HHS Public Access

Author manuscript

*Biochim Biophys Acta Gene Regul Mech*. Author manuscript; available in PMC 2020 October 23.

Published in final edited form as:

*Biochim Biophys Acta Gene Regul Mech*. 2019 October ; 1862(10): 194434. doi:10.1016/j.bbagrm.2019.194434.

# TDP-43 regulates transcription at protein-coding genes and Alu retrotransposons

**Andrés A. Morera**[1,2], **Nasiha S. Ahmed**[1,2], **Jacob C. Schwartz**[2,*]

[(1)]Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ 85721, USA

[(2)]Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ 85721, USA

## Abstract

The 43-kDa transactive response DNA-binding protein (TDP-43) is an example of an RNA-binding protein that regulates RNA metabolism at multiple levels from transcription and splicing to translation. Its role in post-transcriptional RNA processing has been a primary focus of recent research, but its role in regulating transcription has been studied for only a few human genes. We characterized the effects of TDP-43 on transcription genome-wide and found that TDP-43 broadly affects transcription of protein-coding and noncoding RNA genes. Among protein-coding genes, the effects of TDP-43 were greatest for genes less than 30 thousand base pairs in length. Surprisingly, we found that the loss of TDP-43 resulted in increased evidence for transcription activity near repetitive Alu elements found within expressed genes. The highest densities of affected Alu elements were found in the shorter genes, whose transcription was most affected by TDP-43. Thus, in addition to its role in post-transcriptional RNA processing, TDP-43 plays a critical role in maintaining the transcriptional stability of protein-coding genes and transposable DNA elements.

### Keywords

TDP-43; TARDBP; transcription; Alu SINE; retrotransposon; GRO-seq

## INTRODUCTION

Transactive response (TAR) DNA-binding protein (TDP-43; transcribed from the gene *TARDBP*) is a ubiquitously expressed RNA-binding protein whose function impacts RNA transcription, processing, transport, and translation [1-3]. TDP-43 is conserved throughout the vertebrate and invertebrate animal kingdoms, from mammals to *D. melanogaster* and *C. elegans* [4, 5]. As its name implies, TDP-43 activity was originally revealed through to its

*To whom correspondence should be addressed: jcschwartz@email.arizona.edu. 1041 E. Lowell St., Tucson AZ, 85721. 940-367-8186.

**AUTHOR CONTRIBUTIONS:** A.A.M. and J.C.S. contributed to Conceptualization, Formal analysis, and Writing. A.A.M. and N.S.A performed experiments, contributing to the Investigation and Validation. J.C.S. contributed to Supervision and Funding acquisition.

effects on transcription of the *HIV-1* gene [6]. However, few subsequent studies have explored any broader role for TDP-43 in regulating transcription.

Recent studies have provided more insight into TDP-43 activity at the chromatin level and suggested new mechanistic models for the effects of TPD-43 on transcription. First among these was a study in *D. melanogaster* providing evidence that the fly homologue of TDP-43, TBPH, associated with promoter and enhancer regions of genes and regulated their transcription by two mechanisms: recruitment to GU repeats in nascent RNA transcripts and binding to cohesin proteins [4]. Additionally, the ENCODE consortium recently released data mapping TDP-43 localization on chromatin through chromatin-immunoprecipitation and sequencing (CHIP-seq) for multiple human cell lines [7]. While TDP-43 is a member of the heterogenous RNA-binding protein (hnRNP) family of proteins, TDP-43 has an ability to bind DNA *in vitro* but direct binding to DNA in cells has not yet been shown [1, 8-10]. The studies in fly and data provided by ENCODE suggest that TDP-43 belongs to the list of hnRNP proteins that associate with chromatin by binding nascent transcripts, and through protein-protein interactions with other chromatin-binding or DNA-binding proteins [11, 12].

TDP-43 activity and its biological functions are closely tied to the mechanisms underlying several neurological pathologies and diseases [13-15]. Mutations in TDP-43 that drive protein aggregation are tied with the hnRNP protein, FUS, to be the third most common known cause of amyotrophic lateral sclerosis (ALS) [13, 16]. In sharp contrast to all other genes with mutations known to cause ALS, the TDP-43 protein is found aggregated in motor neurons of nearly all ALS patients, including the subset of cases whose genetic causes are known as well as the much larger set with unknown genetic roots. While ALS is a rapidly progressing and fatal disease of motor neurons, TDP-43 proteinopathy has provided a link between ALS and the second most common human dementia, frontal temporal dementia (FTD) [17]. TDP-43 aggregation is the primary pathological indicator for 45% of FTD cases. TDP-43 aggregates have also been identified in patients with Alzheimer's disease and Huntington's disease [18-20]. Most recently, TDP-43 inclusions have been identified in the brains of humans following traumatic brain injury (TBI) and TDP-43 positive stress granules were found to follow TBI in the brains of flies, mice, and humans [20-22].

We determined the effects of TDP-43 on transcription in cultured human cells using global run-on and sequencing (GRO-seq) to directly measure transcription levels following siRNA-mediated knockdown of TDP-43. We analyzed our GRO-seq data alongside previously published ChIP-seq data and ChIP-seq data made available by the ENCODE consortium for the same cell line. However, these analyses did not yield a correlation between the substantial changes in transcription revealed by GRO-seq and associations between TDP-43 and chromatin along those genes. Instead, the magnitude of changes to transcription was linked to gene length. Most unexpectedly, TDP-43 knockdown caused a switch in primate-specific retrotransposons, Alu repeat elements, to show evidence of transcription activity. The highest densities of affected Alu elements were found within genes of shorter lengths, which were also the genes whose transcription was most reduced by TDP-43 knockdown.

## MATERIALS AND METHODS

### Cell lines

Flp-In™-293 cells were purchased from ThermoFisher (catalog # R75007). HEK293T/17 cells were obtained from ATCC (catalog # CRL-11268). Cells were cultured at 37°C and 5% $CO_2$ in Dulbecco's Modification of Eagle's Medium (DMEM) with 4.5 g/L glucose, L-glutamine, and without sodium pyruvate, supplemented with either 10% v/v and 50 μg/mL or 5% v/v fetal bovine serum for Flp-In™-293 cells and HEK293T/17 cells, respectively. Knockdown of TDP-43 was performed using RNAiMax™ (Invitrogen, catalog # 13778150) and 50 nM siRNA annealed from sense and antisense strands: siTDP sense, 5'–GGAUGAGACAGAUGCUUCAUU–3'; siTDP antisense, 5'–UGAAGCAUCUGUCUCAUCCUU–3'; SCR sense, 5'–GAUGCAGACAUUCAGGAUGUU–3'; and SCR antisense, 5'-CAUCCUGAAUGUCUGCAUCUU–3'.

### Co-immunoprecipitation

Co-immunoprecipitation (co-IP) experiments were performed using mouse anti-RNA Pol II C-terminal domain antibody (clone CTD4H8, EMD Millipore, catalog # 05-623) and protein A/G beads (ThermoFisher). HEK293T/17 cells grown to ~90% confluency were lysed in 2 mL of ice-cold Co-IP lysis buffer (25 mM Tris pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 5% glycerol) with added protease inhibitors (Complete™, Mini, EDTA-free, Roche) and 50 U benzonase (EMD Millipore). Lysates were centrifuged for 10 minutes at 13,000 ~ g; 500 μL of the supernatant, diluted to 1 mg/mL, was added to the resin and incubated overnight at 4°C with mixing. Antibodies used in Western blots were anti-TDP-43 (Proteintech, catalog # 10782-2-AP) and anti-RNA Pol II (clone CTD4H8, EMD Millipore, catalog # 05-623); secondaries used were goat anti-rabbit (ThermoFisher, catalog # 31460) and goat anti-mouse (ThermoFisher, catalog # 31432).

### Global-run on and sequencing

Flp-In™-293 cells transfected with either scramble or TDP-43 siRNA were harvested three days after transfection and nuclei of siRNA-treated cells were isolated as previously described by Allen et al. [23]. Aliquots with $5{\sim}10^6$ cells were flash frozen in liquid nitrogen and stored at - 80°C until use in nuclear run-on experiments. Two independent harvests of transfected cells of two or three dishes were validated for knockdown by westerns, run-ons performed, and libraries generated for each dish harvested. Nuclear run-on was performed as described in Allen et al. Each 100μL aliquot of nuclei was thawed on ice in Reaction Buffer (10 mM Tris pH 8.0, 5 mM MgCl2, 300 mM KCl, 1 mM DTT, and 0.5 mM each of rATP, rGTP, rCTP, and Br-UTP). Run-on reactions were stopped with 1 mL of Trizol™ (Invitrogen), mixed thoroughly by vortexing, followed by two rounds of acid phenol:chloroform extraction of the nascent RNA. RNA was precipitated by addition of 1 μL of GlycoBlue™ (ThermoFisher) and three volumes of ice-cold ethanol, followed by storage at −20°C for at least 20 minutes. RNA was fragmented with NEB fragmentation buffer (NEBNext® Fragmentation Module) at 94°C for 5 minutes. Short RNA fragments were selected using 50 μL of Agencourt AMPure XP beads (Beckman Coulter). Anti-BrU agarose beads (Santa Cruz) were used to select for BrU-incorporated RNA.

Nascent RNAs from replicates treated with SCR or siTDP were pooled, and libraries prepared using the NEBNext® Ultra™ Directional RNA Library Prep Kit according to the manufacturer's instructions. The concentration and quality of the obtained nascent RNA libraries were assessed by Qubit fluorometer and Agilent 2100 Bioanalyzer. Samples were sequenced on either an Illumina HiSeq2500 (100 bp paired-end reads) or NovaSeq 6000 (150 bp paired-end reads).

