⌘ *Author's Choice*

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 2, pp. 1204–1215, January 14, 2011
© 2011 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in the U.S.A.

Identification of Neuronal RNA Targets of TDP-43-containing Ribonucleoprotein Complexes*[S]◆

Received for publication, October 4, 2010, and in revised form, October 27, 2010 Published, JBC Papers in Press, November 4, 2010, DOI 10.1074/jbc.M110.190884

Chantelle F. Sephton[‡1], Can Cenik[§1], Alper Kucukural[¶1], Eric B. Dammer[**], Basar Cenik[‡ ‡‡], YuHong Han[‡], Colleen M. Dewey[‡], Frederick P. Roth[§2], Joachim Herz[‡ ‡‡], Junmin Peng[**], Melissa J. Moore[¶‖], and Gang Yu[‡3]

From the Departments of [‡]Neuroscience and [‡‡]Molecular Genetics, the University of Texas Southwestern Medical Center, Dallas, Texas 75390-9111, the [§]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, the [¶]Department of Biological Chemistry and Molecular Pharmacology, University of Massachusetts Medical School and [‖]Howard Hughes Medical Institute, Worcester, Massachusetts 01655, and the [**]Department of Human Genetics, Emory University, Atlanta, Georgia 30322

TAR DNA-binding protein 43 (TDP-43) is associated with a spectrum of neurodegenerative diseases. Although TDP-43 resembles heterogeneous nuclear ribonucleoproteins, its RNA targets and physiological protein partners remain unknown. Here we identify RNA targets of TDP-43 from cortical neurons by RNA immunoprecipitation followed by deep sequencing (RIP-seq). The canonical TDP-43 binding site $(TG)_n$ is 55.1-fold enriched, and moreover, a variant with adenine in the middle, $(TG)_nTA(TG)_m$, is highly abundant among reads in our TDP-43 RIP-seq library. TDP-43 RNA targets can be divided into three different groups: those primarily binding in introns, in exons, and across both introns and exons. TDP-43 RNA targets are particularly enriched for Gene Ontology terms related to synaptic function, RNA metabolism, and neuronal development. Furthermore, TDP-43 binds to a number of RNAs encoding for proteins implicated in neurodegeneration, including TDP-43 itself, FUS/TLS, progranulin, Tau, and ataxin 1 and -2. We also identify 25 proteins that co-purify with TDP-43 from rodent brain nuclear extracts. Prominent among them are nuclear proteins involved in pre-mRNA splicing and RNA stability and transport. Also notable are two neuron-enriched proteins, methyl CpG-binding protein 2 and polypyrimidine tract-binding protein 2 (PTBP2). A PTBP2 consensus RNA binding motif is enriched in the TDP-43 RIP-seq library, suggesting that PTBP2 may co-regulate TDP-43 RNA targets. This work thus reveals the protein and RNA components of the TDP-43-containing ribonucleoprotein complexes and provides a framework for understanding how dysregulation of TDP-43 in RNA metabolism contributes to neurodegeneration.

Gene expression is an essential process common to all living organisms. Regulation of genes in mammals can occur by repression or activation at transcription promoter sites or by regulating aspects of RNA metabolism (1). RNA metabolism is dysregulated in several neurodevelopmental and neurodegenerative diseases (2, 3), and it is plausible that defective RNA metabolism contributes to the pathogenesis and progression of neurodegeneration. Genetic mutations in two RNA-binding proteins, TDP-43[4] and fused in sarcoma/translated in liposarcoma (FUS/TLS), have recently been identified as causative factors of familial and sporadic amyotrophic lateral sclerosis (ALS) (4, 5). TDP-43 and FUS/TLS are also major components of the ubiquitinated neuronal and glial inclusions in affected brain and spinal cord regions of patients with ALS and frontotemporal lobar degeneration with ubiquitin-positive inclusions (6). In animal studies, transgenic mice for wild type (7) and mutant (8) TDP-43 partially phenocopy the human diseases. Genomic deletion of TDP-43 is embryonic lethal, indicating an essential role of TDP-43 in early embryogenesis (9, 10). Its neural function, however, is not known, nor is how alterations of neural TDP-43 lead to neurodegeneration.

TDP-43 is part of the family of heterogeneous nuclear ribonucleoproteins (hnRNPs), containing two highly conserved RNA recognition motifs and a non-conserved C-terminal region that mediates protein-protein interactions (11). It has been implicated in gene transcription, pre-mRNA splicing, mRNA stability, and mRNA transport (12). TDP-43 was shown to have high binding affinity for the $(TG)_n$ motif (13).

* This work was supported, in whole or in part, by National Institutes of Health Grants R01 AG029547 and AG023104 (to G. Y.), HG004233 (to F. P. R.), T32 NS007480 (to E. B. D.), and P30 NS055077 (to J. P.). This work was also supported by the Consortium for Frontotemporal Dementia Research (to G. Y. and J. H.), the Welch Foundation (to G. Y.), the Ted Nash Long Life Foundation (to G. Y.), and a fellowship from the Canadian Institute for Advanced Research (to F. P. R.). M. J. M. is an HHMI investigator.

⌘ *Author's Choice*—Final version full access.

◆ This article was selected as a Paper of the Week.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Data Files S1 and S2, Figs. S1–S3, and Tables S1–S9.

The nucleotide sequences reported in this paper have been submitted to the GEO database under accession number GSE25032.

[1] These authors contributed equally to this work.

[2] Present address: Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Ontario M5S 3E1, Canada and the Samuel Lunenfeld Research Institute, Mt. Sinai Hospital, Toronto, Ontario M5G 1X5, Canada.

[3] To whom correspondence should be addressed: Dept. of Neuroscience, University of Texas Southwestern Medical Center, 6000 Harry Hines Blvd., Dallas, TX 75390-9111. Tel.: 214-648-5157; Fax: 214-648-1801; E-mail: Gang.Yu@UTSouthwestern.edu.

[4] The abbreviations used are: TDP-43, TAR DNA-binding protein; RIP, RNA immunoprecipitation; PTBP2, polypyrimidine tract-binding protein 2; MECP2, methyl CpG-binding protein 2; CFTR, cystic fibrosis transmembrane conductance regulator; FUS/TLS or Fus, fused in sarcoma; SFRS1, splicing factor arginine/serine-rich 1; nt, nucleotide(s); hnRNP, heterogeneous nuclear ribonucleoprotein; SNRNP, small nuclear ribonucleoprotein; PPAR, peroxisome proliferator-activated receptor.

Splicing of the cystic fibrosis transmembrane conductance regulator (*CFTR*) (13), apolipoprotein A-II (*APOAII*) (14), and survival of motor neuron (*SMN*) (15) was reported to be regulated by TDP-43. In addition, TDP-43 has been implicated in regulation of mRNA biogenesis (16) and shown to be localized to sites of mRNA transcription and processing in neurons (17) and to bind directly to miRNAs (18).

TDP-43 is one of ~600 annotated and predicted RNA-binding proteins that function in multiprotein complexes, working in cooperation to perform a collective function in RNA metabolism to regulate protein-coding genes. The constitutions of many RNA-binding protein complexes and their RNA targets are not well characterized, nor is the site specificity of these complexes for their RNA targets known. These questions can now be answered due to recent advances in proteomics, functional genomics, and high throughput sequencing.

This study aims to identify the native protein constituents of TDP-43-containing ribonucleoprotein complexes and their RNA targets in neural cells. We performed a proteomic analysis on proteins co-purified with TDP-43 from brain nuclear extracts and determined that TDP-43 is in complexes with 25 endogenous nuclear proteins, mostly RNA-binding proteins and splicing factors. Notable among them are neural proteins methyl CpG-binding protein 2 (MECP2) and polypyrimidine tract-binding protein 2 (PTBP2). In parallel, we used RIP-seq and a bioinformatics approach, which included producing our own mappability track for calculating density reads within each gene, to identify and analyze novel, *in vivo* TDP-43 RNA targets. We found that TDP-43 binds predominantly to RNAs containing the consensus motif $(UG)_n$. Moreover, our analysis shows that there is often an adenine in the middle of the motif, $(UG)_nUA(UG)_m$. (Because our TDP-43 library generated for deep sequencing is composed of cDNAs that are then mapped to the rat genome, for simplicity, we will use TG instead of UG in describing TDP-43 targets in the rest of the text.) We also demonstrated that our TDP-43 library is significantly enriched in reads containing a PTBP2 consensus motif, suggesting that PTBP2 may co-regulate TDP-43 RNA targets. This work thus reveals the nuclear components of the TDP-43-containing ribonucleoprotein complexes and provides a framework for understanding the neuronal function of TDP-43 and its contribution to neurodegenerative disorders.

EXPERIMENTAL PROCEDURES

Materials—Electrophoresis reagents were from Bio-Rad. All other chemicals were reagent-grade and were as indicated in the following sections. TDP-43 was detected with antibodies 748C (9) or TDP-43 Proteintech Group Inc.; other antibodies include: hnRNPA1, lamin A/C, and MECP2 (Sigma-Aldrich).

TDP-43 Co-immunoprecipitation and Size Exclusion Chromatography—Mice on a mixed background (ages ~3–5 months) were sacrificed, and brains were harvested. Mouse brains were added to homogenization buffer (10 mM HEPES-NaOH, pH 7.4, 1 mM $MgCl_2$, 250 mM sucrose, 1× protease inhibitors) (Roche Diagnostics) and homogenized using a Dounce homogenizer. Homogenates were centrifuged at 1,500 × *g* for 10 min at 4 °C. The pellet was suspended in nuclear isolation buffer (10 mM HEPES-NaOH, pH 7.4, 1 mM $MgCl_2$, 1.42 M sucrose, 1 mM DTT, 1× protease inhibitors), added to Beckman centrifuge tubes, and centrifuged in a SW 45 Ti rotor at 100,000 × *g* for 1 h. The nuclear pellet (P100) was suspended in NT2 lysis buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, 0.05% Nonidet P-40, 20 mM DTT, 1× protease inhibitors).

Precleared nuclear extracts were applied to the AminoLink Plus resin (Pierce) cross-linked with either nonspecific rabbit IgG or TDP-43(748C) antibodies. Briefly, antibodies were diluted in coupling buffer (0.1 M $Na_3C_6H_5O_7$ and 0.05 M Na_2CO_3, pH 10) and coupled to the resin. Antibodies were then cross-linked to the resin, using 0.1 M $NaBH_3CN$. Lysates were incubated overnight at 4 °C and eluted with 50 mM glycine, pH 2.5, and eluents were neutralized with 1 M Tris-HCl, pH 8.0.

HeLa nuclear lysates were untreated or pretreated with RNase A (Roche Diagnostics) and rat brain nuclear lysates were untreated or pretreated with micrococcal nuclease before they were loaded into a Superose 6 or Superdex 200 column (GE Healthcare), respectively, in buffer containing 50 mM Tris-HCl, 150 mM NaCl at pH 7.5. The collected fractions were used for Western blot analysis.

Online Reverse Phase Liquid Chromatography, LC-MS/MS, and Proteomic Analysis—Immunoprecipitation eluates were desalted via loading and briefly resolving protein bands in a 10% polyacrylamide SDS gel. After staining with Coomassie Blue, each gel lane was cut into a band, and bands were subjected to in-gel digestion (12.5 μg/ml trypsin). Extracted peptides were loaded onto a C18 column (100-μm internal diameter, 12 cm long, ~300 nl/min flow rate, 5 μm, 200 Å pore size resin from Michrom Bioresources, Auburn, CA) and eluted during a 10–30% gradient (Buffer A, 0.4% acetic acid, 0.005% heptafluorobutyric acid, and 5% acetonitrile; Buffer B, 0.4% acetic acid, 0.005% heptafluorobutyric acid, and 95% acetonitrile) for 90 min (Experiment Sample 1) or 30 min (Experiment Sample 2). The eluted peptides were detected by Orbitrap (350–1500 m/z, 1,000,000 automatic gating control target, 1,000-ms maximum ion time, resolution 30,000 full width at half maximum) followed by 9–10 data-dependent MS/MS scans in linear trap quadrupole (2 m/z isolation width, 35% collision energy, 5,000 automatic gating control target, 200-ms maximum ion time) on a hybrid mass spectrometer (Thermo Finnigan).

