

Characterizing the RNA targets and position-dependent splicing regulation by TDP-43

James R Tollervey^{1,8}, Tomaž Curk^{2,8}, Boris Rogelj^{3,8}, Michael Briese¹, Matteo Cereda^{1,4}, Melis Kayikci¹, Julian König¹, Tibor Hortobágyi³, Agnes L Nishimura³, Vera Župunski^{3,5}, Rickie Patani⁶, Siddharthan Chandran^{6,7}, Gregor Rot², Blaž Zupan², Christopher E Shaw³ & Jernej Ule¹

TDP-43 is a predominantly nuclear RNA-binding protein that forms inclusion bodies in frontotemporal lobar degeneration (FTLD) and amyotrophic lateral sclerosis (ALS). The mRNA targets of TDP-43 in the human brain and its role in RNA processing are largely unknown. Using individual nucleotide-resolution ultraviolet cross-linking and immunoprecipitation (iCLIP), we found that TDP-43 preferentially bound long clusters of UG-rich sequences *in vivo*. Analysis of RNA binding by TDP-43 in brains from subjects with FTLD revealed that the greatest increases in binding were to the MALAT1 and NEAT1 noncoding RNAs. We also found that binding of TDP-43 to pre-mRNAs influenced alternative splicing in a similar position-dependent manner to Nova proteins. In addition, we identified unusually long clusters of TDP-43 binding at deep intronic positions downstream of silenced exons. A substantial proportion of alternative mRNA isoforms regulated by TDP-43 encode proteins that regulate neuronal development or have been implicated in neurological diseases, highlighting the importance of TDP-43 for the regulation of splicing in the brain.

TAR DNA binding protein (TDP-43) is a predominantly nuclear protein that has been reported to regulate transcription, alternative splicing and RNA stability^{1,2}. TDP-43 has two RNA-recognition motif (RRM) domains, which bind UG repeat sequences with high affinity³. Mislocalisation and aggregation of TDP-43 have been implicated in the pathogenesis of ALS and FTLD. Mutations in the gene that encodes TDP-43 are associated with 1–4% of familial and sporadic ALS and rare cases of FTLD, and confer toxicity^{4,5}. The pathology of ~60% of FTLD and ~90% of ALS cases are characterized by the formation of cytoplasmic inclusion bodies in neurons and astroglia that are rich in ubiquitinated, hyperphosphorylated C-terminal fragments of TDP-43 (refs 5,6). Neurons that contain cytoplasmic TDP-43 inclusions often have markedly reduced TDP-43 staining in the nucleus^{5,6}. Thus, TDP-43 inclusions might have a toxic gain of function in the cytoplasm or sequester nuclear TDP-43 (or both), disrupting its role in RNA processing. TDP-43 cleavage products in the cytoplasmic inclusions have been shown to alter the splicing of a TDP-43-regulated exon in minigene studies, indicating that they can interfere with wild-type TDP-43 function⁷. The toxicity of TDP-43 overexpression depends on it retaining its RNA binding function in yeast⁸, *Drosophila* and chick⁹.

Only a handful of RNAs regulated by TDP-43 have been identified. To understand the function of TDP-43, and its dysfunction in disease, we have comprehensively identified the RNA sites that interact with TDP-43 in the human brain and characterized changes in RNA that follow reduced TDP-43 expression. To identify RNA targets on a global scale, we used individual-nucleotide resolution CLIP (iCLIP), which captures truncated cDNAs and quantifies them using a random barcode contained within the

primer that is used for reverse transcription¹⁰. Duplicate sequences that shared the same random barcode were removed from analysis to avoid any apparent binding-site bias due to PCR artifacts.

In this study, we sought to identify the RNAs that were bound by TDP-43 in healthy brain tissue and that from subjects with FTLD with TDP-43 inclusions (FTLD-TDP). Although this was the first study of protein-RNA interactions in post-mortem human brain, we identified many RNA binding sites using these samples. Most binding mapped to introns, long noncoding (nc)RNAs and intergenic transcripts, which also had the greatest enrichment of UG-rich motifs. We characterized the changes in alternative splicing that followed TDP-43 knockdown in neuroblastoma cells and identified splicing changes in 158 alternative cassette exons. Analysis of TDP-43 iCLIP cross-link sites in the vicinity of these exons provided insights into the mechanisms by which TDP-43 splicing is regulated. The exons regulated by TDP-43 are enriched in genes involved in neuronal development and those implicated in a range of neurodegenerative diseases.

RESULTS

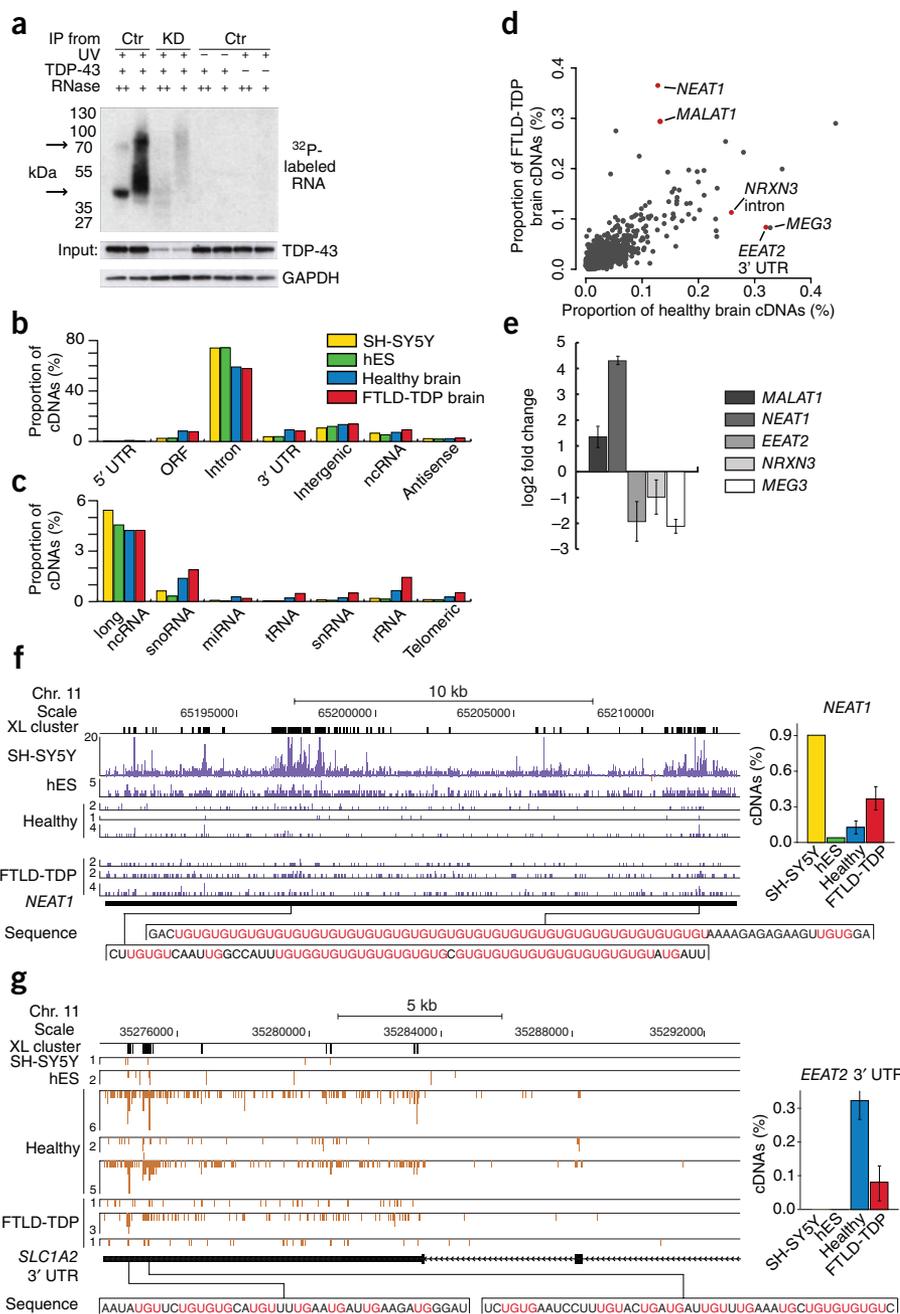
Comparing RNA targets of TDP-43 in healthy and FTLD brains

