nature
neuroscience

Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43

Magdalini Polymenidou[1,2,6], Clotilde Lagier-Tourenne[1,2,6], Kasey R Hutt[2,3,6], Stephanie C Huelga[2,3], Jacqueline Moran[1,2], Tiffany Y Liang[2,3], Shuo-Chien Ling[1,2], Eveline Sun[1,2], Edward Wancewicz[4], Curt Mazur[4], Holly Kordasiewicz[1,2], Yalda Sedaghat[4], John Paul Donohue[5], Lily Shiue[5], C Frank Bennett[4], Gene W Yeo[2,3] & Don W Cleveland[1,2]

We used cross-linking and immunoprecipitation coupled with high-throughput sequencing to identify binding sites in 6,304 genes as the brain RNA targets for TDP-43, an RNA binding protein that, when mutated, causes amyotrophic lateral sclerosis. Massively parallel sequencing and splicing-sensitive junction arrays revealed that levels of 601 mRNAs were changed (including *Fus (Tls)*, progranulin and other transcripts encoding neurodegenerative disease–associated proteins) and 965 altered splicing events were detected (including in sortilin, the receptor for progranulin) following depletion of TDP-43 from mouse adult brain with antisense oligonucleotides. RNAs whose levels were most depleted by reduction in TDP-43 were derived from genes with very long introns and that encode proteins involved in synaptic activity. Lastly, we found that TDP-43 autoregulates its synthesis, in part by directly binding and enhancing splicing of an intron in the 3′ untranslated region of its own transcript, thereby triggering nonsense-mediated RNA degradation.

Amyotrophic lateral sclerosis (ALS) is an adult-onset disorder in which premature loss of motor neurons leads to fatal paralysis. Most cases of ALS are sporadic, with only 10% of affected individuals having a familial history. A breakthrough in understanding ALS pathogenesis was the discovery that TDP-43, which in the normal setting is primarily nuclear, mislocalizes and forms neuronal and glial cytoplasmic aggregates in ALS[1,2], frontotemporal lobar degeneration (FTLD), and in Alzheimer's and Parkinson's diseases (reviewed in ref. 3). Dominant mutations in TDP-43 were subsequently identified as being causative in sporadic and familial ALS cases and in rare individuals with FTLD[4–7]. At present, it is unresolved as to whether neurodegeneration is a result of a loss of TDP-43 function or a gain of toxic property or a combination of the two. However, a notable feature of TDP-43 pathology is TDP-43 nuclear clearance in neurons containing cytoplasmic aggregates, consistent with pathogenesis being driven, at least in part, by a loss of TDP-43 nuclear function[1,7].

Several lines of evidence suggest an involvement of TDP-43 in multiple steps of RNA metabolism, including transcription, splicing or transport of mRNA[3], as well as microRNA metabolism[8]. Misregulation of RNA processing has been described in a growing number of neurological diseases[9]. The recognition of TDP-43's role in neurodegeneration and the recent identification of ALS-causing mutations in FUS/TLS[10,11], another RNA/DNA binding protein, has reinforced a crucial role for RNA processing regulation in neuronal integrity. However, a comprehensive protein-RNA interaction map for TDP-43 and identification of post-transcriptional events that may be crucial for neuronal survival remain to be established.

A common approach for identifying specific RNA-binding protein targets or aberrantly spliced isoforms related to disease has been through selection of candidate genes. However, recent advances in DNA sequencing technology have provided powerful tools for exploring transcriptomes at remarkable resolution[12]. Moreover, cross-linking, immunoprecipitation and high-throughput sequencing (CLIP-seq or HITS-CLIP) experiments have found that a single RNA-binding protein can have previously unrecognized roles in RNA processing and affect many alternatively spliced transcripts[13–15]. We used these approaches to identify a comprehensive TDP-43 protein-RNA interaction map in the CNS. After depletion of TDP-43 *in vivo*, RNA sequencing and splicing-sensitive microarrays were used to determine that TDP-43 is crucial for maintaining normal levels and splicing patterns of >1,000 mRNAs. The most downregulated of these TDP-43–dependent RNAs have pre-mRNAs with very long introns that contain multiple TDP-43–binding sites and encoded proteins related to synaptic activity. The nuclear TDP-43 clearance widely reported in TDP-43 proteinopathies[1,7] will lead to a disruption of this role on long RNAs that are preferentially expressed in brain, thereby contributing to neuronal vulnerability.

RESULTS

Protein-RNA interaction map of TDP-43 in mouse brain

We used CLIP-seq to identify *in vivo* RNA targets of TDP-43 in adult mouse brain (**Supplementary Fig. 1a**). After ultraviolet irradiation to stabilize *in vivo* protein-RNA interactions, we immunoprecipitated TDP-43 with a monoclonal antibody[16] that had a higher immunoprecipitation efficiency than any of the commercial antibodies tested (**Supplementary Fig. 1b**). Complexes representing the expected molecular weight of a single molecule of TDP-43 bound to its target RNAs were excised (**Fig. 1a**) and sequenced. We also observed lower mobility protein-RNA complexes whose abundance was reduced by increased nuclease digestion. Immunoblotting of the same immunoprecipitated samples before radioactive labeling of the target RNAs revealed that TDP-43 protein was a component of both the ~43-kDa complex and more slowly migrating complexes (**Fig. 1a**).
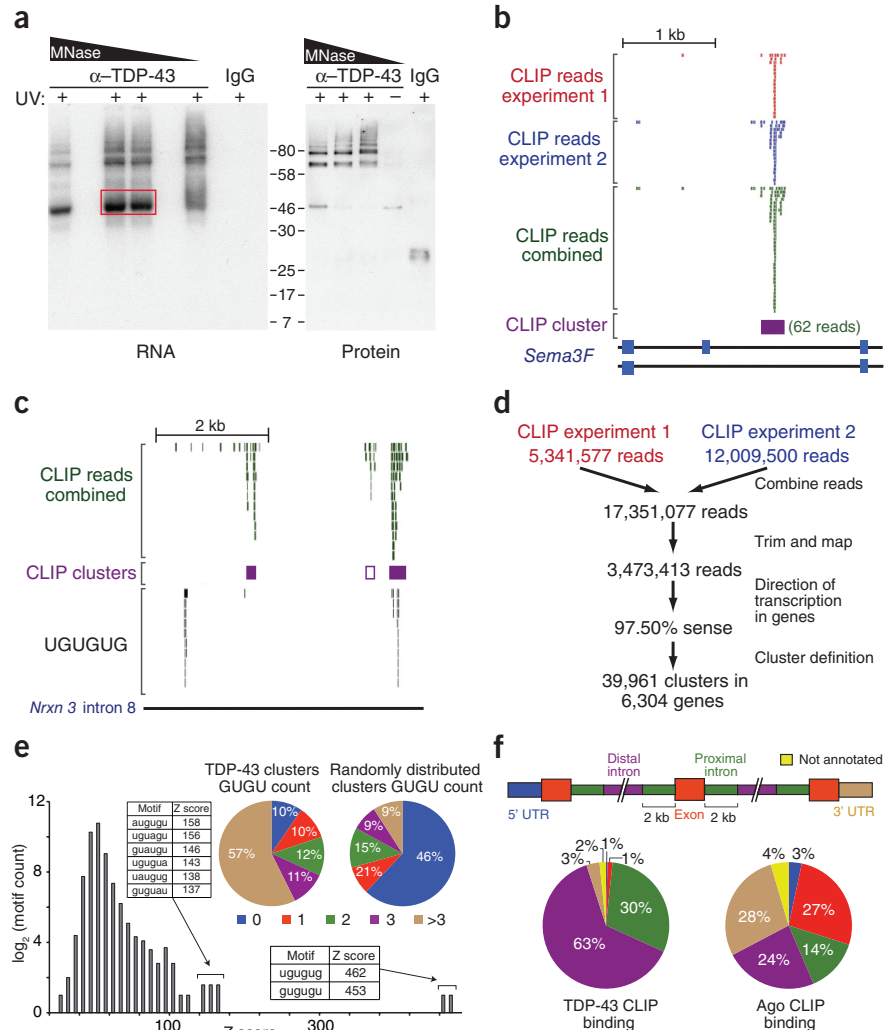
We performed two independent experiments and obtained 5,341,577 and 12,009,500 36-bp sequence reads. Mapped reads of both experiments were predominantly in protein-coding genes, with ~97% of them oriented in the direction of transcription, confirming little DNA contamination. The positions of mapped reads from both experiments were highly consistent, as exemplified by TDP-43 binding on the semaphorin 3F transcript (**Fig. 1b**).

We used a cluster-finding algorithm with gene-specific thresholds that accounted for pre-mRNA length and variable expression levels[15,17] to identify TDP-43–binding sites from clusters of sequence reads (**Fig. 1b**), with a conservative threshold (the number of reads mapped to a cluster had to exceed the expected number by chance at $P < 0.01$). This stringent definition will miss some true binding sites, but was intentionally chosen to identify the strongest bound sites while minimizing false positives. Indeed, additional probable binding sites could be identified by inspection of reads mapped to specific RNAs (see **Fig. 1c**). Moreover, similarly defined clusters from the low mobility complexes (**Supplementary Fig. 1b**) showed a 92% overlap with those from the monomeric complexes (**Fig. 1a**), consistent with the reduced mobility complexes comprising multiple TDP-43s (or other RNA-binding proteins) bound to a single RNA.

Genome-wide comparison of our replicate experiments revealed (**Supplementary Fig. 1c**) that the vast majority (90%) of TDP-43–binding sites in experiment 1 overlapped with those in experiment 2 (compared with an overlap of only 8% ($P \approx 0$, $Z = 570$) when clusters were randomly distributed across the length of the pre-mRNAs containing them). Combining the mapped sequences yielded 39,961 clusters, representing binding sites of TDP-43 in 6,304 annotated protein-coding genes, approximately 30% of the murine transcriptome (**Fig. 1d**). We computationally sampled reads (in 10% intervals) from the CLIP sequences and found a clear logarithmic relationship (**Supplementary Fig. 1d**), from which we calculated that our current dataset contains ~84% of all TDP-43 RNA targets in the mouse brain. Comparison with the mRNA targets identified from primary rat neuronal cells[18] by RNA

Figure 1 TDP-43 binds distal introns of pre-mRNA transcripts through UG-rich sites *in vivo*. (**a**) Autoradiograph of TDP-43-RNA complexes trimmed by different concentrations of micrococcal nuclease (MNase, left). Complexes in red box were used for library preparation and sequencing. Immunoblot showing TDP-43 in ~46 kDa and higher molecular weight complexes dependent on ultraviolet (UV) crosslinking (right). (**b**) Example of a TDP-43–binding site (CLIP cluster) on Semaphorin 3F defined by overlapping reads from two independent experiments surpassing a gene-specific threshold. (**c**) University of California Santa Cruz (UCSC) Genome Browser screenshot of neurexin 3 intron 8 (mm8; chr12:89842000–89847000) displaying three examples of TDP-43–binding modes. The right-most CLIP cluster represents a canonical binding site coinciding GU-rich sequence motifs, whereas the left-most cluster lacks GU-rich sequences and a region containing multiple GU repeats showed no evidence of TDP-43 binding. The second CLIP cluster (middle purple-outlined box) with weak binding was found only when relaxing cluster-finding algorithm parameters. (**d**) Flow-chart illustrating the number of reads analyzed from both CLIP-seq experiments. (**e**) Histogram of *Z* scores indicating the enrichment of GU-rich hexamers in CLIP-seq clusters compared with equally sized clusters, randomly distributed in the same pre-mRNAs. Sequences and *Z* scores of the top eight hexamers are indicated. Pie charts enumerate clusters containing increasing counts of $(GU)_2$ compared with randomly distributed clusters ($P \approx 0$, $\chi^2 = 21,662$). (**f**) Pre-mRNAs were divided into annotated regions (top). Distribution of TDP-43 (left) or previously published Argonaute-binding sites[21] as a control (right) showed preferential binding of TDP-43 in distal introns.

immunoprecipitation (RIP, an approach with the serious caveat that the absence of cross-linking allows re-association of RNAs and RNA-binding proteins after cell lysis, as previously documented[19]) revealed 2,672 of the genes with CLIP-seq clusters in common. As expected from our CLIP-seq analysis in whole brain, we found strong representation of neuronal (see below) and glial mRNA targets, including glutamate transporter 1 (*Glt1*), myelin-associated glycoprotein (*Mag*) and myelin oligodendrocyte glycoprotein (*Mog*).