## GRO-Seq data analysis

Raw sequence data were aligned analyzed using either in-house computational resources or the CyVerse® Discovery Environment resources [24]. Alignments were made to the human genome (GRCh38 from the Genome Reference Consortium) using Bowtie2 version 2.2.1 with default parameters [25]. To calculate fold changes in gene transcription (see Figure 2), reads aligning to hg38 repeat elements (hg38 - Dec 2013 - RepeatMasker open-4.0.5 - Repeat Library 20140131, www.repeatmasker.org) were suppressed from bed files using bedtools. Quantitative analysis of repeat element signals was made by extracting repeat element alignments defined by RepeatMasker. To visually inspect masked data by genome browser, sequences were aligned to the hg38.chromFaMasked assembly file downloaded from UCSC genome browser with repetitive elements masked by N's. Integrative Genomics Viewer (IGV) version 2.3.93 was used for manual inspection of data and creating genome browser track figures [26]. Bedtools version 2.27 was also adjusted according to identify reads mapped to annotated genes [27]. FPKM values for GRO-seq data were normalized for changes in genome coverage (see Figure 3B) after TDP-43 knockdown by multiplying the siTDP read counts by a normalization factor (median density of siTDP reads mapping to REFSEQ genes over that for SCR reads). Fold change of gene transcription was calculated as adjusted FPKM for siTDP-treated samples divided by that for SCR-treated samples. Protein-coding genes called "expressed" had more than 8 GRO-seq reads mapped per kilobase in the gene body (from +500 bp of the TSS to the terminus). Hyperlinks to load GRO-seq bam files for viewing by genome browser can be copied from here: Rep1 (SCR and siTDP); Rep2 with duplicates removed (SCR and siTDP). Fastq files can be downloaded from these hyperlinks: Rep1 (SCR forward reads, SCR reverse reads, siTDP forward reads, and siTDP reverse reads); Rep2 without duplicates removed (SCR1 forward reads, SCR1 reverse reads, siTDP forward reads, and siTDP reverse reads).

## Nuclear RNA sequencing

Nuclear RNA was isolated from cell nuclei harvested by hypotonic lysis as described for GRO-seq above. Total RNA was extracted by Trizol™ and libraries generated using the NEBNext® Ultra™ Directional RNA Library Prep Kit. Nuclear RNA did not undergo poly-A selection. Libraries were sequenced for 150 bp of paired-end reads on an Illumina NovaSeq 6000. Analysis of changes in spliced transcripts was made using TopHat v2.1.1 and Cuffdiff v2.2.1 for REFSEQ annotated transcripts from hg38. Because TopHat is optimized to detect splice junctions and small exons, it was inefficient at aligning reads to repetitive elements. To quantify changes to repetitive elements or intronic transcripts, reads were aligned to hg38 using Bowtie2 v2.3.4.1 and compared by FPKM values. Hyperlinks to bam files of SCR or siTDP-treated nuclear RNA-seq data for viewing by genome browser can be copied from here: TopHat alignments (SCR1, SCR2, siTDP1, and siTDP2); Bowtie

alignments (SCR1, SCR2, siTDP1, and siTDP2). Fastq files can be downloaded at these hyperlinks: SCR1 forward reads, SCR1 reverse reads, SCR2 forward reads, SCR2 reverse reads, siTDP1 forward reads, siTDP1 reverse reads, siTDP2 forward reads, and siTDP2 reverse reads.

### Northern blotting and quantitative PCR of Alu transcripts

HEK293T/17 cells were treated with either scramble or TDP-43 siRNA in 150 mm diameter dishes. After three days, RNA from cells was harvested using Trizol™ (Invitrogen). Probes were $^{32}$P-labeled: ALU, 5'-GTCGCCCAGGCTGGAGTGCAGTGG-3'; or 5S RNA, 5'-AAAGCCTACAGCACCCGGTAT-3' [28]. RNA integrity was confirmed by staining ribosomal RNA using SYBR™ Gold nucleic acid gel stain (Invitrogen). UV crosslinking prior to probing was performed using a UVP HL-2000 HybriLinker™ UV crosslinker at 120 mJ/cm$^2$. Signals from hybridized probes were detected using a phosphor screen and PharosFX™ Plus system. Quantitative PCR primers included those targeting an Alu in the *PSEN1* gene (GRCh38, Chr14:73137801-73138083): 5'-GAGTTCGAGACCAGACAACACGGCG-3'; 5'-GATTCTCCTGCCTCAGCCTCGCTTG-3'. We also used previously published primers specific for the consensus Alu sequence [29]: 5'-CATGGTGAAACCCCGTCTCTA-3'; 5'-GCCTCAGCCTCCCGAGTAG-3'.

## RESULTS

### TDP-43 localized to gene promoters but did not bind RNA Polymerase II.

Previous reports have linked TDP-43 to changes in transcription for the genes *HIV-1* in human and *SP-10* in mouse [6, 30]. More recently, a study in *D. melanogaster* revealed associations of TDP-43 with chromatin through nascent RNA transcripts at promoter and enhancer regions. While changes in transcription itself were not measured, the enrichment of TDP-43 was found to be correlated with enrichment of RNA Pol II phosphorylated at Ser5 (Ser5P) in the repeated heptade amino acid sequence comprising the C-terminal domain (CTD) of the polymerase [4]. To determine the extent of TDP-43 associations with chromatin in human cells, we analyzed TDP-43 ChIP-seq results, which was performed in HEK293T cells and made available by the ENCODE consortium (GSE92026, from the lab of Michael Snyder) [7]. We compared these results to our previously published ChIP-seq of Ser5P RNA Pol II, also in HEK293T cells [31].

Upon inspection of the ChIP-seq data, we noted that genes with TDP-43 enriched at their promoters were also highly enriched for RNA Pol II. Examples of genes were inspected that showed strong enrichment for TDP-43 overlapping RNA Pol II in promoters, such as the hnRNP-family protein *HNRNPQ* (also called *SYNCRIP*, Figure 1A). The promoter for AAA Domain Containing Protein 3A (*ATAD3A*) was bound by TDP-43 and RNA Pol II. RNA Pol II was absent from the promoter of a gene neighboring to *ATAD3A*, transmembrane protein 240 (*TMEM240*), and TPD-43 was also not present. The Rho GDP Dissociation Inhibitor Alpha (*ARHGDIA*) is an exceptionally highly expressed gene with strong RNA Pol II signals in the gene body and terminus, while TDP-43 signals remain enriched at the promoter region. The averaged TDP-43 and RNA Pol II enrichment for all

unique genes annotated by REFSEQ (N = 28,739) revealed enrichment of TDP-43 to be generally localized to gene promoters relative to that for gene bodies or distal flanking regions (Figure 1B). Genes comprising the top 20% for enrichment of RNA Pol II at their promoters showed an even greater enrichment of TDP-43 (Figure 1B, N = 5748).

To test for a correlation between TDP-43 and RNA Pol II enrichment, we summed the reads aligning within ±500 bp of gene transcription start sites (TSS), which revealed a linear relationship reflecting a trend for TDP-43 enrichment at promoters to increase as RNA Pol II enrichment increased (Figure 1C; slope = 1.0, $R^2 = 0.6$). A simple explanation could have been that TDP-43 bound RNA Pol II directly. However, we performed co-IP experiments for RNA Pol II, which failed to show a direct interaction with TDP-43 (Figure 1D). We also noted that colocalization of the two proteins if ChIP-seq was not perfect, as seen for the gene *ARHGDIA* where the polymerase but not TDP-43 was highly enriched within the gene. Because our co-IP protocol involved treatment with nucleases, we concluded that a direct interaction was unlikely, but the possibility remained for an interaction through binding to an RNA, like that reported for *D. melanogaster* [4].

## Loss of TDP-43 reduced the transcription of thousands of genes.

While TDP-43 and RNA Pol II were colocalized to gene promoters, we questioned whether TDP-43 played a role in regulating transcription itself. To answer this, we applied the technique of global run-on followed by next generation sequencing (GRO-seq) to quantify levels of transcription genome-wide. GRO-seq sequences nascent transcripts after a modified nucleotide, bromo-uridine (BrU), is introduced to a nuclear lysate for only a few minutes [23, 32]. The result is a map of all the sites of active transcription. Importantly, GRO-seq is unaffected by post-transcriptional regulation, which can strongly affect steady-state levels of RNA in the cell. Our design allowed us to evaluate the effect on transcription following a loss of TDP-43. We designed RNA duplexes against TDP-43 mRNA and selected siRNA the most effective at knocking down the protein (siTDP) compared to our control siRNA (SCR) of a scrambled sequence (Figure 2A).

We pooled BrU-incorporated RNA recovered from three transfections of each siRNA in HEK293T cells and generated libraries for sequencing. The resulting pair-end sequences of 100 base pairs (bp) in length provided nearly 43 million reads aligned to the human genome, hg38, for SCR-treated cells and 36 million for siTDP-treated cells. We found GRO-seq signals to be enriched near the TSS of genes, which was consistent with results previously reported for mammalian cells (Figure 2B) [33]. After adjusting for differences in aligned read density between the siTDP and SCR samples, GRO-seq signals averaged across expressed and unique protein-coding REFSEQ genes (N = 7332) were generally lower following TDP-43 knockdown compared to SCR-treated cells (Figure 2B).

We chose to focus most of our attention on protein-coding genes; however, we also included genes for noncoding RNA transcripts that were annotated by GENCODE to analyze the relationship of TDP-43 enrichment on chromatin to the changes in transcription following TDP-43 knockdown. Among the expressed protein-coding genes we identified (N = 7332), transcription was reduced by 2-fold or more for one-third (N = 2502, 34%). Conversely, transcription of only 5% of protein-coding genes (N = 238) increased by 2-fold (Figure 2C).

For antisense noncoding genes (N = 313), one-third showed a decrease of >2-fold (N = 110, 35%), and much fewer increased (13%, N = 41) in transcription by >2-fold (Figure 2D). Long intergenic noncoding RNA genes (lincRNA, N = 206) also had more reductions (N = 83, 40%) than increases (N = 20, 9%) in transcription (Figure 2E).