We purified TDP-43 in complex with its direct RNA targets from ultraviolet cross-linked cells and analyzed it on SDS-PAGE gel. The size of the dominant TDP-43-RNA complex in the low and high RNase conditions corresponded to a single TDP-43 molecule bound to the RNA; however, a complex corresponding to two TDP-43 molecules bound to the RNA was also present, particularly under low RNase

¹Medical Research Council (MRC) Laboratory of Molecular Biology, Cambridge, UK. ²Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia. ³MRC Centre for Neurodegeneration Research, King's College London, Institute of Psychiatry, London, UK. ⁴Scientific Institute IRCCS E. Medea, Bosisio Parini (LC), Italy. ⁵Faculty of Chemistry and Chemical Technology, University of Ljubljana, Ljubljana, Slovenia. ⁶MRC Laboratory for Regenerative Medicine, Department of Clinical Neurosciences, Cambridge, UK. ⁷Centre for Clinical Brain Sciences, University of Edinburgh, UK. ⁸These authors contributed equally to this work. Correspondence should be addressed to J.U. (jule@mrc-lmb.cam.ac.uk).

Received 16 December 2010; accepted 2 February 2011; published online 27 February 2011; doi:10.1038/nn.2778

Figure 1 Comparison of TDP-43 RNA binding in brain tissue from subjects with and without FTLD-TDP. **(a)** To validate the specificity of the antibody to TDP-43 for iCLIP, we isolated the 32 P-labeled RNA bound to TDP-43 gel from control (Ctr) or knockdown (KD) HeLa cells in the presence or absence of ultraviolet (UV) cross-linking or antibodies to TDP-43. We used high and low RNase concentrations to confirm the presence of RNA bound to TDP-43. Arrows, positions on gel corresponding to the size of TDP-43 monomer or dimer. TDP-43 western analysis of input extracts confirmed TDP-43 knockdown, and GAPDH was used as a loading control. For full image, see **Supplementary Figure 1a**. **(b)** The proportion of cDNAs (out of all cDNAs that mapped to human genome) from the TDP-43 iCLIP experiments in the four types of samples that mapped to different RNA regions. **(c)** The proportion of cDNAs that mapped to different types of ncRNAs. **(d)** The proportion of cDNAs that mapped to individual RNAs with at least 10 cDNAs in any experiment. Red, RNAs with the largest significant change between samples from control subjects and those with FTLD-TDP (difference in proportion of cDNAs > 0.1% and $P < 0.05$ by Student's t test, one-tailed, unequal variance). The long ncRNA *MEG3* (maternally expressed 3) with the largest decrease in TDP-43 binding in FTLD-TDP is also marked. **(e)** Real-time PCR analysis of transcripts with largest TDP-43 iCLIP changes in total RNA isolated from control brain samples and samples from subjects with FTLD-TDP. **(f)** iCLIP cDNA counts for TDP-43 cross-link positions in *NEAT1* in experiments from SH-SY5Y and hES cell lines, and healthy and FTLD-TDP tissue. Blue bars, sequences on the sense strand of the genome; height of bars corresponds to cDNA count. Positions of significant cross-link clusters (XL cluster) are shown on top; RNA sequence underlying the two main clusters is shown below (pink, UG repeats). **(g)** iCLIP cDNA counts for TDP-43 cross-link positions in *EEAT2*. Orange bars, sequences on the antisense strand of the genome. Data show mean \pm s.d. proportion of cDNAs that map to *NEAT1* transcript in each experiment (out of all cDNAs mapping to the human genome). **(h)** iCLIP cDNA counts for TDP-43 cross-link positions in *EEAT2* 3' UTR. The RNA sequence underlying the peak cross-linking sites is shown below; pink, UG repeats.



conditions (**Fig. 1a**). This result agreed with the earlier finding that TDP-43 binds RNA as a homodimer².

We carried out iCLIP experiments with human cortical tissue from post-mortem brain samples from three cognitively normal (healthy) subjects and three subjects with sporadic FTLD-TDP, as well as from embryonic stem (hES) cells and SH-SY5Y neuroblastoma cells (**Supplementary Table 1**). When we allowed a maximum of one nucleotide mismatch, 17 million sequence reads aligned as single hits to the human genome. After eliminating PCR amplification artifacts by removing sequences that truncated at the same nucleotide in the genome and shared the same random barcode, we identified 3.7 million reads representing unique cDNA molecules (**Supplementary Table 2**).

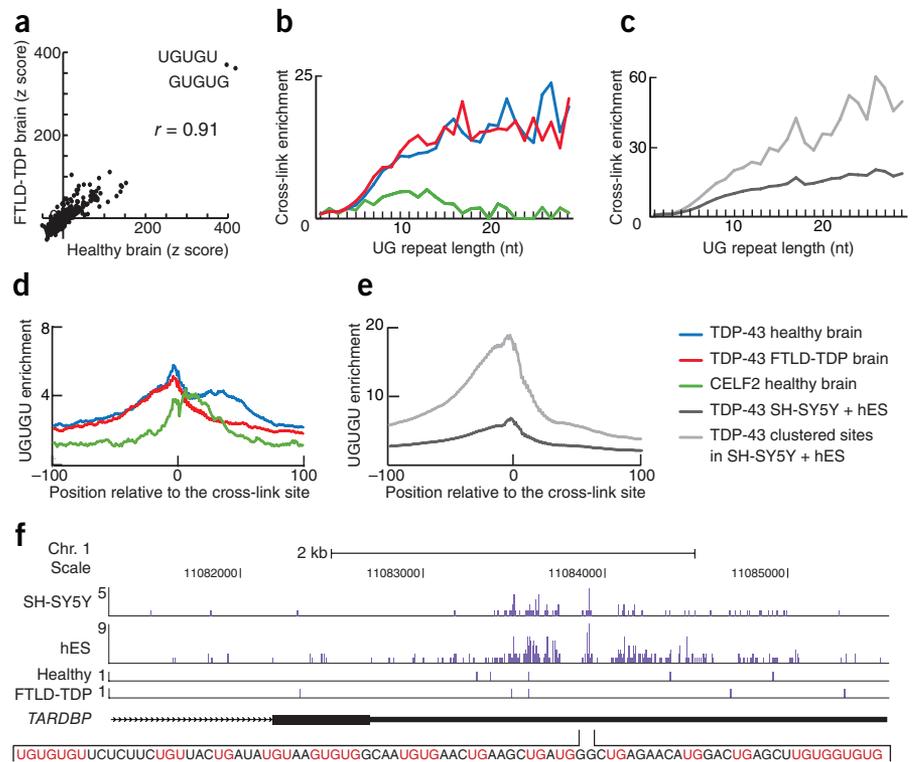
Genomic annotation of the cDNA sequences showed that most TDP-43 binding occurred in introns—74% in the cell lines and 58% in brain samples (**Fig. 1b**). The decreased proportion in brain samples, and

corresponding increase in exonic sequences, might be because intronic RNA is more sensitive to degradation in postmortem samples. About 5% of iCLIP cDNAs mapped to long ncRNAs in all samples (**Fig. 1c**). Samples from subjects with FTLD-TDP showed increased binding of TDP-43 to small nucleolar (sno)RNAs, small nuclear (sn)RNAs, ribosomal (r)RNAs and telomeric ncRNAs (**Fig. 1c**). In total, we found cross-link sites in 2,139 ncRNA genes in control samples and 2,702 in samples from subjects with FTLD-TDP.