TDP-43 binds GU-rich distal intronic sites

Sequence motifs enriched in TDP-43–binding sites were determined by comparing sequences in clusters with randomly selected regions of similar sizes in the same protein-coding genes. *Z* score statistics revealed that the most significantly enriched hexamers consisted of GU-repeats ($Z > 450$), consistent with published *in vitro* results[20], or a GU-rich motif interrupted by a single adenine ($Z = 137$–158) (**Fig. 1e**). The majority (57%) of clusters contained at least four GUGU elements, as compared with only 9% when equally sized clusters were randomly placed in the same pre-mRNAs (**Fig. 1e**). Furthermore, the number of GUGU tetramers correlated with the strength of binding, as estimated by the relative number of reads in each cluster per gene compared with all clusters in other genes (**Supplementary Fig. 2a**). Nevertheless, genome-wide analysis revealed that GU-rich repeats were neither necessary nor sufficient to specify a TDP-43–binding site. One example is the left-most binding site in neurexin 3 (**Fig. 1c**), which does not have a GU motif, whereas a GU-rich motif 2-kb upstream of it was not bound by TDP-43. In fact, only ~3% of all transcribed 300 nucleotide stretches containing more than three GUGU tetramers contained TDP-43 clusters by CLIP-seq, indicating that TDP-43 target genes cannot be identified by simply scanning nucleotide sequences for GU-rich regions.
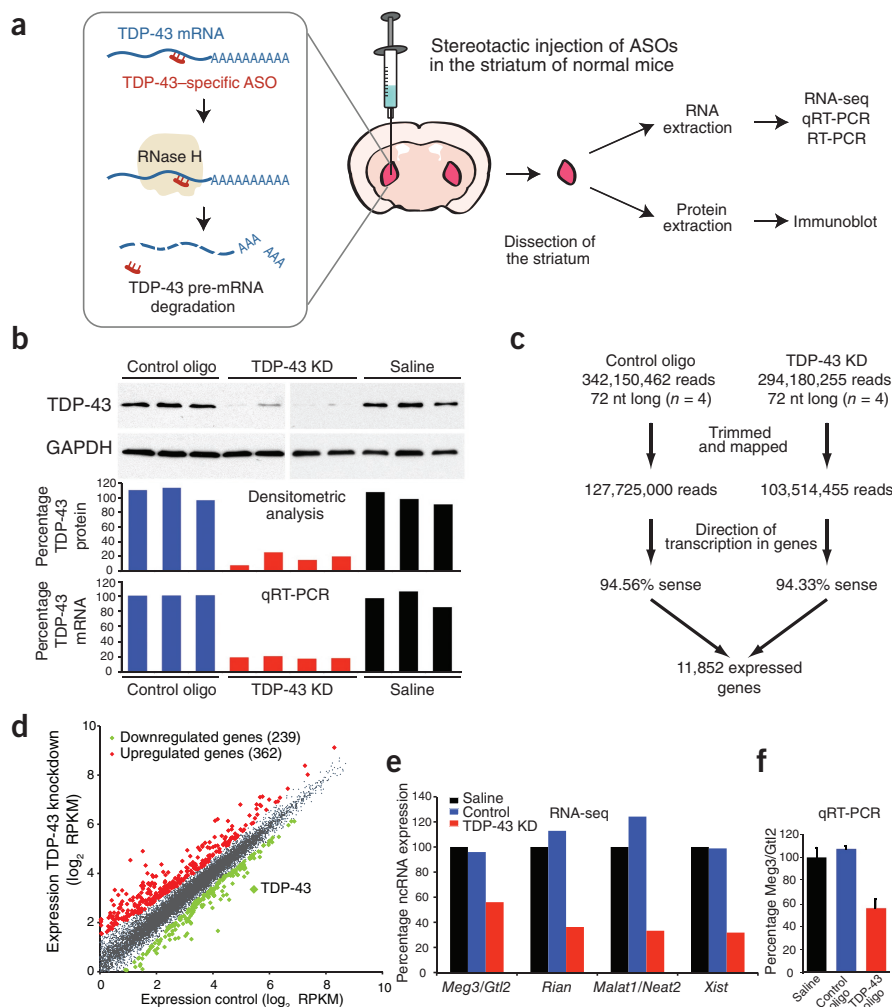
Although the vast majority (93%) of TDP-43 sites were in introns, we identified a binding preference of TDP-43 with most (63%)

intronic clusters being >2 kb from the nearest exon-intron boundary (**Fig. 1f**). This number rose to 82% for clusters >500 bases from the nearest exon-intron boundary (**Supplementary Fig. 2b**). Such distal intronic binding is in sharp contrast with published RNA-binding maps for tissue-specific RNA binding proteins involved in alternative splicing, such as *Nova* or *Fox2*[14,15]. The same analysis on published data in mouse brain for the Argonaute proteins[17,21], which are recruited by microRNAs to the 3′ ends of genes in metazoans[17,21], showed a substantially different pattern of binding. Only 24% (or 30%) of Argonaute clusters resided within 2 kb (or 500 bases) of the nearest exon-intron boundary, whereas 28% were in 3′ untranslated regions (3′ UTRs) (**Fig. 1f** and **Supplementary Fig. 2b**). This prominent concentration of Argonaute binding near 3′ ends is in stark contrast with the uniform distribution of TDP-43–binding sites across the length of pre-mRNAs (**Supplementary Fig. 2c**).

RNAs altered after *in vivo* TDP-43 depletion in mouse brain

To identify the contribution of TDP-43 in maintaining levels and splicing patterns of RNAs, we injected two antisense oligonucleotides (ASOs) directed against TDP-43 and a control ASO with no target in the mouse genome into the striatum of normal adult mice (**Fig. 2a**). Striatum is a well-defined structure that is amenable to accurate dissection and isolation, with TDP-43 expression levels being comparable to other brain regions. Stereotactic injections of ASOs that target TDP-43, control ASO or saline were performed in three groups of age- and sex-matched adult C57BL/6 mice and were tolerated with minimal effects

Figure 2 *In vivo* depletion of TDP-43 in mouse brain with ASOs. (**a**) Strategy for depletion of TDP-43 in mouse striatum. TDP-43–specific or control ASOs were injected into the striatum of adult mice. TDP-43 mRNA was degraded via endogenous RNase H digestion, which specifically recognizes ASO–pre-mRNA (DNA/RNA) hybrids. Mice were killed after 2 weeks and striata were dissected for RNA and protein extraction. (**b**) Semi-quantitative immunoblot demonstrating substantial depletion of TDP-43 protein to 20% of controls in TDP-43 ASO–treated mice compared with saline- and control ASO–treated mice. qRT-PCR showed similar TDP-43 depletion at the mRNA level. KD, knockdown. (**c**) Flow-chart illustrating the number of reads sequenced and aligned from the RNA-seq experiments. (**d**) Differentially regulated genes identified by RNA-seq analysis. Scatter-plot revealed the presence of 362 and 239 genes (diamonds) that were significantly up- (red) or down- (green) regulated after TDP-43 depletion. RNA-seq analysis confirmed that TDP-43 levels were reduced to 20% of control animals. (**e**) Normalized expression (based on RPKM values from RNA-seq) of four noncoding RNAs that were TDP-43 targets and were downregulated after TDP-43 depletion. (**f**) Quantitative RT-PCR validation of noncoding RNA Meg3/Gtl2.

on the survival of the animals. Mice were killed after 2 weeks and total RNA and protein from striata were isolated (**Fig. 2a**). Samples treated with TDP-43 ASO showed a significant and reproducible reduction of TDP-43 RNA and protein to approximately 20% of normal levels when compared with controls ($P < 3 \times 10^{-5}$; **Fig. 2b**).

To explore the effects of TDP-43 down-regulation on its target RNAs, we converted poly-A–enriched RNAs from four biological replicates of TDP-43 or control ASO-treated, as well as three saline-treated animals, to cDNAs and sequenced them in a strand-specific manner[22], yielding an average of >25 million 72-bp reads per library. The number of mapped reads per kilobase of exon, per million mapped reads (RPKM) for each annotated protein-coding gene was determined to establish a metric of normalized gene expression[12]. Hierarchical clustering of gene expression values for the independent samples revealed high correlation ($R^2 = 0.96$) between biological replicates of each condition (TDP-43 and control ASO/saline) (**Supplementary Fig. 3a**). Notably, all control ASO–treated samples were clustered together, as were the samples from TDP-43 ASO–treated animals, consistent with TDP-43 reduction having an appreciable effect on regulation of gene expression.

Reads of each treatment group were combined, yielding greater than 100 million uniquely mapped reads per condition (**Fig. 2c**). Approximately 70% (11,852) of annotated protein-coding genes in mouse satisfied at least 1 RPKM in either condition. Statistical comparison revealed that 362 genes were significantly upregulated and 239 downregulated on reduction of TDP-43 protein ($P < 0.05$; **Fig. 2d** and **Supplementary Tables 1** and **2**). We found that TDP-43 itself was downregulated by RPKM analysis to 20% of the levels in control treatments, consistent with quantitative RT-PCR (qRT-PCR) measurements (**Fig. 2b**). RNAs unique to neurons (including doublecortin, the neuron-specific beta-tubulin and choline O-acetyl transferase (*Chat*)) or glia (including glial fibrillary acidic protein, myelin binding protein, *Glt1* and *Mag*) were highly represented in the RNA-seq data, confirming assessment of RNA levels in multiple cell types, as expected.

Of the set of ~242 literature-curated murine noncoding RNAs[23], expression of 4 increased and expression of 55 decreased by more than twofold on TDP-43 depletion ($P < 10^{-5}$; **Supplementary Table 3**). *Malat1/Neat2*, *Xist*, *Rian* and *Meg3* are noncoding RNA examples that were both decreased (**Fig. 2e,f**) and bound by TDP-43, consistent with TDP-43 having a direct role in regulating their expression.

