. B. Lee, V. M.-Y. Lee, J. Q. Trojanowski, V. M. Van Deerlin, M. Grossman, D. J. Irwin, C. T. McMillan, UNC13A polymorphism contributes to frontotemporal disease in sporadic amyotrophic lateral sclerosis. *Neurobiol. Aging*. **73**, 190–199 (2019).

19. C. Pottier, Y. Ren, R. B. Perkerson, M. Baker, G. D. Jenkins, M. van Blitterswijk, M. DeJesus-Hernandez, J. G. J. van Rooij, M. E. Murray, E. Christopher, S. K. McDonnell, Z. Fogarty, A. Batzler, S. Tian, C. T. Vicente, B. Matchett, A. M. Karydas, G.-Y. R. Hsiung, H. Seelaar, M. O. Mol, E. C. Finger, C. Graff, L. Öijerstedt, M. Neumann, P. Heutink, M.

Synofzik, C. Wilke, J. Prudlo, P. Rizzu, J. Simon-Sanchez, D. Edbauer, S. Roeber, J. Diehl-Schmid, B. M. Evers, A. King, M. M. Mesulam, S. Weintraub, C. Geula, K. F. Bieniek, L. Petrucelli, G. L. Ahern, E. M. Reiman, B. K. Woodruff, R. J. Caselli, E. D. Huey, M. R. Farlow, J. Grafman, S. Mead, L. T. Grinberg, S. Spina, M. Grossman, D. J. Irwin, E. B. Lee, E. Suh, J. Snowden, D. Mann, N. Ertekin-Taner, R. J. Uitti, Z. K. Wszolek, K. A. Josephs, J. E. Parisi, D. S. Knopman, R. C. Petersen, J. R. Hodges, O. Piguet, E. G. Geier, J. S. Yokoyama, R. A. Rissman, E. Rogaeva, J. Keith, L. Zinman, M. C. Tartaglia, N. J. Cairns, C. Cruchaga, B. Ghetti, J. Kofler, O. L. Lopez, T. G. Beach, T. Arzberger, J. Herms, L. S. Honig, J. P. Vonsattel, G. M. Halliday, J. B. Kwok, C. L. White, M. Gearing, J. Glass, S. Rollinson, S. Pickering-Brown, J. D. Rohrer, J. Q. Trojanowski, V. Van Deerlin, E. H. Bigio, C. Troakes, S. Al-Sarraj, Y. Asmann, B. L. Miller, N. R. Graff-Radford, B. F. Boeve, W. W. Seeley, I. R. A. Mackenzie, J. C. van Swieten, D. W. Dickson, J. M. Biernacka, R. Rademakers, Genome-wide analyses as part of the international FTLD-TDP whole-genome sequencing consortium reveals novel disease risk factors and increases support for immune dysfunction in FTLD. *Acta Neuropathol. (Berl.)*. **137**, 879–899 (2019).

20. B. Yang, H. Jiang, F. Wang, S. Li, C. Wu, J. Bao, Y. Zhu, Z. Xu, B. Liu, H. Ren, X. Yang, UNC13A variant rs12608932 is associated with increased risk of amyotrophic lateral sclerosis and reduced patient survival: a meta-analysis. *Neurol. Sci.* **40**, 2293–2302 (2019).

21. R. P. A. van Eijk, M. J. C. Eijkemans, S. Nikolakopoulos, M. D. Jansen, H.-J. Westeneng, K. R. van Eijk, R. A. A. van der Spek, J. J. F. A. van Vugt, S. Piepers, G.-J. Groeneveld, J. H. Veldink, L. H. van den Berg, M. A. van Es, Pharmacogenetic interactions in amyotrophic lateral sclerosis: a step closer to a cure? *Pharmacogenomics J.* **20**, 220–226 (2020).

22. J. R. Tollervey, T. Curk, B. Rogelj, M. Briese, M. Cereda, M. Kayikci, T. Hortobágyi, A. L. Nishimura, V. Župunski, R. Patani, S. Chandran, G. Rot, B. Zupan, C. E. Shaw, J. Ule, Characterising the RNA targets and position-dependent splicing regulation by TDP-43; implications for neurodegenerative diseases. *Nat. Neurosci.* **14**, 452–458 (2011).

23. C. Appocher, F. Mohagheghi, S. Cappelli, C. Stuani, M. Romano, F. Feiguin, E. Buratti, Major hnRNP proteins act as general TDP-43 functional modifiers both in Drosophila and human neuronal cells. *Nucleic Acids Res.* **45**, 8026–8045 (2017).

24. E. Y. Liu, J. Russ, C. P. Cali, J. M. Phan, A. Amlie-Wolf, E. B. Lee Correspondence, Loss of Nuclear TDP-43 Is Associated with Decondensation of LINE Retrotransposons. *CellReports*. **27**, 1409-1421.e6 (2019).

25. J. R. Burrell, G. M. Halliday, J. J. Kril, L. M. Ittner, J. Götz, M. C. Kiernan, J. R. Hodges, The frontotemporal dementia-motor neuron disease continuum. *The Lancet*. **388**, 919–931 (2016).

26. D. Ray, H. Kazan, K. B. Cook, M. T. Weirauch, H. S. Najafabadi, X. Li, S. Gueroussov, M. Albu, H. Zheng, A. Yang, H. Na, M. Irimia, L. H. Matzat, R. K. Dale, S. A. Smith, C. A. Yarosh, S. M. Kelly, B. Nabet, D. Mecenas, W. Li, R. S. Laishram, M. Qiao, H. D. Lipshitz, F. Piano, A. H. Corbett, R. P. Carstens, B. J. Frey, R. A. Anderson, K. W. Lynch, L. O. F. Penalva, E. P. Lei, A. G. Fraser, B. J. Blencowe, Q. D. Morris, T. R. Hughes, A compendium of RNA-binding motifs for decoding gene regulation. *Nature*. **499**, 172–177 (2013).

27. E. H. Corder, A. M. Saunders, W. J. Strittmatter, D. E. Schmechel, P. C. Gaskell, G. W. Small, A. D. Roses, J. L. Haines, M. A. Pericak-Vance, Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*. **261**, 921–923 (1993).

28. J. S. Dittman, Unc13: a multifunctional synaptic marvel. *Curr. Opin. Neurobiol.* **57**, 17–25

439 (2019).

440 29. F. Varoqueaux, A. Sigler, J.-S. Rhee, N. Brose, C. Enk, K. Reim, C. Rosenmund, Total arrest
441   of spontaneous and evoked synaptic transmission but normal synaptogenesis in the absence
442   of Munc13-mediated vesicle priming. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 9037–9042 (2002).

443 30. F. Varoqueaux, M. S. Sons, J. J. Plomp, N. Brose, Aberrant Morphology and Residual
444   Transmitter Release at the Munc13-Deficient Mouse Neuromuscular Synapse. *Mol. Cell.*
445   *Biol.* **25**, 5973–5984 (2005).

446 31. L. A. Gilbert, M. A. Horlbeck, B. Adamson, J. E. Villalta, Y. Chen, E. H. Whitehead, C.
447   Guimaraes, B. Panning, H. L. Ploegh, M. C. Bassik, L. S. Qi, M. Kampmann, J. S.
448   Weissman, Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation.
449   *Cell*. **159**, 647–661 (2014).

450 32. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson,
451   T. R. Gingeras, Sequence analysis STAR: ultrafast universal RNA-seq aligner. **29**, 15–21
452   (2013).

453 33. A. Frankish, M. Diekhans, A.-M. Ferreira, R. Johnson, I. Jungreis, J. Loveland, J. M. Mudge,
454   C. Sisu, J. Wright, J. Armstrong, I. Barnes, A. Berry, A. Bignell, S. Carbonell Sala, J. Chrast,
455   F. Cunningham, T. Di Domenico, S. Donaldson, I. T. Fiddes, C. García Girón, J. M.
456   Gonzalez, T. Grego, M. Hardy, T. Hourlier, T. Hunt, O. G. Izuogu, J. Lagarde, F. J. Martin,
457   L. Martínez, S. Mohanan, P. Muir, F. C. P. Navarro, A. Parker, B. Pei, F. Pozo, M. Ruffier,
458   B. M. Schmitt, E. Stapleton, M.-M. Suner, I. Sycheva, B. Uszczynska-Ratajczak, J. Xu, A.
459   Yates, D. Zerbino, Y. Zhang, B. Aken, J. S. Choudhary, M. Gerstein, R. Guigó, T. J. P.
460   Hubbard, M. Kellis, B. Paten, A. Reymond, M. L. Tress, P. Flicek, GENCODE reference
461   annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).

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

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

466 36. F. Mölder, K. P. Jablonski, B. Letcher, M. B. Hall, C. H. Tomkins-Tinch, V. Sochat, J.
467   Forster, S. Lee, S. O. Twardziok, A. Kanitz, A. Wilm, M. Holtgrewe, S. Rahmann, S.
468   Nahnsen, J. Köster, Sustainable data analysis with Snakemake. *F1000Research*. **10**, 33
469   (2021).

470 37. J. Vaquero-Garcia, A. Barrera, M. R. Gazzara, J. González-Vallinas, N. F. Lahens, J. B.
471   Hogenesch, K. W. Lynch, Y. Barash, A new view of transcriptome complexity and
472   regulation through the lens of local splicing variations. *eLife*. **5**, e11752 (2016).

473 38. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic
474   features. *Bioinformatics*. **26**, 841–842 (2010).

475 39. R. Middleton, D. Gao, A. Thomas, B. Singh, A. Au, J. J.-L. Wong, A. Bomane, B. Cosson,
476   E. Eyras, J. E. J. Rasko, W. Ritchie, IRFinder: assessing the impact of intron retention on
477   mammalian gene expression. *Genome Biol.* **18**, 51 (2017).

478 40. K. J. Livak, T. D. Schmittgen, Analysis of Relative Gene
479   Expression Data Using Real−Time Quantitative PCR and the 2−ΔΔCT
480   Method. *Methods*. **25**, 402–408 (2001).

481 41. A. P. Pereverzev, N. G. Gurskaya, G. V. Ermakova, E. I. Kudryavtseva, N. M. Markina, A.
482   A. Kotlobay, S. A. Lukyanov, A. G. Zaraisky, K. A. Lukyanov, Method for quantitative
483   analysis of nonsense-mediated mRNA decay at the single cell level. *Sci. Rep.* **5**, 7729 (2015).

484 42. J. Humphrey, N. Birsa, C. Milioto, M. McLaughlin, A. M. Ule, D. Robaldo, A. B. Eberle, R.

485   Kräuchi, M. Bentham, A.-L. Brown, S. Jarvis, C. Bodo, M. G. Garone, A. Devoy, G. Soraru,
486   A. Rosa, I. Bozzoni, E. M. C. Fisher, O. Mühlemann, G. Schiavo, M.-D. Ruepp, A. M.
487   Isaacs, V. Plagnol, P. Fratta, FUS ALS-causative mutations impair FUS autoregulation and
488   splicing factor networks through intron retention. *Nucleic Acids Res.* **48**, 6889–6905 (2020).

489 43. L. Blazquez, W. Emmett, R. Faraway, J. M. B. Pineda, S. Bajew, A. Gohr, N. Haberman, C.
490   R. Sibley, R. K. Bradley, M. Irimia, J. Ule, Exon Junction Complex Shapes the
491   Transcriptome by Repressing Recursive Splicing. *Mol. Cell.* **72**, 496-509.e9 (2018).

492 44. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* **9**,
493   357–359 (2012).

494 45. T. Smith, A. Heger, I. Sudbery, UMI-tools: modeling sequencing errors in Unique Molecular
495   Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).