The high proportion of intronic cDNA sequences indicated that the iCLIP data mainly represented nuclear RNA targets of TDP-43. We quantified the cross-linked complexes that were obtained using the iCLIP procedure from nuclear and cytoplasmic fractions of brain tissue from healthy subjects and those with FTLD-TDP. Only 6.5% of the total TDP-43–RNA complexes were obtained from cytoplasmic fractions of brain tissue, with no significant difference between tissue from healthy

Figure 2 TDP-43 binding motif analysis.

(a) z-scores of pentamer occurrence within the 61-nt sequence surrounding all cross-link sites (–30 nucleotides (nt) to +30 nt) for iCLIP of healthy samples and samples from subjects with FTLN-TDP. The sequences of the two most enriched pentamers and the Pearson correlation coefficient (r) between the two samples are given. (b) Enrichment of cross-linking compared to randomized data in UG repeats of different lengths in TDP-43 iCLIP experiments on brain samples from subjects without (blue) and with FTLN-TDP (red), and CELF2 experiments in healthy brain (green). (c) Enrichment of cross-linking compared to randomized data on UG repeats of different lengths in TDP-43 iCLIP experiments from SH-SY5Y and hES cells. Dark gray, all cross-link sites; light gray, only sites that mapped to cross-link clusters. (d) Analysis of positions of UGUGU enrichment compared to randomized data around TDP-43 cross-link sites. Color code as in b. (e) Analysis of positions of UGUGU enrichment compared to randomized data in TDP-43 iCLIP experiments from SH-SY5Y and hES cells. Color code as in c. (f) TDP-43 cross-linking in its own transcript (*TARDBP*). Replicate experiments are summed and shown in a single track. The RNA sequence underlying the peak cross-linking site is shown; pink, UG and GU dinucleotides.



subjects and those with FTLN-TDP (**Supplementary Fig. 1b**). Analysis of the RNAs bound by TDP-43 in the two fractions in SH-SY5Y cells by iCLIP showed 34% 3' UTR binding in the cytoplasm, 3.2% in the nuclear fraction and 3.8% in total cell extract (**Supplementary Fig. 1d** and **Fig. 1b**). As a control, we analyzed cross-link complexes with TIAL1, a protein that shuttles between the nucleus and cytoplasm. We obtained 31.5% of total TIAL1–RNA cross-linked complexes from cytoplasmic fractions of healthy brain tissue and samples from subjects with FTLN-TDP (**Supplementary Fig. 1b**). Thus, even though the TIAL1 analysis showed that iCLIP can efficiently purify proteins bound to cytoplasmic RNAs, our analysis indicates that more than 93% of TDP-43 iCLIP data presented here represents interactions with nuclear RNAs in brain tissue from both healthy subjects and those with FTLN-TDP.

To study the changes in TDP-43 binding to individual transcripts in FTLN-TDP, we analyzed the proportion of cDNAs that mapped to individual co-expressed genomic regions in the different brain samples. We found a significant ($P < 0.05$ by Student's t -test, one-tailed, unequal variance) and greater than 0.025% change in the proportion of cDNAs between control brain tissue and tissue from subjects with FTLN-TDP in 4 ncRNAs, 3' UTRs of 7 transcripts and introns of 48 transcripts (**Supplementary Table 3**). The fact that introns had the greatest number of changes indicates that the changes detected by iCLIP mainly reflect changes in RNA interactions that occur in cell nuclei.

Four transcripts had a significant and greater than 0.1% change in the proportion of cDNAs ($P < 0.05$; **Fig. 1d**). Analysis of the abundance of these transcripts by RNA-seq and real-time PCR showed that the change in TDP-43 binding agreed with a similar extent of change in transcript expression (**Fig. 1e** and **Supplementary Fig. 2**). The most significant increase in TDP-43 binding in FTLN-TDP was observed in nuclear paraspeckle assembly transcript 1 (*NEAT1*) and metastasis-associated lung adenocarcinoma transcript 1 (*MALAT1*, also known as *NEAT2*), which are transcribed from proximal loci on chromosome 11. Whereas *NEAT1* was previously identified as a 4-kb ncRNA¹¹, we detected TDP-43 binding along the length of a 18-kb ncRNA, with the peak binding in a

tandem UG-repeat sequence close to the 3' end (**Fig. 1d,f**). The primary binding sites in these two ncRNAs contained long UG repeats, and were identified in a highly reproducible manner (**Fig. 1f** and **Supplementary Fig. 1f**). The neurexin 3 (*NRXN3*) transcript and the glial excitatory amino acid transporter-2 (*EAAT2*, also called *SLC1A2*) had significantly decreased proportions of cDNA in tissue from subjects with FTLN-TDP (**Fig. 1d,g**). *EAAT2* is necessary for glutamate clearance at the synapse, a process that prevents glutamate excitotoxicity¹².

TDP-43 binds to clusters of UG-rich sequences

To compare the RNA sequence specificity of TDP-43 in the different iCLIP experiments, we assessed the enrichment of all possible pentamers within 30 nucleotides on either side of all cross-link sites. In all iCLIP datasets, the most significantly enriched pentamers were GUGUG and UGUGU (**Fig. 2a** and **Supplementary Fig. 3a–c**). The pentamer enrichment scores were correlated between samples from subjects with and without FTLN-TDP ($r = 0.91$; **Fig. 2a**).

We analyzed the relationship between TDP-43 binding and UG repeat length by monitoring the frequency of cross-linking in the last 10 nucleotides (nt) of UG tandem repeats compared to randomized positions. An example of a UG tandem repeat can be seen in the *NEAT1* ncRNA (**Fig. 1f**). Enrichment of TDP-43 cross-linking to UG repeats increased up to a repeat length of 15 nt, in a manner that was similar in samples from subjects with and without FTLN-TDP (**Fig. 2b,c**). For comparison, we performed iCLIP with CELF2, which also binds UG-rich motifs in the human brain (**Supplementary Fig. 3c**). In contrast to TDP-43, binding of CELF2 to UG repeats increased only up to a length of 10 nt, and dropped once the repeats became longer than 15 nt (**Fig. 2b**). This indicates that TDP-43 has a stronger preference for tandem UG repeats than does CELF2.

Proteins that interact with multiple identical sequences on RNA often allow spacing between different binding sites; for instance, NOVA proteins bind clusters of a minimum of three YCAY tetramers, but can tolerate variable spacing between these¹³. To analyze whether

TDP-43 can similarly bind to more dispersed UG-rich motifs, we monitored the frequency of UGUGU motifs over a 200-nt region surrounding the TDP-43 cross-link sites compared to randomized positions. UGUGU was enriched sevenfold at the TDP-43 cross-link sites, and remained twofold enriched even at a distance of 100 nt from the cross-link sites (Fig. 2d,e). One cause for this phenomenon may be a tendency of TDP-43 to bind to long UG-rich RNA regions, such as the 90-nt region in the 3' UTRs of *TARDBP* mRNA, where TDP-43 binds to autoregulate its mRNA levels^{14,15} (Fig. 2f). By contrast, even though UGUGU is fourfold enriched at the *CELF2* cross-link sites, we found no such enrichment at a distance of 100 nt (Fig. 2d). This indicates that TDP-43 has a unique capacity to recognize dispersed clusters of UG-rich motifs, or to spread its RNA binding to positions proximal to the UG-rich motifs.