We also evaluated the changes to those short noncoding RNA genes identified as expressed by GRO-seq. Similar changes to transcription were found for small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs), and microRNAs (miRNAs). For snoRNAs, 38% decreased more than 2-fold while 27% increased (N = 50 and 35, respectively, among 129 genes, Figure 2F). For miRNAs, 31% decreased in transcription by 2-fold and 21% increased (N = 136 and 91, respectively, among 429 genes, Figure 2G). Finally, for snRNA genes, a more even proportion of genes showed decreased (38%) and increased (33%) transcription after TDP-43 was depleted (N = 55 and 48, respectively, among 143 genes, Figure 2H). From these findings, we concluded that TDP-43 had a much broader ability to affect transcription of coding and noncoding genes than had been previously appreciated.

To conclude our survey of the effects of TDP-43 on transcription, we investigated if the enrichment of TDP-43 found in the ChIP-seq discussed previously was indicative of those genes showing changes in transcription upon TDP-43 knockdown. Taking the top 20% from each class of genes ranked by their enrichment for TDP-43, the median $LOG_2$ fold change for transcription of snRNA genes was lower than that of all snRNA genes (−0.92 and 0.75 respectively, $p = 0.014$, Student's t-test) (Figure 2H). Changes in transcription for all other non-coding genes enriched for TDP-43 was not found to be significantly different from those of all genes in their class ($p > 0.05$, Student's t-test). The difference in the median $LOG_2$ fold change between TDP-43 enriched and all expressed protein-coding genes was small but significant, likely from effects of a large N (−0.51 and −0.41 respectively, $p = 0.002$, Student's t-test, Figure 2C). In conclusion, we found no compelling evidence that the relative abundance of TDP-43 at a gene promoter could predict the degree of changes to transcription following a loss of TDP-43.

## Changes to transcription were dependent on gene length.

We sought to find changes in transcription for protein-coding genes. We used REF-seq annotated genes and removed duplicates of transcripts mapping to the same genomic coordinates produced by splicing and processing variants. We focused our analysis on expressed genes by including the genes in the top 40% according to magnitude of GRO-seq signals (7332 genes) and longer than 2000 basepairs (Figure 3A). While a similar number of reads aligned to hg38 from SCR and siTDP-treated sample, we noted a global change in GRO-seq signals toward other targets and away from annotated genes (Figure 3B). For this reason, we quantified fold changes in GRO-seq signals for protein-coding genes using an adjusted FPKM to account for the change in genome coverage after TDP-43 knockdown (see Methods).

We inspected fold changes (FC) in transcription and the structure of affected genes and found gene length to relate inversely to changes in transcription (Figure 3C). Larger changes after TDP-43 knockdown was seen in shorter genes than in long genes. We inspected the distribution of changes to transcription for protein-coding genes grouped by length: short

genes, 2 to 30 kbp; medium genes, 30 to 80 kbp; and long genes, more than 80 kbp. We observed the median change in transcription for short genes was 2-fold lower, 1.3-fold lower for medium genes, and almost unchanged for long genes (Figure 3D and E). To confirm this surprisingly broad change in transcription, we repeated the GRO-seq experiment in HEK293. The second experiment yielded fewer non-duplicate reads, or lower genome coverage, and fewer genes from which to calculate fold-changes in transcription. Nevertheless, the same relationship was found between gene length and magnitude of change to transcription (Supplemental Figures 1A and B). Fold-changes to transcription for short, medium, or long genes clustered closely for genes quantified from both replicates (Supplemental Figures 1C-E). Therefore, the association of TDP-43 with chromosomal DNA of a gene did not predict the changes in transcription, but these observations suggested that the mechanism at play could be related to the structure of the affected gene.

### Loss of TDP-43 changed transcription near repetitive elements within expressed genes.

During manual inspection of genes affected by TDP-43 loss, we found a high frequency of genes that acquired GRO-seq signals at new sites (Figure 4A). While reads mapped to repetitive elements were masked to calculate the fold change in gene transcription (see Methods), we inspected the unmasked data to determine that new signals originated from repetitive DNA elements. Long, paired-end sequences preserved the high-quality for mapping reads to these repetitive elements.

One example of a short gene with prominent GRO-seq signals at repetitive elements was the nuclear pore protein 62 (*NUP62*), whose has been previously reported to be affected by TDP-43 dysfunction in cells [34]. By this inspection, it can be see that signals are strictly confined within the boundaries of Alu repetitive elements (Figure 4A, 13 of 24 Alu repeats annotated for clarity) [34, 35]. Additionally, a specificity is revealed that increased transcriptional complex occupancy appeared for a subset of Alu elements and others remaining silent.

Alu elements are a primate-specific class of Short Interspersed Nuclear Elements (SINE) that together comprise nearly 11% of the human genome [36, 37]. An example of a long gene (*STAT3*) revealed new signal at numerous Alu elements transcribed following TDP-43 knockdown (Figure 4A). To our knowledge, *STAT3* has no known connection to TDP-43 function. Unlike protein-coding genes that are transcribed by RNA Pol II, Alu elements are known to be transcribed by RNA Pol III [38]. For this reason, we inspected another gene transcribed by RNA Pol III, the 5S rRNA gene (*RNA5S*), finding no measurable changes to their transcription and suggesting that any potential changes to RNA Pol III transcription was not universal.

We tested the level of enrichment for Alu repeats among repetitive DNA features affected by summing the number of reads aligned to repeat elements for samples treated with SCR or siTDP, after adjusting read counts in siTDP for the change in genome coverage (see Methods). In both treatments, the largest fraction of signals was attributable to Alu elements (Figure 4B). We found four times as many reads aligned to Alu elements in siTDP-treated cells compared with SCR treatment. This amount of increase was not found for Long Interspersed Nuclear Elements (LINE) or Long Terminal Repeats (LTR). Transcription of

ribosomal RNA (rRNA) genes was also not increased (Figure 4B, see also Supplemental Figure 2A). The majority of signals originate from Alu elements resided within protein-coding genes and, specifically, those we had defined as expressed (Figure 4C). Genes that we had defined as expressed overlapped approximately 14% of the human genome and produced more than half of GRO-seq reads from Alu elements (Figure 4C, E). This proportion was greater than that found for the SCR-treated control cells. Finally, we counted from REFSEQ genes that roughly half of Alu elements within protein-coding REFSEQ genes (246,090 out of 546,356) are oriented to be transcribed in the same direction as the gene containing them. However, >90% of the Alu elements occupied by the transcription machineries were oriented in the same direction as the gene containing them (23,864 out of 25,418).

## Transcription changes at Alu elements did not involve TDP-43 binding to their DNA or RNA.

Alu elements have evolved among primates into several distinguishable families and subfamilies [39]. We questioned whether a specific Alu family was more enriched with respect to other elements. The AluS family are the most abundant in the human genome, comprising 58% of all Alu elements in hg38 (Table 1). Among Alu elements affected, the AluS family was enriched above its natural abundance in the genome, making up 80% of the Alu elements affected. AluS subfamilies affected were also enriched above their natural abundance. An example is AluSx that represented 23% of all Alu elements but 32% of affected Alu elements. The portion of affected Alu elements in the AluY family was the same as that found in the human genome. The oldest Alu element arisen during primate evolution is the AluJ family and these were 3-fold less abundant among affected Alu elements compared to their abundance in the human genome (Table 1) [39].

We next tested whether the DNA or RNA from affected Alu elements were associated with TDP-43 according to ChIP-seq or eCLIP data. We inspected GRO-seq signals of Alu elements with the highest levels by TDP-43 ChIP-seq (N = 24162 or ~2% of elements, see Methods). Few of these Alu elements had GRO-seq signals or those were increased more than a 2-fold after TDP-43 loss (N = 1343 or 5.5% of Alu elements with high ChIP-seq signals, Supplemental Figure 2B). For Alu elements affected by the loss of TDP-43 (N = 51966), TDP-43 ChIP-seq signals were low, with their average number of reads (average = 11 reads, Supplemental Figure 2C) being 4-fold less than TDP-43 ChIP-seq signals found at promoters of expressed genes (average = 40 reads, see Figure 1B).

Using publicly available eCLIP data of TDP-43 in HEK293T cells [40] (SRA Accession number: SRX3041685), we found little of any signals from Alu elements. Taking any Alu element with more than one eCLIP read observed (N = 15545), few were affected by a loss of TDP-43 (Supplemental Figure 2D). For those elements affected by a loss of TDP-43, most did not have TDP-43 eCLIP reads mapped to them (Supplemental Figure 2E). In agreement with the lack of eCLIP data indicating direct binding to Alu transcripts, the specific RNA-binding motif for TDP-43, GUGUGU, was not found in the consensus sequences for Alu families or subfamilies (Supplemental Figure 3) [1]. We concluded that these findings suggest that the mechanism for TPD-43 to affect transcription near of Alu

elements along the DNA did not involve direct association of TDP-43 to the Alu DNA or RNA.

### Changes to nuclear RNA levels after a loss of TDP-43.

We next investigated whether RNA levels within the nucleus were altered by the loss of TDP-43 in the same way as transcription. We reasoned that nuclear RNA would be less susceptible to post-transcriptional mechanisms of regulating RNA stability that may compensate for the slowing of transcription, as observed by our GRO-seq experiments. We isolated nuclei from SCR or siTDP-treated HEK293T cells using hypotonic lysis, extracted RNA, and generated libraries of the total nuclear RNA recovered. These were aligned by TopHat (v2.1.1) to investigate changes in spliced transcript abundance, or by Bowtie2 (v2.3.4.1) to investigate changes in intronic RNA or Alu transcripts (see Methods).

We noted that spliced transcripts resisted changes to their steady state levels despite a global slowing in transcription. Plotting changes to expressed transcript levels as a function of gene length revealed that 4 times as many transcripts (N=559) were more than 2-fold decreased than those 2-fold increase (N=128) (Supplemental Figure 4A). Despite the reduced transcription found after the loss of TDP-43, the median fold changes were small for short, medium, and long genes (Supplemental Figure 4B).