TDP-43 binding to long pre-mRNAs sustains their levels

RNA-seq and CLIP-seq datasets were integrated by first ranking all 11,852 expressed genes by their degree of change on TDP-43 reduction compared with control treatment. For each group of 100 consecutively ranked genes (starting from the most upregulated gene), we determined the mean number of TDP-43 clusters. No enrichment in TDP-43 clusters in the upregulated genes was identified, indicating that their upregulation was likely an indirect consequence of TDP-43 loss.
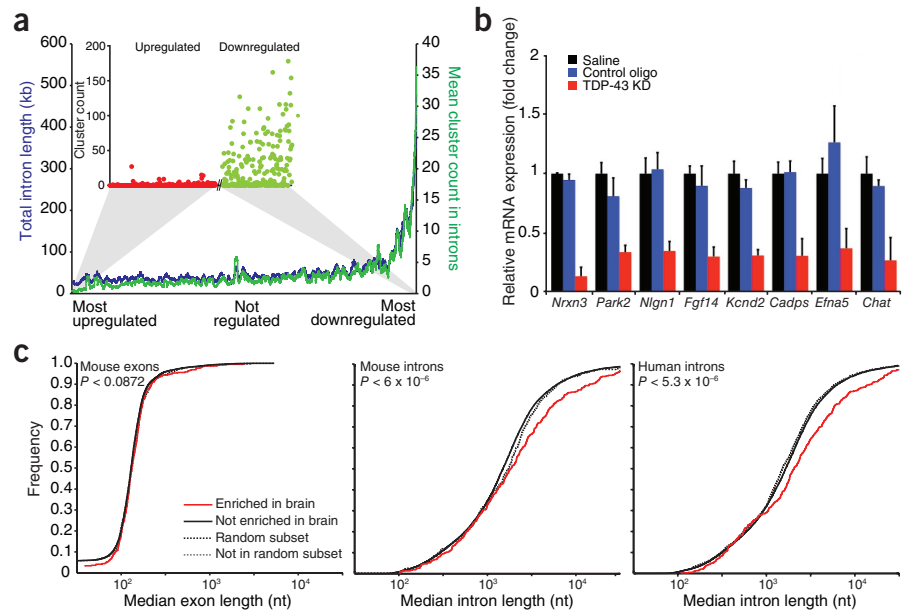

Figure 3 Binding of TDP-43 on long transcripts enriched in brain sustains their normal mRNA levels. (**a**) Correlation between RNA-seq and CLIP-seq data. Genes were ranked on their degree of regulation after TDP-43 depletion (*x* axis) and the mean number of intronic CLIP clusters found in the next 100 genes from the ranked list were plotted (*y* axis, green line). Similarly, the mean total intron length for the next 100 genes was plotted (*y* axis, red line). The cluster count for each upregulated gene (red dots) and each downregulated gene (green dots) was plotted using the same ordering (inset). (**b**) qRT-PCR for selected downregulated genes with long introns (except for Chat) revealed a significant reduction of all transcripts when compared with controls ($P < 8 \times 10^{-3}$). Error bars represent s.d. calculated in each group for 3–5 biological replicates. (**c**) Cumulative distribution plots comparing exon length (left) or intron length (middle) across mouse brain tissue–enriched genes (388 genes) and non-brain tissue–enriched genes (15,153 genes). Genes enriched in brain had significantly longer median intron length compared with genes not enriched in brain (right, solid red line and black lines, $P < 6.2 \times 10^{-6}$ by two-sample Kolmogorov Smirnov goodness-of-fit hypothesis test), whereas a random subset of 388 genes showed no difference in intron length (dashed lines). Similar analysis across human brain tissue–enriched genes (387 genes) and non-brain tissue–enriched genes (17,985 genes) also showed significantly longer introns in brain enriched genes (solid red and black lines, $P < 5.3 \times 10^{-6}$), whereas a random subset of 387 genes showed no difference in intron length (dashed lines).

Of the RNAs containing TDP-43 clusters, 96% were unaffected by TDP-43 depletion, suggesting either that other RNA-binding proteins compensate for TDP-43 loss or that the remaining 20% of TDP-43 protein suffices to regulate these transcripts. For the 239 RNAs downregulated after TDP-43 depletion, a striking enrichment of multiple TDP-43 binding sites was observed (**Fig. 3**). In fact, the 100 most downregulated genes contained an average of ~37 TDP-43–binding sites per pre-mRNA and 12 genes had more than 100 clusters (**Fig. 3a** and **Supplementary Table 2**). We did not observe this bias for multiple TDP-43 binding sites if we randomized the order of genes (**Supplementary Fig. 4a**), if we ordered them by their expression levels (RPKM) in either treatment (**Supplementary Fig. 4b,c**), or if the Argonaute-binding sites were plotted on genes ranked by their expression pattern on TDP-43 depletion (**Supplementary Fig. 4d**). Furthermore, this trend was significant for TDP-43 clusters found in introns (**Fig. 3a**), but not in exons, 5′ or 3′ UTRs (**Supplementary Fig. 4e–g**).

To address whether TDP-43 binding enrichment in the downregulated genes could be attributed to intron size, we performed the same analysis on the ranked list, but calculated total (**Fig. 3a**) or mean (**Supplementary Fig. 4h**) intron size instead of cluster counts. We found that the most downregulated genes after TDP-43 reduction had exceptionally long introns that were more than sixfold longer (average of 28,707 bp; median length of 11,786 bp) than unaffected or upregulated genes (average of 4,532 bp, median length of 2,273 bp,

Table 1 The fraction of downregulated, but not upregulated TDP-43 targets increased with intron size

Mean intron length (kb)	Expressed genes	TDP-43 targets	Downregulated genes	Downregulated TDP-43 targets	Upregulated genes	Upregulated TDP-43 targets
0–1	2,566	510 (20%)	31	8 (26%)	100	7 (7%)
1–10	8,022	4,485 (56%)	80	47 (59%)	252	56 (32%)
10–100	1,238	1,027 (83%)	109	104 (95%)	10	3 (30%)
>100	26	26 (100%)	19	19 (100%)	0	0 (0%)
Total	11,852	6,048 (51%)	239	178 (74%)	362	66 (18%)

Number of TDP-43 targets among expressed, downregulated or upregulated genes, grouped by their mean intron length as indicated. The percentages in parentheses represent the fraction of TDP-43 targets found in each group upon TDP-43 depletion.

$P < 4 \times 10^{-18}$ by t test). Again, this correlation of downregulation with intron size was not observed for any of the control conditions mentioned above (**Supplementary Fig. 4a–g**). Indeed, the enrichment of TDP-43 binding can be largely attributed to intron size differences, as the number of TDP-43–binding sites per kilobase of intron length (cluster density; **Supplementary Fig. 4i**) was only slightly increased ($P < 0.022$) for downregulated versus unaffected or upregulated genes (0.072 sites per kb downregulated genes, 0.059 sites per kb other genes). Dividing all mouse protein-coding genes into four groups on the basis of mean intron length (<1 kb, 1–10 kb, 10–100 kb and >100 kb) confirmed that the fraction of TDP-43 targets increased (from 20% to 100%) with intron size (**Table 1**). Indeed, 83% of genes that contained average intron lengths of 10–100 kb and all 26 genes that contained >100-kb-long introns were direct targets of TDP-43.

A highly significant fraction (74%) of all downregulated genes were direct targets of TDP-43 in comparison with genes that were unchanged (52%, $P < 0.001$) or upregulated (18%, $P < 10^{-17}$) on TDP-43 depletion. Notably, all 19 downregulated genes with >100-kb-long introns were direct TDP-43 targets. In strong contrast, no genes in the same intron length category were upregulated on TDP-43 depletion and only 30% of upregulated genes with 10–100-kb-long introns were TDP-43 targets (**Table 1**). The crucial role of TDP-43 in maintaining the mRNA abundance of long intron-containing genes was also reflected by the downregulation after TDP-43 depletion of ~10% of genes with >10-kb-long introns, the large majority (123 of 128, 96%) of which are direct TDP-43 targets.

Gene ontology analysis showed that TDP-43 targets whose expression was downregulated after TDP-43 depletion were highly enriched for synaptic activity and function (**Supplementary Figs. 5** and **6** and **Supplementary Table 4**). Notably, several genes with long introns targeted by TDP-43 have crucial roles in synaptic function and have been implicated in neurological diseases, such as subunit 2A of the NMDA receptor (*Grin2a*), the ionotropic glutamate receptor 6 (*Grik2/GluR6*), the calcium-activated potassium channel alpha (*Kcnma1*), the voltage-dependent calcium channel (*Cacna1c*), and the synaptic cell-adhesion molecules neurexin 1 and 3 (*Nrxn1*, *Nrxn3*) and neuroligin 1 (*Nlgn1*). We analyzed a compendium of expression array data from different mouse organs and human tissues and found that genes preferentially expressed in brain have significantly longer introns ($P < 6 \times 10^{-6}$), but not exons (**Fig. 3c**). The length of these genes was not correlated with the size of the respective proteins and the prevalence of long introns was largely conserved between the corresponding mouse and human genes. Although binding of TDP-43 in long introns can be explained by the increased likelihood to contain UG repeats, the conservation through evolution of this particular gene structure (**Supplementary Fig. 7**) suggests that these exceptionally long introns contain important regulatory elements.