To specifically analyze the high-affinity RNA binding sites of TDP-43, we grouped all data and searched for cross-link sites clustered with a maximum spacing of 15 nt that contained a substantially greater cDNA count than randomized positions (false discovery rate < 0.05). We identified 111,691 such cross-link clusters. The reproducibility of cross-linking within these clusters between different samples increased with the number of iCLIP cDNA sequences that mapped to a cluster (Supplementary Fig. 3d–g). For instance, 76% of clusters with summed cDNA counts of five or more in brain from healthy subjects were similarly bound in brain from subjects with FTLN (Supplementary Fig. 3d). Clustered cross-link sites were highly enriched in UG tandem repeats, with increased enrichment at repeat lengths up to 30 nt (Fig. 2c). UGUGU was 18-fold enriched at the TDP-43 cross-link sites, and remained fivefold enriched even at a distance of 100 nt from the cross-link sites (Fig. 2e). The increased incidence of UG-rich motifs at the cross-link clusters confirms that these clusters are the likely high-affinity binding sites of TDP-43.

To compare TDP-43 binding to different types of RNAs, we analyzed the enrichment of the UGUGU motif at the cross-link sites. The rRNAs, snRNAs, tRNAs and snoRNAs had the lowest UGUGU enrichment, indicating that cross-linking in these RNAs mainly reflected low-affinity, transient interactions (Supplementary Fig. 4). We found threefold enrichment in 5' UTRs and open reading frames (ORFs), and fourfold UG enrichment in 3' UTRs, micro (mi)RNA precursors and telomeric transcripts (Supplementary Fig. 4). UG was similarly enriched in the 3' UTRs when isolated with nuclear or cytoplasmic mRNA, indicating that TDP-43 binds to RNA in a sequence-specific manner in both the nucleus and the cytoplasm (Supplementary Fig. 1e). There was sevenfold enrichment of UGUGU pentamers in introns, long ncRNAs and intergenic RNAs, which together contain most TDP-43 binding sites (90% in cell lines and 76% in brain samples; Fig. 1b,c). The high UG enrichment in intergenic RNAs is particularly notable, indicating that high-affinity TDP-43 interactions occur with a large number of unannotated transcripts. Together, the iCLIP results indicate that TDP-43 binds pre-mRNAs and several types of ncRNAs in a sequence-specific manner.

The TDP-43 RNA splicing map

As most iCLIP cDNAs mapped to introns, we reasoned that TDP-43 might have an important role in splicing regulation. Therefore, we used Affymetrix high-resolution splice-junction microarrays to evaluate splicing changes in TDP-43 knockdown SH-SY5Y cells. Analysis of splicing

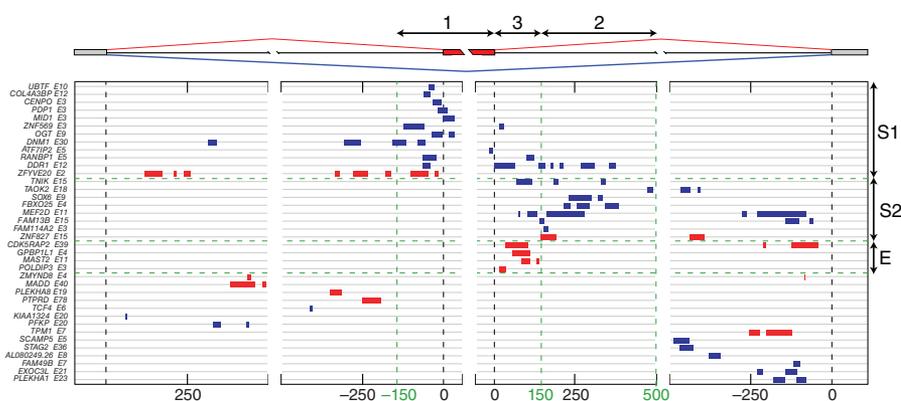


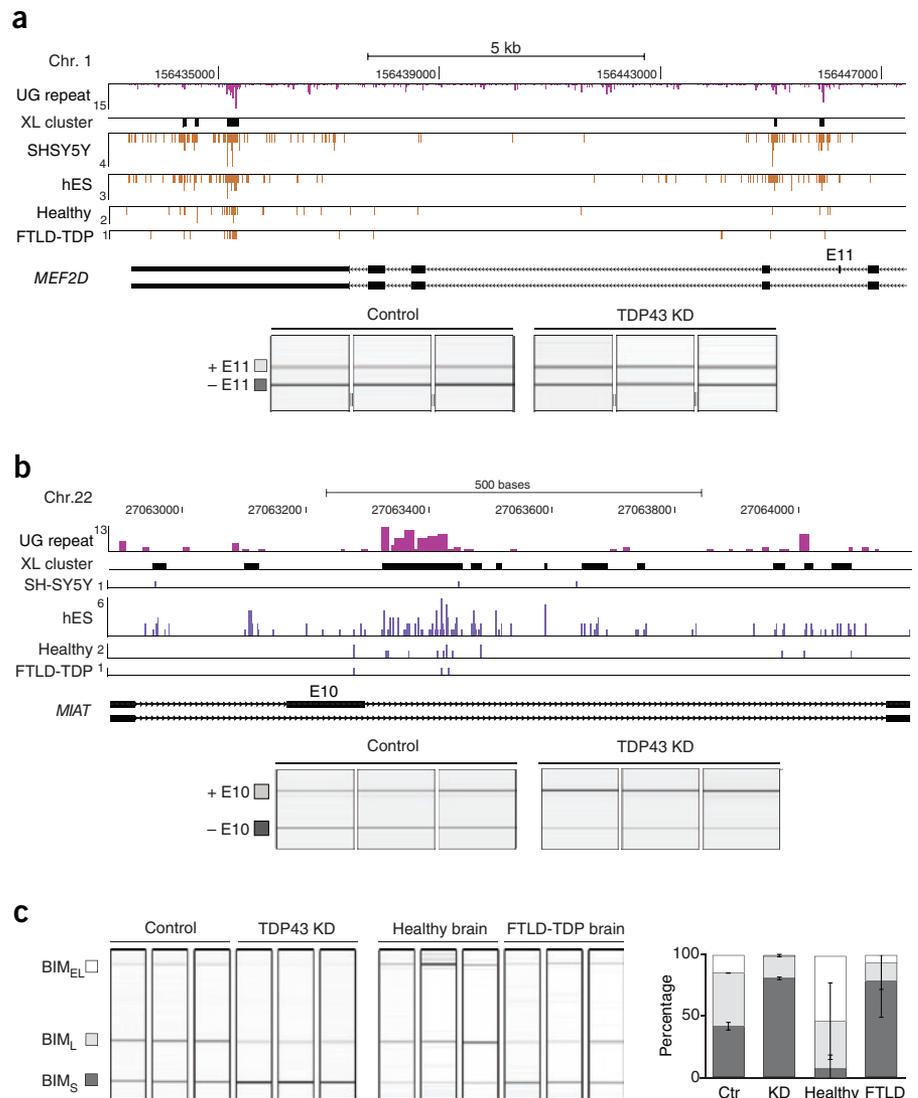
Figure 3 The RNA splicing map of TDP-43. Map of cross-link clusters at positions within 500 nt of alternative exons and flanking exons. Cassette exons with $\Delta\text{rank} > 1$ (enhanced exons, red clusters) or $\Delta\text{rank} < -1$ (silenced exons, blue clusters) with at least one cross-link cluster in these regions are shown. The exons are grouped by sequential analysis of cross-link cluster positions in three regions: group S1 is identified by clusters in region 1, 150 nt upstream of the exon and within the exon; group S2 by clusters in region 2, 150–500 nt downstream of the exon; and group E by clusters 0–150 nt downstream of the exon.

isoform reciprocity using the ASPIRE3 software identified 158 splicing changes in alternative cassette exons and 71 other types of splicing changes ($|\Delta\text{rank}| \geq 1$ where I is exon inclusion) out of 30,154 alternative exons that were detected by the microarray. We used RT-PCR and capillary electrophoresis to further assess 40 of the cassette exons. For 32 splicing events, primer pairs generated both isoforms, and 28 of these reproduced the direction of the observed splicing change, resulting in an 87% (28/32) validation rate of the microarray data (Supplementary Table 4).