Acquired MS/MS spectra were extracted and searched against a mouse reference database from the National Center for Biotechnology Information using the SEQUEST Sorcerer algorithm (version 2.0, SAGE-N). Searching parameters included mass tolerance of precursor ions (±50 ppm) and product ion (±0.5 m/z), partial tryptic restriction, with a dynamic mass shift for oxidized Met (+15.9949), four maximal modification sites, and three maximal missed cleavages. Only b and y ions were considered during the database match. To evaluate false discovery rate, all original protein sequences were reversed to generate a decoy database that was concatenated to the original database. The false discovery rate was estimated by the number of decoy matches (nd) and the total number of assigned matches (na). False discovery rate = 2 × nd/na, assuming the mismatches in the original database were

the same as in the decoy database. To remove false positive matches, assigned peptides were grouped by a combination of trypticity (fully, partial, and non-tryptic) and precursor ion charge state (1+, 2+, and 3+). Considering shift from expected precursor m/z (<10 ppm) and by dynamically increasing XCorr (minimal 1.8) and ΔC_n (minimal 0.05) values, protein false discovery rate was reduced to less than 5% (and less than 3% for proteins identified by a single peptide match). Proteins were quantified using the Abundance Index, which is defined as the spectral counts divided by the number of peptides per protein (19, 20).

TDP-43 RIP—Rat cortical neurons (14 days *in vitro*) were isolated from embryonic day 18 rat brains and cultured as described previously (9). Cells were rinsed with PBS, lysed using polysomal lysis buffer (100 mM KCl, 5 mM $MgCl_2$,10 mM HEPES pH 7.0, 0.5% Nonidet P-40, 1 mM DTT, 100 units ml^{-1} RNase Out, 400 μM vanadyl ribonucleoside complexes, 1× protease inhibitors), and sonicated using a Biorupter® UCD200 to fragment the RNA (30 s on, 30 s off, repeated three times), and then lysates were precleared. Supernatants were diluted 10-fold in NT2 buffer (supplemented with 200 units ml^{-1} RNase Out, 400 μM vanadyl ribonucleoside complexes, 20 mM EDTA) and added to antibody-protein A beads. Both TDP-43 and nonspecific rabbit IgG antibodies were affinity-purified using protein A beads. Immunoprecipitation occurred for 2 h. Beads were washed with ice-cold NT2 buffer before eluting the RNP components and RNA with the addition of RNA-Stat60. The aqueous phase was separated by adding chloroform, and RNA was precipitated from the aqueous phase using 70% ethanol. Isolated RNA was treated with DNase I to remove any genomic DNA contamination. The cDNA libraries were generated as per the Illumina manufacturer's instructions accompanying the RNA sample kit (part number 1004898). Briefly, the isolated RNA was fragmented using the provided fragmentation buffer, and first strand cDNA was generated using SuperScript II followed by second strand synthesis using DNA polymerase I. The cDNA was end-repaired using a combination of T4 DNA polymerase, *Escherichia coli* DNA polymerase I large fragment (Klenow polymerase), and T4 polynucleotide kinase. The blunt, phosphorylated ends were treated with Klenow fragment (3′ to 5′ exonuclease-minus) and dATP to yield a protruding 3-A base for ligation of the Illumina adapters, which have a single T base overhang at the 3′ end. After adapter ligation, cDNA was PCR-amplified with Illumina primers for 15 cycles, and library fragments of ~250 bp (insert plus adaptor and PCR primer sequences) were band-isolated from an agarose gel. The purified cDNA was captured on an Illumina flow cell for cluster generation. Libraries were sequenced on the Illumina GA IIx genome analyzer following the manufacturer's protocols.

Analysis of RIP Reads—The rat genome (build rn4) was downloaded from the University of California, Santa Cruz (UCSC) Genome Browser (21) in May 2010. We used the RefSeq annotations (22) for rat and downloaded the associated table from the UCSC Genome Browser in May 2010. 5′-UTR, 3′-UTR, and coding region as well as intron and exon definitions were based on this RefSeq annotation.

The read lengths from both TDP-43 and control libraries were 36 nt (nucleotides) long. These reads were mapped to the rat genome using the Bowtie software (23). To ensure that the highest possible coverage was obtained, the reads were first truncated 1 nt at a time so that the length range was 12–36 nt, yielding a total of 24 different search files. Next, the error rates were compared in mapping with 0–3 mismatches. It was found that using the whole length of the reads (36 nt) allowing two mismatches gave the best trade-off between coverage and mapping accuracy (supplemental Fig. S1). Given the nature of the sequencing protocol, it was not possible to differentiate between reads coming from the two different strands. Using the mapped Bowtie output, the number of reads that map to a particular region (intron/exon, 5′-UTR, 3′-UTR, coding sequence) of each gene was calculated.

Calculation of Effective Length—The number of reads that map to a particular gene is strongly correlated with its length. Furthermore, due to shared/repetitive sequences between genes, not all 36-nt reads can be mapped uniquely to the genome. Therefore, it is not enough to simply divide the number of uniquely mapped reads by the total length of the gene to calculate the density of reads per unit sequence length.

Currently, there is no existing information about how mappable a given sequence is for the rat genome, so we generated our own mappability track for the accurate calculation of the density of reads within each gene. For every position in the rat genome, we took 36 nt starting at that nucleotide and mapped this sequence using the Bowtie software back to the rat genome, allowing no mismatches. We assigned a mappability score of 1 if this sequence was the only instance in the genome. Otherwise, we assigned a mappability score of 0. Finally, we defined the effective length of a region as the sum of mappability scores of all positions in that region.

Unbiased Search for Frequent Short Sequences in the TDP-43 Library—One way to identify common short sequences that are preferentially bound by TDP-43 is to search for sequence fragments that are enriched in the TDP-43 library when compared with the control library. To determine how many times a short sequence appears in the TDP-43 library, we first found all possible 12-nt sequence fragments by shifting one nucleotide at a time to produce 24-fragments from each 36-nt read in the library. For each read, the same 12-nt sequence could appear more than once. In these cases, we only counted a single occurrence of these 12 nt. Due to the high computational cost of this operation, we decided to use a randomly selected subset of ~10 million reads from the TDP-43 library. After ranking the 12-nt sequences by the number of occurrences among these reads, we took the most frequent 100 fragments, and we counted the total number of occurrences of only these sequences in both the entire TDP-43 and the control libraries. The -fold enrichment is calculated using Fisher's exact test. Because we used the same dataset to generate and test our hypothesis, we corrected for the multiple hypotheses testing problem using Bonferroni correction. We multiplied our p values by 4–12 to obtain the adjusted p values.

Identification of TDP-43 Targets—For each gene, we first determined the number of reads mapping to its exons and

introns and calculated the total intronic and exonic effective length of all genes. Then, we defined the exonic read density of each gene as equal to the number of reads mapped to its exons divided by its total exonic effective length. We also calculated intronic read density analogously.

Using the reads from the TDP-43 library, we ranked all RefSeq genes based on either exonic or intronic read density, obtaining two separate ranked lists. Then, for each of these genes, we calculated the ratio of exonic and intronic read densities in the TDP-43 library to the control library. The top 25% genes from these lists were analyzed further after filtering out genes for which the ratio of read density between the TDP-43 sample and the control library was less than the ratio of total number of reads in the TDP-43 library to that in the control library (1.0547). This filter eliminated more genes from the exonic set than the intronic set. Among the filtered out genes were highly abundant transcripts. We defined three groups within TDP-43 targets. The "exonic targets" of TDP-43 were genes that ranked in the top 25% in exonic read density and had at least 1.0547-fold more exonic reads in the TDP-43 library when compared with the control library. The "intronic targets" of TDP-43 were defined similarly. The genes that appeared in the top 25% of both exonic and intronic read density and had a greater than 1.0547-fold difference in both exonic and intronic ratio of TDP-43 reads to control were defined as the set of "dual targets." For a summary of the computational analysis of the libraries, see supplemental Fig. S2.

Functional and Statistical Analyses—The functional analysis of TDP-43 targets was performed using the FuncAssociate software (24, 25). Briefly, enrichment for each Gene Ontology classification was calculated using Fisher's exact test. The *p* values were adjusted for the multiple hypotheses testing problem using a resampling approach (24). We specified the set of all RefSeq IDs as the universe of all genes unless otherwise specified. All other statistical analyses were carried out using the R 2.9 software package.

Search for Consensus Binding Motifs of Proteins That Co-purify with TDP-43—We searched the TDP-43 library for the consensus binding motifs and their reverse complements of PTPB2 (CTCTCTCTCTCT), hnRNP-A2/B1 (TTAGGGTTAGGG), and hnRNPC (CTTTACATTTG) and as a negative control PPARγ (AGGTCAXAGGTCA). We compared the number of occurrences of these motifs in the TDP-43 library with the control one to determine -fold enrichment.

RESULTS AND DISCUSSION

TDP-43 High Molecular Mass Complexes Are Associated with RNAs—We first determined the native state of TDP-43 in HeLa nuclear extracts using gel filtration. We found that TDP-43 elutes in a high molecular mass range with a peak elution at >500 kDa and less prominently in a lower molecular mass range (Fig. 1*A*). We show that with the addition of RNase A, TDP-43 elution in the high molecular mass range could be shifted to the lower molecular mass range (Fig. 1*A*). A similar pattern was observed for hnRNPA1 in control lysates as well as lysates treated with RNase A but not for lamin A/C (Fig. 1*A*). We noticed that nuclease treatment resulted in the presence of ~35- and ~25-kDa TDP-43 bands, which are thought to be associated with neuropathology. To test whether TDP-43 exists in high molecular mass complexes in the brain, we took rat brain nuclear extracts and applied the extracts to a gel filtration column. Similar trends were observed as in HeLa nuclear extracts (Fig. 1*B*). TDP-43 elution in a lower molecular mass range of ~100 kDa probably reflects a TDP-43 dimer (Fig. 1, *A* and *B*) (26). The observation that TDP-43 prominently exists in high molecular mass complexes (which probably are functional units) is consistent with the association of TDP-43 with RNAs.

Purification and Deep Sequencing of TDP-43-associated RNAs—Native TDP-43 RNA targets in neurons have not been identified. To identify RNAs associating with TDP-43, we used a modified RNA immunoprecipitation method followed by deep sequencing (RIP-seq) (27) from primary cultured rat cortical neurons (Fig. 1*C*). We obtained 30.6 and 28.9 million 36-nt reads from the TDP-43 and control libraries, respectively. Of these, 47.1% (TDP-43) and 17.7% (control) mapped uniquely to the rat genome (build rn4), respectively. This 2.75-fold difference in mappability between the two libraries indicates that the TDP-43 library was enriched for regions of high sequence complexity. Consistent with this, 7.98 million reads in the TDP-43 library mapped uniquely to annotated RefSeq genes *versus* only 1.97 million reads from the control library (supplemental Data Files S1 and S2).