496 46. A. Buniello, J. A. L. MacArthur, M. Cerezo, L. W. Harris, J. Hayhurst, C. Malangone, A.
497   McMahon, J. Morales, E. Mountjoy, E. Sollis, D. Suveges, O. Vrousgou, P. L. Whetzel, R.
498   Amode, J. A. Guillen, H. S. Riat, S. J. Trevanion, P. Hall, H. Junkins, P. Flicek, T. Burdett,
499   L. A. Hindorff, F. Cunningham, H. Parkinson, The NHGRI-EBI GWAS Catalog of
500   published genome-wide association studies, targeted arrays and summary statistics 2019.
501   *Nucleic Acids Res.* **47**, D1005–D1012 (2019).

502 47. R. J. Pruim, R. P. Welch, S. Sanna, T. M. Teslovich, P. S. Chines, T. P. Gliedt, M. Boehnke,
503   G. R. Abecasis, C. J. Willer, LocusZoom: regional visualization of genome-wide association
504   scan results. *Bioinforma. Oxf. Engl.* **26**, 2336–2337 (2010).

505 48. 1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison,
506   H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis, A
507   global reference for human genetic variation. *Nature.* **526**, 68–74 (2015).

508 49. O. H. Tam, N. V. Rozhkov, R. Shaw, J. Ravits, J. Dubnau, M. Gale, H. Correspondence,
509   Postmortem Cortex Samples Identify Distinct Molecular Subtypes of ALS: Retrotransposon
510   Activation, Oxidative Stress, and Activated Glia. *Cell Rep.* **29** (2019),
511   doi:10.1016/j.celrep.2019.09.066.

512 50. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R.
513   Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map
514   format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).

515 51. Picard toolkit. *Broad Inst. GitHub Repos.* (2019) (available at
516   http://broadinstitute.github.io/picard/).

517 52. K. C. Cotto, Y.-Y. Feng, A. Ramu, Z. L. Skidmore, J. Kunisaki, M. Richters, S. Freshour, Y.
518   Lin, W. C. Chapman, R. Uppaluri, R. Govindan, O. L. Griffith, M. Griffith, RegTools:
519   Integrated analysis of genomic and transcriptomic data for the discovery of splicing variants
520   in cancer. *bioRxiv*, 436634 (2021).

521 53. Y. I. Li, D. A. Knowles, J. Humphrey, A. N. Barbeira, S. P. Dickinson, H. K. Im, J. K.
522   Pritchard, Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**,
523   151–158 (2018).

524 54. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
525   *ArXiv13033997 Q-Bio* (2013) (available at http://arxiv.org/abs/1303.3997).

526 55. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A.
527   Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M.
528   Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A
529   framework for variation discovery and genotyping using next-generation DNA sequencing
530   data. *Nat. Genet.* **43**, 491–498 (2011).

56. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

57. P. J. Lukavsky, D. Daujotyte, J. R. Tollervey, J. Ule, C. Stuani, E. Buratti, F. E. Baralle, F. F. Damberger, F. H.-T. Allain, Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struct. Mol. Biol.* **20**, 1443–1449 (2013).

Author contributions:

Conceptualization: ALB,OGW,MJK,SEH,JH,MEW,PF

Data curation: ALB,OGW,MZ,SBS

Formal analysis: ALB,OGW,MJK,MZ,SBS,AB

Funding acquisition: PF,MEW,EB

Investigation: ALB,OGW,MJK,SEH,MZ,FCYL,LM,YAQ,SBS,AB,WCL,AG

Methodology: ALB,OGW,MJK,SEH,JH,MEW,PF

Project administration: PF,MEW

577 Resources: HP,TL,EB
578 Software: ALB,OGW,MZ,SBS,JH
579 Supervision: PF,MEW,JH,JU,MS,TR,TL,EMCF,GS
580 Visualization: ALB,OGW,MJK,WCL
581 Writing – original draft: ALB,OGW,MJK,MEW,PF
582 Writing – review & editing: SEH,WCL,EB,JU,JH
583
584 **Competing interests:**
585 ALB, OGW, MJK, SEH, MEW and PF declare competing financial interest. A patent application
586 related to this work has been filed.
587
588 **Data and materials availability:**
589 Analysis code and data to reproduce figures available:
590 https://github.com/frattalab/unc13a_cryptic_splicing/
591 RNA-Seq Data for i3Neurons, SH-SY5Y and SK-N-DZ[a] are available through the European
592 Nucleotide Archive (ENA) under accession PRJEB42763.
593 Public data was obtained from Gene Expression Omnibus (GEO): iPSC MNs (Klim et al., 2019)-
594 GSE121569, SK-N-DZ[b]-GSE97262, and FACS-sorted frontal cortex neuronal nuclei-
595 GSE126543.
596 *Riboseq:* E-MTAB-10235.
597 *Targeted RNA seq:* E-MTAB-10237
598 *Minigene iCLIP*: E-MTAB-10297
599 *NYGC ALS Consortium RNA-seq:* RNA-Seq data generated through the NYGC ALS Consortium
600 in this study can be accessed via the NCBI's GEO database (GEO GSE137810, GSE124439,
601 GSE116622, and GSE153960). All RNA-Seq data generated by the NYGC ALS Consortium are
602 made immediately available to all members of the Consortium and with other consortia with
603 whom we have a reciprocal sharing arrangement. To request immediate access to new and
604 ongoing data generated by the NYGC ALS Consortium and for samples provided through the
605 Target ALS Postmortem Core, complete a genetic data request form at
606 ALSData@nygenome.org.
607 *NYGC ALS Consortium Whole Genome Seq:* to be released later with companion manuscript.
608

609 **Supplementary Materials**
610 Materials and Methods
611 Figs. S1 to S4
612 Table S1
613 Data S1 to S4
614 References (*31–57*)
615
616
617
618
619

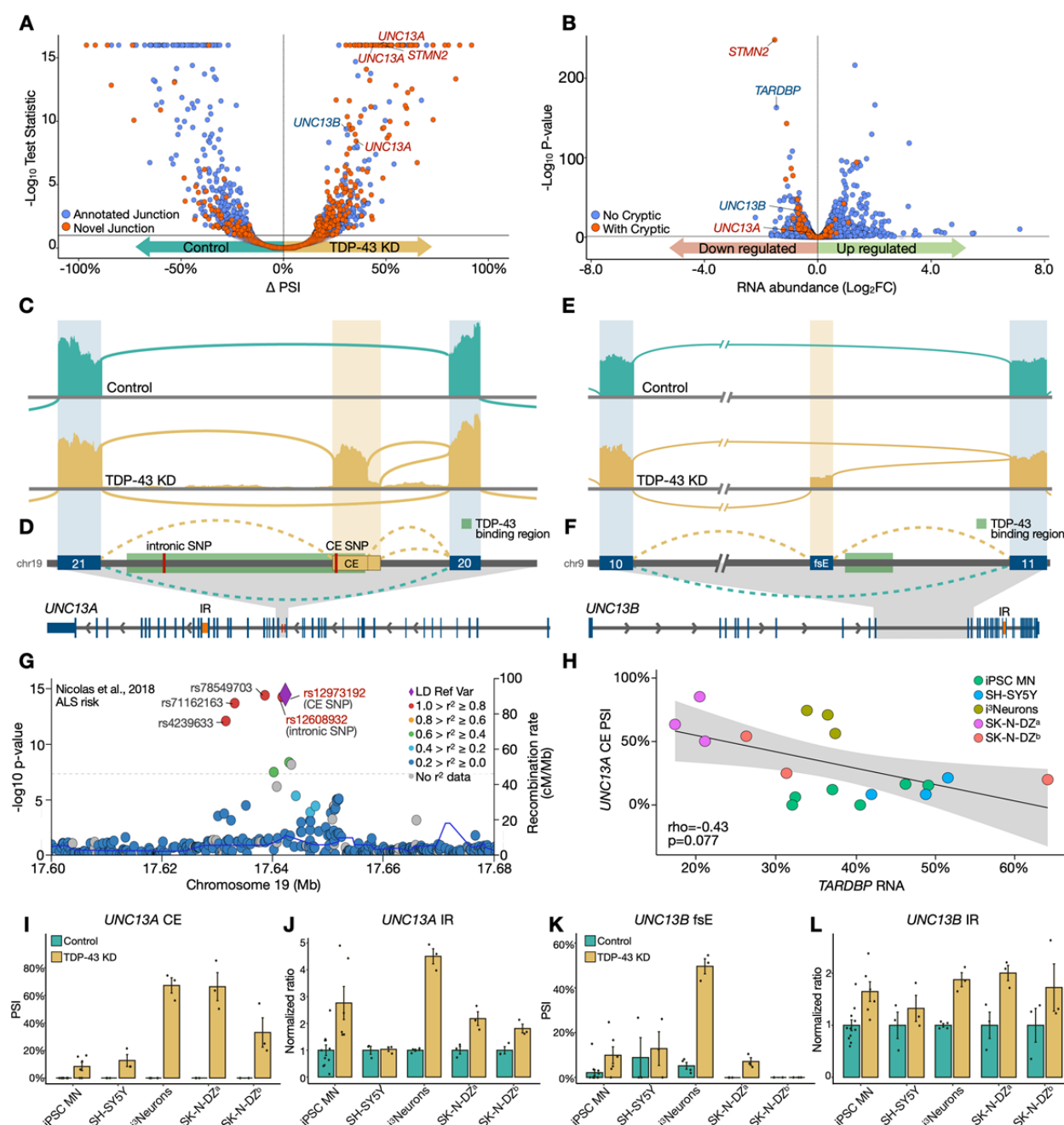14

620
621
622 **Figures**

623
624

Fig. 1. TDP-43 depletion in neurons leads to altered splicing in synaptic genes *UNC13A* and *UNC13B*. **(A)** Differential splicing and **(B)** expression in control (N=4) and CRISPRi TDP-43 depleted (N=3) iPSC-derived cortical-like i[3]Neurons. Each point denotes a splice junction (A) or

15

628   gene (B). **(C)** Sashimi plots showing cryptic exon (CE) inclusion between exons 20 and 21 of

629   *UNC13A* upon TDP-43 knockdown (KD). **(D,F)** Schematics showing intron retention (IR, lower

630   schematic, orange), TDP-43 binding region (*22*)(green), and two ALS/FTLD associated SNPs

631   (red). **(E)** Sashimi plot of *UNC13B* showing inclusion of the frameshifting exon (fsE) upon TDP-

632   43 KD. **(G)** LocusZoom plot of the *UNC13A* locus in the latest ALS GWAS. Lead SNP

633   *rs12973192* plotted as purple diamond, other SNPs coloured by linkage disequilibrium with

634   *rs12973192* in European individuals from 1000 Genomes. **(H)** Correlation between relative

635   *TARDBP* RNA and *UNC13A* CE PSI across five TDP-43 knockdown datasets **(I,K)** PSI of TDP-

636   43 regulated splicing in *UNC13A* and *UNC13B* across neuronal datasets. **(J,L)** Intron retention

637   ratio of TDP-43 regulated retained introns in *UNC13A* and *UNC13B* across neuronal datasets.