To study how TDP-43 binds pre-mRNAs to regulate splicing, we visualized the positions of cross-link clusters in the form of an RNA splicing map, where the cross-link sites are plotted within 500 nt of the silenced (blue clusters) or enhanced (red clusters) cassette exons and the flanking exons (Fig. 3). The cross-link clusters identified three regions on the RNA splicing map where TDP-43 binding was mainly restricted to silenced or enhanced exons (Fig. 3). Binding in region 1, within the alternative exons or 0–150 nt upstream, was present at 11 silenced and 1 enhanced exon (group S1). Binding in region 2, 150–500 nt downstream of the alternative exons, was present at 7 silenced and 1 enhanced exon (group S2). Five of the silenced exons contained multiple cross-link clusters that spanned a sequence of 100 nt or longer in this region. This indicates that ‘deep’ intronic binding involving multiple dispersed binding sites might contribute to the ability of TDP-43 to silence exon inclusion. Finally, binding in region 3, 0–150 nt downstream of the alternative exons (but not in region 1 or 2), was present at 4 enhanced exons (group E).

To evaluate the proportion of TDP-43 iCLIP targets that changed splicing after knockdown, we analyzed splicing of exons with cross-link clusters in the three regions defined by the RNA splicing map. We used RT-PCR to evaluate splicing of 41 putative alternative exons that had highest number of cDNAs mapping to these regions. Of these exons, 29% (12/41) showed no evidence of alternative splicing in SH-SY5Y cells (single band detected by RT-PCR). Of the remaining exons, 79% (23/29) changed splicing after knockdown, indicating that iCLIP data alone has high predictive value (Supplementary Table 4 and Supplementary Fig. 5b). We also assayed splicing of 12 alternative exons that had UG-rich regions within 200 nt of the alternative exons, but were not identified by microarray or iCLIP. Splicing changes in 5 of these exons were validated by RT-PCR (Supplementary Table 4). Similar to the iCLIP binding, UG-rich motifs upstream of the exon or dispersed motifs downstream of the exon were associated with silencing, whereas proximal UG-rich motifs downstream of the exon were associated with enhancing the function of TDP-43 (Supplementary Fig. 5c).

Figure 4 TDP-43 regulates splicing of non-coding and protein-coding RNAs. **(a)** TDP-43 cross-linking in *MIAT* ncRNA. Pink bars, UG repeat lengths; black bars, cross-link clusters. Analysis of *MIAT* exon 10 inclusion (gel electropherogram, left, and quantification, right) in control and TDP-43 knockdown SH-SY5Y cells is shown below. **(b)** TDP-43 cross-linking in *MEF2D* protein-coding transcript. Colours as in **a**. Analysis of *MEF2D* exon 11 inclusion (gel electropherogram, left, and quantification, right) in control and TDP-43 knockdown SH-SY5Y cells is shown below. **(c)** Analysis of *BIM* exon 3 splicing (gel electropherogram, left, and quantification, right) in control and TDP-43 knockdown SH-SY5Y cells, and in brain samples from subjects without (C23, C25, C30) and with FTLN-TDP (F19, F20, F21).



TDP-43 regulates mRNAs involved in neuronal development

We also found TDP-43 cross-link clusters in introns of 7,499 protein-coding genes and 3' UTRs of 1,172 genes. Sixty-seven percent of the genes with 3' UTR clusters (785/1,172) also had intronic clusters. For instance, two cross-link clusters were present in the intron downstream of the alternative exon 11 and three clusters in the 3' UTR of myocyte enhancer factor 2D (*MEF2D*) mRNA (Fig. 4a). We identified a 29% decrease in inclusion of the *MEF2D* exon 11 in TDP-43 knockdown SH-SY5Y cells (Fig. 4a). The additional binding sites in the 3' UTR indicate that TDP-43 can remain associated with the *MEF2D* mRNA after splicing is finished and can thereby regulate additional aspects of RNA processing.

Owing to the strong sequence-specificity in binding to long ncRNAs (Supplementary Fig. 4), we also speculated that TDP-43 might regulate processing of such ncRNAs. Using iCLIP, we identified an alternative exon in the long ncRNA myocardial infarction associated transcript (*MIAT*). Eleven cross-link clusters are dispersed over the 1-kb region of both introns that flank this exon (Fig. 4b). We identified a 43% decrease in inclusion of the *MIAT* exon 10 in TDP-43 knockdown SH-SY5Y cells (Fig. 4b and Supplementary Fig. 5b).

We also identified a cross-link cluster in the *CFTR* pre-mRNA, which was the first identified RNA target of TDP-43 (ref. 3). The cross-link overlapped with the UG-rich element upstream of exon 9, which is required for TDP-43-dependent splicing regulation of this exon (Supplementary Fig. 6). Previous studies identified several candidate RNA targets of TDP-43, for which the direct *in vivo* interaction with TDP-43 has not been validated. Here, we identified TDP-43 binding sites in the *FUS* pre-mRNA¹⁶, *hNFL* 3' UTR¹⁷, the *CDK6* intron and 3' UTR¹⁸, and the intronic region that overlaps with the *mir558* gene¹⁹ (Supplementary Fig. 6). We also observed cross-link clusters in the mRNAs encoding HDAC6 and Casein kinase 1 and 2, which have been reported to regulate the phosphorylation and expression of TDP-43, respectively^{20,21} (Supplementary Fig. 6).

We performed GO term analysis of genes regulated by TDP-43 at the level of alternative splicing, as determined by analysis of SH-SY5Y knockdown cells. This identified several terms related to development, including neural tube formation, organ morphogenesis and chordate embryonic development (Table 1). Fifteen of the exons that are regulated

by TDP-43 are within genes that encode proteins with well-characterized functions in neuronal survival or development (Table 2). TDP-43 promotes inclusion of alternative exon 3 in Bcl-2 interacting mediator of cell death (*BIM*, also called *BCL2L11*), and thereby suppresses production of the most cytotoxic *BIM* isoform (*BIM_S*)²². We have evaluated splicing of *BIM* in FTLN-TDP, and have found that the *BIM_S* isoform increases in tissue from subjects with FTLN-TDP to a similar extent as in TDP-43 knockdown cells (Fig. 4c). This result indicates that aberrant splicing might contribute to neurodegeneration in the human brain.

DISCUSSION

Our results show that TDP-43 interacts with a diverse spectrum of RNAs with important functions in the brain. We identified binding of TDP-43 to non-coding RNAs, introns and 3' UTRs of mRNAs, indicating that TDP-43 has a role in coupling different aspects of gene expression. Even though only a minor portion of the protein was detected in the cytoplasm, we found a tenfold increase in TDP-43 binding to 3' UTR elements when comparing RNAs bound to cytoplasmic and nuclear protein. We also identified changes in alternative splicing associated with decreased TDP-43 protein. Analysis of cross-link clusters in the regulated pre-mRNAs indicated that the splicing function of TDP-43 depends on the position where it interacts with pre-mRNAs.

We found that TDP-43 preferentially bound to UG tandem repeats or long clusters of UG-rich motifs. These results agree with *in vitro* studies in which RNA binding of TDP-43 was shown to increase with the length of the UG repeat^{2,3}. iCLIP results also indicated that the increased affinity for long binding sites differentiates the RNA specificity of TDP-43 from CELF2, which recognizes shorter clusters of UG-rich motifs. As a single RRM domain generally recognizes four nucleotides²³, this RNA specificity for long binding sites indicates that TDP-43 can cooperatively recognize multiple RNA binding sites and thereby achieve high-affinity RNA binding. The apparent RNA binding of TDP-43 in a homodimer state might contribute to its ability to recognize binding sites composed of multiple dispersed UG-rich motifs (Fig. 1a).