Genomic Distribution of TDP-43 Binding Sites—In our TDP-43 library, 1.33 million reads mapped to exonic regions of genes *versus* 6.65 million reads to intronic regions (Fig. 1*D*). We calculated read densities (number of reads per 1,000 mappable nucleotides per million reads (*mRPKM*)), which takes into account differences in lengths of various regions (see "Experimental Procedures"). Within exons, TDP-43 reads exhibited 2.2 and 2.3 times higher read density in 3′-UTRs when compared with 5′-UTRs or open reading frames (coding sequence), respectively (Fig. 1*E*). This suggests that TDP-43 binding sites are enriched within 3′-UTRs. Interestingly, we observed several cases where TDP-43 reads extended beyond the annotated 3′-UTR end (data not shown), suggesting that several of these genes might have alternative isoforms in which use of distal polyadenylation sites results in longer 3′-UTRs. This is consistent with previous analyses suggesting that in differentiated tissues, genes tend to have longer 3′-UTRs (28). With regard to reads mapping to introns, we did not find any overall bias in the distribution of reads (8.33×10^{-7}, 6.94×10^{-7}, and 8.33×10^{-7}, for coding sequence intron, 3′-UTR intron, and 5′-UTR intron read density, respectively).

We then undertook an unbiased search for frequent short sequences in our TDP-43 library (see "Experimental Procedures"). Given our sequencing procedure, it was not possible to differentiate between the sense and antisense strands (for example, $(TG)_6$ and $(AC)_6$ were considered as a single motif). We discovered $(TG)_6$ with its reverse complement $(AC)_6$ to be the most frequent 12-nt sequence in our TDP-43 library such that 1.30 million reads contain one or the other. The number of these sequences in TDP-43 showed a remarkable 55.1-fold enrichment when compared with the control library

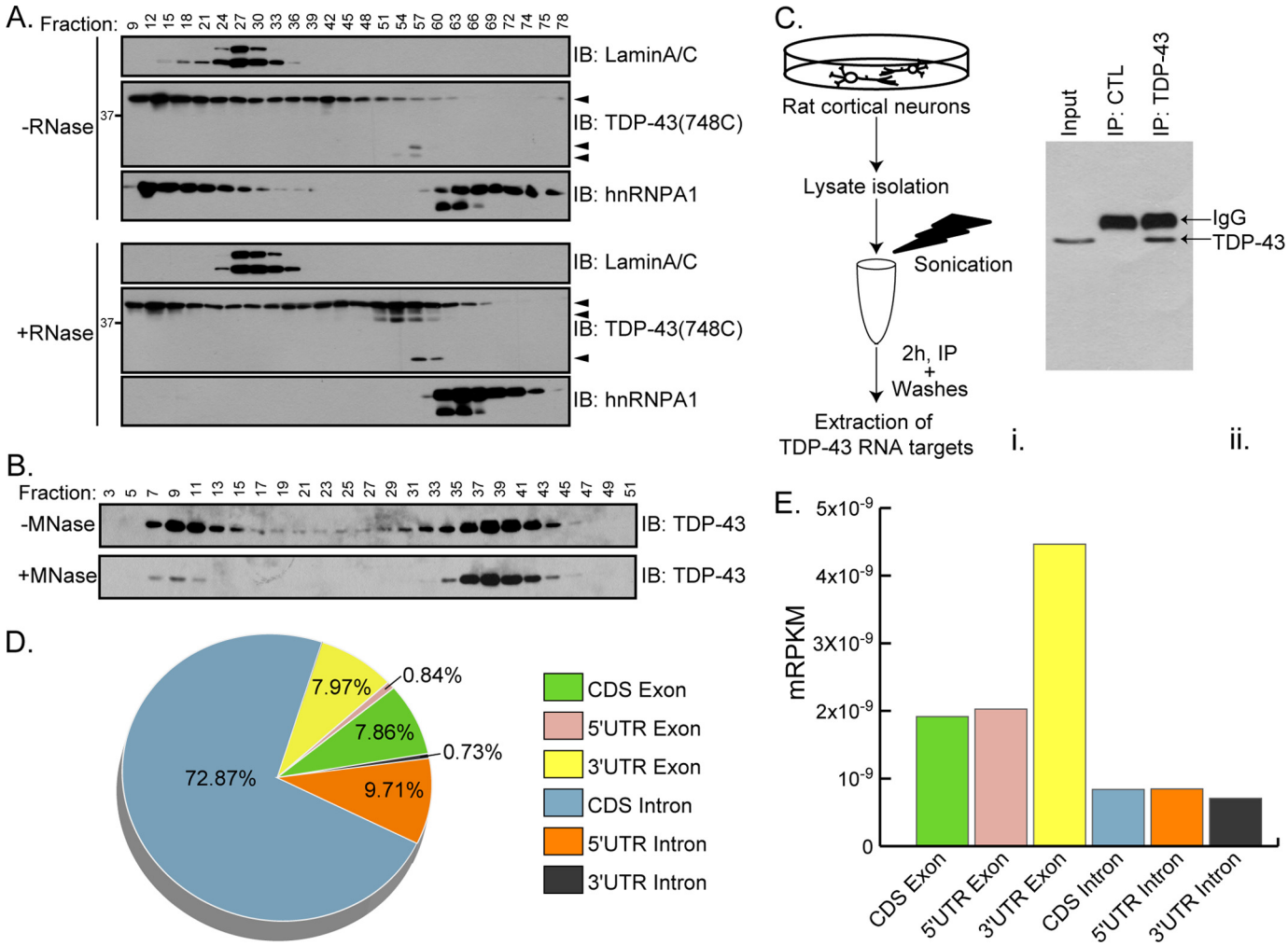

FIGURE 1. **Genomic distribution of reads from TDP-43 RNA library.** *A*, Western blot (*IB*) of fractions from HeLa nuclear extracts (± RNase A) applied to a size exclusion column, blotted for TDP-43, hnRNPA1, and lamin A/C. Fraction 6 = blue dextran (2000 kDa), fraction 30 = apoferritin (443 kDa), fraction 40 = alcohol dehydrogenase (150 kDa), and fraction 54 = bovine serum albumin (54 kDa) (data not shown). *B*, Western blot of fractions from rat brain nuclear extracts (± micrococcal nuclease (± *MNase*)) blotted for TDP-43. Fraction 5 = blue dextran (2,000 kDa), fraction 45 = bovine serum albumin (54 kDa), and fraction 54 = carbonic anhydrase (29 kDa) (data not shown). Note that different size exclusion columns were used in *A* and *B*. *C*, *panel i*, diagram of TDP-43 RIP method. *C*, *panel ii*, representative Western blot of TDP-43 RIP. *IP:CTL*, control immunoprecipitation. *D*, distribution of raw reads from the TDP-43 library mapped to exonic and intronic genes regions. *CDS*, coding sequence. *E*, read density, number of reads per 1,000 mappable nucleotides per million reads (*mRPKM*) of gene regions from the TDP-43 library.

(Fisher's exact test; adjusted p value $< 3.4 \times 10^{-8}$). Previously, TDP-43 was suggested to bind to TG repeats, which is in agreement with our analysis. More surprisingly, we found motifs of class $(TG)_nTA(TG)_m$ (and their reverse complement) to be highly enriched in the TDP-43 library as well (Fig. 2 and supplemental Table S1). For example, for the sequence $(CA)_3TA(CA)_2$, the odds ratio was 95.8 (Fisher's exact test, adjusted p value $< 3.4 \times 10^{-8}$). The most frequent variant was in instances where $n = m + 1$ and $n > 1$, $m > 1$ in the motif $(TG)_nTA(TG)_m$, indicating that adenine is in the middle of the motif.

We also looked at the distribution of reads within each gene from the TDP-43 library. We analyzed exonic and intronic reads separately and observed that TDP-43 reads for most genes are spread across the entirety of the gene for both exons and introns (supplemental Fig. S3). However, there were two minor sets of genes, one that had most of their exonic reads in the 3′-UTR and another that had most of their intronic reads in the 5′-UTR (supplemental Fig. S3, *A*, *panel iii*, and *B*, *panel i*).

We next defined TDP-43 RNA targets. A total of 4,352 genes passed the enrichment filter (see "Experimental Procedures" and supplemental Table S2) and are referred to as TDP-43 RNA targets henceforth. There were 1,971 TDP-43 RNA targets that had predominantly intronic reads and 910 targets that had predominantly exonic reads, whereas 1,471 targets had both exonic and intronic reads (Fig. 3, *A*–*C*). These three categories of genes are henceforth referred as: 1) exonic, 2) intronic, and 3) dual RNA targets of TDP-43.

We then asked whether these three categories of TDP-43 RNA targets differed with respect to their functions. We used the Gene Ontology database and searched for statistically significant enrichment of functional categories within these three sets. TDP-43-targeted RNAs were enriched in diverse functional categories (supplemental Tables S3–S6). Remarkably, all three sets revealed distinct functional enrichment

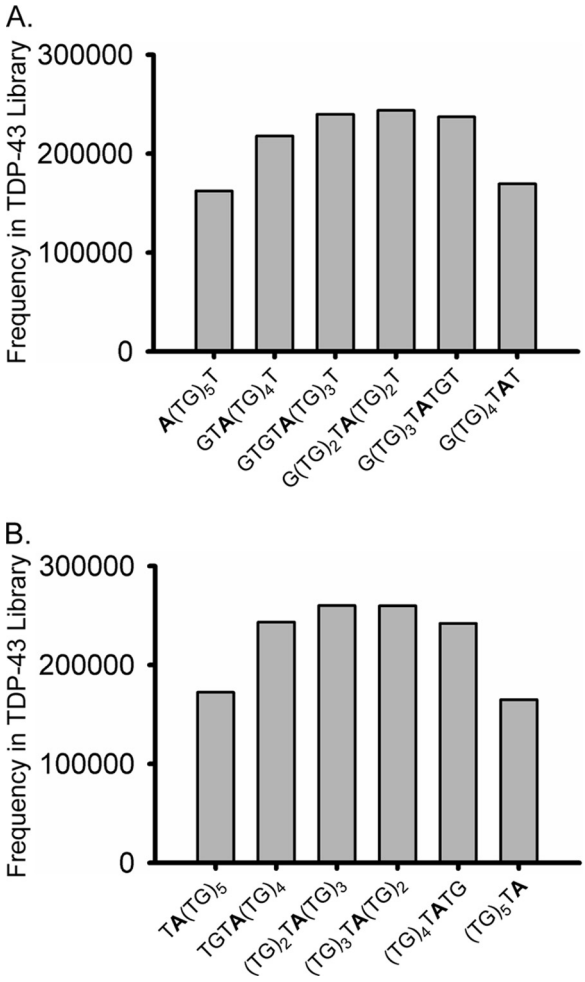

FIGURE 2. **Identification of short nucleotide sequences enriched in TDP-43 library.** *A*, short nucleotide sequences of type $(TG)_nTA(TG)_m$ are highly enriched in the TDP-43 library when compared with the control library. The number of 36-nt reads with at least one occurrence of each variant is shown. The graph depicts sequences where adenine replaces guanine in positions 1, 3, 5, 7, 9, or 11. The shape of the distribution reveals that adenine tends to appear in the middle of the sequence. *B*, same as *panel A*, except that the adenine is in positions 2, 4, 6, 8, 10, or 12.