We investigated fold changes in FPKM for nuclear RNA mapped to Alu elements in expressed genes. We found a reduction in the median level of RNA for Alu elements whose GRO-seq signals were affected (Supplemental Figure 4C). We noted this reduction extended to the adjacent intronic regions in the gene, since the large majority of Alu elements are found in introns. The reduction in intronic RNA for expressed genes was similar to changes mapped to Alu elements and consistent with reductions in GRO-seq signals (Supplemental Figure 4D). We concluded that the signals measured by RNA-seq were predominantly of intronic sequences, whose reduction in levels was consistent with GRO-seq data for expressed genes.

We sought to test whether transcription of Alu elements could be confirmed by detecting an increase in the abundance of Alu transcripts in cells. Due to the limited sequence diversity among Alu elements, protocols requiring hybridization of probes or primers for detection of Alu sequences have a limited ability to distinguish them uniquely. We chose a previously reported probe specific to Alu repeats to perform Northern analysis from three biological replicates of SCR- or siTDP-treated HEK293T cells (Figure 5A, left) [29]. We found constitutive levels of Alu transcripts to be detectable at the expected length of 300 nucleotides. The ALU probe could detect an increase in the 300 nt RNA of 1.5-fold compared to the 5S rRNA and with no evidence of a new transcript of different size that bound the ALU probe (Figure 5A, right, $p = 0.018$, Student's t-test). We also used quantitative PCR analysis with previously published primers for Alu transcripts (Alu_V) and those we designed toward a specific affected Alu element contained in the gene for the protein presenilin-1, *PSEN1* [29]. Using this approach, we found a similar approximate 1.5-fold increase in Alu transcripts, though this change did not reach statistical significance possibly due to interference of the intronic RNAs, which was observed in our RNA-seq data (Figure 5B, N=3).

### The density of affected Alu elements corresponds to changes in gene transcription.

We first hypothesized the magnitude of the averaged fold change among Alu elements within a gene (FC = sum of reads mapped to Alu elements / number of Alu elements, within the gene) might relate to those genes most affected by siTDP treatment. However, neither the range nor the median fold change in GRO-seq signals within Alu elements differed between small, medium, or long genes to the same degree as the observed changes for that of the gene itself (Figure 5C). We hypothesized that the number of Alu elements affected may relate to the change in transcription of the gene. To remove bias resulting from gene length, we determined the density of affected Alu elements per kilobase. Counting the number of Alu elements from siTDP-treated cells with GRO-seq signals meeting a minimum threshold of 80 reads aligned, short genes were found to have a markedly higher density of affected Alu elements. For the long genes whose class had no median change in transcription, the median density of affected Alu elements was nearly 5-fold lower (Figure 5D). These findings were also observed in our repeated GRO-seq experiment (Supplemental Figure 5A-D).

In addition to increased levels in the number of affected Alu elements per kilobase in short genes versus long genes, we also noted that shorter genes are higher transcribed in the control, SCR-treated HEK293T cells (Supplemental Figures 6A-B). Indeed, even FPKM values measured for the abundance of mature transcripts was higher for those expressed from short genes than for long genes (Supplemental Figure 6C). Therefore, in addition to a correlation that densities in affected Alu elements varies according to gene length, it can also be plotted that the density of affected Alu elements varies according levels of gene transcription in the affected genes (Supplemental Figure 6D).

We concluded that a loss of TDP-43 triggers significant changes in gene transcription along with an accumulation of transcription complex occupancy at repetitive Alu elements. Because we did not find a direct interaction between TDP-43 and RNA Pol II or the affected Alu elements themselves, we reason that the ability of TDP-43 to affect transcription at Alu sites in DNA is mediated by more distant interactions. TDP-43 effects on transcription may result from disruptions to chromatin stability at Alu sites, the silencing mechanisms for Alu elements, or the transcription complex itself. These findings present a novel mechanism by which TDP-43 functions to control RNA metabolism in the cell.

## DISCUSSION

We have extended the previously reported effects of TDP-43 on transcription from a few individual genes to a genome-wide effect of TDP-43 activity (Figure 2C-E). Changes in transcription were greatest for protein-coding genes, but many non-coding genes were also affected. Unlike traditional transcription factors or the transcriptional regulation by the hnRNP-family protein FUS, TDP-43 appears to act through a mechanism that does not require a direct interaction with the polymerase (Figure 1D). For most genes affected by the loss of TDP-43, their transcription was reduced; at the same time, the accumulation of transcription activity near repetitive elements was markedly increased (Figure 4A). Finally, changes to gene transcription after the loss of TDP-43 correlated inversely to gene length and the density of affected Alu elements within the gene.

The effects of TDP-43 on transcription have previously only been studied for a few human genes. Nonetheless, TDP-43 is conserved throughout metazoans and a recent study in *D. melanogaster* also found a broad TDP-43 effect on transcription and a genome-wide localization of TDP-43 to regions of chromatin enriched for RNA Pol II [4]. Direct interactions were not found in fly between TDP-43 and RNA Pol II or TDP-43 and chromosomal DNA. Instead, TDP-43 was proposed to localize by two mechanisms: binding to the nascent RNA transcript and via its ability to interact with the proteins cohesin and Nipped-B. A model involving TDP-43 interactions with nascent transcripts is consistent with data from human studies, as TDP-43 has been shown to bind pre-mRNAs and introns [1, 41]. TDP-43 regulation of transcription has also been noted for *C. elegans* and can involve interactions with chromatin-associated proteins [42].

The most unexpected finding in this study is that Alu elements became sites of accumulated transcription activity, which are usually expected to be held in a transcriptionally silent state (Figure 4A-B). Alu elements contained in human genes are primarily found in intronic regions and their distribution is fairly even across all genes [38]. However, short genes possessed a higher density of Alu elements affected by the loss of TDP-43. We found more genes affected at the level of transcription than would be expected based on previous studies of TDP-43 effects on steady-state mRNA levels [1, 41, 43, 44]. Our nuclear RNA-seq agreed with previous reports because most spliced transcripts maintained their steady state abundance, while it was the intronic RNA that was reduced similar to the change in transcription found by GRO-seq (Supplemental Figure 4A-D). We can propose two models that are consistent with these findings. First, cells can respond to disruptions in transcription by stabilizing their mRNA transcripts, particularly as part of the stress response pathway [45]. By stabilizing transcripts, steady state levels can be maintained despite a global slowing of transcription itself. Second, while we did not observe overt cell death at three days post-transfection, a TDP-43 knockout is lethal to most organisms and cell lines [46, 47]. Cell death resulting from destabilized mRNA levels for housekeeping genes, which are essential to survival, would also deplete our sequencing of the effects of mRNA instability. In this way, our sequencing approaches could selectively reveal the balanced state of cells with slowed transcription, so long as the effects to steady state mRNA levels do not deviate too far from those needed for survival.

Repetitive elements are poorly understood regarding their regulation, mechanisms of replication, and effects on cell viability [37, 48-50]. Activity of repetitive elements have been noted to modify transcription. One example is the mouse B2 RNA or human Alu RNA binding RNA Pol II and repressing transcription following heat shock [51-53]. A considerable challenge to cell-based investigations is the large number of nearly identical Alu sequences found throughout the genome, which frustrates many techniques that might shed light into these questions. However, previous studies have also drawn associations between repetitive elements and the activity of TDP-43. It has been noted that genes with mature transcripts containing SINE elements are among those whose abundance is most affected by a loss of TDP-43 [54]. TDP-43 is reported to bind proteins that are involved in A-to-I editing and required for regulation of Alu expression [49]. Additionally, recent studies have noted a striking increase in DNA damage following a loss of TDP-43 in human cells [30, 55]. By extension, DNA damage is a mechanism previously reported to de-repress

Alu transcription to a degree comparable to the changes observed by our study. Finally, TDP-43 dysfunction is most notable for its association with neurogenerative disease. In both worms and flies, increases in transposable element transcripts have been found upon disruption of TDP-43 function, and activation of these elements contributes to neurodegeneration in TDP-43 models of ALS [5, 42, 54, 56]. Increases in Alu transcription have also been noted for several neurodegenerative diseases [36, 57, 58].

Alu elements are known to be transcribed by RNA Pol III from their own promoters [39, 59, 60]. Of course, those affected by TDP-43 are also transcribed by RNA Pol II while producing the mRNA transcript [38]. Since Alu elements affected by the loss of TDP-43 were also contained in expressed genes, RNA-seq signals from these sites must include those from intronic RNA and any from the Alu element. RNA signals within introns decreased uniformly after the loss of TDP-43, despite the presence of Alu elements. This suggests that with or without TDP-43 knockdown, Alu transcripts are no more abundant than pre-mRNA or intronic RNA (Supplemental Figure 4C). The low abundance of Alu transcripts would seem insufficient to block RNA Pol II activity, even locally. Moreover, the individual Alu elements transcribed in long genes had the same fold change found for short genes but failed to produce the same effect on RNA Pol II transcription of the gene (Figure 5C).

To fairly investigate the mechanism of TDP-43 effects on Alu transcription, future studies must consider and investigate several models. The first model is that accumulated signals result from a slowing of RNA Pol II along the DNA near Alu elements. Next, there have been several reports for RNA Pol III to affect RNA Pol II that resulted from RNA transcript produced, polymerase interference with transcription factor binding to DNA, blocking of one polymerase by the occupancy of the other, and transcription-induced changes to chromatin or DNA [61-65]. The simplest model for the effects we found is that RNA Pol II and RNA Pol III are in competition to occupy the same location along the DNA. We observed transcribed Alu elements to be largely limited to active RNA Pol II genes (Figure 4C). The number of Alu elements transcribed related to the length but also the level of transcription of the gene, supporting a model that transcription plays a key role (Supplemental Figure 6D). Interference by RNA Pol II and RNA Pol III has been identified throughout eukaryotes. We find one reported example of interference of RNA Pol II transcription for the gene *Polr3e* caused by RNA Pol III transcription of an internal MIR SINE element [66]. RNA Pol II activity has been reported to “insulate” genomic DNA from binding by RNA Pol III but also to enhance chromatin access to RNA Pol III transcription in other cases [64, 67].