To validate the RNA-seq results, we analyzed a selection of brain-enriched TDP-43 targets containing long introns. Genome browser views of neurexin 3 (*Nrxn3*), Parkin 2 (*Park2*), neuroligin 1 (*Nlgn1*), fibroblast growth factor 14 (*Fgf14*), potassium voltage-gated channel subfamily D member 2 (*Kcnd2*), calcium-dependent secretion activator (*Cadps*) and ephrin-A5 (*Efna5*) revealed a scattered distribution of multiple TDP-43–binding sites across the full length of the pre-mRNA (**Supplementary Fig. 7a**), consistent with the results of the global analysis (**Fig. 3a**). qRT-PCR verified the TDP-43–dependent reduction of all the long transcripts that we tested (**Fig. 3b**). *Chat* had a median intron size of <10 kb, with TDP-43 clusters restricted to a single intronic site (**Supplementary Fig. 7a**). Nevertheless, qRT-PCR confirmed the RNA-seq result of a significant reduction in *Chat* levels after TDP-43 depletion ($P = 0.005$; **Fig. 3b**).
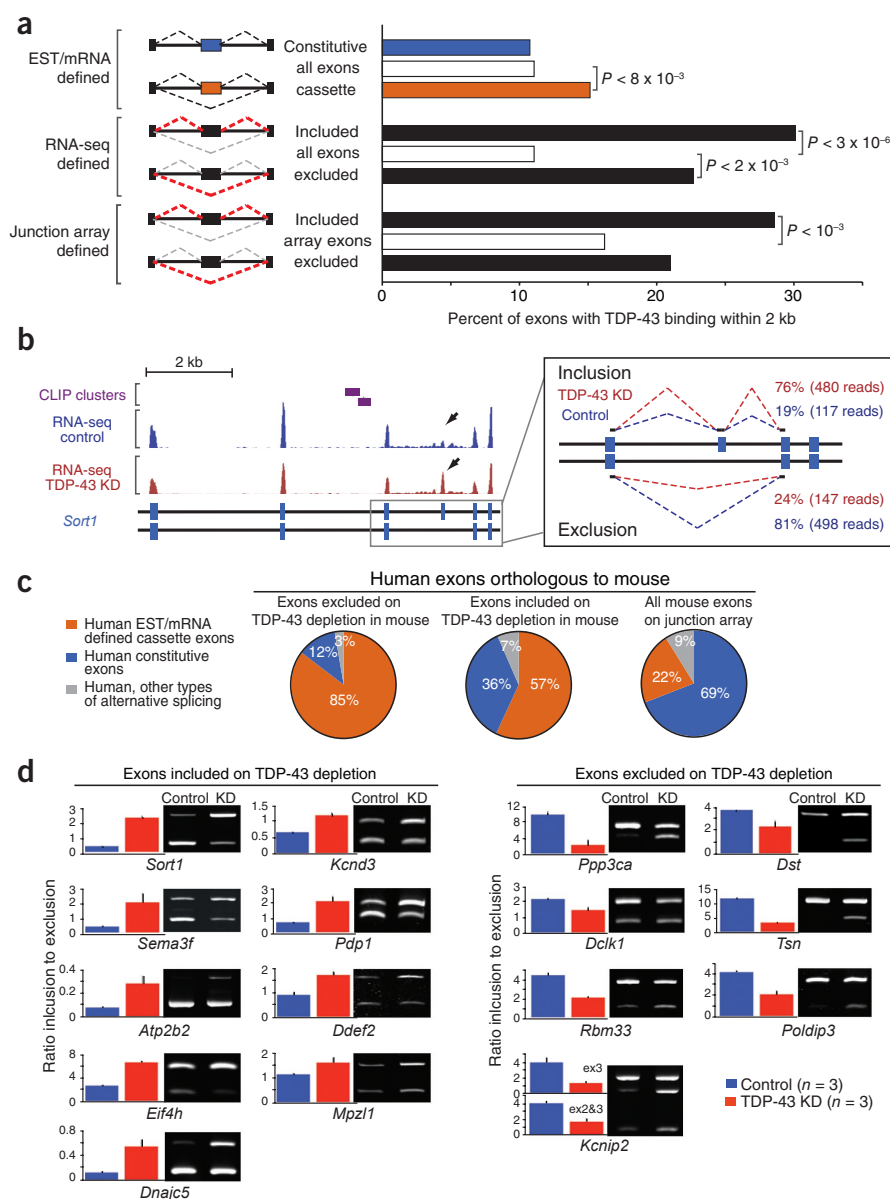
Only 18% of the upregulated genes were direct targets of TDP-43 (**Table 1**) and gene ontology analysis revealed an enrichment for genes involved in the inflammatory response (**Table 2**), suggesting that their differential expression is an indirect consequence of TDP-43 loss. However, of the 66 upregulated RNAs that contained CLIP-seq clusters, 29% harbored TDP-43–binding site(s) in their 3′ UTR, a percentage that is twofold higher than that of downregulated genes (**Supplementary Fig. 8**). This suggests that TDP-43 represses gene expression when bound to 3′ UTRs.

Table 2 Main functional categories enriched within the TDP-43–regulated genes

GO categories	Downregulated genes (239)			Upregulated genes (362)		
		GO term	Corrected P value		GO term	Corrected P value
Molecular function	GO:0006811	Ion transport	2.31×10^{-8}	GO:0006952	Defense response	2.28×10^{-20}
	GO:0007268	Synaptic transmission	1.79×10^{-7}	GO:0006954	Inflammatory response	4.97×10^{-12}
Cellular component	GO:0045202	Synapse	5.33×10^{-8}	GO:0000323	Lytic vacuole	1.70×10^{-10}
	GO:0005886	Plasma membrane	2.35×10^{-13}	GO:0005764	Lysosome	1.70×10^{-10}
Biological process	GO:0005216	Ion channel activity	1.49×10^{-11}	GO:0004197	Cysteine-type endopeptidase activity	6.00×10^{-4}
	GO:0022803	Passive transmembrane transporter activity	7.06×10^{-12}	GO:0003950	NAD+ ADP-ribosyltransferase activity	1.11×10^{-2}

Examples of the most significantly enriched Gene Ontology (GO) terms in the list of up- or downregulated genes on TDP-43 knockdown, as annotated. Downregulated genes, the majority of which were direct targets of TDP-43, were enriched for synaptic activity and function. In contrast, upregulated genes, most of which were not bound by TDP-43, encoded primarily inflammatory and immune response–related proteins. The P value indicated was corrected for multiple testing using the Benjamini-Hochberg method.

Figure 4 TDP-43 mediates alternative splicing regulation of its RNA targets. (**a**) Schematic representation of different exon classes defined by EST and mRNA libraries, RNA-seq or splicing-sensitive microarray data as indicated (left). Right, bar plot displaying the percentage of exons that contain TDP-43 clusters within 2 kb upstream and downstream of the exon-intron junctions. (**b**) Example of alternative splicing change on exon 18 of sortilin 1 analyzed using RNA-seq reads mapping to the exon body (left). Arrows depict increased density of reads in the TDP-43 knockdown samples compared with controls. 76% of the spliced-junction reads in the TDP-43 knockdown samples supported inclusion versus only 19% in the control oligo–treated samples (right). (**c**) Comparison of mouse cassette exons detected by splicing-sensitive microarrays to conserved exons in human orthologous genes. 85% and 57% of human exons corresponding to the excluded and included mouse exons, after TDP-43 depletion, respectively (left and middle pie charts), contained EST and mRNA evidence for alternative splicing. As a control, the percentage of human exons orthologous to all mouse exons represented on the splicing-sensitive array and that have alternative splicing evidence are shown in the right pie chart. (**d**) Semi-quantitative RT-PCR analyses of selected targets confirmed alternative splicing changes in TDP-43 knockdown samples compared with controls. Right, representative acrylamide gel pictures of RT-PCR products from control or knockdown adult brain samples. Quantification of splicing changes from three biological replicates per group (left); error bars represent s.d.

TDP-43 mediates alternative splicing of its mRNA targets

Although TDP-43–binding sites were enriched in distal introns (**Fig. 1f**), 11% (21,041 out of 190,161) of all mouse exons, including both constitutive and alternative exons, contained TDP-43–binding site(s) in a 2-kb window extending from the 5′ and 3′ exon-intron boundaries (**Fig. 4a**). Compared with all exons, TDP-43 clusters were significantly enriched ($P < 8 \times 10^{-3}$) around exons with transcript evidence for either alternative inclusion or exclusion (that is, cassette exons). Of the 8,637 known mouse cassette exons, 15.1% contained TDP-43–binding sites in the exon or intron within 2 kb of the splice sites. A splice index score for all exons, a measure similar to the 'percent spliced in' (or ψ) metric[24], was determined by the number of reads that mapped on exons as well as reads that mapped at exon junctions (**Fig. 4b**). This analysis resulted in the identification of 203 cassette exons that were differentially included (93) or excluded (110) ($P < 0.01$) when TDP-43 was depleted. Notably, *sortilin 1*, the gene encoding the receptor for progranulin[25,26], had the highest splice index score, with exon 18 exclusion requiring TDP-43 (**Fig. 4b**). Included exons ($P < 3 \times 10^{-6}$) and, to a lesser extent, excluded exons ($P < 2 \times 10^{-3}$) identified by RNA-seq were significantly enriched (~2.7-fold and ~2.0-fold, respectively) for TDP-43 binding when compared to all mouse exons (**Fig. 4a**). Only 33% of RNA-seq–verified TDP-43–regulated cassette exons had previous expressed sequence tag (EST)/mRNA evidence for alternative splicing, demonstrating that our approach identified previously unknown alternative splicing events.

As an independent method of identifying TDP-43–regulated exons, RNAs from the same ASO-treated animals were analyzed on custom-designed splicing-sensitive Affymetrix microarrays[27]. Using a conservative statistical cutoff, we detected 779 alternatively spliced events that significantly changed on TDP-43 depletion (**Supplementary Fig. 9**). Notably, included ($P < 10^{-3}$), but not excluded, exons ($P < 0.3$) were significantly enriched for TDP-43 binding (~1.8-fold and ~1.3-fold, respectively) when compared with the unchanged exons on the microarray (**Fig. 4a**), similar to the trend seen by RNA-seq. The combined RNA-seq and splicing-sensitive microarray data defined a set of 512 alternatively spliced cassette exons whose splicing was affected by loss of TDP-43. The majority of human orthologs of these murine exons (85% of those with excluded and 57% with included exons) had prior EST/mRNA evidence for alternative splicing (**Fig. 4c**).

Semi-quantitative RT-PCR on selected RNAs validated splicing alterations with more inclusion or exclusion after TDP-43 reduction (**Fig. 4d** and **Supplementary Fig. 10**). Notably, varying the extent of TDP-43 downregulation (between 40–80%) correlated with the magnitude of splicing changes (**Supplementary Fig. 11**). However, the majority of

altered splicing events observed on TDP-43 depletion do not have TDP-43 clusters within 2 kb of the splice sites, implicating longer-range interactions or indirect effects of TDP-43 through other splicing factors. Consistent with this latter hypothesis, we identified TDP-43 binding on pre-mRNAs of RNA-binding proteins, including *Fus/Tls*, *Ewsr1*, *Taf15*, *Adarb1*, *Cugbp1*, *RBFox2* (*Rbm9*), *Tia1*, *Nova 1* and *2*, *Mbnl*, and neuronal *Ptb* (or *Ptbp2*). After TDP-43 depletion, mRNA levels of *Fus/Tls* and *Adarb1* were reduced and exon 5 in the *Tia1* transcript was more included, whereas exon 10 in *Ptbp2* was more excluded (**Supplementary Table 5**).

TDP-43 autoregulation through binding on its 3′ UTR

We found TDP-43–binding sites in an alternatively spliced intron in the 3′ UTR of the TDP-43 pre-mRNA (**Fig. 5a**). This binding does not coincide with a long stretch of UG repeats, suggesting a lower strength of binding (**Supplementary Fig. 2a**), consistent with a recent report[28]. TDP-43 mRNAs spliced at this site (**Fig. 5a**) are predicted to be substrates for nonsense-mediated RNA decay (NMD), a process that targets mRNAs for degradation when exon-junction complexes (EJCs) deposited during splicing, located 3′ of the stop codon, are not displaced during the pioneer round of translation[29]. In contrast, TDP-43 mRNAs with an unspliced 3′ UTR would not have such a premature termination codon and should escape NMD. This TDP-43 binding implies autoregulatory mechanisms reminiscent to those reported for other RNA-binding proteins[30,31]. Indeed, expression in mice of a TDP-43–encoding transgene without the regulatory 3′ UTR (E.S., S.-C.L. and D.W.C., unpublished observations) led to significant reduction of endogenous TDP-43 mRNA and protein ($P < 4 \times 10^{-3}$; **Fig. 5b,c**) in the CNS.