638
639
640

641

642

643

644

645

646

647

648

649

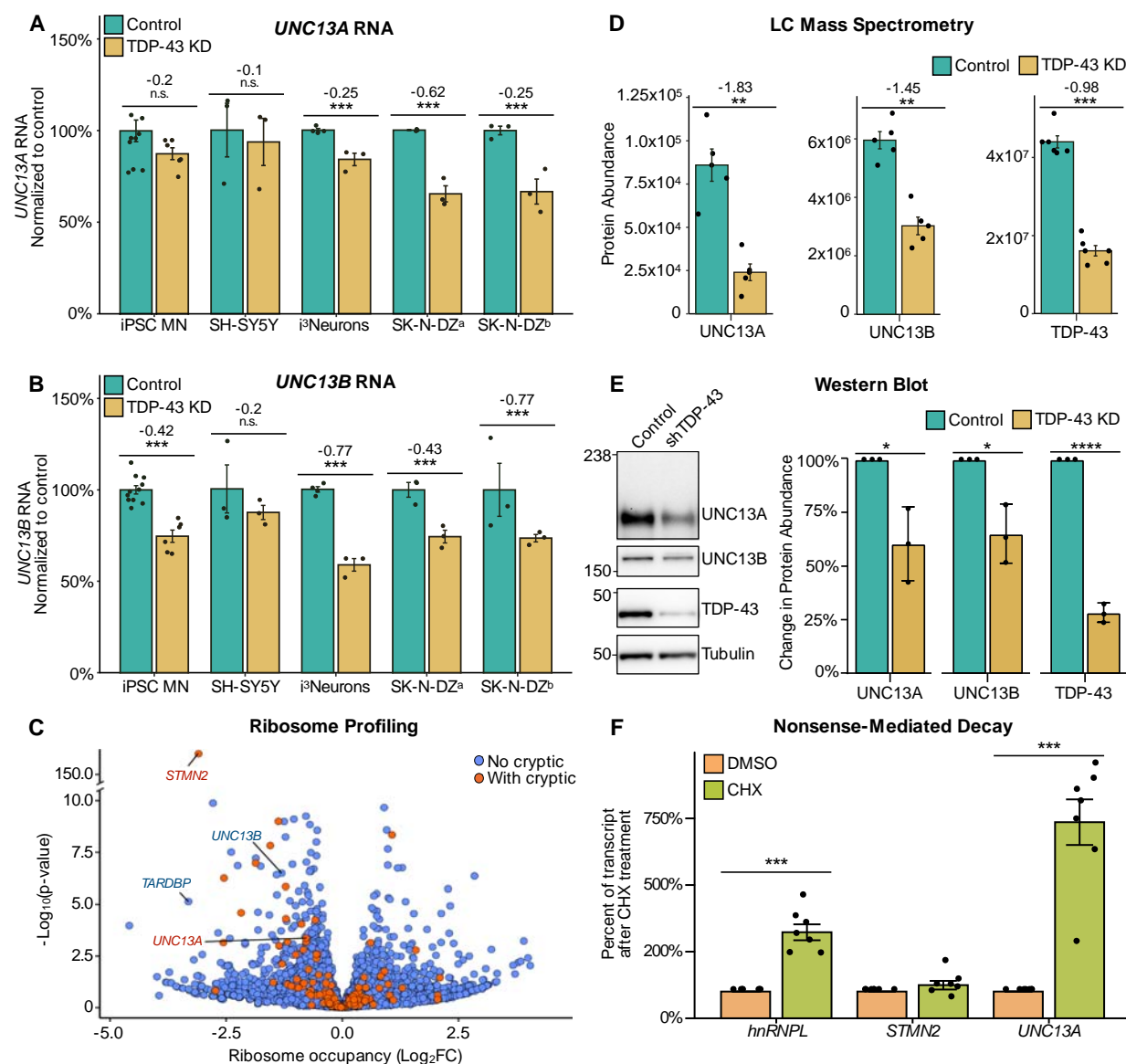650

651

652

653

Fig. 2. *UNC13A* **and** *UNC13B* **are downregulated after TDP-43 knockdown due to the production of NMD-sensitive transcripts.** Relative gene expression levels for *UNC13A (A)* and *UNC13B* **(B)** after TDP-43 knockdown across neuronal cell lines. Normalized RNA counts are shown as relative to control mean. Numbers show log$_2$ fold change calculated by DESeq2. Significance shown as adjusted p-values from DESeq2. **(C)** Ribosome profiling of TDP-43 knockdown in i³Neurons shows reduction in ribosome occupancy of *STMN2, UNC13A* and

660   *UNC13B* transcripts. **(D)** Mass spectrometry-based proteomic analysis shows reduction in

661   protein abundance of UNC13A, UNC13B and TDP-43 upon TDP-43 knockdown in i³Neurons.

662   Numbers refer to $\log_2$ fold change of unique peptide fragments, P-values from Wilcoxon test. **(E)**

663   Western blot analysis of protein lysates from untreated and TDP-43 knockdown SH-SY5Y cells

664   show a significant reduction in UNC13A and UNC13B proteins levels after TDP-43 depletion.

665   Graphs represent the means ± S.E., N=3, One sample t-test, **(F)** Transcript expression upon CHX

666   treatment suggests *UNC13A* but not *STMN2*, are sensitive to nonsense-mediated decay.

667   *HNRNPL* (heterogeneous nuclear ribonucleoprotein L) is a positive control. Significance levels

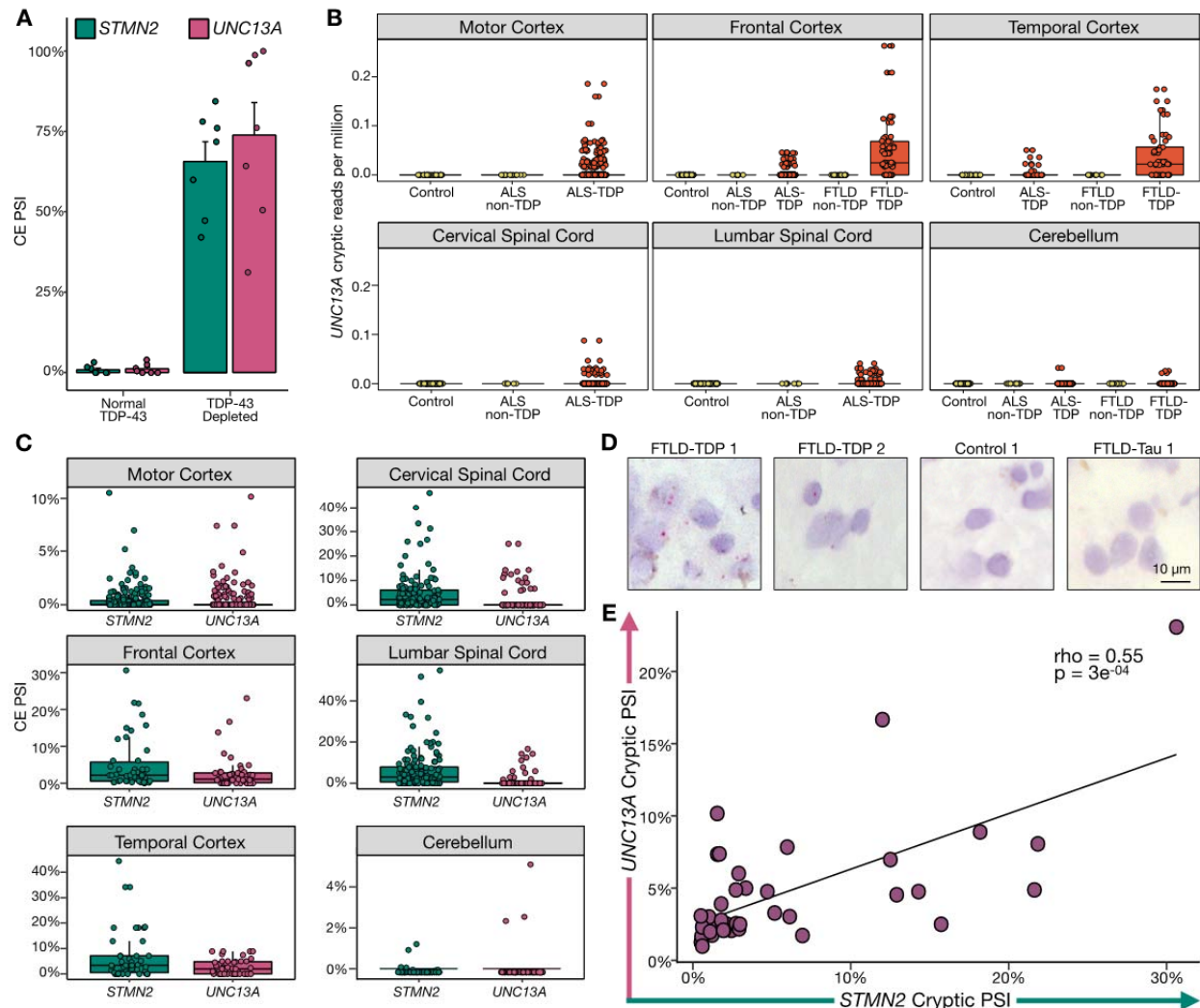668   reported as * (p<0.05) ** (p<0.01) *** (p<0.001) **** (p <0.0001).

669
670
671

672

673

674

675

676

677

678

Fig. 3. *UNC13A* **CE is highly expressed in ALS/FTLD patient tissue and correlates with known markers of TDP-43 loss of function.** (**A**) *UNC13A* and *STMN2* CE expression in ALS/FTLD patient frontal cortex neuronal nuclei from (*24*) sorted according to the expression of nuclear TDP-43. (**B**) *UNC13A* CE expression in bulk RNA-seq from NYGC ALS Consortium normalized by library size across disease and tissue samples. ALS cases stratified by mutation status, FTLD cases stratified by pathological subtype. (**C**) CE expression throughout ALS/FTLD-TDP cases across tissue (**D**) BaseScope detection of *UNC13A* CE (red foci) in FTLD-TDP but not control or FTLD-Tau frontal cortex samples. (**E**) Correlation in ALS/FTLD-

688   TDP cortex between *UNC13A* and *STMN2* CE PSI in patients with at least 30 spliced reads
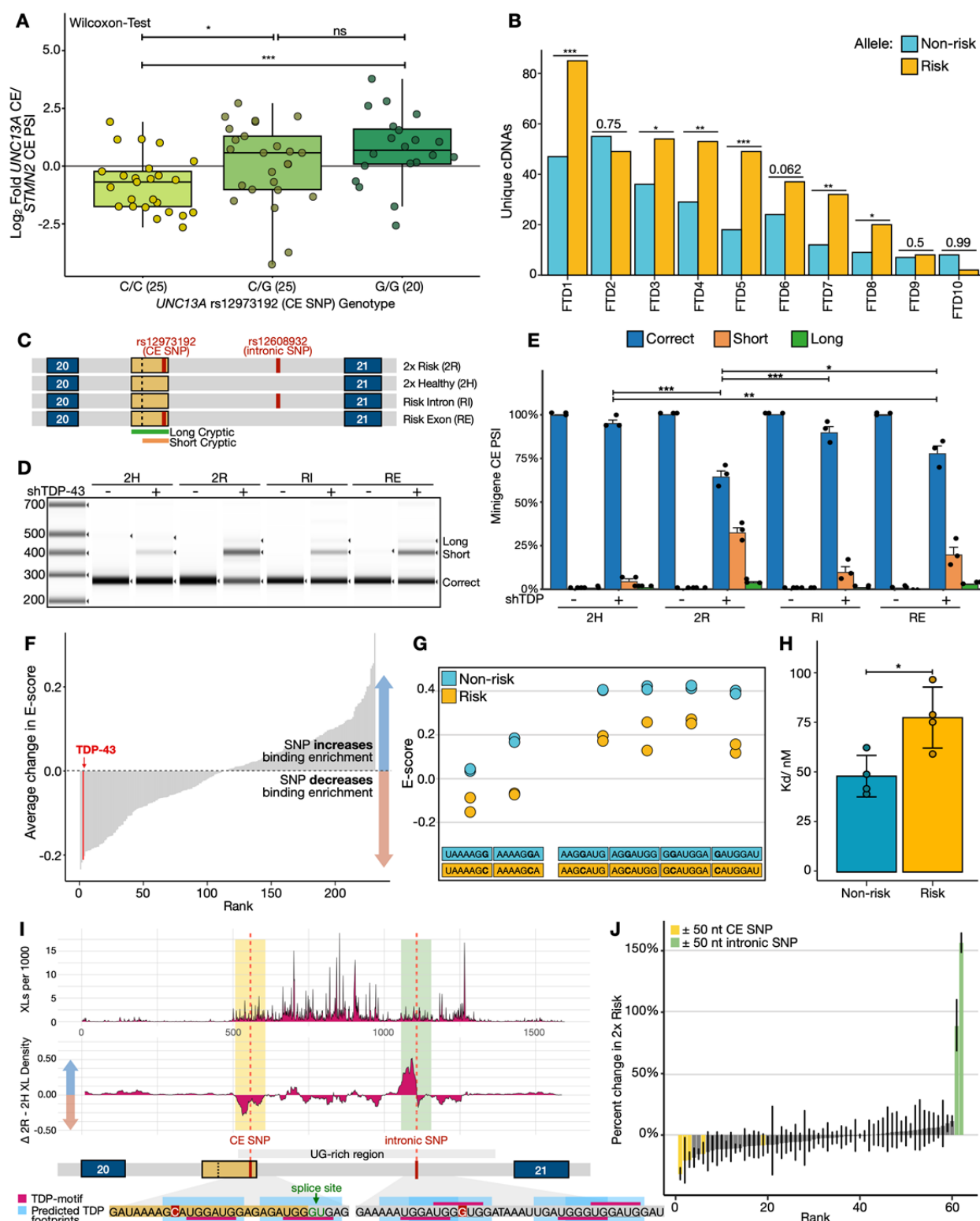
689   across the CE locus.