Lastly, Alu elements have been noted for their high susceptibility to DNA damage and that DNA damage can induce Alu transcription [38]. We and others have noted that a loss of TDP-43 leads to damage of chromosomal DNA (*T. Kawaguchi, J. Schwartz, manuscript in review*) [30, 55]. This can suggest two additional models: that TDP-43 effects on chromatin stability may either stimulate local RNA Pol III transcription, or that GRO-seq signals had resulted from an increased concentration of RNA Pol II stalled by the DNA damage. Transcription by RNA Pol II and RNA Pol III transcription can be challenging to distinguish due to a lack of robust antibodies for RNA Pol III factors, and that many factors may

associate with either polymerase, as suggested by several recent studies [61, 63, 66]. Nevertheless, this is the key question that future studies must resolve to determine where TDP-43 inserts itself into the chain of events that lead to such broad changes in transcription observed in this study.

We conclude that TDP-43 is an important and global contributor for maintaining normal RNA metabolism not only at the level of processing, transport, and translation, but also at the level of transcription. The effects of TDP-43 on transcription can occur independent of post-transcriptional regulation (Figure 3D and Supplemental Figure 4B). Most importantly, a loss of TDP-43 function destabilizes transcription at repetitive Alu element sites in the DNA. Future research should further define more general cellular effects of TDP-43 activity resulting from changes to transcription of genes and repetitive elements. Because dysfunctions of both TDP-43 and repeat elements are robust features of neurodegenerative disease, their relationship to each other is likely important to further understanding of these pathologies [37, 48, 57, 58, 68].

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## FUNDING AND ACKNOWLEDGEMENTS:

This work was funded by the National Institute of Health, NS082376, to J.C.S. Research reported in this publication was also supported by the Office of the Director, National Institutes of Health of the National Institutes of Health under award number S10OD013237. The authors thanks Dr. David R. Corey (UT Southwestern) for helpful comments and Dr. Mary A. Allen (CU Boulder) for training in the GRO-seq protocol used here.

## References

[1]. Lagier-Tourenne C, Polymenidou M, Hutt KR, Vu AQ, Baughn M, Huelga SC, Clutario KM, Ling SC, Liang TY, Mazur C, Wancewicz E, Kim AS, Watt A, Freier S, Hicks GG, Donohue JP, Shiue L, Bennett CF, Ravits J, Cleveland DW, Yeo GW, Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs, Nat Neurosci 15(11) (2012) 1488–97. [PubMed: 23023293]

[2]. Lagier-Tourenne C, Polymenidou M, Cleveland DW, TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration, Hum Mol Genet 19(R1) (2010) R46–64. [PubMed: 20400460]

[3]. Buratti E, Baralle FE, Multiple roles of TDP-43 in gene expression, splicing regulation, and human disease, Front Biosci 13 (2008) 867–78. [PubMed: 17981595]

[4]. Swain A, Misulovin Z, Pherson M, Gause M, Mihindukulasuriya K, Rickels RA, Shilatifard A, Dorsett D, Drosophila TDP-43 RNA-Binding Protein Facilitates Association of Sister Chromatid Cohesion Proteins with Genes, Enhancers and Polycomb Response Elements, PLoS Genet 12(9) (2016) e1006331. [PubMed: 27662615]

[5]. Saldi TK, Ash PE, Wilson G, Gonzales P, Garrido-Lecca A, Roberts CM, Dostal V, Gendron TF, Stein LD, Blumenthal T, Petrucelli L, Link CD, TDP-1, the Caenorhabditis elegans ortholog of TDP-43, limits the accumulation of double-stranded RNA, EMBO J 33(24) (2014) 2947–66. [PubMed: 25391662]

[6]. Ou SH, Wu F, Harrich D, Garcia-Martinez LF, Gaynor RB, Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs, J Virol 69(6) (1995) 3584–96. [PubMed: 7745706]

[7]. E.P. Consortium, An integrated encyclopedia of DNA elements in the human genome, Nature 489(7414) (2012) 57–74. [PubMed: 22955616]

[8]. Lukavsky PJ, Daujotyte D, Tollervey JR, Ule J, Stuani C, Buratti E, Baralle FE, Damberger FF, Allain FH, Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43, Nat Struct Mol Biol 20(12) (2013) 1443–9. [PubMed: 24240615]

[9]. Kuo PH, Doudeva LG, Wang YT, Shen CK, Yuan HS, Structural insights into TDP-43 in nucleic-acid binding and domain interactions, Nucleic Acids Res 37(6) (2009) 1799–808. [PubMed: 19174564]

[10]. Huelga SC, Vu AQ, Arnold JD, Liang TY, Liu PP, Yan BY, Donohue JP, Shiue L, Hoon S, Brenner S, Ares M Jr., Yeo GW, Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins, Cell Rep 1(2) (2012) 167–78. [PubMed: 22574288]

[11]. Ozdilek BA, Thompson VF, Ahmed NS, White CI, Batey RT, Schwartz JC, Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding, Nucleic Acids Res 45(13) (2017) 7984–7996. [PubMed: 28575444]

[12]. Schwartz JC, Cech TR, Parker RR, Biochemical Properties and Biological Functions of FET Proteins, Annu Rev Biochem 84 (2015) 355–79. [PubMed: 25494299]

[13]. Taylor JP, Brown RH Jr., Cleveland DW, Decoding ALS: from genes to mechanism, Nature 539(7628) (2016) 197–206. [PubMed: 27830784]

[14]. Garber K, CELL BIOLOGY. Protein 'drops' may seed brain disease, Science 350(6259) (2015) 366–7. [PubMed: 26494738]

[15]. Da Cruz S, Cleveland DW, Understanding the role of TDP-43 and FUS/TLS in ALS and beyond, Curr Opin Neurobiol 21(6) (2011) 904–19. [PubMed: 21813273]

[16]. Renton AE, Chio A, Traynor BJ, State of play in amyotrophic lateral sclerosis genetics, Nat Neurosci 17(1) (2014) 17–23. [PubMed: 24369373]

[17]. Mackenzie IR, Rademakers R, Neumann M, TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia, Lancet Neurol 9(10) (2010) 995–1007. [PubMed: 20864052]

[18]. Chang XL, Tan MS, Tan L, Yu JT, The Role of TDP-43 in Alzheimer's Disease, Mol Neurobiol 53(5) (2016) 3349–3359. [PubMed: 26081142]

[19]. Tauffenberger A, Chitramuthu BP, Bateman A, Bennett HP, Parker JA, Reduction of polyglutamine toxicity by TDP-43, FUS and progranulin in Huntington's disease models, Hum Mol Genet 22(4) (2013) 782–94. [PubMed: 23172908]

[20]. Schwab C, Arai T, Hasegawa M, Yu S, McGeer PL, Colocalization of transactivation-responsive DNA-binding protein 43 and huntingtin in inclusions of Huntington disease, J Neuropathol Exp Neurol 67(12) (2008) 1159–65. [PubMed: 19018245]

[21]. Anderson EN, Gochenaur L, Singh A, Grant R, Patel K, Watkins S, Wu JY, Pandey UB, Traumatic injury induces stress granule formation and enhances motor dysfunctions in ALS/FTD models, Hum Mol Genet 27(8) (2018) 1366–1381. [PubMed: 29432563]

[22]. Tan XL, Sun M, Brady RD, Liu S, Llanos R, Cheung S, Wright DK, Casillas-Espinosa PM, Sashindranath M, O'Brien TJ, McDonald SJ, Turner BJ, Shultz SR, Transactive Response DNA-Binding Protein 43 Abnormalities after Traumatic Brain Injury, J Neurotrauma (2018).

[23]. Allen MA, Andrysik Z, Dengler VL, Mellert HS, Guarnieri A, Freeman JA, Sullivan KD, Galbraith MD, Luo X, Kraus WL, Dowell RD, Espinosa JM, Global analysis of p53-regulated transcription identifies its direct targets and unexpected regulatory mechanisms, Elife 3 (2014) e02200. [PubMed: 24867637]

[24]. Merchant N, Lyons E, Goff S, Vaughn M, Ware D, Micklos D, Antin P, The iPlant Collaborative: Cyberinfrastructure for Enabling Data to Discovery for the Life Sciences, PLoS Biol 14(1) (2016) e1002342. [PubMed: 26752627]

[25]. Langmead B, Trapnell C, Pop M, Salzberg SL, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol 10(3) (2009) R25. [PubMed: 19261174]

[26]. Thorvaldsdottir H, Robinson JT, Mesirov JP, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration, Brief Bioinform 14(2) (2013) 178–92. [PubMed: 22517427]

[27]. Quinlan AR, Hall IM, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics 26(6) (2010) 841–2. [PubMed: 20110278]

[28]. Rudin CM, Thompson CB, Transcriptional activation of short interspersed elements by DNA-damaging agents, Genes Chromosomes Cancer 30(1) (2001) 64–71. [PubMed: 11107177]

[29]. Vossaert L, O'Leary T, Van Neste C, Heindryckx B, Vandesompele J, De Sutter P, Deforce D, Reference loci for RT-qPCR analysis of differentiating human embryonic stem cells, BMC Mol Biol 14 (2013) 21. [PubMed: 24028740]

[30]. Mitra J, Guerrero EN, Hegde PM, Liachko NF, Wang H, Vasquez V, Gao J, Pandey A, Taylor JP, Kraemer BC, Wu P, Boldogh I, Garruto RM, Mitra S, Rao KS, Hegde ML, Motor neuron disease-associated loss of nuclear TDP-43 is linked to DNA double-strand break repair defects, Proc Natl Acad Sci U S A (2019).