TDP-43 cross-link clusters were located upstream from and within the silenced exons and immediately downstream of enhanced exons, which is also a characteristic of several other RNA-binding proteins, including the Nova proteins²⁴. In addition, dispersed clusters of TDP-43 cross-linking further downstream of exons were associated with splicing silencing. A similar mode of splicing silencing, involving multiple binding sites flanking an alternative exon, has also been reported for Sex lethal (Sxl), heterogeneous nuclear ribonucleoprotein A1, C and L (hnRNP A1, L and C), and polypyrimidine tract binding protein (PTB)^{10,25,26}. The glycine-rich C-terminal domains of Sxl and hnRNP A1 are required for the cooperative assembly of these proteins on multiple proximal binding sites and the associated splicing silencing²⁵. TDP-43 also contains a glycine-rich C terminus. Mutations in the *TARDDBP* gene that were identified in subjects with ALS or FTL are concentrated in the region that encodes the glycine-rich C terminus⁵, indicating that the ability of TDP-43 to assemble cooperatively on long RNA binding sites is involved in the disease mechanisms.

Table 1 Significant GO terms of the exons regulated by TDP-43

GO term	P value	Reference exons	TDP-43 regulated exons
Organ morphogenesis	<0.001	811	13 BCL2L11_E5, COL4A3BP_E12, CTNND1_E30, DLC1_E2, FGFR1_E4, FZD6_E2, MACF1_E68, MEF2D_E11, PLEKHA1_E21, PLEKHA1_E23, TFAP2A_E2, TLE1_E10, TPM1_E7
Neural tube closure	0.001	45	3 DLC1_E2, FZD6_E2, TFAP2A_E2
Homeostasis of number of cells	0.01	187	4 BCL2L11_E5, RPS24_E11, RPS24_E9, SOX6_E9
Mitotic cell cycle	0.016	856	9 CLIP1_E33, HORMAD1_E4, KIF2A_E19, MYO16_E8, PMF1_E2, PSMD6_E4, RBM38_E3, SIRT7_E9, STAG2_E36
Cell surface receptor linked signaling pathway	0.018	2,005	16 ANXA1_E2, BRD8_E24, CNTFR_E2, CTNND1_E30, FZD6_E2, GRM4_E7, HOMER2_E5, IRAK3_E2, MACF1_E68, MADD_E40, PILRB_E25, PLEKHA1_E21, PLEKHA1_E23, PTPRD_E78, TLE1_E10, TNIK_E15
Chordate embryonic development	0.021	475	6 BCL2L11_E5, COL4A3BP_E12, DLC1_E2, FZD6_E2, SOX6_E9, TFAP2A_E2
Lipid transport	0.037	276	4 ANXA1_E2, ATP11C_E36, COL4A3BP_E12, STARD4_E5

The tandem UG repeat sequences that are recognized by TDP-43 correspond to the most common microsatellite (AC) in the human genome. Variations in the lengths of microsatellites can introduce heritable phenotypic variation²⁷ or lead to changes in transcription and alternative splicing²⁸. Genes that are involved in organ morphogenesis and neurogenesis are enriched in the variable tandem repeats²⁹. Similarly, the genes that contain the exons that are regulated by TDP-43 are enriched for neural tube formation and organ morphogenesis. This is consistent with the finding that a TDP-43 knockout is lethal for mice from embryonic day 3.5 (ref. 30). It could be speculated that the ability of TDP-43 to regulate splicing by binding the UG repeats in the human genome might contribute to variation in the splicing regulation of genes that are involved in development.

TDP-43 regulates the splicing of several transcripts that encode proteins involved in neuronal survival, as well as seven mRNAs that encode proteins that are relevant for neurodegenerative diseases (Table 2).

Table 2 Neuronal functions of proteins encoded by the alternative mRNA isoforms regulated by TDP-43

Gene symbol	Gene name	Neurological disease association	Neuronal function	Pubmed PMID
AP2 (TFAP2A)	Transcription factor AP-2 alpha		Cranial neural-tube closure	11137286
BIM (BCL2L11)	Bcl-2 interacting mediator of cell death	Ischemia-induced neuronal death	Apoptotic signaling cascade	16780816 17568190
CNTFR	Ciliary neurotrophic factor receptor	Amyotrophic lateral sclerosis	Neuronal survival	19386761
MADD	MAP-kinase activating death domain	Alzheimer's disease	Neuronal survival under stress conditions	9482930
MEF2D	Myocyte enhancer factor 2D	Parkinson's disease	Neuronal survival	19119233
CDK5RAP2	CDK5 regulatory subunit associated protein 2	Primary autosomal-recessive microcephaly	Maintenance of the neural progenitor pool	20471352
CTNND1	Delta catenin	Mental retardation syndrome Cri-du-Chat	Maintenance of dendritic spines	19914181
DLC1	Deleted in liver cancer 1		Neural tube development	15710412
FZ3 (FZD3)	Frizzled homolog 3		Axonal guidance, neural tube closure	16495441
KIF2A	Kinesin family member 2A		Axonal branching	12887924
KIF1B	Kinesin family member 1B	Charcot-Marie-Tooth neuropathy	Axonal vesicle transport	11389829
SOX6	Transcription factor SOX-6		Cortical interneuron development	19657336
TLE1	Transducin-like enhancer of split 1		Cortical neuron differentiation	16314515
TNIK	TRAF2 and NCK interacting kinase		Dendrite formation	20159449
UNC5C	Unc-5 homolog C (netrin receptor)		Spinal accessory motor neuron development	17543537

Delta-catenin (CTNND1) is involved in the development and maintenance of dendritic spines^{31,32}. Furthermore, CTNND1 interacts with presenilin-1 to inhibit production of β -amyloid peptide, the main constituent of amyloid plaques in the brains of individuals with Alzheimer's disease³³. MEF2D is a transcription factor that has been reported to contribute to neuronal survival in Parkinson's disease³⁴. BIM is a core factor in the neuron-specific JNK-mediated apoptotic pathway, and loss of BIM protects mice from ischemia-induced damage^{35,36}. The BIM splice isoforms that are regulated by TDP-43 have been functionally characterized; BIM_S antagonizes the pro-survival Bcl-2 family members more effectively than BIM_L, whereas BIM_{EL} is the least potent²². Production of the most cytotoxic BIM isoform (BIM_S) was increased in TDP-43 knockdown cells and in tissue from subjects with FTLTDP. However, analysis of other exons regulated by TDP-43 did not indicate a significant correlation with the splicing changes present in samples from subjects with FTLTDP (unpublished observation).

We also found that TDP-43 bound to long ncRNAs in highly sequence-specific manner in tissue from subjects with or without FTLTDP. The greatest increase in TDP-43 binding in samples from subjects with FTLTDP is seen in the *NEAT1* ncRNA. Expression of *NEAT1* and *MALAT1* is markedly increased in FTLTDP, which might be the primary cause of the increased association of TDP-43 with these RNAs. *NEAT1* functions in paraspeckle assembly¹¹ and the *MALAT1* ncRNA recruits splicing factors to nuclear speckles and affects phosphorylation of serine/arginine-rich splicing factor proteins³⁷. In conclusion, we have characterized the splicing function of TDP-43 and quantified the changes in its RNA binding in FTLTDP. Analysis of these RNAs in model organisms and in neurons containing TDP-43 inclusions will be required for full understanding of the mechanisms of loss or gain of TDP-43 function in the neurodegenerative process.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/natureneuroscience/>.