690

691   **Fig. 4.** *UNC13A* **ALS/FTD risk variants enhance** *UNC13A* **CE splicing in patients and in**

692   **vitro by altering TDP-43 pre-mRNA binding. (A)** Ratio *UNC13A / STMN2* CE PSI, split by

693   genotype for *UNC13A* risk alleles. **(B)** Unique cDNAs from targeted RNA-seq in ten CE SNP

694   heterozygous FTLD-TDP patients. p-values from single-tailed binomial tests. FTD1, 5, and 7 are

695   *C9orf72 hexanucleotide repeat* carriers **(C)** Illustration of *UNC13A* minigenes containing exon

696   20, intron 20, and exon 21 with both risk SNPs (2R), both healthy SNPs (2H), or risk SNP in CE

697   (RE) or intron (RI). **(D)** Representative image of RT-PCR products from *UNC13A* minigenes in

698   SH-SY5Y ± TDP-43 KD. **(E)** Quantification of (D) plotted as means ± S.E. N=3, One-way

699   ANOVA analysis; **(F)** Average change in E-value (measure of binding enrichment) across

700   proteins for heptamers containing risk/healthy CE SNP allele; red - TDP-43. **(G)** Each CE SNP

701   heptamer's TDP-43 E-value. **(H)** Binding affinities between TDP-43 and 14-nt RNA containing

702   the healthy or risk sequence measured by ITC; 4 replicates. **(I)** TDP-43 iCLIP of 2R and 2H

703   minigenes: top - average crosslink density; bottom - average density change 2R - 2H (rolling

704   window = 20 nt, units = crosslinks per 1,000). Cartoon - predicted TDP-43 binding footprints

705   (UGNNUG motif). **(J)** Fractional changes at iCLIP peaks for 2R versus 2H minigene (mean and

706   75% confidence interval shown). Peaks that are within 50nt of each SNP are highlighted. ***

707   (p<0.001) ** (p<0.01) * (p<0.05)**.**

708

709
710
711
712

Supplementary materials

714
715   Materials and Methods
716   Figs. S1 to S4
717   Table S1
718   Captions for Data S1 to S4

719
720 **Other Supplementary Materials for this manuscript include the following:**

721
722   Data S1 to S4
723
724
725
726

Materials and Methods:

Human iPSC culture

All policies of the NIH Intramural research program were followed for the procurement and use of induced pluripotent stem cells (iPSCs). The iPSCs used in this study were from the WTC11 line, derived from a healthy thirty-year old male, and obtained from the Coriell cell repository. All culture procedures were conducted as previously (*11*). In short, iPSCs were grown on tissue culture dishes coated with hESC-qualified matrigel (Corning, REF 354277). They were maintained in Essential 8 Medium (E8; Thermo Fisher Scientific, Cat. No. A1517001) supplemented with 10 μM ROCK inhibitor (RI; Y-27632; Selleckchem, Cat. No. S1049) in a 37°C, 5% CO_2 incubator. Media was replaced every 1-2 days as needed. Cells were passaged with accutase (Life Technologies, Cat. No. A1110501), 5-10 minutes treatment at 37°C. Accutase was removed and cells were washed with PBS before re-plating. Following dissociation, cells were plated in E8 media supplemented with 10 μM RI to promote survival. RI was removed once cells grew into colonies of 5-10 cells.

The following cell lines/DNA samples were obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research: GM25256

TDP-43 knockdown in human iPSCs

The human iPSCs used in this study were previously engineered (*11, 13*) to express mouse Neurogenin-2 (NGN2) under a doxycycline-inducible promoter integrated at the AAVS1 safe harbor, as well as an enzymatically dead Cas9 (+/- CAG-dCas9-BFP-KRAB) integrated at a safe harbor at the CLYBL promoter (*12*).

To achieve knockdown, sgRNAs targeting either *TARDBP*/TDP-43 or a non-targeting control guide were delivered to iPSCs by lenti-viral transduction. To make the virus, Lenti-X Human Embryonic Kidney (HEK) cells were transfected with the sgRNA plasmids using Lipofectamine 3000 (Life Technologies, Cat. No. L3000150), then cultured for 2-3 days in the following media: DMEM, high glucose GlutaMAX Supplement media (Life Technologies, Cat. No. 10566024) with 10% FBS (Sigma, Cat. No. TMS-013-B), supplemented with viral boost reagent (ALSTEM, Cat. No. VB100). Virus was then concentrated from the media 1:10 in PBS using Lenti-X concentrator (Takara Bio, Cat. No. 631231), aliquoted and stored at -80°C for future use.

The sgRNAs were cloned into either pU6-sgRNA EF1Alpha-puro-T2A-BFP vector (gift from Jonathan Weissman; Addgene #60955) (*12, 31*) or a modified version containing a human U6 promoter, a blasticidin (Bsd) resistance gene, and eGFP. sgRNA sequences were as follows: non-targeting control: GTCCACCCTTATCTAGGCTA and *TARDBP*: GGGAAGTCAGCCGTGAGACC.

Virus was delivered to iPSCs in suspension following an accutase split. Cells were plated and cultured overnight. The following morning, cells were washed with PBS and media was changed to E8 or E8+RI depending on cell density. Two days post lentiviral delivery, cells were selected overnight with either puromycin (10 μg/ml) or blasticidin (100 μg/ml). iPSCs were then expanded 1-2 days before initiating neuronal differentiation. Knockdown efficiency was tested at iPSC and neuronal stages using immunofluorescence, QT-PCR and observed in RNA-seq data.

iPSC-derived i³Neuron differentiation and culture

To initiate neuronal differentiation, 20-25 million iPSCs per 15 cm plate were individualized using accutase on day 0 and re-plated onto matrigel-coated tissue culture dishes in N2 differentiation media containing: knockout DMEM/F12 media (Life Technologies Corporation, Cat. No. 12660012) with N2 supplement (Life Technologies Corporation, Cat. No. 17502048), 1x GlutaMAX (Thermofisher Scientific, Cat. No. 35050061), 1x MEM nonessential amino acids (NEAA) (Thermofisher Scientific, Cat. No. 11140050), 10 μM ROCK inhibitor (Y-27632; Selleckchem, Cat. No. S1049) and 2 μg/mL doxycycline (Clontech, Cat. No. 631311). Media was changed daily during this stage.

On day 3 pre-neuron cells were replated onto dishes coated with freshly made poly-L-ornithine (PLO; 0.1 mg/ml; Sigma, Cat. No. P3655-10MG), either 96-well plates (50,000 per well), 6-well dishes (2 million per well), or 15 cm dishes (45 million per plate), in i³Neuron Culture Media: BrainPhys media (STEMCELL Technologies, Cat. No. 05790) supplemented with 1x B27 Plus Supplement (ThermoFisher Scientific, Cat. No. A3582801), 10 ng/mL BDNF (PeproTech, Cat. No. 450-02), 10 ng/mL NT-3 (PeproTech, Cat. No. 450-03), 1 μg/mL mouse laminin (Sigma, Cat. No. L2020-1MG), and 2 ug/mL doxycycline (Clontech, Cat. No. 631311). i³Neurons were then fed

780   three times a week by half media changes. i³Neuron were then harvested on day 17 post addition of doxycycline or
781   14 days after re-plating.
782

Generation of Stable TDP-43 knockdown cell line
784   SH-SY5Y and SK-N-DZ cells were transduced with SmartVector lentivirus (V3IHSHEG_6494503)
785   containing a doxycycline-inducible shRNA cassette for TDP-43. Transduced cells were selected with puromycin (1
786   µg/mL) for one week.
787

Depletion of TDP-43 from immortalised human cell lines
789   SH-SY5Y cells for RT-qPCR validations and western blots were grown in DMEM/F12 containing
790   Glutamax (Thermo) supplemented with 10% FBS (Thermo). For induction of shRNA against TDP-43 cells were
791   treated with 5 µg/mL Doxycyline Hyclate (Sigma D9891). After 3 days media was replaced with Neurobasal
792   (Thermo) supplemented with B27 (Thermo) to induce differentiation. After a further 7 days, cells were harvested for
793   protein or RNA. SH-SY5Y and SK-N-DZ cells for RNA-seq experiments were treated with siRNA, as previously
794   described (*23*).
795

RNA-sequencing, differential gene expression and splicing analysis
797   For RNA-seq experiments of i³Neurons, the i³Neurons were grown on 96-well dishes. To harvest on day
798   17, media was completely removed, and wells were treated with tri-reagent (100 µL per well) (Zymo research
799   corporation, Cat. No. R2050-1-200). Then 5 wells were pooled together for each biological replicate: control (n=3);
800   TDP-43 knockdown (n=4). To isolate RNA, we used a Direct-zol RNA miniprep kit (Zymo Research Corporation,
801   Cat. No. R2052), following manufacturer's instructions including the optional DNAse step. Note: one control
802   replicate did not pass RNA quality controls and so was not submitted for sequencing. Total RNA was then enriched
803   for polyA and sequenced 2x75 bp on a HiSeq 2500 machine.
804   Samples were quality trimmed using Fastp with the parameter "qualified_quality_phred: 10", and aligned
805   to the GRCh38 genome build using STAR (v2.7.0f) (*32*) with gene models from GENCODE v31 (*33*). Gene
806   expression was quantified using FeatureCounts (*34*) using gene models from GENCODE v31. Any gene which did
807   not have an expression of at least 0.5 counts per million (CPM) in more than 2 samples was removed. For
808   differential gene expression analysis, all samples were run in the same manner using the standard DESeq2 (*35*)
809   workflow without additional covariates, except for the Klim MNs dataset, where we included the day of
810   differentiation. DESeq2's median of ratios, which controls for both sequencing depth and RNA composition, was
811   used to normalize gene counts. Differential expression was defined at a Benjamini-Hochberg false discovery rate <
812   0.1. Our alignment pipeline is implemented in Snakemake version 5.5.4 (*36*) and available at:
813   https://github.com/frattalab/rna_seq_snakemake.
814   Differential splicing was performed using MAJIQ (v2.1) (*37*) using the GRCh38 reference genome. A
815   threshold of 0.1 ΔPSI was used for calling the probability of significant change between groups. The results of the
816   deltaPSI module were then parsed using custom R scripts to obtain a PSI and probability of change for each
817   junction. Cryptic splicing was defined as junctions with PSI < 0.05 in control samples, ΔPSI > 0.1, and the junction
818   was unannotated in GENCODE v31. Our splicing pipeline is implemented in Snakemake version 5.5.4 and available
819   at: https://github.com/frattalab/splicing.
820   Counts for specific junctions were tallied by parsing the STAR splice junction output tables using bedtools
821   (*38*). Splice junction parsing pipeline is implemented in Snakemake version 5.5.4 and available at:
822   https://github.com/frattalab/bedops_parse_star_junctions
823

824    $$\text{Percent spliced in (PSI)} = \frac{inclusion\ reads}{inclusion\ reads\ +\ exclusion\ reads}$$ using coordinates from Table S1.