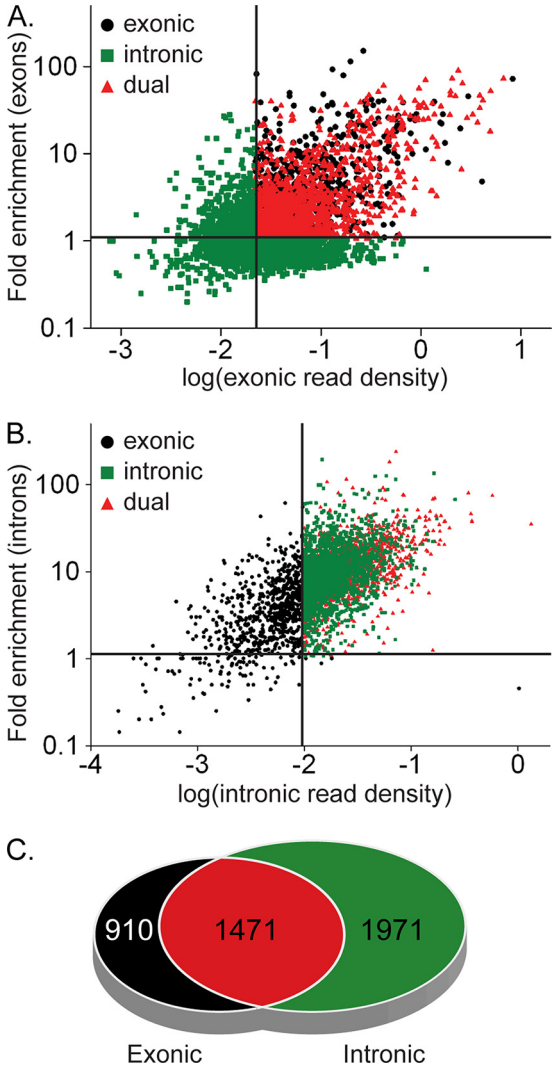

FIGURE 3. **Distribution of the read density for the top 25% of TDP-43 RNA targets.** The scatter plot depicts exonic (*A*) and intronic (*B*) read density of TDP-43 RNA targets and -fold enrichment of reads in the TDP-43 library relative to the control library. *C*, summary of the number of genes in each of the TDP-43 RNA target categories.

profiles (Fig. 4, *A–C*). In particular, genes with TDP-43 exonic reads were enriched for Gene Ontology terms related to splicing and RNA processing and maturation (Fig. 4*A*, *panel i*, and supplemental Table S3), whereas genes with intronic TDP-43 reads were enriched for terms associated with synaptic formation and function and in regulation of neurotransmitter processes (Fig. 4*B*, *panel i*, and supplemental Table S4); genes with dual TDP-43 reads were enriched for terms related to various aspects of development (Fig. 4*C*, *panel i*, and supplemental Table S5). These results provide an important perspective about how TDP-43 regulates different biological and pathobiological processes.

One major group of TDP-43 exonic targets consists of transcripts for proteins involved in RNA metabolism (Table 1), for example splicing factor arginine/serine-rich 1 (SFRS1) and RNA-binding proteins: TDP-43 itself (Fig. 4*A*, *panel ii*), FUS/TLS, hnRNPs (A1, A2/B1, C, D, Dl, F, H1, K, M, R, U), and poly(A)-binding protein cytoplasmic 1 (PABPC1). Notably, there was a particular enrichment of reads in 3′-UTRs of some of these genes, like TDP-43 (Fig. 4*A*, *panel ii*). Together with the observation that in *Tardbp*$^{+/-}$ heterozygous mice there is a compensatory increase in TDP-43 RNA levels (9), the current work supports a model wherein TDP-43 binds to the 3′-UTR and regulates the stability or translational efficiency of its own RNA transcript. Moreover, our RIP-seq studies, when combined with our proteomics analysis (see below), suggest that TDP-43 and other factors involved in RNA metabolism mediate post-transcriptional regulation in a complex regulatory network analogous to gene regulation by transcriptional factors.

Transcripts with predominantly intronic reads had enriched Gene Ontology terms related to synaptic formation and function and in regulation of neurotransmitter processes (Table 2). Prominent examples included transcripts for neurexin (*Nrxn1–3*) and neuroligin (*Nlgn1–3*), alternative pre-mRNA splicing of which specifies a trans-synaptic signaling code (29, 30) and slit homolog (*Slit1,3*) (Fig. 4*B*, *panel ii*), *Slit3* being the primary transcript for miR218-2, which is also

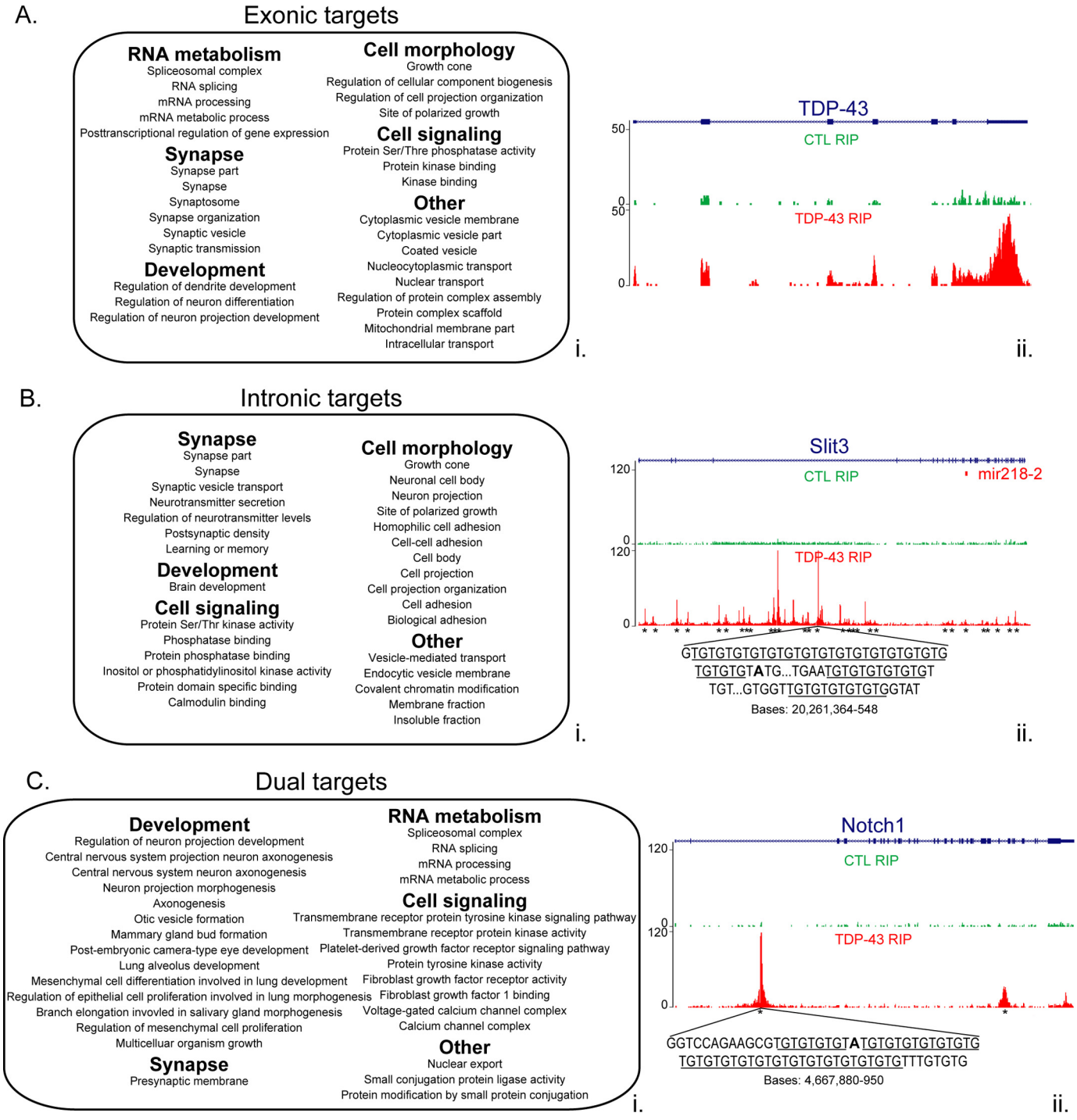

FIGURE 4. **Functional categorization of top TDP-43 RNA targets.** *A–C*, summary of the top 30 most enriched Gene Ontology terms in TDP-43 RNA targets in exonic (*A*, *panel i*); intronic (*B*, *panel i*); and dual sets (*C*, *panel i*). For the complete functional listing, see supplemental Tables S3–S5. *A–C*, *panel ii*, snapshots of genes representing each category of binding. *A*, *panel ii*, exonic-TDP-43; *B*, *panel ii*, intronic-Slit3; and *C*, *panel ii*, dual Notch1. The number of uniquely mapped reads to the gene were shown for both the TDP-43 library and the control (*CTL*) library. The *asterisk* indicates TG-rich regions. Note the adenine (*bolded*) in the $(TG)_n$ motifs shown for Slit3 and Notch1.

involved in neuron differentiation (31). We also noticed that reads from the TDP-43 library were mapped to genomic regions where a number of known miRNAs are annotated (the data are deposited in the National Center for Biotechnology Information (NCBI) GEO database, GSE25032).

TDP-43 RNA targets bound in both intronic and exonic regions were particularly enriched for genes involved in CNS development and differentiation (Table 3). Genes from this category include: notch homolog 1 (*Notch1*) (Fig. 4*C*, *panel ii*), neurotrophic tyrosine kinase receptor types 2 and 3 (*Ntrk2,3*), myelin transcription factor 1-like (*Myt1l*), and dual specificity tyrosine phosphorylation-regulated kinase 1A (*Dyrk1a*). In our previous study addressing the biological impact of deleting *Tardbp* in mice, we determined that TDP-43 is necessary for embry-

TABLE 1

TDP-43 RNA targets associated with RNA metabolism

For complete listing of genes, see supplemental Table S1 and NCBI GEO accession number GSE25032.