Accession codes. iCLIP and microarray data were deposited in the ArrayExpress archive and are accessible at E-MTAB-527 and E-MTAB-530.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENT

We thank K. Zarnack for advice and comments on the manuscript, F. Baralle, Y. Ayala and A.D. Ambrogio for HeLa knockdown RNA, and J. Hadfield and the genomic team at CRI for Illumina sequencing. This work was supported by the European Research Council (206726-CLIP), the MRC, Slovenian Research Agency (P2-0209, J2-2197, L2-1112, Z7-3665), Wellcome Trust and MRC Strategic Grant Award (089701/Z/09/Z), Motor Neuron Disease Association, Heaton-Ellis Trust and Psychiatry Research Trust.

AUTHOR CONTRIBUTIONS

J.R.T. carried out TDP-43 iCLIP, microarray and PCR experiments. M.B. carried out CELF2 iCLIP. T.C. and G.R. mapped the iCLIP sequence reads to genome, evaluated random barcodes, determined cross-link clusters and annotated the data. T.C. analyzed the reproducibility, sequence and positioning of TDP-43 cross-link sites and performed gene ontology analysis. B.R., A.L.N. and V.Z. prepared RNA from knockdown cells and brain tissue. T.H. selected, sampled and analyzed the brain samples. M.C. and M.K. analyzed splice-junction microarray data and generated the RNA splicing map. R.P. prepared the embryonic stem cells. S.C., C.E.S., B.Z., J.K. and J.U. supervised the project. J.R.T., T.C., B.R., J.K., C.E.S. and J.U. prepared the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/natureneuroscience/>.

Reprints and permissions information is available online at <http://www.nature.com/reprintsandpermissions/>.

- Buratti, E. & Baralle, F.E. Multiple roles of TDP-43 in gene expression, splicing regulation, and human disease. *Front. Biosci.* **13**, 867–878 (2008).
- Kuo, P.H., Doudeva, L.G., Wang, Y.T., Shen, C.K. & Yuan, H.S. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.* **37**, 1799–1808 (2009).
- Buratti, E. & Baralle, F.E. Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J. Biol. Chem.* **276**, 36337–36343 (2001).
- Sreedharan, J. *et al.* TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science* **319**, 1668–1672 (2008).
- Lagier-Tourenne, C., Polymenidou, M. & Cleveland, D.W. TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. *Hum. Mol. Genet.* **19**, R46–R64 (2010).
- Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).
- Igaz, L.M. *et al.* Expression of TDP-43 C-terminal fragments *in vitro* recapitulates pathological features of TDP-43 proteinopathies. *J. Biol. Chem.* **284**, 8516–8524 (2009).
- Johnson, B.S., McCaffery, J.M., Lindquist, S. & Gitler, A.D. A yeast TDP-43 proteinopathy model: exploring the molecular determinants of TDP-43 aggregation and cellular toxicity. *Proc. Natl. Acad. Sci. USA* **105**, 6439–6444 (2008).
- Voigt, A. *et al.* TDP-43-mediated neuron loss *in vivo* requires RNA-binding activity. *PLoS ONE* **5**, e12247 (2010).
- König, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* **17**, 909–915 (2010).
- Clemson, C.M. *et al.* An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol. Cell* **33**, 717–726 (2009).
- Cleveland, D.W. & Rothstein, J.D. From Charcot to Lou Gehrig: deciphering selective motor neuron death in ALS. *Nat. Rev. Neurosci.* **2**, 806–819 (2001).
- Ule, J. *et al.* CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212–1215 (2003).
- Ayala, Y.M. *et al.* TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J.* **30**, 277–288 (2011).
- Polymenidou, M. *et al.* Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat. Neurosci.* advance online publication, doi:10.1038/nn.2779 (27 February 2011).
- Sephton, C.F. *et al.* Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes. *J. Biol. Chem.* **286**, 1204–1215 (2011).
- Strong, M.J. *et al.* TDP43 is a human low molecular weight neurofilament (hNFL) mRNA-binding protein. *Mol. Cell. Neurosci.* **35**, 320–327 (2007).
- Ayala, Y.M., Misteli, T. & Baralle, F.E. TDP-43 regulates retinoblastoma protein phosphorylation through the repression of cyclin-dependent kinase 6 expression. *Proc. Natl. Acad. Sci. USA* **105**, 3785–3789 (2008).
- Buratti, E. *et al.* Nuclear factor TDP-43 can affect selected microRNA levels. *FEBS J.* **277**, 2268–2281 (2010).
- Kametani, F. *et al.* Identification of casein kinase-1 phosphorylation sites on TDP-43. *Biochem. Biophys. Res. Commun.* **382**, 405–409 (2009).
- Fiesel, F.C. *et al.* Knockdown of transactive response DNA-binding protein (TDP-43) downregulates histone deacetylase 6. *EMBO J.* **29**, 209–221 (2010).
- O'Connor, L. *et al.* Bim: a novel member of the Bcl-2 family that promotes apoptosis. *EMBO J.* **17**, 384–395 (1998).
- Cléry, A., Blatter, M. & Allain, F.H. RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* **18**, 290–298 (2008).
- Witten, J.T. & Ule, J. Understanding splicing regulation through RNA splicing maps. *Trends Genet.* published online, doi:10.1016/j.tig.2010.12.001 (11 January 2011).
- Black, D.L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**, 291–336 (2003).
- Hung, L.H. *et al.* Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. *RNA* **14**, 284–296 (2008).
- Rockman, M.V. & Wray, G.A. Abundant raw material for cis-regulatory evolution in humans. *Mol. Biol. Evol.* **19**, 1991–2004 (2002).
- Kashi, Y. & King, D.G. Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* **22**, 253–259 (2006).
- Legendre, M., Pochet, N., Pak, T. & Verstrepen, K.J. Sequence-based estimation of minisatellite and microsatellite repeat variability. *Genome Res.* **17**, 1787–1796 (2007).
- Wu, L.S. *et al.* TDP-43, a neuro-pathosignature factor, is essential for early mouse embryogenesis. *Genesis* **48**, 56–62 (2010).
- Matter, C., Pribadi, M., Liu, X. & Trachtenberg, J.T. Delta-catenin is required for the maintenance of neural structure and function in mature cortex *in vivo*. *Neuron* **64**, 320–327 (2009).
- Ishiyama, N. *et al.* Dynamic and static interactions between p120 catenin and E-cadherin regulate the stability of cell-cell adhesion. *Cell* **141**, 117–128 (2010).
- Kouchi, Z. *et al.* p120 catenin recruits adherins to gamma-secretase and inhibits production of A β peptide. *J. Biol. Chem.* **284**, 1954–1961 (2009).
- Yang, Q. *et al.* Regulation of neuronal survival factor MEF2D by chaperone-mediated autophagy. *Science* **323**, 124–127 (2009).
- Ness, J.M. *et al.* Selective involvement of BH3-only Bcl-2 family members Bim and Bad in neonatal hypoxia-ischemia. *Brain Res.* **1099**, 150–159 (2006).
- Becker, E.B. & Bonni, A. Pin1 in neuronal apoptosis. *Cell Cycle* **6**, 1332–1335 (2007).
- Tripathi, V. *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* **39**, 925–938 (2010).