825
826   Intron retention was assessed using IRFinder (*39*) with gene models from GENCODE v31.
827

Analysis of published iCLIP data
829   Cross-linked read files from TDP-43 iCLIP experiments in SH-SY5Y and human neuronal stem cells (*22*)
830   were processed using iCount v2.0.1.dev implemented in Snakemake version 5.5.4, available at
831   https://github.com/frattalab/pipeline_iclip . Sites of cross-linked reads from all replicates were merged into a single
832   file using iCount group command. Significant positions of cross-link read density with respect to the same gene
833   (GENCODE v34 annotations) were then identified using the iCount peaks command with default parameters. The
834   pipeline

835

Western Blot

837 SH-SY5Y cells were lysed directly in the sample loading buffer (Thermo NP0008). Lysates were heated at
838 95°C for 5 min with 100 mM DTT. If required lysates were passed through a QIAshredder (Qiagen) to shear DNA.
839 Lysates were resolved on 4-12% Bis-Tris Gels (Thermo) or homemade 6% Bis-Tris gels and transferred to 0.45 µm
840 PVDF (Millipore) membranes. After blocking with 5% milk, blots were probed with antibodies [Rb anti-UNC13A
841 (Synaptic Systems 126 103); Rb anti-UNC13B (abcam ab97924); Rat anti-Tubulin (abcam ab6161), Mouse anti-
842 TDP-43 (abcam ab104223)] for 2 hours at room temperature. After washing, blots were probed with HRP
843 conjugated secondary antibodies and developed with Chemiluminescent substrate (Thermo) on a ChemiDoc
844 Imaging System (Bio-Rad). Band intensity was measured with ImageJ (NIH).

845

RT-qPCR

847 RNA was extracted from SH-SY5Y and SK-N-DZ cells with a RNeasy kit (Qiagen) using the
848 manufacturer's protocol including the on column DNA digestion step. RNA concentrations were measured by
849 Nanodrop and 1 µg of RNA was used for reverse transcription. First strand cDNA synthesis was performed with
850 SSIV (Thermo 18090050) or RevertAid (Thermo K1622) using random hexamer primers and following the
851 manufacturer's protocol including all optional steps. Gene expression analysis was performed by qPCR using
852 Taqman Multiplex Universal Master Mix (Thermo 4461882) and TaqMan assays (UNC13A-Fam Hs00392638_m1,
853 UNC13B-Fam Hs01066405_m1, TDP-43-Vic Hs00606522_m1, GAPDH-Jun asay 4485713) on a QuantStudio 5
854 Real-Time PCR system (Applied Biosystems) and quantified using the ΔΔCt method (*40*).

855

Nonsense-mediated decay (NMD) inhibition

857 Ten days post induction of shRNA against TDP-43 with 1 µg/ml doxycyline hyclate (Sigma D9891-1G),
858 SH-SY5Y cells were treated either with 100 µM cycloheximide (CHX) or DMSO for 6 hours (*41*) before harvesting
859 the RNA through RNeasy Minikit (Qiagen). Reverse transcription was performed using RevertAid cDNA synthesis
860 kit (Thermo), and transcript levels were quantified by qPCR (QuantStudio 5 Real-Time PCR system, Applied
861 Biosystems) using the ΔΔCt method and *GAPDH* as reference (*40*). Since it proved to undergo NMD (*42*), *hNRNPL*
862 NMD transcript was used as a positive control.

863

Quantification of TDP-43, UNC13A, and UNC13B using quantitative proteomics

865 i^{3}Neurons were harvested from 6-well plates on day 17 post initiation of differentiation. Two wells were
866 pooled for each biological replicate, n=6 for each control and TDP-43 knockdown neurons. To harvest, wells were
867 washed with PBS, and then SP3 protein extraction was performed to extract intercellular proteins. Briefly, we
868 harvested and lysed using a very stringent buffer (50 mM HEPES, 50 mM NaCl, 5 mM EDTA 1% SDS, 1% Triton
869 X-100, 1% NP-40, 1% Tween 20, 1% deoxycholate and 1% glycerol) supplemental with cOmplete protease
870 inhibitor cocktail at 1 tablet/10 ml ratio. The cell lysate was reduced by 10 mM dithiothreitol (30 min, 60°C) and
871 alkylated using 20 mM iodoacetamide (30min, dark, room temperature). The denatured proteins were captured by
872 hydrophilic magnetic beads, and tryptic on-beads digestion was conducted for 16 hours at 37°C. We injected 1 µg
873 resulting peptides to a nano liquid chromatography (LC) for separation, and subsequently those tryptic peptides were
874 analyzed on an Orbitrap Eclipse mass spectrometer (MS) coupled with a FAIMS interface using data-dependent
875 acquisition (DDA) and data-independent acquisition (DIA). The peptides were separated on a 120 minute LC
876 gradient with 2-35% solvent B (0.1% FA, 5% DSMO in acetonitrile), and FAIMS's compensation voltages were set
877 to -50, -65 and -80. For DDA, we used MS1 resolution at 12000 and cycle time was selected for 3 seconds, MS2
878 fragments were acquired by linear ion trap. For DIA, we used 8 m/z isolation windows (400-1000 m/z range), cycle
879 time was set to 3 seconds, and MS2 resolution was set to 30000. The DDA and DIA MS raw files were searched
880 against Uniprot-Human-Proteome_UP000005640 database with 1% FDR using Proteome Discoverer (v2.4) and
881 Spectronaut (v14.1), respectively. The raw intensity of quantified peptides was normalized by total peptides
882 intensity identified in the same sample. The DDA quantified TDP-43- and UNC13A-derived unique and sharing
883 peptides were parsed out and used for protein quantification. Specifically, we visualized and quantified the unique
884 peptides of UNC13A using their MS/MS fragment ion intensity acquired by DIA.

885

Ribosome profiling

887 For ribosome profiling experiments, i^{3}Neurons were grown on 15 cm plates, one plate per biological
888 replicate for control (n=4) and TDP-43 knockdown (n=4) neurons. On day 17, i^{3}Neuron Culture Medium was
889 replaced 90 minutes prior to harvesting the neurons to boost translation. Then the medium was removed, cells were

890 washed with cold PBS, PBS was removed and 900 μL of cold lysis buffer (20 mM Tris pH 7.4, 150 mM NaCl,
891 5mM MgCl$_2$, 1 mM DTT freshly made, 100 ug/mL Cycloheximide, 1% TX100; 25 U/ml Turbo DNase I) was added
892 to each 15 cm plate. Lysed cells were scraped and pipetted into microcentrifuge tubes on ice. Cells were then passed
893 through a 26-gauge needle 10 times, and then centrifuged twice at 19,000xg 4°C, for 10 minutes, each time moving
894 the supernatant to a fresh tube. Tubes containing supernatant were flash frozen in liquid nitrogen and stored at -80°C
895 until processing.
896   Ribosome footprints from 3x TDP-43 knockdown and 3x control samples were generated and purified as
897 described, using a sucrose cushion (McGlincy and Ingolia, 2017) and a customised library preparation method based
898 on revised iCLIP (*43*). No rRNA depletion step was performed, and libraries were sequenced on an Illumina Hi-Seq
899 4000 machine (SR100). Reads were demultiplexed and adaptor/quality trimmed using Ultraplex
900 (https://github.com/ulelab/ultraplex), then aligned with Bowtie2 against a reference file containing abundant
901 ncRNAs that are common contaminants of ribosome profiling, including rRNAs (*44*). Reads that did not pre-map
902 were then aligned against the human genome with STAR (*32*) and the resulting BAM files were deduplicated with
903 UMI-tools (*45*). Multi-mapping reads were discarded and reads 28-30nt in length were selected for analysis.
904 FeatureCounts (*34*) was used to count footprints aligning to annotated coding sequences, and DESEQ2 (*35*) was
905 used for differential expression analysis, using default parameters in both cases. Periodicity analysis was performed
906 using a custom R script, using transcriptome-aligned bam files. Raw data has been uploaded to E-MTAB-10235.
907

Genome-wide association study data

909   Harmonised summary statistics for the latest ALS GWAS (*17*) were downloaded from the NHGRI-EBI
910 GWAS Catalog (*46*) (accession GCST005647). Locus plots were created using LocusZoom (*47*), using linkage
911 disequilibrium values from the 1000 Genomes European superpopulation (*48*).
912

NYGC ALS Consortium RNA-seq cohort

914   Our analysis contains 377 patients with 1349 neurological tissue samples from the NYGC ALS dataset,
915 including non-neurological disease controls, FTLD, ALS, FTD with ALS (ALS-FTLD), or ALS with suspected
916 Alzheimer's disease (ALS-AD). Patients with FTD were classified according to a pathologist's diagnosis of FTD
917 with TDP-43 inclusions (FTLD-TDP), or those with FUS or Tau aggregates. ALS samples were divided into the
918 following subcategories using the available Consortium metadata: ALS with or without reported SOD1 or FUS
919 mutations. All non-SOD1/FUS ALS samples were grouped as "ALS-TDP" in this work for simplicity, although
920 reporting of postmortem TDP-43 inclusions was not systematic and therefore not integrated into the metadata.
921 Confirmed TDP-43 pathology postmortem was reported for all FTLD-TDP samples.
922   Sample processing, library preparation, and RNA-seq quality control have been extensively described in
923 previous papers (*10*, *49*). In brief, RNA was extracted from flash-frozen postmortem tissue using TRIzol (Thermo
924 Fisher Scientific) chloroform, and RNA-Seq libraries were prepared from 500 ng total RNA using the KAPA
925 Stranded RNA-Seq Kit with RiboErase (KAPA Biosystems) for rRNA depletion. Pooled libraries (average insert
926 size: 375 bp) passing the quality criteria were sequenced either on an Illumina HiSeq 2500 (125 bp paired end) or an
927 Illumina NovaSeq (100 bp paired end). The samples had a median sequencing depth of 42 million read pairs, with a
928 range between 16 and 167 million read pairs.
929   Samples were uniformly processed, including adapter trimming with Trimmomatic and alignment to the
930 hg38 genome build using STAR (2.7.2a) (*32*) with indexes from GENCODE v30. Extensive quality control was
931 performed using SAMtools (*50*) and Picard Tools (*51*) to confirm sex and tissue of origin.
932   Uniquely mapped reads within the *UNC13A* locus were extracted from each sample using SAMtools. Any
933 read marked as a PCR duplicate by Picard Tools was discarded. Splice junction reads were then extracted with
934 RegTools (*52*) using a minimum of 8 bp as an anchor on each side of the junction and a maximum intron size of 500
935 kb. Junctions from each sample were then clustered together using LeafCutter (*53*) with relaxed junction filtering
936 (minimum total reads per junction = 30, minimum fraction of total cluster reads = 0.0001). This produced a matrix
937 of junction counts across all samples.
938   As the long CE acceptor was detected consistently in control cerebellum samples, as part of an unannotated
939 cerebellum-enriched 35 bp exon containing a stop codon between exons 20 and 21 (sup fig 3C,D), we excluded the
940 long CE acceptor for quantification of *UNC13A* CE PSI in patient tissue. Only samples with at least 30 spliced reads
941 at the exon locus were included for correlations.
942

BaseScope assay

944   Frozen tissue from the frontal cortex of FTLD-TDP (n = 5), FTLD-Tau (n = 3) and control (n = 3) cases
945 were sectioned at 10 μm thickness onto Plus+Frost microslides (Solmedia). Immediately prior to use, sections were

946 dried at RT and fixed for 15 minutes in pre-chilled 4 % paraformaldehyde. Sections were then dehydrated in
947 increasing grades of ethanol and pre-treated with RNAscope® hydrogen peroxide (10 mins, RT) and protease IV (30
948 mins, RT). The BaseScope™ v2-RED assay was performed using our UNC13A CE target probe (BA-Hs-UNC13A-
949 O1-1zz-st) according to manufacturer guidelines with no modifications (Advanced Cell Diagnostics, Newark, CA).
950 Sections were nuclei counterstained in Mayer's haematoxylin (BDH) and mounted (VectaMount). Slides were also
951 incubated with a positive control probe (Hs-PPIB-1 ZZ) targeting a common housekeeping gene and a negative
952 control probe (DapB-1 ZZ) which targets a bacterial gene to assess background signal (< 1-2 foci per ~ 100 nuclei).
953 Representative images were taken at x60 magnification.
954 Hybridised sections were graded, blinded to disease status, according to the relative frequency of red foci
955 which should identify single transcripts with the *UNC13A* CE event. Grades were prescribed by relative comparison
956 with the negative control slide. - = Less signal than negative control probe; + = similar signal strength to negative
957 control; ++ = visibly greater signal than negative control, +++ = considerably greater signal than negative
958 control.We identified a signal level above background (++ or +++) in 4 of 5 FTLD-TDP cases and a signal
959 considerably above background (+++) level in 2 cases. All FTLD-Tau and control cases were graded as exhibiting
960 either reduced (-) or comparable (+) signal relative to background.