Gene ID	Refseq ID	Gene ID	Refseq ID	Gene ID	Refseq ID	Gene ID	Refseq ID	Gene ID	Refseq ID	Gene ID	Refseq ID
Aars	NM_001100517	Dgcr8	NM_001105865	Hif3a	NM_022528	Nop58	NM_021754	Rnaseh2a	NM_001013234	Srp19	NM_001106157
Aars2	NM_001106891	Dhx8	NM_001047844	Hnrnpa1	NM_017248	Npm1	NM_012992	Rnasen	NM_001107655	Srp54a	NM_053871
Aarsd1	NM_001034109	Dhx9	NM_001107184	Hnrnpa2b1	NM_001104613	Nudt21	NM_001039004	Rngtt	NM_001107923	Srpk2	NM_001106575
Aco1	NM_017321	Dhx30	NM_001013249	Hnrnpc	NM_001025633	Nxf1	NM_021579	Rnps1	NM_001011890	Srrm1	NM_001107986
Adar	NM_031006	Dimt1l	NM_001106408	Hnrnpdl	NM_001033696	Pabpc1	NM_134353	Rpl9	NM_001007598	Ssb	NM_031119
Adarb1	NM_001111055	Dis3l	NM_001008380	Hnrnpf	NM_001037286	Pabpc2	NM_001106151	Rpl22	NM_031104	Ssu72	NM_001025657
Aimp1	NM_053757	Dnajb11	NM_001015021	Hnrnpk	NM_057141	Pabpc4	NM_001100538	Rpl37	NM_031106	Stau1	NM_053436
Akap7	NM_001001801	Dnajc8	NM_001013168	Hnrnpr	NM_175603	Pabpn1	NM_001135008	Rps7	NM_031570	Stau2	NM_001007149
Ankrd17	NM_001105999	Dusp11	NM_001025650	Hnrnpu	NM_057139	Paip1	NM_001108937	Rps9	NM_031108	Strap	NM_001011969
Apobec4	NM_001017492	Eif1ay	NM_001106963	Hnrpd	NM_001082540	Papd4	NM_001008372	Rps13	NM_130432	Strbp	NM_053416
Ascc3l1	NM_001037766	Eif2ak4	NM_001105744	Hnrph1	NM_080896	Papd5	NM_001107416	Rps17	NM_017152	Supt5h	NM_001107497
Atxn1	NM_012726	Eif2c2	NM_021597	Hnrpm	NM_053876	Papola	NM_001108056	Rps18	NM_213557	Surf6	NM_001015014
Auh	NM_001108407	Eif2s1	NM_019356	Hsp90b1	NM_001012197	Pdcd11	NM_001107604	Rps24	NM_031112	Syf2	NM_133417
Bat1	NM_133300	Eif3b	NM_001031640	Ilf2	NM_001047886	Pdcd4	NM_022265	Rpusd2	NM_001135845	Syncrip	NM_001047916
Baz2a	NM_001107158	Eif3e	NM_001011990	Ilf3	NM_053412	Pes1	NM_001044228	Rrp1	NM_001012073	Taf13	NM_001107716
Bcas2	NM_001106458	Eif4a2	NM_001008335	Ints2	NM_001100630	Phax	NM_173133	Rtcd1	NM_001004227	Taf1a	NM_001037204
Bmpr2	NM_080407	Eif4e	NM_053974	Ints6	NM_001047904	Phf5a	NM_138888	Safb	NM_022394	Taf9	NM_001012463
Brunol4	NM_001107400	Eif4e2	NM_001108808	Ints10	NM_001134416	Phrf1	NM_139093	Sarnp	NM_001033070	Tardbp	NM_001011979
Carm1	NM_001030041	Eif4e3	NM_001106612	Ireb2	NM_022863	Plrg1	NM_021757	Sart1	NM_031596	Tars	NM_001006976
Cars	NM_001106319	Eif4g3	NM_001106693	Isy1	NM_001014188	Pnpt1	NM_001142371	Sart3	NM_001107156	Tarsl2	NM_001014020
Casc3	NM_147144	Eif4h	NM_001006957	Jakmip1	NM_001033894	Pnrc2	NM_001103360	Sbds	NM_001008289	Tdrkh	NM_001014038
Ccar1	NM_001108535	Eif5a	NM_001033681	Jund	NM_138875	Polr1e	NM_001107938	Serbp1	NM_145086	Tgs1	NM_001107904
Ccnh	NM_052981	Elac1	NM_001107406	Khdrbs2	NM_133318	Polr2g	NM_053948	Serinc1	NM_182951	Thoc5	NM_001012153
Ccnt1	NM_001108110	Elavl2	NM_173309	Khdrbs3	NM_022249	Ppargc1a	NM_031347	Sf1	NM_001110793	Thra	NM_031134
Cd2bp2	NM_001106297	Elavl3	NM_172324	Khsrp	NM_133602	Ppp1r10	NM_022951	Sf3a3	NM_001025698	Tia1	NM_001012096
Cdc40	NM_001108538	Elavl4	NM_001077651	Larp6	NM_001108154	Prim1	NM_001008768	Sf3b1	NM_053426	Tial1	NM_001013193
Cdc5l	NM_053527	Elp2	NM_001034145	Lars	NM_001009637	Prkcsh	NM_001106806	Sf3b2	NM_001106326	Tnrc4	NM_001109190
Cdk105	NM_134415	Eprs	NM_001024238	Lgtn	NM_001017489	Prkra	NM_001024780	Sfpq	NM_001025271	Top1	NM_022615
Cdkal1	NM_001108413	Eral1	NM_001013229	Lkap	NM_133421	Prkrip1	NM_001098793	Sfrs1	NM_001109552	Tra2a	NM_001126296
Cebpg	NM_012831	Ets1	NM_012555	Lmo4	NM_001009708	Prpf4b	NM_001011923	Sfrs2	NM_001009720	Tra2b	NM_057119
Cirbp	NM_031147	Exosc3	NM_001107936	LOC498453	NM_001025735	Prpf6	NM_001079766	Sfrs3	NM_001047907	Trit1	NM_001108676
Cnbp	NM_022598	Exosc7	NM_001100725	Lrpprc	NM_001008519	Prpf18	NM_138523	Sfrs4	NM_001108685	Trmt1	NM_001013870
Cnot4	NM_001037782	Exosc9	NM_001025406	Luc7l3	NM_001108291	Prpf38b	NM_001024305	Sfrs5	NM_019257	Tsn	NM_021762
Cpeb2	NM_001108361	Farsa	NM_001024237	Magoh	NM_001100536	Prpf39	NM_001108026	Sfrs6	NM_001014185	U2af1l4	NM_001008775
Cpsf2	NM_001106753	Fbl	NM_001025643	Mapkapk2	NM_178102	Pspc1	NM_001025672	Sfrs8	NM_001034924	Uhmk1	NM_017293
Cpsf6	NM_001106785	Fech	NM_001108434	Mars	NM_001127659	Ptbp2	NM_001005555	Sfrs9	NM_001009255	Upf3b	NM_001135873
Cpsf7	NM_001014245	Fip1l1	NM_001008295	Matr3	NM_019149	Pum1	NM_001108684	Sfrs11	NM_001035255	Urm1	NM_001137562
Crcp	NM_053670	Fto	NM_001039713	Mcts1	NM_001044237	Pum2	NM_001106715	Sfrs12	NM_020092	Usf1	NM_031777
Cryz	NM_001012183	Fubp1	NM_001037653	Mecp2	NM_001115025	Pus7	NM_001170589	Sfrs15	NM_001037347	Wbp4	NM_053766
Csda	NM_031979	Fus	NM_001012137	Mex3c	NM_001107377	Qk	NM_001115021	Sirt7	NM_001107073	Wibg	NM_001108986
Csdc2	NM_001170542	Fusip1	NM_001025738	Mov10	NM_001107711	Rab3a	NM_013018	Slc4a1ap	NM_001106709	Xab2	NM_139109
Cstf2t	NM_001107586	Fxr1	NM_001012179	Mphosph10	NM_001106340	Rbm3	NM_053696	Slu7	NM_001100550	Xpo1	NM_053490
Cstf3	NM_001077672	Fxr2	NM_001100647	Mrm1	NM_001108832	Rbm5	NM_001100548	Smc1a	NM_031683	Xrn2	NM_001108596
Cugbp1	NM_001025421	Fyttd1	NM_001047899	Mrpl44	NM_001031650	Rbm9	NM_001079895	Smndc1	NM_001025400	Ythdc1	NM_133423
Cugbp2	NM_017197	Gm672	NM_001108891	Msi1	NM_148890	Rbm10	NM_152861	Snrnp35	NM_001014127	Zc3h14	NM_138920
Ddit3	NM_024134	Gtf2a1	NM_022208	Mtpap	NM_001107359	Rbm16	NM_139094	Snrnp48	NM_001106107	Zcchc11	NM_001107953
Ddx1	NM_053414	Gtf2e2	NM_001107318	Naf1	NM_001024772	Rbm34	NM_001014015	Snrnp70	NM_001108483	Zcchc17	NM_001109267
Ddx21	NM_001037201	Gtf2f1	NM_001007711	Nars	NM_001025635	Rbm39	NM_001013207	Snrpd1	NM_001106163	Zcrb1	NM_001034940
Ddx23	NM_001106793	Gtf3c6	NM_001108537	Ncbp1	NM_001014785	Rg9mtd1	NM_001008337	Snrpg	NM_001135085	Zfp219	NM_001007681
Ddx25	NM_031630	Gtfc2	NM_001025120	Ncl	NM_012749	RGD1307890	NM_001037192	Snupn	NM_001004270	Zfp346	NM_001107338
Ddx26	NM_139098	Hars2l	NM_001014012	Nip7	NM_138847	RGD1309888	NM_001014243	Snw1	NM_001109279	Zranb2	NM_031616
Ddx50	NM_001013198	Hdlbp	NM_172039	Nop56	NM_001025732	RGD1560124	NM_031065	Srp9	NM_001126095	Zrsr1	NM_001017504

onic development and is highly expressed in the developing CNS of embryos and into adulthood (9). The RIP-seq data from the current study have allowed us to appreciate the broad spectrum of transcripts regulated by TDP-43 in the CNS and the functional importance of TDP-43 during development and maintenance of the CNS.

Some of the TDP-43 RNA targets have also been associated with neurodegenerative diseases, for example, *Tardbp*, *Fus*, progranulin (*Grn*), α-synuclein (*Scna*), microtubule-associated protein Tau (*Mapt*), adenosine deaminase, RNA-specific B1 (*Adarb1*), and ataxin 1 and -2 (*Atxn1,2*) (3, 4, 32–35) (Table 4). Interestingly, a recent study showed a positive correlation between ADARB1-absent neurons and TDP-43-positive cytoplasmic inclusions (34). Given the numerous transcripts TDP-43 regulates, including an apparent self-regulation mechanism, it is conceivable that the TDP-43 inclusions that are present in ALS and frontotemporal lobar degeneration with ubiquitin-positive inclusions and subsequent neurodegeneration could be a result of, or alternatively could result in, a loss of TDP-43 function for a subset of its RNA targets.

TABLE 2

TDP-43 RNA targets associated with synaptic function

For complete listing of genes, see supplemental Table S1 and NCBI GEO accession number GSE25032. Genes present in multiple categories are marked with * (RNA metabolism) or + (nervous system development).