ONLINE METHODS

iCLIP analysis. The iCLIP protocol was performed as described¹⁰, with the following modifications. SH-SY5Y neuroblastoma or H9 human embryonic stem cells were irradiated once with 150 mJ cm⁻² in a Stratlinker 2400 at 254 nm, brain tissue was dissociated in cold PBS and the suspension was cross-linked four times with 100 mJ cm⁻². TDP-43 was immunoprecipitated with protein A Dynabeads (Invitrogen) conjugated to rabbit-anti TDP-43 (Proteintech, 10782-2-AP). For iCLIP of CELF2, protein G Dynabeads conjugated to mouse anti-CELF2 (Sigma, C9367) were used. In both cases, the region corresponding to 55–100-kDa complexes was excised from the membrane to isolate the RNA. High-throughput sequencing using Illumina GA2 was done using 54 or 72 cycles (**Supplementary Table 1**). The barcode sequences corresponding to the individual experiment were as described (**Supplementary Table 1**). The random barcodes were registered and the barcodes were removed before mapping the sequences to the human genome sequence (version GRCh37/hg19) allowing one mismatch using Bowtie version 0.10.1 (command line: -a -m 1 -v 1).

Nucleocytoplasmic fractionation. We resuspended 50 mg of cross-linked SH-SY5Y cells or brain tissue in 1 ml of cold cytoplasmic lysis buffer (50 mM Tris-HCl, pH 7.4, 10 mM NaCl, 0.5% NP-40 (vol/vol), 0.25% Triton X-100 (vol/vol), 1 mM EDTA, 0.5% RNasin (vol/vol)). After rotation at 4 °C for 5 min, homogeniser was used to complete the lysis. After centrifugation at 4 °C, 3,000 g for 5 min, supernatant was collected and centrifugation was repeated at 4 °C, 10,000 g for 10 min. Supernatant was collected and 200 µl of 0.5% SDS (vol/vol), 0.25% sodium deoxycholate (vol/vol), 0.5 M NaCl was added; this was used as cytoplasmic fraction for CLIP or western blot analyses. The pellet (after 3,000 g spin) was resuspended in 1 ml of CLIP lysis buffer¹⁰, sonicated and used as nuclear fraction for CLIP or western blot analyses. The radioactive CLIP protein–RNA complexes on nitrocellulose membrane were quantified using phosphorimager.

Sequence analyses. Analysis of reproducibility of cross-link sites was done as described¹⁰. Identification of significant iCLIP cross-link clusters and *z*-score analysis of enriched pentamers was done as described³⁸. For analysis of cross-linking to UG repeats of different lengths, the number of UG repeats in transcribed genome with a cross-link site in at least one of the last ten nucleotides was determined for each dataset. Enrichment was represented by the ratio of the number of UG repeats with a cross-link site in real data compared to randomized data. The pentamer with the highest *z*-scores in TDP-43 iCLIP (UGUGU) was used to assess the positions at which the TDP-43 binding sites were enriched relative to the positions of cross-link sites. The genomic sequences at positions within 100 nucleotides of the iCLIP cross-link site were filtered to retain only the sequence containing the pentamer. The sequences were aligned at each iCLIP cross-link site, and the number of sequences containing the pentamer was determined at each nucleotide position between –100 and +100 nt. Enrichment was represented by the ratio of the number of cross-link sites containing the pentamer in real data compared to randomized data.

siRNA knockdown. Acell Smartpool oligos against *TARDBP* (E-012394-00-0005, Dharmacon) were transfected according to the manufacturer's instructions (Acell, Dharmacon). A-012394-14, 5'-GGUCAAGCAUGGAUUCUA-3'; A-012394-15, 5'-GUCUCAAGUCAAAUGGAUU-3'; A-012394-16, 5'-GUGUUAAGUGAAAUGAUAC-3'; A-012394-17, 5'-GGGUGAUGUUCUAAUUUACA-3'. Acell nontargeting pool oligos (D-001910-10-05, Dharmacon) were used as controls.

Western blotting. SH-SY5Y cells were lysed in iCLIP lysis buffer and sonicated. To digest DNA, 10 µl of DNase (Ambion) was added per 1 ml of lysate and the mixture was incubated at 37 °C for 3 min. The samples were centrifuged and the supernatants were collected for western analysis. The protein concentration was determined using Lowry's assay (Bio-RAD). Equal amounts of protein were loaded on 4–12% Bis-Tris gels (Invitrogen) and transferred to a nitrocellulose membrane. The membrane was incubated with rabbit anti-TDP-43 anti-

body (Proteintech, 10782-2-AP; 1:1,000). Rabbit anti-GAPDH (Cell Signaling; 1:5,000) or anti-actin antibodies were used as a loading control.

Splice-junction microarray. A total of 6 samples were used, 3 from the siRNA control and 3 from knockdown. The high-resolution AltSplice splice-junction microarrays were produced by Affymetrix, the cDNA samples were prepared using the GeneChip WT cDNA Synthesis and Amplification Kit (Affymetrix) and data were analyzed with version 3 of ASPIRE (Analysis of Splicing Isoform REciprocity)¹⁰.

RT-PCR. Total RNA was extracted using the RNeasy Kit (Qiagen) and 200 ng of RNA was used for reverse transcription using Superscript III (Invitrogen) according to the manufacturer's instructions. For analysis of transcript levels, real-time PCR was performed using SYBR Green Fast PCR master mix (Applied Biosystems). For analysis of splicing, PCR was performed using Immomix (Bioline) using primers listed in **Supplementary Table 4**. The PCR products were visualized using QIAxcel capillary electrophoresis system (Qiagen). To calculate exon inclusion (*I*), the percentage of the peak representing exon inclusion was divided by the total percentage of peaks representing exon inclusion and skipping. Splicing change was calculated by subtracting the exon inclusion in the knockdown cells from the inclusion in wild-type cells (thus, a positive ΔI represents exons enhanced by TDP-43, and negative those silenced by TDP-43).

RNA splicing map. To analyze the effect of TDP-43 positioning on splicing regulation, we assessed the positioning of cross-link clusters in the proximity of the exon-intron boundaries of alternative exons and flanking exons, including 50 nucleotides of exonic and 500 nucleotides of intronic sequence. All cassette exons with identified splicing changes in knockdown SH-SY5Y cells and at least one significant cross-link cluster in the analyzed regions are shown in the RNA splicing map.

Gene ontology (GO) analysis. To make GO annotation exon-centric, we multiplied each gene annotation record to include all exons on the splicing microarray for the gene. For example, if gene X is annotated to GO term T, and there are probes for five exons in gene X, the expanded annotation would include five records (gene X, exon 1), ... (gene X, exon 5) that would all be assigned to term T. Binomial distribution was then used to compute the probability ($\Pr(K \geq k)$) of observing *k* or more exons annotated to GO term among all *n* exons identified as being differentially expressed in the experiment. If *m* is the number of all exons assigned to a GO term, and *N* is the total number of exons measured in

the experiment (control set), then $\Pr(K \geq k) = \sum_{i=k}^n \binom{n}{i} p^i (1-p)^{n-i}$, for $p = \frac{m}{N}$.

Only molecular processes and term levels greater than 2 were evaluated (to avoid the terms that are too general); the redundancy of each term with all other terms with lower *P* value was evaluated. For **Table 1**, all terms with *P* < 0.04, redundancy < 50%, and at least three matching exons are shown. In addition, three terms with redundancy higher than 50% are shown: neural tube closure, homeostasis of number of cells and chordate embryonic development.

Brain tissue. Frozen brain tissues were obtained from the MRC London Neurodegenerative Diseases Brain Bank (King's College London). Consent for autopsy and neuropathological assessment was obtained from all subjects in accordance with the guidelines for organ donation approved by The Joint South London and Maudsley and the Institute of Psychiatry Research Ethics Committee (#231/01) and work with human brain tissue was approved by the Cambridgeshire Research Ethics Committee (07/Q0108/80). The superior sulcal margin of the middle temporal gyrus (Brodmann area 21) at the level of the mamillary body was used. As healthy controls, we isolated samples from donated brains of individuals with no clinical history of a neurodegenerative condition and lacking any signs of disease neuropathology.

38. Wang, Z. *et al.* iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLoS Biol.* **8**, e1000530 (2010).