To identify the molecular basis of this mechanism, we generated HeLa cells in which *GFP-myc–TDP-43–HA* mRNA lacking introns and 3′ UTR was transcribed from a single copy, tetracycline-inducible transgene inserted at a predefined locus by site-directed (Flp) recombinase[16]. After 24 or 48 h of *GFP-myc–TDP-43–HA* induction, a substantial reduction of endogenous TDP-43 protein was observed, accompanied by accumulation of a shorter, ~30-kDa product (**Fig. 5d**) recognized by four different TDP-43–specific antibodies. Although this ~30-kDa band could be derived from the transgene encoding TDP-43, it was not recognized by antibodies to myc or HA and its size is compatible with the endogenous TDP-43 isoform 3. Using qRT-PCR with primers spanning the exon junctions of TDP-43 isoform 3, we found a ~100-fold increase of the spliced isoform 3 on overexpression of GFP-myc–TDP-43–HA protein (**Fig. 5e**).

To test whether TDP-43 drives splicing of its pre-mRNA through binding to its 3′ UTR, we cloned a long unspliced version (containing the TDP-43–binding sites) and a short spliced version (without TDP-43–binding
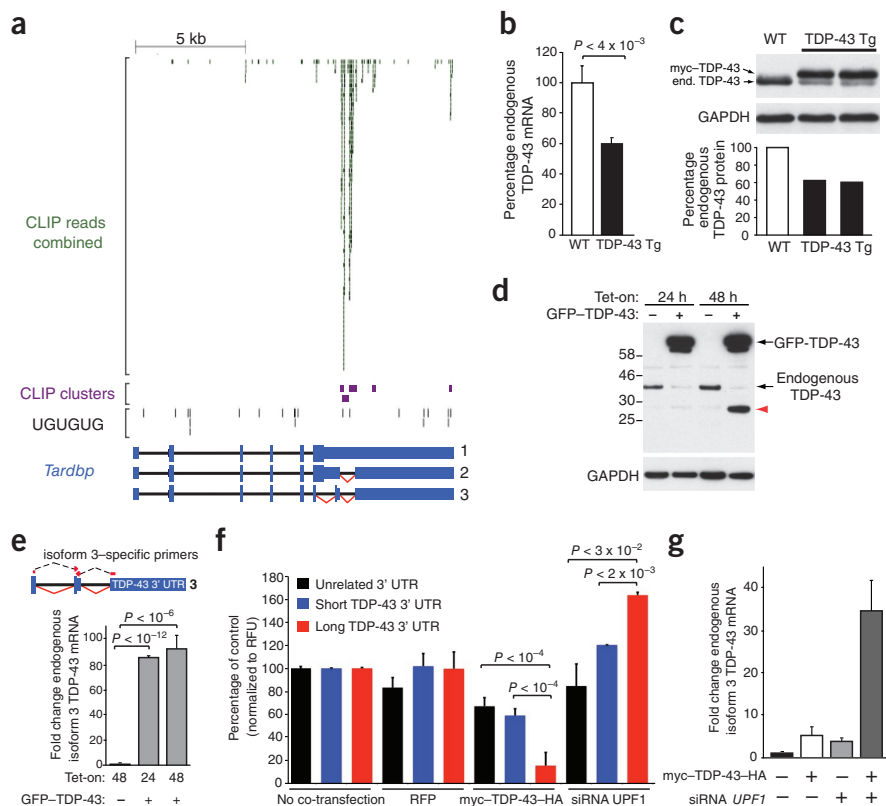

Figure 5 Autoregulation of TDP-43 through binding on the 3′ UTR of its own transcript. (**a**) CLIP-seq reads and clusters on TDP-43 transcript showing binding mainly in an alternatively spliced part of the 3′ UTR, lacking long uninterrupted UG repeats. (**b**) qRT-PCR showing ~50% reduction of endogenous TDP-43 mRNA in transgenic mice overexpressing human myc–TDP-43 not containing introns and 3′ UTR. (**c**) Immunoblots confirming the reduction of endogenous TDP-43 protein (upper panel) to 50% of control levels (by densitometry, lower panel) in response to human myc–TDP-43 overexpression. (**d**) Immunoblot showing reduction of endogenous TDP-43 protein in HeLa cells on tetracycline induction of a transgene encoding *GFP-myc–TDP-43–HA* (annotated GFP–TDP-43). The ~30-kDa product accumulating after 48 h (red arrow) was immunoreactive with four TDP-43–specific antibodies. (**e**) qRT-PCR using primers spanning the junctions of TDP-43 isoform 3 (upper panel), showed ~100-fold increase in response to tetracycline induction of GFP–TDP-43. TDP-43 isoform 3 was present at very low levels before tetracycline induction. (**f**) Luciferase assays showing significant reduction of relative fluorescence units (RFUs) in cells expressing renilla luciferase under the control of long TDP-43 3′ UTR when co-transfected with *myc–TDP-43–HA*. Cells treated with siRNA against *UPF1* showed a significant increase of RFUs when expressing Luciferase with the long TDP-43 3′ UTR, but not in controls. (**g**) qRT-PCR scoring levels of endogenous TDP-43 isoform 3 in HeLa cells transiently transfected with *myc–TDP-43–HA*, *UPF1* siRNA or both simultaneously. TDP-43 isoform 3 was increased in response to elevated TDP-43 protein levels, blocking of NMD (by *UPF1* siRNA) and there was a synergistic effect in the combined condition.

sites) of the TDP-43 3′ UTR downstream of the stop codon of a renilla luciferase gene (**Supplementary Fig. 12a**). Both unspliced and spliced 3′ UTRs were determined to be present in brain RNAs from mouse and human CNSs (**Supplementary Fig. 12a**). Both variants, as well as an unaltered luciferase reporter, were transfected into HeLa cells along with plasmids driving either increased TDP-43 expression or red fluorescent protein (RFP; **Supplementary Fig. 12b**). Increased levels of TDP-43 protein led to a significant reduction of luciferase produced from the gene carrying the long, intron-containing TDP-43 3′ UTR when compared with the short or unrelated 3′ UTR ($P < 10^{-4}$; **Fig. 5f**). Moreover, co-transfection of the reporters with siRNAs targeting *UPF1* (**Supplementary Fig. 12c**), an essential component that marks an NMD substrate for degradation[32], enhanced luciferase produced by the intron-containing 3′ UTR by ~1.5-fold, indicating that UPF1-dependent degradation of this mRNA occurred (**Fig. 5f**). Lastly, the endogenous spliced isoform 3 of TDP-43

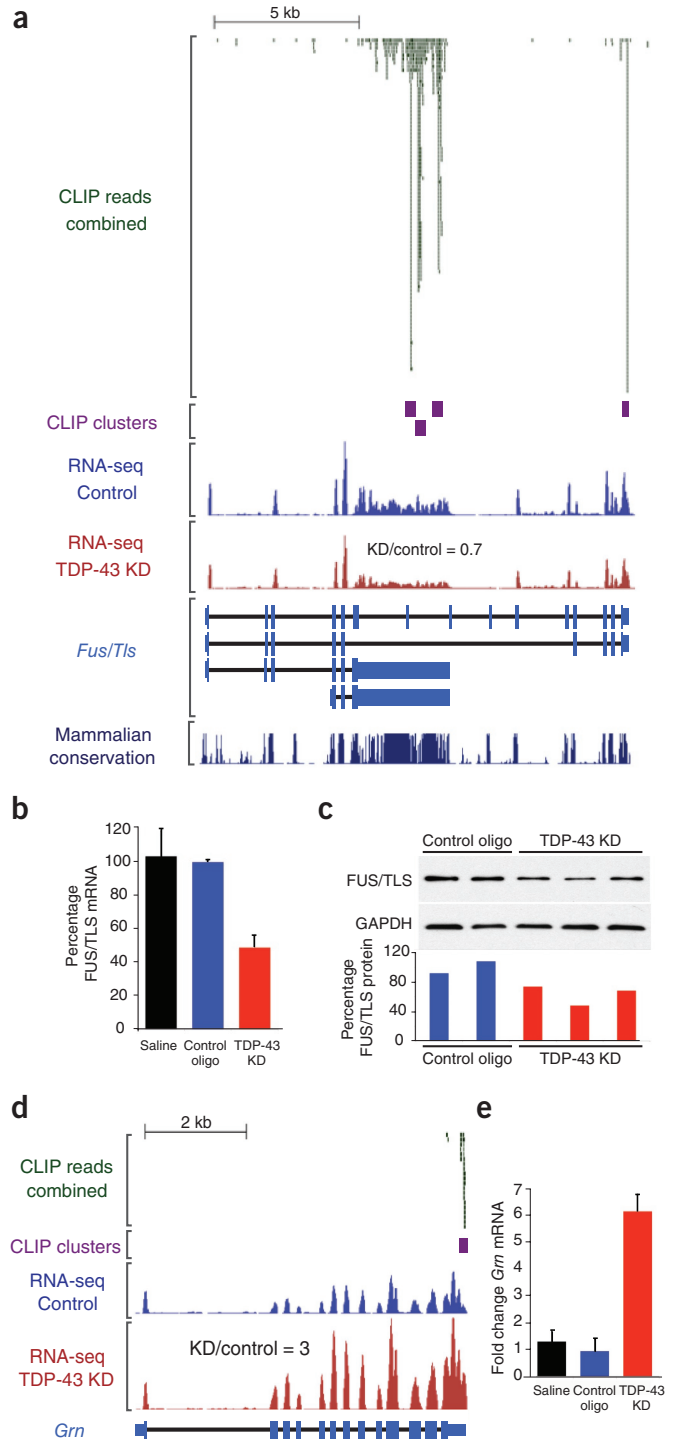**Figure 6** TDP-43 regulates expression of FUS/TLS and progranulin. (**a**) CLIP-seq reads and clusters on *Fus/Tls* transcript showing TDP-43 binding in introns 6 and 7 (also annotated as an alternative 3′ UTR) and the canonical 3′ UTR. RNA-seq reads from control or TDP-43 knockdown samples (equal scales) showed a slight reduction in *Fus/Tls* mRNA in the TDP-43 knockdown group and expression values from RNA-seq confirmed that *Fus/Tls* mRNA was reduced to 70% of control on TDP-43 reduction. (**b**) qRT-PCR confirmed *Fus/Tls* mRNA downregulation on TDP-43 depletion. s.d. was calculated within each group for three biological replicas. (**c**) Semi-quantitative immunoblot (upper panel) revealed a slight, but consistent, reduction of FUS/TLS protein in ASO-treated mice to ~70% of control levels, as quantified by densitometric analysis (lower panel). (**d**) CLIP-seq reads and clusters on progranulin transcript (*Grn*) showing a sharp binding in the 3′ UTR. RNA-seq reads from control or TDP-43 knockdown samples (equal scales) showed a significant increase in progranulin mRNA in the TDP-43 knockdown group compared with controls. (**e**) qRT-PCR confirmed the statistically significant increase ($P < 3 \times 10^{-4}$) in progranulin mRNA in samples with reduced TDP-43 levels when compared with controls. s.d. was calculated in each group for three biological replicas.

was increased, not only on elevated TDP-43 expression (by transient transfection), but also on blocking of NMD, with a synergistic effect in the combined conditions (**Fig. 5g**).