961

962 ## *UNC13A genotypes in the NYGC ALS Consortium*

963 Whole Genome Sequencing (WGS) was carried out for all donors, from DNA extracted from blood or
964 brain tissue. Full details of sample preparation and quality control will be published in a future manuscript. Briefly,
965 paired-end 150bp reads were aligned to the GRCh38 human reference using the Burrows-Wheeler Aligner (BWA-
966 MEM v0.7.15) (*54*) and processed using the GATK best-practices workflow. This includes marking of duplicate
967 reads by the use of Picard tools(*51*) (v2.4.1), followed by local realignment around indels, and base quality score
968 recalibration using the Genome Analysis Toolkit (*55*, *56*) (v3.5). Genotypes for *rs12608932* and *rs12973192* were
969 then extracted for the samples.

970 ## *Targeted RNA-seq*

971 RNA was isolated from temporal cortex tissue of 10 FTLD-TDP and four control brains (6M, 4F, average
972 age at death 70.6±5.8y, average disease duration 10.98±5.9y). 50 mg of flash-frozen tissue was homogenised in 700
973 µl of Qiazol (Qiagen) using a TissueRuptor II (Qiagen). Chloroform was added and RNA subsequently extracted
974 following the spin-column protocol from the miRNeasy kit with DNase digestion (Qiagen). RNA was eluted off the
975 column in 50 µl of RNAse-free water. RNA quantity and quality were evaluated using a spectrophotometer.

976 Purified RNA was reverse transcribed with Superscript IV (Thermo Fisher Scientific) using either
977 sequence-specific primers containing sample-specific barcodes or random hexamers, following the manufacturer
978 recommendations. Unique molecular identifiers (UMIs) and part of the P5 Illumina sequence were added either
979 during first- or second-strand-synthesis (with Phusion HF 2x Master Mix) respectively. Barcoded primers were
980 removed with exonuclease I treatment (NEB; 30 min) and subsequently bead/size selection of RT/PCR products
981 (TotalPure NGS, Omega Biotek). Three rounds of nested PCR using Phusion HF 2x Master Mix (New England
982 Biolabs) were used to obtain highly specific amplicons for the UNC13A cryptic, followed by gel extraction and a
983 final round of PCR in which the full length P3/P5 Illumina sequences were added. Samples were sequenced with an
984 Illumina HiSeq 4000 machine (SR100).

985 Raw reads were demultiplexed, adaptor/quality trimmed and UMIs were extracted with Ultraplex
986 (https://github.com/ulelab/ultraplex), then aligned to the hg38 genome with STAR (*32*); for the hexamer data, a
987 subsample of reads was used to reduce the number of PCR duplicates during analysis. Reads were deduplicated via
988 analysis of UMIs with a custom R script; to avoid erroneous detection of UMIs due to sequencing errors, UMI
989 sequences with significant similarity to greatly more abundant UMIs were discarded - this methodology was tested
990 using simulated data, and final results were manually verified. Raw reads for targeted RNA-seq are available at E-
991 MTAB-10237.

992

993 Primers used:

Name	Sequence	Purpose	Sample
Specific_RT_1	ATCACGACGCTCTTCCGATCT NNNN TCATC ACC NNNN CATTGTTCTGCACGTCGGTC	Method 1 Reverse Transcription	P45/15
Specific_RT_2	ATCACGACGCTCTTCCGATCT NNNN TCATC GGA NNNN CATTGTTCTGCACGTCGGTC	Method 1 Reverse Transcription	P28/07
Specific_RT_3	ATCACGACGCTCTTCCGATCT NNNN TCATC ATA NNNN CATTGTTCTGCACGTCGGTC	Method 1 Reverse Transcription	P56/13

Specific_RT_4	ATCACGACGCTCTTCCGATCT NNNN TCATC TGG NNNN CATTGTTCTGCACGTCGGTC	Method 1 Reverse Transcription	P40/04
Specific_RT_5	ATCACGACGCTCTTCCGATCT NNNN TCATC GCT NNNN CATTGTTCTGCACGTCGGTC	Method 1 Reverse Transcription	P63/05
Specific_RT_6	ATCACGACGCTCTTCCGATCT NNNN TCATC GTG NNNN CATTGTTCTGCACGTCGGTC	Method 1 Reverse Transcription	P64/11
Specific_RT_7	ATCACGACGCTCTTCCGATCT NNNN TCATC CAA NNNN CATTGTTCTGCACGTCGGTC	Method 1 Reverse Transcription	P86/08
Specific_RT_8	ATCACGACGCTCTTCCGATCT NNNN TCATC TCA NNNN CATTGTTCTGCACGTCGGTC	Method 1 Reverse Transcription	P17/07
Specific_RT_9	ATCACGACGCTCTTCCGATCT NNNN TCATC GAC NNNN CATTGTTCTGCACGTCGGTC	Method 1 Reverse Transcription	P47/11
Specific_RT_10	ATCACGACGCTCTTCCGATCT NNNN TCATC CTT NNNN CATTGTTCTGCACGTCGGTC	Method 1 Reverse Transcription	P35/07
Specific_RT_11	ATCACGACGCTCTTCCGATCT NNNN TCATC TAT NNNN CATTGTTCTGCACGTCGGTC	Method 1 Reverse Transcription	P16/09
Specific_RT_12	ATCACGACGCTCTTCCGATCT NNNN TCATC AGT NNNN CATTGTTCTGCACGTCGGTC	Method 1 Reverse Transcription	P11/07
Specific_RT_13	ATCACGACGCTCTTCCGATCT NNNN TCATC TTC NNNN CATTGTTCTGCACGTCGGTC	Method 1 Reverse Transcription	P13/13
Specific_RT_14	ATCACGACGCTCTTCCGATCT NNNN TCATC CCG NNNN CATTGTTCTGCACGTCGGTC	Method 1 Reverse Transcription	P07/15
p5_sol_AT_V_V_short	ATCACGACGCTC	Method 1&2 Nested PCR 1	
p5_sol_AT_vshort	ATCACGACGCTCTTC	Method 1&2 Nested PCR 2	
p5_sol_AT	ATCACGACGCTCTTCCGATCT	Method 1&2 Nested PCR 3	
Fwd1	CAAGCGAACTGACAAATC	Method 1 Nested PCR 1	
Fwd2	GGCTCCACATCAGTGTG	Method 1 Nested PCR 2	
Fwd3	GTCCAGTACACCTGTCTGC	Method 1 Nested PCR 3	
UNC_TAR_FWD_V2	TACTGAACCGCTCTTCCGATCT GTCCAGTACACCTGTCTGC	Method 1 - add P3 solexa sequence	
unc_tg3_SSS_1	ATCACGACGCTCTTCCGATCT NNNN AACTC ACC NNNN CAGATGAATGAGTGATGAGTAG	Method 2 second strand synthesis	P45/15
unc_tg3_SSS_2	ATCACGACGCTCTTCCGATCT NNNN AACTC GGA NNNN CAGATGAATGAGTGATGAGTAG	Method 2 second strand synthesis	P28/07
unc_tg3_SSS_3	ATCACGACGCTCTTCCGATCT NNNN AACTC ATA NNNN CAGATGAATGAGTGATGAGTAG	Method 2 second strand synthesis	P56/13
unc_tg3_SSS_4	ATCACGACGCTCTTCCGATCT NNNN AACTC TGG NNNN CAGATGAATGAGTGATGAGTAG	Method 2 second strand synthesis	P40/04
unc_tg3_SSS_5	ATCACGACGCTCTTCCGATCT NNNN AACTC GCT NNNN CAGATGAATGAGTGATGAGTAG	Method 2 second strand synthesis	P63/05
unc_tg3_SSS_6	ATCACGACGCTCTTCCGATCT NNNN AACTC GTG NNNN CAGATGAATGAGTGATGAGTAG	Method 2 second strand synthesis	P64/11
unc_tg3_SSS_7	ATCACGACGCTCTTCCGATCT NNNN AACTC CAA NNNN CAGATGAATGAGTGATGAGTAG	Method 2 second strand synthesis	P86/08
unc_tg3_SSS_8	ATCACGACGCTCTTCCGATCT NNNN AACTC TCA NNNN CAGATGAATGAGTGATGAGTAG	Method 2 second strand synthesis	P17/07
unc_tg3_SSS_9	ATCACGACGCTCTTCCGATCT NNNN AACTC GAC NNNN CAGATGAATGAGTGATGAGTAG	Method 2 second strand synthesis	P47/11
unc_tg3_SSS_10	ATCACGACGCTCTTCCGATCT NNNN AACTC CTT NNNN CAGATGAATGAGTGATGAGTAG	Method 2 second strand synthesis	P35/07
unc_tg3_SSS_11	ATCACGACGCTCTTCCGATCT NNNN AACTC TAT NNNN CAGATGAATGAGTGATGAGTAG	Method 2 second strand synthesis	P16/09

unc_tg3_SSS_12	ATCACGACGCTCTTCCGATCT NNNN AACTC AGT NNNN CAGATGAATGAGTGATGAGTAG	Method 2 second strand synthesis	P11/07
unc_tg3_SSS_13	ATCACGACGCTCTTCCGATCT NNNN AACTC TTC NNNN CAGATGAATGAGTGATGAGTAG	Method 2 second strand synthesis	P13/13
unc_tg3_SSS_14	ATCACGACGCTCTTCCGATCT NNNN AACTC CCG NNNN CAGATGAATGAGTGATGAGTAG	Method 2 second strand synthesis	P07/15
unc_tg3_nest_1	CTGGGATCTTCACGACC	Method 2 Nested PCR 1	
unc_tg3_nest_2	ACGACCCCATTGTTCTGC	Method 2 Nested PCR 2	
unc_tg3_nest_3	GTTCTGCACGTCGGTCAC	Method 2 Nested PCR 3	
unc_tg3_add_p3	TACTGAACCGCTCTTCCGATCT GGTCACGAAGTGGAACAGG	Method 2 - add P3 solexa sequence	

Splicing reporters

One variant of the UNC13A exon 20, intron 20 and exon 21 sequence was synthesised and cloned into a pIRES-EGFP vector (Clontech) by BioCat. Plasmids with all four possible combinations of SNPs were generated by whole-plasmid PCR using primers with 5' mismatches, followed by phosphorylation and ligation. Stbl3 bacteria grown at 30°C were used due to the observed instability of the plasmids in DH5alpha cells grown at 37°C. Sequences were verified by Sanger sequencing.