[31]. Schwartz JC, Ebmeier CC, Podell ER, Heimiller J, Taatjes DJ, Cech TR, FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2, Genes Dev 26(24) (2012) 2690–5. [PubMed: 23249733]

[32]. Lladser ME, Azofeifa JG, Allen MA, Dowell RD, RNA Pol II transcription model and interpretation of GRO-seq data, J Math Biol (2016).

[33]. Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M, Global quantification of mammalian gene expression control, Nature 473(7347) (2011) 337–42. [PubMed: 21593866]

[34]. Chou CC, Zhang Y, Umoh ME, Vaughan SW, Lorenzini I, Liu F, Sayegh M, Donlin-Asp PG, Chen YH, Duong DM, Seyfried NT, Powers MA, Kukar T, Hales CM, Gearing M, Cairns NJ, Boylan KB, Dickson DW, Rademakers R, Zhang YJ, Petrucelli L, Sattler R, Zarnescu DC, Glass JD, Rossoll W, TDP-43 pathology disrupts nuclear pore complexes and nucleocytoplasmic transport in ALS/FTD, Nat Neurosci 21(2) (2018) 228–239. [PubMed: 29311743]

[35]. Khosravi B, Hartmann H, May S, Mohl C, Ederle H, Michaelsen M, Schludi MH, Dormann D, Edbauer D, Cytoplasmic poly-GA aggregates impair nuclear import of TDP-43 in C9orf72 ALS/FTLD, Hum Mol Genet 26(4) (2017) 790–800. [PubMed: 28040728]

[36]. Krestel H, Meier JC, RNA Editing and Retrotransposons in Neurology, Front Mol Neurosci 11 (2018) 163. [PubMed: 29875629]

[37]. Payer LM, Steranka JP, Yang WR, Kryatova M, Medabalimi S, Ardeljan D, Liu C, Boeke JD, Avramopoulos D, Burns KH, Structural variants caused by Alu insertions are associated with risks for many human diseases, Proc Natl Acad Sci U S A 114(20) (2017) E3984–E3992. [PubMed: 28465436]

[38]. Deininger P, Alu elements: know the SINEs, Genome Biol 12(12) (2011) 236. [PubMed: 22204421]

[39]. Price AL, Eskin E, Pevzner PA, Whole-genome analysis of Alu repeat elements reveals complex evolutionary history, Genome Res 14(11) (2004) 2245–52. [PubMed: 15520288]

[40]. Van Nostrand EL, Shishkin AA, Pratt GA, Nguyen TB, Yeo GW, Variation in single-nucleotide sensitivity of eCLIP derived from reverse transcription conditions, Methods 126 (2017) 29–37. [PubMed: 28790018]

[41]. Colombrita C, Onesto E, Megiorni F, Pizzuti A, Baralle FE, Buratti E, Silani V, Ratti A, TDP-43 and FUS RNA-binding proteins bind distinct sets of cytoplasmic messenger RNAs and differently regulate their post-transcriptional fate in motoneuron-like cells, J Biol Chem 287(19) (2012) 15635–47. [PubMed: 22427648]

[42]. Saldi TK, Gonzales P, Garrido-Lecca A, Dostal V, Roberts CM, Petrucelli L, Link CD, The Caenorhabditis elegans Ortholog of TDP-43 Regulates the Chromatin Localization of the Heterochromatin Protein 1 Homolog HPL-2, Mol Cell Biol 38(15) (2018).

[43]. Jeong YH, Ling JP, Lin SZ, Donde AN, Braunstein KE, Majounie E, Traynor BJ, LaClair KD, Lloyd TE, Wong PC, Tdp-43 cryptic exons are highly variable between cell types, Mol Neurodegener 12(1) (2017) 13.

[44]. Ling JP, Pletnikova O, Troncoso JC, Wong PC, TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD, Science 349(6248) (2015) 650–5. [PubMed: 26250685]

[45]. Buchan JR, Parker R, Eukaryotic stress granules: the ins and outs of translation, Mol Cell 36(6) (2009) 932–41. [PubMed: 20064460]

[46]. Mitchell JC, Constable R, So E, Vance C, Scotter E, Glover L, Hortobagyi T, Arnold ES, Ling SC, McAlonis M, Da Cruz S, Polymenidou M, Tessarolo L, Cleveland DW, Shaw CE, Wild type human TDP-43 potentiates ALS-linked mutant TDP-43 driven progressive motor and cortical

neuron degeneration with pathological features of ALS, Acta Neuropathol Commun 3 (2015) 36. [PubMed: 26108367]

[47]. Dib S, Xiao S, Miletic D, Robertson J, Gene targeting of mouse Tardbp negatively affects Masp2 expression, PLoS One 9(4) (2014) e95373. [PubMed: 24740308]

[48]. Savage AL, Schumann GG, Breen G, Bubb VJ, Al-Chalabi A, Quinn JP, Retrotransposons in the development and progression of amyotrophic lateral sclerosis, J Neurol Neurosurg Psychiatry 90(3) (2019) 284–293. [PubMed: 30305322]

[49]. Quinones-Valdez G, Tran SS, Jun HI, Bahn JH, Yang EW, Zhan L, Brummer A, Wei X, Van Nostrand EL, Pratt GA, Yeo GW, Graveley BR, Xiao X, Regulation of RNA editing by RNA-binding proteins in human cells, Commun Biol 2 (2019) 19. [PubMed: 30652130]

[50]. Castro-Diaz N, Friedli M, Trono D, Drawing a fine line on endogenous retroelement activity, Mob Genet Elements 5(1) (2015) 1–6.

[51]. Ponicsan SL, Houel S, Old WM, Ahn NG, Goodrich JA, Kugel JF, The non-coding B2 RNA binds to the DNA cleft and active-site region of RNA polymerase II, J Mol Biol 425(19) (2013) 3625–38. [PubMed: 23416138]

[52]. Kassube SA, Fang J, Grob P, Yakovchuk P, Goodrich JA, Nogales E, Structural insights into transcriptional repression by noncoding RNAs that bind to human Pol II, J Mol Biol 425(19) (2013) 3639–48. [PubMed: 22954660]

[53]. Yakovchuk P, Goodrich JA, Kugel JF, B2 RNA and Alu RNA repress transcription by disrupting contacts between RNA polymerase II and promoter DNA within assembled complexes, Proc Natl Acad Sci U S A 106(14) (2009) 5569–74. [PubMed: 19307572]

[54]. Li W, Jin Y, Prazak L, Hammell M, Dubnau J, Transposable elements in TDP-43-mediated neurodegenerative disorders, PLoS One 7(9) (2012) e44099. [PubMed: 22957047]

[55]. Hill SJ, Mordes DA, Cameron LA, Neuberg DS, Landini S, Eggan K, Livingston DM, Two familial ALS proteins function in prevention/repair of transcription-associated DNA damage, Proc Natl Acad Sci U S A 113(48) (2016) E7701–E7709. [PubMed: 27849576]

[56]. Krug L, Chatterjee N, Borges-Monroy R, Hearn S, Liao WW, Morrill K, Prazak L, Rozhkov N, Theodorou D, Hammell M, Dubnau J, Retrotransposon activation contributes to neurodegeneration in a Drosophila TDP-43 model of ALS, PLoS Genet 13(3) (2017) e1006635. [PubMed: 28301478]

[57]. Guo C, Jeong HH, Hsieh YC, Klein HU, Bennett DA, De Jager PL, Liu Z, Shulman JM, Tau Activates Transposable Elements in Alzheimer's Disease, Cell Rep 23(10) (2018) 2874–2880. [PubMed: 29874575]

[58]. Larsen PA, Lutz MW, Hunnicutt KE, Mihovilovic M, Saunders AM, Yoder AD, Roses AD, The Alu neurodegeneration hypothesis: A primate-specific mechanism for neuronal transcription noise, mitochondrial dysfunction, and manifestation of neurodegenerative disease, Alzheimers Dement 13(7) (2017) 828–838. [PubMed: 28242298]

[59]. Wang C, Huang S, Nuclear function of Alus, Nucleus 5(2) (2014) 131–7. [PubMed: 24637839]

[60]. Burns KH, Boeke JD, Human transposon tectonics, Cell 149(4) (2012) 740–52. [PubMed: 22579280]

[61]. Policarpi C, Crepaldi L, Brookes E, Nitarska J, French SM, Coatti A, Riccio A, Enhancer SINEs Link Pol III to Pol II Transcription in Neurons, Cell Rep 21(10) (2017) 2879–2894. [PubMed: 29212033]

[62]. Wang Q, Nowak CM, Korde A, Oh DH, Dassanayake M, Donze D, Compromised RNA polymerase III complex assembly leads to local alterations of intergenic RNA polymerase II transcription in Saccharomyces cerevisiae, BMC Biol 12 (2014) 89. [PubMed: 25348158]

[63]. Duttke SH, RNA polymerase III accurately initiates transcription from RNA polymerase II promoters in vitro, J Biol Chem 289(29) (2014) 20396–404. [PubMed: 24917680]

[64]. Alla RK, Cairns BR, RNA polymerase III transcriptomes in human embryonic stem cells and induced pluripotent stem cells, and relationships with pluripotency transcription factors, PLoS One 9(1) (2014) e85648. [PubMed: 24465633]

[65]. Lukoszek R, Mueller-Roeber B, Ignatova Z, Interplay between polymerase II- and polymerase III-assisted expression of overlapping genes, FEBS Lett 587(22) (2013) 3692–5. [PubMed: 24113658]

[66]. Yeganeh M, Praz V, Cousin P, Hernandez N, Transcriptional interference by RNA polymerase III affects expression of the Polr3e gene, Genes Dev 31(4) (2017) 413–421. [PubMed: 28289142]

[67]. Listerman I, Bledau AS, Grishina I, Neugebauer KM, Extragenic accumulation of RNA polymerase II enhances transcription by RNA polymerase III, PLoS Genet 3(11) (2007) e212. [PubMed: 18039033]