Symbol	RefSeq ID	Symbol	RefSeq ID	Symbol	RefSeq ID	Symbol	RefSeq ID	Symbol	RefSeq ID	Symbol	RefSeq ID	Symbol	RefSeq ID
Abat	NM_031003	Cadps	NM_013219	Dvl1*	NM_031820	Grin2b	NM_012574	Nlgn1+	NM_053868	Rims1	NM_052829	Sptbn2	NM_019167
Abi1	NM_024397	Camk2a	NM_012920	Efna2	NM_001168670	Grip2	NM_138535	Nlgn2+	NM_053992	Rims2	NM_053945	Strn	NM_019148
Actr3+	NM_031068	Camk2d	NM_012519	Epha4	NM_001162411	Grm5+	NM_017012	Nlgn3+	NM_134336	Rims3	NM_022931	Stx1a	NM_053788
Add1	NM_016990	Cask	NM_022184	Epha7	NM_134331	Homer2	NM_053309	Nos1ap	NM_138922	Rims4	NM_170666	Stx7	NM_021869
Adrbk2	NM_012897	Cd24+	NM_012752	Erbb4	NM_021687	Igsf9	NM_001107197	Nptn+	NM_019380	Rps6kb1	NM_031985	Stxbp1	NM_013038
Agrn+	NM_175754	Cdh2+	NM_031333	Erc2	NM_170787	Insr	NM_017071	Npy1r	NM_001113357	Samd4a	NM_001107254	Stxbp5	NM_178346
Agtpbp1	NM_001106100	Cdk5+	NM_080885	Fmr1*	NM_052804	Itgb1	NM_017022	Nrcam	NM_013150	Scamp1	NM_001100636	Sumo1	NM_001009672
Akap1*	NM_053665	Chrm1	NM_080773	Freq+	NM_024366	Itsn1	NM_001136096	Nrxn1+	NM_021767	Scamp5	NM_031726	Sv2a	NM_057210
Akap5+	NM_133515	Chrna7	NM_012832	Gabbr1	NM_031028	Kcnd2	NM_031730	Nrxn2+	NM_053846	Sema4c	NM_001106902	Sv2b	NM_057207
Amph	NM_022217	Cln3	NM_001006971	Gabbr2	NM_031802	Kcnma1	NM_031828	Nrxn3+	NM_053817	Sema4f+	NM_019272	Svop	NM_134404
Ank3	NM_031805	Clstn3	NM_134376	Gabra2	NM_001135779	Kif3a	NM_053377	Nsf	NM_021748	Sept11	NM_001107208	Syn2	NM_001034020
Apba1	NM_031779	Cnn3	NM_019359	Gabra3	NM_017069	L1cam	NM_017345	Ntm+	NM_017354	Sept3	NM_019375	Syn3	NM_017109
Apba2	NM_031780	Cplx2	NM_053878	Gabrb1	NM_012956	Lin7b	NM_021758	Nufip1*	NM_001007758	Sh3kbp1	NM_053360	Syngap1+	NM_181092
Apbb1	NM_080478	Ctnnb1	NM_053357	Gabrb2	NM_012957	Lin7c	NM_021851	P2rx6	NM_012721	Shank1	NM_031751	Synj1*+	NM_053476
App*+	NM_019288	Ctnnd1	NM_001107740	Gabrb3	NM_017065	Lphn1	NM_022962	Pdlim5+	NM_053326	Shank2	NM_001004133	Synpr	NM_023974
Arhgef11	NM_023982	Dab1+	NM_153621	Gad2	NM_012563	Lrrc7+	NM_057142	Phactr1	NM_214457	Shank3	NM_021676	Sypl1	NM_001014263
Arrb1*	NM_012910	Dlg1	NM_012788	Gap43	NM_017195	Lzts1+	NM_153470	Picalm+	NM_053554	Siah1a	NM_080905	Syt1	NM_001033680
Atp1a2	NM_012505	Dlg2	NM_022282	Glrb+	NM_053296	Magi2	NM_053621	Pick1	NM_053460	Slc17a7	NM_053859	Syt11	NM_031667
Atp6v0d1	NM_001011927	Dlg3	NM_031639	Gnas	NM_001159653	Map1b+	NM_019217	Pja2	NM_138896	Slc1a2	NM_017215	Syt4	NM_031693
Axin1	NM_024405	Dlg4	NM_019621	Gria2	NM_017261	Map1s	NM_001106070	Ppp1cc	NM_022498	Slc1a3	NM_019225	Syt5	NM_019350
Bcan	NM_001033665	Dlgap1	NM_022946	Gria3	NM_032990	Mdm2	NM_001108099	Ppp1r9b+	NM_053474	Slit1+	NM_022953	Syt7	NM_021659
Bin1	NM_053959	Dlgap2	NM_053901	Gria4	NM_017263	Mib1+	NM_001107405	Ppt1	NM_022502	Slit3+	NM_031321	Trim9	NM_130420
Bsn	NM_019146	Dlgap3	NM_173138	Grid1	NM_024378	Myo5a	NM_022178	Prkaca	NM_001100922	Snap25+	NM_030991	Unc13b	NM_022862
Cabp1	NM_001033676	Dlgap4	NM_173145	Grik2	NM_019309	Myrip	NM_182844	Ptprf+	NM_019249	Snap91+	NM_031728	Usp48	NM_198785
Cacna1b	NM_147141	Dnm2	NM_013199	Grik3	NM_001112716	Nefm+	NM_017029	Rab14	NM_053589	Snca	NM_019169	Vamp2	NM_012663
Cacna1c	NM_012517	Doc2a	NM_022937	Grik4	NM_012572	Neto1	NM_001107371	Rab3c	NM_133536	Snip+	NM_019378	Vamp3	NM_057097
Cacnb4	NM_001105733	Doc2b	NM_031142	Grik5	NM_031508	Nf1+	NM_012609	Rab8a	NM_053998	Snph	NM_001106525	Vti1a	NM_023101
Cadm1	NM_001012201	Dpysl2+	NM_001105717	Grin2a	NM_012573	Nf2	NM_013193	Rasgrp2	NM_001082977	Sos1	NM_001100716	Wnt7a	NM_001100473

We noticed a pronounced representation of exonic reads in the 3′-UTR and intronic reads in the 5′-UTR in a subset of TDP-43 targets (data not shown). We found genes with intronic TDP-43 binding in 5′-UTRs to be enriched among several regulatory functions including regulation of transcription and metabolism (supplemental Table S6). However, we realized that the enrichment of 5′-UTR introns in regulatory genes holds true in the rat genome irrespective of TDP-43 binding (supplemental Table S7), consistent with a similar enrichment of 5′-UTR introns in regulatory genes in the human genome (36).

These results, taken together, suggest that TDP-43 regulates genes in three different modes. One is through binding sites in introns, one is through binding sites in exons, and another is through binding across both introns and exons. Our analysis also supports that, consistent with the "post-transcriptional operon" theory (37), TDP-43 regulates functionally coherent sets of genes via binding to distinct modalities.

This study is also the first report showing that endogenous TDP-43 RNA targets in a genome-wide manner. Previous studies reporting TDP-43 binding to RNAs used overexpression models or showed a correlation between knockdown of TDP-43 and changes in transcript and protein levels (13–18, 38, 50). Previously identified TDP-43 RNA targets, *HDAC6*, *APOAII*, *SMN*, and neurofilament (*NEF*), were not identified in our study. These TDP-43 RNA targets may be context-specific. Several TDP-43 RNA targets have been predicted from microarray analysis of altered cellular transcripts upon siRNA knockdown of TDP-43 (18). TDP-43 targets identified from our RIP-seq data set corresponded to some of these altered transcripts including: *Dyrk1a*, cyclin-dependent kinase 6 (*Cdk6*), insulin-like growth factor 1 receptor (*Igf1r*), laminin γ1 (*Lamc1*), structural maintenance of chromosomes protein (*Smc1a*), Rho-related BTB domain-containing (*Rhobtb2*), and protein CDV3 homolog (*Cdv3*) (18) (Tables 1 and 3 and supplemental Table S1). The altered transcripts listed by Buratti *et al.* (18) are also down-regulated following *let-7b* overexpression, which the authors interpreted as being the result of the TDP-43-miRNA interaction. Interestingly, our TDP-43 RIP-seq results indicate that there is a direct interaction between TDP-43 and these transcripts, which indicates a dual means of transcript regulation.

Identification of TDP-43 Nuclear Interactome—We immunoprecipitated endogenous TDP-43 from rodent brain nuclear extracts and analyzed the resultant precipitation products with semiquantitative mass spectrometry. Taking into consideration the abundance index (>3.33) and consistency of spectral count trend (Fig. 5*A*, under spectral counts) of two independent experiments, we reliably identified 25 co-precipitating proteins highly enriched in the TDP-43 precipitate relative to control (Fig. 5*A*). There were 34 co-purified proteins that did not meet our criteria (supplemental Table S8), which nevertheless may represent transient interacting proteins of TDP-43.

Of the 25 proteins we identified as part of a TDP-43 nuclear interactome, 16 had been previously shown to co-purify with TDP-43 (39, 40). The nine new proteins not previously reported to co-purify with TDP-43 are GM9242 (similar to hnRNPA3), MECP2, SFRS1, peroxiredoxin (PRDX1 and -2), calmodulin-like 3 (CALML3), U1 small nuclear ribonucleoprotein (SNRNP), eukaryotic translation initiation factor 5A (EIF5A), and splicing factor 3a, subunit 1 (SF3A). Many of these proteins are ubiquitously expressed, except MECP2 and PTBP2, which are highly enriched in the CNS. The TDP-43 nuclear interactome reveals that the majority of co-purified proteins are RNA-binding proteins, splicing factors, and translation factors involved in aspects

TABLE 3

TDP-43 RNA targets associated with nervous system development

For complete listing of genes, see supplemental Table S1 and NCBI GEO accession number GSE25032. Genes present in multiple categories are marked with * (RNA metabolism) or # (synaptic function).

Symbol	RefSeq ID	Symbol	RefSeq ID	Symbol	RefSeq ID	Symbol	RefSeq ID	Symbol	RefSeq ID
Acs16	NM_130739	Dixdc	NM_001037654	Kit	NM_022264	Pcm1	NM_031076	Rufy3	NM_001025127
Actb	NM_031144	Dll1	NM_032063	Klf7	NM_001108800	Pias2*	NM_053337	Ryk	NM_080402
Adcy1	NM_001107239	Dll3	NM_053666	LOC312831	NM_001014060	Picalm	NM_053554	Scn3b	NM_139097
Adnp	NM_022681	Dscam	NM_133587	Lppr4	NM_001001508	Pip5k1c	NM_001033970	Sema3f	NM_001108185
Afg3l2	NM_001134864	Dst	NM_001108208	Lrrc4c	NM_001107753	Plxnb1	NM_001108188	Sema4d	NM_001170563
Akt1	NM_033230	Dyrk1a	NM_012791	Mapk8ip3	NM_001100673	Plxnb2	NM_001108106	Sema6b	NM_053471
Apc	NM_012499	Ephb1	NM_001104528	Mapt	NM_017212	Pou3f2	NM_172085	Serpine2	NM_019197
Arhgap4	NM_144740	Ephb2	NM_001127319	Mef2d	NM_030860	Pou3f3	NM_138837	Sh2b2	NM_053669
Arhgef1	NM_021694	Ephb3	NM_001105868	Meis1	NM_001134702	Prex1	NM_001135718	Slitrk1	NM_001107283
Ascl1	NM_022384	Etv5	NM_001107082	Mycbp2	NM_001106055	Prkca	NM_001105713	Smarca4	NM_134368
Atf1	NM_001100895	Fam5b	NM_173115	Myt1l	NM_053888	Psen1#	NM_019163	Smo	NM_012807
Atf2	NM_031018	Fez1	NM_031066	Nav2	NM_138529	Pten	NM_031606	Snx27	NM_152847
Atf5	NM_172336	Fgfr2	NM_001109894	Ndel1	NM_133320	Ptk2	NM_013081	Socs2	NM_058208
Atn1	NM_017228	Fig4	NM_001047096	Ndrg2	NM_133583	Ptpn11	NM_013088	Sox11	NM_053349
Atxn10	NM_133313	Fnbp1	NM_138914	Nme1	NM_138548	Ptpra	NM_012763	Ss18l1	NM_138918
Bcl11b	NM_001108057	Gas7	NM_053484	Notch1	NM_001105721	Ptprg	NM_134356	Stk11	NM_001108069
Bcl2	NM_016993	Ghrl	NM_021669	Nr2f1	NM_031130	Ptprz1	NM_001170685	Tbce	NM_001012161
Bhlhe40	NM_053328	Gng8	NM_139185	Nrep	NM_178096	Rbm45*	NM_153306	Tiam1	NM_001100558
Bzw2	NM_134402	Gpsm1	NM_144745	Nrp1	NM_145098	RGD1306622	NM_001170487	Tp53*	NM_030989
Camk1	NM_134468	Gsk3b	NM_032080	Nrp2	NM_030869	RGD1309707	NM_001109598	Tpp1	NM_031357
Ccnd2	NM_022267	Hap1	NM_177982	Ntng2	NM_001107825	RGD1311558	NM_001079705	Ttc3	NM_001108315
Cdk5rap1*	NM_145721	Hes5	NM_024383	Ntrk2#	NM_012731	Rgnef*	NM_001108542	Ube2v2	NM_183052
Cdk5rap2	NM_173134	Id4	NM_175582	Ntrk3	NM_019248	Rhoc	NM_001106461	Ulk1	NM_001108341
Cnr1	NM_012784	Igf1r	NM_052807	Numbl	NM_001033888	Rnf10	NM_001011904	Vcan	NM_001170560
Dbn1	NM_031024	Inppl1	NM_022944	Olig2	NM_001100557	Rnf6	NM_001107118	Vegfa*	NM_031836
Dcc	NM_012841	Jak2	NM_031514	Omg	NM_001005898	Robo1	NM_022188	Wint7a	NM_001100473
Dclk1	NM_053343	Kalrn	NM_032062	Pafah1b1	NM_031763	Rtn4	NM_031831	Xrcc4	NM_001006999
Dguok	NM_001106602	Katnb1	NM_001024746	Pbx1	NM_001100681	Rtn4r	NM_053613	Ywhah	NM_013052

TABLE 4

TDP-43 RNA targets associated with neurodegenerative diseases

The genes listed in this table are a partial representation of the genes associated with disease. For complete list of genes, see supplemental Table S1 and NCBI GEO accession number GSE25032.