TDP-43 inducible knockdown SH-SY5Y cells were electroporated with 2 μg of DNA with the Ingenio electroporation kit (Mirus) using the A-023 setting on an Amaxa II nucleofector (Lonza). The cells were then left untreated or treated for 6 days with 1 μg/mL doxycycline before RNA extraction. Reverse transcription was performed with RervertAid (Thermo Scientific) and cDNA was amplified by nested PCR with miniGene specific primers 5'-TCCTCACTCTCTGACGAGG-3' and 5'-CATGGCGGTCGACCTAG-3' followed by UNC13A specific primers 5'-CAAGCGAACTGACAAATCTGCCGTGTCG-3' and 5'-CGACACGGCAGATTTGTCAGTTCGCTTG-3'. PCR products were resolved on a TapeStation 4200 (Agilent) and bands were quantified with TapeStation Systems Software v3.2 (Agilent).

TDP-43 protein purification

His-tagged TDP-43 was expressed in BL21-DE3 Gold *E. coli* (Agilent) as previously described (*57*). Bacteria were lysed by two hours of gentle shaking in lysis buffer (50 mM sodium phosphate pH 8, 300 mM NaCl, 30 mM imidazole, 1 M urea, 1% v/v Triton X-100, 5 mM beta-mercaptoethanol, with Roche EDTA-free cOmplete protease inhibitor) at room temperature. Samples were centrifuged at 16,000 rpm in a Beckman 25.50 rotor at 4°C for 10 minutes, and the supernatant was clarified by vacuum filtration (0.22 μm).

The clarified lysate was loaded onto a 5 ml His-Trap HP column (Cytiva) equilibrated with Buffer A (50 mM sodium phosphate pH 8, 300 mM NaCl, 20 mM imidazole) using an AKTA Pure system, and eluted with a linear gradient of 0-100% Buffer B (50 mM sodium phosphate pH 8, 300 mM NaCl, 500 mM imidazole) over 90 column volumes. The relevant fractions were then analysed by SDS-PAGE and then extensively dialysed (3.5 kDa cutoff) against ITC buffer (50 mM sodium phosphate pH 7.4, 100 mM NaCl, 1 mM TCEP) at 4°C.

Isothermal titration calorimetry

RNAs with sequences 5'-AAGGAUGGAUGGAG-3' (healthy) and 5'-AAGCAUGGAUGGAG-3' (risk) were synthesised by Merck, resuspended in Ultrapure water, then dialysed against the same stock of ITC buffer overnight at 4°C using 1 kDa Pur-a-lyzer tubes (Merck). Protein and RNA concentrations after dialysis were calculated by A280 and A260 absorbance respectively. ITC measurements were performed on a MicroCal PEAQ-ITC calorimeter (Malvern Panalytical). Titrations were performed at 25°C with TDP-43 (9.6-12 μM) in the cell and RNA (96-120 μM) in the syringe. Data were analysed using the MicroCal PEAQ-ITC analysis software using nonlinear regression with the One set of sites model. For each experiment, the heat associated with ligand dilution was measured and subtracted from the raw data.

iCLIP of minigene-transfected cells

HEK293T cells were transfected with either the 2x Healthy or 2x Risk minigenes using Lipofectamine 3000 (Thermofisher Scientific). Each replicate consisted of 2x 3.5 cm dishes, with two replicates per sample, for eight dishes total. 48 h after transfection, cells were crosslinked with 150 mJ/cm^2 at 254 nm on ice, pelleted and flash frozen. Immunoprecipitations were performed with 4ug of TDP-43 antibody (proteintech 10782-2-AP) with

1036   100ul of protein G dynabeads per sample, and iCLIP sequencing libraries were prepared as described in (*43*).
1037   Libraries were sequenced on an Illumina HiSeq4000 machine (SR100).
1038   After demultiplexing the reads with Ultraplex, we initially aligned to the human genome using STAR (*32*),
1039   which showed that >5% of uniquely aligned reads mapped solely to the genomic region that is contained in the
1040   minigene. Given the high prior probability of reads mapping to the minigene, we therefore instead used Bowtie2 to
1041   align to the respective minigene sequences alone, thus minimising mis-mapping biases that could be caused by the
1042   SNPs(*44*) with settings "--norc --no-unal --rdg 50,50 --rfg 50,50 --score-min L,-2,-0.2 --end-to-end -N 1", then
1043   filtered for reads with no alignment gaps, and length >25 nt. Due to the exceptional read depth and high library
1044   complexity, we did not perform PCR deduplication to avoid UMI saturation at signal peaks. All downstream
1045   analysis was performed using custom R scripts. Raw data is available at E-MTAB-10297.
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092

1093
1094
1095
1096
1097
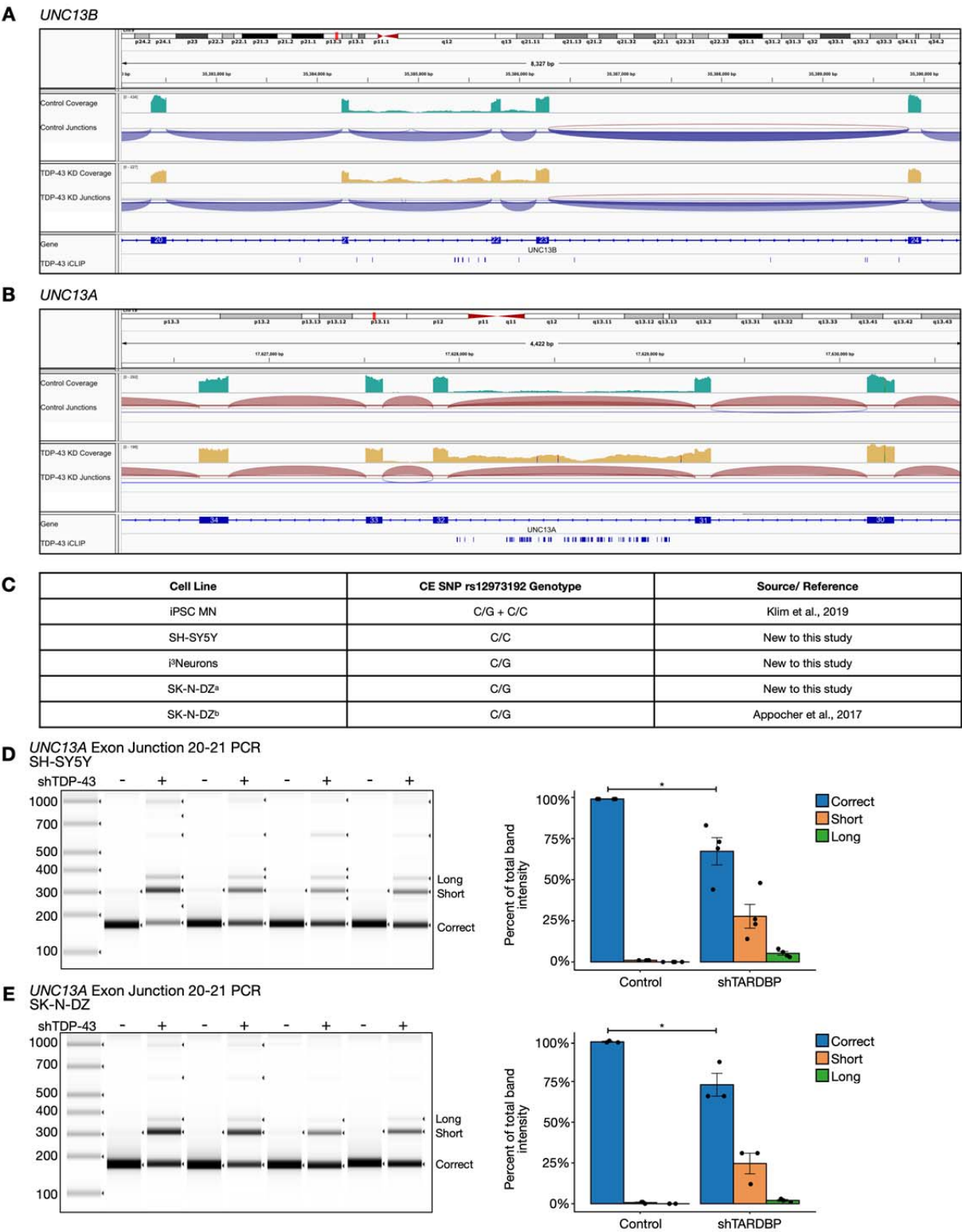1098   **Supplementary Figures**

1099

1100 **Fig. S1. Characterization and validation of *UNC13A* and *UNC13B* mis-splicing after TDP-43 KD. (A)**
1101 Genotypes and references of cell lines used in RNA-Seq analysis. **(B,C)** RNA-seq traces from IGV of representative
1102 samples from control (top) and *TARDBP* KD (bottom) in i³Neurons showing intron retention in *UNC13A* (A) (top)
1103 and *UNC13B* (B), overlaid with TDP-43 iCLIP peaks. **(D,E)** Capillary electrophoresis image of RT-PCR products
1104 for the *UNC13A* CE in SH-SY5Y (D) and SK-N-DZ (E) human cell lines upon *TARDBP* shRNA knockdown shows
1105 three products corresponding to correct splicing, a shorter and a longer CE. Replicates from each cell line are from 4
1106 independent experiments. Graphs represent the means ± S.E. of the quantification of each PCR product as a
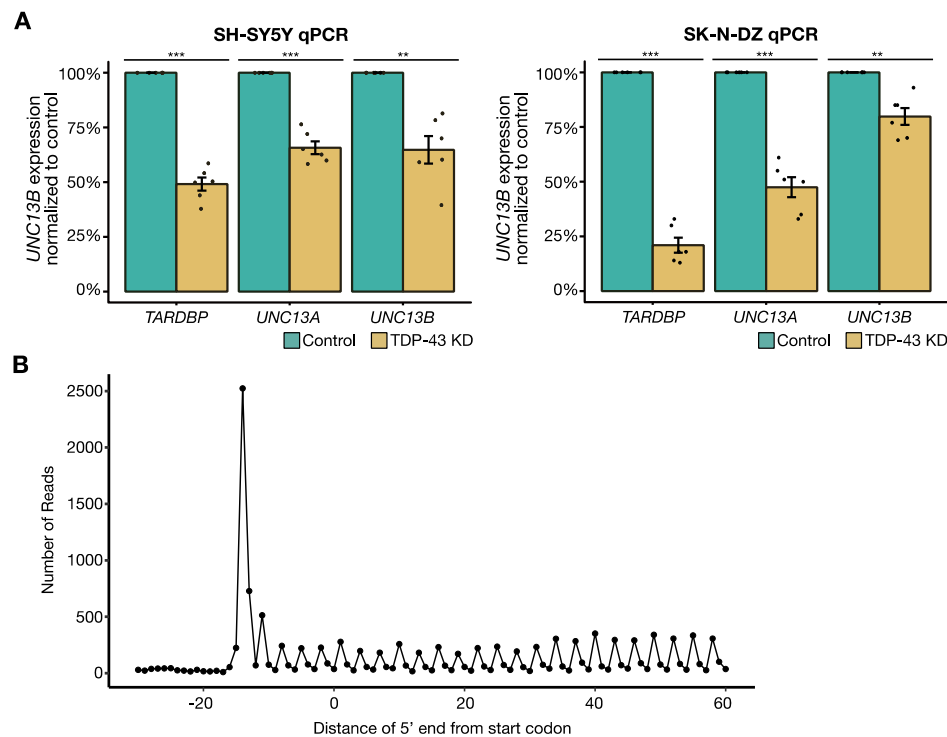1107 percentage of the total product. N=4, Student's t-test, *(p<0.05)
1108