[68]. Prudencio M, Gonzales PK, Cook CN, Gendron TF, Daughrity LM, Song Y, Ebbert MTW, van Blitterswijk M, Zhang YJ, Jansen-West K, Baker MC, DeTure M, Rademakers R, Boylan KB, Dickson DW, Petrucelli L, Link CD, Repetitive element transcripts are elevated in the brain of C9orf72 ALS/FTLD patients, Hum Mol Genet 26(17) (2017) 3421–3431. [PubMed: 28637276]

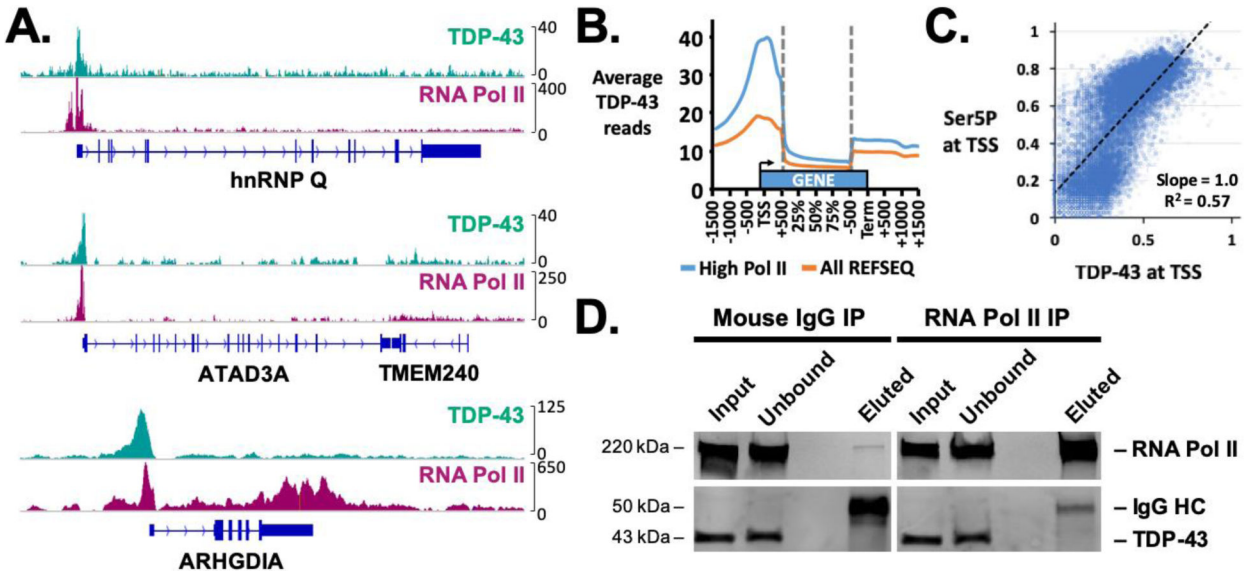


**Figure 1: TDP-43 associates with chromatin for actively expressed genes but does not bind RNA Pol II directly.**

(**A**) ChIP-seq data from HEK293T cells for TDP-43 (GSE92026) or RNA Pol II with Ser5 phosphorylated in its CTD [31] were aligned to the genome to reveal colocalization in gene promoters, as shown here for the genes *HNRNPQ, ATAD3A*, and *ARHGDIA*. (**B**) The average number of reads are from the TDP-43 ChIP-seq for all REFSEQ genes and adjacent regions for those annotated to unique, non-overlapping regions of the genome (orange, N = 28,739) and those ranking in the top 20% according to the level of RNA Pol II bound to their promoters (blue, N = 5748). (**C**) Levels of RNA Pol II or TDP-43 were plotted from the sum of reads within ±500 base pairs surrounding the TSS of each gene and normalized as relative levels from 0 to 1. (**D**) RNA Pol II was immunoprecipitated (IP) from HEK293T cells, and the eluted proteins were probed for TDP-43 by Western analysis. The negative control is an IP using a non-specific mouse IgG. The large excess of cell lysate used to improve the detection of even low TDP-43 levels is demonstrated by the levels of unbound RNA Pol II remaining in the supernatant after IP of the polymerase.

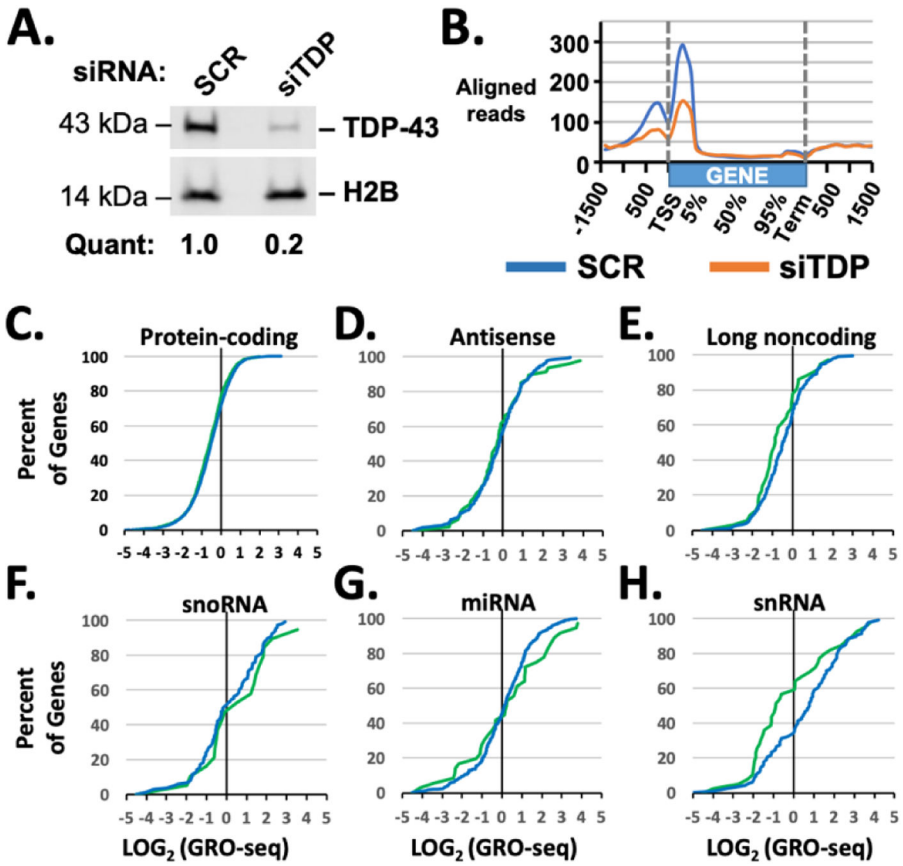


**Figure 2: Loss of TDP-43 broadly affects gene transcription.**

(**A**) TDP-43 protein was depleted from HEK293T cells using an siRNA targeting the TDP-43 mRNA (siTDP) compared to a control siRNA (SCR). Shown here is a representative Western of 3 siTDP and SCR transfections used in GRO-seq experiments. (**B**) GRO-seq signals for expressed genes in HEK293T cells and adjacent regions were normalized for genome coverage and the averages plotted (N=7280). Averaged transcription of siTDP-treated cells (orange) was lower than SCR-treated cells (blue). Cumulative percentile plots show the percentage of expressed genes (blue) versus the $LOG_2$ for fold-changes in GRO-seq signals. In green is the cumulative plots for genes in the top 20% according to levels of TDP-43 provided by ChIP-seq. Included are protein-coding genes (**C**), antisense transcripts (**D**), long noncoding RNAs (**E**), small nucleolar RNAs (snoRNA, **F**), microRNAs (miRNA, **G**), and small nuclear RNAs (snRNA, **H**). Negative $LOG_2$ values represent reductions in transcription and more than one $LOG_2$ changes (>1 or < −1) are those more than 2-fold.

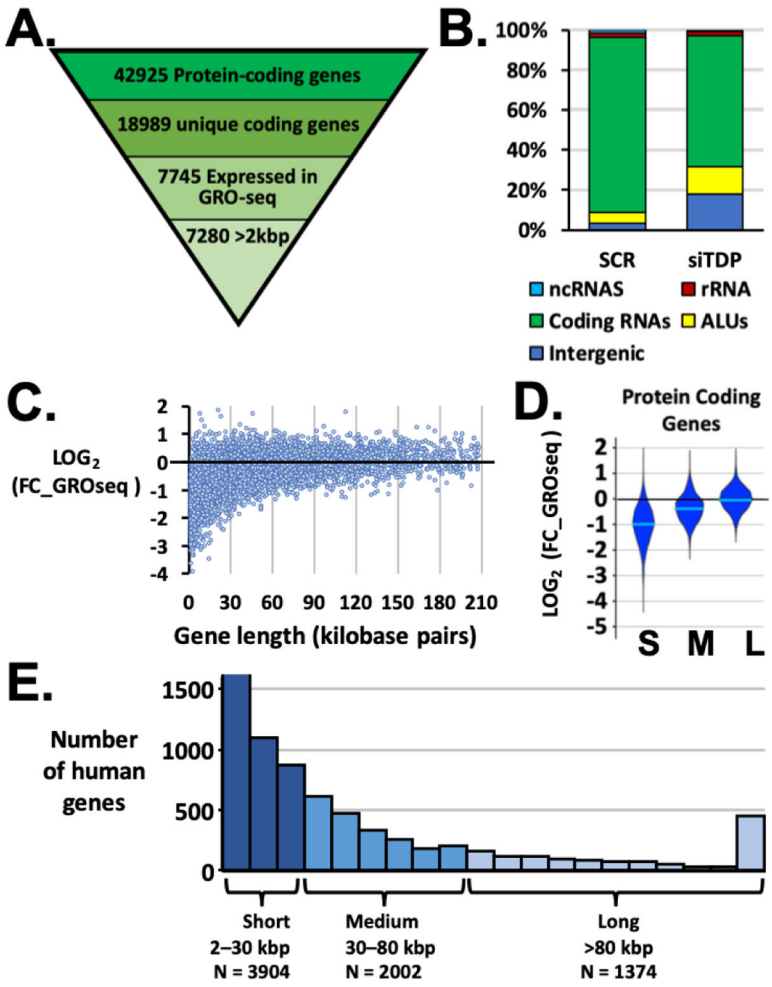


**Figure 3: Changes to transcription due to a loss of TDP-43 vary according to gene length.**

(**A**) Protein-coding genes were screened to ensure these mapped uniquely to the genome in order to avoid overrepresentation of genes with numerous annotated splice variants. Those with the highest GRO-seq signals were selected for further analysis. (**B**) Approximately equal numbers of SCR and siTDP GRO-seq reads aligned with hg38 but a smaller proportion of reads from siTDP-treated cells aligned to genic regions, including protein-coding genes (green). Subsequent analysis of changes to transcription for protein-coding genes were normalized to account for this change in the proportion of reads sequenced. (**C**) By plotting the $LOG_2$ for fold-change (FC) in transcription according to gene length, shorter genes were noted to be most affected by the loss of TDP-43. (**D–E**) Grouped according to lengths of short, medium, and long, short genes were most affected with a median 2-fold reduction in transcription, and few long genes underwent a change more than 2-fold. Part D shows violin plots, whose widths reflect the number of genes changed and horizontal lines indicate the median $LOG_2$ fold-change.