TDP-43 regulates expression of disease-related transcripts

TDP-43 protein bound and directly regulated a variety of transcripts involved in neurological diseases (**Supplementary Table 6** and **Supplementary Figs. 8** and **13**), including *Fus/Tls* and *Grn* (**Fig. 6**), which encode FUS/TLS and progranulin, mutations in which cause ALS[10,11] or FTLD[33,34], respectively. TDP-43 bound to the 3′ UTR of *Fus/Tls* mRNA and to introns 6 and 7, all of which are highly conserved between mammalian species (**Fig. 6a**). Gene annotation and the presence of RNA-seq reads in these introns were consistent with either an alternative 3′ UTR or intron retention. *Fus/Tls* mRNA and protein were reduced to approximately 40% of their normal levels (**Fig. 6b,c**). *Progranulin* mRNA, on the other hand, was markedly increased by ~3–6-fold compared with controls (**Fig. 6d,e**).

CLIP-seq data also confirmed TDP-43 binding to two RNAs that were previously reported to be associated with TDP-43: histone deacetylase 6 (*Hdac6*)[35] and low–molecular weight neurofilament subunit (*Nefl*)[36]. Our RNA-seq data indicate that HDAC6, which functions to promote the degradation of polyubiquitinated proteins, was reduced on TDP-43 depletion (**Supplementary Fig. 14a,b**), albeit to a lesser degree *in vivo* than previously reported in cell culture[35]. It has been known for many years that *Nefl* mRNA levels are reduced in degenerating motor neurons from individuals with ALS[37]. We identified TDP-43 clusters in the 3′ UTR of *Nefl* (**Supplementary Fig. 14c**). In addition, RNA-seq data confirmed that the mouse *Nefl* 3′ UTR was longer than annotated and *Nefl* mRNA levels were slightly reduced on TDP-43 depletion (**Supplementary Fig. 14d**). Multiple TDP-43–binding sites were also present in the pre-mRNA from the *Mapt* gene encoding tau, whose mutation or altered splicing of exon 10 has been implicated in FTD[38]. However, neither the levels nor the splicing pattern of *Mapt* RNA were affected by TDP-43 depletion (**Supplementary Table 6**). Moreover, we identified multiple TDP-43 intronic binding sites in the *Hdh* transcript (**Supplementary Fig. 13**), which encodes huntingtin, the protein whose polyglutamine expansion causes Huntington's disease in humans[39], accompanied by cytoplasmic TDP-43 accumulations[40]. Moreover, *Hdh* levels were decreased in mouse brain on TDP-43 depletion (**Supplementary Table 6**). In contrast, we found no evidence for direct binding or TDP-43 regulation of the *Sod1* transcript (**Supplementary Table 6**), whose aberrant splicing in familial ALS cases[41,42] has raised the possibility that *SOD1* missplicing may be involved in the pathogenesis of sporadic ALS.

DISCUSSION

TDP-43 is a central component in the pathogenesis of an ever-increasing list of neurodegenerative conditions. We created a genome-wide RNA map of >39,000 TDP-43–binding sites in the mouse transcriptome and determined that levels of 601 mRNAs and splicing patterns of 965 mRNAs were altered following TDP-43 reduction in the adult nervous system. Thus, although earlier efforts have implicated TDP-43 as a splicing regulator of a few candidate genes[20,43,44], our RNA-seq and microarray results establish that TDP-43 regulates the largest set (512) of cassette exons thus far reported, demonstrating its broad role in

alternative splicing regulation. We also found that TDP-43 was required for maintenance of 42 non-overlapping, noncoding RNAs.

These findings also provide a direct test for how the nuclear loss of TDP-43 widely reported in the remaining motor neurons in ALS autopsy samples[1] may contribute to neuronal dysfunction, independent of potential damage from TDP-43 aggregates. The results of our experiments using *in vivo* reduction of TDP-43 coupled with RNA sequencing indicate that TDP-43 is crucial for sustaining levels of 239 mRNAs, including those encoding synaptic proteins, the choline acetyltransferase, and the disease-related proteins FUS/TLS and progranulin. A substantial proportion of these pre-mRNAs are directly bound by TDP-43 at multiple sites in exceptionally long introns, a feature that we found most prominently in brain-enriched transcripts (**Fig. 3**), thereby identifying one component of neuronal vulnerability from TDP-43 loss.

A plausible model for the role of TDP-43 in sustaining the levels of mRNAs derived from long pre-mRNAs is that TDP-43 binding in long introns prevents unproductive splicing events that would introduce premature stop codons and thereby promote RNA degradation. Our results thus identify a previously unknown conserved role for TDP-43 in regulating a subset of these long intron-containing brain-enriched genes. None of our evidence eliminates the possibility that TDP-43 affects RNA levels by additional mechanisms, such as through transcription regulation or by facilitation of RNA polymerase elongation, similar to what has been shown for another splicing regulator, SC35 (ref. 45).

FUS/TLS is another RNA-binding protein whose mutation causes ALS[10,11] and, in some rare cases, FTLD-U. Similar to TDP-43, FUS/TLS aggregation has been observed in different neurodegenerative conditions, including Huntington's disease and spinocerebellar ataxia (reviewed in ref. 3). We have now shown that *Fus/Tls* mRNA is a direct target of TDP-43 and its level is reduced on TDP-43 depletion (**Fig. 6**), thereby identifying a previously unknown FUS/TLS dependency on TDP-43. The latter is also true for two additional disease-relevant proteins, progranulin and its proposed receptor sortilin. Progranulin levels were sharply increased on TDP-43 reduction and splicing of sortilin was altered. In fact, TDP-43 directly binds and regulates the levels and splicing patterns of transcripts implicated in various neurologic diseases[46] (**Supplementary Table 6** and **Supplementary Fig. 13**), consistent with a broad role of TDP-43 in these conditions[3].

Finally, in contrast with a recent study that reported an autoregulatory mechanism for TDP-43 that is independent of pre-mRNA splicing[28], our results indicate that TDP-43 acts as a splicing regulator to reduce its own expression level by binding to the 3′ UTR of its own pre-mRNA. Although there may be additional mechanisms beyond NMD[28], we found that TDP-43 enhanced splicing of an alternative intron in its own 3′ UTR, thereby autoregulating its levels through a mechanism that involves splicing-dependent RNA degradation by NMD (**Fig. 5**). TDP-43 autoregulation occurs in the mammalian CNS, as seen by a significant reduction of endogenous TDP-43 mRNA and protein in response to the expression of a TDP-43 transgene lacking the regulatory intron, as we found and others have reported previously[47,48]. Both spliced and unspliced TDP-43 RNAs were found in human and mouse brain, consistent with autoregulation at normal TDP-43 levels that substantially attenuates TDP-43 synthesis through production of unstable, spliced RNA.

TDP-43-dependent splicing of its 3′ UTR intron as a key component of a TDP-43 autoregulatory loop could participate in a feedforward mechanism enhancing the cytoplasmic TDP-43 aggregates that are hallmarks of familial and sporadic ALS. Following an initiating insult (for example, one that traps some TDP-43 in initial cytoplasmic aggregates), the reduction in nuclear TDP-43 levels would decrease splicing of its 3′ UTR, which would in turn produce an elevated pool of stable TDP-43 mRNA. Repeated translation of that TDP-43 mRNA would increase

synthesis of new TDP-43 in the cytoplasm whose subsequent co-aggregation into the initial complexes would drive their growth. Disrupted autoregulation, by any event that lowers nuclear TDP-43, thus provides a mechanistic explanation for what may be a critical, intermediate step in the molecular mechanisms underlying age-dependent degeneration and death of neurons in TDP-43 proteinopathies.

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

1. Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).
2. Arai, T. *et al.* TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem. Biophys. Res. Commun.* **351**, 602–611 (2006).
3. 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).
4. Gitcho, M.A. *et al.* TDP-43 A315T mutation in familial motor neuron disease. *Ann. Neurol.* **63**, 535–538 (2008).
5. Kabashi, E. *et al.* TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat. Genet.* **40**, 572–574 (2008).
6. Sreedharan, J. *et al.* TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science* **319**, 1668–1672 (2008).
7. Van Deerlin, V.M. *et al.* TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. *Lancet Neurol.* **7**, 409–416 (2008).
8. Buratti, E. *et al.* Nuclear factor TDP-43 can affect selected microRNA levels. *FEBS J.* **277**, 2268–2281 (2010).
9. Cooper, T.A., Wan, L. & Dreyfuss, G. RNA and Disease. *Cell* **136**, 777–793 (2009).
10. Kwiatkowski, T.J. Jr. *et al.* Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* **323**, 1205–1208 (2009).

11. Vance, C. *et al.* Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* **323**, 1208–1211 (2009).

12. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).

13. Ule, J. *et al.* CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212–1215 (2003).

14. Licatalosi, D.D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 (2008).

15. Yeo, G.W. *et al.* An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat. Struct. Mol. Biol.* **16**, 130–137 (2009).

16. Ling, S.C. *et al.* ALS-associated mutations in TDP-43 increase its stability and promote TDP-43 complexes with FUS/TLS. *Proc. Natl. Acad. Sci. USA* **107**, 13318–13323 (2010).

17. Zisoulis, D.G. *et al.* Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. *Nat. Struct. Mol. Biol.* **17**, 173–179 (2010).

18. Sephton, C.F. *et al.* Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes. *J. Biol. Chem.* **286**, 1204–1215 (2011).

19. Mili, S. & Steitz, J.A. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **10**, 1692–1694 (2004).

20. Buratti, E. *et al.* Nuclear factor TDP-43 and SR proteins promote *in vitro* and *in vivo* CFTR exon 9 skipping. *EMBO J.* **20**, 1774–1784 (2001).

21. Chi, S.W., Zang, J.B., Mele, A. & Darnell, R.B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460**, 479–486 (2009).

22. Parkhomchuk, D. *et al.* Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* **37**, e123 (2009).

23. Pang, K.C. *et al.* RNAdb—a comprehensive mammalian noncoding RNA database. *Nucleic Acids Res.* **33**, D125–D130 (2005).

24. Wang, E.T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).

25. Hu, F. *et al.* Sortilin-mediated endocytosis determines levels of the frontotemporal dementia protein, progranulin. *Neuron* **68**, 654–667 (2010).

26. Carrasquillo, M.M. *et al.* Genome-wide screen identifies rs646776 near sortilin as a regulator of progranulin levels in human pasma. *Am. J. Hum. Genet.* **87**, 890–897 (2010).

27. Du, H. *et al.* Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nat. Struct. Mol. Biol.* **17**, 187–193 (2010).

28. Ayala, Y.M. *et al.* TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J.* 30, 277–288 (2011).

29. Le Hir, H., Moore, M.J. & Maquat, L.E. Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev.* **14**, 1098–1108 (2000).

30. Wollerton, M.C., Gooding, C., Wagner, E.J., Garcia-Blanco, M.A. & Smith, C.W. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol. Cell* **13**, 91–100 (2004).

31. Sureau, A., Gattoni, R., Dooghe, Y., Stevenin, J. & Soret, J. SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *EMBO J.* **20**, 1785–1796 (2001).