Gene Name	Symbol	RefSeq ID
Adenosine deaminase, RNA-specific	Adar	NM_031006
Amyloid β (A4) precursor protein	App	NM_019288
α-Synuclein	Snca	NM_019169
β-Synuclein	Sncb	NM_080777
Ataxin 1	Atxn1	NM_012726
Ataxin 2	Atxn2	NM_001105930
Ataxin 10	Atxn10	NM_133313
Chromatin- modifying protein 2A	Chmp2a	NM_001108906
CUG triplet repeat, RNA-binding protein 1	Cugbp1	NM_001025421
Cyclin-dependent kinase 5	Cdk5	NM_080885
Fused in sarcoma	Fus	NM_001012137
Granulin	Grn	NM_001145842
Huntingtin	Htt	NM_024357
Microtubule-associated protein Tau	Mapt	NM_017212
Neurexin 2	Nrxn2	NM_053846
Niemann-Pick disease, type C2	Npc2	NM_173118
Presenilin 1	Psen1	NM_019163
Presenilin 2	Psen2	NM_031087
Prion protein	Prnp	NM_012631
Sirtuin	Sirt2	NM_001008368
Superoxide dismutase 2	Sod2	NM_017051
TAR DNA-binding protein	Tardbp	NM_001011979
Valosin-containing protein	Vcp	NM_053864

of RNA metabolism (Fig. 5*B* and supplemental Table S9), but some are considered antioxidants (PDX1 and -2) or a calcium-binding protein (CALML3). Previously, SFRS1, which promotes exon skipping of CFTR, was shown to work with TDP-43 additively to promote CFTR exon skipping (13). The brain-specific PTBP2 binds intronic clusters of RNA regulatory elements and controls the assembly of other splicing regulatory factors, including RNA-binding proteins (41). From this list, RNA binding motif protein, X chromosome retrogene (RBMXRT) and hnRNPH2 are RNA-binding proteins involved in RNA splicing, transport, and stability (42). The predicted hnRNP A3 isoform 4 homolog (GM9242) is likely to have a similar role in RNA metabolism, but its function is unknown. MECP2 is an X-linked gene, mutations of which cause Rett syndrome, a progressive neurodevelopmental disorder (43). MECP2 is known for binding methylated DNA repressing translation, but its interaction with TDP-43 was confirmed by co-immunoprecipitation plus Western blotting (Fig. 5*C*). There is one report that demonstrates that MECP2 interacts with the RNA-binding protein YBX1 (Y box-binding protein 1) and that together they regulate splicing of reporter minigenes (44). Otherwise its role in RNA metabolism is not well characterized.

Several recent studies have reported identification of many proteins that interact with TDP-43 from peripheral cells overexpressing tagged TDP-43 protein (39, 40, 45). The native nuclear interactome of TDP-43 that we isolated from mouse brain nuclear extracts only partially overlaps with those from the other studies. Future studies will be needed to examine whether the 16 commonly co-purified proteins from our study and the other studies reflect ubiquitous TDP-43-interacting proteins. Moreover, the nine novel proteins identified in our study may represent

A.

Protein name	Gene symbol	Mass (kDa)	RefSeq ID	IP: Total peptides			IP: Spectral counts			IP: Abundance index		Total IP: TDP-43 spectral counts
				CTL	TDP-43(1)	TDP-43(2)	CTL	TDP-43(1)	TDP-43(2)	TDP-43	TDP-43/CTL (approx.)	
TAR DNA binding protein iso.1	Tardbp	45	NP_663531.1	0	6	8	0	12	34	1.02	76.67	46
Similar to heterogeneous nuclear ribonucleoprotein A3, iso.4	GM9242	37	XP_001004765.1	0	5	6	0	11	18	0.78	48.33	29
Heterogeneous nuclear ribonucleoprotein A2/B1 iso.2	Hnrnpa2/b1	32	NP_872591.1	0	6	6	0	12	13	0.78	41.67	25
Polypyrimidine tract binding protein 2	Ptbp2	58	NP_062423.1	0	5	6	0	11	13	0.41	40.00	24
RNA binding motif protein, X chromosome retrogene	Rbmxrt	42	NP_033059.2	0	3	6	0	9	11	0.48	33.33	20
Heterogeneous nuclear ribonucleoprotein K	Hnrnpk	51	NP_079555.1	0	2	5	0	5	13	0.35	30.00	18
Heterogeneous nuclear ribonucleoprotein A/B iso.2	Hnrnpab	31	NP_034578.1	0	3	4	0	6	9	0.48	25.00	15
Methyl CpG binding protein 2 isoform 2	Mecp2	52	NP_034918.1	0	1	6	0	2	12	0.27	23.33	14
Heterogeneous nuclear ribonucleoprotein D-like	Hnrpdl	46	NP_057899.2	0	3	3	0	5	8	0.28	21.67	13
Interleukin enhancer binding factor 2	Ilf2	43	NP_080650.1	0	1	4	0	1	9	0.23	16.67	10
Peroxiredoxin 1	Prdx1	22	NP_035164.1	0	2	3	0	3	6	0.41	15.00	9
Heterogeneous nuclear ribonucleoprotein C	Hnrnpc	34	NP_058580.1	0	2	2	0	4	4	0.24	13.33	8
Eukaryotic translation elongation factor 2	Eef2	95	NP_031933.1	0	3	1	0	6	2	0.08	13.33	8
Similar to heterogeneous nuclear ribonucleoprotein A0	Hnrnpa0	31	XP_001001311.1	0	1	2	0	3	4	0.23	11.67	7
Splicing factor proline/glutamine rich	Sfpq	75	NP_076092.1	0	0	4	0	0	6	0.08	10.00	6
Splicing factor, arginine/serine-rich 1 iso.1	Sfrs1	28	NP_775550.2	0	0	2	0	0	6	0.21	10.00	6
Peroxiredoxin 2	Prdx2	22	NP_035693.3	0	1	2	0	1	4	0.23	8.33	5
Heterogeneous nuclear ribonucleoprotein L	Hnrnpl	64	NP_796275.3	0	0	1	0	0	4	0.06	6.67	4
Calmodulin-like 3	Calml3	17	NP_081692.1	0	1	1	0	1	3	0.24	6.67	4
Small nuclear ribonucleoprotein D3	Snrpd3	14	NP_080371.1	0	1	1	0	1	2	0.21	5.00	3
Small nuclear ribonucleoprotein N	Snrpn	25	NP_038698.1	0	1	1	0	1	2	0.12	5.00	3
Heterogeneous nuclear ribonucleoprotein R	Hnrnpr	71	NP_083147.1	0	0	1	0	0	2	0.03	3.33	2
Eukaryotic translation initiation factor 5A	Eif5a	17	NP_853613.1	0	1	0	0	2	0	0.12	3.33	2
Splicing factor 3a, subunit 1	Sf3a1	88	NP_080451.4	0	1	0	0	2	0	0.02	3.33	2
Heterogeneous nuclear ribonucleoprotein H2	Hnrnph2	49	NP_063921.1	0	0	1	0	0	2	0.04	3.33	2

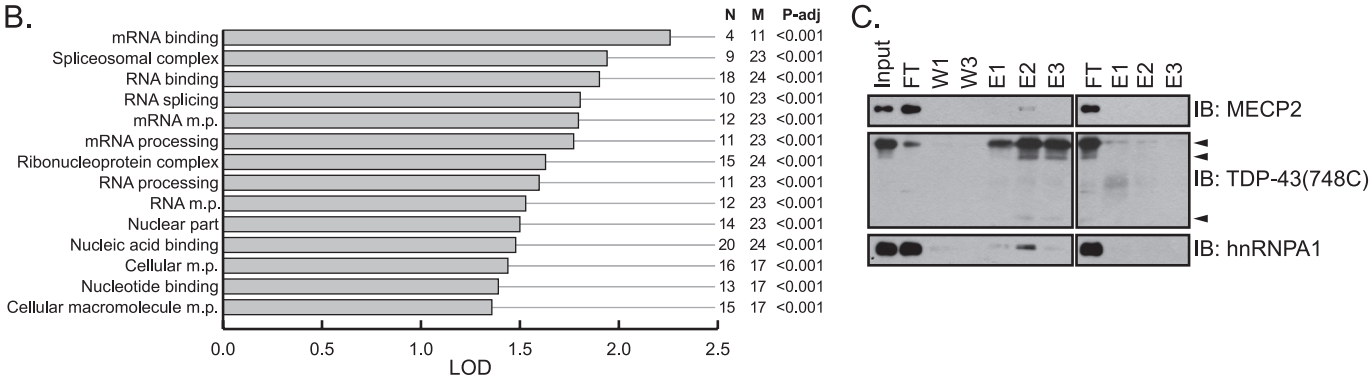

FIGURE 5. **TDP-43 nuclear interactome.** *A*, 25 proteins co-purified with TDP-43 from mouse brain in two independent immunoprecipitation experiments, *IP (1)* and *IP (2)*. *B*, functional classification of TDP-43 nuclear interactome using Gene Ontology terms (*m.p.*, metabolic process; *N*, number of proteins from the TDP-43 IP that are in the functional category; *M*, total proteins involved in that functional category, *LOD*, logarithm (base 10) of odds ratio; *P-adj*, *p*-value adjusted for multiple hypothesis testing). *C*, Western blot of TDP-43 co-immunoprecipitation products from mouse brain nuclear extracts showing co-precipitated proteins hnRNPA1 and MECP2. *FT*, flow through; *W*, wash; *E*, elution (1% of total elution was loaded). *Arrows* indicate TDP-43-specific bands. *IB*, immunoblot.

unique constituents of nuclear TDP-43-containing complexes in the brain.

Correlation between TDP-43 RNA Targets and PTBP2 Binding Motifs—Our proteomics analysis (Fig. 5, *A* and *B*) indicated that the bulk of TDP-43 is physically associated with other proteins, and many of these proteins are RNA-binding proteins with known RNA binding motifs (46, 47). Therefore, we wondered whether there was any enrichment for reads containing motifs for TDP-43-associated RNA-binding proteins. We focused on binding motifs for three TDP-43-associated proteins, PTBP2, hnRNPA2/B1, and hnRNPC, that have well defined and relatively long binding sites. Binding motifs for other associated proteins were not searched because their binding motifs either are not well defined or are too short. For example, the known binding site of hnRNPH2 is 4 nt long (GGGA). The consensus motif of hnRNPL, $(CA)_n$, on the other hand, could not be searched independently due to the nature of our TDP-43 library generation, which does not account for strandedness. Therefore, it is not possible to determine what fraction of the significant enrichment of $(TG)_n$/$(CA)_n$ in our TDP-43 library is explained by the affinity of hnRNPL to $(CA)_n$ or by the affinity of TDP-43 to $(TG)_n$.