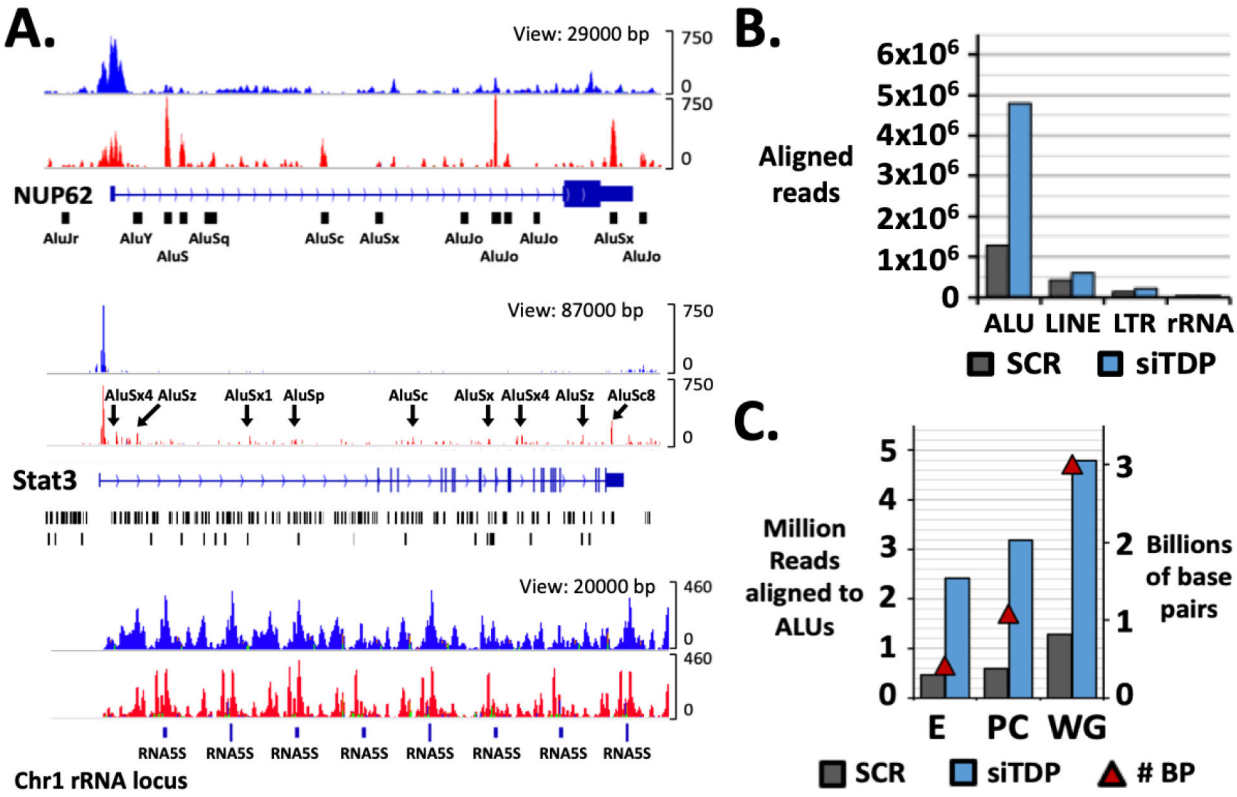


**Figure 4: Transcription at repetitive DNA elements is affected by the loss of TDP-43.**

(**A**) Examples of GRO-seq data are shown for SCR- (blue) and siTDP- (red) treated samples for a short gene, *NUP62*, and one long gene, *STAT3*. For clarity, 13 of the 24 Alu elements are shown for *NUP62*, while all are shown for *STAT3*. At the same time 5S rRNA gene transcription remained nearly constant. (**B**) The number of GRO-seq reads are shown that aligned to Alu, LINE, LTR, and rRNA repetitive elements from SCR- or siTDP-treated samples (see also Supplemental Figure 2A). (**C**) The number of reads aligning to Alu elements are shown within expressed genes (E), all protein-coding genes (PC), and the whole genome (WG). The number of base pairs, bps, from the human genome spanned by each class (red triangles) corresponds to the right axis: expressed genes span 0.4 billion bps, all protein-coding genes span 1.0 billion bps, and the whole genome is 3.0 billion bps.

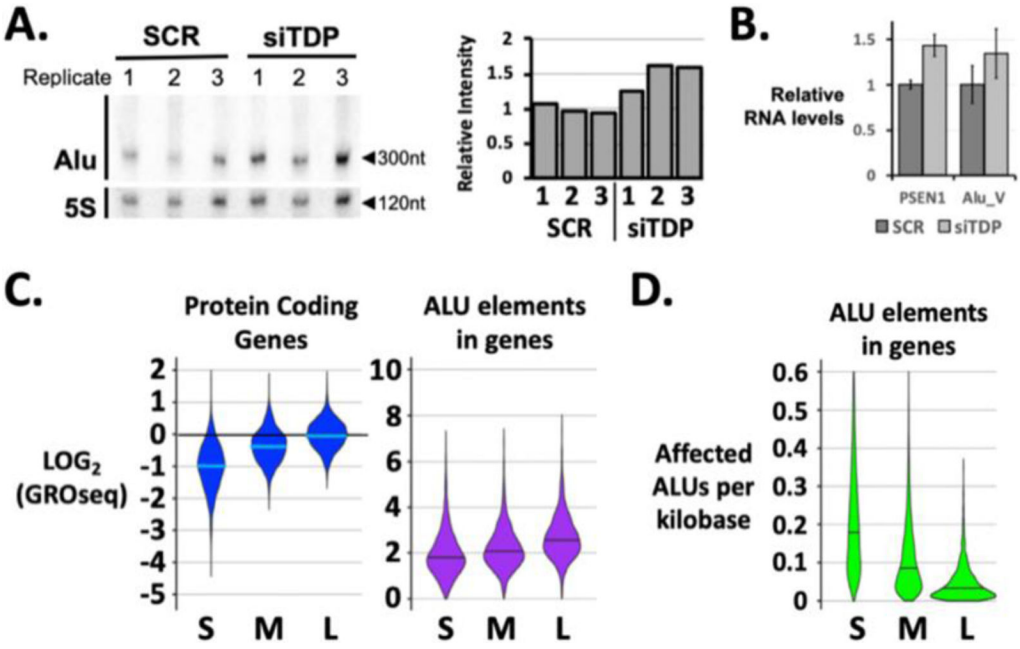


**Figure 5: Affected Alu elements are concentrated within genes with repressed transcription.**

(**A**) Northern analysis from three replicates of SCR- and siTDP-treated HEK293T cells for Alu transcripts and normalized to 5S rRNA (left) were quantified to reveal up to a 1.5-fold increase in cellular levels (right, $p = 0.018$, Student's t-test). (**B**) Relative quantitative PCR using primers designed against a specific Alu element in the *PSEN1* gene (PSEN1) or Alu transcripts in general (Alu_V) revealed the same 1.5-fold increase in cellular levels observed in the Northern analysis ($p > 0.05$, Student's t-test). (**C**) Changes in GRO-seq for small, medium, and long genes did not translate to similar differences in the magnitude of increased signal for affected Alu elements with the gene. Horizontal bars indicate the median $LOG_2$ fold-change. The median increase in Alu signals is approximately 4-fold within the short, medium, or long genes. (**D**) The number of affected Alu elements per kilobase in siTDP-treated cells was greatest within short genes (see also Supplemental Figure 5). Horizontal lines indicate the median value for the number of affected Alu elements per kilobase, which was higher for short genes than medium or long genes.

**Table 1:**

Alu families affected by TDP-43 knockdown.

| Alu Family | Number affected | % of total | Number in hg38 | % of total |
|---|---|---|---|---|
| **AluS** | **41603** | **80.0** | **713351** | **57.6** |
| AluSx | 16916 | 32.5 | 292252 | 23.6 |
| AluSq | 5409 | 10.4 | 87182 | 7.0 |
| AluSc | 3502 | 6.7 | 66384 | 5.4 |
| AluSg | 3862 | 7.4 | 55443 | 4.5 |
| **AluY** | **5797** | **11.2** | **146308** | **11.8** |
| **AluJ** | **4390** | **8.4** | **322603** | **26.0** |
| AluJb | 2740 | 5.3 | 131759 | 10.6 |
| AluJr | 813 | 1.6 | 109469 | 8.8 |
| AluJo | 837 | 1.6 | 81375 | 6.6 |
| **Other** | **60** | **0.1** | **12333** | **1.0** |
| **All ALUs** | **51976** | | **1238897** | |