32. Perlick, H.A., Medghalchi, S.M., Spencer, F.A., Kendzior, R.J. Jr. & Dietz, H.C. Mammalian orthologs of a yeast regulator of nonsense transcript stability. *Proc. Natl. Acad. Sci. USA* **93**, 10928–10932 (1996).

33. Baker, M. *et al.* Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. *Nature* **442**, 916–919 (2006).

34. Cruts, M. *et al.* Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. *Nature* **442**, 920–924 (2006).

35. Fiesel, F.C. *et al.* Knockdown of transactive response DNA-binding protein (TDP-43) downregulates histone deacetylase 6. *EMBO J.* **29**, 209–221 (2010).

36. Strong, M.J. *et al.* TDP43 is a human low molecular weight neurofilament (hNFL) mRNA-binding protein. *Mol. Cell. Neurosci.* **35**, 320–327 (2007).

37. Bergeron, C. *et al.* Neurofilament light and polyadenylated mRNA levels are decreased in amyotrophic lateral sclerosis motor neurons. *J. Neuropathol. Exp. Neurol.* **53**, 221–230 (1994).

38. Hutton, M. *et al.* Association of missense and 5′-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* **393**, 702–705 (1998).

39. MacDonald, M.E. *et al.* A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**, 971–983 (1993).

40. Schwab, C., Arai, T., Hasegawa, M., Yu, S. & McGeer, P.L. Colocalization of transactivation-responsive DNA-binding protein 43 and huntingtin in inclusions of Huntington disease. *J. Neuropathol. Exp. Neurol.* **67**, 1159–1165 (2008).

41. Valdmanis, P.N. *et al.* A mutation that creates a pseudoexon in SOD1 causes familial ALS. *Ann. Hum. Genet.* **73**, 652–657 (2009).

42. Birve, A. *et al.* A novel SOD1 splice site mutation associated with familial ALS revealed by SOD activity analysis. *Hum. Mol. Genet.* **19**, 4201–4206 (2010).

43. Mercado, P.A., Ayala, Y.M., Romano, M., Buratti, E. & Baralle, F.E. Depletion of TDP 43 overrides the need for exonic and intronic splicing enhancers in the human apoA-II gene. *Nucleic Acids Res.* **33**, 6000–6010 (2005).

44. Dreumont, N. *et al.* Antagonistic factors control the unproductive splicing of SC35 terminal intron. *Nucleic Acids Res.* **38**, 1353–1366 (2009).

45. Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S. & Fu, X.D. The splicing factor SC35 has an active role in transcriptional elongation. *Nat. Struct. Mol. Biol.* **15**, 819–826 (2008).

46. Tollervey, J.R. *et al.* Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.* advance online publication, doi:10.1038/nn.2778 (27 February 2011).

47. Xu, Y.F. *et al.* Wild-type human TDP-43 expression causes TDP-43 phosphorylation, mitochondrial aggregation, motor deficits, and early mortality in transgenic mice. *J. Neurosci.* **30**, 10851–10859 (2010).

48. Igaz, L.M. *et al.* Dysregulation of the ALS-associated gene TDP-43 leads to neuronal death and degeneration in mice. *J. Clin. Invest.* **121**, 726–738 (2011).

ONLINE METHODS

CLIP-seq library preparation and sequencing. Brains from 8-week-old female C57Bl/6 mice were rapidly dissociated by forcing through a cell strainer with a pore size of 100 μm (BD Falcon) before ultraviolet crosslinking. CLIP-seq libraries were constructed as previously described[15], using a custom-made mouse monoclonal antibody to TDP-43 (ref. 16, 40 μl of ascites per 400 μl of beads per sample). Libraries were subjected to standard Illumina GA2 sequencing protocol for 36 cycles.

Generation of transgenic mice. All animal procedures were conducted in accordance with the guidelines of the Institutional Animal Care and Use Committee of the University of California. cDNA containing N-terminal myc-tagged full-length human TDP-43 were amplified and digested by SalI and cloned into the XhoI cloning site of the MoPrP.XhoI vector (ATCC #JHU-2). The resultant MoPrP.XhoI-myc-hTDP-43 construct was then digested upstream of the minimal *Prnp* promoter and downstream of the *Prnp* exon 3 using BamHI and NotI and cloned into a shuttle vector containing *loxP* flanking sites. The final construct was then linearized using XhoI, injected into the pro-nuclei of fertilized C56Bl6/C3H hybrid eggs and implanted into pseudopregnant female mice.

Stereotactic injections of antisense oligonucleotides. We anesthetized 8–10-week-old female C57Bl/6 mice with isofluorane. Using stereotaxic guides, 3 μl of ASO solution, corresponding to a total of 75 μg or 100 μg ASOs, or saline buffer was injected using a Hamilton syringe directly into the striatum. Mice were monitored for any adverse effects for the next 2 weeks until they were killed. The striatum and adjacent cortex area were dissected and frozen at −80 °C in 1 ml Trizol (Invitrogen). Trizol extraction of RNA and protein was performed according to the manufacturer's instructions.

RNA quality and RNA-seq library preparation. RNA quality was measured using the Agilent Bioanalyzer system according to the manufacturer's recommendations. RNA-seq libraries were constructed as described previously[22]. We used 8 pM of amplified libraries for sequencing on the Illumina GA2 for 72 cycles.

RT and qRT-PCR. cDNA of total RNA extracted from striatum was generated using oligodT and Superscript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions. To test candidate splicing targets, RT-PCR amplification using between 24 and 27 cycles were performed from at least three mice treated with a control ASO and three mice with treated with a TDP-43 ASO. Products were separated on 10% polyacrylamide gels followed by staining with SYBR gold (Invitrogen). Quantification of the different isoforms was performed with ImageJ software (US National Institutes of Health). Intensity ratio between products with included and excluded exons were averaged for three biological replicates per group.

qRT-PCR for mouse *Tardbp* and *Fus/Tls* were performed using the Express One-Step SuperScript qRT-PCR kits (Invitrogen) and the thermocycler ABI Prism 7700 (Applied Biosystems). cDNA synthesis and amplification were performed according to the manufacturer's instruction using specific primers and 5′ FAM, 3′ TAMRA-labeled probes. The *cyclophilin* gene was used to normalize the expression values.

qRT-PCR for all other genes tested were performed with 3–5 mice for each group (treated with saline, control ASO or ASO to TDP-43) and two technical replicates using the iQ SYBR green supermix (Bio-Rad) on the IQ5 multicolor real-time PCR detection system (Bio-Rad). The analysis was done using the IQ5 optical system software (Bio-Rad; version 2.1). Expression values were normalized to at least two of the following control genes: *Actb*, *Actg1* and *Rsp9*. Expression values were expressed as a percentage of the average expression of the saline treated samples. Inter-group differences were assessed by two-tailed Student's *t* test. Primers for RT-PCR and qRT-PCR were designed using Primer3 software (http://frodo.wi.mit.edu/primer3/) and sequences are available on request.

Immunoblots. Proteins were separated on custom-made 12% SDS page gels and transferred to nitrocellulose membrane (Whatman) following standard protocols. Membranes were blocked overnight in Tris-buffered saline Tween-20 (TBST) and 5% non-fat dry milk (wt/vol) at 4 °C, incubated for 1 h at 20 °C with primary antibodies and then again with horseradish peroxidise–linked secondary antibodies to rabbit or mouse (GE Healthcare) in TBST with 1% milk. For primary antibodies, we used rabbit antibody to FUS/TLS (Bethyl Laboratories,

cat #A300-302A; 1:5,000), rabbit antibody to TDP-43 (Proteintech, cat #10782; 1:2,000), rabbit antibody to TDP-43 (Aviva System Biology, cat #ARP38942_T100; 1:2,000), custom-made mouse antibody to TDP-43 (1:1,000)[16], custom-made rabbit antibody to RFP raised against the full-length protein (1:7,000), mouse DM1α antibody to tubulin (1:10,000) and mouse antibody to GAPDH (Abcam, cat #AB8245; 1:10,000).

Cells, cloning and luciferase assays. HeLa Flp-In cells expressing *GFP-myc–TDP-43–HA* were generated as previously described[16]. Isogenic cell lines were grown at 37 °C and 5% CO_2 in Dulbecco's modified Eagle medium (DMEM), supplemented with 10% tetracycline-free fetal bovine serum (vol/vol) and penicillin/streptomycin. Expression of *GFP-myc–TDP-43–HA* was induced with 1 mg ml^{-1} tetracycline for 24–48 h.

To assess the mechanism of TDP-43 autoregulation, we cloned the proximal part of mouse TDP-43 3′ UTR into the psiCHECK-2 vector (Promega), which contains a *Renilla* luciferase and a *Firefly* luciferase reporter expression cassettes. The following primers were used to amplify 1.7 kb of TDP-43 3′ UTR using cDNA from mouse brain: 5′-AAA CTC GAG CAG GCT TTT GGT TCT GGA AA-3′ and 5′-AAA GCG GCC GCA CCA TTT TAG GTG CGG TCA C-3′. We obtained two products of 1.7 kb and 1.1 kb, corresponding, respectively, to an unspliced and a spliced isoform of TDP-43 3′ UTR (**Supplementary Fig. 13a**). Both products were purified on 1% agarose gels and cloned independently in the psiCHECK-2 vector using NotI and XhoI restriction sites located 3′ to the *Renilla* luciferase translational stop. Given that binding sites for TDP-43 mainly lie in the alternative intron, the spliced isoform was used as a control to assess the effect of TDP-43 protein on its own RNA.

Human *myc-TDP-43-HA* cDNA (a generous gift from C. Shaw, King's College London) was cloned into mammalian expression vector, pCl-neo (Promega). RFP was cloned in the vector pcDNA3 (Invitrogen).

250 ng of psiCHECK-2 vector with or without TDP-43 3′ UTR and 250 ng of the vector expressing TDP-43 or RFP were co-transfected in Hela cells using Fugene 6 transfection reagent (Roche) in 12-wells cell culture plates. Luciferase assays were performed 48 h after transfection using the Dual-Luciferase Reporter 1000 assay system (Promega) according to the manufacturer's instructions. Five independent experiments were performed and 20 μl of lysate were used in duplicate for each condition. RFU for *Renilla* luciferase were normalized to *firefly* luciferase to control for transfection efficiency. Duplicates were averaged and each condition was expressed as a percentage of the samples without transfection of TDP-43 or RFP cDNA. Inter-group differences were assessed by two-tailed Student's *t* test.

Mouse gene structure annotations. The mouse genome sequence (mm8) and annotations for protein-coding genes were obtained from the UCSC Genome Browser. Known mouse genes (knownGene containing 31,863 entries) and known isoforms (knownIsoforms containing 31,014 entries in 19,199 unique isoform clusters) with annotated exon alignments to the mouse genomic sequence were processed as follows. Known genes that were mapped to different isoform clusters were discarded. All mRNAs aligned to mm8 that were greater than 300 nucleotides were clustered together with the known isoforms. For the purpose of inferring alternative splicing, genes containing fewer than three exons were removed from further consideration. A total of 1.9 million spliced ESTs were mapped onto the 16,953 high-quality gene clusters to identify alternative splicing events. Final annotated gene regions were clustered together so that any overlapping portion of these databases was defined by a single genomic position. Promoter regions were arbitrarily defined as 1.5 kb upstream of the transcriptional start site of the gene and intergenic regions as unannotated regions in the genome. To identify 5′ and 3′ UTRs, we relied on the coding annotation in UCSC Genome Browser known genes that we extended 1.5 kb downstream or upstream the start and stop codons, respectively.