We found a 18.9-fold enrichment for reads containing a consensus binding site motif for PTBP2, $(CT)_6$, in our TDP-43 library when compared with the control (Fisher's exact test; p value $< 2 \times 10^{-16}$; 95,300 reads in the TDP-43 library *versus* 5,100 in the control library). These results suggest that PTBP2 binding sites are in proximity of TDP-43 binding sites and that PTBP2 may co-regulate TDP-43 RNA targets. However, we did not find any significant enrichment of reads containing the binding sites of hnRNPA2/B1 or hnRNPC. One possible explanation is that these RNA-binding proteins do not directly bind TDP-43 transcripts but work as co-regulators through their direct association with the C-terminal region of TDP-43 (48). The other possibility is that hnRNPA2/B1 and hnRNPC bind to regions distal to TDP-43 binding sites. As an additional negative control, we used the DNA binding motif for PPARγ (49), which has not been suggested to associate with TDP-43 in previous analyses. As expected, there was no enrichment for reads containing the PPARγ binding motif in our TDP-43 library.

Concluding Remarks—This study reveals the nuclear components of the TDP-43-containing ribonucleoprotein complexes in the nervous system. We uncovered 25 protein con-

stituents of TDP-43 nuclear protein complexes, referred to as the TDP-43 nuclear interactome. Using RIP-seq, we identified 4,352 RNA targets of TDP-43 and revealed distinct regulatory roles of TDP-43 in post-transcriptional regulation. We also observed similar profiles of TDP-43 RNA targets using cross-linking immunoprecipitation followed by deep sequencing (data not shown). Our work on the TDP-43 nuclear interactome and RNA targets provides a framework for uncovering the biochemical principle of TDP-43-dependent regulation of pre-mRNA splicing and RNA stability and transport, for revealing the neural functions of TDP-43, and for understanding how dysregulation of TDP-43 in RNA metabolism contributes to neurodegeneration.

Acknowledgments—We thank Dr. Edward K. Wakeland and the staff at University of Texas Southwestern Medical Center Microarray Core Facility and the members of the Dr. Jane Johnson laboratory for assistance with the deep sequencing and its analysis.

REFERENCES

1. Keene, J. D. (2007) *Nat. Rev. Genet.* **8,** 533–543
2. Nelson, P. T., and Keller, J. N. (2007) *J. Neuropathol. Exp. Neurol.* **66,** 461–468
3. Lagier-Tourenne, C., Polymenidou, M., and Cleveland, D. W. (2010) *Hum. Mol. Genet.* **19,** R46–64
4. Vance, C., Rogelj, B., Hortobágyi, T., De Vos, K. J., Nishimura, A. L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P., Ganesalingam, J., Williams, K. L., Tripathi, V., Al-Saraj, S., Al-Chalabi, A., Leigh, P. N., Blair, I. P., Nicholson, G., de Belleroche, J., Gallo, J. M., Miller, C. C., and Shaw, C. E. (2009) *Science* **323,** 1208–1211
5. Sreedharan, J., Blair, I. P., Tripathi, V. B., Hu, X., Vance, C., Rogelj, B., Ackerley, S., Durnall, J. C., Williams, K. L., Buratti, E., Baralle, F., de Belleroche, J., Mitchell, J. D., Leigh, P. N., Al-Chalabi, A., Miller, C. C., Nicholson, G., and Shaw, C. E. (2008) *Science* **319,** 1668–1672
6. Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., Bruce, J., Schuck, T., Grossman, M., Clark, C. M., McCluskey, L. F., Miller, B. L., Masliah, E., Mackenzie, I. R., Feldman, H., Feiden, W., Kretzschmar, H. A., Trojanowski, J. Q., and Lee, V. M. (2006) *Science* **314,** 130–133
7. Wils, H., Kleinberger, G., Janssens, J., Pereson, S., Joris, G., Cuijt, I., Smits, V., Ceuterick-de Groote, C., Van Broeckhoven, C., and Kumar-Singh, S. (2010) *Proc. Natl. Acad. Sci. U.S.A.* **107,** 3858–3863
8. Wegorzewska, I., Bell, S., Cairns, N. J., Miller, T. M., and Baloh, R. H. (2009) *Proc. Natl. Acad. Sci. U.S.A.* **106,** 18809–18814
9. Sephton, C. F., Good, S. K., Atkin, S., Dewey, C. M., Mayer, P., 3rd, Herz, J., and Yu, G. (2010) *J. Biol. Chem.* **285,** 6826–6834
10. Wu, L. S., Cheng, W. C., Hou, S. C., Yan, Y. T., Jiang, S. T., and Shen, C. K. (2010) *Genesis* **48,** 56–62
11. Buratti, E., and Baralle, F. E. (2001) *J. Biol. Chem.* **276,** 36337–36343
12. Buratti, E., and Baralle, F. E. (2008) *Front. Biosci.* **13,** 867–878
13. Buratti, E., Dörk, T., Zuccato, E., Pagani, F., Romano, M., and Baralle, F. E. (2001) *EMBO J.* **20,** 1774–1784
14. Mercado, P. A., Ayala, Y. M., Romano, M., Buratti, E., and Baralle, F. E. (2005) *Nucleic Acids Res.* **33,** 6000–6010
15. Bose, J. K., Wang, I. F., Hung, L., Tarn, W. Y., and Shen, C. K. (2008) *J. Biol. Chem.* **283,** 28852–28859
16. Strong, M. J., Volkening, K., Hammond, R., Yang, W., Strong, W., Leystra-Lantz, C., and Shoesmith, C. (2007) *Mol. Cell. Neurosci.* **35,** 320–327
17. Casafont, I., Bengoechea, R., Tapia, O., Berciano, M. T., and Lafarga, M. (2009) *J. Struct. Biol.* **167,** 235–241
18. Buratti, E., De Conti, L., Stuani, C., Romano, M., Baralle, M., and Baralle, F. (2010) *FEBS J.* **277,** 2268–2281
19. Zhang, Y., Wen, Z., Washburn, M. P., and Florens, L. (2009) *Anal Chem.* **81,** 6317–6326
20. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005) *Mol. Cell. Proteomics* **4,** 1265–1272
21. Rhead, B., Karolchik, D., Kuhn, R. M., Hinrichs, A. S., Zweig, A. S., Fujita, P. A., Diekhans, M., Smith, K. E., Rosenbloom, K. R., Raney, B. J., Pohl, A., Pheasant, M., Meyer, L. R., Learned, K., Hsu, F., Hillman-Jackson, J., Harte, R. A., Giardine, B., Dreszer, T. R., Clawson, H., Barber, G. P., Haussler, D., and Kent, W. J. (2010) *Nucleic. Acids Res.* **38,** D613–619
22. Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2007) *Nucleic. Acids Res.* **35,** D61–65
23. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009) *Genome. Biol.* **10,** R25
24. Berriz, G. F., King, O. D., Bryant, B., Sander, C., and Roth, F. P. (2003) *Bioinformatics* **19,** 2502–2504
25. Berriz, G. F., Beaver, J. E., Cenik, C., Tasan, M., and Roth, F. P. (2009) *Bioinformatics* **25,** 3043–3044
26. Kuo, P. H., Doudeva, L. G., Wang, Y. T., Shen, C. K., and Yuan, H. S. (2009) *Nucleic. Acids Res.* **37,** 1799–1808
27. Keene, J. D., Komisarow, J. M., and Friedersdorf, M. B. (2006) *Nat. Protoc.* **1,** 302–307
28. Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A., and Burge, C. B. (2008) *Science* **320,** 1643–1647
29. Boucard, A. A., Chubykin, A. A., Comoletti, D., Taylor, P., and Südhof, T. C. (2005) *Neuron* **48,** 229–236
30. Chih, B., Gollan, L., and Scheiffele, P. (2006) *Neuron* **51,** 171–178
31. Sempere, L. F., Freemantle, S., Pitha-Rowe, I., Moss, E., Dmitrovsky, E., and Ambros, V. (2004) *Genome Biol.* **5,** R13
32. Elden, A. C., Kim, H. J., Hart, M. P., Chen-Plotkin, A. S., Johnson, B. S., Fang, X., Armakola, M., Geser, F., Greene, R., Lu, M. M., Padmanabhan, A., Clay-Falcone, D., McCluskey, L., Elman, L., Juhr, D., Gruber, P. J., Rüb, U., Auburger, G., Trojanowski, J. Q., Lee, V. M., Van Deerlin, V. M., Bonini, N. M., and Gitler, A. D. (2010) *Nature* **466,** 1069–1075
33. Lücking, C. B., and Brice, A. (2000) *Cell Mol. Life Sci.* **57,** 1894–1908
34. Aizawa, H., Sawada, J., Hideyama, T., Yamashita, T., Katayama, T., Hasebe, N., Kimura, T., Yahara, O., and Kwak, S. (2010) *Acta Neuropathol.* **120,** 75–84
35. Strong, M. J. (2010) *J. Neurol. Sci.* **288,** 1–12
36. Cenik, C., Derti, A., Mellor, J. C., Berriz, G. F., and Roth, F. P. (2010) *Genome Biol.* **11,** R29
37. Keene, J. D., and Lager, P. J. (2005) *Chromosome Res.* **13,** 327–337
38. Ou, S. H., Wu, F., Harrich, D., García-Martínez, L. F., and Gaynor, R. B. (1995) *J. Virol.* **69,** 3584–3596
39. Freibaum, B. D., Chitta, R. K., High, A. A., and Taylor, J. P. (2010) *J. Proteome Res.* **9,** 1104–1120
40. Ling, S. C., Albuquerque, C. P., Han, J. S., Lagier-Tourenne, C., Tokunaga, S., Zhou, H., and Cleveland, D. W. (2010) *Proc. Natl. Acad. Sci. U.S.A.* **107,** 13318–13323
41. Markovtsov, V., Nikolic, J. M., Goldman, J. A., Turck, C. W., Chou, M. Y., and Black, D. L. (2000) *Mol. Cell. Biol.* **20,** 7463–7479
42. Dreyfuss, G., Matunis, M. J., Piñol-Roma, S., and Burd, C. G. (1993) *Annu. Rev. Biochem.* **62,** 289–321
43. Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., and Zoghbi, H. Y. (1999) *Nat. Genet.* **23,** 185–188
44. Young, J. I., Hong, E. P., Castle, J. C., Crespo-Barreto, J., Bowman, A. B., Rose, M. F., Kang, D., Richman, R., Johnson, J. M., Berget, S., and Zoghbi, H. Y. (2005) *Proc. Natl. Acad. Sci. U.S.A.* **102,** 17551–17558
45. Kim, S. H., Shanware, N. P., Bowler, M. J., and Tibbetts, R. S. (2010) *J. Biol. Chem.* **285,** 34097–34105
46. Gabut, M., Chaudhry, S., and Blencowe, B. J. (2008) *Cell* **133,** 192 e191
47. Singh, R., Valcárcel, J., and Green, M. R. (1995) *Science* **268,** 1173–1176
48. Buratti, E., Brindisi, A., Giombi, M., Tisminetzky, S., Ayala, Y. M., and Baralle, F. E. (2005) *J. Biol. Chem.* **280,** 37572–37584
49. Desvergne, B., and Wahli, W. (1999) *Endocr. Rev.* **20,** 649–688
50. Fiesel, F. C., Voigt, A., Weber, S. S., Van den Haute, C., Waldenmaier, A., Gorner, K., Walter, M., Anderson, M. L., Kern, J. V., Rasse, T. M., Schmidt, T., Springer, W., Kirchner, R., Bonin, M., Neumann, M., Baekelandt, V., Alunni-Fabbroni, M., Schulz, J. B., and Kahle, P. J. (2010) *EMBO J.* **29,** 209–221