Fig. S2. Orthogonal validation of *UNC13A* and *UNC13B* loss after TDP-43 KD. **(A)** RT-qPCR analysis shows *TDP-43, UNC13A* and *UNC13B* gene expression is reduced by *TARDBP* shRNA knockdown in both SH-SY5Y and SK-N-DZ human cell lines. Graphs represent the means ± S.E., N=6, One sample t-test, ***(p<0.001), **(p<0.01). **(B)** The 5' ends of 29 nt reads relative to the annotated start codon from a representative ribosome profiling dataset (TDP-43 KD replicate B). As expected, we detected strong three-nucleotide periodicity, and a strong enrichment of reads across the annotated coding sequence relative to the upstream untranslated region.
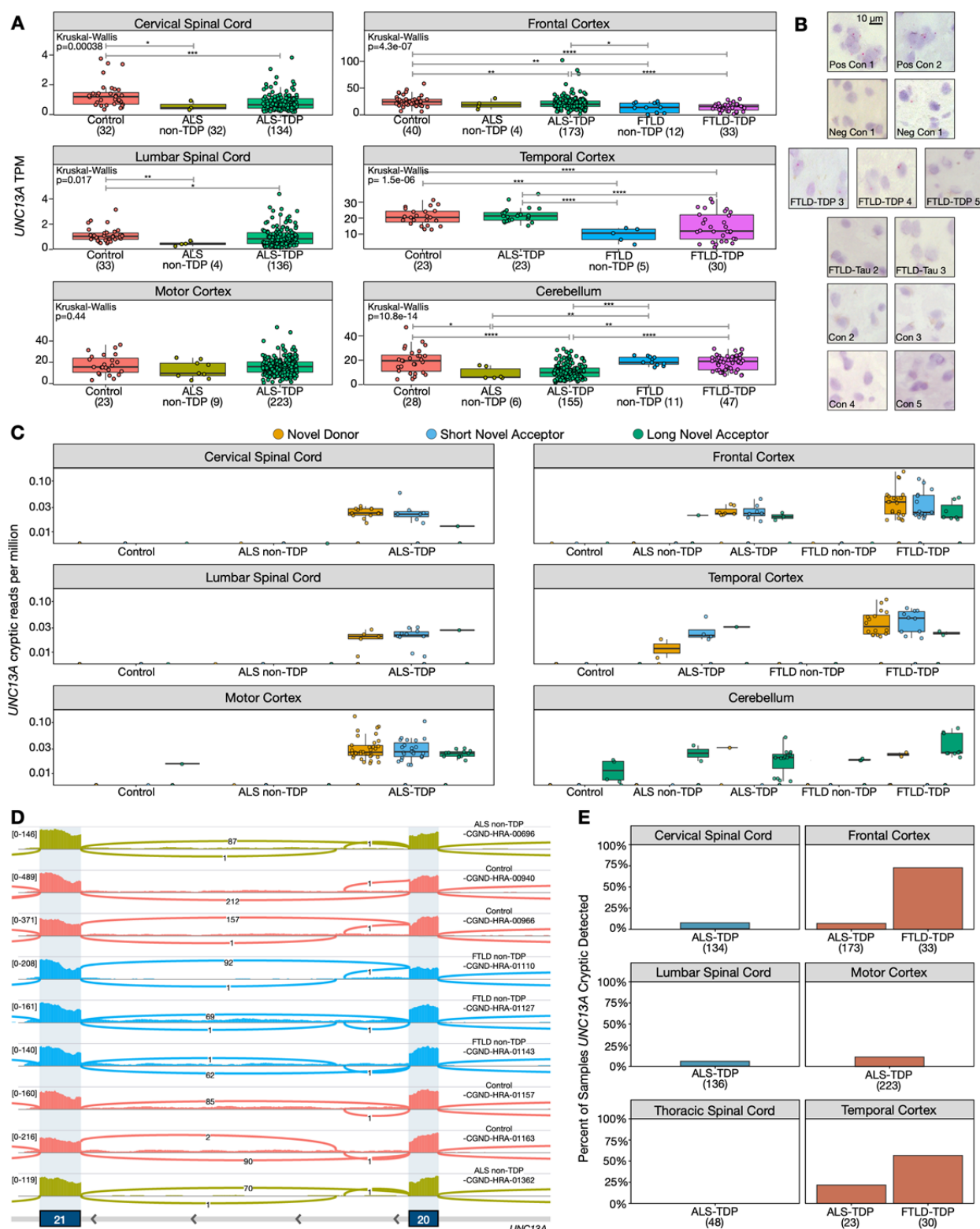
Fig. S3. *UNC13A* transcript and CE expression across neuronal tissues. (A) *UNC13A* expression across tissues and disease subtypes in the NYGC ALS Consortium RNA-seq dataset. Expression normalised as transcripts per million (TPM). Cortical regions have noticeably higher *UNC13A* expression than the spinal cord. **(B)** Representative images from RNA-based in situ hybridisation on frontal cortex brain tissue. Top and second rows show positive (PPIB-targeting) and negative (DapB-targeting) probe signal respectively. Other images are representative FTLD-TDP, FTLD-Tau and control cases validated with an UNC13a CE-targeting probe that were not shown in Fig. 3 **(C)**

1124 Expression of splice junction reads supporting the UNC13A CE across tissues and disease subtypes. Junction counts
1125 are normalised by library size in millions (junctions per million). The long novel acceptor junction is expressed
1126 across all disease subtypes in the cerebellum. **(D)** Example RNA-seq traces from IGV showing *UNC13A* cerebellar
1127 exon which shares the long novel acceptor junction as the *UNC13A* CE **(E)** Percentage disease relevant tissue
1128 samples with detectable *UNC13A* CE (1 supporting spliced read), split by disease and tissue.
1129
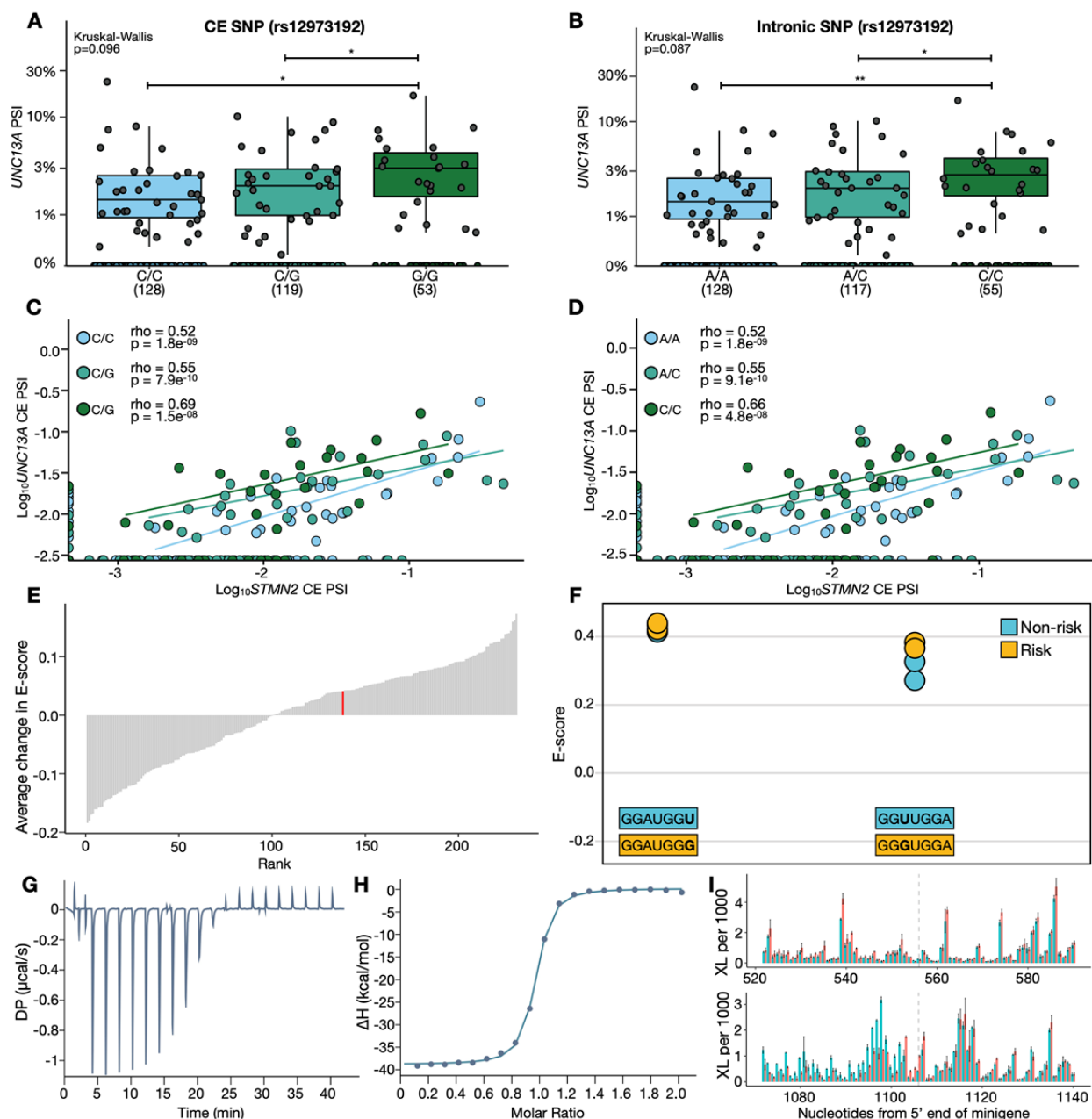1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141

Fig. S4. CE and intronic SNPs in *UNC13A* make the UNC13A CE more sensitive to TDP-43 depletion by altering TDP-43 binding affinity across the UNC13A CE-containing intron. (A-B) *UNC13A* CE PSI by genotype (Wilcoxon test) **(C-D)** Effect of CE or intronic SNP on the correlation between *STMN2* and *UNC13A* CE PSI in ALS/FTD cortex in samples with at least 30 junction reads across the CE locus. **(E)** The average changes in E-score for heptamers containing the intronic SNP; red = TDP-43. **(F)** Individual TDP-43 E-scores for the two heptamers for which there was data (average change shown in red in **(D)**). Significance levels reported as *** (P<0.001); ** (P<0.01); * (P<0.05); ns (P > 0.05). **(G-H)** ITC measurement of the interaction of TDP-43 with 14-nt RNA containing the healthy sequence. A representative data set is reported, with raw data **(G)** and integrated heat plot **(H).** Circles indicate the integrated heat, the curve represents the best fit. **(I)** Mean crosslink density around the exonic (top) and intronic (bottom) SNPs in the 2H (red) and 2R (blue) minigenes, relative to the 5' end of minigene (error bars = standard deviation; dashed lines show SNP positions).

1158 **Table S1. Hg38 coordinates for splice junctions used to calculate PSI.**

1159

1160

chromosome	start	end	name	score_left_blank_for_bed_format	strand
chr9	35313989	35366947	UNC13B_annotated	0	+
chr9	35313989	35364545	UNC13B_nmd_junction_5prime	0	+
chr9	35364567	35366947	UNC13B_nmd_junction_3prime	0	+
chr8	79611214	79636802	STMN2_annotated	0	-
chr8	79611214	79616822	STMN2_cryptic	0	-
chr19	17641556	17642414	UNC13A_NovelDonor	0	-
chr19	17641556	17642845	UNC13A_annotated	0	-
chr19	17642541	17642845	UNC13A_ShortNovelAcceptor	0	-
chr19	17642591	17642845	UNC13A_LongNovelAcceptor	0	-
chr19	17641556	17642557	UNC13A_cerebellum_donor	0	-

1161 **Data S1. (separate file)**

1162 List of differentially spliced junctions between control and TDP-43 KD i3Neurons (Fig. 1A).

1163 **Data S2. (separate file)**

1164 List of differentially expressed genes between control and TDP-43 KD i3Neurons (Fig. 1B).

1165 **Data S3. (separate file)**

1166 List of differentially ribosomal profiling genes between control and TDP-43 KD i3Neurons (Fig

1167 2C).

1168

1169 **Data S4. (separate file)**

1170 Individual and average thermodynamic parameters obtained from ITC experiments

1171

1172