Comparison to human alternatively spliced exons. To find evidence for conserved alternative splicing patterns in humans, we used the UCSC LiftOver tool to obtain the corresponding human coordinates of the mouse exons that had evidence for differential splicing either from RNA-seq or splicing-sensitive microarray data. These human genome coordinates were then compared with human gene structure annotations constructed analogously to mouse annotations as described above to determine whether the exon in the human ortholog was alternatively or constitutively spliced based on transcript evidence in human EST or mRNA databases[49].

Computational identification of CLIP-seq clusters. CLIP-seq reads were trimmed to remove adaptor sequences and homopolymeric runs, and mapped to the repeat-masked mouse genome (mm8) using the bowtie short-read alignment software (version 0.12.2) with parameters -q -p 4 -e 100 -y -a -m 10 --best --strata (-q for fastq file, -p for cpus, -e for mismatch threshold, -y to do a sensitive search, -a to report all alignments, --best for ordered output, --strata report only the best group of alignments), incorporating the base-calling quality of the reads. To eliminate redundancies created by PCR amplification, we considered all reads with identical sequence to be a single read. Significant clusters were calculated by first determining read number cutoffs using the Poisson distribution, $f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$, where λ is the frequency of reads mapped over an interval of nucleotide sequence, k is the number of reads being analyzed for significance and $f(k;\lambda)$ return the probability that exactly k reads would be found. For any desired P value, P cutoff, a read number cutoff was calculated by summing the probabilities for finding k or more tags, and determining the minimum value of k that satisfies $i = k$ such that $f(k;\lambda) > P$ cutoff. The frequency λ was calculated by dividing the total number of mapped reads by the number of non-overlapping intervals present in the transcriptome. The interval size was chosen on the basis of the average size of the CLIP product, which includes only the selected RNA fragments, but not any ligated adapters (150 bp). A global and a local cutoff were determined using the whole transcriptome frequency or gene-specific frequency, respectively. The gene-specific frequency was the number of reads overlapping that gene divided by the pre-mRNA length. A sliding window of 150 bp was used to determine where the read numbers exceed both the global and local cutoffs. At each significant interval, we attempted to extend the region by adding in the next read, ensuring continued significance at the same P cutoff.

Log curve analysis to determine target saturation of CLIP-derived clusters. We used a curve-fitting approach with various sampling rates to estimate the number of CLIP-seq reads required to discover additional CLIP clusters. The target rate was calculated by determining the new set of CLIP-derived gene targets found at each step-wise increase in the number of sequenced reads. A scatter plot of targets found versus sampling rate was fitted with a log curve, which was then used to extrapolate the number of targets expected to be found by increasing read counts.

Splicing-sensitive microarray data analysis. Microarray data analysis was performed as previously described, selecting significant, high-quality events using a q value > 0.05 and an absolute separation score (sepscore) > 0.5 (ref. 27). The equation for sepscore = $\log_2$[TDP-43 depleted (Skip/Include)/ Control treated (Skip/Include)]. For each replicate set, the $\log_2$ ratio of skipping to inclusion was estimated using robust least-squares analysis. Previously published work using similar cutoffs have validated about 85% of splicing events by RT-PCR[27].

Events on the array were defined in the mm9 genome annotation, and for proper comparison to RNA-seq and CLIP-seq data, cassette events were converted to the mm8 genome annotation using the UCSC LiftOver tool. If an event did not exactly overlap an mm8 annotated exon, it was left out of further analyses.

Genomic analysis of CLIP clusters. Two biologically independent TDP-43 CLIP-seq libraries were generated and sequenced on the Illumina GA2. Subjecting reads from each library to our cluster-identification pipeline described above defined 15,344 and 30,744 clusters for CLIP experiments 1 and 2, respectively. A gene was considered to be a TDP-43 target that overlapped in both experiments if it contained an overlapping cluster. We generated ten randomly distributed cluster sets and compared each to the original clusters. To compute the significance of the overlap, we calculated the standard or Z-score as follows: (percent overlap in the two experiments – mean percent overlap in ten randomly distributed cluster sets)/ s.d. of percent overlap in the randomly distributed cluster sets. A P value was computed from the standard normal distribution and assigned significance if it was lower than $P < 0.01$. To generate the final set of TDP-43 CLIP-seq clusters, unique reads from both experiments were combined and subjected to our cluster-identification pipeline. Overall, clusters were at most 300 bases in length, with the median of 142 bases. As a comparison to a published HITS-CLIP/CLIP-seq dataset of a RNA binding protein in mouse brain, we downloaded Ago-HITS-CLIP reads (Brain[A–E]_130_50_fastq.txt) from http://ago.rockefeller.edu/rawdata.php and subjected the combined 1,651,104 reads to our cluster-identification and generated 33,390 clusters in 7,745 genes[21]. To identify enriched motifs in cluster regions, Z scores were computed for hexamers as previously published[15].

Functional annotation analysis. We used the Database for Annotation, Visualization and Integrated Discovery (DAVID Bioinformatic Resources 6.7; http://david.abcc.ncifcrf.gov/). For all genes downregulated or upregulated on TDP-43 depletion, a background corresponding to all genes expressed in brain was used.

Transcriptome and splicing analysis. Strand-specific RNA-seq reads each from control oligonucleotide, saline- and TDP-43 oligonucleotide–treated animals were generated and ~50% mapped uniquely to our annotated gene structure database using the bowtie short-read alignment software (version 0.12.2, with parameters -m 5 -k 5 --best --un --max -f) incorporating the base-calling quality of the reads. To eliminate redundancies created by PCR amplification, all reads with identical sequence were considered single reads. The expression value of each gene was computed by RPKM. Each RNA-seq sample was summarized by a vector of RPKM values for every gene and pair-wise correlation coefficients were calculated for all replicates using a linear least-squares regression against the log RPKM vectors. Hierarchical clustering revealed that the three treated conditions clustered into similar groups. The reads in each condition were combined to identify genes that were significantly up- and downregulated on TDP-43 depletion. Local mean and s.d. were calculated for the nearest 1,000 genes, as determined by log average RPKM values between knockdown and control and a local Z score was defined. The resulting Z scores were used to assign significantly changed genes ($Z > 2$ upregulated, $Z < –2$ downregulated), as well as ranking the entire gene list for relative expression changes. Specific parameters, such as intron length or TDP-43–binding sites, were plotted for the next 100 genes.

Exons with canonical splice signals (GT-AG, AT-AC, GC-AG) were retained, resulting in a total of 190,161 exons. For each protein-coding gene, the 50 bases at the 3′ end of each exon were concatenated with the 50 bases at the 5′ end of the downstream exon producing 1,827,013 splice junctions. An equal number of 'impossible' junctions was generated by joining the 50-base exon junction sequences in reverse order. To identify differentially regulated alternative cassette exons, we employed a modification of a published method[24]. In short, the read count supporting inclusion of the exon (overlapping the cassette exon and splice junctions including the exon) were compared with the read count supporting exclusion of the exon (overlapping the splice junction of the upstream and downstream exon). For a splice junction read to be enumerated, we required that at least 18 nucleotides of the read aligned and 5 bases of the read extended over the splice junction with no mismatches 5 bases around the splice junction. For the TDP-43 and control ASOs comparison, we constructed a 2 × 2 contingency table using of the counts of the reads supporting the inclusion and exclusion of the exon, in two conditions. Every cell in the 2 × 2 table had to contain at least five counts for a χ^2 statistic to be computed. At $P < 0.01$, 110 excluded and 93 included single cassette exons were detected to be differentially regulated by TDP-43. As an estimate of false discovery, we observed that ~20 single cassette exons were detected by using the impossible junction database.

Pre-mRNA features and tissue-specificity analysis. Affymetrix microarray data representing 61 mouse tissues were downloaded from the Gene Expression Omnibus repository (http://www.ncbi.nih.gov/geo) under accession number GSE1133 (ref. 50). Probes on the microarrays were cross-referenced to 15,541 genes in our database using files downloaded from the UCSC Genome Browser (knownToGnf1m and knownToGnfAtlas2). The expression value for each gene was represented by the average value of the two replicate microarray experiments for each tissue. To identify genes enriched in brain, we grouped 13 tissues as 'brain' (substantia nigra, hypothalamus, preoptic, frontal cortex, cerebral cortex, amygdala, dorsal striatum, hippocampus, olfactory bulb, cerebellum, trigeminal, dorsal root ganglia, pituitary) and the remaining 44 tissues as 'non-brain', excluding the five embryonic tissues (embryonic days 10.5, 9.5, 8.5, 7.5 and 6.5). For each gene, the t statistic was computed as $t = \frac{(\mu_{brain} - \mu_{non\text{-}brain})}{\sqrt{\sigma^2_{brain} + \sigma^2_{non\text{-}brain}}}$, where $\mu_{brain}(\sigma_{brain})$ and $\mu_{non\text{-}brain}(\sigma_{non\text{-}brain})$ were the average (s.d.) of the gene expression values in brain and non-brain tissues, respectively. At a t statistic value cutoff of ≥1.5, 388 genes were categorized as being brain enriched. Concurrently, at a cutoff of <1.5, 15,153 genes were categorized as being non-brain enriched. Random sets of 388 genes were selected from the 15,541 genes as controls for the brain-enriched set. To determine whether pre-mRNA features were significantly different in the set of brain-enriched genes (or randomly chosen genes)

compared with non-brain–enriched genes, we performed the two-sample Kolmogorov-Smirnov goodness-of-fit hypothesis test, which determines whether the distribution of features were drawn from the same underlying continuous population. Cumulative distribution plots of pre-mRNA features were generated to illustrate the differences.

Affymetrix microarray data representing 79 human tissues were downloaded from the Gene Expression Omibus repository under the same accession number GSE1133 (ref. 50). Probes on the human microarrays were cross-referenced to 18,372 genes in our database using files downloaded from the UCSC Genome Browser (knownToGnfAtlas2 – hg18). The same analysis as done for the mouse array data was repeated for these human array data. We grouped 17 tissues as 'brain' (temporal lobe, globus pallidus, cerebellum peduncles, cerebellum, caudate nucleus, whole brain, parietal lobe, medulla oblongata, amygdala, prefrontal cortex, occipital lobe, hypothalamus, thalamus, subthalamic nucleus, cingulated cortex, pons, fetal brain), and the remaining 62 tissues as 'non-brain'. At the same t statistic cutoffs, 387 and 17,985 genes were categorized as brain enriched and non-brain enriched, respectively. Random sets of 387 genes were selected from the 18,372 genes as controls for the brain-enriched set.

49. Yeo, G.W., Van Nostrand, E.L. & Liang, T.Y. Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. *PLoS Genet.* **3**, e85 (2007).
50. Su, A.I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA* **101**, 6062–6067 (2004).